Identifizierung, Sequenzierung und Charakterisierung des *Dmxl*1-Gen in *Mus musculus* sowie die funktionelle Analyse durch Knock-Out

Dissertation zur Erlangung des Grades
„Doktor der Naturwissenschaften“

am Fachbereich Biologie
der Johannes Gutenberg-Universität
in Mainz

Sven-Ernö Bikár
geb. in Wiesbaden

Oktober 2009

Dekan:
1. Berichterstatter:
2. Berichterstatter:

Tag der mündlichen Prüfung: 06.10.2010

Inhaltsverzeichnis

Abbildungsverzeichnis v

Tabellenverzeichnis vii

Abkürzungen viii

1- und 3- Buchstabencode der Aminosäuren x

1 Einleitung 1
1.1 Vergleichende Genomanalyse und Bioinformatik 2
1.2 Die Maus als Modellsystem – Funktionsanalyse und Genomik 6
1.2.1 Die klassische Mausgenetik 6
1.2.2 Entwicklungen in der modernen Mausgenetik 8
1.2.3 Funktionsanalysen in der Maus 12
1.3 Das DmX-Gen und seine Homologen 14
1.3.1 *CpY*, *DmX* und *DMXL*1 14
1.3.2 DmX in anderen Spezies 17
1.3.3 DmX gehört zu den WD-Repeat-Proteinen 18
1.4 Zielsetzung 20

2 Material und Methoden 22
2.1 Versuchstiere und Zellen 22
2.1.1 *Mus musculus* 22
2.1.2 Embryonale Stammzellen Typ 129-R1 22
2.2 Standardmethoden 23
2.2.1 Restriktion von DNA 23
2.2.2 Fällung von DNA 23
2.2.3 Elektrophorese von DNA 23
2.2.4 Inhibition von *Pseudomonas*-Kontaminationen 24
2.3 DNA-Präparation 24
2.3.1 Präparation von Plasmid-DNA 24
2.3.2 Präparation von Cosmid-, PAC- und P1-DNA 25
2.3.3 DNA-Gewinnung aus Geweben 25
2.3.4 DNA-Isolierung aus Zellkulturen 26
2.3.5 DNA-Gewinnung aus Zellkulturen im 96'er Maßstab 26
2.4 Klonierung in Plasmidvektoren 27
2.4.1 Klonierung in pUC 18 27
2.4.2 Klonierung in pGEM®-T Easy 27
2.4.3 Klonierung in den Oligo-Targeting-Vektor 29

2.4.4 Herstellung einer „shotgun“ Klonbibliothek 30
2.5 Hybridisierungstechniken 31
2.5.1 Southern-Hybridisierung 31
2.5.2 Koloniefilter-Hybridisierung 32
2.5.3 Hybridisierung an hoch-dichte Klonfiltern 33
2.5.4 Markierung von DNA durch „random primed oligo-labelling“ 33
2.6 Southern-Hybridisierung an DNA aus embryonalen Stammzelllinien 34
2.6.1 Restriktion genomischer DNA aus embryonalen Stammzellen 34
2.6.2 Elektrophorese und Southern-Transfer 34
2.6.3 Hybridisierung 34
2.7 DNA-Sequenzierung 35
2.8 Polymerasekettenreaktion (PCR) 36
2.8.1 PCR-Technik 36
2.8.2 Standard-PCR 36
2.8.3 „X-large” und „High Fidelity“ PCR 37
2.8.4 Reverse Transkriptase-PCR (RT-PCR) 38
2.9 RNA-Präparationen 39
2.9.1 RNA-Isolierung mit Guanidinthiocyanat 39
2.9.2 RNA Isolierung mit Trizol Reagent 39
2.9.3 Isolierung von polyA$^+$ -RNA 40
2.10 Transformation von Bakterien 40
2.11 Knock out Technik bei Mäusen 40
2.11.1 Routinekultur von ES-Zellen 41
2.11.2 Elektroporation und Selektion von ES-Zellen 41
2.11.3 Isolierung von ES-Zellklonen und Kryokonservierung 41
2.11.4 Isolierung von Blastozysten 42
2.11.5 Injektion von ES-Zellen in Blastozysten 42
2.11.6 Embryo-Transfer in pseudoträchtige Mäuse 43
2.11.7 Identifizierung von Chimären 44
2.12 In Silico-Analysen von Nukleotid- und Aminosäuresequenzen 45
2.13 Nomenklatur von DmX 48
2.14 Puffer und Lösungen 49
2.15 *E. coli*-Stämme 52
2.16 Bezugsquellen 52

3 Ergebnisse 53

3.1 Funktionelle Analyse durch „Knock-Out“ 53
3.1.1 Identifikation von *Dmxl*1 in *Mus musculus* 53
3.1.1.1 Cosmid-Klone aus der Maus 53
3.1.1.2 PAC-Klone aus der Maus 54
3.1.1.3 Lagen der PAC- und Cosmid-Klone 56
3.1.2 Sequenzierung der Maus-Klone Cos-N10467Q3, PAC-F13324Q2 und PAC-I02348Q2 57
3.1.2.1 Sequenzierungsstrategie für Cos-N10467Q3 57

3.1.2.2 Selektive Sequenzierung von PAC-F13324Q2 und PAC-I02348Q2 57
3.1.3 Die cDNA des murinen *Dmxl*1 61
3.1.4 Konstruktion des Targeting-Vektors 62
3.1.4.1 Auswahl der Targeting-Sequenzen 62
3.1.4.2 Klonierung der Targeting-Sequenzen 64
3.1.5 *Dmxl*1: Gen-Targeting 68
3.1.5.1 Transformationen des Targeting-Konstruktes 68
3.1.5.2 Detektion der endogenen Bande durch Restriktion und Southern-Analysen 68
3.1.5.3 Blastozysten Injektionen 73
3.1.5.4 *Dmxl*1-„Knock Out“: chimäre Mäuse 74
3.2 *In silico*-Analysen der Sequenzdaten 84
3.2.1 Exonvorhersage für Cos-N10467Q3 durch Kombination von DOTPLOT Analysen (Mensch cDNA vs genomische Maussequenz), Genvorhersageprogrammen und Datenbanksuchen 84
3.2.2 Repetitive Elemente in Cosmid N10467Q3 87
3.2.3 Genomische Lokalisation von Cosmid N10467Q3 88
3.2.4 Analyse der Sequenzdaten von PAC-F13324Q2 und PAC-I02348Q293 89
3.2.5 Genomische Lokalisation und Genstruktur des murinen *Dmxl*1 und humanen *DMXL*1 89
3.2.6 GC-Gehalt und „CpG-Inseln“ 102
3.2.7 Identifikation des Promotors von *Dmxl*1 durch Kombination von Promotorvorhersageprogrammen Und Maus/Mensch-Sequenzvergleich 105
3.2.8 Identifikation möglicher regulatorischer Sequenzen im ersten Intron von *Dmxl*1 und *DMXL*1 113
3.2.9 Identifizierung von Polyadenylierungssignalen 115
3.2.10 Vergleich der cDNA und Aminosäuresequenzen von DmX zwischen Maus, Mensch und anderen Spezies 118
3.2.11 Analyse der abgeleiteten Aminosäuresequenz von *Dmxl*1 126

4 Diskussion 132

4.1 Bewertung der Strategie zur Identifikation von *Dmxl*1 in *Mus musculus* und Konstruktion des Targeting-Vektors 132
4.2 Identifikation von rekombinanten embryonalen Stammzellen – ein spezifisch entwickeltes Restriktions- / Detektionssystem 135
4.3 *Dmxl*1 „Knock-Out“ 136
4.3.1 Gen-Targeting 136
4.3.2 Hochchimäre Tiere zeigen einen neuen Phänotyp 137

4.3.3 Vergleich mit „Knock-Out“-Experimenten in anderen Spezies 144
4.4 In silico-Analysen des *Dmxl*1-Gens 146
4.4.1 GC-Gehalt, Isochoren, GC3-Gehalt und „CpG-Inseln“ der genomischenRegion von *Dmxl*1 und *DMXL*1 146
4.4.2 Repetitive Elemente der genomischen Region von Dmxl1 150
4.4.3 Genvorhersage durch Kombinationsstrategien 155
4.4.4 Die Genstruktur von *Dmxl*1 und *DMXL*1 156
4.4.5 Promotorstrukturen und Enhancer von *Dmxl*1 162
4.4.6 Dmxl1 besitzt möglicherweise zwei alternative Polyadenylierungssignale 178
4.4.7 Strukturelle Motive in der abgeleiteten Aminosäuresequenz von Dmxl1 182
4.4.8 Rekonstruktion des C-terminalen WD-Propeller und Diskussion möglicher Faltungsstrukturen von Dmxl1 188
4.4.9 Ist Dmxl1 ein Multifunktionsprotein? 194

5 Zusammenfassung 197

Literaturverzeichnis

Anhang

Danksagung

Lebenslauf

Veröffentlichungen und Abstracts

Verzeichnis der Lehrtätigkeit

Abbildungsverzeichnis

Abb. 1.1: Syntäne Regionen in Maus, Mensch und Ratte 3
Abb. 1.2: Bioinformatik in der biologischen und medizinischen Forschung 6
Abb. 1.3: Transgene Riesen-Maus 8
Abb. 1.4: Systematischer Ablauf eines Gen-Targeting Experiments 10
Abb. 1.5: Gen-Targeting im hoch-Durchsatz Verfahren durch VelociGene 13
Abb. 1.6: Struktur eines repräsentativen Propellerblatts aus der Propellerstruktur eines typischen WD-Repeat-Proteins 20
Abb. 2.1: Vektorkarte pUC 18 28
Abb. 2.2: Vektorkarte pGEM T Easy 28
Abb. 2.3: Targeting-Vektor 29
Abb. 2.4: Funktion und Detektion einer „Imaging Plate“ 31
Abb. 2.5: Vergleich der Empfindlichkeit von „Imaging Plates“ und Röntgenfilmen (Fuji HR-S) 32
Abb. 2.6: Beispiel einer Blastozysteninjektion 42
Abb. 2.7: Implantation manipulierter Blastozysten 44
Abb. 3.1: Restriktions- und Southern-Analyse der Cosmid-Klone 54
Abb. 3.2: Southern-Analyse der PAC-Klone PAC-O18380Q2, PAC-F19112Q2, PAC-F13324Q2, PAC-I02348Q2 und PAC-H15193Q2 55
Abb. 3.3: Position der PAC- und Cosmid-Klone 56
Abb. 3.4: Sequenzierungsstrategie für Cos-N10467Q3 59
Abb. 3.5: Selektive Sequenzierung von PAC-F13324Q2 und PAC-I02348Q2 60
Abb. 3.6: Genomische Organisation der ausgewählten Targeting-Sequenzen 64
Abb. 3.7a: Restriktionsanalyse nach Klonierung von Integrat K7 65
Abb. 3.7b: Klonierungsstrategie des zweiten Targeting-Arms 66
Abb. 3.8: Ergebnisse der Sequenzierungen nach der Klonierung in den Targeting-Vektor, dargestellt als Sequencher-Projekt 67
Abb. 3.9: Endogene Bande 69
Abb. 3.10: Schematische Darstellung der homologen Rekombination mit dem Targeting-Konstrukt 70
Abb. 3.11: Detektion der positiv transformierten Stammzelllinie 4.P14-D4 71
Abb. 3.12: Southernblot der Zelllinie 4.P14-D4 73
Abb. 3.13: Chimäre *Dmx l*1 „Knock-Out“-Mäuse 76
Abb. 3.14: Optische Unterschiede in Größe und Gewicht bei differierenden Chimärismusgraden 82
Abb. 3.15: Vergleich der Testis zwischen einer niederchimären (25 %, Maus 1+5) und hochchimären (75 %, Maus 2+3) Maus 83
Abb. 3.16: DOTPLOT Analyse und Exon-Identifikation 85
Abb. 3.17: Computergestützte und webbasierte Exonvorhersage 86
Abb. 3.18: Cosmid N10467Q3 ist ein *in vitro*-Ligat von DNA-Abschnitten aus Chromosom 18 und Chromosom 4 89
Abb. 3.19: Genomische Lokalisation von *Dmxl*1 und *DMXL*1 91
Abb. 3.20: PIP-Vergleich der murinen genomischen *Dmxl*1-Sequenz mit der humanen 96
Abb. 3.21: PIP-Vergleich der humanen genomischen *DMXL*1-Sequenz mit der murinen 99
Abb. 3.22: Vergleich der murinen genomischen *Dmxl*1-Sequenz (vertikal) mit der humanen Sequenz(horizontal) mit Hilfe der DOTPLOT-Option des Programms PipMaker 102

Abb. 3.23: „GC-Plot“, *Hpa* II Restriktionskarte und Lage von *Dmxl*1 bzw. *DMXL*1 104
Abb. 3.24: DOTPATH-Analyse 105
Abb. 3.25: Konservierte colineare Sequenzblöcke der 5’-Stromaufwärts gelegenen genomischen Region von *Dmxl*1 und *DMXL*1 106
Abb. 3.26: Ergebnisse der Promotoranalyse mit dem Programm ElDorado 109
Abb. 3.27: Positionierung der Transkriptionsfaktoren 112
Abb. 3.28: zPicture-Analyse des ersten Introns von Maus und Mensch 114
Abb. 3.29: Alternative Polyadenylierungssignale von *Dmxl*1 116
Abb. 3.30: Vergleich der umgebenden Sequenz von polyA1 zwischen Maus und Mensch 117
Abb. 3.31: EST-Datenbanksuche mit der cDNA von *Dmxl*1 und *DMXL*1 118
Abb. 3.32: Globale Übereinstimmung der Aminosäuresequenz zwischen Maus und Mensch 119
Abb. 3.33: Vorhersage möglicher Transmembranhelices 130
Abb. 4.1: Aberrantes Spleißen führt zu einer Verschiebung und Unterbrechung des offenen Leserahmens 135
Abb. 4.2: Chromosomale Lage von Genen und QTL’s, die Körpergewicht und Fettleibigkeit bei Mäusen beeinflussen 141
Abb. 4.3: Vergleich verschiedener „Knock-Out“-Modelle, die einen Phänotyp mit verkleinerten Testes aufweisen 143
Abb. 4.4: *in situ* Expressionsprofil von *F54E4.1* 145
Abb. 4.5: Varianz des GC-Gehalts verschiedener Genome und Isochoren-Klassen des humanen Genoms 146
Abb. 4.6: Durchschnittliche prozentuale Verteilung repetitiver Elemente in Maus und Mensch 154
Abb. 4.7: Genomische Organisation von DmX verschiedener Spezies im Vergleich 161
Abb. 4.8: FrameWorker-Analyse hochkonservierter Bindungsstellen von Transkriptionsfaktoren zwischen Maus und Mensch 165
Abb. 4.9: Grafische Darstellung der potentiellen Promotor- und Enhancerstruktur von *Dmxl*1 169
Abb. 4.10: Affymetrix-Expressionsspektren des murinen und humanen DmX 178
Abb. 4.11: Sekundärstruktur des 3’-UTR von Maus und Mensch 181
Abb. 4.12: Identifizierte Motive in den abgeleiteten Aminosäuresequenzen von *Dmxl*1, *DMXL*1 und Soi3p/Rav1p 184
Abb. 4.13: Vergleich des identifizierten peroxisomalen Motivs mit der redefinierten humanen Konsensussequenz 188
Abb. 4.14: Dreidimensionale Rekonstruktion von 10 C-terminalen WD-Einheiten des Dmxl1-Proteins 191
Abb. 4.15: β-Blätter bilden das Rückgrat der Propellerstruktur 192
Abb. 4.16: Rekonstruktion von 300 C-terminalen Aminosäuren durch ESyPred3D 193

Tabellenverzeichnis

Tab. 2.1: Standardprotokoll für die DNA-Sequenzierung mit BigDye™ Version 2.0 und BigDye™ Version 3.1 35
Tab. 2.2: Übersicht der im Standardprotokoll für die DNA-Sequenzierung eingesetzten Template-DNA 35
Tab. 2.3: Parameter des 2step55 Protokolls 36
Tab. 2.4: 3-Schritt Standard PCR Bedingungen 37
Tab. 2.5: Standardbedingungen für eine „X-large“ PCR 38
Tab. 2.6: Standardbedingungen einer „High Fidelity“ PCR 38

Inhaltsverzeichnis

Tab. 3.1: Daten aller Injektionstermine 75
Tab. 3.2 Zusammenfassung von Chimärismus und Gewicht der untersuchten Tiere 81
Tab. 3.3: Prozentuales Übergewicht der Chimären Tiere 81
Tab. 3.4: Exons und Introns von Cosmid N10467Q3 87
Tab. 3.5: SIM4 Analyse der murinen genomischen Region von *Dmxl*1 92
Tab. 3.6: SIM4 Analyse der humanen genomischen Region von *DMXLl*1 93
Tab. 3.7: Repetitive Elemente der genomischen Region von *Dmxl*1 und *DMXL*1 94
Tab. 3.8: Tabelle der identifizierten Bindungsstellen für Transkriptionsfaktoren 111
Tab. 3.9: Interspeziesvergleich konservierter Transkriptionsfaktoren 115
Tab. 3.10: Identifizierte Motive in der Aminosäuresequenz von *Dmxl*1 und ihre relative Positionierung auf der CDS 125
Tab. 3.11: Vergleich der Homologen von DMX 126
Tab. 3.12: Durch das Programm REP identifizierte Protein-Repeatsequenzen 128
Tab. 4.1: Zusammenfassung der wichtigsten Gendefekte, die bei der Ausprägung von Fettleibigkeit eine Rolle spielen 140
Tab. 4.2: GC- und GC_3-Gehalt *Dmxl*1 homologer Gene verschiedener Spezies 149
Tab. 4.3: Genomische Organisation von Genen mit einer cDNA >8500 bp 158
Tab. 4.4: Konservierte Intron-Positionen von DmX zwischen Maus, Mensch, Drosophila, Chironomus und Caenorhabditis 160
Tab. 4.5: Durch FrameWorker-Analysen identifizierte Transkriptionsfaktoren die im Promotorbereich von *Dmxl*1 binden 166
Tab. 4.6: Identifizierte Transkriptionsfaktoren in konservierten Sequenzblöcken stromaufwärts des Promotors von *Dmxl*1 170
Tab. 4.7: Durch kombinatorische Analyse identifizierte Transkriptionsfaktoren 173
Tab. 4.8: Durch Interspeziesvergleich identifizierte Transkriptionsfaktoren 175
Tab. 4.9: Gewebe mit erhöhter transkriptioneller Aktivität 179

Abkürzungen:

3D	dreidimensional
Abb.	Abbildung
abs.	absolut
A. bidest.	Aqua bidestillata
Amp	Ampicillin
AP	alkalische Phosphatase
BLAST	„Basic Local Alignment Tool“
bp	Basenpaare
BSA	Rinderserumalbumin
bzw.	beziehungsweise
ca.	circa
cDNA	komplementäre DNA
CDS	translatierter DNA-Abschnitt
Chr	Chromosom
d.h.	das heißt
DNA	„desoxyribonucleic acid“
dATP	2’-Desoxyadenosintriphosphat
dCTP	2’-Desoxycytidintriphosphat
dGTP	2’-Desoxyguanosintriphosphat
dTTP	2’-Desoxythymidintriphosphat
dNTP	2’-Desoxynucleosidtriphosphat
dpc.	Tage post coitum
EDTA	Ethylendiamintetraessigsäure
E-Puffer	Elektrophoresepuffer
ES-Zellen	embryonale Stammzellen
et al.	und andere (et altera)
EtOH	Ethanol
g	Erdbeschleunigung
ges.	gesättigt
h	Stunden
het oder +/-	heterozygote Mutante
HR	Homologe Rekombination
IPTG	Isopropyl-β-D-thio-galaktopyranosid
kb	Kilobasenpaare
kDa	Kilodalton
KO oder -/-	homozygot negative Mutante (Knock-Out)
konz.	konzentriert
Lsg.	Lösung
M	molar
MCS	multiple Klonierungsschnittstelle
min	Minuten

ml	Milliliter
mM	millimolar
NaAC	Natriumacetat
NaCl	Natriumchlorid
NaH_2PO_4	Natriumdihydrogenphosphat
NaOH	Natronlauge/Natriumhydroxid
Neo	Neomycin
ng	Nanogramm
nM	nanomolar
nmol	Nanomol
ORF	offener Leserahmen („open reading frame“)
PCR	„polymerase chain reaction“
pH	potentium hydrogenium
pmol	Pikomol
PM	Präinkubationsmedium
RNA	„ribonucleic acid“
RNase	Ribonuklease
RNasin	Ribonuklease-Inhibitor
RNAi	„RNA interference“
mRNA	„messenger ribonucleic acid“
rRNA	„ribosomal ribonucleic acid“
rpm	Umdrehungen pro Minute
RT-PCR	Reverse Transkriptase- PCR
RT	Raumtemperatur
RZPD	Deutsches Ressourcenzentrum für Genomforschung GmbH
SDS	Natrium-Dodecylsulfat
sek	Sekunden
SSC	Standard-Saline-Citrat Puffer
Tab.	Tabelle
TBE	Tris-Borat-Elektrophorese-Puffer
TE	Tris-EDTA-Puffer
TELT	Tris-EDTA-Lithiumchlorid-Triton X-Puffer
TgVektor	Targetingvektor
TM	„Trademark“
TmD	Transmembrandomäne
Tris	Tris (hydroxymethyl)aminomethan
U	Units, Einheiten
unit	definierte Enzymaktivität
ü.N.	über Nacht
UTR	nicht-translatierter Bereich („untranslated region“)
UV	ultraviolett
VE	voll entsalzt
Vol.	Volumen
WD-Repeat	Tryptophan-Asparaginsäure Wiederholungseinheit
WWW	„world wide web“, Internet

X-Gal	5-Brom-4-chlor-3-indolyl-β-D-galakto-pyranosid
z.B.	zum Beispiel

1- und 3- Buchstabencode der Aminosäuren:

A	Ala	Alanin
C	Cys	Cystein
D	Asp	Asparaginsäure
E	Glu	Glutaminsäure
F	Phe	Phenylalanin
G	Gly	Glycin
H	His	Histidin
I	Ile	Isoleucin
K	Lys	Lysin
L	Leu	Leucin
M	Met	Methionin
N	Asn	Asparagin
P	Pro	Prolin
Q	Gln	Glutamin
R	Arg	Arginin
S	Ser	Serin
T	Thr	Threonin
V	Val	Valin
W	Trp	Tryptophan
Y	Tyr	Tyrosin

1 Einleitung

Die Identifizierung von Genen in einer DNA-Sequenz und die Analyse ihrer Funktionen, also die Verknüpfung von Genotyp und Phänotyp, sind grundlegende Ziele der genetischen Forschung. Um diese Ziele zu erreichen, ist es notwendig, viele verschiedene experimentelle und bioinformatische Techniken, die die moderne aber auch klassische biologische Forschung bietet, miteinander zu verknüpfen. Über einen langen Zeitraum hinweg erfolgte die Analyse von Genen über die Entdeckung von Phänotypen, die sich erkennbar vom Wildtyp unterschieden. Solche Mutanten konnten isoliert und näher charakterisiert werden. Später wurde diese Technik durch eine unspezifische Mutagenese der Untersuchungsobjekte verfeinert. Dadurch war es möglich, eine große Zahl von Phänotypen zu erzeugen, die dann eine Untersuchung verschiedener Gene eines definierten Organismus möglich machten. Vergleichende Analysen sind eine grundsätzliche Strategie in der gesamten biologischen und medizinischen Forschung. Ein entscheidender Fortschritt der letzten Jahrzehnte besteht vor allem in der Entwicklung molekulargenetischer Techniken, die es ermöglichten, die eher unspezifische Erforschung von Phänotyp und Genotyp durch eine gerichtete Vorgehensweise zu ergänzen und zu ersetzten. Die rasanten Entwicklungen im Bereich der Klonierung, Amplifikation und vor allem Sequenzierung (Sanger et al., 1977b) von DNA gestattet es in relativ kurzer Zeit, bei überschaubaren Kosten, Gene bekannter Phänotypen zu isolieren und die zugrunde liegenden Sequenzinformationen zu ermitteln. Die Kenntnis der Nukleotidabfolge einer DNA erlaubt aber auch eine umgekehrte Vorgehensweise bei der Analyse von Genen. Anstatt den Weg vom Phänotyp über den Genotyp zur DNA-Sequenz zu gehen, kann hier eine Identifizierung der Gene ohne einen bekannten Phänotyp erfolgen. Die Aufklärung ihrer Funktion und Beschreibung des Phänotyps erfolgt erst anschließend. Diese Vorgehensweise wird als „reverse genetics“ bezeichnet (Plasterk, 1992; Kaiser, 1990; van Ommen et al., 1989; Brown et al., 1988; Reeders, 1987). Dabei wird der klassische Ansatz des Vergleichs von der modernen genetischen Forschung adaptiert, da die Identifizierung von Genen in einer Nukleotidabfolge, ohne die Kenntnis ihrer Existenz durch die Beschreibung eines Phänotyps, gerade bei den Eukaryoten aufgrund ihrer Exon/Intron-Struktur und gewaltiger DNA-Abschnitte ohne erkennbare Funktion äußerst schwierig ist. Die vergleichende Gen- und Genomanalyse, unter Zuhilfenahme der Bioinformatik, hat sich hierbei zu einem unverzichtbaren Werkzeug entwickelt. Mittels Sequenzvergleich verschieden nah verwandter Spezies können durch Identifizierung konservierter und/oder homologer DNA-Abschnitte kodierende Bereiche und funktionale aber nicht kodierende Abschnitte orthologer Gene erkannt werden (Haubold et al., 2004; Miller et al., 2004; Nobrega et al., 2003; Ureta-Vidal et al., 2003; Wei et al., 2003; Gottgens et al., 2002; Pennacchio et al., 2001; Loots et al., 2000; Hardison, 2000; Duret et al., 1997; Hardison et al., 1997a,b). Je mehr sequenzierte Gene verschiedener Spezies zur Verfügung stehen, desto sicherer können solche Voraussagen getroffen werden. Dies gilt vor allem für die weniger gut konservierten Bereiche der genregulatorischen Netzwerke.

Neben der Definition kodierender Sequenzabschnitte und regulatorischer Elemente ist die Analyse der Funktion der Gene, möglicher Transkripte und resultierender Proteine

vor allem bei höheren Eukaryoten ein noch komplexeres Aufgabenfeld. Hier wurden seit Ende der 80er Jahre durch die Weiterentwicklung der Gentechnologie und Etablierung der ersten großen Genomprojekte enorme Fortschritte erzielt. Aktuell werden 1386 Genomprojekte verzeichnet (Stand Januar 2005), davon fallen 654 auf Prokaryoten, 474 auf Eukaryoten sowie 256 auf fertig gestellte und publizierte Genome aller Reiche (http://www.genomesonline.org/, Bernal et al., 2001). Basierend auf den gewonnenen Sequenzinformationen konnten seither zahlreiche eukaryotische Gene gezielt manipuliert und so ihre Funktion in einem Organismus aufgeklärt werden. Die Kenntnis der genomischen Sequenz ist in diesem Zusammenhang eine wichtige Voraussetzung für die Funktionsanalyse. Dennoch beschränkt sich dieser Forschungsbereich derzeit auf wenige gut charakterisierte Modellorganismen. Diese Limitierung basiert vor allem auf der Zugänglichkeit eines Organismus für eine gezielte genetische Manipulation mit den derzeit zur Verfügung stehenden Techniken, der Generationszeit des Organismus, dem Haltungsaufwand und der damit verbundenen Kosten sowie auf schon vorhandenen Erfahrungen aus dem Bereich der Physiologie und Genetik des Organismus. Die klassischen Untersuchungsobjekte der Biologie stehen derzeit noch bei den meisten Forschungsprojekten deutlich im Vordergrund: darunter *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus*, *Rattus norvegicus*, *Xenopus laevis, Escherichia coli*, *Arabidopsis thaliana*. Es sind neuerdings jedoch auch Organismen für diese Techniken zugänglich, die erst durch die Genomprojekte zu Modellorganismen geworden sind: u.a. *Takifugu (Fugu) rubripes*, *Tetraodon nigroviridis*, *Danio rerio* und *Neurospora crassa*. Dies zeigt, dass gerade auf dem Gebiet der genetischen Manipulation große Forschungsanstrengungen unternommen werden. Es ist zu erwarten, dass in den nächsten Jahren eine Vielzahl weiterer Organismen mit Hilfe gentechnologischer Methoden manipuliert werden können und so der funktionellen Analyse von Genen zur Verfügung stehen.

1.1 Vergleichende Genomanalyse und Bioinformatik

Die Verwendung der vergleichenden Sequenzanalyse als Werkzeug zur Identifizierung funktioneller DNA-Abschnitte basiert auf der Hypothese, dass biologisch wichtige Sequenzabschnitte aufgrund ihrer fundamentalen Funktion zwischen verschiedenen Spezies konserviert sind. Sie evolvieren also mit einer geringeren Geschwindigkeit als jene Bereiche im Genom, die keine Funktion ausüben (Gibbs et al., 2004; Nardone et al., 2004; Wystub et al., 2004; Frazer et al., 2003; Hardison et al., 2003; Waterston et al., 2002; Hardison et al., 1997; Gelfand et al., 1996; Koop et al., 1995). Die Konservierung homologer Sequenzabschnitte zeigt sich bei einem Interspeziesvergleich durch Blöcke evolutionär verwandter Sequenzen auf verschiedenen Chromosomen, innerhalb derer Position und Orientierung der Gene konserviert sind. Die Genome von Spezies wie z.B. Maus, Ratte und Mensch haben aufgrund ihrer divergenten Evolution im Laufe der Zeit eine Vielzahl von solchen Chromosomenrearrangements angesammelt (Abb. 1.1).

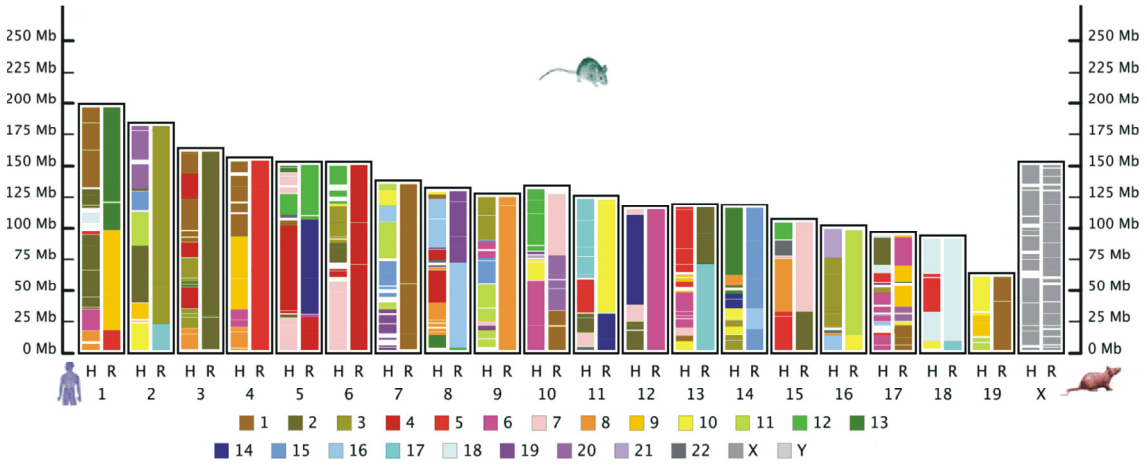


Abb. 1.1: Syntäne Regionen in Maus, Mensch und Ratte.

Bei dieser Darstellung handelt es sich um einen Syntänie-Vergleich von Mensch und Ratte mit der Maus. Auf der Y-Achse ist die Größe der Chromosomen in Megabasen angegeben, auf der X-Achse sind die Chromosomen der Maus aufgetragen. Die Kolorierung zeigt die Syntänie zu den Chromosomen in Mensch (H) und Ratte (R) entsprechend dem darunter befindlichen Farbcode an. Es handelt sich um eine modifizierte Darstellung von www.genboree.org (Gibbs et al., 2004).

Die Identifizierung orthologer Gene und funktionaler DNA-Abschnitte ist eine Grundvoraussetzung für eine erfolgreiche Funktionsanalyse von Genen (Clark, 1999). Idealer Weise werden solche Vergleiche zwischen Spezies gezogen, die eine ähnliche Physiologie oder Biologie aufweisen. Entscheidend für die Anwendung der vergleichenden Genomanalyse ist der Grad der Konservierung zwischen den verglichenen Spezies. Bei zu starker Konservierung ist eine Unterscheidung funktionaler von nicht-funktionalen Bereichen unmöglich; ist die Homologie zu schwach, unterscheiden sich diese Bereiche zu stark und werden nicht erkannt (Bergmann et al., 2002). Der Maus-Mensch Vergleich erfüllt viele der genannten Voraussetzungen. Basierend auf der Beobachtung, dass die Divergenz unabhängig voneinander evolvierender Vertebraten-Genome 0.1-0.5 % pro Millionen Jahre beträgt, wird angenommen, dass sich Maus und Mensch vor ~75-80 Millionen Jahren von einem gemeinsamen Vorläufer getrennt haben (Tamura et al., 2002; Waterston et al., 2002, Murphy et al., 2001, Kumar et al., 1998). Diese evolutionäre Distanz erwies sich aufgrund verschiedener Studien als akzeptabel, wenn auch nicht optimal, da das Genom beider Spezies nicht uniform evolviert, dennoch können funktionale Sequenzen sehr gut identifiziert werden (Pennacchio et al., 2003; Gottgens et al., 2002; Amid et al., 2001; Pennacchio et al., 2001; Kappen et al., 2000; Loots et al., 2000; Tautz, 2000). Für den Vergleich beider Spezies lassen sich u.a. folgende Ziele definieren: 1. Annotation vorher unbekannter Gene, 2. Aufklärung der Genstruktur in verwandten Spezies, wenn lediglich die cDNA aus einer Spezies bekannt ist, 3. Identifikation von regulatorischen Elementen und Protein-Motiven, 4. Charakterisierung der Bindungsmotive von Transkriptionsfaktoren, 5. Syntenie, Isochorenklassen, evolutionäre Entwicklung.

Die Identifikation neuer Gene gehört sicher zu den wichtigsten dieser Ziele. Durch den Vergleich zweier Genome mit ausreichendem evolutionärem Abstand zueinander, lassen sich Regionen finden, die unter einer starken negativen Selektion stehen. Sehr häufig handelt es sich hierbei um kodierende Abschnitte aktiver Gene in beiden Spezies. Die Identifizierung solcher homologer Abschnitte kann ein guter Anhaltspunkt für weitere Untersuchungen sein, wenn es sich hierbei um völlig neue und unbekannte Gene handelt, die auch in den EST-Datenbanken nicht verzeichnet sind, also eine *de-novo* Genvorhersage. Die Umkehr dieses Verfahrens ist ebenfalls eine häufige Anwendung. Vorhandene EST-Daten können verwendet werden um die Genstruktur homologer Gene in den verglichenen Spezies zu annotieren, zudem können die EST-Einträge selbst überprüft werden (Bergman et al., 2002; Haas et al., 2002). Konservierte Sequenzelemente (CSEs; Xuan et al., 2002) weisen aber nicht nur auf Gene und ihre Exonstrukturen hin, sondern sind ganz allgemein mit einer möglichen Funktion verbunden. Dies bedeutet für nicht-kodierende funktionale DNA-Abschnitte, dass es sich bei diesen konservierten Sequenzen um regulatorische Elemente wie „Enhancer“, „Silencer“ oder Bindungsmotive für Transkriptionsfaktoren in Promotoren, aber auch um 5’- und 3’-untranslatierte Regionen sowie wichtige strukturelle Bereiche, die die Integrität der DNA gewährleisten, handeln kann (Smith et al., 2004; Elnitski et al., 2003; Glazko et al., 2003; Bergman et al., 2002; Dermitzakis et al., 2002; Ludwig, 2002; Majewski et al., 2002; Shabalina et al., 2001; Fickett et al., 2000). Die Möglichkeit, solche Motive durch den Vergleich evolutionär entfernt verwandter Spezies zu identifizieren, ist nicht sofort einsichtig, da z.B. die Bindungsmotive für Transkriptionsfaktoren meist aus kurzen 6-12 bp langen Sequenzabschnitten bestehen und solche kurzen Motive im Hintergrundrauschen zweier Spezies, die durch 80 Millionen Jahre Evolution getrennt sind, untergehen würden. Da aber die meisten „Enhancer“, „Silencer“ und Promotoren aus einer Abfolge multipler Bindungsstellen verschiedener Transkriptionsfaktoren bestehen, die in modularen Blöcken angeordnet sind, ähneln sie in Aufbau und Größe den Exon/Intron-Strukturen von Genen. Der Interspeziesvergleich wurde schon lange vor der Verfügbarkeit der Sequenzinformationen ganzer Genome verwendet, um Gene und ihre regulatorischen Elemente zu Identifizieren. Schon 1995 erkannte Koop bei Sequenzvergleichen der homologen Gene für Myosin, Immunoglobulin, Crystallin, Globulin und dem T-Zell-Rezeptor von Maus und Mensch, das nicht alle nicht-kodierende Sequenzen im gleichen Maße divergieren. Er folgerte, dass diese Regionen möglicherweise für die strukturelle Funktion der Chromosomen wichtig sein könnten. Zwei Jahre später nutzten Oeltjen et al. diese Technik, um für fünf Gene des BTK-Lokus flankierende regulatorische Regionen vorherzusagen, die potenzielle Bindungsregionen für Transkriptionsfaktoren darstellen (Oeltjen et al., 1997). Es folgte eine Vielzahl weiterer Arbeiten, die zeigten, dass die vergleichende Genomanalyse neben der *de novo* Identifizierung bisher unbekannter Gene vor allem bei der Beschreibung cis-regulatorischer Elemente und nicht-kodierender Sequenzabschnitte von Genen zu einer deutlich verbesserten Vorhersagegenauigkeit geführt hat (u.a. Kloster et al., 2005; Ryder et al., 2005; Gaffney et al., 2004; Sinha et al., 2004; Kloster et al., 2004; Kolbe et al., 2004; Boffelli et al., 2003; Eisen et al., 2003; Hardison et al., 2003; Thomas et al., 2003; Nobrega et al., 2003; DeSilva et al. 2002; Mural et al., 2002; Dehal et al., 2001; Ansari-Lari et al.,

1998). Gerade diese Bereiche waren zuvor kaum zu definieren und nur experimentell zu charakterisieren.

Vergleichende Genomanalysen sind untrennbar mit ihrem wichtigsten Werkzeug verbunden, der Bioinformatik. Was aber ist Bioinformatik? Das Ziel der modernen Biologie ist es, Daten aus Genomik und Proteomik zu komplexen Netzwerken zu verbinden, die dann Stoffwechselwege und Physiologie eines Organismus beschreiben können. Biologische Systeme basieren auf Molekülen, die Bioinformatik beschreibt diese Moleküle und die damit verbundenen Fragestellungen mit Hilfe der Mathematik, Statistik und Informatik und versucht, alle verfügbaren Informationen zu integrieren und mit Hilfe von Algorithmen zu verarbeiten, zu organisieren und zu verknüpfen um so *de novo* Vorhersagen für uncharakterisierte Sequenzen zu ermöglichen (Abb. 1.2; Yu et al., 2003; Luscombe et al., 2001). Die Bioinformatik selbst unterliegt ebenfalls einem integrativen Prozess. Von Mitte der 60er (z.B. Zuckerkandl et al., 1965) bis in die frühen 90er Jahre (z.B. Altschul et al., 1990) des letzten Jahrhunderts beschränkten sich bioinformatische Lösungen auf ganz spezifische einzelne Fragestellungen. Dies lässt sich vor allem mit der erst beginnenden Entwicklung der Computertechnologie begründen, die notwendige leistungsstarke Hard- und Software stand nicht zur Verfügung (Ouzounis et al., 2003). Erst die Entwicklung und der Ausbau des Internets sowie die Verdoppelung der Anzahl der Transistoren auf integrierten Schaltungen alle 18 Monate (Mooresches Gesetz) in Kombination mit benutzerfreundlichen Betriebssystemen schafften die Vorraussetzungen, um überhaupt die anfallenden Datenmengen bearbeiten zu können. Das bedeutet aktuell eine Verdoppelung der Rechenleistung von Computern alle 12 Monate und ist insofern von Bedeutung, da auch die Zahl der Sequenzdaten noch immer exponentiell zunimmt. Im Jahr 1982 lag die Menge der sequenzierten Nukleotide bei 680.338 bp, bis Ende 2004 ist sie auf 44.5 Milliarden angestiegen (http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html). Allerdings befindet sich die Umsetzung bestehender Erkenntnisse zu erfolgreich arbeitenden Programmen, die eine umfassende computergestützte Analyse der Sequenzdaten und Vorhersage aller relevanten Daten funktionaler DNA-Bereiche erlauben würden, also die Bioinformatik, noch in der Entwicklung. So zeigte sich in vielen Arbeiten, dass eine ganze Reihe von Vorhersageprogrammen und bioinformatischen Ansätze für z.B. Genstruktur, Promotoren und Charakterisierung nicht kodierender funktionaler Sequenzabschnitte nur unzureichend arbeiten (de Val et al., 2004; Jamet, 2004; Mujica, 2004; Frazer et al., 2003; Amid, 2002; Bahr, 1999; Bikar, 1997). Dieser Umstand hat sich generell durch die Integration der vergleichenden Genomanalyse und die Zusammenfassung vieler verschiedener bioinformatischer Werkzeuge zu komplexen Vorhersageprogramme geändert, die Vorhersagegenauigkeit hat sich damit dramatisch verbessert (Fraser et al., 2004; Hsu et al., 2004; Sinha et al., 2004; Frazer et al., 2003; Hotz-Wagenblatt et al., 2003; Yu et al., 2003; Leung et al., 2002). Dennoch ist es derzeit noch notwendig, manuell möglichst viele verschiedene Programme für die Lösung einer Fragestellung zu kombinieren, um eine akzeptable und fundierte Vorhersagen treffen zu können.

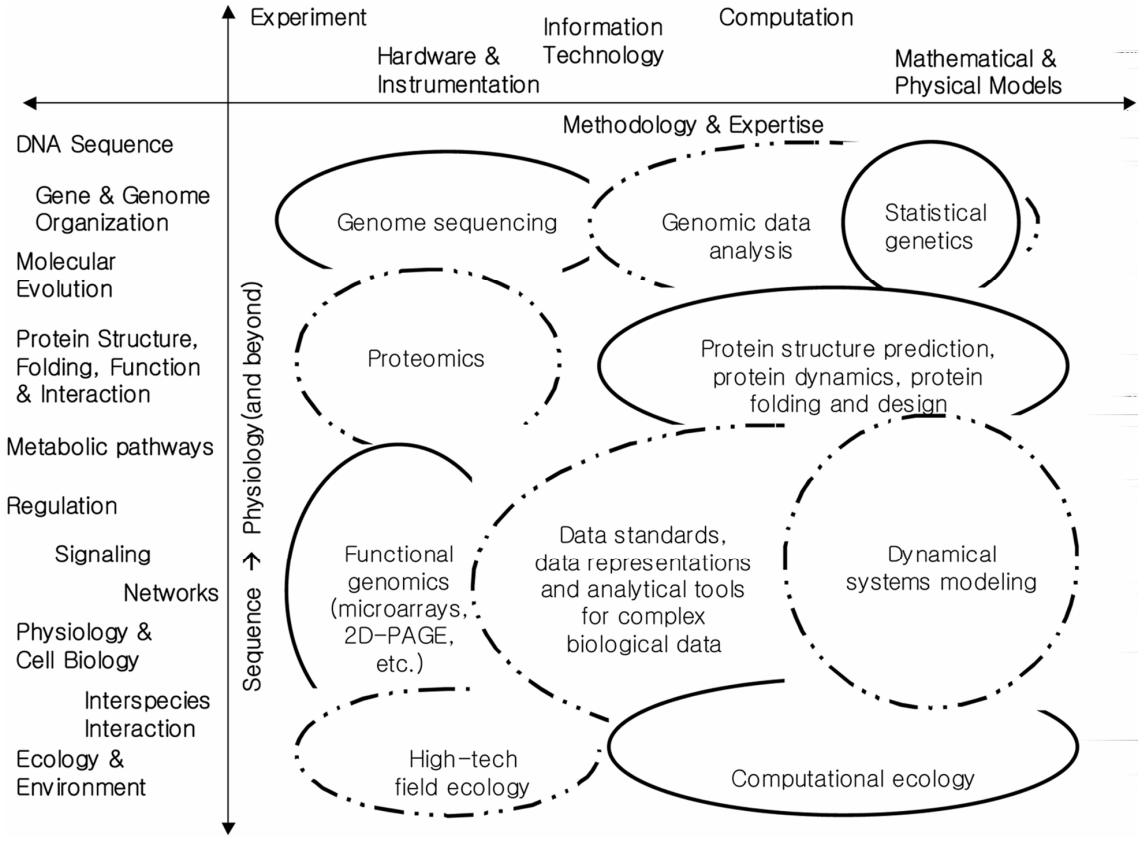


Abb. 1.2: Bioinformatik in der biologischen und medizinischen Forschung (Quelle: Yu et al., 2003)

1.2 Die Maus als Modellsystem – Funktionsanalyse und Genomik

1.2.1 Die klassische Mausgenetik

Theoretisch könnte die Maus bereits mehr als 150 Jahre der Genetik als Modellsystem dienen, wenn Mendel seinen ursprünglichen Plan, Erbgänge anhand der Fellfarbe von Mäusen zu untersuchen, hätte umsetzen können. Bischof Anton Ernst Schaffgotsch empfand es jedoch als unangemessen, dass ein Mönch in seinem Quartier Tiere studiert, die sich fortpflanzen (Hennig, 2002). So dauerte es weitere fünfzig Jahre, bis 1902 Lucien Cuénot nach der Wiederentdeckung der von Mendel formulierten Vererbungsregeln zeigen konnte, dass sich diese Gesetze auch auf die Vererbung der Fellfarbe bei Mäusen anwenden lässt (Cuénot, 1903, 1902). 1905 beschrieb Cuénot mit dem A^y-Allel (Agouti-Lokus) die erste letale Mutation bei Mäusen die zu einer Segregation führte, die nicht den Mendel-Regeln entsprach. 1910 konnte dies durch Castle und Little korrigiert werden, sie fanden die „fehlenden“ Mäuse als tote Embryos in den Muttertieren. Ein Jahr zuvor begann Little mit Kreuzungen für einen Inzuchtstamm (DBA-Stamm). Zusammen mit Tyzzer arbeitete Little in den folgenden Jahren daran, einen genetisch uniformen Inzuchtstamm zu erzeugen, um an diesem die

Mechanismen der Tumortransplantation zu untersuchen. 1916 konnten sie zeigen, dass die Akzeptanz von Tumortransplantaten Mendel'schen Regeln unterliegt und schlossen aus ihren Untersuchungen, dass 14-15 Gene an der Abstoßung der Transplantate beteiligt sein müssen. 1929 gründete Little das „Jackson Laboratory", das noch heute zu den führenden Instituten im Bereich der Maus-Genetik gehört. Bis 1980 wurden weltweit 300 Inzuchtstämme erzeugt (Staats, 1980). Für keinen anderen Vertebraten wurde jemals eine solche Vielfalt diverser uniformer Stämme gezüchtet. Die meisten der heute verwendeten Labormäuse basieren auf den von Little etablierten Stämmen.

Die ersten 50 Jahre der Maus-Genetik wurden von zwei Themen beherrscht: Identifizierung der genetischen Faktoren, die eine Abstoßung oder Akzeptanz eines Tumortransplantats bestimmen und die Erforschung der genetischen Basis spontan auftretender Neoblastome, die letztendlich zur Entdeckung der Retroviren und ihrer Rolle in der Tumorentwicklung führte (Hickman und Cairns, 2003; Paigen, 2003). In den 60er Jahren änderte sich der Fokus der Mausforschung, man erkannte, dass sich die entwickelten Inzuchtstämme für viele unterschiedliche Fragestellungen eignen. 1961 publizierte Mary Lyon ihre Hypothese, dass bei Mammaliern ein X-Chromosom im frühen Embryo zufällig inaktiviert wird. Diese Hypothese basiert auf der Beobachtung der Expression X-chromosomaler Mutationen bei Mausinzuchtstämmen, die zu einem sichtbaren Phänotyp führten. Untersuchungen an Chromosomenaberrationen der Maus zeigten, dass das Y-Chromosom das Geschlecht bestimmt. Dies machte klar, dass Geschlechtsbestimmung und Dosiskompensation bei Säugern nach einem völlig anderen Mechanismus erfolgten als die zu diesem Zeitpunkt anerkannte klassische Erklärung, die auf Untersuchungen an *Drosophila melanogaster* basierte. Die Analyse einer großen Zahl von Maus-Mutanten erbrachte völlig neue genetische und physiologische Einblicke. Eine wichtige Entdeckung war die Erkenntnis, dass der genetische Hintergrund sowie weitere regulierende Gene einen unterschiedlichen Einfluss auf einen bestimmten Lokus ausüben können (Hummel et al., 1972). Hier konnte gezeigt werden, dass die Mutationen „*obese*" und „*diabetic*" in zwei Inzuchtstämmen jeweils unterschiedliche Phänotypen erzeugten, sobald sie sich aber im gleichen genetischen Hintergrund befanden, erzeugten sie einen identischen Phänotyp. Coleman (1978) zeigte später, dass zwei unterschiedliche Mutationen zwar den gleichen Phänotyp erzeugen können, sich jedoch physiologisch voneinander unterscheiden. Ein weiterer wichtiger Meilenstein in der Mausgenetik war die Kartierung des Mausgenoms. Die Verfügbarkeit von Markern ist für jedes genetische Untersuchungsobjekt eine wichtige Voraussetzung, um bei Kreuzungsversuchen die Erbgänge definierter Chromosomenregionen verfolgen zu können. Eine genetische Karte, die die Abfolge dieser Marker und ihre Entfernungen zueinander darstellen kann, ist ein wichtiges genetisches Werkzeug. Haldane et al. publizierten 1915 die erste Kopplungsgruppe in der Maus zwischen „*albino*" und „*pink-eyed*". 1935 waren erst 11 Marker in der Maus bekannt, in den folgenden Jahren beschleunigten sich die Kartierungsbemühungen und ereichten schließlich ein exponentielles Wachstum (Eicher, 1981). Die Identifizierung von genetischen Markern lässt sich in drei experimentelle Bereiche einteilen. Die Beschreibung der ersten Marker basierte auf Untersuchungen morphologisch sichtbarer Mutationen, die z.B. die Fellfarbe oder

Skelettanomalien betrafen. Durch die Entwicklung biochemischer Untersuchungsmethoden (z.B. elektrophoretische Auftrennung von Proteinen) und die Entwicklung rekombinanter Inzuchtstämme (Bailey, 1971; Taylor, 1978) wurde es dann möglich, Proteinvarianten zu detektieren und zu kartieren. Das jüngste Feld in den Kartierungsbemühungen des Mausgenoms ist die Identifizierung von Polymorphismen in der DNA-Sequenz (RFLPs, SSLPs, SNPs). Die Möglichkeit, auf molekularer Ebene Gene zu kartieren, führte zu einem immensen Anstieg bekannter Marker im Mausgenom (Lyon, 1990). Diese Entwicklungen in der klassischen Mausgenetik erzeugten fundamentale Einblicke in die Erbgänge und genetischen Zusammenhänge der Säugergenome im speziellen und Eukaryotengenome allgemein.

1.2.2 Entwicklungen in der modernen Mausgenetik

Ohne Zweifel können die Publikationen, die in den Monaten um den Jahreswechsel 1980/1981 erschienen, als Meilenstein der genetischen Forschung bezeichnet werden. Nahezu gleichzeitig wurde von fünf Arbeitsgruppen gezeigt, dass virale oder artfremde Säuger-DNA in das Mausgenom integriert werden kann (Brinster et al., 1981; Costantini et al., 1981; Habers et al., 1981; Wagner et al., 1981; Gordon et al., 1980). Das entscheidende Experiment wurde von der Gruppe um Wagner durchgeführt, die zeigte, dass ein intaktes β-Globin-Gen aus Kaninchen in das Mausgenom integriert werden kann und dort spezifisch und korrekt exprimiert wird. Die erste transgene Maus mit einem Kaninchen-Protein in Maus-Erythrozyten war entstanden. Dieses Experiment zeigte, dass ein Gen nicht an einen bestimmten Lokus auf einem Chromosom gebunden sein muss, sondern sich fast überall befinden kann, solange alle regulatorischen Elemente, die für eine korrekte Expression notwendig sind, kompatibel sind und die chromosomale Region nicht eine Expression ausschließt (z.B. heterochromatische Bereiche). Zudem wurde klar, dass die entwicklungs- und gewebsspezifischen regulatorischen Systeme zwischen verschiedenen Säugern konserviert sein müssen. Die wichtigste Erkenntnis war jedoch, dass es nun möglich war, die Gene aus Mensch und anderen Mammaliern in dem durch die klassische Genetik gut etablierten Modellsystem der Maus zu untersuchen. Diese Publikationen führten später zu einer wahren Flut weiterer transgener Tiere. Schon ein Jahr später fusionierten Palmiter und Brinster das Gen für das Wachstumshormon aus der Ratte mit dem mausspezifischen Metallothionein-Promotor und erzeugten mit diesem Konstrukt eine transgene Riesen-Maus (Palmiter et al., 1982, Abb. 1.3).

Abb. 1.3: Transgene Riesen-Maus (links)

(Palmiter et al., 1982; Foto von Brinster)

Im Jahr 1981 wurde noch eine weitere Entdeckung gemacht, die einige Jahre später zu einer der wichtigsten Entwicklungen in der Mausforschung führte. Evans, Kaufmann und Martin etablierten die ersten Kulturen pluripotenter embryonaler Stammzellen (Evans et al., 1981; Martin, 1981). 1984 konnte gezeigt werden, dass diese Zellen nach vielen Generationen in Zellkultur, injiziert in Blastozysten, an der Entwicklung des Organismus teilhaben und ihr Genom auch vererbt werden kann (Bradley et al., 1984). Die ersten chimären Mäuse waren entstanden. Zwei Jahre später wurden genetisch manipulierten ES-Zellen entwickelt, die eine Neomycin-Resistenz trugen oder eine retrovirale Insertion aufwiesen. Beide Zelllinien konnten die erworbene Information über chimäre Mäuse an Nachkommen stabil vererben (Doetschman et al., 1986; Gossler et al., 1986; Robertson et al., 1986). Parallel zu dieser Entwicklung etablierte Smithies et al. (1985) eine Technik, die als Gen-Targeting bezeichnet wird. Seine Arbeitsgruppe klonierte Sequenzabschnitte des humanen β-Globin-Gens in ein Plasmid und integrierte dieses Konstrukt mittels homologer Rekombination in den β-Globin-Lokus von Kulturzellen. Zwei Jahre später wurde in den Arbeitsgruppen von Smithies und Capecchi ein erfolgreiches Gen-Targeting an embryonalen Stammzellen publiziert. Hierbei wurde in das *HPRT*-Gen (Hypoxanthin-Phosphoribosyl-Transferase) durch Gen-Targeting eine Mutation eingeführt und die manipulierten Zellen durch natürliche Selektion isoliert (Thomas et al., 1986, 1987). Smithies und Kollegen korrigierten eine *HPRT*-Mutation durch Gen-Targeting in ES-Zellen, die zuvor auf klassischem Weg auf diesen Defekt hin selektiert wurden (Doetschman et al., 1987; Hooper et al., 1987; Kuehn et al., 1987). Capecchis Arbeitsgruppe entwickelte zudem eine Technik, mit der sich die transformierten ES-Zellen positiv und negativ selektieren lassen. Dieses System steigerte die Effizienz des Gen-Targetings durch die gezielte Selektion erfolgreich transformierter Zellen um das 2000fache (Mansour et al., 1988). Ab diesem Zeitpunkt war es möglich, durch Gen-Targeting auch nicht-selektierbare Gene zu manipulieren und die Folgen dieses Eingriffs in den Nachkommen zu studieren. Ein weiteres Jahr später wurden von drei Gruppen die ersten „Knock-Out"-Mäuse vorgestellt. Disruptiert bzw. korrigiert (*HPRT*) wurden die Gene *β_2-Microglobulin* (Koller et al., 1989), *HPRT* (Thompson et al., 1989) und *int-1* (Thomas et al., 1990). Alle drei Manipulationen transduzierten erfolgreich. Evans, Smithies und Capecchi erhielten für ihre Entwicklungen auf diesen Gebieten 2001 den Lasker-Preis und 2007 den Nobelpreis. Der Ablauf eines typischen Gen-Targeting-Experiments ist in Abb. 1.4 gezeigt und wird in der Abbildungslegende detailliert beschrieben. Gene, die sich nach einem „Knock-Out" als letal erwiesen, konnten jedoch auf diese Weise nicht näher untersucht werden. Dies änderte sich, als 1992 das Cre/loxP-System von zwei Arbeitsgruppen nahezu gleichzeitig publiziert wurde (Lakso et al., 1992; Orban et al., 1992). Es handelt sich bei dieser Methode um ein konditionales Gen-Targeting. Das Prinzip besteht darin, dass das vom Bakteriophagen P1 stammende Cre-Rekombinase-Gen verwendet wird, um einen transgenen Mausstamm zu erzeugen, in dem dieses unter Kontrolle eines wählbaren Promotors steht. Die Zielsequenzen (lox) der Cre-Rekombinase, es handelt sich hierbei um kurze 13 bp lange invertierte Repeats, flankierten ein Targeting-Konstrukt, das mittels Gen-Targeting die endogene Zielsequenz in einem anderen Mausstamm ersetzt. Durch Kreuzung des Cre-transgenen

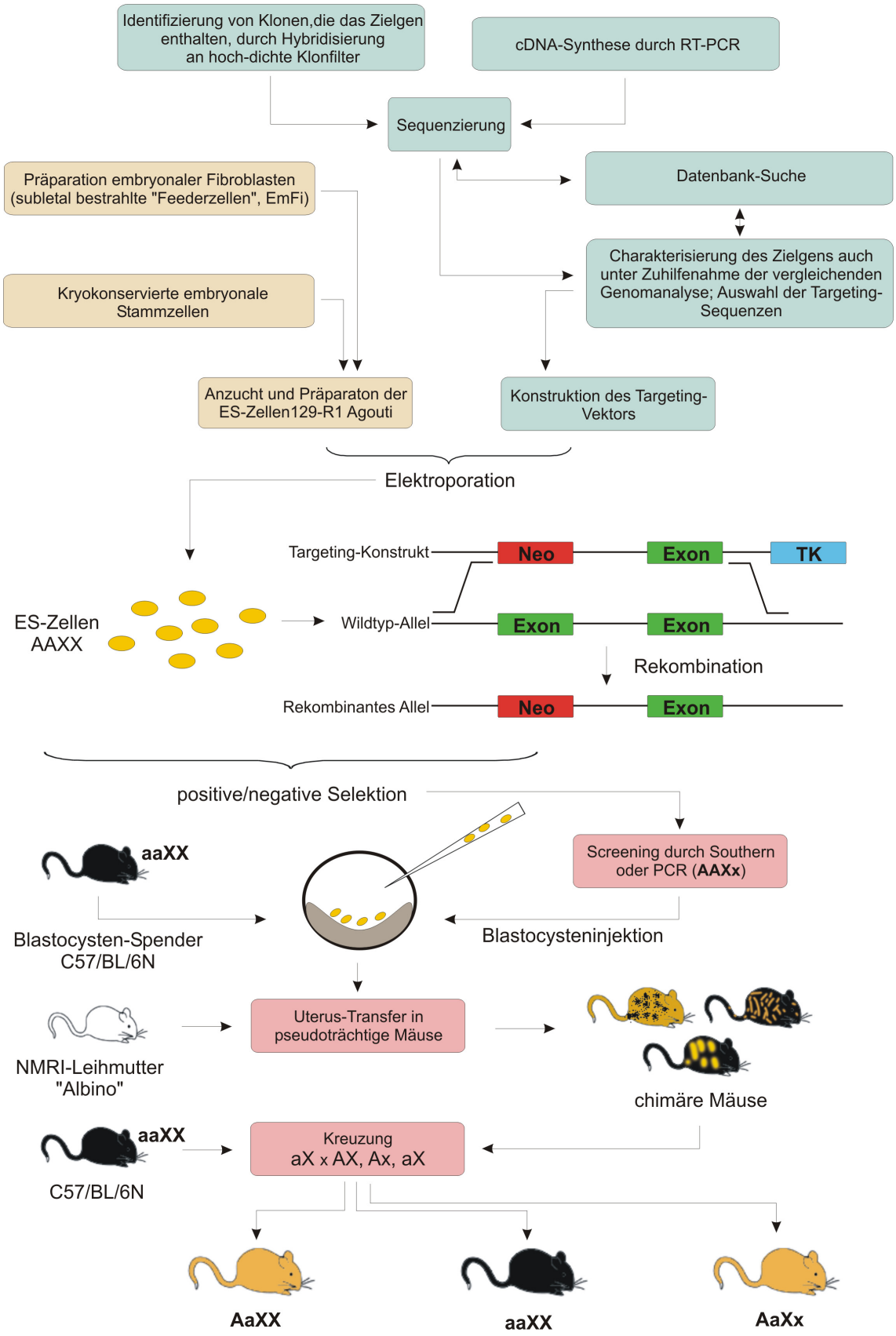

Identifizierung von Klonen,die das Zielgen enthalten, durch Hybridisierung an hoch-dichte Klonfilter
cDNA-Synthese durch RT-PCR
Sequenzierung
Präparation embryonaler Fibroblasten (subletal bestrahlte "Feederzellen", EmFi)
Datenbank-Suche
Charakterisierung des Zielgens auch unter Zuhilfenahme der vergleichenden Genomanalyse; Auswahl der Targeting-Sequenzen
Kryokonservierte embryonale Stammzellen
Anzucht und Präparaton der ES-Zellen129-R1 Agouti
Konstruktion des Targeting-Vektors
Elektroporation
Targeting-Konstrukt
Neo
Exon
TK
ES-Zellen AAXX
Wildtyp-Allel
Exon
Exon
Rekombination
Rekombinantes Allel
Neo
Exon
positive/negative Selektion
aaXX
Screening durch Southern oder PCR (AAXx)
Blastocysten-Spender C57/BL/6N
Blastocysteninjektion
Uterus-Transfer in pseudoträchtige Mäuse
NMRI-Leihmutter "Albino"
chimäre Mäuse
aaXX
Kreuzung aX x AX, Ax, aX
C57/BL/6N
AaXX
aaXX
AaXx

Abb. 1.4: Systematischer Ablauf eines Gen-Targeting Experiments

In der gezeigten Abbildung ist der gesamte Ablauf eines „Knock-Out"-Experiments wiedergegeben. Um ein Gen-Targeting erfolgreich durchzuführen, ist es notwendig, die exakte Struktur des Gens und mögliche funktionale Motive in der abgeleiteten Aminosäuresequenz zu kennen. Sollte die gesuchte Sequenz nicht oder nur teilweise in den öffentlichen Datenbanken vorhanden sein, sind zunächst Sequenzierungsarbeiten notwendig. Um die Genstruktur aufklären zu können, ist es z.T. unumgänglich, auch die cDNA zu synthetisieren und zu sequenzieren. Nach der Sequenzierung bzw. Recherche in öffentlichen Datenbanken kann die Gesamtstruktur des Gens mit Hilfe der vergleichenden Genomanalyse und Bioinformatik beschrieben werden. Basierend auf diesen Kenntnissen ist es dann möglich, die Targeting-Sequenzen zu bestimmen und den Vektor zu konstruieren. Um die später verwendeten ES-Zellen zu ernähren, werden embryonale Fibroblasten subletal bestrahlt (EmFi), so dass sich diese nicht mehr teilen können aber überleben. Kryokonservierte Stammzellen (z.B. 129-R1, Fellfarbe „Agouti") werden mit den EmFi in Kultur gebracht und angezüchtet. Der konstruierte linearisierte Targeting-Vektor wird durch Elektroporation in die ES-Zellen tranfiziert und diese danach wieder in Kultur gebracht. In diesen Zellen findet dann in seltenen Fällen eine Rekombination zwischen einem Allel des Zielgens und dem manipulierten Targeting-Konstrukt statt. Der Targeting-Vektor ist so konstruiert, das ein im Verhältnis kurzer manipulierter Abschnitt (z.B. Austausch oder Disruption eines Exons durch eine Neo-Kassette mit integriertem Stopp-Codon in allen Leserahmen) von längeren flankierenden Bereichen mit Sequenzübereinstimmung zum Zielgen umgeben ist. Diese sollen eine sequenzspezifische Integration in das Genom gewährleisten. In den meisten Fällen ereignet sich jedoch keine oder eine unspezifische Integration in das Genom. Anschließend erfolgt daher eine positiv/negativ-Selektion über den verwendeten Resistenzmarker (Neo) und die von der TK-Kassette kodierte Thymidin-Kinase. Zellen die das Targeting-Konstrukt durch Rekombination nicht in das Genom integriert haben, sterben nach Selektion mit G418 durch die fehlende Neomycin-Resistenz ab. Die negative Selektion erfolgt durch die Zugabe des Substrats Gancyclovir. Zellen, bei denen eine Integration in das Genom nicht sequenzspezifisch erfolgte, enthalten die Information zur Kodierung der Thymidin-Kinase. Das resultierende Protein wandelt Gancyclovir in ein toxisches Triphosphat um, die Zellen mit einer unspezifischen Integration sterben ab. Diese Selektionen sind nur zum Teil erfolgreich, so dass anschließend eine Typisierung der überlebenden ES-Zell-Klone durch Southern-Hybridisierung oder PCR notwendig ist. Die Integrationsfrequenz ist von sehr vielen Faktoren abhängig, liegt aber sehr häufig zwischen 1:200 bis 1:1000 der nach Selektion untersuchten Zellklone. Zellen, bei denen ein erfolgreiches Targeting nachgewiesen werden konnte, werden anschließend amplifiziert. Der Genotyp der Zellen ist hier mit AAXx charakterisiert, wobei das Allel „A" für die Fellfarbe „Agouti" steht und „X" für das Zielgen. Das disruptierte Zielgen ist durch ein „x" symbolisiert. Für die Blastocysteninjektion werden Embryos aus Spendertieren entnommen, die für eine andere Fellfarbe kodieren (z.B. C57/BL6/N, Fellfarbe schwarz, Symbol „a"). Nach der Injektion erfolgt ein Transfer der Blastocysten in scheinschwangere Leihmütter, die wiederum eine andere Fellfarbe haben (hier weiß), um eine Kontamination befruchteter Blastocysten bei unvollständiger Vasektomie der verwendeten Männchen zu erkennen. Die durch diesen Prozess generierten Mäuse zeigen einen unterschiedlichen Chimäritätsgrad. Das Verhältnis der Fellfarbe „Agouti" zu der Fellfarbe schwarz zeigt näherungsweise die Gesamtzusammensetzung des Organismus aus der Zelllinie der Blastocystenspenderin und den rekombinanten ES-Zellen. Je höher der Chimärismus zu Gunsten der verwendeten ES-Zellen ist, desto höher ist die Wahrscheinlichkeit einer Keimbahntransformation. Zur Segregation der Keimbahntransformanten werden die Chimären mit Tieren rückgekreuzt, deren Genotyp der Blastocystenspenderin entspricht. Die Nachkommen dieser Kreuzung können drei mögliche Genotypen aufweisen (AaXX, Fellfarbe „Agouti"; aaXX, Fellfarbe schwarz; AAXx, Fellfarbe „Agouti"), nur eine davon ist heterozygot für das disruptierte Allel.

Mausstamms mit dem durch Gen-Targeting manipulierten Mausstamm kommt es zu einer kontrollieren Rekombination zwischen den beiden lox-P Sequenzen, die zwischenliegenden Sequenzabschnitte werden deletiert. Auf diese Weise wurde es

möglich, normalerweise embryonal letale Mutationen in adulten Tieren zu untersuchen. Die Kombination von Gen-Targeting und dem Cre/loxP-System ermöglicht nicht nur einen konditionalen „Knock-Out“, sondern auch die gezielte Umgestaltung ganzer Chromosomenabschnitte und die Herstellung von Balancer-Chromosomen in der Maus (Yu et al., 2001).

Eine weitere Variante dieser „Knock-Out“-Experimente ist das so genannte „Knock-In“. Hierbei wird das Zielgen durch Fusion mit der cDNA eines anderen oder des gleichen Gens, das sich im gleichen offenen Leserahmen befinden muss, deletiert. Durch geschickte Wahl der flankierenden homologen Sequenzen kann entweder das komplette Zielgen eliminiert werden, oder nur Teilbereiche von ihm. Mit dieser Technik lassen sich funktionale Redundanzen ähnlicher Gene untersuchen oder ganz gezielt Mutationen wie Substitutionen bestimmter funktionaler Domänen, Deletionen oder Punktmutationen in das endogene Gen einfügen. Auf diese Weise wurde z.B. die funktionale Redundanz von Cyclin E und Cyclin D1 nachgewiesen. Die konstruierte „Knock-In“-Maus exprimierte Cyclin E mit dem gleichen Muster wie Cyclin D1 und rettete alle phänotypischen Defekte einer Cyclin D1-defizienten Maus (Geng et al., 1999).

1.2.3 Funktionsanalysen in der Maus

Die vollständige Sequenzierung des Maus- und Menschgenoms (Waterston et al., 2002; Lander et al., 2001; Venter et al., 2001) und das Potenzial der vergleichenden Genomanalyse haben das Interesse an der Funktionsanalyse von Genen in der Maus erneut aufleben lassen. Für die Aufklärung der Funktion einzelner Gene *in vivo* ist die Analyse von Phänotypen mutanter Tiere eines der wichtigsten Werkzeuge. Die Erzeugung einer möglichst großen Zahl von Phänotypen in Mäusen ist inzwischen ein wichtiges Forschungsfeld in der post-genomischen Ära (Sung et al., 2004). Der Fokus dieser Bemühungen liegt derzeit bei Techniken, die in kurzer Zeit, ohne viel Aufwand eine sehr große Anzahl an Mutationen erzeugen kann, hierzu gehören ENU (N-Ethyl-N-Nitrosourea; Shibuya et al., 1993), „Gene Trap“-Mutagenese (Friedrich et al., 1991), RNAi (RNA interference; Fire et al., 1998, 1991, Montgomery et al., 1998; Nellen et al., 1993; Izant et al., 1984). Einige dieser Methoden erzeugen jedoch nur zufällige Mutanten, die für das präzise Studium physiologischer und molekularer Funktionen nicht geeignet sind (Valenzuela et al., 2003). Daher sind nach wie vor die spezifischen Methoden wie Gen-Targeting und RNAi (Abb. 1.5) von herausragender Bedeutung (Fujita et al., 2007; Matveeva et al., 2007; De Souza et al., 2006; Dykxhoorn et al., 2006; Sen und Blau, 2006; Sung et al., 2004; Klysik, 2002; Perkins, 2002). Vor allem die jüngste der beiden Methoden, RNAi, wird in Zukunft eine wichtige Rolle spielen. Erstmalig untersucht und optimiert, mit dem Ziel einer umfassenden Funktionsanalyse vieler Gene, wurde RNAi von Fire und Kollegen (Fire et al., 1998; Montgomery et al., 1998).

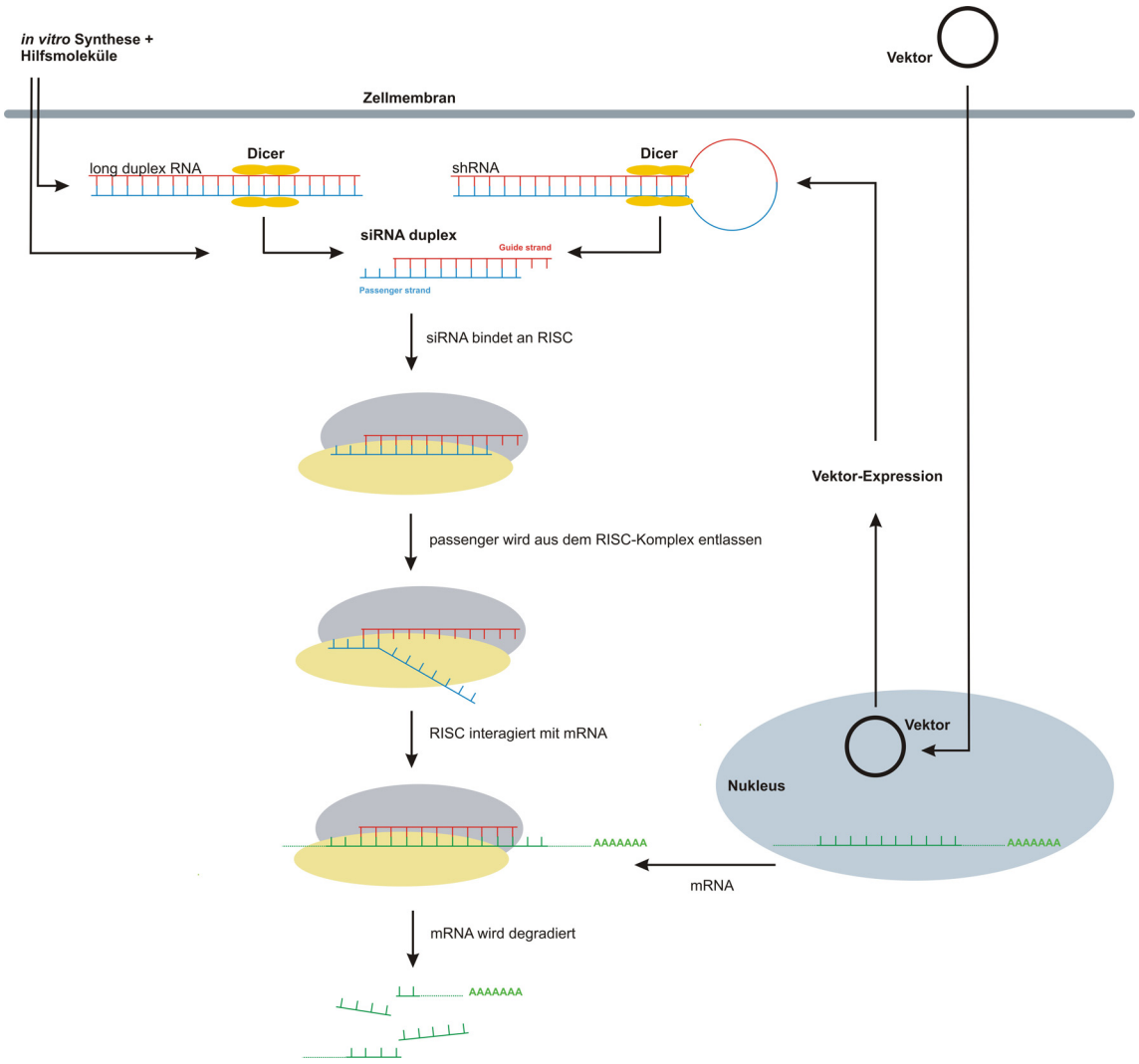


Abb. 1.5: Wirkmechanismus von RNAi

Doppelsträngige RNA (dsRNA) wird von Dicer spezifisch geschnitten, es entstehen kurze doppelsträngige Fragmente mit einer Länge von 20 bis 25 Basenpaaren, die als siRNA (small interfering RNA) bezeichnet werden (MacRea et al. 2006). Bei Dicer handelt es sich um eine hochkonservierte Ribonuklease. Dicer ist auch bei der Beladung des RISC-Komplexes (RNA-induced silencing complex) mit der geschnittenen doppelsträngigen siRNA beteiligt (MacRea et al., 2007). siRNA kann auch in vitro hergestellt werden. Durch Konjugation mit geeigneten Träger-Molekülen ist es möglich, siRNA in Zielzellen einzuschleusen. Eine weitere Möglichkeit spezifische siRNA gezielt in einer Zelle entstehen zu lassen, besteht darin, mit Hilfe eines DNA basierten Vektorsystems durch Expression shRNA (short hairpin RNA) in der Zelle zu synthetisieren, die dann von Dicer geschnitten werden kann (Grim und Kay, 2007). Nachdem die doppelsträngige siRNA in den RISC-Komplex aufgenommen wurden, entwinden sich die beiden komplementären Stränge und nur der Antisense-Strang zur Ziel-RNA verbleibt in dem Komplex. Die einzelsträngige siRNA ist nun in der Lage im Komplex mit RISC und mRNA mit dem komplementären Sequenzabschnitt der mRNA zu paaren. Die bewirkt, dass RISC spezifisch die mRNA in diesem Bereich schneidet, die mRNA wird degradiert.

Entscheidend für einen Einsatz dieser Technik in der Maus war der Nachweis, dass kurze doppelsträngige RNA-Fragmente (small interfering RNA, siRNA) auch in Säugerzellen die Genexpression spezifisch inhibieren können (Hannon, 2002; Caplen et al., 2001; Elbashir et al., 2001). Derzeit sind vor allem Systeme etabliert, die von der RNA-Polymerase III kontrolliert werden und kurze RNAs mit Haarnadelstrukturen exprimieren (Wang et al., 2006; Brummelkamp et al., 2006, 2002; Paddison et al., 2002). Mit diesem System konnte inzwischen die Expression von zahlreichen Genen erfolgreich unterdrückt werden (Dykxhoorn et al., 2006; Kunath et al., 2003). Tiere mit siRNA induziertem „Knock-Down" zeigten den gleichen Phänotyp wie Tiere mit einem entsprechenden „Knock-Out" des untersuchten Gens (De Souza et al., 2006). Problematisch ist bei diesem System, dass es auf einer Transfektion der Zielzellen beruht und für die Expression die ubiquitär aktive RNA-Polymerase III nutzt. Um Stammzellen erfolgreich zu erreichen, wurde jüngst diverse Vektor-Systeme entwickelt, die eine Expression der siRNA auch in schwer zugänglichen Zellen ermöglicht (Fujita et al., 2007; Wang et al., 2006; Rubinson et al., 2003; Tiscornia et al., 2003). Für die Verwendung von RNAi zur funktionellen Analyse vieler Gene *in vivo* ist es notwendig, wie für das Gen-Targeting auch, ein konditionales System zu entwickeln, dass einen gewebsspezifischen und kontrollierten „Knock-Down" der Zielgene möglich macht und im Hochdurchsatz angewendet werden kann. Erste Arbeiten mit Tetracyclin-induzierten U6 Promotoren (Matsukura et al., 2003; Wiznerowicz et al., 2003) und die Verwendung des Cytomegalovirus-Promotors für die gewebsspezifische RNA-Polymerase II-Expression der doppelsträngigen RNA-Fragmente (Silva et al., 2008; Wang et al., 2006; Sen et al, 2006; Xia et al., 2006; Shinagawa et al., 2003) zeigen, dass die RNAi-Technik routinemäßig für die Funktionsanalyse in der Maus eingesetzt werden kann (De Souza et al., 2006).

1.3 Das *DmX*-Gen und seine Homologen

1.3.1 *CpY*, *DmX* und *DMXL*1

CpY

Erstmalig beschrieben wurde ein *DmX*-Gen von Kraemer (1994). Es handelte sich um das Homologe aus *Chironomus piger*, *CpY* (*C. piger* Y-chromosomales Gen). Die Entdeckung dieses Gens beruht auf einer Untersuchung der geschlechtsbestimmenden Region von Chironomiden. Ziel war es, Hinweise auf den dominanten männlichen Geschlechtsdeterminator M zu finden. M liegt in den Spezies *C. thummi* und *C. piger* in der Region D3e1 des Chromosomenarms F (Hägele, 1985, 1986). Zu näheren Charakterisierung dieser Region wurden durch einen Chromosomenmarsch 100 kb der geschlechtsbestimmenden Region in *C. thummi* (Klon Cla1.1) und 70 kb in *C. piger* (Klon Cla1.8) sequenziert (Kunz, 2001; Kraemer et al., 1993). Um Gene in dieser Region zu identifizieren, wurde eine embryonale cDNA-Bibliothek von *C. piger* mit

Sequenzen aus dem klonierten Bereich abgesucht. Hierbei konnte u.a. das *CpY*-Gen identifiziert werden. Die nähere Charakterisierung des offenen Leserahmens zeigte, dass es sich um ein ungewöhnlich großes Gen handelt, die vollständige cDNA von *CpY* ist 11,8 kb groß, genomisch erstreckt sich das Gen über mehr als 15,5 kb und besteht aus 15 Exons und 14 Introns. Die Exons haben eine Länge zwischen 70 bp und 4337 bp, die Introngröße variiert zwischen 90 bp und 781 bp. Die Exons 4 und 8 werden differenziell gespleißt, *CpY* besitzt zudem zwei alternative Polyadenylierungssignale (Kraemer, 1994). Der Sequenzabschnitt von Exon 4 aus *CpY* ist ein integraler Bestandteil des wesentlich größeren Exons 4 von *DmX* und *F54E4.1* (Abb. 4.6), homologe Sequenzabschnitte zu Exon 8 fehlen dagegen in diesen beiden Spezies. Das *CpY*-Gen wird zwar in Männchen und Weibchen gleich stark transkribiert, dennoch gibt hier geschlechtsspezifische Unterschiede. Die Transkriptionsanalyse der differenziell gespleißten Exons zeigt eine mehr oder weniger gleichmäßige Verteilung der Transkripte mit und ohne Exon 4 in den männlichen und meisten weiblichen Geweben. Im Ovar allerdings ist die Anzahl der Transkripte mit enthaltenem Exon 4 überrepräsentiert. Es ist jedoch nicht bekannt, ob Exon 4 für den Transport von RNA in die Oozyte verantwortlich ist. Exon 8 enthaltende Transkripte sind im Männchen deutlich häufiger nachzuweisen als in weiblichen Tieren, im Ovar werden sie überhaupt nicht gefunden. Transkripte mit Exon 8 sind gegenüber Transkripten mit Exon 4 unterrepräsentiert, in Zellkulturzellen können keine Exon 8-Transkripte nachgewiesen werden. Dies kann auf eine spezifische, für die Zelle aber nicht lebenswichtige Funktion von Exon 8 hinweisen. Das differenzielle Spleißen von Exon 8 ist zudem nicht an das differenzielle Spleißen von Exon 4 gekoppelt. Da bei beiden Exons eine Geschlechtsspezifität festgestellt werden konnte, wird ein Zusammenhang der geschlechtsspezifischen Diversifikation mit der Lokalisation von *CpY* vermutet (Weil, 2000). *CpY* kodiert für ein Protein von 3512 Aminosäuren mit einem Molekulargewicht von mehr als 390 kDa (Kraemer, 1994). Northern-Analysen zeigten, dass *CpY* konstitutiv, schwach und in allen Entwicklungsstadien exprimiert wird. „Whole mount" *in situ*-Hybridisierungen an Ovarien von Chironomus zeigten darüber hinaus, dass große Mengen an *CpY*-Transkripten in den Oozyten vorhanden sind. Ob es sich bei *CpY* um M handelt, kann nicht abschließend beurteilt werden, weiterführende Untersuchungen von Kraemer (persönliche Mitteilung) und Weil (2000) machen eine Beteiligung von *CpY* an der Geschlechtsbestimmung jedoch eher unwahrscheinlich. Ein besonderes Strukturmerkmal der abgeleiteten Aminosäuresequnz sind die zahlreichen WD-Wiederholungseinheiten (Kraemer et al., 1998).

DmX

Das *DmX*-Gen (*Drosophila melanogaster* X-chromosomales Gen) wurde durch RT-PCR mit degenerierten Primern von *CpY*-Sequenzen in *Drosophila melanogaster* identifiziert und ist in der Region 5D5/6-5E1 auf dem X-Chromosom lokalisiert (Kraemer et al., 1998). Die cDNA von *DmX* ist 11,5 kb groß, ein Hinweis auf ein alternatives Polyadenylierungssignal, wie bei *CpY,* konnte nicht gefunden werden. Der offene Leserahmen kodiert für 3427 Aminosäuren, das Molekulargewicht des

abgeleiteten Proteins beträgt 380 kDa. Genomisch erstreckt sich *DmX* über ca. 16 kb und enthält, wie *CpY*, 15 Exons und 14 Introns. Dennoch gibt es Unterschiede: das differenziell gespleißte kleine Exon 4 von *CpY* ist bei *DmX* Bestandteil des größten Exons 4, für das differenziell gespleißte Exon 8 aus *CpY* existieren bei *D. melanogaster* keine homologen Sequenzabschnitte. Exon 8 könnte also eine speziespezifische Funktion in *CpY* ausüben. Die Größen der Exons liegen zwischen 67 bp und 6980 bp, die Größen der Introns zwischen 55 bp und 2,4 kb (Weil, 2000; Kraemer et al., 1998). Auffällig ist der sehr kurze 5'-Promotorbereich von *DmX*, zwischen den Transkriptionsstartpunkten von *DmX* und dem benachbarten Gen *DmSPX*, die eine Kopf an Kopf-Orientierung aufweisen, liegen nur 174 bp. Weil (2000) konnte mit Konstrukten, die verschiede Abschnitte möglicher regulatorischer Elemente des *DmX*-Gens enthalten und mit einem Reportergen gekoppelt wurden, nachweisen, dass dieser kurze Bereich nicht ausreicht, um die Transkription von *DmX* zu initiieren. Untersuchungen mit dem Rettungsvektor pCosPerDmX-13 zeigten, dass erst die Anwesenheit von 2,6 kb stromaufwärts und 830 bp des 3' untranslatierten Bereichs von *DmX* den Mutantenphänotyp retten und die wildtypische Expression ermöglichen (Kraemer, pers. Mitteilung; Weil, 2000). Northern-Analysen zeigen, dass *DmX* ubiquitär, konstitutiv, aber mit gewebsspezifischen Unterschieden exprimiert wird. Weil (2002) beobachtete eine erhöhte spezifische Transkriptionsrate in den Vorläuferstrukturen des Mitteldarms, dem ZNS und dem Antenna-Maxillar-Komplex. Eine geschlechtsspezifische Regulation konnte nicht nachgewiesen werden, ebenso wenig eine Funktion als Numerator-Element. Es wurden jedoch durch *In-situ*-Hybridisierungen an RNA von Ovarien während der Oogenese massive Transporte von *DmX*-Transkripten in die Oozyte beobachtet (Weil, 2000). Auch die Genprodukte konnten in der Oozyte durch spezifische Antikörperfärbung nachgewiesen werden, so dass vermutet werden kann, dass es sich bei *DmX* um ein maternales Effektgen handelt (Weil, 2000, Jüttner, 2001). Gestützt wird diese Annahme durch beobachtete larvale Phänotypen von EMS induzierten Mutationen bei *DmX* in *Drosophila melanogaster* (Kraemer, pers. Mitteilung; Jüttner, 2001). Untersucht wurden hier vier Stämme, die nach einer EMS-Mutagenese isoliert werden konnten. Die verwendeten Stämme waren durch den X-chromosomalen Balancer FM7-GFP balanciert, so dass männliche Larven, die das X-Chromosom mit der Mutation tragen, über die fehlenden GFP-Expression selektiert werden konnten. In allen vier Stämmen wurden durch Sequenzierung der cDNA an jeweils unterschiedlichen Positionen „nonsense"-Mutationen im *DmX*-Gen nachgewiesen, die zu einem vorzeitigen Terminationssignal führen und die Synthese eines stark verkürzten Genprodukts zur Folge haben (Jüttner, 2001). Alle Larven der verschiedenen Stämme sterben unabhängig von der Position der Mutationen ungefähr zum selben Zeitpunkt ab. Zuvor verläuft die Entwicklung der mutanten Tiere bis zum Eintritt in das erste Larvenstadium normal. Die Weiterentwicklung in das zweite Larvenstadium kann jedoch von diesen Tieren nicht vollzogen werden. Ihr Körper ist im Vergleich zum Wildtyp deutlich kürzer und dicker, ihre Bewegung träge. Obwohl sie weiter Nahrung aufnehmen ist der Fettkörperanteil gering (Schmidt, pers. Mitteilung). Im weiteren Verlauf verliert der Körper dann an Turgor und Volumen und die Larven sterben. Zusammen mit dem beobachteten massiven Transport von *DmX*-Transkripten in die Oozyte kann angenommen werden, dass die funktionalen maternalen

Genprodukte in der Lage sind, die Funktion von *DmX* während der Embryonalphase und im frühen Larvenstadium aufrecht zu erhalten. Sobald die Larve damit beginnt. auf einen autonomen Stoffwechsel umzuschalten, produziert sie nicht-funktionsfähige DMX-Proteine, ihre Lebensfunktion wird aber noch durch das immer weiter ausdünnende maternale Genprodukt aufrechterhalten. Die Abnahme der maternalen DMX-Konzentration im Organismus führt vermutlich dann zu einem langsamen Tod der Larve (Kraemer pers. Mitteilung; Jüttner, 2001).
Wie bei *CpY* auch, können in der abgeleiteten Aminosäuresequenz viele WD-Repeats identifiziert werden. Je nach akzeptierten Homologiegrad sind bis zu 30 WD-Wiederholungseinheiten denkbar (Weil, 2000; Kraemer et al., 1998)

***DMXL*1**

Das *DMXL*1-Gen (*Dmx-like* 1) wurde, wie *CpY* und *DmX*, erstmalig von Kraemer beschrieben (Kraemer et al., 2000), seine Existenz in *Homo sapiens* ist jedoch durch Datenbanksuchen mit der cDNA-Sequenz von *DmX* schon länger bekannt (Kraemer, persönliche Mitteilung). *DMXL*1 konnte durch eine Fluoreszenz-*in situ*-Hybridisierung (FISH) auf Chromosom 5q22 lokalisiert werden und kodiert für eine cDNA von 11 kb Länge. Die abgeleitete Aminosäuresequnz umfasst 3027 aa mit einem Molekulargewicht von mehr al 380 kDa. Kraemer (2002) konnte in dieser Sequenz bis zu 28 WD-Wiederholungseinheiten identifizieren. Bei der Darstellung der cDNA von *DMXL*1 konnten auch Klone identifiziert werden, die sich von *DMXL*1 in ihrer Sequenz deutlich unterscheiden. Es muss davon ausgegangen werden, dass im Menschen eine zweite Kopie existiert, die als offiziell als *DMXL*2 bezeichnet wird. Weder in Chironomus noch in Drosophila konnte eine solche zweite Kopie identifiziert werden. Dies ist jedoch nicht ungewöhnlich, da viele Gene wie z.B. Ras, Raf und Notch bei *D. melanogaster* nur mit einer Kopie, in Vertebraten jedoch mit mehreren Kopien vertreten sind (Miklos et al., 1996). Die Aminosäuresequenzen beider Gene zeigen eine Similarität von 71,8 % und eine Identität von 55,2 %. Die Genstruktur von *DMXL1* konnte erst in der vorliegenden Arbeit aufgeklärt werden.

1.3.2 DmX in anderen Spezies

Gene, die für die Funktion eines Organismus von grundlegender Bedeutung sind, zeigen in der Regel einen hohen Grad evolutionärer Konservierung. Homologe Gene von DmX wurde bisher in allen komplett sequenzierten eukaryotischen Organismen identifiziert, beginnend mit der Hefe bis zu Maus und Mensch. Zudem existieren in nahezu allen unvollständig sequenzierten Eukaryoten zahlreiche EST-Daten mit Sequenzhomologien zu *DmX*, so dass mit großer Wahrscheinlichkeit angenommen werden kann, dass DmX ein essenzieller Bestandteil der Transkriptome aller eukaryotischen Organismen ist. Bis heute konnte kein prokaryotisches Gen identifiziert werden, das Ähnlichkeiten zu DmX aufweist. Es ist anzunehmen, dass es sich bei DmX um ein eukaryoten-spezifisches Gen handelt und früh in der evolutionären Entwicklung der Eukaryoten entstanden sein

muss. Derzeit sind für 19 verschiedene Spezies die homologen Gene in den öffentlichen Datenbanken vermerkt. Da es sich bei diesen Einträgen in vielen Fällen um reine computergestützte Vorhersagen handelt, können für einen phylogenetischen Vergleich der DmX-Homologen nur annotierte Sequenzen verwendet werden. Evolutionär am weitesten entfernt von Maus und Mensch ist der offene Leserahmen YJR033c von *Saccharomyces cerevisiae*, er kodiert für das Protein Soi3p/Rav1p, das eine wichtige Rolle bei der Zusammensetzung und Kontrolle der vakuolaren H^+-ATPase sowie bei Transportvorgängen zwischen frühen Endosomen und dem prävakuolarem Kompartement/späten Endosomen im Trans-Golgi-Netzwerk spielt (Sipos et al., 2004; Seol et al., 2001). Obwohl das Protein fast zwei Drittel kürzer als DmX ist, enthält es wesentliche konservierte Strukturmerkmale. Vor allem die Positionen der im nächsten Kapitel näher besprochenen WD-Wiederholungseinheiten sind im Vergleich zu Maus und Mensch konserviert. Basierend auf der Annahme, dass sich die Funktionen der Homologen von DmX innerhalb der verschiedenen Spezies ähneln, ist gerade dieser am weitesten entfernte Verwandte von besonderem Interesse. Da für Soi3p/Rav1p zwei Funktionen in der Hefe bereits bekannt sind, ist es nicht unwahrscheinlich, dass es sich hierbei um die ursprünglichen Aufgaben handelt, die auch in höheren Organismen nach wie vor von DmX übernommen werden. Zusätzlich sind hier aber auch erweiterte Funktionen für DmX denkbar.

1.3.3 DmX gehört zu den WD-Repeat-Proteinen

Bioinformatische Analysen der DmX-Homologen verschiedener Spezies zeigen immer ein prädominantes strukturelles Motiv in der abgeleiteten Aminosäuresequnz. Es handelt sich hierbei um Tryptophan-Asparaginsäure-Wiederholungseinheiten (WD-Repeat, WD-40, Abb. 1.6; Neer et al., 1994; van der Voorn et al., 1992), die in unterschiedlicher, ungewöhnlich hoher Anzahl in den DmX-Proteinen gefunden werden. Bei WD-Repeats handelt es sich um minimal konservierte Domänen mit einer Länge von 40-60 Aminosäuren. Häufig befinden sich 11-14 Aminosäuren entfernt vom N-Terminus des Repeats ein Glycin-Histidin (GH)-Dipeptid und ein WD-Dipeptid am C-Terminus der Wiederholungseinheit. In sehr vielen Fällen sind jedoch auch diese beiden Dipeptide nicht konserviert. Zwischen beiden Dipeptiden liegt die insgesamt am stärksten konservierte, ungefähr 40 Aminosäuren lange Kernsequenz. Grundsätzlich zeigen WD-Repeats jedoch eine starke Varianz in allen Positionen. Dies ist vor allem mit der Funktion der WD-Repeat Proteine zu begründen, deren prinzipielle Aufgabe darin besteht, die Zusammensetzung von Multiproteinkomplexen zu koordinieren (Smith et al., 1999). Die WD-Repeats bilden hierbei eine Art Rückgrat für Proteininteraktionen, die simultan und mit vielen verschiedenen Proteinen erfolgen können. Die Konsensussequenz einer WD-Wiederholungseinheit ist in Abb. 1.6 gezeigt. Variabilitäten in jeder einzelnen Wiederholungseinheit erhöhen also das Bindungsspektrum des Proteins. Die Kristallstruktur des bisher am besten untersuchten WD-Repeat-Proteins, die β-Untereinheit des G-Proteins, wurde von zwei Gruppen nahezu gleichzeitig publiziert (Sondek et al., 1996; Wall et al., 1995). Hier konnte

erstmalig die Faltung eines WD-Repeat-Proteins dargestellt werden. Die WD-Wiederholungseinheiten bilden zusammen eine hochsymmetrische Struktur, einen so genannten β-Propeller. Jedes Repeat besteht aus vier antiparallelen β-Faltblättern, die allerdings nicht zusammen für die Ausbildung eines Propellerblatts zuständig sind. Jeder Propeller besteht zwar aus vier β-Faltblattstrukturen, wobei jedoch das erste außen liegende β-Faltblatt durch den d-Strang des vorher gehenden WD-Repeats gebildet wird (Sondek et al., 1996; Ghosh et al., 1995). Es wird angenommen, dass diese Verknüpfung der Propellerblätter die molekulare Struktur stabilisiert (Ghosh et al., 1995; Neer et al., 1994). Dieser prinzipielle Aufbau einer Propellerstruktur konnte durch Röntgenstrukturanalysen an bis heute insgesamt 20 untersuchten WD-Repeat-Proteien bestätigt werden (u.a. Cheng et al., 2004; Madrona et al., 2004; Sengupta et al., 2004; Voegtli et al., 2003; Ono et al., 2003; Orlicky et al., 2003; Pickles et al., 2002; Carr et al., 1999). Aufgrund der Variabilität der einzelnen WD-Repeats dieser Proteine muss davon ausgegangen werden, dass weniger die Konservierung der Aminosäuresequenz der Wiederholungseinheiten eine Rolle spielt, sondern vielmehr die resultierende Sekundär- und Tertiärstruktur für eine Faltung des Proteins zu einem Propeller wichtig ist. Erstaunlicherweise zeigten fast alle bisher kristallographisch untersuchten WD-Repeat-Proteine Propellerstrukturen mit sieben Propellerblättern, wobei im Falle von Aip1p (Voegtli et al., 2003) zwei Propeller mit jeweils sieben WD-Repeats beobachtet werden. Lediglich das prokaryotische WD-Protein TOLB bildet einen sechsblättrigen Propeller aus, obwohl mehr als sieben WD-Wiederholungseinheiten in der Aminosäuresequenz gefunden werden. Die übrigen WD-Repeats erzeugen hier eine separate, mit helikalen Abschnitten gemischte Struktur (Carr et al., 1999). Auch wenn alle anderen untersuchten Proteine 7 oder ein Vielfaches von 7 Wiederholungseinheiten aufweisen, ist es denkbar, dass diese siebenblättrige Konstruktion die prädominante Struktur darstellt (Mikyas et al., 2004; Voegtli et al., 2003; Stöhr et al., 2002; Li et al., 2001; Smith et al., 1999; Murzin et al., 1992). Zukünftige Röntgenstrukturanalysen an Proteinen mit einer Anzahl WD-Repeats, die nicht dem Vielfachen von Sieben entsprechen, werden mit großem Interesse erwartet. Derzeit sind 6268 Proteine, die WD-Repeats enthalten, in den öffentlichen Datenbanken aufgeführt. Die größte Gruppe bilden hier die Eukaryoten mit 6111 Proteinen (Fungi (973), Metazoa (4040), Viridiplantae (712)) gefolgt von den Bakteria mit 101 Proteinen und den Archaea mit 11 Proteinen, zudem sind noch 431 Proteine undefiniert (Quelle: http://smart.embl-heidelberg.de). Allein im Menschen werden 882 WD-Repeat-Proteine gefunden, in der Maus sind es 738. Für keines dieser Proteine werden jedoch so viele WD-Repeats vorausgesagt, wie für DmX. Ähnlich umfangreich wie die Anzahl der Proteine scheint auch die Funktion der WD-Repeat-Proteine zu sein. Aktuell liegen keine verlässlichen Daten über die Funktionsvielfalt der WD-Repeat-Proteine vor, letzte Untersuchungen zeigen jedoch, das die Proteine in weit mehr als 30 funktionelle Gruppen eingeteilt werden können (Li et al., 2001; Yu et al., 2000). Hier seien nur einige genannt: Signaltransduktion, RNA-Synthese, Aufbau von Chromatinstrukturen, vesikularer Transport, Aufbau des Zytoskeletts, Regulation des Zellzyklus und programmierter Zelltod. Dies zeigt, wie bedeutungsvoll die WD-Repeat-Proteine sind, da sie in nahezu allen wichtigen zellulären Prozessen involviert sind.

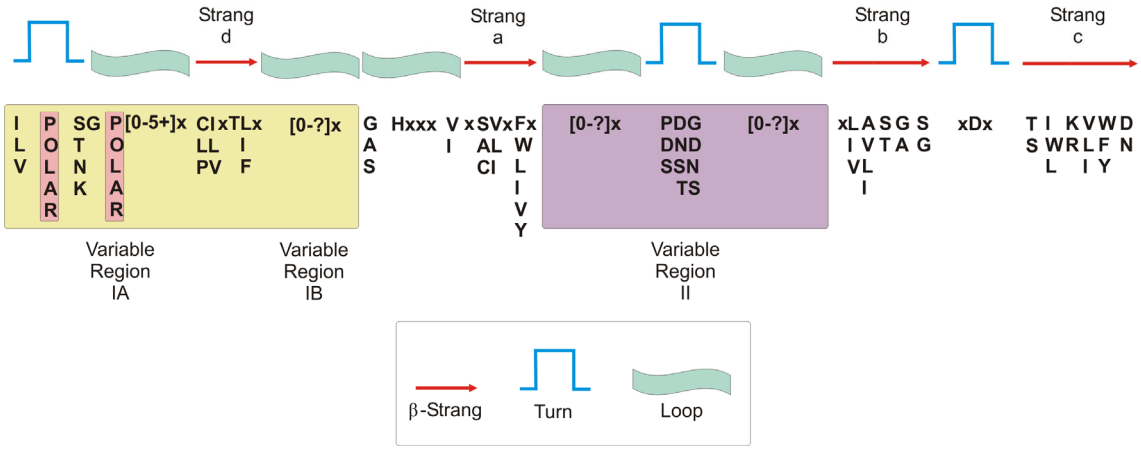


Abb. 1.6: Struktur eines repräsentativen Propellerblatts aus der Propellerstruktur eines typischen WD-Repeat-Proteins

Die Struktur eines Propellerblatts unterscheidet sich von einer WD-Wiederholungseinheit. In einer Wiederholungseinheit gibt es vier Beta-Faltblätter in der Abfolge a, b, c, d, ein Propellerblatt beginnt mit dem β-Strang d der nachfolgenden WD-Wiederholungseinheit und endet mit dem β-Strang c. Der β-Strang d wird dem vorhergehenden Propeller zur Verfügung gestellt. Jeder Propeller hat eine variable Region I (unterteilt in A und B) und II, die vor allem ausgeprägte Längenvariabilität zeigen. Die Zahlen in den eckigen Klammern zeigen die erlaubte Anzahl von Aminosäuren bis zur nächsten Aminosäure der Propellerstruktur. Die oberste Aminosäureabfolge repräsentiert die Konsensussequenz, wobei die mit x gekennzeichneten Positionen von jeder beliebigen Aminosäure besetzt werden können. Die Symbole unterhalb der Konsensussequenz repräsentieren die häufigsten Varianten an bestimmten Aminosäurepositionen.

1.4 Zielsetzung der Arbeit

Obwohl schon umfangreiche experimentelle Arbeiten mit den Homologen von *Dmxl*1 in Chironomous, Drosophila und dem Menschen durchgeführt wurden, konnte die Funktion von DmX in diesen Organismen bislang nicht erkannt werden. Da die bisher untersuchten Spezies evolutionär sehr weit voneinander entfernt sind, ist es auch möglich, dass die Funktion von DmX, obgleich als hochkonserviertes Gen bekannt, bei Chironomus und Drosophila nicht direkt mit der Funktion des DmX-Gens in z.B. Säugetieren vergleichbar ist. Ziel der vorliegenden Arbeit war es daher, die Funktion von *Dmxl*1 in *Mus musculus* experimentell durch Gen-Targeting zu bestimmen.
Da zu Beginn der Arbeit nur Fragmente der cDNA-Sequenz aus Maus und Mensch, aber keine Informationen über die genomische Struktur von *Dmxl*1 in der Maus bekannt waren, sollte zunächst die murine cDNA-Sequenz von DmX durch reverse Transkription und Sequenzierung vollständig aufgeklärt werden. Daten von Kraemer (persönliche Mitteilung) zeigten, dass das humane *DMXL*1 zwar ubiquitär, aber auf sehr niedrigem Niveau exprimiert wird und eine mit *Chironomus* und *Drosophila* vergleichbare Länge von mehr als 11 kb aufweist. Es war anzunehmen, dass eine voll-

längen Darstellung experimentell kaum möglich sein würde. Die Aufklärung der cDNA-Sequenz sollte daher durch eine Kombination von Interspezies-Hybridisierung, Bioinformatik und Primer-Walking erfolgen, wobei vor allem die Daten der humanen cDNA-Sequenz (Kraemer et al., 2000) Verwendung fanden. Mit Hilfe dieser Information sollte in einem zweiten Schritt die genomische Struktur des *Dmxl*1-Gens durch Identifizierung und Sequenzierung der Exon-/Intron-Grenzen aufgeklärt werden.
Bei ausreichender Kenntnis der genomischen Struktur von *Dmxl*1 sollte dann ein Targeting-Vektor konstruiert werden, mit dessen Hilfe ein murines *Dmxl*1-Null Allel in einer geeigneten Maus-Stammzelllinie etabliert werden kann, um schließlich *Dmxl1* „Knock Out"-Mäuse zu erzeugen.
Neben diesen experimentellen Arbeiten sollte das *Dmxl*1-Gen umfassend *in silico* charakterisiert werden. Ziel war es, chromosomale Lokalisation, Struktur der chromosomalen Region, genomische Organisation sowie Gen- und Promotorstruktur bioinformatisch zu untersuchen, möglichst erschöpfend zu beschreiben und mit allen bekannten Homologen von DmX zu vergleichen. Zusätzlich zur DNA-Sequenz sollte auch die abgeleitete Aminosäuresequenz mit Hilfe der Bioinformatik näher charakterisiert werden
Durch Vergleich der experimentellen Ergebnisse des „Knock-Out" von *Dmxl*1 und der bioinformatischen Charakterisierung sollte dann ein mögliches Modell der Funktion von *Dmxl*1 und *DMXL*1 erarbeitet werden.

2 Material und Methoden

2.1 Versuchstiere und Zellen

2.1.1 *Mus musculus*

Die in dieser Arbeit verwendeten Inzuchtstämme der Art *Mus musculus musculus* stammen aus der Zentralen Versuchstiereinrichtung (ZVTE) des Naturwissenschaftlich-Medizinischen Forschungszentrums der Johannes Gutenberg-Universität. Die exakte Bezeichnung der Tiere erfolgt nach der Nomenklatur von Snell (1931), dem „International Committee on Standardized Genetic Nomenclature for Mice“ (1952, 1960, 1976, 1981, 1989, 1996), Festing (1972, 1976, 1979, 1993), Staats (1985) und Maltais et al. (1997). Als Blastozystenspender und für Rückkreuzungen wurden in dieser Arbeit Tiere des Mausstamms C57/BL/6 verwendet. Der Stamm C57/BL wurde 1921 von Little durch die Kreuzung der weiblichen Maus Nr. 57 und der männlichen Maus Nr. 52 aus „Miss Abbie Lathrop's stock“ etabliert. 1937 wurde dieser Stamm durch weitere Inzucht in die Stämme C57/BL/6 und C57/BL/10 aufgespalten. C57/BL/6 ist weltweit der am weitesten verbreitete und genetisch sowie physiologisch am besten charakterisierte Inzuchtstamm. Als Ammen-Tiere dienten Weibchen des Inzuchtstamms NMRI. Dieser Stamm wurde 1937 von Poiley (National Institutes of Health (NIH)) durch Inzucht etabliert und trug den Namen NIH/P1. Diese Linie wanderte anschließend an das Naval Medical Research Institute, USA (NMRI). Von dort wurde die Linie am Zentralinstitut für Versuchstierzucht in Hannover durch zufällige Kreuzungen als Auszuchtstamm etabliert (Rapp, 1972) Die Tiere sind Albinos und eignen sich daher sehr gut als Leihmütter, eine Kontamination würde durch die Fellfarbe sofort erkannt werden. Für die Verpaarung mit den Ammen-Tieren wurden durch Vasektomie sterilisierte männliche Tiere des Stamms NMRI verwendet, auch hier könnte anhand der Fellfarbe eine unvollständige Sterilisation erkannt werden.

2.1.2 Embryonale Stammzellen Typ 129-R1

Alle verwendeten ES-Zellen entsprechen dem Typ 129-R1. Die genaue Bezeichnung erfolgt nach den Regeln des „International Committee on Standardized Genetic Nomenclature for Mice“ (1952, 1960, 1976, 1981, 1989, 1996) basierend auf den Charakterisierungen von Substämmen durch Simpson et al. (1997) und Threadgill et al. (1997). Für alle Arbeiten an embryonalen Stammzellen wurden Kulturen des Substammes R1 (Passage 16) verwendet. Etabliert wurde dieser Stamm 1991 aus 129/Sv x 129/Sv-CP F1 3.5 Tage alten Blastocysten. Diese Zelllinie ist heterozygot für den C Locus (Fellfarbe: +/c(ch)) und den pink-eye Locus (+/p). Die Abkürzung des Stamms zu 129-R1 ist legitim. Die Fellfarbe der adulten Tiere ist in der F1-Generation uniform „agouti“ (gelbbraun), in der F2-Generation erfolgt eine Segregation der Gene für die Fellfarbe in Abhängigkeit vom Partner der Keimbahn-Chimäre (Nagy et al., 1993).

2.2 Standardmethoden

2.2.1 Restriktion von DNA

Bei Verdau von DNA mit Restriktionsendonukleasen wurde nach den Angaben der Hersteller verfahren. Das Volumen der Restriktionsansätze richtete sich nach Art und Konzentration der verwendeten DNA (genomische DNA, Plasmid-DNA). Üblicherweise wurden 20 µl- (analytisch) oder 150 µl-Ansätze (präparativ) verwendet. Zur Aufreinigung der geschnittenen DNA wurden Phenolextraktionen durchgeführt. Verwendet wurde eine Phenollösung, die mit 50 mM Tris, pH 7,6 gesättigt, oder zusätzlich mit Chloroform und Isoamylalkohol im Verhältnis 25:24:1 gemischt war. Die Anzahl der Extraktionen richtete sich nach der Reinheit der DNA-Lösung. Es wurde immer eine Chloroformextraktion angeschlossen, um verbliebenes Phenol aus der wässrigen Phase zu entfernen. Alternativ wurden die Aufreinigungen mit dem „QIAquick PCR Purification Kit“ der Firma Qiagen (Hilden) oder mit „CONCERT^{TM} Rapid PCR Purification System“ der Firma GIBCO**BRL** Life Technologies (Karlsruhe) durchgeführt.

2.2.2 Fällung von DNA

Eine DNA-Fällung erfolgte normalerweise durch Zugabe von 1/10 Vol. 10x Dialysepuffer und 2,5 Vol. 100% Ethanol und 20 minütiger Inkubation bei –20°C bzw. –80°C. Anschließend wurde die DNA 15-30 min bei 12000g und 4°C sedimentiert. Das DNA-Pellet wurde anschließend mit 70%igem Ethanol gewaschen und in einer Vakuumzentrifuge (UNIVAPO 100H/ UNIJET II; LTF Labortechnik, Wasserburg/Bodensee) getrocknet.

2.2.3 Elektrophorese von DNA

Die elektrophoretische Auftrennung von DNA-Restriktionsfragmenten bzw. PCR-Produkten im Bereich von 50 bp bis 23 kb erfolgte nach Zugabe von 0,2 Vol. Stoppuffer auf vertikal, bzw. für hohen Probendurchsatz horizontal orientierten verschieden großen Agarosegelen, deren Konzentration -je nach gewünschtem Trennbereich- zwischen 0,7% und 2,4% (Sambrook and Russell, 2001, S. 5.6) betrug. Als Elektrophorespuffer dienten entweder 1x TBE- oder 1x E-Puffer. Verwendete Größenstandards waren *Hind III* geschnittene DNA des Phagen λ (aus *E. coli* WA312), *Eco RI* / *HindIII* geschnittene DNA des Plasmids pF (Bahr, 1995) oder *Alu I* restringierte pBR322-DNA. Die Gele wurden in Ethidiumbromidlösung (5 µg/ml in 1x E-Puffer) gefärbt, unter UV-Licht bei 312 nm auf einem UV-Transilluminator betrachtet und gegebenenfalls mit dem digitalen EDAS-System (Kodak) dokumentiert. Die Wiedergewinnung von DNA-Fragmenten aus Agarosegelen erfolgte entweder durch Elektroelution (McDonell et al.,

1977) oder durch Verwendung von Baukastensystemen der Firmen GIBCO**BRL** Life Technologies (Karlsruhe, CONCERT™ Rapid Gel Extraction System) und QIAGEN (Hilden, QIAquick™ Gel Extraction Kit). Es wurde jeweils nach Angaben der Hersteller verfahren.

2.2.4 Inhibition von *Pseudomonas*-Kontaminationen

Bei allen PAC-Klonen, die über das RZPD (Deutsches Ressourcenzentrum für Genomforschung GmbH, Berlin) bestellt wurden, konnte eine massive Kontamination mit *Pseudomonas* festgestellt werden. Eine saubere Plasmid-DNA-Präparation wurde so unmöglich. Kontaminierte Klone mussten zunächst einer Behandlung mit dem Antibiotikum Cefsulodin unterzogen werden. Cefsulodin inhibiert die Synthese bakterieller Zellwände von Pseudomonaden und führt zu einer Deformation und Lyse der Bakterien. Das Antibiotikum zeichnet sich durch eine Resistenz gegen verschiedene, durch Pseudomonaden sekretierten Beta-Lactamasen aus. Die kontaminierten Klone wurden zur Dekontamination auf Cefsulodin- (10 µg/ml) und dem korrespondierenden antibiotikumhaltigen Nährboden bei 37°C ÜN angezogen. Eine einzelne Kolonie wurde dann in 3 ml Flüssigmedium überführt, die das Cefsulodin (10 µg/ml) und das selektierende Antibiotikum enthielt und wieder bei 37°C ÜN inkubiert. Aus der Flüssigkultur wurden 30 µl entnommen und auf Cefsulodin-freien, aber antibiotikumhaltigen Nährboden ausplattiert. Nach Wachstum ÜN konnten die Klone für die DNA-Präparation verwendet werden (Kapitel 2.3.2)

2.3 DNA-Präparation

2.3.1 Präparation von Plasmid-DNA

Die schnelle Isolierung von Plasmid-DNA zur Identifikation rekombinanter Bakterien-Klone erfolgte nach der TELT-Methode von He et al. (1990). Für Arbeiten, bei denen sehr reine Plasmide in geringen Mengen gewonnen werden mussten, wurden sogenannte „Kits für die schnelle Plasmidisolierung“ verschiedener Hersteller („QIAPrep Spin Miniprep Kit“, QIAGEN (Hilden); CONCERT™ Rapid Plasmid Miniprep System“, GIBCO**BRL** Life Technologies (Karlsruhe); E.Z.N.A.® Plasmid Miniprep Kit I, peqlab (Erlangen)) verwendet. Für die Isolierung großer Mengen Plasmid-DNA wurde das QIAGEN Plasmid Midi/Maxi Kit (QIAGEN, Hilden) verwendet. Für spezielle Anwendungen (Kapitel 3.3) wurden Plasmide mit dem EndoFree® Plasmid Maxi Kit (QIAGEN, Hilden) isoliert. Alle Kits der Hersteller basieren auf dem Prinzip der alkalischen Lyse mit anschließender Ionenaustauschchromatographie. Die Methoden und Anwendungen richten sich nach den Angaben der Hersteller.

2.3.2 Präparation von Cosmid-, PAC- und P1-DNA

Für Sequenzierung, Restriktionsverdau sowie Subklonierung wurde Cosmid-/PAC-/P1-DNA verwendet, die durch eine modifizierte, auf dem Prinzip der alkalischen Lyse basierenden Methode (Birnboim und Doly, 1979) gewonnen wird. Die Modifikationen wurden von der Arbeitsgruppe um Trevor Hawkins (Whitehead Institute/MIT Center for Genome Research, http://www.genome.wi.mit.edu) entwickelt. Aus einer Einzelkolonie wurde eine Übernachtkultur bei 37°C und 325 rpm inkubiert. Hierbei sollte ein Volumen von 250 ml für Plasmide mit hoher Kopienzahl und 500 ml – 1 l für Plasmide mit geringer Kopienzahl gewählt werden. Die Kultur wurde nach 12-18 h in Zentrifugenröhrchen überführt und die Wirtszellen bei 5000 g sedimentiert, die erhaltenen Pellets vereinigt und in 20 ml Sol I resuspendiert, danach für 5 min bei RT inkubiert. Im Folgenden wurden 40 ml Sol II hinzugefügt, durch vorsichtiges Invertieren gemischt und die Lösung dann für 10 min auf Eis gestellt. Nach der angegebenen Zeit erfolgte eine zweimalige Zentrifugation bei 10.000 g („clearing spin"), wobei nach dem ersten Zentrifugationsschritt die Lösung ohne verbleibende Zellbestandteile in neue Zentrifugenröhrchen überführt wurde. Anschließend wurde die Lösung mit 45 ml Isopropanol versetzt, gemischt, die so gefällte DNA bei 5000 g sedimentiert und der Überstand verworfen. Das Pellet wurde nun in 9 ml TE (10 mM Tris, pH8; 50 mM EDTA) gelöst und 4,5 ml Kaliumacetat (7,5 mM) zugefügt. Diese Suspension wurde bei –80°C für 30 min eingefroren. Nach dem Auftauen (Lösung sollte noch viskös sein) erfolgte die sofortige Zentrifugation bei 5000 g für 15 min, der Überstand wurde dann mit 27 ml 100% Ethanol versetzt und nochmals für 10 min bei 5000 g zentrifugiert. Das so erhaltene Präzipitat wurde in 0,7 ml TE-50 Puffer (50 mM Tris, pH8; 50 mM EDTA) gelöst, mit 10 µl RNase (10 mg/ml) versetzt und 30-60 min bei 37°C in einem Mikrozentrifugen-Reaktionsgefäß inkubiert. Die Lösung wurde darauf folgend mit 0,7 ml Phenol versetzt, invertiert, bei 10.000 g zentrifugiert und die Phenolphase verworfen. Dieser Schritt wurde 2-5x wiederholt. In den folgenden zwei Schritten wurde das Phenol zunächst durch Phenol/Chloroform/Isoamyl-Alkohol und dann durch Chloroform ersetzt. Im Anschluss konnte dann ein finaler Fällungsschritt, wie unter Kapitel 2.2.2 beschrieben, durchgeführt werden. Das so erhaltene Pellet wurde mit 70% Ethanol gewaschen und in der Vakuumzentrifuge getrocknet. Die auf diese Weise hergestellte sehr reine DNA wurde in einer adäquaten Menge A. bidest (100-500 µl) resuspendiert. Die zu erzielenden Ausbeuten bewegen sich zwischen 150-500 µg DNA pro 500 ml eingesetzter Bakterienkultur.

2.3.3 DNA-Gewinnung aus Geweben

Bei der Isolierung genomischer DNA aus Geweben oder Mausembryonen musste vor der Aufreinigung der DNA zunächst das Gewebe mechanisch aufgeschlossen, die äußere Zellmembran lysiert und die Zellkerne isoliert werden. Hierzu wurde das Gewebe in Anwesenheit von Homogenisierungspuffer und 1/10 Vol. 10%iger Triton-X-100-Lösung in einem Potter-Gefäß bei 4°C homogenisiert. Anschließend wurde das Homo-

genat durch Gaze filtriert und die Zellkerne 15-30 min. bei 5000 x g sedimentiert. Das Zellkernsediment wurde erneut in Homogenisierungspuffer aufgenommen und sedimentiert. Dieser Waschschritt wurde drei- bis sechsmal wiederholt. Die Lyse der Zellkerne erfolgte bei 60°C für 2 Stunden in Homogenisierungspuffer, dem 1/10 Vol. 10%ige SDS-Lösung, 1/10 Vol. Dialysepuffer und eine Spatelspitze Proteinase K zugesetzt wurden. Anschließend wurde die DNA wie unter 2.2.1 beschrieben aufgereinigt. Für die Gewinnung hochmolekularer genomischer DNA wurde die wässrige Phase nach der Chloroformextraktion in ein großvolumiges Gefäß überführt und mit 1/10 Vol. Dialysepuffer versetzt. Die Lösung wurde dann auf Eis mit 2,5 Vol. eiskaltem Ethanol überschichtet. Die an der Grenzschicht von Ethanol und wässriger Lösung ausfallende DNA wurde mit einem sterilen Glasstab aufgewickelt, in 70% EtOH gewaschen und in einem adäquaten Volumen A. bidest abgespult. Alternativ kann nach der Zugabe des Ethanols invertiert und anschließend bei 5000 g für 60 min zentrifugiert werden. Auch hier schließt sich ein Waschschritt mit 70% EtOH an.

2.3.4 DNA-Isolierung aus Zellkulturen

Die Anzucht erfolgte in Zellkulturflaschen mit zellspezifischem Medium. Für die Aufarbeitung der Zellen wurde zunächst das Medium verworfen und die Kultur zwei- bis dreimal mit PBS gewaschen. Bei adhärenten Zelllinien wurde trypsinisiert. Das Trypsin wurde daraufhin durch mehrmaliges Waschen mit PBS entfernt und die Zellen wie unter 2.3.5 beschrieben mit SDS und Proteinase K aufgeschlossen und die DNA gegebenenfalls durch Phenol und Chloroform extrahiert.

2.3.5 DNA-Gewinnung aus Zellkulturen im 96'er Maßstab

Um die genomische DNA aus Zellkulturen zu gewinnen, die in 96'er Microwell-Platten (200 µl Volumen/Well) angezogen wurden, musste zunächst das Nährmedium durch schnelles Umdrehen der Platten verworfen werden. Die konfluent gewachsenen Zellen wurden dann zwei- bis viermal mit PBS gewaschen. Pro Well wurde anschließend 100 µl Proteinase-K Puffer zugegeben und ü.N. in einer feuchten Kammer bei 55°C inkubiert. Am nächsten Morgen wurde das Volumen des Lysates mit A. bidest auf 200 µl erhöht und in neue Microwell-Platten (2 ml Volumen/Well) überführt. Für die Fällung der DNA wurde das Volumen mit A. bidest auf 500 µl erhöht und 55 µl 3M Na-Acetat und 440 µl Isopropanol zugegeben. Die Platten wurden kurz geschwenkt und 30-60 min. bei 4°C inkubiert. Anschließend wurde 60 min. bei 8000 g zentrifugiert, der Überstand verworfen und das Pellet mit 1 ml 70% EtOH gewaschen. Die Pellets konnten dann in einer Vakuumzentrifuge getrocknet und in 35 µl TE bzw. A. bidest gelöst werden.

2.4 Klonierung in Plasmidvektoren

In der Vorliegenden Arbeit wurden die in den Abbildungen 2.1-2.3 gezeigten Plasmidvektoren für Klonierung, Subklonierung und zur Konstruktion des Targetingvektors verwendet.

2.4.1 Klonierung in pUC 18

Zur Sequenzierung und Amplifikation von DNA-Fragmenten wurde der Plasmidvektor pUC 18 (Yanisch-Perron et al., 1985; Abb. 2.1) verwendet. Dieser Vektor besitzt eine multiple Klonierungsschnittstelle und kodiert für ein lacZ'-Peptid. Durch α-Komplementation mit der vom Wirtsstamm kodierten, teilweise deletierten β-Galaktosidase kann sich ein funktionsfähiges Protein bilden (Langley et al., 1975), das eine blau/weiß-Selektion auf X-Gal-haltigen Platten ermöglicht. Für die Transformation von pUC 18 wurde in dieser Arbeit der *E. coli*-Stamm RRIΔM15 (Rüther, 1982) verwendet. Weiter kodiert der Vektor für das β-Lactamase-Gen, dessen Genprodukt dem Wirtsorganismus eine Resistenz gegen Ampicillin verleiht. Die zur multiplen Klonierungsschnittstelle inkompatiblen Enden von DNA-Fragmenten wurden mit dem Klenow-Enzym der DNA-Polymerase I aus *E. coli* aufgefüllt und in die *Sma* I- bzw. *Hind II*-Restriktionsschnittstelle ligiert. Die Ligation erfolgte mit T4-Ligase. Die Selektion auf positive Klone erfolgte entweder durch Klonanalyse (Kapitel 2.2.3) oder durch Koloniefilterhybridisierung (Kapitel 2.5.2).

2.4.2 Klonierung in pGEM®-T Easy

Neben der Klonierung in pUC 18 wurden PCR-Produkte auch mit Hilfe des pGEM®-T Easy Vector System I der Firma Promega (Mannheim) kloniert (Abb.2.2). Dieses Vektorsystem eignet sich unter anderem besonders für die Klonierung von PCR-Fragmenten, die durch Polymerasen synthetisiert wurden, die keine Korrekturfunktion aufweisen. Diese Polymerasen addieren matrizenunabhängig ein Nukleotid, meistens ein Adenin, an das terminale 3'-Ende des Amplifikats (Clark et al., 1988). Der Vektor pGEM®-T Easy liegt in linearer Form vor und weist seinerseits ein 3'-terminal addiertes Deoxythymidin auf. Diese 3'überhängenden Enden sind kompatibel und erhöhen die Klonierungseffiziens, eine Selbstligation des Vektors wird unterdrückt. Die Ligation wurde nach den Angaben des Herstellers durchgeführt. Eine Selektion auf positive Klone erfolgte wie unter 2.2.1 und 2.4.1 beschrieben.

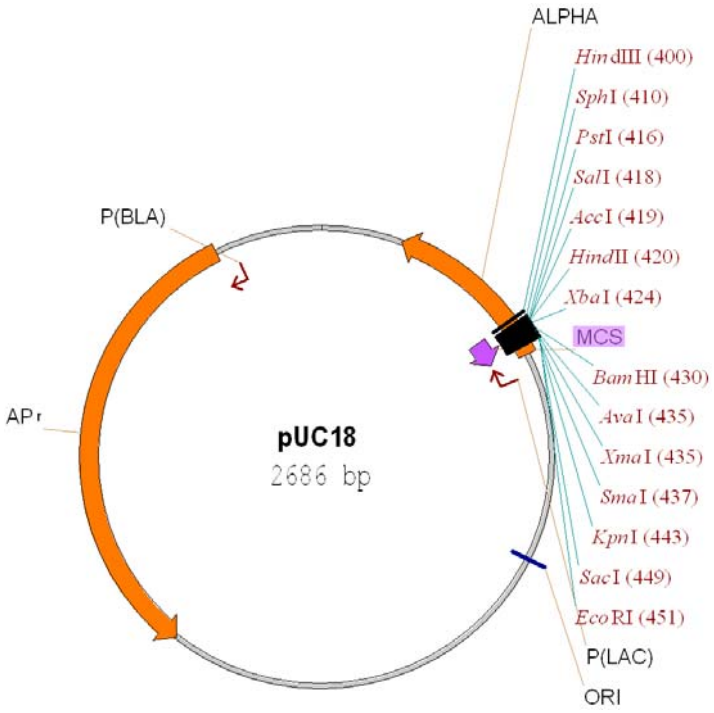
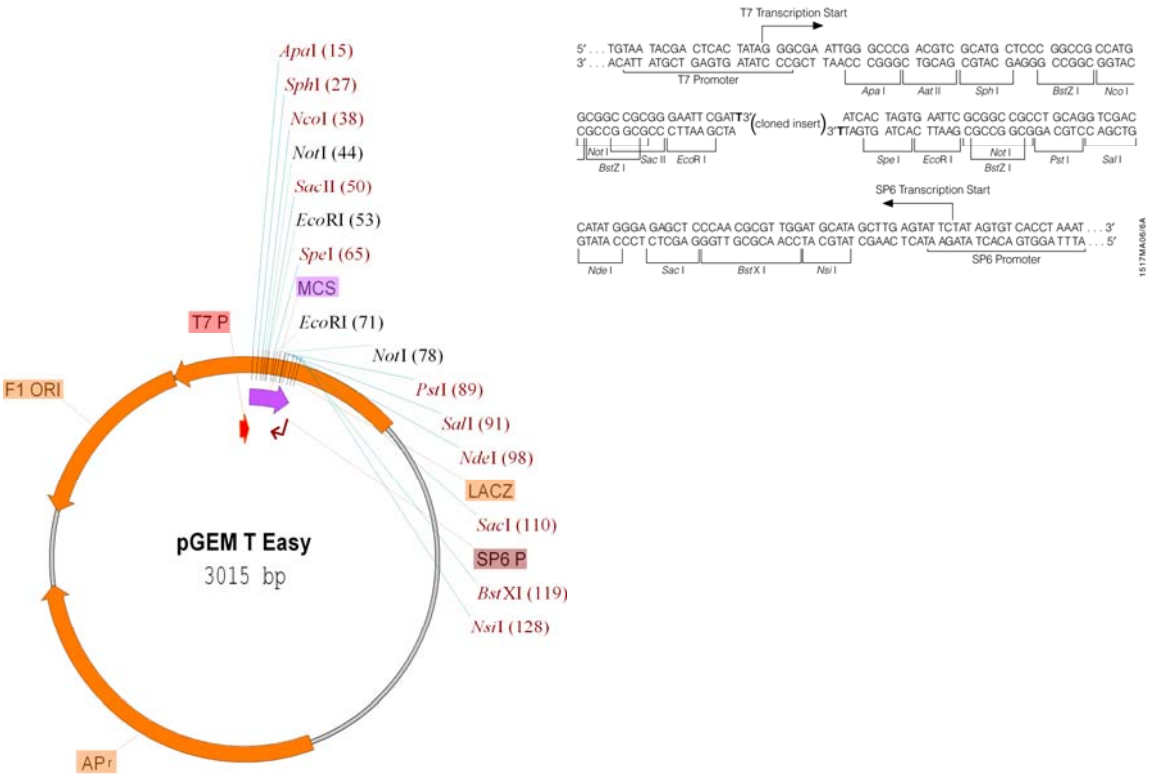


Abb. 2.1: Vektorkarte pUC 18

Abb. 2.2: Vektorkarte pGEM T Easy

Neben der Karte ist hier die Sequenz der multiplen Klonierungsschnittstelle inkl. des terminal addierten Deoxythymidin wiedergegeben.



2.4.3 Klonierung in den Oligo-Targeting-Vektor

Der Oligo-Targeting-Vektor (TV), (Abb. 2.3), basiert auf einem modifizierten pBR322-Derivat (Blessing, 1997). Der pBR322-Anteil ist ein *Nde I* - *Aat II* Fragment, das den „origin of replication“ (ORI) enthält und für das β-Lactamase-Gen kodiert (Blessing, 1993), so dass eine Amplifikation in *E. coli* RRIΔM15 und eine Selektion auf ampicillinhaltigen Medien möglich ist, jedoch keine blau/weiß-Selektion. Der Vektor besitzt weiter eine multiple Klonierungsschnittstelle 1, eine pMC1TKA^{+}-Kassette (Mansour, 1988) in entgegengerichteter transkriptioneller Orientierung zur angrenzenden pPGKNeo^{R}A^{+}-Kassette, eine multiple Klonierungsschnittstelle 2 sowie eine pGKTKA^{+} -Kassette in gleicher transkriptioneller Orientierung wie die davor liegende pPGKNeo^{R}A^{+}-Kassette (Abb. 2.3) (Rudnicki, 1992). Eine Klonierung erfolgte in zwei Schritten in die beiden unterschiedlichen multiplen Klonierungsschnittstellen, nach jeder Klonierung wurde eine Screening auf positive Klone im Hochdurchsatzverfahren durchgeführt (Kapitel 2.5.2, 3.3.2).

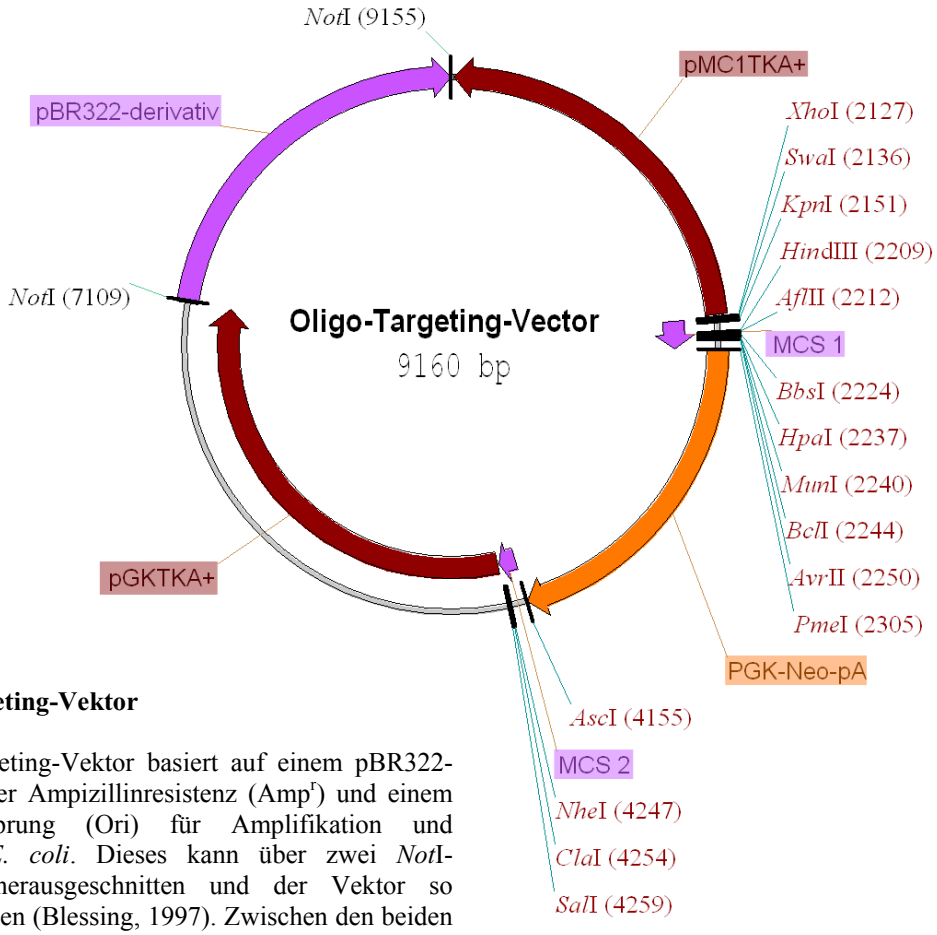


Abb. 2.3: Targeting-Vektor

Der Oligo-Targeting-Vektor basiert auf einem pBR322-Derivat mit einer Ampizillinresistenz (Amp^{r}) und einem Replikationsursprung (Ori) für Amplifikation und Selektion in *E. coli*. Dieses kann über zwei *Not*I-Schnittstellen herausgeschnitten und der Vektor so linearisiert werden (Blessing, 1997). Zwischen den beiden multiplen Klonierungsschnittstellen (MCS 1 und MCS 2) liegt eine pPGKNeo^{R}A^{+}-Kassette, die eine positive Selektion mit G418 in eukaryotischen Zellen ermöglicht. Zwei weitere Kassetten (pMC1TKA^{+} (Mansour, 1988), pGKTKA^{+} (Rudnicki, 1992)) in entgegengerichteter transkriptioneller Orientierung ermöglichen eine negative Selektion durch das Substrat Gancyclovir, das durch die kodierte Thymidin-Kinase in ein toxisches Triphosphat umgewandelt wird.

2.4.4 Herstellung einer „shotgun“ Klonbibliothek

Die Herstellung einer „shotgun“ Klonbibliothek aus PAC-Klonen in den Plasmidvektor pUC 18 (Yanisch-Perron et al., 1985) beruht auf einer Weiterentwicklung einer von Amid (2002) etablierten Methode. Bei der Präparation von PAC-Klonen kommt es immer ungewollt zu einer Isolierung geringer Mengen kontaminierender linearer genomischer DNA aus *E. coli*. Um diese Verunreinigungen zu entfernen, wurde die ATP-abhängige DNase Plasmid-Safe™ der Firma EPICENTRE TECHNOLOGIES (USA) verwendet. Hierbei handelt es sich um eine DNase, die spezifisch lineare Einzel- und Doppelstrang-DNA und zirkuläre Einzelstrang-DNA abbaut, den zirkulären Doppelstrang der PAC-DNA jedoch unbeeinträchtigt läßt. 50 µg PAC-DNA wurden mit 100 U Plasmid-Safe™ in 150 µl 1x Plasmid-Safe™-Reaktionspuffer und 1 mM ATP für 16 Stunden bei 37°C inkubiert. Nach Inaktivierung der Reaktion bei 70°C für 15 min. wurde der Ansatz mit TE-Puffer auf 2 ml aufgefüllt und die DNA mit Hilfe eines „DNA-Nebulizer“ (GATC, Konstanz) für 30-60 Sekunden bei einem Druck von 1 bar geschert. Bei nachfolgender Anwendung einer selektiven Sequenzierungsstrategie (Kapitel 3.2.2) kann gegebenenfalls auf die Behandlung mit Plasmid-Safe™ verzichtet werden. Nach dieser Behandlung lag der Größenbereich der daraus resultierenden DNA-Fragmente bei 0,5-4,5 kb. Anschließend wurde die DNA gefällt (Kapitel 2.2.2) und in 30 µl A. bidest gelöst. Die Größe der DNA-Fragmente und die Konzentration der DNA wurde durch gelelektrophoretische Auftrennung von $^1/_{10}$ Vol. überprüft. Das Auffüllen überhängender Enden erfolgte durch Inkubation von 10 µg der gescherten DNA mit 12 U Klenow-Enzym, 4 U T4-DNA-Polymerase in 60 µl 1x Kombinationspuffer und 40 µM dNTPs für 30 Minuten bei RT. Die DNA und ein geeigneter Marker wurden anschließend auf einem präparativen Gel gelelektrophoretisch aufgetrennt, der Markerbereich des Gels abgeschnitten und nur dieser in Ethidiumbromid-Färbelösung gefärbt. Die interessierenden Markerbanden wurden unter UV-Licht durchstochen und das Gel anschließend wieder zusammengesetzt. Der Bereich von 05-4,5 kb konnte nun „blind“, also ohne Beeinträchtigung der aufgetrennten DNA durch UV-Licht, ausgeschnitten und die enthaltenen DNA-Fragmente mit Hilfe der „QIAquick Gel Extraction Kit“ der Firma QIAGEN (Hilden) nach Angaben des Herstellers wiedergewonnen werden. Die Ligation von 100-300 ng der wiedergewonnenen DNA-Fragmente der Größenbereiche 0,5-1,5 kb, 1,5-2,5 kb und 2,5-4,5 kb mit 50 ng dephosphoriliertem, Sma I-geschnittenem pUC 18-Vektor (Pharmacia, Freiburg) erfolgte unter Verwendung von 10 U T4-DNA-Ligase (Boehringer, Mannheim) für 16 Stunden bei 4°C. Die Ligationsansätze wurden mit Hilfe des „QIAquick PCR Purification Kit“ (Qiagen, Hilden) bzw. „Concert™ Rapid PCR Purification System“ (GIBCO**BRL** Life Technologies, Karlsruhe) nach Angaben der Hersteller aufgereinigt. Jeweils $^1/_{50}$ Vol. der Ligationsansätze wurden für die Elektroporation von kompetenten *E. coli*-Zellen (Kapitel 2.10) des Stammes RRIΔM15 (Rüther, 1982) verwendet, $^1/_{10}$ Vol. eines Elektroporationsansatzes wurden jeweils auf X-Gal- und IPTG-haltigen Nährböden ausplattiert (Kapitel 2.4.1).

2.5 Hybridisierungstechniken

2.5.1 Southern-Hybridisierung

Der Transfer von DNA-Restriktionsfragmenten auf Nitrozellulose- bzw. Nylonmembranen erfolgte nach der Methode von Southern (1975) in der Anordnung nach Sambrook und Russell (2001, S.6.33-6.49). Nach erfolgtem Transfer wurde die DNA für 2 h bei 80°C (Nylon: 120°C) oder alternativ bzw. ergänzend durch zweimalige Bestrahlung mit UV-Licht im UV Stratalinker 2400 (Stratagene, La Jolla, California) mit jeweils 120.000 µ Joule fixiert. Anschließend erfolgte eine drei- bis vierstündige Präinkubation in 2x PM in 6x SSC (Denhardt, 1966) mit 1% SDS bei 60°C. Für die Hybridisierung mit einer radioaktiven Sonde wurde der Southern-Blot ü.N. in 1x PM in 3x SSC mit 0,5% SDS bei 58-65 °C inkubiert. Als „Carrier" diente Kalbsthymus-DNA (Endkonzentration bei 222 µg/ml). Durch mehrmaliges Waschen in 0,1-2x SSC bei 58-65°C für 4-6 Stunden wurden nicht gebundene Moleküle der Sonde entfernt. Anschließend wurde der Filter getrocknet und bei -80°C mit einem Röntgenfilm (Kodak „X-omat"(USA), Fuji RX (Japan)) und Verstärkerfolie (Dupont cronex High-Plus, Dr.Groos-Suprema) ü.N. oder bis zu 14 Tage exponiert. Alternativ wurden zur Detektion „Imaging Plates" der Firma Fuji (BAS-IP MP 2025 E) verwendet. Diese Platten sind mit einem kristallinen Material beschichtet, jedes Kristall (Körnung ~5 µm) besteht aus vielen Barium-Fluorobromid Verbindungen, die um eine zentrales bivalentes Europium angeordnet sind. Hochenergetische Elektronenstrahlung wirkt auf das zentrale Europium-Ion und überführt dieses in einen dreiwertigen Zustand zu Eu^{3+}. Das Elektron kann dann von freien Valenzen des Bariums aufgenommen werden und so einen metastabilen Zustand bilden. Jedes Kristall ist somit in der Lage eine große Anzahl Elektronen zu speichern. Wird das Material anschließend mit einer Energiequelle einer anderen Wellenlänge bestrahlt (hier Laserstrahlung), werden die Elektronen in Form von Luminiszens (photostimulierte Luminiszens) freigesetzt. Durch Verstärkung des ausgestrahlten Lichtes im Photomultiplier kann eine positionsabhängige Abbildung eingestrahlter Elektronen erstellt werden (Abb. 2.4). Der Vorteil dieses Systems ist seine hohe Empfindlichkeit und damit verbunden sehr kurze Expositionszeiten im Vergleich zu herkömmlichen Röntgenfilmen (Abb. 2.5). Die „Imaging Plates" können nach dem Auslesen in einem so genannten „Eraser" wieder gelöscht und bis zu 1000x wieder verwendet werden. Nachteilig wirkt sich bei diesem System, bedingt durch die Größe der Kristalle, die geringe Auflösung aus. In dieser Arbeit wurde ein Radioluminographie- und Floureszens-Laser-Scanner (BAS 1800 II) und Eraser der Firma Raytest (Straubenhardt, Deutschland) verwendet.

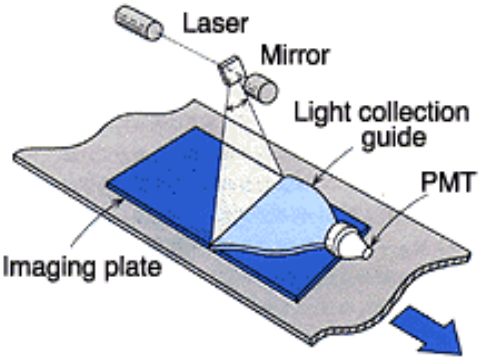


Abb. 2.4: Funktion und Detektion einer „Imaging Plate"

Eine „Imaging Plate" wird durch ein Lasersystem abgetastet. Entstehende photo-stimulierte Luminiszens wird durch ein Verstärkersystem geleitet und anschließend in elektrische Signale umgewandelt (Quelle: Fujifilm).

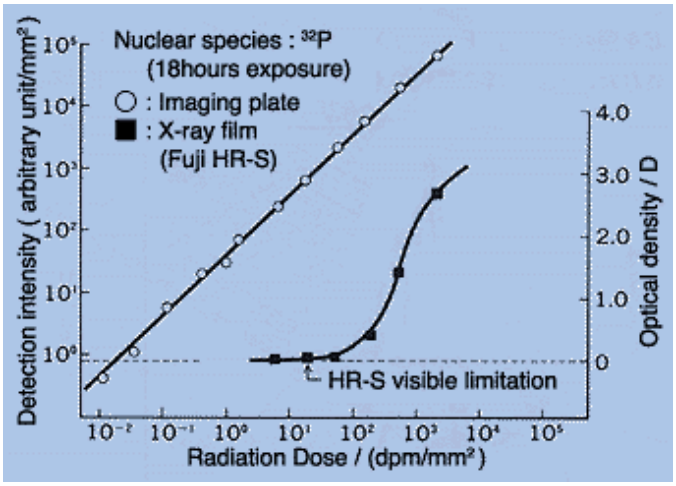


Abb. 2.5: Vergleich der Empfindlichkeit von „Imaging Plates“ und Röntgenfilmen (Fuji HR-S)

Quelle: Fujifilm

2.5.2 Koloniefilter-Hybridisierung

Für die integratspezifische Identifikation rekombinanter Plasmide in Bakterienklonen nach erfolgter Transformation wurde eine Koloniefilter-Hybridisierung durchgeführt. Bei dieser von Grunstein und Hogness (1975) beschriebenen Methode werden Bakterienklone in identischer Weise auf zwei gerasterte Nitrozellulosefilter überimpft, die vorher auf einem antibiotikahaltigen Nährboden ausgelegt wurden. Auch andere Selektionsverfahren, z.B. die Blau-Weiß Selektion werden durch die Nitrozellulose nicht behindert (siehe auch 2.4.1). Alternativ wurde eine modifizierte Stempelmethode verwendet, bei der ein angefeuchteter Nitrozellulosefilter auf einen unter selektiven Bedingungen dicht bewachsenen Bakterienrasen gelegt wurde. Ein Durchstechen des Nitrozellulosefilters bis in den Nährboden erleichtert die spätere Repositionierung auf der Originalplatte. Anschließend wurden die Filter auf den Selektivnährböden bei 37°C ü.N. bebrütet. Einer der Replikafilter wurde dann einer Koloniefilterbehandlung unterzogen. Hierbei werden die Bakterienzellen durch Verwendung von 0,5 M NaOH-Lösung und folgender Inkubation mit Proteinase K (1 mg/ml in 1x SSC) lysiert. Durch Pressen der Koloniefilter zwischen zwei Blatt Filterpapier können Bakterienreste entfernt werden. Die Plasmid-DNA kann anschließend durch Anbacken bei 80°C auf den Nitrozellulosefiltern fixiert werden. Die Hybridisierung und Detektion der Filter mit einer radioaktiv markierten Sonde kann dann wie unter 2.5.1 und 2.5.4 beschrieben erfolgen. Ein spezifisches Signal lässt sich so einem definierten Bakterienklon zuordnen. Bei der Verwendung der Stempelmethode sind weitere Schritte notwendig. Zunächst wird der Bereich, in dem ein spezifisches Signal beobachtet wurde, von der Originalplatte großzügig abgenommen und in einer entsprechenden Verdünnung erneut auf einem Nährboden ausplattiert. Die Bakterienklone werden dann auf gerasterte Nitrozellulosefilter übertragen und wie oben beschrieben weiter behandelt. Es können so leicht 3000-5000 Bakterienklone pro Platte aufgearbeitet werden. Diese Methode eignet sich besonders für ein Screening im Hochdurchsatzverfahren (Kapitel 3.3.2).

2.5.3 Hybridisierung an hoch-dichte Klonfiltern

Die Identifikation der in dieser Arbeit verwendeten Cosmid-, P1- und PAC-Klone erfolgte an hoch-dichte Klonfiltern des Deutschen Ressourcenzentrum für Genomforschung GmbH (RZPD, Ressourcen Zentrum / Primär Datenbank) in Berlin nach den Empfehlungen des Herstellers. Die Filter wurden zunächst für 2 h bei 65°C in einem Hybridisierungspuffer nach Church et al. (1994) präinkubiert. Die Hybridisierung mit einer radioaktiv markierten Sonde erfolgte für 16 h ebenfalls in diesem Puffer bei 65°C. Zur Identifizierung der Koordinaten der positiven Klone mußte durch ergänzende Zugabe von $^{1}/_{10}$ Vol. einer radioaktiv markierten Sonde des jeweiligen Vektors ein Hintergrundrauschen erzeugt werden, dass das Auftragsmuster der Klone auf den Klonfiltern erkennen ließ. Das Waschen der Filter erfolgte einmal für 1 min bei 25°C und zweimal für 30 min bei 65°C in einem Waschpuffer nach Church et al. (1994). Die getrockneten Filter wurden dann mit einem Röntgenfilm (Fuji RX, Japan) und Verstärkerfolie (Dupont cronex Hi-Plus, Dr. Groos -Suprema) bei -80°C autoradiographiert. Es wurden folgende DNA-Bibliotheken verwendet: Nr. 703 (ICRFP, MP1 Mouse P1 library), hergestellt von Fiona Francis aus C57/BL/ 6 Maus-DNA ligiert in pAd10SacBII. Nr. 121 (129/Ola), hergestellt von Burgtorf, Poch und Wiles aus 129/Ola Maus-DNA ligiert in Cosmid Lawrist 7. Nr. 711 (RPCI21 Mouse PAC), hergestellt von Kazutoyo Osoegawa und Pieter de Jong (Roswell Park Cancer Institut) aus 129/SvevTACfBr (Leber) Maus-DNA. Auf die genaue Bezeichnung der Klonbibliothek wird im Folgenden bei der Nennung der Klone verzichtet, die Kürzel P1- (Klonbibliothek ICRFP703), Cos- (MPMGc121) und PAC- (RPCI21) ersetzen diese.

2.5.4 Markierung von DNA durch „random primed oligo-labelling“

Die Markierung der DNA wurde radioaktiv mit dem „Random Primed DNA Labelling Kit“ der Firma Roche (Mannheim) nach der Methode von Feinberg und Vogelstein (1983) entsprechend den Angaben des Herstellers durchgeführt. Für genomische Southern-Analysen wurden ca. 1µg Sonden-DNA und 50-100 µCi α^{32}PdATP eingesetzt. Zur Entfernung nicht eingebauter radioaktiver Nukleotide wurde die DNA wahlweise gefällt (Kapitel 2.2.2) oder mit dem „QIAquick PCR Purification Kit“ (Qiagen, Hilden) aufgereinigt. Anschließend wurde die DNA mit Kalbsthymus-DNA als „carrier“ (Endkonzentration 25 µg/ml) vermischt und nach Denaturierung zur Hybridisierung eingesetzt.

2.6 Southern-Hybridisierung an DNA aus embryonalen Stammzelllinien

Das nachfolgende Protokoll ist eine Adaption von Kevin P. Foley, Dept. of Genetics, ZymoGenetics Inc., Seattle, WA 98102, basierend auf einem Protokoll von Phil Soriano's Lab, Fred Hutchinson Cancer Research Center, Seattle, WA (Soriano et al., 1991; Foley et al., 1998). Es wurden, basierend auf eigenen Kenntnissen, weitere Modifikationen vorgenommen. Es ist so möglich in 0,5-15 µg genomischer DNA einen spezifischen DNA-Abschnitt bei einer ÜN-Exposition nach erfolgter Southern-Hybridisierung nachzuweisen.

2.6.1 Restriktion genomischer DNA aus embryonalen Stammzellen

0,5-15 µg isolierte genomische DNA aus Stammzellkulturen wurden mit A. bidest auf ein Volumen von 99 µl verdünnt. Anschließend wurden 100 µl 2x BCS-Puffer zugegeben und für 2-3 Stunden bei 4°C inkubiert. Die Lösung sollte stündlich invertiert werden. Es folgte eine kurze Erwärmung auf 37°C und dann die Zugabe von 1-2 µl des Restriktionsenzyms und Inkubation für 6 Stunden bzw. ü.N. bei vorgeschriebener Restriktionstemperatur. Am nächsten Morgen konnte, je nach Restriktionseffiziens, 1-2 µl des Restriktionsenzyms nachgefüttert und weitere 3 Stunden inkubiert werden. Anschließend wurde das Volumen mit A. bidest auf 400 µl erhöht und 1x mit Phenol/Chloroform/ Isoamylalkohol, wie unter 2.2.1 beschrieben, extrahiert. Nach der Präzipitation (Kapitel 2.2.2) konnte die DNA in 20-30 µl A. bidest gelöst werden. Wahlweise kann auf die Phenol/Chloroform/Isoamylalkohol-Extraktion verzichtet werden.

2.6.2 Elektrophorese und Southern-Transfer

Die Auftrennung der DNA erfolgte auf vertikalen 0,8%igen Agarose-Gelen mit 1x E-Puffer als Laufpuffer (20,8 x 14,8 cm) bei 15 mA über 18 - 26 Stunden bei 4°C. Als DNA-Stoppuffer wurde Ficoll-Ladepuffer verwendet, dieser erzeugt wesentlich schärfere Banden als ein Standard-Ladepuffer. Der Transfer der DNA auf die Nitrozellulose- bzw. Nylonmembran erfolgte wie unter 2.5.1 beschrieben.

2.6.3 Hybridisierung

Die Präinkubation erfolgte in 2x PM in 6x SSC (Denhardt, 1966) mit 1% SDS bei 60°C. Für die Hybridisierung mit einer radioaktiven Sonde wurde der Southern-Blot ü.N. in FBI-Puffer (entwickelt in den Crime Labs, Federal Bureau of Investigation, USA) unter

stringenten Bedingungen (63°C) inkubiert. Durch mehrmaliges Waschen in 0.1 x SSC mit 0,1% SDS über 4-8 h wurden die nicht gebundenen Moleküle der Sonde entfernt. Anschließend wurde der Filter getrocknet und mit einer „Imaging Plate“ ü.N. exponiert. Am nächsten Tag wurde die „Imaging Plate“ entwickelt (siehe 2.5.1) und durch einen Röntgenfilm und Verstärkerfolie ersetzt und für weitere 7-14 Tage bei -80°C exponiert.

2.7 DNA-Sequenzierung

Die für die Sequenzierung von DNA eingesetzte Methode beruht auf dem Prinzip der Kettenabbruchsynthese (Sanger et al., 1977; modifiziert von Chen und Seeburg, 1985). Hierbei werden vier mit unterschiedlichen Floureszensfarbstoffen gekoppelte Didesoxinukleotide (Dye-Terminator™) in Kombination mit dNTPs verwendet. In der vorliegenden Arbeit wurden sowohl BigDye-Premix Version 3.1™ als auch BigDye-Premix Version 2.0™ verwendet, die entsprechenden Protokolle zur DNA-Sequenzierung unterscheiden sich zwischen den Versionen und können Tab. 2.1 und Tab 2.2 entnommen werden.

DNA-Sequenzierung BigDye™ Version 2.0	**DNA-Sequenzierung BigDye™ Version 3.1**
x µl Matrizen DNA	x µl Matrizen DNA
+1 µl Primer (10 pmol/µl)	+1 µl Primer (10 pmol/µl)
+4 µl BigDye-Premix Version 2.0	+2 µl BigDye-Premix Version 3.1
A. bidest ad 20 µl	+2 µl Sequencing buffer v3.1
	A. bidest ad 10 µl

Tab 2.1: Standardprotokoll für die DNA-Sequenzierung mit BigDye™ Version 2.0 und BigDye™ Version 3.1

DNA	**BigDye™ Version 2.0**	**BigDye™ Version 3.1**
Plasmid-DNA	500 ng – 1 µg	170 ng - 350 ng
PCR-Produkte:		
1000 bp	200 ng	70 ng
500 bp	100 ng	35 ng
200 bp	40 ng	15 ng
Cosmid/BACs/PACs	1-3 µg	350 ng – 1 µg

Tab 2.2: Übersicht der im Standardprotokoll für die DNA-Sequenzierung eingesetzten Template-DNA

Für die Sequenzierung von Plasmiden wurde DNA verwendet, die mit Hilfe der „Kits für die schnelle Plasmidisolierung“ (siehe 2.3.1) gewonnen wurde. Sequenzierungsreaktionen an Cosmid-, PAC- und BAC-Klonen erfolgte an DNA, die nach dem „Whitehead“-Protokoll (Kapitel 2.3.2) isoliert wurde. PCR-Produkte und DNA-

Fragmente aus Agarosegelen wurden wie unter 2.2.3 und 2.8.2 beschrieben aufgereinigt. Für die Sequenzierung hat sich das „2-step-cycle-sequencing-protocol“ (2step55) mit den in Tab. 2.3 aufgeführten Parametern bewährt. Der Ansatz wurde dann mit 1/10 Vol. 3M Natriumacetat pH 6.0 und 2 Vol. Ethanol gefällt, gewaschen und getrocknet. Alternativ erfolgte eine Aufreinigung über Sephadex G50 Fine™. Für die Elektrophorese der Proben wurden diese in Formamid gelöst, wenn die Detektion auf einem ABI 377 (Perkin-Elmer) erfolgte. Bei Elektrophorese auf einem ABI 3730 DNA ANALYSER (Applied Biosystems/Hitachi) erfolgte zunächst ein eine SDS-Behandlung. Hierbei wurde der Sequenzierungsreaktion 1µl einer 2,2% SDS-Lösung zugesetzt und für 2 min auf 94°C erhitzt. Es folgte eine Aufreinigung über Sephadex G50 Fine™ und Elution mit 10 µl A. bidest. Die Rohdaten, die von den Sequenzierautomaten erzeugt wurden, konnten anschließend mit der mitgelieferten Analysis-Software automatisch ausgewertet und durch die Programme Sequencher 4.1 (Gene Codes Corporation, Ann Arbor USA), DNASTAR (DNASTAR, Inc., USA) und BLAST (Altschul et al., 1990) weiterverarbeitet werden.

Initialdenaturierung:	96°C 2 min
Denaturierung:	96°C 10 sec.
„Anealing“ und Elongation:	55°C 4 min
	30-40 Zyklen

Tab 2.3: Parameter des 2step55 Protokolls

2.8 Polymerasekettenreaktion (PCR)

2.8.1 PCR-Technik

Die Amplifikation von DNA erfolgte entweder auf einem Thermal Cycler PHC-3 der Firma Techne (England), einem Autogene II-Gerät (GRANT Instruments, England), einem PTC-200 DNA-Engine der Firma MJ-Research Inc. (USA) oder einem PCR-Sprint Modell der Firma Hybaid (Heidelberg).

2.8.2 Standard-PCR

Für DNA-Amplifikate bis 2,5 kb wurde die PCR (Saiki et al., 1988) mit 50-500 ng genomischer, Cosmid-, PAC/BAC-DNA oder 1-10 ng Plasmid-DNA durchgeführt. Neben der angegebenen Menge DNA wurden dem Reaktionsansatz noch 100-300 ng Primer, dNTP’s zu einer Endkonzentration von 0,2 mM, 1x Taq-Puffer (Endkonzentration) und 2,5 U Taq-Polymerase zugefügt. Im Standardfall wurde die PCR unter den in Tabelle

2.4 aufgeführten Bedingungen durchgeführt. Für optimale Ergebnisse wurde dieses Protokoll in Bezug auf DNA-Menge, $MgCl_2$-Konzentration und Temperaturen variiert. Die verwendete Taq-Polymerase wurde nach Pluthero (1993) und BioTechniques (19:780-784, 1995) aus *E. coli* isoliert.

Denaturierung:	10 sec. – 1min	94°C
„Annealing“:	30 sec.	55-58°C
Elongation:	1 – 3 min	72°C
		33 Zyklen

Tab 2.4: 3-Schritt Standard PCR Bedingungen

Die Aufreinigung der Amplifikate erfolgte entweder durch eine Fällung (siehe 2.2.2.) oder mittels „QIAquick PCR Purification Kit“ (Qiagen, Hilden) bzw. „Concert™ Rapid PCR Purification System“ (GIBCO**BRL** Life Technologies, Karlsruhe). Bei Verunreinigung mit unspezifischen PCR-Produkten wurden die Amplifikate auf einem präparativen Gel aufgetrennt, die gewünschten DNA-Banden ausgeschnitten und mittels Elektroelution wiedergewonnen. Alternativ erfolgte die Wiedergewinnung mit dem „QIAquick Gel Extraction Kit“ (Qiagen, Hilden) bzw. dem „Concert™ Rapid Gel Extraction System“ (GIBCO**BRL** Life Technologies, Karlsruhe).

2.8.3 „X-large” und „High Fidelity“ PCR

Zur Amplifikation von DNA-Fragmenten ab 2,5 kb wurde ein Gemisch verschiedener thermostabiler Polymerasen verwendet. In diesen Gemischen werden DNA-Polymerasen mit und ohne Korrekturlesefunktion kombiniert. Dadurch wird die Amplifikation längerer Fragmente überhaupt erst ermöglicht und die Replikationsgenauigkeit erhöht. Verwendet wurde in dieser Arbeit das „Expand Long Template PCR System“ der Firma Roche (Mannheim) und Taq/Pwo Mix (Proof-Mix) der Hybaid-AG (Heidelberg). Mit diesen Systemen sind Amplifikate von 0,5 kb bis 30 kb möglich. Die Bedingungen für Amplifikate zwischen 5 und 10 kb können Tabelle 2.5 entnommen werden. Für besonders hohe Ausbeuten und Genauigkeit von DNA-Amplifikaten (z.B. Amplifikation nach erfolgter Erststrangsynthese, Kapitel 2.8.4) zwischen 0,5 – 2,5 kb wurde ein Gemisch aus Taq- und ProofStart™-Polymerase der Firma QIAGEN (Hilden) in einem 12:1 Verhältnis hergestellt. Teilweise wurde ProofStart™-Polymerase auch mit DNA-Polymerasen anderer Hersteller und selbst isolierter Taq-Polymerase (siehe 2.8.2) im Verhältnis 12:1 vermischt. In jedem Fall wurden gute Ergebnisse erzielt. Die Bedingungen für eine Amplifikation können Tabelle 2.6 entnommen werden. Die Zusammensetzung der Reaktionskomponenten basieren auf den jeweiligen Angaben des Herstel-

lers, jedoch wurden zahlreiche Modifikationen der Komponenten und Bedingungen realisiert.

Initialdenaturierung:	2 min	94°C	
„Annealing“:	30 sec.		55-65°C
Elongation:	5 – 10 min		68°C
Denaturierung:	10 sec. – 1min		94°C
			10 Zyklen
„Annealing“:	30 sec.		55-65°C
Elongation:	5– 10 min (+20 sec. /Zyklus)		68°C
Denaturierung:	10 sec. – 1min		94°C
			20-25 Zyklen

Tab 2.5: Standardbedingungen für eine „X-large“ PCR

Initialdenaturierung:	2 min	95°C	
„Annealing“:	30 sec.		56-58°C
Elongation:	1 – 3 min		68°C
Denaturierung:	10 sec.		94°C
			40-45 Zyklen
Finale Elongation:	6 min.	56-58°C	

Tab 2.6: Standardbedingungen einer „High Fidelity“ PCR

2.8.4 Reverse Transkriptase-PCR (RT-PCR)

Um die Expression des *Dmxl*1-Gens in der Maus zu untersuchen, mußte zunächst Gesamt-RNA und nachfolgend gegebenenfalls Poly-A^+ RNA aus verschiedenen Geweben isoliert werden (Kapitel 2.9). Für eine Erststrangsynthese wurden 500 ng - 5 µg dieser RNA eingesetzt. Verwendet wurden die Reversen Transkriptasen M-MLV Reverse Transcriptase RNase H Minus Point Mutant (Promega GmbH, Mannheim), ImProm-II™ Reverse Transcriptase (Promega GmbH, Mannheim), Omniscript™ Reverse Transcriptase (Qiagen, Hilden)und SuperScript™ III (Invitrogen, GmbH, Karlsruhe). Die Synthesen erfolgten jeweils nach den Angaben der Hersteller, es wurden für die Erststrangsynthese immer genspezifische Primer verwendet (Die Positionen der Primer sind in den Sequencher™-Projekten im beigefügten Datenträger zu finden). Um eine hohe Spezifität der Synthesereaktion zu gewährleisten, wurden die „Anealing“-Temperaturen jeweils auf 55°C erhöht. Typischerweise wurden 5 µg Gesamt-RNA zusammen mit 20 pmol genspezifischem Primer und RNase freiem A. bidest für 5-10

min. bei 65-70°C inkubiert und sofort auf Eis gekühlt. Nach Zugabe von Erststrang-Puffer, 1 µl 0,1 M DTT, RNase Inhibitor, 1µl 10 mM dNTP Mix und 200 Units Reverse Transkriptase auf ein Endvolumen von 20-25 µl wurde für 60 min. bei 55°C inkubiert. Danach erfolgte eine Inaktivierung der Komponenten bei 70°C für 10 min und anschließendem Verdau komplementärer RNA durch RNase H bei 37°C für 20 min. Nachfolgend wurden 5-10 µl ungereinigter Synthesereaktion für eine X-large PCR, wie unter 2.8.2 beschrieben, eingesetzt. In den meisten Fällen war es notwendig an den Produkten der X-large PCR ein bis zwei weitere Amplifikationsschritte mit weiter innen liegenden Primerpaaren („Nested-PCR“) durchzuführen. Wenn weiter unspezifische Amplifikate beobachtet wurden, erfolgte eine Klonierung der gewünschten DNA-Fragmente mit Hilfe des pGEM®-T Easy Vector System I der Firma Promega (Mannheim) (Kapitel 2.4.2).

2.9 RNA-Präparationen

2.9.1 RNA-Isolierung mit Guanidinthiocyanat

Die Isolierung von RNA wurde nach einer veränderten Methode von Chomczynski und Sacchi (1987) durchgeführt. Dabei wurden 100 mg Gewebe in 1 ml Guanidinlösung homogenisiert, mit 0,1 Vol. 2M Natriumacetat pH4 versetzt und gemischt. Nach Zugabe von 1 Vol. wassergesättigtem Phenol und 0,2 Vol. Chloroform erfolgte eine viertelstündige Inkubation auf Eis. Anschließend wurde das Gemisch 30 Minuten mit 14000 g bei 4°C zentrifugiert. Die obere Phase wurde dann mit 1 Vol. Isopropanol bei –20 °C gefällt und wieder bei 14000 g und 4°C für 30 Minuten zentrifugiert. Das resultierende Pellet wurde in 0,3 Vol. Guanidinlösung aufgenommen und nochmals mit Isopropanol gefällt. Danach wurde das Pellet in 75% Ethanol resuspendiert, für 10 Minuten bei Raumtemperatur inkubiert und abzentrifugiert. Nach Trocknung in einer Vakuumzentrifuge wurde die RNA in 100-200 µl A. bidest aufgenommen.

2.9.2 RNA Isolierung mit Trizol Reagent

Die Trizol Reagent-Methode basiert auf den kommerziell erhältlichen Produkten TRIzol™ von GIBCO**BRL** Life Technologies (Karlsruhe) und TRI REAGENT™ der Firma SIGMA (USA). Das in dieser Arbeit verwendete Trizol Reagent wurde selbst hergestellt und entspricht den kommerziellen Produkten näherungsweise. Das verwendete Protokoll ist eine Kombination der Protokolle von SIGMA und GIBCO**BRL** Life Technologies und hat den Vorteil, das sehr reine RNA kostengünstig in weniger als zwei Stunden isoliert werden kann. Für die Isolierung der RNA wurden pro 50-100 mg Gewebe 1ml Trizol zugegeben, homogenisiert und für 5 Minuten bei Raumtemperatur inkubiert. Nach Zugabe von 0,2 Vol. Chloroform pro Vol. Trizol Reagent erfolgte eine weitere 15 minütige Inkubation bei Raumtemperatur mit anschließender Zentrifugation bei 5000g für 30 Minuten. Danach wurde die wässrige

Phase mit 1/2 Vol. Isopropanol, 1/2 Vol. 1,2 M Natriumchlorid und 1/2 Vol. 0,8 M Natriumcitrat versetzt, 5 Minuten bei Raumtemperatur inkubiert, 30 Minuten bei 4500g zentrifugiert und das Pellet in 2 ml 70% Ethanol resuspendiert. Nach Zentrifugation bei 14000g für 10 Minuten folgten zwei weitere Waschschritte mit 2 ml 70% Ethanol. Das Pellet wurde dann in einer Vakuumzentrifuge getrocknet und in 100 µl A. bidest gelöst, erneut für 5 Minuten bei 14000g zentrifugiert und der Überstand in ein neues 1,5 ml Reaktionsgefäß überführt.

2.9.3 Isolierung von polyA$^+$ -RNA

Zur Aufreinigung von polyA$^+$ -RNA aus 250-1000 µg Gesamt-RNA wurde der „mRNA Isolation Kit“ (Boehringer, Mannheim) und der „Oligotex mRNA Kit“ (Qiagen, Hilden) entsprechend den Empfehlungen der Hersteller verwendet.

2.10 Transformation von Bakterien

Die Transformation von Plasmid-DNA in kompetente Wirtszellen erfolgte durch Elektroporation. Um Wirtszellen kompetent für die Elektroporation zu machen, wurde eine 5 ml Kultur ü.N. angezogen und am nächsten Tag in 500 ml frisches Medium überführt. Bei Erreichen einer OD_{600} von 0,5-0,7 wurde die Kultur bei 4°C pelletiert und dreimal mit 100 ml eiskaltem A.bidest gewaschen. Der letzte Waschschritt erfolgte in 10%igem Glyzerin, anschließend wurden die Zellen in 1 Vol. dieser Lösung resuspendiert, in 50 µl-Aliquots in flüssigem Stickstoff schockgefrohren und bei –80°C gelagert. Die Transformation erfolgte in 2 mm Küvetten mit 2000-2500 Volt auf Puls-Generatoren der Bezeichnung EPI 2500 (Dr. Fischer, Heidelberg) und EasyjecT Prima (EQUIBIO, USA). Sofort nach dem Hochvolt-Puls wurde 1 ml gekühltes L-Medium zugesetzt und 50-300 µl ausplattiert. Positive Klone wurden durch Plasmidisolierung nach He et al. (1990) und anschließendem Restriktionsverdau identifiziert (Kapitel 2.2.1). 200 µl jeder positiven Kultur wurden in 2x L-Medium/Glycerin im Verhältnis 1:1 bei –80°C bzw. –20°C als Dauerkultur gelagert.

2.11 Knock-Out Technik bei Mäusen

Die Transformation und Selektion embryonaler Stammzellen des Typs 129-R1, Passage16, sowie damit zusammenhängende weitere Arbeiten erfolgten in Kooperation mit Univ.-Prof. Dr. rer. nat. Manfred Blessing und Mitarbeiter, I. Med. Klinik und Poliklinik, Abteilung Pathophysiologie des Klinikums der Johannes Gutenberg-Universität. Die Isolierung und Injektion von Blastozysten, der Embryotransfer sowie Züchtung und Haltung daraus resultierender Tiere erfolgte in Kooperation mit Dr. Reifenberg (Leiter der Zentralen Versuchstiereinrichtung (ZVTE)). Diese Arbeiten wurden im Sicherheitsbereich der Zentralen Versuchstiereinrichtung des

Naturwissenschaftlich-Medizinischen Forschungszentrums der Johannes Gutenberg-Universität durchgeführt

2.11.1 Routinekultur von ES-Zellen

Die Isolierung und gezielte Manipulation von pluripotenten embryonalen Stammzellen (ES-Zellen) ist ein entscheidener Fortschritt im Rahmen der entwicklungsbiologischen Forschung bei Säugetieren. Die ersten ES-Zellen wurden von Evans und Kaufmann (1981) und Martin (1981) aus der inneren Zellmasse von Blastozysten der 129/Sv-Maus isoliert. Bald danach wurde von Bradley et al. (1984) gezeigt, daß sich diese Zellen an der Entwicklung sämtlicher Gewebe des resultierenden Embryos, einschließlich der Keimdrüsen, beteiligen, wenn sie in die Blastocyste injiziert werden. Die erste genetische Manipulation an ES-Zellen wurde von Robertson et al. (1986) und Gossler et al. (1986) durchgeführt. Diese Gruppen konnten zeigen, daß Manipulationen im Genom von ES-Zellen von diesen durch die Keimbahn transmittiert werden. Die ersten erfolgreichen homologen Rekombinationen in ES-Zellen wurden fast gleichzeitig von Thomas und Capecchi (1987) sowie Doetschman et al. (1987) publiziert. In der vorliegenden Arbeit wurden R1-ES-Zellen (Nagy et al., 1993) zusammen mit subletal bestrahlten embryonalen „Feederzellen“ (Fibroblasten, EmFi) auf 0,2 % Gelatine beschichteten Kulturschalen ausgesät. Das Kulturmedium wurde aufgrund der starken Ansäuerung durch die ES-Zellen täglich gewechselt. Eine Passagierung wurde je nach Dichte der Zellen alle zwei bis drei Tage durchgeführt und die ES-Zellen in Trypsin/EDTA-Puffer 1:3 bis 1:8 geteilt. Vor der erneuten Plattierung wurde die Trypsinlösung durch Zentrifugation mit 270g für 5 min bei RT entfernt. Bei jeder Passage wurden Stocks der ES-Zellen in flüssigem Stickstoff kryokonserviert.

2.11.2 Elektroporation und Selektion von ES-Zellen

Die ES-Zellen wurden durch Elektroporation von 150 µg mit *Not* I linearisiertem Vektor (Nagy et al., 1993) transformiert. Anschließend wurde eine positiv und negativ Selektion mit G418 (300 µg/ml Endkonzentration) und Gancyclovir (2 µM Endkonzentration) an den Zellen durchgeführt. Von den resistenten Klonen wurden Replikaplatten für die DNA-Präparation (Kapitel 2.11.3) angefertigt und die Originalplatten bei -20°C gelagert.

2.11.3 Isolierung von ES-Zellklonen und Kryokonservierung

Resistente Klone wurden nach einer G418-Selektion über 8 bis 9 Tage isoliert. Dazu wurden die Kulturen einmal in PBS-Puffer gewaschen und mit 10 ml PBS-Puffer überschichtet. Die Klone wurden mechanisch mit Hilfe einer sterilen Pipettenspitze von den embryonalen Feederzellen getrennt und bei 37°C in 1 x Trypsin/EDTA in 96-Loch-Platten dissoziiert. Der enzymatische Verdau wurde nach 5 min durch Zugabe von 50 µl

ES-Zellmedium gestoppt. Die vereinzelten Klone wurden auf 24-Loch-Platten mit Feederzellen und frischem ES-Zellmedium übertragen und bei 37°C kultiviert. Nach zwei bis drei Tagen waren die Einzelklone zur Konfluenz gewachsen und wurden erneut mit Trypsin behandelt. Die eine Hälfte der Zellen wurde erneut auf „Feederzellen" ausgesäht (master plate), während die andere Hälfte der Zellen auf einer Gelatine-behandelten Platte (duplicate plate) inkubiert wurde. Die auf Feederzellen wachsenden ES-Zellen wurden nach Erreichen der Konfluenz mit 100 µl Trypsin/EDTA bei 37°C für 5 min dissoziiert, in 1 ml Kryogefäße (Nunc) überführt und nach Zugabe von 100 µl 2 x Gefriermedium gemischt und bei -80°C eingefroren. Die für die DNA-Isolierung vorgesehenen Zellen wurden ebenfalls bis zur Konfluenz kultiviert und anschließend die DNA-Extraktion (Kapitel 2.3.5) durchgeführt. Homolog rekombinierte ES-Zellen wurden nach dem Auftauen bei 37°C in 10 ml kaltes ES-Zellmedium gegeben, bei 270 x g für 5 min abzentrifugiert und auf Feederzellen in 24-Loch-Platten kultiviert. Nach Erreichen der Konfluenz wurden die Zellen trypsinisiert und in 3-4 Aliquots eingefroren.

2.11.4 Isolierung von Blastozysten

Um möglichst viele Mausembryonen des Stammes C57/BL/6N zu erhalten, wurde eine Superovulation bei Spenderweibchen dieses Stammes durch Hormongabe (intraperitoneal) induziert. Durch diese Behandlung kommen die Weibchen in den Ovarialzyklus und es erfolgt eine vermehrte Abgabe von Eizellen in den Eileiter. Superovulierte Weibchen wurden anschließend mit einem Männchen verpaart. Eine erfolgreiche Verpaarung läßt sich durch die Anwesenheit des Vaginalpfropfes (VP) bei den Weibchen nachweisen. Der Zeitpunkt des Auftretens des Vaginalpfropfes wird definitionsgemäß als Tag 0,5 der Embryonalentwicklung berechnet. Für die Injektion rekombinanter ES-Zellen wurden Mausembryonen vom Tag 3,5 nach der Befruchtung benötigt. Hierzu wurden VP-positive Weibchen am Tag 3,5 abgetötet, die Ovidukte präpariert.

2.11.5 Injektion von ES-Zellen in Blastozysten

In die isolierten Blastozysten wurden mittels einer Hohlnadel ausgewählte ES-Zellen injiziert. Hierbei ist besonders darauf zu achten, dass die Injektion der embryonalen Stammzellen durch das Trophoectoderm in das Blastocoel erfolgt und dabei der Epiblast nicht verletzt wird (Abb. 2.6).

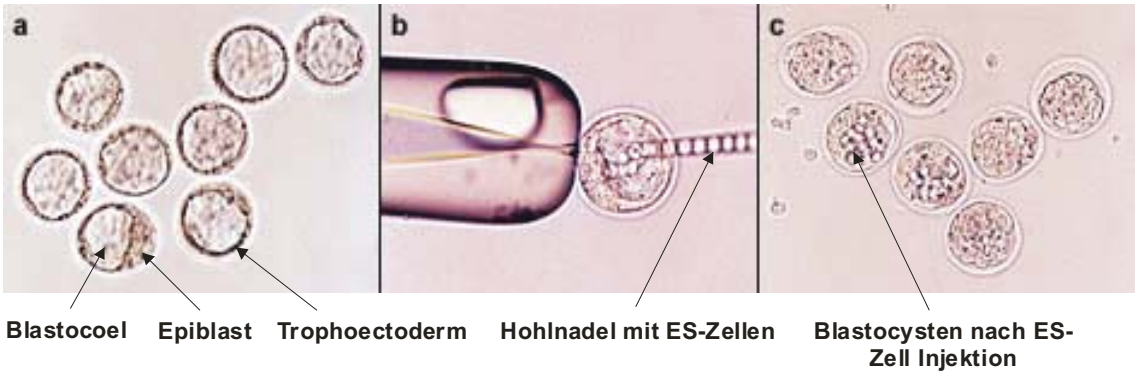


Abb 2.6: Beispiel einer Blastozysteninjektion

Gezeigt ist der Vorgang einer Blastozysteninjektion. Die Injektion erfolgt an isolierten Blastozysten (a) durch das Blastocoel in den Epiblasten (b). Hierzu wird in eine Hohlnadel zunächst eine Luftblase eingezogen, darauf folgen ca. 15 ES-Zellen und eine weitere Luftblase. Die Luftblasen dienen der Orientierung. Die ES-Zellen werden dann in das Blastocoel injiziert (b). Die Injizierten ES-Zellen können in den Blastozysten deutlich erkannt werden (c), sie integrieren später in den Epiblasten (die gezeigten Abbildungen stammen von MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, UK).

2.11.6 Embryo-Transfer in pseudoträchtige Mäuse

Pseudoträchtige Weibchen des Stammes NMRI wurden durch die Kopulation mit vasektomierten NMRI-Männchen gewonnen. Nach der Kopulation besitzen diese Weibchen einen Vaginalpfropf. Der Hormonhaushalt dieser Weibchen entspricht dem eines trächtigen Tieres, obwohl sich keine befruchteten Eizellen in den Reproduktionstrakten befinden. Für den Embryonen-Retransfer wurden Weibchen vom Tag 3,5 nach der Kopulation verwendet. Nach Betäubung der Maus wurde auf einer Seite der Rückenhaut ein 1 cm langer Schnitt auf Höhe der Niere gesetzt. Über dem rötlich durchscheinenden Ovar wurde ein kleines Loch in die Unterhaut geschnitten und das Ovar mit einer Pinzette am anliegenden Fettgewebe herausgezogen. Mit Hilfe einer feinen Kanüle wurde in das nun freiliegende Uterushorn ein kleines Loch gebohrt (Abb. 2.7). Die Blastozysten wurden mit der fein ausgezogenen Mundpipette aus den Vertiefungen der Kulturschale herausgespült und mit möglichst wenig Medium in die Spitze der Pipette aufgenommen. Die Pipette wurde dann durch das Loch in den Uterus eingeführt und vorsichtigt ausgeblasen. Durchschnittlich wurden etwa 10-12 Blastozysten pro Maus übertragen. Nach erfolgtem Retransfer wurden die Wundränder aufeinander gelegt und mit einer Klammer verschlossen. Nach Desinfektion der Wunde wurde die Maus bis zum Erwachen auf eine Wärmeplatte gelegt.

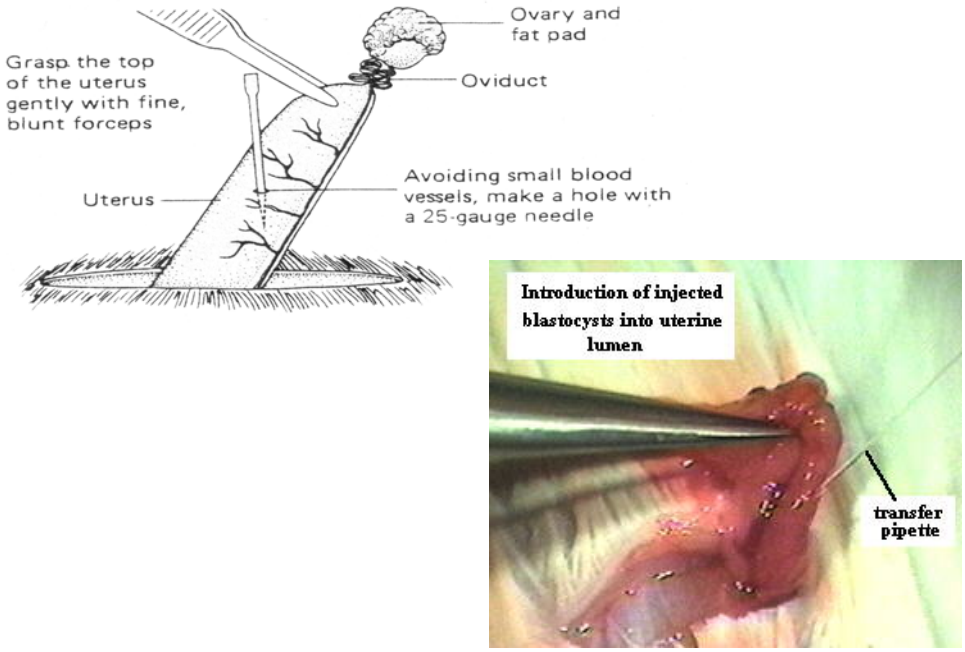


Abb. 2.7: Implantation manipulierter Blastozysten

Für die Präparation eines Uterus-Horns wird ein Schnitt 0,5 cm rechts oder links des Rückgrat in der Mitte zwischen dem Rückenbuckel und dem Schnittpunkt von hinterem Lauf und Abdomen platziert. Das Uterushorn wird freigelegt und die Blastozysten durch eine Hohlschliffnadel direkt in den Uterus injiziert. Der Vorgang ist sowohl schematisch (aus „Manipulating the Mouse Embryo“) als auch als Foto (The Microinjection Workshop v2K1, Gary A.J. Brown, University of Florida's Transgenic Core Facility) dargestellt.

2.11.7 Identifizierung von Chimären

Die ES-Zellen wurden aus Tieren des Mausstammes 129/Sv x 129/Sv-CP isoliert, deren Fell eine hellbraune (Okker) Färbung aufweist (Agouti-Fellmarker). Die bei der Aggregation eingesetzten Blastozysten wurden aus einem anderen Mausstamm, dem C57/BL/6N-Stamm isoliert, dessen Tiere eine schwarze Fellfarbe haben. Sind ES-Zellen an der Entwicklung der retransferierten Embryonen involviert, so können diese chimären Tiere anhand einer gescheckten Fellfärbung identifiziert werden. Die nichtchimären Tiere sind schwarz. Männliche Mäuse mit einem Chimärismus von mindestens 25 % wurden mit den C57/BL/6N-Weibchen zurückgekreuzt.

2.12 In Silico-Analysen von Nukleotid- und Aminosäuresequenzen

Viele der Analysen erfolgten durch Programme, die im weltweiten Netz (WWW) zur Verfügung stehen. Als Schnittstelle in das WWW wurden die Programme NETSCAPE NAVIGATOR™, NETSCAPE COMUNICATOR™ (Netscape Communications Corporation, USA) und Microsoft (R) Internet Explorer (Microsoft Corporation, USA) auf den Plattformen der Betriebssysteme von Apple und Microsoft verwendet. In Tabelle 2.7 sind die WWW-Adressen aller im weltweiten Netz verwendeten Programme aufgeführt. Neben diesen Analysen wurden auch lokale Programme zur Sequenzauswertung und Interpretation auf Einzelplatzrechnern mit Betriebssystemen von Microsoft und Apple verwendet. Die Anpassung aller Programmparameter erfolgte datenspezifisch und zielorientiert.

Bearbeitung von Rohdaten

Unbearbeitete Sequenzdaten wurden mit Hilfe des Programms DATA ANALYSIS (PE Applied Biosystems, Weiterstadt) editiert. Die Weiterverarbeitung der Einzelsequenzen aus Cosmid-, PAC- und cDNA-Klonen sowie genomische Sequenzinformationen zu komplexen Gesamtdaten erfolgte mit Hilfe des Programms SEQUENCHER™ 3.0-4.1 (Gene Codes Corporation, USA). Das Programm SEQUENCHER™ 3.0-4.1 diente unter anderem auch zur Erstellung von Restriktionskarten, Übersetzung von Nukleotid- in Aminosäuresequenzen und Visualisierung verarbeiteter Sequenzdaten. Weiter wurde das Programmpaket Lasergene (DNASTAR™, USA) verwendet, das die Bearbeitung von Rohdaten, Management von Sequenzierungprojekten, Sequenzanalysen, Erstellung von Restriktionskarten, Vergleiche von Sequenzdaten und Proteinanalysen ermöglicht.

Datenbanksuchen

Datenbanksuchen mit Nukleotid- und Aminosäuresequenzen wurden mit den BLAST-Algorithmen (Basic Local Alignment Search Tool) BLASTN, BLASTP, BLASTX, TBLASTN, TBLASTX, GAPPED BLAST, PSI-BLAST, MEGABLAST, RPS-BLAST, BLAST2, GENOMIC BLAST, VecScreen-BLAST und TRACE BLAST (Altschul et al., 1990; Gish und States, 1993; Madden et al., 1996; Altschul et al., 1996; Zhang und Madden, 1997; Pearson und Lipman, 1988) und den Datenbanken „nr“, „est“, est human“, est mouse“, „gss“, „htgs“, „pdb“, „mounth“, „alu_repeat“, „dbsts“, „chromosome“, „wgs“, „swissprot“, „pat“, „RNA“ und „Protein“ im WWW durchgeführt. Repetitive Elemente wurden vorher maskiert. Die Datenbanksuchen wurden im Februar 2004 abgeschlossen.

Bestimmung des GC-Gehalts und Suche nach CpG-Inseln

Der GC-Gehalt der Sequenzen wurde mit dem Programm WINDOW aus der HUSAR-Programmgruppe ermittelt. Die graphische Darstellung erfolgte mit dem Programm CpGPLOT aus dem EMBOSS-Programmpaket. Für die Vorhersage möglicher CpG-Inseln wurden die Programme CpG-Report, NEWCPGSEEK, CPGPLOT (alle EMBOSS-Programmpaket), Dragon GC + Promoter Finder (DRAGON GENOME EXPLORER) sowie CpGisland-revealing (WEBGENE) und CpG-FINDER (Softberry) herangezogen.

Exonvorhersage und Darstellung der genomischen Organisation

Für die computergestützte Vorhersage der Exons von *Dmxl*1 und *DMXL*1 wurden die Programme DNASweep (HUSAR-Programmpaket), GENSCAN (Burge et al., 1997, 1998), FGENES(H) -M -GC (Softberry), cdna2genome (HUSAR-Programmpaket), HmmGENE (CBS Prediction Servers), GeneID (Blanco et al., 2002, 2001; Parra et al., 2000), XPOUND (Thomas et al., 1994) und DOTPLOT (MegAlign, DNASTAR) verwendet. Besonders geeignet für eine Exonvorhersage, bei der die cDNA homologer Gene anderer Organismen bereits bekannt ist, ist DNASweep. Das Programm maskiert eine Sequenz zunächst mit dem Werkzeug RepMask, um unerwünschte Homologien repetitiver Elemente bei nachfolgenden Datenbanksuchen zu verhindern. Die so maskierte Sequenz wird anschliessend mit GENSCAN auf seine Genstruktur hin untersucht. Parallel erfolgen Datenbanksuchen mit BlastN2 in allen öffentlichen Datenbanken (NR, EST, HTG). Mit FastA wird die Datenbank eukaryotischer Promotoren (EPD) durchsucht; das Programm FACTOR durchsucht die TRANSFAC-Datenbank nach potenziellen Transkriptionsfaktoren. Alle Ergebnisse werden dann von DNASweep verarbeitet und eine Vorhersage der möglichen Genstruktur getroffen. DNASweep ist das derzeit fortschrittlichste Programm in diesem Bereich. Das Programm DOTPLOT eignet sich ebenfalls für eine Exonvorhersage, wenn die entsprechende cDNA eines nahe verwandten Organismus bekannt ist. Hier müssen die Exon/Intron-Grenzen jedoch per Auge definiert werden. Für die Analyse der gesamten genomischen Organisation von *Dmxl*1 wurde das Programm SIM4 bzw. cDNA2genome (beide HUSAR-Programmpaket) verwendet. Beide Programme vergleichen eine vorhandene cDNA mit der zugehörigen genomischen Sequenz. Hierbei werden mit Hilfe des BLAST-Algorithmus Segmentpaare erzeugt und bewertet. cDNA2genome verwendet zusätzlich noch die öffentlichen EST-Datenbanken. Anschliessend ermitteln beide Programme die am besten passenden Segmentketten und definieren die Exon/Intron-Grenzen nach der GT-AG oder CT-AC Regel.

Paarweise Sequenzvergleiche

Die Nukleotid- und Aminosäuresequenzen des orthologen humanen und murinen DmX Gens wurden mit den Programmen ALIGN „water“ bzw. „needle“ (beide EMBOSS-Programmpaket), MEGALIGN (DNASTAR), MALIGN, PRETTYPLOT (HUSAR-Programmpaket) ECR Browser (Ovcharenko et al., 2004) und MULAN (http://mulan.dcode.org) verglichen. Ein Vergleich der abgeleiteten Aminosäuresequenzen von DmX verschiedener Spezies wurde mit dem Programm CLUSTALW durchgeführt.

Phylogenie

Phylogenetische Analysen von Nukleotid- und Aminosäureaequenzen wurden mit den Programmen MULAN (http://mulan.dcode.org) und eShadow (Ovcharenko et al., 2004) erarbeitet.

„percent identity plot“ bzw. zPicture-Analysen

Für einen graphischen Vergleich von humanen und murinen Sequenzabschnitten wurden die Programme VISTA (Dubchak et al., 2000; Mayor et al., 2000; Bray et al., 2003), Multi PipMaker (Schwartz et al., 2000), Laj (local alignments with java Wilson et al., 2001) und zPicture (Ovcharenko et al., 2004) verwendet.

Identifizierung und Maskierung repetitiver Elemente

Position und Klasse der repetitiven Elemente wurden durch das Programm REPEATMASKER2 (Smit, AFA & Green, unveröffentlicht) ermittelt, zugleich wurden repetitive Elemente in den übermittelten Sequenzen maskiert und diese dann für weitere Analysen herangezogen. Alternativ wurde das Programm RepMask (HUSAR-Programmpaket) verwendet.

Promotorvorhersage, *cis*-Regulatorische Elemente und Transkriptionsfaktoren

Zur Vorhersage von Promotor- und Enhancerstrukturen und möglichen Bindungsstellen von Transkriptionsfaktoren wurden jeweils 10000 bp aus dem 5’-Bereich des murinen und humanen DmX mit den Programmen rVISTA (Loots et al., 2002, 2004), Promotor 2.0 Prediction Server (Knudsen, 1999), NNPP (Reese et al., 1996), TSSW (Solovyev und Skolnick, 1997), DOTPATH (HUSAR-Programmpaket), DNA Block Aligner Form (DBA) und PromoterWise (beide EMBL-EBI Toolbox), TSSG und PromH(W) (beide Softberry), CONPRO (Liu et al., 2002), Dragon Promoter Finder ver.1.4 (DRAGON GENOME EXPLORER; Bajic et al., 2002, 2003) und GenomatixSuite (Unterprogramme ElDorado, Gen2Promoter, GEMS Launcher, MatInspector und

PromoterInspector; Genomatix Software GmbH) analysiert. Teilweise erfolgten die Analysen unter Einbeziehung des Maus/Mensch-Sequenzvergleichs.

Primer-Design

Das Primer-Design erfolgte durch webbasierte Tools wie Primer3 (Steve Rozen und Helen J. Skaletsky, 2000), PrimerFinder, DoPrimer, PRIMO, WEBPRIMER, PCP primer selection.

Polyadenylierungssignale und funktionelle Signale im 3'-UTR

Neben der Suche per Auge wurden die Programme POLYA SCAN (Hornischer und Blöcker, unpubliziert), ERPIN (Gautheret et al., 2001), POLYADQ (Tabaska et al., 1999) und UTRscan (Pesole et al., 1999, 2002) verwendet.

Analyse der Aminosäuresequenz

Die abgeleitete Aminosäuresequenz von Dmxl1 wurde mit den Programmen SMART (Schultz et al., 1998; Letunic et al., 2002), REP (Andrade et al., 2000) und PSA (Stultz et al., 1993, 1997; White et al., 1994) auf mögliche WD-Wiederholungseinheiten hin untersucht. Mit dem Programm SMART konnte die Aminosäuresequenz zusätzlich auf mögliche homologe Strukturen und Domänen bereits bekannter Proteine, auf Signalpeptide und interne Repeats überprüft werden. Mit REP (Andrade et al., 2000) wurde zusätzlich die Aminosäuresequenz auf bekannte Repeats hin analysiert. Die Suche nach konservierten Aminosäuremotiven wurde mit den Programmen PPSearch (EMBL-EBI Toolbox), Predict Protein Server (Rost et al., 2003), MotifScan (Pagni et al., 2001) sowie diversen Programmen des NPS@-Pakets (Network Protein Sequence Analysis; Pôle Bio-Informatique Lyonnais) durchgeführt. Für die Identifikation von Transmembranhelices wurden die Programme TMpred (Hofmann et al., 1993), TM-Finder (Deber et al., 2001), TOPPRED (von Heijne et al., 1992), PHDhtm (Rost et al., 1995), TMHMM Server 2.0 (CBS Prediction Servers), tmap (EMBOSS-Programmpaket), orienTM (Liakopoulos et al., 2001) und SOSUI (Hirokawa et al., 1998) verwendet. Die subzellulare Lokalisation wurde mit dem Programm ESLpred (Bhasin et al., 2004) bestimmt. Eine Einordnung der Aminosäuresequenz in eine Proteinfamilie erfolgte mit dem Programm Superfamily (Madera et al., 2004). Proteinfaltungen wurden mit dem Programm RasTop visualisiert (http://www.geneinfinity.org/rastop/).

2.13 Nomenklatur von DmX

Die Nomenklatur von DmX ist durch die Richtlinien des „Human Gene Nomenclature Committee" und des „MDG Mouse Nomenclature Committee" festgelegt. Nach diesen Richtlinien werden humane Gene groß und kursiv geschrieben, bei murinen Genen wird

nur der Anfangsbuchstabe groß geschrieben und alle Buchstaben kursiv. Aufgrund der Erstbeschreibung von *DmX* (*D. melanogaster X* gene) bei *Drosophila melanogaster* durch Kraemer et al. (1998), basieren die homologen humanen und murinen Gene auf diesem Namen. Wird im Allgemeinen von DmX und seinen Homologen gesprochen, so wird der erstbeschriebene Name ohne kursive Buchstaben, hier also DmX für „*D. melanogaster X* gene“, verwendet. Soll das humane homologe Gen angesprochen werden, so wird dieses als *DMXL*1(DmX-like 1) bezeichnet. Gleiches gilt für das murine Gen, also *Dmxl*1. Da sowohl beim Menschen als auch bei der Maus zwei unterschiedliche Kopien von DmX existieren (Kreamer et al., 1999) werden diese z.B. mit *DMXL*1 und *DMXL*2 bezeichnet. Bei der Beschreibung der zugehörigen Proteine entfällt die kursive Schrift und das Wort „Protein“ wird ergänzend angefügt (z.B. Dmxl1-Protein für das murine System).

2.14 Puffer und Lösungen

BCS-Puffer	2x RE-Puffer (Enzym-Originalpuffer) 0,2 mg/ml BSA 0,2 mg/ml Casein 0,2 mg/ml Spermidin
Denaturierungspuffer (Southern-Blot)	1,5 M NaCl 0,5 M NaOH
10x Dialysepuffer	0,25 M Tris 3 M NaCl 0,1 M Na_2EDTA
DNA-Stoppuffer	4 M Harnstoff 0,1 M Na_2EDTA 0,1 % w/v Bromphenolblau 50 % w/v Saccharose
10x E-Puffer	0,36 M Tris 0,3 M NaH_2PO_4 0,1 M Na_2EDTA
Ethidiumbromid Färbelösung	5 µg Ethidiumbromid / ml in 1x E-Puffer
FBI-Puffer	10% PEG 8000 7% SDS 1,5x SSPE-Puffer

Ficoll-Ladepuffer	20% w/v Ficoll 100 mM Na_2EDTA 1% w/v SDS 0.25% w/v Bromphenolblau 0.25% w/v Xylene Cyanol
Guanidin-Lösung	4 M Guanidinthiocyanat 25 mM Natriumcitrat 0,5 % w/v N-Laurylsarcosin 0,1 M β-Mercaptoethanol
Hybridisierungslösung nach Curch et al.	7% w/v SDS 0,5 M Na_3PO_4 pH 7,2 1 mM Na_2EDTA
Homogenisierungspuffer	11,6% w/v Saccharose 2 mM Na_2EDTA pH 7,4 0,5 mM EGTA pH 7,4 0,06 mM KCl 15 mM NaCl 1,5 mM Spermin 0,5 mM Spermidin 15 mM Tris-HCl pH 7,4
Kombinationspuffer (5x)	0,05 M $MgSO_4$ 0,25 M Tris pH7,5 0,5 mM DTT
L-Medium	10 g Trypton 5 g Hefeextrakt 5 g NaCl ad 1000 ml A. bidest
Neutralisierungspuffer (Southern Blot)	3 mM NaCl 0,5 mM Tris-HCl pH 7,5
1x PM	0,02% w/v Ficoll 400 0,02% w/v Polivinylpyrrolidon 0,02% w/v BSA in 3x SSC
Proteinase K-Puffer	100 mM Tris pH 7,6 50 mM Na_2EDTA 0,5% w/v SDS 1 mg/ml Proteinase K

SDS-Mix	2,5% w/v SDS 0,25 M Na_2EDTA 0,5 M Tris-HCl pH 8,0
Sol I	10 mM Na_2EDTA
Sol II	0,2 M NaOH 1% w/v SDS
Sol III	50 ml 7,5 M Kaliumacetat 23 ml Essigsäure (konz.) 127 ml A. bidest
20x SSC	3 M NaCl 0,3 M Natriumcitrat
20x SSPE-Puffer	3 M NaCl 0,2 M NaH_2PO_4 0,02 M Na_2EDTA pH 7,4
10x Taq-Puffer	400 mM KCl 100 mM Tris-HCl pH 8,3 15 mM $MgCl_2$
20x TBE	1,8 M Tris 1,8 mM Borsäure 25 mM Na_2EDTA
TE-Puffer	10 mM Tris-HCl pH 8,0 1 mM Na_2EDTA
TE-50-Puffer	50 mM Tris, pH8 50 mM Na_2EDTA
TELT-Puffer	50 mM Tris-HCl 62 mM Na_2EDTA 2,5 M LiCl 4% w/v Triton X-100
Triton X-100 Lösung	50 mM Tris-HCl pH 8,0 62,5 mM Na_2EDTA pH 8,0 0,2% w/v Triton X-100
Waschpuffer nach Curch et al.	0,1% w/v SDS 0,04 M Na_3PO_4 pH 7,2

2.15 *E. coli*-Stämme

HB101: *sup*E44 *ara*14 *gal*K2 *lac*Y1 Δ(*gpt-pro*A)62 *rps*L20 (Str^r) *xyl*-5 *mtl*1 *rec*A13 Δ(*mcrC-mrr*) *hsd*S20$^-$(r$^-$m$^-$) (Boyer et al., 1969; Smith et al., 1989)

DH10B: F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80 *lacZ*Δ*M* 15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara,leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ$^-$ (Lorow et al.,1990)

RRI: HB101 *recA*$^+$ (Bolivar et al., 1977; Peacock et al., 1981)

Dh5α: *sup*E44 Δ*lac*U169 (ϕ80 *lacZ*Δ*M* 15) *hsd*R17 *rec*A1 *end*A1 *gyrA*96*thi*-1 *rel*A1(Hanahan, 1985)

WA312: W3110 (λC1857 Sam7) *su*

2.16 Bezugsquellen

Wenn nicht andes angegeben, wurden Enzyme und Reaktionssysteme der Firmen Amersham Life Science (USA), Boehringer (Mannheim) bzw. Roche (Mannheim), Pharmacia Biotech Gmbh (Freiburg), Stratagene Cloning Systems (Heidelberg), GIBCO**BRL** Life Technologies (Karlsruhe), Invitrogen (Karlsruhe), MBI Fermentas GmbH (St.Leon-Rot) und Promega (Mannheim) verwendet. Chemikalien und Feinchemikalien wurden von den Firmen Roth (Karlsruhe), Merck (Darmstadt), Sigma-Aldrich (Deisenhofen) und ICN Biomedicals (Eschwege) bezogen, Kunstoff-Verbrauchsmaterialien von den Firmen Biozym (Oldendorf), Roth (Karlsruhe), Thermo Life Sciences (Dreieich), Greiner bio-one (Frickenhausen) und Starlab (Ahrensburg).

Weitere Bezugsquellen:

Bacto-Hefeextrakt	Difco, USA
Bacto-Trypton	Difco, USA
Tris	Appli Chem (Darmstadt)

3 Ergebnisse

3.1 Funktionelle Analyse durch „Knock-Out“

3.1.1 Identifikation von *Dmxl*1 in *Mus musculus*

Zu Beginn der vorliegenden Arbeit war durch einen EST-Klon aus dem 3’-UTR bekannt, dass das DmX-Gen in *Mus musculus* (*Dmxl*1) existent sein musste. Aufgrund von EST-Daten und cDNA-Synthesen konnte gezeigt werden, dass auch ein homologes Gen in *Homo sapiens* (*DMXL*1) existiert und dort auch exprimiert wird (Kraemer, persönliche Mitteilung). Da es sich bei DmX um ein evolutionär hochkonserviertes Gen handelt und zu diesem Zeitpunkt neben dem Menschen homologe Gene in *D. melanogaster* (*DmX*), *C. elegans* (*F54E4.1*) und *C. thummi* (*CpY*) bekannt waren (Kraemer et al., 1997), konnte davon ausgegangen werden, dass eine Identifikation in der Maus Erfolg haben würde. Vergleichende Sequenzanalysen zur Genidentifizierung zwischen Mensch und verschiedenen Nagetieren (Koop, 1995; Oeltjen et al. 1997) und der Vergleich genetischer Karten von Mensch und Maus (DeBry und Seldin, 1996) stützten diese Annahme. Zur Identifizierung von *DmxL*1 in der Maus wurden zunächst zwei Genbanken, die zu diesem Zeitpunkt zur Verfügung standen, auf positive Klone hin untersucht (Kapitel 2.5.3), später stand eine dritte Genbank zur Verfügung. Aufgrund der zu erwartenden Polymorphismen, vor allem in nicht kodierenden Sequenzabschnitten (Thomas et al., 2003), zwischen dem Inzuchtstamm C57/BL6 und den verwendeten ES-Zellklonen 129-R1 aus der 129/Sv Maus-Linie (Munclinger et al., 2003; Greene-Till et al., 2000; Zhao et al.,1995; International Human Genome Consortium, 2001), wurden positiv detektierte Klone der Genbank Nr. 703 (Kapitel 2.5.3) nicht weiter bearbeitet. Positiv detektierte Klone mit Sonde H19442 der DNA-Bibliothek Nr. 703 (Kapitel 3.1): P1-F15150Q5, P1-N06163Q5, P1-N14151Q5, P1-A12164Q5, P1-P03325Q5. Der Klon P1-P03325Q5 zeigte in einer Southern-Analyse mit Sonde H19442 (Kapitel 3.1.1) eine Hybridisierung mit einem Restriktionsfragment.

3.1.1.1 Cosmid-Klone aus der Maus

Da aus dem murinen *Dmxl*1 nur wenig Sequenzinformationen zu Verfügung standen, sollten die Klonbibliotheken Nr. 121 (129/Ola) durch Xeno-Hybridisierung mit Sonden aus dem kodierenden Bereich des humanen *DMXL*1 abgesucht werden. Als Sonden wurden das Integrat des cDNA Klon’s IMAGp998H19442 (Sonde H19442) und ein PCR-Amplifikat (Sonde Homo34/35) der cDNA vom humanen *DMXL*1 verwendet, beide decken zusammen die Exons 5-17 des kodierenden Sequenzabschnitts von *DMXL*1 ab. Durch Koloniefilter-Hybridisierung wurden aus der genomischen Maus-Cosmid-Bibliothek fünf Klone (Cos-H23233Q4, Cos-N10467Q3, Cos-N06467Q3, Cos-C23546Q3, Cos-121I23534Q3) isoliert. Von den fünf Cosmid-Klonen zeigten nach Southern-Analyse zwei Klone (Cos-N10467Q3, Cos-N06467Q3) eine Hybridisierung

von Restriktionsfragmenten mit beiden Sonden (Abb. 3.1b/c). Das Restriktionsmuster der Klone ist Identisch, beide Klone enthalten das gleiche Integrat (Abb. 3.1a). Es wurde Cosmid Cos-N10467Q3 für eine Sequenzierung ausgewählt.

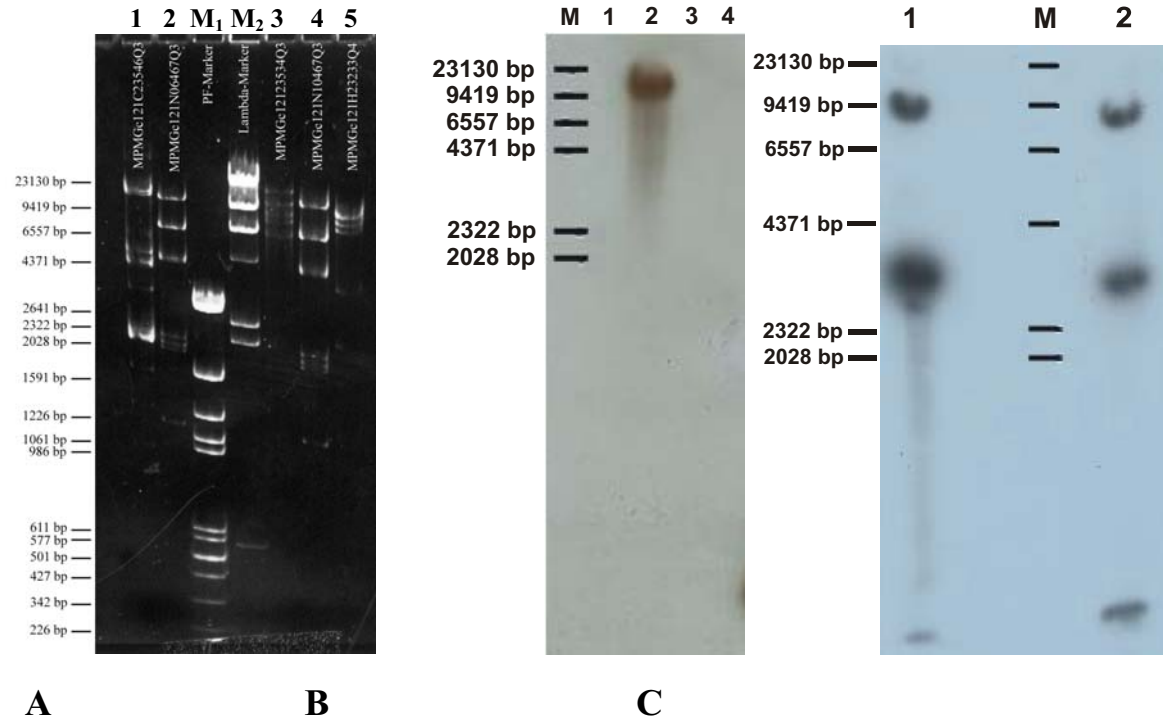


Abb 3.1: Restriktions- und Southern-Analyse der Cosmid-Klone

A: Restriktionsanalyse der Cosmid-Klone Cos-C23546Q3 (1), Cos-N06467Q3 (2), Cos-I23534Q3 (3), Cos-N10467Q3 (4) und Cos-H23233Q4 (5) mit dem Enzym *Hind* III, M1 = pF-Marker, M2 = Marker λ/*Hind* III. Die Klone Cos-N10467Q3 und Cos-N06467Q3 zeigen ein identisches Restriktionsmuster. Die daraus ermittelte Größe der Klone liegt bei ~ 33400 bp, davon sind 5380 bp Vektoranteil.

B: Autoradiogramm des Southern Blots der Klone, Cos-I23534Q3 (1), Cos-N10467Q3 (2), Cos-H23233Q4 (3) und Cos-23546Q3 (4) nach Restriktion mit *Bam H*I und Hybridisierung mit Sonde H19442, M = Marker λ/*Hind* III. Nur bei Klon Cos-N10467Q3 konnte ein Restriktionsfragment von ~ 16 kb nachgewiesen werden.

C: Autoradiogramm des Southern Blots der Klone Cos-N06467Q3 (1) und Cos-N10467Q3 (2) nach Restriktion mit *Hind* III und Hybridisierung mit Sonde Homo34/35, M = Marker λ/*Hind* III. Bei beiden Klonen konnten identische Restriktionsfragmente nachgewiesen werden.

3.1.1.2 PAC-Klone aus der Maus

Durch die Sequenzierung von Cosmid Cos-N10467Q3 wurden ausreichend Sequenzinformationen von *Dmxl*1 gewonnen, die es ermöglichten, die genomische PAC-Klonbibliothek Nr. 711 (RPCI21 Mouse PAC) zu durchsuchen. Als Sonde wurde ein PCR-Produkt (BE 1) aus Exon 18 des *Dmxl*1-Gens verwendet. Es konnten mit dieser Sonde

neun PAC-Klone identifiziert werden: PAC-H15193Q2, PAC-L19656Q2, PAC-K07584Q2, PAC-O18380Q2, PAC-F19112Q2, PAC-C05346Q2, PAC-M04307Q2, PAC-F13324Q2 und PAC-I02348Q2. Hierbei zeigte sich, dass alle diese Klone, die vom RZPD angefordert und zugeschickt wurden, mit *Pseudomonas* kontaminiert waren. Nach Dekontamination mit Cefsulodin (Kapitel 2.3.3) erfolgte eine Restriktion der Klon-DNA mit Hind III. Die Klone PAC-F13324Q2, PAC-I02348Q2, PAC-O18380Q2, PAC-F19112Q2 und PAC-H15193Q2 wurden für eine Southern Blot-Analyse ausgewählt (Abb. 3.2), die verworfenen Klone hatten kein oder ein sehr kleines Integrat. (Daten nicht gezeigt). Nach Hybridisierung mit Sonde BE1 zeigte sich bei den Klonen PAC-O18380Q2, PAC-F19112Q2, PAC-F13324Q2 und PAC-I02348Q2 ein positives Signal (Abb. 3.3). Aufgrund der geringen Integratgrößen von PAC-O18380Q2 und PAC-F19112Q2 wurden nur die Klone PAC-F13324Q2 und PAC-I02348Q2 für weitere Analysen verwendet.

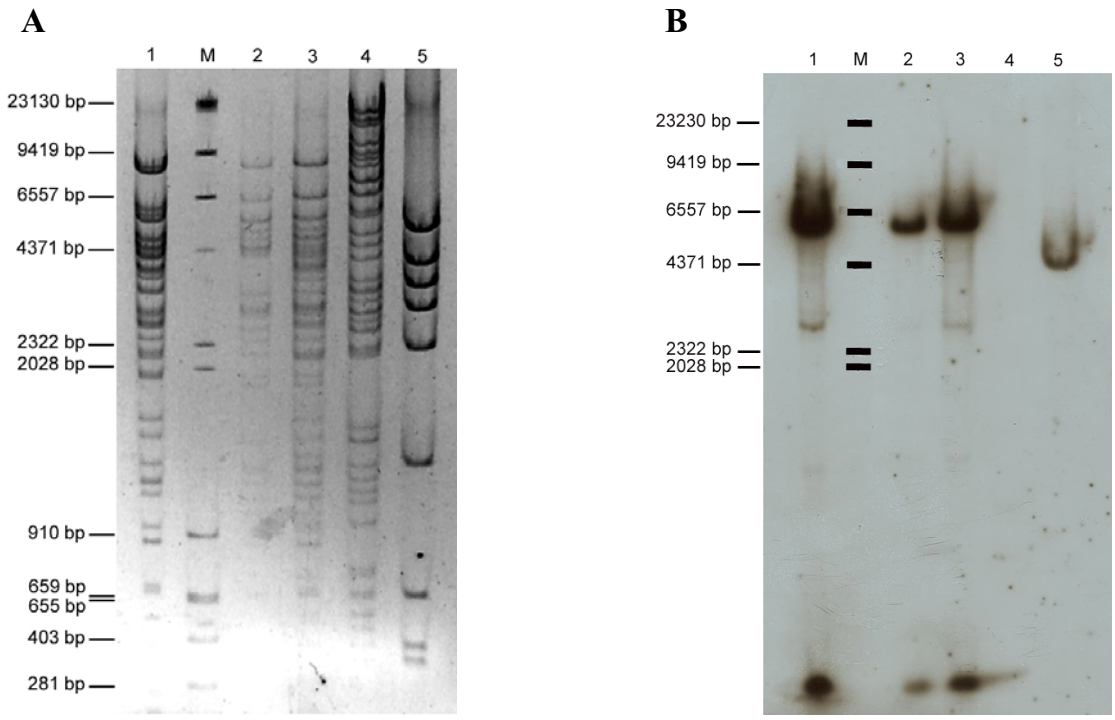


Abb. 3.2: Southern-Analyse der PAC-Klone PAC-O18380Q2, PAC-F19112Q2, PAC-F13324Q2, PAC-I02348Q2 und PAC-H15193Q2.

A. Restriktionsmuster der Klone mit dem Enzym Hind III (Spur 1: PAC-F13324Q2; Spur M: λ- + pBR-Marker; Spur 2: PAC-O18380Q2; Spur 3: PAC-I02348Q2; Spur 4: PAC-H15193Q2; Spur 5: PAC-F19112Q2).
B. Autoradiogramm von A nach Southern-Hybridisierung mit Sonde BE1. In Spur 1, 2 und 3 konnten die gleichen Restriktionsfragmente nachgewiesen werden. Klon PAC-H15193Q2 (Spur 4) zeigte keine Hybridisierung mit der Sonde. Bei Klon PAC-F19112Q2 zeigt sich eine Hybridisierung, das hybridisierende Fragment ist jedoch deutlich kleiner als das der Klone PAC-F13324Q2, PAC-O18380Q2 und PAC-I02348Q2.

3.1.1.3 Lage der PAC- und Cosmid-Klone

Ausgangspunkt für die Analyse von *Dmxl*1 waren die identifizierten PAC- und Cosmid-Klone PAC-F13324Q2 und PAC-I02348Q2 sowie Cos-N10467Q3. Alle Klone enthalten Teile von *Dmxl*1, jedoch keiner von ihnen den vollständigen genomischen Abschnitt. Durch Sequenzierung und Vergleich mit Angaben in Datenbanken konnten die Klone auf Chromosom 18 Bande C lokalisiert werden. Die relative Lage der Klone zueinander und ihre chromosomale Positionierung sind in Abb 3.3 wiedergegeben.

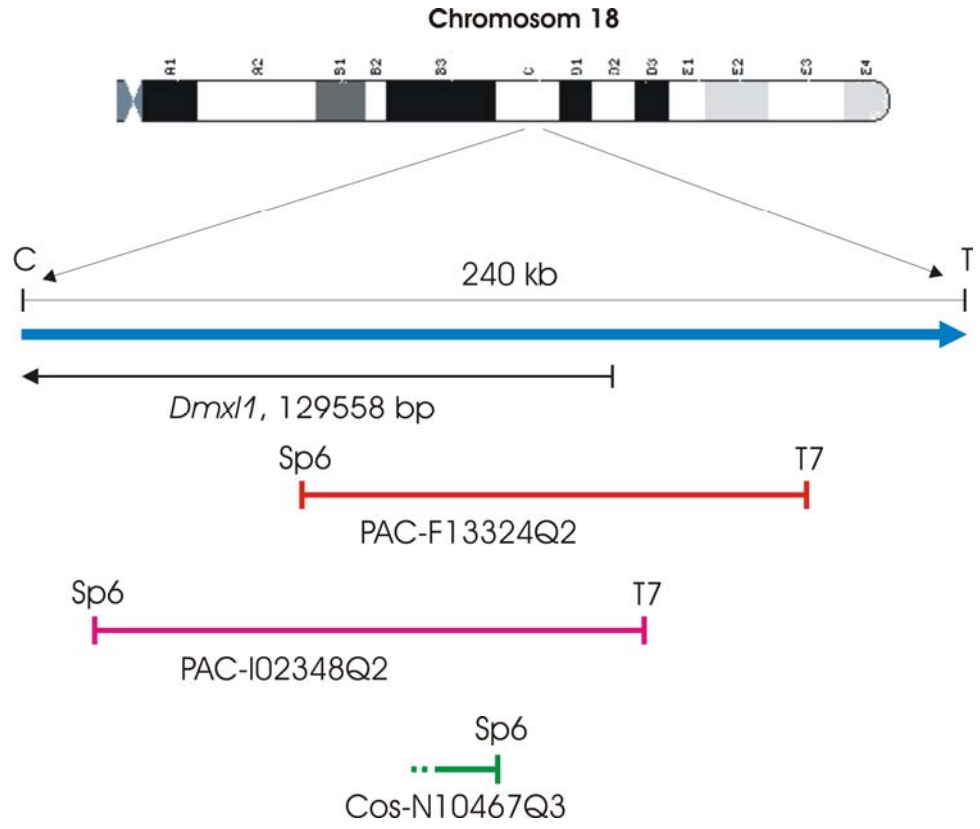


Abb 3.3: Position der PAC- und Cosmid-Klone

Gezeigt ist die genomische Lage und Orientierung der beiden PAC-Klone PAC-F13324Q2 und PAC-I02348Q2 sowie von Cosmid N10467Q3 zu *Dmxl*1. Berücksichtigt wurde hierbei nur der translatierte Abschnitt von *Dmxl*1 (Kapitel 4.5.4). Eine vollständige Abdeckung der Region mit PAC-Klonen konnte nicht erreicht werden. Da es sich bei Cosmid N10467Q3 vermutlich um ein *in vitro*-Ligat handelt (Kapitel 3.5.3), wurde der T3-Terminus nicht berücksichtigt.

3.1.2 Sequenzierung der Maus-Klone Cos-N10467Q3, PAC-F13324Q2 und PAC-I02348Q2

3.1.2.1 Sequenzierungsstrategie für Cos-N10467Q3

Ziel der Sequenzierungen war die Ermittlung der kodierender Abschnitte von *Dmxl*1, um mit diesen Sequenzinformationen ein System zu entwickeln, das ein gezieltes Gen-Targeting durch die Konstruktion eines Target-Vektors sowie nachfolgend eine Detektion der manipulierten genomischen Sequenzabschnitte ermöglicht. Für die Sequenzierung von Cos-N10467Q3 wurde eine Kombination aus „primer-walking", Subklonierung und Religationsklonierung gewählt (Abb. 3.4). Durch Subklone und Religationsklone sollten möglichst viele Startpunkte für die anschließende „primer-walking"-Strategie generiert werden, zudem konnten Subklone gegebenenfalls direkt zur Konstruktion des Targeting-Vektors Verwendung finden. Das „primer-walking" erfolgte sowohl an den Subklonen/ Religationsklonen als auch direkt an Cos-N10467Q3. Die ermittelte Sequenz von Cos-N10467Q3 befindet sich auf dem beigefügten Datenträger. Die Sequenzauswertung erfolgt in Kapitel 3.5.1. Bei gegebenen technischen Schwierigkeiten der direkten Sequenzierung von Cosmid-DNA wurde zunächst der entsprechende DNA-Abschnitt mittels PCR amplifiziert und anschließend sequenziert. Insgesamt wurden 137809 bp Rohdaten mit 250 erfolgreichen Sequenzierungen ermittelt, das entspricht einer durchschnittlichen Leseweite von 551 bp und einer Redundanz von 4,95. Die für diese Strategie hohe Redundanz begründet sich auf Mehrfachsequenzierungen der kodierenden Abschnitte von *Dmxl*1 in Cos-N10467Q3. Es konnten zehn Exons identifiziert werden.

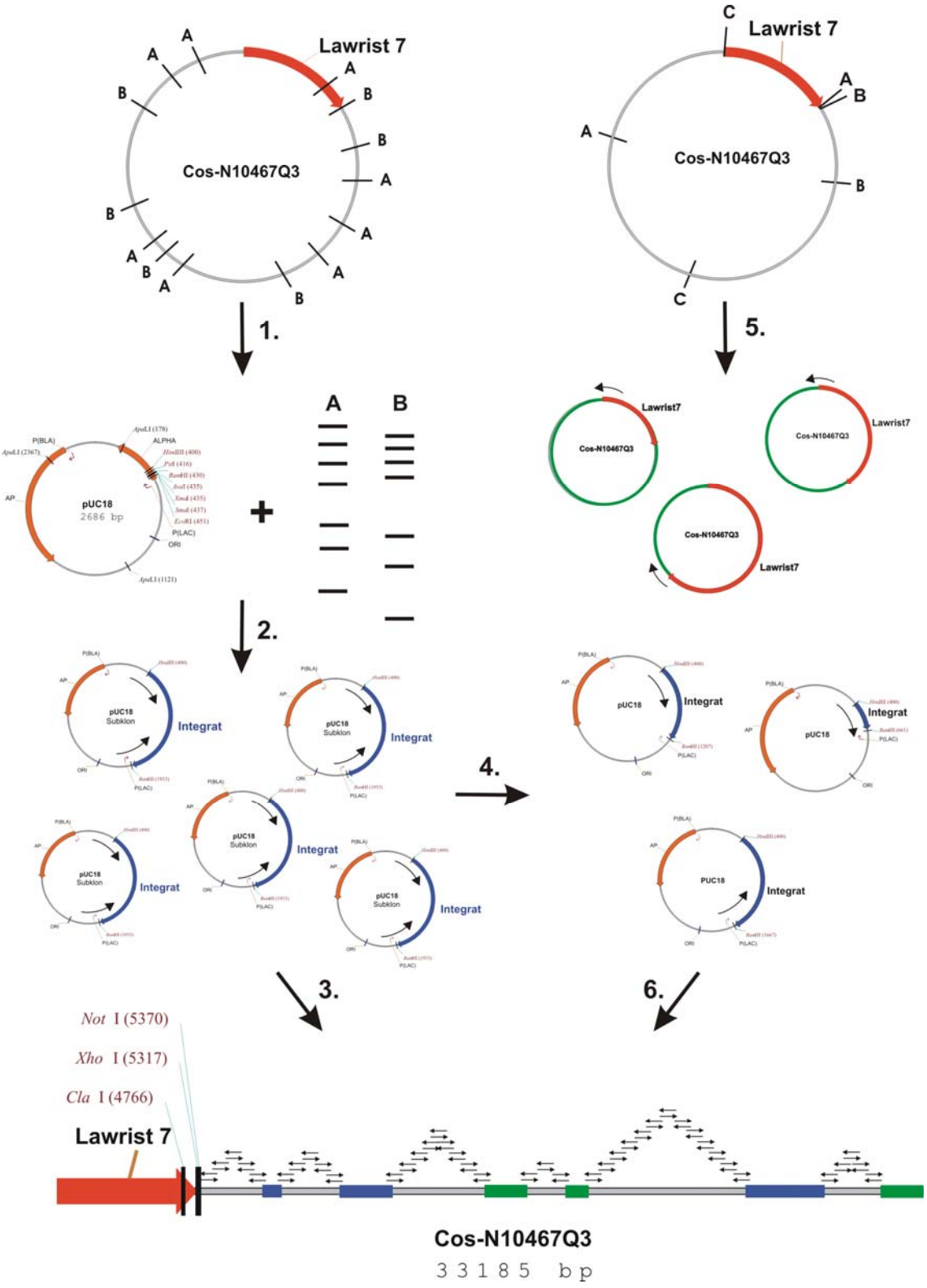
3.1.2.2 Selektive Sequenzierung von PAC-F13324Q2 und PAC-I02348Q2

Durch Vergleich der menschlichen cDNA von *DMXL*1 und der genomischen Maus-Sequenz in Cos-N10467Q3 konnten die enthaltenen kodierenden Abschnitte des murinen *Dmxl*1 ermittelt werden (Kapitel 3.5.1). Es wurde dabei schnell klar, dass in Cos-N10467Q3 nur ein Bruchteil der Exons des *Dmxl*1-Gens zu finden waren (Kapitel 3.5.3). Durch den Vergleich der Exon- und Intronlängen in Bezug auf die Anzahl der kodierenden Sequenzabschnitte wurde eine Verteilung der Exons auf mehr als 120 kb geschätzt. Da es sich bei dieser Arbeit nicht um ein Sequenzierungsprojekt handeln sollte, wurde zur Identifikation der noch fehlenden kodierenden Abschnitte und zur Ermittlung der Genstruktur von *Dmxl*1 eine neue Strategie gewählt. Zunächst sollte, wie unter 3.1.2 beschrieben, eine PAC-Bank nach positiven Klonen durchsucht werden. Die hierbei identifizierten Klone PAC-F13324Q2 und PAC-I02348Q2 wurden dann mittels „shotgun"-Verfahren (Kapitel 2.4.4) statistisch in Fragmente zerlegt und subkloniert auf gerasterte Filter übertragen. Als Sonden für eine Koloniefilter-Hybridisierung (Kapitel 2.5.2) wurden RT-PCR-Produkte verwendet, die das humane *DMXL*1-Gen abdeckten,

dessen cDNA zu diesem Zeitpunkt gerade bekannt war (zur Verfügung gestellt von Frau Dr. C. Kraemer). Ausgenommen wurden hier die kodierenden Bereiche, die schon durch die Sequenzierung von Cos-N10467Q3 bekannt waren, um Mehrfach-Sequenzierungen zu vermeiden. Auf diese Weise sollten durch Interspezies-Hybridisierung alle Klone detektiert werden, in denen kodierende Abschnitte des murinen *Dmxl*1 enthalten waren, um aus den Sequenzdaten Primer zu generieren, die dann eine Synthese der cDNA über RT-PCR zulassen. Zunächst wurden die Filter mit der kleinsten Fraktion untersucht, da eine überlappende bzw. vollständige Sequenzierung mit den Standard-Primern möglich war. Die so gewonnenen Sequenzinformationen wurden mit der humanen cDNA von *DMXL*1 abgeglichen und erneut Sonden hergestellt, die nur noch fehlende oder unvollständige Abschnitte des murinen *Dmxl*1 in der Genbank detektieren sollten. Mit diesen Sonden wurden dann die beiden verbleibenden Fraktionen untersucht. Auf diese Weise konnten insgesamt 60 Klone identifiziert und sequenziert werden. Diese Klone enthielten kodierende Sequenzen des murinen *Dmxl*1 die 23 Exons repräsentieren, die nicht in Cos-N10467Q3 enthalten waren. Weitere Exons konnten nicht detektiert werden, da PAC-F13324Q2 und PAC-I02348Q2 nicht den kompletten codierenden Bereich von *Dmxl*1überspannen (Abb. 3.3). Ziel dieser Arbeiten waren zum einen die Ermittlung von Sequenzinformationen für eine nachfolgende cDNA-Synthese durch RT-PCR und zum anderen die Festlegung von Exon/Intron-Grenzen. Eine grafische Darstellung der Sequenzierungsstrategie findet sich in Abbildung 3.5, die ermittelten Sequenzdaten befinden sich als Sequencher-Projekt auf dem beigefügten Datenträger.

Abb. 3.4: Sequenzierungsstrategie für Cos-N10467Q3 (fogende Seite)

Schematische Darstellung der Sequenzierungsstrategie. Eine Cosmid-DNA Präparation wurde mit geeigneten Enzymen geschnitten (1.), der Restriktionsansatz gefällt und komplett mit dem vorher durch die gleichen Enzyme geschnittenen pUC18-Vektor vermischt und ligiert. Die daraus resultierenden Subklone (2.) wurden mit den Standardprimern „M13-Universal“ und „M13-Reverse“ ansequenziert. Mit den gewonnenen Sequenzinformationen konnten vorwärts und rückwärts gerichtete Primer entwickelt werden, die wieder auf den Subklonen für eine weitere Sequenzierungsrunde verwendet wurden. Dieser Vorgang wiederholte sich so oft, bis der gesamte Subklon sequenziert war. Gleichzeitig wurden die Sequenzinformationen und rückwärts gerichtete Primer verwendet, um durch „Primer-Walking“ direkt an der Cosmid-DNA weitere Sequenzen zu generieren (3.). Zusätzlich wurden von Cos-N10467Q3 Religationsklone durch die Verwendung von Restriktionsenzymen, die nur einmal in der multiplen Klonierungsschnittstelle schneiden, hergestellt (5.). Durch Sequenzierung mit einem der Standardprimern T7 oder Sp6 wurden weitere Sequenzinformationen ermittelt, die wieder als Startpunkte zur Sequenzierung von Cos-N10467Q3 dienten (6.). Von Subklonen, in die große Intergrate ligiert waren, wurden ebenfalls, wie oben beschrieben, Religationsklone hergestellt (4.). Auch diese Sequenzen dienten als Startpunke für Sequenzierungsreaktionen an Cos-N10467Q3, aber auch um die Subklone schneller komplett zu sequenzieren. In blau dargestellt sind Integrate der Sub- und Religationsklone sowie die Sequenzierungen, die an den Subklonen und deren Religationsklonen erfolgten, grüne Balken symbolisieren Sequenzinformationen, die von Religationsklonen des Cosmids stammen.


Cos-N10467Q3
Lawrist 7
1.
2.
3.
4.
5.
6.
pUC18
2686 bp
ALPHA
P(BLA)
P(LAC)
ORI
A
B
C
pUC18 Subklon
Integrat
PUC18
Lawrist7
Not I (5370)
Xho I (5317)
Cla I (4766)
Cos-N10467Q3
33185 bp

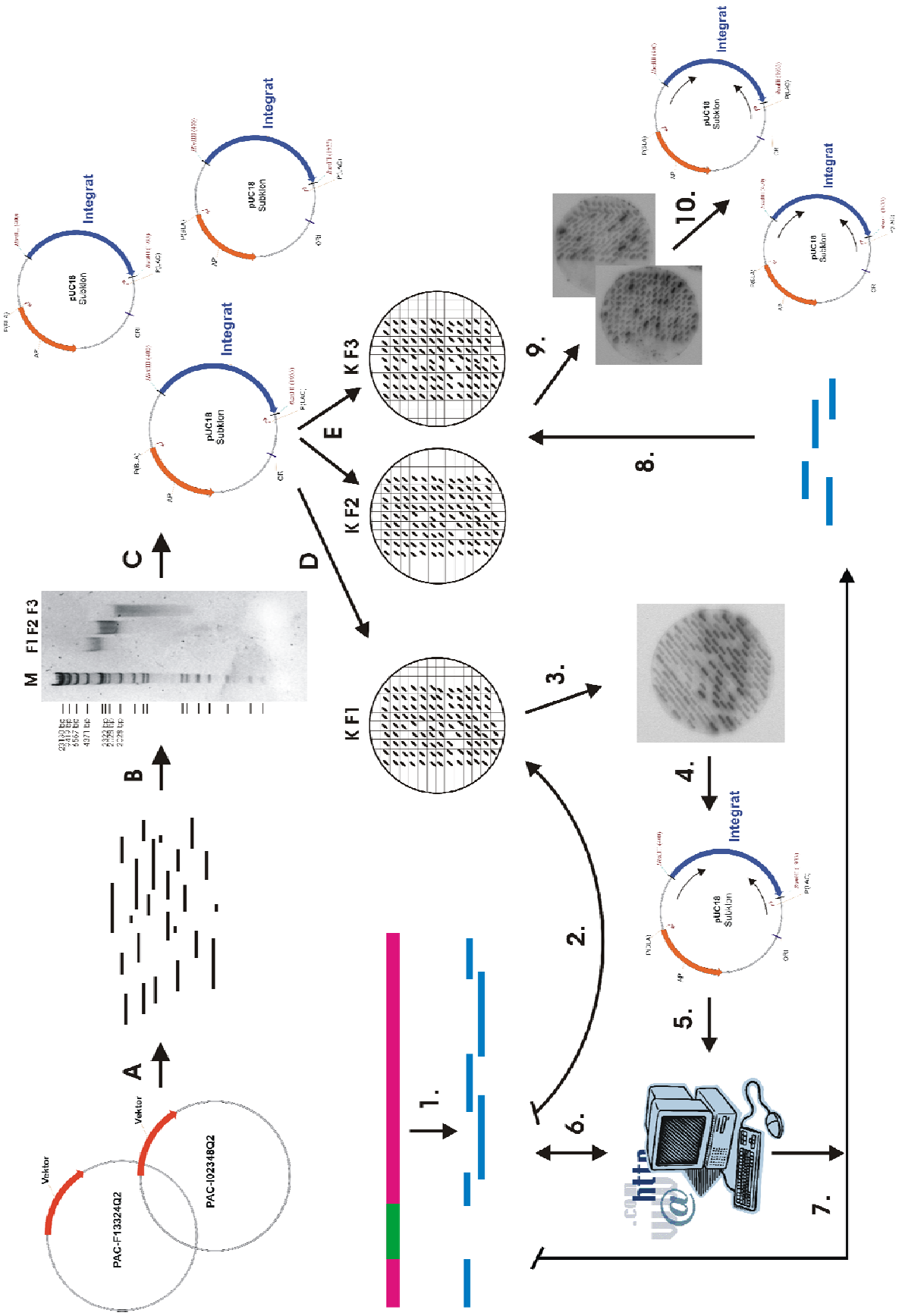

PAC-F13324Q2
Vektor
PAC-I02318Q2
Vektor
A
B
M F1 F2 F3
C
pUC18 Subklon
Integrat
D
E
K F1
K F2
K F3
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

Abb. 3.5: Selektive Sequenzierung von PAC-F13324Q2 und PAC-I02348Q2 (vorhergehende Seite)

Dargestellt ist die selektive Sequenzierungsstrategie, eine effiziente Methode, um durch Xeno-Hybridisierung und Interspeziesvergleich homologe kodierende Sequenzabschnitte mit minimalem Sequenzierungsaufwand zu identifizieren. Zunächst wurden die PAC-Klone mittels „Nebulizer“ (Kapitel 2.4.4) statistisch geschert (A). Die DNA-Fragmente wurden dann gelelektrophoretisch aufgetrennt und in Fraktionen der Größen 0,5-1,5 kb (F1), 1,5-2,5 kb (F2) und 2,5-4,5 kb (F3) aufgeteilt (B). Jede Fraktion wurde getrennt in pUC 18 kloniert (C) und auf gerasterte Duplikatfilter übertragen (D, E). Aus der humanen cDNA von *DMXL*1 wurden dann durch PCR Sonden hergestellt (1.), ausgenommen wurden nur schon bekannte Sequenzabschnitte (grüner Balken). Die Filter mit der kleinsten Fraktion wurden dann einer Koloniefilter-Hybridisierung (Kapitel 2.5.2) unterzogen (2.) und mit einer „Imager-Plate“ (Kapitel 2.5.1) exponiert (3.). Klone mit einem positiven Signal wurden aufgearbeitet (4.) und sequenziert. Die gewonnenen Daten wurden mit den Sequenzinformationen der menschlichen cDNA abgeglichen (6.). Für noch fehlende Sequenzabschnitte wurden neue Sonden durch PCR an der humanen cDNA generiert (7.) oder direkt Integrate mit unvollständigen Exon-Informationen der Subklone verwendet (8.). Durch Koloniefilter-Hybridisierung der Fraktionen F2 und F3 (9.) konnten weitere Klone mit kodierenden Sequenzabschnitten des murinen *Dmxl*1 identifiziert (10.) und sequenziert werden (11.).

3.1.3 Die cDNA des murinen *Dmxl*1

Neben der genomischen Organisation von *Dmxl*1 (Kapitel 3.2, Kapitel 3.5.4) ist die kodierende Sequenz der funktionellen RNA von Wichtigkeit. Zu Beginn dieser Arbeit lagen Sequenzinformationen aus *Chironomus piger* (Kraemer et al., 1993) und *Drosophila melanogaster* (Kraemer et al., 1998) vollständig und aus *Homo sapiens* teilweise vor (Kraemer, persönliche Mitteilung). Die Synthese der cDNA aus menschlicher Poly $(A)^+$ RNA (Kraemer et al., 2000) erwies sich als äußerst schwierig. Zwar wird *DMXL*1 nahezu ubiquitär exprimiert (Kraemer et al., 2000), dies jedoch vermutlich in sehr geringer Kopienzahl (Kraemer, persönliche Mitteilung). Da nur wenig aus dem 3’-UTR stammende murine Sequenzinformationen von *Dmxl*1 zur Verfügung standen, wurde zunächst die genomische Sequenz der kodierenden Abschnitte von *Dmxl*1 durch selektive Sequenzierung ermittelt (Kapitel 3.2.2). Der Vergleich dieser Sequenzdaten und ergänzende Sequenzinformationen aus den öffentlichen Datenbanken mit der von Kraemer ermittelten menschlichen cDNA-Sequenz ermöglichte eine Identifizierung fast aller kodierenden Abschnitte des murinen *Dmxl*1. Basierend auf diesen Daten konnten Primer erstellt werden, die für die cDNA-Synthese Verwendung fanden. Zunächst wurde zu Testzwecken Total-RNA und Poly $(A)^+$ RNA aus verschiedenen Geweben von Mäusen der Stämme C57/BL/6 und 129/sv isoliert (Kapitel 2.9). Diese RNA wurde im Folgenden revers transkribiert (Kapitel 2.8.4) und durch eine anschließende PCR-Reaktion amplifiziert (Kapitel 2.8.2, 2.8.3). Es zeigte sich sehr schnell, dass eine direkte Amplifikation der revers transkribierten Produkte nur in sehr wenigen Fällen und nur bei bestimmten Primerkombinationen erfolgreich war. Um dieses Problem zu umgehen, wurde die „Nested“-PCR-Technik (Kapitel 2.8.4) bei fast allen Amplifikationen angewendet. So war es möglich, Amplifikate bei allen untersuchten Geweben zu erhalten. Für die Synthese der cDNA wurden aufgrund der „Nested“-Strategie 63 Primer benötigt. Untersucht wurden Gehirn, Lunge, Leber, Magen, Darm, Herz, Niere, Milz, Testis/Ovar, Knochen, Muskel

und embryonale Stammzellen aus der Linie 129-R1 (Daten nicht gezeigt). Diese Befunde decken sich mit EST-Daten der öffentlichen Datenbanken. Ein weites Expressionsspektrum von *Dmxl*1 (Kraemer et al., 1998, 2000; Weil, 2000), das viele Gewebetypen einschließt, konnte so bestätigt werden. In diesem Zusammenhang wurde vielfach vergeblich versucht, das Transkript von *Dmxl*1 in den verschiedenen Geweben durch eine Northern-Hybridisierung nachzuweisen. Da hier unterschiedliche Northern-Variationen ohne Erfolg angewendet wurden, kann dies in Zusammenhang mit der meist notwendigen „Nested"-Strategie als Hinweis gewertet werden, das *Dmxl*1 auch in der Maus nur schwach exprimiert wird. Für die komplette Darstellung der cDNA wurde Total-RNA bzw. Poly $(A)^+$ RNA der embryonalen Stammzelllinie 129-R1 verwendet. Ergänzend wurden vier cDNA-Klone (IMAGp998B022563Q2, IMAGp998E173005Q2, IMAGp998J193130Q2, IMAGp998M019196Q2 (RZPD, I.M.A.G.E. Consortium)) sequenziert, die aus dem 3'-UTR stammen. Die zugehörigen Klone wurden vollständig sequenziert und überlappen sowohl miteinander als auch mit der synthetisierten cDNA. Mit Hilfe der synthetisierten überlappenden cDNA-Fragmente und den ergänzenden cDNA-Klonen konnte die Sequenz einer „full-length" cDNA von 10991 bp bzw. 12210 bp Länge erstellt werden. Die ermittelte Sequenz umfasst 94 bp des 5'-UTR, den vermutlichen Translationsstartpunkt, einen 9042 bp ORF (offener Leserahmen), der für eine abgeleitete Sequenz von 3014 Aminosäuren codiert, sowie 1855 bp bzw. 3073 bp des 3'-UTR der zwei alternative Polyadenylierungsstellen enthält (siehe Anhang, Abb. A1). Ein Vergleich der ermittelten cDNA mit der veröffentlichten genomischen Sequenz ergab, dass das murine *Dmxl*1 von mindestens 43 Exons codiert wird. Alternativ gespleißte Exons konnten nicht identifiziert werden. Eine umfassende Analyse dieser Sequenzdaten findet sich in Kapitel 3.2, ein Sequencher-Projekt der Sequenzierungsreaktionen befindet sich auf dem beigefügten Datenträger. Die ermittelte cDNA Sequenz wurde am Ende der Fertigstellung der vorliegenden Arbeit in den öffentlichen Datenbanken publiziert. Der Eintrag kann unter der Nummer (Accession-Number) AY590892 aufgerufen werden. Die abgeleitete Aminosäuresequenz von *Dmxl*1 ist unter der Nummer AAT01618 in den Proteindatenbanken vermerkt.

3.1.4 Konstruktion des Targeting-Vektors

3.1.4.1 Auswahl der Targeting-Sequenzen

Nach der Sequenzierung von Cosmid N10467Q3 konnte eine nähere Analyse der Sequenzdaten durchgeführt werden (Kapitel 3.5.1), um eine Auswahl der Targeting-Sequenzen zu treffen. Um eine hohe Effizienz der homologen Rekombination zu gewährleisten, sollten die Targeting-Sequenzen folgende Eigenschaften aufweisen:

- Die Länge der beiden Targeting-Arme sollte zusammen ca. 10-12 kb (Thomas et al., 1987; Shulman et al., 1990; Deng, 1992; Blessing, persönliche Auskunft) betragen.
- Um zu verhindern, dass nach dem Targeting-Ereignis ein funktionales bzw. teilfunktionales Protein entsteht, sollten die auszuwählenden Sequenzabschnitte

möglichst weit im 5'-Bereich des Gens liegen oder, bei bekannter Proteinstruktur, die funktionalen Bereiche des Proteins treffen.
- Es müssen neben geeigneten Klonierungsschnittstellen auch Restriktionsschnittstellen gefunden werden, die eine spätere Detektion der endogenen Bande (Abb. 3.9) nach erfolgter homologer Rekombination ermöglichen.
- Es sollten wenige bzw. keine repetitiven Elemente in den Targeting-Sequenzen enthalten sein, da diese eine unspezifische Integration in das Genom des Zielorganismus fördern und die Anzahl der spezifischen Rekombinationsereignisse stark herabsetzen würden.
- Ergänzend kann durch das Rekombinationsereignis eine Deletion einiger kodierender Abschnitte erfolgen, um bei aberranten Splicevorgängen (Abb 4.1) durch gezielte Leserasterverschiebung ein funktionsuntüchtiges Protein, bzw. ein Stop-Codon in der resultierenden mRNA zu erzeugen.
- Die flankierenden Sequenzen außerhalb von mindestens einem der beiden Targeting-Arme müssen sich für die Konstruktion einer Sonde eignen, die in einem genomischen Southernblot spezifisch die endogene Bande nachweisen.

Zum Zeitpunkt der Targeting-Vektor Konstruktion waren keine Einträge des entsprechenden genomischen Abschnitts in den öffentlichen Datenbanken vorhanden und die Klon-Bibliothek der später isolierten PAC-Klone noch nicht verfügbar. Die Sequenzanalyse von Cos-N10467Q3 zeigte, dass sich die vorgegebenen Kriterien nicht alle einhalten ließen. Eine besonders ungünstige Verteilung der Restriktionsschnittstellen machte eine direkte Klonierung über die Schnittstellen von potentiellen Targeting-Sequenzen unmöglich. Weiter wurde eine hohe Dichte von repetitiven Elementen in der Sequenz von Cos-N10467Q3 beobachtet (Kapitel 3.5.2). Für die Klonierung wurden schließlich zwei Sequenzabschnitte ausgewählt, die bereits über *Hind* III-Schnittstellen in pUC 18 kloniert vorlagen (Kapitel 3.2.1). Subklon K2 enthält ein 4347 bp großes Integrat, mit zwei kodierenden Abschnitten von *Dmxl*1 (Exon 8, 9), das Integrat von Subklon K7 ist 6786 bp lang und enthält drei Exons (Exon 15,16,17). Die beiden Sequenzabschnitte flankieren einen 6594 bp Abschnitt, der die Exons 10-14 enthält und bei erfolgreicher homologer Rekombination deletiert wird (Abb. 3.6). Das Konstrukt wurde so gewählt, dass ein aberranter Splicevorgang, bei dem Exon 9 auf Exon 15 gespleißt wird, zu einer Leserasterverschiebung führt und damit ein frühes Stop-Codon erzeugt wird. Eine Integration in ein Exon und somit eine direkte Disruption des Leserahmens war aufgrund fehlender Schnittstellen nicht möglich. In den beiden ausgewählten Tageting-Sequenzen gibt es keine großen und komplexen repetitiven Elemente, jedoch einige kurze Abschnitte, die mit dem Programm REPEATMASKER2 (Kapitel 2.11) als „simple repeats" identifiziert wurden (Kapitel 3.5.2). Ein völliger Ausschluss repetitiver Sequenzen war nicht möglich.

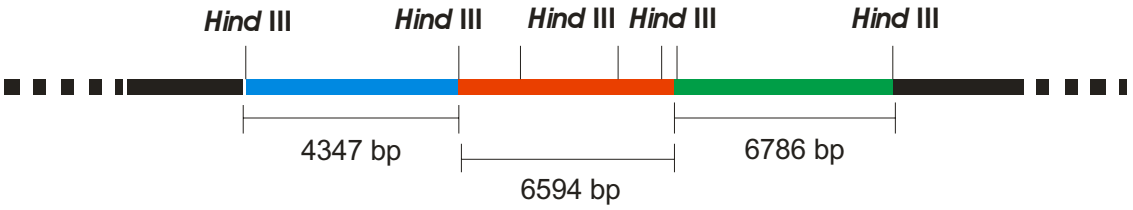


Abb 3.6: Genomische Organisation der ausgewählten Targeting-Sequenzen

Dargestellt ist ein Abschnitt aus Cos-N10467Q3, blau und grün unterlegt sind die Sequenzabschnitte, die als Subklone K2 und K7 bereits in pUC 18 über Hind III Schnittstellen kloniert vorlagen. Der rot gezeichnete Bereich würde durch homologe Rekombination deletiert und durch die pPGKNeo^{R}A^{+}-Kassette ersetzt (Kapitel 3.3.4).

3.1.4.2 Klonierung der Targeting-Sequenzen

Da eine Klonierung über komplementäre Schnittstellen des Targeting-Vektors nicht mit beiden Targeting-Armen möglich war, wurde aufgrund seiner Größe und der damit erwarteten schwierigeren Klonierung das Integrat von K7 ausgesucht, um über die kompatible *Hind* III –Schnittstelle in die erste MCS des Targeting Vektors kloniert zu werden. Weil es sich um das 3'-Fragment der ausgewählten Targeting-Sequenzen handelte, musste die Klonierung in entgegengerichteter Orientierung zur pPGKNeo^{R}A^{+}-Kassette erfolgen (Abb 3.7a). Hierzu wurde das Integrat von K7 mit *Hind* III herausgeschnitten und in den ebenfalls mit *Hind* III restringierten Targeting-Vektor ligiert. Das Screening erfolgte im Hochdurchsatzverfahren (Kapitel 2.5.2). Unter ~25000 getesteten Klonen wurden fünf gefunden, die eine Integration in richtiger Orientierung aufwiesen (Abb. 3.7a). Einer der Klone wurde aufgearbeitet und die Orientierung des Integrats durch Sequenzierung verifiziert. Die Klonierung des Integrats K2 erwies sich aufgrund der ungünstigen Schnittstellenverhältnisse als wesentlich komplexer, eine schematische Darstellung der Klonierung ist in Abbildung 3.7b gezeigt. Zunächst wurde das Integrat von Subklon K2 mit den Enzymen *Sal* I und *Nde* I herausgeschnitten und der gesamte Ansatz gefällt. Anschließend erfolgte eine weitere Restriktion mit *Sma* I, die ein 19 bp Fragment aus der multiplen Klonierungsschnittstelle des verbliebenen pUC 18 herausschneiden sollte. Dieses Fragment sollte dann durch eine Aufreinigung mit dem „QIAquick PCR Purification Kit" gezielt verloren gehen. Eine Intramolekulare Reaktion des Vektors und Religation des Integrats in den Vektor sollte damit verhindert werden. Der Targeting-Vektor mit integriertem Fragment K7 wurde mit *Sal* I und *Nhe* I geschnitten und ebenfalls mit Hilfe des „QIAquick PCR Purification Kit" aufgereinigt, um eine intermolekulare Reaktion mit dem herausgeschnittenen Abschnitt der MCS 2 zu verhindern.

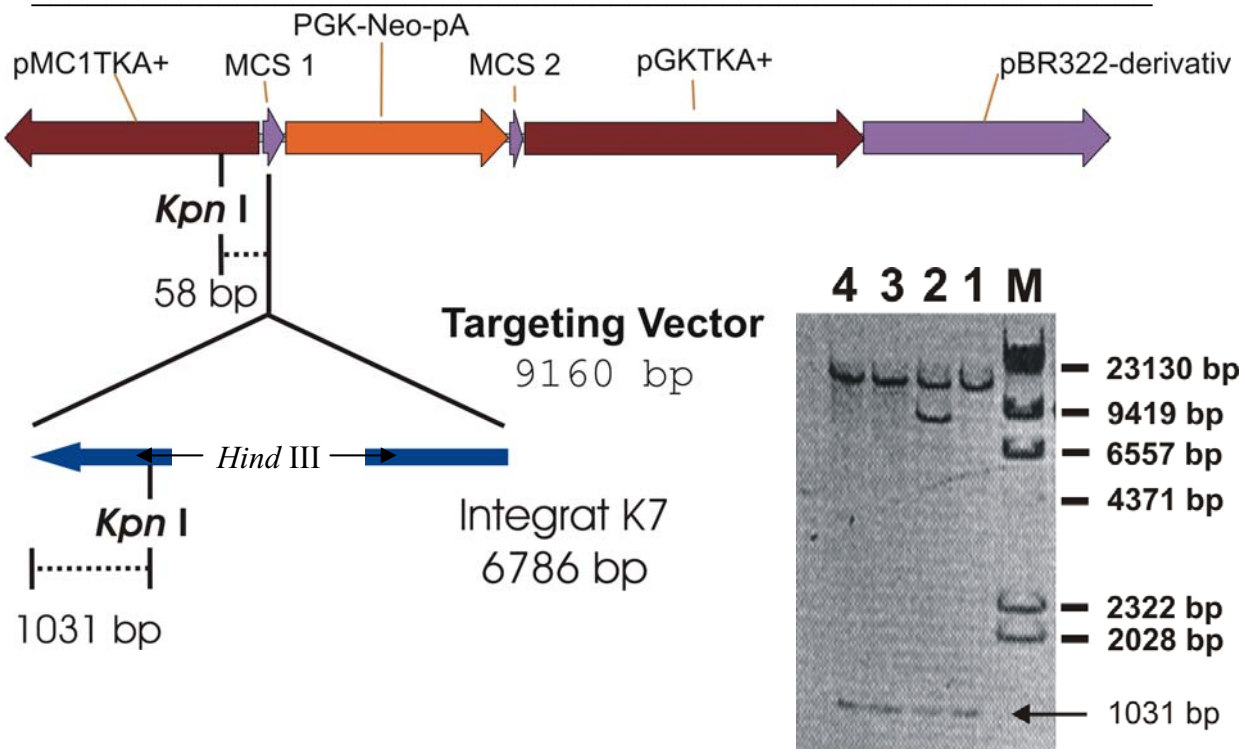


Abb. 3.7a: Restriktionsanalyse nach Klonierung von Integrat K7

Durch Klonierung des Integrates von Subklon K7 in entgegengerichteter Orientierung zur pPGKNeo^{R}A^{+}-Kassette entsteht bei Restriktion des Konstruktes mit Kpn I ein 14857 bp-Fragment aus Vektor und Teilen des Integrats K7 sowie ein 1089 bp-Fragment. Bei falscher Orientierung entstehen zwei Fragmente mit 5755 bp und 10191 bp. Die Gesamtgröße von Integrat und Vektor beträgt 15946 bp. Auf dem Gelfoto sind die Restriktionsanalysen mit Kpn I von positiven Klonen in den Spuren 1, 3 und 4 zu erkennen, Spur 2 zeigt eine Kontamination mit leerem Vektor.

Beide Ansätze wurden dann über die *Sal* I Schnittstelle und glatte Enden ligiert. Durch diese Strategie wurde auch ein 216 bp Anteil des pUC Vektors (Basenposition 237-418 in pUC 18) in den Targeting Vektor überführt. Anschließend erfolgten die Transformation und ein erneutes Screening im Hochdurchsatzverfahren. Hierbei wurden ~35000 Klone untersucht, es konnte nur ein positiver Klon identifiziert und durch Sequenzierung verifiziert werden (Abb. 3.8). Eine schematische Darstellung des Targeting-Konstruktes ist in Abb. 3.9 zu sehen. Anschließend wurde von dem Konstrukt eine endotoxinfreie DNA-Präparation hergestellt (Kapitel 2.3.1) und mit *Not* I geschnitten. Durch diese Restriktion wurde der pBR 322-Anteil des Targeting-konstruktes herausgeschnitten und der Vektor linearisiert (Mansour et al., 1988).

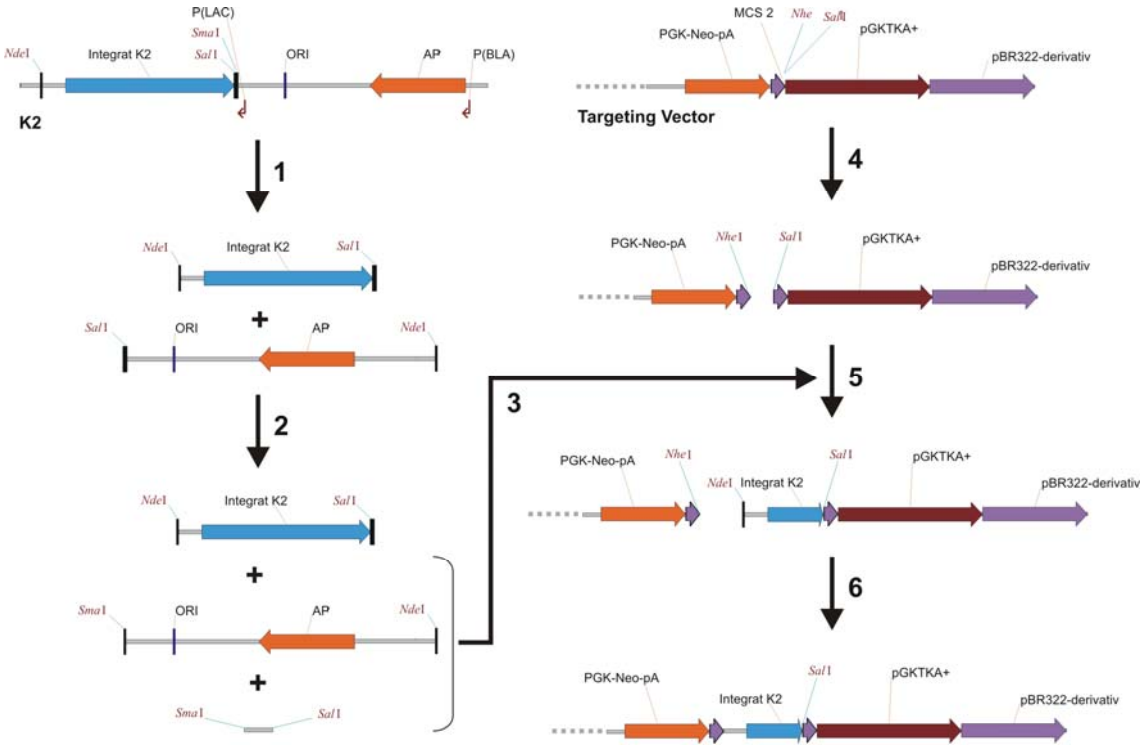
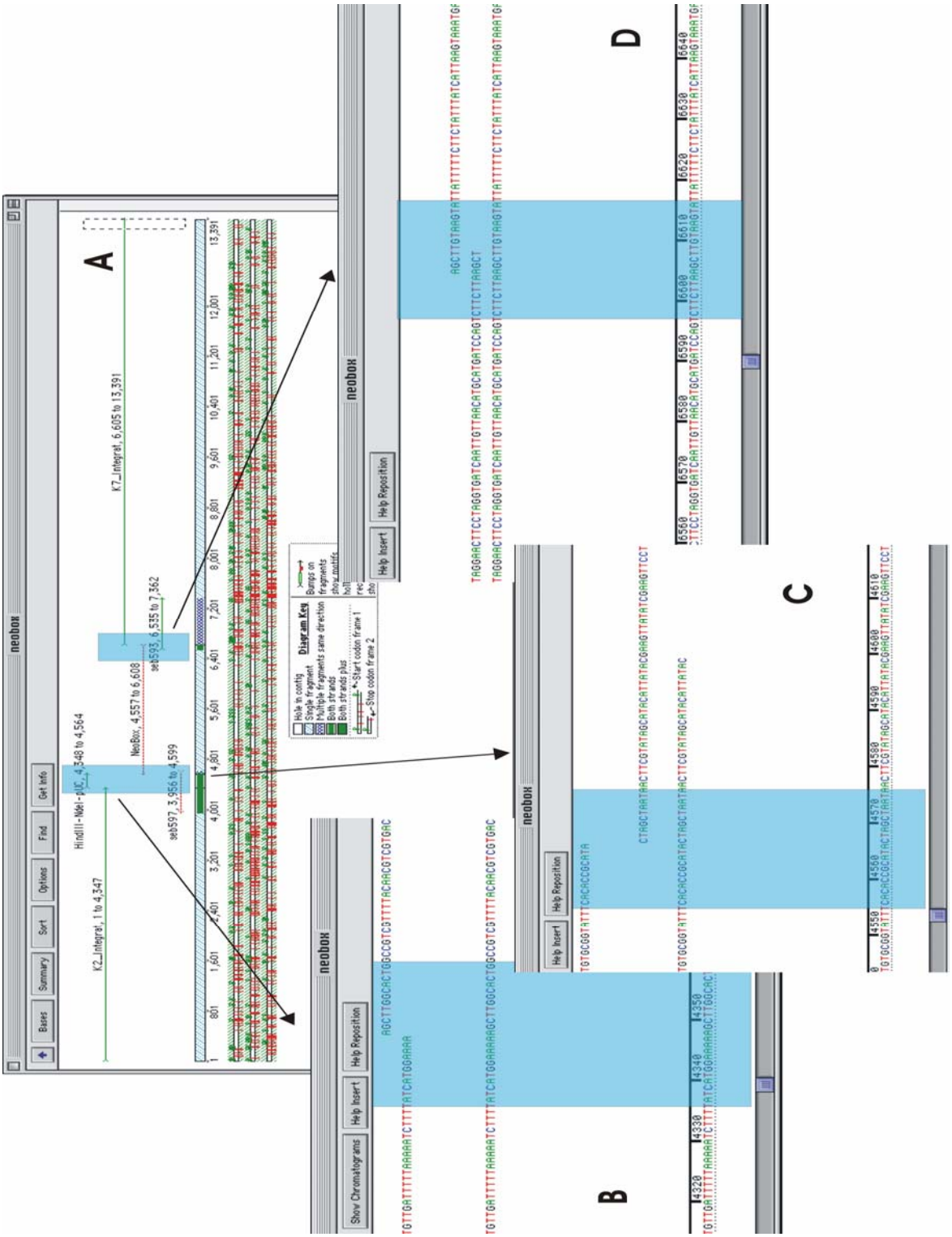


Abb. 3.7b: Klonierungsstrategie des zweiten Targeting-Arms

1: Restriktion von Subklon K2 und Excision des Integrats mit den Enzymen *Sal* I und *Nde* I mit anschließender Aufreinigung.
2: Restriktion des Gemisches aus Integrat K2 und pUC18 mit dem Enzym *Sma* I und Excision eine 19 bp Fragments aus pUC18
3: Aufreinigung des Restriktionsansatzes mit Hilfe des „QIAquick PCR Purification Kit"
4: Restriktion des Targetingvektors mit Integrat K7 durch die Enzyme *Sal* I und *Nhe* I
5: Ligation des Gemisches aus Integrat K2 (*Sal* I/*Nde* I), pUC18 (*Nde* I/*Sma* I) und Targetingvektor mit Integrat K7 (*Sal* I/*Nhe* I).
6: Auffüllung der inkompatiblen Schnittstellen *Nde* I und *Nhe* I, Aufreinigung dieses Reaktionsansatzes und anschließende Ligation der glatten Enden

Abb. 3.8: Ergebnisse der Sequenzierungen nach der Klonierung in den Targeting-Vektor, dargestellt als Sequencher-Projekt (folgende Seite).

A: Übersicht der Klonierungsstrategie und Ergebnisse der Sequenzierung. Zentral ist die pPGKNeo^{R}A^{+}-Kassette in umgekehrter Orientierung zur Leserichtung von *Dmxl* 1 zu sehen (roter Pfeil mit Beschriftung NeoBox). Das Integrat K7 wurde über eine *Hind* III Schnittstelle einkloniert, das Integrat K2 wurde zunächst mit *Sal* I und *Nde* I aus dem Subklon K2 herausgeschnitten und dann über die *Sal* I Schnittstelle und glatte Enden kloniert. Die Sequenzierungen über die entscheidenden Bereiche sind mit sebx 597 und sebx 593 bezeichnet.
B: Sequenzierung in das Integrat K2. Ursprünglich wurde K2 über die *Hind* III Schnittstelle (AAGCTT) in pUC 18 kloniert, der Übergang des durch die Klonierungsstrategie verbliebenen Vektoranteils zu Integrat K2 ist im Sequenzierungsergebnis (untere Sequenz) zu erkennen.
C: Sequenzierung des Übergangs aus der pPGKNeo^{R}A^{+}-Kassette in den klonierten Vektoranteil von pUC 18. Durch die Klonierungsstrategie wird die *Nde* I Schnittstelle (CATATG) nicht regeneriert, ist jedoch fragmentarisch (CATA) zu erkennen.
D: Die Klonierung von K7 erfolgte über *Hind* III Schnittstellen der MCS 1 des Targeting Vektors, auch hier zeigte die Sequenzierung einen korrekten Übergang von Vektor zu Integrat K7.


neobox
Bases
Summary
Sort
Options
Find
Get Info
A
K2_integral, 1 to 4,347
HindIII - NheI - pUC, 4,348 to 4,564
NeoBox, 4,557 to 6,608
seb597, 3,956 to 4,599
seb593, 6,535 to 7,362
K7_integral, 6,605 to 13,391
Diagram Key
Help Reposition
Help Insert
Show Chromatograms
B
C
D

3.1.5 *Dmxl*1: Gen-Targeting

3.1.5.1 Transformation des Targeting-Konstruktes

Alle Arbeiten an embryonalen Stammzellen wurden in Kooperation mit Univ.-Prof. Dr. rer. nat. Manfred Blessing, I. Med. Klinik und Poliklinik, Abteilung Pathophysiologie im Verfügungsbau des Klinikums der Johannes Gutenberg-Universität Mainz durchgeführt.

Für die Transformation der ES-Zellen des Typs 129- R1 wurden 150 µg des linearisierten Targeting-Vektors eingesetzt. Zunächst erfolgten die DNA-Präparationen des Vektors mit Hilfe der Whitehead-Methode (Kapitel 2.3.2). Mit dieser DNA wurden drei Transformationsversuche durchgeführt, die Effizienz war sehr niedrig (Blessing, persönliche Mitteilung). Um eine höhere Effiziens der Transformationen zu erreichen, wurde für die vierte Transformation der Targeting-Vektor mit Hilfe des EndoFree® Plasmid Maxi Kit der Firma Qiagen nach Angaben des Herstellers isoliert, die Transformation der embryonalen Stammzellen erzeugte hier 391 Klone. Zum Ende der vorliegenden Arbeit wurde eine fünfte Transformation durchgeführt, wobei 395 Klone isoliert werden konnten. Alle Transformationen und Selektionsschritte erfolgten mit den gleichen Parametern (Kapitel 2.11.2, 2.11.3).

3.1.5.2 Detektion der endogenen Bande durch Restriktion und Southern-Analysen

Um eine erfolgreiche homologe Rekombination der klonierten Targeting-Sequenzen mit den entsprechenden genomischen Sequenzabschnitten nachzuweisen, wurden Restriktionsanalysen der genomischen DNA aus ES-Zellen mit verschiedenen Enzymen durchgeführt. Durch die Integration der pPGKNeo^{R}A^{+}-Kassette in das Genom wurden zusätzliche Schnittstellen eingefügt, die ein zum Original divergentes Restriktionsmuster erzeugen und durch radioaktiv markierte Sonden nachgewiesen werden können. Bei der endogenen Bande handelt es sich um das wildtypische Restriktionsfragment, das durch eine Sonde sichtbar gemacht wird (Abb. 3.9). Ein sicherer Nachweis der endogenen Bande konnte nur mit Sonden erfolgen, die außerhalb der rekombinanten Sequenzabschnitte spezifisch binden. Hierdurch wird die Detektion zufällig in das Genom intergrierter Targeting-Konstrukte mit ungefähr gleich langen Restriktionsbanden verhindert. Da zu diesem Zeitpunkt lediglich die Sequenzinformationen von Cosmid N10467Q3 zur Verfügung standen, mussten das Sonden-Design und die Auswahl der Restriktionsenzyme auf dieser Basis erfolgen. Es konnten nur Enzyme verwendet werden, die insensitiv gegenüber Methylierungsereignissen in den entsprechenden Palindromsequenzen waren.

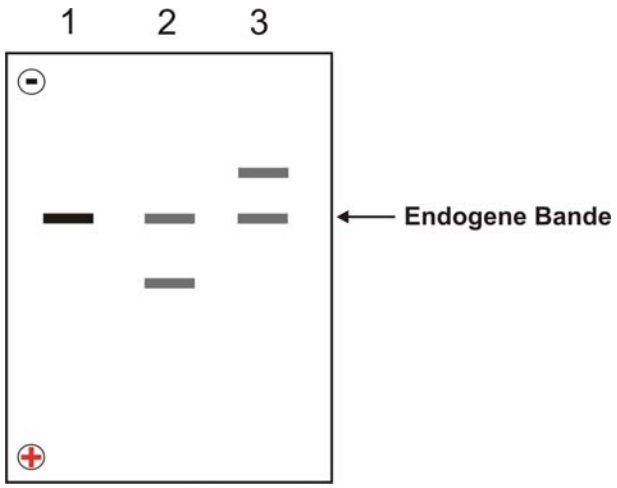


Abb. 3.9: Endogene Bande

Gezeigt ist das zu erwartende Restriktionsmuster nach erfolgreichem Einbau des Konstrukts. In Spur 1 ist der Wildtyp gezeigt, beide Allele erzeugen das gleiche Restriktionsmuster. In Spur 2 und 3 sind neben der endogenen Bande weitere Banden zu erkennen, deren Restriktionsmuster durch die Manipulation eines der beiden Allele auf genomischer Ebene erreicht wurde. Dies kann zu einer größeren oder kleineren Bande führen.

Der „linke“ Targeting-Arm umfasst 4347 bp. Für die Sondenkonstruktion wurde unzählige Varianten von Restriktions/Sonden- Systemen an embryonalen Stammzellen gestestet, erzeugten jedoch immer unspezifische Signale. Erst die Verfügbarkeit der PAC-Klonbibliothek Nr. 711 und die nachfolgende Identifikation positiver PAC-Klone (Kapitel 3.1.2) ermöglichte den Test von weiteren Restriktions/Sonden-Systemen. Durch Interspeziesvergleich mit der cDNA-Sequenz des humanen *DMXL*1 konnte ein Klon (4.206) in der „shotgun“-Bank identifiziert werden, der das vorausgehende Exon zu den kodierenden Sequenzen in Fragment K2 enthielt (Kapitel 3.3.1) und 4796 bp entfernt zu K2 liegt. Die Integratlänge von Klon 4.206 beträgt 1897 bp. Im Anschluss wurde durch Restriktions- und Southernanalysen ein Enzym (*Eco R*V) identifiziert, das vor der ausgewählten Sondensequenz schneidet und einen detektierbaren Unterschied nach homologer Rekombination zum Wildtyp erzeugt. Nach Angaben von Einträgen in öffentlichen Datenbanken würde der Wildtyp nach Restriktion genomischer DNA aus embryonalen Stammzellen des Typs 129 ein 19059 bp Fragment zeigen. Eine erfolgreiche homologe Rekombination führt eine Schnittstelle ein, die nach Restriktion ein 18044 bp–Fragment zeigen sollte (Abb 3.10). Allerdings befand sich bis zur Fertigstellung der vorliegenden Arbeit in der Sequenz des Restriktionsfragments eine Lücke unbekannter Größe. Eigene Analysen durch Southernblot und anschließende Auswertung der Molekulargewichte mit dem digitalen EDAS-System (Kodak) zeigten eine Hybridisierung der *Eco R*V-Bande eher im Bereich zwischen 20300-20600 bp (Abb.3.11). Es ist daher davon auszugehen, dass die Lücke eine Länge von 1,3-1,6 kb aufweist. Um einen Unterschied zwischen Wildtyp-Allel und „Knock-Out“-Konstrukt von 1015 bp in diesem Molekulargewichtbereich überhaupt detektieren zu können,

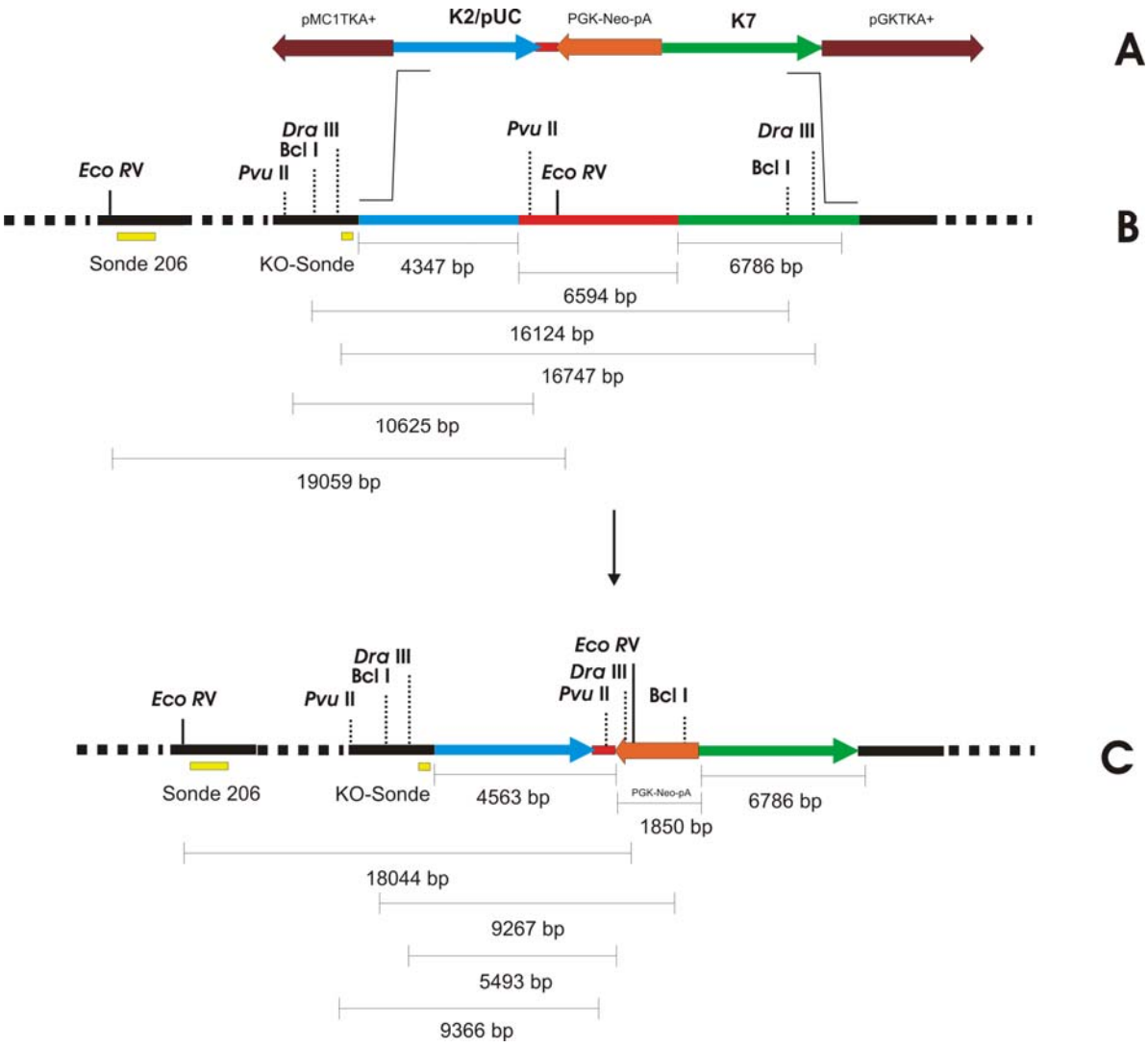


Abb 3.10: Schematische Darstellung der homologen Rekombination mit dem Targeting-Konstrukt .

A: Finale Struktur des Targeting-Vektors. In blau (K2) und grün (K7) dargestellt sind die beiden *Dmxl1*-Arme, die später an der homologen Rekombination beteiligt sind. Der rot markierte Abschnitt zeigt den pUC18-Anteil, der durch die Klonierungsstrategie mit in das Konstrukt übertragen wurde (Kapitel 3.3.2). Die pPGKNeo^{R}A^{+}-Kassette positioniert sich entgegen der Orientierung von *Dmxl* 1 und ist als oranger Pfeil gezeigt. Flankierend befinden sich die pMC1TKA^{+}-Kassette (Mansour, 1988) in entgegengerichteter transkriptioneller Orientierung zur pPGKNeo^{R}A^{+}-Kassette sowie eine pGKTKA^{+} -Kassette in gleicher transkriptioneller Orientierung wie die davor liegende pPGKNeo^{R}A^{+}-Kassette (Rudnicki, 1992). Beide TK-Kasetten werden bei einer erfolgreichen homologen Rekombination nicht in das Genom integriert (Kapitel 2.4.3).
B: Genomische Organisation der Zielsequenz im Wildtyp. In blau und grün dargestellt sind die Sequenzabschnitte, die für die Konstruktion des Tartgeting-Vektors subkloniert wurden und die Paarungssequenz für eine homologe Rekombination bilden. In gelb dargestellt sind die verwendeten Sonden und ihre Position relativ zu den Schnittstellen der verwendeten Enzyme. Die Balken geben schematisch die Länge der Restriktionsfragmente der verwendeten Enzyme im Wildtyp wieder.
C: Genomische Organisation in transformierten embryonalen Stammzellen nach erfolgter homologer Rekombination. Gezeigt sind die Positionen der Schnittstellen und die daraus resultierenden veränderten Größen der Restriktionsfragmente nach Integration der pPGKNeo^{R}A^{+}-Kassette. Erfolgreiche Restriktio-

nen mit sicher detektierbaren Signalen im Southernblot konnten nur mit dem Enzym *Eco R*V in Kombination mit Sonde 206 durchgeführt werden. Die Angaben der Fragmentgrößen des *Eco R*V-Verdaus vor und nach homologer Rekombination basieren auf den Einträgen in den öffentlichen Datenbanken. Wahrscheinlicher sind allerdings Restriktionsfragmente, die 1,2-1,5 kb größer als die angegebenen Werte (Kapitel 3.3.4) sind. Der Unterschied von 1015 bp nach der homologen Rekombination wird dadurch jedoch nicht beeinflusst.

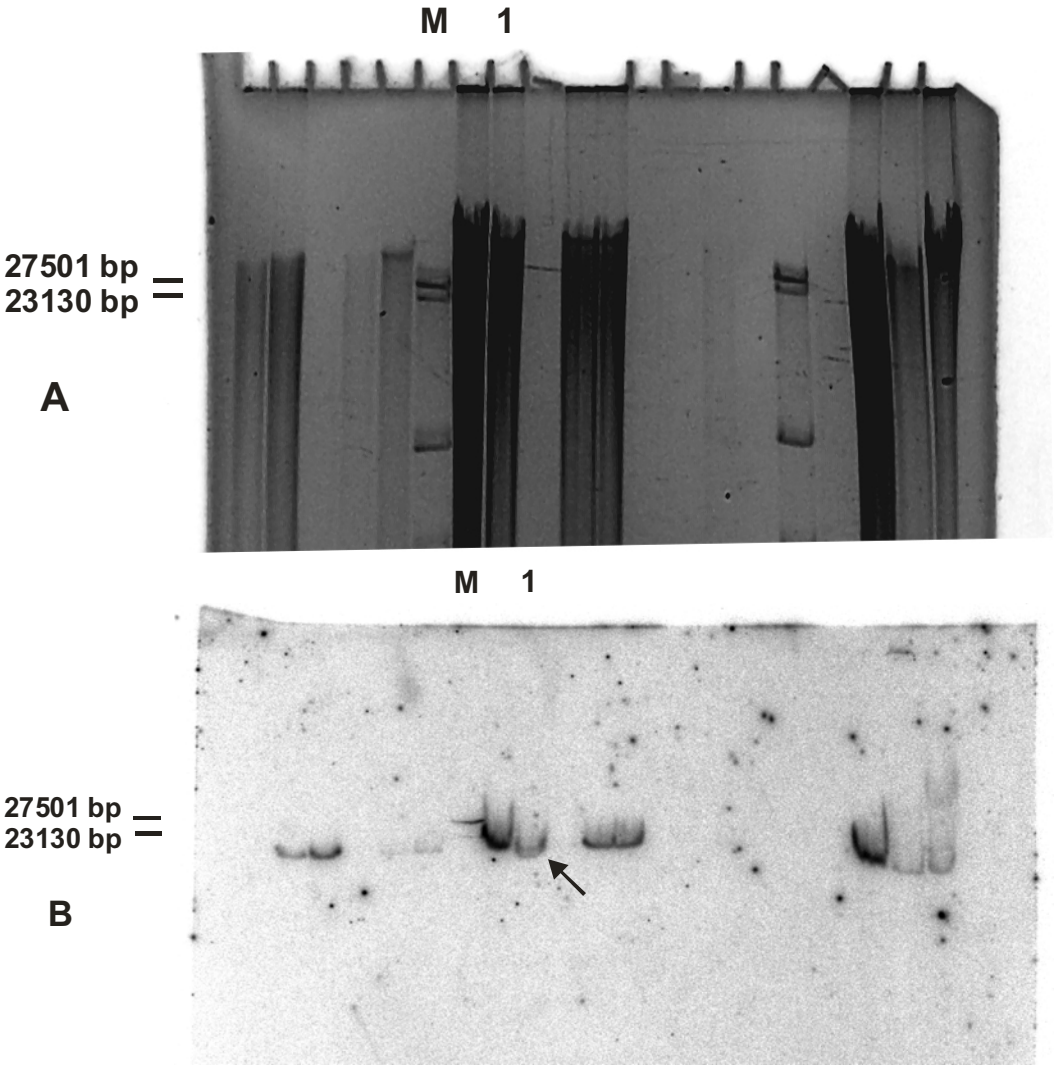


Abb. 3.11: Detektion der positiv transformierten Stammzelllinie 4.P14-D4

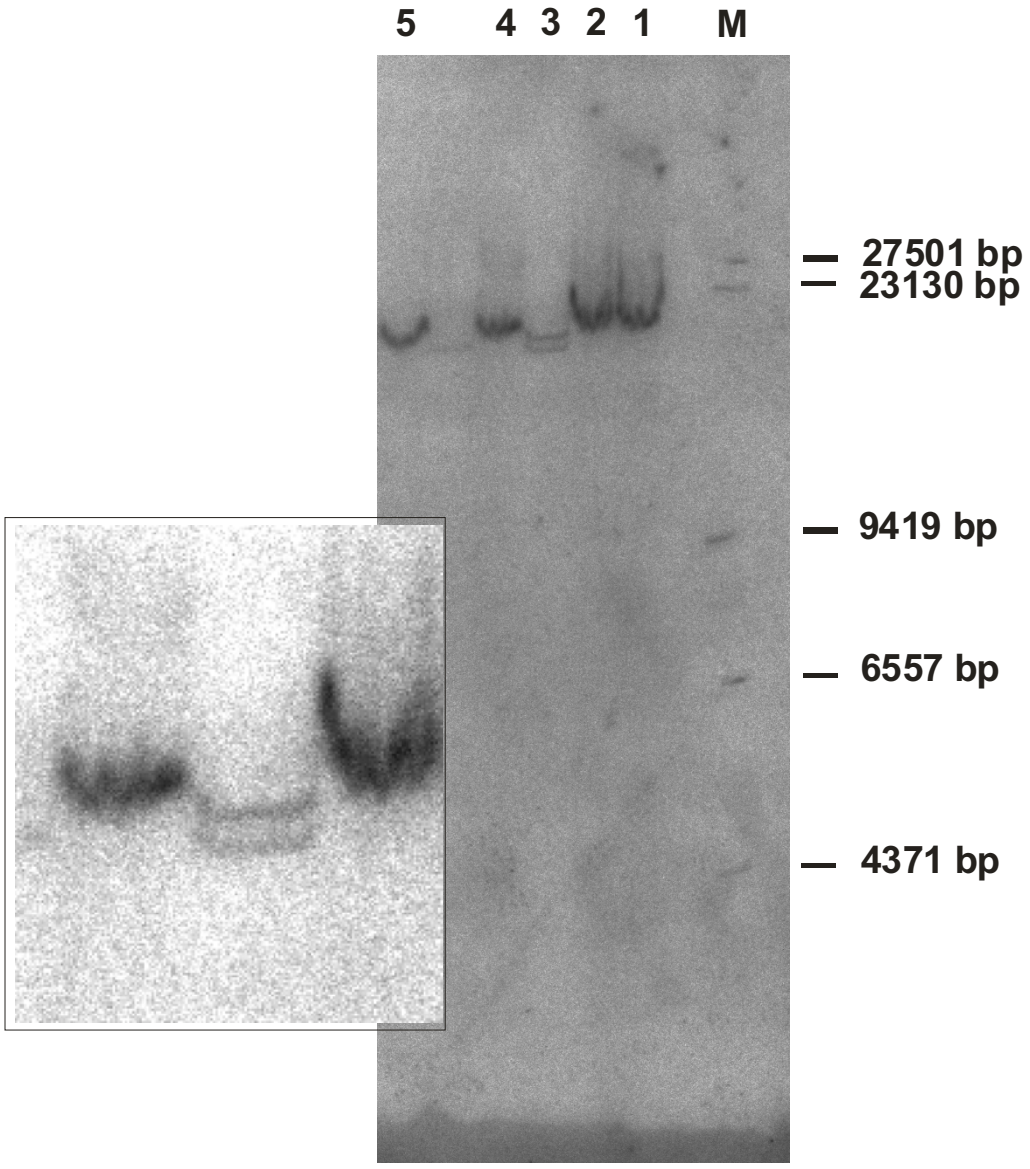
Dargestellt ist eine typische Southern Analyse embryonaler Zellklone. Die digitalen Bilder, die durch den BAS 1800 II (Raytest) erzeugt wurden, konnten direkt für weitere Analysen mit dem EDAS-System (Kodak) verwendet werden.
A: Hochauflösendes Agarosegel mit der Zelllinie 4.P14-D4 (1). Die obere Bande des Markers (M) repräsentiert 27501 bp. Das Gel wurde an mehreren Positionen mit der isolierten DNA der Zelllinien deutlich überladen.
B: Southernblot der Zelllinie 4.P14-D4 (1), detektiert mit einer Imager-Platte. Obwohl das Gel in diesem Bereich überladen wurde, kann eine Doppelbande erkannt werden (Pfeil). Der starke „Smilie"-Effekt hat keine Auswirkungen auf die Detektion der endogenen Bande. Eine Auswertung mit dem digitalen EDAS-

System (Kodak) zeigte, das die beiden Fragmente („Allele“) sich in ihrer länge um ca. 1000 bp unterschieden. Dies entspricht exakt den Erwartungen.

mussten hochauflösende Agarosegele und ein speziell auf diese Bedingungen abgestimmtes Restriktions-, Markierungs- und Hybridisierungssystem verwendet werden (Kapitel 2.6). In der vierten Transformation embryonaler Stammzellen konnte schließlich mit dem oben dargestellten Restriktions-/Sondensystem aus *Eco R*V und Sonde 206 eine erfolgreich transformierte Stammzelllinie detektiert werden (Abb. 3.11, 3.12). Die Zelllinie 4.P14-D4 zeigte die erwartete Doppelbande. Eine Analyse der markierten Restriktionsfragmente mit dem digitalen EDAS-System (Kodak) bestätigte eine Übereinstimmung mit dem theoretisch ermittelten Zielwert innerhalb der Fehlertoleranzen des Auswertungssystem, bedingt durch die Lauf- und Trenneigenschaften der Agarosegele für Restriktionsfragmente sehr hoher Molekulargewichte (Abb.3.12).

Abb 3.12: Southernblot der Zelllinie 4.P14-D4 (folgende Seite)

Nach Anzucht der vermutlich transfizierten Zelllinie 4.P14-D4 (3) wurde die DNA wie unter 2.4.3 beschrieben aufgearbeitet, restringiert und erneut ein hochauflösender Southernblot hergestellt. Als Vergleich wurden Stammzellpräparationen ohne erfolgreiche homologe Rekombination parallel analysiert (1,2,4,5). Für die Auftrennung wurde hier ein 0,7%iges Gel verwendet. Durch die optimale Beladung und die verbesserte Auftrennung des Gels konnte die erfolgreiche Transformation und erfolgte homologe Rekombination von 4.P14-D4 nachgewiesen werden, beide Banden zeigen eine etwa gleiche Intensität der Hybridisierung. Eine Analyse mit dem EDAS-System wies den erwarteten Größenunterschied von 1 kb aus (Daten nicht gezeigt).



3.1.5.3 Blastozysten Injektionen

Für diese Arbeit wurden drei unabhängige Blastozysteninjektionen durchgeführt (26.08.2002, 09.09.2002, 16.09.2002; Tab 3.1). Hierzu wurden je 20 (Injektionstermine 1+2) bzw. 28 (Injektionstermin 3) C57/BL/6N-Weibchen eines Alters von 8-12 Wochen hormonell behandelt. Hiervon wurden 11, 14 bzw. 21 Tiere von C57/BL/6N-Männchen gedeckt (Hogan et al. 1994). Der Erfolg wurde über das Vorhandensein eines Kopulationspfropfes verifiziert. Die hier verwendeten Blastozysten wurden dann operativ entnommen, es konnten so je 70 (Injektionstermine 1+2) bzw. 60 (Injektionstermin 3) Blastozysten gewonnen werden. Insgesamt wurden davon je 45 (Injektionstermine 1+2) bzw. 30 (Injektionstermin 3) Blastozysten erfolgreich injiziert. Dabei wurden jeweils ca. 15 ES-Zellen der Zelllinie 4.P14-D4 in jedes Blastocoel einer

Blastozyste implantiert (Abb. 2.6). Es wurden in der vorliegenden Arbeit insgesamt 120 Blastozysteninjektionen vorgenommen. Die Implantation der Blastozysten in die scheinschwangeren Mäuse erfolgte nach der Deckung durch vasectomierte NMRI-Männchen. Die Retransfers wurden auf 5, 4 bzw. 3 NMRI-Empfängerweibchen durchgeführt und jeweils in beide Uterus-Hörner implantiert (Abb. 2.7). Die Isolierung und Injektion von Blastozysten, der Embryotransfer sowie Züchtung und Haltung daraus resultierender Tiere erfolgte in Kooperation mit Dr. Reifenberg (Leiter der Zentralen Versuchstiereinrichtung (ZVTE)). Die Arbeiten wurden im Sicherheitsbereich der Zentralen Versuchstiereinrichtung des Naturwissenschaftlich-Medizinischen Forschungszentrum der Johannes Gutenberg-Universität in Kooperation mit Dr. Reifenberg durchgeführt.

3.1.5.4 *Dmxl*1-„Knock-Out“: chimäre Mäuse

Die in Kapitel 3.3.5 beschriebenen Blastozysteninjektionen erzeugten 8, 7 bzw. 3 Nachkommen, hiervon zeigten insgesamt fünf Tiere einen Chimärismus bis 25 % (2♂/3♀). Ein Tier (♀) zeigte einen Chimäritätsgrad zwischen 25 % und 50 % und bei vier Tieren lag der Chimärismus bei >75 % (alle ♂), ermittelt durch optische Kontrolle der Fellfarbe (Abb 3.13) durchgeführt von Dr. Reifenberg (Leiter der Zentralen Versuchstiereinrichtung (ZVTE) des Naturwissenschaftlich-Medizinischen Forschungszentrum der Johannes Gutenberg-Universität). Der Chimärismus des gesamten Organismus wird bei der Beurteilung anhand des prozentualen Anteils der Fellfarbe „agouti“ ab (Kapitel 2.1.2) hergeleitet. Alle chimären ♂ Tiere wurden nach der Geschlechtsreife mit vielen weiblichen C57/BL/6N Mäusen verpaart. Keines der Tiere transduzierte, d.h. es wurden keine heterozygoten Nachkommen gezeugt und es konnte keine stabile Keimbahntransformation etabliert werden. Die Tiere waren aber eingeschränkt fertil, sie zeugten reinerbige C57/BL/6N-Nachkommen (insgesamt 67 Nachkommen). Die hochchimären Tiere zeugten hier pro Wurf nur 1-3, die niederchimären 6-8 Nachkommen. Für jede Paarung wurden unterschiedliche weibliche Tiere verwendet. Eine komplette Statistik aller Kreuzungen und der entsprechenden Nachkommen kann Tabelle 3.1 entnommen werden. Bei allen Tieren erfolgten die Kreuzungen über mindestens acht Monate, danach wurden die noch lebenden Tiere (acht von 10) eingeschläfert und näher untersucht.

	Injektionstermine			
	26.08.2002	09.09.2002	16.09.2002	Gesamt
C57/BL/6N ♀, hormonell behandelt	20	20	28	68
♀ Tiere mit Kopulationspfropf	11	14	21	46
Anzahl isolierter Blastozysten	70	70	60	200
Anzahl injizierter Blastozysten	45	45	30	120
NMRI-Retransfers	5	4	3	12
Anzahl der Nachkommen	8	7	3	18
davon bis 25 % ♂/♀	1/0	1/3	0/0	2/3
davon 25-50 % ♂/♀	0/1	0/0	0/0	0/1
davon 50-75 % ♂/♀	0/0	0/0	0/0	0/0
davon > 75 % ♂/♀	3/0	1/0	0/0	4/0
Anzahl der Nachkommen nach Rückkreuzung	∑ 67, alle C57/BL/6N			

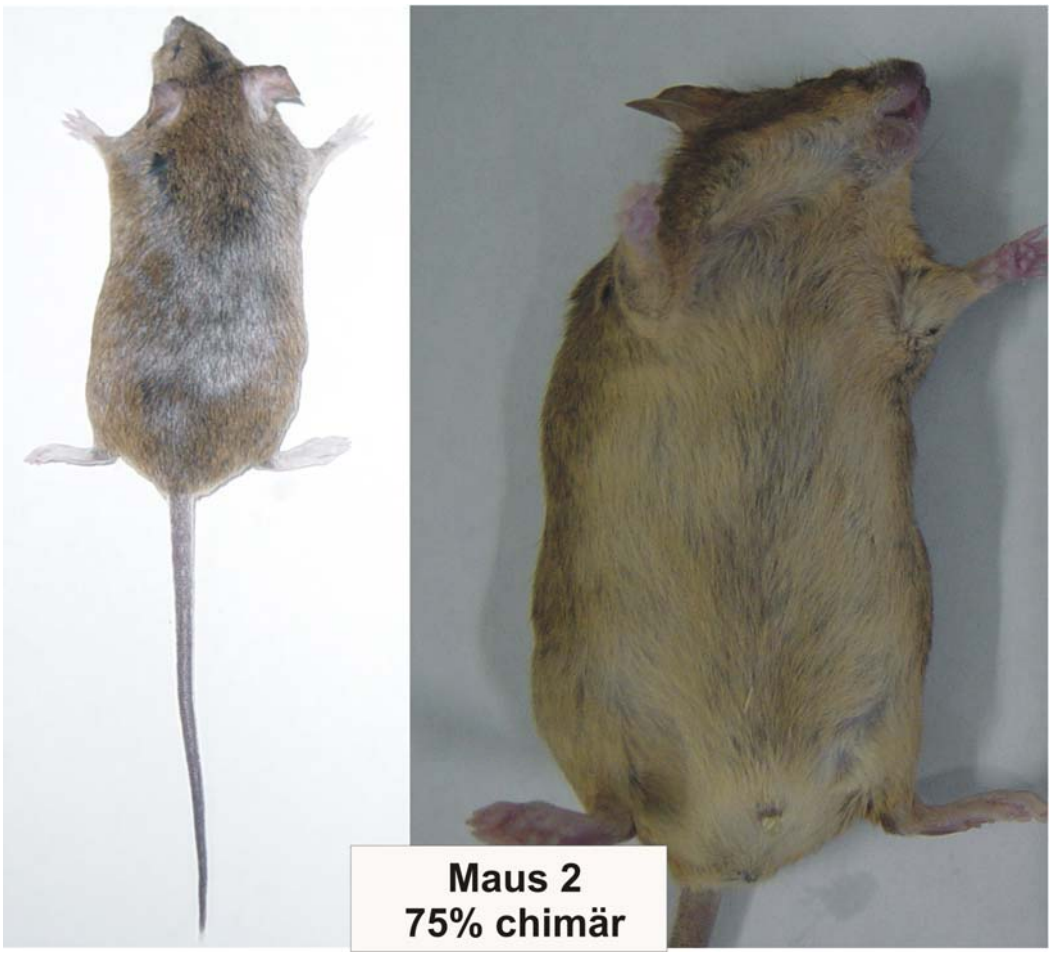
Tab. 3.1: Daten aller Injektionstermine

Zusammenfassung der Ergebnisse der drei Injektionen.

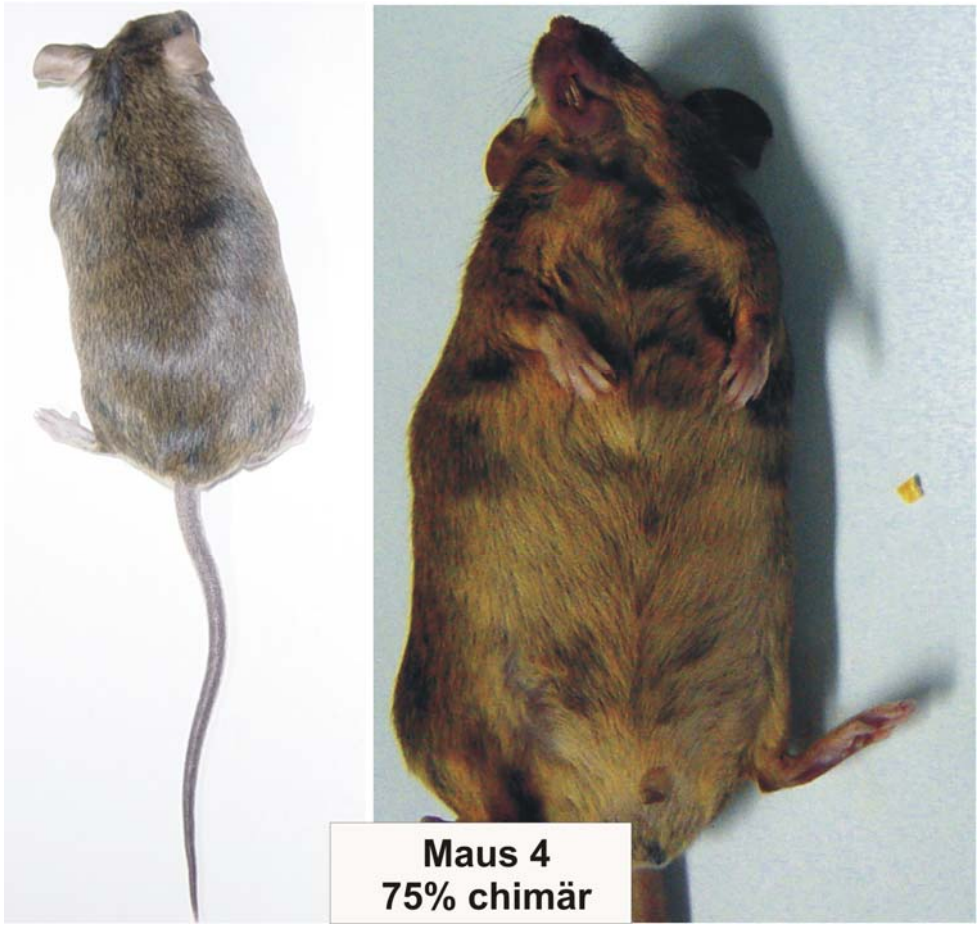
Abb. 3.13: Chimäre *Dmx l*1 „Knock-Out"-Mäuse (folgende Seiten)

Die Fotos zeigen alle untersuchten chimären Tiere in dorsaler bzw. ventraler Orientierung. Die prozentuale Abschätzung des Chimärismus eines Tieres erfolgte aufgrund der gesamten Zusammensetzung der Fellfarben „Schwarz" und „Agouti" in einem Tier. Die Einordnung eines Tieres in eine Gruppe mit definiertem Chimärismusgrad ist grob, die Übergänge in der Gruppe eher fließend.

Maus 1
25% chimär


Maus 2
75% chimär

Maus 3
75% chimär


Maus 4
75% chimär

Maus 5
25% chimär

Maus 6
25% chimär

Maus 7
25% chimär

Maus 8
25% chimär

Die Untersuchung zeigte schon beim bloßen Augenschein Unterschiede und eine deutliche Veränderung. Die Chimären waren größer und schwerer als Tiere des Ausgangsstamms (Abb. 3.13, 3.14). Je höher der Chimärismus, desto fettleibiger wirkten die Tiere. Das Gewicht (Tabelle 3.2) scheint mit dem Grad des Chimärismus zuzunehmen, im Vergleich zum Wildtyp wurde ein Übergewicht von 30-120 % beobachtet. Für den Vergleich wurden statistische Angaben reiner Linien von C57/BL/6 (Mouse Genome Database, Mouse Genome Informatics, Mouse Phenome Database, The Jackson Laboratory, 2003; Bachmanov et al., 1998, 2001; Barsh et al., 2000; Rhees et al., 2000) und 129/Sv (Mouse Genome Database, Mouse Genome Informatics, Mouse Phenome Database, The Jackson Laboratory, 2003; Crabbe et al.; Naveilhan et al., 1999) herangezogen. Zusätzlich dienten die Gewichtswerte eines gemischten genetischen Hintergrunds aus C57/BL/6 und 129/Sv (Dong et al., 1997) als Richtwerte für die chimären Tiere. Die Gewichte von Mäusen sind stark abhängig vom Alter, Geschlecht und Diät der Tiere, die hier angegebenen Durchschnitts- und Vergleichswerte entsprechen näherungsweise dem Alter, Geschlecht, die männlichen Tiere sind in der Regel schwerer als die weiblichen. Die in der Literatur angegebenen Maximalwerte der Lebendgewichte würden bei normaler Fütterung der Tiere, wie es bei den chimären Tieren der Fall war (Dr. Reifenberg, persönliche Mitteilung), nicht erreicht werden. Für die hochchimären Mäuse ergaben sich Werte (Tab 3.3) die nach den Literaturangaben annähernd nur von Tieren erbracht wurden, bei denen gezielt Gene manipuliert wurden, die eine Rolle im Fettstoffwechsel spielen, z.B. „diabetes mouse“, „fat mouse“ oder „tubby mouse“, (Dong et al., 1997; Yaswen et al. 1999; Beattie et al., 1998; Stubdal et al., 2000; Bachmanov et al., 2001; Cohen et al., 2001; Miller et al., 2002;). Als Maximum erzielten homozygote „Knock-Out“-Tiere in den entsprechenden Publikationen eines ObR^{synl}-„Knock-Out“ 58 Gramm Körpergewicht und damit einen mit den hier erzeugten Chimären vergleichbaren Wert. Angaben über Körpergewichte chimärer Tiere dieser Experimente existieren in der Literatur nicht. Die Zahl der chimären Tiere ist leider zu klein, um eine statistisch signifikante Aussage zu treffen. Allerdings ist es die z.T. dramatische Gewichtszunahme und die nicht zu leugnende Beobachtung, die auf einen Zusammenhang mit der Ausschaltung des *Dmxl1*-Gens hinweist.

Tab. 3.2 Korrelation von Chimärismus und Gewicht der untersuchten Tiere

Zusammenfassung von Geschlecht, Chimärismus und gemessenem Lebendgewicht der untersuchten Tiere. Als Referenz sind Durchschnitts-, Mini- und Maximalgewichte aus der Literatur aufgeführt. Die Maximalwerte der C57/BL/6-Tiere und des gemischten genetischen Hintergrunds aus C57/BL/6J / 129/SV wurden durch eine Hochfettdiät erreicht (Dong et al., 1997; Barsh et al., 2000; Rhees et al., 2000) und können so nur bedingt als Referenz gelten.

	Geschlecht	Chimärismus	Gewicht
Maus 1	♀	>25 %	32,3 g
Maus 2	♂	~75 %	47,8 g
Maus 3	♂	>75 %	58,2 g
Maus 4	♂	~75 %	47,6 g
Maus 5	♂	>25 %	33,5 g
Maus 6	♀	~25 %	30,4 g
Maus 7	♂	>25 %	34,1 g
Maus 8	♂	~25 %	33,9 g
∅ C57/BL/6	♂	-	17-33 g, ∅ 26 g
∅ C57/BL/6	♀	-	14-24 g, ∅ 20 g
∅ 129/Sv	♂	-	∅ 28 g
∅ 129/Sv	♀	-	∅ 22 g
C57BL/6J / 129/Sv	♂	-	∅ 36 g ± 1,9 g
C57BL/6J / 129/Sv	♀	-	∅ 33 g ± 1,2 g

	Prozentuales Übergewicht verglichen mit		
	129/Sv	**C57/BL/6**	**C57BL/6J / 129/Sv**
Maus 1 ♀	146 %	161 %	98 %
Maus 2 ♂	170 %	183 %	132 %
Maus 3 ♂	208 %	223,8 %	161 %
Maus 4 ♂	170 %	183 %	132 %
Maus 5 ♂	119 %	128 %	93 %
Maus 6 ♀	138 %	152 %	92 %
Maus 7 ♂	121 %	131 %	94 %
Maus 8 ♂	121 %	130 %	94 %

Tab. 3.3: Prozentuales Übergewicht der chimären Tiere

Der Vergleich mit dem gemischten genetischen Hintergrund kann nur bedingt gewertet werden, da die Angaben für den Referenzstamm auf einer Fütterung mit einer Hoch-Fett-Diät basieren (Dong et al., 1997), die hier untersuchten Chimaren aber normal gefüttert wurden. Zudem zeigt der Stamm C57/BL/6J im Gegensatz zu C57/BL/6N per se eine hohe Neigung zu futterinduzierter Fettleibigkeit und Diabetes Typ II (Mouse Genome Database, Mouse Genome Informatics, Mouse Phenome Database, The Jackson Laboratory, 2003). Außerdem kann ein 25 %iger Chimärismus nur teilweise zu einem Vergleich mit einem gemischten genetischen Hintergrund herangezogen werden. Bei einer Gegenüberstellung mit normal gefütterten Tieren zeigt sich auch hier eine deutliche Gewichtszunahme.

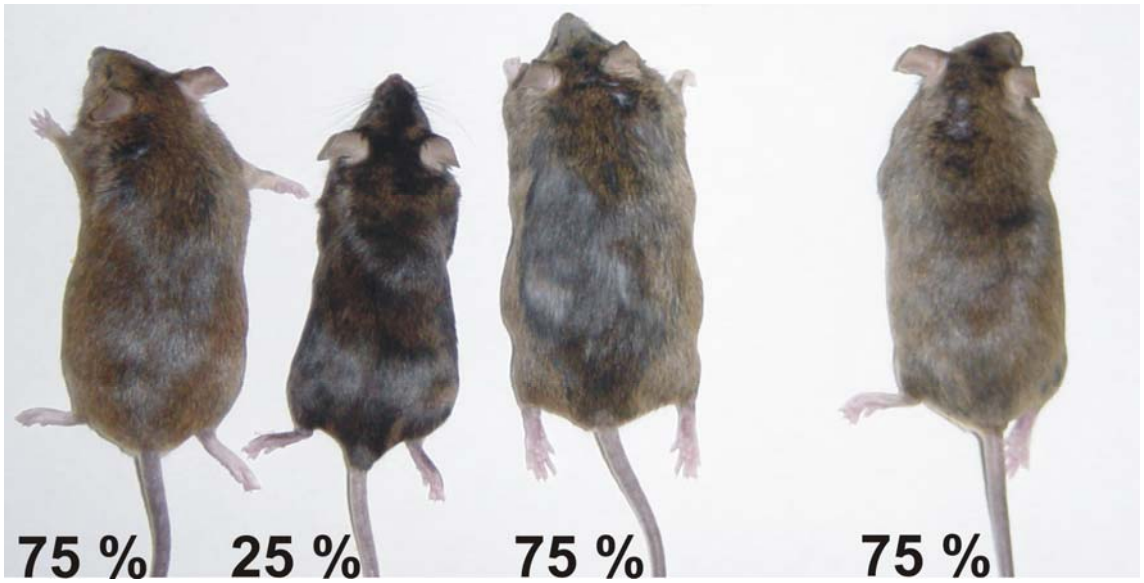


Abb. 3.14: Optische Unterschiede in Größe und Gewicht bei differierenden Chimärismusgraden

Bei der anschließenden Sektion der hochchimären Tieren zeigte sich ein deutlicher Unterschied in der Größe der Hoden, verglichen mit den männlichen Tieren, die nur einen 25 %igen Chimärismus aufwiesen oder unveränderten Vergleichstieren. Die Hoden der hochchimären Tiere waren ca. 70-80 % kleiner als bei den niederchimären bzw. unveränderten Vergleichstieren und im umliegenden Fettgewebe kaum auszumachen (Abb. 3.15). Da die Tiere 1-3 Nachkommen pro Wurf zeugten (alle C57/BL/6N) müssen die Testes jedoch als reduziert funktional angesehen werden. Eine ähnlich dramatische Reduktion der Testes konnte bei ARKO-Mäusen (Yeh et al., 2002) beobachtet werden, allerdings kommt es bei diesen Tieren zu einer Gewichtsreduktion und nicht wie bei den hier untersuchten Chimaren zu einer Gewichtszunahme. Bei anderen Versuchen wurde zwar eine Abnahme des Testisgewichts beobachtet, das Körpergewicht entsprach aber dem Wildtyp (Schürmann et al., 2002). Eine Reduktion der Testes bei gleichzeitiger Gewichtszunahme wurde so bisher nicht beschrieben. Eine Feminisierung des Genitalapparates konnte ebenfalls nicht beobachtet werden, auch die übrige Morphologie der Reproduktionsorgane erschien normal. Es kann angenommen werden, dass bei den 25 %igen Chimären gegebenenfalls keine Keimbahntransformation stattgefunden hat und daher die Hoden eine normale Größe aufwiesen. Es konnte nicht festgestellt werden, ob heterozygote Embryonen entstanden, dann aber in der Embryonalentwicklung abgestorben sind.

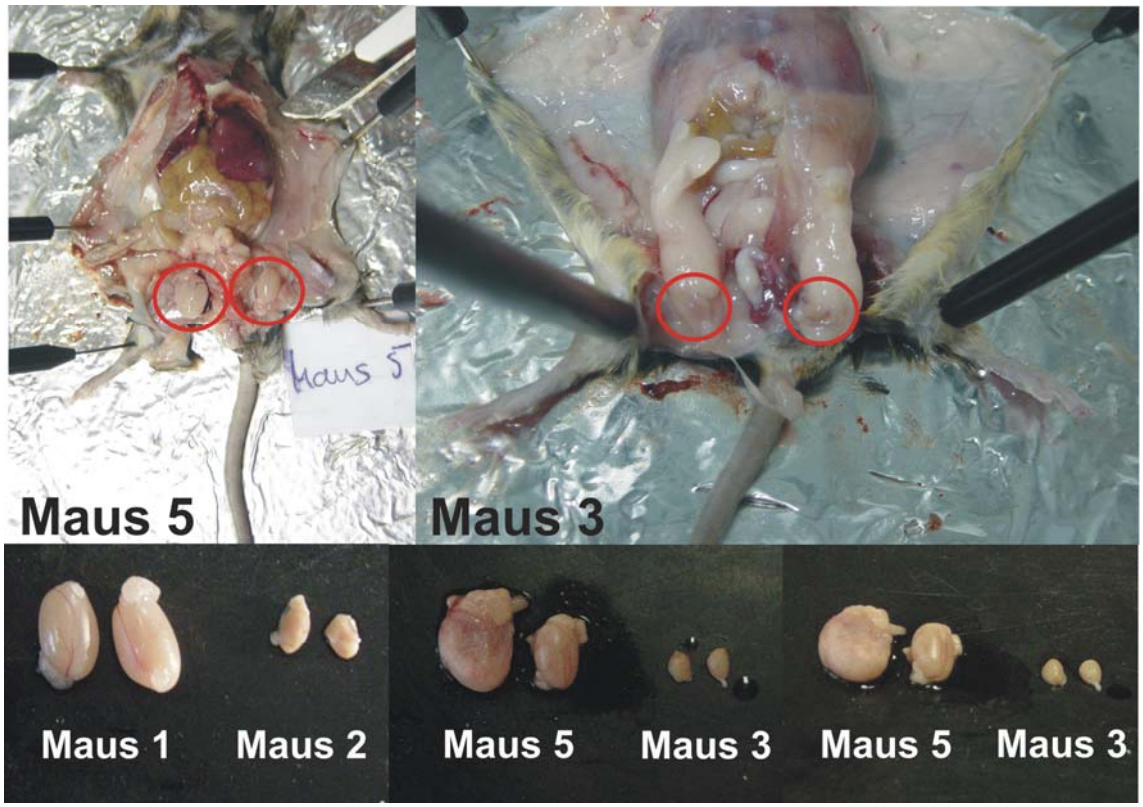


Abb 3.15: Vergleich der Testis zwischen einer niederchimären (25 %, Maus 1+5) und hochchimären (75 %, Maus 2+3) Maus

Gezeigt sind zwei sezierte Chimären. Im linken oberen Bildabschnitt ist ein Tier (Maus 5) mit 25 % Chimärismus zu erkennen, die Testes wurden mit roten Kreisen markiert und zeigen eine normale Größe. Der rechte Bildabschnitt zeigt ein Tier (Maus 3) mit 75 % Chimärismus. Auch hier wurden die Testes mit roten Kreisen markiert. Die Hoden waren im Fettgewebe kaum zu erkennen. Die untere Bildhälfte zeigt einen Vergleich der Testes zwischen 25 %igen und 75 %igen Chimären nach Präparation. In der linken unteren Bildhälfte sind die Testes von Maus1 (25 %) und Maus 2 (75 %) gegenübergestellt, die untere Bildmitte zeigt den Vergleich zwischen Maus 5 und Maus 3. In der rechten unteren Bildhälfte ist eine perspektivische Sicht auf die präparierten Testes zu sehen, die den Größenunterschied noch deutlicher macht. Es kann angenommen werden, dass evtl. bei den niederchimären Tieren keine Keimbahntransformation stattgefunden hat.

3.2 *In silico*-Analysen der Sequenzdaten

3.2.1 Exonvorhersage für Cos-N10467Q3 durch Kombination von DOTPLOT Analysen (Mensch cDNA vs genomische Maussequenz), Genvorhersageprogrammen und Datenbanksuchen

Die ermittelten Sequenz-Rohdaten von Cosmid N10467Q3 umfaßten 137809 bp und konnten zu zwei Contigs mit 26347 bp (Contig1) und 2358 bp (Contig2) assembliert werden. Primäres Ziel war die Identifikation der zu *DMXL*1 homologen Sequenzabschnitte. Zu diesem Zweck wurden die ermittelten genomischen Sequenzdaten durch eine DOTPLOT Analyse (Abb 3.16) mit dem Programm MegAlign (DNASTAR/Lasergene) mit der cDNA des menschlichen *DMXL*1 verglichen. Auf Grundlage dieser Daten wurden - unter Berücksichtigung der GT-AG-Regel – die Exon-/Intron-Grenzen „per Hand“ festgelegt (Abb 3.16). Auf diese Weise konnten die in der Sequenz enthaltenen Exons sicher identifiziert werden. Gleichzeitig wurden verschiedene Programme zur Gen- und Exonvorhersage verwendet. Die zum Zeitpunkt der Cosmidsequenzierung verfügbaren Programmversionen (GRAIL, GENSCAN, FGENES, XPOUND (alles ältere Versionen)) erzeugten nur bedingt verlässliche Daten (Bikár, 1997; Bahr, 1999; Amid, 2002; Frazer et al., 2003) und zeichneten sich besonders durch falsch positiv vorhergesagte Exons und ungenau prognostizierte Exon/Intron-Grenzen aus. Sie wurden jedoch zur Verifizierung der DOTPLOT Analysen herangezogen. Auf diese Weise konnten in Cosmid N10467Q3 zehn Exons (Exon 8-17 der cDNA) identifiziert werden (Tab. 3.4). Neuere Programme zur Exonvorhersage (DNASweep, GENSCAN, FGENES (alle Abb. 3.17), cdna2genome, HmmGENE, Geneld, XPOUND) überzeugen durch verbesserte Verlässlichkeit. Von allen getesteten Programmen erkannte allerdings nur DNASweep alle enthaltenen Exons, definierte die Exon/Intron-Grenzen korrekt (Kapitel 3.5.4) und lieferte keine falschpositiven Exonvorhersagen. DNASweep ermöglichte auch die gleichzeitige gerichtete Suche in diversen öffentlichen Datenbanken. Die Länge der identifizierten Exons und Introns variiert stark und kann Tabelle 3.4 entnommen werden. Der Homologiegrad der Exonsequenzen von *Dmxl*1/*DMXL*1 zwischen Mensch und Maus liegt zwischen 84,5 % und 95,1 %, der Durchschnitt liegt bei 89,27 %. Im Schnitt werden bei Vergleichen der CDS zwischen Maus und Mensch Werte zwischen 84,7 % und 86 % ermittelt (Makalowski et al., 1996, 1998; Onyango et al., 2000; Waterston et al., 2002). Dies zeigt, dass die *Dmxl*1 Nukleotidsequenz stärker zwischen Maus und Mensch konserviert ist als der Duchschnitt aller Gene.

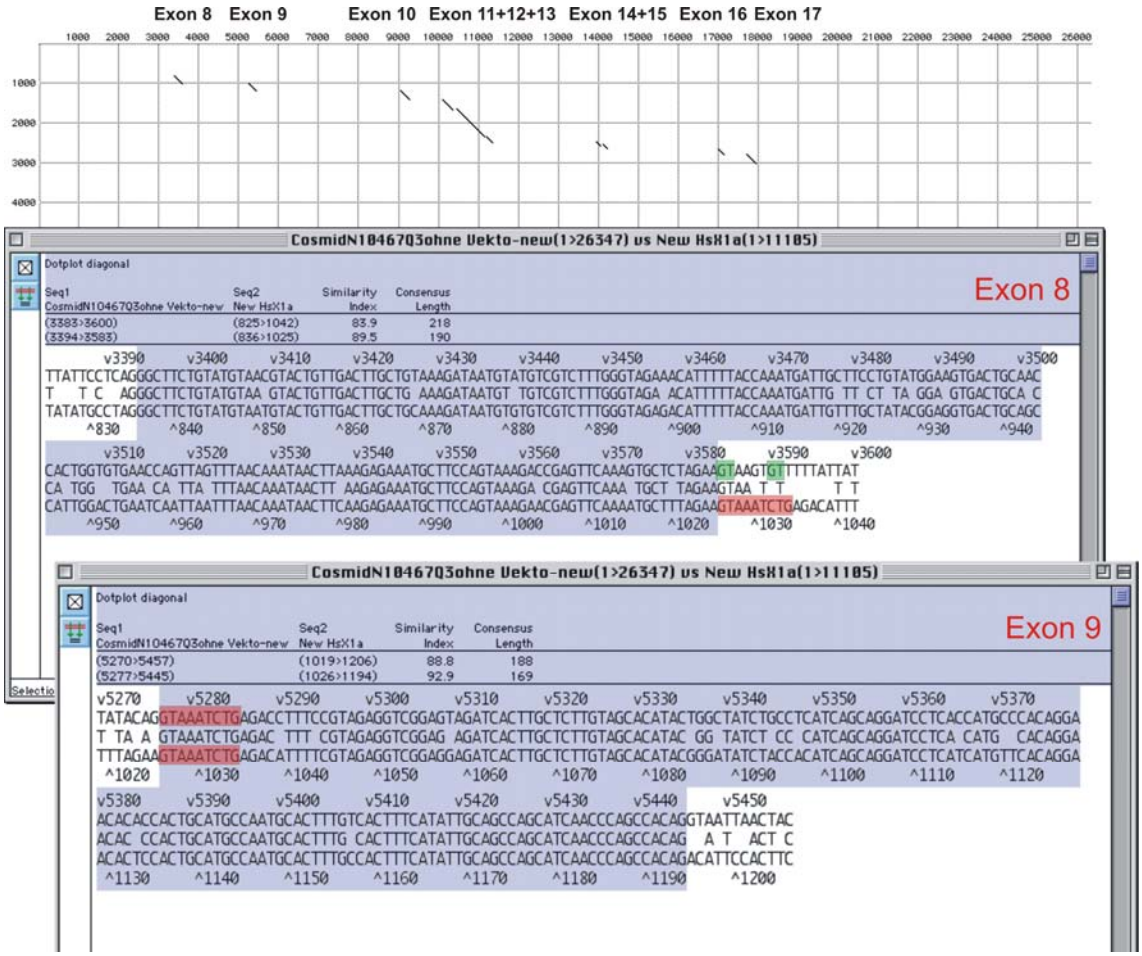
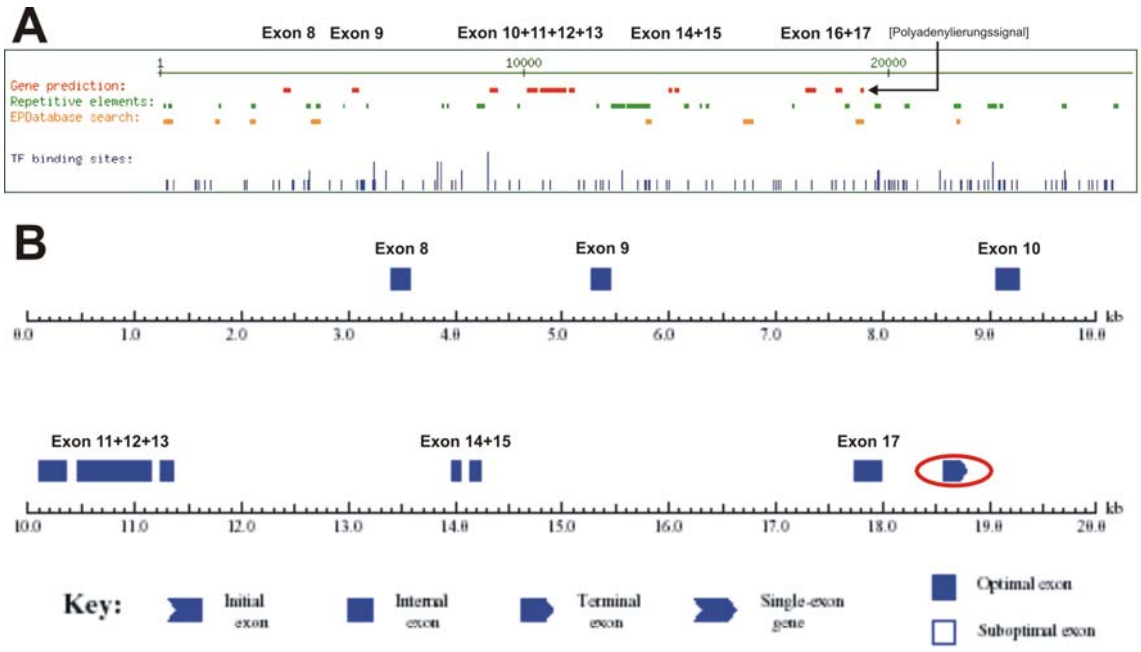


Abb. 3.16: DOTPLOT Analyse und Exon-Identifikation

Vergleich der humanen *DMXL*1 cDNA (vertikal) gegen die Sequenz von Cosmid N10467Q3 (horizontal). Die Analyse wurde mit der DOTPLOT-Option des Programms MEGALIGN durchgeführt (Parameter sind eine mindestens 80 %ige Übereinstimmung, eine Länge von mindestens 50 bp und ein Suchfenster von 50 bp). Die identifizierten hochkonservierten Bereiche sind als Linien auf dem Raster zu erkennen. Jede Linie entspricht einem Exon. Die Exonpositionen sind horizontal über dem Längenmaßstab der Cosmidsequenz vermerkt. Jeder konservierte Abschnitt kann als Sequenzvergleich dargestellt werden, hier am Beispiel von Exon 8 und Exon 9. Zunächst wurde für jeden Vergleich die Exon- und Intron-Position per Hand festgelegt. Der Abbruch der Sequenzübereinstimmung am Ende bzw. Anfang der beiden Exons ist deutlich zu erkennen. Gesucht wurde hier jeweils nach den Introngrenzen GT/AG. Für Exon 8 ergeben sich zwei mögliche Introngrenzen, die sich um zwei Codonpositionen unterscheiden und grün unterlegt sind. Da es sich hierbei um einen Vergleich von cDNA- und genomischer Sequenz handelt, lässt sich durch den Sequenzvergleich mit dem darauf folgenden Exon 9 eine eindeutige Positionierung der Grenzen festlegen. Der rot unterlegte Sequenzbereich wird in Exon 9 als hochkonservierter Abschnitt wiedergefunden, so dass das Ende des Exons 8 mit dem ersten grün unterlegten Splicesignal definiert werden kann. Das Exon kann dann markiert werden (blau unterlegter Hintergrund) und die entsprechenden Parameter für Homologie, Länge und Position auf Cosmid N10467Q3 bei Bedarf ausgelesen werden. Die älteren Versionen der Programme zur Exonvorhersage (Kapitel 3.5.1) konnten hier keine genaue Positionierung der Exon/Intron-Grenzen darstellen, vor allem da sich beide Splicesignale im selben Leserahmen befinden.



C

Positions of predicted genes and exons:

G	Str		Feature	Start		End	Score
1	+		TSS	1630			-7.89
1	+	1	CDSf	2133	-	2200	-1.10
1	+	2	CDSi	3394	-	3583	5.49
1	+	3	CDSi	5277	-	5445	8.57
1	+	4	CDSi	9066	-	9278	13.46
1	+	5	CDSi	10109	-	10362	14.40
1	+	6	CDSi	10471	-	11155	11.09
1	+	7	CDSi	11244	-	11365	4.64
1	+	8	CDSi	13965	-	14054	5.42
1	+	9	CDSi	14135	-	14237	4.47
1	+	10	CDSi	17038	-	17157	3.72
1	+	11	CDSi	17735	-	17956	4.23
1	+	12	CDSl	24844	-	25136	8.93
1	+		PolA	25953			-3.98

Abb. 3.17: Computergestützte und internetbasierte Exonvorhersage

Gezeigt sind drei Beispiele der computergestützten Exonvorhersage. Verwendet wurden hier die Programme DNAsweep (A), GENSCAN (B) und FGENES (C).

A: DNAsweep ist ein Bestandteil des HUSAR Programmpakets (Heidelberg, Kapitel 2.12). Bei diesem Programm ist es möglich, Suchparameter für eine Fülle unterschiedlicher öffenlicher Datenbanken selektiv einzubinden. Auf diese Weise kann einerseits speziesspezifisch gesucht werden, aber auch Interspeziesvergleiche können so realisiert werden. Als Parameter wurden hier eine Exonvorhersage für die Spezies Maus gewählt, zusätzlich wurden diverse Datenbanken hinzugezogen. Repetitive Sequenzabschnitte wurden automatisch von dem Programm RepMask (Kapitel 2.12) maskiert. Alle Exons wurden korrekt erkannt, in der Grafik sind sie als rote Balken dargestellt. Das Programm sucht standardmäßig nach Polyadenylierungssignalen, Promotorstrukturen (gelbe Balken) und Bindestellen für

Transkriptionsfaktoren, diese Ergebnisse können hier ignoriert werden. Die Verteilung repetitiver Elemente (grüne Balken) ist ebenfalls dargestellt.
B: Für die Arbeit mit GENSCAN wurden die in der Sequenz enthaltenen repetitiven Elemente zunächst mit Hilfe des Programms RepeatMasker (Kapitel 2.12) maskiert, anschließend konnten die Daten auf dem GENSCAN-Server bearbeitet werden. GENSCAN detektierte 9 von 10 in der Sequenz enthaltenen Exons. Exon 16 wurde nicht detektiert, ein falschpositives Exon (rot markiert) wurde für den Bereich 18574-18715 vorgeschlagen. Eine Legende der Zeichen ist in der Abbildung wiedergegeben.
C: Bei der Analyse der Cosmidsequenz wurden mit dem Subprogramm FGENESH des FGENES-Pakets, das eine separate Einstellung für die Exonvorhersage bei Maus aufweist, alle enthaltenen Exons korrekt erkannt und zwei falschpositive Exons vorhergesagt.

	Länge des Exons	Position auf Cosmid	Position auf cDNA	Länge des folgenden Introns	Homologie Maus/Mensch
Exon 8	189 bp	3394-3583	831-1020	1692 bp	89,8 %
Exon 9	168 bp	5277-5445	1021-1189	3619 bp	92,9 %
Exon 10	212 bp	9066-9278	1190-1402	829 bp	89,7 %
Exon 11	253bp	10109-10362	1403-1656	107 bp	87,4 %
Exon 12	684 bp	10471-11155	1657-2341	87 bp	90.5 %
Exon 13	121 bp	11244-11365	2342-2463	2598 bp	95,1 %
Exon 14	89 bp	13965-14054	2464-2553	79 bp	92,2 %
Exon 15	102 bp	14135-14237	2554-2656	2799 bp	84,5 %
Exon 16	119 bp	17038-17157	2657-2776	576 bp	85,0 %
Exon 17	221 bp	17735-17956	2777-2998		85,6 %
∅					89,27 %

Tab. 3.4: Exons und Introns von Cosmid N10467Q3

In der Tabelle sind alle Exons mit ihrer Länge, Position in der Cosmidsequenz, Position in der cDNA und ihrem Homologiegrad im Vergleich mit der humanen Sequenz von *DMXL*1 aufgeführt.

3.2.2 Repetitive Elemente in Cosmid N10467Q3

Für die Analyse der repetitiven Elemente wurde das Programm REPEATMASKER (Smit & Green, unpubliziert) in der Version 2002/07/13 verwendet. Die Ergebnisse dieser Analyse sind in die Aufstellung aller repetitiven Elemente der genomischen Region von *Dmxl*1 eingearbeitet (Kapitel 3.2.5, Tab 3.8). Insgesamt beläuft sich der Anteil an repetitiven Elementen inklusive der DNA-Abschnitte mit niedriger Komlexität und einfachen Sequenzwiederholungen auf 26,65 %. Die Verteilung der Elemente ist in Abb. 3.17 A (grüne Balken) wiedergegeben.

3.2.3 Genomische Lokalisation von Cosmid N10467Q3

Im Laufe dieser Arbeit wurde das humane *DMXL*1-Gen von Kraemer et al. (1999) in der Region 5q22 auf Chromosom 5 durch *in situ* Hybridisierung lokalisiert. Die syntäne Region zu diesem Abschnitt liegt in der Maus auf Chromosom 18 (Waterston et al., 2002), es konnte daher davon ausgegangen werden, dass sich das orthologe Gen der Maus auf diesem Chromosom befindet. Datenbanksuchen bestätigten diese Annahme, die genomische Sequenz aus Cosmid N10467Q3 konnte dem murinen Chromosom 18 Bande C zugeordnet werden (Abb 3.18). Es wurden die Basen 1-18334 der Cosmidsequenz durch Abgleich mit öffentlichen Datenbanken vollständig lokalisiert. Die Basen 18335 bis 26347 sowie Contig2 der Cosmidsequenz zeigten keine Homologie zu Einträgen in den öffentlichen Datenbanken. Durch eine später erfolgte BLAST-Suche konnten die verbleibenden Sequenzabschnitte auf Chromosom 4 lokalisiert werden. Eine nähere Untersuchung der Cosmidsequenz zeigte, dass sich exakt an dem Übergang der Datenbankeinträge von Chromosom 18 und Chromosom 4 eine *Mbo* I Schnittstelle befindet (Abb 3.18). Die Genbank 129, aus der auch Cosmid N10467Q3 stammt, wurde von Burgtorf, Poch und Wiles durch einen unvollständigen genomischen Verdau mit *Mbo* I konstruiert. Es ist daher davon auszugehen, dass es sich bei Cos N10467Q3 um eine *in vitro*-Ligat verschiedener DNA-Abschnitte unterschiedlicher Chromosomen handelt. Für die Konstruktion des Targeting-Vektors wurde u.a. ein Hind III-Fragment verwendet, das den Sequenzbereich von 14049 bis 20835 bp umfasst und somit 2358 bp enthält, die nicht mit der Zielsequenz des Tagetings übereinstimmt. Der Vergleich der ermittelten Sequenzdaten mit den Einträgen in den öffentlichen Datenbanken zeigt eine 99.93 %ige Übereinstimmung auf Chromosom 18 (zwölf unterschiedliche Basen bzw. Lücken) und eine 99,94 %ige Homologie auf Chromosom 4 (fünf untertschiedliche Basen bzw. Lücken). Ein Vergleich mit den erhobenen Rohdaten ergab, dass es sich hierbei offensichtlich um Polymorphismen handelt. Die Einträge in den öffentlichen Datenbanken basieren auf einer Sequenzierung des Mausstamms C57BL6/J (Waterston et al., 2002), Cosmid N10467Q3 stammt aus einer Maus mit gemischtem genetischen Hintergrund aus 129- und Ola-Stamm (RZPD, 1997). Die beobachtete Anzahl der SNP's („single nucleotide polymorphisms") entpricht einer Genregion mit einer niedrigen Polymorphismusrate. Ungefähr 58 % des Genoms zeigten diese niedrige Varianz bei einem Sequenzvergleich zwischen 129/Ola- und C57BL6/J-Maus (Wade et al., 2002).

Abb. 3.18: Cosmid N10467Q3 ist ein *in vitro*-Ligat von DNA-Abschnitten aus Chromosom 18 und Chromosom 4 (folgende Seite)

1: Mbo I Restriktionskarte von Cos N10467Q3. Der blau dargestellte Abschnitt liegt auf Chromosom 18 und enthält alle identifizierten Exons von *Dmxl*1 (Tabelle 3.5). In rot ist der Bereich abgebildet, der auf Chromosom 4 lokalisiert wurde.

2: Sequencher-Projekt von Cosmid N10467Q3. Dargestellt ist der Übergang von Chromosom 18 auf Chromosom 4, die blau markierte Mbo I-Schnittstelle mit der Sequenz GATC ist an dieser Stelle in den Referenzsequenzen der Datenbanken auf Chromosom 4 und auf Chromosom 18 zu finden, die davor liegenden Sequenzen befinden sich auf Chromosom 18, dahinterliegende auf Chromosom 4.

3.2.4 Analyse der Sequenzdaten von PAC-F13324Q2 und PAC-I02348Q2

Die durch selektive Sequenzierung aus PAC-F13324Q2 und PAC-I02348Q2 gewonnenen Rohdaten (Kapitel 3.2.2) wurden wie unter Kapitel 3.5.1, 3.5.2 und 3.5.3 beschrieben verarbeitet. Die Identifikation der Exons erfolgte primär „per Auge“ und durch Maus/Mensch-Vergleich mit dem Programm MEGALIGN. Andere Programme zur Exonvorhersage wurden nur zur Verifikation der ermittelten Exons herangezogen. Insgesamt wurden mit dieser Strategie 121993 bp Rohdaten ermittelt, die teilweise zu Contigs zusammengefügt werden konnten. Durch Analyse wurden dann die Exons 1-7 und 18-34 inklusive angrenzender Intronsequenzen ermittelt. Weitere Exons konnten auf diese Weise nicht ermittelt werden, da die isolierten PACs nicht den kompletten genomischen Bereich von *Dmxl*1 abdeckten (Abb. 3.3). Mit Hilfe dieser Daten konnten dann Primer abgeleitet werden, die für die Amplifikation der cDNA über RT-PCR verwendet wurden. Der Vergleich der in Abb. 3.6 dargestellten cDNA mit den öffentlichen Datenbanken zeigte eine 100 %ige Übereinstimmung der ermittelten Exonsequenzen.

3.2.5 Genomische Lokalisation und Genstruktur des murinen *Dmxl*1 und humanen *DMXL*1

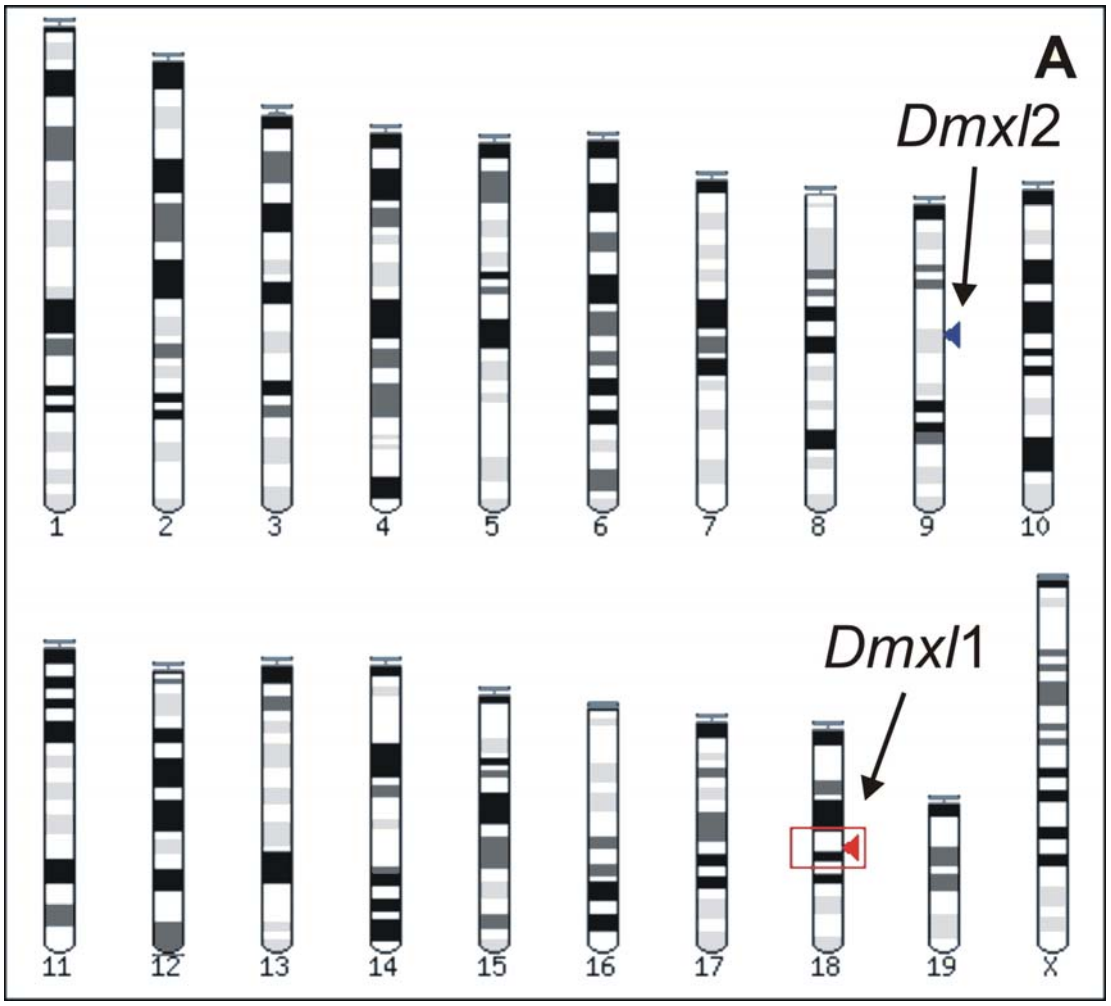
Um die komplette Genstruktur von *Dmxl*1 aufklären zu können, musste zunächst der entsprechende Genbereich von *Dmxl*1 aus den öffentlichen Datenbanken herausgesucht

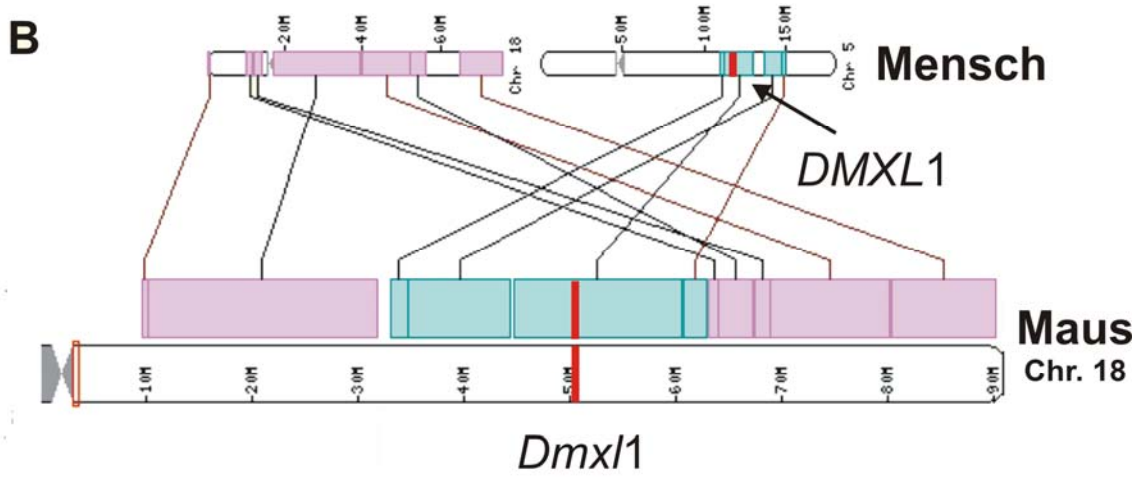
werden. Hierzu wurden Datenbanksuchen mit dem ersten Exon (beginnend mit dem Startcodon) und dem letzten Exon (endend mit dem Stopcodon) durchgeführt. Auf diese Weise konnten das erste Exon auf Chromosom 18 Bande C in dem Chromosomenbereich 50488667-50488753, der kodierende Bereich des letzten Exons in dem Chromosomenabschnitt 50359444-50359666 (beide Angaben basieren auf der Mausgenom-Version 19.32a.2 vom 09.02.2004) lokalisiert werden (Abb. 3.19). Demzufolge liegt *Dmxl*1 auf dem Crick-Strang, der translatierte Bereich ersteckt sich auf genomischer Ebene über 129558 bp und die Orientierung verläuft in Richtung Centromer. Für weitere Untersuchungen wurde zusätzlich zu dem oben angegebenen DNA-Abschnitt jeweils 10000 bp vor dem Startcodon und nach dem Stopcodon aus den Datenbanken heruntergeladen und lokal zur Verfügung gestellt. Mit diesen Daten wurde anschliessend eine Analyse der cDNA und der genomischen Region von *Dmxl*1 und *DMXL*1 mit dem Programm SIM4 durchgeführt (Tab. 3.5 und Tab. 3.6). Diese Analyse dient zur Überprüfung der Sequenzdaten, die bereits durch die selektive Sequenzierung der PAC-Klone PAC-F13324Q2 und PAC-I02348Q2 und durch Datenbankanalysen mit der humanen cDNA von *DMXL*1 gewonnen wurden (Kapitel 3.2.2). Mit diesen Sequenzdaten konnten, wie unter 3.5.1 beschrieben, DOTPLOT-Auswertungen mit dem Programm MEGALIGN durchgeführt und so weitere 33, nicht in Cos N10467Q3 enthaltene Exons identifiziert werden (Daten nicht gezeigt, ein Sequencher-Projekt mit den entsprechenden Sequenzdaten befindet sich auf dem beigefügten Datenträger).

Abb 3.19: Genomische Lokalisation von *Dmxl*1 und *DMXL*1 (folgende Seite)

A: Nach einer Datenbanksuche (ENSEMBL Version 19.32a.2 vom 09.02.2004) mit Exon 1 konnte das *Dmxl*1-Gen Maus-Chromosom 18 Bande C zugeordnet werden. Eine Suche mit Exon 43 (inkl. 3'-UTR) bestätigte dies. Ein weiterer Eintrag in den Datenbanken zeigt eine 75 %ige Homologie zu *Dmxl1* auf Chromosom 9. Es handelt sich hierbei um eine zweite Kopie von *Dmxl* (*Dmxl*2, auch bekannt unter Rabconnectin 3), eine Besprechung dieses Eintrags erfolgt in Kapitel 3.2.10 und 4.4.6

B: Darstellung aller syntänen Chromosomenregionen des Menschen zu Maus Chromosom 18. Der Abschnitt von Chromosom 18, auf dem das murine *Dmxl*1 lokalisiert ist, entspricht dem humanen Chromosom 5q23, Base 118473481-118659130.


A
Dmxl2
Dmxl1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X


B
Mensch
DMXL1
Maus
Chr. 18
Dmxl1

Tab. 3.5	Lokalisation auf genomischer DNA (lokal)	Abschnitt der cDNA	Übereinstimmung mit genomischen Daten	Länge des Exons	Länge des folgenden Introns
5'UTR, Exon1	9908-10087	1-180	100%	~ 180 bp	7210 bp
Exon2	17299-17424	181-306	100%	125 bp	3272 bp
Exon3	20698-20769	307-378	100%	71 bp	2726 bp
Exon4	23497-23575	379-457	100%	78 bp	2003 bp
Exon5	25580-25706	458-584	100%	126 bp	2634 bp
Exon6	28342-28408	585-651	100%	66 bp	748 bp
Exon7	29158-29336	652-830	100%	178 bp	5055 bp
Exon8	34393-34582	831-1020	100%	189 bp	1692 bp
Exon9	36276-36444	1021-1189	100%	168 bp	3619 bp
Exon10	40065-40277	1190-1402	100%	212 bp	829 bp
Exon11	41108-41361	1403-1656	100%	253 bp	107 bp
Exon12	41470-42154	1657-2341	100%	684 bp	87 bp
Exon13	42243-42364	2342-2463	100%	121 bp	2599 bp
Exon14	44965-45054	2464-2553	100%	89 bp	79 bp
Exon15	45135-45237	2554-2656	100%	102 bp	2794 bp
Exon16	48033-48152	2657-2776	100%	119 bp	576 bp
Exon17	48730-48951	2777-2998	100%	221 bp	5890 bp
Exon18	54843-56525	2999-4681	100%	1682 bp	1368 bp
Exon19	57895-58002	4682-4789	100%	107 bp	7877 bp
Exon20	65881-66050	4790-4959	100%	169 bp	576 bp
Exon21	66628-66725	4960-5057	100%	97 bp	1054 bp
Exon22	67781-67946	5058-5223	100%	165 bp	608 bp
Exon23	68556-68817	5224-5485	100%	261 bp	1565 bp
Exon24	70384-71462	5486-6564	100%	1078 bp	683 bp
Exon25	72147-72338	6565-6756	100%	191 bp	3529 bp
Exon26	75869-75945	6757-6833	100%	76 bp	3479 bp
Exon27	79426-79553	6834-6961	100%	127 bp	390 bp
Exon28	79945-80193	6962-7210	100%	248 bp	9613 bp
Exon29	89808-89962	7211-7365	100%	154 bp	4780 bp
Exon30	94744-94886	7366-7508	100%	142 bp	2165 bp
Exon31	97053-97138	7509-7594	100%	85 bp	855 bp
Exon32	97995-98196	7595-7796	100%	201 bp	1565 bp
Exon33	99763-99880	7797-7914	100%	117 bp	9194 bp
Exon34	109076-109162	7915-8001	100%	86 bp	2585 bp
Exon35	111749-111870	8002-8123	100%	121 bp	485 bp
Exon36	112357-112498	8124-8265	100%	141 bp	4215 bp
Exon37	116714-116774	8266-8326	100%	60 bp	11585 bp
Exon38	128361-128488	8327-8454	100%	127 bp	4491 bp
Exon39	132981-133073	8455-8547	100%	92 bp	1391 bp
Exon40	134466-134557	8548-8639	100%	91 bp	1003 bp
Exon41	135562-135614	8640-8692	100%	52 bp	2607 bp
Exon42	138223-138440	8693-8910	100%	217 bp	895 bp
Exon43, 3' UTR	139337-142636	8910-9132, 9133-12209	99,73%	222 bp, 3076 bp	

Tab. 3.5: SIM4 Analyse der murinen genomischen Region von *Dmxl*1

Die Tabelle zeigt eine Zusammenfassung der SIM4 Analyse. Lokalisation, Übereistimmung mit den genomischen Daten aus den öffentlichen Datenbanken sowie die Länge aller 43 Exons inklusive des 3'- und 5'- UTR werden in tabellarischer Form dargestellt. Zusätzlich wird die Länge aller 42 Introns angegeben. Die 99,73 %ige Übereinstimmung von Exon 43 beruht auf auf dem Sequenzunterschied an neun Positionen im 3'- UTR und ist auf Polymorphismen zurückzuführen (Wade et al., 2002). Die Länge der Introns unterscheidet sich an zwei Positionen von Cos N10467Q3 (dunkelgrau unterlegt), Intron 13 ist in der Cosmidsequenz um eine Base länger, Intron 15 unterscheidet sich um +5 Basen. Intron 11 und 12 erscheinen auffällig (hellgrau unterlegt), da sie in ihrer Länge mit der humanen Sequenz übereinstimmen (Intron 12 minus eine Base im Vergleich zum Menschen).

Tab. 3.6	Lokalisation auf genomischer DNA (lokal)	Abschnitt der cDNA	Übereinstimmng mit genomischen Daten	Länge des Exons	Länge des folgenden Introns
5'UTR, Exon1	9870-10048	1-179	100%	~179 bp	26321 bp
Exon2	36371-36496	180-305	100%	125 bp	3829 bp
Exon3	40327-40398	306-377	100%	71 bp	3172 bp
Exon4	43572-43650	378-456	100%	78 bp	4891 bp
Exon5	48543-48675	457-589	100%	132 bp	4190 bp
Exon6	52867-52933	590-656	100%	66 bp	1615 bp
Exon7	54550-54728	657-835	100%	178 bp	2477 bp
Exon8	57207-57396	836-1025	100%	189 bp	1965 bp
Exon9	59363-59531	1026-1194	100%	168 bp	8070bp
Exon10	67603-67815	1195-1407	100%	212 bp	3707 bp
Exon11	71524-71777	1408-1661	100%	253 bp	107 bp
Exon12	71886-72570	1662-2346	100%	684 bp	88 bp
Exon13	72660-72781	2347-2468	100%	121 bp	9450 bp
Exon14	82233-82322	2469-2558	100%	89 bp	604 bp
Exon15	82928-83030	2559-2661	100%	102 bp	2197 bp
Exon16	85229-85348	2662-2781	100%	119 bp	291 bp
Exon17	85641-85862	2782-3003	100%	221 bp	1267 bp
Exon18	87131-88813	3004-4686	100%	1682 bp	1506 bp
Exon19	90321-90428	4687-4794	100%	107 bp	12469 bp
Exon20	102899-103068	4795-4964	100%	169 bp	505 bp
Exon21	103575-103672	4965-5062	100%	97 bp	1334 bp
Exon22	105008-105173	5063-5228	100%	165 bp	820 bp
Exon23	105995-106256	5229-5490	100%	261 bp	2324 bp
Exon24	108582-109672	5491-6581	100%	1090 bp	496 bp
Exon25	110170-110361	6582-6773	99%	191 bp	3290 bp
Exon26	113653-113729	6774-6850	100%	76 bp	2018 bp
Exon27	115749-115876	6851-6978	100%	127 bp	510 bp
Exon28	116388-116636	6979-7227	100%	248 bp	11462 bp
Exon29	128100-128278	7228-7406	100%	178 bp	3940 bp
Exon30	132220-132362	7407-7549	100%	142 bp	2399 bp
Exon31	134763-134848	7550-7635	100%	85 bp	1297 bp
Exon32	136147-136348	7636-7837	100%	201 bp	5361 bp
Exon33	141711-141828	7838-7955	100%	117 bp	13463 bp
Exon34	155293-155379	7956-8042	100%	86 bp	3483 bp
Exon35	158864-158985	8043-8164	100%	121 bp	345 bp
Exon36	159332-159473	8165-8306	100%	141 bp	3526 bp
Exon37	163101-163161	8307-8367	100%	60 bp	8569 bp
Exon38	171732-171859	8368-8495	100%	127 bp	3853 bp
Exon39	175714-175806	8496-8588	100%	92 bp	1573 bp
Exon40	177381-177472	8589-8680	100%	91 bp	1337 bp
Exon41	178811-178863	8681-8733	98%	52 bp	3886 bp
Exon42	182751-182968	8734-8951	100%	217 bp	2417 bp
Exon43, 3' UTR	185387-187519	8952-9174, 9175-11084	99%	222 bp, 1909 bp	

Tab. 3.6: SIM4 Analyse der humanen genomischen Region von *DMXL*1

Die Darstellungsweise dieser Tabelle entspricht Tab. 3.6. Die 99 %ige Übereinstimmung von Exon 43 beruht auf auf dem Sequenzunterschied an zehn Positionen im 3'- UTR und ist entweder auf Polymorphismen oder Fehler in den öffentlichen Datenbanken zurückzuführen. In Exon 24 und Exon 41 wurde jeweils ein Unterschied in einer Base beobachtet, die Gründe hierfür sind unbekannt. Die Introns 11 und 12 (hellgrau unterlegt) sind im Vergleich zur Maussequenz hochkonserviert, Intron 12 ist im Vergleich zur Maus eine Base länger. Alle Exons entsprechen in Ihrer Länge exakt den Exons in der Maus, Ausnahmen (dunkelgrau unterlegt) bilden nur die Exons 5 (+2 Codonpositionen), 24 (+ 4 Codonpositionen) und 29 (+ 8 Codonpositionen).

		Anzahl der Elemente		Länge (bp)		% der Sequenz	
		Maus	**Mensch**	**Maus**	**Mensch**	**Maus**	**Mensch**
SINE		116	237	15635	33035	10.45	16,88
	B1	52	220	6437	30535	4,30	15,60
	B2-B4	53		8348		5,58	
	ID	8		604		0,40	
	MIR	3	17	246	2500	0,16	1,28
LINE		34	51	15630	37238	10,45	19,03
	LINE1	30	32	14510	29487	9,70	15,07
	LINE2	4	18	1120	7696	0,75	3,93
	L3/CR1		1		55		0,03
LTR-Elemente		32	20	10507	9426	7,03	4,82
	MaLR	24	13	8450	4862	5,65	2,48
	ERVL	2	4	223	1043	0,15	0,53
	ERV I	1	3	596	3521	0,40	1,80
	ERV II	3		568		0,38	
DNA-Elemente		7	18	2056	3895	1,37	1,99
	MER1	6	17	978	2919	0,65	1,49
	MER2	1	1	1078	976	0,72	0,50
„small RNA“		5	3	500	244	0,33	0,12
einfache Repetetive		44	23	1917	1331	1,28	0,68
„low complexity“		56	43	2060	1920	1,38	0,98
Gesamt		294	395	48301	87091	32,30	44,51

Tabelle 3.7: Repetitive Elemente der genomischen Region von *Dmxl*1 und *DMXL*1 (umseitig)

SINE: „short interspersed nuclear element“, LINE: „long interspersed nuclear element“, LTR: „long terminal repeat“, MIR: „mammalian-wide interspersed repetitive element“, MER: „medium reinteration frequency interspersed repeat“, MaLR: mammalian LTR-retrotransposon“, ERV: „endogenous retroviral sequences class I/II“, ERVL: endogenous retroviral-like sequences (class III).

SIM4 verwendet den BLAST Algorithmus (Altschul et al., 1997) in Kombination mit einer Suchfunktion für Splice-Stellen in der genomischen Sequenz. SIM4 identifiziert alle Exons aus den Sequenzierungen der Cosmid- und PAC-Klone und bestätigt die

vorher definierten Exon/Intron-Grenzen (Tabelle 3.5), die komplette Analyse befindet sich auf dem beigefügten Datenträger. Alle Spleißstellen entsprechen der GT-AG-Regel. Für eine umfassendere Analyse der Genstruktur von *Dmxl*1 wurde die syntäne Genregion des Menschen auf die gleiche Weise wie oben beschrieben identifiziert und die Sequenz lokal zur weiteren Analyse gespeichert. Hierbei konnte die Annahme, dass sich die homologen Gene *Dmxl*1 und *DMXL*1 in den entsprechenden syntänen Regionen von Maus und Mensch befinden, nochmals bestätigt werden (Abb 3.20). Mit der ermittelten humanen genomischen Sequenz wurde ebenfalls eine SIM4-Analyse durchgeführt (Tabelle 3.6), auch hier ist das ausführliche Ergebnis auf dem beiligenden Datenträger zu finden. Grundsätzlich scheint vor allem die Länge der Exons zwischen Maus und Mensch hochkonserviert zu sein. Der Mensch unterscheidet sich hier nur in den Exons 6 (+2 Codonpositionen), 24 (+4 Codonpositionen) und 29 (+8 Codonpositionen) von der Maus. Die Intronlängen sind weniger gut konserviert und unterscheiden sich in vielen Fällen deutlich von der Maus, im Schnitt sind die Introns in der humanen Sequenz aufgrund einer höheren Zahl repetitiver Elemente länger als in der murinen Sequenz (Tabelle 3.6, 3.7). Nur die Längen von Intron 12 und Intron 13 (+1 bp) zeigen sich hier hochkonserviert. Um alle repetitiven Elemente in der murinen und humanen Sequenz zu identifizieren, wurden beide Sequenzen mit dem Programm REPEATMASKER untersucht (Tabelle 3.7). Insgesamt beläuft sich der Anteil an repetitiven Elementen in der murinen Sequenz auf 34,49 % (=51580 bp), beim Menschen sind es 51,52 % (= 100566 bp). Die Abbildungen 3.21 und 3.22 geben einen Überblick über die Verteilung der repetitiven Elemente. Vor allem im Bereich der LINE- und SINE-Elemente zeigt der Mensch einen höheren prozentualen Anteil dieser Elemente als die Maus. Daten, die mit beiden Analysen gewonnen wurden, konnten anschließend von dem Programm Advanced PipMaker in eine vergleichende grafische Darstellung, „percent identity plot“ (PIP, Schwartz et al., 2000), der genomischen Region von *Dmxl*1 in Mensch und Maus umgewandelt werden (Abb. 3.20 und 3.21). Hierbei sind die relativen prozentualen Übereinstimmungen beider Sequenzen miteinander verglichen, das Alignment basiert auf der Verwendung des BlastZ-Algorythmus (Altschul et al., 1997; Zhang et al., 1998). Zusätzlich sind in dieser Grafik alle Exons, 5’- und 3’-UTR, alle identifizierten repetitiven Elemente sowie GC-reiche Regionen und mögliche tRNA-Gene aufgeführt. Die Abbildungen zeigen, dass die meisten Exons hoch konserviert sind, die Übereinstimmung liegt im Durchschnitt bei 89,9 % (Kapitel 3.5.5, Tab 3.9). Auch die Speißstellen und der angrenzende Intronbereich zeigen in den meisten Fällen eine sehr hohe Konservierung und sind als distinkte Bereiche durchgehender Linien neben den markierten Exonabschnitten deutlich zu erkenen. Die 5’- und 3’- nicht-translatierten Bereiche zeigen ebenfalls eine hohe Sequenzübereinstimmung. Unterschiede finden sich erwartungsgemäß vor allem in den Intronbereichen, bei denen die Länge zwischen beiden Spezies variiert. Ursächlich hierfür ist eine große Anzahl repetitiver Elemente, in denen sich beide Arten voneinander unterscheiden, obwohl auch hier die Lage einiger Elemente konserviert scheint (SINE-Elemente zwischen Exon 8 und Exon 9, LINE2-Elemente, zwischen Exon 13 und Exon 14, LINE1-Elemente zwischen Exon 19 und 20, LINE2-Element zwischen Exon 23 und Exon 24, LTR-Element zwischen Exon 25 und Exon 26, ein anderes repetitives Element zwischen Exon 29 und Exon 30, LINE1-Elemente zwischen

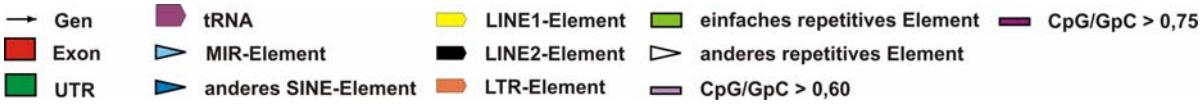
Exon 30, 31 und 32, sowie ein LINE1-Element zwischen Exon 41 und Exon 42). Ein weiteres Ziel dieser Analyse war es, konservierte Nicht-Exonbereiche zu identifizieren und anschließend näher zu untersuchen. Neben möglichen alternativ gespleißten Exons sollten so Kandidaten für Promotorstrukturen und regulatorische Elemente gefunden werden (Kapitel 3.5.7). Das Programm PipMaker ist zusätzlich in der Lage DOTPLOTS der jeweiligen Sequenzbereiche zu erstellen (Abb. 3.22). Die Abschnitte beider Genome sind syntän, da man deutlich eine Diagonale erkenen kann. Die „dotplot"-Diagonale wird allerdings in den Intronbereichen an 101 Stellen unterbrochen. Es haben hier offensichtlich Insertionen bzw. Deletionen in einem der beiden Genome stattgefunden

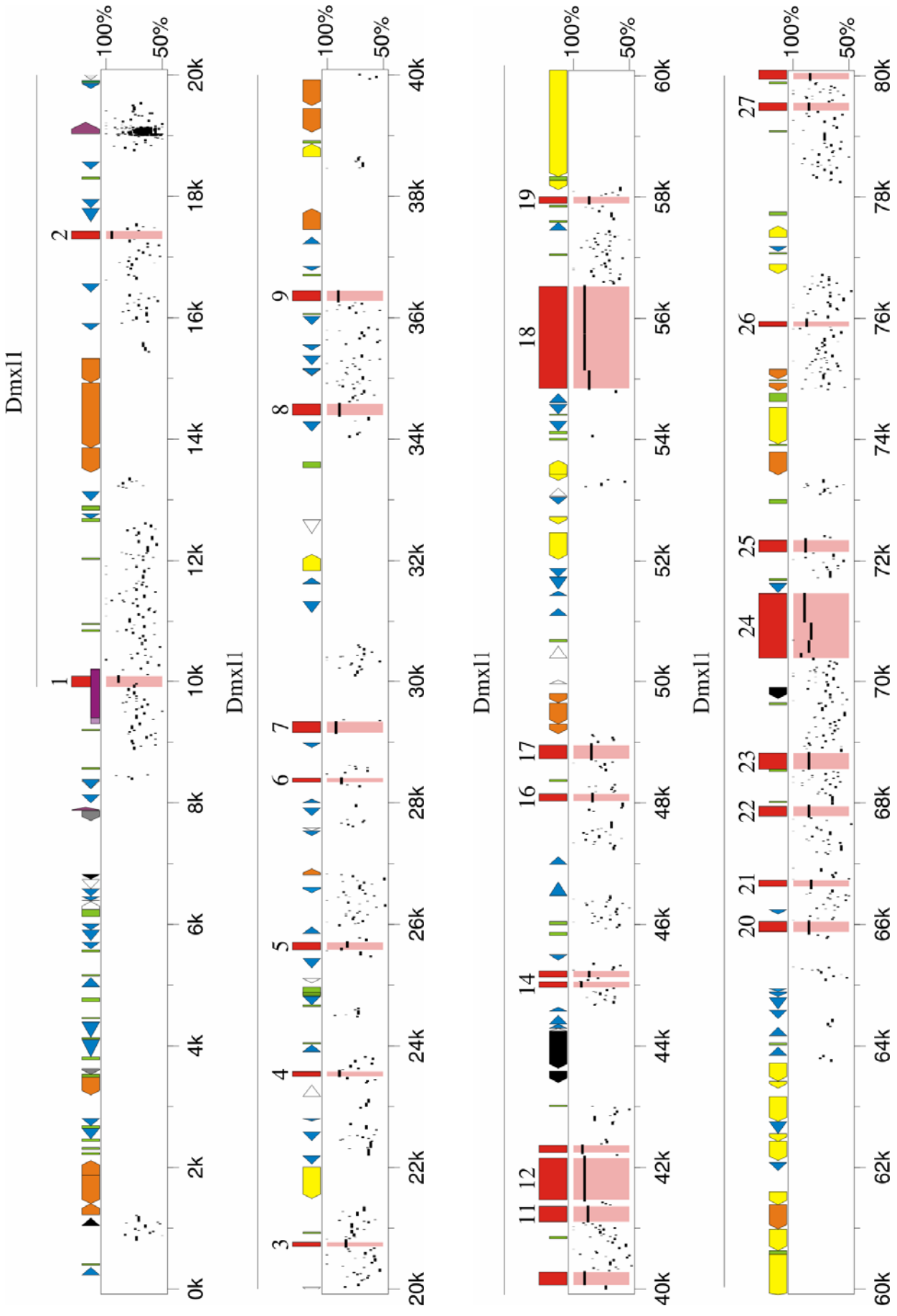
Abb. 3.20: PIP-Vergleich der murinen genomischen *Dmxl*1-Sequenz mit der humanen (folgende Seiten).

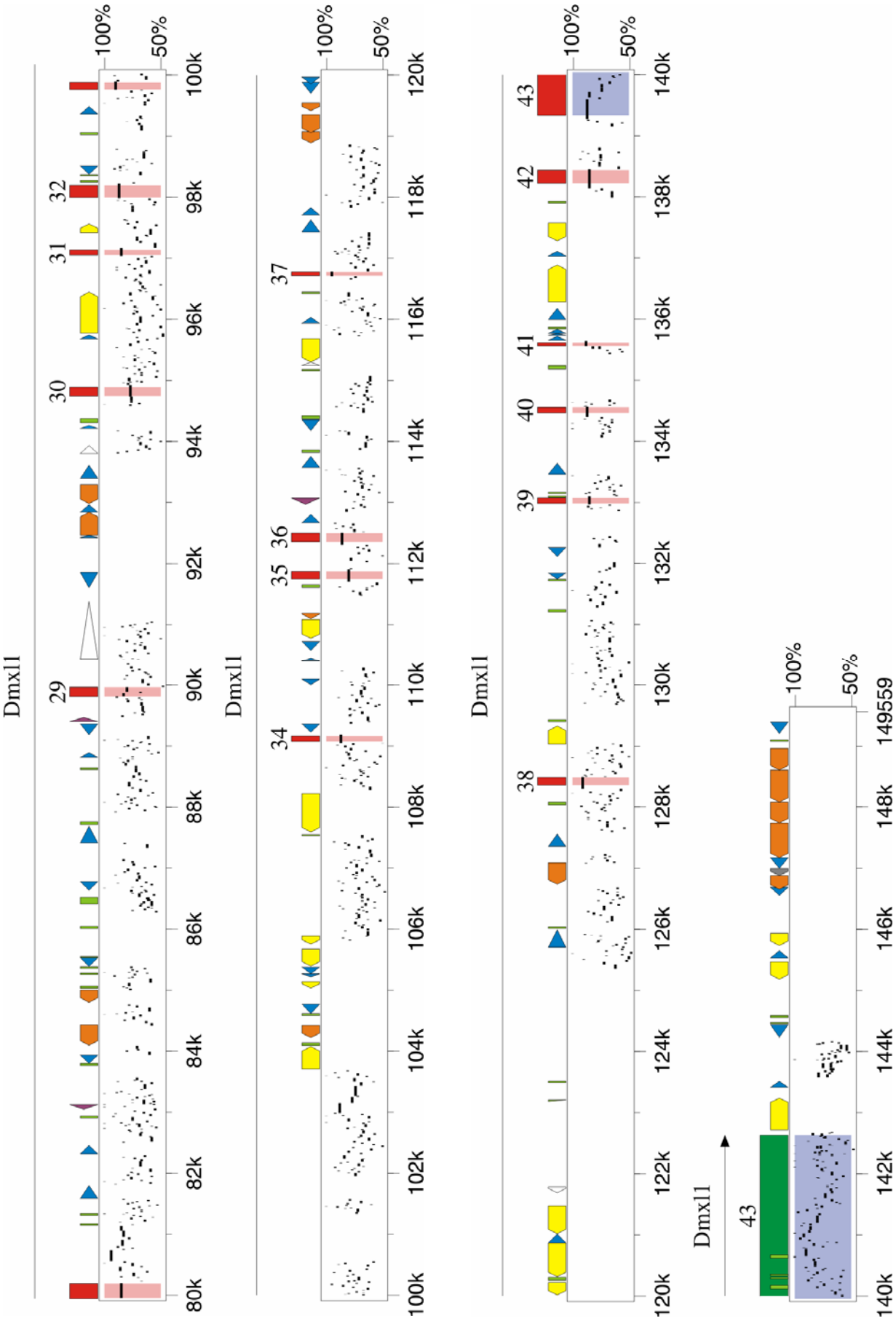
Der Vergleich der Maus/Mensch-Sequenzen erfolgte zunächst mit dem Programm SIM4. Auf diese Weise wurden alle Exons von *Dmxl*1 sowie 5'- und 3'-UTR des Gens in der genomischen Sequenz identifiziert. Anschließend wurde die genomische Sequenz mit dem Programm REPEATMASKER untersucht. Alle Daten wurden dann mit dem Programm Advanced PipMaker verarbeitet und durch einen „percent identity plot" (PIP) visualisiert. Hierbei ist die prozentuale Übereinstimmung der humanen genomischen Sequenz relativ zur murinen genomischen Sequenz dargestellt. Neben den Exons von *Dmxl*1 sind auch diverse repetitive Elemente, 5'- und 3'- UTR, GC-reiche Regionen sowie mögliche tRNA-Gene eingezeichnet. Eine Abbildungslegende befindet sich unten.

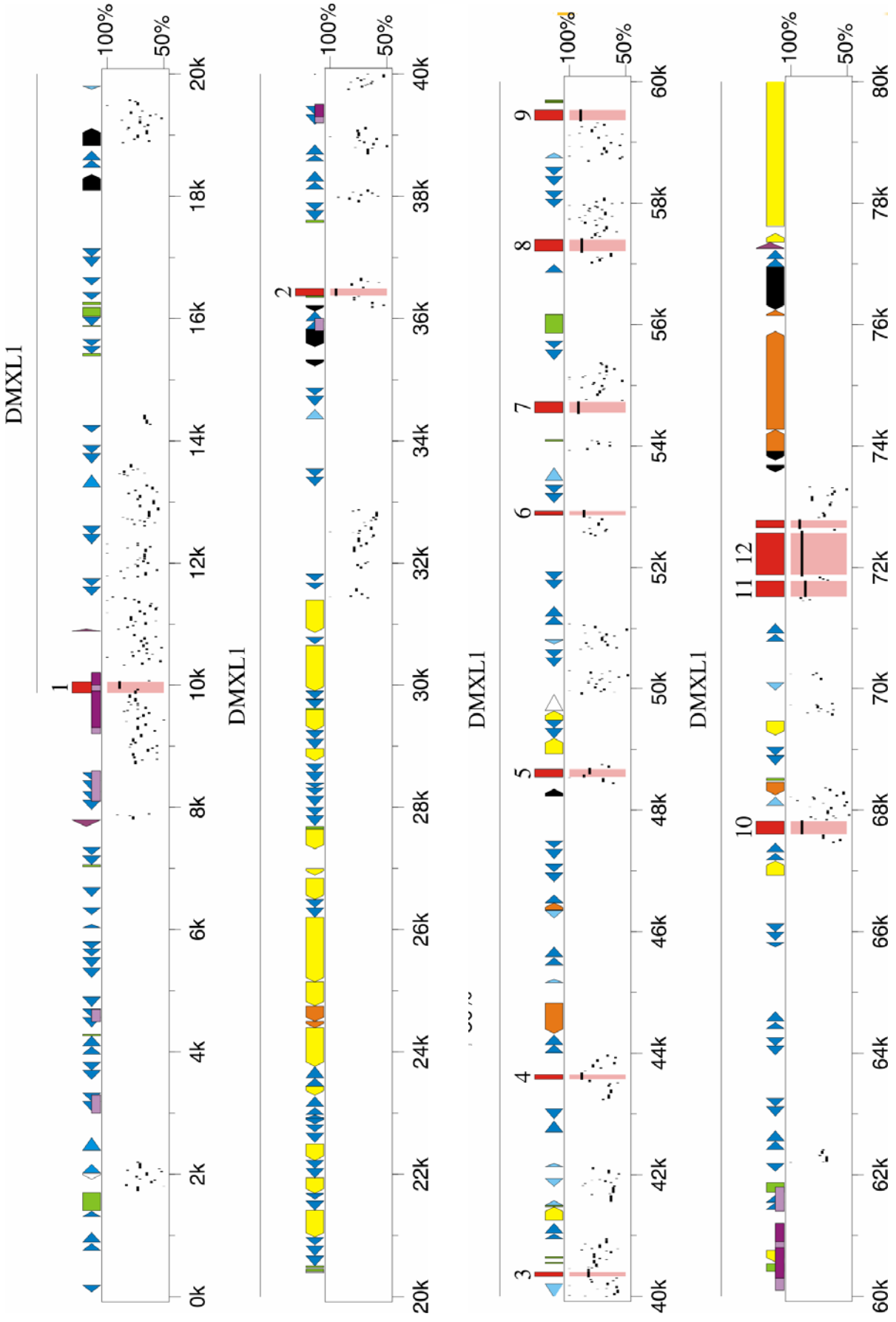
Abb. 3.21: PIP-Vergleich der humanen genomischen *DMXL*1-Sequenz mit der murinen (folgende Seiten).

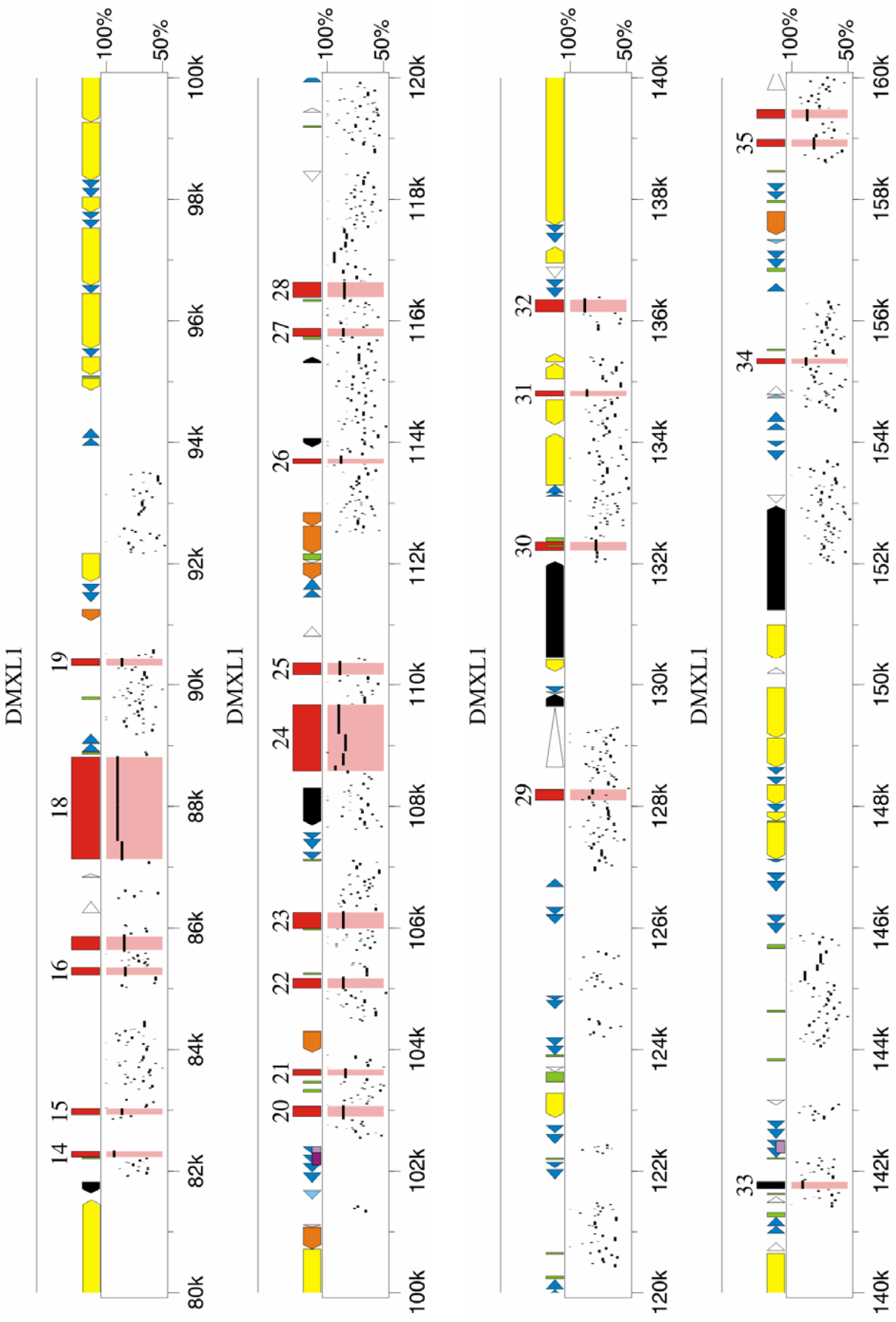
Der Vergleich mit den Mensch/Maus-Sequenzen erfolgte zunächst mit dem Programm SIM4. Auf diese Weise wurden alle Exons von *DMXL*1 sowie 5'- und 3'-UTR des Gens in der genomischen Sequenz identifiziert. Anschließend wurde die genomische Sequenz mit dem Programm REPEATMASKER untersucht. Alle Daten wurden dann mit dem Programm Advanced PipMaker verarbeitet und durch einen „percent identety plot" (PIP) visualisiert. Hierbei ist die prozentuale Übereinstimmung der murinen genomischen Sequenz relativ zur humanen genomischen Sequenz dargestellt. Neben den Exons von *DMXL*1 sind auch diverse repetitive Elemente, 5'- und 3'- UTR, GC-reiche Regionen sowie mögliche tRNA-Gene eingezeichnet. Eine Abbildungslegende befindet sich jeweils unter den Darstellungen.

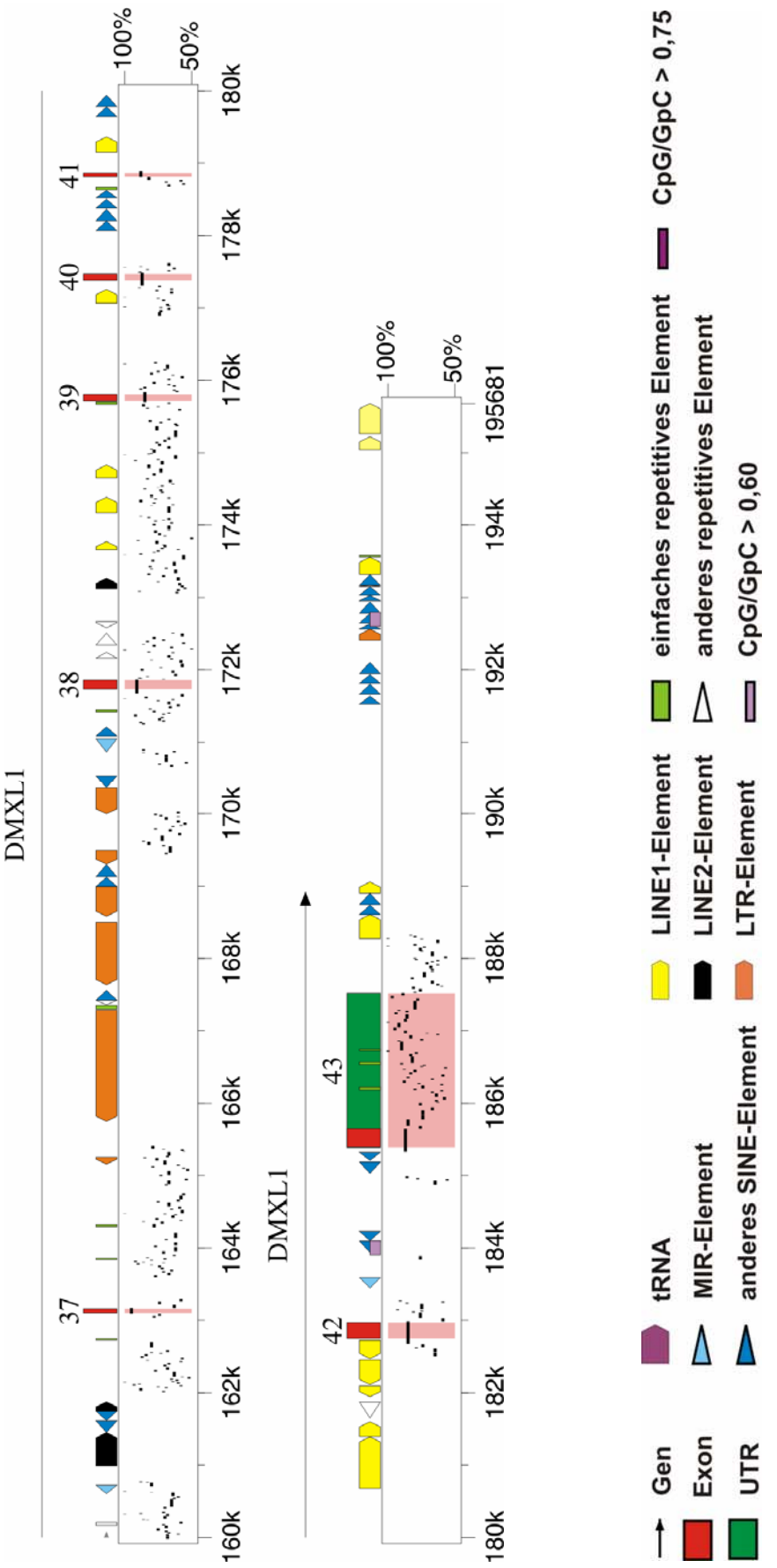



Dmxl1
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
1
2
Dmxl1
100%
50%
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
3
4
5
6
7
8
9
Dmxl1
100%
50%
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
11
12
14
16
17
18
19
Dmxl1
100%
50%
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k
20
21
22
23
24
25
26
27


Dmxl1
100%
50%
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k
29
30
31
32
Dmxl1
100%
50%
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k
34
35
36
37
Dmxl1
100%
50%
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k
38
39
40
41
42
43
Dmxl1
43
100%
50%
140k
142k
144k
146k
148k
149559


DMXL1
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
1
DMXL1
100%
50%
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
2
DMXL1
100%
50%
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
3
4
5
6
7
8
9
DMXL1
100%
50%
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k
10
11
12


DMXL1
100%
50%
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k
14
15
16
18
19
DMXL1
100%
50%
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k
20
21
22
23
24
25
26
27
28
DMXL1
100%
50%
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k
29
30
31
32
DMXL1
100%
50%
140k
142k
144k
146k
148k
150k
152k
154k
156k
158k
160k
33
34
35


DMXL1
37
38
39
40
41
100%
50%
160k
162k
164k
166k
168k
170k
172k
174k
176k
178k
180k
DMXL1
42
43
100%
50%
180k
182k
184k
186k
188k
190k
192k
194k
195681
Gen
Exon
UTR
tRNA
MIR-Element
anderes SINE-Element
LINE1-Element
LINE2-Element
LTR-Element
einfaches repetitives Element
anderes repetitives Element
CpG/GpC > 0,60
CpG/GpC > 0,75

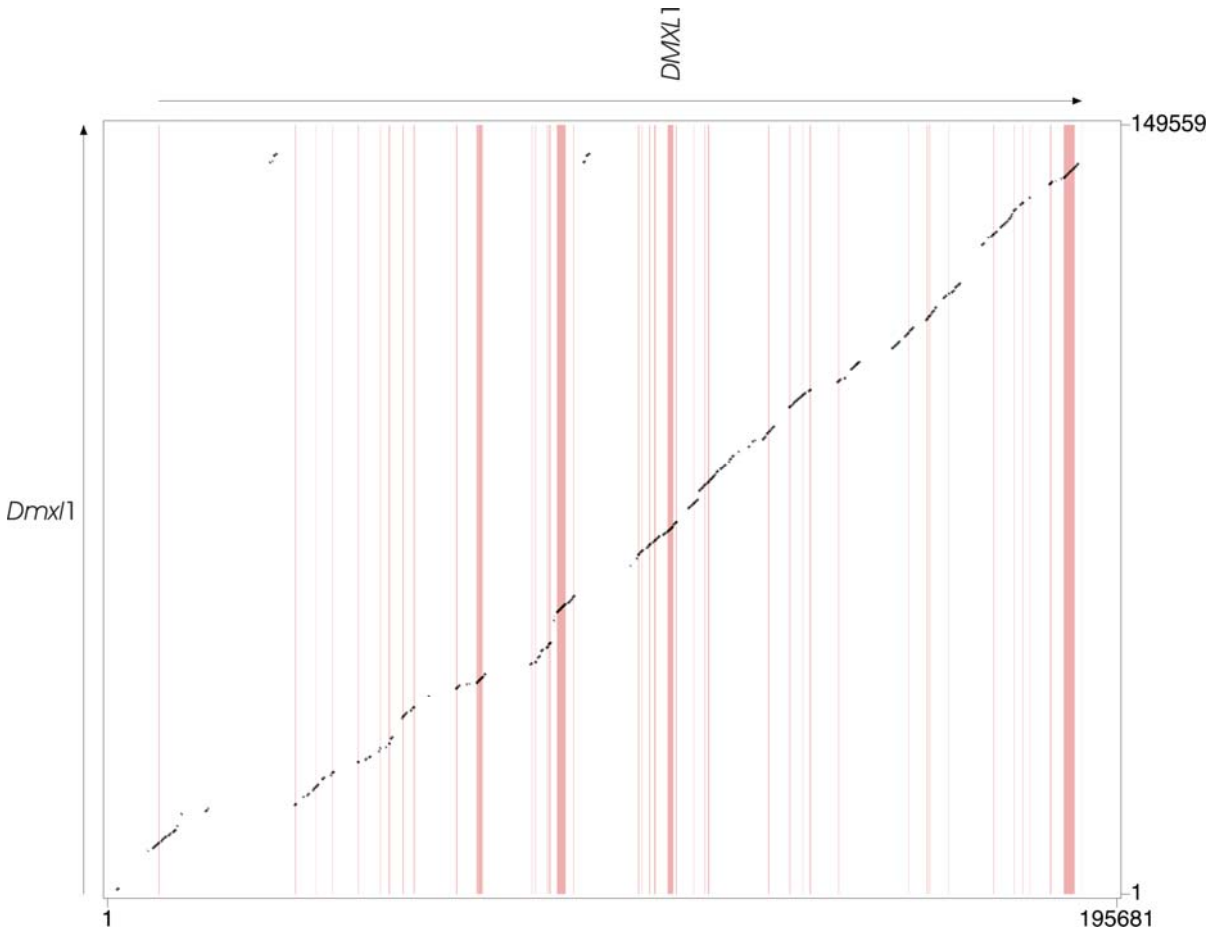


Abb.3.22: Vergleich der murinen genomischen *Dmxl*1-Sequenz (vertikal) mit der humanen Sequenz (horizontal) mit Hilfe der DOTPLOT-Option des Programms PipMaker.

Der Vergleich der beiden Sequenzen erfolgte mit den gleichen Parametern (mind. 50 % Übereinstimmung), mit denen auch der „percent identity plot“ (PIP) erstellt wurde. Die Orientierung der homologen Gene ist durch Pfeile am Rand der Abbildung gekennzeichnet, alle Exons der humanen Sequenz und ihre Position sind durch rote Balken dargestellt. Die Auswertung des Plots zeigt eine Unterbrechung der Diagonalen an 101 Positionen. Hier befinden sich entweder in der murinen oder humanen Sequenz Lücken, bedingt vor allem durch die unterschiedliche Länge und Sequenz der Intronbereiche und die höhere Anzahl repetitiver Elemente in der humanen Sequenz.

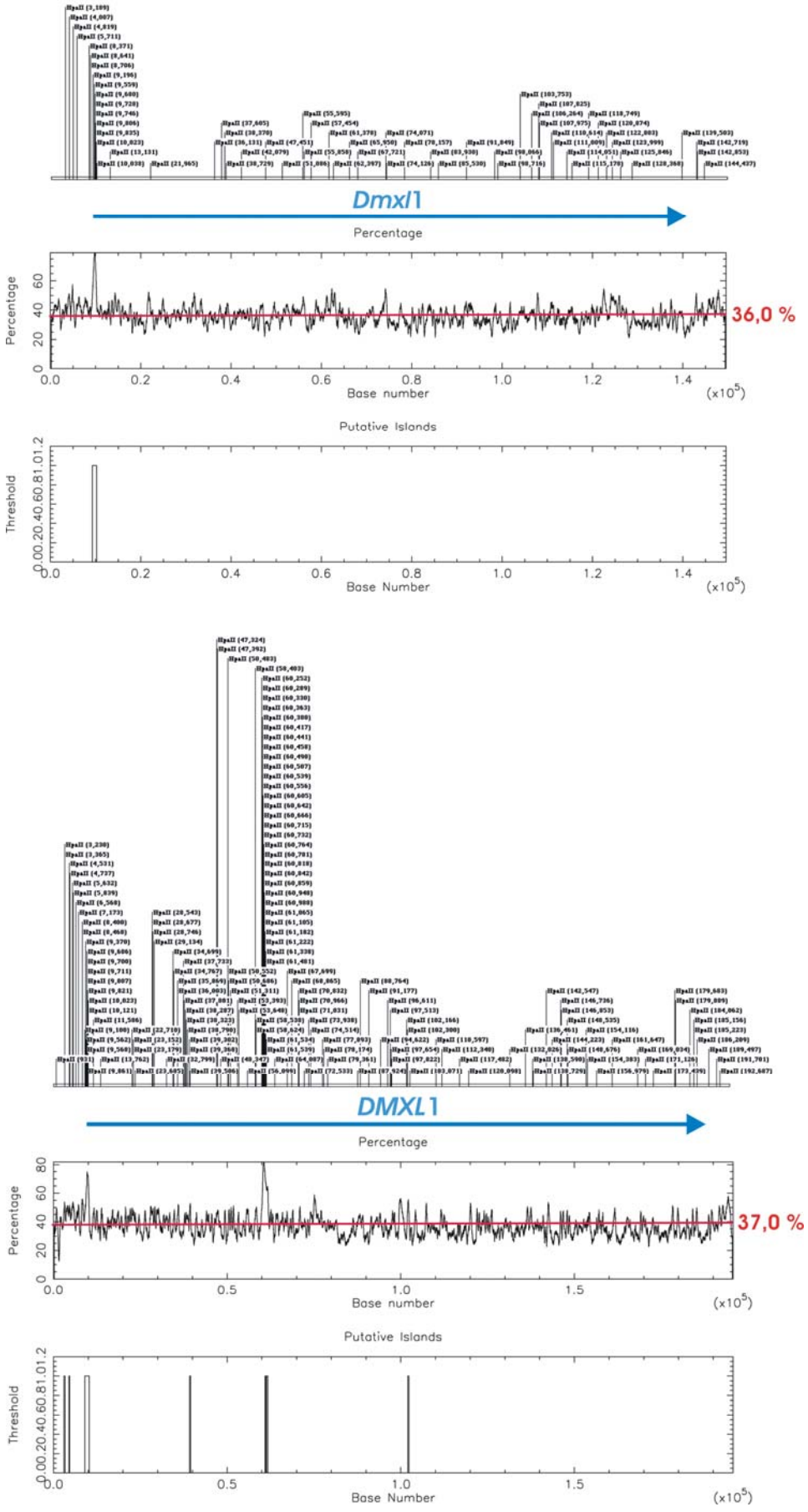
3.2.6 GC-Gehalt und „CpG-Inseln“

Der durchschnittliche GC-Gehalt der murinen Sequenz der *Dmxl*1-Region beträgt 36 %, für die humane *DMXL*1-Region wurden 37 % ermittelt. Die Abb 3.23 zeigt jeweils die Verteilung des GC-Gehalts und eine *Hpa* II-Restriktionskarte für die murine und humane Sequenz. Der GC-Gehalt beider Sequenzen zeigt deutliche Schwankungen, so variieren die Maus- und Mensch-Sequenzen in ihrem GC-Gehalt zwischen 20-60 %. In der Maus zeigt ein Bereich direkt vor dem ersten Exon von *Dmxl*1 einen GC-Anteil von

über 80 % und geht mit einer Häufung potenzieller *Hpa* II-Schnittstellen einher. Bei *DMXL*1 sind zwei Bereiche mit erhöhtem GC-Gehalt von 75 % bzw. 80 % zu erkennen, eine Vorhersage für potenzielle „CpG-Inseln“ erfolgt jedoch für sechs Positionen. Auch hier liegt der erste Bereich vor dem ersten Exon von *DMXL*1, der zweite Bereich liegt zwischen Exon 9 und Exon 10. Hierbei handelt es sich um ein GC-reiches repetitives Element niedriger Komplexität, dicht gefolgt von einer Reihe weiterer repetitiver Sequenzen. Ein hierzu homologer Abschnitt in der murinen Sequenz konnte nicht identifiziert werden, dieser Abschnitt ist zwischen Maus und Mensch nicht konserviert. Bei allen anderen vorhergesagten „CpG-Inseln“ handelt es sich ebenfalls um kurze repetitive Elemente mit hohem GC-Gehalt. Eine Überprüfung dieser Regionen mittels Datenbanksuchen zeigte keine relevanten Einträge. Übereinstimmungen wurden lediglich für die Abschnitte der repetitiven Elemente gefunden. Bei den in beiden Spezies vorhandenen konservierten Abschnitten mit erhöhtem GC-Gehalt vor dem ersten Exon von *Dmxl*1/*DMXL*1 handelt es sich wahrscheinlich um sogenannte „CpG-Inseln“ (Tykocinski et al., 1984; Bird et al., 1986). Diese nicht-methylierten, zweihundert bis zwei kb langen Bereiche im Säugergenom mit hohem GC-Gehalt sind in der Regel mit den 5’-Bereichen von Genen assoziiert (Cross et.al., 1995) und eignen sich daher als Anhaltspunkt für das Vorhandensein von Genen und markieren gegebenenfalls Promotorbereiche (Gardiner-Garden et al., 1987). Die Länge der identifizierten „GpG-Inseln“ variiert mit den verwendeten Programmen, so wird für die murine Sequenz von den Programmen CpG-Report, NEWCPGSEEK und CPGPLOT (alle EMBOSS-Programmpaket) eine „CpG-Insel“ für den Bereich 9419-10155 bp vorausgesagt. Das Programm CpGisland-revealing (WEBGENE) bestimmt einen Bereich von 9280-10292 bp, CpG-FINDER (Softberry) lokalisiert die „CpG-Insel“ in dem Bereich von 9433-10131 bp. Für die humane Sequenz verhält es sich ähnlich, die Programme CpG-Report, NEWCPGSEEK und CPGPLOT identifizieren eine „CpG-Insel“ im Bereich 9388-10143 bp, mit dem Programm von WEBGENE wurde ein Bereich von 8919-10272 bp bestimmt, der CpG-Finder von Softberry definierte den Abschnitt von 9259-10139 bp als „CpG-Insel“.

Abb 3.23: „GC-Plot“, *Hpa* II Restriktionskarte und Lage von *Dmxl*1 bzw. *DMXL*1 (folgende Seite)

Die Darstellung der „GC-Plots“ erfolgte mit dem Programm CPGPLOT (EMBOSS). Neben der Kalkulation des GC-Gehalts der genomischen Regionen von Maus und Mensch erfolgte eine Vorhersage möglicher „CpG-Inseln“. Die Voraussage für die murine Sequenz ist eindeutig, für die humane Sequenz werden insgesamt sechs mögliche „CpG-Inseln“ vorhergesagt. Fünf der möglichen „CpG-Inseln“ sind repetitive Elemente mit hohem GC-Gehalt, nur der Sequenzabschnitt vor dem Translationsstartpunkt von *DMXL*1 mit einem homologen Partner in der murinen Sequenz zeigt die Charakteristika einer „CpG-Insel“. Die *Hpa* II-Restriktionskarten wurden mit dem Programm SEQUENCHER ermittelt. Hier zeigt sich eine deutliche Korrelation der *Hpa* II-Schnittstellen mit dem GC-Gehalt der untersuchten Sequenz. Die Position von *Dmxl*1 und *DMXL*1 ist als blauer Pfeil eingezeichnet.


Dmxl1
Percentage
Percentage
36,0 %
Base number
(x10^5)
Putative Islands
Threshold
Base Number
DMXL1
Percentage
Percentage
37,0 %
Base number
(x10^5)
Putative Islands
Threshold
Base Number

3.2.7 Identifikation des Promotors von *Dmxl*1 durch Kombination von Promotorvorhersageprogrammen und Maus/Mensch-Sequenzvergleich

Zunächst wurden für beide Spezies 10000 bp des 5'-stromaufwärts gelegenen Sequenzbereichs vor dem Translationsstartpunkt von *Dmxl*1 und *DMXL*1 isoliert. Beide Sequenzabschnitte wurden anschließend mit dem Programm DOTPATH (EMBOSS-Programmpaket) miteinander verglichen. Da die Bindungsstellen für Transkriptionsfaktoren meist nur 10-15 Bp lang sind, wurde für diese Analyse das Suchfenster auf „Wordsize" 7 eingestellt. Auf diese Weise konnten hoch konservierte kurze Sequenzabschnitte für den Bereich 8500-10000 identifiziert werden (Abb.3.24). Weiter stromabwärts befindliche homologe Abschnitte erwiesen sich durch Abgleich mit erhobenen REPEATMASKER-Daten als repetitive Elemente oder Sequenzabschnitte niedriger Komplexität. Der so ermittelte Sequenzbereich wurde in Folge mit den Programmen DNA Block Aligner Form (DBA) und PromoterWise (beide EMBL-EBI Toolbox) untersucht. Beide Programme wurden speziell für die Analyse nichtkodierender syntäner Sequenzabschnitte mit konservierten kurzen colinearen Sequenzblöcken, die durch nichthomologe Abschnitte variabler Längen unterbrochen sind, entwickelt. Zudem können auch Sequenzen als Blöcke erkannt werden, die durch mehrere Lücken, Inversionen oder Translokationen unterbrochen wurden. Auch bei dieser Analyse wurde ein Abgleich mit den REPEATMASKER-Ergebnissen durchgeführt. Es konnten zehn homologe Sequenzblöcke identifiziert werden (Abb 3.25).

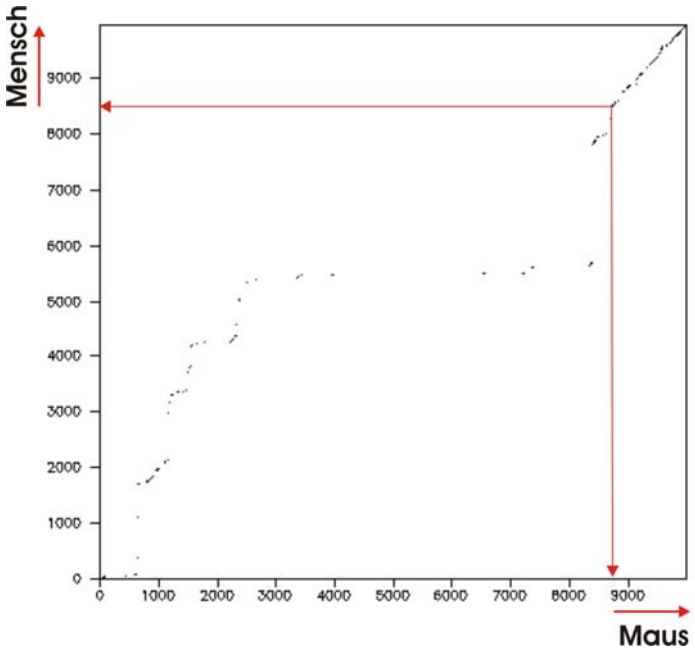


Abb. 3.24: DOTPATH-Analyse

Analysiert wurden 10000 bp 5'-stromabwärts des Transkriptionsstartpunkts von *Dmxl*1 und *DMXL*1. Eine deutliche Diagonale mit Sequenzblöcken hoher Homologie ist für den Bereich 8500-10000 zu erkenen. Alle weiteren homologen Abschnitte beruhen auf repetitiven Elementen oder Sequenzabschnitten niedriger Komplexität. Ein potenzieller Promotor sollte sich im Bereich bis -1500 vom Translationsstart befinden.

```
promomus    8473    AGTTTCTTGATTTATTT-GTTTCGAGACTGGTTCTCTTTCTGTAGCCCATGCTG
                B   AGTTT   G  TT TTT GTTT GAGAC GG TCTC  TCTGT  CCCA GCTG
promohum    8268    AGTTTTGAGGGTTGTTTGGTTTTGAGACGGGATCTCACTCTGTCACCCAGGCTG

>-----------<

promomus    8872    GGAAGCCATTTGAGCAGATTTTTGTTAATAATAAGACGAAAGAAGTGTCAAAA
                B   GG AG CAT T A CAGAT TTTGT AAT ATAA AC AAA   GTGTCAAAA
promohum    8706    GGGAGACATATAAACAGATGTTTGTCAATTATAATACAAAACTTGTGTCAAAA

>-----------<

promomus    8961    TCATTTTGAGATTCTTGGGTTCTTTTTTAAAA
                B   TCATT TG  ATTC T GGTTCTTTT  AAAA
promohum    8800    TCATTCTGGAATTCCTAGGTTCTTTTCCAAAA

>-----------<


promomus    8994    CAAAGTAGGACTGGTGGGTCATTGTTATCGGCACTATTTTAAA
                B   CAA G AGG CTGGTG GT  TT TTA C GCACTATTT AAA
promohum    8840    CAATGCAGGGCTGGTGTGTTGTTATTACCAGCACTATTTCAAA

>-----------<

promomus    9209    GGGGAAGTGGTGCATCTTGGGACTTGTAGTCATACTAGGAGTAAGGAACCTGCACCTCC
                C   GGGG AG GGTGCAT TTGGGACTTGTAGTC TAC AG A   AGGAA CTGCA CTCC
promohum    9063    GGGGGAGCGGTGCATATTGGGACTTGTAGTCCTACCAGAACCGAGGAAACTGCAACTCC

>-----------<

promomus    9274    TTGAACTTTCTGTCCTAACTGTTAGGTTGATCCCTTACCCGAAGCACAGCT
                B   TT AA TTTCTG CC AAC GTTAG T    CCCT ACC  AAGCACA CT
promohum    9144    TTAAATTTTCTGGCCCAACAGTTAGATGACGCCCTAACCTAAAGCACAACT

>-----------<
promomus    9388    CTCCTCCACAAGCCTCAAAGAGATTGGCTTTTTCT-C-GGAAGGGGGCGGAGCAGTAGGT-CGTTTG
                C   CTCCT  A AAGCCTCAAAGAGATTGGCTT TTCT C GG  G GGGCGGAGC  TAGGT  G   G
promohum    9270    CTCCTGTAGAAGCCTCAAAGAGATTGGCTT-TTCTCCGGGTGG-GGGCGGAGCCCTAGGTGGGACCG


promomus    9452    CGCCTCGGGATTGGC-CACTCACGGACTAGAAAAGGCGTT
                C    G  TCGGGATTGGC CA T  C GACTAGAA AGGCGTT
promohum    9335    GGGGTCGGGATTGGCCCAGTTTC-GACTAGAAGAGGCGTT

>-----------<


promomus    9535    TCTCGCGATACTTTGGCCGCAAGGGCCCGGAGCCGCGTAGCGAGCGCTGGCAGCCTG-TGAGGGCGG
                B   TCTCGCGATA TT GG CGCA GG  CC GAG CGCGTAG GAG G  GG  G CTG TG GGG GG
promohum    9451    TCTCGCGATATTTCGG-CGCAGGGA-CCAGAGGCGCGTAGTGAGTGGCGGAGG-CTGTTGGGGGTGG


promomus    9601    CGGGGGCCGCGGCTGGGCTGGGGCCGCTCTGGTGAGGAGGGCAAGCGCGGACTGCCGGCGCTTCTCG
                B   CGGG  CCG  GC GGG  GGG C G TC GG G GGA  GC AGCGCG ACTGC  GCGC  CTCG
promohum    9515    CGGG-ACCGGAGCCGGGAAGGGACAGGTCAGGCGGGGA--GC-AGCGCGTACTGCT-GCGCCACTCG


promomus    9668    GGCTGTGCCGGGCCCACCGACCCGGGCTCGGGAGGAATCGTGTGGACAGCTCGG-CCCCGCGGGCCG
                B   GGCTG G C GG  C  C  CCCG G TC  GAGG A  GTGTGGACAGC CGG CC   C GGCCG
promohum    9578    GGCTGGG-C-GGAGCGGCTGCCCG-GGTCCAGAGG-AGTGTGTGGACAGCGCGGCCCTTCCTGGCCG

>-----------<

promomus    9767    GTGACAGGTGCAGGAGCGAGCGCGGCGCCGCACTCGCTCCCCGGCCCGCCCCCCTCAGGCCTCGGCC
                B   GTGACAGGTGC  GA  GAGCGCGGCGC  C C C CTC CCG CCCGCCCCC  C GGCCTCG CC
promohum    9714    GTGACAGGTGCG-GAAGGAGCGCGGCGCTCCGCCCTCTCGCCGACCCGCCCCCTCCGGGCCTCGCCC
```

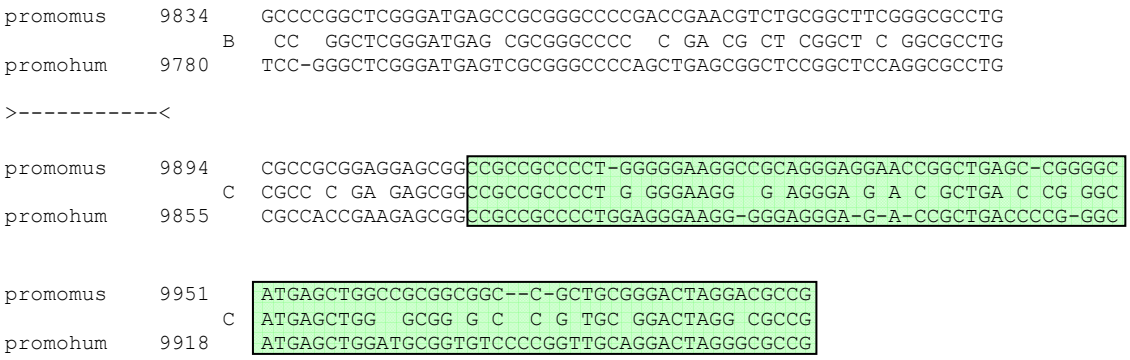


Abb. 3.25: Konservierte kolineare Sequenzblöcke der 5'-Stromaufwärts gelegenen genomischen Region von *Dmxl*1 und *DMXL*1

Durch die Analyse der syntänen genomischen Abschnitte von *Dmxl*1/*DMXL*1 in Maus und Mensch mit den Programmen DBA und PromoterWise konnten zehn hochkonservierte Sequenzblöcke im 5'-untranslatierten Bereich von DmX identifiziert werden. Diese Abschnitte stimmen mit den identifizierten „CpG-Inseln" in Maus und Mensch überein. Der grün unterlegte Abschnitt zeigt den 5'-untranslatierten Abschnitt der bekannten cDNA von DmX.

Neben der Identifikation konservierter Strukturen im 5'-stromaufwärts von *Dmxl*1 und *DMXL*1 gelegenen Sequenzbereich durch Interspeziesvergleich wurden zahlreiche Analysen mit Programmen zur Promotoridentifikation durchgeführt. Alle verwendeten Programme identifizierten Promotor und Bindungsstellen für verschiedene Transkriptionsfaktoren bei Maus und Mensch im Bereich der „CpG-Insel" bzw. der 5'-stromaufwärts gelegenen konservierten Strukturen. Grundsätzlich ist bei allen verwendeten Programmen, die Bindungsstellen für Transkriptionsfaktoren identifizieren, mit den höchsten Homologieeinstellungen gearbeitet worden, um die Anzahl falsch positiver Vorhersagen zu minimieren. Transkriptionsfaktoren binden an kurze 5-20 bp lange, z.T variable Sequenzen, die daher auch mit einer gewissen Wahrscheinlichkeit per se im Genom zu finden sind, ohne eine Funktion auszuüben. Es ist anzunehmen, dass bei den verwendeten Einstellungen nicht alle Transkriptionsfaktoren identifiziert wurden, die Qualität der vorhergesagten regulatorischen Faktoren jedoch sehr hoch ist. Das zurzeit umfangreichste und beste Programm auf diesem Gebiet ist GenomatixSuite der Genomatix Software GmbH. Dieses kommerziell vertriebene Programm kann von akademischen Nutzern in gewissem Umfang unentgeldlich verwendet werden. Es umfasst die Programme ElDorado, Gen2Promoter, GEMS Launcher, MatInspector und PromoterInspector. Die Analysen wurden sowohl für die Maus als auch für die Menschsequenz durchgeführt.

ElDorado: Die Promotoridentifikation bei diesem Programm basiert vor allem auf der Verwendung der vergleichenden Genomanalyse unter Einbeziehung von Datenbanksuchen (EST, nr, Proteine) und den Programmen MatInspector, ModelInspector und

PromoterInspector. Ein identifizierter Promotorbereich wird gegebenenfalls mit direkter Zuordnung zu einem Gen angezeigt. Zusätzlich werden Module (ModelInspector) von Bindungsstellen für Promotormodule mit höchster Homologie und Wahrscheinlichkeit für diesen Bereich aufgeführt (Abb. 3.26). Diese Module und die zugehörige Matrix basieren auf bekannten und experimentell verifizierten Promotoren. Für die Mausse-quenz wird ein Promotor für den Bereich 9501-10101 bp vorhergesagt. Der integrierte PromoterInspector spezifiziert einen DNA-Abschnitt von 9586-9961 bp als Promotor. Es wurden weiter zwei Module für diesen Abschnitt definiert. Im Bereich 9682-9731 bp binden SP1-Dimere („stimulating protein 1“, Seguin et al., 1987). Der Bereich 9766-9819 bp zeigt Bindestellen für SP1 und MyoD („myogenic determination protein“, Tapscott et al., 1988). Für die Menschsequenz wird ein Promotor für den Bereich 9382-9982 bp vorhergesagt, PromoterInspector identifiziert den Bereich 9653-9844 bp als Promotor. Es wurden acht Module in diesem Bereich gefunden, Sp1-Dimere können von 9706-9752 bp, MyoD und SP1 von 9714-9768 bp binden und finden in der Maus homologe Module. Weiter bindet MyoD in der Region 9689-9727 bp, SP1 und ETS in der Region 9706-9895 bp (plus-Strang) bzw. 9754-9789 bp (minus-Strang), SP1 und EGR („early growth response factor“, Joseph et al., 1988; Mora-Garcia et al., 2000) im Bereich 9738-9754 bp (plus-Strang) bzw. 9754-9770 bp (minus-Strang) sowie IKRS (Ikaros-Protein, Georgopoulos et al., 1992; Molnar et al., 1994) und AP-2 (activator protein 2, Courtois et al., 1990) im Abschnitt 9762-9799 bp. Viele Mitglieder der ETS-Familie gehören zu den basalen Transkriptionsfaktoren und sind an Embryonal- und Organentwicklung beteiligt (Kola et al., 1993; Wasylyk et al., 1993). Die Mitglieder der EGR-Familie zeigen Zinc-Finger Motive und binden in der Promotorregion vieler Haushaltsgene (Chavrier et al., 1990; Cao et al., 1990; Lemaire et al., 1990; Dorn et al., 1999). In der Maus werden diese beiden Module nicht gefunden.

MatInspector: Dieses Programm verwendet eine sehr umfangreiche Datenbank mit Matrixbeschreibungen von Transkriptionsfaktoren. Jeder gefundenen Übereinstimmung wird eine Wertschätzung zugeordnet. Diese Qualitätsbewertung beinhaltet drei separat definierbare Bereiche: Die „Core similarity“ wertet die Übereinstimmung mit den am höchsten konservierten Bereichen der Matrix. Die „Matrix similarity“ vergleicht jede Position der Bindungsstelle mit mit den am höchsten konservierten Nukleotiden dieser Position in der Matrix. Das „Optimized matrix threshold“ wird für jede Matrix separat bestimmt. Die Einstellung wird so gewählt, dass ein Minimum an Übereinstimmungen mit nicht-regulatorischen Sequenzen erzeugt wird, also die Anzahl falsch positiver Signale minimiert werden kann. Die Ergebnisse für die murine Sequenz dieser Untersuchung sind in Tabelle 3.8 und Abb. 3.27 dargestellt. Die humane Sequenz wurde ebenfalls untersucht, diese Ergebnisse befinden sich aus Platzgründen auf den beigefügten Datenträger.

Maus:

58971 (47064076)	59571 (47064676)	+	601 bp	**Promoter Region**	Promoter for Loc240283	No transcript assigned. Promoter based on Comparative Genomics	☐ Analyse Promoter
59056 (47064161)	59431 (47064536)	n/a	376 bp	**PromoterInspector Prediction**	PI008870	-	☐
59152 (47064257)	59201 (47064306)	±	50 bp	**Module**	SP1F_SP1F_01	-	☐
59236 (47064341)	59289 (47064394)	+	54 bp	**Module**	MYOD_SP1F_01	-	☐

Mensch:

Start rel. (abs.)	End rel. (abs.)	Strand	Length	Type of element	Name	Annotation	Extraction / Analysis
11541 (20821698)	12141 (20822298)	+	601 bp	**Promoter Region**	Promoter for Loc1657	Promoter for transcript(s): DMXL1/NM_005509	☐ Analyse Promoter
11812 (20821969)	12003 (20822160)	n/a	192 bp	**PromoterInspector Prediction**	PI006089	-	☐
11848 (20822005)	11886 (20822043)	+	39 bp	**Module**	MYOD_MYOD_01	-	☐
11865 (20822022)	12054 (20822211)	+	190 bp	**Module**	SP1F_ETSF_02	-	☐
11865 (20822022)	11911 (20822068)	±	47 bp	**Module**	SP1F_SP1F_01	-	☐
11873 (20822030)	11927 (20822084)	+	55 bp	**Module**	MYOD_SP1F_01	-	☐
11897 (20822054)	11913 (20822070)	-	17 bp	**Module**	EGRF_SP1F_01	-	☐
11913 (20822070)	11948 (20822105)	-	36 bp	**Module**	ETSF_SP1F_03	-	☐
11913 (20822070)	11929 (20822086)	-	17 bp	**Module**	EGRF_SP1F_01	-	☐
11921 (20822078)	11958 (20822115)	-	38 bp	**Module**	IKRS_AP2F_01	-	☐
12041 (20822198)	189678 (20999835)	+	177638 bp	Primary Transcript	DMXL1 NM_005509 Loc1657	Dmx-like 1	☐ • More Gene Info • Comparative Genomics • PubMed

Abb 3.26: Ergebnisse der Promotoranalyse mit dem Programm ElDorado

Für die Analyse der genomischen Region von *Dmxl*1 und *DMXL*1 durch das Programm ElDorado wurde automatisch auf Einträge in den öffentlichen Datenbanken zurückgegriffen. Hierbei wurde von dem Programm die untersuchte Sequenz genomisch lokalisiert bzw. einer Position in einem Contig zugeordnet. Da es sich nicht um die aktuellsten Datensätze handelte, differieren die Angaben der genomischen Lokalisation in beiden Spezies von den in dieser Arbeit ermittelten aktuellen Daten. Die erzielten Ergebnisse wurden dann auf den in dieser Arbeit verwendeten genomischen Datensatz umgerechnet. Die Analyse der Maussequenz zeigte keine Zuordnung zu einem bekannten Transkript, da die murine Sequenz von *Dmxl*1 erst am Ende dieser Arbeit annotiert wurde. Dennoch konnte für den Bereich, in dem auch für das humane *DMXL*1 ein Promotor vorhergesagt wurde, in der Maussequenz ein Promotorbereich identifiziert werden. Die in der Maus erkannten Module mit Bindestellen für Transkriptionsfaktoren finden sich auch in der humanen Sequenz. Eine TATA-Box konnte in keiner Sequenz ermittelt werden.

Family / matrix	Further Information	Opt.	Position	Str.	Core sim.	Matrix sim.	Sequence
V$NF1F/NF1.01	Nuclear factor 1	0.94	426 - 434	(+)	1.000	0.937	cttTGGCgcaagggc cgga
V$SIXF/SIX3.01	SIX3 / SIXdomain (SD) and Homeodomain (HD) transcription factor	1.00	445 - 451	(-)	1.000	0.966	cttATTAtt
V$TALE/TGIF.01	TG-interacting factor belonging to TALE class of homeodomain factors	1.00	446 - 458	(+)	1.000	1.000	tGTCAaa
V$AREB/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	0.98	454 - 466	(-)	1.000	0.906	tccatGTTTtgac
V$AREB/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	0.98	480 - 486	(-)	1.000	0.961	atccaGTTTccat
V$TALE/TGIF.01	TG-interacting factor belonging to TALE class of homeodomain factors	1.00	515 - 527	(-)	1.000	0.999	tGTCAgg
V$AREB/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	0.98	517 - 533	(-)	1.000	0.903	actttGTTTttaa
V$SORY/SRY.01	Sex-determining region Y gene product	0.95	534 - 544	(+)	1.000	0.923	aaaaACAAagtaggaact
V$AP1F/AP1FJ.01	Activator protein 1	0.97	737 - 749	(-)	1.000	0.940	aaTGACcacc
V$NKXH/NKX32.01	Homeodomain protein NKX3.2 (BAPX1, NKX3B, Bagpipe homolog)	0.96	737 - 749	(+)	1.000	0.950	ggggaAGTGgtgc
V$EBOR/DELTAEF1.01	deltaEF1	0.99	802 - 812	(-)	1.000	0.942	gaggAGGTgcaggtt
V$HEAT/HSF1.01	Heat shock factor 1	0.93	821 - 841	(-)	1.000	0.973	AGAAagttcaa
V$MOKF/MOK2.02	Ribonucleoprotein associated zinc finger protein MOK-2 (human)	0.98	827 - 839	(+)	1.000	0.907	tgttaggttgatcCCTTaccc
V$IKRS/IK2.01	Ikaros 2, potential regulator of lymphocyte differentiation	0.98	861 - 873	(-)	1.000	0.928	gtaaGGGAtcaac
V$GATA/GATA1.02	GATA-binding factor 1	0.99	971 - 983	(+)	1.000	0.992	tttaGATAactcc
V$AREB/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	0.98	983 - 995	(+)	1.000	0.948	ggtacGTTTgcgc
V$IKRS/IK2.01	Ikaros 2, potential regulator of lymphocyte differentiation	0.98	988 - 998	(+)	1.000	0.949	cctcGGGAttggc

Family / matrix	Further Information	Opt.	Position	Str.	Core sim.	Matrix sim.	Sequence
V$PCAT/CAAT.01	Cellular and viral CCAAT box	0.90	1073 - 1091	(-)	1.000	0.918	tgagCCAAtcc
V$SIXF/SIX3.01	SIX3 / SIXdomain (SD) and Homeodomain (HD) transcription factor	1.00	1078 - 1098	(-)	1.000	0.966	cttATTAtt
V$MOKF/MOK2.02	Ribonucleoprotein associated zinc finger protein MOK-2 (human)	0.98	1100 - 1110	(-)	1.000	0.970	acgcggctccggcCCTTgcgc
V$ZF5F/ZF5.01	Zinc finger / POZ domain transcription factor	0.95	1106 - 1122	(+)	1.000	0.951	gcgaGCGCtgg
V$AP4R/AP4.01	Activator protein 4	0.97	1108 - 1124	(+)	1.000	0.970	gctggCAGCtgctgagg
V$AP4R/AP4.01	Activator protein 4	0.97	1295 - 1303	(-)	1.000	0.934	gccctCAGCagctgcca
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	0.95	1294 - 1308	(+)	1.000	1.000	gtGACAggt
V$MYOD/MYOD.02	Myoblast determining factor	0.98	1296 - 1310	(-)	1.000	0.997	tctgCACCtgtcacg
V$EBOR/DELTAEF1.01	deltaEF1	0.99	1309 - 1319	(+)	1.000	0.959	tgacAGGTgcagagc
V$ZF5F/ZF5.01	Zinc finger / POZ domain transcription factor	0.95	1334 - 1348	(+)	1.000	0.957	gcgaGCGCggc
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	0.89	1449 - 1455	(-)	1.000	0.957	aggggGGCGggccgg
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	0.98		(+)	1.000	0.968	agGGGGa

Tab 3.8: Tabelle der identifizierten Bindungsstellen für Transkriptionsfaktoren.

Für diese Analyse wurde der genomische Bereich der konservierten Sequenzblöcke (Abb 3.25, 1528 bp) ausgewählt. In der Tabelle sind nur Faktoren dargestellt, die in allen drei Parametern („Core similarity“, „Matrix similarity“, „Optimized matrix threshold“, ohne Reihenfolge) eine Wichtung von >0,95 erreichten, um die Anzahl vorrausgesagter falsch positiver Faktoren zu minimieren. Allerdings kann es bei solcher Verfahrensweise auch zum Ausschluss von „richtigen“ potenziellen Bindungsstellen kommen. Aus diesem Grund befindet sich die komplette Tabelle auf dem beigefügten Datenträger. Ausgenommen wurden von dieser Regelung blockartig aufeinanderfolgende Bindestellen für Transkriptionsfaktoren und Bindungsstellen, die auch von anderen Programmen mit hoher Wahrscheinlichkeit erkannt wurden. Besonders auffällig ist bei den gelisteten Faktoren, dass es sich bei der überwiegenden Anzahl um basale Transkriptionsfaktoren handelt, die ihre Funktion in der Gruppe der Haushaltsgene ausüben. Die Kernsequenz („core“) der jeweiligen Bindungsstelle ist durch rote Großbuchstaben gekennzeichnet.

Abb. 3.27: Positionierung der Transkriptionsfaktoren

Dargestellt ist die Lokalisation aller identifizierten Transkriptionsfaktoren in der untersuchten Region (Abb. 3.25) von Maus und Mensch, erzeugt von dem Programm MatInspector. Die Farbkodierung entspricht der von Tabelle 3.8. Die untersuchte Maussequenz ist 1528 bp, die Menschsequenz 1743 bp lang. Die roten Balken symbolisieren den Promotorbereich und enthalten den potenziellen Transskriptionsstartpunkt (TSS). Der blaue Balken zeigt einen von dem Programm TSSG prognostizierten Enhancer (Maus) bzw. weiteren Promotorbereich (Mensch) an.

Weitere Analysen der Promotorregion wurden mit den Programmen DRAGON GENOME EXPLORER, TSSG und PromH(W) (beide Softberry) durchgeführt.

DRAGON GENOME EXPLORER: Das integrierte Programm DRAGON PROMOTOR FINDER v.1.4 sagt für die Maussequenz einen Promotorbereich von 9489 bis 9838 vorraus. Es wurde ein Transskriptionsstartpunkt (TSS) für die Position 9739 identifiziert. In der humanen Sequenz wird ein Promotor für den Bereich 9588-9957 angezeigt, der Transskriptionsstart für die Position 9838 vorrausgesagt. Eine Liste der identifizierten Bindungsstellen für Transkriptionsfaktoren befindet sich auf dem beigefügten Datenträger.

TSSG: In der Maus wurde ein Promotor für die Position 9453-9740, Transkriptionsstart 9740) bestimmt, ein Enhancerbereich für den Abschnitt 9095-9386. Für die Menschsequenz wurde der Bereich 9512-9810, Transskriptionsstart 9810, als Promotor definiert. Ein weiterer hochkonservierter Bereich mit potenzieller Promotorfunktion wurde für den Abschnitt 9043-9318 prognostiziert.

PromH(W): Dieses Programm detektiert Promotoren mit Hilfe der vergleichenden Genomanalyse. In der murinen Sequenz konnte ein Promotor für den Bereich 9819 (TSS) bestimmt werden, die lokale Übereinstimmung mit der verglichenen humanen Sequenz beträgt 78 % (-100, TSS+40), die lokale Übereinstimmung direkt am Transskriptionsstartpunkt beträgt 66 %, die homologie stromabwärts vom TSS beträt 87 % und die Homologie der regulatorischen Elemente im verglichenen Sequenzabschnitt beträgt 79 %. Für die Menschsequenz wird ein Promotor für den Bereich 9828 (TSS) identifiziert, die Übereinstimmungen betragen 73 %, 77 %, 61% und 82 % (gleiche Reihenfolge wie oben).

Alle verwendeten Programme identifizierten einen Promotor in der Maussequenz für das Gebiet 9400-9950, in der Menschsequenz von 9500-9880. Der Bereich des „Core"-Promotors wurde allerdings von den verwendeten Programmen jeweils innerhalb des angegebenen Sequenzabschnitts in unterschiedlichen Längen und Positionen identifiziert. Diese liegen jedoch alle im Gebiet der homologen Sequenzblöcke und der CpG-Inseln. Die meisten verwendeten Programme, bis auf TSSG, waren nicht in der Lage zwischen möglichen Promotorbereichen und regulatorischen bzw. transkriptionsverstärkenden Sequenzabschnitten (Enhancer) zu unterscheiden. Die gezeigten Ergebisse weisen jedoch darauf hin, dass sich direkt vor dem Promotorbereich ein weiterer regulatorisch wirksamer Sequenzabschnitt befindet, dies würde sich auch mit den Ergebnissen von Weil (2000) decken, der Untersuchungen regulatorische Breiche von *DmX* und *CpY* durchführte. Eine ausfühliche Besprechung des Promotors und der identifizierten Transkriptionsfaktoren erfolgt in Kapitel 4.4.5.

3.2.8 Identifikation möglicher regulatorischer Sequenzen im ersten Intron von *Dmxl*1 und *DMXL*1

Die Identifikation *cis*-regulatorischer Elemente gestaltet sich deutlich schwieriger als das Erkennen von Promotorstrukturen, da hier aktuell keine Programme existieren, die zufriedenstellend arbeiten. Aufgrund früherer Untersuchungen regulatorischer Bereiche von *DmX* und *CpY* (Weil, 2000), der konservierten Position des ersten Introns zwischen *DmX*, *CpY*, *F54E4.1*, *Dmxl*1 und *DMXL*1 (Kapitel 4.4.4) und der Größe des Introns in Maus und Mensch werden jedoch mögliche regulatorisch wirksame Sequenzabschnitte innerhalb des ersten Introns vermutet. Die in Kapitel 3.5.8 beschriebenen Analysemethoden können teilweise auch auf das erste Intron angewendet werden. Allerdings zeigen neuere Untersungungen mit bekannten Enhancer-Elementen, dass eine Identifikation der dort bindenenden Transkriptionsfaktoren durch Interspeziesvergleiche nicht ausreichend ist. Viele regulatorische Elemente zeigen speziesspezifische Unterschiede der dort bindenden Faktoren und auch die Abfolge der Bindungspositionen ist nicht immer konserviert (Velverde-Garduno et al., 2004; Long et al., 2003; Follows et al., 2003; Sasaki et al., 2002; Scemama et al., 2002; Silverman et al., 2002). Daher wurde hier eine zweigleisige Strategie aus kombinatorischer Vorhersage und Interspeziesvergleich gewählt. Zunächst wurde der „percent identity plot" (Abb. 3.20, 3.21) als Basis für weitere Untersuchungen verwendet. Hier zeigte ein Abschnitt von 3,5 kb des ersten Introns von *Dmxl*1 einen deutlich erhöhten Homologiegrad zu einem 4 kb großen Sequenzabschnitt des ersten Introns von *DMXL*1. Diese DNA-Sequenz wurde mit dem Programmen DBA und zPicture untersucht. Colineare Sequenzblöcke mit mehr als 75% Homologie zwischen Maus und Mensch konnten hierbei für die Bereiche 94-140 (271-316), 254-297 (395-436), 335-415 (472-553), 650-751 (704-805), 754-844 (891-974), 888-963 (1011-1086), 966-1044 (1113-1191), 1155-1263 (1305-1420), 1268-1361 (1724-1817), 1428-1457 (1898-1927), 1582-1764 (2012-2199), 2042-2140 (2701-2799), 2234-2417 (2908-3091) und 2442-2540 (3137-3237) identifiziert werden. Die Angaben beziehen sich auf die Nukleotidposition innerhalb des ersten Introns, die Positionierung im ersten Intron der Menschsequenz ist

in Klammern angegeben. Die visuelle Darstellung mit zPicture ist in Abb. 3.28 gezeigt. Eine Analyse dieser Region mit den Programmen ElDorado und FrameWorker zeigt sechs regulatorische Module für die Maus-Region: NF-κB/Sox-9 (488-472/322-338), AP-1/NFAT (630-646/588-598), MEF2/T3R (730-708/708-690), FoxF2/FoxI1a (743-727/726-710), Sox-5/SF1F (1201-1185/1233-1245) und HLF/TEF (2661-2677/3115-3099). Für die humane Sequenz werden ebenfalls Module vorrausgesagt. Hier binden die Faktoren AP-1/ELK-1 (1358-1348/1367-1351), STAT5/NFY (1712-1730 bzw. 1730-1712/1756-1724), NFAT/AP-1 (2112-2102/2102-2092), Sox-9/ELK-1 (2312-2328/2354-2338), SMAD3/AP-1 (2478-2486/2474-2464), MEF2/T3R (2567-2589/2501-2483) und OCT1/CEBP (3327-3341 bzw. 3382-3391/3404-3390). Bei dieser Analyse wurde die DNA-Sequenz auf ein definiertes Netzwerk von Bindestellen für Transkriptionsfaktoren hin untersucht. Hierbei spielt vor allem die Kombination und Entfernung der Bindungspositionen verschiedener Transkriptionsfaktoren zueinander eine Rolle, die Stringenz der Programme lässt sich nicht beeinflussen, ein Interspezies-Vergleich wird nicht durchgeführt. Bei einer hochstringenten Suche mit MathInspector oder DRAGON PROMOTOR FINDER wurden einzelne Faktoren vorausgesagter Module nicht erkannt, betroffene Module wurden daher als nicht relevant eingestuft und sind hier nicht aufgeführt. Gleiche Module sind oben im Text rot gekennzeichnet, gleiche Transkriptionsfaktoren, die mit einem anderen Partner kombinieren sind blau dargestellt. Einige der gezeigten Module bzw. Transkriptionsfaktoren liegen außerhalb der zwischen Maus und Mensch konservierten Sequenzabschnitte. Neben dieser kombinatorischen Analyse wurden zusätzlich Faktoren, die innerhalb der Sequenzblöcke binden, mit verschiedenen Programmen (MathInspector, Dragon Genome Explorer, Match Search, rVista) in Maus und Mensch identifiziert. Alle Ergebnisse wurden anschließend miteinander verglichen und nur die zwischen Maus und Mensch hochkonservierten Transkriptionsfaktoren, die von mehreren Programmen identifiziert wurden und sich in einer vergleichbaren Position befinden, als potenzielle *cis*-regulatorischer Elemente extrahiert (Tab. 3.9).

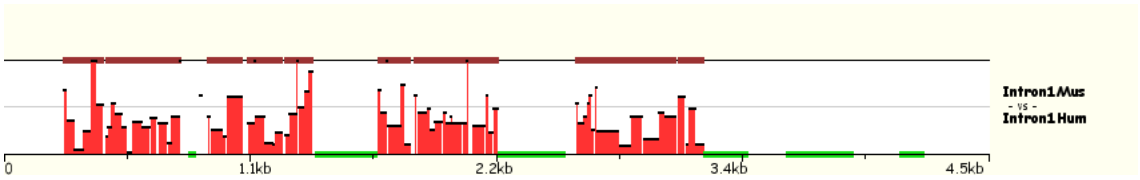


Abb. 3.29: zPicture-Analyse des ersten Introns von Maus und Mensch.

Gezeigt ist ein blastz-Alignment des ersten Introns von Maus und Mensch. Nur Sequenzabschnitte mit einer Homologie über 50% werden hier angezeigt, die mittlere Linie entspricht 75%. Übereinstimmungen konnten bei dieser Analyse für die ersten 3,5 kb gefunden werden, danach bricht die Homologie zwischen Maus und Mensch ab. Aus Platzgründen sind hier nur 4,5 kb des Alignments gezeigt. Grüne Bereiche repräsentieren repetitive Elemente in der DNA-Sequenz.

Transkriptionsfaktor	Position Maus	Position Mensch
NF I	82-100 (-)	259-277 (-)
RORα1	250-262 (+)	391-403 (+)
LEF-1	265-271 (+)	408-414 (+)
SMAD3	282-290 (-)	424-432 (-)
FoxA2	391-407 (+)	487-503 (-)
STAT5A	1247-1261 (+)	1391-1405 (+)
FoxO4 (AFX)	1815-1825 (-)	2262-2272 (-)
TCF4	2052-2059 (+)	2710-2717 (+)

Tab. 3.9: Interspeziesvergleich konservierter Transkriptionsfaktoren

Bindungspositionen der identifizierten konservierten Transkriptionsfaktoren im ersten Intron von Maus und Mensch.

3.2.9 Identifikation von Polyadenylierungssignalen

Die Identifizierung des 3' nichttranslatierten Abschnitts von *Dmxl*1 erfolgte teilweise durch Sequenzierung von cDNA-Klonen (Kapitel 3.1.3). Hierbei zeigte sich, dass drei der Klone (IMAGp998B022563Q2, IMAGp998E173005Q2, IMAGp998M019196Q2 (RZPD, I.M.A.G.E. Consortium)) zwar eine unterschiedliche Länge aufwiesen, jedoch an gleicher Position einen PolyA-Abschnitt zeigten (polyA1m). Der cDNA-Klon (IMAGp998J193130Q2 (RZPD, I.M.A.G.E. Consortium)) überspannte diesen Bereich und zeigte erst 1218 bp später eine charakteristische Abfolge von Adenin-Nukleotiden (polyA2). Beide Bereiche wurden anschließend auf genomischer Ebene per Auge auf mögliche Polyadenylierungssignale (AAUAAA bzw. AUUAAA, Proudfoot et al., 1976; Wickens, 1990) durchsucht: vierzehn Basenpaare vor polyA1 (Abb 3.29; genomisch Position 141419 bp, Position auf der cDNA 10991 bp) befindet sich die charakteristische Hexamersequenz AAUAAA (Position -20 – -15). Ein weiteres wichtiges Signal für eine Polyadenylierung sind ein oder mehrere U(G)-reiche Sequenzelemente, die sich <20-60 bp stromabwärts der Hexamersequenz befinden (Gil et al., 1984,1986; McDevitt et al., 1984; McLauchlan et al., 1985; Chou et al., 1994; MacDonald et al., 1994; Takagaki et al., 1997; Zarudnaya et al., 2003). Für polyA1 konnten in dem Bereich 15-37 bp hinter der Hexamersequenz drei Abschnitte aufeinanderfolgender U(G)-reicher Regionen identifiziert werden (Position +1 – +23). Ein CA-Dinukleotid ist ein weiterer, aber nicht so häufig vorkommender Bestandteil des Polyadenylierungssignals und befindet sich normalerweise in Position -1 und +1 der Polyadenylierungsstelle (Chen et al., 1995). Ein solches Element konnte für polyA1 nicht identifiziert werden. Die Untersuchung der polyA2 (Abb 3.29; genomische Position 142639, Position auf der cDNA 12210 bp) umgebenden Sequenz zeigte in Position -20 – -26 die Hexamersequenz AUUAAA, exakt an der Polyadenylierunggsstelle befindet sich ein CA-Dinukleotid und in Position +18 – +52 finden sich sich mehrere U(G)-reiche Elemente.

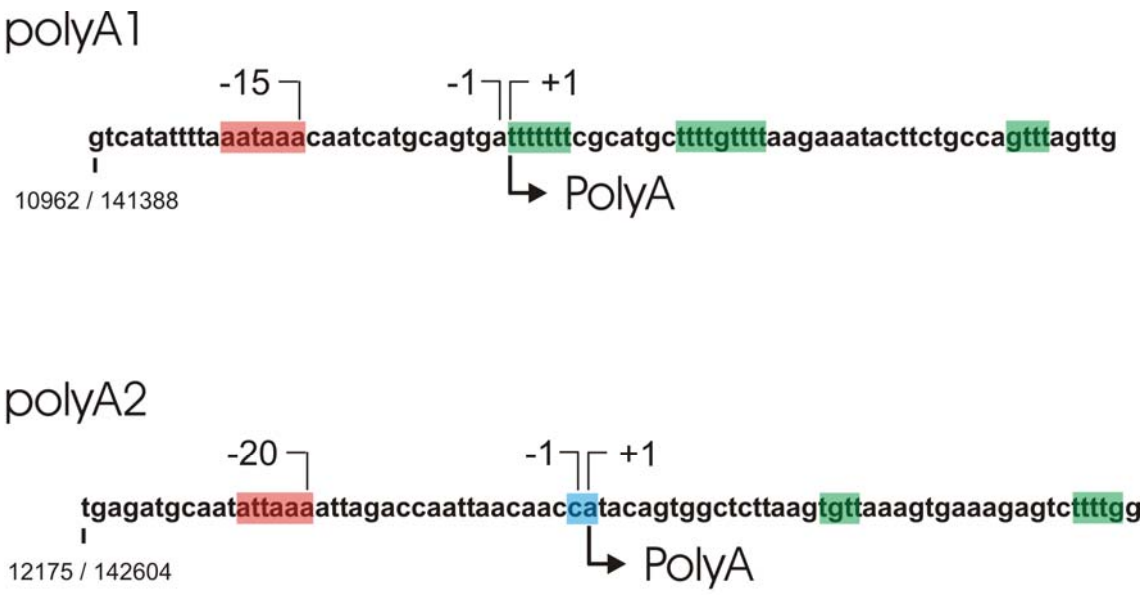


Abb. 3.30: Alternative Polyadenylierungssignale von *Dmxl*1

Dargestellt ist die genomische Sequenz von beiden potenziellen Polyadenylierungssignalen. Die Positionsangabe der jeweils ersten aufgeführten Base beschreibt ihre Lage in der cDNA bzw. genomische Lokalisation. Die Polyadenylierungsstelle ist mit PolyA gekennzeichnet. Das Polyadenylierungssignal ist rot, U(G)-reiche Sequenzelemente grün und das CA-Dinucleotid blau unterlegt. Die weiteren Positionsangaben beziehen sich auf die Polyadenylierungsstelle.

Beide Sequenzabschnitte mit den potenziellen Polyadenylierungssignalen wurden anschließend mit den Programmen POLYA SCAN (Hornischer und Blöcker, unpubliziert), ERPIN (Gautheret et al., 2001) und POLYADQ (Tabaska et al., 1999) untersucht. Alle drei Programme erkannten mit den Standardeinstellungen nur polyA1 als vollwertiges PolyA-Signal, polyA2 wurde nicht erkannt. Bei Veränderung der Einstellung (Erhöhung) bezüglich der Entfernung der U(G)-reichen Elemente zur Polyadenylierungsposition wurde auch polyA2 von POLYA SCAN erkannt, POLYADQ erkannte das Signal nur mit einer geringeren Wahrscheinlichkeit. ERPIN ließ keine Veränderungen der Einstellungen zu. Mit der Sequenz der alternativen Transkripte wurden anschließend EST-Datenbanksuchen durchgeführt. Hier zeigte sich, dass die meisten eingetragenen 3'-cDNA-Fragmente mit polyA1 identisch sind, es findet sich jedoch auch ein Eintrag für das alternative Ende polyA2 (Abb. 3.29a), es handelt sich hierbei um den sequenzierten cDNA-Klon. Die humane cDNA-Sequenz, die mit 2 kb Sequenzdaten aus den öffentlichen Datenbanken 3'-terminal verlängert wurde, zeigt bei einer EST-Datenbanksuche ebenfalls alternative Transkripte. Die identifizierten EST-Datenbankeinträge der cDNA-Klone enden jedoch an unterschiedlichen Positionen in der A/T-reichen Region eines Alu-Elements, das sich genomisch in gleicher transkriptionellen Orientierung wie *DMXL*1 befindet. Da keiner der EST-Klone mit dem humanen polyA1 überlappt, und es sich bei den meisten EST-Sequenzen in den öffentlichen Datenbanken nur um ähnliche Transkripte von Alu-Elementen handelt,

muss die Möglichkeit in Betracht gezogen werden, dass es bei der Herstellung der cDNA-Klone zu einer Kontamination mit genomischer DNA gekommen ist und die verwendeten Primer mit der A/T-reichen Region des Alu-Elements hybridisiert haben. Im 3'-Bereich dieser A/T-reichen Region ist allerdings durch eine A→T Transversion die charakteristische Hexamersequenz AAUAAA entstanden. Die murine genomische Sequenz der umgebenden Region von polyA1 und polyA2 wurde anschliessend mit der humanen Sequenz verglichen (Abb. 3.30), hierbei zeigte sich, dass nur der Sequenzabschnitt von polyA1 zwischen beiden Spezies hochkonserviert ist. Stromabwärts von polyA1 bricht die homologie zwischen Maus und Mensch ab.

```
Maus : 7  acattatgtaagtccatatgtatttacatccagagtc---atattttaaataaacaatca 63
          ||||||||||||| ||||||||||||||||||||||||   |||||||||||||||||||
Mensch:3  acattatgtaagcccatatgtatttacatccagagtcataatattttaaataaacaatca 62


Maus : 64 tgcagtgaTTTTTTTCGCATGC
          |||||
Mensch:63 tgcagaaacttttTTTAGGGGGT
```

Abb. 3.31: Vergleich der umgebenden Sequenz von polyA1 zwischen Maus und Mensch

Die durchschnittliche Sequenzübereinstimmung des 3'-UTR zwischen Maus und Mensch beträgt 78 % (Daten nicht gezeigt), in der direkten Umgebung von polyA1m und polyA1h steigt diese auf 94 %. Das Polyadenylierungssignal ist in diese Abbildung rot unterlegt, die Grossbuchstaben mit oranger Unterlegung zeigen die genomische Sequenz, in dieser Position befindet sich in der prozessierten RNA der PolyA-Schwanz.

Abschliessend wurde der gesamte 3'-untranslatierte Abschnitt von *Dmxl*1 auf weitere funktionelle Elemente hin untersucht, hierfür wurden das Programm UTRscan (Pesole et al., 1999, 2002) eingesetzt. In der Maussequenz konnte das Heptamermotiv AGCTTTA (Position 9907) 1060 bp vor dem ersten Polyadenylierungssignal identifiziert werden. Es handelt sich hierbei um eine Brd-Box (Lai et al., 1997, 1998), einen negativen Regulator, der die Anzahl der Transkripte durch Deadenylierung oder Inhibition der Polyadenylierung reguliert. Eine solche Box kann auch im 3'-UTR der humanen cDNA gefunden werden, allerdings in einer anderen Position (10904), 152 bp vor dem Polyadenylierungssignal. Weitere Motive wurden nicht identifiziert. Die Sekundärstruktur und freie Energie aller 3'-UTR-Bereiche wurde mit dem Programm Vienna RNA secondary stucture server (Hofacker, 2003) ermittelt, die Ergebnisse befinden sich auf dem beigefügten Datenträger.

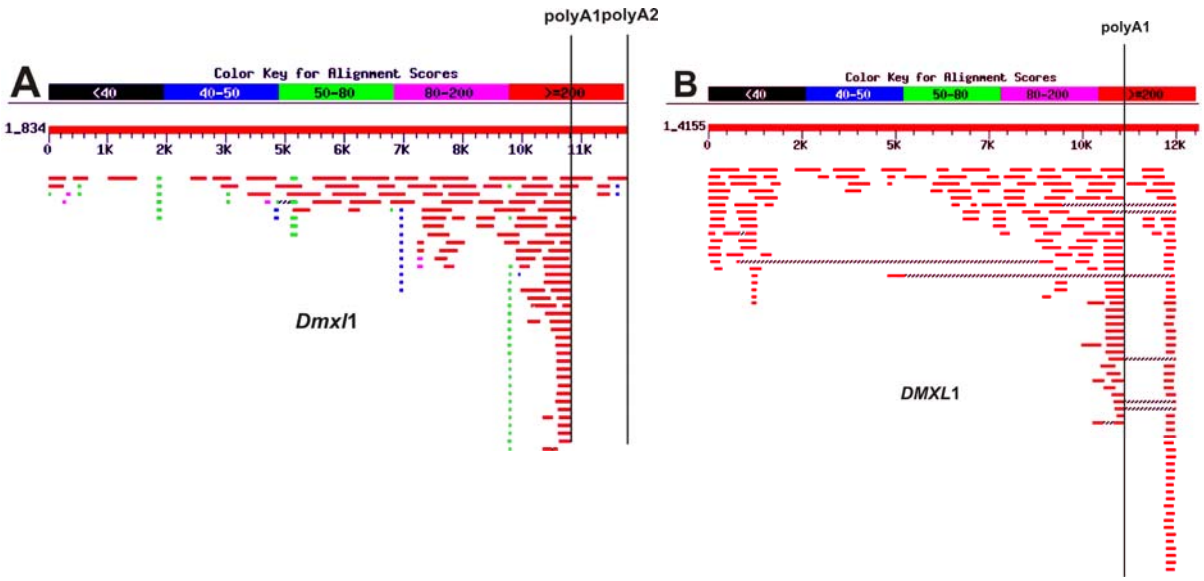


Abb. 3.32: EST-Datenbanksuche mit der cDNA von *Dmxl*1 und *DMXL*1

Für die mausspezifische EST-Datenbanksuche wurde die 12210 bp lange murine cDNA verwendet, die beide möglichen Polyadenylierungssignale enthält. Die humane cDNA wurde für Datenbanksuche durch genomische Sequenzdaten ergänzt. Die meisten Einträge für *Dmxl*1 enden mit einer Sequenz, die einer Polyadenylierung durch polyA1 entspricht. Nur ein Eintrag endet mit der alternativen Polyadenylierungsstelle polyA2, es handelt sich hierbei um den in dieser Arbeit sequenzierten cDNA-Klon IMAGp998J193130Q2 (RZPD, I.M.A.G.E. Consortium). Der Klon stammt aus einer „mamary gland" cDNA-Bibliothek nach Soares (Bonaldo et al., 1996; Lennon et al., 1996). Für die humane Sequenz findet sich eine vielzahl alternativer Einträge, jedoch enden diese an unterschiedlichen Positionen in den A/T-reichen Regionen eines Alu-Elements. Bei den zahlreichen kurzen EST-Datenbankeinträgen stromabwärts des humanen polyA1 handelt es sich um Transkripte von Alu-Elementen.

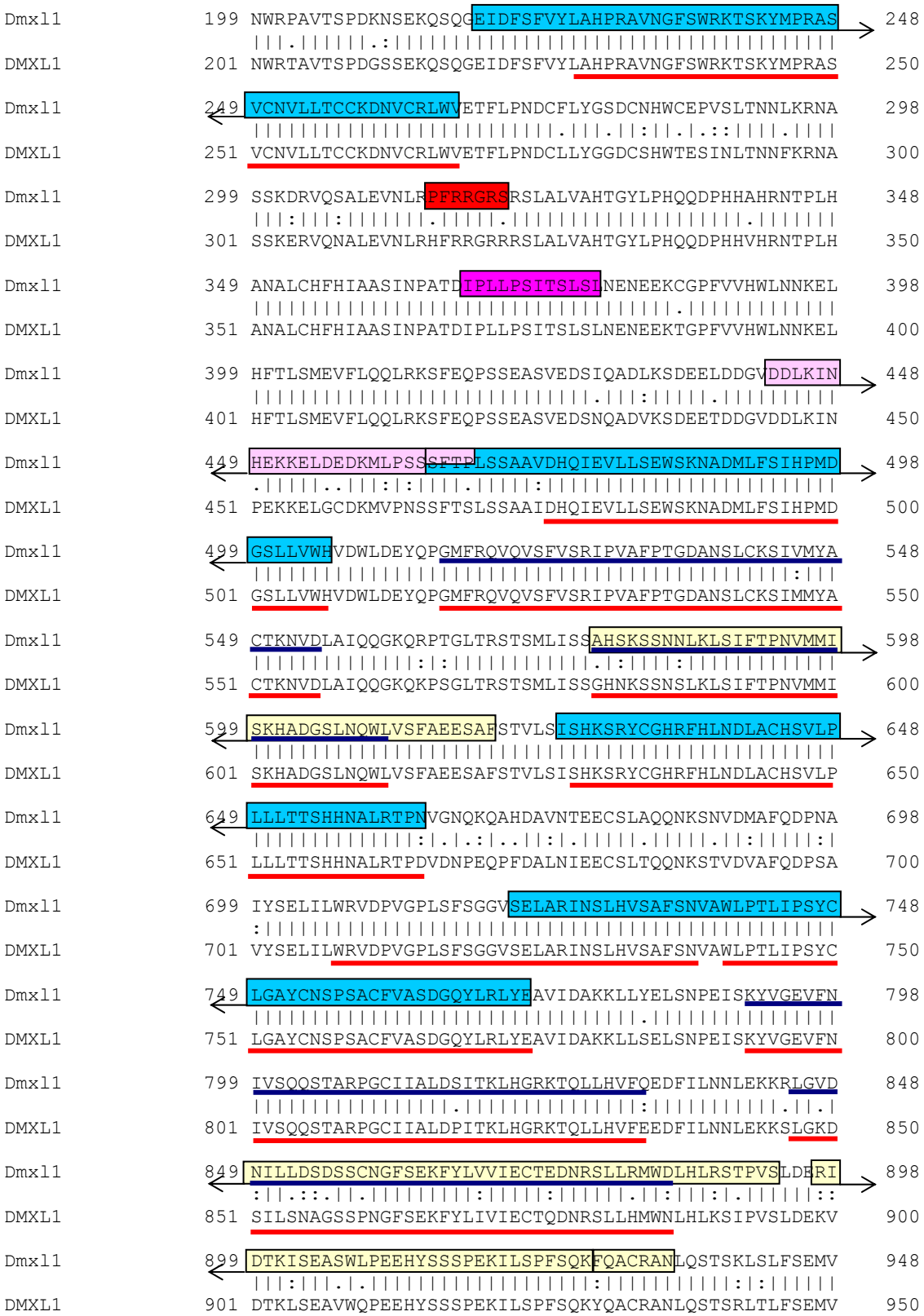
3.2.10 Vergleich der cDNA und Aminosäuresequenzen von DmX zwischen Maus, Mensch und anderen Spezies

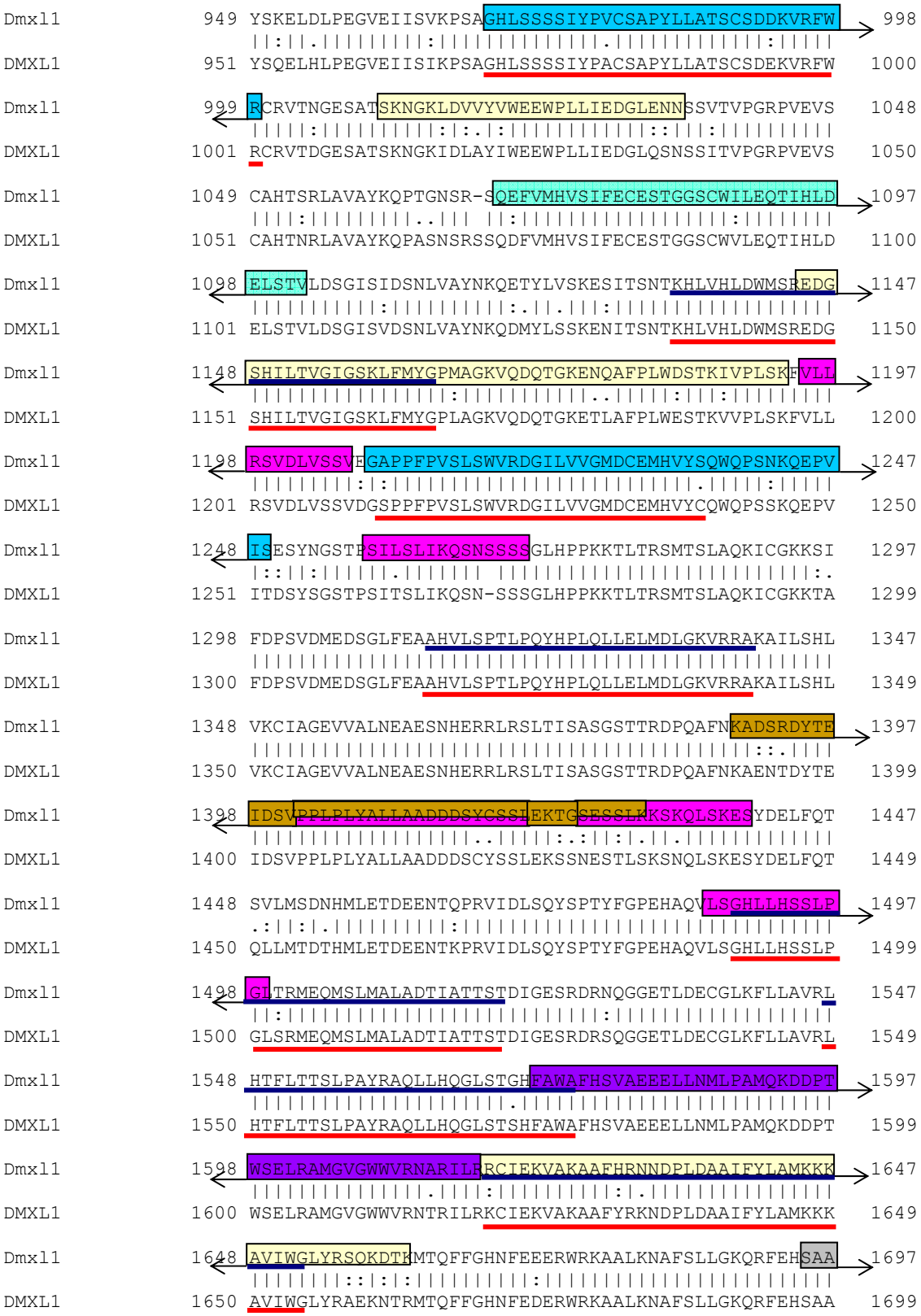
Für den Vergleich der codierenden Sequenzabschnitte bzw. der abgeleiteten Aminosäuresequenzen von DmX wurden die Programme ALIGN „water" bzw. „needle" (beide EMBOSS-Programmpaket), MEGALIGN (DNASTAR), MALIGN und PRETTYPLOT (HUSAR-Programmpaket) verwendet. Aus Platzgründen befinden sich die Vergleiche der Nukleotidsequenzen und die multispezies Aminosäurevergleiche (MALIGN, PRETTYPLOT) auf dem beigefügten Datenträger. Zwischen Maus und Mensch besteht nach Analyse mit dem Programm "Align-water" (Abb.3.33) über die gesamte Aminosäuresequenz eine Identität von 89,3 % (2707/3030) und eine Similarität von 94,7 % (2869/3030) bei 20 Lücken. Auf Nukleotidebene beläuft sich die Identität der CDS auf 88,9 % (8096/9105) mit 84 Lücken. Da es sich hierbei um Durchschnittswerte handelt und diese nur bedingt aussagekräftig sind, wurde eine Analyse der Nukleotidsequenz mit dem Programm MEGALIGN durchgeführt und der Homologiegrad für jedes Exon extrahiert (Tab. 3.10). Hier zeigte sich ein wesentlich differenzierteres Bild, so reicht die Spannne von 66,3 % Identität für Exon 29, das in der Maus im Vergleich zum Menschen zwei Lücken mit 24 fehlenden Nukleotiden

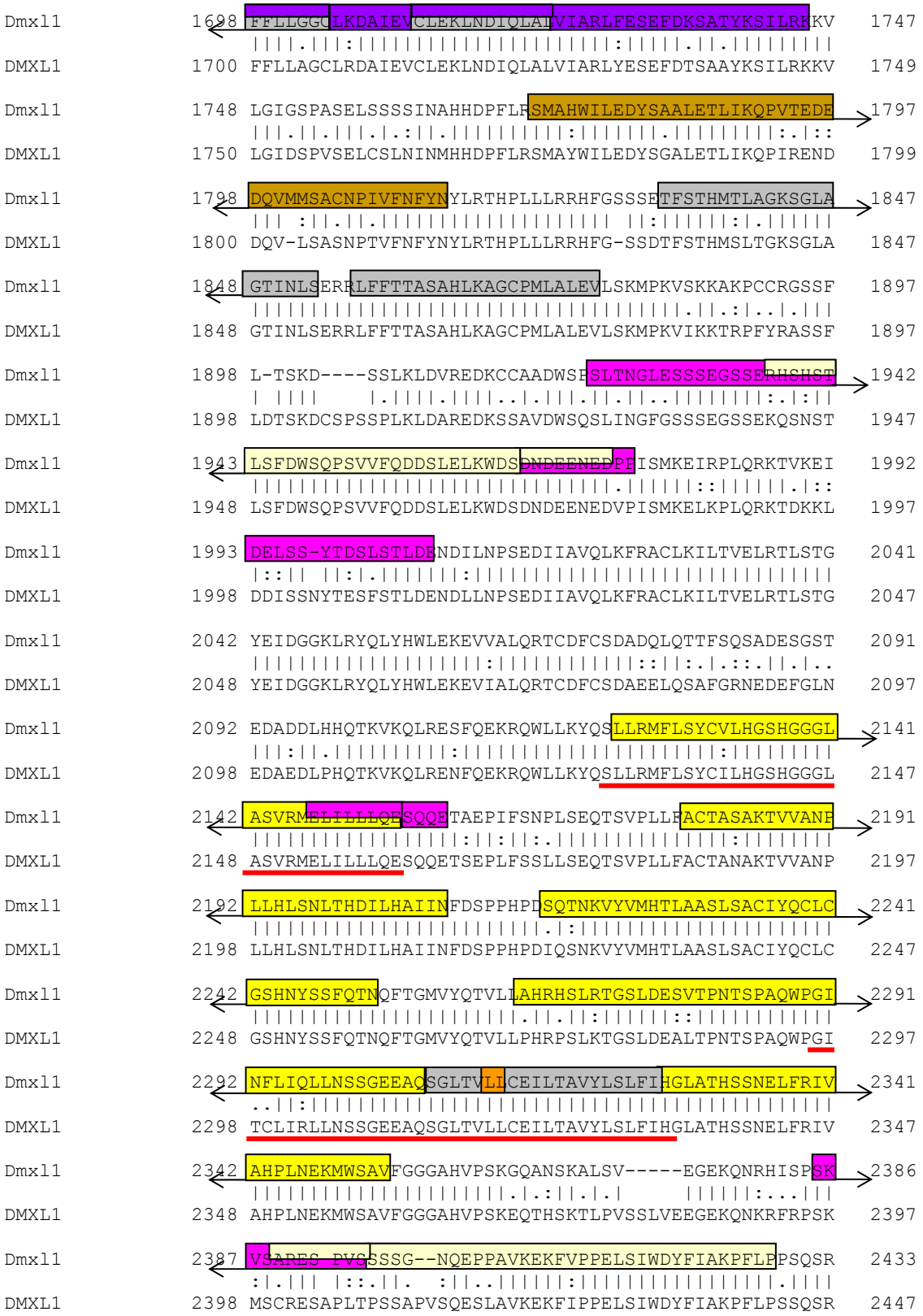
aufweist, bis 97,4 % in Exon 26. Vor allem Exons die (teilweise) für WD-Wiederholungseinheiten kodieren (Kapitel 3.5.11) zeigen sich in diesen Sequenzabschnitten überdurchschnittlich konserviert (Tab.3.10, Abb 3.32).

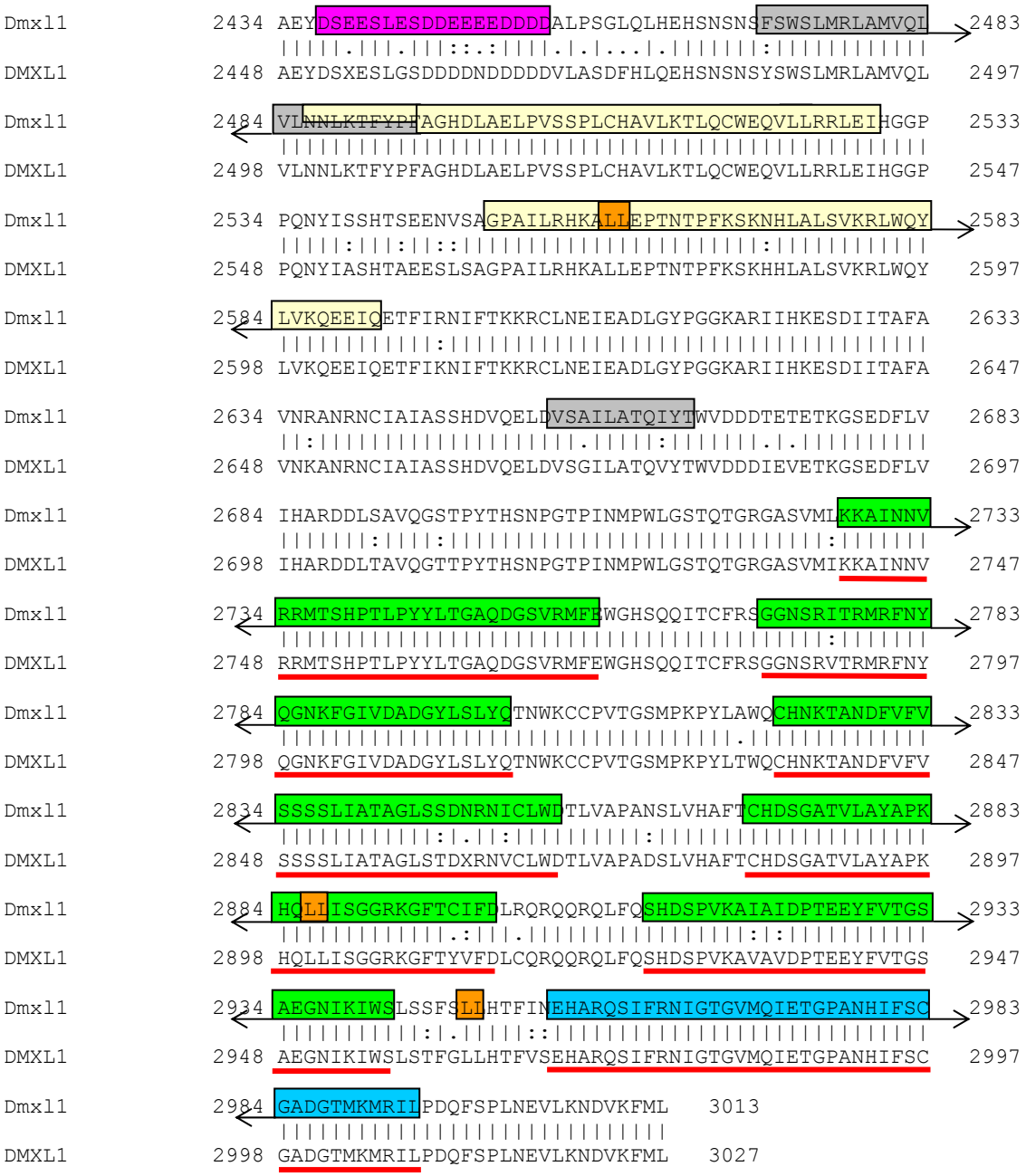
Abb. 3.33: Globale Übereinstimmung der Aminosäuresequenz zwischen Maus und Mensch (folgende Seiten).
Das dargestellte Alignment der abgeleiteten Aminosäuresequenz wurde mit dem Needleman-Wunsch Algorithmus (Needleman et al. 1970) des Programms ALIGN (EMBOSS) erstellt. Verwendet wurde hier die Matrix Blosum62, die Werte „Gap Open“ und „Gap Extend“ wurden mit jeweils 10 bzw. 0,5 Punkten negativ bewertet. Ein lokaler Vergleich mit dem Smith-Waterman Algorythmus zeigte exakt das gleiche Ergebnis (Daten nicht gezeigt). Die Identität des gezeigten Sequenzvergleichs beträgt 89,3 %, die Similarität der verglichenen Aminosäuren 94,7 %. Striche zeigen eine Übereinstimmung der untersuchten Aminosäuresequenzen an, Doppelpunkte konservierte Substitutionen und Punkte keine Übereinstimmung. Identifizierte Motive in der Sequenz wurden farbig unterlegt (Kapitel 3.5.11). Eine Legende der Farbcodierung findet sich am Ende der Abbildung:

```
Dmxl1          1 MNLHQVLTGAVNPGDHCFAVGSVGEQRFTAYASGCDIVILGSNFERLQII     50
                 |||||||||||||||||||:|||:|:||||||||||||||||||:|||||||
DMXL1          1 MNLHQVLTGAVNPGDHCFSVGSIGDQRFTAYASGCDIVILGSDFERLQII     50

Dmxl1         51 PGAKHGNIQVGCVDCSMQQGKIAASYGNVISVFEPVSIPKKRKNLEFYSQ    100
                 ||||||||||||||||||||||||||||||||:||||:|||::||||.|||
DMXL1         51 PGAKHGNIQVGCVDCSMQQGKIAASYGNVISIFEPVNLPKQKKNLELYSQ    100

Dmxl1        101 WQKSGQFFLDSIAHNITWDPAGNRLLTGSSCLQLWCNS--RKQTEDENPD    148
                 |||||||||:||||||||||.|:|||||||.||||.|:  .|.||||||.:
DMXL1        101 WQKSGQFFLESIAHNITWDPTGSRLLTGSSYLQLWSNTNLEKPTEDENLN    150

Dmxl1        149 KTDLNFGNWMCIWHCKTASQVHLMKFSPDGEFFATAGKDDCLLKVWYNVE    198
                 |||||||:|.||||||||||||||||||||||||||||||||||||||||
DMXL1        151 KTDLNFGDWKCIWHCKTASQVHLMKFSPDGEFFATAGKDDCLLKVWYNVE    200
```


Dmxl1 199 NWRPAVTSPDKNSEKQSQGEIDFSFVYLAHPRAVNGFSWRKTSKYMPRAS 248
DMXL1 201 NWRTAVTSPDGSSEKQSQGEIDFSFVYLAHPRAVNGFSWRKTSKYMPRAS 250
Dmxl1 249 VCNVLLTCCKDNVCRLWVETFLPNDCFLYGSDCNHWCEPVSLTNNLKRNA 298
DMXL1 251 VCNVLLTCCKDNVCRLWVETFLPNDCLLYGGDCSHWTESINLTNNFKRNA 300
Dmxl1 299 SSKDRVQSALEVNLRPFRRGRSRSLALVAHTGYLPHQQDPHHAHRNTPLH 348
DMXL1 301 SSKERVQNALEVNLRHFRRGRRRSLALVAHTGYLPHQQDPHHVHRNTPLH 350
Dmxl1 349 ANALCHFHIAASINPATDIPLLPSITSLSLNENEEKCGPFVVHWLNNKEL 398
DMXL1 351 ANALCHFHIAASINPATDIPLLPSITSLSLNENEEKTGPFVVHWLNNKEL 400
Dmxl1 399 HFTLSMEVFLQQLRKSFEQPSSEASVEDSIQADLKSDEELDDGVDDLKIN 448
DMXL1 401 HFTLSMEVFLQQLRKSFEQPSSEASVEDSNQADVKSDEETDDGVDDLKIN 450
Dmxl1 449 HEKKELDEDKMLPSSSFTPLSSAAVDHQIEVLLSEWSKNADMLFSIHPMD 498
DMXL1 451 PEKKELGCDKMVPNSSFTSLSSAAIDHQIEVLLSEWSKNADMLFSIHPMD 500
Dmxl1 499 GSLLVWHVDWLDEYQPGMFRQVQVSFVSRIPVAFPTGDANSLCKSIVMYA 548
DMXL1 501 GSLLVWHVDWLDEYQPGMFRQVQVSFVSRIPVAFPTGDANSLCKSIMMYA 550
Dmxl1 549 CTKNVDLAIQQGKQRPTGLTRSTSMLISSAHSKSSNNLKLSIFTPNVMMI 598
DMXL1 551 CTKNVDLAIQQGKQKPSGLTRSTSMLISSGHNKSSNSLKLSIFTPNVMMI 600
Dmxl1 599 SKHADGSLNQWLVSFAEESAFSTVLSISHKSRYCGHRFHLNDLACHSVLP 648
DMXL1 601 SKHADGSLNQWLVSFAEESAFSTVLSISHKSRYCGHRFHLNDLACHSVLP 650
Dmxl1 649 LLLTTSHHNALRTPNVGNQKQAHDAVNTEECSLAQQNKSNVDMAFQDPNA 698
DMXL1 651 LLLTTSHHNALRTPDVDNPEQPFDALNIEECSLTQQNKSTVDVAFQDPSA 700
Dmxl1 699 IYSELILWRVDPVGPLSFSGGVSELARINSLHVSAFSNVAWLPTLIPSYC 748
DMXL1 701 VYSELILWRVDPVGPLSFSGGVSELARINSLHVSAFSNVAWLPTLIPSYC 750
Dmxl1 749 LGAYCNSPSACFVASDGQYLRLYEAVIDAKKLLYELSNPEISKYVGEVFN 798
DMXL1 751 LGAYCNSPSACFVASDGQYLRLYEAVIDAKKLLSELSNPEISKYVGEVFN 800
Dmxl1 799 IVSQQSTARPGCIIALDSITKLHGRKTQLLHVFQEDFILNNLEKKRLGVD 848
DMXL1 801 IVSQQSTARPGCIIALDPITKLHGRKTQLLHVFEEDFILNNLEKKSLGKD 850
Dmxl1 849 NILLDSDSSCNGFSEKFYLVVIECTEDNRSLLRMWDLHLRSTPVSLDERI 898
DMXL1 851 SILSNAGSSPNGFSEKFYLIVIECTQDNRSLLHMWNLHLKSIPVSLDEKV 900
Dmxl1 899 DTKISEASWLPEEHYSSSPEKILSPFSQKFQACRANLQSTSKLSLFSEMV 948
DMXL1 901 DTKLSEAVWQPEEHYSSSPEKILSPFSQKYQACRANLQSTSRLTLFSEMV 950

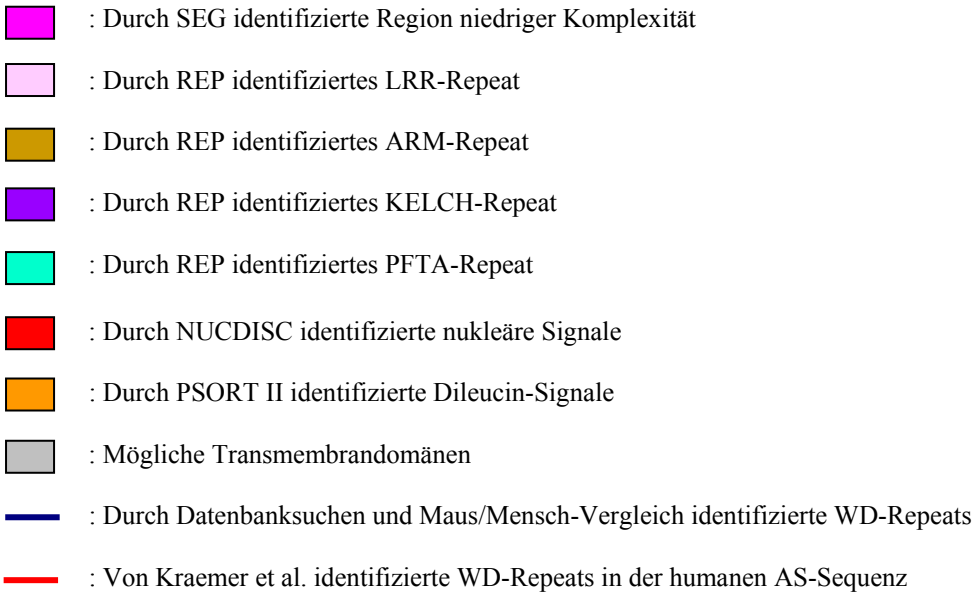

Dmxl1 949 YSKELDLPEGVEIISVKPSAGHLSSSSIYPVCSAPYLLATSCSDDKVRFW 998
DMXL1 951 YSQELHLPEGVEIISIKPSAGHLSSSSIYPACSAPYLLATSCSDEKVRFW 1000
Dmxl1 999 RCRVTNGESATSKNGKLDVVYVWEEWPLLIEDGLENNSSVTVPGRPVEVS 1048
DMXL1 1001 RCRVTDGESATSKNGKIDLAYIWEEWPLLIEDGLQSNSSITVPGRPVEVS 1050
Dmxl1 1049 CAHTSRLAVAYKQPTGNSR-SQEFVMHVSIFECESTGGSCWILEQTIHLD 1097
DMXL1 1051 CAHTNRLAVAYKQPASNSRSSQDFVMHVSIFECESTGGSCWVLEQTIHLD 1100
Dmxl1 1098 ELSTVLDSGISIDSNLVAYNKQETYLVSKESITSNTKHLVHLDWMSREDG 1147
DMXL1 1101 ELSTVLDSGISVDSNLVAYNKQDMYLSSKENITSNTKHLVHLDWMSREDG 1150
Dmxl1 1148 SHILTVGIGSKLFMYGPMAGKVQDQTGKENQAFPLWDSTKIVPLSKFVLL 1197
DMXL1 1151 SHILTVGIGSKLFMYGPLAGKVQDQTGKETLAFPLWESTKVVPLSKFVLL 1200
Dmxl1 1198 RSVDLVSSVEGAPPFPVSLSWVRDGILVVGMDCEMHVYSQWQPSNKQEPV 1247
DMXL1 1201 RSVDLVSSVDGSPPFPVSLSWVRDGILVVGMDCEMHVYCQWQPSSKQEPV 1250
Dmxl1 1248 ISESYNGSTPSILSLIKQSNSSSSGLHPPKKTLTRSMTSLAQKICGKKSI 1297
DMXL1 1251 ITDSYSGSTPSITSLIKQSN-SSSGLHPPKKTLTRSMTSLAQKICGKKTA 1299
Dmxl1 1298 FDPSVDMEDSGLFEAAHVLSPTLPQYHPLQLLELMDLGKVRRAKAILSHL 1347
DMXL1 1300 FDPSVDMEDSGLFEAAHVLSPTLPQYHPLQLLELMDLGKVRRAKAILSHL 1349
Dmxl1 1348 VKCIAGEVVALNEAESNHERRLRSLTISASGSTTRDPQAFNKADSRDYTE 1397
DMXL1 1350 VKCIAGEVVALNEAESNHERRLRSLTISASGSTTRDPQAFNKAENTDYTE 1399
Dmxl1 1398 IDSVPPLPLYALLAADDDSYCSSLEKTGSESSLKKSKQLSKESYDELFQT 1447
DMXL1 1400 IDSVPPLPLYALLAADDDSCYSSLEKSSNESTLSKSNQLSKESYDELFQT 1449
Dmxl1 1448 SVLMSDNHMLETDEENTQPRVIDLSQYSPTYFGPEHAQVLSGHLLHSSLP 1497
DMXL1 1450 QLLMTDTHMLETDEENTKPRVIDLSQYSPTYFGPEHAQVLSGHLLHSSLP 1499
Dmxl1 1498 GLTRMEQMSLMALADTIATTSTDIGESRDRNQGGETLDECGLKFLLAVRL 1547
DMXL1 1500 GLSRMEQMSLMALADTIATTSTDIGESRDRSQGGETLDECGLKFLLAVRL 1549
Dmxl1 1548 HTFLTTSLPAYRAQLLHQGLSTGHFAWAFHSVAEEELLNMLPAMQKDDPT 1597
DMXL1 1550 HTFLTTSLPAYRAQLLHQGLSTSHFAWAFHSVAEEELLNMLPAMQKDDPT 1599
Dmxl1 1598 WSELRAMGVGWWVRNARILRRCIEKVAKAAFHRNNDPLDAAIFYLAMKKK 1647
DMXL1 1600 WSELRAMGVGWWVRNTRILRKCIEKVAKAAFYRKNDPLDAAIFYLAMKKK 1649
Dmxl1 1648 AVIWGLYRSQKDTKMTQFFGHNFEEERWRKAALKNAFSLLGKQRFEHSAA 1697
DMXL1 1650 AVIWGLYRAEKNTRMTQFFGHNFEDERWRKAALKNAFSLLGKQRFEHSAA 1699


Dmxl1 1698 FFLLGGCLKDAIEVCLEKLNDIQLAIVIARLFESEFDKSATYKSILRKKV 1747
DMXL1 1700 FFLLAGCLRDAIEVCLEKLNDIQLALVIARLYESEFDTSAAYKSILRKKV 1749
Dmxl1 1748 LGIGSPASELSSSSINAHHDPFLRSMAHWILEDYSAALETLIKQPVTEDE 1797
DMXL1 1750 LGIDSPVSELCSLNINMHHDPFLRSMAYWILEDYSGALETLIKQPIREND 1799
Dmxl1 1798 DQVMMSACNPIVFNFYNYLRTHPLLLRRHFGSSSETFSTHMTLAGKSGLA 1847
DMXL1 1800 DQV-LSASNPTVFNFYNYLRTHPLLLRRHFG-SSDTFSTHMSLTGKSGLA 1847
Dmxl1 1848 GTINLSERRLFFTTASAHLKAGCPMLALEVLSKMPKVSKKAKPCCRGSSF 1897
DMXL1 1848 GTINLSERRLFFTTASAHLKAGCPMLALEVLSKMPKVIKKTRPFYRASSF 1897
Dmxl1 1898 L-TSKD----SSLKLDVREDKCCAADWSPSLTNGLESSSEGSSERHSHST 1942
DMXL1 1898 LDTSKDCSPSSPLKLDAREDKSSAVDWSQSLINGFGSSSEGSSEKQSNST 1947
Dmxl1 1943 LSFDWSQPSVVFQDDSLELKWDSDNDEENEDPPISMKEIRPLQRKTVKEI 1992
DMXL1 1948 LSFDWSQPSVVFQDDSLELKWDSDNDEENEDVPISMKELKPLQRKTDKKL 1997
Dmxl1 1993 DELSS-YTDSLSTLDENDILNPSEDIIAVQLKFRACLKILTVELRTLSTG 2041
DMXL1 1998 DDISSNYTESFSTLDENDLLNPSEDIIAVQLKFRACLKILTVELRTLSTG 2047
Dmxl1 2042 YEIDGGKLRYQLYHWLEKEVVALQRTCDFCSDADQLQTTFSQSADESGST 2091
DMXL1 2048 YEIDGGKLRYQLYHWLEKEVIALQRTCDFCSDAEELQSAFGRNEDEFGLN 2097
Dmxl1 2092 EDADDLHHQTKVKQLRESFQEKRQWLLKYQSLLRMFLSYCVLHGSHGGGL 2141
DMXL1 2098 EDAEDLPHQTKVKQLRENFQEKRQWLLKYQSLLRMFLSYCILHGSHGGGL 2147
Dmxl1 2142 ASVRMELILLLQESQQETAEPIFSNPLSEQTSVPLLFACTASAKTVVANP 2191
DMXL1 2148 ASVRMELILLLQESQQETSEPLFSSLLSEQTSVPLLFACTANAKTVVANP 2197
Dmxl1 2192 LLHLSNLTHDILHAIINFDSPPHPDSQTNKVYVMHTLAASLSACIYQCLC 2241
DMXL1 2198 LLHLSNLTHDILHAIINFDSPPHPDIQSNKVYVMHTLAASLSACIYQCLC 2247
Dmxl1 2242 GSHNYSSFQTNQFTGMVYQTVLLAHRHSLRTGSLDESVTPNTSPAQWPGI 2291
DMXL1 2248 GSHNYSSFQTNQFTGMVYQTVLLPHRPSLKTGSLDEALTPNTSPAQWPGI 2297
Dmxl1 2292 NFLIQLLNSSGEEAQSGLTVLLCEILTAVYLSLFIHGLATHSSNELFRIV 2341
DMXL1 2298 TCLIRLLNSSGEEAQSGLTVLLCEILTAVYLSLFIHGLATHSSNELFRIV 2347
Dmxl1 2342 AHPLNEKMWSAVFGGGAHVPSKGQANSKALSV-----EGEKQNRHISPSK 2386
DMXL1 2348 AHPLNEKMWSAVFGGGAHVPSKEQTHSKTLPVSSLVEEGEKQNKRFRPSK 2397
Dmxl1 2387 VSARES PVSSSSG--NQEPPAVKEKFVPPELSIWDYFIAKPFLPPSQSR 2433
DMXL1 2398 MSCRESAPLTPSSAPVSQESLAVKEKFIPPELSIWDYFIAKPFLPSSQSR 2447



Legende der Farbcodierung:

- : Durch SMART und REP identifizierte WD-Wiederholungseinheiten
- : Durch SMART, REP und PSA identifizierte WD-Wiederholungseinheiten
- : Durch PSA identifizierte WD-Wiederholungseinheiten
- : Durch REP identifizierte WD-Wiederholungseinheiten



Mit dem Programm ALIGN wurden ergänzende Analysen mit Spezies durchgeführt, in denen die abgeleitete Aminosäuresequenz der DmX-Homologen vollständig bekannt ist. Weiter wurde ein Vergleich mit der humanen Aminosäure- bzw. Nucleotidsequenz von *DMXL2* (Rabconnectin-3; Kawabe et al., 2003; Nagano et al., 2002; Kraemer et al., 2000) vorgenommen. Eine parallele Untersuchung der murinen Sequenz von *Dmxl2* war nur bedingt möglich, da die Nukletidsequenz der CDS in der Maus nicht bekannt ist. Durch Datenbanksuchen mit der menschlichen cDNA von *DMXL2* konnte jedoch der genomische Bereich des murinen *Dmxl2* auf Chromosom 9 Bande C im Bereich der Basen 54478246-54681959 lokalisiert werden. Der entsprechende Abschnitt wurde dann mit dem Programm SIM4 (Kapitel 3.5.6) analysiert und die Nukleotidsequenz der vorausgesagten 44 Exons extrahiert. Eine exakte Definition der Exon/Intron-Grenzen war jedoch aufgrund der unzureichenden Homologie zwischen der humanen cDNA von *DMXL2* und der murinen genomischen Sequenz nicht möglich, folglich konnte auch kein ORF und darauf basierende abgeleitete Aminosäuresequenz etabliert werden. Weiter muß davon ausgegangen werden, dass einige Nukleotide nicht identifiziert werden konnten und daher eine Abweichung der unten dargestellten Werte bis zu 1 % denkbar ist. Die humane CDS von *DMXL2* umfasst 9108 bp, die mit der SIM4-Analyse identifizierten homologen Abschnitte in der Maus ergeben eine 9012 bp lange Sequenz. Diese wurde für die in Tabelle 3.11 gezeigten Vergleiche verwendet. Ein Multispezies-Alignment der abgeleiteten Aminosäuresequenzen von DmX befindet sich auf dem beigefügten Datenträger.

	Maus/ Mensch	Identifizierte Strukturen in der AS-Sequenz		Maus/ Mensch	Identifizierte Strukturen in der AS-Sequenz
Exon1	92,0 %	WD40	**Exon23**	90,3 %	WD40/KELCH/ARM
Exon2	95,2 %	WD40/NUC	**Exon24**	85,3 %	WD40/LOW/TM
Exon3	79,2 %	WD40	**Exon25**	91,2 %	WD40
Exon4	94,9 %	WD40	**Exon26**	97,4 %	WD40
Exon5	80,7 %	WD40	**Exon27**	85,2 %	WD40
Exon6	92,5 %	WD40	**Exon28**	87,1 %	WD40/TM/DL
Exon7	93,3 %	WD40	**Exon29**	66,3 %	WD40/LOW
Exon8	89,5 %	NUC	**Exon30**	79,2 %	WD49/LOW/TM
Exon9	92,6 %	LOW	**Exon31**	85,9 %	WD40
Exon10	89,7 %	LRR	**Exon32**	89,4 %	WD40
Exon11	86,7 %	WD40	**Exon33**	95,7 %	WD40/DL
Exon12	89,9 %	WD40	**Exon34**	95,4 %	WD40
Exon13	95,1 %	WD40	**Exon35**	88,6 %	TM
Exon14	94,4 %	WD40	**Exon36**	90,8 %	WD40
Exon15	84,5 %	WD40	**Exon37**	93,4 %	WD40
Exon16	85,0 %		**Exon38**	96,9 %	WD40
Exon17	85,6 %	WD40	**Exon39**	94,8 %	WD40
Exon18	89,2 %	WD40/LOW/PFTA	**Exon40**	96,7 %	WD40
Exon19	91,2 %	WD40/KELCH	**Exon41**	86,8 %	WD40
Exon20	82,9 %	WD40/KELCH	**Exon42**	91,7 %	WD40/DL
Exon21	94,9 %	WD40	**Exon43**	92,4 %	WD40
Exon22	81,9 %	KELCH/TM			

Tab. 3.10: Identifizierte Motive in der Aminosäuresequenz von *Dmxl*1 und ihre relative Positionierung auf der CDS

Gezeigt sind die Exons des murinen *Dmxl*1 und der Grad ihrer Homologie (Identität) zu den entsprechenden Exons in der humanen Sequenz. Die identifizierten Motive (Kapitel 3.5.11) der abgeleiteten Aminosäuresequenz von *Dmxl*1 liegen zum größten Teil in hochkonservierten DNA-Abschnitten. Motive, die in weniger gut konservierten Exons liegen, zeigen meist in den entsprechenden Bereichen einen Homologiegrad, der deutlich über der durchschnittlichen Identität des entsprechenden Exons liegt (Abb3.33). WD40: WD-Wiederholungseinheit, LOW: Regionen mit Aminosäureabfolgen geringer Komplexität (Wootton et al., 1996), KELCH: KELCH-Repeat (Xue et al., 1993), LRR: LRR-Repeat („leucine-rich repeat“ Kobe et al., 1994), ARM: ARM-Repeat („armadillo“; Pfeifer et al., 1994), PFTA: PFTA-Repeat („ protein farnesyl transferase α-subunit repeats“; Boguski et al., 1992), TM: potenzielle Transmembrandomäne, DL: potenzielles Dileucin-Signal, NUC: durch NUCDISC identifiziertes nukleäres Signal.

	Maus *Dmxl1*	Mensch *DMXL1*	Drosophila *DmX*	Chironomus *CpY*	Caenorhabditis *F54E4.1*	Maus *Dmxl2*	Mensch *DMXL2*
Maus *Dmxl1*	-	94,7	55,2	55,9	52,0		72,3
Mensch *DMXL1*	89,3	-	54,9	54,8	52,5		71,8
Drosophila DmX	37,2	36,5	-	72,3	52,2		55,8
Chironomus CpY	37,8	37,6	58,1	-	52,1		56,1
Caenorhabdis F54E4.1	31,9	32,3	32,9	32,6	-		50,8
Maus *Dmxl2*	59,4	58,3	47,5	48,0	45,9	-	83,7
Mensch *DMXL2*	55,2	55,2	37,7	38,1	31,2	83,7	-

Tab. 3.11: Vergleich der homologen Proteine von DMX

Mit dem Programm ALIGN "water" (EMBOSS, Blossum62) wurde eine multiple Aminosäuresequenzanalyse der DMX-Proteine der Spezies *Homo sapiens* (DMXL1, 3027 aa), *Mus musculus* (Dmxl1, 3013aa), *Drosophila melanogaster* (DmX, 3429 aa), *Chironomus thummi* (CpY, 3512 aa) und *Caenorhabditis elegans* (F54E4.1, 2948 aa) durchgeführt. Zusätzlich wurde die Nucleotidsequenz von *DMXL2* (Rabconnectin-3) von Mensch und Maus in die Untersuchung mit eingebunden (ALIGN "water", EMBOSS, DNAfull). Blaue Zahlenwerte zeigen die Identität der Aminosäuresequenz, rote Zahlen geben die Similarität der untersuchten Aminosäuresequenz wieder. Schwarze Zahlen belegen die Identität des codierenden Bereichs der untersuchten Nukleotidsequenz. Interessant ist hierbei, dass die humane Aminosäuresequenz von DMXL2 zu der murinen Sequenz von Dmxl1 eine höhere Similarität zeigt, als zu der humanen Sequenz von DMXL1.

3.2.11 Analyse der abgeleiteten Aminosäuresequenz von Dmxl1

Da es sich bei den Produkten der homologen Genen zu *Dmxl*1 in *Drosophila melanogaster* und *Homo sapiens* um WD-Repeat Proteine handelt (Kraemer et al., 1998; 2000), wurde zunächst die abgeleitete Aminosäuresequenz auf WD-Wiederholungseinheiten hin untersucht. Verwendet wurden hierfür die Programme SMART (Schultz et al., 1998; Letunic et al., 2004, 2002) und PSA (Stultz et al., 1993, 1997; White et al., 1994). Mit Hilfe der SMART-Analyse wurden 14 WD-Wiederholungseinheiten und 12 Regionen niedriger Komplexität identifiziert (Unterprogramm SEG). Das Programm PSA identifizierte im C-terminalen WD-Repeat-Komplex zehn Wiederholungseinheiten, ein von SMART identifziertes WD-Repeat wurde nicht erkannt. Demnach befinden innerhalb der 1000 C-terminalen Aminosäuren 6 dicht aufeinander folgende WD-Repeats, gefolgt von 373 Aminosäuren ohne erkennbares Motiv sowie 5 weitere dicht aufeinanderfolgende WD-Wiederholungseinheiten. Im N-terminalen Bereich erkannte PSA keine der von SMART identifizierten WD-Repeats. SMART konnte über einen Abschnitt von 1236 Aminosäuren 9 WD-Wiederholungseinheiten erfassen. Insgesamt wurden mit Hilfe dieser Analyseprogramme 19 WD-Wiederholungseinheiten in der abgeleiteten Aminosäuresequenz gefunden. Von Kraemer et al. (2000) konnten in der humanen Aminosäuresequenz von DMXL1 per Auge und Datenbanksuchen 28 WD-Repeats

identifiziert werden, daher wurde der Maus/Mensch-Vergleich (Abb. 3.32) der abgeleiteten Aminosäuresequenzen ebenfalls per Auge und BLASTP- bzw. Pfam-Datenbanksuchen auf Tryptophan-Asparaginsäure-Wiederholungseinheiten hin untersucht. Hierbei konnten die meisten von Kraemer et al. identifizierten Wiederholungseinheiten auch in der murinen Sequenz aufgefunden werden. Unterschiede zeigen sich in den Regionen 706-772 und 2131-2353 aa. Im Bereich 706-772 wurden von Kraemer et al. 2 WD-Repeats identifiziert, die Analysen der murinen Aminosäuresequenz deuten jedoch eher auf nur eine Wiederholungseinheit hin. Für den Abschnitt 2131-2353 wurden von Kraemer et al. zwei WD-Wiederholungseinheiten vorrausgesagt, eine Untersuchung mit dem Programm PSA und Datenbanksuchen identifizieren in dieser Region sowohl für die murine als auch für die humane Aminosäuresequenz fünf WD-Repeats. Da es sich bei den WD-Wiederholungseinheiten um Protein-Repeats handelt, wurde die Aminosäuresequenz mit dem Programme REP (Andrade et al., 2000) untersucht. Das Programm REP erzeugt ein lokales Alignment mit bekannten repetitiven Elementen, die in einer Datenbank hinterlegt wurden. Auf diese Weise konnten 25 WD-Repeats in der abgeleiteten Aminosäuresequenz identifiziert werden, darunter 6, die weder durch eines der anderen Vorhersageprogramme noch mittels Maus/Mensch-Vergleich identifiziert werden konnten. Das Programm liegt mit dieser Vorhersagegenauigkeit damit deutlich vor SMART und PSA. Neben den WD-Wiederholungseinheiten konnten weitere repetitive Einheiten anderer Superfamilien identifiziert werden: KELCH (Xue et al., 1993) und LRR („leucine-rich repeat“; Kobe et al., 1994). Weiter wurden ARM- („armadillo“; Pfeifer et al., 1994) und PFTA-Repeats („protein farnesyl transferase α-subunit repeats“; Boguski et al., 1992). Identität, Sequenz und Position der Repeats sind in Tabelle 3.12, Abb. 3.32 und 4.11 aufgeführt. Insgesamt wurden mit Hilfe der Analyseprogramme, Datenbanksuchen und dem Maus/Mensch-Vergleich 36 Tryptophan-Asparaginsäure-Wiederholungseinheiten in der abgeleiteten Aminosäure-sequenz von Dmxl1 identifiziert (Abb 3.32), von denen 24 eine gute Homologie zu WD-Wiederholungseinheiten anderer Proteine aufweisen und entweder durch SMART, PSA, Datenbanksuchen oder REP identifiziert werden konnten. Die aufgefundenen Unterschiede zwischen Maus und Mensch resultieren vermutlich aus der Weiterentwicklung der Analyseprogramme und der gestiegenen Anzahl WD-Repeat Protein spezifischer Datenbankeinträge.

Für die Identifikation weiterer Motive in der abgeleiteten Aminosäuresequenz fanden diverse Programme Verwendung:

PSORT II: Mit dem Unterprogramm NUCDISK konnte zwei Signale identifiziert werden, die eine nukleäre Lokalisation wahrscheinlich machen, in Position 89 die Sequenz PKKRKNL und in Position 314 das Motiv PFRRGRS. Weiter konnten fünf Dileucin-Motive (LL) im C-Terminus in den Positionen 2312, 2523, 2559, 2886 und 2948 identifiziert werden. Allerdings liegt eines der Signale in einer vorgeschlagenen Transmembrandomäne, ein weiteres ist Teil eines WD-Repeats (Abb. 3.32). Dileucin-Signale können ein Hinweis auf eine selektive Inklusion in endocytotische Vesikel sein

Repeat	Position in der Dmxl1-AS-Sequenz von	bis	Sequenz
WD40	30	71	AYASGCDIVILGSNFERLQIIPGAKHGNIQVGCVDCSMQQGK
WD40	108	145	FLDSIAHNITWDPAGNRLLTGS..SCLQLWCNSRKQTEDE
WD40	164	204	KTASQVHLMKFSPDGEFFATAGKDDCLLKVWYNVENWRPAV
WD40	227	275	AHPRAVNGFSWRKTSKYMPRASVCNVLLTCCKDNVCRLWVETFLPNDCF
LRR	443	467	DDL.KINHEKKELDEDKMLPSSSFTP
WD40	474	514	DHQIEVLLSEWSKNADMLFSIHPMDGSLLVWHVDWLDEYQP
WD40	578	619	AHSKSSNNLKLSIFTPNVMMISKHADGSLNQWLVSFAEESAF
WD40	729	781	LHVSAFSNVAWLPTLIPSYCLGAYCNSPSACFVASDGQYLRLYEAVIDAKKLL
WD40	842	893	KKRLGVDNILLDSDSSCNGFSEKFYLVVIECTEDNRSLLRMWDLHLRSTPVS
WD40	897	934	RIDTKISEASWLPEEH..YSSSPEKILSPFSQKFQACRAN
WD40	970	1008	HLSSSSIYPVCSAPYLLATSCSDDKVRFWRCRVTNGESA
WD40	1010	1035	SKNGKLDVVYVWEEWPLLIEDGLENN
PFTA	1069	1102	QEFVMHVSIFECESTGGSCWILEQTIHLDELSTV
WD40	1145	1193	EDGSHILTVGIGSKLFMYGPMAGKVQDQTGKENQAFPLWDSTKIVPLSK
WD40	1208	1248	GAPPFPVSLSWVRDGILVVGMDCEMHVYSQWQPSNKQEPVI
ARM	1393	1436	KADSRDYTEIDSVPPLPLYALLAADDDSYCSSLEKTGSESSLKK
KELCH	1572	1618	FAWAFHSVAEEELLNMLPAMQKDDPTWSELRAMGVGWWVRNARILRR
WD40	1618	1661	RCIEKVAKAAFHRNNDPLDAAIFYLAMKKKAVIWGLYRSQKDTK
KELCH	1697	1745	AFFLLGGCLKDAIEVCLEKLNDIQLALVIARLFESEFDKSATYKSILRK
ARM	1776	1818	SMAHWILEDYSAALETLIKQPVTEDEDQVMMSACNPIVFNFYN
WD40	1937	1973	RHSHSTLSFDWSQPS...VVFQDDSLELKWDSDNDEENED
WD40	2389	2428	ARESPVSSSSGNQEPPAVKEKFVPPELSIWDYFIAKPFLP
WD40	2486	2529	NNLKTFYPFAGHDLAELPVSSPLCHAVLKTLQCWEQVLLRRLEI
WD40	2550	2591	GPAILRHKALLEPTNTPFKSKNHLALSVKRLWQYLVKQEEIQ
WD40	2728	2769	KAINNVRRMTSHPTLPYYLTGAQDGSVRMFEWGHSQQITCFR
WD40	2771	2810	GGNSRITRMRFNYQGNKFGIVDADGYLSLYQTNWKCCPVT
WD40	2822	2864	CHNKTANDFVFVSSSSLIATAGLSSDNRNICLWDTLVAPANSL
WD40	2870	2909	CHDSGATVLAYAPKHQLLISGGRKGFTCIFDLRQRQQRQL
WD40	2912	2951	SHDSPVKAIAIDPTEEYFVTGSAEGNIKIWSLSSFSLLHT
WD40	2964	3002	NIGTGVMQIETGPANH.IFSCGADGTMKMRILPDQFSPLN

Tab. 3.12: Durch das Programm REP identifizierte Protein-Repeatsequenzen

Aufgelistet wurden nur Repeats, die mit einer „hohen“ Wahrscheinlichkeit identifiziert wurden. Hierbei erfolgte die Orientierung an „Score“-Werten von WD-Wiederholungseinheiten, die auch von anderen Programmen wie SMART oder PSA identifiziert wurden und keine oder geringe Überlappung zu anderen Wiederholungseinheiten zeigten.

(Haft et al., 1998, 1994; Letourneur et al., 1992). Ein peroxisomales Signal (SKL2, PTS2; Petriv et al., 2004; Rachubinski et al.1995; Marzioch et al., 1994; Kragler et al., 1993)) mit der Sequenz KLHGRKTQL wurde für die Position 819 vorausgesagt, liegt damit aber innerhalb eines durch Datenbanksuchen identifizierten WD-Repeat mit einer schwächeren Homologie zur WD-Konsensussequenz. Da keine weiteren cytosolischen Signale gefunden wurden, berechnete PSORT II eine nukleäre Lokalisation des Dmxl1-Proteins mit einer Wahrscheinlichkeit von 69, 6 % (Human: 65,2 %). Alle Signale wurden auch für das humane DMXL1 vorausgesagt.

ESLpred: Dieses Programm wurde speziell für die zellulare Lokalisation von Proteinen entwickelt (Bhasin et al., 2004). Für die abgeleiteten Aminosäuresequenzen von Dmxl1 und DMXL1 wird jeweils eine nukleäre Lokalisation mit einer Genauigkeit von 98%

bestimmt. Für DmX wird eine nukleäre Lokalisation mit 53 % Genauigkeit angegeben, CpY hingegen soll mit einer 94%igen Genauigkeit im Cytoplasma gefunden werden. Für das homologe Protein F54E4.1 wird eine nukleäre Lokalisation mit 98% Genauigkeit bestimmt.

Identifikation von Transmembranhelices: Das Programm **TMpred** identifizierte 20 potenzielle Transmembranhelices mit einer Innen- nach Außen-Orientierung und 18 Helices mit einer Außen- nach Innen-Orientierung. Von diesen wurden sieben als signifikant eingestuft („Score"-Werte über 500), drei davon mit „Score"-Werten über 1000 müssen als hochsignifikant betrachtet werden (Abb. 3.34-A). Bei dem berechneten Modell befindet sich der N-Terminus im Innenbereich. **TM-Finder** identifizierte 9 transmembranspannende Segmente, von denen 5 mit einer Ladung versehen sind (Abb. 3.34-B). Zwei der vorausgesagten Helices wurden auch von TMpred erkannt, eine von TOPPRED. **TOPPRED** konnte zwölf Transmembranregionen lokalisieren, von denen sechs mit einer sehr hohen Wahrscheinlichkeit bewertet wurden (Abb. 3.34-C). **PHDhtm** sagt eine Transmembranhelix in der Position 2312-2329 voraus, nach diesem Modell befindet sich der N-terminale Bereich außerhalb, der C-Terminus innerhalb. Diese Transmembrandomäne wurde von allen verwendeten Programmen identifiziert und befindet sich innerhalb eines Blocks aus fünf WD-Wiederholungseinheiten. Die Programme **TMHMM Server 2.0, tmap** (EMBOSS), **orienTM** und **SOSUI** konnten keine Transmembrandomänen identifizieren. Ein Vergleich der Positionen der Transmembrandomänen und der identifizierten WD-Repeats zeigte jedoch, dass die meisten projizierten Transmembranhelices mit den WD-Wiederholungseinheiten überlappen oder innerhalb eines Blocks der WD-Repeats liegen (Abb. 3.32, Abb. 3.33).

SRS-Tool (LION bioscience AG): Die abgeleitete Aminosäuresequenz von Dmxl1 wurde mit dem Unterprogramm PATMATMOTIVES untersucht, hierbei wurden drei Amidierungspositionen identifiziert: 821-824, 1292-1295 und 2890-2893. Da Amidierungen C-terminal erfolgen (Kreil, 1984; Bradbury et al., 1987), würde nur die letzte Position für eine solche Modifikation in Frage kommen. Diese liegt jedoch innerhalb einer als sicher geltenden WD-Repeat-Einheit. Weiter konnte ein SUBTILASE_ASP-Motiv in der Position 1100-1110 identifiziert werden. Es handelt sich hierbei um eine 100 %ige Übereinstimmung zum aktiven Zentrum der Konsensussequenz einiger Serin-Proteasen. Dieses Motiv wurde auch von dem Programm PPSearch (EMBL-EBI) erkannt.

MotifScan (Pagni et al., 2004; Falquet et al., 2002): Das Programm identifizierte bei Datenbanksuchen in „Prosite patterns", „Prosite profiles" und „Pfam collection" neben einigen von SMART und PSA identifizierten WD-Wiederholungseinheiten in der Position 328-356 eine Histidin-reiche Region und von 1100-1110 das auch von PATMATMOTIFS identifizierte SUBTILASE_ASP-Motiv. Beide Motive wurden ebenfalls von dem Programm ScanProsite (Gattiker et al., 2002) erkannt.

Superfamily (Wilson et al., 2007; Madera et al., 2004;): Die abgeleitete Aminosäuresequenz von Dmxl1 wurde von diesem Programm eindeutig als „WD-40 repeat Protein“ eingestuft.

A:

```
-----> STRONGLY prefered model: N-terminus inside
 7 strong transmembrane helices, total score : 6107
 # from   to length score orientation
 1  729  749 (21)    1153 i-o -----→ WD
 2  971  987 (17)     667 o-i -----→ WD
 3 2119 2142 (24)     510 i-o -----→ WD
 4 2222 2243 (22)    1108 o-i -----→ WD
 5 2311 2331 (21)    1305 i-o
 6 2470 2494 (25)     553 o-i
 7 2827 2845 (19)     811 i-o -----→ WD
```

B:

```
Helix   Beginn  -       End     charge content
1       1192    -       1201    high* -----→ WD
2       1219    -       1228    high* -----→ WD
3       1503    -       1512    high* -----→ WD
4       1541    -       1550    high* -----→ WD
5       1695    -       1704    --
6       1718    -       1729    high*
7       2308    -       2327    --
8       2474    -       2487    --
9       2654    -       2664    --

* high charge content represents there being ≥ 15% charged amino acids
in the TM segment
```

C:

```
Helix   Beginn  -       End     Score   Certainity

1       733     -       753     1.223   Certain  -----→ WD
2       971     -       991     1.173   Certain  -----→ WD
3       1147    -       1167    1.017   Certain  -----→ WD
4       1561    -       1581    0.669   Putative -----→ WD
5       1833    -       1853    0.621   Putative
6       1857    -       1877    0.653   Putative
7       2124    -       2144    0.744   Putative -----→ WD
8       2172    -       2192    0.952   Putative -----→ WD
9       2222    -       2242    1.164   Certain  -----→ WD
10      2307    -       2327    1.396   Certain
11      2827    -       2847    1.210   Certain  -----→ WD
12      2936    -       2956    0.684   Putative -----→ WD
```

Abb. 3.34: Vorhersage möglicher Transmembranhelices

A: Von TMpred vorhergesagte Transmembrandomänen. Fünf der sieben Helices liegen innerhalb identifizierter WD-Wiederholungseinheiten, Helix Nr. 5 erreicht die höchste Wertung und wurde von allen Programmen als Transmembrandomäne erkannt.

B: Durch das Programm TM-Finder vorhergesagte Transmembranregionen. Vier von neun identifizierten Bereichen sind innerhalb von WD-Repeats lokalisiert, eine qualitative Wertung der Vorhersage ist mit diesem Programm nicht möglich. Helix Nr. 8 wurde auch von dem Programm TMpred erkannt.
C: TOPPRED-Vorhersage von Transmembranhelices. Neun von zwölf möglichen Domänen befinden sich im Bereich der WD-Repeats. Helix Nr. 10, die auch von anderen Programmen identifiziert wurde, ist hierbei mit der höchsten Wahrscheinlichkeit bewertet worden.

4 Diskussion

In der vorliegenden Arbeit wurde mit Hilfe der „Knock-Out"-Technik das murine *Dmxl*1-Gen auf genomischer Ebene disruptiert. Weiter wurden die vollständige cDNA und die genomische Organisation von *Dmxl*1 aufgeklärt und mit diesen Informationen umfangreiche computergestützte Analysen durchgeführt. In diesem Abschnitt sollen zunächst die in dieser Arbeit verwendeten Techniken und Resultate zur Sequenzierung, Konstruktion des Targeting-Vektors und „Knock-Out" von *Dmxl*1 diskutiert werden. Anschließend erfolgt eine Bewertung der *In Silico*-Analysen und eine Interpretation der ermittelten Daten in Hinblick auf den resultierenden Phänotyp der chimären Tiere des „Knock-Out"-Experiments.

4.1 Bewertung der Strategie zur Identifikation von *Dmxl*1 in *Mus musculus* und Konstruktion des Targeting-Vektors

Eine grundlegende Voraussetzung für die erfolgreiche Durchführung eines „Knock-Out"-Experiments in der Maus ist die detaillierte Kenntnis der genomischen Struktur des zu disruptierenden Gens (Galli-Taliadoros et al., 1995). Zu Beginn dieser Arbeit lagen keine Sequenzdaten der genomischen Region von *Dmxl*1 bzw. *DMXL*1 in Maus und Mensch vor, da das Humane Genomprojekt erst 1998 mit dem dritten Fünfjahresplan in die entscheidende Phase der Sequenzierung im Hochdurchsatz-Verfahren (Collins et al., 1998) trat. Die Sequenzierung des Mausgenoms wurde 1998 beschlossen und erst 1999 begonnen (Battey et al., 1999). Außer einem cDNA-Klon (IMAGE: 905860), der Sequenzen aus dem 3'-UTR enthält, lagen somit keine weiteren Sequenzinformationen für das murine *Dmxl*1 vor. Aus dieser Tatsache definierte sich das erste Ziel dieser Arbeit: Zunächst sollten durch Hybridisierung mit Sequenzabschnitten der menschlichen cDNA Klone aus einer murinen genomischen DNA-Bibliothek herausgesucht werden. Für die Wahl einer geeigneten Bibliothek ist dabei Folgendes zu bedenken: Verschiedene Experimente zeigten, dass die Effizienz der homologen Rekombination zwischen zwei Sequenzen ganz entscheidend von der Länge der verwendeten homologen Sequenzen abhängig ist (Thomas et al., 1987). Bei einer Variation der Länge der für die homologe Rekombination verwendeten Sequenzen von 1,2 zu 9,5 kb konnte eine 25-40 Fach höhere Effizienz für die längeren Sequenzen beobachtet werden (Deng, 1992; Shulman et al., 1990; Thomas et al., 1987). Für ein erfolgreiches Gen-Targeting ist außerdem der Grad der Homologie zwischen zwei Sequenzabschnitten von grundlegender Bedeutung (Smith, 2001). Isogene Sequenzen (genetisch identische Homologieblöcke) aus syngenen (genetisch identischen) Tieren oder Zellinien zeigen eine 20-25 Fach höhere Wirkkraft in der homologen Rekombination als nicht-isogene Sequenzabschnitte (geringe Sequenzunterschiede innerhalb der Homologieblöcke) (te Riele et al., 1992; Van Deursen et al, 1992; Hasty et al., 1991; te Riele et al., 1990). Auch konnte eine Verschiebung der zufälligen zu gerichteten Genomintegrationen um den Faktor vier bei

Verwendung von isogenen Tieren beobachtet werden (Riele et al., 1989). Verglichen wurden dabei Targeting-Sequenzen der Mausstämme BALB/c und 129. Das Gen-Targeting dieser Arbeit sollte in embryonalen Stammzellen des Typs 129-R1 (Nagy et al., 1993) erfolgen. Eine syngene DNA-Bibliothek stand jedoch zu diesem Zeitpunkt ebenfalls nicht zur Verfügung. Um die Anzahl möglicher Polymorphismen gering zu halten, fand daher zunächst die Cosmid-Bank Nr. 121 (129/Ola) Verwendung, hergestellt von Burgtorf, Poch und Wiles aus einem gemischten genetischen Hintergrund aus 129/Ola Maus-DNA ligiert in Cosmid Lawrist 7 (Kapitel 3.1.1). Aus diesem Grund wurden auch positiv detektierte Klone einer vorliegenden P1-DNA-Bank aus C57/BL/6 (Kapitel 3.1) nicht weiter bearbeitet, da der zu erwartende Grad möglicher polymorpher Sequenzabschnitte schon von vornherein eine Reduktion der Rekombinationseffizienz bedeutet hätte (Munclinger et al., 2003; Wade et al., 2002; Thomas et al., 2002; Lander et al., 2001; Green-Till et al., 2000; Zhao et al., 1995).
Von grundlegender Bedeutung für ein erfolgreiches Gen-Targeting ist weiter die Auswahl geeigneter Zielsequenzen für die später stattfindende homologe Rekombination. Um zu verhindern, dass nach der Disruption ein trunkiertes aber teilfunktionales Protein entsteht, sollten die ausgewählten Sequenzabschnitte für den „Knock-Out“ möglichst weit im 5’-Bereich des Zielgens lokalisiert sein (Transgenic Mouse Core Facility: Instruction Manual for Users, Purdue University, 2004; Deng, 1992; Mansour et al., 1988). Basierend auf diesen Vorgaben erfolgte die Suche nach positiven Klonen auf den hoch-dichten Klonfiltern der Cosmid-Bank durch Inter-spezies-Hybridisierung (Karasawa et al., 1993; Smith et al., 1976) mit zwei Sonden aus dem 5’-Bereich der humanen cDNA von *DMXL*1: Das Integrat des cDNA Klons IMAGp998H19442 (Sonde H19442) und ein PCR-Amplifikat (Sonde Homo34/35), das sich an H19442 anschließt (Kapitel 3.1.1). Das Integrat des Klons IMAGp998H19442 überspannt dabei die Exons 5-10, das PCR-Produkt die Exons 11-17 (zusammen bp 586-3085) der menschlichen cDNA von *DMXL*1 und repräsentierten damit die am weitesten 5’-wärts gelegenen Sequenzen, die zu diesem Zeitpunkt im Menschen bekannt waren. Das mit diesen Sequenzen detektierte Cosmid N10467Q3 enthielt die Exons 8-17 (bp 831-2998). Weitere Klone, welche die Exons fünf bis sieben enthalten, konnten in der vorhandenen Genbank nicht identifiziert werden. Die Gründe hierfür sind unbekannt, jedoch nicht methodischer Art. Es kann angenommen werden, dass diese Sequenzabschnitte bei der Herstellung der Genbank unterrepräsentiert waren und/oder nicht kloniert wurden.
Die Sequenzierungsstrategie von Cosmid N10467Q3 muss aus heutiger Sicht kritisch beurteilt werden. Die in dieser Arbeit verwendete Kombination aus „primer-walking“, Subklonierung und Religationsklonierung (Kapitel 3.2.1) erwies sich als sehr zeit- und arbeitsaufwendig. Der Vorteil einer niedrigeren Redundanz, und damit geringeren Kosten für die erforderlichen Sequenzierungsreaktionen, gegenüber einer „Shotgun“-Sequenzierung (Chissoe et al., 1997; Andersson et al., 1996; Ansari-Lari et al., 1996; Chen et al., 1996; Venter et al., 1996; Adams et al., 1995; Fleischmann et al., 1995; Bodenteich et al., 1994), bei der jeder DNA-Abschnitt sechs- bis elfmal sequenziert wird, wurde durch die notwendigen Synthesen der Primer und den Zeitaufwand relativiert. Mehr als ein Viertel der Cosmidsequenz besteht aus repetitiven Elementen (Tab. 3.5), welche das Primer-Design erheblich erschwerten bzw. unmöglich machten

und unter anderem zu fehlgeschlagenen Sequenzierungsreaktionen führten. Eine „Shotgun"-Sequenzierung hätte vermutlich die gleichen Kosten verursacht, aber den Zeit- und Arbeitsaufwand halbiert. Ein Vorteil dieser Strategie lag allerdings darin, dass sequenzierte DNA-Abschnitte zum Teil in klonierter Form vorlagen und somit für die Erzeugung des Targeting-Vektors zur Verfügung standen.

Für die Konstruktion des Targeting-Vektors standen somit nur die durch die Sequenzierung von Cosmid N10467Q3 ermittelten Daten zur Verfügung. Die Herstellung des Targeting-Vektors orientierte sich an publizierten Vorgaben und der persönlichen Beratung durch Prof. Dr. Manfred Blessing. Die in vielen Arbeiten verwendete Integration der „Neo-Box" des Targeting-Konstrukts in ein Exon ist wünschenswert, aber für eine Disruption des Zielgens nicht zwingend erforderlich (u.a. Zhang et al., 2001; Postic et al., 1999; Zhu et al., 1999; Ho et al., 1998; Markkula et al., 1996; Umans et al., 1995; Blessing, persönliche Auskunft). In dieser Arbeit wurde mit einem „replacement"-Vektor gearbeitet, also einem Konstrukt, das durch homologe Rekombination gegen die entsprechenden genomischen Sequenzabschnitte ausgetauscht wird (Smith, 2002). Für das Vektor-Design bestand daher auch die Möglichkeit, Exons zu deletieren. Neben der Integration des Neomycin-Resistenzgens *neo* (Stewart et al., 1987; Gossler et al., 1986) und einem damit assoziierten Polyadenylierungssignal in allen Leserahmen, kann zusätzlich auch eine Verschiebung des ORFs herbeigeführt werden. Bei einem möglichen aberranten Spleißvorgang führt dies zu einer veränderten Aminosäureabfolge und/oder einem Stop-Codon. Weitere Voraussetzungen für die Konstruktion des Targeting-Vektors waren: Kompatible genomische Schnittstellen zu den beiden multiplen Klonierungsschnittstellen des Vektorsystems, zwei möglichst lange Sequenzabschnitte, die sich als Targeting-Arme eignen und keine repetitiven Elemente enthalten, passende Schnittstellen außerhalb des auszutauschenden Sequenzabschnitts und innerhalb des später genomisch integrierten Resistenzmarkers zur Detektion der endogenen Bande mittels Southernblot sowie ein geeigneter Sequenzabschnitt innerhalb der externen Schnittstelle, aber außerhalb des rekombinanten Bereichs zur Herstellung einer Sonde (Abb. 3.10) (Transgenic Mouse Core Facility: Instruction Manual for Users, Purdue University, 2004; Blessing et al., 1997; Rossant et al., 1995; Galli-Taliadoros et al., 1994; Morrow et al., 1993; Deng, 1992; Koller et al., 1992; Shulman et al., 1990; Mansour et al., 1988; Thomas et al., 1987). Der verfügbare genomische Sequenzabschnitt aus Cosmid N10467Q3 erfüllte diese Vorgaben nur teilweise, daher waren fünfzehn molekulare Manipulationen und die Überprüfung von ca. 60.000 Klonen für die Konstruktion des Vektors notwendig. In der vorliegenden Arbeit konnte ein „replacement"-Vektor mit einer Gesamtgröße von 20293 bp hergestellt werden, der zwei homologe Sequenzabschnitte (Targeting-Arme) von 4347 bp (5') und 6786 bp (3') besitzt. Eine Integration des Resistenzmarkers $pPGKNeo^{R}A^{+}$ in eines der Exons war aufgrund inkompatibler Schnittstellen nicht möglich. Eine erfolgreiche homologe Rekombinationzwischen dem Targeting-Vektor und dem endogenen *Dmxl*1 führt zu einer Deletion der Exons 10-14 (genomisch 6594 bp). Der zu deletierende genomische Abschnitt wurde so gewählt, dass es bei einem aberranten Spleißvorgang von Exon 9 auf Exon 15 zu einer Verschiebung des offenen Leserahmens und zur Erzeugung eines Stop-Codons kommen würde, da auf Exon 9 ein

Phase 1-Intron folgt und vor Exon 15 ein Phase 0-Intron lokalisiert ist (Abb. 4.1). Diese Konstruktion diente als Sicherungsmaßnahme, die Integration eines starken Polyadenylierungssignals in den transkribierten genomischen Abschnitt ist normalerweise ausreichend, um eine Disruption zu bewirken, da das Polyadenylierungssignal schon wärend der Transkription terminierend auf die RNA-Polymerase II wirkt (Proudfoot, 2000; Bentley, 1999; Minvielle-Sebastia et al., 1999; Colgan et al., 1997). Die Länge der Targeting-Arme wurde so gewählt, dass sie den von Thomas et al. (1987), Shulman et al. (1990) und Deng et al. (1992) experimentell ermittelten optimalen Längenmaß entsprechen.

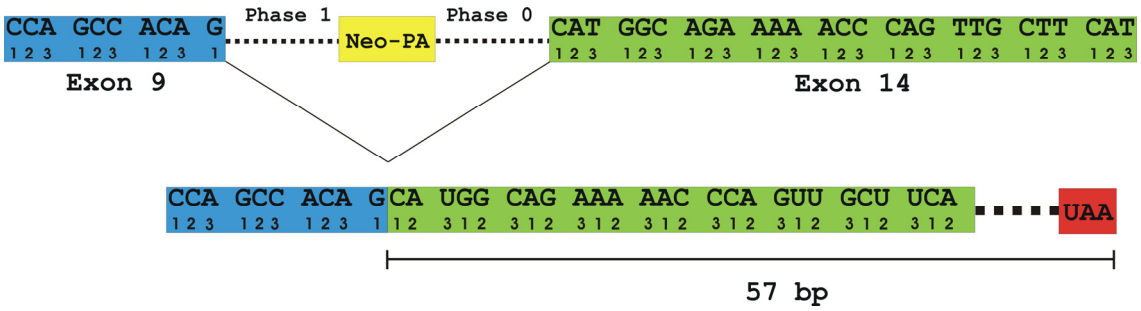


Abb. 4.1: Aberrantes Spleißen führt zu einer Verschiebung und Unterbrechung des offenen Leserahmens.

In der oberen Bildhälfte ist ein Ausschnitt der genomischen Organisation nach erfolgtem Rekombinationsereignis gezeigt. Auf Exon 9 (blau) folgt ein Phase 1-Intron, die gelbe Box symbolisiert den ausgetauschten genomischen DNA-Abschnitt mit integriertem Resistenzmarker und dem Polyadenylierungssignal. Vor Exon 14 (grün) liegt ein Phase 0-Intron, die Codierung des offenen Leserahmens wird durch die Zahlenabfolge 123 dargestellt, diese sind der entsprechenden Basenabfolge zugeordnet. In der unteren Bildhälfte ist die prozessierte RNA nach einem aberranten Spleißvorgang dargestellt. Die ursprünglich codierte Aminosäureabfolge ist ab der Spleißstelle um eine Position verschoben, daraus ergibt sich eine veränderte Aminosäuresequenz und nach 54 bp ein Stop-Codon (rote Box).

4.2 Identifikation von rekombinanten embryonalen Stammzellen – ein spezifisch entwickeltes Restriktions- / Detektionssystem

Für das zu entwickelnde Restriktions- / Detektionssystem konnten zunächst anhand der Sequenzinformationen aus Cosmid N10467Q3 kaum geeignete Schnittstellen in direkter Umgebung der späteren Rekombinationsposition identifiziert werden. Eine geeignete Sondensequenz, die zu einem potenziellen Restriktionssystem passen würde, konnte aufgrund der ungünstigen Verteilung der repetitiven Elemente in diesem DNA-Abschnitt oder unspezifischer Bindungen der getesteten Sonden nicht gefunden werden (Kapitel 3.3.4). Erst die Sequenzierung (Kapitel 3.2.2) ermöglichte die Identifikation

geeigneter Sequenzabschnitte, die dann zur Herstellung einer Sonde Verwendung fanden, aber gleichzeitig die Entwicklung eines neuen Restriktionssystems mit dem Enzym *Eco R*V notwendig machte. Das aus dem Verdau resultierende Restriktionsfragment ließ sich durch Hybridisierung als diskrete Bande im Molekulargewichtsbereich von 20300-20600 bp nachweisen. Das erwartete Rekombinationsereignis sollte eine zweite Bande erzeugen, die genau 1015 bp kleiner als die Wildtyp-Bande im Southernblot erscheint. Da sich die nachzuweisenden DNA-Fragmente hinsichtlich ihrer Größe an der Auflösungsgrenze von linearen DNA-Fragmenten einer Standard-Gelelektrophorese befinden (Rill et al., 2002; Doggett et al., 1992; Heller et al., 1989) und die für die Restriktion zur Verfügung stehenden DNA-Mengen mit 0,5-15 µg sehr gering waren, musste ein speziell auf diese Bedingungen abgestimmtes Restriktions- / Detektionssystem entwickelt werden (Kapitel 2.6). Die Aufarbeitung der embryonalen Stammzellen erfolgte aus Zeitgründen im Schnellverfahren in 96'er Mikrotiterplatten. Daraus resultierten zunächst relativ unsaubere DNA-Präzipitate, die sich mit Restriktionsenzymen in enzymspezifischen Puffern der Hersteller teilweise nur unbefriedigend schneiden ließen. Dieses Problem konnte mit einem selbst gemischten Puffersystem umgangen werden (Kapitel 2.6.1). Auch das zunächst verwendete Standardprotokoll für die Southern-Hybridisierung (Kapitel 2.5.1) lieferte keine reproduzierbaren Ergebnisse. Erst die Adaption eines Protokolls von Foley (Soriano et al., 1991; Foley et al., 1998) an eigene Erfahrungen und die Kombination mit weiteren Protokollen ermöglichte ein standardisiertes Verfahren zur Detektion der endogenen Bande mittels Southernblot unter Verwendung hochauflösender Agarosegele.

4.3 *Dmxl*1 „Knock-Out"

4.3.1 Gen-Targeting

Das Targeting von *Dmxl*1 erfolgte an embryonalen Stammzellen des Typs 129-R1 durch Elektroporation der Zellen mit dem Targeting-Vektors in vier unabhängigen experimentellen Ansätzen. Die ersten drei Versuche zeichneten sich durch eine extrem niedrige Transformationseffizienz aus (Kapitel 3.3.3). Dies konnte im Folgenden auf die Art der DNA-Isolierung zurückgeführt werden. Für die ersten drei Transformationen wurde die DNA nach der Whitehead-Methode isoliert (Kapitel 2.3.2). Bei dieser Art der Plasmid-Präparation werden massiv Endotoxine freigesetzt (Cotten et al., 1994). Bei Endotoxinen handelt es sich um Zellwandbestandteile gramnegativer Bakterien, so genannte Lipopolysaccharide (LPS). Die chemische Struktur der LPS-Moleküle bewirkt die Ausbildung von micellaren Strukturen, die sich wie große DNA-Moleküle verhalten, diese werden bei der DNA-Aufarbeitung coisoliert. Endotoxine führen zum Absterben sensitiver eukaryotischer Zellen, bei embryonalen Stammzellen handelt es sich um einen endotoxin-sensitiven Zelltyp (Cotten et al., 1994). Eine LPS-Kontamination der DNA-Präparation führt daher zu einer deutlichen Reduktion der Transfektionseffizienz von ES-Zellen (Weber et al.,

1995). Um eine endotoxin-freie Präparation des Tageting-Vektors zu erhalten, wurde für die vierte Transformation der „EndoFree® Plasmid Maxi Kit“ der Firma Qiagen für die DNA-Isolierung verwendet. Da jedoch Endotoxine stark an Glas- und Plastikwänden adhärieren, wurden zusätzlich alle verwendeten Materialien durch zweifaches Sterilisieren bei 230 °C für 4h bzw. mehrfaches Autoklavieren endotoxin-frei gemacht. Die vierte Transformation zeigte dann zwar eine um 200% gesteigerte Transfektionsrate, lag aber mit 391 Klonen immer noch deutlich niedriger als die unter den beschriebenen experimentellen Bedingungen erwarteten 600-800 Klonen (Blessing, persönliche Auskunft).

Aus zusammen 759 untersuchten Klonen konnte eine positiv-rekombinante Zelllinie durch Southernanalyse identifiziert werden. Dies entspricht 0,13 % und liegt damit im Bereich publizierter Analysen der Rekombinationsfrequenz bei Targeting-Experimenten an ES-Zellen bei Mäusen (Ohtsuka et al., 2006; Saito et al., 2004; Vasquez et al., 2001; Liang et al., 1998; Thomas et al., 1986). In diesem Zusammenhang muss die Möglichkeit eines ungewollten Targetings auf Chromosom 4 diskutiert werden. Da es sich bei Cosmid N10467Q3 um ein *in vitro*-Ligat handelt (Kapitel 3.5.3) und die terminalen 2358 bp des rechten Targetingarms von Chromosom 4 stammen, ist eine Reduktion der Transformationseffizienz durch diesen Umstand nicht auszuschließen. Datenbanksuchen zeigten, dass das Fragment in dem Bereich 135651528-135653885 (Datenbanklokalisation AL 627214.15.1.182829) auf Chromosom 4 lokalisiert ist. Die Übereinstimmung mit den Datenbankeinträgen beträgt 100 %, es gibt keine Hinweise auf kodierende Sequenzen in diesem Bereich. Eine theoretisch denkbare Integration des homologen Sequenzabschnitts würde somit vermutlich keinen Effekt haben und auch zu keinem veränderten Phänotyp führen. Die Länge des für eine mögliche homologe Rekombination zur Verfügung stehenden Sequenzabschnitts ist mit 2358 bp suboptimal (Deng et al., 1992; Shulman et al., 1990; Thomas et al., 1987). Es kann davon ausgegangen werden, dass insgesamt die Effizienz des spezifischen Gen-Targetings durch fehlgeleitete DNA-Moleküle, die dann für die entsprechende Zielsequenz nicht mehr zur Verfügung standen, um wenige Prozent gesenkt wurde, erklärt aber nicht die grundsätzlich niedrige Transformationseffizienz. Die Wirkung eines Targeting-Experiments hängt von sehr vielen individuellen Faktoren ab (Lukacsovich et al., 2001), eine exakte Darstellung der Gründe für die reduzierte Effizienz ist somit ohne weitere Untersuchung nicht möglich und würde im Rahmen dieser Arbeit zu weit führen.

4.3.2 *Dmxl*1 „Knock-Out“ – hochchimäre Tiere zeigen einen neuen Phänotyp

In der vorliegenden Arbeit wurden 120 Blastocysten mit jeweils ca. 15 ES-Zellen der positiven Zelllinie 4.P14-D4 injiziert. Diese wurden dann in 12 NMRI-Empfängerweibchen implantiert, aus denen sich 18 Nachkommen entwickelten. Von diesen konnten 10 Tiere als chimär eingestuft werden. Im Allgemeinen liegt die Frequenz chimärer Nachkommen bei 2-10% (Nagy et al., 1993; Mombaerts et al., 1992;

DeChiara et al., 1990; Schwartzberg et al., 1989), in dieser Arbeit wurden 8% erreicht. Die männlichen Chimären wurden mit weiblichen Mäusen verpaart, aus diesen Züchtungen entstanden 67 reinerbige C57/BL/6N-Nachkommen, heterozygote Tiere wurden nicht geboren. Phänotypisch zeigten die Chimären, in Abhängigkeit vom Grad des Chimärismus, Fettleibigkeit und Hypogonadismus. In diesem Zusammenhang muss zunächst geklärt werden, ob die beobachtete Obesitas möglicherweise auf die in dieser Arbeit verwendeten Mausstämme bzw. embryonale Stammzellen und deren genetischen Hintergrund zurückzuführen sein kann. In unterschiedlichen Arbeiten konnte in der Tat gezeigt werden, dass die Ausprägung von Fettleibigkeit ganz entscheidend von dem genetischen Hintergrund der verwendeten Mausstämme abhängt (Leiter, 2002). Verschiedene Untersuchungen zeigten, dass vor allem die genetische Diversität der 129-Substämme sehr groß ist (JAX Bulletin # 1, 2001, JAX® Mice Nomenclature; Simpson et al., 1997; Threadgill et al., 1997). Da die Ursachen für Obesitas häufig auf eine komplexes Zusammenspiel vieler verschiedener Gene und alleler Ausprägungen zurückzuführen ist (Kumar et al., 2008; Diament et al., 2003), kann eine zufällige allele Kombination plötzlich zu einer verstärkten Ausprägung einer vorhandenen Fettleibigkeit führen (Itoi-Babaya et al., 2007; Paracchini et al., 2005; Leiter 2002; Kido et al., 2000). Die in dieser Arbeit verwendeten Mausstämme bzw. Zellen des Typs C57/BL/6N und 129-R1 können jedoch in dieser Hinsicht als unbedenklich eingestuft werden. Während z.B. der Stamm C57/BL/6J und Zellen des Typs 129/J eine massive Ausprägung von Fettleibigkeit bzw. Diabetes Typ II vor allem vor einem gemischten genetischen Hintergrund zeigen (Mouse Genome Database, Mouse Genome Informatics, Mouse Phenome Database, The Jackson Laboratory, 2003; Leiter, 2002;), ist eine solche Voraussetzung bei den Mausstämmen bzw. Zellen des Typs C57/BL/6N und 129-R1 nicht gegeben (Mouse Genome Database, Mouse Genome Informatics, Mouse Phenome Database, The Jackson Laboratory, 2003; Bachmanov et al., 2001,1998; Barsh et al., 2000; Rhees et al., 2000; Crabbe et al.,1999; Naveilhan et al., 1999; Dong et al., 1997). Es kann somit festgestellt werden, dass die gezeigte dramatische Gewichtszunahme nicht auf den genetischen Hintergrund zurückzuführen ist.

Zusammenfassend kann der beobachtete Phänotyp der Tiere wie folgt beschrieben werden: Die hier erzeugten männlichen hochchimären Mäuse zeigten in Abhängigkeit des prozentualen Chimärismus eine extreme Obesitas und die Testes der Tiere waren in ihrer Größe um ca. 80 % reduziert (Hypogonadismus) ohne eine Feminisierung der männlichen Mäuse zu bewirken. Der Reproduktionsapparat war prinzipiell funktionstüchtig, da alle Tiere reinerbige C57/BL/6N-Nachkommen zeugten. Weibliche Tiere wiesen ebenfalls keine erkennbare Beeinträchtigung der Ovarien oder Genitalien auf, allerdings zeigten beide untersuchten weiblichen Mäuse nur einen Chimärismus von 25 %. Es ist aber durchaus denkbar, dass in weiblichen chimären Tieren der Knock-Out Effekt wesentlich größer ist als in männlichen Tieren und alle hochchimären weiblichen Tiere schon in der frühen Embryonalentwicklung abgestorben sind. Aussagen über mögliche Veränderungen bei weiblichen Tieren können nicht getroffen werden.

In der Literatur sind inzwischen zahlreiche Mausmodelle für Fettleibigkeit beschrieben worden. Dazu gehören zum einen Mutationen einzelner Gene, die bei Inzuchtstämmen

identifiziert wurden und als Modelle für Obesitas Verwendung finden, zum anderen „Knock-Out“ und transgene Mausmodelle, bei denen gezielt Kandidatengene, die eine vermutete Rolle bei der Entstehung von Fettleibigkeit spielen, ausgeschaltet oder überexprimiert wurden. In Tabelle 4.1 ist eine Auflistung aller wichtigen Mausmodelle für Fettleibigkeit wiedergegeben (Butler et al., 2001; Brockmann et al., 2002). Weiter gibt es zahlreiche so genannte QTL’s („quantitative trait loci“), die für eine polygene Obesitas verantwortlich sind und einen spezifischen Phänotyp durch additive Effekte verschiedener Allele und in Abhängigkeit von Fütterung, Alter und Geschlecht und genetischem Hintergrund erzeugen (Brockmann et al., 2002). In Abb. 4.2 sind alle bekannten chromosomalen Loci aufgezeigt, die bei der Entstehung von Fettleibigkeit eine Rolle spielen.
Alle bekannten Maus-Phänotypen, die mit Obesitas assoziiert sind, wurden mit dem Phänotyp der vorliegenden *Dmxl*1-„Knock-Out“ Chimären verglichen. Nur in einem Fall wurde eine Fettleibigkeit in Verbindung mit leicht verkleinerten Testes beschrieben, hierbei handelt es sich um die so genannte ob/ob-Maus (Zhang et al., 1994; Jones et al., 1957; Ingalls et al., 1950). Bei diesen Tieren liegt eine homozygote Mutation des Leptin-Gens vor (Zhang et al., 1994). Heterozygote Tiere zeigen diese Defekte nicht. Die Reduktion der Testes ist moderat (Mounzih et al., 1997) und in keinem Fall mit dem vorliegenden Phänotyp vergleichbar, zudem sind die Tiere völlig steril. In einigen Publikationen wurde Fettleibigkeit in Verbindung mit einer Haploinsuffiziens beschrieben (Holder et al., 2004; Lubrano-Berthelier et al., 2003; Wolfrum et al., 2003; Xu et al., 2002; Michaud et al., 2001; Cone, 2000; Dhar et al., 2000; Farooqi et al., 2000; Vaisse et al., 2000; Sina et al., 1999). Bei Untersuchungen an Mäusen führte hier in allen Fällen ein „Knock-Out“ schon bei heterozygoten Tieren ($X^{+/-}$) zu einer Zu- oder Abnahme von Fettleibigkeit. Haploinsuffiziens kann auftreten wenn: a.) ein Genprodukt in ein quantitatives Signaltranduktionssystem eingebettet ist, b.) es sich um miteinander konkurrierende Genprodukte handelt, oder c.) Genprodukte in einer festen Stöchiometrie miteinander interagieren. Gewichtszunahmen, wie sie in dieser Arbeit für chimäre *Dmxl*1-„Knock-Out“-Mäuse beschrieben wurden, konnten, bei einem vergleichbaren genetischen Hintergrund von C57/Bl/6-129/Sv, nur bei den haploinsuffizienten Tieren durch eine Hoch-Fett Diät erreicht werden (Michaud et al., 2001). Tiere mit einem hohen Chimärismusgrad haben einen entsprechend hohen Anteil von Zellen mit einem Null-Allel am Gesamtorganismus, im Falle einer bestehenden Haploinsuffiziens kann angenommen werden, dass sich der Grad der Auswirkung auf den Phänotyp mit dem Grad des Chimärismus ändert. Genau das konnte in dieser Arbeit nachgewiesen werden (Kapitel 3.3.6). Es zeigte sich ein deutlicher Zuwachs des Gewichts in Abhängigkeit eines ansteigenden Chimärismusgrads, wobei auch die niederchimären Tieren bei einem Vergleich mit Referenzstämmen ein erkennbares Übergewicht zeigten. Da bei dem vorliegenden Phänotyp eine Haploinsuffiziens als Ursache für die bestehende Obesitas in Betracht gezogen werden muss, wurde auch im Falle der morphologisch extrem verkleinerten Testes nach vergleichbaren Publikationen gesucht.

Gen-Symbol	Gen-Produkt	Funktion bzw. Gen-Defekt	Publikation
Mutationen einzelner Gene, die bei Inzuchtstämmen identifiziert wurden			
A^y	Agouti Signalpeptid	Antagonist des Melanocortin Signalwegs	Miller et al., 1993
Atrn	Attractin	Für Funktion des Agouti Signalpeptids nötig	Gunn et al., 1999
Cpe	Carboxypeptidase E	Wandelt Proinsulin und Proopiomelanocortin in die biologisch aktive Form um	Naggert et al., 1995
Lep	Leptin	Sättigungshormon, von Fettzellen produziert	Zhang et al., 1994
Lepr	Leptin-Rezeptor	Leptin Rezeptor	Tartaglia et al., 1995
Tub	Insulin Signalprotein	Involviert im Insulin Signalweg	Kleyn et al., 1996
KO oder TG einzelner Gene auf dem Hintergrund von Wild-Typ Stämmen und verschiedener Stämme mit einer Prädisposition für Fettleibigkeit			
Adr3b	Adrenorezeptor β-3	Metabolismus?	Susulic et al., 1995, Grujic et al., 1997
Brs3	Bombesin-Rezeptor Subtyp 3	Hyperphagie, Metabolismus	Ohki-Hamazaki et al., 1997
Cyp 19	Cytochrom 450, Aromatase	Reduzierte Aktivität, sekundär	Jones et al., 2000
Dgat1	Diacylglycerol O-acetyltransferase	An der Entwicklung des Hypothalamus beteiligt	Smith et al., 2000
Esr1	Östrogen-Rezeptor-α	Reduzierte Aktivität, Metabolismus	Heine et al., 2000
Fshr	FSH-Rezeptor	-	Danilovich et al., 2000
Gabt1	Γ-Amino-buttersäure (GABA-A) Transporter 1	Zentraler Signaltransfer	Ma et al., 2000
Hcrt	Hypocretin (Orexin)	Neurotransmitter	Hara et al., 2001
Hsd11b1	Hydroxysteroid 11-β Dehydrogenase 1	Triglycerid Synthese	Masuzaki et al., 2001
Htr2c	5-Hydroxytryptamin (Seratonin)-Rezeptor 2C	Hyperphagie	Tecott et al., 1995; Nonogaki et al., 1998; Heisler et al., 1998
Insr	Insulin-Rezeptor	Insulin Signalweg	Bruning et al., 2000
Mc3r	Melanocortin-3 Rezeptor	Reduzierte Aktivität	Chen et al., 2000; Butler et al., 2000
Mc3r	Melanocortin-4 Rezeptor	Reduzierte Aktivität, Hyperphagie, Metabolismus	Marie et al., 2000; Butler et al., 2001; Huszar et al., 1997; Marsh et al., 1999
Npy	Neuropeptid Y (NPY)	Kein Defekt	Erickson et al., 1996; Erickson et al., 1996
Npy1r	NPY1-Rezeptor	Reduzierte Aktivität, Metabolismus	Pedrazzini et al., 1998; Kushi et al., 1998
Npy2r	NPY2-Rezeptor	Hyperphagie, reduzierte Aktivität	Naveilhan et al., 1999
Npy5r	NPY5-Rezeptor	Hyperphagie	Marsh et al., 1998
Pmch	Promelanin Konzentrations-Hormon	Gewichtsregulation	Ludwig et al., 2001
Pomc1	Pro-opiomelanocortin	Hyperphagie	Yaswen et al., 1999; Forbes et al., 2001
Ppara	Peroxisom-Proliferator, aktivierter Rezeptor-α	Fetteinlagerung	Lee et al., 1995
Ptpn1	Protein-Tyrosin-Phosphatase, non-Rezeptor 1	Regulation der Insulinwirkung	Klamann et al., 2000

Tab 4.1: Zusammenfassung der wichtigsten Gendefekte, die bei der Ausprägung von Fettleibigkeit eine Rolle spielen
KO: „Knock-Out“
TG: Transgen

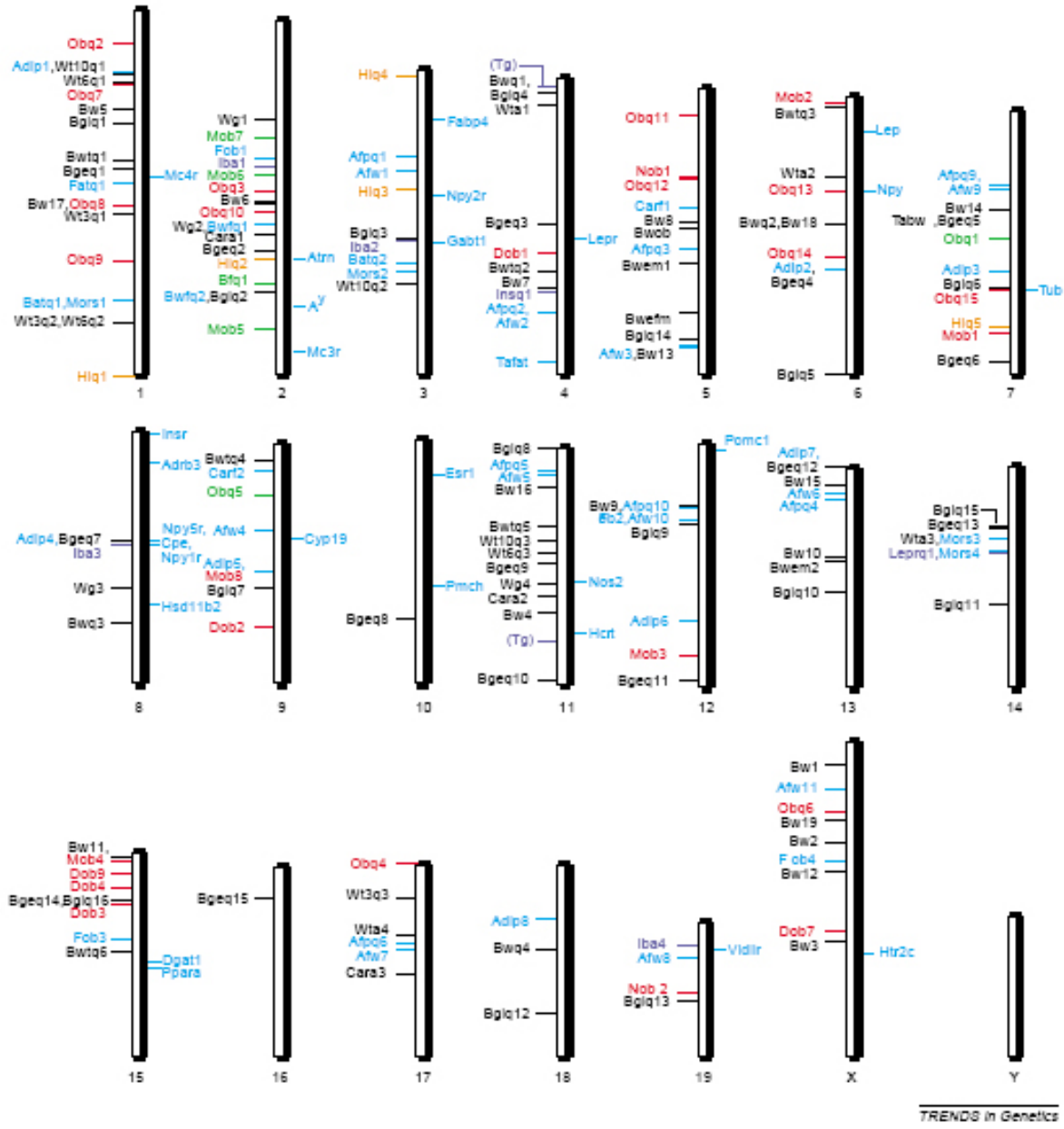


Abb. 4.2: Chromosomale Lage von Genen und QTL's, die Körpergewicht und Fettleibigkeit bei Mäusen beeinflussen (Brockmann et al., 2002)

Auf der linken Seite jedes Chromosoms sind die QTL-Positionen (Quantitative Trait Loci) dargestellt: Körpergewicht allgemein (schwarz); Fettleibigkeit und „heat loss" (erhöhte Stoffwechselaktivität mit reduzierter Fetteinlagerung und verstärkter Nahrungsaufnahme, orange); Diät-induzierte Fettleibigkeit (rot); Diät-induzierte Fettleibigkeit, die auch Subphänotypen beeinflusst (grün); mit Fettleibigkeit assoziierte Subphänotypen, die durch die QTL's nicht direkt beeinflusst werden (violett); andere Ursachen für Fettleibigkeit (blau). Auf der rechten Seite der Chromosomen sind die Positionen von Genen wiedergegeben, die bei Mutanten oder durch Gen-Targeting erzeugten Mausstämmen eine Zu- oder Abnahme von Fettleibigkeit bewirken. Auf Chromosom 18 wurden bisher drei QTL-Positionen identifiziert, keine dieser QTL's befindet sich in der Nähe von *Dmxl*1. Ein direkter Zusammenhang der beobachteten Obesitas mit dem Targeting-Ereignis selbst ist unwahrscheinlich.

Als Ursache für die Größenreduktion der Hoden bei den hochchimären Tieren könnte auch ein grundsätzlicher Defekt auf DNA-Ebene angenommen werden. Ein solcher Effekt wäre unabhängig vom Target-Ereignis und bedingt durch die vielfältigen Manipulationen der verwendeten embryonalen Stammzellen, der zu einem prinzipiellen Ausschluss der injizierten ES-Zellen an der Keimbahnentwicklung führen könnte. Dies erklärt jedoch nicht die Reduktion der Testes um ca. 80 % bei den Hochchimären und die Gewichtszunahme aller chimären Tiere. Vorstellbar ist hier eher der ursächliche Zusammenhang mit dem Target-Ereignis und einer daraus resultierenden Haploinsuffiziens (Cho et al., 2003, 2001; Li et al., 2002; Seidman et al., 2002; Veitia et al., 2002, 2001) in der Keimbahn- oder der Embryonalentwicklung. Da die chimären Tiere nicht in der Lage waren, das *Dmxl*1-Null-Allel zu transduzieren, kann neben einer Haploinsuffiziens in der Spermatogenese oder bei der Keimzellentwicklung auch eine Haploletalität (Veitia, 2002) bei einem 50 %igen Verlust der *Dmxl*1-Transkripte in Erwägung gezogen werden. Es gibt zahlreiche Publikationen, bei denen ein „Knock-Out“ zu Hypogonadismus führte (u.a. Yan et al., 2004; Kasai et al., 2003; Yeh et al., 2002; Veitia et al., 2001; Zhang et al., 2001; Zhang et al. 2001; Kumar et al., 2000), jedoch nur wenige Arbeiten, die in Verbindung mit einer Haploinsuffiziens einen solchen Phänotyp aufzeigten. In der Arbeit von Yan et al. (2004) ist eine solche Kombination beschrieben: Hier wurde ein „Knock-Out“ des Gens erzeugt, das für das KLHL10-Protein (kelch-like protein homolog 10) kodiert. Es handelt sich dabei um ein bei Mammaliern hochkonserviertes Testes-spezifisches Gen. Das resultierende Protein gehört interessanterweise zu der Superfamilie der Kelch-Proteine und besitzt C-Terminal sechs Kelch-Wiederholungseinheiten, die in der Lage sind, genau wie WD-Wiederholungseinheiten, einen Propeller zu formen (Kapitel 4.4.8). Das Protein ist entweder Bestandteil des E3-Ubiquitin-Ligase-Komplexes, der Ubiquitinierungsvorgänge während der Spermiogenese katalysiert, oder übt seine Funktion als cytoskeletales Protein bei der Formgebung der Spermatiden aus. Die aus dem „Knock-Out“-Experiment resultierenden Chimären waren teilsteril bis steril und die Testesgröße betrug 50-60 % des Wildtyps, also deutlich größer als bei den *Dmxl1*-Chimären, eine Feminisierung der Genitalien wurde nicht beobachtet, ebenso wenig eine Veränderung des Körpergewichts. Der Hypogonadismus ist auf apoptotische Vorgänge in Spermatogonien und Spermatocyten sowie degenerative haploide Keimzellen zurückzuführen. In der Arbeit von Kasai et al. (2003) erzeugte ein „Knock-Out“ des *Bcl-x* Gens bei heterozygoten Tieren eine Reduktion der Testes auf 40 % des Wildtyps, allerdings in Zusammenhang mit einer deutlichen Abnahme des Körpergewichts. Andere Organe des Reproduktionsapparates, wie seminale Vesikel oder Epididymis waren nicht betroffen. Bei Bcl-x handelt es sich um ein ubiquitär exprimiertes anti-apoptotisches Molekül, das unter anderem bei der gezielten Apoptose bestimmter Zellen in der Keimbahnentwicklung eine Rolle spielt (Kasai et al., 2003; Print et al., 2000; Adams et al., 1998; Print et al., 1998). Die Tiere waren vermindert fertil, resultierende homozygote Embryonen starben im Stadium E13.5. Bei der so genannten ARKO-Maus (Yeh et al., 2002) konnte bei männlichen Tieren eine Reduktion der Testes um 80 % beobachtet werden, also ein vergleichbarer Wert mit den in dieser Arbeit erzeugten Chimären. Die ARKO-Mäuse waren vollständig infertil und zeigten eine deutliche Abnahme des Körpergewichts. Die Tiere wiesen eine klare

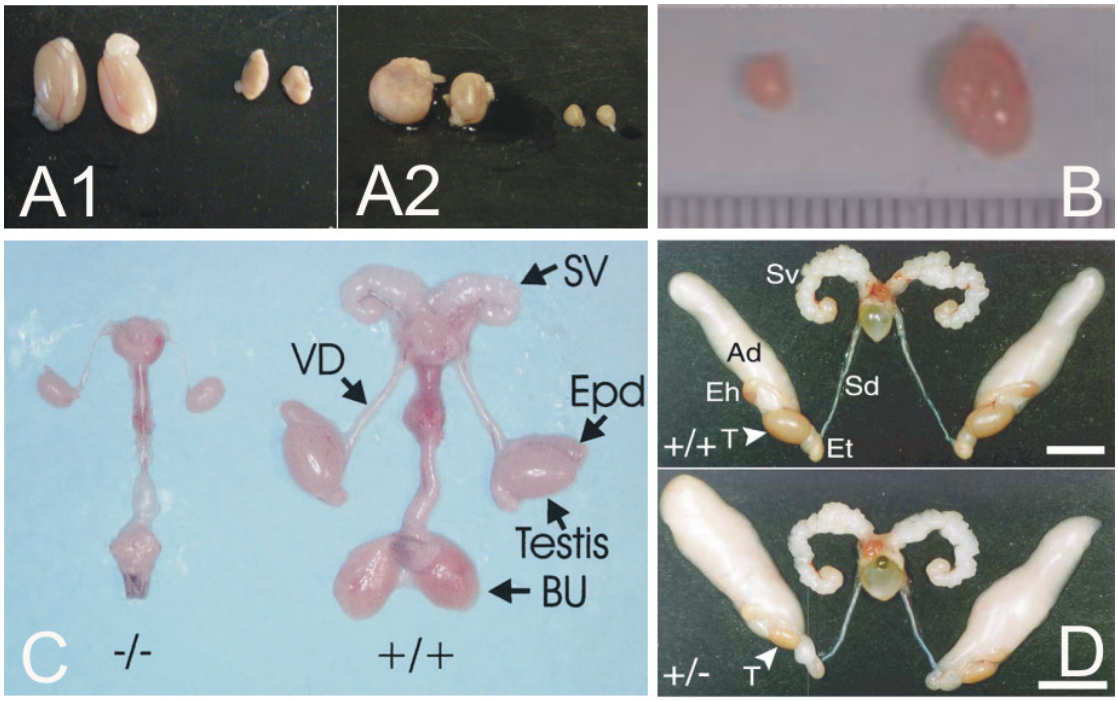


Abb. 4.3: Vergleich verschiedener „Knock-Out"-Modelle, die einen Phänotyp mit verkleinerten Testes aufweisen (folgende Seite).

A1: Präparat aus der Dmxl1$^{+/-}$-chimären Maus Nr. 1 (links, 25 % chimär, Größe der Testes wie Wildtyp) und Maus Nr. 2 (rechts, 75 % chimär,), vergl. Kapitel 3.3.6
A2: Präparat aus der Dmxl1$^{+/-}$-chimären Maus Nr. 5 (links, 25 % chimär, Größe der Testes wie Wildtyp) und Maus Nr. 3 (rechts, >75 % chimär), vergl. Kapitel 3.3.6
B: Vergleich der Testes eines gemischten genetischen Hintergrunds aus 129/B6. In der rechten Bildhälfte ist der Wildtyp zu erkennen, in der linken Bildhälfte ist ein Hoden aus der ARKO-Maus gezeigt, Körper und Genitalien zeigen eine deutliche Feminisierung, die Testes sind gegenüber dem Wildtyp um 80 % verkleinert, die Tiere sind infertil (aus Yeh et al., 2002)
C: Präparat des Reproduktionsapparates aus der LuRKO-Maus (linke Bildhälfte) verglichen mit dem Wildtyp (rechte Bildhälfte), die Testes sind gegenüber dem Wildtyp um 80 % verkleinert, die Tiere sind steril und bleiben in ihrer Entwicklung vor der Geschlechtsreife stehen, eine Veränderung des Körpergewichts konnte nicht beobachtet werden (aus Zhang et al., 2001). VD: Vas deferens; SV: seminal vesicle; Epd: epididymis; BU: bulbo-urethral gland
D: Vergleich eines Präparats aus einer Bcl-x $^{+/-}$-Maus (untere Bildhälfte) mit dem Wildtyp (obere Bildhälfte), die Testes des haploinsuffizienten „Knock-Out" sind gegenüber dem Wildtyp um 40-50 % verkleinert, das Körpergewicht reduziert, die Tiere sind fertil und gleichen sonst morphologisch dem Wildtyp (aus Kasai et al., 2003). AD: adipose tissue; Eh: head of epididymis; Sd: Spermiduct; Sv: seminal vesicle; T: testis

Feminisierung des gesamten Körpers auf. Das betroffene Gen kodiert für den Androgen-Rezeptor (AR) und wird ebenfalls in zahlreichen Geweben exprimiert (Heinlein et al., 2002). Ein Vergleich reduzierter Testes verschiedener „Knock-Out"-Modelle mit den Testes von *Dmxl*1-Chimären ist in Abbildung 4.3 dargestellt. Die Beispiele der Publikationen zeigen, dass ubiquitär exprimierte Gene bei einem „Knock-Out" zu einer Haploinsuffiziens in Kombination mit Hypogonadismus führen können. Allerdings wurden diese morphologischen Merkmale bisher nicht in Verbindung mit einer ausgeprägten Obesitas beschrieben. In keiner Publikation konnte also ein

ähnlicher Phänotyp wie bei den in dieser Arbeit erzeugten Chimären gefunden werden. Ein grundsätzliches Problem besteht darin, die beobachtete Obesitas und den Hypogonadismus in Zusammenhang zu bringen. Es kann hier nicht entschieden werden, ob die Fettleibigkeit eine Folge der unterentwickelten Testes ist, oder ob es sich bei beiden gezeigten morphologischen Veränderungen um unabhängige und möglicherweise jeweils sekundäre Effekte handelt. Hierzu müssten umfangreiche Verhaltensstudien zur Nahrungsaufnahme, histologische und histochemische Untersuchungen sowie Blutbilduntersuchungen durchgeführt werden. In Kapitel 4.4.9 wird jedoch über einen Zusammenhang zwischen dem beobachteten Phänotyp und einer möglichen Funktion von *Dmxl*1 diskutiert.

4.3.3 *Dmxl*1 „Knock-Out" – Vergleich mit „Knock-Out"-Experimenten in anderen Spezies

Bisher wurden drei Versuche unternommen, *Dmxl*1 homologe Gene anderer Spezies zu disruptieren bzw. die Umsetzung in ein funktionsfähiges Protein zu verhindern. Bei *Drosophila melanogaster* wurden von Weil (2000), in Zusammenarbeit mit Kraemer, verschiedene Methoden zur Erzeugung von *DmX*-Mutanten angewendet. Die Verwendung einer Antisense-Strategie sowie die Konstruktion eines *DmX*-spezifischen Ribozyms führten dabei zu keinem erkennbaren Phänotyp der resultierenden Fliegen. Eine EMS-Mutagenese in Verbindung mit einem spezifisch konstruierten Rettungsvektor erzeugte schließlich *DmX* $^{-}$-Tiere. Die *DmX*-Mutanten führen eine normale Embryogenese durch und sind zum Anfang des L1-Stadiums nicht vom Wildtyp zu unterscheiden. In der folgenden Entwicklung läßt jedoch Agilität deutlich nach, bis die Larven nur noch auf der Stelle verharren. Durch mechanische Reize konnten sie jedoch zur Bewegung veranlasst werden. Vor dem Erreichen des L2-Stadiums sterben die meisten Tiere, sie wirken zu diesem Zeitpunkt transparent und schlaff. Abgeschwächte Phänotypen, die eher mit den *Dmxl*1-Chimären vergleichbar wären, können nicht identifiziert werden (Weil, 2000), eine direkte Korrelation des Phänotyps der *DmX*-Mutanten mit den beobachteten morphologischen Charakteristika der chimären Mäuse kann somit nicht hergestellt werden, da heterozygote *DmX*-Mutanten keinen erkennbaren Phänotyp aufweisen und der Phänotyp von *DmX* $^{-}$-Tieren durch Einkreuzen einer transgenen Kopie von *DmX* revertiert werden (Weil, 2000). Bei dem Nematoden *Caenorhabditis elegans* wurde im Verlauf der systematischen Funktionsanalyse des Genoms durch RNAi versucht, die Transkription des zu *Dmxl*1 homologen Gens F54E4.1 durch Fütterung mit Bakterien, die eine für dieses Gen spezifische doppelsträngige RNA (dsRNA) exprimieren, zu unterbinden (Kamath et al., 2003). F54E4.1 wird in *C. elegans* ubiquitär in allen Stadien, vom 100-Zell-Embryo bis zum adulten Tier, zum Teil mit einer leichten Mosaikstruktur, exprimiert und zeigt eine deutliche nukleäre Lokalisation (Mounsey, Bauer, Hope, WormBase, 2000; Abb 4.4). Der resultierende Phänotyp entsprach in allen Stadien dem Wildtyp, verwendet wurde das Konstrukt sjj_F54E4.1 (1163 bp, genomische Lokalisation 1466269-14617432 auf dem X-Chromosom), weiter konnte keine erhöhte embryonale Letalität festgestellt werden (Kamath et al., 2003). Grundsätzlich ist hierbei anzumerken, dass nur ein

einziges Experiment durchgeführt wurde und nur 63,5 % aller möglichen Phänotypen mit dieser Methode überhaupt erzeugt werden können (Kamath et al., 2003, 2000; Mounsey et al., 2002). Hinzu kommt, dass die erfolgreiche Anwendung von RNAi ganz entscheidend von zellspezifischen (Organismus oder Zelllinie, Degradation, Kreuzhybridisierung oder Konformation des Zielgens usw.) und molekularen Faktoren (Länge des Konstrukts, Duplex-Konformation, thermodynamische Stabilität, Konzentration und Lokalisation auf dem Zielgen usw.) abhängt. Diese Faktoren lassen sich nur mit komplexen mathematischen Berechnungen quantifizieren (Pancoska et al., 2004), tragen aber ganz entscheidend zu Erfolg oder Misserfolg eines Experimentes bei (Pancoska et al., 2004; Schütze, 2004; Kamath et al., 2000). Da *DmX*-Mutanten bei *Drosophila melanogaster* einen letalen Phänotyp bei einer ebenso ubiquitären Expression zeigen (Weil, 2000), ist es eher unwahrscheinlich, dass F54E4.1 bei einer erfolgreichen Unterdrückung der Expression keinen erkennbaren Phänotyp erzeugt. Es kann vielmehr angenommen werden, das F54E4.1 entweder zu den 36,5 % jener Gene gehört, die sich mit der Fütterungs-Methode nicht erfassen lassen, oder einige der oben genannten Faktoren könnten dieses eine Experiment negativ beeinflusst haben.

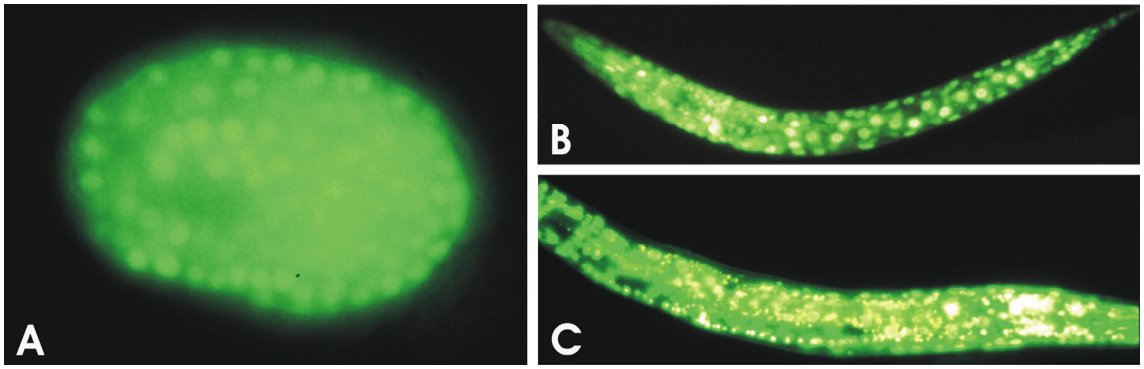


Abb 4.4: *in situ* Expressionsprofil von *F54E4.1*

Gezeigt ist die Expression von F54E4.1 in drei verschiedenen Entwicklungsstadien, erzeugt durch ein GFP-Fusionsprodukt aus einem *Hind* III – *Sal* I-Restriktionsfragment aus Cosmid *F54E4* kloniert in den Leserahmen 2 eines GFP-Vektors (Mounsey et al., euwm2000ab73, Mounsey et. al, WormBase, 2000). In allen Entwicklungsstadien sind eine ubiquitäre Expression und eine Lokalisation im Zellkern zu finden.

A: Embryo, 16E-Zell-Stadium, anterior auf der rechten Seite
B: L1 Larve, anterior auf der linken Seite
C: Adult, mittlerer Körperbereich, anterior rechts

Der offene Leserahmen YJR033c (Standardname *SOI*3/*RAV*1 (Suppressors of rapid „onset of Impotence“/Regulator of (H+)-ATPase in Vacuolar membrane)) aus *Saccharomyces cervisiae* zeigt begrenzte aber signifikante Ähnlichkeiten u.a. zu einem zentralen Abschnitt von *Dmxl*1 (AS 1534-1821). *RAV*1 bildet zusammen mit *RAV*2 und *SKP*1 einen Komplex, der die Glukose-induzierte Organisation des V-ATPase Holoenzyms unterstützt (Sipos et al., 2004; Seol et al., 2001) und somit am

Stoffwechsel beteiligt ist. Es wurden bei der funktionellen Untersuchung durch systematische Deletion von Genen in der Hefe (Winzeler et al., 1999) verschiedene Mutanten für *RAV*1 erzeugt (Giaever et al., 2002; Steinmetz et al., 2002). Die resultierenden Organismen sind voll lebensfähig (Giaever et al., 2002), zeigen aber auf respiratorischen Kohlenstoffquellen einen Wachstumsdefekt (Steinmetz et al., 2002). Hier kann möglicherweise ein direkter Zusammenhang des beobachteten Phänotyps mit den *Dmxl*1-Chimären hergestellt werden (Kapitel 4.4.9).

4.4 *In silico*-Analysen

4.4.1 GC-Gehalt, Isochoren, GC_3-Gehalt und „CpG-Inseln“ der genomischen Region von *Dmxl*1 und *DMXL*1

GC-Gehalt und Isochoren

Vertebratengenome besitzen in der Regel einen durchschnittlichen GC-Gehalt von 40% und variieren um diesen Wert nur in engen Grenzen, im Gegensatz zu Invertebraten, Protisten und Prokaryoten, wobei hier noch zwischen Kalt- und Warmblütern unterschieden werden muss (Abb. 4.4; Bernardi, 2000). Dennoch weisen unterschiedliche Genomabschnitte der Vertebraten einen unterschiedlichen GC-Gehalt auf, der zwischen 30% und 60% schwankt (Bernardi, 2000), sie bestehen also aus einem Mosaik so genannter Isochoren (Bernardi et al., 1985).

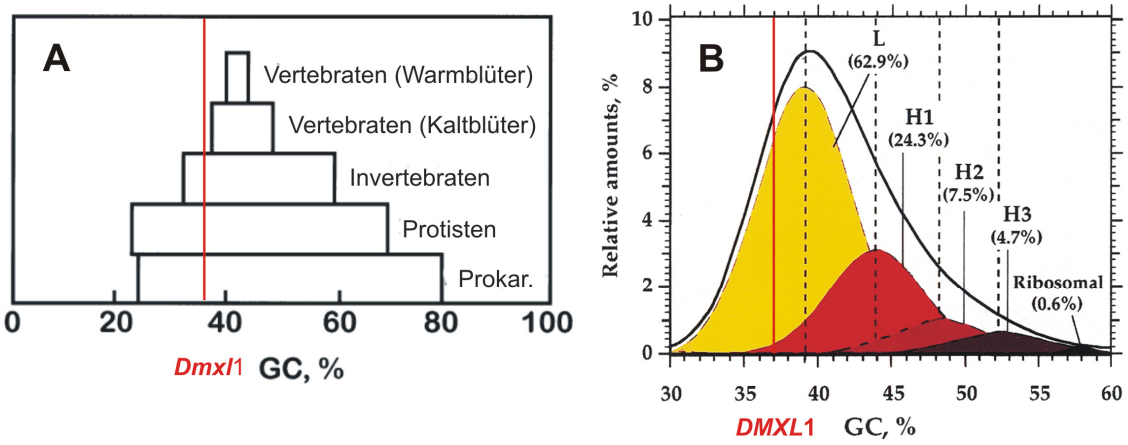


Abb. 4.5: Varianz des GC-Gehalts verschiedener Genome und Isochoren-Klassen des humanen Genoms

A: Die Balken zeigen die Grenzen an, innerhalb derer der durchschnittliche GC-Gehalt verschiedener Genome schwankt (Bernardi, 2000). Zur Orientierung ist der GC-Gehalt der genomischen Region des murinen *Dmxl*1 mit einer roten Linie gekennzeichnet, dieser Bereich zeigt einen unterdurchschnittlichen GC-Gehalt.
B: Verteilung des prozentualen GC-Gehalts des humanen Genoms auf die einzelnen Isochoren-Klassen (Bernardi, 1999). Die genomische Region des humanen *DMXL*1 zeigt sich mit einem GC-Gehalt von 37 % der L1-Klasse zugehörig.

Hierbei handelt es sich um Abschnitte von >300 kb bis zu 1,4 Mb, die in sich einen uniformen GC-Gehalt aufweisen (Paces et al., 2004; Zhang et al., 2003; Bernardi et al., 2000). Das humane Genom wird in fünf Isochoren-Klassen eingeteilt: Die GC-armen Klassen L1 (38% GC) und L2 (41% GC) repräsentieren 33% bzw. 30% des Genoms, die GC-reichen Abschnitte H1 (44% GC), H2 (49% GC) und H3 (53%) haben einen Anteil von 24% bzw. 7,5% bzw. 4-5% am gesamten humanen Genom (Alvarez-Valin et al., 2002; Bernardi, 2001, 2000; Bernardi, 1993). Das murine Genom zeigt in Hinblick auf den GC-Gehalt wesentlich weniger Varianz (Mouchiroud et al., 1988; Salinas et al., 1986) und verhält sich bei der Betrachtung ganzer Chromosomen uniformer als das menschliche Genom, das hier deutlich mehr Extreme aufweist (Waterstone et al., 2002). Insgesamt liegt der GC-Gehalt des Maus-Genoms mit 42% etwas höher als der des Menschen (41%; Waterston et al., 2002). Dies spiegelt sich auch in der Länge der Isochoren wider, die Abschnitte reichen von >300 kb bis zu 21,99 Mb (Zhang et al., 2003). Hierbei kann zwischen AT-Isochoren mit einem geringeren und ansteigenden GC-Gehalt als der Durchschnitt und GC-Isochoren mit einem höheren und abfallenden GC-Gehalt als der Durchschnitt unterschieden werden. Berechnet werden diese im Verhältnis zum genomischen GC-Gehalt (Zhang et al., 2003). Die Einteilung in Klassen nach Bernardi beruht auf dem reinen GC-Gehalt der Isochoren (Pavlicek, 2002; Zoubak, et al., 1996) und bleibt davon unbenommen. AT-Isochoren entsprechen den Klassen L1 und L2, während GC-Isochoren den Klassen H1, H2 und H3 entsprechen (Zhang et al., 2003; Zoubak et al., 1996). Das Maus-Genom hat 28 Isochoren, die länger als 1 Mb sind. Auf Chromosom 18 liegt nur ein AT-Isochor mit 5,12 Mb Länge (Position 79.22 bis 84.34) und einem GC-Gehalt von 37% (Klasse L1), alle anderen Isochoren des Chromosoms sind kleiner als 1 Mb. *Dmxl*1 liegt demzufolge nicht innerhalb dieses AT-Isochors. Die genomische Region von *Dmxl*1 zeigt einen durchschnittlichen GC-Gehalt von 36%, der Bereich kann damit der Isochoren-Klasse L1 zugeordnet werden. Für die homologe Region des Menschen auf Chromosom 5 konnte ein durchschnittlicher GC-Gehalt von 37% ermittelt werden, sie gehört damit ebenfalls der Isochoren-Klasse L1 an (Kapitel 3.5.6, Abb. 3.22). Der GC-Gehalt der Isochoren-Klassen steht auch im Zusammenhang mit der beobachtete Gendichte innerhalb dieser Bereiche. Je höher der GC-Gehalt der Isochoren, desto mehr Gene werden hier beobachtet (Oliver et al., 2002; Bernardi, 2001; Saccone et al., 2001; Gardiner, 1996; Zoubak et al., 1996; D'Onofrio et al., 1991). Dabei reicht die Spanne von 0,0025 Gene pro kb bei einem GC-Gehalt von 34% bis zu 0,04 Gene pro kb für einen beobachteten GC-Gehalt von 57% (Oliver et al., 2002). Diese Werte beziehen sich auf das humane Genom, das murine Genom verhält sich aber in dieser Hinsicht gleich (Zhang et al., 2003; Smith et al., 2002). Es kann daher postuliert werden, dass sich *Dmxl*1 und *DMXL*1 jeweils in einer genarmen Region befinden. In der Tat wurden bisher in Maus und Mensch neben *Dmxl*1 bzw. *DMXL*1 nur 7 weitere Gene in einer 1 Mb umfassenden Region, in deren Mitte *Dmxl*1 liegt, identifiziert. Dies entspricht einer Gendichte von 0,008/kb und kann damit als sehr genarm eingestuft werden. Über einen möglichen Zusammenhang von Isochoren-Klassen und gewebsspezifischer bzw. ubiquitärer Expression wird aktuell ebenfalls diskutiert (Vinogradov, 2003; D'onofrio, 2002; Galtier et al., 2001; Ponger et al., 2001; Goncalves et al, 2000; Pesole et al.,

1999). Es wird angenommen, dass gewebsspezifische Gene primär in GC-reichen Isochoren gefunden werden und Gene mit einem ubiquitären Expressionsspektrum entweder in GC-armen Isochoren oder unabhängig von der Isochorenklasse exprimiert werden. Es besteht damit zumindest eine Tendenz für *Dmxl*1 mit seiner Lokalisation in einem GC-armen Isochor nicht zu den gewebsspezifischen, sondern eher zu den ubiquitär exprimierten Genen zu gehören. Diese Annahme passt auch sehr gut zu dem beobachteten Expressionsspektrum (Kapitel 4.4.5).

GC_3-Gehalt

In diesem Kontext ist auch der GC_3-Gehalt, d.h. der prozentuale Anteil von G und C Nukleotiden in der dritten Codonposition, von *Dmxl*1 interessant. Der GC_3-Gehalt steht nicht nur in direktem Zusammenhang mit der Isochorenklasse, sondern kann auch Auskunft über die Expressionsrate geben (Konu et al.,, 2001; Bernardi, 2000; Bellgard et al., 1999; Bernardi et al., 1985). Die Einteilung von Genen, basierend auf ihrem GC_3-Gehalt, erfolgt in vier Klassen: GC_3-arme Gene ($GC_3 < 45\%$), Gene mit mittlerem GC_3-Gehalt ($45\% < GC_3 < 60\%$), Gene mit hohem GC_3-Gehalt ($60\% < GC_3 < 75\%$) und Gene mit extrem hohen GC_3-Gehalt ($GC_3 > 75\%$) (Alvarez-Valin et al., 2002). Ein Vergleich von Maus und Mensch zeigte, dass der GC_3-Gehalt von GC-reichen Genen im Menschen höher ist als bei der Maus. Umgekehrt verhält es sich bei GC-armen Genen, hier ist der GC_3-Gehalt beim Menschen niedriger als in der Maus (Bernardi, 2000; Mouchiroud et al., 1988). Bei den Säugergenomen ist eine grundsätzliche Tendenz zur Abnahme des GC_3-Gehalts zu beobachten (Duret et al., 2002; Piganeau et al., 2002). Verantwortlich hierfür ist zum einen die Anhäufung von GC → AT-Mutationen in der dritten Codonposition (stille Mutationen) unter Verwendung synonymer Codons und zum anderen die Anpassung des GC-Gehalts von Genen an die sie umgebende Isochorenklasse bei gleichzeitigem Verlust GC-reicher Isochoren durch ein Ungleichgewicht von AT ↔ GC-Mutationen (Duret et al., 2002; Piganeau et al., 2002; Lander et al., 2001; Hurst et al., 2000; Galtier et al., 1998; Clay et al., 1996; Aissani et al., 1991). Erstaunlicherweise spielt hierbei die Hypermutation von CpG-Dinukleotiden durch Deaminierung methylierter Cytosine (Bird, 1980; Coulondre et al., 1978) nur eine untergeordnete Rolle (Duret et al., 2002). *DmX* weist sowohl in der Maus als auch beim Menschen einen extrem niedrigen GC-Gehalt in der dritten Codonposition auf (Tabelle 4.2). Beide homologen Gene liegen in einem GC-armen Isochor und *DMXL*1 zeigt in der Tat einen niedrigeren GC_3-Gehalt als *Dmxl*1, dies bei einem nahezu identischen GC-Gehalt beider Gene und einem höherem GC-Gehalt des humanen Isochors. Die Erosion des GC- und GC_3-Gehalts ist, verglichen mit den homologen Genen anderer Spezies (Tabelle 4.2), demnach sehr weit fortgeschritten. Neben der evolutionären Komponente des GC_3-Gehalts lässt sich mit diesem Wert auch eine Aussage über die Transkriptionsrate des jeweiligen Gens in Maus und Mensch treffen. In der Arbeit von Konu et al. (2001) konnte bei den Rodentia eine positive Korrelation zwischen der dritten Codonposition und der mRNA-Expression festgestellt werden, eine solche Abhängigkeit wird hier auch für andere Mammalia postuliert. Der 5'- bzw. 3'-UTR scheint dabei von der evolutionären Entwicklung des kodierenden Bereichs des Gens abgekoppelt zu sein. Der niedrige GC_3-Gehalt von *Dmxl*1 führt zu

der Annahme, dass es sich hierbei um ein minimal exprimiertes Gen handeln könnte, dies passt auch zu den experimentell ermittelten Daten (Kapitel 3.4).

	GC	GC_3	A_3	T_3	C_3	G_3	Isochor
***Dmxl*1 (*M. musculus*)**	0.408	0.294	0.4184	0.4583	0.1952	0.1808	0.36
***DMXL*1 (*H. sapiens*)**	0.401	0.277	0.4108	0.4857	0.1742	0.1811	0.37
***DmX* (*D. melanogaster*)**	0.584	0.543	0.3002	0.2045	0.2892	0.3232	-
***CpY* (*C. thummi*)**	0.385	0.473	0.3825	0.3051	0.2780	0.3984	-
***At2g46560* (*A. thaliana*)**	0.427	0.327	0.3556	0.4644	0.1865	0.2328	-
***F54E4.1* (*C. elegans*)**	0.452	0.416	0.3319	0.3961	0.2520	0.2884	-

Tab. 4.2: GC- und GC_3-Gehalt *Dmxl*1 homologer Gene verschiedener Spezies

Neben dem GC- und GC_3-Gehalt ist die Nukleotidzusammensetzung der dritten Codonposition aufgeschlüsselt. Während der GC-Gehalt zwischen den Spezies relativ homogen ist, kann für die dritte Codonposition in Maus und Mensch eine deutliche Abnahme des GC_3-Gehaltes beobachtet werden. *Cpy* zeigt als einziger Vertreter einen höheren GC_3-Gehalt im Vergleich zum Gesamtgehalt an G- und C-Nukleotiden.

„CpG-Inseln“

Die Genome von Maus und Mensch unterscheiden sich neben GC-Gehalt und Isochoren auch in der Anzahl und Ausprägung von so genannten „CpG-Inseln“. Es handelt sich hierbei um nicht-methylierte, >200 bp lange Bereiche im Säugergenom, die einen signifikant höheren GC-Gehalt mit einem überdurchschnittlichen Anteil an CpG-Dinukleotiden im Vergleich zum übrigen Genom aufweisen. Konkret bedeutet dies einen GC-Gehalt >0,5 und einem CpG_{obs}/CpG_{exp}-Verhältnis (beobachtet zu erwartet basierend auf den GC-Gehalt) >0,6 (Tycocinski et al., 1984; Bird et al., 1985; Gardiner-Garden et al., 1987; Cross et al., 1995). Im Falle von epigenetisch kontrollierten bzw. X-chromosomal assoziierten Genen können diese Abschnitte auch methyliert sein (Razin et al., 1994; Strichman-Almashanu, 2002). „CpG-Inseln“ sind häufig mit dem 5’-Bereich von Genen verknüpft (Cross et al., 1995, 2000) und können somit als Anhaltspunkt für das Vorhandensein von Genen dienen. Die Zahl der „CpG-Inseln“ liegt im humanen Genom bei ca. 27.000, bei einer erwarteten Gesamtzahl von geschätzten 30000-40000 Genen (International Human Sequencing Consortium, 2001). Es wird angenommen, dass ca. 56 % aller Gene mit einer „CpG-Insel“ assoziiert sind, darin enthalten nahezu alle Haushaltsgene und 40 % aller gewebsspezifischen Gene (Cross et al., 2000; Antequera et al., 1993). Die Zahl der beobachteten „CpG-Inseln“

liegt in der Maus mit 15500 niedriger als im humanen Genom. Ein einleuchtender Grund hierfür liegt in der Tatsache, dass das murine Genom einen deutlich geringeren Anteil mit extremem GC-Gehalt aufweist und diesbezüglich eine wesentlich homogenere Verteilung zeigt (Waterston et al., 2002; Mouchiroud et al., 1988; Salinas et al., 1986). Weiter ist die präzise Definition einer „CpG-Insel" ein ausschlaggebender Faktor. Zusammengenommen können diese Größen zu unterschiedlichen Ergebnissen bei den verwendeten Algorithmen zur Identifizierung von „CpG-Inseln" führen, und somit die niedrigere Anzahl der beobachteten „CpG-Inseln" im murinen Genom erklären (Waterston et al., 2002). Es ist daher möglich, dass ein Verlust der CpG-Inseln im murinen Genom (Cross et al., 1995; Ansari-Lari et al., 1998) nicht oder nur in geringerem Maße stattgefunden hat. In der vorliegenden Arbeit konnte in der genomischen Region von *Dmxl*1 vor dem ersten Exon von mehreren Programmen eine CpG-Insel identifiziert werden. Die Länge variierte dabei, je nach verwendetem Programm, zwischen 736 bp (EMBOSS-Programmpaket), 743 bp (DRAGON GENOME EXPLORER) und 1012 bp (CpG-FINDER) bei einem GC-Gehalt von deutlich über 80 %. In der homologen humanen Region wurden sechs potenzielle CpG-Inseln identifiziert, nur zwei dieser Abschnitte zeigten einen deutlich erhöhten GC-Gehalt von 75 % bzw. 80 %. Fünf der möglichen CpG-Inseln beim Menschen konnten durch Analyse GC-reichen repetitiven Elementen zugeordnet werden, die in der Maus nicht vorhanden sind. Lediglich die direkt vor dem ersten Exon liegende CpG-Insel mit 75% GC-Gehalt findet ihren homologen Partner in der Maus. Die Länge variierte, je nach verwendetem Programm, zwischen 755 bp (EMBOSS-Programmpaket), 880 bp (DRAGON GENOME EXPLORER) und 1353 bp (CpG-FINDER). Damit zeigt sich die CpG-Insel in der humanen Sequenz nur unwesentlich länger als in der homologen murinen Region. Der GC-Gehalt ist in der humanen Sequenz in diesem Bereich >5 % niedriger als in der Maus. Für *Dmxl*1 kann daher die Annahme, dass CpG-Inseln beim Menschen eine 3% höheren GC-Gehalt aufweisen (Antequera et al., 1993; Matsuo et al., 1993) und 30% länger sind als in der Maus (Cross et al., 1997), nicht bestätigt werden. Die vergleichbare Lage und Länge der CpG-Insel bei Maus und Mensch ist jedoch ein starkes Indiz, dass sich in dieser Region in beiden Spezies der Promotor und das erste Exon befinden.

4.4.2 Repetitive Elemente der genomischen Region von *Dmxl*1

Ein erheblicher Anteil des Genoms von Säugern besteht aus repetitive DNA-Sequenzen, viele davon sind interspergierte repetitive Elemente, die als „Fossilien" transposabler Elemente im Genom verteilt sind (Prak et al, 2000; Smit et al., 1999, 1996). Tandem-repetitive DNA-Elemente sind vor allem auf Zentromer- und Telomer-Bereiche des Genoms beschränkt und wurden auch in den analysierten Sequenzabschnitten nicht gefunden. Das humane Genom besteht zu 46% aus interspergierten repetitiven Sequenzelementen, die als transposable Elemente in den letzten 150-200 Millionen Jahren aktiv waren und in das Genom integrierten. Es wird jedoch angenommen, dass der wirkliche prozentuale Anteil noch wesentlich höher liegt. Es ist augenblicklich nicht möglich, solche Elemente über ein gewisses

Zeitfenster hinaus zu bestimmen, da die Sequenzen durch Erosion zu sehr divergieren. Das murine Genom besteht „nur“ zu 37,5% aus solchen gestrandeten transposablen Elementen, der Grund hierfür liegt in einer höheren Nukleotid-Substitutionsrate, ancestrale repetitive Elemente werden nicht mehr als solche erkannt (Lander et al., 2001; Wu et al., 1985). Da in der untersuchten genomischen Region von Maus und Mensch viele repetitive Elemente identifiziert wurden, sollen diese im Folgenden besprochen werden. Maus und Mensch teilen prinzipiell die gleichen vier Klassen interspergierter Elemente:

- Die in Maus und Mensch vorherrschende Klasse werden als L1 der LINE-Elemente („long interspersed nuclear element“; Singer,1982) bezeichnet. In der Maus-Linie scheinen diese Elemente weiterhin hochaktiv zu sein, während im Menschen eine abnehmende Aktivität festgestellt werden konnte. Das murine Genom enthält ungefähr 3000 potenziell aktive L1-Elemente (Goodier et al, 2001; Lander et al., 2001).
- Die 100-400 bp langen SINE-Elemente („short interspersed nuclear element“; Singer, 1982). Während sich in der humanen Linie nur ein SINE-Element (Alu) aktiv im Genom ausbreiten konnte, wurden im murinen Genom vier aktive Klassen (B1, B2, ID, B4) gefunden. Alle vier Klassen benötigen die Präsenz von L1-Elementen zur Transposition, zeigen aber untereinander keine Sequenzhomologien, so wie es für andere LINE-SINE-Paare oft der Fall ist (Ohshima et al., 1996; Smit, 1996). Im Maus-Genom gibt es insgesamt mehr SINE-Elemente als Alu-Elemente im humanen Genom (1,4 zu 1,1 Millionen) dennoch ist der prozentuale Anteil im menschlichen Genom höher, da die Elemente dort größer sind (10,7% zu 7,6%; Waterston et al., 2002). Im Gegensatz zu den meisten Mitgliedern der SINE-Familie, die in der Regel nur in nahe verwandten Spezies zu finden sind, kommen die MIR-Elemente („mammalien-wide interspersed repeat“) in allen Säugergenomen vor (Smit et al., 1995). Ihre Ausbreitung im Genom muss vor der Abspaltung der Säugetiere stattgefunden haben, da sie häufig in orthologen Positionen verschiedener Spezies zu finden sind (Jurka et al., 1995) und dort möglicherweise eine Funktion ausüben (Murnane et al., 1995).
- Alle LTR-Elemente („long terminal repeat“) sind Abkömmlinge des vertebratenspezifischen Retrovirusstamms der Retrotransposons. Jüngere endogene retrovirale Elemente stammen vermutlich wiederum von Infektionen durch Retroviren ab. Endogene Retroviren werden in drei Klassen (I-III) eingeteilt:
  - Klasse III beinhaltet ca. 80% aller LTR-Elemente, ihre Entstehung liegt vor der Speziation von Maus und Mensch. Sie stellen in der Maus mit 388000 Kopien die größte Gruppe der Klasse III (Waterston et al., 2002). In der Maus sind diese Elemente nach wie vor aktiv (MERVL-, MT-, ORR1- MaRLs), im Primatengenom stellten sie ihre Aktivität vor ca. 50 Millionen Jahren ein (Smit et al., 1993).
  - In der Klasse II existieren zwei aktive Gruppen in der Maus, die „intracisternal-A particles“ (IAP) und „early transposons“ (ETn), ihre Anzahl ist in der Maus um den Faktor 10 höher als im Menschen. 10% aller spontan entstehenden Maus-Mutanten haben ihre Ursache in einer Insertion von IAP oder ETn.

- Die Klasse I-ERVs werden im humannen Genom häufiger gefunden als in der Maus, obwohl sie dort, im Gegensatz zur Maus, nicht mehr aktiv sind. Zu dieser Klasse gehört z.B. das Murine Leukämie Virus (Waterston et al., 2002).

- DNA-Transposons transponieren durch Exzision und Reintegration ohne sich über ein RNA-Intermediat zu verbreiten. In der Primaten-Linie sind 14 Vertreter dieser Gruppe bekannt, in der Maus nur vier. Die Mariner-ähnlichen Elemente sind zahlenmäßig in beiden Genomen die häufigsten Vertreter dieser Gruppe. Die Mariner-Elemente scheinen unabhängig voneinander die Nager- und Primatengenome infiltriert zu haben (Waterston et al., 2002).

Die vorgestellten Klassen der interspergierten repetitiven Elemente sind nicht zufällig über das Genom verteilt, vielmehr zeigen diese Elemente eine deutliche Prädisposition für bestimmte genomische Regionen. Während LINE-Elemente vor allem in AT-reichen Regionen zu finden sind, bevorzugen SINEs GC-reiche genomische Abschnitte (Jurka et al., 2004; Waterston et al., 2002).
Die Analyse der genomischen Region von *Dmxl*1 und *DMXL*1 zeigte einen Anteil an repetitiven Elementen von 34,49% in der Maus und 51,53% im Menschen (Tab. 3.8). Dieser passt damit nur bedingt zu dem ermittelten statistischen Durchschnitt für das Maus- und Menschgenom. Bei genauerer Betrachtung kann für ein L1-Isochor mit 36 % in der Maus ein wesentlich höherer Wert von > 40% erwartet werden, damit liegt dieser Abschnitt > 6% unter dem prognostizierten Anteil repetitiver Elemente. Für das humane Genom stimmen ermittelte und erwartete Werte eher überein, der Mensch liegt hier ca. 4% über dem erwarteten Wert, was allerdings konträr zu der Entwicklung in der Maus zu sein scheint. Eine Erklärung hierfür könnte in der zweifachen (Durchschnitt seit der Trennung von Maus und Mensch) bzw. fünffachen (aktuelle Substitutionsrate in der Maus) höheren Mutationsrate des Mausgenoms, auch bedingt durch die längere Generationszeit des Menschen (20fach) liegen. Als Folge dieser unterschiedlichen Mutationsraten erkennt das Programm REPEATMASKER oberhalb von 37 % Divergenz einen großen Teil der repetitiven Elemente nicht mehr (Waterston et al., 2002). Die prozentuale Verteilung der verschiedenen Klassen entspricht in Maus und Mensch ebenfalls nicht den Annahmen (Abb. 4.5). Der Anteil der SINE-Elemente ist in beiden Genomen mehr als doppelt so hoch wie das erwartete statistische Mittel. Der Zahl der L1-Elemente ist vor allem in der Maus sehr viel geringer als es für eine AT-reiche Sequenz erwartet würde. Dafür ist aber der prozentuale Anteil der DNA-Transposons und Repetitionseinheiten einfacher Komplexität merklich erhöht. In der humanen Sequenz ist der Anteil der LINE2-Elemente höher und die Zahl der LTR-Elemente sehr viel niedriger als es für eine Sequenz in einer GC-armen Region erwartet würde. Der erhöhte Anteil der SINE-Elemente in beiden Spezies kann durch obige Ausführungen einer GC-unabhängigen Insertion erklärt werden, die niedrige Zahl der LINE-Elemente scheint also ein besonderes Merkmal dieses DNA-Abschnitts zu sein.

Die Zahl der SINE-Elemente ist in der untersuchten humanen Sequenz mit 141 zu 112 höher als in der Maus (Abb 3.20, Abb. 3.21) Dies ist äußerst ungewöhnlich, da normalerweise im Mausgenom doppelt so viele SINEs erwartet werden wie Alu-Elemente im menschlichen Genom. Da die SINE-Elemente in der Maus nur halb so

lang sind wie die Alu-Elemente im Menschen, ist ihr prozentualer Anteil an einer untersuchten Sequenz entsprechend kleiner. In der genomischen Region von *DMXL*1 ist der beobachtete Anteil der Alu-Elemente an der Gesamtsequenz mehr als doppelt so hoch (35284 bp zu 15452 bp) als der Anteil der SINE-Elemente in der orthologen Region der Maus. Dies entspricht somit wiederum den statistischen Erwartungen. Auffällig sind hierbei zahlreiche gut erhaltene SINE/Alu-Elemente in Maus und Mensch mit Substitutionsraten unter 9% bzw. unter 7%. So konnte in der murinen Sequenz 15 SINE-Elemente mit einer Substitutionsrate unter 9% gefunden werden, in Intron 37 wurde ein B1_MM-Element identifiziert, dass bei voller Länge eine Substitutionsrate von nur 2% aufweist, zwei C→T Transitionen sowie eine A→C Transversion und zwei Integrationen im Poly A-Bereich. Im Menschen konnten neun Alu-Elemente mit einer Substitutionsrate von unter 7% gefunden werden. Bei diesen Elementen in Maus und Mensch handelt es sich um junge SINE/Alu-Repeats, die ganz offensichtlich unabhängig voneinander bevorzugt in diese Regionen integriert sind. In der humanen Region von *DMXL*1 wurden 11 junge AluY-Elemente gefunden, ihre Zahl ist damit fast doppelt so hoch wie für eine Region mit einem niedrigen GC-Gehalt erwartet würde (Jurka et al., 2004; Medstrand et al., 2002). Eine Besonderheit dieser jungen Alu/SINE-Elemente stellt ihr Methylierungsstatus dar, die meisten Mitglieder dieser Gruppe sind komplett unmethyliert und können möglicherweise *in cis* den Methylierungsstatus der umgebenden Sequenzabschnitte beeinflussen (Hasse et al., 1994). In der humanen Sequenz liegen vier AluY-Elemente innerhalb der analysierten 10 kb vor dem Translationsstartpunkt, in der Maus konnte eine vergleichbare Organisation nicht identifiziert werden. Dennoch ist vorstellbar, dass die beobachteten SINE/Alu-Repeats die grundlegende transkriptionelle Aktivität von *Dmxl*1 und *DMXL*1 beeinflussen. In den untersuchten Sequenzabschnitten von Maus und Mensch spielen die identifizierten MIR-Elemente nur eine untergeordnete Rolle. In der murinen Sequenz wurden drei dieser Elemente gefunden, in der humanen Sequenz 18. In Maus und Mensch finden sich jeweils zwei dieser Elemente vor der Promotorregion von DmX, allerdings nicht in einer konservierten Position. Das dritte Element liegt in der Maus in Intron 19, im Menschen befinden sich dort drei Elemente, der Grad der Konservierung ist bei allen Elementen schwach. In Maus und Mensch finden sich in diesem Intron in orthologer Position eine große Anzahl LINE-Elemente. Die Komposition des Introns scheint zwischen den beiden Spezies konserviert zu sein. Der prozentuale Anteil der LINE-Elemente ist in der untersuchten humanen Region mit 20,27% fast doppelt so hoch wie in der orthologen murinen Sequenz mit 10,94%, obwohl die Anzahl von 49 zu 35 dieses Verhältnis nicht widerspiegelt. Dies beruht darauf, dass die LINE-Elemente der humanen Sequenz weniger trunkiert zu sein scheinen als in der murinen Sequenz. Auch hier spielt möglicherweise die höhere Mutationsrate in der Maus eine Rolle. Es kann angenommen werden, dass nicht alle LINE-Elemente identifiziert wurden. Eine orthologe Zuordnung der LINE-Elemente zwischen beiden Spezies ist aufgrund der hohen Divergenz schwierig, vor allem da die Intronlänge zwischen Maus und Mensch stark differiert. Prinzipiell lassen sich jedoch einige LINE-Elemente (L1 und L2) orthologen Lagen (Introns 14, 19, 23, 25, 30, 31, 33, 37, 41; Abb. 3.20, 3.21) in der humanen Sequenz zuordnen. Die Gruppe der LTR-Elemente ist in der analysierten Region gegenüber der humanen Sequenz

überrepräsentiert, sie nehmen 8% der untersuchten Region ein. Für die MaLR-Elemente ist hier eine stärkere Vertretung in der Maus bekannt (Onyango et al., 2000; Ansari-Lari et al, 1998).

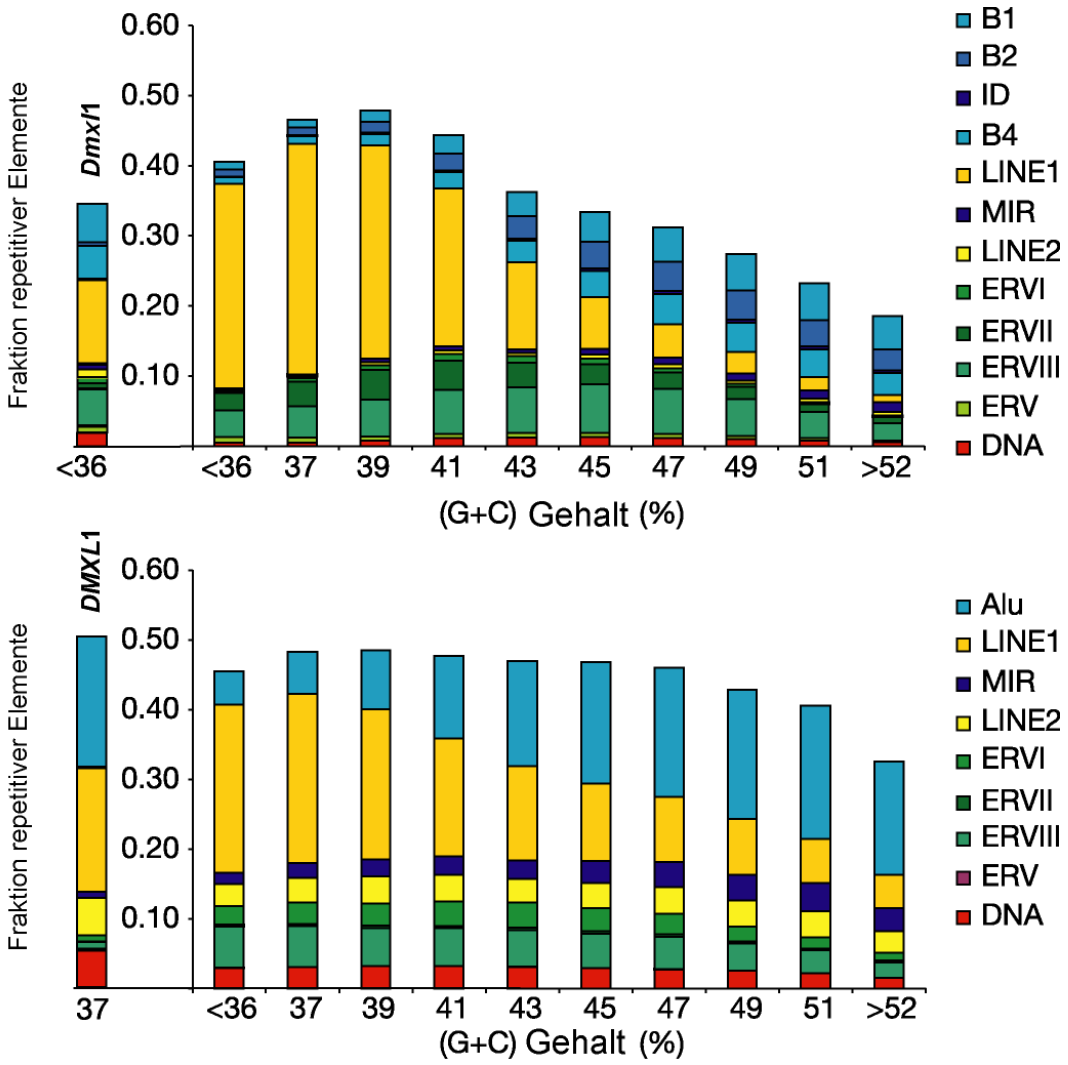


Abb 4.6: Durchschnittliche prozentuale Verteilung repetitiver Elemente in Maus und Mensch.

Die Verteilung der unterschiedlichen Klassen der repetitiven DNA-Elemente ist abhängig vom GC-Gehalt der Sequenz. Maus und Mensch unterscheiden sich ebenfalls in der Zusammensetzung bezogen auf den jeweiligen GC-Gehalt. In der oberen Hälfte der Abbildung ist der prozentuale Anteil von repetitiven Elementen verschiedener Klassen an der Gesamtsequenz in Abhängigkeit vom GC-Gehalt des Mausgenoms gezeigt. In der unteren Hälfte der Abbildung ist dies für das humane Genom dargestellt. Die Verteilung von repetitiven Sequenzen für die genomische Region von *Dmxl*1 und *DMXL*1 ist als separater Balken gezeigt. Die Abbildung ist eine Modifikation der Abbildung 11 aus der Publikation von Waterston et al. (2002).

Der Anteil an DNA-Transposons beträgt in der murinen Sequenz 1,09% und in der humanen Sequenz 5,12%, damit ist auch diese Klasse stärker vertreten, als es für eine

AT-reiche Region erwartet werden kann (Abb. 4.5). In Intron 29 konnte bei beiden untersuchten Spezies in orthologer Position ein *Tigger6a*-Element identifiziert werden. In der Maus ist dieses Transposon an drei Stellen unterbrochen und in vier Elemente zerteilt, verursacht wurde dies durch die Integration mehrerer SINE und LTR Elemente, die in sich ebenfalls durch die Integration von SINE- bzw. LTR-Elementen unterbrochen wurden. In der humanen Sequenz ist das Repeat fast vollständig erhalten und nicht durch andere repetitive Elemente unterbrochen worden. Bei den *Tigger*-Elementen handelt es sich um fossile DNA-Transposons, die dem *pogo*-Transposon aus *Drosophila melanogaster* ähnlich sind, und deutliche Homologien zu dem Zentromer-Protein CENP-B der Säugetiere zeigen (Kipling et al., 1997; Smit et al., 1996). Sowohl *Pogo*- als auch *Tigger*-Elemente enthalten Motive für PCNA (proliferating cell nuclear antigen; Warbrick et al., 1998) In der genomischen Region von *DmX* bei *Drosophila melanogaster* findet sich allerdings kein *pogo*-Element (Daten nicht gezeigt). Möglicherweise stellen die Tigger-Elemente Bindungsstellen für CENP-B bereit und wirken so transaktivierend (Kapitel 4.4.5).

4.4.3 Genvorhersage durch Kombinationsstrategien

Eine der schwierigsten Aufgaben der Bioinformatik besteht in der korrekten *de novo* Vorhersage und Identifikation von Genen. In den meisten Fällen sind die aktuell verfügbaren Programme auf diesem Gebiet noch nicht in der Lage, ein Gen und seine Struktur zufrieden stellend und sicher vorauszusagen (Mujica, 2004; Frazer et al., 2003; Amid, 2002; Bahr, 1999; Bikár, 1997). Von allen in dieser Arbeit verwendeten Programmen (DNASweep, GENSCAN, FGENES, HmmGENE, Geneid, XPOUND, Grail,) konnte nur DNASweep überzeugen. Zwar zeigten die neuesten Versionen der anderen Genvorhersage-Programme eine deutlich verbesserte Zuverlässigkeit, dennoch konnte keines dieser Programme, bis auf DNASweep, *Dmxl*1 und *DMXL*1 korrekt identifizieren. Da diese Probleme im Vorfeld bekannt waren, wurde für die Analyse der genomischen Organisation von *Dmxl*1/*DMXL*1 in Maus und Mensch eine Kombinationsstrategie aus DOTPLOT Analysen, Genvorhersageprogrammen, Interspeziesvergleich und die Suche per Auge gewählt. Zwar ist diese Vorgehensweise extrem aufwendig, da für jede identifizierte Sequenz vielfache unterschiedliche Analysen notwendig sind, dennoch ist dies aktuell der einzige Weg, eine nahezu 100%ige Vorhersagegenauigkeit zu erreichen. Mit Hilfe der experimentell ermittelten cDNA-Sequnz von *Dmxl*1 und der von Kraemer zur Verfügung gestellten cDNA-Sequenz des humanen *DMXL*1 wurde die vorher *in silico* festgelegte Genstruktur überprüft. Hierzu wurde das Programm DNASweep verwendet, das neben einem lokalen BLAST-Alignment auch nach möglichen Spleißstellen sucht. Das Ergebnis dieser Untersuchung deckt sich zu 100% mit den vorher festgelegten Exon/Intron-Grenzen. Erst nach dieser Überprüfung konnte gesagt werden, das DNASweep alle Exons richtig erkannt hatte und keine falsch positiven Exons vorausgesagt hat. Dennoch bleibt anzumerken, dass DNASweep vor allem Daten aus den öffentlichen Datenbanken für eine Vorhersage verwendet und insofern zum einen von den vielen EST-Daten für *DMXL*1 und zum anderen von der großen Homologie zwischen *Dmxl*1

und *DMXL*1 profitieren konnte. Es ist unsicher, ob DNASweep diese Genauigkeit bei weniger gut konservierten Genen erreichen kann. In dieser Arbeit konnte durch die Kombination verschiedener Strategien eine abgesicherte Strukturvorhersage sowohl für das murine *Dmxl*1- als auch für das humane *DMXL*1-Gen getroffen werden.

4.4.4 Die Genstruktur von *Dmxl*1 und *DMXL*1

Für das murine Genom werden aktuell 23149 proteinkodierende Gene (inkl. 1221 Pseudogene) vorhergesagt, diese bestehen aus 248388 Exons und werden durch insgesamt 40288 Transkripte repräsentiert (Ensembl, build 48.37a). Für das humane Genom werden zurzeit 23622 proteinkodierende Gene (inkl. 2081 Pseudogene) prognostiziert, die von 275708 Exons kodiert und durch 48400 Transkripte vertreten werden (Ensembl build 48.36j). Aktuellen Schätzungen zufolge sind für das humane und murine Genom deutlich weniger als 30000 proteinkodierender Gene zu erwarten (Birney et al., 2007; Clamp 2007; Southan, 2004; Waterston et al., 2002). Durchschnittlich können 8,8 (Mensch) bzw. 8,4 (Maus) Exons pro proteinkodierendem Transkript beobachtet werden. Aktuelle Statistiken über die Komposition von Genen liegen derzeit nur für das humane, nicht jedoch für das murine Genom vor. Die meisten menschlichen Gene bestehen aus acht bis zehn Exons, ihre Länge liegt im Mittel bei 170 bp, in 80-85% der Fälle sind diese kleiner als 200 bp (Sakharkar et al., 2004; 2002). Die initialen und terminalen Exons und Introns sind meist größer als interne Exons bzw. Introns (Chen et al., 2001). Der Sequenzumfang der Introns divergiert wesentlich stärker, ihr Wert schwankt zwischen 4741,54 und 23527,35 bp, das Mittel liegt bei 5419 bp. Es ist interessant, dass sehr große und sehr kleine Introns hierbei nur einen geringen Anteil haben. Weniger als 5,24% aller Introns sind größer als 200.000 bp und weniger als 10% größer als 11000 bp. Nur 0,01% der Introns sind kleiner als 20 bp. Die scheint auf eine Limitierung beim Spleißen sehr großer und sehr kleiner Introns hinzuweisen. Ein durchschnittliches humanes Gen hat 6-9, im Schnitt 7,8 Introns. Die Anzahl schwankt aber zwischen Null (Gene mit nur einem Exon, beobachtet bei 3362 Genen) und 147 Introns bei *NEB* (Nebulin) auf Chromosom 2 (Sakharkar et al., 2004; 2002). Es scheint einen grundlegenden Zusammenhang zwischen der Länge von Introns und der Expressionsrate der zugehörigen Gene zu geben. Je höher die Expressionsrate, desto kürzer sind die Intronsequenzen der zugehörigen Gene, im Schnitt umfassen sie nur noch 342,5 bp, also 6,3% der durchschnittlichen Intronlänge. Eine Veränderung der Exonlängen bei hochexprimierten Genen wurde nicht beobachtet, allerdings reduziert sich die Gesamtlänge der Exons dieser Gene um bis zu bis zu 50% (Castillo-Davis et al., 2002). Die Maus unterscheidet sich vom Menschen vor allem in der Länge der Introns, im Durchschnitt sind diese um 50% kürzer. Die Exons wiederum sind nur geringfügig kleiner als beim Menschen (Deutsch et al., 1999). Auch die Phasen der Introns sind nicht gleichmäßig verteilt, in der Maus wurde eine Aufteilung von 44% (Phase 0): 38% (Phase 1): 18% (Phase 2) beobachtet. Die Überrepräsentation der Phase 0-Introns gegenüber den Phase 2-Introns kann durch konservierte Spleißstellen in den Exons erklärt werden, bei Phase 0-Introns bleibt die Codonstruktur intakt, hier ist der höchste Grad der Konservierung zu beobachten. Die Phase der Introns scheint

abhängig von der Konservierung der umgebenden Exonsequenzen zu sein, Phase 2 umgebende Exons zeigen die höchste Sequenzvariabilität, Phase 0 umgebende Exons den höchsten Konservierungsgrad (Long et al., 1999). Die oben aufgeführten Zahlen beziehen sich alle auf proteinkodierende Gene. Jüngste Untersuchungen von 1% des humanen Genoms durch ENCODE und anderen assoziierten und unabhängigen Untersuchungen (Clamp et al., 2007; Margulies et al., 2007; ENCODE Project Consortiom 2007, Wong et al., 2001) zeigen, dass die bisherige Vorstellung von Genen und kodierenden Sequenzabschnitten möglicherweise nicht mehr gültig sind (Gerstein et al., 2007; Gingeras, 2007). Es konnte gezeigt werden, dass das menschliche Genom umfassend transkribiert wird. Es gibt zudem eine gewaltige Anzahl funktionaler, aber nicht-proteinkodierender Transkripte und es existieren sehr viele funktionale DNA-Elemente die nicht transkribiert werden. Vor diesem Hintergrund muss vor allem die Reduktion des Genoms auf die im Verhältnis wenigen kodierenden Transkripte revidiert werden. Die Anzahl der Gene wird vermutlich dramatisch ansteigen, da die Zahl funktionaler aber nicht-proteinkodirender Transkripte bisher völlig unterschätzt wurde.

Die Exon/Intron-Struktur von *Dmxl*1/*DMXL*1 wurde durch den Vergleich der ermittelten cDNA-Sequenz (Maus) bzw. unter Verwendung der durch Kraemer et al. (2000) ermittelten humanen cDNA mit den genomischen Sequenzen der öffentlichen Datenbanken und den Sequenzdaten der „shot-gun"-Sequenzierung ermittelt. Um die oben genannten statistischen Werte in einen sinnvollen Zusammenhang mit den hier besprochenen Genen *Dmxl*1 und *DMXL*1 zu bringen, wurde zusätzlich die Datenbank H-Inv DB (http://www.h-invitational.jp) für Gegenüberstellungen verwendet (Yamasaki et al., 2008). Diese Datenbank besteht aus ca. 187156 „full-length" cDNA-Klonen, die 36073 humane Gencluster repräsentieren und die nach einer Vielzahl von verschiedenen Kriterien durchsucht werden kann (Imanishi et al., 2004). Für einen Vergleich wurden aus dieser Datenbank alle Gene isoliert, deren Aminosäuresequenz länger als 3000 aa ist. Insgesamt konnten so 89 Gene identifiziert werden. Dies entspricht 0,047% aller dort erfassten Transkripte. 69 Transkripte sind länger als DMXL1.
Das murine und humane *Dmxl*1/*DMXL*1 bestehen je aus 43 Exons und 42 Introns, die für eine cDNA von 10991 bp bzw. 12210 bp (Maus) und 11082 bp (Mensch) kodieren (Abb. 4.6). Das größte Exon (Exon 18) hat in Maus und Mensch eine Länge von 1682 bp, das kleinste Exon (Exon 41) umfasst in beiden Spezies nur 52 bp. Die durchschnittlichen Längen der Exons von DmX betragen in der Maus 210,36 bp und im Menschen 211,44 bp und liegen damit 23% über dem Median humaner Exons. Zwischen beiden Spezies sind die Exonpositionen und ihre Längen hoch konserviert und unterscheiden sich in der Länge nur in den Exons 5 (+ 2 Codons in der humanen Sequenz), 24 (+ 4 Codons in der humanen Sequenz) und 29 (+ 8 Codons in der humanen Sequenz) voneinander (Tab. 3.7). Die Intronpositionen zwischen Maus und Mensch sind ebenfalls hoch konserviert, die Introns unterscheiden sich jedoch in Ihrer Länge zum Teil erheblich. Die durchschnittliche Intronlänge beträgt in der murinen Sequenz 2868,52 bp und in der humanen Sequenz 3874,14 bp, damit sind die Introns von *Dmxl*1 im Schnitt 26% kleiner als im Menschen. Erstaunlicher Weise entsprechen

diese 26% zusammen fast genau jenen 48986 bp, die die genomische Region von *DMXL*1 reicher an repetitiven Sequenzen ist (Kapitel 3.2.5; 4.4.2). Die Differenzen der Intronlängen zwischen Maus und Mensch basieren somit also fast ausschließlich auf der unterschiedlichen Zusammensetzung und Länge von repetitiven Elementen in den Introns.
In der Maus ist Intron 36 mit 11585 bp das größte und Intron 14 mit 79 bp das kleinste. Für *Dmxl*1 kann somit nicht bestätigt werden, dass das angenommene initiale bzw. terminale Exon/Intron größer ist als die internen Exons bzw. Introns. Die Intronsequenzen von *DMXL*1 sind deutlich kürzer als die durchschnittliche Intronlänge humaner Gene, das größte Intron von *DMXL*1 umfasst 26321 bp (Intron 1), das kleinste ist Intron 12 mit 88 bp. Im Menschen ist damit das initiale Intron in der Tat das größte, allerdings trifft dies nicht auf das erste Exon und das terminale Exon/Intron zu.

Gen	Anzahl der Exons	Länge der cDNA	Länge im Genom
KIAA0570	80	11269 bp	283300 bp
***DMXL*1**	**43**	**11084 bp**	**177699 bp**
KIAA0399	54	10861 bp	119490 bp
KIAA1773	21	10759 bp	34500 bp
KIAA0311	14	10327 bp	303700 bp
KIAA1302	49	9944 bp	87400 bp
Protein Prenyltransferase	5	9935 bp	50500 bp
KIAA1506	33	9929 bp	90000 bp
KIAA1109	50	9765 bp	113000 bp
KIAA0283	30	9525 bp	813160 bp
Similar to ALR-like protein	23	9332 bp	43000 bp
KIAA1289	45	9078 bp	14900 bp
KIAA1776	63	8966 bp	82100 bp
KIAA0279	34	8924 bp	24000 bp
KIAA0342	13	8859 bp	32000 bp
*TACC*2	22	8821 bp	259900 bp
KIAA1675	17	8671 bp	369400 bp
KIAA1302	8	8624 bp	48800 bp
KIAA1920	4	8556 bp	13390 bp
KIAA0219	59	8515 bp	67500 bp

Tab. 4.3: Genomische Organisation von Genen mit einer cDNA >8500 bp

Die gezeigten Gene sind eine Auswahl und ihre genomische Organisation wurden durch Datenbanksuchen in H-Inv DB ermittelt. Hierbei konnten insgesamt 89 Gene identifiziert werden, deren cDNA größer als 10000 bp ist. Bei 69 Genen ist die cDNA größer als bei *DMXL*1. Die meisten der aufgeführten Gene sind, gemessen an Ihrer Größe und der Anzahl ihrer Exons, als kompakt zu bezeichnen. Bis auf vier Ausnahmen (grau hinterlegt) erreichen sie nicht die genomische Ausdehnung, die statistisch erwartet würde. Der Vergleich zeigt zudem, dass das humane *DMXL*1 zu den größten Genen im menschlichen Genom gehört, auch wenn diese Daten nicht vollständig sind.

*DMXL*1 ist also ein eher kompakt organisiertes Gen. Allerdings zeigt eine Gegenüberstellung der genomischen Organisation mit Genen, deren cDNA mehr als 10000 bp umfasst, ein noch differenzierteres Bild (Tab 4.3). Die meisten Gene mit einer vergleichbaren Länge, deren genomische Organisation verfügbar und verifiziert in den Datenbanken vorliegt, sind teilweise noch dichter organisiert als *DMXL*1. Die Mehrzahl dieser großen Gene nimmt genomisch weniger Platz ein, als statistisch erwartet werden kann. Basierend auf diesen Datenbankeinträgen muss *DMXL*1 zu den größten Genen im menschlichen Genom gerechnet werden.
Die Phasenverteilung der Introns von *Dmxl*1 in der Maus entspricht ebenfalls nicht den ermittelten statistischen Werten (Long et al., 1999). Für *Dmxl*1 wurde eine Verteilung von 45,23% (Phase 0): 30,95% (Phase 1): 23,81% (Phase 2) beobachtet. Es zeigt sich eine deutliche Verschiebung von Phase 1 zu Phase 2 Introns. Da bei *Dmxl*1 vor allem an den Exongrenzen die höchste Konservierung beobachtet werden kann (Daten nicht gezeigt), muss davon ausgegangen werden, dass die Beobachtungen von Long et al. (1999) gegebenenfalls nicht für hochkonservierte Gene gelten.

Ein Vergleich der Genstruktur von *Dmxl*1 mit den Homologen anderer Spezies (Abb. 4.6) zeigt, dass auch die genomische Organisation zum Teil konserviert ist. *DmX* (Drosophila) und *CpY* (Chironomus) bestehen jeweils aus 15 Exons und 14 Introns, *CpY* umfasst eine Länge von ca. 14,8 kb und *DmX* ist ca. 14,9 kb lang. Das bei *CpY* differenziell gespleißte kleine Exon 4 ist bei *DmX* Bestandteil des größten Exon 4. Das zweite differenziell gespeißte Exon 8 von *CpY* fehlt bei *DmX* völlig (Weil, 2000). Das homologe Gen aus *C. elegans* (*F54E4.1*) besteht nach neuesten Analysen aus 31 Exons und 29 Introns. Bei dem in der Arbeit von Weil (2000) beschriebenen Exon 1 handelt es sich nicht um das initiale Exon, sondern um Exon 2. Das erste Exon von *F54E4.1* findet auch nach Datenbanksuchen in den umliegenden genomischen Regionen der anderen Spezies keinen homologen Sequenzabschnitt. Demnach ist das zweite Intron in homologer Position zu dem ersten Intron in Drosphila und Chironomus. Der zum differenziell gespleißten Exon 4 aus *C. piger* homologe Abschnitt ist wie bei *D. melanogaster* Bestandteil eines größeren Exon 17. Eine homologe Region zu dem differenziell gespleißten Exon 8 von *C. piger* fehlt, wie auch bei Drosophila, in *C. elegans*. Zwischen diesen drei Spezies scheinen neben dem vorher erwähnten Intron 1 bzw. 2 noch zwei weitere Intronposition konserviert zu sein: Intron 9 aus *DmX* entspricht dem Intron 28 in *F54E4.1* und den Introns 7 und 8 aus *CpY*, zwischen denen sich das in den anderen Arten nicht vorhandenen Exon 8 befindet. Die Position von Intron 10 aus *DmX/CpY* entspricht dem Intron 29 des *F54E4.1* Gens. Zwischen den Spezies Maus und Mensch sind alle Intron- und Exon-Positionen konserviert, hier bestehen keine Unterschiede. Ein Vergleich von *Dmxl*1/*DMXL*1 mit dem evolutionär weit entfernten Gen *F54E4.1* aus *C. elegans* zeigt die erstaunliche Zahl von zehn konservierten Intron-Positionen (Tab. 4.4). Zwischen allen fünf Spezies sind die Positionen des Intron 1 (*DmX*, *CpY*, *Dmxl*1, *DMXL*1) zu Intron 2 (*F54E4.1*) sowie Intron 9 (*DmX*), Intron 7+8 (*CpY*), Intron 35 (*Dmxl*1, *DMXL*1) zu Intron 27 (*F54E4.1*) homolog.

Intron-Position in *F54E4.1*	**Homologe Intron-Position in *Dmxl*1/*DMXL*1**
Intron 2	Intron 1 (*DmX*, *CpY*,*Dmxl*1/*DMXL*1)
Intron 9	Intron 14 (*Dmxl*1/*DMXL*1)
Intron 10	Intron 17 (*Dmxl*1/*DMXL*1)
Intron 18	Intron 18 (*Dmxl*1/*DMXL*1)
Intron 19	Intron 20 (*Dmxl*1/*DMXL*1)
Intron 22	Intron 24 (*Dmxl*1/*DMXL*1)
Intron 23	Intron 25 (*Dmxl*1/*DMXL*1)
Intron 24	Intron 28 (*Dmxl*1/*DMXL*1)
Intron 25	Intron 31 (*Dmxl*1/*DMXL*1)
Intron 27	Intron 9 (*DmX*), Intron 7+8 (*CpY*), Intron 35 (*Dmxl*1, *DMXL*1)
Intron 29	Intron 39 (*Dmxl*1/*DMXL*1)

Tab. 4.4: Konservierte Intron-Positionen der DmX-Homologen zwischen Maus, Mensch, Drosophila, Chironomus und Caenorhabditis

Für die Berechnung homologer Positionen wurde das letzte Codon des vorhergenden Exons der DNA-Sequenz aus C. elegans in die entsprechende Positon auf der abgeleiteten Aminosäuresequnz umgerechnet. Zu dieser Position wurde der homologe Abschnitt in Maus und Mensch durch ein ClustalW-Alignment zwischen Maus, Mensch und C. elegans bestimmt. Anschließend wurde diese Position auf die entsprechende DNA-Sequenz in Maus und Mensch umgerechnet und mit den Intronpositionen aus Tab. 3.6 bzw. 3.7 verglichen (Daten nicht gezeigt).

DmX
Drosophila melanogaster
CpY
Chironomus piger
F54E4.1
Caenorhabditis elegans
Dmxl1
Mus musculus
DMXL1
Homo sapiens
AUG
UAA
pA
pA1
pA2

Abb. 4.7: Genomische Organisation von DmX verschiedener Spezies im Vergleich .

Dargestellt ist ein Vergleich der genomischen Organisation von *DmX*, *CpY*, *F54E4.1*, *Dmxl*1 und *DMXL*1. Die Farbkodierung basiert auf einer Darstellung von Kraemer (persönliche Mitteilung). Die Farben zeigen homologe Sequenzabschnitte sowie die Organisation von Exon- und Intron-Strukturen zwischen *CpY* und *DmX*. Diese Kodierung wurde auf die homologen Gene von Maus und Mensch übertragen. Für jedes der dargestellten Gene sind jeweils die Exon- bzw. Intron-Strukturen in sich proportional, sie sind jedoch aus Platzgründen nicht mit den jeweils anderen Spezies vergleichbar. Konservierte Intron-Positionen zwischen *DmX*, *CpY*, *F54E4.1*, *Dmxl*1 und *DMXL*1 sind in rot dargestellt, zwischen *Dmxl*1/*DMXL*1 und *F54E4.1* sind diese in blau gezeigt. Konservierte Introns zwischen *DmX*, *CpY* und *F54E4.1* sind in grün dargestellt. Bekannte differenziell gespleißte Exons sind durch Linien oberhalb der farbcodierten Region verbunden.

4.4.5 Promotorstrukturen und Enhancer von *Dmxl*1

Der Promotor:

Die Analyse der transkriptionellen Kontrolle, das heißt, die Koordination der Transkription von Genen in Abhängigkeit von Ort und Zeit, ist einer der wichtigsten Schlüssel für das Verständnis der Funktion und des Zusammenwirkens von Genen in der Zelle. Die *in silico* Identifikation von Promotorstrukturen gehört derzeit, neben der Genvorhersage, zu den schwierigsten Aufgaben der Bioinformatik (Das et al., 2007; Zhu et al., 2007; Guha Thaakurta et al., 2006; Gustincich et al., 2006; Davuluri et al., 2001; Hannenhalli et al., 2001). Dies liegt daran, dass es sich bei Promotoren nicht um klar definierbare Einheiten handelt. Die Region um den Transskriptionsstartpunkt, die alle notwendigen Strukturen für die basale Initiation der Transkription enthält, wird als proximaler Promotor bezeichnet (Werner, 1999). Diese Region überlappt oder beinhaltet häufig das erste Exon eines Gens (Werner et al., 2003). Unglücklicherweise zeigen diese Abschnitte in Säugern nur eine geringe Sequenzübereinstimmung beim Vergleich verschiedener Gene, auch wenn sie in Ihrer Funktion korrelieren. Eine BLAST basierte Suche kann daher auf diese Abschnitte nicht ohne weiteres angewendet werden (Werner, 2002; Scherf et al., 2000). Innerhalb einer Promotorregion sind nur wenige kurze Abschnitte wirklich funktional, es handelt sich hierbei um Bereiche, an denen Transkriptionsfaktoren binden können. Diese zeigen meist eine Länge von ca. 12 bp und sind in ihrer Sequenzabfolge höchst variabel. Die Abschnitte sind wiederum durch so genannte „Spacer“ unterschiedlicher Länge und variabler Sequenz voneinander getrennt. Der Aufbau eines Promotors erinnert somit stark an den Aufbau eukaryotischer Gene mit einer Exon/Intron-Struktur. Die Variabilität des gesamten Promotorbereichs führt bei den computergestützten Vorhersagemodellen zu einer großen Anzahl falsch positiver Vorraussagen, so wird von den meisten Programmen ein Promotor pro 1000 bp identifiziert (Waterston et al., 2002; Werner et al., 2002; Fickett et al., 1997). Auch wenn die Qualität der verfügbaren Programme in den letzten Jahren deutlich gestiegen ist (Bajic et al., 2002; Down et al., 2002; Davuluri et al., 2001; Scherf et al., 2000) und die Vorhersage eines Promotors bei einigen Programmen auf 1 in 10000-50000 gefallen ist, muss noch immer mit einer großen Zahl falsch positiver Vorhersagen gerechnet werden (Werner et al., 2003). Um

eine möglichst genaue Aussage über die Promotorstruktur von *Dmxl*1 treffen zu können, wurde in der vorliegenden Arbeit, wie schon für die Genvorhersage, eine Kombination verschiedener Programme und Vorhersagemodelle gewählt. Zunächst erfolgte die Eingrenzung eines möglichen Promotorbereichs durch den Interspeziesvergleich von Maus und Mensch. Da durch die Amplifikation der cDNA in Maus und Mensch ein putatives erstes Exon vorausgesagt werden konnte, wurden diese Sequenzabschnitte als Ausgangspunkt für eine stromaufwärtsgerichtete Analyse von weiteren 10000 bp gewählt. Dieser ausgewählte Sequenzbereich wurde in eine PIP-Analyse (Abb. 3.20, 3.21) einbezogen und separat mit dem Programm DOTPATH untersucht. Das Ergebnis dieses Sequenzvergleichs zeigt, dass die Homologie in einem Bereich bis 1,5 kb vor dem putativen ersten Exon in Maus und Mensch erwartungsgemäß zwar leicht abnimmt (Waterston et al., 2002), aber kurze Sequenzabschnitte enthält, die eine signifikante Homologie aufweisen. Weiter stromaufwärts nimmt die Ähnlichkeit zwischen beiden Sequenzen dramatisch ab, lediglich einige kurze Bereiche zeigten wieder eine Übereinstimmung, die jedoch mit Hilfe einer REPEATMASKER-Analyse als repetitive Sequenzen in beiden Spezies identifiziert wurden. Im Folgenden konnte in dem kurzen 1,5 kb Sequenzabschnitt von verschiedenen Programmen eine „CpG-Insel“ in beiden Spezies identifiziert werden (Kapitel 3.5.6; 4.4.1). Diese nicht-methylierten, 200 bp bis 2 kb langen Bereiche im Säugergenom mit hohem GC-Gehalt sind in der Regel mit den 5’-Bereichen von Genen assoziiert (Cross et.al., 1995) und eignen sich daher als Anhaltspunkt für Promotorbereiche (Gardiner-Garden et al., 1987). Die durch diese Analysen identifizierte potenzielle Promotorregion wurde dann intensiver mit den Programmen DBA und PromoterWise untersucht (Birney et al., 2004). Es handelt sich hierbei nicht um klassische Promotorvorhersageprogramme, vielmehr identifizieren sie sehr kurze colineare Sequenzblöcke zwischen zwei Sequenzen. Auch wenn Promotorregionen nur eine geringe Sequenzübereinstimmung zwischen Maus und Mensch zeigen, so sind doch kurze Abschnitte konserviert, die möglicherweise Bindestellen für Transkriptionsfaktoren darstellen. In der Regel sind diese Blöcke jedoch aufgrund von Inversionen, Translokationen oder Insertionen für ein klassisches BLAST-Alignment nicht zugänglich. Mit Hilfe der beiden spezialisierten Programme konnten zehn kurze, hochkonservierte Sequenzblöcke in der ausgewählten Region ausfindig gemacht werden (Abb. 3.24). Anschließend wurde die potenzielle Promotorregion in Maus und Mensch von fünf weiteren Programmen zur Promotorvorhersage näher untersucht (Kapitel 3.5.7), es wurde dabei jeweils die maximale Vorhersagegenauigkeit ausgewählt. Hierbei wurde von allen Programmen mit sehr großer Wahrscheinlichkeit ein Promotor in Maus und Mensch identifiziert. In der Maus konnte durch den Abgleich aller Daten der Bereich 9400-9950 bp, im Menschen der Bereich 9.500-9.880 bp der untersuchten genomischen Sequenz (Kapitel 3.5.7) als Promotor identifiziert werden. Die Analysen zeigten, dass es sich in beiden Spezies um einen Promotor ohne TATA-Box handelt. Dies ist nicht ungewöhnlich, da vermutlich nur 33% aller Säuger-Promotoren diese Box aufweisen (Suzuki et al., 2001), neueren Untersuchungen zufolge ist das Vorhandensein einer TATA-Box abhängig vom umgebenden GC-Gehalt und schwankt zwischen 42% bei hohem GC-Gehalt und 7% bei niedrigem GC-Gehalt (Bajic et al., 2004). Der Transskriptionsstartpunkt (TSS) wurde von den verwendeten

Programmen allerdings jeweils unterschiedlich positioniert. In der Maus zeigten sich drei potenzielle Starpunkte: bei 9739 bp, 9740 bp und 9818 bp. Im Menschen wurden ebenfalls drei unterschiedliche Startpositionen bestimmt: bei 9810 bp, 9828 bp und 9838 bp. Für eine exakte Festlegung des TSS ist eine experimentelle Bestimmung unumgänglich, dies kann allerdings nur in weiterführenden Arbeiten geleistet werden. Ein weiteres großes Problem ist die Bestimmung der Transskriptionsfaktoren (TF), die in dem ermittelten Promotorbereich binden. Auch wenn die Analysen mit höchster Stringenz durchgeführt wurden, konnten in der Maus 111 Transkriptionsfaktoren identifiziert werden, die Bindestellen in dem Promotorbereich finden. Eine Nachselektion per Hand war daher unumgänglich. Hierbei wurden folgende Kriterien angewendet: Die Bindestellen der Transkriptionsfaktoren mussten innerhalb der von DBA und PromoterWise identifizierten Blockregionen liegen und es wurden nur Transkriptionsfaktoren berücksichtigt, die von mehreren Programmen erkannt wurden und zwischen Maus und Mensch konserviert im Promotorbereich vorliegen. Um die auferlegte Stringenz der Analysemethoden aufrechtzuerhalten, wurde der Interspeziesvergleich auf Basis von Transkriptionsfaktor-Netzwerken (Abb. 4.7) durchgeführt. Bei dieser Analyse spielt die Kopplung verschiedener Transkriptionsfaktoren eine wichtige Rolle. Einige Faktoren entfalten ihre aktivierende oder reprimierende Funktion nur in Kombination mit anderen Transkriptionsfaktoren, zudem unterliegen sie untereinander einer bestimmten Abfolge und räumlicher Orientierung. Wenn diese Netzwerke dann zwischen Maus und Mensch konserviert in vergleichbarer Position zu finden sind, kann mit großer Wahrscheinlichkeit davon ausgegangen werden, dass diese Bindungsstellen im Promotor auch funktional sind. Durchgeführt wurden diese Analysen mit dem Programm FrameWorker (Genomatix Software GmbH, München). Auf diese Weise konnte in der murinen Sequenz die Anzahl der Faktoren auf 15 reduziert werden. In Tabelle 4.5 werden alle Transkriptionsfaktoren, die nach den o.g. Kriterien im Promotorbereich von *Dmxl*1 binden, besprochen.

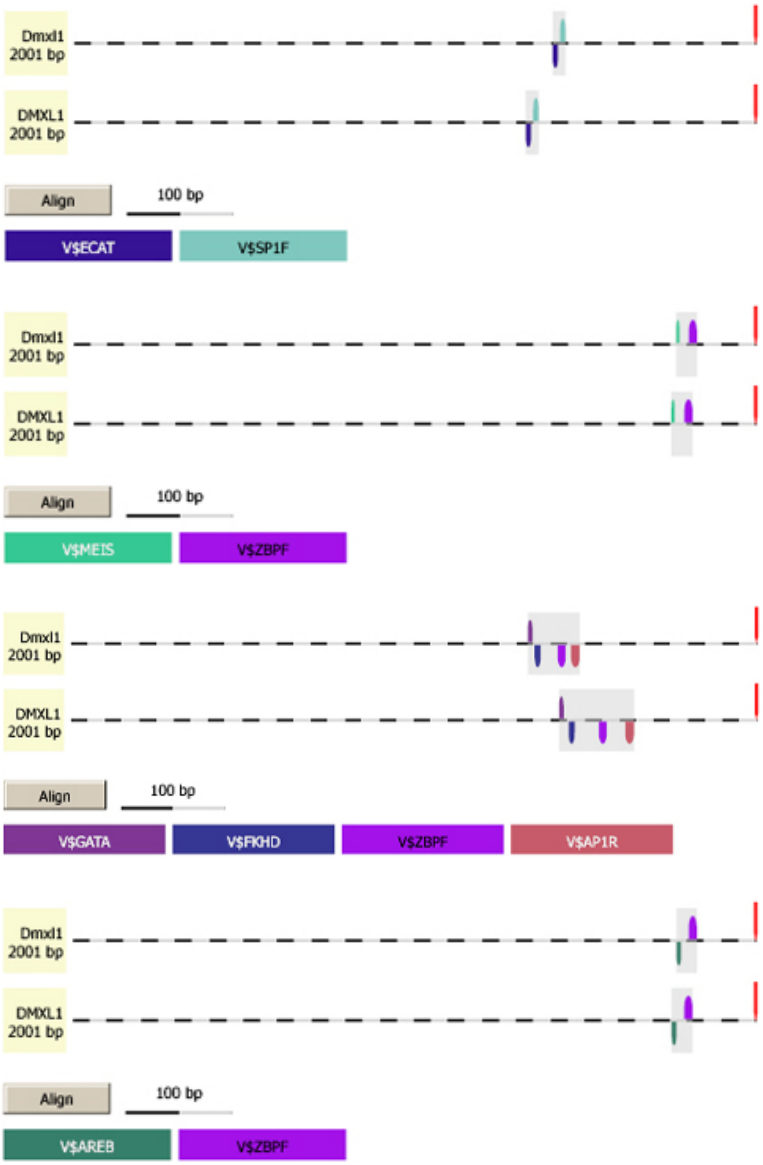


Abb. 4.8: FrameWorker-Analyse hochkonservierter Bindungsstellen von Transkriptionsfaktoren zwischen Maus und Mensch.

Analysiert wurden mit dem Programm FrameWorker 2001 bp stromaufwärts vom Translationsstartpunkt (roter Balken). Zwischen Maus und Mensch konnten vier hochkonservierte Netzwerke von zwei und mehr Transkriptionsfaktoren identifiziert werden. Alle liegen innerhalb des definierten Promotorbereichs und der identifizierten Sequenzblöcke. Gezeigt sind hier jeweils die homologen Netzwerke mit den interagierenden farbcodierten Transkriptionsfaktoren.

Transkriptionsfaktor	Bindungsposition	Expression/Funktion/Analyse	Literatur
AREB6 (ZEB)	9443-9455 bp(+)	Bei diesem Faktor handelt es sich um ein E-Box bindendes „zinc-finger homeodomain"-Protein, dass eine duale Funktion als Repressor oder Aktivator ausübt. Der Transkriptionsfaktor kann mit δEF1 interagieren und wirkt dabei als Repressor in hematopoetischen Zellen, in anderen Zelltypen kann diese Kombination jedoch aktivierend wirken. AREB6 kann die Expression gewebsspezifischer Gene während der Embryonalentwicklung regulieren.	Ikeda et al., 1995, 1998; Postigo, 2003; Naomi et al., 2004)
AP-1	9460-9484 bp (-)	Der AP-1 Komplex besteht typischerweise aus einem Jun- und einem Fos-Element, AP-1 Bindungsstellen sind sowohl in Promotoren als auch Enhancern zu finden. Der Faktor reguliert Zell-Proliferation, Zell-Differenzierung, Immunantwort, Onkogenese und Apoptose. AP-1 agiert als Biosensor in Zellen und ist häufig in stressinduzierten Transkriptionsfaktorkomplexen zu finden. Der Faktor befindet sich meist am Ende einer Transduktionskaskade und konvertiert extrazelluläre Signale in eine Veränderung der Genexpression.	Cottage et al., 2003; Wagner et al., 2003; Shaulian et al., 2002; Jochum et al., 2001
AP-4	9579-9585 bp (+) und 9581-9587 bp (-).	AP-4 gehört zur „helix-loop-helix" Proteinfamilie (HLH). Mitglieder dieser Familie sind vor allem bei Differenzierung und Proliferation embryonaler Zellen von Bedeutung. AP-4 besitzt neben einem HLA-Motiv zwei „leucine-zipper", die ebenfalls eine Dimerisierung des Proteins vermitteln. Der Transkriptionsfaktor gehört zu den ubiquitär exprimierten Faktoren.	Andriamanalijaona et al., 2003; Aranburu et al., 2001; Littlewood et al., 1995; Hu et al., 1990
CPBP („Core promoter-binding protein", ZBPF)	9421-9443 bp (-) und 9804-9827 bp (+)	Bei diesem Transkriptionsfaktor handelt es sich um einen transaktivierenden Faktor, der ubiquitär ab dem Embryonalstadium exprimiert wird und Gene ohne TATA-Box reguliert. Er gehört zur Familie der Krüppel-Zink-Finger-Proteine. Neben drei carboxyterminalen Zink-Finger-Motiven besitzt das Protein eine Bindungsdomäne für GC-reiche Sequenzen und ist in der Lage zusammen mit SP1 einen minimalen Transkriptionsfaktorkomplex zu bilden). CPBP bildet zusammen mit MEIS 1 oder AREB6 ein Netzwerk, die Bindungspositionen sind zwischen Maus und Mensch hoch konserviert (Abb. 4.7).	(Slavin et al., 1999; Kim et al., 1998
FKHD/Foxi1	9349-9361 bp(-)	Foxi1 (Fkh 10) wird in der Maus zuerst zwischen Tag E 9,5 und 10,5 im Epithel des endolymphatischen Ductus exprimiert. Foxi1-mutante Mäuse zeigen schwere Gehörschäden und eine Dysfunktion des Gleichgewichtssinns. Ab Tag E 16.5 wird eine Expression in den Nieren beobachtet, Foxi1$^{-/-}$-Mutanten zeigen hier eine fehlende Differenzierung der Epithelzellen des distalen Nephrons, die zu einer dRTA („distal renal tubular acidosis") führen. Neben der Steuerung von Genen, die die Pufferung des Urins kontrollieren, wird angenommen, dass Foxi1 in der Lage ist, die Expression einer ganzen Reihe weiterer Gene, die für die Entwicklung von Zelltypen, Biogenese von Organellen, apicaler Strukturen (Mikrovilli und Mikroplicae) sowie apicaler Funktionen (Endozytose und Exocytose) zu gewährleisten.	Blomquist et al., 2004; Al-Awquati et al., 2004; Ohyama et al., 2004; Hulander et al., 2003, 1998; Overdier et al., 1997
GATA 1	9334-9346 bp (+)	Die Bindungsstelle für diesen Transkriptionsfaktor befindet sich eigentlich außerhalb der vorher definierten Promotorregion, allerdings wurde dieser Faktor ebenfalls von verschiedenen Programmen erkannt und mit einem sehr hohen „Score" belegt. GATA 1 gehört zur Familie der GATA-Zink-Finger-Proteine und wird in der Maus zuerst in den sich entwickelnden Sertoli-Zellen während der ersten Spermatogenese exprimiert. Sertoli-Zellen sind die ersten Zellen, die sich in den fetalen Gonaden differenzieren und die Vorraussetzung für die Entwicklung der Leyding-Zellen schaffen. Zudem reguliert der Faktor die Expression anderer Gene hoch, die in Sertoli-Zellen exprimiert werden. Der Verlust der GATA1-Expression verhindert die Reifung der Sertoli-Zellen und kann in Folge zu verkleinerten oder kryptischen Testes führen. Weiter ist dieser Transkriptionsfaktor an Entwicklung und Differenzierung des hämatopoetischen Systems in der frühen Embryonalentwicklung beteiligt (). GATA1 interagiert mit dem Transkriptionsfaktor SP1, Bindungsstellen für GATA1 werden häufig in räumlicher Nähe zu SP1-Bindungsstellen gefunden.	Sharpe et al., 2003; Ohneda et al., 2002; Tang et al., 2001; Feng et al., 2000; Merika et al., 1995; Yomogida et al., 1994
IK 2 (Ikaros 2)	9455-9467 bp (+)	Die IKAROS-Gene codieren für eine Familie von funktionell unterschiedlichen Transkriptionsfaktoren mit Zinc-Finger-Motiven und spielen eine Rolle in der frühen Embryonalentwicklung. Ikaros 2 ist, wie NF1, ein modularer Transkriptionsfaktor, der sowohl reprimierend als auch aktivierend wirkt und zur Krüppel-Familie der Zink-Finger-Proteine gehört. Mitglieder der Ikaros-Familie spielen eine Rolle in der frühen Embryonalentwicklung sind aber auch bei der Entwicklung der Lymphozyten beteiligt. Ikaros ist ein integraler Bestandteil des Chromatin-Reorganisations-Komplexes in Thymocyten und T-Zellen und vermutlich in der Lage, bestimmte Chromatinstrukturen, die entweder aktivierend oder reprimierend auf bestimmte Zelltypen wirken, zu induzieren und aufrechtzuerhalten.	(Liu et al., 2004; Georgopoulos, 2002; Harker et al., 2002; Koipally et al., 2002; Kim et al., 1999; Winandy et al., 1995; Hahm et al., 1994; Molnar et al., 1994
MEIS 1	9768-9776 bp (+)	Dieser Transkriptionsfaktor gehört zur Familie der TALE Homeobox-Proteine. MEIS wird nur in der Embryonalentwicklung und dort ubiquitär exprimiert. MEIS ist in der Lage mit HOX- und PBX-Proteinen Dimere zu bilden und spielt vor allem in der Organogenese und Zelldifferenzierung sowie in der neuronalen Entwicklung eine wichtige Rolle.	Maeda et al., 2001; Thorsteinsdottir et al., 2001; Capdevila et al., 1999; Shanmugan et al., 1999
MOK 2	9550-9560 (-)	MOK 2 gehört ebenfall zur Krüppel-Familie der Zink-Finger-Proteine und wird primär in Testes und Gehirn exprimiert. Der Transkriptionsfaktor ist in der Lage sowohl an DNA als auch RNA zu binden, aufgrund dieser dualen Affinität und einer beobachteten Assoziation mit RNP („Ribonucleoprotein")-Komponenten wird ihm neben seiner regulatorischen Wirkung bei der Transkription auch eine Funktion in der posttranskriptionalen Regulation zugeschrieben.	Arranz et al., 2001, 1997
MyoD	9766-9790 bp (-)	Bei MyoD handelt es sich um einen basalen „helix-loop-helix"-Transkriptionsfaktor (bHLH) der früh in der Embryonalentwicklung exprimiert wird. Der Transkriptionsfaktor ist in der Lage den Präinitiationskomplex zu stabilisieren. Eine Interaktion von MyoD mit SP1 ist für die Entwicklung der Muskelzellen essentiell. MyoD hat zudem bei der Entwicklung des gesamten skelettalen Muskelsystems und im Zellzyklus der beteiligten Zellen eine wichtige Aufgabe.	Buckingham et al., 2003, 2001; Pownall et al., 2002; Heller et al., 1998; Sartorelli et al., 1990
MZF 1	9922-9928 bp (+)	MZF1 gehört zur Krüppel-Familie der Zink-Finger-Proteine und wird in totipotenten Zellen des hämatopoetischen Systems und myeloiden Vorläuferzellen exprimiert und spielt bei der Differenzierung dieser Zellen eine entscheidende Rolle.	Costanzi et al., 2003; Gaboli et al., 2001; Shivdasani et al., 1996; Morris et al., 1994; Bavisotto et al., 1991

Transkriptionsfaktor	Bindungsposition	Expression/Funktion/Analyse	Literatur
NFY (CBF, CP1, ECAT)	9405-9419 bp (-)	NFY ist ein ubiquitärer Transkriptionsfaktor, der aus drei Untereinheiten besteht und auf DNA-Ebene eine invertierte CCAAT-Box erkennt. NFY hat eine angenommene Funktion in der Regulation von Genen, die an der Cholesterol-Biosynthese und am Fettsäurestoffwechsel beteiligt sind. Der Transkriptionsfaktor ist zudem in der Lage, reprimierend und aktivierend zu wirken.	Peng et al., 2003, 2002; Gilthrope et al., 2002; Ericsson et al., 1997, 1996; Kim et al., 1996; Lopez et al., 1996; Bennett et al., 1995; Jackson et al., 1995; Maity et al., 1992;
SP1 („GC box element", „stimulating protein 1", Seguin et al., 1987)	9682-9731 bp (+/-) und 9807-9819 bp (+)	Das SP1-Protein ist ein basaler, ubiquitär exprimierter Transkriptionsfaktor und gehört zur Familie der Krüppel-Zink-Finger-Proteine, der sowohl an TATA- als auch an GC-Boxen bindet und im Vergleich zu TFIID ein 5-10 fach höhere Transkriptionsaktivität aufweist. Seine GC-reiche Bindungsstelle auf DNA-Ebene (GGGCG) wird vor allem in Promotoren ohne TATA-Box gefunden. Der Faktor positioniert die basale Transkriptionsmaschinerie an Promotoren ohne TATA-Motiv. SP1 ist in fast 60 % aller humanen Promotoren zu finden und in ein weites Spektrum zellulärer Prozesse eingebunden, darunter Regulation von Zellzyklusprozessen, Erhaltung der Genaktivität während der Embryonalentwicklung durch Unterbindung der *de novo* Methylierung von „CpG-Inseln" von Haushaltsgenen und Erhaltung der Chromatinstruktur. SP1 kann mit zahlreichen gewebsspezifischen Transkriptionsfaktoren kooperieren. Weiter ist bekannt, das SP1 an Promotoren von Haushaltsgenen bindet und ihre Transkription durch Interaktion mit der RelA-Untereinheit des enhancerbindenden Faktors NFκB verstärkt wird. Im Bereich 9682-9731 bp binden SP1-Dimere, der Bereich 9766-9819 bp zeigt Bindestellen für SP1 und MyoD („myogenic determination protein". Diese Abschnitte wurden auch von dem Programm ElDorado erkannt und als Promotor-Module definiert. Dies ist insofern interessant, da bei diesem Programm ein erkanntes Modul zwei oder mehr konservierte Elemente mit definierten und experimentell verifizierten Entfernungen untereinander und ihre Orientierung auf dem DNA-Strang repräsentiert. Diese erkannten Module sind hochspezifisch und die Wahrscheinlichkeit, dass es sich hierbei zusammen mit CPBP um die Kernkomponenten des Promotors handelt, ist sehr hoch. Auch dieser Transkriptionsfaktor fügt sich sehr gut in das Bild von *Dmxl*1 ein. SP1 gewährleistet die basale ubiquitäre Transkription und ist in der Lage mit Faktoren, die im Promotorbereich und potenziellen Enhancern von *Dmxl*1 binden, zu interagieren.	Klingenhoff et al., 2002Tang et al., 2001; Scherf et al., 2000Philipsen et al., 1999 Braun et al., 1998; Han et al., 1998 Naar et al., 1998 Pasceri et al., 1998 (Marin et al., 1997;Karlseder et al., 1996; Lin et al., 1996; Birnbaum et al., 1995 Krey et al., 1995; Look et al., 1995 Merika et al., 1995 Brandeis et al., 1994 Macleod et al., 1994 Pecorino et al., 1991Somma et al., 1991Su et al., 1991; Pugh et al., 1990Schmidt et al., 1989 Tapscott et al., 1988
ZF5	9573-9583 bp und 9782-9792 bp (beide +)	Der Transkriptionsfaktor ZF5 kann reprimierend und aktivierend zugleich wirken, er gehört zur Krüppel-Familie der Zink-Finger-Proteine und besitzt zwei funktionale Domänen, carboxyterminal ein Zink-Finger-Motiv, N-terminal wurde eine POZ-Domäne identifiziert. Das Zink-Finger-Motiv ist für eine aktivierende Funktion zuständig, die POZ-Domäne wirkt durch Protein-Protein Interaktion reprimierend. Die Bindungssequenz ähnelt der von SP1 und ist GC-reich). Es gibt Hinweise auf eine Kooperation mit SP1. In mehr als 65 % aller humanen Promotoren ist ZF5 fester Bestandteil des „Core"-Promotors, er bindet prozentual zu gleichen Teilen auf dem Plus- oder Minus-Strang.	Bajic et al., 2004; Numoto et al., 1999; Kaplan et al., 1997
δEF1	9769-9784 bp (+)	Dieser Transkriptionsfaktor besitzt ungewöhnlicherweise multiple Zink-Finger-Motive und eine Homeo-Domäne und ist essentiell für die embryonale Entwicklung. Er wird in multiplen Geweben exprimiert und steuert u.a. die Entwicklung des Nervensystems, reguliert die Expression von Ovalbumin und wird durch Östrogen induziert, zudem ist der Faktor in der Lage in Kombination mit ELF-2 einen Repressor zu bilden.	Darling et al., 2003; Dillner et al., 2002a, 2002b; Remacle et al., 1999

Tab. 4.5: Durch FrameWorker-Analysen identifizierte Transkriptionsfaktoren die im Promotorbereich von *Dmxl*1 binden

Mit Hilfe der identifizierten Transkriptionsfaktoren des murinen Promotors von *Dmxl*1 kann ein mögliches Szenario der Regulation von *Dmxl*1 erarbeiten werden. So können im Promotorbereich Transkriptionsfaktoren binden, die eine basale ubiquitäre Expression von *Dmxl*1 in jedem Entwicklungsstadium ermöglichen (AREB-6, AP-1, CPBP, NFY, SP1, ZF5). Dies passt sehr gut zu den experimentellen Daten dieser Arbeit und zu einer Transkriptionsanalyse von *DmX* in Drosophila durch Weil (2000), hier zeigte sich ebenfalls eine übiquitäre Grundexpression. Vor allem scheint die basale Expression von *DmX* im adulten Entwicklungsstadium ausreichend für die Funktion des resultierenden Proteins zu sein. Besonders zu erwähnen ist in diesem Zusammenhang der Transkriptionsfaktor NFY, der eine entscheidende Rolle in der Regulation von Genen spielt, die an der Cholesterol-Biosynthese und am Fettsäurestoffwechsel beteiligt sind. Möglicherweise besteht hier ein Zusammenhang mit der beobachteten Obesitas der chimären Tiere. Viele weitere identifizierte Transkriptionsfaktoren üben vor allem in der Embryonalentwicklung eine wichtige Funktion aus (AREB6, AP-4, FKHD, GATA 1, IK 2, MEIS, MOK 2, MyoD, SP1, δEF1) und sind z. T. nur in dieser Zeit, oder mit veränderten Aufgaben aktiv. Teilweise handelt es sich hierbei um gewebsspezifische Faktoren, so ist GATA 1 an der Entwicklung der Testes beteiligt. Dieser Faktor könnten somit eine Rolle bei der Interpretation des beobachteten Phänotyps der *Dmxl*1$^{-/+}$-Chimären spielen. Möglicherweise muss *Dmxl*1 in der Entwicklungsphase der Testes besonders reguliert werden und die Anzahl der *Dmxl*1-Transkripte in den entsprechenden Zelltypen höher sein als in anderen Geweben. Eine Reduktion der Transkripte um 50% könnte so theoretisch zu einer Haploinsuffizienz führen. Auch die Transkriptionsanalysen von *DmX* während der Embryonalentwicklung von Drosophila zeigen eine Abweichung vom der ubiquitären Grundexpression hin zu einer differenziellen Verteilung der Transkripte (Weil, 2000). Hierbei wurde eine intensive Transkription im Vorder-, Mittel- und Enddarm sowie im ZNS beobachtet. Ach hier lassen sich Parallelen zu den identifizierten Transkriptionsfaktoren ziehen: MOK 2 und δEF1 sind an der Entwicklung des Gehirns und des ZNS beteiligt und möglicherweise spielt auch der Transkriptionsfaktor MyoD eine Rolle, da der gesamte Darm von Drosophila mit einer viszeralen Darmmuskulatur umgeben ist und diese Muskulatur einen Teil zu der beobachteten intensiven Transkription in Drosophila beigetragen haben könnte. Interessant ist auch der relativ große Anteil an Transkriptionsfaktoren, der sowohl eine aktivierende als auch reprimierende Funktion ausüben kann (AREB6, IK 2, NFY, ZF5, δEF1). Reine Repressoren konnten nicht identifiziert werden. Zusammenfassend lässt sich aufgrund dieser Beobachtungen sagen, das *Dmxl*1 in der Embryonalentwicklung vermutlich sehr komplex und z.T. gewebsspezifisch reguliert wird, im adulten Organismus jedoch eine niedrige Grundexpression ausreichend zu sein scheint. Ein Modell des Promotors und der im Folgenden besprochenen Enhancer ist in Abb. 4.8 gezeigt.

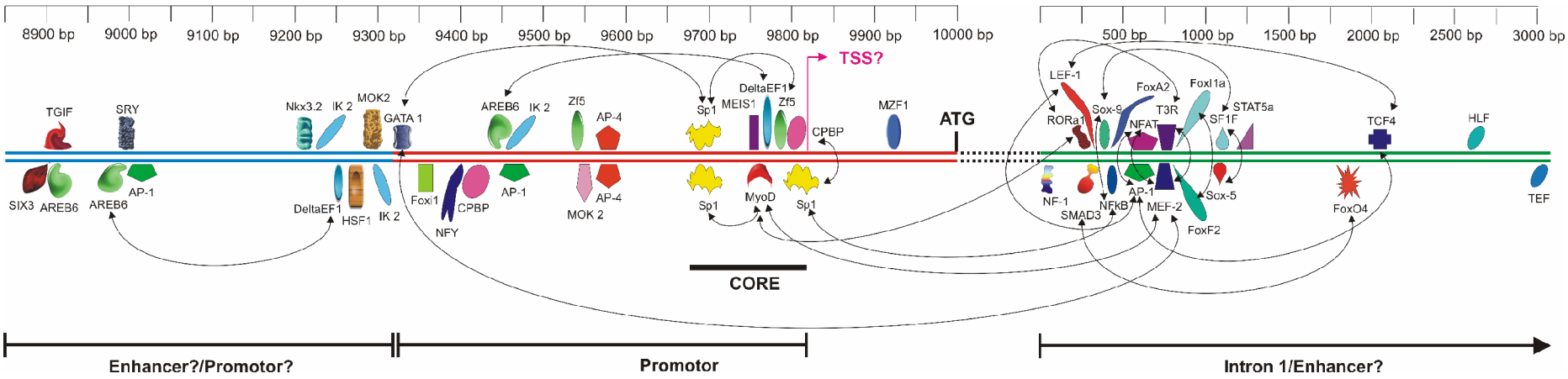


Abb. 4.9: Grafische Darstellung der potentiellen Promotor- und Enhancerstruktur von *Dmxl*1

Die durch Computervorhersage identifizierten regulatorischen DNA-Abschnitte von *Dmxl1* sind in drei Bereiche gegliedert. Der blaue DNA-Doppelstrang symbolisiert eine stromaufwärts gelegene Region, bei der es sich möglicherweise um einen Enhancer handeln könnte. Aufgrund der vorliegenden Daten ist eine eindeutige Klassifizierung jedoch nicht möglich. Der rot markierte DNA-Abschnitt zeigt den mit großer Wahrscheinlichkeit identifizierten Promotorbereich von *Dmxl*1, der „Core"-Promotor ist durch einen schwarzen Balken angedeutet. Der putative Transskriptionsstart und der Translationsstart sind ebenfalls gezeigt. Der grüne DNA-Abschnitt zeigt einen Teil des ersten Introns, der möglicherweise weitere Ebhancer–Elemente enthält. Durch Literaturquellen belegte Interaktionen zwischen verschiedenen Transkriptionsfaktoren sind durch Pfeile markiert, bei mehreren vorhandenen gleichen Kooperationspartnern ist nur eine der möglichen Interaktionen aufgezeigt.

Enhancer:

Untersuchungen zur Regulation der Transkription von *DmX* in Drosophila zeigten, dass sich im ersten Intron möglicherweise regulatorische Abschnitte befinden, die die Transkription regeln, nicht jedoch für die Initiation der Transkription ausreichen. Für eine wildtypische Expression von *DmX* sind zusätzlich stromaufwärts liegende Sequenzabschnitte notwendig (Weil, 2000). Bei der Analyse des Promotorbereichs von *Dmxl*1 wurden in direkter Nachbarschaft zu dem identifizierten Promotor weitere promotorähnliche Strukturen erkannt. Das Programm TSSG identifizierte diesen Bereich klar als Enhancer und grenzte ihn vom Promotor ab, der ebenfalls erkannt wurde. Andere verwendete Programme sind nicht in der Lage zwischen Promotor und Enhancer zu unterscheiden, hier obliegt es dem Nutzer dies zu interpretieren. Kein anderes Programm identifizierte in diesem Bereich eine Promotor-ähnliche Struktur, dennoch wurde die Region mit gleichen stringenten Bedingungen wie bei der Promotoranalyse untersucht. Vor der eigentlichen Promotorregion befinden sich fünf weitere durch die Programme DBA und PromoterWise erkannte Blockregionen (Abb. 3.26). Innerhalb dieser Blockregionen sind zwischen Maus und Mensch mehrere hochkonservierte Bindungsstellen für folgende Transkriptionsfaktoren zu finden: AREB6, TGIF, SIX3, AP1, SRY, IK2, Nkx3.2, MOK 2, HSF1. Alle Faktoren wurden unabhängig voneinander von verschiedenen Programmen erkannt. Aufgrund der räumlichen Nähe zu dem ermittelten Promotor ist anhand der vorliegenden Daten nicht zu entscheiden, ob es sich hierbei um ein „Enhancer"-Element handelt oder dieser Bereich strukturell noch zum Promotor zu rechnen ist. Einige der gefundenen Transkriptionsfaktoren binden zusätzlich auch im vorher beschriebenen Promotor von *Dmxl*1, neue Faktoren werden in Tabelle 4.6 gezeigt:

Transkriptionsfaktor	Expression/Funktion/Analyse	Literatur
TGIF (TALE, „TG-interacting factor")	Dieser Transkriptionsfaktor gehört zur Familie der TALE Homeobox-Proteine. Er wirkt primär als Repressor in TGFβ-abhängigen Signalkaskaden und blockiert auch SMAD-Proteine. Mutationen in TGIF erzeugen beim Menschen die Krankheit Holoprosencephalie, eine häufig vorkommende Störung der Hirnentwicklung. Allerdings konnte die Subklasse des hier vorliegenden TGIF nicht näher identifiziert werden. Interessanter Weise reguliert TGIF in Drosophila die Spermatogenese und auch in der Maus ist eine Subklasse des TGIF (TEX1) an der Spermatogenese beteiligt.	Dubourg et al., 2004; Ayyar et al., 2003
SIX3	Hierbei handelt es sich ebenfalls um ein Homeobox-Protein, das wie TGIF in enger Verbindung mit dem Krankheitsbild der Holoprosencephalie steht. SIX3 steuert die Entwicklung der Vorderhirns und der Augen.	Dubourg et al., 2004; Laflamme et al., 2004; Zuber et al., 2003; Toy et al., 1998; Kawakami et al., 1996
SRY (Sex-determining Region gene)	Dieser Faktor bestimmt in den sich entwickelnden Gonaden durch seine An- oder Abwesenheit die Entwicklung zu Testes oder Ovarien und wird in pre-Sertoli Zellen exprimiert. SRY gehört zu der Gruppe der SOX („SRY-related HMB box")-Proteine. Da SRY eine hochaffine DNA-Bindungsdomäne aufweist, wird im allgemeinen angenommen, dass er als Transkriptionsfaktor agiert und die Ausbildung der männlichen Geschlechtsorgane steuert. Das SRY-Gen wird nur bei Säugern gefunden.	(Knower et al., 2003 Clarkson et al., 2002Albrecht et al., 2001 Bergstrom et al., 2000 Soullier et al., 1998; Just et al., 1995

Transkriptionsfaktor	Expression/Funktion/Analyse	Literatur
Nkx3.2 (Babx1)	Der Transkriptionsfaktor gehört zur Familie der Homeobox-Proteine und wird in der frühen Embryonalentwicklung im Mesoderm und anderen Geweben exprimiert und ist an der Skelettalentwicklung beteiligt.	Kim et al., 2003; Herbrand et al., 2002; Kempf et al., 2002; Schneider et al., 2000, 1999; Tanaka et al., 1999; Sciavolino et al., 1997; Tribioli et al., 1997a, 1997b
HSF1	„heat shock"-Faktoren sind die hauptsächlichen transaktivierenden Faktoren, die die Regulation von „heat shock"-Proteinen gewährleisten. HSF1 wird ubiquitär in allen Geweben und nahezu allen Zelltypen exprimiert, ist zudem essenziell für die Entwicklung der Testes, er reguliert dort Zielgene, die zu bestimmten Zeitpunkten während der Entwicklung der Keimzellen aktiv sind.	Wang et al., 2004; Huang et al., 2001; Morimoto, 1998; Nakai et al., 1997, 1995; Wu, 1995; Fernandes, 1994

Tab 4.6: Identifizierte Transkriptionsfaktoren in konservierten Sequenzblöcken stromaufwärts des Promotors von *Dmxl*1

Die in Tabelle 4.6 gezeigten Transkriptionsfaktoren binden in einem zwischen Maus und Mensch konservierten Sequenzabschnitt, der sich vor dem eigentlichen putativen Promotor von *Dmxl*1 befindet. Gezeigt sind nur Transkriptionsfaktoren, für die Bindungsstellen in orthologer Position in beiden Spezies exitieren. Auf Transkriptionsfaktoren, die schon in Tabelle 4.5 besprochen wurden, wird hier nicht näher eingegangen

Auch wenn die vorliegenden Daten nicht klar auf eine Enhancer-Struktur hinweisen, scheint diese Region auch hinsichtlich der konservierten Blöcke zwischen Maus und Mensch eine Funktion auszuüben. Unabhängig von ihrer Einordnung als Enhancer oder noch zum Promotor gehörend, ist aufgrund der dort bindenden Transkriptionsfaktoren u.a. eine Testes-spezifische Regulation von *Dmxl*1 und damit auch eine spezifische Funktion in der Keimzellentwicklung zu erwarten. Ein weiteres mögliches Enhancer-Element wurde im ersten Intron von *Dmxl*1 identifiziert. Der für die vorliegende Arbeit erstellte PIP- und zPicture-Vergleich (Abb. 3.21, 3.22, 3.29) zeigt außerhalb von Regionen mit repetitiven Elementen einige ungewöhnlich homologe Sequenzabschnitte im ersten Intron. Diese Regionen wurden zunächst ebenfalls mit den Programmen DBA und PromoterWise näher untersucht. Hierbei konnte ein Abschnitt von 3,5 kb identifiziert werden, der aus zwischen Maus und Mensch konservierten colinearen Sequenzblöcken besteht. Eine Analyse dieser Region mit dem Programm ElDorado und FrameWorker zeigte spezifische Module für die Maus-Region. Der Bereich wurde zusätzlich mit den Programmen MatInspector, Dragon Genome Explorer, Match Search und rVista untersucht und die Daten wie oben beschrieben untereinander abgeglichen. Auf diese Weise sollte wieder eine möglichst hohe Stringenz erreicht werden. Allerdings wurde bewusst der Interspeziesvergleich der in dieser Region bindenden Transkriptionsfaktoren (Tab. 4.8) separat behandelt, da bekannt ist, dass vor allem Enhancer teilweise speziesspezifische Unterschiede in den dort bindenden Faktoren und ihrer Lokalisation zeigen können. Es werden häufig überlappende, aber nicht identische Sätze von Transkriptionsfaktoren in homologen Enhancern gefunden, zudem ist ihre Anordnung nicht zwingend konserviert (Valverde-Garduno et al., 2004; Long et al.,

2003; Follows et al., 2003; Sasaki et al., 2002; Scemama et al., 2002; Silverman et al., 2002). Es ist daher wahrscheinlich, dass nicht alle der gezeigten Transkriptionsfaktoren *in vivo* in der untersuchten DNA-Sequenz binden. Eine noch stringentere Untersuchung war jedoch aus oben genannten Gründen nicht sinnvoll. Die kombinatorische Analyse (Tab. 4.7) zeigt sechs Module mit den Faktorkombinationen NF-κB/Sox-9, AP-1/NFAT, MEF2/T3R, FoxF2/FoxI1a, Sox-5/SF1F, und HLF/TEF. Für die humane Sequenz sind ebenfalls sieben Module für DNA-bindende Faktoren vorausgesagt: AP-1/ELK-1, STAT5/NFY, NFAT/AP-1, Sox-9/ELK-1, SMAD3/AP-1, MEF2/T3R und OCT1/CEBP. Die Module AP-1/NFAT und MEF2/T3R sind in beiden Spezies zu finden, der Transkriptionsfaktor Sox-9 kann ebenfalls, jeweils mit einem anderen Partner, in der potenziellen trans-regulatorischen Region in Maus und Mensch identifiziert werden. Bei einem Interspeziesvergleich zwischen Maus und Mensch werden elf konservierte Transkriptionsfaktoren (Tab 3.10) in orthologer Position von verschiedenen Programmen erkannt. Im Folgenden werden die identifizierten Transkriptionsfaktoren des potenziellen Enhancerbereichs der Maus vorgestellt, die noch nicht bei der Besprechung des Promotors beschrieben wurden.

Die ermittelten konservierten Bindungsstellen der putativen Enhancerregionen und die damit assoziierten Transkriptionsfaktoren passen meist sehr gut zu dem identifizierten Promotor von *Dmxl*1. Die regulatorische Wirkung der Faktoren liegt teilweise in der frühen Embryonalentwicklung und dort unter anderem in der Testes-spezifischen Expression sowie einer komplex regulierten gewebsspezifischen Expression im adulten Organismus. Um die hier gezeigten computergestützten Analysen des Promotors und der Enhancerregionen zu überprüfen, müssten umfangreiche experimentelle Untersuchungen durchgeführt werden, die in dieser Arbeit nicht geleistet werden konnten. Dennoch ist es möglich, über ein Expressionsspektrum von *Dmxl*1 die identifizierten Transkriptionsfaktoren des Promotors und der Enhancer zumindest teilweise zu überprüfen. Hierzu wurden öffentlich zugängliche Daten genutzt. Su et al. (2004, 2002) verwendete die Affymetrix-Technologie, um eine globale Analyse des humanen und murinen Transkriptoms zu erstellen. In ihrer ersten Arbeit wurden DNA-Chips auf der Grundlage von Transkripten von ~10000 humanen und ~7000 murinen annotierten Genen sowie uncharakterisierten EST-Daten hergestellt. Diese wurden mit Proben von 46 humanen und 45 murinen Geweben hybridisiert und die Ergebnisse in einer Datenbank (Gene Expression Atlas 2 (Datensatz U95A), Mouse Gene Atlas (Datensatz U72a), http://symatlas.gnf.org) öffentlich zugänglich gemacht. In einer zweiten Arbeit wurden 44775 humane und 36182 murine annotierte Gene und uncharakterisierte EST-Daten mit RNA aus 79 humanen und 61 murinen Geweben hybridisiert. Auch diese Daten sind öffentlich zugänglich (Mensch: U133A+GNF1H; Maus: GNF1M). Das Expressionsspektrum von *DMXL*1 ist in beiden Datensätzen vorhanden, das murine *Dmxl*1 findet sich nur im GNF1M-Datensatz. In Abb 4.9 ist das Expressionsspektrum U133A des humanen *DMXL*1 im Vergleich mit *Dmxl*1 (GNF1M) gezeigt. Alle Hybridisierungsexperimente wurden primär mit adultem Ausgangsmaterial durchgeführt (Su et al., 2004, 2002), so dass eine Beurteilung der gewebsspezifischen Expression der Transkriptionsfaktoren von *Dmxl*1, die vor allem in

Transkriptionsfaktor	Expression/Funktion/Analyse	Literatur
NF-κB	Der Faktor bindet an regulatorische Elemente (Enhancer) der κB-Familie und wird zu den basalen transkriptionsverstärkenden Faktoren gerechnet, er reguliert eine ständig ansteigende Zahl von Zielgenen und wird ubiquitär exprimiert. Er spielt vor allem in der Entwicklung der Testes, Immunantwort, Apoptose, Stressantwort, Zelldifferenzierung und nicht zuletzt bei der Entstehung von Krebs eine zentrale Rolle. Die Regulation eines prostataspezifischen Gens konnte ebenfalls nachgewiesen werden. NF-κB ist in der Lage mit SP1 zu interagieren.	Torchinski et al., 2004; Zelivianski et al., 2004 Kucharczak et al., 2003; Senftleben, 2003; Budde et al., 2002; Ghosh et al., 2002; Senftleben et al., 2002; Persson et al., 2001; Nishikimi et al., 1999; May et al., 1997; Perkins et al., 1994; Molitor et al., 1990
Sox-9 (SRY-related high-mobility group (HMG) box 9)	Sox-9 gehört wie SRY und SF1 zu den wichtigsten geschlechtsbestimmenden Faktoren und wird primär in den sich entwickelnden Knochen und Testes exprimiert. Er fungiert als enhancerbindender Transkriptionsfaktor und interagiert u.a. mit SF1 und NF-κB. Bisher konnten nur wenige Gene identifiziert werden, die nachweislich von Sox-9 reguliert werden. Bekannt ist eine essenzielle Rolle in der Zellkondensation und Differenzierung von Prechondrozyten und Chondroblasten sowie der Ausbildung des männlichen Geschlechts. Der Transkriptionsfaktor bindet auch in der potenziellen homologen Enhancerregion im Menschen.	Smits et al., 2004; Harley et al., 2003; Knower et al., 2003; Pannetier et al., 2003; Akiyama et al., 2002; Bronwyn et al., 2002; Clarkson et al., 2002; Lefebvre, 2002, 1997; Bowles et al., 2001; Koopman et al., 2001; Veitia et al., 2001; Bi et al., 1999; Mertin et al., 1999; Bell et al., 1997
NFAT („nuclear factor of activated T cells“)	Transkriptionsfaktoren der NFAT-Familie zeigen eine Rel-ähnliche Bindungsdomäne und werden durch die Kalzium/Calmodulin-abhängige Phosphatase Calcineurin aktiviert. Die NFAT-Familie hat vier Mitglieder (C1-C4) die während der Embryonalentwicklung die Zellentwicklung, Zelladaption und Zelldifferenzierung u.a. von Herz, vaskularen Endothelzellen, Skelettmuskeln, Chondrozyten, Keratinozyten und Adipozyten steuern. NFAT bildet zusammen mit AP-1 ein regulatorisches Netzwerk, benachbarte Bindungspositionen ermöglichen die Ausbildung eines hochstabilen ternären Komplexes, der in immunologisch wirksamen Zellen eine Vielzahl von Genen reguliert, darunter Interleukine, Cytokine, Chemokine und membranständuge Rezeptoren. Der Komplex ist vermutlich auch in anderen Geweben regulatorisch aktiv. Die potenzielle Enhancerregion von DmX zeigt in Maus und Mensch Bindungsbereiche für AP-1/NFAT-regulatorische Netzwerke.	Schulz et al., 2003; Crabtree et al., 2002; Horsley et al., 2002, 2001; Chineneov et al., 2001; Graef et al., 2001; Friday et al., 2001; Kegley et al., 2001; Marcian et al., 2001; Liu et al., 1999; Aramburu et al., 1998; Beals et al., 1997; Rao et al., 1997; Logan et al., 1996; Luo et al, 1996; Shibaski et al., 1996; Jain et al., 1993
MEF2 (myocyte enhancer factor-2)	Dieser faktor wird in Muskelzellen und dem ZNS exprimiert und bindet dort spezifisch an Enhancersequenzen. Er gehört zur Familie der „basic helix-loop-helix“-Proteine (bHLH). Neben MyoD ist MEF2 der prominenteste Transkriptionsfaktor, der die Entwicklung des skelettalen Muskelsystems steuert. Beide Faktoren interagieren miteinander und es wird angenommen, das MyoD diese Interaktion benötigt, um transkriptionell aktiv zu sein. MEF2 ist zudem in der Lage mit GATA und NFAT und T3R zu kooperieren und wird durch Calmodulin kontrolliert. Ein MEF2/T3R-Modul wird in Mensch und Maus für den möglichen Enhancerbereich vorausgesagt.	Pan et al., 2004; Buckingham et al., 2003; Dodou et al., 2003; McKinsey et al., 2002; Naya et al., 1999 Lin et al., 1996; Gossett et al., 1989
T3R	T3R gehört zur Familie der hormonregulierten Transkriptionsfaktoren. Der „thyroid hormone receptor“ ist in der Lage, durch Interaktion mit verschiedenen Cofaktoren die Transkription von Genen zu reprimieren oder zu aktivieren. Typischer weise reprimiert T3R in Abwesenheit des Hormons die Transkription, während vorhandenes Hormon die Transkription aktiviert. Dies geschieht durch die Rekrutierung von Cofaktoren, die dann reprimierend oder aktivierend auf den Promotor einwirken. T3Rs spielen eine wichtige Rolle in der Homöostase, Differenzierung und Entwicklung von Vertebraten. Sie werden ubiquitär in jedem Entwicklungsstadium exprimiert und in multiplen Isoformen gefunden, die zell- und entwicklungsspezifische Rollen bei der Regulation von Genen übernehmen. T3R ist in der Lage mit MEF2 kooperativ die Transkription zu verstärken. Beide Faktoren werden in einer modularen Anordnung in der potenziellen Enhancerregion in Maus und Mensch gefunden.	Yang et al., 2001; Zhang et al., 2000; Privalski et al., 1999; Murata et al., 1998; Torchia et al., 1998; Meier et al., 1997; Moriscot et al., 1997; Horwitz et al., 1996; Mangelsdorf et al., 1995; Tsai et al., 1994Lazar et al., 1993; Ribeiro et al., 1993
FKHD/FoxF2	bezeichnet die „Fork Head Domain“-Familie, auch diese Faktoren binden an Enhancer-Elemente. Mitglieder der Familie werden in nahezu allen Geweben und Embryonalstadien exprimiert. Hier konnte ein Modul mit den Faktoren Foxi1 und FoxF2 (FREAC-2) identifiziert werden. FoxF2 bindet an Enhancer-Elemente und wird im an das Epithel angrenzenden Mesenchym der Atmungsorgane, Darm und Genitalien exprimiert. Während der Embryogenese ist zudem eine starke Expression im zentralen Nervensystem, Auge und Ohr zu beobachten.	Ormestad et al., 2004; Wang et al., 2003; Aitola et al., 2000; Hellqvist et al., 1998; Kaestner et al., 1996, Weigel et al., 1990
SORY/Sox-5	SORY ist der Oberbegriff der „SOx/sRY-sex/testis determining and related HMG box factors“-Familie. Hierzu gehören z.B. SOX5, SOX9 und SRY. Bindungsstellen für diese Faktoren werden vor allem in Enhancern gefunden und sind essenziell für die Entwicklung der Testes. Bei der hier identifizierten Bindungsstelle handelt es sich um die Konsensussequenz von Sox-5. Sox-5 wird als kurzes (SOX5) oder langes (L-SOX5) Transkript exprimiert, kurze Transkripte werden in den Testes während der Spermatogenese gefunden, lange Transkripte dagegen im Hirn, in Chondrozyten und anderen Geweben während der Embryonalentwicklung.	Smits et al., 2004; Lioubinski et al., 2003; Clarkson et al., 2002; Lefebvre et al., 2002, 1998; Mertin et al., 1999; Hiraoka et al., 1998; Takamatsu et al., 1995; Connor et al., 1994;

Transkriptionsfaktor	Expression/Funktion/Analyse	Literatur
SF1F	Als SF1F bezeichnet den „steroidogenic factor 1 F“, dieser Faktor bindet meist in Kombination mit anderen Faktoren wie Sox-9, Sox-5 oder GATA4 an DNA-Bereiche und wirkt transkriptionsverstärkend. Der Faktor wird primär in Steroid-produzierenden Organen, wie dem Adrenocortex und den Gonaden exprimiert, aber auch im ventromedialen Hypothalamus und den gonadotrophen Zellen des vorderen Hypophysenlappens gefunden. SF1 nimmt eine Schlüsselstellung in der Regulation steroidogener und gonadotropher Gene ein und ist essenziell an der Entwicklung der Testes beteiligt. Neuere Untersuchungen zeigen eine Beteiligung von SF1 bei der Entwicklung einer im Hypothalamus ausgelösten Fettleibigkeit.	Fowkes et al., 2003; Clarkson et al., 2002; Majdic et al., 2002; Achermann et al., 1999; Tremblay et al., 1999; Oba et al., 1996; Bernhart et al., 1994; Ingraham et al., 1994; Ikeda et al., 1993; Morohashi et al., 1993
HLF	Der Transkriptionsfaktor „hepatic leukemia factor“ HLF gehört zu einer Subfamilie der an Prolin und sauren Aminosäuren reichen bZIP-Proteine. Weitere Mitglieder dieser Familie sind TEF und DBP. HLF wurde ursprünglich bei der Untersuchung von Patienten mit akuter lymphatischer Leukämie entdeckt, bei denen eine Translokation t(17:19) zu einem Fusionsprotein aus E2A und HLF führte. HLF wird in Leber, Niere, Lunge und dem adulten Nervensystem exprimiert. Im frühen Embryonalstadium ist die Expression von HLF auf die Hypophyse beschränkt und weitet sich postnatal auf das gesamte Nervensystem aus. Es wird angenommen, dass dieser Faktor eine wichtige Rolle in der Regulation des programmierten Zelltods ausübt. HLF gehört zudem zu einer Gruppe von Transkriptionsfaktoren, die eine tagesrhythmusabhängige Expression zeigen und durch das Hormon TSHβ gesteuert werden.	Young et al., 2001; Hitzler et al., 1999; Inukai et al., 1996; Hunger et al., 1992; Inaba et al., 1992; Drolet et al., 1991
TEF	Der zur bZip-Proteinfamilie gehörende Faktor wird in der Maus in zwei Isoformen, TEFα und TEFβ („thyrotroph enhancer factor“), gefunden. Die Isoformen unterscheiden sich N-terminal und resultieren aus einer alternativen Verwendung des Promotors. Beide Isoformen werden sehr stark in Lunge, Harnblase, Niere, Darm und Hirn exprimiert, eine schwächere Expression ist ubiquitär zu beobachten. DmX wird bei Maus und Mensch ebenfalls in Harnblase, Niere und Hirn stärker exprimiert, als in anderen Geweben (Tab. 4.9). Der Faktor ist in der Lage, die Transkription konstitutiv durch Bindung an ein „TFE-response element“ (TFERE) zu aktivieren. Neuere Untersuchungen zeigen eine kalziumabhängige Aktivierung von TFE durch CaMK IV. Wie bei HLF ist die Expression von TEF vom Tagesrhythmus abhängig und korreliert mit der der Ausschüttung von TSHβ, einem durch die Hypophyse abgegebenen Hormon.	Allaman-Pillet et al., 2004; Zhou et al., 2004; Young et al., 2001; Krueger et al., 2000; Hunger et al., 1996; Drolet et al., 1991

Tab 4.7: Durch kombinatorische Analyse identifizierte Transkriptionsfaktoren.

Transkriptionsfaktor	Expression/Funktion/Analyse	Literatur
FKHD/FoxA2	FoxA2, auch als HNF3β („hepatocyte nuclear factor 3“) bekannt, gehört zur „Fork Head Domain“-Familie und reguliert die Expression verschiedener Gene in Leber, Pankreas und dem ZNS. Zudem spielt der Faktor eine wichtige Rolle in der embryonalen Entwicklung, FoxA2-Null-mutante Mäuse sterben in der frühen Embryonalentwicklung durch Defekte der vom Endoderm abgeleiteten Organe. FoxA2 bindet an regulatorische Elemente und wirkt transaktivierend.	Ceelie et al., 2003; Fouchet et al., 2003; Wan et al., 2003; Benoit et al., 2002; Wang et al., 2002; Zaret, 1999; Duncan et al., 1998; Levinson-Duchnik et al., 1997; Sharma et al., 1997; Philippe, 1995; Weinstein et al., 1994; Ang et al., 1993
FKHD/FoxO4	Die O-Subfamilie der „Fork Head Domain“-Superfamilie spielt eine wichtige Rolle in Zellproliferation und Tumorgenese und interagiert mit Smad3. Die O-Famile ist erst in jüngster Zeit Ziel der Forschung, daher konnten bisher nur wenige Gene identifiziert werden, bei denen FoxO-Transkriptionsfaktoren eine wichtige Rolle spielen. FoxO4 (AFX, „acute-lymphocytic-laukaemia-1 fused gene from chromosome X“) reguliert die Transkription von Bcl-6 und suprimiert die zelluläre Antwort auf Hypoxie.	Arden, 2004; Seoane et al., 2004; Hu et al., 2004, Birkenkamp et al., 2003; So et al., 2003; Tzu-Ling et al., 2003; Tang et al., 2002
LEF-1	Der „lymphoid enhancer factor“-1 gehört zur Familie der TCF/LEF-DNA-bindenden Transkriptionsfaktoren (HMG-Box Superfamilie), die eine wichtige Rolle in der Embryogenese, Tumorgenese und Erneuerung bzw. Aufrechterhaltung von regenerativen Geweben spielen. TCF und LEF sind nukleäre Effektoren des wnt-Signalwegs. Während der Embryogenese und in einigen adulten Geweben werden beide Faktoren Co-Exprimiert und zeigen redundante Funktionen. LEF-1 ist in der Lage, synergistisch mit AP-1 und TCF-4 zu kooperieren.	Rivat et al., 2003; Hurlstone et al., 2002; Wortman et al., 2002; Oosterwegel et al., 1993
NFI	Der „Nuclear Factor I“ gehört zur Familie der CTF-Proteine und reguliert sowohl adenovirale DNA-Replikation als auch zellulare Genexpression. Inzwischen wurden mehr als hundert zelluläre und virale Gene identifiziert, die NFI-Bindungspositionen in Promotor, Enhancer und Silencer aufweisen. Mutationsanalysen zeigten, das NFI essenziell für die Expression der meisten dieser Gene zu sein scheint. Der Faktor ist in der Lage, reprimierend oder aktivierend auf diese Gene einzuwirken. Die NFI-Familie besteht aus vier Genen (-A, -B, -C, -X) die ubiquitär exprimiert werden. Durch alternatives Spleißen und Dimerisierung werden jedoch viele spezielle Formen gewebsspezifisch exprimiert.	Murthag et al., 2003; Gronostajski, 2000; Cooke et al., 1999; Osada et al., 1997; Adams et al., 1995; Monaci et al., 1995; Seipel et al., 1992; Martinez et al., 1991; Santoro et al., 1988; Jones et al., 1987; Gronostajski et al., 1985; Nagata et al., 1982, 1983
RORα1	Die RORα-Proteine („retinoic acid receptor-relatet orphan receptor“) gehören zur Superfamilie der nukleären Hormonrezeptoren (NR) und fungieren als konstitutive ubiquitär exprimierte Transaktivatoren). RORα existiert in vier verschiedenen Formen (α1, α2, α3, z), die durch alternatives Spleißen entstehen. Die Isoformen werden besonders stark in Gehirn, Blut, Skelettmuskeln und Fettzellen exprimiert und kontrollieren u.a. Gene, die am Lipidstoffwechsel beteiligt sind. Mutationen des Transkriptionsfaktors können zu Fettleibigkeit führen. RORα ist zudem in der Lage, mit T3R einen Komplex zu bilden, der dann eine potenzierte transaktivierende Funktion ausübt. Eine Interaktion mit MyoD wurde ebenfalls nachgewiesen.	Lau et al., 2004; Besnard et al., 2002; Raspe et al., 2001; Koibuchi et al., 1999; Lau et al., 1999; Adelmant et al., 1996; Bonneleye et al., 1994; Giguere et al., 1994; Forman et al., 1994; Retnakaran et al., 1994; Beckerandre et al., 1993; Harding et al., 1993; Evans, 1988; Green et al., 1988
SMAD3	SMAD3 wird durch den TGF-β-Rezeptor aktiviert und wandert assoziiert mit SMAD4 in den Zellkern und ist dort in der Lage aktivierend oder reprimierend auf die Transkription von Zielgenen einzuwirken. SMAD3 steuert die Seneszenz von Keratinozyten und wirkt als Tumorsupressor. SMAD3 defiziente Mäuse zeigen eine beschleunigte Wundheilung und eine mukosale Immunschwäche.	Zhu et al., 2004; Vijayachandra et al., 2003; Attisano et al., 2002; Mehra et al., 2002; Kumar et al., 2001; 2003; Datto et al., 1999; Yang et al., 1999; Zhu et al., 1998
STAT5A	STAT-Proteine („signal transducers and activators of transcription“) werden durch eine ganze Reihe von Wachstumsfaktoren, Zytokine und Hormone aktiviert und binden nach Dimerisierung spezifisch an DNA-Sequenzen. STAT5 wurde ursprünglich als Prolactin-induzierter Transkriptionsfaktor im Gewebe der Brustdrüsen identifiziert. STAT5A-Knock-Out Mäuse zeigen Defekte bei der Differenzierung des Brustgewebes, der Entwicklung des Gelbkörperchens und einer damit verbundenen Sterilität und des blutbildenden Systems.	Lin et al., 2000; Akira, 1999; Davey et al., 1999; Leonard et al., 1998; Teglund et al., 1998; Liu et al., 1997; Wakao et al., 1994
TCF-4	Der „cytoplasmic transcription factor 4“ gehört ebenfalls zur Familie der TCF/LEF-DNA-bindenden Transkriptionsfaktoren und ist in der Lage mit β-catenin und LEF-1 eine Komplex zu bilden. TCF-4 bindet an Gene, die Wachstum und Proliferation von Zellen steuern und ist ein Effektor des wnt/ β-catenin-Signalwegs. Die Expression von TCF-4 wird durch p53 reguliert.	Rother et al., 2004; Hülsken et al., 2002, Roose et al., 1999; Cho et al., 1998; Korinek et al., 1997

Tab 4.8: Durch Interspeziesvergleich identifizierte Transkriptionsfaktoren

der Embryogenese und dort gegebenenfalls in einem schmalen Zeitfenster eine Rolle spielen, nicht erfolgen kann. Zwischen Maus und Mensch können zudem nur 32 Gewebe miteinander verglichen werden (Su et al., 2004). Dennoch ist es möglich einige prinzipielle Rückschlüsse zu ziehen. *Dmxl*1 und *DMXL*1 werden demnach in beiden Spezies in allen untersuchten adulten Geweben auf einem grundsätzlich sehr niedrigen Level exprimiert. Diese Beobachtung passt zu den identifizierten basalen Transkriptionsfaktoren, die an den Promotor binden. In einigen Geweben ist jedoch eine deutliche Erhöhung der Transkription über den Median zu beobachten (Tab. 4.9). Erstaunlicherweise konnte für jedes dieser Gewebe im potenziellen Promotor- bzw. Enhancerbereich eine Bindungsstelle für einen Transkriptionsfaktor identifiziert werden, der zu einer verstärkten Transkription von *Dmxl*1 führen könnte. Dennoch muss diese Expressionsanalyse mit einer gewissen Skepsis betrachtet werden, da sich z.B. die Expression von *DMXL*1 in gleichen Geweben unterschiedlicher Datensätze zum Teil erheblich unterscheidet (Abb. 4.9c, d). Es ist anzunehmen, dass aufgrund der geringen Expression von *Dmxl*1/*DMXL*1 die Qualität der verwendeten Chips und der präparierten RNA einen ganz wesentlichen Einfluss auf die beobachteten Ergebnisse hat. Zudem unterscheiden sich die präparierten Gewebe der jeweiligen Datensätze in Alter und Zahl der Spender. Möglicherweise beeinflussen diese Faktoren und die gegebenen Lebensumstände der Spender die Expression von *Dmxl*1/*DMXL*1 ebenfalls, so dass eine differenzierte Aussage basierend auf den gezeigten Datensätzen nur bedingt möglich ist.

Die Daten der Promotor und Enhancer-Analyse können zu einem vorläufigen Modell der regulatorischen Regionen von *Dmxl*1 zusammengefasst werden (Abb. 4.10). Das Modell zeigt einen strukturierten Aufbau der Promotorregion in der Maus. An den als „Core"-Promotor identifizierten Bereich binden die basalen Transkriptionsfaktoren SP1, CPBP und ZF5, die zusammen in der Lage sind, eine Grundexpression von *Dmxl*1 zu gewährleisten. In diesem Bereich bindet zudem MyoD, dass die Assoziation von TFIID an den Präinitiationskomplex massiv stabilisieren kann (Heller et al., 1998). Stromaufwärts des „Core"-Promotors liegen Bindungspositionen für Transkriptionsfaktoren, die eine zell-, gewebs- oder entwicklungsspezifische Expression bzw. Repression ermöglichen. Diese Region erstreckt sich über mehrere hundert Basenpaare, es besteht die Möglichkeit, dass nicht mehr alle dort bindenden Transkriptionsfaktoren dem Promotor zugerechnet werden können, sondern es sich hierbei um einen stromaufwärts gelegenen Enhancer handeln könnte. Eine abschließende Einordnung dieser Region ist aber aufgrund unterschiedlicher Literaturangaben über die Länge von Promotoren nicht möglich (Suzuki et al., 2004, 2002, 2001; Giardine et al., 2003; Ureta-Vidal et al., 2003; Loots et al., 2002). Da sich jedoch die Region der colinearen Sequenzblöcke über den gesamten in Abb. 4.10 gezeigten stomaufwärts vom TSS gelegenen DNA-Abschnitt erstreckt, ist eher davon auszugehen, dass es sich um einen zum Promotor gehörende Bereich handelt. Die Analyse der dort bindenden Transkriptionsfaktoren lässt auf eine sehr komplexe Regulation schließen. Die Untersuchung des ersten Introns von *Dmxl*1 zeigte eine zwischen Maus und Mensch konservierte Region von 3,5 kb Länge, hier wurde ein mögliches Enhancer-Element identifiziert. Die hypothetisch in diesem DNA-Abschnitt bindenden Transkriptionsfaktoren sind teilweise zwischen Maus und Mensch

hochkonserviert, unterscheiden sich aber auch in Zusammensetzung und Abfolge. Besonders auffällig ist die Zahl der möglichen Interaktionen der identifizierten Transkriptionsfaktoren.

Gewebe	Transkriptionsfaktor	Gewebe	Transkriptionsfaktor
Ovarien (M/H)	δEF1, STAT5A	Geruchsorgane (M/H)	LEF-1, Foxi1
Prostata (M/H)	FoxA2, NF-κB, SP1, NFY	Oocyte (M)	TCF4
Harnblase (M)	TEF, NF-κB	B-Zellen (M/H)	RORα1, LEF-1, NF-κB, IK2
Niere (M)	HLF, TEF, Foxi1	Leber (M)	FoxA2, HLF
Fettgewebe (M/H)	RORα1, NFAT, NFY	Retina (M)	RORα1, MEF2, NFAT, MyoD
Knochenmark (M/H)	STAT5A, Sox-5, Sox-9	Milz (M)	NF-κB
Nebenniere (M)	Foxi1	Trachea (M)	SMAD3, FoxF2, LEF-1
verschiedene Hirnbereiche (M/H)	RORα1, FoxA2, TEF, HLF, SF1F, Sox-5, MEF2, Six3, TGIF, NFI	skelettales Muskelsystem (M)	FoxF2, Six3
Herz (M)	NF-κB	befruchtete Eizelle (M)	NF-κB

Tab. 4.9: Gewebe mit erhöhter transkriptioneller Aktivität

Die Tabelle basiert auf den Expressionsdaten der Affymetrix-Chips von Maus und Mensch (Abb. 4.9). Gewebe mit einer über den Median erhöhten Expression von *Dmxl*1 wurden mit allen identifizierten Transkriptionsfaktoren verglichen. Transkriptionsfaktoren, die ein Potenzial für eine Expressionsverstärkung in den jeweiligen Geweben besitzen, wurden diesen zugeordnet. Die Buchstaben M (Maus) und H (Mensch) zeigen, ob sich eine verstärkte Transkription nur in der Maus oder auch im Menschen beobachten lässt. Hierbei ist zu beachten, dass nicht jedes Gewebe der Maus auch im Menschen untersucht wurde.

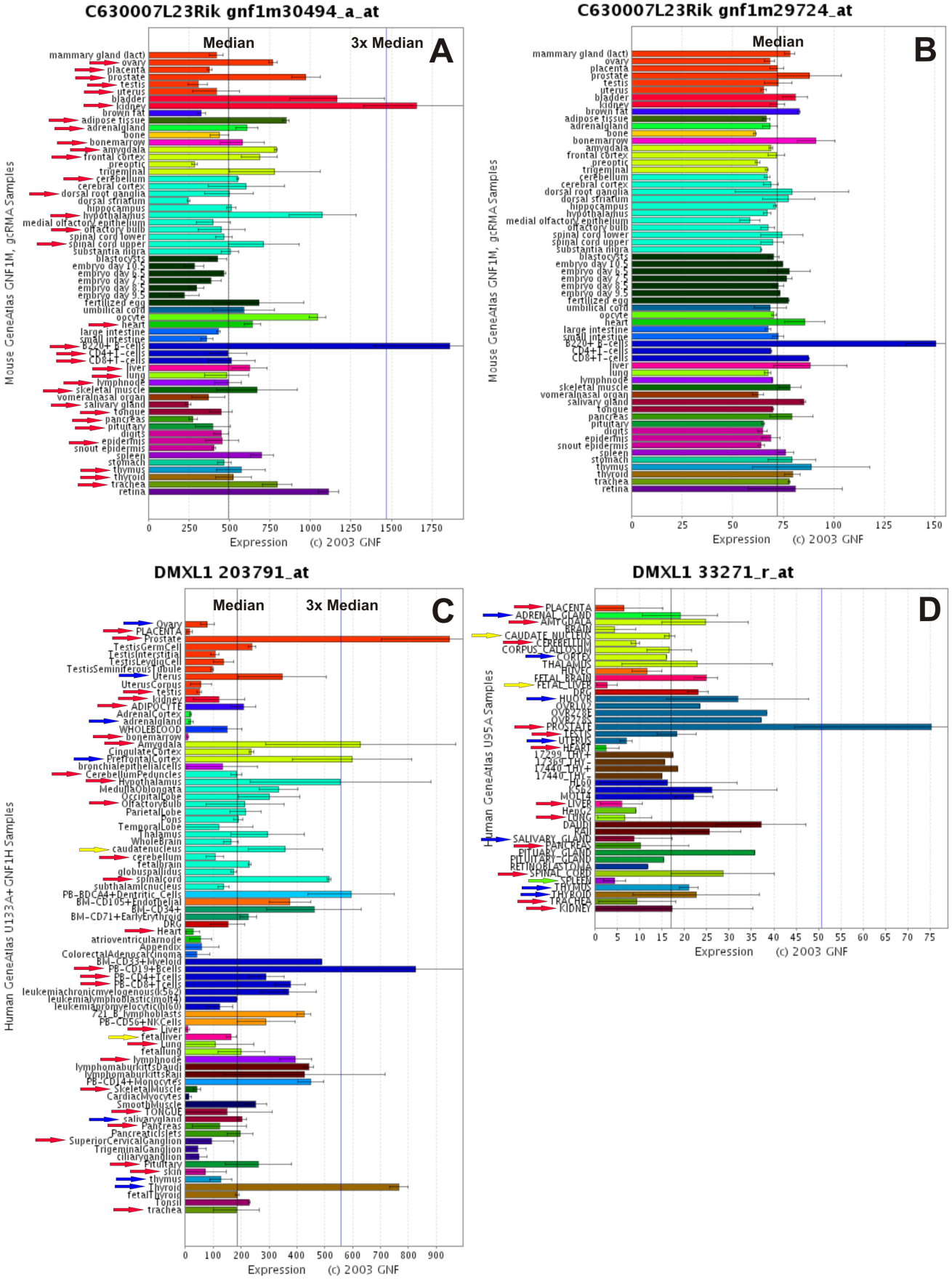

C630007L23Rik gnf1m30494_a_at
Median
3x Median
A
Mouse GeneAtlas GNF1M, gcRMA Samples
Expression
(c) 2003 GNF
C630007L23Rik gnf1m29724_at
Median
B
Mouse GeneAtlas GNF1M, gcRMA Samples
Expression
(c) 2003 GNF
DMXL1 203791_at
Median
3x Median
C
Human GeneAtlas U133A+GNF1H Samples
Expression
(c) 2003 GNF
DMXL1 33271_r_at
D
Human GeneAtlas U95A Samples
Expression
(c) 2003 GNF

Abb. 4.10: Affymetrix-Expressionsspektren des murinen und humanen DmX

Die Expression von *Dmxl*1 wurde im GNF1M-Datensatz mit zwei unabhängigen Proben untersucht, die resultierenden Ergebnisse unterscheiden sich dramatisch (A, B). Mit Hilfe des Affymetrix-Werkzeugs „NetAffx query tool" können die Expressionsspektren den Probensequenzen zugeordnet und ihre Position auf der cDNA von *Dmxl*1 bestimmt werden. Hierbei zeigte sich, dass die Probe gnf1m29724_at auf einem cDNA-Klon basiert, die Probe gnf1m30494_a_at spezifisch für das Experiment entwickelt wurde. Die auf dem cDNA-Klon basierende Probe ist laut Affymetrix in mehrerer Hinsicht suboptimal, daher wurden für die Analyse nur die Daten der spezifischen Probe (A) verwendet. Als Vergleich dient das Expressionsspektrum von *DMXL*1 der Datensätze U133A+GNF1H (C) und U95A (D). Zwischen Maus und Mensch vergleichbare Gewebe sind durch rote Pfeile gekennzeichnet, die Reihenfolge der aufgeführten Gewebe unterscheidet sich zwischen beiden Spezies. Blaue Pfeile zeigen Unterschiede in der Expression zwischen den beiden humanen Datensätzen und dem mit der Maus vergleichbaren Gewebe an. Grüne Pfeile markieren mit der Maus vergleichbare Gewebe, die nur im U95A-Datensatz vorhanden sind. Gelbe Pfeile zeigen Gewebe, die nicht mit der Maus vergleichbar sind und eine unterschiedliche Expression zwischen den beiden humanen Datensätzen zeigen. Der Median zeigt lediglich den Quotienten aller ermittelten Daten, er kann nicht als Maßstab für eine Grundexpression von *Dmxl*1 gelten. Die Daten wurden durch GCRMA („Robust Multi-Chip Analysis") normalisiert, hierbei wird die Intensität der Probe in Abhängigkeit von GC-Gehalt berechnet (Wu et al., 2003), die gezeigten Werteskalen der Expression sind nur innerhalb eines Chips und Experiments direkt vergleichbar. Die Daten für beide Spezies zeigen eine grundsätzlich sehr niedrige Expression von *Dmxl*1/*DMXL*1 und eine etwas erhöte Expression in einigen in Tabelle 4.9 zusammengefassten Geweben.

4.4.6 *Dmxl*1 besitzt möglicherweise zwei alternative Polyadenylierungssignale

Differenzielle Polyadenylierung ist bei höheren Eukaryoten ein weit verbreiteter Mechanismus, um mRNA's unterschiedlicher Länge für unterschiedliche Zwecke zu synthetisieren. An diesem Prozess sind unter anderem alternative Polyadenylierungssignale im 3'-UTR beteiligt, jedes von ihnen mit einer unabhängigen spezifischen Stärke des Signals (Beaudoing et al., 2000, Edwalds-Gilbert et al., 1997). Von alternativer Polyadenylierung wird dann gesprochen, wenn zwei oder mehr vorhandene PolyA-Signale im 3'-UTR zu zwei oder mehr unterschiedlich langen Isoformen eines Transkriptes führen (Legendre et al., 2003). Circa 22% aller mRNA-Transkripte unterliegen einer alternativen Polyadenylierung (Beaudoing et. al., 2000), meist erfolgt diese gewebs- bzw. entwicklungsspezifisch (Edwalds-Gilbert et al., 1997). In der vorliegenden Arbeit konnten für *Dmxl*1 zwei mögliche Polyadenylierungssignale identifiziert werden. EST-Datenbanksuchen und eigene Sequenzierungen von cDNA-Klonen zeigten, dass die meisten murinen Transkripte mit dem distalen PolyA-Signal (polyA1, Kapitel 3.2.9) enden. Es handelt sich hierbei um ein nahezu „perfektes" Signal (Abb. 3.30), mit der charakteristischen Hexamersequenz AAUAAA (Proudfood et al., 1976; Wickens, 1990) und drei U(G)-reichen Elementen in einem idealen Abstand zu dieser Hexamersequenz (Gil et al., 1984,1986; McDevitt et al., 1984; McLauchlan et al., 1985; Chou et al., 1994; MacDonald et al., 1994; Takagaki et al., 1997; Legendre et al., 2003; Zarudnaya et al., 2003). Ein Vergleich des murinen Sequenzabschnitts mit dem genomischen Bereich des humanen Polyadenylierungssignals zeigt, dass das erste PolyA-Signal zwischen beiden Spezies hochkonserviert ist (Abb. 3.31). Für den

Sequenzbereich vor polyA1 konnten weder in Maus noch in Mensch Hinweise auf ein mögliches weiteres Polyadenylierungssignal gefunden werden. Es ist daher davon auszugehen, dass es sich bei polyA1 um das erste PolyA-Signal von *Dmxl*1 handelt.
Bei der Sequenzierung diverser cDNA-Klone zeigte sich, dass der Klon IMAGp998J193130Q2 nicht mit polyA1 endet, sondern erst 1220 bp später einen PolyA-Schwanz aufweist (polyA2). Für *Dmxl*1 existieren nur vier EST-Datenbankeinträge mit Transkripten, die nicht auf polyA1 enden, der sequenzierte cDNA-Klon ist einer von ihnen, und nur dieser endet auf polyA2. Die Untersuchung des entsprechenden genomischen Sequenzabschnitts und der cDNA-Klone per Auge und mit Hilfe diverser Programme des Internets zeigt ein weiteres, jedoch nicht ganz so eindeutiges Polyadenylierungssignal wie polyA1. Die Hexamersequenz dieses Signals entspricht der ebenfalls häufigen Abfolge AUUAAA, allerdings befindet sich diese in Position -20 – -26 und damit im Grenzbereich für ein funktionsfähiges PolyA-Signal (Chen et al., 1995; Colgan et al., 1997). Die weiteren identifizierten Elemente dieses Polyadenylierungssignals sprechen wiederum dafür, dass es sich hier um ein reales Signal handelt. Vor allem das C/A-Dinukleotid-Motiv in der Position -1/+1 muss hier als Hinweis auf ein existierendes PolyA-Signal gewertet werden (Chen et al., 1995), eine zufällige Polyadenylierung in dieser Position ist eher unwahrscheinlich. EST-Datenbanksuchen mit genomischen Sequenzdaten des humanen *DMXL*1, die downstream der polyA1-Sequnz liegen, zeigen ebenfalls alternative Transkripte (Abb. 3.29 b), die jedoch teilweise an unterschiedlichen Positionen enden. Datenbankeinträge weisen darauf hin, dass es sich hierbei um *DMXL*1-Transkripte handelt, da keine signifikante Sequenzhomologie zu der genomischen Region von *DMXL*2 festgestellt werden konnte. Diese Variationen sind vermutlich auf eine Hybridisierung der bei der Herstellung der cDNA-Klone verwendeten Primer (PolyT) mit einer A/T-reichen Region (Alu-Element) mit bis zu 21 Adeninnukleotiden zurückzuführen, die sich in der genomischen Sequenz direkt im Anschluss der EST-Datenbankeinträge befindet. Es ist also wahrscheinlich, dass es sich hierbei um eine genomische Kontamination durch Fehlpriming handelt. In der Maussequenz wurde eine solche Anhäufung von Adeninnukleotiden im Bereich von polyA2 nicht beobachtet, ein Mispriming ist für die Maus also eher unwahrscheinlich. Die Überprüfung des 3'-UTR der genomischen humanen Sequenz mit Programmen zur Identifikation von PolyA-Signalen zeigt dennoch zwei mögliche Signale in dieser Region. Das eine befindet sich 36 bp stromabwärts der A/T-reichen Region und zeigt die charakteristische Hexamersequenz AAUAAA, ein C/A-Dinukleotid sowie drei U(G)-reiche Regionen, jedoch endet keiner der humanen EST-Klone in dieser Region. Das zweite erkannte Signal befindet sich in der 5'-terminalen A/T-reichen Region eines in der humanen genomischen Sequenz befindlichen Alu-Elements. Hier ist durch eine Transversion die Hexamersequenz AAUAAA entstanden, welches von den unterschiedlichen Programmen als PolyA-Signal erkannt wird. Zwar wird dieses Signal von den verschiedenen Programmen mit einer hohen Wahrscheinlichkeit bewertet, die EST-Daten sind jedoch nicht eindeutig genug, um eine genomische Kontamination auszuschließen. Gegen ein zweites Polyadenylierungssignal in der humanen Sequenz spricht vor allem, dass diese Region zwischen Maus und Mensch nicht konserviert ist, alle bisherigen Daten von

*Dmxl*1/*DMXL*1 zeigten einen hohen Konservierungsgrad aller wichtigen Sequenzelemente.

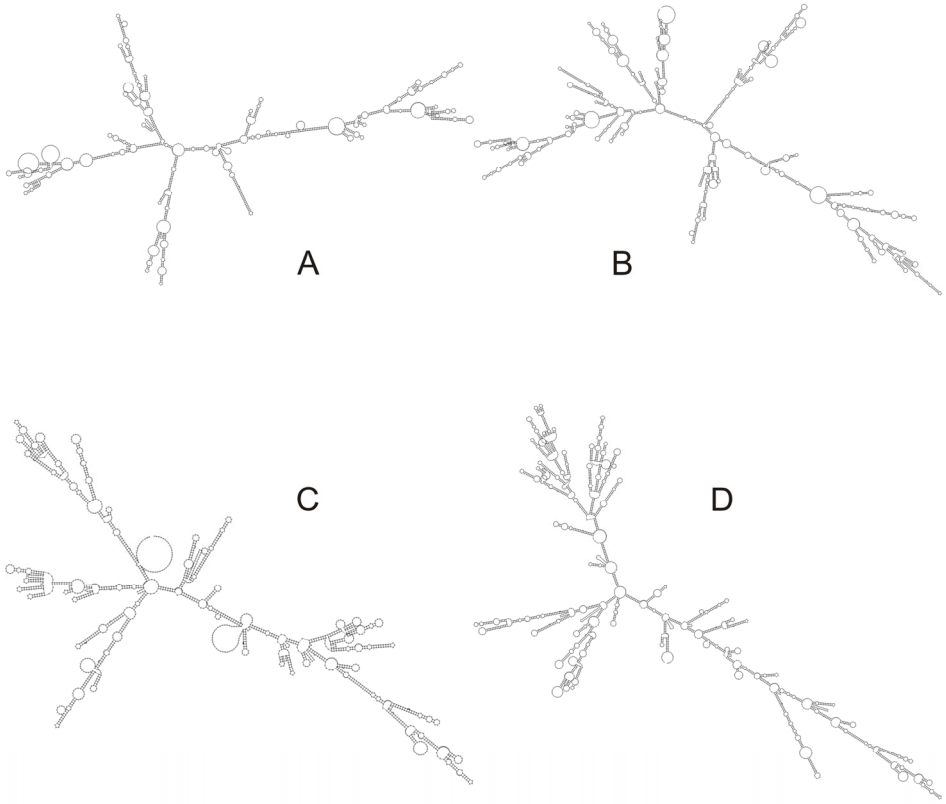


Abb. 4.11: Sekundärstruktur des 3'-UTR von Maus und Mensch.

Die Sekundärstruktur des 3'-UTR beider Spezies wurde für jeweils beide potenziellen Polyadenylierungssignale berechnet. Die Faltungen sind äußerst komplex, zeigen aber keine signifikante Ähnlichkeit. Die berechneten freien Energien sind jedoch fast identisch. A: Sekundärstruktur der murinen Sequenz, die mit polyA1 endet. B: Sekundärstruktur der murinen Sequenz die mit polyA2 endet. C: Sekundärstruktur der humanen Sequenz, die mit polyA1 endet. D: Sekundärstruktur der humanen Sequenz, die mit polyA2 endet.

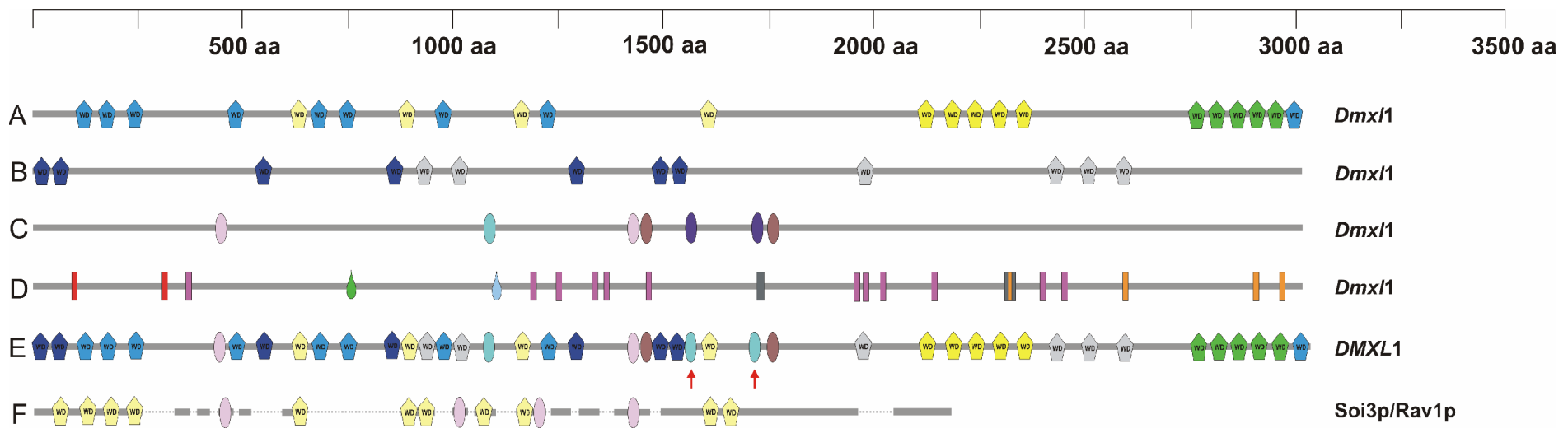
Das dies bei einem zweiten Polyadenylierungssignal nicht der Fall sein soll, ist eher unwahrscheinlich. Aufgrund dieser Daten ist daher eher davon auszugehen, dass in der humanen genomischen Region von *DMXL*1 kein zweites PolyA-Signal existiert.
Die DNA-Sequenz zwischen dem ersten und zweiten murinen Polyadenylierungssignal (Abb. 3.30) wurde mit diversen Programmen (Kapitel 3.29) untersucht, ein weiteres funktionelles Signal konnte jedoch nicht identifiziert werden. Neben solchen Signalen kann aber auch die Sequenz selbst eine Funktion ausüben, indem sie durch die Ausbildung von Sekundärstrukturen zur Stabilität der Transkripte beiträgt (Wu et al.,

2004; Hans et al., 2000). Mit dem Programm Vienna RNA secondary structure server (Hofacker, 2003; Hofacker et al., 2002) wurde die wahrscheinliche Sekundärstruktur der Maus- und Mensch-Sequenz für jeweils beide denkbaren Polyadenylierungssignale ermittelt. Die Faltung der Maussequenz mit polyA1 und polyA2 erzeugt äußerst komplexe Sekundärstrukturen (Abb. 4.10a, b) mit einer freien Energie von -467,46 kcal/mol (polyA1) bzw. -752,51 kcal/mol (polyA1 + polyA2). Auch die untersuchte humane Sequenz bildet komplizierte Sekundärstrukturen (Abb. 4.10c, d), die ermittelte freie Energie liegt bei -460,35 kcal/mol (polyA1hum) und -790,28 kcal/mol (polyA1hum + polyA2hum). Dass sich die Werte für das erste Polyadenylierungssignal aufgrund der Homologie zwischen Maus und Mensch ähneln, war zu erwarten, dass jedoch die freien Energien für die potenziellen zweiten Polyadenylierungssignale beider Spezies ebenfalls sehr ähnlich sind, konnte auch in Hinblick auf die völlig unterschiedlichen Sequenzen und Sekundärstrukturen nicht unbedingt erwartet werden. Dies könnte als Hinweis auf ein vorhandenes zweites Polyadenylierungssignal auch im Menschen gewertet werden.

4.4.7 Strukturelle Motive in der abgeleiteten Aminosäuresequenz von Dmxl1

Um weitere Hinweise auf eine mögliche Funktion des Dmxl1-Proteins zu erhalten, wurde die abgeleitete Aminosäuresequenz mit diversen bioinformatischen Methoden untersucht. Basierend auf den Arbeiten von Kraemer et al. (1998, 2000) und Weil (2000) an *DmX*, *CpY* und *DMXL*1 wurde die abgeleiteten Proteinsequenzen der homologen Gene mit dem Programm Superfamily (Madera et al., 2004) überprüft. Es konnten die abgeleiteten Aminosäuresequenzen des murinen und humanen Dmxl1/DMXL1 eindeutig als „WD40 repeat-Protein“ (Neer et al., 1994) eingestuft werden. Damit kann *Dmxl*1 der Klasse der WD-Repeat Proteine zweifelsfrei zugeordnet werden.
Die Untersuchung der abgeleiteten Aminosäuresequenz des DmX- und DMXL1-Proteins von Kraemer et al. (2000,1998) per Auge und Datenbanksuchen zeigte die enorme Anzahl von 30 bzw. 28 WD-Wiederholungseinheiten. Dies ist die höchste jemals in einem Protein gefundene Anzahl. Allerdings zeigten diese Bereiche einen höchst unterschiedlichen Konservierungsgrad bezüglich der WD-Repeat Konsensussequenz (Smith et al., 1999). Die bioinformatischen Ansätze, die eine bewertende Analyse ermöglichen, waren zu diesem Zeitpunkt noch unzureichend entwickelt. Mit einer rasch zunehmenden Anzahl identifizierter WD-Proteine wurden auch die Analyseprogramme deutlich verbessert. Neben der Untersuchung per Auge und Datenbanksuche wurden in der vorliegenden Arbeit die Programme SMART (Letunic et al., 2002; Schultz et al., 1998), PSA (Stultz et al., 1993; White et al., 1994) und REP (Andrade et al., 2000) verwendet. Insgesamt wurden auf diese Weise 36 WD-Wiederholungseinheiten in der abgeleiteten Aminosäuresequenz von *Dmxl*1 identifiziert. Hiervon zeigen 24 Wiederholungseinheiten eine sehr gute Homologie zu der Konsensussequenz und konnten neben einer hohen Bewertung durch entweder mindestens zwei Analyseprogramme oder durch ein Analyseprogramm und

Datenbanksuchen verifiziert werden. Weitere 12 WD-Wiederholungseinheiten zeigen eine weniger gute Konservierung und wurden entweder nur durch eines der Programme, per Auge, durch Interspeziesvergleich oder Datenbanksuchen identifiziert. Das Programm REP erscheint bei dieser Analyse besonders glaubwürdig, da es alle von SMART oder PSA identifizierten WD-Repeats erkannte, SMART und PSA aber jeweils unterschiedliche Ergebnisse lieferten. Zusätzlich wurden sechs weitere WD-Wiederholungeinheiten mit einer hohen Wahrscheinlichkeit von REP erfasst, die durch keine andere verwendete Methode identifiziert werden konnten. Eine Festlegung auf eine bestimmte Anzahl dieser Einheiten in der abgeleiteten Aminosäuresequenz ist aber aufgrund der heterogenen Ergebnisse nicht möglich. Es kann davon ausgegangen werden, dass das Dmxl1-Protein zwischen 24 und 36 WD-Wiederholungseinheiten besitzt. Auch bei der Identifikation der WD-Repeats zeigt sich, dass neben der Weiterentwicklung bioinformatischer Werkzeuge die Kombination verschiedener verfügbarer Programme zusammen mit einem Interspeziesvergleich die Signifikanz, Glaubwürdigkeit und Stringenz einer Vorhersage deutlich verbessern kann. Die Verteilung aller identifizierten WD-Repeats ist in Abb. 3.29 und 4.11 wiedergegeben. Da es sich bei dem Programm REP um ein Werkzeug handelt, das auf eine umfangreiche Datenbank verschiedenster Protein-Repeats zugreift, wird standardmäßig eine zu untersuchende Sequenz auf alle in dieser Datenbank vorhandenen Motive überprüft. Hierbei werden auch solche Motive angegeben, die in anderen Proteinen in vielfacher Kopie vorhanden sind, in der untersuchten Sequenz jedoch nur einmal vorkommen. Bei dieser Analyse zeigte sich, dass die abgeleitete Aminosäuresequenz des Dmxl1-Proteins neben den WD-Wiederholungseinheiten noch weitere Repeat-Motive enthält. Dies ist insofern erstaunlich, da bisher nicht bekannt war, das DmX noch andere Klassen von Protein-Repeats einschließt. Insgesamt konnten sieben zusätzliche repetitive Einheiten anderer Superfamilien in der Dmxl1-Aminosäuresequenz identifiziert werden (Kapitel 3.2.11). Die Einordnung dieses Befundes ist jedoch äußerst schwierig, da bisher kaum Proteine beschrieben wurden, die gleich mehrere Klassen repetitiver Motive in sich vereinen. Meist handelt es sich dabei um eng verwandte Strukturen mit gemeinsamen Vorläufern wie HEAT (AAA, ADB und IMB-Subklassen) und ARM („armadillo“; Pfeifer et al., 1994), die zusammen in Proteinsequenzen gefunden werden (Andrade et al., 2001a), oder um Kombinationen von zwei unterschiedlichen Repeats (Andrade et al., 2001b). Die Zusammensetzung eines Proteins aus fünf verschiedenen Repeat-Klassen wurde bisher nicht beschrieben. Zudem konnten die neuen Repeat-Motive nur durch das Programm REP identifiziert werden, eine Überprüfung dieser Ergebnisse mit weiteren Programmen war nicht möglich. Um die Genauigkeit dieser Vorhersage zu erhöhen, wurde die REP-Analyse auch mit der abgeleiteten Aminosäuresequenz von *H.sapiens* (Abb. 4.11) durchgeführt und nur die zwischen beiden Spezies konservierten Motive in die Interpretation einbezogen. Hier zeigte sich, dass fünf Motive zwischen Maus und Mensch konserviert sind, zwei Repeat-Einheiten werden von REP als LRR-Repeats (Leucine Rich Repeats) erkannt und befinden sich in gleicher Position der in der Maus identifizierten Kelch-Motive. Dies ist insofern erstaunlich, da sich beide Motive erheblich unterscheiden. LLR-Wiederholungseinheiten bilden bogenförmige bzw. hufeisenförmige Strukturen



- von SMART, PSA und REP erkannte WD-Wiederholungseinheit
- von SMART und REP erkannte WD-Wiederholungseinheit
- von REP und durch Datenbanksuchen erkannte WD-Wiederholungseinheit
- von REP erkannte WD-Wiederholungseinheit
- durch Datenbanksuchen erkannte WD-Wiederholungseinheit
- von REP erkannte LRR-Wiederholungseinheit
- von REP erkannte PFTA-Wiederholungseinheit
- von REP erkanntes Armadillo-Repeat
- von REP erkanntes KELCH-Repeat
- von PSORT II identifiziertes peroxisomales Signal-Motiv
- von PATMATMOTIVES, MotifScan und PPsearch erkanntes Subtilase_ASP-Motiv
- Von TMpred, TOPPRED und TM-Finder identifizierte Transmembrandomäne
- von SMART erkannte Region niedriger Komplexität
- von PSORT II identifiziertes Dileucin-Signal

Abb. 4.12: Konservierte Motive in den abgeleiteten Aminosäuresequenzen von *Dmxl*1, *DMXL*1 und Soi3p/Rav1p

Um die Darstellung übersichtlicher zu gestalten, wurden die Positionen der identifizierten Motive in der abgeleiteten Aminosäuresequenz von *Dmxl*1 in vier Bereiche unterteilt. In A sind alle WD-Reapeats, die von mehreren Programmen identifiziert wurden, gezeigt. In B sind zusätzliche WD-Wiederholungseinheiten positioniert, die nur von einem Programm oder durch Datenbanksuchen erkannt wurden. In C sind die Positionen von anderen Repeats, die von REP erkannt wurden (Kap. 3.2.11), gezeigt. In D sind weiter Motive und ihre Lage in der abgeleiteten Aminosäuresequenz, die in Dmxl1 von diversen Programmen gefunden wurden (Kap. 3.2.11), gezeigt. Aus Platzgründen wurde auf die Darstellung dieser Motive in der ebenfalls gezeigten Aminosäuresequenz von *DMXL*1 (E) weitgehend verzichtet. Die Motive wurden, bis auf einige Abschnitte niederiger Komplexität in identischer Position auch in DMXL1 gefunden. DMXL1 unterscheidet sich von Dmxl1 vor allem in zwei Repeats (rote Pfeile). Hier wurde in der humanen Aminosäuresequenz in gleicher Position von REP jeweils ein PFTA-Repeat anstatt eines KELCH-Motifs identifiziert. In F sind die zu Dmxl1 und DMXL1 homologen Abschnitte von Soi3p/Rav1p gezeigt. Hier sind die identifizierten WD-Wiederholungseinheiten und andere, von REP erkannte Repeats, aufgeführt.

und sind in Signaltransduktion, Zelladhäsion, DNA-Reparatur, Rekombination, Transkription, RNA-Prozessierung, Immunantwort und vielen weiteren Prozessen involviert (Enkhabayar et al., 2004; Kajava et al., 2002; Kobe et al., 2001). Kelch-Wiederholungseinheiten bilden im Gegensatz dazu, genau wie WD-Repeats, eine Propeller-Struktur aus. Auch die biologischen Funktionen der Proteine mit Kelch-Motiven unterscheiden sich hier von LLR-Repeat-Proteinen, so werden Kelch-Proteine vor allem bei Zellmorphologie und Zellorganisation sowie bei Aktin-assoziierten Proteinen gefunden (Adams et al., 2000). In *Mus musculus* werden fünf der sieben nicht-WD-Repeats im zentralen Abschnitt des Dmxl1-Proteins gefunden. In diesem Bereich weisen die meisten der identifizierten WD-Wiederholungseinheiten eine deutlich geringere Homologie zu der WD-Konsensussequenz auf und konnten entweder nur durch ein Vorhersageprogramm oder durch Datenbanksuchen aufgefunden werden. Dieser zentrale Abschnitt der Aminosäuresequenz von Dmxl1 und DMXL1 zeigt wiederum eine hohe Übereinstimmung zu dem offenen Leserahmen YJR033c (*SOI*3/*RAV*1) der Hefe (der Proteinvergleich von Maus und Hefe befindet sich auf dem beigefügten Datenträger; Kraemer et al., 2000), YJR033c enthält sieben oder acht WD-Wiederholungseinheiten (Sipos et al., 2004, Kane et al., 2003) und nach eigenen Analysen sogar bis zu elf WD-Wiederholungseinheiten (davon zehn in konservierter Position zu Maus und Mensch) und ebenfalls Repeats anderer Protein-Superfamilien (Abb. 4.11). YJR033c kodiert für das Protein Rav1p/Soi3p (Sipos et al., 2004; Smardon et al., 2002; Seol et al., 2001), das eine wichtige Rolle bei der Zusammensetzung der vakuolären ATPase, der Regulation des endozytotischen Transfers früher Endosomen zur Vakuole und der Lokalisation von Transmembranproteinen im Trans-Golgi Netzwerk spielt (Sipos et al., 2004; Seol et al., 2001). Die abgeleitete Aminosäuresequenz umfasst 1357 aa (Zagulski et al., 1995) und ist damit deutlich kleiner als Dmxl1. Aufgrund der Ähnlichkeit kann angenommen werden, dass beide Proteine in den homologen Positionen möglicherweise gemeinsame funktionale Domänen besitzen und die Repeats hier eine Rolle spielen (Abb. 4.11). Zwischen Maus, Mensch und Hefe sind zwei LRR-Repeats in ihrer Position konserviert. Weiter werden

zwischen Maus und Mensch zwei Armadillo- und ein PFTA-Motiv in homologer Position gefunden. Armadillo-Repeats formen Super-Helices aus Helices und werden vor allem in Proteinen gefunden, die eine Rolle bei der intrazellulären Signaltransduktion und cytoskelettalen Regulation spielen. Proteine mit PFTA-Motiven (auch PPTA, „protein farnesyl transferase α-subunit repeats“; Boguski et al., 1992) gehören zur Superfamilie der Tetratricopeptid-Repeats (TPR). Proteine mit TPRs sind bei der Regulation des Zellzyklus, mitochondrialem und peroxisomalen Proteintransport, Neurogenese, Proteinfaltung und Transkriptionskontrolle involviert. Es konnten noch weitere Repeats identifiziert werden, die zwar zu unterschiedlichen Superfamilien gehören, aber in ihrer Position innerhalb der abgeleiteten Aminosäuresequenz zwischen allen drei Spezies konserviert sind. Von allen identifizierten nicht-WD-Repeats zeigen die LRR-Wiederholungseinheiten die größte Homologie zur Konsensussequenz dieser Superfamilie. Eine Motivsuche in verschiedenen Datenbanken mit dem Programm SMART zeigte, dass bisher bei elf angenommenen Proteinen eine Kombination von LRR- und WD-Repeats gefunden wurden, sechs dieser Proteine enthalten nur drei oder weniger LRR-Motive. Die Kombination aus LRR und Armadillo in einem Protein ist bei drei abgeleiteten Aminosäuresequenzen gezeigt, eine Zusammensetzung aus WD-, LRR- und Armadillo-Motiven wurde bisher für kein Protein beschrieben (Daten nicht gezeigt). Grundsätzlich muss dieses Ergebnis mit einer gewissen Skepsis betrachtet werden. Auch wenn diese Motive in der abgeleiteten Aminosäurersequenz gefunden werden, liegt ihre Natur vor allem in einer tandemartigen Abfolge der Wiederholungseinheiten, die dadurch sekundäre bzw. tertiäre Strukturen formen können und häufig erst so ihre funktionelle Bedeutung erhalten (Murray et al., 2004; Andrade et al., 2001; Kajava, 2001). Diese Voraussetzung wird für das Dmxl1-Protein nach strukturellen Berechnungen nur durch einige WD-Wiederholungseinheiten erfüllt, die anderen identifizierten Klassen sind jeweils durch andere Repeats und größere Entfernungen räumlich getrennt (Abb. 4.11). Dennoch ist es denkbar, dass diese Motive im korrekt gefalteten Dmxl1-Protein in räumliche Nähe gelangen und so funktionelle Einheiten bilden, die das Bindungsspektrum von Dmxl1 bei seiner angenommenen Funktion im zellulären Transport (Kapitel 4.4.9) erhöhen.

Bei der Analyse der abgeleiteten Aminosäuresequenz des Dmxl1-Proteins auf mögliche Transmembrandomänen (TmD, Kapitel 3.2.11) zeigen die hierzu verwendeten Programme sehr heterogene Ergebnisse. Besonders auffällig ist dabei, dass die meisten identifizierten TmD innerhalb von WD-Wiederholungseinheiten liegen, ganz offensichtlich sind die Ladungseigenschaften der β-Faltblätter denen von Transmembrandomänen recht ähnlich. Es kann jedoch davon ausgegangen werden, dass es sich hierbei um falsch Positive Ergebnisse handelt, da es sich bei diesen WD-Repeats um „sichere“ Voraussagen handelt. Nach Ausschluss dieser vermutlichen falsch Positiven wird von drei Programmen jeweils eine hochsignifikante TmD erkannt. TMpred und TOPPRED (von Heijne et al., 1992) identifizierten eine Transmembrandomäne in gleicher Position (2307-2331 aa), allerdings befindet sich diese innerhalb eines Blocks aus mindestens fünf WD-Wiederholungseinheiten. Das Programm TM-Finder (Deber et al., 2001) konnte eine wahrscheinliche Transmembrandomäne in der Position 1718-1729 aa identifizieren, also im zentralen

Abschnitt der abgeleiteten Aminosäuresequenz von Dmxl1, der weitgehend frei von WD-Wiederholungseinheiten ist. Diese Domäne überlappt nicht mit den anderen identifizierten Repeats. Für Rav1p/Soi3p wird von TM-Finder in ähnlicher Position wie in Maus und Mensch, ebenfalls mit hoher Signifikanz, eine TmD vorausgesagt (Abb. 4.11). Die anderen verwendeten Programme konnten keine Transmembrandomänen identifizieren. Aufgrund dieser sehr unterschiedlichen Ergebnisse und der vielen falsch Positiven ist es nicht möglich festzustellen, ob und in welcher Position Dmxl1 Transmembrandomänen besitzt. In PubMed- und Entrez-Recherchen konnten jedoch zahlreiche Proteine gefunden werden, die sowohl WD-Repeats als auch Transmembrandomänen besitzen, eine Kombination dieser strukturellen Merkmale ist daher nicht ungewöhnlich.
Interessanterweise konnten im C-terminalen Abschnitt von Dmxl1 durch PSORT II fünf Dileucin-Motive gefunden werden, hiervon müssen drei als signifikant eingestuft werden. Diese Motive werden häufig in membrangebundenen oder G-Protein gekoppelten Proteinen gefunden und spielen dort eine wichtige Rolle bei der Proteinverteilung im Trans-Golgi Netzwerk bzw. der Internalisierung in frühe und späte Endosomen und Lysosomen (Storch et al., 2004; Kil et al., 2000; Geisler et al., 1998; Haft et al., 1998; Warren et al., 1998; Gabilondo et al., 1997; Heilker et al., 1996; Mauxion et al., 1996; Dietrich et al., 1994; Letourneur et al., 1992). Dies könnte ein wichtiger Hinweis auf die Funktion des Dmxl1-Proteins sein und wird in Kapitel 4.4.9 diskutiert. Neben den Dileucin-Motiven wurden von PSORT II im N-terminalen Bereich zwei nukleäre Signale außerhalb der gefundenen WD-Wiederholungseinheiten identifiziert, ihre Funktionalität muss als hoch eingestuft werden. In Position 819 aa wird darüber hinaus ein peroxisomales Signal (PTS2, „peroxisomal targeting signal“ Albertini et al., 1997) vorausgesagt. Dieses Signal entspricht zwar sehr gut der redefinierten Konsensussequenz von PTS2 (Abb. 4.12; Petriv et al., 2004), liegt aber innerhalb einer putativen WD-Wiederholungseinheit, die nur durch Datenbanksuche identifiziert werden konnte und eine schwächere Homologie zu der WD-Konsensussequenz zeigt. Es kann daher nicht mit Sicherheit gesagt werden, ob es sich hierbei um ein „echtes“ PTS2-Signal handelt. Zumal in der Literatur einige WD-Repeat Proteine und Proteine mit Tetratricopeptid-Repeats beschrieben sind, die diese Signale mit ihrer Propellerstruktur bzw. Superhelices erkennen und als Transporter die signaltragenden Proteine zu den Peroxisomen leiten, auch unter Zuhilfenahme weiterer Proteine, selbst aber diese Signale nicht tragen (Mukai et al., 2002; Zhang et al., 1996, 1995). Auf der anderen Seite werden Peroxisomen vor allem in Zellen gefunden, die Lipide und/oder Steroidhormone synthetisieren, metabolisieren oder speichern, hierzu gehören z.B. auch Fettzellen und Leydig-Zellen. Ein möglicher Zusammenhang zu dem beobachteten Phänotyp der chimären Mäuse ist daher durchaus denkbar. Ein solcher Zusammenhang, wenn auch nur sekundär, würde jedoch zumindest eine zeitweilige cytoplasmatische Lokalisation von Dmxl1 voraussetzen. Von PSORT II wird für Dmxl1 eine 69,6%ige (Human: 65,2%) nukleäre Lokalisation berechnet. Dieser Wert ist jedoch ebenfals nicht eindeutig genug, um hier eine klare Aussage treffen zu können. Das neu entwickelte spezialisierte Programm ESLpred liefert hierzu eine nicht wesentlich eindeutigere Vorhersage. Für die Proteine Dmxl1, DMXL1 und F54E4.1 wird zwar eine nukleäre Lokalisation mit einer Genauigkeit von 98% bestimmt, für das homologe aus

Drosophila kann eine nukleäre Lokalisation mit nur noch 53% Genauigkeit vorrausgesagt werden, CPY hingegen wird mit einer Genauigkeit von 94% als cytoplasmatisches Protein definiert. Es ist äußerst unwahrscheinlich, dass die homologen Proteine ihre Funktion in einem jeweils völlig unterschiedlichen zellularen Kompartiment ausüben. Demzufolge kann auch diese Analyse nicht ohne Skepsis betrachtet werden. Grundsätzlich jedoch scheint eine primär nukleäre Lokalisation der besprochenen homologen Proteine wahrscheinlicher zu sein. Ein weiteres Strukturmerkmal ist ein SUBTILASE_ASP-Motiv. Bei diesen 10 Aminosäuren langen Abschnitt handelt es sich um eine Sequenz mit 100%iger Homologie zum aktiven Zentrum verschiedener Serinproteasen. Das Motiv überlappt mit zwei Aminosäuren mit einem identifizierten PFTA-Repeat. Da jedoch die Faltungsstruktur in diesem Abschnitt von Dmxl1 unklar ist, kann nicht gesagt werden, ob dieser möglicherweise katalytische Bereich für Substrate überhaupt zugänglich ist.

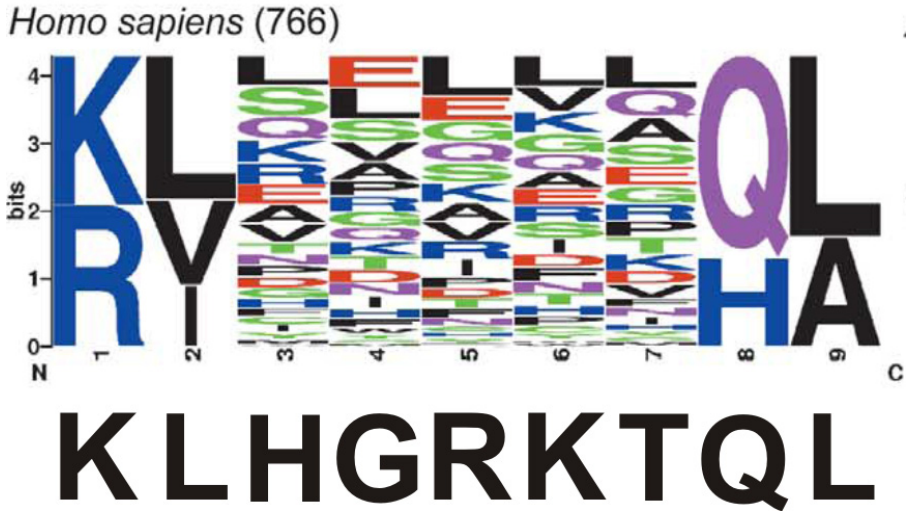


Abb. 4.13: Vergleich des identifizierten peroxisomalen Motivs mit der redefinierten humanen Konsensussequenz

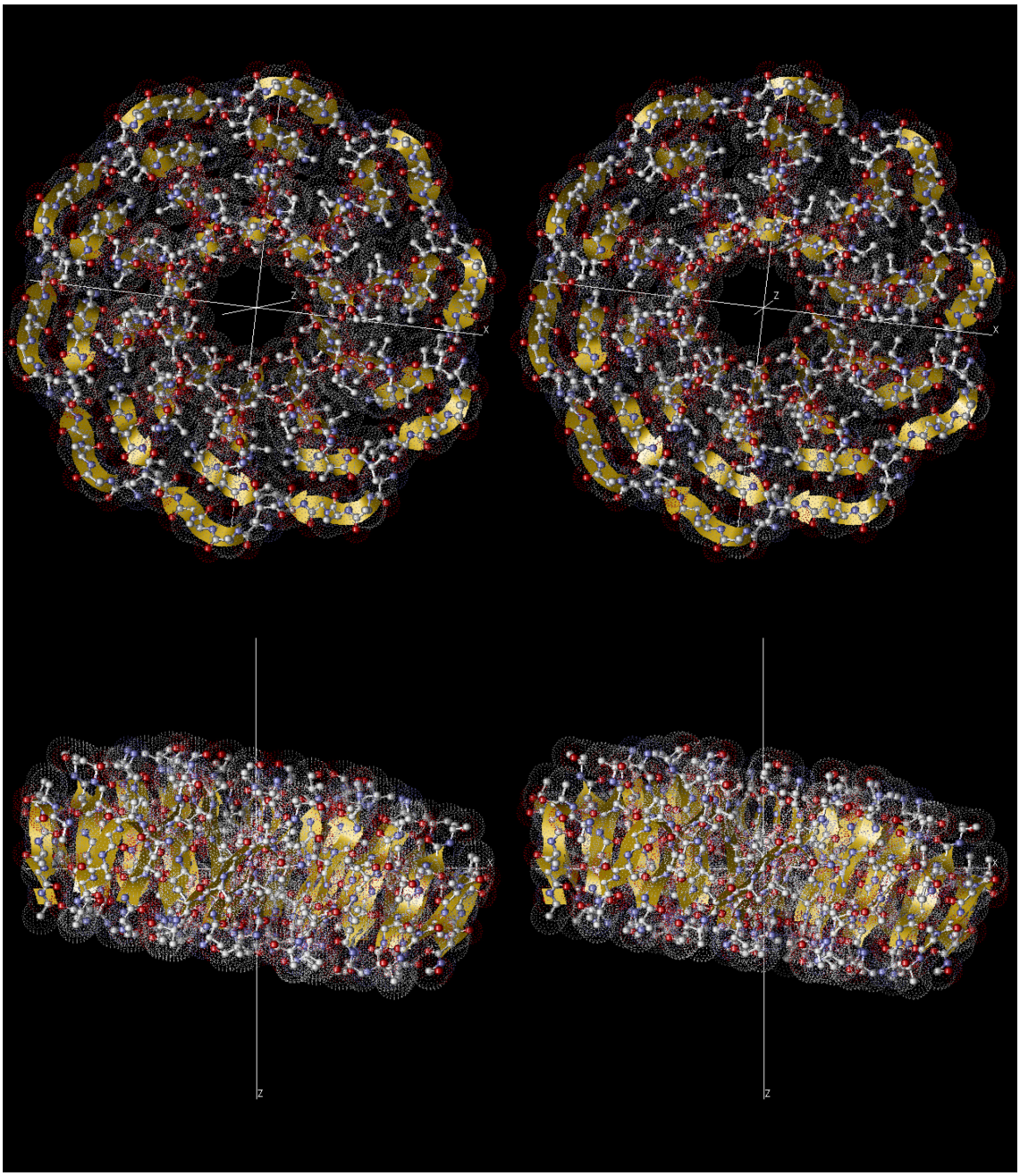
Neuere Untersuchungen zum PTS2-Signal von Petriv et al. (2004) erlauben eine verbesserte Definition peroxisomaler Motive. Gezeigt sind die Konsensussequenz des humanen PTS2-Signals und die statistische Verteilung der einzelnen Aminosäuren aus 766 untersuchten Proteinen. Die Zusammensetzung der Aminosäuren eines PTS2-Signals unterscheidet sich zwischen den Tierstämmen. Die untersuchten humanen Proteine können zumindest als repräsentativ für die Mammalia gelten. Das identifizierte Signal befindet sich als Buchstabenabfolge unter der Grafik der statistischen Verteilung.

4.4.8 Rekonstruktion des C-terminalen WD-Propeller und Diskussion möglicher Faltungsstrukturen von Dmxl1

Für eine direkte und spezialisierte *in silico* Rekonstruktion der WD-Wiederholungseinheiten steht derzeit weltweit nur das Programmpaket PSA am

BMERC (BioMolecular Engineering Research Center) zur Verfügung, Faltungsrekonstruktionen sind allerdings nur mit den von PSA identifizierten WD-Repeats möglich. Für jede der zehn identifizierten WD-Wiederholungseinheiten wurden von PSA drei β-Faltblattbereiche (Strang A-C) vorhergesagt, der vierte β-Strang (Strang D) konnte von PSA nicht klar eingegrenzt werden, da diese β-Stränge die größte Variabilität zeigen. D-Stränge befinden sich zudem innerhalb der variablen Region 1A und 1B (Kapitel 1.3.3), deren Länge zwischen 20 und 150 Aminosäuren schwankt (Li et al., 2001; Smith et al., 1999; Chothia et al., 1997; Neer et al., 1994). Die Vorhersage steht in guter Übereinstimmung mit der in der Einleitung beschriebenen Konsensusstruktur. Als Faltungsvorlage für eine β-Propellerstruktur stehen derzeit zwanzig kristallographisch untersuchte WD-Proteine zur Verfügung (nach Suche in PROSITE-Datenbank), darunter die Proteine Skip8p (Cheng et al., 2004; Madrona et al., 2004), RACK1 (Sengupta et al., 2004), Aip1p (Voegtli et al., 2003; Ono et al., 2003), CBF3 (Orlicky et al., 2003), TLE1 (Pickles et al., 2002), TOLB (Carr et al., 1999), TUP1 Wiliams et al., 1990) und verschiedene Varianten der Beta-Untereinheit des G-Proteins (Sondek et al., 1996; Wall et al., 1995). Die Proteine zeigen in diesen Untersuchungen, mit Ausnahme von TOLB, jeweils eine siebenblättrige Struktur eines Propellers. Um so erstaunlicher ist das Ergebnis der Rekonstruktion des C-terminalen WD-Propellers von Dmxl1 und DMXL1: PSA berechnete für beide abgeleiteten Aminosäuresequenzen eine zehn-blättrige Propellerstruktur (Abb. 4.13,14). Hier ist anzumerken, dass PSA in beiden Spezies nicht in der Lage war, die letzte C-terminale, als sicher geltende WD-Wiederholungseinheit, zu erkennen. Eigene Untersuchungen durch eine virtuelle Verlängerung des C-Terminus um eine von PSA erkannte WD-Einheit zeigen, dass das Programm grundsätzlich auch in der Lage ist, eine Propellerrekonstruktion mit elf WD-Repeats durchzuführen. Strukturen über elf WD-Wiederholungseinheiten können aktuell nicht rekonstruiert werden. Die abgeleiteten Aminosäuresequenzen von Maus und Mensch besitzen C-Terminal mindestens elf WD-Repeats. Für alle weiteren in dieser Arbeit angesprochenen homologen DmX-Proteine wurden auch Rekonstruktionen durchgeführt. Bei F54E4.1 konnte eine Struktur mit ebenfalls zehn Propellerblättern rekonstruiert werden, für die abgeleiteten Aminosäureaequenzen von DMX und CPY wurden jeweils neun WD-Repeats von PSA vorhergesagt und rekonstruiert. Bei diesen drei Spezies konnten von dem Programm, im Gegensatz zu Maus und Mensch, auch die terminalen WD-Wiederholungseinheiten erkannt werden, dafür wurden die weiter innen liegenden WD-Repeats aufgrund der Eingabebeschränkung von PSA auf 1000 Aminosäuren nicht bei der Berechnung der Faltungsstruktur berücksichtigt. Analysen zeigen aber, dass alle angesprochenen Spezies C-Terminal mindestens elf WD-Wiederholungseinheiten besitzen. Diese Analysen sowie die zugehörigen PDB-Dateien befinden sich auf dem beigefügten Datenträger. Bei der Betrachtung dieser Faltungsrekonstruktionen stellt sich die grundlegende Frage, ob WD-Propellerstrukturen mit mehr als sieben Einheiten überhaupt stabil sind und *in vivo* realisiert werden. Die meisten bisher untersuchten und kristallisierten Proteine formen einen oder mehrere siebenblättrige Propeller, dies mag jedoch vor allem daran liegen, das alle untersuchten Proteine n * 7 WD-Wiederholungseinheiten in der Aminosäuresequenz aufweisen. In der Arbeit von Murzin et al. (1992) wird eine Propellerstruktur mit sieben Propellerblättern als die

grundsätzlich stabilste Form beschrieben, weitere Arbeiten zu diesem Thema existieren nicht. Bis heute ist völlig unklar, ob Proteine mit mehr als sieben WD-Repeats größere Propeller bilden oder sich zu mehreren kleineren Propellern formen und „überzählige" WD-Wiederholungseinheiten nicht in diese integriert werden. Es ist jedoch denkbar, dass ein siebenblättriger Propeller die prädominante Struktur darstellt (Mikyas et al., 2004; Voegtli et al., 2003; Stöhr et al., 2002; Li et al., 2001; Smith et al., 1999). In der Arbeit von Carr et al. (1999) wurde das prokaryotische Protein TOLB kristallographisch untersucht. Obwohl mehr als sieben WD-Wiederholungseinheiten in der Aminosäuresequenz gefunden werden, wird hier nur ein Propeller mit sechs Propellerblättern ausgebildet. Die anderen WD-Repeats werden nicht zur Ausbildung eines sieben-blättrigen Propellers herangezogen, sondern erzeugen eine separate, mit helikalen Abschnitten gemischte Struktur. Dies kann als Anhaltspunkt gewertet werden, dass sich anhand der Anzahl vorhandener WD-Repeats nicht auf die Zusammensetzung der Propellerstruktur schließen lässt. Erschwert wird eine Vorhersage auch aufgrund experimenteller Daten, die zeigen, dass vermutlich verschiedene Chaperone die Faltung von WD-Proteinen *in vivo* assistieren und stabilisieren (Craig, 2003; Ho et al., 2002; Valpuesta et al., 2002). Zudem ist es denkbar, dass es weitere, bisher nicht bekannte Proteine gibt, die eine solche Funktion bei WD-Proteinen ausüben können. Grundsätzlich muss aber davon ausgegangen werden, dass aufgrund der sterischen Beschränkungen in einer Propellerstruktur nicht beliebig viele β-Faltblätter angeordnet werden können, eine Obergrenze lässt sich aber derzeit nicht genau bestimmen. Neben der Analyse des C-terminalen WD-Propellers mit PSA wurde auch eine Rekonstruktion dem Programm ESyPred3D versucht. ESyPred3D war jedoch nur in der Lage, die Positionen der terminalen 300 Aminosäuren, die sechs WD-Repeats enthalten, zu berechnen (Abb. 4.15). Indes zeigen weitere Rekonstruktionen mit den abgeleiteten Aminosäuresequenzen von DMXL1, CPY, DMX, F54E4.1 und Soi3p/Rav1p mit diesem Programm teilweise große Unterschiede. Obwohl das DMXL1-Protein im C-terminalen Abschnitt eine sehr hohe Konservierung zu der abgeleiteten Aminosäuresequenz von Dmxl1 zeigt, konnte für die humane Sequenz kein Propeller rekonstruiert werden, lediglich die Berechnung des C-Terminus von CPY erzeugt eine Struktur mit erkennbarer Übereinstimmung zu Dmxl1. Aufgrund der oben beschriebenen Probleme bei der Voraussage der Propellerstruktur für Proteine mit mehr als sieben WD-Repeats oder mit einer Anzahl, die nicht einem Vielfachen von sieben WD-Wiederholungseinheiten entspricht, kann über die Faltung der WD-Repeats der abgeleiteten Aminosäuresequenz von Dmxl1 nur spekuliert werden. Auffällig ist jedoch die blockartige Anordnung der WD-Repeats im C-Terminus (Abb. 4.11). Bei der Betrachtung der mit hoher Wahrscheinlichkeit identifizierten WD-Einheiten lassen sich zwei Blöcke unterscheiden, die durch 371 Aminosäuren voneinander getrennt sind. PSA bildet aus beiden Blöcken einen Propeller mit zehn WD-Wiederholungseinheiten, wobei die letzte C-terminale Einheit nicht erkannt wird. Hier muss also von einer Propellerstruktur mit elf WD-Repeats ausgegangen werden. Ob sich die Blöcke *in vivo* wirklich zu dieser Anordnung zusammenschließen, kann nicht vorausgesagt werden, ebenso wenig die Faltungsstruktur der dazwischenliegenden 371 Aminosäuren. Eine weitere denkbare Möglichkeit wäre eine Faltung, die zwei unabhängige fünf- und



sechsblättrige Propeller darstellt, diese Variante mutet allerdings nicht sehr überzeugend an. Interessanterweise werden innerhalb der 371 Aminosäuren von REP drei weitere

Abb. 4.14: Dreidimensionale Rekonstruktion von 10 C-terminalen WD-Einheiten des Dmxl1-Proteins.

Die Berechnung der atomaren Koordinaten erfolgte mit dem Programmpaket PSA am BMERC. Als Faltungsvorlage wurde dabei die β-Untereinheit des G-Proteins verwendet. Die Visualisierung der Daten erfolgte mit dem frei erhältlichen Programm RasTop von Philippe Valadon, welches auf der Open-Source Version von RasMol (Roger Sayle) programmiert wurde. Bei dieser Abbildung handelt es sich um eine

komplexe Darstellung der berechneten C-terminalen Struktur von Dmxl1 bei der sowohl die Polypeptidstränge, β-Faltblätter mit Orientierung sowie die Van der Waals Oberflächenladung gezeigt sind. Für die Aminosäureabschnitte, die die Propellerblätter verbinden, existieren keine Faltungsvorlagen, sie sind daher nicht gezeigt. Die zentrale Pore hat eine konische Form und misst an ihrer engsten Stelle 11,82Å und am äußeren Rand 19.31Å, sie ist damit größer als bei einer Propellerstruktur mit sieben Propellerblättern. Dennoch ist davon auszugehen, dass die Pore zum einen zu klein ist, um komplexere Moleküle zu beherbergen, durchzulassen oder als aktives Zentrum eine katalytische Aktivität auszuüben, zum anderen zeigen Kristallstrukturanalysen anderer WD-Proteine, dass die Pore durch hereinragende Polypeptidabschnitte, die die Propeller verbinden, meist versperrt ist (u.a. Cheng et al., 2004; Madrona et al., 2004; Lodowski et al., 2003; Pickles et al., 2002). Bei der gezeigten Abbildung handelt es sich um Stereobilder, diese können bei normalem Leseabstand durch Schielen oder unter Verwendung einer Stereolupe zur Deckung gebracht werden.

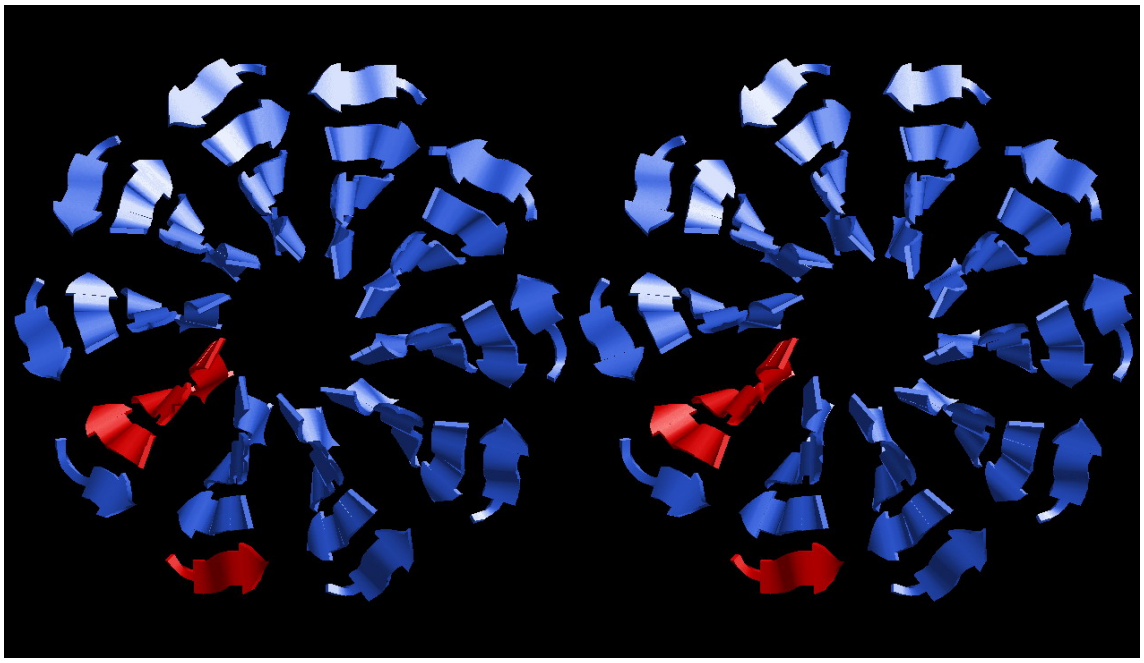


Abb 4.15: β-Blätter bilden das Rückgrat der Propellerstruktur.

Zur Verdeutlichung der Propellerstruktur der dreidimensionalen Rekonstruktion der zehn terminalen WD-Repeats des Dmxl1-Proteins wurde hier nur das Rückgrat der β-Blätter dargestellt. Eine virtuelle Lichtquelle mittig oberhalb der Moleküle verdeutlicht die Neigung der einzelnen β-Blätter zur Z-Achse. In rot dargestellt ist eine WD-Wiederholungseinheit. Stränge A-C (β1-3) stellen die inneren drei Blätter, Strand D (β4) stellt die äußere Struktur des folgenden Propellerblatts. Es handelt sich hierbei ebenfalls um eine Stereodarstellung.

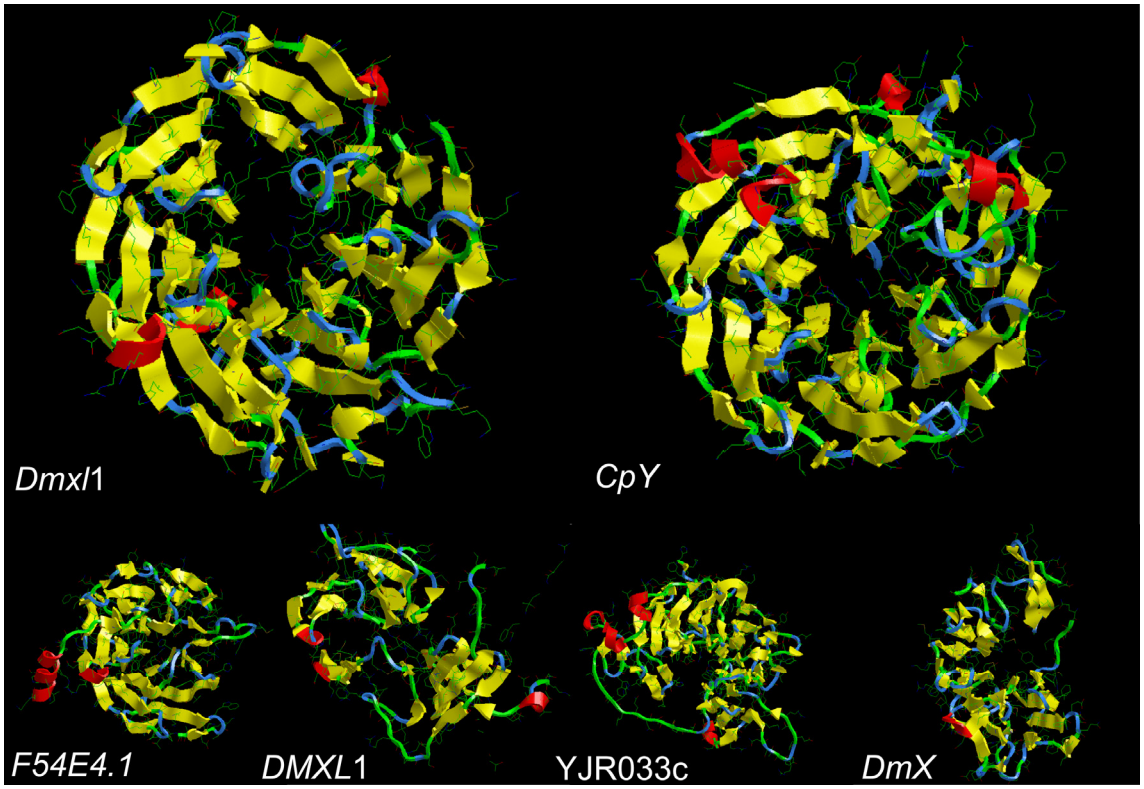


Abb. 4.16: Rekonstruktion von 300 C-terminalen Aminosäuren durch ESyPred3D

Die Rekonstruktion wurde mit allen abgeleiteten Aminosäuresequenzen von DmX-Homologen, die in dieser Arbeit angesprochen wurden, durchgeführt. Obwohl dieses Programm nicht auf WD-Repeats und ihre Faltung spezialisiert ist, konnte zumindest für die abgeleiteten Aminosäuresequenzen von Dmxl1 und CPY eine annähernde Propellerstruktur mit sieben Blättern berechnet werden. Für die anderen Homologen wurde eine solche Faltung nicht vorhergesagt. Bei den Berechnungen wurden, im Gegensatz zur Vorhersage durch PSA, auch die die WD-Repeats verbindenden Aminosäuresequenzen mit einbezogen (blaue Abschnitte). β-Blatt-Strukturen sind in gelb dargestellt, helikale Bereiche in rot. Auch wenn die Darstellung optisch einem Propeller ähnelt, sind die β-Blätter wesentlich ungeordneter als es für eine Propellerstruktur aus WD-Wiederholungseinheiten erwartet werden kann, daher weist die für Dmxl1 und CPY ermittelte Faltung auch deutlich Deformationen auf. Dies legt den Schluss nahe, das es derzeit kein Programm gibt, das sowohl die Faltung der WD-Repeats als auch der dazwischenliegenden Aminosäureabschnitte korrekt berechnen kann.

WD-Wiederholungseinheiten identifiziert, die von keinem anderen Programm gefunden wurden. Dies würde die Anzahl der C-terminalen Repeats auf 14 erhöhen und damit zwei Propellerblätter mit jeweils sieben Propellerblättern wahrscheinlich machen. In der Arbeit von Voegtli et al. (2003) konnten durch *in silico* Analysen der Sequenzdaten von Aip1p zunächst nur zehn WD-Repeats identifiziert werden (Ono, 2001). Erst die Kristallstrukturanalyse des Proteins zeigte, dass in der Sequenz noch mindestens vier weitere WD-Repeats enthalten waren, die aufgrund einer deutlich geringeren Homologie zur WD-Konsensussequenz unentdeckt geblieben waren.
Dies zeigt zum einen, dass die derzeit zur Verfügung stehenden Programme aufgrund der Diversität der WD-Wiederholungseinheiten nur einen Teil der wirklich vorhandenen

Repeats identifizieren können und zum anderen, dass sich zumindest WD-Repeats, deren Zahl in einem Protein einem Vielfachen von sieben entspricht, möglicherweise eher zu mehreren siebenblättrigen Propellerstrukturen falten, als einen große Propeller zu bilden. Die Vorhersage der Faltungsstruktur der N-terminalen WD-Repeats ist noch schwieriger. In diesem Bereich werden wesentlich mehr „unsichere“ WD-Wiederholungseinheiten (Abb. 4.10) gefunden als im C-Terminus, die Anzahl schwankt daher zwischen zwölf WD-Repeats, die von mehreren Programmen identifiziert wurden und 21 Einheiten mit unsicheren Kandidaten. Auch hier ist es nicht möglich, eine Vorhersage über eine mögliche Propellerstruktur zu treffen, dennoch sind C-terminal zwei bis drei siebenblättrige Propeller denkbar. Für Abschnitte ohne WD-Repeats gibt es keine Faltungsvorlage, so dass über die Konformation dieser Bereiche keine Aussage getroffen werden kann. Letztendlich ist Aufklärung der dreidimensionalen Struktur des Dmxl1-Proteins nur durch eine Kristallstrukturanalyse denkbar. Es ist anzunehmen, dass die Strukturanalyse von Dmxl1 erst in einigen Jahren erfolgreich durchgeführt werden kann.

4.4.9 Ist *Dmxl*1 ein Multifunktionsprotein?

In der Arbeit von Weil (2000) wurde die Funktion der beiden homologen Gene *CpY* und *DmX* näher untersucht. Hierbei wurde vor allem eine Aufgabe der beiden Proteine in der Endozytose und dem zellulären Transport diskutiert. Diese Annahme beruht vor allem auf einer beobachteten Kopräzipitation des DmX-Homologen aus *Rattus norvegicus* mit Amphiphysin II (Ramjaun et al., 1997), einer Isolierung dieses homologen Proteins durch die Bindungsaffinität zu Endophilin und einer Übereinstimmung des in der Arbeit von Weil untersuchten Transkriptionsmusters von *DmX* zu einer Antikörperfärbung gegen α-Adaptin. Zudem zeigt der Phänotyp der von Kraemer (pers. Mitteilung) und Weil (2000) erzeugten *DmX* $^{-}$-Mutanten eine Ähnlichkeit zu *Synaptotagmin*- (DiAntonio et al., 1993; Prokop, pers. Mitteilung an Weil) und *α-Adaptin*^{06694} -Mutanten (Gonzalez-Gaitan et al., 1997). Alle angesprochenen Gene spielen eine wichtige Rolle in der Endozytose. Da bis heute die cytologische Lokalisation und mögliche Interaktionspartner der DmX-Homologen nicht bekannt sind, soll im Folgenden diese Hypothese in Hinblick auf den in dieser Arbeit beobachteten Phänotyp, die erstellten *in silico*-Analysen und die inzwischen bekannte Funktion des homologen Gens *SOI*3/*RAV*1 von *Saccharomyces cerevisiae* diskutiert werden. Für den offenen Leserahmen YJR033c der Hefe war schon bekannt, dass er die größte Ähnlichkeit zu Genen höherer Eukaryoten mit den DmX-Homologen zeigt (Kane et al., 2003; Kraemer et al. 2000, 1998), seine Funktion in der Hefe konnte aber bisher nicht aufgeklärt werden. Erst in der Arbeit von Sipos et al. (2004) wurde die Identität von YJR033c mit den Genen *SOI*3 (Sipos et al., 2004; Brickner et al., 1997; Redding et al., 1996) und *RAV*1 (Seol et al., 2001) nachgewiesen. Wie die Homologen von DmX codiert auch *SOI*3/*RAV*1 für ein WD-Repeatprotein, das jedoch aufgrund seiner deutlich kürzeren Aminosäuresequnz nur 7-11 WD-Wiederholungseinheiten besitzt (Kapitel 4.4.7; Sipos et al., 2004; Kane et al., 2003).

Soi3p/Rav1p übt offensichtlich mehrere Funktionen aus. Zum einen ist es integraler Bestandteil des RAVE-Komplexes (Seol et al., 2001), der aus den Proteinen Rav1, Skp1 und Rav2 besteht. Dieser zytosolische Komplex interagiert mit Untereinheiten der V_1-ATPase, die sich ebenfalls frei im Zytosol befindet und ermöglicht bzw. reguliert den Zusammenbau der V_1-ATPase mit der membrangebundenen V_0-ATPase-Einheit zu einer funktionalen Protonenpumpe (Kane et al., 2003; Smardon et al., 2002; Seol et al., 2001). Die Kontrolle über die Anzahl dissoziierter V_1-ATPase Einheiten ist ganz entscheidend für die Regulation lokaler pH-Werte (Nishi et al., 2002; Seol et al., 2001; Forgac, 1999). Vakuolare H^+-ATPasen werden in allen eukaryotischen Organismen gefunden. Es handelt sich um eine Familie ATP-abhängiger Protonenpumpen, die aus vielen Untereinheiten zusammengesetzt und in ein weites Spektrum physiologischer Prozesse involviert sind. Sie werden in Endomembran-Organellen wie Vakuolen, Lysosomen, Endosomen, dem Golgi-Apparat, Granula und anderen Vesikeln gefunden. Ihre Hauptaufgabe liegt in der Ansäuerung des luminalen pH-Wertes intrazellularer Kompartemente. Die Protonenpumpen sorgen auch für konstante pH-Werte in spezialisierten Zellen wie Osteoklasten, Epithelzellen, Nierenzellen und dem männlichen Reproduktionstrakt (Sun-Wada et al., 2004). Aufgrund der Homologie zu DmX und der Untersuchungen von Weil (2000) ist eine zu Soi3p/Rav1p analoge Funktion von DmX sehr wahrscheinlich. Es ist nicht bekannt, ob sich ein vergleichbarer RAVE-Komplex auch in höheren Eukaryoten bildet. Klar ist, dass bisher keine Homologen zu *RAV*2 in höheren Eukaryoten identifiziert werden konnten. Ob die *Skp*1-Homologen höherer Eukaryoten mit DmX komplexieren, kann derzeit nicht gesagt werden. Da die DmX-Proteine wesentlich größer und komplexer als Soi3p/Rav1p sind, ist es durchaus denkbar, dass die Funktion des RAVE-Komplexes hier von DmX teilweise oder allein übernommen werden könnte. Es ist also vorstellbar, dass DmX die Assoziation der vakuolaren ATPase aus den Untereinheiten V_1 und V_0 ermöglicht, steuert und kontrolliert. Dies würde auch die niedrige Expression und komplexe Regulation von *Dmxl*1 erklären (Kapitel 4.4.5). Interessant ist in diesem Zusammenhang, dass die H-Untereinheit des V1-Sektors an den AP-2 Komplex bindet, α-*Adaptin* und *AP50* sind wiederum Untereinheiten des AP-2 Komplexes, die ein ähnliches Transkriptionsmuster wie *DmX* zeigen (Weil, 2000). Ein besonderes Augenmerk muss auf die Funktion der V-ATPase im männlichen Reproduktionstrakt geworfen werden. Bei der Entwicklung der Testes, Reifung und Lagerung der Spermien sowie der Synthese des Prostata-Sekrets spielt der luminale pH-Wert eine entscheidende Rolle. Nach Untersuchungen an der Ratte werden in Epididymis und proximalem Vas deferens bis zu 80% der gesamten H^+-Sekretion von der V-ATPase geleistet (Herak-Kramberger et al., 2001; Breton et al., 2000, 1996; Brown et al., 1997, 1992). Eine starke Expression der H^+-ATPase wird zudem in seminalen Vesikeln, Prostata, Ductus deferens und Urethra beobachtet (Herak-Kramberger, 2001). Neben der Regulation des pH-Wertes im Reproduktionstrakt spielt die Vakuoläre ATPase eine wichtige Rolle bei der Endozytose und lysosomaler Aktivität in der Epididymis durch Ansäuerung der Lumina (Andonian et al., 2001). Eine Veränderung des pH-Wertes um 0.3 Einheiten kann dabei schon zu einer totalen Sterilität führen (Yeung et al., 2004). Es könnte also möglicherweise eine Verbindung zum beobachteten Phänotyp der *Dmxl*1^{-}-Chimären existieren. Die Reduktion der *Dmxl*1-Transkripte um 50% in Zellen mit einem *Dmxl*1-

Null Allel könnte zu Störungen bei der Entwicklung der Testes in der frühen Embryonalentwicklung führen. Da gerade in der Keimbahnentwicklung schon geringe Beeinträchtigungen durch ein restriktives Kontrollsystem eine Apoptose der betroffenen Zellen einleiten (Print et al., 2000), ist es vorstellbar, dass der beobachtete Hypogonadismus auf apoptotische Abläufe während der Keimzellentwicklung zurückzuführen sein könnte. In der Arbeit von Zhang et al. (2004) führte eine Disruption des WD-Repeat-Proteins PF20 zu Hypogonadismus, Haploinssuffiziens und, wenn auch stärker ausgeprägt als bei den *Dmxl*1-Cimären, zu einer Teilsterilität. Es wurden nur Nachkommen gezeugt, die nicht das disruptierte Allel trugen. Eine Obesitas konnte in diesem Zusammenhang nicht beobachtet werden. Wie aber lässt sich nun die Fettleibigkeit der chimären *Dmxl*1-Mäuse begründen? Ein direkter Zusammenhang mit der Funktion der V-ATPase und dem Fettsäurestoffwechsel ist nicht gegeben. Eine Erklärungsmöglichkeit liegt in der Verknüpfung der Regulation des Haupt-Glukosetransporters GLUT4 durch den pH-Wert in Endosomen von Adipozyten (Chinni et al., 1999). GLUT4 ist mit endosomal-tubovesikularen Strukturen des Trans-Golgi-Netzwerks in braunen und weißen Fettzellen assoziiert und wird unter normalen Bedingungen nur intrazellular gefunden (Malide et al., 1997; Millar et al., 1997; Rodnick et al., 1992; Smith et al., 1991; Slot et al., 1991). Die Ansäuerung der Endosomen spielt eine wichtige Rolle im Recycling internalisierter Rezeptoren zurück auf die Zelloberfläche (Merzendorfer et al., 1997; van Weert et al., 1995). In der Arbeit von Chinni et al. (1999) konnte gezeigt werden, dass eine Inhibition der V-ATPase durch Bafilomycin A_1 zu einer Translokation von GLUT4 aus intrazellularen Membranen an die Zelloberfläche von Adipozyten führt. Eine Modifikation des pH-Wertes der Endosomen von Adipozyten bewirkt eine Externalisation von GLUT4 an die Zelloberfläche. Der Fettstoffwechsel und Glucosestoffwechsel ist über Acetyl-CoA gekoppelt, eine gesteigerte Verfügbarkeit eines der beiden Energieträger supprimiert die Oxidation des anderen (Randle et al., 1965). Eine verstärkte Aufnahme der Glucose in Adipozyten durch vermehrte Externalisierung von GLUT4 inhibiert die Fettsäureoxidation und aktiviert die Fettsäuresynthese (Oh et al., 2004). Bei den in dieser Arbeit gezeigten hochchimären Tieren könnte also der Mangel an *Dmxl*1-Transkripten zu einer verminderten Zusammensetzung der Untereinheiten V_0 und V_1 zur funktionalen V-ATPase führen und so eine reduzierte Ansäuerung der Vesikel bewirken die dann eine gesteigerte Externalisierung von GLUT4 ermöglicht. Die Ausprägung der Fettleibigkeit müsste daher in Abhängigkeit vom Grad des Chimärismus auftreten, was auch in dieser Arbeit gezeigt werden konnte. Die Obesitas der *Dmxl*1-Chimären wäre demzufolge ebenfalls ein sekundärer Effekt.

Eine weitere Funktion von Soi3p/Rav1p besteht aus Transportvorgängen zwischen frühen Endosomen und dem prävakuolarem Kompartiment/späten Endosomen im Trans-Golgi-Netzwerk (Sipos et al., 2004). Es wird angenommen, dass Soi3p/Rav1p mit multiplen Transmembranproteinen des Trans-Golgi-Netzwerks interagiert (Redding et al., 1996a) und möglicherweise auch für Transportvorgänge vom prävakuolarem Kompartiment zur Vakuole verantwortlich ist (Sipos et al., 2004). Eine wichtige Rolle zahlreicher WD-Repeat Proteine ist der zelluläre und vesikuläre Transport (Jin et al., 2004; Rybakin et al., 2004; Lin et al., 2003; Li et al., 2001; Smith et al., 1999). Die

Daten von Weil (2000) und die in dieser Arbeit gewonnenen Erkentnisse weisen darauf hin, dass die DmX-Homologen vermutlich hier eine entscheidende Rolle spielen könnten. Dass eine solche Funktion nun auch für Soi3p/Rav1p nachgewiesen wurde, bestärkt diese Annahme. Es ist jedoch anzunehmen, dass ein mögliches Bindungsspektrum von DmX gegenüber dem homologen Protein der Hefe wesentlich ausgeweitet ist. Allein die Anzahl der identifizierten WD-Reapeats, die daraus resultierenden Propellerstrukturen und die zusätzlichen Sequenzabschnitte machen ein erweitertes Aufgabengebiet von DmX wahrscheinlich.

5 Zusammenfassung

Bei *Dmxl*1 handelt es sich um ein neuartiges Gen aus *Mus musculus*. Das ebenfalls in der vorliegenden Arbeit bioinformatisch untersuchte Gen *DMXL*1 ist das zu *Dmxl*1 homologe Gen des Menschen. Beide Gene bestehen aus 43 Exons, das murine *Dmxl*1 codiert für eine mRNA von 10992 bp bzw. 12210bp, das humane *DMXL*1 kodiert für eine cDNA von 11082 bp, der offene Leserahmen umfasst bei der Maus 9042 bp. In der Maus konnte ein mögliches alternatives Polyadenylierungssignal identifiziert werden. Zwischen beiden Spezies sind die Exonpositionen und ihre Längen hoch konserviert. *Dmxl*1 liegt auf dem Crick-Strang von Chromosom 18 Bande C, der translatierte Bereich erstreckt sich auf genomischer Ebene über 129558 bp und die Orientierung verläuft in Richtung Centromer. *Dmxl*1 und *DMXL*1 gehört damit zu den größten bekannten Genen in Maus und Mensch. Bei beiden Spezies liegen die DmX-Homologen genomisch innerhalb eines Bereichs der Isochoren-Klasse L1 in einer genarmen Region. Die Anzahl der repetitiven Elemente innerhalb der Genregion von *Dmxl*1 liegt 6% unter dem erwarteten Wert eines L1 Isochors, die Anzahl beim Menschen liegt 4% über dem erwarteten Wert. Um die mögliche Promotorstruktur von *Dmxl*1 darzustellen, wurden umfangreiche *in silico*-Analysen der Region um den putativen Transkriptionsstart vorgenommen. Mit Hilfe der gewonnenen Daten konnte ein Transkriptionstartpunkt identifiziert werden. Zumdem wurde eine Promotorstruktur erarbeitet, bei der angenommen werden kann, dass sie eine gute Näherung an die tatsächlich vorhandenen Bindungsstellen von Transkriptionsfaktoren darstellt. Die mit bioinformatischen Werkzeugen erzeugte virtuelle Promotor- und Enhancerstruktur zeigt das Potenzial, *Dmxl*1 basal und ubiquitär zu exprimieren. Gleichzeitig zeigen diese Daten, dass *Dmxl*1 vermutlich in einigen Geweben der Keimbahn, im Fettgewebe, dem blutbildenen System und während der Embryogenese hochkomplex reguliert werden kann. Eine regulierte Expression zur Steuerung des Energiestoffwechsels ist ebenfalls wahrscheinlich. Diese Ergebnisse passen sehr gut zu den experimentell ermittelten Daten und den beobachteten Phänotypen *Dmxl1*-chimärer Mäuse.
Die abgeleitete Aminosäuresequenz umfasst in der Maus 3013 AS, im Menschen 3027 AS, der Vergleich der abgeleiteten Aminosäuresequenzen zeigt eine Identität von 89,3 % und eine Similarität von 94,7 % zwischen beiden Spezies. Im Dmxl1/DMXL1-Protein von Maus und Mensch konnten mindestens 24 und maximal 36 WD-Wiederholungseinheiten identifiziert werden, zudem wurden eine Reihe weiterer

konservierter Proteinmotive gefunden. Die *in silico*-Strukturanalysen beider abgeleiteter Aminosäuresequenzen lässt vermuten, dass sich C- und N-terminal WD-Propellerstrukturen befinden. In dieser Arbeit gelang eine C-terminale Rekonstruktion einer 10-blättrigen Propellerstruktur, denkbar ist jedoch auch eine Struktur mit mindestens drei WD-Propellern, wenn eine prädominante Struktur mit Propellern aus jeweils sieben Propellerblättern angenommen wird.
Das primäre Ziel dieser Arbeit, die Etablierung einer stabilen Mauslinie mit diruptiertem *Dmxl1*-Gen konnte aufgrund einer beobachteten Haploinsuffiziens nicht erreicht werden. Trotz zahlreicher Transformationen von Maus-Stammzelllinien konnte letztlich nur eine stabil transformierte Linie mit einem *Dmxl*1-Null-Allel identifiziert werden, was auch zu den theoretischen Daten und den angenommenen Aufgaben von Dmxl1 als komplex und diffizil reguliertes Multifunktions-Protein passt. Aus der transformierten Mauszelllinie konnten chimäre Mäuse entwickelt werden, die in Abhängigkeit von dem Ausmaß des Chimärismus phänotypisch massive Schädigungen aufwiesen. Neben einer Teilsterilität wurden massive Fettleibigkeit und ein ausgeprägter Hypogonadismus beobachtet. Keines der Tiere war in der Lage das *Dmxl*1-Null-Allel zu transduzieren. Die Tiere waren nur sehr eingeschränkt fertil, die wenigen Nachkommen entsprachen genotypisch und phänotypisch ausschließlich den verwendeten Blastocysten. Um die in dieser Arbeit dargestellte Funktion von *Dmxl*1 zu bestätigen und besser zu verstehen, müsste versucht werden, entweder einen konditionalen Knock-Out durchzuführen bzw. das Gen mit Hilfe von siRNA auszuschalten und/oder transgene Tiere herzustellen, um durch Überexpression von *Dmxl*1 weitere Hinweise auf die Funktion des Gens zu erhalten.

Adams, J., R. Kelso, et al. (2000). The kelch repeat superfamily of proteins: propellers of cell function. Trends Cell Biol. **10:** 17-24.

Adams, J. M. and S. Cory (1998). "The Bcl-2 protein family: arbiters of cell survival." Science **281**(5381): 1322-6.

Aissani, B. and G. Bernardi (1991). "CpG islands, genes and isochores in the genomes of vertebrates." Gene **106**(2): 185-95.

Aissani, B. and G. Bernardi (1991). "CpG islands: features and distribution in the genomes of vertebrates." Gene **106**(2): 173-83.

Aissani, B., G. D'Onofrio, et al. (1991). "The compositional properties of human genes." J Mol Evol **32**(6): 493-503.

Albertini, M., P. Rehling, et al. (1997). "Pex14p, a peroxisomal membrane protein binding both receptors of the two PTS-dependent import pathways." Cell **89**(1): 83-92.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.

Altschul, S. F. and D. J. Lipman (1990). "Protein database searches for multiple alignments." Proc Natl Acad Sci U S A **87**(14): 5509-13.

Alvarez-Valin, F., G. Lamolle, et al. (2002). "Isochores, GC3 and mutation biases in the human genome." Gene **300**(1-2): 161-8.

Amid, C. (2002) Vergleichende Genomanalyse: die Primärstruktur der Chromosomenregion 11p15.3 des Menschen und des homologen Abschnitts der Maus. Mainz,Univ., Diss.

Amid, C., A. Bahr, et al. (2001). "Comparative genomic sequencing reveals a strikingly similar architecture of a conserved syntenic region on human chromosome 11p15.3 (including gene ST5) and mouse chromosome 7." Cytogenet Cell Genet **93**(3-4): 284-90.

Andrade, M. A. and P. Bork (2000). "Automated extraction of information in molecular biology." FEBS Lett **476**(1-2): 12-7.

Ansari-Lari, M. A., J. C. Oeltjen, et al. (1998). "Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6." Genome Res **8**(1): 29-40.

Antequera, F. and A. Bird (1993). "CpG islands." Exs **64**: 169-85.

Antequera, F. and A. Bird (1993). "Number of CpG islands and genes in human and mouse." Proc Natl Acad Sci U S A **90**(24): 11995-9.

Bachmanov, A. A., X. Li, et al. (2001). "High-resolution genetic mapping of the sucrose octaacetate taste aversion (Soa) locus on mouse Chromosome 6." Mamm Genome **12**(9): 695-9.

Bachmanov, A. A., X. Li, et al. (2001). "Positional cloning of the mouse saccharin preference (Sac) locus." Chem Senses **26**(7): 925-33.

Bachmanov, A. A., M. G. Tordoff, et al. (1998). "Voluntary sodium chloride consumption by mice: differences among five inbred strains." Behav Genet **28**(2): 117-24.

Bahr, A. (1999) Die molekulare Struktur der Chromosomenregion 11p15.3 des Menschen und des homologen Abschnitts der Maus. Nukleotidsequenz, neue Gene und Interspeziesvergleich. Mainz,Univ., Diss.

Bailey, D. W. (1971). "Allelic forms of a gene controlling the female immune response to the male antigen in mice." Transplantation **11**(4): 426-8.

Bailey, D. W. (1971). "Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes." Transplantation **11**(3): 325-7.

Bailey, D. W. and J. Hoste (1971). "A gene governing the female immune response to the male antigen in mice." Transplantation **11**(4): 404-7.

Bajic, V. B. and V. Brusic (2003). "Computational detection of vertebrate RNA polymerase II promoters." Methods Enzymol **370**: 237-50.

Bajic, V. B. and S. H. Seah (2003). "Dragon Gene Start Finder identifies approximate locations of the 5' ends of genes." Nucleic Acids Res **31**(13): 3560-3.

Bajic, V. B. and S. H. Seah (2003). "Dragon gene start finder: an advanced system for finding approximate locations of the start of gene transcriptional units." Genome Res **13**(8): 1923-9.

Bajic, V. B., S. H. Seah, et al. (2003). "Computer model for recognition of functional transcription start sites in RNA polymerase II promoters of vertebrates." J Mol Graph Model **21**(5): 323-32.

Bajic, V. B., S. H. Seah, et al. (2002). "Dragon Promoter Finder: recognition of vertebrate RNA polymerase II promoters." Bioinformatics **18**(1): 198-9.

Bajic, V. B., S. L. Tan, et al. (2003). "Dragon ERE Finder version 2: A tool for accurate detection and analysis of estrogen response elements in vertebrate genomes." Nucleic Acids Res **31**(13): 3605-7.

Barsh, G. S., I. S. Farooqi, et al. (2000). "Genetics of body-weight regulation." Nature **404**(6778): 644-51.

Battey, J., E. Jordan, et al. (1999). "An action plan for mouse genomics." Nat Genet **21**(1): 73-5.

Beattie, J. H., A. M. Wood, et al. (1998). "Obesity and hyperleptinemia in metallothionein (-I and -II) null mice." Proc Natl Acad Sci U S A **95**(1): 358-63.

Beaudoing, E., S. Freier, et al. (2000). "Patterns of variant polyadenylation signal usage in human genes." Genome Res **10**(7): 1001-10.

Beck, B., S. Richy, et al. (2003). "Ghrelin and Body Weight Regulation in the Obese Zucker Rat in Relation to Feeding State and Dark/Light Cycle." Experimental Biology and Medicine **228**(10): 1124-1131.

Bellgard, M. I. and T. Gojobori (1999). "Significant differences between the G+C content of synonymous codons in orthologous genes and the genomic G+C content." Gene **238**(1): 33-7.

Bergmann, C., K. Zerres, et al. (2002). "Allelic variants in the 5' non-coding region of the connexin32 gene: possible pitfalls in the diagnosis of X linked Charcot-Marie-Tooth neuropathy (CMTX)." J Med Genet **39**(9): e58.

Bernal, A., U. Ear, et al. (2001). "Genomes OnLine Database (GOLD): a monitor of genome projects

world-wide." Nucleic Acids Res **29**(1): 126-7.

Bernardi, G. (1985). "The organization of the vertebrate genome and the problem of the CpG shortage." Prog Clin Biol Res **198**: 3-10.

Bernardi, G. (1993). "The vertebrate genome: isochores and evolution." Mol Biol Evol **10**(1): 186-204.

Bernardi, G. (2000). "Isochores and the evolutionary genomics of vertebrates." Gene **241**(1): 3-17.
Bernardi, G. (2001). "Misunderstandings about isochores. Part 1." Gene **276**(1-2): 3-13.

Bernardi, G. and G. Bucciarelli (1999). "Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes)." Mol Phylogenet Evol **13**(1): 77-81.

Bernardi, G., B. Olofsson, et al. (1985). "The mosaic genome of warm-blooded vertebrates." Science **228**(4702): 953-8.

Bhasin, M. and G. P. Raghava (2004). "ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST." Nucleic Acids Res **32**(Web Server issue): W414-9.

Bikar, S.-E. (1997). Sequenzierung eines Abschnitts des menschlichen Chromosoms 11 und computergestützte Auswertung der Sequenzdaten. Diplomarbeit, Johannes Gutenberg-Universität Mainz

Bird, A. P. (1980). "DNA methylation and the frequency of CpG in animal DNA." Nucleic Acids Res **8**(7): 1499-504.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-13.

Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**(6): 1513-23.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Blanco, A. G., M. Sola, et al. (2002). "Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator." Structure **10**(5): 701-13.

Blanco, P. V., G. Theile, et al. (2001). "Novel HLA allele, HLA-B*4013, identified from a potential bone marrow recipient. Putative gene conversion events with donor sequence." Tissue Antigens **57**(4): 380-3.

Blessing, M., L. B. Nanney, et al. (1993). "Transgenic mice as a model to study the role of TGF-beta-related molecules in hair follicles." Genes Dev **7**(2): 204-15.

Boffelli, D., J. McAuliffe, et al. (2003). "Phylogenetic shadowing of primate sequences to find functional regions of the human genome." Science **299**(5611): 1391-4.

Boguski, M. S. (1992). "Computational sequence analysis revisited: new databases, software tools, and the research opportunities they engender." J Lipid Res **33**(7): 957-74.

Boguski, M. S., R. C. Hardison, et al. (1992). "Analysis of conserved domains and sequence motifs in cellular regulatory proteins and locus control regions using new software tools for multiple alignment and visualization." New Biol **4**(3): 247-60.

Boguski, M. S., A. W. Murray, et al. (1992). "Novel repetitive sequence motifs in the alpha and beta

subunits of prenyl-protein transferases and homology of the alpha subunit to the MAD2 gene product of yeast." New Biol **4**(4): 408-11.

Bolivar, F., M. C. Betlach, et al. (1977). "Origin of replication of pBR345 plasmid DNA." Proc Natl Acad Sci U S A **74**(12): 5265-9.

Bolivar, F., R. L. Rodriguez, et al. (1977). "Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9." Gene **2**(2): 75-93.
Bolivar, F., R. L. Rodriguez, et al. (1977). "Construction and characterization of new cloning vehicles. II. A multipurpose cloning system." Gene **2**(2): 95-113.

Bonaldo, M. F., G. Lennon, et al. (1996). "Normalization and subtraction: two approaches to facilitate gene discovery." Genome Res **6**(9): 791-806.

Boyer, H. W. and D. Roulland-Dussoix (1969). "A complementation analysis of the restriction and modification of DNA in Escherichia coli." J Mol Biol **41**(3): 459-72.

Bradbury, A. F. and D. G. Smyth (1987). "Biosynthesis of the C-terminal amide in peptide hormones." Biosci Rep **7**(12): 907-16.

Bradbury, A. F. and D. G. Smyth (1987). "Enzyme-catalysed peptide amidation. Isolation of a stable intermediate formed by reaction of the amidating enzyme with an imino acid." Eur J Biochem **169**(3): 579-84.

Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-6.

Bradley, A. and E. Robertson (1986). "Embryo-derived stem cells: a tool for elucidating the developmental genetics of the mouse." Curr Top Dev Biol **20**: 357-71.

Bray, G. A. (1997). "Obesity and reproduction." Hum Reprod **12 Suppl 1**: 26-32.

Bray, G. A. and D. A. York (1997). "Leptin and Clinical Medicine: A New Piece in the Puzzle of Obesity." J Clin Endocrinol Metab **82**(9): 2771-2776.

Bray, N., I. Dubchak, et al. (2003). "AVID: A global alignment program." Genome Res **13**(1): 97-102.

Bray, N. and L. Pachter (2003). "MAVID multiple alignment server." Nucleic Acids Res **31**(13): 3525-6.

Breton, S., P. J. Smith, et al. (1996). "Acidification of the male reproductive tract by a proton pumping (H+)-ATPase." Nat Med **2**(4): 470-2.

Breton, S., T. Wiederhold, et al. (2000). "The B1 subunit of the H+ATPase is a PDZ domain-binding protein. Colocalization with NHE-RF in renal B-intercalated cells." J Biol Chem **275**(24): 18219-24.

Brickner, J. H. and R. S. Fuller (1997). "SOI1 encodes a novel, conserved protein that promotes TGN-endosomal cycling of Kex2p and other membrane proteins by modulating the function of two TGN localization signals." J Cell Biol **139**(1): 23-36.

Brinster, R. L., H. Y. Chen, et al. (1981). "Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs." Cell **27**(1 Pt 2): 223-31.

Brinster, R. L., H. Y. Chen, et al. (1981). "Mouse oocytes transcribe injected Xenopus 5S RNA gene." Science **211**(4480): 396-8.

Brockmann, G. A. and M. R. Bevova (2002). "Using mouse models to dissect the genetics of obesity." Trends Genet **18**(7): 367-76.

Brown, D., B. Lui, et al. (1992). "A plasma membrane proton ATPase in specialized cells of rat epididymis." Am J Physiol **263**(4 Pt 1): C913-6.

Brown, D., P. J. Smith, et al. (1997). "Role of V-ATPase-rich cells in acidification of the male reproductive tract." J Exp Biol **200**(Pt 2): 257-62.

Brown, T. R., D. B. Lubahn, et al. (1988). "Deletion of the steroid-binding domain of the human androgen receptor gene in one family with complete androgen insensitivity syndrome: evidence for further genetic heterogeneity in this syndrome." Proc Natl Acad Sci U S A **85**(21): 8151-5.

Brummelkamp, T. R., R. Bernards, et al. (2002). "A system for stable expression of short interfering RNAs in mammalian cells." Science **296**(5567): 550-3.

Brummelkamp, T. R., A. W. Fabius, et al. (2006). "An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors." Nat Chem Biol **2**(4): 202-6.

Burge, C. and S. Karlin (1997). "Prediction of complete gene structures in human genomic DNA." J Mol Biol **268**(1): 78-94.

Burge, C. B. and S. Karlin (1998). "Finding the genes in genomic DNA." Curr Opin Struct Biol **8**(3): 346-54.

Burgtorf, C., K. Welzel, et al. (1998). "Gridded genomic libraries of different chordate species: a reference library system for basic and comparative genetic studies of chordate genomes." Genomics **52**(2): 230-2.

Butler, A. A. and R. D. Cone (2001). "Knockout models resulting in the development of obesity." Trends Genet **17**(10): S50-4.

Cao, X. M., R. A. Koski, et al. (1990). "Identification and characterization of the Egr-1 gene product, a DNA-binding zinc finger protein induced by differentiation and growth signals." Mol Cell Biol **10**(5): 1931-9.

Caplen, N. J., S. Parrish, et al. (2001). "Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems." Proc Natl Acad Sci U S A **98**(17): 9742-7.

Carr, P. D., D. Verger, et al. (1999). "Chloroplast NADP-malate dehydrogenase: structural basis of light-dependent regulation of activity by thiol oxidation and reduction." Structure **7**(4): 461-75.

Castle, W. E. and C. C. Little (1910). "On a Modified Mendelian Ratio among Yellow Mice." Science **32**(833): 868-870.

Chavrier, P., C. Vesque, et al. (1990). "The segment-specific gene Krox-20 encodes a transcription factor with binding sites in the promoter region of the Hox-1.4 gene." Embo J **9**(4): 1209-18.

Chen, E. Y. and P. H. Seeburg (1985). "Supercoil sequencing: a fast and simple method for sequencing plasmid DNA." DNA **4**(2): 165-70.

Chen, F., C. C. MacDonald, et al. (1995). "Cleavage site determinants in the mammalian polyadenylation

signal." Nucleic Acids Res **23**(14): 2614-20.

Chen, R., J. B. Bouck, et al. (2001). "Comparing vertebrate whole-genome shotgun reads to the human genome." Genome Res **11**(11): 1807-16.

Cheng, H., X. He, et al. (2004). "The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3." Mol Cell Biol **24**(7): 2932-43.
Cheng, Z., Y. Liu, et al. (2004). "Crystal structure of Ski8p, a WD-repeat protein with dual roles in mRNA metabolism and meiotic recombination." Protein Sci **13**(10): 2673-84.

Chinni, S. R. and A. Shisheva (1999). "Arrest of endosome acidification by bafilomycin A1 mimics insulin action on GLUT4 translocation in 3T3-L1 adipocytes." Biochem J **339 (Pt 3)**: 599-606.

Cho, B. N., M. L. McMullen, et al. (2001). "Reproductive deficiencies in transgenic mice expressing the rat inhibin alpha-subunit gene." Endocrinology **142**(11): 4994-5004.

Cho, C., H. Jung-Ha, et al. (2003). "Protamine 2 deficiency leads to sperm DNA damage and embryo death in mice." Biol Reprod **69**(1): 211-7.

Cho, C., W. D. Willis, et al. (2001). "Haploinsufficiency of protamine-1 or -2 causes infertility in mice." Nat Genet **28**(1): 82-6.

Cho, K. J., H. E. Moon, et al. (2003). "Alpha-lipoic acid inhibits adipocyte differentiation by regulating pro-adipogenic transcription factors via mitogen-activated protein kinase pathways." J Biol Chem **278**(37): 34823-33.

Chomczynski, P. and N. Sacchi (1987). "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." Anal Biochem **162**(1): 156-9.

Chothia, C. and M. Gerstein (1997). "Protein evolution. How far can sequences diverge?" Nature **385**(6617): 579, 581.

Chothia, C., T. Hubbard, et al. (1997). "Protein folds in the all-beta and all-alpha classes." Annu Rev Biophys Biomol Struct **26**: 597-627.

Chou, Z. F., F. Chen, et al. (1994). "Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals." Nucleic Acids Res **22**(13): 2525-31.

Church, D. M., C. J. Stotler, et al. (1994). "Isolation of genes from complex sources of mammalian genomic DNA using exon amplification." Nat Genet **6**(1): 98-105.

Clamp, M., B. Fry, et al. (2007). "Distinguishing protein-coding and noncoding genes in the human genome." Proc Natl Acad Sci U S A **104**(49): 19428-33.

Clark, A. B., M. E. Cook, et al. (1999). "Functional analysis of human MutSalpha and MutSbeta complexes in yeast." Nucleic Acids Res **27**(3): 736-42.

Clark, J. M. (1988). "Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases." Nucleic Acids Res **16**(20): 9677-86.

Claros, M. G. and G. von Heijne (1994). "TopPred II: an improved software for membrane protein structure predictions." Comput Appl Biosci **10**(6): 685-6.

Clay, O., S. Caccio, et al. (1996). "Human coding and noncoding DNA: compositional correlations." Mol Phylogenet Evol **5**(1): 2-12.

Cohen, P., C. Zhao, et al. (2001). "Selective deletion of leptin receptor in neurons leads to obesity." J Clin Invest **108**(8): 1113-21.

Coleman, D. L. (1978). "Diabetes and obesity: thrifty mutants?" Nutr Rev **36**(5): 129-32.

Coleman, D. L. (1978). "Obese and diabetes: two mutant genes causing diabetes-obesity syndromes in mice." Diabetologia **14**(3): 141-8.

Colgan, D. F. and J. L. Manley (1997). "Mechanism and regulation of mRNA polyadenylation." Genes Dev **11**(21): 2755-66.

Collins, F. S., A. Patrinos, et al. (1998). "New goals for the U.S. Human Genome Project: 1998-2003." Science **282**(5389): 682-9.

Cone, R. D. (2000). "Haploinsufficiency of the melanocortin-4 receptor: part of a thrifty genotype?" J Clin Invest **106**(2): 185-7.

Conrad, S., A. Viertelhaus, et al. (2001). "Sequencing and tissue distribution of the canine MRP2 gene compared with MRP1 and MDR1." Toxicology **156**(2-3): 81-91.

Costantini, F. and E. Lacy (1981). "Introduction of a rabbit beta-globin gene into the mouse germ line." Nature **294**(5836): 92-4.

Cotten, M., A. Baker, et al. (1994). "Lipopolysaccharide is a frequent contaminant of plasmid DNA preparations and can be toxic to primary human cells in the presence of adenovirus." Gene Ther **1**(4): 239-46.

Coulondre, C., J. H. Miller, et al. (1978). "Molecular basis of base substitution hotspots in Escherichia coli." Nature **274**(5673): 775-80.

Courtois, S. J., D. A. Lafontaine, et al. (1990). "Nuclear factor-I and activator protein-2 bind in a mutually exclusive way to overlapping promoter sequences and trans-activate the human growth hormone gene." Nucleic Acids Res **18**(1): 57-64.

Crabbe, J. C., A. Kosobud, et al. (1983). "Genetic selection for ethanol withdrawal severity: differences in replicate mouse lines." Life Sci **33**(10): 955-62.

Crabbe, J. C., D. Wahlsten, et al. (1999). "Genetics of mouse behavior: interactions with laboratory environment." Science **284**(5420): 1670-2.

Craig, E. A. (2003). "Eukaryotic chaperonins: lubricating the folding of WD-repeat proteins." Curr Biol **13**(23): R904-5.

Crawley, J. N., J. K. Belknap, et al. (1997). "Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies." Psychopharmacology (Berl) **132**(2): 107-24.

Cross, S. H. and A. P. Bird (1995). "CpG islands and genes." Curr Opin Genet Dev **5**(3): 309-14.

Cross, S. H., V. H. Clark, et al. (2000). "CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes." Mamm Genome **11**(5): 373-83.

Cross, S. H., M. Lee, et al. (1997). "The chromosomal distribution of CpG islands in the mouse: evidence for genome scrambling in the rodent lineage." Genomics **40**(3): 454-61.

Cue´not, L., 1902 La loi de Mendel et l'he´re´dite´ de la pigmentation chez les Souris. Arch. Zool. Exp. Gen. 3: xxvii–xxx.

Cue´not, L., 1903 L'he´re´dite´ de la pigmentation chez les Souris. Arch. Zool. Exp. Gen. 4: xxxiii–xli.
Davisson, M. T. (1994). "Rules and guidelines for nomenclature of mouse genes. International Committee on Standardized Genetic Nomenclature for Mice." Gene **147**(2): 157-60.

Davisson, M. T. (1997). "Rules and guidelines for genetic nomenclature in mice: excerpted version. Committee on Standardized Genetic Nomenclature for Mice." Transgenic Res **6**(5): 309-19.

Davuluri, R. V., I. Grosse, et al. (2001). "Computational identification of promoters and first exons in the human genome." Nat Genet **29**(4): 412-7.

De Souza, A. T., X. Dai, et al. (2006). "Transcriptional and phenotypic comparisons of Ppara knockout and siRNA knockdown mice." Nucleic Acids Res **34**(16): 4486-94.

De Val, S., J. P. Anderson, et al. (2004). "Mef2c is activated directly by Ets transcription factors through an evolutionarily conserved endothelial cell-specific enhancer." Dev Biol **275**(2): 424-34.

Deber, C. M., C. Wang, et al. (2001). "TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales." Protein Sci **10**(1): 212-9.

DeBry, R. W. and M. F. Seldin (1996). "Human/mouse homology relationships." Genomics **33**(3): 337-51.

DeChiara, T. M., A. Efstratiadis, et al. (1990). "A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting." Nature **345**(6270): 78-80.

Dehal, P., P. Predki, et al. (2001). "Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution." Science **293**(5527): 104-11.

Deng, C. and M. R. Capecchi (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**(8): 3365-71.

Denhardt, D. T. (1966). "A membrane-filter technique for the detection of complementary DNA." Biochem Biophys Res Commun **23**(5): 641-6.

Dermitzakis, E. T. and A. G. Clark (2002). "Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover." Mol Biol Evol **19**(7): 1114-21.

DeSilva, U., L. Elnitski, et al. (2002). "Generation and comparative analysis of approximately 3.3 Mb of mouse genomic sequence orthologous to the region of human chromosome 7q11.23 implicated in Williams syndrome." Genome Res **12**(1): 3-15.

Deutsch, M. and M. Long (1999). "Intron-exon structures of eukaryotic model organisms." Nucleic Acids Res **27**(15): 3219-28.

Dhar, M., L. S. Webb, et al. (2000). "A novel ATPase on mouse chromosome 7 is a candidate gene for

increased body fat." Physiol Genomics **4**(1): 93-100.

Diament, A. L., J. S. Fisler, et al. (2003). "Studies of natural allele effects in mice can be used to identify genes causing common human obesity." Obes Rev **4**(4): 249-55.

DiAntonio, A., R. W. Burgess, et al. (1993). "Identification and characterization of Drosophila genes for synaptic vesicle proteins." J Neurosci **13**(11): 4924-35.

DiAntonio, A., K. D. Parfitt, et al. (1993). "Synaptic transmission persists in synaptotagmin mutants of Drosophila." Cell **73**(7): 1281-90.

Dietrich, J., X. Hou, et al. (1994). "CD3 gamma contains a phosphoserine-dependent di-leucine motif involved in down-regulation of the T cell receptor." Embo J **13**(9): 2156-66.

Doetschman, T., A. Gossler, et al. (1986). "Introduction of genes into mouse embryonic stem cells." Prog Clin Biol Res **217A**: 47-50.

Doetschman, T., R. G. Gregg, et al. (1987). "Targetted correction of a mutant HPRT gene in mouse embryonic stem cells." Nature **330**(6148): 576-8.

Doggett, N. A., C. L. Smith, et al. (1992). "The effect of DNA concentration on mobility in pulsed field gel electrophoresis." Nucleic Acids Res **20**(4): 859-64.

Dong, Z. M., J. C. Gutierrez-Ramos, et al. (1997). "A new class of obesity genes encodes leukocyte adhesion receptors." Proc Natl Acad Sci U S A **94**(14): 7526-30.

D'Onofrio, G., D. Mouchiroud, et al. (1991). "Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins." J Mol Evol **32**(6): 504-10.

Dorn, G. W., 2nd, N. M. Tepe, et al. (1999). "Low- and high-level transgenic expression of beta2-adrenergic receptors differentially affect cardiac hypertrophy and function in Galphaq-overexpressing mice." Proc Natl Acad Sci U S A **96**(11): 6400-5.

Down, T. A. and T. J. Hubbard (2002). "Computational detection and location of transcription start sites in mammalian genomic DNA." Genome Res **12**(3): 458-61.

Dubchak, I., M. Brudno, et al. (2000). "Active conservation of noncoding sequences revealed by three-way species comparisons." Genome Res **10**(9): 1304-6.

Duret, L. and P. Bucher (1997). "Searching for regulatory elements in human noncoding sequences." Curr Opin Struct Biol **7**(3): 399-406.

Duret, L., M. Semon, et al. (2002). "Vanishing GC-rich isochores in mammalian genomes." Genetics **162**(4): 1837-47.

Dykxhoorn, D. M. and J. Lieberman (2006). "Knocking down disease with siRNAs." Cell **126**(2): 231-5.

Dykxhoorn, D. M. and J. Lieberman (2006). "Running interference: prospects and obstacles to using small interfering RNAs as small molecule drugs." Annu Rev Biomed Eng **8**: 377-402.

Dykxhoorn, D. M. and J. Lieberman (2006). "Silencing viral infection." PLoS Med **3**(7): e242.

Dykxhoorn, D. M., D. Palliser, et al. (2006). "The silent treatment: siRNAs as small molecule drugs." Gene Ther **13**(6): 541-52.

Dykxhoorn, D. M., L. D. Schlehuber, et al. (2006). "Determinants of specific RNA interference-mediated silencing of human beta-globin alleles differing by a single nucleotide polymorphism." Proc Natl Acad Sci U S A **103**(15): 5953-8.

Edwalds-Gilbert, G., K. L. Veraldi, et al. (1997). "Alternative poly(A) site selection in complex transcription units: means to an end?" Nucleic Acids Res **25**(13): 2547-61.

Eicher, E. M. (1981). "Foundation for the future: formal genetics of the mouse." Prog Clin Biol Res **45**: 1-49.

Eisen, J. A. and C. M. Fraser (2003). "Phylogenomics: intersection of evolution and genomics." Science **300**(5626): 1706-7.

Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-8.

Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.

Elbashir, S. M., J. Martinez, et al. (2001). "Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate." Embo J **20**(23): 6877-88.

Elnitski, L., R. C. Hardison, et al. (2003). "Distinguishing regulatory DNA from neutral sites." Genome Res **13**(1): 64-72.

Enkhbayar, P., M. Kamiya, et al. (2004). "Structural principles of leucine-rich repeat (LRR) proteins." Proteins **54**(3): 394-403.

Evans, M. (1981). "Origin of mouse embryonal carcinoma cells and the possibility of their direct isolation into tissue culture." J Reprod Fertil **62**(2): 625-31.

Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-6.

Falquet, L., M. Pagni, et al. (2002). "The PROSITE database, its status in 2002." Nucleic Acids Res **30**(1): 235-8.

Farooqi, I. S., G. S. Yeo, et al. (2000). "Dominant and recessive inheritance of morbid obesity associated with melanocortin 4 receptor deficiency." J Clin Invest **106**(2): 271-9.

Feinberg, A. P. and B. Vogelstein (1983). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity." Anal Biochem **132**(1): 6-13.

Festing, M., K. Kondo, et al. (1972). "International standardized nomenclature for outbred stocks of laboratory animals." Z Versuchstierkd **14**(4): 215-24.

Festing, M. F. (1976). "Phenotypic variability of inbred and outbred mice." Nature **263**(5574): 230-2.

Festing, M. F. (1979). "Properties of inbred strains and outbred stocks, with special reference to toxicity testing." J Toxicol Environ Health **5**(1): 53-68.

Festing, M. F. (1993). "Genetic variation in outbred rats and mice and its implications for toxicological screening." J Exp Anim Sci **35**(5-6): 210-20.

Fickett, J. W. and A. G. Hatzigeorgiou (1997). "Eukaryotic promoter recognition." Genome Res **7**(9): 861-78.

Fickett, J. W. and W. W. Wasserman (2000). "Discovery and modeling of transcriptional regulatory regions." Curr Opin Biotechnol **11**(1): 19-24.

Fire, A., D. Albertson, et al. (1991). "Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle." Development **113**(2): 503-14.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-11.

Foley, K. P., G. A. McArthur, et al. (1998). "Targeted disruption of the MYC antagonist MAD1 inhibits cell cycle exit during granulocyte differentiation." Embo J **17**(3): 774-85.

Follows, G. A., H. Tagoh, et al. (2003). "Differential transcription factor occupancy but evolutionarily conserved chromatin features at the human and mouse M-CSF (CSF-1) receptor loci." Nucleic Acids Res **31**(20): 5805-16.

Forgac, M. (1999). "Structure and properties of the vacuolar (H+)-ATPases." J Biol Chem **274**(19): 12951-4.

Fraser, A. G. and E. M. Marcotte (2004). "Development through the eyes of functional genomics." Curr Opin Genet Dev **14**(4): 336-42.

Fraser, H. B. and A. E. Hirsh (2004). "Evolutionary rate depends on number of protein-protein interactions independently of gene expression level." BMC Evol Biol **4**: 13.

Fraser, H. B., A. E. Hirsh, et al. (2004). "Coevolution of gene expression among interacting proteins." Proc Natl Acad Sci U S A **101**(24): 9033-8.

Fraser, J. A., S. Diezmann, et al. (2004). "Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms." PLoS Biol **2**(12): e384.

Frazer, K. A., X. Chen, et al. (2003). "Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates." Genome Res **13**(3): 341-6.

Frazer, K. A., L. Elnitski, et al. (2003). "Cross-species sequence comparisons: a review of methods and available resources." Genome Res **13**(1): 1-12.

Friedrich, G. and P. Soriano (1991). "Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice." Genes Dev **5**(9): 1513-23.

Fujita, S. and H. Iba (2008). "Putative promoter regions of miRNA genes involved in evolutionarily conserved regulatory systems among vertebrates." Bioinformatics **24**(3): 303-8.

Fujita, S., E. Ota, et al. (2007). "Highly efficient reverse transfection with siRNA in multiple wells of microtiter plates." J Biosci Bioeng **104**(4): 329-33.

Gabilondo, A. M., J. Hegler, et al. (1997). "A dileucine motif in the C terminus of the beta2-adrenergic receptor is involved in receptor internalization." Proc Natl Acad Sci U S A **94**(23): 12285-90.

Gaffney, D. J. and P. D. Keightley (2004). "Unexpected conserved non-coding DNA blocks in mammals." Trends Genet **20**(8): 332-7.

Galli-Taliadoros, L. A., J. D. Sedgwick, et al. (1995). "Gene knock-out technology: a methodological overview for the interested novice." J Immunol Methods **181**(1): 1-15.

Galtier, N., G. Piganeau, et al. (2001). "GC-content evolution in mammalian genomes: the biased gene conversion hypothesis." Genetics **159**(2): 907-11.
Gardiner, K. (1996). "Base composition and gene distribution: critical patterns in mammalian genome organization." Trends Genet **12**(12): 519-24.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.

Gattiker, A., W. V. Bienvenut, et al. (2002). "FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification." Proteomics **2**(10): 1435-44.

Gattiker, A., E. Gasteiger, et al. (2002). "ScanProsite: a reference implementation of a PROSITE scanning tool." Appl Bioinformatics **1**(2): 107-8.

Gautheret, D. and A. Lambert (2001). "Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles." J Mol Biol **313**(5): 1003-11.

Geisler, C., J. Dietrich, et al. (1998). "Leucine-based receptor sorting motifs are dependent on the spacing relative to the plasma membrane." J Biol Chem **273**(33): 21316-23.

Gelfand, M. S., A. A. Mironov, et al. (1996). "Gene recognition via spliced sequence alignment." Proc Natl Acad Sci U S A **93**(17): 9061-6.

Gelfand, M. S., L. I. Podolsky, et al. (1996). "Recognition of genes in human DNA sequences." J Comput Biol **3**(2): 223-34.

Geng, Y., W. Whoriskey, et al. (1999). "Rescue of cyclin D1 deficiency by knockin cyclin E." Cell **97**(6): 767-77.

Georgopoulos, K., D. D. Moore, et al. (1992). "Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment." Science **258**(5083): 808-12.

Gerstein, M. B., C. Bruce, et al. (2007). "What is a gene, post-ENCODE? History and updated definition." Genome Res **17**(6): 669-81.

Ghosh, D., V. Z. Pletnev, et al. (1995). "Structure of human estrogenic 17 beta-hydroxysteroid dehydrogenase at 2.20 A resolution." Structure **3**(5): 503-13.

Giaever, G., A. M. Chu, et al. (2002). "Functional profiling of the Saccharomyces cerevisiae genome." Nature **418**(6896): 387-91.

Giardine, B., L. Elnitski, et al. (2003). "GALA, a database for genomic sequence alignments and annotations." Genome Res **13**(4): 732-41.

Gibbs, R. A., G. M. Weinstock, et al. (2004). "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature **428**(6982): 493-521.

Gil, A. and N. J. Proudfoot (1984). "A sequence downstream of AAUAAA is required for rabbit beta-

globin mRNA 3'-end formation." Nature **312**(5993): 473-4.

Gil, G., M. S. Brown, et al. (1986). "Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the hamster. II. Isolation of the gene and characterization of the 5' flanking region." J Biol Chem **261**(8): 3717-24.

Gil, G., J. L. Goldstein, et al. (1986). "Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the hamster. I. Isolation and sequencing of a full-length cDNA." J Biol Chem **261**(8): 3710-6.
Gingeras, T. R. (2007). "Origin of phenotypes: genes and transcripts." Genome Res **17**(6): 682-90.

Glazko, G. V., E. V. Koonin, et al. (2003). "A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions." Trends Genet **19**(3): 119-24.

Glazko, G. V. and M. Nei (2003). "Estimation of divergence times for major lineages of primate species." Mol Biol Evol **20**(3): 424-34.

Goncalves, I., L. Duret, et al. (2000). "Nature and structure of human genes that generate retropseudogenes." Genome Res **10**(5): 672-8.

Gonzalez-Gaitan, M. and H. Jackle (1997). "Role of Drosophila alpha-adaptin in presynaptic vesicle recycling." Cell **88**(6): 767-76.

Goodier, J. L., E. M. Ostertag, et al. (2001). "A novel active L1 retrotransposon subfamily in the mouse." Genome Res **11**(10): 1677-85.

Gordon, J. W., G. A. Scangos, et al. (1980). "Genetic transformation of mouse embryos by microinjection of purified DNA." Proc Natl Acad Sci U S A **77**(12): 7380-4.

Gossler, A., T. Doetschman, et al. (1986). "Transgenesis by means of blastocyst-derived embryonic stem cell lines." Proc Natl Acad Sci U S A **83**(23): 9065-9.

Gottgens, B., L. M. Barton, et al. (2002). "Transcriptional regulation of the stem cell leukemia gene (SCL)--comparative analysis of five vertebrate SCL loci." Genome Res **12**(5): 749-59.

Greene-Till, R., Y. Zhao, et al. (2000). "Gene flow of unique sequences between Mus musculus domesticus and Mus spretus." Mamm Genome **11**(3): 225-30.

Grunstein, M. and D. S. Hogness (1975). "Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene." Proc Natl Acad Sci U S A **72**(10): 3961-5.

GuhaThakurta, D. (2006). "Computational identification of transcriptional regulatory elements in DNA sequence." Nucleic Acids Res **34**(12): 3585-98.

Gustincich, S., A. Sandelin, et al. (2006). "The complexity of the mammalian transcriptome." J Physiol **575**(Pt 2): 321-32.

Haas, B. J., N. Volfovsky, et al. (2002). "Full-length messenger RNA sequences greatly improve genome annotation." Genome Biol **3**(6): RESEARCH0029.

Haber, D. A., S. M. Beverley, et al. (1981). "Properties of an altered dihydrofolate reductase encoded by amplified genes in cultured mouse fibroblasts." J Biol Chem **256**(18): 9501-10.

Haft, C. R., M. De La Luz Sierra, et al. (1998). "Analysis of the juxtamembrane dileucine motif in the insulin receptor." Endocrinology **139**(4): 1618-29.

Haft, C. R., R. D. Klausner, et al. (1994). "Involvement of dileucine motifs in the internalization and degradation of the insulin receptor." J Biol Chem **269**(42): 26286-94.

Hagele, K. (1985). "Identification of a polytene chromosome band containing a male sex determiner of Chironomus thummi thummi." Chromosoma **91**(3-4): 167-71.

Haldane, J. B. S., A. D. Sprunt and N. M. Haldane, 1915 Reduplication in mice. J. Genet. 5: 133–135. Haldane, J. B. S., 1933 The genetics of cancer. Nature 132: 265–267.

Hanahan, D. (1985). "Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes." Nature **315**(6015): 115-22.

Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-51.

Hans, H. and J. C. Alwine (2000). "Functionally significant secondary structure of the simian virus 40 late polyadenylation signal." Mol Cell Biol **20**(8): 2926-32.

Hardison, R., J. L. Slightom, et al. (1997). "Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights." Gene **205**(1-2): 73-94.

Hardison, R. C. (2000). "Conserved noncoding sequences are reliable guides to regulatory elements." Trends Genet **16**(9): 369-72.

Hardison, R. C. (2003). "Comparative genomics." PLoS Biol **1**(2): E58.

Hardison, R. C., J. Oeltjen, et al. (1997). "Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome." Genome Res **7**(10): 959-66.

Hardison, R. C., K. M. Roskin, et al. (2003). "Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution." Genome Res **13**(1): 13-26.

Hasse, A. and W. A. Schulz (1994). "Enhancement of reporter gene de novo methylation by DNA fragments from the alpha-fetoprotein control region." J Biol Chem **269**(3): 1821-6.

Hasty, P., J. Rivera-Perez, et al. (1991). "The length of homology required for gene targeting in embryonic stem cells." Mol Cell Biol **11**(11): 5586-91.

Hasty, P., J. Rivera-Perez, et al. (1991). "Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells." Mol Cell Biol **11**(9): 4509-17.

Haubold, B. and T. Wiehe (2004). "Comparative genomics: methods and applications." Naturwissenschaften **91**(9): 405-21.

Hausberger, F. X. and B. C. Hausberger (1966). "Castration-induced obesity in mice. Body composition, histology of adrenal cortex and islets of Langerhans in castrated mice." Acta Endocrinol (Copenh) **53**(4): 571-83.

He, M., A. Wilde, et al. (1990). "A simple single-step procedure for small-scale preparation of Escherichia coli plasmids." Nucleic Acids Res **18**(6): 1660.

Heilker, R., U. Manning-Krieg, et al. (1996). "In vitro binding of clathrin adaptors to sorting signals correlates with endocytosis and basolateral sorting." Embo J **15**(11): 2893-9.

Heinlein, C. A. and C. Chang (2002). "Androgen receptor (AR) coregulators: an overview." Endocr Rev **23**(2): 175-200.

Heinlein, C. A. and C. Chang (2002). "The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions." Mol Endocrinol **16**(10): 2181-7.

Heller, C. and F. M. Pohl (1989). "A systematic study of field inversion gel electrophoresis." Nucleic Acids Res **17**(15): 5989-6003.

Heller, H. and E. Bengal (1998). "TFIID (TBP) stabilizes the binding of MyoD to its DNA site at the promoter and MyoD facilitates the association of TFIIB with the preinitiation complex." Nucleic Acids Res **26**(9): 2112-9.

Hennig, Wolfgang. Genetik; Springer-Lehrbuch 3., überarb. u. erw. Aufl., 2002, XXVIII, 853 S. 462 illus., Geb. ISBN: 3-540-42958-1.

Herak-Kramberger, C. M., S. Breton, et al. (2001). "Distribution of the vacuolar H+ atpase along the rat and human male reproductive tract." Biol Reprod **64**(6): 1699-707.

Hirokawa, T., S. Boon-Chieng, et al. (1998). "SOSUI: classification and secondary structure prediction system for membrane proteins." Bioinformatics **14**(4): 378-9.

Ho, D. N., G. A. Coburn, et al. (2002). "The crystal structure and mutational analysis of a novel RNA-binding domain found in the human Tap nuclear mRNA export factor." Proc Natl Acad Sci U S A **99**(4): 1888-93.

Hofacker, I. L. (2003). "Vienna RNA secondary structure server." Nucleic Acids Res **31**(13): 3429-31.

Hofacker, I. L., M. Fekete, et al. (2002). "Secondary structure prediction for aligned RNA sequences." J Mol Biol **319**(5): 1059-66.

Hofmann, K. & Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments; Biol. Chem. Hoppe-Seyler 374,166

Holder, J. L., Jr., L. Zhang, et al. (2004). "Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice." Am J Physiol Endocrinol Metab **287**(1): E105-13.

Hooper, D. C., J. L. Young, et al. (1987). "Murine T cells reactive against autologous erythrocytes: evidence for in vitro and in vivo priming with mouse and rat red blood cells." Cell Immunol **106**(1): 53-61.

Hotz-Wagenblatt, A., T. Hankeln, et al. (2003). "ESTAnnotator: A tool for high throughput EST annotation." Nucleic Acids Res **31**(13): 3716-9.

Hummel, K. P., D. L. Coleman, et al. (1972). "The influence of genetic background on expression of mutations at the diabetes locus in the mouse. I. C57BL-KsJ and C57BL-6J strains." Biochem Genet **7**(1): 1-13.

Hurst, L. D. and E. J. Williams (2000). "Covariation of GC content and the silent site substitution rate in rodents: implications for methodology and for the evolution of isochores." Gene **261**(1): 107-14.

Imanishi, T., T. Itoh, et al. (2004). "Integrative annotation of 21,037 human genes validated by full-length

cDNA clones." PLoS Biol **2**(6): e162.

Ingalls, A. M., M. M. Dickie, et al. (1950). "Obese, a new mutation in the house mouse." J Hered **41**(12): 317-8.

Itateyama, E., S. Chiba, et al. (2003). "Hypothalamic Neuronal Histamine in Genetically Obese Animals: Its Implication of Leptin Action in the Brain." Experimental Biology and Medicine **228**(10): 1132-1137.

Itoi-Babaya, M., H. Ikegami, et al. (2007). "Fatty liver and obesity: phenotypically correlated but genetically distinct traits in a mouse model of type 2 diabetes." Diabetologia **50**(8): 1641-8.

Izant, J. G. and H. Weintraub (1984). "Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis." Cell **36**(4): 1007-15.

Jamet, E. (2004). "Bioinformatics as a critical prerequisite to transcriptome and proteome studies." J Exp Bot **55**(405): 1977-9.

Jin, F., J. Dai, et al. (2004). "A novel human gene (WDR25) encoding a 7-WD40-containing protein maps on 14q32." Biochem Genet **42**(11-12): 419-27.

Jones, N. and G. A. Harrison (1957). "Genetically determined obesity and sterility in the mouse." Proc Soc Study Fertil **9**: 51-64.

Joseph, L. J., M. M. Le Beau, et al. (1988). "Molecular cloning, sequencing, and mapping of EGR2, a human early growth response gene encoding a protein with "zinc-binding finger" structure." Proc Natl Acad Sci U S A **85**(19): 7164-8.

Jurka, J. (2004). "Evolutionary impact of human Alu repetitive elements." Curr Opin Genet Dev **14**(6): 603-8.

Jurka, J., O. Kohany, et al. (2004). "Duplication, coclustering, and selection of human Alu retrotransposons." Proc Natl Acad Sci U S A **101**(5): 1268-72.

Jurka, J. and C. Pethiyagoda (1995). "Simple repetitive DNA sequences from primates: compilation and analysis." J Mol Evol **40**(2): 120-6.

Jurka, J., E. Zietkiewicz, et al. (1995). "Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era." Nucleic Acids Res **23**(1): 170-5.

Kaiser, M. W., N. Lyamicheva, et al. (1999). "A comparison of eubacterial and archaeal structure-specific 5'-exonucleases." J Biol Chem **274**(30): 21387-94.

Kajava, A. V. (2001). "Review: proteins with repeated sequence--structural prediction and modeling." J Struct Biol **134**(2-3): 132-44.

Kajava, A. V. and B. Kobe (2002). "Assessment of the ability to model proteins with leucine-rich repeats in light of the latest structural information." Protein Sci **11**(5): 1082-90.

Kamath, R. S. and J. Ahringer (2003). "Genome-wide RNAi screening in Caenorhabditis elegans." Methods **30**(4): 313-21.

Kamath, R. S., A. G. Fraser, et al. (2003). "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi." Nature **421**(6920): 231-7.

Kamath, R. S., M. Martinez-Campos, et al. (2001). "Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans." Genome Biol **2**(1): RESEARCH0002.

Kane, P. M. and A. M. Smardon (2003). "Assembly and regulation of the yeast vacuolar H+-ATPase." J Bioenerg Biomembr **35**(4): 313-21.

Kappen, C. (2000). "Analysis of a complete homeobox gene repertoire: implications for the evolution of diversity." Proc Natl Acad Sci U S A **97**(9): 4481-6.
Kasai, S., S. Chuma, et al. (2003). "Haploinsufficiency of Bcl-x leads to male-specific defects in fetal germ cells: differential regulation of germ cell apoptosis between the sexes." Dev Biol **264**(1): 202-16.

Kawabe, H., T. Sakisaka, et al. (2003). "A novel rabconnectin-3-binding protein that directly binds a GDP/GTP exchange protein for Rab3A small G protein implicated in Ca(2+)-dependent exocytosis of neurotransmitter." Genes Cells **8**(6): 537-46.

Kemnitz, J. W., R. W. Goy, et al. (1977). "Effects of gonadectomy on hypothalamic obesity in male and female rats." Int J Obes **1**(3): 259-70.

Kido, Y., D. J. Burks, et al. (2000). "Tissue-specific insulin resistance in mice with mutations in the insulin receptor, IRS-1, and IRS-2." J Clin Invest **105**(2): 199-205.

Kido, Y., N. Philippe, et al. (2000). "Genetic modifiers of the insulin resistance phenotype in mice." Diabetes **49**(4): 589-96.

Kil, S. J. and C. Carlin (2000). "EGF receptor residues leu(679), leu(680) mediate selective sorting of ligand-receptor complexes in early endosomal compartments." J Cell Physiol **185**(1): 47-60.

Kipling, D. and P. E. Warburton (1997). "Centromeres, CENP-B and Tigger too." Trends Genet **13**(4): 141-5.

Kloster, M. (2005). "Analysis of evolution through competitive selection." Phys Rev Lett **95**(16): 168701.

Kloster, M. and C. Tang (2004). "Simulation and analysis of in vitro DNA evolution." Phys Rev Lett **92**(3): 038101.

Kloster, M., C. Tang, et al. (2005). "Finding regulatory modules through large-scale gene-expression data analysis." Bioinformatics **21**(7): 1172-9.

Klysik, J. (2002). "Mice and humans: chromosome engineering and its application to functional genomics." Acta Biochim Pol **49**(3): 553-69.

Knudsen, S. (1999). "Promoter2.0: for the recognition of PolII promoter sequences." Bioinformatics **15**(5): 356-61.

Kobe, B. and J. Deisenhofer (1994). "The leucine-rich repeat: a versatile binding motif." Trends Biochem Sci **19**(10): 415-21.

Kobe, B. and A. V. Kajava (2001). "The leucine-rich repeat as a protein recognition motif." Curr Opin Struct Biol **11**(6): 725-32.

Kola, I., S. Brookes, et al. (1993). "The Ets1 transcription factor is widely expressed during murine embryo development and is associated with mesodermal cells involved in morphogenetic processes such as organ formation." Proc Natl Acad Sci U S A **90**(16): 7588-92.

Kolbe, D., J. Taylor, et al. (2004). "Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat." Genome Res **14**(4): 700-7.

Kolbe, J. J., R. E. Glor, et al. (2004). "Genetic variation increases during biological invasion by a Cuban lizard." Nature **431**(7005): 177-81.

Koller, B. H., L. J. Hagemann, et al. (1989). "Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells." Proc Natl Acad Sci U S A **86**(22): 8927-31.

Koller, B. H. and O. Smithies (1989). "Inactivating the beta 2-microglobulin locus in mouse embryonic stem cells by homologous recombination." Proc Natl Acad Sci U S A **86**(22): 8932-5.

Konu, O., J. K. Kane, et al. (2001). "Region-specific transcriptional response to chronic nicotine in rat brain." Brain Res **909**(1-2): 194-203.

Koop, B. F. (1995). "Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution." Trends Genet **11**(9): 367-71.

Kraemer, C., T. Enklaar, et al. (2000). "Mapping and structure of DMXL1, a human homologue of the DmX gene from Drosophila melanogaster coding for a WD repeat protein." Genomics **64**(1): 97-101.

Kraemer, C. (1994) Molekulare Analyse der geschlechtsbestimmenden Region von Chironomiden Mainz,Univ., Diss.

Kraemer, C. and E. R. Schmidt (1993). "The sex determining region of Chironomus thummi is associated with highly repetitive DNA and transposable elements." Chromosoma **102**(8): 553-62.

Kraemer, C., B. Weil, et al. (1998). "The new gene DmX from Drosophila melanogaster encodes a novel WD-repeat protein." Gene **216**(2): 267-76.

Kragler, F., A. Langeder, et al. (1993). "Two independent peroxisomal targeting signals in catalase A of Saccharomyces cerevisiae." J Cell Biol **120**(3): 665-73.

Kreil, G. (1984). "Occurrence, detection, and biosynthesis of carboxy-terminal amides." Methods Enzymol **106**: 218-23.

Kuehn, M. R., A. Bradley, et al. (1987). "A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice." Nature **326**(6110): 295-8.

Kumar, K. G., L. O. Byerley, et al. (2008). "Genetic variation in Glp1r expression influences the rate of gastric emptying in mice." Am J Physiol Regul Integr Comp Physiol **294**(2): R362-71.

Kumar, S. and S. B. Hedges (1998). "A molecular timescale for vertebrate evolution." Nature **392**(6679): 917-20.

Kumar, T. R., A. L. Wiseman, et al. (2000). "Reproductive defects in gamma-glutamyl transpeptidase-deficient mice." Endocrinology **141**(11): 4270-7.

Kunath, T., G. Gish, et al. (2003). "Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype." Nat Biotechnol **21**(5): 559-61.

Kunz, S. (2000), Struktur der geschlechtsbestimmenden Region von Chironomus thummi (Diptera, Insecta) und Vergleich der Gene CtY und CpY. Mainz,Univ., Diss.

Lai, E. C., C. Burks, et al. (1998). "The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts." Development **125**(20): 4077-88.

Lai, E. C. and J. W. Posakony (1997). "The Bearded box, a novel 3' UTR sequence motif, mediates negative post-transcriptional regulation of Bearded and Enhancer of split Complex gene expression." Development **124**(23): 4847-56.

Lakso, M., B. Sauer, et al. (1992). "Targeted oncogene activation by site-specific recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(14): 6232-6.

Lakso, M., P. S. Steeg, et al. (1992). "Embryonic expression of nm23 during mouse organogenesis." Cell Growth Differ **3**(12): 873-9.

Lambright, D. G., J. Sondek, et al. (1996). "The 2.0 A crystal structure of a heterotrimeric G protein." Nature **379**(6563): 311-9.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Langley, K. E., M. R. Villarejo, et al. (1975). "Molecular basis of beta-galactosidase alpha-complementation." Proc Natl Acad Sci U S A **72**(4): 1254-7.

Laporte, L., J. Stultz, et al. (1997). "Solution conformations and interactions of alpha and beta subunits of the Oxytricha nova telomere binding protein: investigation by Raman spectroscopy." Biochemistry **36**(26): 8053-9.

Legendre, M. and D. Gautheret (2003). "Sequence determinants in human polyadenylation site selection." BMC Genomics **4**(1): 7.

Leiter, E. H. (2002). "Mice with targeted gene disruptions or gene insertions for diabetes research: problems, pitfalls, and potential solutions." Diabetologia **45**(3): 296-308.

Lemaire, P., C. Vesque, et al. (1990). "The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptional activator." Mol Cell Biol **10**(7): 3456-67.

Lennon, G., C. Auffray, et al. (1996). "The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression." Genomics **33**(1): 151-2.

Letourneur, F. and R. D. Klausner (1992). "A novel di-leucine motif and a tyrosine-based motif independently mediate lysosomal targeting and endocytosis of CD3 chains." Cell **69**(7): 1143-57.

Letunic, I., L. Goodstadt, et al. (2002). "Recent improvements to the SMART domain-based sequence annotation resource." Nucleic Acids Res **30**(1): 242-4.

Leung, Y. F. and C. P. Pang (2002). "EYE on bioinformatics: dissecting complex disease traits in silico." Appl Bioinformatics **1**(2): 69-80.

Li, B., J. C. Ruiz, et al. (2002). "CUL-4A is critical for early embryonic development." Mol Cell Biol **22**(14): 4997-5005.

Li, D. and R. Roberts (2001). "WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases." Cell Mol Life Sci **58**(14): 2085-97.

Liakopoulos, T. D., C. Pasquier, et al. (2001). "A novel tool for the prediction of transmembrane protein topology based on a statistical analysis of the SwissProt database: the OrienTM algorithm." Protein Eng **14**(6): 387-90.

Liang, F., M. Han, et al. (1998). "Homology-directed repair is a major double-strand break repair pathway in mammalian cells." Proc Natl Acad Sci U S A **95**(9): 5172-7.

Lin, B., J. T. White, et al. (2003). "Isolation and characterization of human and mouse WDR19,a novel WD-repeat protein exhibiting androgen-regulated expression in prostate epithelium." Genomics **82**(3): 331-42.

Liu, R. and D. J. States (2002). "Consensus promoter identification in the human genome utilizing expressed gene markers and gene modeling." Genome Res **12**(3): 462-9.

Liu, X., R. R. Driskell, et al. (2004). "Characterization of Lef-1 Promoter Segments that Facilitate Inductive Developmental Expression in Skin." J Invest Dermatol **123**(2): 264-274.

Lodowski, D. T., J. F. Barnhill, et al. (2003). "Purification, crystallization and preliminary X-ray diffraction studies of a complex between G protein-coupled receptor kinase 2 and Gbeta1gamma2." Acta Crystallogr D Biol Crystallogr **59**(Pt 5): 936-9.

Lodowski, D. T., J. A. Pitcher, et al. (2003). "Keeping G proteins at bay: a complex between G protein-coupled receptor kinase 2 and Gbetagamma." Science **300**(5623): 1256-62.

Long, J., I. Matsuura, et al. (2003). "Repression of Smad transcriptional activity by PIASy, an inhibitor of activated STAT." Proc Natl Acad Sci U S A **100**(17): 9791-6.

Long, M. and M. Deutsch (1999). "Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns." Mol Biol Evol **16**(11): 1528-34.

Long, M. B. and B. A. Sullenger (1999). "Evaluating group I intron catalytic efficiency in mammalian cells." Mol Cell Biol **19**(10): 6479-87.

Loots, G. G., R. M. Locksley, et al. (2000). "Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons." Science **288**(5463): 136-40.

Loots, G. G. and I. Ovcharenko (2004). "rVISTA 2.0: evolutionary analysis of transcription factor binding sites." Nucleic Acids Res **32**(Web Server issue): W217-21.

Loots, G. G., I. Ovcharenko, et al. (2002). "rVista for comparative sequence-based discovery of functional transcription factor binding sites." Genome Res **12**(5): 832-9.

Lubrano-Berthelier, C., M. Cavazos, et al. (2003). "The human MC4R promoter: characterization and role in obesity." Diabetes **52**(12): 2996-3000.

Ludwig, M. Z. (2002). "Functional evolution of noncoding DNA." Curr Opin Genet Dev **12**(6): 634-9.

Lukacsovich, T., Z. Asztalos, et al. (2001). "Dual-tagging gene trap of novel genes in Drosophila melanogaster." Genetics **157**(2): 727-42.

Lukacsovich, T., B. C. Waldman, et al. (2001). "Efficient recruitment of transfected DNA to a

homologous chromosomal target in mammalian cells." Biochim Biophys Acta **1521**(1-3): 89-96.

Lukacsovich, T. and D. Yamamoto (2001). "Trap a gene and find out its function: toward functional genomics in Drosophila." J Neurogenet **15**(3-4): 147-68.

Luscombe, N. M., D. Greenbaum, et al. (2001). "What is bioinformatics? A proposed definition and overview of the field." Methods Inf Med **40**(4): 346-58.

Lyon, M. F. (1961). "Gene action in the X-chromosome of the mouse (Mus musculus L.)." Nature **190**: 372-3.

Lyon, M. F. (1990). "Genetics. Evolution of the X chromosome." Nature **348**(6302): 585-6.

Lyon, M. F. (1990). "L. C. Dunn and mouse genetic mapping." Genetics **125**(2): 231-6.
MacDonald, C. C., J. Wilusz, et al. (1994). "The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location." Mol Cell Biol **14**(10): 6647-54.

Macdonald, L. E., R. K. Durbin, et al. (1994). "Characterization of two types of termination signal for bacteriophage T7 RNA polymerase." J Mol Biol **238**(2): 145-58.

Madera, M., C. Vogel, et al. (2004). "The SUPERFAMILY database in 2004: additions and improvements." Nucleic Acids Res **32**(Database issue): D235-9.

Madrona, A. Y. and D. K. Wilson (2004). "The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure." Protein Sci **13**(6): 1557-65.

Majewski, J. and J. Ott (2002). "Distribution and characterization of regulatory elements in the human genome." Genome Res **12**(12): 1827-36.

Makalowski, W. and M. S. Boguski (1998). "Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences." Proc Natl Acad Sci U S A **95**(16): 9407-12.

Makalowski, W., J. Zhang, et al. (1996). "Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences." Genome Res **6**(9): 846-57.

Malide, D., N. K. Dwyer, et al. (1997). "Immunocytochemical evidence that GLUT4 resides in a specialized translocation post-endosomal VAMP2-positive compartment in rat adipose cells in the absence of insulin." J Histochem Cytochem **45**(8): 1083-96.

Malide, D., J. F. St-Denis, et al. (1997). "Vp165 and GLUT4 share similar vesicle pools along their trafficking pathways in rat adipose cells." FEBS Lett **409**(3): 461-8.

Maltais, L. J., J. A. Blake, et al. (1997). "Rules and guidelines for mouse gene nomenclature: a condensed version. International Committee on Standardized Genetic Nomenclature for Mice." Genomics **45**(2): 471-6.

Mansour, S. L., K. R. Thomas, et al. (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**(6197): 348-52.

Margulies, E. H., G. M. Cooper, et al. (2007). "Analyses of deep mammalian sequence alignments and

constraint predictions for 1% of the human genome." Genome Res **17**(6): 760-74.

Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proc Natl Acad Sci U S A **78**(12): 7634-8.

Marzioch, M., R. Erdmann, et al. (1994). "PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes." Embo J **13**(20): 4908-18.

Matsukura, S., P. A. Jones, et al. (2003). "Establishment of conditional vectors for hairpin siRNA knockdowns." Nucleic Acids Res **31**(15): e77.

Matsuo, K., O. Clay, et al. (1993). "Evidence for erosion of mouse CpG islands during mammalian evolution." Somat Cell Mol Genet **19**(6): 543-55.
Matveeva, O., Y. Nechipurenko, et al. (2007). "Comparison of approaches for rational siRNA design leading to a new efficient and transparent method." Nucleic Acids Res **35**(8): e63.

Mauxion, F., R. Le Borgne, et al. (1996). "A casein kinase II phosphorylation site in the cytoplasmic domain of the cation-dependent mannose 6-phosphate receptor determines the high affinity interaction of the AP-1 Golgi assembly proteins with membranes." J Biol Chem **271**(4): 2171-8.

Mayor, C., M. Brudno, et al. (2000). "VISTA : visualizing global DNA sequence alignments of arbitrary length." Bioinformatics **16**(11): 1046-7.

McDevitt, M. A., M. J. Imperiale, et al. (1984). "Requirement of a downstream sequence for generation of a poly(A) addition site." Cell **37**(3): 993-9.

McDonell, M. W., M. N. Simon, et al. (1977). "Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels." J Mol Biol **110**(1): 119-46.

McLauchlan, J., D. Gaffney, et al. (1985). "The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini." Nucleic Acids Res **13**(4): 1347-68.

Medstrand, P., L. N. van de Lagemaat, et al. (2002). "Retroelement distributions in the human genome: variations associated with age and proximity to genes." Genome Res **12**(10): 1483-95.

Merzendorfer, H., R. Graf, et al. (1997). "Regulation of proton-translocating V-ATPases." J Exp Biol **200**(Pt 2): 225-35.

Michaud, J. L., F. Boucher, et al. (2001). "Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus." Hum Mol Genet **10**(14): 1465-73.

Miklos, G. L. and G. M. Rubin (1996). "The role of the genome project in determining gene function: insights from model organisms." Cell **86**(4): 521-9.

Mikyas, Y., M. Makabi, et al. (2004). "Cryoelectron microscopy imaging of recombinant and tissue derived vaults: localization of the MVP N termini and VPARP." J Mol Biol **344**(1): 91-105.

Millar, C. A., L. C. Campbell, et al. (1997). "Compartment-ablation studies of GLUT4 distribution in adipocytes: evidence for multiple intracellular pools." Biochem Soc Trans **25**(3): 974-7.

Millar, C. A., D. E. James, et al. (1997). "Characterisation of proteins associated with the GLUT4

intracellular compartment." Biochem Soc Trans **25**(3): 465S.

Miller, R. A., J. M. Harper, et al. (2002). "Big mice die young: early life body weight predicts longevity in genetically heterogeneous mice." Aging Cell **1**(1): 22-9.

Miller, W., K. D. Makova, et al. (2004). "Comparative genomics." Annu Rev Genomics Hum Genet **5**: 15-56.

Molnar, A. and K. Georgopoulos (1994). "The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins." Mol Cell Biol **14**(12): 8292-303.

Molnar, P. and L. J. Murphy (1994). "Effects of oestrogen on rat uterine expression of insulin-like growth factor-binding proteins." J Mol Endocrinol **13**(1): 59-67.

Mombaerts, P., J. Iacomini, et al. (1992). "RAG-1-deficient mice have no mature B and T lymphocytes." Cell **68**(5): 869-77.

Montgomery, M. K. and A. Fire (1998). "Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression." Trends Genet **14**(7): 255-8.

Montgomery, M. K., S. Xu, et al. (1998). "RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans." Proc Natl Acad Sci U S A **95**(26): 15502-7.

Mora-Garcia, P. and K. M. Sakamoto (2000). "Granulocyte colony-stimulating factor induces Egr-1 up-regulation through interaction of serum response element-binding proteins." J Biol Chem **275**(29): 22418-26.

Mouchiroud, D., C. Gautier, et al. (1988). "The compositional distribution of coding sequences and DNA molecules in humans and murids." J Mol Evol **27**(4): 311-20.

Mounsey, A., P. Bauer, et al. (2002). "Evidence suggesting that a fifth of annotated Caenorhabditis elegans genes may be pseudogenes." Genome Res **12**(5): 770-5.

Mounzih, K., R. Lu, et al. (1997). "Leptin treatment rescues the sterility of genetically obese ob/ob males." Endocrinology **138**(3): 1190-3.

Mujica, A. O. 2004 Comparative genomic sequencing analysis of a region in human chromosome 11p15.3, mouse chromosome 7 and analysis of the novel STK33, Stk33 gene. Mainz,Univ., Diss.

Mukai, S., K. Ghaedi, et al. (2002). "Intracellular localization, function, and dysfunction of the peroxisome-targeting signal type 2 receptor, Pex7p, in mammalian cells." J Biol Chem **277**(11): 9548-61.

Munclinger, P., P. Boursot, et al. (2003). "B1 insertions as easy markers for mouse population studies." Mamm Genome **14**(6): 359-66.

Mural, R. J., M. D. Adams, et al. (2002). "A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome." Science **296**(5573): 1661-71.

Murnane, J. P. and J. F. Morales (1995). "Use of a mammalian interspersed repetitive (MIR) element in the coding and processing sequences of mammalian genes." Nucleic Acids Res **23**(15): 2837-9.

Murphy, W. J., R. Stanyon, et al. (2001). "Evolution of mammalian genome organization inferred from comparative gene mapping." Genome Biol **2**(6): REVIEWS0005.

Murray, K. B., W. R. Taylor, et al. (2004). "Toward the detection and validation of repeats in protein structure." Proteins **57**(2): 365-80.

Murzin, A. G. (1992). "Structural principles for the propeller assembly of beta-sheets: the preference for seven-fold symmetry." Proteins **14**(2): 191-201.

Nagano, F., H. Kawabe, et al. (2002). "Rabconnectin-3, a novel protein that binds both GDP/GTP exchange protein and GTPase-activating protein for Rab3 small G protein family." J Biol Chem **277**(12): 9629-32.

Nagy, A., J. Rossant, et al. (1993). "Derivation of completely cell culture-derived mice from early-passage embryonic stem cells." Proc Natl Acad Sci U S A **90**(18): 8424-8.
Nardone, J., D. U. Lee, et al. (2004). "Bioinformatics for the 'bench biologist': how to find regulatory regions in genomic DNA." Nat Immunol **5**(8): 768-74.

Naveilhan, P., H. Hassani, et al. (1999). "Normal feeding behavior, body weight and leptin response require the neuropeptide Y Y2 receptor." Nat Med **5**(10): 1188-93.

Neer, E. J., C. J. Schmidt, et al. (1994). "The ancient regulatory-protein family of WD-repeat proteins." Nature **371**(6495): 297-300.

Nellen, W. and C. Lichtenstein (1993). "What makes an mRNA anti-sense-itive?" Trends Biochem Sci **18**(11): 419-23.

Nishi, T. and M. Forgac (2002). "The vacuolar (H+)-ATPases--nature's most versatile proton pumps." Nat Rev Mol Cell Biol **3**(2): 94-103.

Nobrega, M. A., I. Ovcharenko, et al. (2003). "Scanning human gene deserts for long-range enhancers." Science **302**(5644): 413.

Oeltjen, J. C., T. M. Malley, et al. (1997). "Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains." Genome Res **7**(4): 315-29.

Oh, S. Y., M. Y. Lee, et al. (2005). "Alternative usages of multiple promoters of the acetyl-CoA carboxylase beta gene are related to differential transcriptional regulation in human and rodent tissues." J Biol Chem **280**(7): 5909-16.

Ohtsuka, M., K. Ishii, et al. (2006). "Construction of mouse 129/Ola BAC library for targeting experiments using E14 embryonic stem cells." Genes Genet Syst **81**(2): 143-6.

Oliver, J. L., P. Carpena, et al. (2002). "Isochore chromosome maps of the human genome." Gene **300**(1-2): 117-27.

Ono, S. (2003). "Regulation of actin filament dynamics by actin depolymerizing factor/cofilin and actin-interacting protein 1: new blades for twisted filaments." Biochemistry **42**(46): 13363-70.

Onyango, P., W. Miller, et al. (2000). "Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain." Genome Res **10**(11): 1697-710.

Orban, P. C., D. Chui, et al. (1992). "Tissue- and site-specific DNA recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(15): 6861-5.

Orlicky, S., X. Tang, et al. (2003). "Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase." Cell **112**(2): 243-56.

Ouzounis, C. A. and A. Valencia (2003). "Early bioinformatics: the birth of a discipline--a personal view." Bioinformatics **19**(17): 2176-90.

Ovcharenko, I., D. Boffelli, et al. (2004). "eShadow: a tool for comparing closely related sequences." Genome Res **14**(6): 1191-8.

Ovcharenko, I., G. G. Loots, et al. (2005). "Mulan: multiple-sequence local alignment and visualization for studying function and evolution." Genome Res **15**(1): 184-94.

Ovcharenko, I., G. G. Loots, et al. (2004). "zPicture: dynamic alignment and visualization tool for analyzing conservation profiles." Genome Res **14**(3): 472-7.

Ovcharenko, I., M. A. Nobrega, et al. (2004). "ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes." Nucleic Acids Res **32**(Web Server issue): W280-6.

Ovcharenko, I., L. Stubbs, et al. (2004). "Interpreting mammalian evolution using Fugu genome comparisons." Genomics **84**(5): 890-5.

Paces, J., R. Zika, et al. (2004). "Representing GC variation along eukaryotic chromosomes." Gene **333**: 135-41.

Paddison, P. J., A. A. Caudy, et al. (2002). "Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells." Genes Dev **16**(8): 948-58.

Paddison, P. J., A. A. Caudy, et al. (2002). "Stable suppression of gene expression by RNAi in mammalian cells." Proc Natl Acad Sci U S A **99**(3): 1443-8.

Paddison, P. J. and G. J. Hannon (2002). "RNA interference: the new somatic cell genetics?" Cancer Cell **2**(1): 17-23.

Pagni, M., V. Ioannidis, et al. (2004). "MyHits: a new interactive resource for protein annotation and domain identification." Nucleic Acids Res **32**(Web Server issue): W332-5.

Pagni, M., C. Iseli, et al. (2001). "trEST, trGEN and Hits: access to databases of predicted protein sequences." Nucleic Acids Res **29**(1): 148-51.

Paigen, K. (2003). "One hundred years of mouse genetics: an intellectual history. I. The classical period (1902-1980)." Genetics **163**(1): 1-7.

Paigen, K. (2003). "One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981-2002)." Genetics **163**(4): 1227-35.

Palmiter, R. D., R. L. Brinster, et al. (1982). "Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes." Nature **300**(5893): 611-5.

Palmiter, R. D., H. Y. Chen, et al. (1982). "Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring." Cell **29**(2): 701-10.

Pancoska, P., Z. Moravek, et al. (2004). "Efficient RNA interference depends on global context of the target sequence: quantitative analysis of silencing efficiency using Eulerian graph representation of

siRNA." Nucleic Acids Res **32**(4): 1469-79.

Paracchini, V., P. Pedotti, et al. (2005). "Genetics of leptin and obesity: a HuGE review." Am J Epidemiol **162**(2): 101-14.

Parra, G., E. Blanco, et al. (2000). "GeneID in Drosophila." Genome Res **10**(4): 511-5.

Pavlicek, A., O. Clay, et al. (2002). "Isochore conservation between MHC regions on human chromosome 6 and mouse chromosome 17." FEBS Lett **511**(1-3): 175-7.

Pavlicek, A., J. Paces, et al. (2002). "A compact view of isochores in the draft human genome sequence." FEBS Lett **511**(1-3): 165-9.

Peacock, S. L., C. M. McIver, et al. (1981). "Transformation of E. coli using homopolymer-linked plasmid chimeras." Biochim Biophys Acta **655**(2): 243-50.

Peifer, M., S. Berg, et al. (1994). "A repeating amino acid motif shared by proteins with diverse cellular roles." Cell **76**(5): 789-91.

Pennacchio, L. A., M. Olivier, et al. (2001). "An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing." Science **294**(5540): 169-73.

Perkins, T. W., B. Faha, et al. (2002). "Adenovirus-mediated gene therapy using human p21WAF-1/Cip-1 to prevent wound healing in a rabbit model of glaucoma filtration surgery." Arch Ophthalmol **120**(7): 941-9.

Pesole, G., G. Bernardi, et al. (1999). "Isochore specificity of AUG initiator context of human genes." FEBS Lett **464**(1-2): 60-2.

Pesole, G. and S. Liuni (1999). "Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNAs." Trends Genet **15**(9): 378.

Pesole, G., S. Liuni, et al. (1999). "UTRdb: a specialized database of 5' and 3' untranslated regions of eukaryotic mRNAs." Nucleic Acids Res **27**(1): 188-91.

Pesole, G., S. Liuni, et al. (2002). "UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002." Nucleic Acids Res **30**(1): 335-40.

Petriv, O. I. and R. A. Rachubinski (2004). "Lack of peroxisomal catalase causes a progeric phenotype in Caenorhabditis elegans." J Biol Chem **279**(19): 19996-20001.

Petriv, O. I., L. Tang, et al. (2004). "A new definition for the consensus sequence of the peroxisome targeting signal type 2." J Mol Biol **341**(1): 119-34.

Pickles, L. M., S. M. Roe, et al. (2002). "Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor." Structure **10**(6): 751-61.

Piganeau, G., D. Mouchiroud, et al. (2002). "Expected relationship between the silent substitution rate and the GC content: implications for the evolution of isochores." J Mol Evol **54**(1): 129-33.

Plasterk, R. H. (1992). "Reverse genetics of Caenorhabditis elegans." Bioessays **14**(9): 629-33.

Pluthero, F. G. (1993). "Rapid purification of high-activity Taq DNA polymerase." Nucleic Acids Res **21**(20): 4850-1.

Ponger, L., L. Duret, et al. (2001). "Determinants of CpG islands: expression in early embryo and isochore structure." Genome Res **11**(11): 1854-60.

Prak, E. T. and H. H. Kazazian, Jr. (2000). "Mobile elements and the human genome." Nat Rev Genet **1**(2): 134-44.

Print, C. G. and K. L. Loveland (2000). "Germ cell suicide: new insights into apoptosis during spermatogenesis." Bioessays **22**(5): 423-30.

Print, C. G., K. L. Loveland, et al. (1998). "Apoptosis regulator bcl-w is essential for spermatogenesis but appears otherwise redundant." Proc Natl Acad Sci U S A **95**(21): 12424-31.

Proudfoot, N. J. (1976). "Nucleotide sequence from the coding region of rabbit beta-globin messenger RNA." Nucleic Acids Res **3**(7): 1811-21.

Proudfoot, N. J. (1976). "Sequence analysis of the 3' non-coding regions of rabbit alpha- and beta-globin messenger RNAs." J Mol Biol **107**(4): 491-525.

Proudfoot, N. J. and G. G. Brownlee (1976). "3' non-coding region sequences in eukaryotic messenger RNA." Nature **263**(5574): 211-4.

Proudfoot, N. J. and G. G. Brownlee (1976). "Nucleotide sequences of globin messenger RNA." Br Med Bull **32**(3): 251-6.

Proudfoot, N. J., C. C. Cheng, et al. (1976). "Sequence analysis of eukaryotic mRNA." Prog Nucleic Acid Res Mol Biol **19**: 123-34.

Proudfoot, N. J. and J. I. Longley (1976). "The 3' terminal sequences of human alpha and beta globin messenger RNAs: comparison with rabbit globin messenger RNA." Cell **9**(4 PT 2): 733-46.

Rachubinski, R. A. and S. Subramani (1995). "How proteins penetrate peroxisomes." Cell **83**(4): 525-8.

Ramjaun, A. R., K. D. Micheva, et al. (1997). "Identification and characterization of a nerve terminal-enriched amphiphysin isoform." J Biol Chem **272**(26): 16700-6.

Randle, P. J., P. B. Garland, et al. (1965). "The glucose fatty acid cycle in obesity and maturity onset diabetes mellitus." Ann N Y Acad Sci **131**(1): 324-33.

Rapp, J. P. and L. K. Dahl (1972). "Mendelian inheritance of 18- and ll beta-steroid hydroxylase activities in the adrenals of rats genetically susceptible or resistant to hypertension." Endocrinology **90**(6): 1435-46.

Razin, A. and H. Cedar (1994). "DNA methylation and genomic imprinting." Cell **77**(4): 473-6.

Razin, A. and T. Kafri (1994). "DNA methylation from embryo to adult." Prog Nucleic Acid Res Mol Biol **48**: 53-81.

Redding, K., M. Seeger, et al. (1996). "The effects of clathrin inactivation on localization of Kex2 protease are independent of the TGN localization signal in the cytosolic tail of Kex2p." Mol Biol Cell **7**(11): 1667-77.

Reeders, S. T. (1987). "A "reverse genetic" approach to autosomal dominant polycystic kidney disease." Pediatr Nephrol **1**(3): 405-10.

Reeders, S. T., M. H. Breuning, et al. (1987). "A study of genetic linkage heterogeneity in adult polycystic kidney disease." Hum Genet **76**(4): 348-51.

Reese, M. G., O. Lund, et al. (1996). "Distance distributions in proteins: a six-parameter representation." Protein Eng **9**(9): 733-40.

Rhees, B. K. and W. R. Atchley (2000). "Body weight and tail length divergence in mice selected for rate of development." J Exp Zool **288**(2): 151-64.

Rill, R. L., A. Beheshti, et al. (2002). "DNA electrophoresis in agarose gels: effects of field and gel concentration on the exponential dependence of reciprocal mobility on DNA length." Electrophoresis **23**(16): 2710-9.

Robertson, E., A. Bradley, et al. (1986). "Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector." Nature **323**(6087): 445-8.

Rodnick, K. J., R. C. Piper, et al. (1992). "Interaction of insulin and exercise on glucose transport in muscle." Diabetes Care **15**(11): 1679-89.

Rodnick, K. J., J. W. Slot, et al. (1992). "Immunocytochemical and biochemical studies of GLUT4 in rat skeletal muscle." J Biol Chem **267**(9): 6278-85.

Rost, B., J. Liu, et al. (2003). "Automatic prediction of protein function." Cell Mol Life Sci **60**(12): 2637-50.

Rozen, S. and H. Skaletsky (2000). "Primer3 on the WWW for general users and for biologist programmers." Methods Mol Biol **132**: 365-86.

Rubinson, D. A., C. P. Dillon, et al. (2003). "A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference." Nat Genet **33**(3): 401-6.

Rudnicki, M. A., T. Braun, et al. (1992). "Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development." Cell **71**(3): 383-90.

Ruther, U., M. Koenen, et al. (1982). "Exon cloning: immunoenzymatic identification of exons of the chicken lysozyme gene." Proc Natl Acad Sci U S A **79**(22): 6852-5.

Rybakin, V., M. Stumpf, et al. (2004). "Coronin 7, the mammalian POD-1 homologue, localizes to the Golgi apparatus." FEBS Lett **573**(1-3): 161-7.

Ryder, O. A. (2005). "Conservation genomics: applying whole genome studies to species conservation efforts." Cytogenet Genome Res **108**(1-3): 6-15.

Saccone, S. and G. Bernardi (2001). "Human chromosomal banding by in situ hybridization of isochores." Methods Cell Sci **23**(1-3): 7-15.

Saccone, S., A. Pavlicek, et al. (2001). "Genes, isochores and bands in human chromosomes 21 and 22." Chromosome Res **9**(7): 533-9.

Saiki, R. K., D. H. Gelfand, et al. (1988). "Primer-directed enzymatic amplification of DNA with a

thermostable DNA polymerase." Science **239**(4839): 487-91.

Saito, S., B. Liu, et al. (2004). "Animal embryonic stem (ES) cells: self-renewal, pluripotency, transgenesis and nuclear transfer." Hum Cell **17**(3): 107-15.

Sakharkar, M., F. Passetti, et al. (2002). "ExInt: an Exon Intron Database." Nucleic Acids Res **30**(1): 191-4.

Sakharkar, M. K., V. T. Chow, et al. (2004). "A report on single exon genes (SEG) in eukaryotes." Front Biosci **9**: 3262-7.

Sakharkar, M. K., V. T. Chow, et al. (2004). "Distributions of exons and introns in the human genome." In Silico Biol **4**(4): 387-93.
Sakharkar, M. K. and P. Kangueane (2004). "Genome SEGE: a database for 'intronless' genes in eukaryotic genomes." BMC Bioinformatics **5**: 67.

Salinas, J., M. Zerial, et al. (1986). "Gene distribution and nucleotide sequence organization in the mouse genome." Eur J Biochem **160**(3): 469-78.

Sambrook, J.; MacCallum, P. *Molecular Cloning: A Laboratory Manual (Third Edition).* Cancer Institute, Melbourne, Australia; *David Russell,* University of Texas Southwestern Medical Center, Dallas.

Sanger, F., G. M. Air, et al. (1977). "Nucleotide sequence of bacteriophage phi X174 DNA." Nature **265**(5596): 687-95.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.

Sasaki, H., H. Sato, et al. (2002). "Electrophile response element-mediated induction of the cystine/glutamate exchange transporter gene expression." J Biol Chem **277**(47): 44765-71.

Sasaki, Y., S. Ishida, et al. (2002). "The p53 family member genes are involved in the Notch signal pathway." J Biol Chem **277**(1): 719-24.

Scemama, J. L., M. Hunter, et al. (2002). "Evolutionary divergence of vertebrate Hoxb2 expression patterns and transcriptional regulatory loci." J Exp Zool **294**(3): 285-99.

Scherf, M., A. Klingenhoff, et al. (2000). "Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach." J Mol Biol **297**(3): 599-606.

Schultz, J., F. Milpetz, et al. (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proc Natl Acad Sci U S A **95**(11): 5857-64.

Schurmann, A., S. Koling, et al. (2002). "Reduced sperm count and normal fertility in male mice with targeted disruption of the ADP-ribosylation factor-like 4 (Arl4) gene." Mol Cell Biol **22**(8): 2761-8.

Schutze, N. (2004). "siRNA technology." Mol Cell Endocrinol **213**(2): 115-9.

Schwartz, S., Z. Zhang, et al. (2000). "PipMaker--a web server for aligning two genomic DNA sequences." Genome Res **10**(4): 577-86.

Schwartzberg, P. L., S. P. Goff, et al. (1989). "Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells." Science **246**(4931): 799-803.

Seguin, C. and D. H. Hamer (1987). "Regulation in vitro of metallothionein gene binding factors." Science **235**(4794): 1383-7.

Seidman, J. G. and C. Seidman (2002). "Transcription factor haploinsufficiency: when half a loaf is not enough." J Clin Invest **109**(4): 451-5.

Sen, G. L. and H. M. Blau (2006). "A brief history of RNAi: the silence of the genes." Faseb J **20**(9): 1293-9.

Sengupta, J., J. Nilsson, et al. (2004). "Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM." Nat Struct Mol Biol **11**(10): 957-62.

Seol, J. H., A. Shevchenko, et al. (2001). "Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly." Nat Cell Biol **3**(4): 384-91.

Shabalina, S. A., A. Y. Ogurtsov, et al. (2001). "Selective constraint in intergenic regions of human and mouse genomes." Trends Genet **17**(7): 373-6.

Shibuya, T. and K. Morimoto (1993). "A review of the genotoxicity of 1-ethyl-1-nitrosourea." Mutat Res **297**(1): 3-38.

Shibuya, T., T. Murota, et al. (1993). "The induction of recessive mutations in mouse primordial germ cells with N-ethyl-N-nitrosourea." Mutat Res **290**(2): 273-80.

Shinagawa, T. and S. Ishii (2003). "Generation of Ski-knockdown mice by expressing a long double-strand RNA from an RNA polymerase II promoter." Genes Dev **17**(11): 1340-5.

Shulman, M. J., L. Nissen, et al. (1990). "Homologous recombination in hybridoma cells: dependence on time and fragment length." Mol Cell Biol **10**(9): 4466-72.

Silva, J. M., K. Marran, et al. (2008). "Profiling essential genes in human mammary cells by multiplex RNAi screening." Science **319**(5863): 617-20.

Silverman, E. S., L. Le, et al. (2002). "Cloning and functional analysis of the mouse 5-lipoxygenase promoter." Am J Respir Cell Mol Biol **26**(4): 475-83.

Simpson, E. M., C. C. Linder, et al. (1997). "Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice." Nat Genet **16**(1): 19-27.

Sina, M., A. Hinney, et al. (1999). "Phenotypes in three pedigrees with autosomal dominant obesity caused by haploinsufficiency mutations in the melanocortin-4 receptor gene." Am J Hum Genet **65**(6): 1501-7.

Singer, M. F. (1982). "Highly repeated sequences in mammalian genomes." Int Rev Cytol **76**: 67-112.

Sinha, S. (2004). "Regulation of intermediate filament gene expression." Methods Cell Biol **78**: 267-96.

Sipos, G., J. H. Brickner, et al. (2004). "Soi3p/Rav1p functions at the early endosome to regulate endocytic trafficking to the vacuole and localization of trans-Golgi network transmembrane proteins." Mol Biol Cell **15**(7): 3196-209.

Siri, P. W. and H. N. Ginsberg (2003). "Ovariectomy leads to increased insulin resistance in human apolipoprotein B transgenic mice lacking brown adipose tissue." Metabolism **52**(6): 659-661.

Slot, J. W., H. J. Geuze, et al. (1991). "Translocation of the glucose transporter GLUT4 in cardiac myocytes of the rat." Proc Natl Acad Sci U S A **88**(17): 7815-9.

Smardon, A. M., M. Tarsio, et al. (2002). "The RAVE complex is essential for stable assembly of the yeast V-ATPase." J Biol Chem **277**(16): 13831-9.

Smit, A. F. (1993). "Identification of a new, abundant superfamily of mammalian LTR-transposons." Nucleic Acids Res **21**(8): 1863-72.

Smit, A. F. (1996). "The origin of interspersed repeats in the human genome." Curr Opin Genet Dev **6**(6): 743-8.

Smit, A. F. (1999). "Interspersed repeats and other mementos of transposable elements in mammalian genomes." Curr Opin Genet Dev **9**(6): 657-63.

Smit, A. F. and A. D. Riggs (1995). "MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation." Nucleic Acids Res **23**(1): 98-102.

Smith, G. R. (1989). "Homologous recombination in E. coli: multiple pathways for multiple reasons." Cell **58**(5): 807-9.

Smith, K. (2001). "Theoretical mechanisms in targeted and random integration of transgene DNA." Reprod Nutr Dev **41**(6): 465-85.

Smith, N. G., M. Brandstrom, et al. (2004). "Evidence for turnover of functional noncoding DNA in mammalian genome evolution." Genomics **84**(5): 806-13.

Smith, R. M., M. J. Charron, et al. (1991). "Immunoelectron microscopic demonstration of insulin-stimulated translocation of glucose transporters to the plasma membrane of isolated rat adipocytes and masking of the carboxyl-terminal epitope of intracellular GLUT4." Proc Natl Acad Sci U S A **88**(15): 6893-7.

Smith, T. F., C. Gaitatzes, et al. (1999). "The WD repeat: a common architecture for diverse functions." Trends Biochem Sci **24**(5): 181-5.

Smithies, O., R. G. Gregg, et al. (1985). "Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination." Nature **317**(6034): 230-4.

Snell, G. D. (1931). "Inheritance in the House Mouse, the Linkage Relations of Short-Ear, Hairless, and Naked." Genetics **16**(1): 42-74.

Solovyev, V. and A. Salamov (1997). "The Gene-Finder computer tools for analysis of human and model organisms genome sequences." Proc Int Conf Intell Syst Mol Biol **5**: 294-302.

Solovyev, V. V., A. A. Salamov, et al. (1995). "Identification of human gene structure using linear discriminant functions and dynamic programming." Proc Int Conf Intell Syst Mol Biol **3**: 367-75.

Sondek, J., A. Bohm, et al. (1996). "Crystal structure of a G-protein beta gamma dimer at 2.1A resolution." Nature **379**(6563): 369-74.

Soriano, P., G. Friedrich, et al. (1991). "Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells." J Virol **65**(5): 2314-9.

Soriano, P., C. Montgomery, et al. (1991). "Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice." Cell **64**(4): 693-702.

Southan, C. (2004). "Has the yo-yo stopped? An assessment of human protein-coding gene number." Proteomics **4**(6): 1712-26.

Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." J Mol Biol **98**(3): 503-17.

Staats, J. (1980). "Standardized nomenclature for inbred strains of mice: seventh listing for the International Committee on Standardized Genetic Nomenclature for Mice." Cancer Res **40**(7): 2083-128.

Steinmetz, L. M. and A. M. Deutschbauer (2002). "Gene function on a genomic scale." J Chromatogr B Analyt Technol Biomed Life Sci **782**(1-2): 151-63.

Steinmetz, L. M., C. Scharfe, et al. (2002). "Systematic screen for human disease genes in yeast." Nat Genet **31**(4): 400-4.

Stohr, H., N. Mohr, et al. (2002). "Cloning and characterization of WDR17, a novel WD repeat-containing gene on chromosome 4q34." Biochim Biophys Acta **1579**(1): 18-25.

Storch, S., S. Pohl, et al. (2004). "A dileucine motif and a cluster of acidic amino acids in the second cytoplasmic domain of the batten disease-related CLN3 protein are required for efficient lysosomal targeting." J Biol Chem **279**(51): 53625-34.

Strichman-Almashanu, L. Z., R. S. Lee, et al. (2002). "A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes." Genome Res **12**(4): 543-54.

Stubdal, H., C. A. Lynch, et al. (2000). "Targeted Deletion of the tub Mouse Obesity Gene Reveals that tubby Is a Loss-of-Function Mutation." Mol. Cell. Biol. **20**(3): 878-882.

Stultz, C. M., J. V. White, et al. (1993). "Structural analysis based on state-space modeling." Protein Sci **2**(3): 305-14.

Su, A. I., M. P. Cooke, et al. (2002). "Large-scale analysis of the human and mouse transcriptomes." Proc Natl Acad Sci U S A **99**(7): 4465-70.

Su, A. I., T. Wiltshire, et al. (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proc Natl Acad Sci U S A **101**(16): 6062-7.

Sung, Y. H., J. Song, et al. (2004). "Functional genomics approach using mice." J Biochem Mol Biol **37**(1): 122-32.

Sun-Wada, G. H., Y. Wada, et al. (2004). "Diverse and essential roles of mammalian vacuolar-type proton pump ATPase: toward the physiological understanding of inside acidic compartments." Biochim Biophys Acta **1658**(1-2): 106-14.

Suzuki, Y., T. Tsunoda, et al. (2001). "Identification and characterization of the potential promoter regions of 1031 kinds of human genes." Genome Res **11**(5): 677-84.

Suzuki, Y., R. Yamashita, et al. (2002). "DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs." Nucleic Acids Res **30**(1): 328-31.

Suzuki, Y., R. Yamashita, et al. (2004). "Large-scale collection and characterization of promoters of

human and mouse genes." In Silico Biol **4**(4): 429-44.

Suzuki, Y., R. Yamashita, et al. (2004). "Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions." Genome Res **14**(9): 1711-8.

Tabaska, J. E. and M. Q. Zhang (1999). "Detection of polyadenylation signals in human DNA sequences." Gene **231**(1-2): 77-86.

Takagaki, Y. and J. L. Manley (1997). "RNA recognition by the human polyadenylation factor CstF." Mol Cell Biol **17**(7): 3907-14.

Tamashiro, K. L. K., T. Wakayama, et al. (2003). "Phenotype of Cloned Mice: Development, Behavior, and Physiology." Experimental Biology and Medicine **228**(10): 1193-1200.
Tamura, K. and S. Kumar (2002). "Evolutionary distance estimation under heterogeneous substitution pattern among lineages." Mol Biol Evol **19**(10): 1727-36.

Tamura, M., T. Kasuga, et al. (2002). "Phylogenetic characterization of Histoplasma capsulatum strains based on ITS region sequences, including two new strains from Thai and Chinese patients in Japan." Nippon Ishinkin Gakkai Zasshi **43**(1): 11-9.

Tapscott, S. J., R. L. Davis, et al. (1988). "MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts." Science **242**(4877): 405-11.

Tautz, D. (2000). "Evolution of transcriptional regulation." Curr Opin Genet Dev **10**(5): 575-9.

Taylor, M. W., M. K. Tokito, et al. (1978). "Regulation of de novo purine biosynthesis in normal and 8-azaguanine-resistant Chinese hamster cells." Biochim Biophys Acta **517**(1): 1-13.

te Riele, H., E. R. Maandag, et al. (1992). "Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs." Proc Natl Acad Sci U S A **89**(11): 5128-32.

te Riele, H., E. R. Maandag, et al. (1990). "Consecutive inactivation of both alleles of the pim-1 proto-oncogene by homologous recombination in embryonic stem cells." Nature **348**(6302): 649-51.

Thomas, A. and M. H. Skolnick (1994). "A probabilistic model for detecting coding regions in DNA sequences." IMA J Math Appl Med Biol **11**(3): 149-60.

Thomas, J. W., J. W. Touchman, et al. (2003). "Comparative analyses of multi-species sequences from targeted genomic regions." Nature **424**(6950): 788-93.

Thomas, K. R. and M. R. Capecchi (1986). "Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene." Nature **324**(6092): 34-8.

Thomas, K. R. and M. R. Capecchi (1986). "Targeting of genes to specific sites in the mammalian genome." Cold Spring Harb Symp Quant Biol **51 Pt 2**: 1101-13.

Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell **51**(3): 503-12.

Thomas, K. R. and M. R. Capecchi (1990). "Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development." Nature **346**(6287): 847-50.

Thomas, K. R., K. R. Folger, et al. (1986). "High frequency targeting of genes to specific sites in the mammalian genome." Cell **44**(3): 419-28.

Thompson, K. L. and M. R. Rosner (1989). "Regulation of epidermal growth factor receptor gene expression by retinoic acid and epidermal growth factor." J Biol Chem **264**(6): 3230-4.

Thompson, S., A. R. Clarke, et al. (1989). "Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells." Cell **56**(2): 313-21.

Threadgill, D. W., A. Matin, et al. (1997). "SSLPs to map genetic differences between the 129 inbred strains and closed-colony, random-bred CD-1 mice." Mamm Genome **8**(6): 441-2.

Timmons, L. and A. Fire (1998). "Specific interference by ingested dsRNA." Nature **395**(6705): 854.
Tiscornia, G., O. Singer, et al. (2003). "A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA." Proc Natl Acad Sci U S A **100**(4): 1844-8.

Tykocinski, M. L. and E. E. Max (1984). "CG dinucleotide clusters in MHC genes and in 5' demethylated genes." Nucleic Acids Res **12**(10): 4385-96.

Ureta-Vidal, A., L. Ettwiller, et al. (2003). "Comparative genomics: genome-wide analysis in metazoan eukaryotes." Nat Rev Genet **4**(4): 251-62.

Vaisse, C., K. Clement, et al. (2000). "Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity." J Clin Invest **106**(2): 253-62.

Valenzuela, D. M., A. J. Murphy, et al. (2003). "High-throughput engineering of the mouse genome coupled with high-resolution expression analysis." Nat Biotechnol **21**(6): 652-9.

Valpuesta, J. M., J. Martin-Benito, et al. (2002). "Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT." FEBS Lett **529**(1): 11-6.

Valverde-Garduno, V., B. Guyot, et al. (2004). "Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification." Blood **104**(10): 3106-16.

van der Voorn, L. and H. L. Ploegh (1992). "The WD-40 repeat." FEBS Lett **307**(2): 131-4.

van Deursen, J., J. Schepens, et al. (1992). "Genetic variability of the murine creatine kinase B gene locus and related pseudogenes in different inbred strains of mice." Genomics **12**(2): 340-9.

van Deursen, J. and B. Wieringa (1992). "Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors." Nucleic Acids Res **20**(15): 3815-20.

van Ommen, G. J. and P. L. Pearson (1989). "Long-range mapping in the research and diagnosis of genetic disease." Genome **31**(2): 730-6.

Van Ommen, G. J., A. Sterk, et al. (1989). "Studies on the structures of the normal and abnormal goat thyroglobulin genes." Biochimie **71**(2): 211-21.

van Weert, A. W., K. W. Dunn, et al. (1995). "Transport from late endosomes to lysosomes, but not sorting of integral membrane proteins in endosomes, depends on the vacuolar proton pump." J Cell Biol **130**(4): 821-34.

Vasquez, K. M., K. Marburger, et al. (2001). "Manipulating the mammalian genome by homologous

recombination." Proc Natl Acad Sci U S A **98**(15): 8403-10.

Veitia, R. A. (2002). "Exploring the etiology of haploinsufficiency." Bioessays **24**(2): 175-84.

Veitia, R. A., L. Salas-Cortes, et al. (2001). "Testis determination in mammals: more questions than answers." Mol Cell Endocrinol **179**(1-2): 3-16.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Vinogradov, A. E. (2003). "Isochores and tissue-specificity." Nucleic Acids Res **31**(17): 5212-20.

Voegtli, W. C., A. Y. Madrona, et al. (2003). "The structure of Aip1p, a WD repeat protein that regulates Cofilin-mediated actin depolymerization." J Biol Chem **278**(36): 34373-9.

von Heijne, G. (1992). "Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule." J Mol Biol **225**(2): 487-94.

Wade, C. M., E. J. Kulbokas, 3rd, et al. (2002). "The mosaic structure of variation in the laboratory mouse genome." Nature **420**(6915): 574-8.

Wagner, E. F., T. A. Stewart, et al. (1981). "The human beta-globin gene and a functional viral thymidine kinase gene in developing mice." Proc Natl Acad Sci U S A **78**(8): 5016-20.

Wagner, T. E., P. C. Hoppe, et al. (1981). "Microinjection of a rabbit beta-globin gene into zygotes and its subsequent expression in adult mice and their offspring." Proc Natl Acad Sci U S A **78**(10): 6376-80.

Wall, M. A., D. E. Coleman, et al. (1995). "The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2." Cell **83**(6): 1047-58.

Wang, S., Z. Shi, et al. (2006). "Development and validation of vectors containing multiple siRNA expression cassettes for maximizing the efficiency of gene silencing." BMC Biotechnol **6**: 50.

Warbrick, E., W. Heatherington, et al. (1998). "PCNA binding proteins in Drosophila melanogaster : the analysis of a conserved PCNA binding domain." Nucleic Acids Res **26**(17): 3925-32.

Warren, R. A., F. A. Green, et al. (1998). "Distinct saturable pathways for the endocytosis of different tyrosine motifs." J Biol Chem **273**(27): 17056-63.

Wasylyk, B., S. L. Hahn, et al. (1993). "The Ets family of transcription factors." Eur J Biochem **211**(1-2): 7-18.

Wasylyk, C. and B. Wasylyk (1993). "Oncogenic conversion of Ets affects redox regulation in-vivo and in-vitro." Nucleic Acids Res **21**(3): 523-9.

Waterston, R. H., E. S. Lander, et al. (2002). "On the sequencing of the human genome." Proc Natl Acad Sci U S A **99**(6): 3712-6.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Weber, M., K. Moller, et al. (1995). "Short technical reports. Effects of lipopolysaccharide on transfection

efficiency in eukaryotic cells." Biotechniques **19**(6): 930-40.

Wei, J., M. B. Goldberg, et al. (2003). "Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T." Infect Immun **71**(5): 2775-86.

Weil, B. (2000). Funktionelle Analyse des CpY/DmX Gens aus Chironomus und Drosophila melanogaster. Doktorarbeit, Johannes Gutenberg-Universität Mainz.

Werner, T. (1999). "Models for prediction and recognition of eukaryotic promoters." Mamm Genome **10**(2): 168-75.

Werner, T. (2002). "Finding and decrypting of promoters contributes to the elucidation of gene function." In Silico Biol **2**(3): 249-55.
Werner, T. (2002). "Promoter analysis." Ernst Schering Res Found Workshop(38): 65-82.

Werner, T. (2003). "The state of the art of mammalian promoter recognition." Brief Bioinform **4**(1): 22-30.

Werner, T., S. Fessele, et al. (2003). "Computer modeling of promoter organization as a tool to study transcriptional coregulation." Faseb J **17**(10): 1228-37.

White, J. V., C. M. Stultz, et al. (1994). "Protein classification by stochastic modeling and optimal filtering of amino-acid sequences." Math Biosci **119**(1): 35-75.

Wickens, M. (1990). "How the messenger got its tail: addition of poly(A) in the nucleus." Trends Biochem Sci **15**(7): 277-81.

Wickens, M. (1990). "In the beginning is the end: regulation of poly(A) addition and removal during early development." Trends Biochem Sci **15**(8): 320-4.

Williams, F. E. and R. J. Trumbly (1990). "Characterization of TUP1, a mediator of glucose repression in Saccharomyces cerevisiae." Mol Cell Biol **10**(12): 6500-11.

Wilson, D., M. Madera, et al. (2007). "The SUPERFAMILY database in 2007: families and functions." Nucleic Acids Res **35**(Database issue): D308-13.

Wilson, M. D., C. Riemer, et al. (2001). "Comparative analysis of the gene-dense ACHE/TFR2 region on human chromosome 7q22 with the orthologous region on mouse chromosome 5." Nucleic Acids Res **29**(6): 1352-65.

Winzeler, E. A., B. Lee, et al. (1999). "Whole genome genetic-typing in yeast using high-density oligonucleotide arrays." Parasitology **118 Suppl**: S73-80.

Winzeler, E. A., D. D. Shoemaker, et al. (1999). "Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis." Science **285**(5429): 901-6.

Wiznerowicz, M. and D. Trono (2003). "Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference." J Virol **77**(16): 8957-61.

Wolfrum, C., D. Q. Shih, et al. (2003). "Role of Foxa-2 in adipocyte metabolism and differentiation." J Clin Invest **112**(3): 345-56.

Wong, G. K., D. A. Passey, et al. (2001). "Most of the human genome is transcribed." Genome Res

11(12): 1975-7.

Wootton, J. C. and S. Federhen (1996). "Analysis of compositionally biased regions in sequence databases." Methods Enzymol **266**: 554-71.

Wu, C. I. and W. H. Li (1985). "Evidence for higher rates of nucleotide substitution in rodents than in man." Proc Natl Acad Sci U S A **82**(6): 1741-5.

Wystub, S., B. Ebner, et al. (2004). "Interspecies comparison of neuroglobin, cytoglobin and myoglobin: sequence evolution and candidate regulatory elements." Cytogenet Genome Res **105**(1): 65-78.

Xu, Z. P., E. F. Wawrousek, et al. (2002). "Transketolase haploinsufficiency reduces adipose tissue and female fertility in mice." Mol Cell Biol **22**(17): 6142-7.
Xuan, Z., J. Wang, et al. (2003). "Computational comparison of two mouse draft genomes and the human golden path." Genome Biol **4**(1): R1.

Xue, F. and L. Cooley (1993). "kelch encodes a component of intercellular bridges in Drosophila egg chambers." Cell **72**(5): 681-93.

Yamasaki, C., K. Murakami, et al. (2008). "The H-Invitational Database (H-InvDB), a comprehensive annotation resource for human genes and transcripts." Nucleic Acids Res **36**(Database issue): D793-9.

Yan, W., L. Ma, et al. (2004). "Haploinsufficiency of kelch-like protein homolog 10 causes infertility in male mice." Proc Natl Acad Sci U S A **101**(20): 7793-8.

Yanisch-Perron, C., J. Vieira, et al. (1985). "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors." Gene **33**(1): 103-19.

Yaswen, L., N. Diehl, et al. (1999). "Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin." Nat Med **5**(9): 1066-70.

Yeh, S., M. Y. Tsai, et al. (2002). "Generation and characterization of androgen receptor knockout (ARKO) mice: an in vivo model for the study of androgen functions in selective tissues." Proc Natl Acad Sci U S A **99**(21): 13498-503.

Yen, T., A. Gill, et al. (1994). "Obesity, diabetes, and neoplasia in yellow A(vy)/- mice: ectopic expression of the agouti gene." FASEB J. **8**(8): 479-488.

Yeung, C. H., S. Breton, et al. (2004). "Increased luminal pH in the epididymis of infertile c-ros knockout mice and the expression of sodium-hydrogen exchangers and vacuolar proton pump H+-ATPase." Mol Reprod Dev **68**(2): 159-68.

Yu, L. C., Y. C. Twu, et al. (2003). "The molecular genetics of the human I locus and molecular background explain the partial association of the adult i phenotype with congenital cataracts." Blood **101**(6): 2081-8.

Yu, Y. and A. Bradley (2001). "Engineering chromosomal rearrangements in mice." Nat Rev Genet **2**(10): 780-90.

Zagulski, M., B. Babinska, et al. (1995). "The sequence of 24.3 kb from chromosome X reveals five complete open reading frames, all of which correspond to new genes, and a tandem insertion of a Ty1 transposon." Yeast **11**(12): 1179-86.

Zarudnaya, M. I., I. M. Kolomiets, et al. (2003). "Downstream elements of mammalian pre-mRNA

polyadenylation signals: primary, secondary and higher-order structures." Nucleic Acids Res **31**(5): 1375-86.

Zhang, D., T. L. Penttila, et al. (2001). "Spermiogenesis deficiency in mice lacking the Trf2 gene." Science **292**(5519): 1153-5.

Zhang, F. P., M. Poutanen, et al. (2001). "Normal prenatal but arrested postnatal sexual development of luteinizing hormone receptor knockout (LuRKO) mice." Mol Endocrinol **15**(1): 172-83.

Zhang, J. W. and P. B. Lazarow (1995). "PEB1 (PAS7) in Saccharomyces cerevisiae encodes a hydrophilic, intra-peroxisomal protein that is a member of the WD repeat family and is essential for the import of thiolase into peroxisomes." J Cell Biol **129**(1): 65-80.

Zhang, J. W. and P. B. Lazarow (1996). "Peb1p (Pas7p) is an intraperoxisomal receptor for the NH2-terminal, type 2, peroxisomal targeting sequence of thiolase: Peb1p itself is targeted to peroxisomes by an NH2-terminal peptide." J Cell Biol **132**(3): 325-34.

Zhang, Y., R. Proenca, et al. (1994). "Positional cloning of the mouse obese gene and its human homologue." Nature **372**(6505): 425-32.

Zhang, Z., P. Berman, et al. (1998). "Alignments without low-scoring regions." J Comput Biol **5**(2): 197-210.

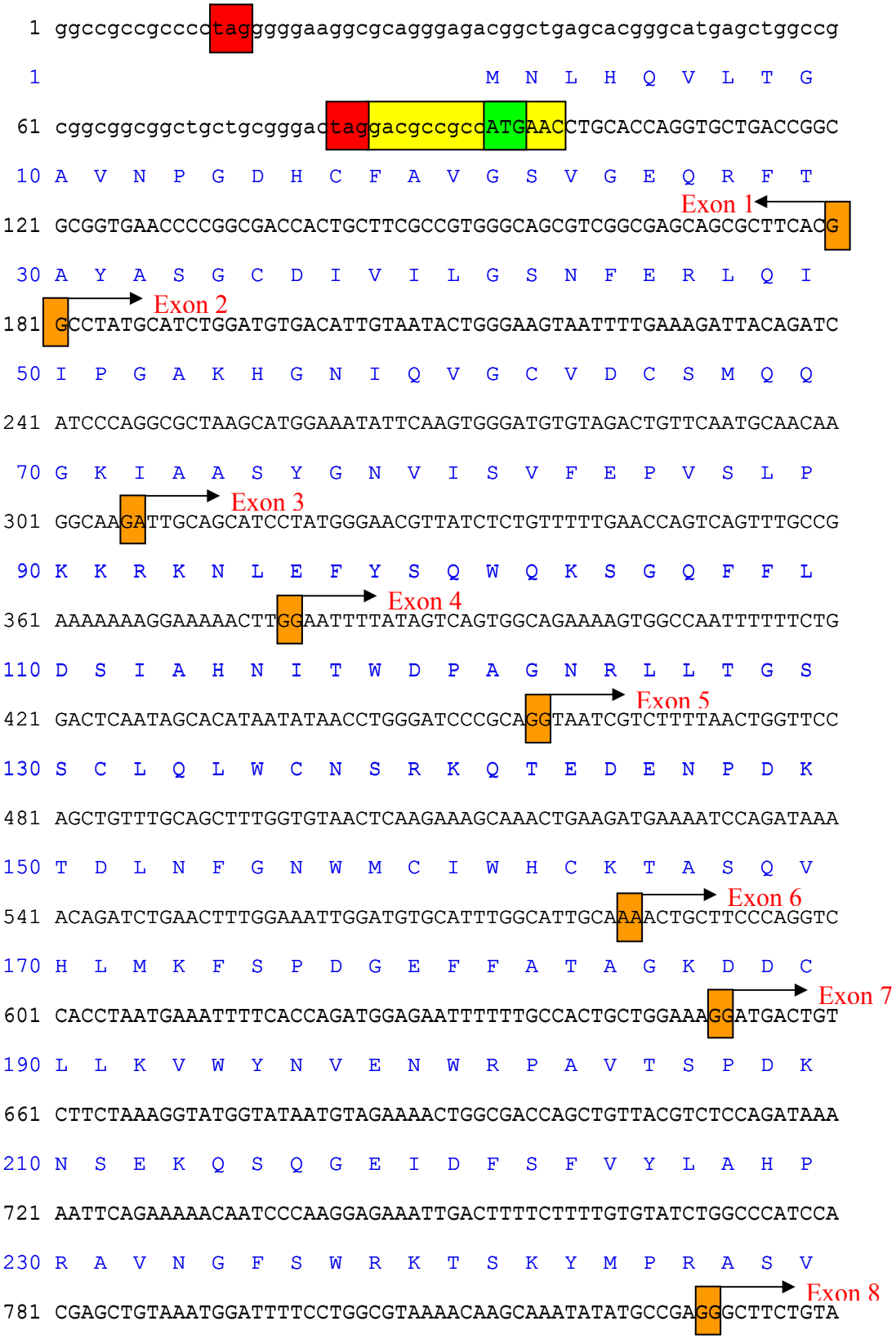
Zhang, Z., P. M. Harrison, et al. (2003). "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome." Genome Res **13**(12): 2541-58.

Zhao, Z., S. Ravid, et al. (1995). "Chromosomal mapping of the mouse A3 adenosine receptor gene, Adora3." Genomics **30**(1): 118-9.

Zoubak, S., O. Clay, et al. (1996). "The gene distribution of the human genome." Gene **174**(1): 95-102.

Zuckerkandl, E. (1965). "The Evolution of Hemoglobin." Sci Am **212**: 110-8.

Zuckerkandl, E. and L. Pauling (1965). "Molecules as documents of evolutionary history." J Theor Biol **8**(2): 357-66.


1 ggccgccgcccctaggggaaggcgcagggagacggctgagcacgggcatgagctggccg
1 M N L H Q V L T G
61 cggcggcggctgctgcgggactaggacgccgccATGAACCTGCACCAGGTGCTGACCGGC
10 A V N P G D H C F A V G S V G E Q R F T
Exon 1
121 GCGGTGAACCCCGGCGACCACTGCTTCGCCGTGGGCAGCGTCGGCGAGCAGCGCTTCACG
30 A Y A S G C D I V I L G S N F E R L Q I
Exon 2
181 GCCTATGCATCTGGATGTGACATTGTAATACTGGGAAGTAATTTTGAAAGATTACAGATC
50 I P G A K H G N I Q V G C V D C S M Q Q
241 ATCCCAGGCGCTAAGCATGGAAATATTCAAGTGGGATGTGTAGACTGTTCAATGCAACAA
70 G K I A A S Y G N V I S V F E P V S L P
Exon 3
301 GGCAAGATTGCAGCATCCTATGGGAACGTTATCTCTGTTTTTGAACCAGTCAGTTTGCCG
90 K K R K N L E F Y S Q W Q K S G Q F F L
Exon 4
361 AAAAAAAGGAAAAACTTGGAATTTTATAGTCAGTGGCAGAAAAGTGGCCAATTTTTTCTG
110 D S I A H N I T W D P A G N R L L T G S
Exon 5
421 GACTCAATAGCACATAATATAACCTGGGATCCCGCAGGTAATCGTCTTTTAACTGGTTCC
130 S C L Q L W C N S R K Q T E D E N P D K
481 AGCTGTTTGCAGCTTTGGTGTAACTCAAGAAAGCAAACTGAAGATGAAAATCCAGATAAA
150 T D L N F G N W M C I W H C K T A S Q V
Exon 6
541 ACAGATCTGAACTTTGGAAATTGGATGTGCATTTGGCATTGCAAAACTGCTTCCCAGGTC
170 H L M K F S P D G E F F A T A G K D D C
Exon 7
601 CACCTAATGAAATTTTCACCAGATGGAGAATTTTTTGCCACTGCTGGAAAGGATGACTGT
190 L L K V W Y N V E N W R P A V T S P D K
661 CTTCTAAAGGTATGGTATAATGTAGAAAACTGGCGACCAGCTGTTACGTCTCCAGATAAA
210 N S E K Q S Q G E I D F S F V Y L A H P
721 AATTCAGAAAAACAATCCCAAGGAGAAATTGACTTTTCTTTTGTGTATCTGGCCCATCCA
230 R A V N G F S W R K T S K Y M P R A S V
Exon 8
781 CGAGCTGTAAATGGATTTTCCTGGCGTAAAACAAGCAAATATATGCCGAGGGCTTCTGTA
250 C N V L L T C C K D N V C R L W V E T F

Anhang

841 TGTAACGTACTGTTGACTTGCTGTAAAGATAATGTATGTCGTCTTTGGGTAGAAACATTT
270 L P N D C F L Y G S D C N H W C E P V S
901 TTACCAAATGATTGCTTCCTGTATGGAAGTGACTGCAACCACTGGTGTGAACCAGTTAGT
290 L T N N L K R N A S S K D R V Q S A L E
961 TTAACAAATAACTTAAAGAGAAATGCTTCCAGTAAAGACCGAGTTCAAAGTGCTCTAGAA
310 V N L R P F R R G R S R S L A L V A H T
Exon 9
1021 GTAAATCTGAGACCTTTCCGTAGAGGTCGGAGTAGATCACTTGCTCTTGTAGCACATACT
330 G Y L P H Q Q D P H H A H R N T P L H A
1081 GGCTATCTGCCTCATCAGCAGGATCCTCACCATGCCCACAGGAACACACCACTGCATGCC
350 N A L C H F H I A A S I N P A T D I P L
Exon 10
1141 AATGCACTTTGTCACTTTCATATTGCAGCCAGCATCAACCCAGCCACAGATATCCCACTT
370 L P S I T S L S L N E N E E K C G P F V
1201 CTTCCATCTATTACATCCTTGAGTTTAAATGAAAATGAAGAGAAATGTGGACCTTTTGTG
390 V H W L N N K E L H F T L S M E V F L Q
1261 GTACACTGGTTAAACAATAAAGAGTTACATTTTACTTTGTCCATGGAAGTCTTCTTACAG
410 Q L R K S F E Q P S S E A S V E D S I Q
1321 CAACTTAGAAAAAGTTTTGAACAGCCCTCATCTGAAGCCAGTGTAGAAGATTCTATTCAG
430 A D L K S D E E L D D G V D D L K I N H
Exon 11
1381 GCAGATTTAAAATCTGATGAAGAACTGGATGATGGTGTTGATGATTTAAAAATAAATCAT
450 E K K E L D E D K M L P S S S F T P L S
1441 GAAAAGAAGGAATTGGATGAAGACAAAATGTTACCAAGTTCAAGTTTTACACCATTATCT
470 S A A V D H Q I E V L L S E W S K N A D
1501 TCAGCCGCCGTCGATCATCAGATTGAAGTGCTTCTGTCTGAATGGAGTAAAAACGCAGAC
490 M L F S I H P M D G S L L V W H V D W L
1561 ATGCTTTTCAGTATCCATCCCATGGATGGGTCTTTACTAGTTTGGCATGTAGATTGGCTG
510 D E Y Q P G M F R Q V Q V S F V S R I P
Exon 12
1621 GATGAATACCAGCCTGGTATGTTTCGTCAAGTTCAGGTGTCCTTTGTTTCCAGAATTCCT
530 V A F P T G D A N S L C K S I V M Y A C

1681 GTAGCTTTCCCCACAGGTGATGCAAACTCTCTTTGTAAAAGCATAGTGATGTATGCTTGT

550 T K N V D L A I Q Q G K Q R P T G L T R

1741 ACCAAGAACGTTGACTTGGCTATTCAACAAGGAAAACAGAGACCCACTGGCCTCACTCGT

570 S T S M L I S S A H S K S S N N L K L S

1801 TCCACATCAATGCTTATCTCCTCTGCTCACAGTAAATCATCAAATAACTTAAAATTAAGT

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

1861 ATTTTTACCCCTAATGTTATGATGATTTCAAAGCATGCTGATGGTTCTTTGAATCAGTGG

610 L V S F A E E S A F S T V L S I S H K S

1921 CTGGTTAGTTTTGCTGAGGAATCTGCTTTTTCTACAGTGCTTAGTATTTCTCACAAATCC

630 R Y C G H R F H L N D L A C H S V L P L

1981 AGATATTGTGGTCATCGTTTTCATCTTAATGACTTAGCTTGCCACTCGGTATTACCTTTA

650 L L T T S H H N A L R T P N V G N Q K Q

2041 TTACTGACAACGTCACACCATAATGCATTAAGGACACCAAATGTTGGTAACCAAAAGCAA

670 A H D A V N T E E C S L A Q Q N K S N V

2101 GCTCATGATGCTGTAAATACTGAAGAATGCTCCTTGGCACAACAAAATAAAAGCAACGTT

690 D M A F Q D P N A I Y S E L I L W R V D

2161 GATATGGCATTTCAAGATCCCAATGCAATTTACAGTGAGCTTATTCTTTGGAGGGTCGAT

710 P V G P L S F S G G V S E L A R I N S L

2221 CCAGTTGGGCCACTATCTTTTTCTGGAGGAGTATCTGAACTTGCCCGGATTAATTCTCTT

730 H V S A F S N V A W L P T L I P S Y C L

2281 CATGTTTCTGCCTTTTCCAATGTGGCATGGCTCCCTACTCTCATACCCAGTTACTGTCTA

750 G A Y C N S P S A C F V A S D G Q Y L R

Exon 13

2341 GCTGCATACTGCAACTCTCCTAGTGCATGCTTTGTAGCTAGTGATGGACAATATCTGAGG

770 L Y E A V I D A K K L L Y E L S N P E I

2401 TTATATGAAGCAGTTATTGATGCCAAGAAACTTTTATATGAACTTTCCAACCCTGAAATT

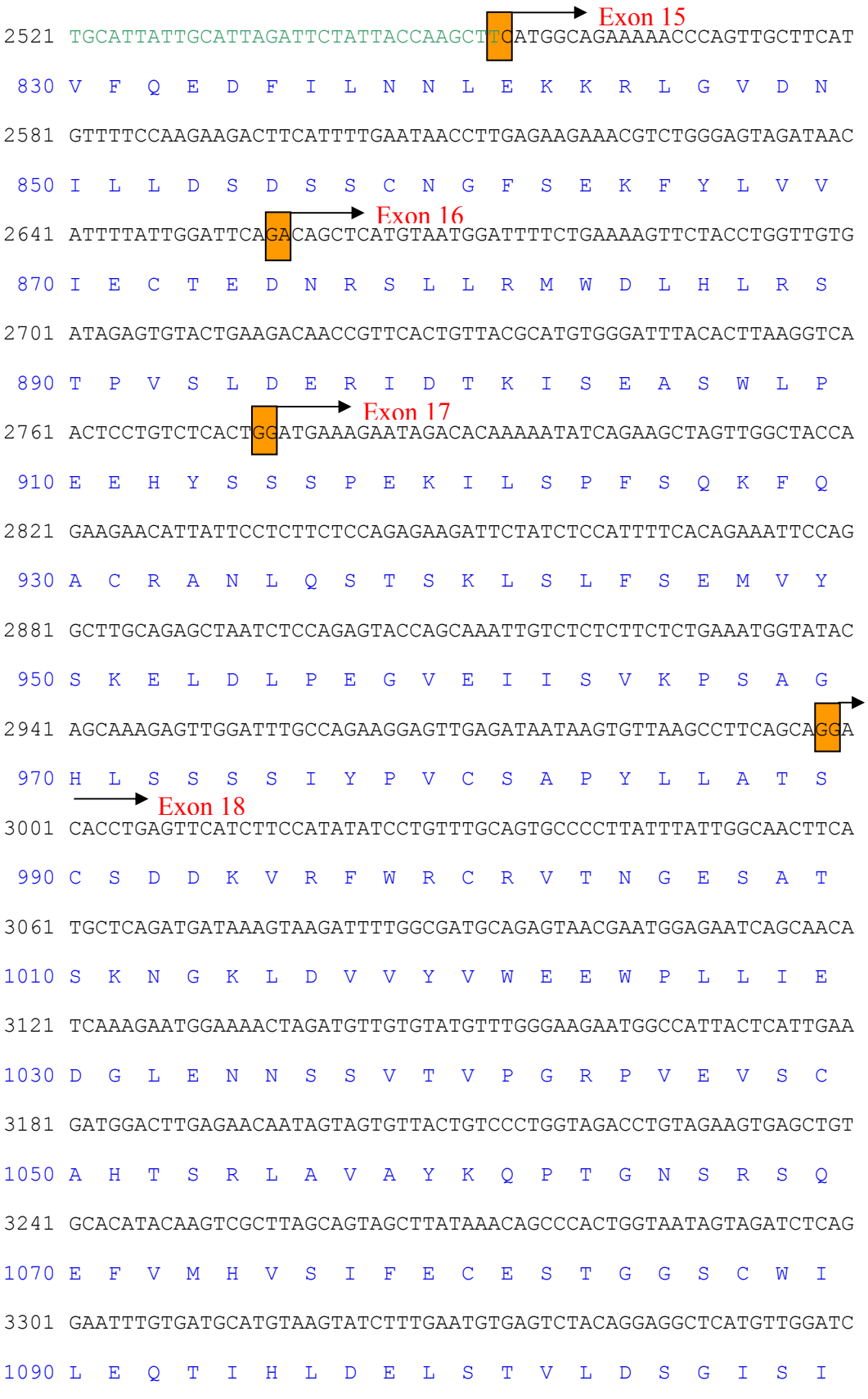
790 S K Y V G E V F N I V S Q Q S T A R P G

Exon 14

2461 TCTAAATATGTTGGTGAAGTGTTTAACATTGTCAGTCAGCAATCAACAGCCAGACCAGGA

810 C I I A L D S I T K L H G R K T Q L L H

Anhang



3361 CTTGAGCAGACAATTCATTTAGATGAGTTAAGTACAGTGTTGGATTCTGGCATTAGTATT

1110 D S N L V A Y N K Q E T Y L V S K E S I

3421 GATAGCAATTTAGTGGCTTATAATAAACAGGAGACATATTTAGTCAGTAAAGAAAGTATT

1130 T S N T K H L V H L D W M S R E D G S H

3481 ACATCAAACACAAAACATTTGGTTCACTTAGATTGGATGTCTAGAGAAGATGGTTCTCAT

1150 I L T V G I G S K L F M Y G P M A G K V

3541 ATCTTAACTGTAGGGATTGGCTCAAAACTTTTTATGTATGGACCTATGGCTGGCAAGGTA

1170 Q D Q T G K E N Q A F P L W D S T K I V

3601 CAAGATCAGACTGGAAAGGAAAACCAGGCATTTCCTCTCTGGGACAGTACTAAAATTGTA

1190 P L S K F V L L R S V D L V S S V E G A

3661 CCTCTTTCTAAATTTGTACTGTTACGAAGTGTGGACTTGGTTTCTTCTGTAGAAGGTGCC

1210 P P F P V S L S W V R D G I L V V G M D

3721 CCACCTTTTCCTGTTTCTTTATCATGGGTCCGGGATGGCATCCTTGTGGTAGGAATGGAT

1230 C E M H V Y S Q W Q P S N K Q E P V I S

3781 TGTGAAATGCATGTATATTCCCAATGGCAGCCATCTAATAAACAAGAACCTGTTATATCA

1250 E S Y N G S T P S I L S L I K Q S N S S

3841 GAGTCATACAATGGAAGCACTCCATCTATACTAAGTTTAATAAAACAAAGTAACTCATCA

1270 S S G L H P P K K T L T R S M T S L A Q

3901 AGTTCTGGGCTACATCCTCCAAAGAAGACTCTGACTCGATCCATGACCAGTCTTGCACAG

1290 K I C G K K S I F D P S V D M E D S G L

3961 AAAATCTGTGGAAAGAAAAGTATATTTGACCCCTCAGTGGATATGGAAGATTCCGGTCTC

1310 F E A A H V L S P T L P Q Y H P L Q L L

4021 TTTGAAGCCGCTCATGTACTTTCCCCAACTTTACCTCAGTACCATCCCCTGCAGCTTTTG

1330 E L M D L G K V R R A K A I L S H L V K

4081 GAACTCATGGATCTTGGCAAAGTTCGGAGAGCCAAGGCCATCTTGTCACATCTTGTCAAA

1350 C I A G E V V A L N E A E S N H E R R L

4141 TGCATTGCTGGGGAAGTTGTGGCTCTGAATGAAGCTGAATCCAATCATGAGCGCCGCCTT

1370 R S L T I S A S G S T T R D P Q A F N K

4201 AGATCTCTCACCATCAGTGCTAGTGGAAGCACCACCAGAGACCCGCAGGCATTCAACAAA

1390 A D S R D Y T E I D S V P P L P L Y A L

4261 GCTGACAGCAGAGACTACACAGAGATAGACTCTGTCCCCCCGCTTCCTTTGTACGCCTTA

1410 L A A D D D S Y C S S L E K T G S E S S

4321 CTTGCAGCAGATGATGACAGCTATTGCTCATCTTTGGAAAAAACTGGGAGTGAGAGTTCC

1430 L K K S K Q L S K E S Y D E L F Q T S V

4381 TTAAAGAAATCAAAACAGTTATCAAAAGAAAGTTATGATGAGCTTTTTCAGACTTCAGTT

1450 L M S D N H M L E T D E E N T Q P R V I

4441 CTGATGTCTGATAATCACATGTTAGAAACAGATGAAGAAAATACTCAGCCCAGAGTCATT

1470 D L S Q Y S P T Y F G P E H A Q V L S G

4501 GACCTTTCACAATACAGTCCAACTTATTTTGGTCCTGAGCATGCTCAGGTTCTTTCTGGC

1490 H L L H S S L P G L T R M E Q M S L M A

4561 CACTTACTTCATTCTAGCTTACCAGGGCTCACCAGGATGGAGCAGATGTCTTTGATGGCC

1510 L A D T I A T T S T D I G E S R D R N Q

4621 TTAGCCGATACAATTGCAACTACAAGCACTGATATTGGAGAAAGTAGAGACAGAAATCAA

1530 G G E T L D E C G L K F L L A V R L H T

Exon 19

4681 GGTGGGGAAACTCTAGACGAGTGCGGATTAAAGTTTCTTTTGGCTGTTCGACTCCATACC

1550 F L T T S L P A Y R A Q L L H Q G L S T

Exon 20

4741 TTTCTTACAACTTCCCTTCCAGCCTACCGAGCTCAACTCCTTCACCAAGGCCTTTCTACT

1570 G H F A W A F H S V A E E E L L N M L P

4801 GGTCATTTTGCTTGGGCTTTTCATTCAGTAGCAGAGGAAGAACTACTGAATATGTTGCCG

1590 A M Q K D D P T W S E L R A M G V G W W

4861 GCCATGCAGAAAGATGATCCAACTTGGTCTGAACTGAGAGCAATGGGTGTGGGGTGGTGG

1610 V R N A R I L R R C I E K V A K A A F H

Exon 21

4921 GTTCGCAATGCTCGCATCCTGCGCAGATGCATAGAAAAAGTAGCCAAAGCAGCCTTTCAT

1630 R N N D P L D A A I F Y L A M K K K A V

4981 AGAAATAATGATCCTTTAGATGCTGCAATTTTTTACCTTGCAATGAAAAAAAAAGCTGTG

1650 I W G L Y R S Q K D T K M T Q F F G H N

Exon 22
5041 ATTTGGGGATTATATAGATCTCAAAAAGACACCAAAATGACACAGTTTTTTGGGCACAAT
1670 F E E E R W R K A A L K N A F S L L G K
5101 TTTGAGGAAGAGAGGTGGCGTAAAGCAGCTTTGAAAAATGCCTTTTCTTTATTAGGCAAG
1690 Q R F E H S A A F F L L G G C L K D A I
5161 CAAAGGTTTGAACATTCTGCAGCATTTTTTCTTCTAGGTGGTTGCCTCAAAGATGCAATT
1710 E V C L E K L N D I Q L A L V I A R L F
Exon 23
5221 GAGGTATGTCTTGAGAAACTGAATGACATTCAGTTGGCTCTTGTAATAGCAAGACTATTT
1730 E S E F D K S A T Y K S I L R K K V L G
5281 GAATCTGAATTTGATAAGTCAGCAACATACAAGTCTATTTTACGCAAAAAAGTTTTGGGA
1750 I G S P A S E L S S S S I N A H H D P F
5341 ATTGGTTCTCCAGCCAGTGAGCTCAGTTCGTCAAGCATAAATGCACATCATGACCCCTTT
1770 L R S M A H W I L E D Y S A A L E T L I
5401 CTTCGGAGTATGGCACATTGGATTTTAGAAGATTATAGTGCTGCTCTGGAAACATTAATA
1790 K Q P V T E D E D Q V M M S A C N P I V
Exon 24
5461 AAGCAACCAGTTACAGAGGATGAAGATCAAGTCATGATGTCAGCCTGTAATCCTATAGTT
1810 F N F Y N Y L R T H P L L L R R H F G S
5521 TTTAATTTCTACAATTACCTAAGAACACATCCTCTTTTGCTAAGAAGACATTTTGGATCA
1830 S S E T F S T H M T L A G K S G L A G T
5581 TCATCGGAAACATTTTCAACACACATGACTTTAGCAGGAAAAAGTGGACTGGCAGGAACA
1850 I N L S E R R L F F T T A S A H L K A G
5641 ATTAATCTAAGTGAAAGAAGATTATTTTTTACTACTGCTAGTGCTCACCTAAAAGCTGGC
1870 C P M L A L E V L S K M P K V S K K A K
5701 TGCCCAATGTTGGCTTTGGAAGTATTGTCAAAGATGCCCAAAGTCAGTAAGAAAGCAAAA
1890 P C C R G S S F L T S K D S S L K L D V
5761 CCGTGTTGTAGAGGATCTAGTTTTTTAACTAGTAAAGATTCTTCACTAAAATTGGATGTA
1910 R E D K C C A A D W S P S L T N G L E S
5821 AGAGAAGACAAGTGTTGTGCTGCTGACTGGTCACCGTCGCTGACAAATGGTCTTGAATCT
1930 S S E G S S E R H S H S T L S F D W S Q

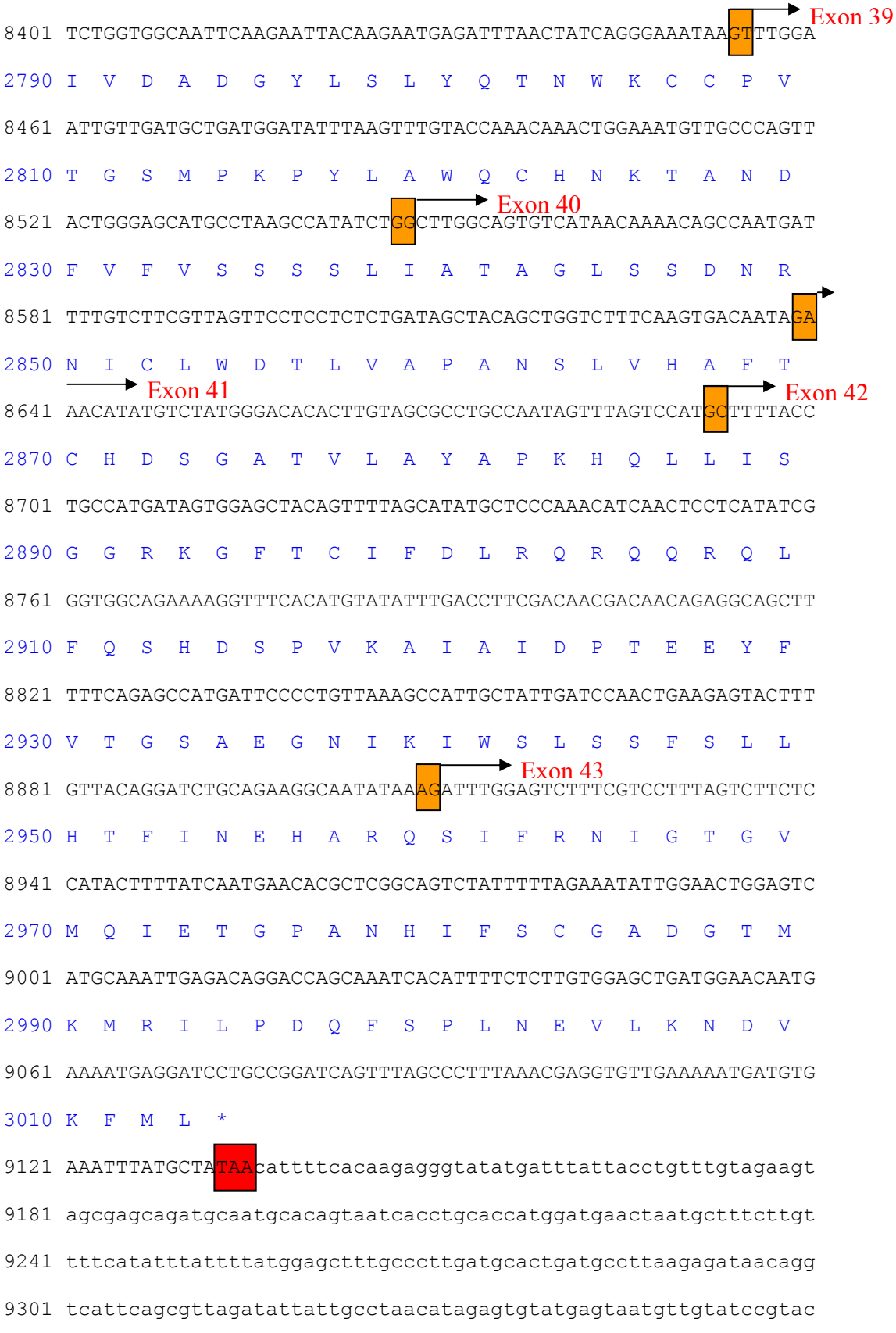
Anhang

```
5881 TCTTCAGAGGGTTCCTCAGAGAGGCACTCACATTCTACCCTTTCTTTTGATTGGAGCCAA
1950 P  S  V  V  F  Q  D  D  S  L  E  L  K  W  D  S  D  N  D  E
5941 CCGAGTGTGGTATTTCAAGATGACTCTTTGGAGTTGAAATGGGATAGTGATAATGATGAA
1970 E  N  E  D  P  P  I  S  M  K  E  I  R  P  L  Q  R  K  T  V
6001 GAAAATGAAGATCCCCCTATTTCAATGAAAGAAATAAGACCTTTGCAGAGAAAGACAGTT
1990 K  E  I  D  E  L  S  S  Y  T  D  S  L  S  T  L  D  E  N  D
6061 AAAGAAATAGATGAATTGAGCAGTTACACAGACTCTTTAAGCACACTAGATGAAAATGAC
2010 I  L  N  P  S  E  D  I  I  A  V  Q  L  K  F  R  A  C  L  K
6121 ATTTTAAATCCATCAGAAGATATAATTGCTGTTCAACTAAAATTTAGAGCATGTTTGAAG
2030 I  L  T  V  E  L  R  T  L  S  T  G  Y  E  I  D  G  G  K  L
6181 ATTCTCACCGTAGAACTTCGTACTTTATCTACTGGCTATGAAATAGATGGTGGGAAATTG
2050 R  Y  Q  L  Y  H  W  L  E  K  E  V  V  A  L  Q  R  T  C  D
6241 CGTTACCAACTCTACCACTGGCTTGAGAAAGAGGTAGTAGCTCTTCAGAGGACTTGTGAC
2070 F  C  S  D  A  D  Q  L  Q  T  T  F  S  Q  S  A  D  E  S  G
6301 TTTTGCTCAGACGCAGACCAGCTGCAGACCACATTTAGCCAGAGTGCAGATGAATCTGGA
2090 S  T  E  D  A  D  D  L  H  H  Q  T  K  V  K  Q  L  R  E  S
6361 TCAACTGAGGATGCTGATGATTTGCATCACCAAACAAAAGTGAAACAACTGAGAGAAAGT
2110 F  Q  E  K  R  Q  W  L  L  K  Y  Q  S  L  L  R  M  F  L  S
6421 TTCCAGGAAAAAAGGCAGTGGCTCTTGAAATATCAGTCACTCTTAAGAATGTTTCTTAGT
2130 Y  C  V  L  H  G  S  H  G  G  G  L  A  S  V  R  M  E  L  I
6481 TACTGTGTGCTTCATGGCTCTCATGGTGGGGGTCTTGCATCAGTGAGAATGGAATTAATT
2150 L  L  L  Q  E  S  Q  Q  E  T  A  E  P  I  F  S  N  P  L  S
6541 TTACTTTTGCAAGAATCTCAGCAGGAAACAGCAGAACCAATATTTTCTAACCCTTTGTCA
2170 E  Q  T  S  V  P  L  L  F  A  C  T  A  S  A  K  T  V  V  A
6601 GAACAAACTTCAGTGCCTCTTCTCTTTGCTTGTACAGCCAGTGCCAAAACAGTAGTTGCC
2190 N  P  L  L  H  L  S  N  L  T  H  D  I  L  H  A  I  I  N  F
6661 AATCCATTATTGCACCTTAGTAATTTAACACATGATATTCTACATGCCATAATAAATTTT
2210 D  S  P  P  H  P  D  S  Q  T  N  K  V  Y  V  M  H  T  L  A
```

6721 GATTCACCACCACATCCTGATAGCCAAACCAATAAAGTATATGTAATGCATACTTTAGCA
Exon 26
2230 A S L S A C I Y Q C L C G S H N Y S S F
6781 GCCTCACTTTCTGCTTGTATTTATCAGTGCCTTTGTGGTAGTCATAACTACAGTTCATTT
Exon 27
2250 Q T N Q F T G M V Y Q T V L L A H R H S
6841 CAAACAAATCAATTTACTGGAATGGTTTATCAAACAGTACTTCTTGCCCATCGGCATTCT
2270 L R T G S L D E S V T P N T S P A Q W P
6901 TTGCGAACTGGAAGTTTGGATGAATCAGTAACTCCCAACACATCACCAGCTCAATGGCCA
2290 G I N F L I Q L L N S S G E E A Q S G L
6961 GGAATAAATTTTCTAATTCAGCTTTTGAATTCCTCTGGGGAAGAAGCTCAGTCAGGACTT
Exon 28
2310 T V L L C E I L T A V Y L S L F I H G L
7021 ACAGTCTTGCTTTGTGAAATTCTCACAGCAGTATATCTTAGTCTGTTCATTCATGGGCTG
2330 A T H S S N E L F R I V A H P L N E K M
7081 GCAACACACTCTAGTAATGAGCTATTTCGGATTGTAGCCCACCCCCTAAATGAAAAAATG
2350 W S A V F G G G A H V P S K G Q A N S K
7141 TGGTCTGCAGTATTTGGTGGAGGAGCACATGTTCCCAGCAAAGGACAGGCAAACTCAAAA
2370 A L S V E G E K Q N R H I S P S K V S A
7201 GCTTTATCTGTTGAAGGAGAAAAACAAAACAGACATATTAGTCCATCAAAAGTGTCTGCC
Exon 29
2390 R E S P V S S S S G N Q E P P A V K E K
7261 AGAGAATCTCCTGTATCCTCTTCCTCTGGAAATCAGGAGCCACCTGCAGTCAAGGAGAAG
2410 F V P P E L S I W D Y F I A K P F L P P
7321 TTCGTCCCGCCTGAACTAAGTATCTGGGACTACTTCATAGCCAAGCCTTTTCTACCACCT
Exon 30
2430 S Q S R A E Y D S E E S L E S D D E E E
7381 TCCCAAAGTAGAGCAGAGTATGATTCTGAGGAAAGTCTAGAAAGTGATGATGAAGAAGAA
2450 E D D D D A L P S G L Q L H E H S N S N
7441 GAGGACGACGATGATGCTTTGCCTTCTGGTCTCCAGCTCCATGAGCATTCAAACTCAAAT
2470 S F S W S L M R L A M V Q L V L N N L K
7501 TCATTCAGTTGGTCTTTGATGAGATTGGCTATGGTCCAGTTGGTGCTCAACAATTTGAAG
Exon 31
2490 T F Y P F A G H D L A E L P V S S P L C

Anhang

7561 ACTTTCTATCCCTTTGCAGGTCATGATCTTGCAGAACTACCAGTTAGTTCACCTCTTTGT
Exon 32
2510 H A V L K T L Q C W E Q V L L R R L E I
7621 CATGCAGTTCTAAAAACTCTTCAATGCTGGGAACAAGTTCTTCTCCGGCGACTTGAAATC
2530 H G G P P Q N Y I S S H T S E E N V S A
7681 CATGGTGGACCACCTCAAAACTATATCTCAAGTCACACTTCTGAGGAGAATGTGTCTGCA
2550 G P A I L R H K A L L E P T N T P F K S
7741 GGCCCTGCAATACTCCGTCACAAAGCTTTACTAGAACCTACAAACACTCCTTTCAAATCC
Exon 33
2570 K N H L A L S V K R L W Q Y L V K Q E E
7801 AAAAATCATCTCGCACTATCTGTGAAGAGGCTTTGGCAGTACTTGGTGAAACAGGAAGAA
2590 I Q E T F I R N I F T K K R C L N E I E
7861 ATTCAAGAAACTTTTATCAGAAATATATTCACAAAGAAACGATGTCTAAATGAGATAGAG
Exon 34
2610 A D L G Y P G G K A R I I H K E S D I I
7921 GCAGATCTGGGATATCCTGGAGGTAAAGCAAGAATTATTCATAAGGAATCTGATATCATT
2630 T A F A V N R A N R N C I A I A S S H D
7981 ACTGCCTTTGCTGTTAATAGGGCAAATAGAAATTGCATAGCAATTGCCTCCAGCCATGAT
Exon 35
2650 V Q E L D V S A I L A T Q I Y T W V D D
8041 GTACAAGAACTGGATGTTTCTGCAATTCTGGCCACACAGATCTATACTTGGGTGGATGAT
2670 D T E T E T K G S E D F L V I H A R D D
8101 GATACAGAGACAGAAACCAAAGGATCAGAAGACTTCTTGGTTATACATGCGCGTGATGAT
Exon 36
2690 L S A V Q G S T P Y T H S N P G T P I N
8161 TTATCAGCTGTCCAGGGTTCAACTCCATACACACACAGCAACCCTGGCACTCCAATCAAC
2710 M P W L G S T Q T G R G A S V M L K K A
8221 ATGCCGTGGCTTGGAAGTACACAGACGGGCAGAGGAGCATCAGTGATGCTTAAGAAAGCC
Exon 37
2730 I N N V R R M T S H P T L P Y Y L T G A
8281 ATTAATAACGTTAGAAGAATGACATCACATCCAACTCTTCCTTACTATCTGACCGGAGCT
Exon 38
2750 Q D G S V R M F E W G H S Q Q I T C F R
8341 CAAGATGGAAGTGTCAGAATGTTTGAATGGGGCCATTCTCAGCAAATAACCTGTTTTCGG
2770 S G G N S R I T R M R F N Y Q G N K F G


8401 TCTGGTGGCAATTCAAGAATTACAAGAATGAGATTTAACTATCAGGGAAATAAGTTTGGA
Exon 39
2790 I V D A D G Y L S L Y Q T N W K C C P V
8461 ATTGTTGATGCTGATGGATATTTAAGTTTGTACCAAACAAACTGGAAATGTTGCCCAGTT
2810 T G S M P K P Y L A W Q C H N K T A N D
8521 ACTGGGAGCATGCCTAAGCCATATCTGGCTTGGCAGTGTCATAACAAAACAGCCAATGAT
Exon 40
2830 F V F V S S S S L I A T A G L S S D N R
8581 TTTGTCTTCGTTAGTTCCTCCTCTCTGATAGCTACAGCTGGTCTTTCAAGTGACAATAGA
2850 N I C L W D T L V A P A N S L V H A F T
Exon 41
8641 AACATATGTCTATGGGACACACTTGTAGCGCCTGCCAATAGTTTAGTCCATGCTTTTACC
Exon 42
2870 C H D S G A T V L A Y A P K H Q L L I S
8701 TGCCATGATAGTGGAGCTACAGTTTTAGCATATGCTCCCAAACATCAACTCCTCATATCG
2890 G G R K G F T C I F D L R Q R Q Q R Q L
8761 GGTGGCAGAAAAGGTTTCACATGTATATTTGACCTTCGACAACGACAACAGAGGCAGCTT
2910 F Q S H D S P V K A I A I D P T E E Y F
8821 TTTCAGAGCCATGATTCCCCTGTTAAAGCCATTGCTATTGATCCAACTGAAGAGTACTTT
2930 V T G S A E G N I K I W S L S S F S L L
8881 GTTACAGGATCTGCAGAAGGCAATATAAAGATTTGGAGTCTTTCGTCCTTTAGTCTTCTC
Exon 43
2950 H T F I N E H A R Q S I F R N I G T G V
8941 CATACTTTTATCAATGAACACGCTCGGCAGTCTATTTTTAGAAATATTGGAACTGGAGTC
2970 M Q I E T G P A N H I F S C G A D G T M
9001 ATGCAAATTGAGACAGGACCAGCAAATCACATTTTCTCTTGTGGAGCTGATGGAACAATG
2990 K M R I L P D Q F S P L N E V L K N D V
9061 AAAATGAGGATCCTGCCGGATCAGTTTAGCCCTTTAAACGAGGTGTTGAAAAATGATGTG
3010 K F M L *
9121 AAATTTATGCTATAAcattttcacaagagggtatatgatttattacctgtttgtagaagt
9181 agcgagcagatgcaatgcacagtaatcacctgcaccatggatgaactaatgctttcttgt
9241 tttcatatttatttatggagctttgcccttgatgcactgatgccttaagagataacagg
9301 tcattcagcgttagatattattgcctaacatagagtgtatgagtaatgttgtatccgtac

```
 9361 ttattcaaagtatattttagtaaacatgccatcgtgttagagagactttgttttctcatt
 9421 gttataaattctgcaagcactgctgcatgaagtccacagacagtaaaacccttaagagag
 9481 gtgccgaccaagagaacactcaccactgaactctccaccttctcacttctctgtcacaca
 9541 gacagctctaattgttttcctgttaataatgaaaacttcatataaatacagataagataa
 9601 gacattgtcactaaaactgctgttgggacttcagtgataagagaatccaagagtttattc
 9661 gtttttctctttccctcctataacccagaagtgtatatgtaaatgtatttatgtgtatgt
 9721 atatatgtgtgcatatgtgtatatatatatatatcttatacatcttacttgatattgtat
 9781 atacagaaacagaagctacaagtcttttcatgtattttgtgccataacacctactgccgc
 9841 agtataaaactcttggtttttttttaagtttttaaattattttaaccttaatgttacttt
 9901 taagaagctttaaatttattattattattttaaagatactgaatcaaataagtctcagta
 9961 ggcaataaagatgttgtgtggcacacaaggcattttatattaaaaatttgtcagcagttt
10021 tgatgatcagacattttagttaaattttgttttaattttatttcttttaagatgaattta
10081 gtctaaaaattctaaaaatgttatagcagtgtttttagaatcaaatgtgaaaacagtcaa
10141 tttaagtatagtatttacaaatgtacaatgccctgcagatctacattgataggaaataaa
10201 tgttatattttattattttgccaaaatattattttattataaaatttgaccaaaatattt
10261 aaaatttacaatgcctctatcttaaatatgataaccctatttctgtctgttttgtgctat
10321 atcttttctgattcagaattataaaaaaaacttgaaaaaaatacttgatttttagccaat
10381 gagactggctgtagaataagtgtttaagggacagttacatactatttagctcaaatttat
10441 ttttagttagaaccagaaatggccatttttatgtaatgaaatatggaaaactattaggaa
10501 ttgaagagtcaaagcaaagaaatgaagggctggtaaatgaattttgtaatatcctcagga
10561 tacttttattataaaagtgtattgttaaaggttttgtaattgtatttcatattttaaata
10621 tagtgagcaagttacagtgaaaacactgtcattatagagagtttatatgcacagagtaaa
10681 ctgtacctgtaaattgtgcaatagcaatatatgactatttgccatggaggaatctgtgac
10741 aacattacaaacaatattttctttgatgtagctaagttatttttttagtaatctcttcaca
10801 aatcaagattcaaaaggctttaaacactttccataagaaagaaaatttactatcgtgttt
10861 attaaaatcaaaagcttcagggatgatctattgagcattttaattgagagtttatatgaa
10921 taacaacattatgtaagtccatatgtatttacatccagagtcatattttaaataaacaat
10981 catgcagtgatttttttcgcatgcttttgttttaagaaatacttctgccagtttagttga
```

```
11041 ttttaattatgtgacttttaaattcagaatttctatatgtgatgattgaagagataattt
11101 aatttcctcaaataccttcttgttattttctacaatatttttacttaatgggtgctcat
11161 taaataattgaaatgtgtatagaaaataatacaaatacacttaaattttcattatccata
11221 caaatgaaatggtcacatcacaattaacttacatgctctgttttcatagagtcatgttac
11281 agagatttataacatggcaggtctcatgttctaaaatactggatttctcttaccaccgtg
11341 aaacactgcttacttgattttcctttggtcccacagtttgcttttagctaagatagaaca
11401 tttgctgttgtgcttaccagaactttaaagggattggtgtaaaccaagtcagcaggtcac
11461 ccttacagacattctttagttaaaaatgaggtttgcttgtttttaaaacagcaaaaatg
11521 ttcaagtggatctcattgactgaaacaatataattagacatttcagaaaatgcttataca
11581 gcctgaaagtggcaatggtatgattaaaaggctagaaacctaggcaaacatacgttacaa
11641 aacaagaaatttaaatatcttatgagaaagaacctcagccagttaaaaaccaagtcaact
11701 acaagaaatatttgtggagaggtgtgttagtgactctgtatgtggtataacttttatgtt
11761 attagagaaagtcaaactgaataacagggcagctcattaaaagttaataaaatattaaat
11821 ggccttaatagtagttagtgtaatctctcaagaaaaaaattatgaatagacataacattt
11881 tattgtgctttgaagataatatgaatgttttacaaatttaaggtttgtctcaactgcgca
11941 tgactatcaactccattttcaaatggtgcaaagttgcttcatagctgtcacattatagta
12001 attcttacaatatttcaaactataattatctctgtattggtttgatcacattattttggg
12061 atatcctgacctatacacatagaacactgcaaacttactcacataaatgcattttgattg
12121 ctccatgactctttgttccccagtctttgtctctttagatcttcctatccattgagatgc
12181 aatattaaaattagaccaattaacaaccaaaaaaaaaaaaaaaaaaaaaaaaa
```

Abb 3.6: Die cDNA- und Aminosäure-Sequenz des murinen *Dmxl*1

: Stopp-Codon

: Bereich der Consensusequenz des Translationsstartpunkts

: Translationsstart

: Exon/Intron-Grenze, markiert ist jeweils das letzte Nukleotid des vorhergehenden Exons und das erste Nukleotid des folgenden Exons

: Polyadenylierungssignal

Anhang

Die Abbildung wurde mit dem Tool „Show Translation“ aus „The Sequence Manipulation Suite“ (Stothard, 2000) erstellt. Die CDS wird durch große Buchstaben wiedergegeben, 5’-UTR und 3’-UTR durch kleine. Die Aminosäuresequenz (blau) ist im Einbuchstaben-Code dargestellt und steht jeweils am Anfang eines Codons. Exons, die durch homologe Rekombination während des Targeting-Ereignisses (Kapitel 3.x.x) deletiert wurden, sind durch grüne Buchstaben gekennzeichnet.

Danksagung

Zur Person

Veröffentlichungen:

Amid C, Bahr A, Mujica A, Sampson N, **Bikar SE**, Winterpacht A, Zabel B, Hankeln T, Schmidt ER.
Comparative genomic sequencing reveals a strikingly similar architecture of a conserved syntenic region on human chromosome 11p15.3 (including gene ST5) and mouse chromosome 7.
Cytogenet Cell Genet. 2001;93(3-4): 284-90.

Thomas Hankeln, Erwin R. Schmidt, Andreas Winterpacht, Bernhard Zabel sowie Clara Amid, André Bahr, **Sven-Ernö Bikar**, Thomas Brückmann, Andrea Cichutek, Alejandro Mujica, Natalie Sampson, Silke Schlaubitz
Die vergleichende Genomanalyse in Mensch und Maus als Werkzeug zur Identifizierung von Genen
Xpress, Informationen aus dem Deutschen Humangenomprojekt, 2000 (Nr.8)

ABSTRACTS:

Bahr, A., Amid, C., **Bikar, S.E**., Brückmann, T., Cichutek, A., Hankeln, T., Seipel, B., Schmidt, E.R., Winterpacht, A., Zabel, B. (1998)
Comperativ genomics: Sequencing of a 1Mb syntenic region in man (HsaC11p15) and mouse (MmuC7). Abstracts of the Human Genome Meeting 1998, Turin, Italy: 3

Schmidt, E.R., Bahr, A., Amid, C., **Bikar, S.E.**, Brückmann, T., Cichutek, A., Seipel, B., Sampson, N., Winterpacht, A., Hankeln, T., Zabel, B. (1998)
Comparative sequencing of a 1Mb region in man (chromosome 11p15) and mouse (chromosome 7)
Abstract of the 10th Genome Sequencing and Analysis Conference. Miami, Microbial and Comparative Genomics 3, C-37

Schmidt, E.R., Bahr, A., Amid, C., **Bikar, S.E.**, Brückmann, T., Cichutek, A., Seipel, B., Sampson, N., Winterpacht, A., Hankeln, T., Zabel, B. (1998)
Comparative sequencing of a 1Mb region in man (chromosome 11p15) and mouse (chromosome 7)
Abstracts of the International Symposium on Genomics and Proteomics – Functional and computational Aspects and Anual Meeting of the GfG, 4-7.10.1998 Heidelberg, Vol.6

Zabel, B., Amid C., Bahr A., **Bikar S.**, Brückmann T., Cichtek A., Mujica A., Sampson N., Schlaubitz S., Hankeln T., Winterpacht A., Schmidt E.R. (1999)
Identification of Genes and putative Gene Regulatory Sequences by comparative sequencing between man and mouse
Abstracts of the DHGP Genome Meeting 1999, Munich, Germany: 45

E.R. Schmidt, Hankeln, T., Amid, C., Bahr, A., **Bikar, S.**, Kraemer, C., Mujica, A., Zabel, B. (2001)
Comparative genomics and chromosome evolution.
Abstracts of the 14th International Chromosome Conference, Würzburg, 04.-08.09.2001

Verzeichnis der Lehrtätigkeit

Eingeladene Vorträge:

„Die Folgen des Humangenomprojekts: Menschenzüchtung oder Krankheitsbekämpfung? Therapeutisches Klonen, Stammzell-Forschung und Präimplantationsdiagnostik“
Vortrag zum Thema „Rote Gentechnik“ , Bertha von Suttner/ Kurfürst-Salentin-Gymnasium, Andernach, 04.2002

Übungen:

1999-2003	Übung zur Grundvorlesung in allgemeiner und molekularer Genetik 1 SWS

Kurse und Praktika

1996-2003	Analyse der Struktur und Funktion von Eukaryotengenen, FII-Praktikum, 6 Wo, ganztägig
1996-2003	Molekulare Genetik FI-Praktikum, 2 Wo, ganztägig

Betreuung von Diplom- und Doktorarbeiten:

Wiesbaden Oktober 2009

Eidesstattliche Versicherung

Hiermit gebe ich die eidesstattliche Versicherung ab, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Wiesbaden im Oktober 2009

Sven-Ernö Bikár