

Inhaltsverzeichnis

Vorwort.....	i
Danksagungen.....	ii
Veröffentlichungen.....	iii
A. Zusammenfassung	5
B. Einleitung	6
B.1 Eukaryotische Genstruktur.....	6
B.1.1 Transkription und Prozessierung eukaryotischer Gene.....	6
B.1.2 Intronklassen.....	7
B.2 DNA-Mutationen und veränderte Genstruktur	9
B.2.1 Übersicht über DNA-Mutationen.....	9
B.2.2 Beispiele für Veränderung der Exon/Intron-Struktur	11
B.2.2.1 Veränderung der Genstruktur durch veränderte Introns.....	11
B.2.2.2 Veränderung der Genstruktur durch Punktmutationen.....	12
B.2.2.3 Veränderung der Genstruktur durch Insertionen von Retroposons	12
B.2.2.4 Veränderung der Genstruktur über Rekombination	12
B.3 Serpine	13
B.3.1 Übersicht über die Serpine.....	13
B.3.2 Genstruktur der Serpine.....	14
B.4 Heparinkofaktor II.....	20
B.4.1 Das Heparinkofaktor II-Protein.....	20
B.4.2 Die Struktur des Heparinkofaktor II-Gens.....	22
B.4.3 Struktur des menschlichen HCII-Promotors	23
B.5 Evolutive Aspekte	25
B.5.1 Entwicklung und Verwandtschaft der untersuchten Arten.....	25
B.5.2 Evolution und HCII.....	25
B.6 Zielsetzung der Arbeit	27
C. Material und Methoden	28
C.1 Material.....	28
C.1.1 Geräte.....	28
C.1.2 Kits.....	28
C.1.3 Dokumentation.....	28
C.1.4 Chemikalien und Verbrauchsmaterialien.....	29
C.1.5 Software und Internet-Datenbanken.....	30
C.1.6 Enzyme.....	30
C.1.7 Primer	31
C.1.8 Plasmide	34
C.1.9 Nukleinsäuren.....	35
C.1.10 Bakterienstämme.....	35
C.1.11 Antibiotika	35
C.1.12 Medien.....	36
C.1.13 Puffer.....	37
C.1.14 Lösungen	38
C.1.15 DNA-Längenstandards	40

C.2 Methoden	41
C.2.1 Basic-Programme.....	41
C.2.2 PCR.....	44
C.2.2.1 DNA-Polymerasen.....	45
C.2.2.2 PCR-Ansatz.....	46
C.2.2.3 Nested PCR.....	47
C.2.2.4 Kolonie-PCR.....	47
C.2.3 Reverse Transkriptase gekoppelte PCR.....	47
C.2.3.1 Genescript reverse Transkriptase.....	48
C.2.3.2 Omniscript reverse Transkriptase.....	48
C.2.3.3 'C. therm. Polymerase'.....	49
C.2.3.4 MuLV reverse Transkriptase.....	50
C.2.4 Anker-PCR (SLIC-Strategie).....	51
C.2.5 Tailing-PCR.....	52
C.2.6 Techniken zur Analyse von DNA und RNA.....	53
C.2.6.1 Agarosegelelektrophorese von DNA.....	53
C.2.6.2 Reinigung von DNA.....	54
C.2.6.3 Konzentrationsbestimmung von DNA und RNA.....	55
C.2.6.4 Spaltung von DNA mit Restriktionsenzymen.....	56
C.2.6.5 Ligation von DNA-Restriktionsfragmenten.....	56
C.2.6.6 Isolierung von Restriktionsfragmenten.....	57
C.2.6.7 Dephosphorylierung von DNA.....	57
C.2.6.8 DEPC-Behandlung von Puffern und Gefäßen.....	58
C.2.6.9 Isolierung von Gesamt-RNA aus tierischen Zellen.....	58
C.2.7 Anzucht und Lagerung von Zellen.....	58
C.2.7.1 Anzucht und Lagerung von <i>E. coli</i> -Stämmen.....	58
C.2.7.2 Anzucht und Lagerung von eukaryotischen Zelllinien.....	59
C.2.8 Plasmidisolierung.....	60
C.2.8.1 'S-Lyse' (Plasmid-DNA für Restriktionsspaltungen und Klonierungen).....	60
C.2.8.2 'AS-Lyse' (Plasmid-DNA für Sequenzierungen).....	61
C.2.9 Klonierung von Plasmid-DNA.....	61
C.2.9.1 Insertionsinaktivierung eines Antibiotika-Resistenzgens.....	61
C.2.9.2 Insertionsinaktivierung des lacZ' -Gens.....	61
C.2.9.3 Plasmid-Schnell-Charakterisierung anhand der Größe der ccc-Form.....	62
C.2.9.4 Kolonie-PCR.....	62
C.2.10 Transformation von Bakterienzellen.....	62
C.2.10.1 Herstellung der kompetenten Zellen.....	62
C.2.10.2 Transformation.....	62
C.2.11 Sequenzierung.....	63
C.2.11.1 Sequenzierreaktionen.....	64
C.2.11.2 Gellauf für GATC 1500 DNA Sequencer ('Long Run System').....	64
C.2.12 Southernblot.....	65
C.2.12.1 Vorbereitung des Agarosegels.....	65
C.2.12.2 DNA-Transfer.....	65
C.2.12.3 Herstellung der Sonden.....	66
C.2.12.4 Fixierung und Hybridisierung.....	67
C.2.12.5 Chemolumineszenz-Nachweisreaktion.....	67
C.2.13 Promotoranalyse.....	68
C.2.13.1 Transfektion von HepG2-Zellen.....	68
C.2.13.2 Bestimmung der Promotoraktivität.....	68

D. Ergebnisse	71
D.1 Analyse des 5'-Bereichs vom Heparinkofaktor II-Gen der Ratte	71
D.1.1 Sequenzierung des 5'-Bereichs vom HCII-Gen der Ratte	71
D.1.2 Theoretische Exon/Introngrenzen im 5'-Bereich des rHCII-Gens	74
D.2 Identifizierung von zwei verschiedenen HCII-mRNAs in Mausleber	75
D.3 Analyse des 5'-Bereichs vom Heparinkofaktor II-Gen der Maus.....	76
D.3.1 Sequenzierung des 5'-Bereichs vom HCII-Gen der Maus.....	76
D.3.2 Alternatives Spleißen von HCII-mRNAs der Maus	78
D.3.3 Theoretische Exon/Introngrenzen in dem sequenzierten Bereich	79
D.4 Alignment der Sequenzen von Intron 1 des Heparinkofaktor II-Gens	81
D.4.1 Alignment.....	81
D.4.2 Suche nach repetitiven Elementen in den Sequenzen.....	89
D.4.3 Überprüfung auf ein Exon 1B im menschlichen Gen.....	91
D.5 Beteiligung eines repetitiven Elements an einer Introngrenze.....	92
D.6 Alignment von Heparinkofaktor II-cDNA-Sequenzen.....	93
D.7 Heparinkofaktor II-Transkripte in verschiedenen Gewebetypen	98
D.8 Startpunkte von menschlichen Heparinkofaktor II-Transkripten.....	100
D.8.1 Methodik zur Identifizierung von Transkriptionsstartpunkten.....	100
D.8.2 Startpunkte der HCII-Transkripte in menschlicher Leber.....	102
D.8.2.1 Vergleich von A-PCR- und T-PCR-Produkten	102
D.8.2.2 Sequenzierung des 5'-Endes der cDNA aus Leber.....	104
D.8.3 Startpunkte von HCII-Transkripten in menschlicher Lunge.....	105
D.8.3.1 Vergleich von A-PCR- und T-PCR-Produkten	105
D.8.3.2 Sequenzierung des 5'-Endes der cDNA aus Lunge	107
D.8.4 Gegenüberstellung der gefundenen Startpunkte.....	109
D.9 Untersuchungen zur Promotorstruktur des Heparinkofaktor II-Gens	110
D.9.1 Bindungsstellen von Transkriptionsfaktoren	110
D.9.1.1 Mögliche Bindungsstellen vor dem 2,3 kB-Transkript	112
D.9.1.2 Mögliche Bindungsstellen vor den längeren Transkripten	114
D.9.2 Repetitive Elemente vor dem Exon 1 des Heparinkofaktor II-Gens	115
D.9.3 Promotoranalyse der 2,3 kB-Transkripte.....	117
E. Diskussion	119
E.1 Nachweis von Exon/Introngrenzen.....	119
E.1.1 Exon/Introngrenzen im 5'-Bereich des HCII-Gens der Ratte.....	119
E.1.2 Exon/Introngrenzen im 5'-Bereich des HCII-Gens der Maus.....	120
E.1.3 Exon/Introngrenzen im 5'-Bereich des HCII-Gens des Menschen.....	120
E.1.4 Vergleich der Exon/Introngrenzen im 5'-Bereich des HCII-Gens.....	121
E.2 Alignment von Intron 1-Sequenzen von Heparinkofaktor II	121
E.2.1 Statistische Daten des Intron 1	122
E.2.2 Vergleich der theoretischen mit den natürlichen Introngrenzen	123
E.2.3 Vergleich der Gesamtstruktur von Intron 1.....	124
E.2.3.1 Vergleich der Intronstrukturen bezüglich Exon 1B.....	128
E.2.3.2 Vergleich der Intronstrukturen bezüglich des 3'-Intronbereichs.....	128
E.2.4 Alignment der Exon/Introngrenzen	128
E.2.4.1 Alignment der Exon/Introngrenze am 3'-Ende von Exon 1	128
E.2.4.2 Alignment der Exon 1B-Region.....	129
E.2.4.3 Alignment des 5'-Bereichs von Exon 2.....	131
E.2.5 Zusammenfassung.....	133

E.3 Vergleich verschiedener cDNAs von Heparinkofaktor II.....	134
E.3.1 Vergleich der Heparinkofaktor II-cDNAs.....	134
E.3.1.1 Evolutiver Vergleich	134
E.3.1.2 Größere Unterschiede zwischen verschiedenen HCII-cDNA-Sequenzen	136
E.3.2 Vergleich der HCII-Exon/Introngrenzen	139
E.3.3 Evolution des Heparinkofaktor II: Vergleich von Intron 1 und cDNA.....	139
E.4 Transkriptionsstartpunkte des menschlichen Heparinkofaktor II-Gens.....	140
E.4.1 Startbereich der 2,3 kB-Transkripte.....	140
E.4.2 Startbereich der langen Transkripte (Exon 1*).....	141
E.4.3 Promotoranalyse.....	143
E.4.3.1 Promotorregion der 2,3 kB-Transkripte	144
E.4.3.2 Promotorregion der langen Transkripte mit Start im Intron 1.....	147
E.5 Zusammenfassung der HCII-Genstruktur	149
F. Verzeichnisse	151
F.1 Literaturverzeichnis	151
F.2 Abkürzungen	166
F.3 Tabellen-Verzeichnis	169
F.4 Abbildungs-Verzeichnis.....	170
G. Anhang.....	172
G.1 Plasmidkarten.....	172
G.1.1 pPBK14-Derivate.....	172
G.1.2 pRL-TK (<i>Renilla</i> Luciferase/Kontroll-Vektor).....	173
G.1.3 pPBK60 (HCII-Promotor/Luciferase-Vektor)	174
G.2 Intron 1-Sequenzierungen des HCII-Gens von Maus und Ratte.....	175
G.2.1 Sequenzierungen des Maus HCII-Intron 1	175
G.2.2 Sequenzierungen des Ratten HCII-Intron 1.....	179

A. Zusammenfassung

Die vorliegende Arbeit beschreibt neue Erkenntnisse zur Organisation und Expression des Heparinkofaktor II-Gens (HCII-Gen).

Die Struktur des 5'-gelegenen Bereichs des HCII-Gens von Maus und Ratte wurde aufgeklärt. Der Intron 1-Bereich umfaßt beim HCII-Gen der Maus 4,8 kB. Das HCII-Intron 1 der Ratte weist dagegen eine Länge von 4,3 kB auf.

Die Intron 1 Sequenz bestätigt, daß die zwei verschiedenen HCII-Transkripte der Ratte¹ auf differentiellem Spleißen beruhen.

Darüber hinaus wurde nachgewiesen, daß auch beim HCII-Gen der Maus zwei unterschiedliche HCII-Transkripte im Lebergewebe existieren. Diese Transkripte entsprechen nicht den HCII-mRNAs der Ratte. Sie werden jedoch ebenfalls durch differentielles Spleißen gebildet.

Eine Gegenüberstellung der ermittelten HCII-Sequenzen und der entsprechenden menschlichen Sequenz² legt nahe, daß das zusätzliche Exon 1B der Ratte vor allem durch einfache Punktmutationen *de novo* aus Intronsequenzen entstand. Das differentielle Spleißen im 5'-Bereich des HCII-Gens der Maus kann durch Punktmutationen, durch Rekombinationsereignisse von (CA)_n-Wiederholungen oder durch unterschiedliche Enden eines Retroposons verursacht sein. Die Exon/Intronstruktur innerhalb der α_1 -Antitrypsin-Serpinfamilie ist stark konserviert. Ein Vergleich der bekannten HCII-cDNA-Sequenzen zeigt die übereinstimmende Genstruktur im kodierenden Sequenzbereich. Variabel scheint dagegen das 5'-Ende zu sein, vergleichbar dem α_1 -Antitrypsin.

Der Transkriptionsstartbereich menschlicher 2,3 kB HCII-Transkripte ließ sich über die T-PCR-Methodik identifizieren. Die Startpunkte der häufigsten Transkripte erstrecken sich über einen Bereich von 45 Bp vor Exon 1.

Die Suche nach bekannten proximalen Promotorelementen in einer Region vor dem Startbereich der mRNAs legt nahe, daß die Transkription nicht von einer TATA-Box, sondern von INR-Elementen gesteuert wird.

Weitere Untersuchungen ergaben, daß HCII- und auch Antithrombin-Transkripte nicht nur in menschlicher Leber vorkommen, sondern auch in unterschiedlichen Konzentrationen im Gewebe des Gehirns, des Herzens, der Lunge, der Niere und der Plazenta vorliegen.

In poly(A)⁺-RNAs der genannten Gewebe ließen sich längere HCII-Transkripte nachweisen, die Schmidt zuvor in einer Lungen-cDNA-Bank entdeckte.³ Eine Charakterisierung dieser Transkripte ergab, daß das Exon 1 der 2,3 kB-Transkripte nicht vorhanden ist. Die mRNAs beginnen in einem über 1 kB großen Bereich vor dem Exon 2.

Mit Hilfe eines Reportergenkonstruktes wurde gezeigt, daß ein 2,2 kB großer, 5'-gelegener Bereich des menschlichen HCII-Gens keine Expression der Transkriptionseinheit in einer Leberzelllinie bewirkt.

¹ Westrup & Ragg, 1994

² Herzog *et al.*, 1991

³ Schmidt, 1997

B. Einleitung

B.1 Eukaryotische Genstruktur

B.1.1 Transkription und Prozessierung eukaryotischer Gene

In dieser Arbeit werden die Genstrukturen, die für das gleiche Protein in unterschiedlichen Eukaryoten (hauptsächlich Mammalier) kodieren, untersucht. Aus diesem Grund wird hier kurz auf die Genstruktur typischer eukaryotischer Gene eingegangen.

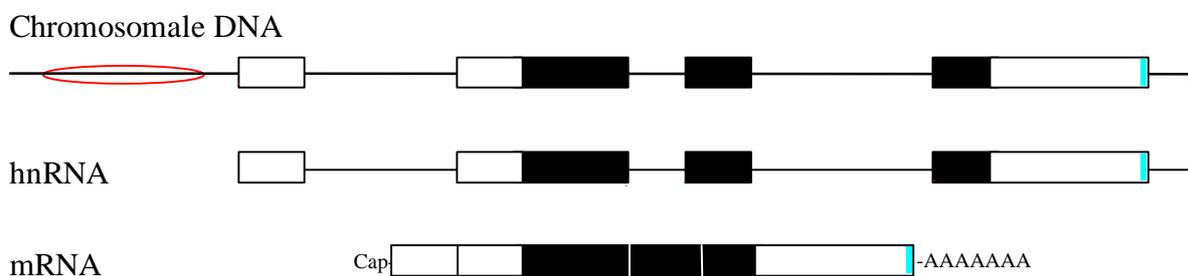
In Eukaryoten existieren drei verschiedene Klassen von Genen, die von drei unterschiedlichen RNA-Polymerasen durch spezifische Transkriptionssignale erkannt und transkribiert werden. In Tabelle 1 ist eine Übersicht über die eukaryotischen RNA-Polymerasen und die von ihnen transkribierten Gene gegeben.

Tabelle 1: RNA-Polymerasen

RNA-Polymerase	Vorkommen	transkribierte Gene
RNA-Polymerase I	Nukleolus	ribosomale Gene
RNA-Polymerase II	Kernplasma	proteinkodierende Gene
RNA-Polymerase III	Kernplasma	Gene für kleine RNAs

In dieser Arbeit werden proteinkodierende Gene analysiert, also Gene, die von der RNA-Polymerase II in hnRNA transkribiert werden. Im Gegensatz zu den Prokaryoten und Viren (Ausnahmen: z. B. manche Gene in Archaeobakterien) sind die meisten proteinkodierenden Gene in Eukaryoten durch Introns unterbrochen. Die von der chromosomalen DNA transkribierte hnRNA (prä-mRNA) wird im Verlauf der weiteren Prozessierung bei den meisten Transkripten am 5'-Ende mit einer 'Cap'-Struktur versehen, und am 3'-Ende wird eine Reihe von Adenin-Nukleotiden angehängt (Polyadenylierung). Darüber hinaus werden von einem als Spleißosom bezeichneten Komplex aus mehreren Proteinen und kleinen RNAs (U1, U2 ...) Stücke aus der prä-mRNA herausgeschnitten, die als Introns bezeichnet werden. Die verbleibenden Stücke, die vom Spleißosom wieder zusammengesetzt werden, werden als Exons bezeichnet. Die auf diese Art prozessierte, fertige mRNA wird schließlich im Zytosol von Ribosomen in Proteine umgeschrieben bzw. translatiert. In der Abbildung 1 ist eine Skizze zur eukaryotischen Genstruktur von proteinkodierenden Genen gegeben.

Abbildung 1: Struktur von eukaryotischen proteinkodierenden Genen



Linien stellen Introns bzw. nicht transkribierte Bereiche dar. Kästen symbolisieren Exons (bzw. als Exons erkannte Sequenzbereiche). Ausgefüllte Kästen deuten den Bereich der kodierenden Sequenz an. Das rote Oval stellt einen von der RNA-Polymerase II erkannten Promotorbereich dar und die blaue Linie ein poly(A)-Signal.

Die Verzweigungsstelle ist nur wenig charakterisiert. Das 5'-Ende des Introns wird an eine 2'-Position eines A-Nukleotids ligiert, welches ca. 18 bis 40 Nukleotide stromaufwärts von der 3'-Spleißstelle liegt. Das Nukleotid ist von der folgenden Sequenz umgeben:⁶

Y₈₀ N Y₈₀ Y₈₇ U₇₅ **A**₁₀₀ Y₉₅

Über alternatives Spleißen ist eine Kodierung unterschiedlicher Proteine aus dem gleichen chromosomalen Genbereich möglich. Der Grund für ein unterschiedliches, spezifisches Spleißen in verschiedenen Geweben könnte in einer Balance zwischen positiven und negativen Exon-Kontroll-Signalen auf der mRNA liegen. Für neuronspezifische Exons sind positive Signale bekannt, die sich in der Nähe von oder im Exon erstrecken sowie negative Signale, die im Polypyrimidin-Bereich oder in einer Intron-Verstärker-Region liegen.⁷

Der Spleißvorgang beginnt, indem sich das U1 snRNP an die 5'-Spleißstelle und das U2 snRNP an den Polypyrimidinbereich anlagert. Anschließend binden U4-, U5- und U6 snRNPs an die 3'-Spleißstelle. Nach dem Schnitt an der 5'-Spleißstelle des Introns wird das 5'-Ende der Intronsequenz mit der 2'-Position des A-Nukleotids an der Verzweigungsstelle verbunden und das U1 snRNP wird freigesetzt. Zuletzt erfolgt der Schnitt an der 3'-Spleißstelle und die Ligation der Exonsequenzen.

Klasse II Introns

Neben den bekannteren und weitaus häufigeren Klasse I Introns wurde eine weitere Klasse von Introns in Eukaryoten entdeckt ('ATAC'-Introns).⁸ Die essentiellen Signale, die für diese Introngruppe wichtig sind, sind in Abbildung 3 dargestellt.

Abbildung 3: Intron signale der Klasse II Introns

```

5'          3'          5'          3'
<-----Intron-----> <-----Intron----->

```

¾¾ **AC** █████ **AT**¾¾¾¾¾¾¾¾¾¾¾¾¾¾¾¾ **A**¾¾¾¾¾¾ **YAC** █████ **AT**¾¾¾¾¾¾ **A**¾¾¾¾¾¾ **YAC** █████ **AT**¾¾¾

GT: 5'-Spleißstelle, erkannt von U11 snRNP; AG: 3'-Spleißstelle, erkannt von U5+U4/U6atac⁹ snRNPs; Verzweigungs-Nukleotid (A): Schleifenbildung, Bindung von U12 snRNP; Y: Pyrimidin

Die Erkennung und Prozessierung der Klasse II Intron-Signale verläuft im Vergleich zu Klasse I Introns ähnlich (siehe Klasse I Introns). Die für Klasse I Introns beschriebenen snRNPs sind beim Spleißvorgang durch Klasse II snRNPs wie folgt ersetzt: U1 durch U11, U2 durch U12 und U4-U6 durch U4-U6atac.

Ein Polypyrimidin-Bereich ist in diesen Introns nicht vorhanden. Die invarianten Dinukleotidpaare an den Enden von Klasse I Introns (GT und AG) sind hier durch AT und AC ersetzt. Bei den Klasse II Introns sind zusätzliche Erkennungssequenzen stark konserviert und im folgenden sind die Matrizen für das 5'- und 3'-Ende eines Introns aufgeführt.⁹

⁶ Lewin, 1991

⁷ Grabowski, 1998

⁸ Hall & Padgett, 1994

⁹ Hall & Padgett, 1996; Tarn & Steitz, 1996

Introngrenzen der Klasse II ('AT-AC') Introns

5'-Introngrenze:

+1	+2	+3	+4	+5	+6	+7	+8
A	T	A	T	C	C	T	T

3'-Introngrenze:

-5	-4	-3	-2	-1
Y	C	C	A	C

Die Verzweigungsstelle ist etwas stärker konserviert als bei den Klasse I Introns. Das 'A'-Verzweigungsnukleotid ist von der folgenden Sequenz umgeben:

T C C T T U A C

Ein 'Y' steht allgemein für Pyrimidine, ein 'U' für Purine und ein 'N' für ein beliebiges Nukleotid. Da bisher nur sehr wenige Introns dieser Klasse bekannt sind, wurden noch keine Aussagen über die Häufigkeit der Nukleotide in den einzelnen Positionen getroffen.

Es gibt mittlerweile eine Reihe von Beispielen von Klasse II Introns. Zuerst wurden vier Gene beschrieben, die Introns dieser Klasse enthielten. So enthält das CMP Gen (cartilage matrix protein) bei Mensch und Huhn, das p120 Gen (proliferating cell nucleolar protein) von Mammaliern, das Maus Rep-3 Gen (Genprodukt ist wahrscheinlich beteiligt an der DNA-Reparatur) und das *Drosophila* Prospero Gen (Homöobox-Gen, Produkt ist beteiligt an der Kontrolle der neuronalen Identität) jeweils ein AT-AC-Intron (neben Introns der Klasse I).⁸ Mittlerweile wurden weitere Gene mit Introns dieser Klasse gefunden.

An der DNA-Synthese oder am Zellzyklus sind Gene der folgenden Produkte beteiligt (neben P120 und Rep-3): das menschliche E2F und das pflanzliche ArLIM15. Im Menschen werden die Gene SCN4A, SCN5A und SCN8A (Produkte kodieren Untereinheiten von Natrium-Kanälen), die AT-AC-Introns enthalten, gewebsspezifisch exprimiert. In der Maus wurde ein Klasse II-Intron im Cdk5-Gen gefunden, das ebenfalls gewebsspezifisch exprimiert wird. Darüber hinaus wurde im HPS-Gen (beim Menschen und der Maus) und im pflanzlichen AtG5-Gen ein Intron dieser Klasse gefunden.¹⁰

B.2 DNA-Mutationen und veränderte Genstruktur

B.2.1 Übersicht über DNA-Mutationen

Viele Mechanismen zur Veränderungen einer DNA-Sequenz sind bekannt. Neben einfachen Punktmutationen, die in einer Veränderung von nur einem Basenpaar (Austausch, Insertion oder Deletion) resultieren, können Duplett- oder Triplet-Wiederholungen der gleichen Basen zu größeren Längenveränderungen in der DNA führen (durch zusätzliche oder fehlende Basen dieses Motivs).

Größere Insertionen und Deletionen im Genom werden hauptsächlich durch eine Reihe von verschiedenen Sequenzen begünstigt. Zum einen können dies längere (TG)_n/(CA)_n-Repeats bewirken. Unter bestimmten Bedingungen kommt es hier zu einer linksgewundenen DNA (Z-DNA).

¹⁰ Tarn & Steitz, 1997

Bei einer $d(TG)_{30}$ -Wiederholung wurde eine 20fache Stimulation der homologen Rekombination beobachtet. Es ist Rekombination zwischen zwei $d(TG)_n$ -Wiederholungen möglich, jedoch auch im schwächeren Umfang mit anderen Sequenzen. Das Z-DNA-Motiv scheint in diesem Fall als Rezipient für genetische Information zu fungieren.¹¹ An den Übergangsbereichen von linksgewundener Z-DNA zur normalen, rechtsgewundenen B-DNA wird in den verbindenden Bereichen ein aufgewundener DNA-Doppelstrang präsentiert. Dadurch wird einer Reihe von Enzymen der Zugang zur DNA erleichtert. Darüber hinaus begünstigen diese Sequenzen auch ohne Proteine Rekombinationsereignisse über die Bildung von viersträngigen Strukturen. Diese werden von speziellen Proteinen (sie gehören zu den HMG-Proteinen) erkannt.¹²

Zum anderen liegen in den Genomen höherer Eukaryonten eine Vielzahl verschiedener, mobiler DNA-Elemente vor. 'Springt' ein solches Element an eine andere Stelle im Genom, wird dort diese zusätzliche Sequenz inseriert und zusätzlich werden die flankierenden Bereiche verändert. Normalerweise kommt es zu einer zusätzlichen Insertion von wenigen Basen an den jeweiligen Enden. Es werden jedoch auch fehlende Bereiche der umgebenen DNA an den Enden des inserierten Elements beobachtet. Es wird spekuliert, daß mobile Elemente u. a. dazu dienen können, Chromosomenbrüche zu schließen.¹³

Neben Insertionen dieser Elemente kann es auch über Rekombinationsereignisse und eigenkodierte Enzyme zur Deletion bzw. Excision von mobilen Elementen kommen.

Eine ständig wachsende Zahl von mobilen Elementen ist bekannt. Sie werden nach den strukturellen Merkmalen der Enden, einer eigenen Transposase und dem Replikationsmechanismus (noch nicht einheitlich) klassifiziert. Zwei der bekanntesten und weitverbreitetsten Elemente sind die LINE1- und die *Alu*-Familie (bzw. Bezeichnung: B-Elemente in der Maus). Beide Elemente verbreiten sich im Genom über Transkription in RNA und reverse Transkription zurück in eine cDNA, die integriert wird (siehe Abbildung 4). Aus diesem Grund werden sie zu den Retro(trans)posons gezählt. Im haploiden menschlichen Genom gibt es ca. 300 000 *Alu*-Kopien. B1-Elemente kommen im Mausgenom in ca. 50 000 Kopien vor. Ein *Alu*-Element besteht normalerweise aus einer tandemförmigen Duplikation von etwa 130 Bp. Aufgrund der Ähnlichkeit zur 7SL-RNA wird eine Abstammung von dieser für möglich gehalten.^{14,15} Die Gruppe dieser Elemente wird auch als SINEs (short interspersed elements) bezeichnet. Sie werden von der RNA-Polymerase III transkribiert, können jedoch trotzdem polyadenyliert sein.¹⁶

Elemente der LINE-Familie (long interspersed elements) sind wesentlich größer als SINEs. Das vollständige L1-Element ist etwa 7 kB groß. Die meisten Elemente sind allerdings stark deletiert, so daß an vielen Stellen im Genom nur kurze Fragmente vorliegen. Vollständige L1-Elemente haben 2 offene Leserahmen (ORFs) und einen poly(A)-Schwanz. ORF2 kodiert für eine Endonuklease und reverse Transkriptase. Maus L1-Elemente unterscheiden sich von den menschlichen L1 durch ihre Tandem-Wiederholungen am 5'-Ende.¹⁷ Die L1-Elemente sind größtenteils sehr alt (mit Ausnahme von neueren Transpositionseignissen, s.u.). Mehr als die Hälfte der untersuchten Elemente wurde vor der Säugetierauftrennung ins menschliche Genom inseriert.¹⁸

¹¹ Hamada *et al.*, 1984; Wahls *et al.*, 1990

¹² Gaillard & Strauss, 1994

¹³ Boeke, 1996; Moore & Haber, 1996; Teng *et al.*, 1996

¹⁴ Lewin, 1991

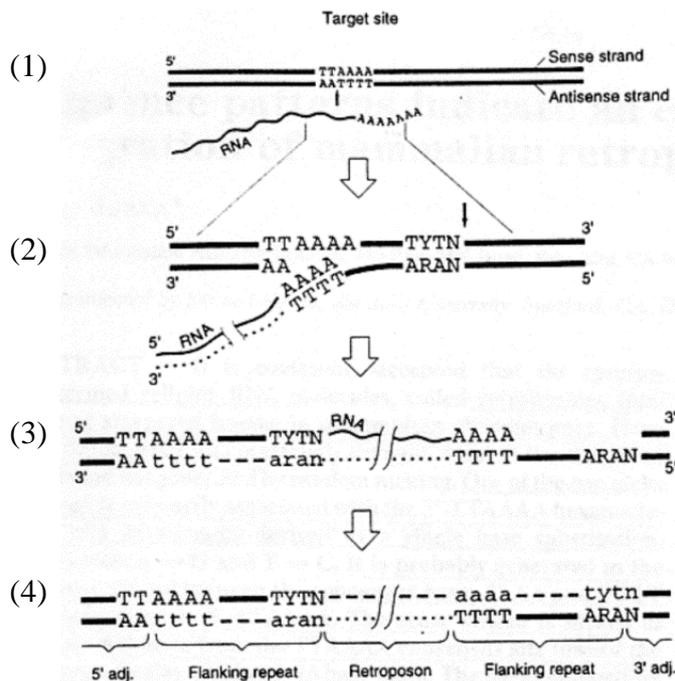
¹⁵ Darnell *et al.*, 1990

¹⁶ Kramerov *et al.*, 1990

¹⁷ Naas *et al.*, 1998

¹⁸ Smit *et al.*, 1995

Abbildung 4: Modell zur Verbreitung von Retroposons



Die Abbildung zeigt ein Modell zur Verbreitung von Retroposons im Genom (das Schema wurde von Jurka¹⁹ übernommen). Zuerst kommt es an einer speziellen Zielsequenz [(Y)_n(R)_n] zu einem Einzelstrangbruch der chromosomalen DNA, eventuell in Anwesenheit der mRNA (1). Diese Zielsequenz wird von der L1-ORF2-Endonuklease erkannt. (2) Das Poly(A)-Ende der mRNA hybridisiert mit T-Nucleotiden aus der DNA und die Reverse Transkription beginnt. Parallel wird der andere, zuvor intakte DNA-Strang geschnitten. (3) Über DNA-abhängige DNA-Synthese werden die einzelsträngigen Enden aufgefüllt (neusynthetisierte Nucleotide sind durch kleine Buchstaben und gestrichelte Linien dargestellt). (4) Die Insertion wird über DNA-Reparaturmechanismen vervollständigt (Austausch der RNA-Nucleotide gegen DNA und Ligation der Enden).

Spezifische Retroposons scheinen sich quantitativ im Genom in zeitlichen Wellen zu verbreiten, jeweils ausgelöst durch ausgetauschte Promotorsequenzen.²⁰ Das B2-Element kommt zum Beispiel bei mausähnlichen Nagetieren in einer Kopienzahl von bis zu 100 000 vor. Beim Eichhörnchen kommt das gleiche Element nur in 2 500 Kopien vor. Die große Zahl von Kopien in ersterem muß also in einem evolutiv gesehen sehr kurzem Zeitraum nach Aufteilung der Arten entstanden sein.²¹

B.2.2 Beispiele für Veränderung der Exon/Intron-Struktur

Größere Veränderungen wie die Insertion von repetitiven DNA-Elementen oder selbst nur einzelne Punktmutationen an essentiellen Stellen in der Nähe von Genen oder in deren Introns können dazu führen, daß sich die Struktur von Genen durch zusätzliche oder fehlende Exons stark ändert.

B.2.2.1 Veränderung der Genstruktur durch veränderte Introns

Eine Veränderung der Genstruktur kann nicht nur über unterschiedliche Exons eines Gens verursacht werden, sondern auch über zusätzliche oder fehlende Introns. So sind in manchen Fällen im Verlauf der Evolution (über revers transkribierte mRNAs und Reintegration der gebildeten cDNA ins Genom) Gene entstanden, denen Intronbereiche fehlen (Beispiel: Globin-Gene in Insekten,²² B14-Legumin-Gen²³). In anderen Fällen ist es im Verlauf der Evolution zu einer Insertion von zusätzlichen Introns gekommen (Beispiele: SRY-Gen von *S. macrourea*,²⁴ COX1-Gen der Hefe²⁵).

¹⁹ Jurka & Klonowski, 1996

²⁰ Brookfield, 1995

²¹ Kass *et al.*, 1997

²² Gruhl *et al.*, 1997

²³ Häger *et al.*, 1996

²⁴ O'Neill *et al.*, 1998

²⁵ Yang *et al.*, 1996

Die Transkriptionseffizienz kann über zusätzliche oder fehlende Introns beeinflusst werden. Diese Veränderungen wirken sich allerdings nicht auf die fertige mRNA aus und können somit auch keine Unterschiede im kodierten Protein bewirken. Im Anschluß werden Beispiele gegeben, in denen sich durch Mutationen im Genom auch die mRNA ändert.

B.2.2.2 Veränderung der Genstruktur durch Punktmutationen

Bisher gaben Kudo und Fukuda ein Beispiel für den Verlust eines Exons (Exon-Skipping) durch eine Punktmutation in der 5'-Spleißstelle des dritten Introns des Glycophorin B-Gens (im Vergleich zum Glycophorin A).²⁶

B.2.2.3 Veränderung der Genstruktur durch Insertionen von Retroposons

Für eine Veränderung der Genstruktur durch Insertionen von beweglichen DNA-Elementen (zumeist Retroposons in Eukaryoten) in fernerer Vergangenheit oder auch bei neu aufgetretenen Mutationen wurden mittlerweile eine große Anzahl von Beispielen gefunden. Dabei lassen sich zwei Arten von veränderter Genstruktur unterscheiden:

Zum einen wird ein veränderter Promotor über zusätzliche, durch die Insertion vermittelte (teilweise gewebsspezifische) Bindungsstellen von Transkriptionsfaktoren und somit auch ein eventuell veränderter 5'-Bereich der mRNA generiert und zum anderen kann ein verändertes Spleißen der prä-mRNA hervorgerufen werden. Einige repetitive Elemente können kryptische Spleiß- und poly(A)-Signale enthalten, die nach einer Insertion verwendet werden und zu einer veränderten mRNA führen. Beispiele für einen veränderten Promotor sind der Amylase-Gencluster des Menschen und das Sex-linked Protein der Maus.²⁷ Poly(A)-Signale werden von unterschiedlichen Retroposons im menschlichen Thymidilatsynthase-Gen²⁸ und im MHC-Klasse 1-Gen der Maus geliefert.²⁹

Ein verändertes Spleißmuster ist zum Beispiel bei Genen der Maus bekannt für das C4-Komplement-Gen,³⁰ für das β -Glycin-Rezeptor-Gen³¹ und für das CIC-1-Gen.³² Daneben können Retroposons auch Enhancer- und Silencer-Bindungsstellen inserieren, allerdings wirken sich diese nur auf ein verändertes Expressionsverhalten des in der Nähe kodierten Proteins aus und nicht auf die eigentliche Genstruktur.

B.2.2.4 Veränderung der Genstruktur über Rekombination

Die Genstruktur kann von mobilen DNA-Elementen nicht nur durch Insertionen beeinflusst werden, sondern indirekt über Rekombination gleicher Elemente an unterschiedlichen Positionen in der Nähe des gleichen Gens. Dies konnte u. a. anhand des Glycophorin-Genclusters gezeigt werden.³³ In diesem Beispiel wird der folgende Mechanismus vorgeschlagen. Nach erfolgter Genduplikation (durch mehrfache Rekombination zwischen zwei *Alu*-Elementen des potentiellen singulären Vorläufers GPA) kommt es zu Rekombinationen zwischen den so entstandenen drei Genen.

²⁶ Kudo & Fukuda, 1989

²⁷ Robins *et al.*, 1992

²⁸ Britten, 1997

²⁹ Kress *et al.*, 1984

³⁰ Pattanakitsakul *et al.*, 1992

³¹ Mülhardt *et al.*, 1994

³² Steinmeyer *et al.*, 1991

³³ Huang *et al.*, 1995

Von den zwei evolutiv gesehen jüngeren Genen (GPB, GPE) werden mRNAs kodiert, die nicht das Exon VI und VII enthalten und in denen einige Exons durch Punktmutationen nicht erkannt werden. Falls ein Fusionsgen aus GPA und GPB bzw. GPA und GPE durch Rekombination gebildet wird, kann durch Übertragung von intakten oder defekten Spleißstellen eine neue Genstruktur entstehen.

Ein Beispiel für eine direkte Veränderung der Genstruktur durch Rekombination von repetitiven Elementen ist für das Antithrombingen bekannt. Typ I Antithrombindefizienz kann durch eine homologe Rekombination zwischen zwei *Alu*-Elementen verursacht sein. In diesem Fall ist ein 2,8 kB großes Fragment des Gens deletiert, welches zwischen den beteiligten *Alu*-Elementen liegt und das Exon 5 enthält.³⁴

B.3 Serpine

B.3.1 Übersicht über die Serpine

Die Serpin-Superfamilie entwickelte sich über einen Zeitraum von 500 Millionen Jahren.^{35,36} Aufgrund der häufigen Funktion als *Serin-Proteinase-Inhibitor* wurde den Proteinen dieser Name verliehen.³⁷ Über 100 dieser Proteine konnten bisher in Viren und Eukaryoten, nicht aber in Prokaryoten identifiziert werden.^{36,38} Sie bestehen aus etwa 400 Aminosäuren mit einem Molekulargewicht von 38-70 Kilodalton (kD). Die Identität in der Aminosäure-Sequenz zwischen den Serpinen beträgt manchmal nur etwa 30 %, ³⁹ doch ihre Tertiärstruktur ist hoch konserviert. Funktionell lassen sich Serin-Proteinase-Inhibitoren, Cystein-Proteinase-Inhibitoren⁴⁰ und nicht inhibierende Proteine⁴¹ unterscheiden. In Tabelle 2 sind exemplarisch einige Serpine und ihre Funktion aufgeführt.

Tabelle 2: Inhibierende und Nichtinhibierende Serpine⁴²

Inhibierende Serpine	Ziel-Proteinase	Funktion der inhibierten Proteinase
α_1 -Antitrypsin	Elastase	Proteolyse von Elastin
α_1 -Antichymotrypsin	Cathepsin G	Entzündung
α_2 -Antiplasmin	Plasmin	Fibrinolyse
Antithrombin	Thrombin	Blutgerinnung
C1-Inhibitor	C1, Kallikrein, Faktor XIIa	Komplement-Aktivierung
Heparinkofaktor II (HCII)	Thrombin	Blutgerinnung, Fibrinolyse, Mitogenese
Plasminogenaktivator-Inhibitor-1 (PAI-1)	Gewebsplasminogen-aktivator (tPA)	Fibrinolyse
Nicht inhibierende Serpine	Funktion des Serpins	
Angiotensinogen	Peptid-Hormon-Vorläufer, Blutdruck-Regulation	
Corticosteroid-bindendes Globulin	Hormon-Transport	
Ovalbumin	Unbekannt	
Thyroxin-bindendes Globulin	Hormon-Transport	

³⁴ Olds *et al.*, 1993

³⁵ Hunt & Dayhoff, 1980; Gils & Declerck, 1998

³⁶ Bartuski *et al.*, 1997

³⁷ Carrell & Travis, 1985

³⁸ Marshall, 1993; Mc Carthy & Warrel, 1997

³⁹ Huber & Carrell, 1989

⁴⁰ Takeda *et al.*, 1995

⁴¹ Wright, 1984

⁴² Krepulat, 1999; Potempa *et al.*, 1994

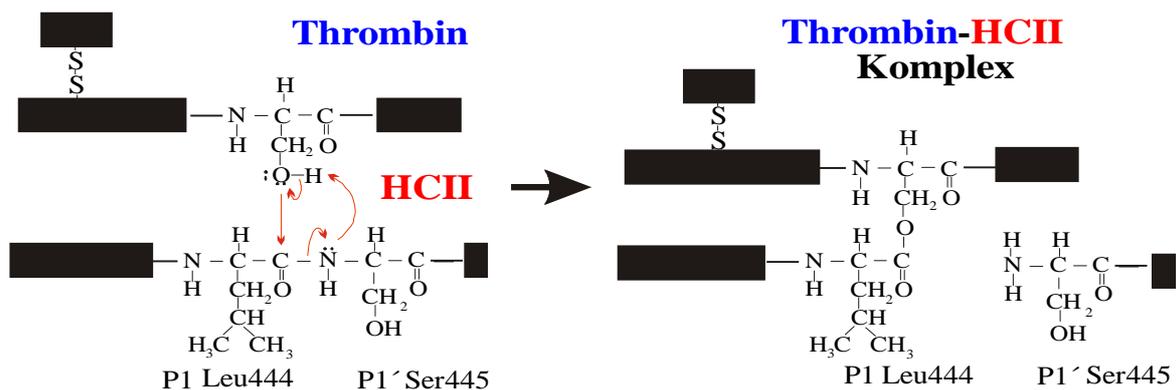
Einen weiteren Hinweis auf den Verwandtschaftsgrad von unterschiedlichen Serpinen einer Art oder aber gleicher Serpine aus unterschiedlichen Arten liefert ihre jeweilige Genstruktur (siehe B.3.2).

Das Schema der Proteinase-Hemmung durch Serpine verläuft für die meisten Serpine einheitlich. Nachdem in einem ersten Schritt ein Michaelis-Komplex ($E \cdot I$) zwischen Inhibitor (I) und Enzym (E) gebildet wird, kommt es anschließend zur Ausbildung eines kinetisch stabilen tetrahedralen Komplexes ($E \equiv I$).⁴³ Die relativen Mengen der einzelnen Komponenten hängen von den Bedingungen für die Komplexbildung, den speziellen Reaktionsbedingungen, der Struktur der sogenannten Serpin-Schleifenregion und der Ziel-Proteinase ab. Die Bildung des Komplexes läßt sich wie folgt schematisch darstellen:



Die exponierte Schleife des reaktiven Zentrums enthält die als P1-P1' bezeichneten Aminosäure-Positionen, deren Peptidbindung durch die Ziel-Proteinase angegriffen und gespalten wird. In Abbildung 5 ist als Beispiel ein Modell der Inaktivierung von Thrombin durch HCII dargestellt. Die Aminosäure an Position P1 hat einen entscheidenden Einfluß darauf, welche Proteinase inhibiert werden kann.

Abbildung 5: Modell der Inaktivierung von Thrombin durch HCII⁴⁴



Die Serin-Hydroxyl-Gruppe im katalytischen Zentrum von Thrombin greift das Carbonyl-Kohlenstoffatom der P1-Einheit im HCII an. Nach Spaltung der Peptidbindung wird eine neue Ester-Bindung mit dem Serinrest vom Thrombin geschlossen.⁴²

B.3.2 Genstruktur der Serpine

Die Anzahl und die Position der Introns ist in den Serpingenen variabel. Aus diesem Grund lassen die Serpine sich nicht nur durch ihre Proteinstruktur und -funktion, sondern auch nach ihrer Genstruktur einteilen. Die Exon/Introngrenzen der sieben möglichen Untergruppen stimmen mit der GT/AG-Regel überein.^{45,46} In der folgenden Tabelle ist eine Übersicht über die Serpingenfamilien gegeben, und im Anschluß an die Tabelle werden Mitglieder der Genfamilien exemplarisch beschrieben.

⁴³ Potempa *et al.*, 1994

⁴⁴ Tollefsen, 1988

⁴⁵ Breathnach & Chambon, 1981

⁴⁶ Mount, 1982

Tabelle 3: Serpingen-Familien

Bezeichnung	Struktur	Beispiele	Zielenzym*	Funktion*
„Ovalbumin Related Serpin“-Familie (Ov-Serpine)	8 Exons, 7 Introns	Ovalbumin ^{47,79} PI-5 (Maspin) ³⁶ PI-8 (CAP2) ³⁶ PI-10 (Bomapin) ³⁶ Plasminogen-Aktivator-Inhibitor-2 (PAI-2) ⁴⁸ SCCA-1, SCCA-2 ^{36,49} Y Gen ⁵⁰	- unbekannt unbekannt unbekannt uPA, tPA unbekannt unbekannt	Unbekannt 5, 10 unbekannt unbekannt 1,5,12 unbekannt unbekannt
1. Subfamilie				
2. Subfamilie	7 Exons, 6 Introns	PI-2 (M/NEI, ELANH2) ^{65,66} PI-6 (placental t. inhibitor, CAP, p35, PTI) [SPI-3] ⁶⁵ PI-9 (CAP3, GrnB-Inh.) ^{64,65}	N. Elastase, Cat.G Thrombin unbekannt	8 14 unbekannt
α_1 -Antitrypsin-Familie	5 Exons, 4 Introns (bei PI-1: 7 Exons)	PI-1 (α_1 -Antitrypsin, ^{51,52} α_1 -Proteinase-Inhibitor) α_1 -Antitr.-related-P. (ATR) Corticosteroid-bindendes Globulin (CBG) ^{68,72} Heparinkofaktor II (Leuserpin-2) ⁵³ Protein C Inhibitor (Plasminogen-Aktivator-Inhibitor-3; PAI-3) ⁶⁷ Angiotensinogen (AGT) ⁵⁴ [rANG bei Ratte] α_1 -Antichymotrypsin ^{55,56} PI-4 (Kallistatin) ⁶⁹ Thyroxin-bindendes Globulin (TBG) ⁵⁶	Elastase unbekannt -, (Elastase) Thrombin Protein C, Kallikrein - Cat. G, Chymase Kallikrein -	1 unbekannt 6 14 14 4 1,2,8 unbekannt 6
1. Subfamilie	1. Intron >5 kB			
2. Subfamilie	1. Intron ca. 2 kB			
α_2 -Antiplasmin-Familie	10 Exons, 9 Introns	α_2 -Antiplasmin (PLI) ⁷³ C1-Inhibitor ⁷⁴	Plasmin C1 Esterase	12 13
Antithrombin-Familie	7 Exons, 6 Introns	Antithrombin (Antithrombin III) ^{57,75}	Thrombin	14
Maus J6-Familie	5/6 Exons, 4/5 Introns	J6 [HS47] ⁵⁸	-	11
<i>Manduca</i> -Serpine-Familie	10 Exons (12 alter. Varianten von Exon 9), 21 Introns (davon 12 altern.)	<i>Manduca sexta</i> -Serpine-1 (Alaserpin) ⁷⁷	Chymotrypsin, Elastase, Trypsin	unterschiedlich
PAI-1-Familie	9 Exons, 8 Introns	PI-7 (Nexin-1) [SPI-4] ⁵⁹ PI-12 (Axo., Neuroserpin) ⁶⁰ Plasminogen-Aktivator-Inhibitor-1 (PAI-1) ⁷⁸	Plasmin, Thrombin uPA, tPA uPA, tPA	1,9 9 ? 1,10,12,14
<i>Paz1</i> -Familie	2 Exons, 1 Intron	<i>PazI</i> -Gen (Barley Protein) ⁶¹	unbekannt	Unbekannt

Erläuterungen zu der Tabelle auf der nächsten Seite

⁴⁷ Heilig *et al.*, 1982; Woo *et al.*, 1981

⁴⁸ Ye *et al.*, 1989; Genbank-Zugriffskodierung: J04606 / M24651 - M24657

⁴⁹ Schneider *et al.*, 1995; Genbank-Zugriffskodierung: U19558 - U19568 / U19569 - U19576

⁵⁰ Heilig *et al.*, 1982

⁵¹ Long *et al.*, 1984

⁵² Morgan & Kalsheker, 1997

⁵³ Kamp & Ragg, 1999; Ragg, 1986; Tollefsen, 1988; Genbank-Zugriffskodierung: J05309

⁵⁴ Fukamizu *et al.*, 1990; Genbank-Zugriffskodierung: X15323 - X15327

⁵⁵ Poller *et al.*, 1993; Genbank-Zugriffskodierung: X68733 - X68737

⁵⁶ Hayashi *et al.*, 1993; Genbank-Zugriffskodierung: L13470

⁵⁷ Prochownik *et al.*, 1978

⁵⁸ Wang, 1992; Genbank-Zugriffskodierung: J05609

⁵⁹ McGrogan *et al.*, 1990; Voss *et al.*, 1996

⁶⁰ Berger *et al.*, 1998

⁶¹ Brand *et al.*, 1990

Erläuterungen zu der Tabelle 3 (vorhergehende Seite)

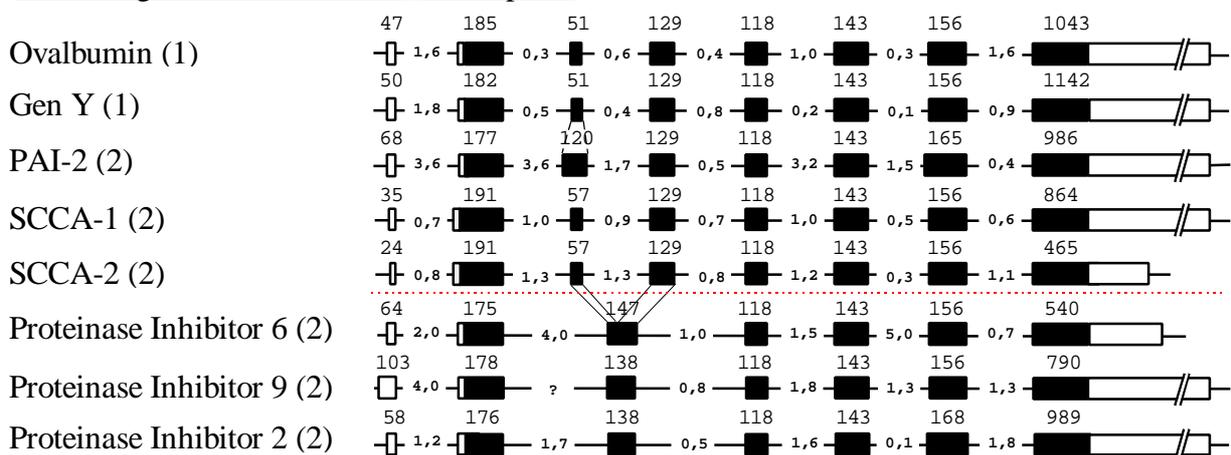
* Das angegebene Zielenzym (und somit auch die Funktion) bezieht sich nur auf die bekanntesten inhibierten Proteinasen. Die meisten Serpine haben eine Reihe von Zielenzymen. Die Angaben wurden von Potempa⁶² übernommen und modifiziert. Alternative Bezeichnungen für Gene wurden in runde Klammern gesetzt, während eckige Klammern für äquivalente Gene in anderen Arten verwendet wurden. Die Numerierung der unterschiedlichen funktionellen Bereiche wurde wie folgt durchgeführt: (1) Remodellierung der extrazellulären Matrix, (2) Umwandlung von Prohormonen, (3) Intrazellulärer Proteinabbau, (4) Kontrolle des Blutdrucks, (5) Tumor Supression, (6) Hormon Transport, (7) Virale oder parasitäre Pathogenität, (8) Modulation der Entzündungsreaktion, (9) Zelldifferenzierung, (10) Zellwanderung, (11) Proteinfaltung, (12) Fibrinolyse, (13) Komplementkaskade und (14) Blutgerinnung. Abkürzungen: Axo.: Axonin, CAP: cytoplasmic antiproteinase, Cat.G: Cathepsin G, GrnB-Inh.: Granzyme B Inhibitor, M/NEI: Monocyten/Neutrophile Elastase Inhibitor, N. Elastase: Neutrophile Elastase, Plasm.: Plasmin, SCCA: squamous cell carcinoma antigen, Thromb.: Thrombin, tPA: Gewebplasminogenaktivator, uPA: Urokinasetyp Plasminogenaktivator

Das pflanzliche *PazI*-Gen (3,1 kB) kodiert das im Endosperm der Gerste vorhandene Protein Z_4 , setzt sich aus zwei Exons und einem 334 Bp großem Intron zusammen und gehört somit in keine der anderen Serpin-Familien.⁶¹ Die Aminosäureidentität gegenüber anderen Serpinen liegt bei 25-30 %. Im folgenden werden die Mitglieder der übrigen Familien beschrieben.

Ovalbumin-Related Serpin-Familie („Ov-Serpine“)

Zu den „Ov-Serpinen“ gehören Serpingene, die aus acht Exons und sieben Introns bestehen.⁶³ Eine Reihe dieser Serpingene (PI-5/8/10) liegt beim Menschen in einem Cluster im Chromosomenbereich 18q21.3 vor.³⁶ Weitere Gene der ersten Subfamilie (PAI-2, SCCA-1, SCCA-2) konnten ebenfalls der Region 18q21-23 zugeordnet werden.

Den Serpinen PI-6, PI-9 und PI-2 (ELANH2) fehlt jeweils das dritte Intron der anderen Serpingene aus dieser Familie. Die Gene sind in einem Cluster auf dem menschlichen Chromosom 6p25 lokalisiert.^{64,65,66} Aufgrund dieser Unterschiede sind sie in Tabelle 3 als eine Subfamilie aufgeführt. In der folgenden Abbildung ist die Genstruktur von einigen Ov-Serpinen schematisch dargestellt.

Abbildung 6: Genstruktur der „Ov-serpine“

Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. Die unteren drei Gene werden in die zweite Subfamilie eingeteilt. Die Länge des 3'-gelegenen Exons ist bis zum Beginn der poly(A)-Sequenz angegeben. Die Zahlen in Klammern bedeuten: (1) Gene vom Huhn und (2) menschliche Gene.

⁶² Potempa *et al.*, 1994

⁶³ Coughlin *et al.*, 1993; Riewald *et al.*, 1995; Sprecher *et al.*, 1995; Zou *et al.*, 1994

⁶⁴ Eyre *et al.*, 1996; Genbank-Zugriffskodierung: U71364

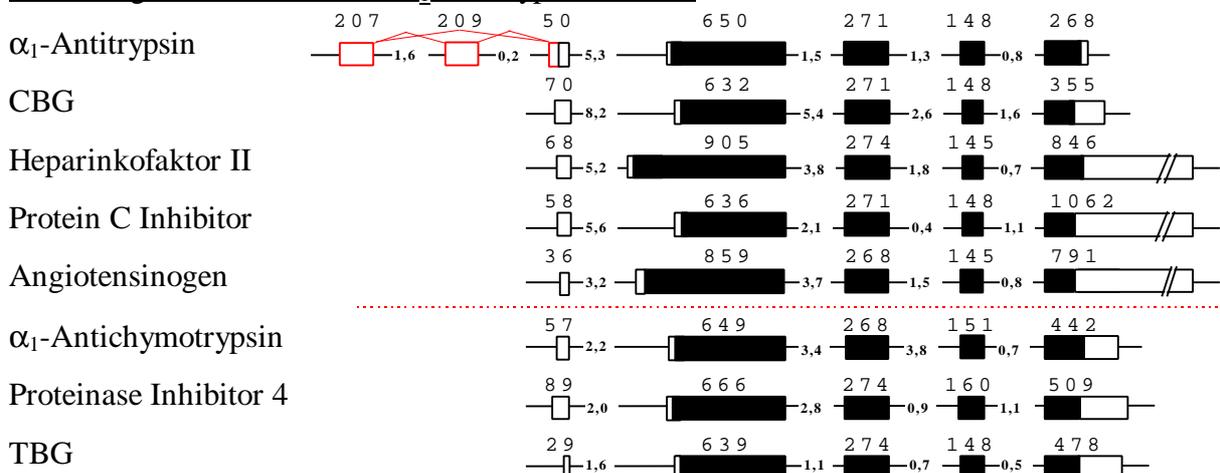
⁶⁵ Sun *et al.*, 1995; Sun *et al.*, 1998; Genbank-Zugriffskodierung: Z22658

⁶⁶ Evans *et al.*, 1995; Zeng *et al.*, 1998; Genbank-Zugriffskodierung: AF053630

α_1 -Antitrypsin-Familie

Die α_1 -Antitrypsin-Familie ist durch eine Kernstruktur von fünf Exons und vier Introns gekennzeichnet. Beim α_1 -Antitrypsin wurden zwei weitere Exons in Makrophagen gefunden (es werden drei verschiedene Transkripte über alternatives Spleißen der Exons 1A/1B/1C gebildet),⁵² so daß vermutet werden kann, daß auch bei anderen Mitgliedern dieser Familie zusätzliche Exons vorkommen. Diese Serpingene beinhalten ein kleines nicht kodierendes erstes Exon⁶⁷ und die drei Kodon-unterbrechenden Introns befinden sich exakt an den gleichen Positionen. Die Exons 3 und 5 des CBG Gens sind nur schwach konserviert. Eine abweichende Funktion dieses Serpins von der verbreiteten Proteinase-Hemmung könnte der Grund sein.⁶⁸ Bei den Mitgliedern dieser Genfamilie sind die Introns gegenüber den anderen Serpingenfamilien unterschiedlich positioniert. Eine Ausnahme bildet das zweite Intron, welches an der äquivalenten Position zum Intron 4 von Antithrombin und zum siebten Intron von α_2 -Antiplasmin liegt (siehe Abbildung 10).⁶⁵ Aufgrund der Länge des ersten Introns werden die Serpine dieser Familie in zwei Unterfamilien eingeteilt,⁶⁹ Serpine mit einem Intron 1 von ca. 2 kB Länge und solche mit über 5 kB Länge. Ein Vergleich der Serpin-Peptidsequenzen führt ebenfalls zu dieser Abgrenzung.⁷⁰ Der Abbildung 7 kann allerdings entnommen werden, daß (auf der Ebene der Genstruktur) diese Einteilung Gene zusammenfaßt, die sich nicht wesentlich von der jeweils anderen Subfamilie unterscheiden. Eine Analyse der chromosomalen Region 14q32.1 des Menschen zeigt, daß hier die Gene für α_1 -Antichymotrypsin, Protein C Inhibitor, Proteinase Inhibitor 4 (Kallistatin), α_1 -Antitrypsin und Corticosteroid-bindendes Globulin in einem Cluster vorliegen. In dem Cluster befindet sich ebenfalls das Gen für das α_1 -Antitrypsin-ähnliche Protein.⁷¹ Nach der Theorie der Entstehung neuer Gene über Genduplikation⁷² spricht dieses Gencluster ebenfalls gegen eine Einteilung in die beschriebenen Subfamilien. In der folgenden Abbildung ist die Genstruktur der Mitglieder der α_1 -Antitrypsin-Familie dargestellt.

Abbildung 7: Genstruktur der α_1 -Antitrypsin-Familie



Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. Die unteren vier Gene werden einer eigenen Subfamilie zugeordnet. Die Länge des 3'-gelegenen Exons ist bis zum Beginn der poly(A)-Sequenz angegeben. Die ersten beiden Exons (rot) des α_1 -Antitrypsingens werden alternativ gespleißt und kommen nur in einem Teil der mRNAs vor.

⁶⁷ Hayashi & Suzuki, 1993; Genbank-Zugriffskodierung: M64880 - M64884 / M68516 / S69366

⁶⁸ Hammond *et al.*, 1987; Underhill & Hammond, 1989; Genbank-Zugriffskodierung: J02943

⁶⁹ Chai *et al.*, 1994

⁷⁰ Marshall, 1993

⁷¹ Rollini & Fournier, 1997

⁷² Seralini *et al.*, 1990

α_2 -Antiplasmin-Familie

Zu dieser Gruppe gehören Serpingene, die aus zehn Exons und neun Introns bestehen. Bisher werden die Gene für α_2 -Antiplasmin und den C1-Inhibitor (obwohl bei diesem Gen bisher nur 8 Exons gefunden sind) zu dieser Familie gerechnet. Bei den beiden Genen sind fünf Intronpositionen identisch.⁷³ Beim Vergleich mit anderen Mitgliedern der Serpingen-Superfamilie stimmt die Position des siebten Introns (von α_2 -Antiplasmin bzw. des fünften Introns des C1-Inhibitorgens) mit der Position des zweiten Introns der Gene für Angiotensinogen der Ratte, des humanen α_1 -Antitrypsin und mit der Position des vierten Introns (mit einer Verschiebung um ein Nukleotid) von Antithrombin überein (siehe Abbildung 10). Dies deutet darauf hin, daß sich α_2 -Antiplasmin von einem gemeinsamen Vorläufer dieser Serin Protease Inhibitoren ableitet.^{73,74} In der folgenden Abbildung ist die Genstruktur dargestellt.

Abbildung 8: Genstruktur der α_2 -Antiplasmin-Familie



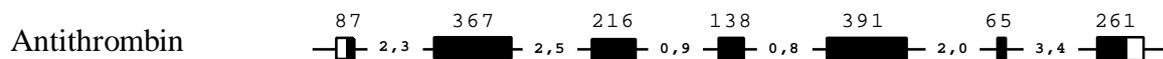
Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. In der Abbildung sind ähnliche Sequenzbereiche untereinander dargestellt. Nur die Introns 5-9 vom α_2 -Antiplasmin-Gen haben identische Position zu den Introns 3-7 des C1-Inhibitor-Gens. Die Länge des 3'-gelegenen Exons ist bis zum Beginn der poly(A)-Sequenz angegeben. Die Angaben beziehen sich auf die menschlichen Gene.

Antithrombin-Familie

Das Antithrombingen besteht aus sieben Exons und sechs Introns. Bisher kann nur das Antithrombingen dieser Struktur zugeordnet werden.⁷⁵

Das zweite und das vierte Intron befindet sich an äquivalenten Positionen anderer Serpine. Das zweite Intron liegt an der gleichen Position wie das zweite Intron des Ovalbumingens und das vierte Intron befindet sich an ähnlicher Position zum Intron 2 des α_1 -Antitrypsins⁶⁵ (siehe Abbildung 10) und des J6-Gens.⁷⁶ Nachfolgend ist die Struktur des Antithrombingens schematisch dargestellt.

Abbildung 9: Antithrombin Genstruktur



Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. Die Länge des 3'-gelegenen Exons ist bis zum Beginn der poly(A)-Sequenz angegeben. Die Angaben beziehen sich auf das menschliche Gen.

⁷³ Hirosawa *et al.*, 1988; Genbank-Zugriffskodierung: D00174 / J03830

⁷⁴ Carter *et al.*, 1991; Genbank-Zugriffskodierung: X54486

⁷⁵ Bock *et al.*, 1982; Bock *et al.*, 1988; Olds *et al.*, 1991; Prochownik *et al.*, 1985; Genbank-Kodierung: X68793

Abbildung 10: Vergleich von Exon/Introngrenzen an homologen Proteinpositionen

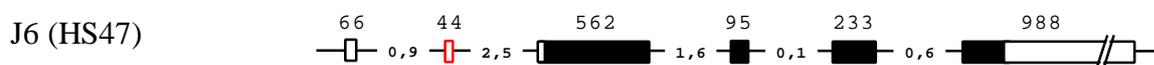
	Exon	Intron	Exon
Ovalbumin	CAA ATA AAT AAG	gtgagcctac...tcaattacag	GTT GTT CGC TTT
Exon 2	Gln Ile Asn Lys		Val Val Arg Phe ₆₀
Antithrombin	CAA CTG ATG GAG	gtacgaccaa...acttctatag	GTA TTT AAG TTT
Exon 2	Gln Leu Met Glu		Val Phe Lys Phe ₁₀₈
α_1 -Antitrypsin	ATC TTC TTT AAA G	gtaagggtt...cctctctccag	GC AAA TGG GAG
Exon 2	Ile Phe Phe Lys G		ly Lys Trp Glu ₉₅
α_2 -Antiplasmin	GCC ATC CAC TTC G	gtgcgctc...ctgggtttcag	GT TTC TGG AGG
Exon 7	Ala Ile His Phe G		ly Phe Trp Arg ₁₀₃
Antithrombin	CTG GAC TTC AAG	gtgagttgca...ttctcctcag	GGC CTG TGG AAG
Exon 4	Leu Asp Phe Lys		Gly Leu Trp Lys ₂₂₆

In der Abbildung sind zwei Exon/Introngrenzen von Mitgliedern unterschiedlicher Serpinfamilien vergleichend dargestellt. Die gezeigten Exonbereiche weisen bzgl. ihrer Aminosäuresequenz Ähnlichkeiten auf. Somit existiert zum einen ein Intron an einer vergleichbaren Positionen nach dem Exon 2 des Ovalbumins und Antithrombins und zum anderen ein Intron an nahezu gleicher Position beim Antitrypsin, Antiplasmin und beim Antithrombin.^{65,73,76} Das Antithrombin-Intron 4 ist um 1 Nukleotid gegenüber den beiden anderen Serpinen verschoben. Die angegebenen Sequenzen stammen vom Menschen, mit Ausnahme von Ovalbumin (Huhn).

Maus J6-Familie

Zu dieser Familie gehören Serpingene, die aus fünf Exons, einem optionalen Exon 1A und vier bzw. fünf Introns bestehen. Bisher kann nur das J6-Gen der Maus dieser Gruppe zugeordnet werden. Die Positionen der Introns sind mit Ausnahme von Intron 2 nicht identisch im Vergleich zu anderen Serpinen.⁷⁶ In Abbildung 11 ist die Genstruktur dargestellt.

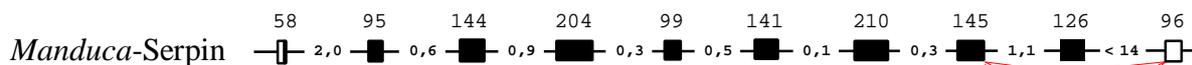
Abbildung 11: Colligin-2 (J6) Genstruktur



Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. Die Länge des 3'-gelegenen Exons ist bis zum Beginn der poly(A)-Sequenz angegeben. Das rot markierte Exon 1A kommt nur in einem Teil der Transkripte vor.

Manduca-Serpin-Familie

Zu der *Manduca*-Serpin-Familie gehören Serpingene, die aus zehn Exons und neun Introns bestehen. Bisher kann nur das Serpin von *Manduca sexta* dieser Gruppe zugeordnet werden. Exon 9, welches für das aktive Zentrum kodiert, kommt dabei in 11 Varianten vor. Alternatives Spleißen der prä-mRNA erzeugt eine Inhibitordiversität, über die eine Vielzahl von Proteinase reguliert werden kann. Jede Variante von Exon 9 hat verschiedene P1-Reste.⁷⁷ Im Anschluß ist die Genstruktur dieses Serpins gezeigt.

Abbildung 12: *Manduca*-Serpin Genstruktur

Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. Die Exon 9-Varianten werden alternativ gespleißt (rot markiert).

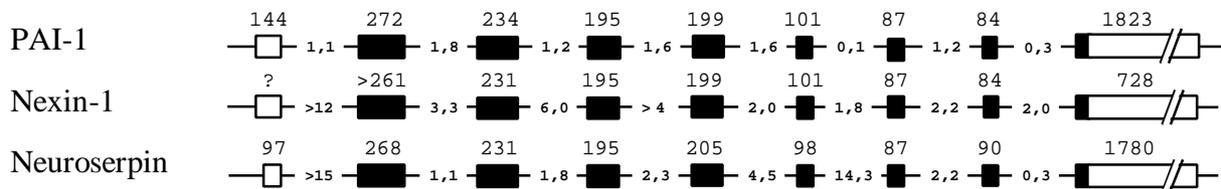
⁷⁶ Wang, 1992

⁷⁷ Jiang *et al.*, 1996; Genbank-Kodierung: U58361

PAI-1-Familie

Zu der PAI-1-Familie gehören Serpingene, die aus neun Exons und acht Introns bestehen.⁷⁸ Die acht Exon/Introngrenzen von PAI-1 sind konform zur GT/AG-Regel.⁷⁹ Zu dieser Familie gehören ebenfalls die Gene für Nexin-1 (PI-7) und für Neuroserpin (PI-12). Diese Proteine werden im Nervensystem exprimiert. Im Anschluß ist die Genstruktur der PAI-1-Familie schematisch dargestellt.

Abbildung 13: Genstruktur der PAI-1-Familie



Exons sind als Balken dargestellt. Schwarze Balken zeigen den translatierten Bereich. Große Zahlen über den Balken geben die Exongröße in [Bp] an, während kleine Zahlen zwischen den Exons die Introngröße in [kB] bezeichnen. Die Länge des 3'-gelegenen Exons ist bis zum Beginn der poly(A)-Sequenz angegeben. Die Angaben beziehen sich auf die menschlichen Gene mit Ausnahme vom Neuroserpin (Maus). Die Exongrößen des Nexingens wurden nach Angaben abgeschätzt, die auf Aminosäurepositionen beruhen.

B.4 Heparinkofaktor II

B.4.1 Das Heparinkofaktor II-Protein

Heparinkofaktor II (HCII) wird auch als Leuserpin-2 bezeichnet⁸⁰ und gehört zu der α_1 -Antitrypsin-Familie der Serpine⁸¹ (Tabelle 3, Abbildung 7). HCII ist ein effektiver Inhibitor der Serin-Proteinasen Thrombin und Chymotrypsin. Es ist wahrscheinlich, daß HCII weitere Aktivitäten besitzt, da bisher nur wenige Patienten mit einer totalen HCII-Defizienz bekannt sind. Thrombin selbst ist durch Aktivierung verschiedener Faktoren (Faktor V, VIII, XI und XIII) ein Schlüsselenzym der Blutgerinnungskaskade, während Chymotrypsin eine Endopeptidase im Dünndarm der Wirbeltiere ist.

Das Molekulargewicht vom glykosylierten HCII des Menschen wurde anhand der Sedimentationsgeschwindigkeit mit 65,6 kD⁸² oder mit Hilfe von SDS-Polyacrylamidgelelektrophoresen zu 66-78 kD (je nach Elektrophoresebedingungen⁸³) bestimmt. Die Polypeptidkette besteht aus 480 Aminosäuren mit einem vorausgehenden aminoterminalen Signalpeptid von 19 Aminosäuren.^{80,84} Das Protein umfaßt drei potentielle N-Glykosylierungsstellen und drei Cysteinreste, die jedoch keine Disulfidbrücken bilden. Die Gesamtmasse besteht zu etwa 10 % aus Kohlehydraten.^{85,86}

⁷⁸ Bosma *et al.*, 1988; Loskutoff *et al.*, 1987; Genbank-Zugriffskodierung: X13338 - X13345

⁷⁹ Breathnach *et al.*, 1978; Breathnach & Chambon, 1981

⁸⁰ Ragg, 1986

⁸¹ Tollefsen, 1995

⁸² Tollefsen *et al.*, 1982

⁸³ Pratt *et al.*, 1989

⁸⁴ Blinder *et al.*, 1988

⁸⁵ Kim, 1988

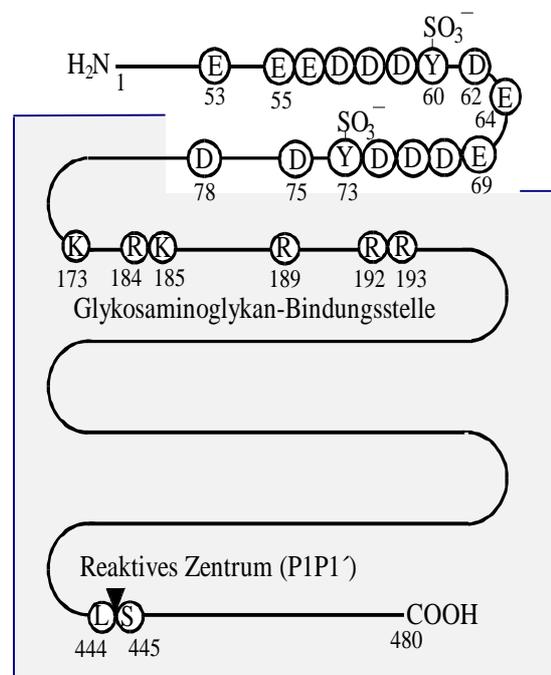
⁸⁶ Zhang *et al.*, 1994

Das menschliche HCII-Protein (hHCII) zeigt zu den Serpinen Antithrombin, α_1 -Antitrypsin und α_1 -Antichymotrypsin eine Aminosäureidentität von mindestens 30 %, ⁸⁷ wobei die größte Ähnlichkeit im carboxyterminalen Bereich zu finden ist. Das reaktive Zentrum (Positionen Leu444–Ser445) liegt in dieser Region. Die Glykosaminoglykan-Bindungsstelle befindet sich in dem zentralen Bereich des Proteins (Abbildung 14, schattierter Bereich). ⁸⁹ Die 80 aminoterminalen Aminosäuren des HCII zeigen Ähnlichkeit zu der carboxyterminalen Domäne von Hirudin, ⁹⁰ jedoch keine Homologie zu anderen Serpinen. Aus diesem Grund war es bisher nicht möglich, einen Vorschlag für die gesamte 3D-Struktur von HCII anhand von anderen Serpinen zu machen.

Der Thrombininhibitor Hirudin des Parasiten *Hirudo medicinalis* (Blutegel) bindet wahrscheinlich an Thrombin ähnlich wie HCII über einen sauren Proteinbereich. ⁹⁰ Der C-Terminus zeigt strukturelle Ähnlichkeit zu den sauren Domänen von HCII und bindet ebenfalls an die Exosite I von Thrombin (Abbildung 15). Die P1-Position des reaktiven Zentrums scheint normalerweise bei Serpinen die Spezifität für die Zielproteinase zu bestimmen. Ein Leucinrest an Position P1 von HCII läßt eine Inhibierung chymotrypsinähnlicher Enzyme und nicht vom Thrombin erwarten. Die aminoterminalen Region im HCII (Abbildung 15) enthält jedoch zwei aufeinanderfolgende saure Sequenzen (AS 53-65 und 66-78 im hHCII), die die Inhibierung von Thrombin bei Anwesenheit von Glykosaminoglykanen ⁹⁰ ermöglichen (Abbildung 15). Die Thrombin-Inhibierungsrate wird durch Glykosaminoglykane, wie z. B. Heparin oder Dermatansulfat, um etwa das 1000fache ⁸³ erhöht, dagegen haben sie keinen Effekt auf die Chymotrypsin-Inhibierung. Auch andere sulfatierte Polyanionen, wie Pentosan Polysulfat, fukosyliertes Chondroitinsulfat, Chondroitinsulfat E oder Dextransulfat können die HCII-Aktivität erhöhen. ⁸⁸

HCII-Defizienz wird beim Menschen autosomal dominant vererbt. ³⁵ Ein Mangel an HCII wurde bei Patienten mit Thrombose beobachtet, obwohl normale Antithrombin-Konzentrationen vorlagen. Aus diesen Daten folgt, daß HCII möglicherweise eine physiologische Bedeutung in der Thromboseprävention besitzt. ⁹¹ Andererseits wird berichtet, daß eine homozygote HCII-Defizienz nur zu einem geringen erhöhten Thromboserisiko führt, wenn nicht parallel eine Antithrombin-Defizienz vorliegt. ⁹²

Abbildung 14: Schematisches HCII-Modell



Die graue Fläche des Modells ^{89,90} markiert den Bereich, der Homologie mit anderen Serpinen aufweist.

⁸⁷ Tollefsen, 1988

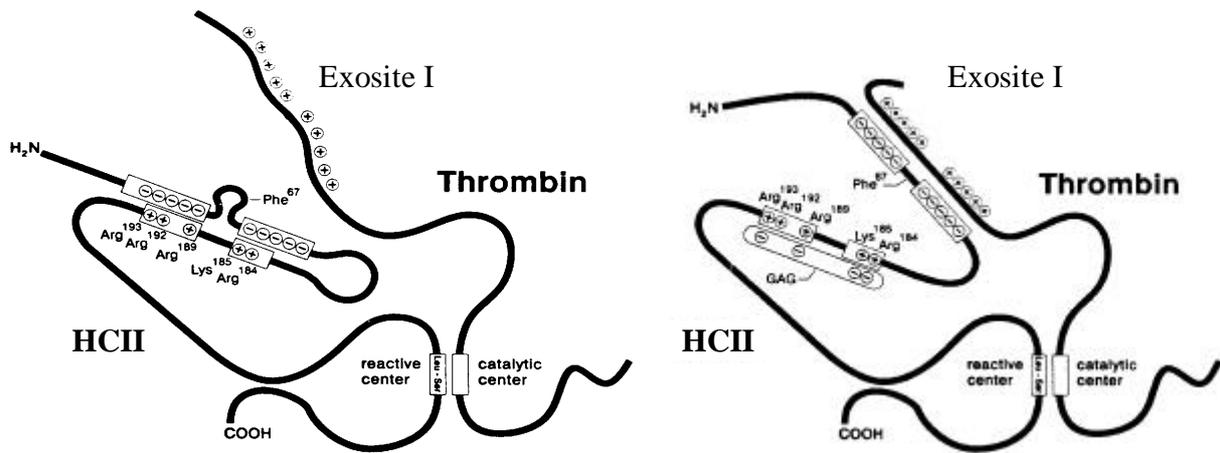
⁸⁸ Colwell *et al.*, 1999

⁸⁹ Tollefsen, 1997

⁹⁰ Ragg *et al.*, 1990

⁹¹ Simioni *et al.*, 1990; Tran *et al.*, 1985

⁹² Villa *et al.*, 1999

Abbildung 15: Inhibierung von Thrombin durch HCII (Modell)⁹⁰

In Abwesenheit von Glykosaminoglykanen (GAG) ist die saure Domäne von HCII über ionische Bindungen intramolekular gebunden. Eine Interaktion mit Thrombin ist nur über das reaktive Zentrum möglich.

Nach GAG-Bindung kann die saure Domäne von HCII mit der Thrombin-Exosite I interagieren und den Komplex stabilisieren.

B.4.2 Die Struktur des Heparinkofaktor II-Gens

Das humane HCII-Gen (hHCII) besteht aus fünf Exons und vier Introns und befindet sich in einem über 14 kB großen Bereich innerhalb des Chromosomenbereichs 22q11.^{90,93,94} Die bekannte Exon/Intron-Struktur von HCII stimmt mit der Gen-Grundstruktur der α_1 -Antitrypsin-Familie überein (Abbildung 7). Im folgenden sind die bekannten Daten zum HCII-Gen von Mensch, Maus und Ratte aufgeführt.

Das menschliche HCII-Gen (*Homo sapiens* HCII, hHCII):

Der chromosomale Bereich, der in Leberzellen in eine hHCII-prä-mRNA transkribiert wird, umfaßt beim Menschen etwa 13,7 kB.⁹⁴ Die cDNA (2,3 kB) wurde synthetisiert, indem humane Hepatozyten-RNA als Matrize verwendet wurde.^{80,84,94} Die mRNA von hHCII konnte bisher in Leber mRNA und Lungen mRNA nachgewiesen werden.⁹⁵

Die in einer Lungen-cDNA-Bank gefundene HCII-cDNA ist am 5'-Ende des zweiten Exons verlängert (Exon 1*). Der genaue Start des mRNA-Transkripts ist unbekannt, allerdings scheint das aus Leber bekannte Exon 1 nicht vorhanden zu sein. Trotz des unterschiedlichen 5'-Endes ist der kodierende Bereich der mRNAs aus Lunge und Leber identisch.⁹⁵

Das HCII-Gen der Maus (*Mus musculus* HCII, mHCII)

Aus Leber-cDNA-Banken der Maus wurde eine mHCII-cDNA mit der Länge von 2,0 kB gewonnen und mit Northernblots eine 2,3 kB große mRNA identifiziert. Keine Transkripte konnten dagegen in Herz, Gehirn, Milz, Lunge, Skelettmuskel, Niere, Testis, Plazenta, Pankreas oder Darm gefunden werden.⁸⁶ Das Gen für HCII liegt auf dem Chromosom 16 der Maus zwischen den Genen *Prm-1* und *Igl* vor. Die ermittelte Gesamtlänge des mHCII-Gens beträgt 7,1 kB. Die publizierte Sequenz enthält jedoch keine Bereiche, die dem menschlichen Exon 1 entsprechen.⁸⁶ Das Protein ist bezüglich seiner Aminosäuresequenzen zu 83 % identisch zum menschlichen HCII.⁹⁶

⁹³ Kim *et al.*, 1996; Ragg & Preibisch, 1988

⁹⁴ Herzog *et al.*, 1991

⁹⁵ U. Schmidt, unveröffentlichte Daten

⁹⁶ Colwell & Tollefsen, 1998

Das HCII-Gen der Ratte (*Rattus norvegicus* HCII, rHCII)

Northernblots ergaben eine Länge der cDNA des rHCII von 2,1 kB. Mit RT-PCR wurden zwei Transkripte unterschiedlicher Länge gefunden.⁹⁷ Southernblottergebnisse zeigen, daß diese Transkripte Produkte ein- und desselben Gens sind. Eines der Transkripte enthält im 5'-Bereich eine kurze, zusätzliche Sequenz, die bei anderen HCII-mRNAs (so beim Mensch und auch bei der Maus) nicht gefunden wird und auf einem zusätzlichen, alternativ gespleißten Exon beruht.⁹⁷ Das Protein ist bezüglich seiner Aminosäuren zu 83 % identisch zum menschlichen HCII.⁹⁶

Das HCII-Gen des Kaninchens (rabHCII), des Frosches und des Huhns

HCII konnte ebenfalls über cDNAs im Kaninchen (2,2 kB⁹⁸), im Frosch (1,8 kB⁹⁶) und im Huhn (1,7 kB⁹⁶) nachgewiesen werden. Die Aminosäureähnlichkeit im Vergleich zum Mensch liegt beim Kaninchen bei 85 %, beim Frosch bei 59 % und beim Huhn bei 65 %.⁹⁶ Eine vergleichende Analyse der einzelnen mRNA-Sequenzen wurde bisher nicht publiziert.

B.4.3 Struktur des menschlichen HCII-Promotors

Die regulatorischen Elemente zur Kontrolle der Expression des HCII-Gens sind weitgehend unbekannt, und es fehlen Kenntnisse über eine gewebsspezifische Kontrolle, wie auch über veränderte Transkriptionslevel durch Hormone oder exogene Faktoren. Nach heutigen Kenntnissen sind Promotoren aus proximalen und distalen DNA-Elementen bzw. regulativen Sequenzen zusammengesetzt, an die spezifische Transkriptionsfaktoren binden können.¹⁵ Es sind mittlerweile viele Bindungsstellen und die daran bindenden Faktoren bekannt und in Datenbanken katalogisiert. In verschiedenen Geweben (sowie bei verändertem Hormonspiegel u. a.) können jeweils unterschiedliche Konzentrationen dieser Transkriptionsfaktoren vorhanden sein, die somit eine gewebsspezifische Expression von Proteinen ermöglichen. Die Faktoren können einen positiven oder negativen Einfluß auf die Transkription bewirken. Somit werden die distalen, d. h. vom Startbereich der mRNA weiter 5'-gelegenen Elemente in 'Silencer' und 'Enhancer' unterschieden.

Unmittelbar vor dem Startpunkt der mRNA liegt normalerweise die TATA-Box. Allerdings wurden auch Transkriptionsstartpunkte gefunden, die ohne dieses Signal auskommen und stattdessen ein anderes Element aufweisen (INR, Initiatorelement).⁹⁹ Normalerweise werden mRNAs in Eukaryoten von der RNA-Polymerase II synthetisiert. Es gibt jedoch auch Hinweise auf durch Retroposons vermittelte Promotoraktivität, die interne Signale für RNA-Polymerase III aufweisen können.¹⁰⁰ Auch bei RNA-Polymerase II-Transkripten können Sequenzen wichtig sein, die innerhalb des transkribierten Bereichs liegen. So wurde kürzlich bei dem humanen Angiotensinogen im Exon 1 eine HNF-3 Bindungsstelle gefunden, die offensichtlich die Transkription in Leberzellen deutlich erhöht. Demgegenüber scheint eine benachbarte AP4-Bindungsstelle die Transkription negativ beeinflussen zu können, indem sie die Bindung von transkriptionsfördernden Faktoren an die DNA erschwert. Insgesamt scheint bei diesem Gen ein Bereich von -32 bis +44 für eine leberspezifische Expression des Proteins ausreichend zu sein.¹⁰¹

Hinweise auf gewebsspezifische- und induzierte Expression und somit eventuell auch auf spezifische Transkription des hHCII-Gens geben die folgenden Erkenntnisse.

⁹⁷ Westrup & Ragg, 1994

⁹⁸ Sheffield *et al.*, 1994

⁹⁹ Smale, 1997

¹⁰⁰ Okada *et al.*, 1997

¹⁰¹ Cui *et al.*, 1998

Bei akuten Entzündungsreaktionen ist die HCII-Konzentration im Plasma durch einen bisher unbekanntem Mechanismus erhöht. Versuche zeigten allerdings, daß die HCII-Synthese in Hepatomazellen nicht durch inflammatorische Cytokine (Interleukin-1 β , Interleukin-6, Tumor Nekrosis Faktor- α) stimuliert wird.¹⁰² Andererseits wurde eine erniedrigte HCII-Aktivität im Plasma bei HIV-infizierten Menschen beobachtet.¹⁰³ Demgegenüber wird die Expression von α_1 -Antitrypsin aus der gleichen Serpin-Genfamilie als 'acute phase'-Protein in Leber unter inflammatorischen Bedingungen (IL6-Gabe) induziert.¹⁰⁴ In Geweben der Maus (Herz, Gehirn, Lunge, Milz, Niere, Testis) wurde darüber hinaus gezeigt, daß neben α_1 -Antitrypsin auch das Plasma Albumin, ebenfalls ein Serpin, in der „acute phase“ spezifisch induziert wird.¹⁰⁵ Somit läßt sich folgern, daß inflammatorische Bedingungen spezifisch die HCII-Konzentration gegenüber ähnlichen Serpinen im Organismus in verschiedenen Geweben beeinflussen. Da eine andere HCII-Konzentration auch durch einen veränderten Metabolismus verursacht sein kann, fehlt bisher ein direkter Hinweis auf eine veränderte Transkription.

Ein erhöhter HCII-Spiegel wird ebenfalls im Verlauf einer Schwangerschaft gemessen. Am Ende der Schwangerschaft, während und kurz nach der Geburt ist die Konzentration von Thrombin-HCII-Komplexen um etwa das 2-3fache erhöht.¹⁰⁶ Einige Tage später erreicht die HCII-Konzentration wieder das ursprüngliche Niveau. Im Gegensatz dazu sinkt während einer Präeklampsie (schwangerschaftsbedingter Bluthochdruck) der HCII-Spiegel ab. Eine vermehrte Thrombin-Komplex-Bildung mit HCII, durch ansteigende Thrombin-Konzentration, scheint unter anderem der Grund zu sein.¹⁰⁷ Aus diesen Befunden folgt, daß HCII wahrscheinlich eine spezielle Funktion in der Plazenta hat, allerdings wurde HCII(-cDNA) auch in weiteren Geweben und in Nicht-Plazentaliern nachgewiesen (Protein im Plasma und cDNA in einer Leber-Bank),⁹⁶ so daß HCII dort andere Funktionen zu übernehmen scheint.

In HepG2-Zellen, die mit Konstrukten transfiziert wurden, die den Ratten-Serum-Albumin-Promotor enthalten, lassen sich funktionelle Bindungsstellen für Hepatozyten Nuklearfaktor 3 nachweisen (HNF3).¹⁰⁸ HNF-Bindungsstellen sind vor leberspezifischen Genen angereichert, vor anderen Genen jedoch nicht ausgeschlossen.¹⁰⁹ Eine Überprüfung auf HNF-Bindungsstellen wurde bei dem hHCII bisher noch nicht durchgeführt. Dem gegenüber wurde bereits eine Analyse des HCII-Gens auf mögliche Transkriptionsstartpunkte mit Hilfe von S1-Nuklease- und RNaseA/T1-Kartierung von Ragg und Preibisch durchgeführt.¹¹⁰ Den Ergebnissen von beiden Methoden lassen sich mehrere Startpositionen der Transkripte entnehmen, sowie eine Exon 1-Länge von 25-115 Bp. Der potentielle, am weitesten stromaufwärts gelegenen Transkriptionsstartpunkt (entsprechend einer Länge des Exon 1 von etwa 115 Bp), wird nur mit der S1-Kartierung erhalten.¹¹⁰ Darüber hinaus liegen noch keine weiterreichenden Kenntnisse über die Transkriptionsstartbereiche des hHCII-Gens in Leber und anderen Geweben vor. Eine Suche allein nach TATA-Boxen reicht dabei nicht aus, da für schwach transkribierte mRNAs auch TATA-Box-lose Promotoren (INR-Elemente) bekannt sind.¹¹¹ Es war schon zu Beginn der Arbeit bekannt, daß sich das in Lungen-RNA gefundene hHCII-Transkript von dem bekannten Transkript aus Leberzellen unterscheidet und daß ein Startbereich innerhalb des ersten Introns in Frage kommt.

¹⁰² Koike *et al.*, 1996

¹⁰³ Toulon *et al.*, 1993

¹⁰⁴ Hafeez *et al.*, 1992

¹⁰⁵ Kalmovarin *et al.*, 1991

¹⁰⁶ Andersson *et al.*, 1996

¹⁰⁷ Bellart *et al.*, 1998

¹⁰⁸ Hsiang *et al.*, 1999

¹⁰⁹ Tronche *et al.*, 1997

¹¹⁰ Ragg & Preibisch, 1988

¹¹¹ Smale, 1997

B.5 Evolutive Aspekte

B.5.1 Entwicklung und Verwandtschaft der untersuchten Arten

In dieser Arbeit wird hauptsächlich die Struktur des HCII-Gens von Maus und Ratte verglichen. Zusätzlich werden Vergleiche mit dem menschlichen Gen durchgeführt und auch Daten von anderen Säugetieren und Nicht-Säugetieren mit einbezogen. Daher ist die verwandtschaftliche Beziehung zwischen diesen Arten von Interesse.

Während sich die Maus und die Ratte erst vor ca. 41 Millionen Jahren in Arten aufgetrennt haben, liegt die Aufspaltung der maus- und hamsterartigen Nagetiere etwa 65 Millionen Jahre zurück.¹¹² Die Stammbäume der Primaten und Lagomorpha ('Kaninchen-artige') trennten sich vor 91 Millionen Jahren. Viel früher, vor etwa 310 Millionen Jahre, lebten die letzten gemeinsamen Vorfahren von Säugetieren und Vögeln.¹¹² Bemerkenswert ist, daß die Lagomorpha, verwandtschaftlich gesehen, den Primaten näher stehen als den Nagetieren.¹¹³

Normalerweise sind die Genome der Arten sehr stabil. Zum Beispiel wurden bei einem Vergleich von Chromosomen des Menschen und von Teleostfischen (Fischen) 25 konservierte Chromosomenfragmente gefunden. Im Gegensatz zu anderen Tierklassen erscheinen aber die Genome der Säugetiere, und vor allem der Nagetiere, stark von Umlagerungen bzw. Rekombinationen und somit einer erhöhten Evolutionsrate betroffen zu sein.¹¹⁴

B.5.2 Evolution und HCII

Essentielle Enzyme und deren Gene sind normalerweise, evolutiv betrachtet, relativ stabil. Mutationen im translatierten Bereich dieser Gene bewirken häufig ein nicht mehr funktionelles Enzym, und Mutationen in den nicht translatierten Bereichen können eine veränderte Transkription oder ein verändertes Spleißen, also quantitativ gesehen, ebenfalls eine unterschiedliche Enzymaktivität, bedeuten. Obwohl nicht bekannt ist, welche HCII-Funktion für den Organismus elementar wichtig ist, scheint es ein wichtiges Gen zu sein, da bisher nur wenige HCII-Merkmalsträger mit homozygoter Defizienz bekannt sind.¹¹⁵ Aus diesem Grund sind für dieses Gen (und den umgebenen Bereich) relativ wenig Veränderungen im Verlauf der Evolution gegenüber dem restlichen Genom zu erwarten.

Im Gegensatz dazu sind Marker ('Krankheits-Loci') in dem chromosomalen Bereich 22q11 und 22q12, in dem u. a. auch HCII liegt, überrepräsentiert. Auf dem menschlichen Chromosom 22 wurden zum Beispiel im Bereich 22q11 das DiGeorge-Syndrom, das Velo-cardial-facial-Syndrom, das 'cat-eye'-Syndrom und verschiedene Tumore kartiert. Der Grund könnte in einer Reihe von mittel- und hochrepetitiven Elementen bestehen, die hier vorliegen.¹¹⁶

Die in der Bande 22q12 vermehrt vorliegenden TG-Sequenz-Wiederholungen könnten ebenfalls der Grund für die verstärkte Mutationsrate sein.¹¹⁷ Eventuell wird auch der in etwa 2 MBp Entfernung kartierte BCR, einem Marker für eine Chromosomen-Bruchstelle,¹¹⁷ durch TG-Sequenzen hervorgerufen (aufgrund der Eigenschaft einzelsträngige und rekombinationsanfällige DNA-Bereiche zu generieren).

¹¹² Kumar & Hedges, 1998

¹¹³ Graur *et al.*, 1996

¹¹⁴ Workshop on Comparative Genome Organization, 1996

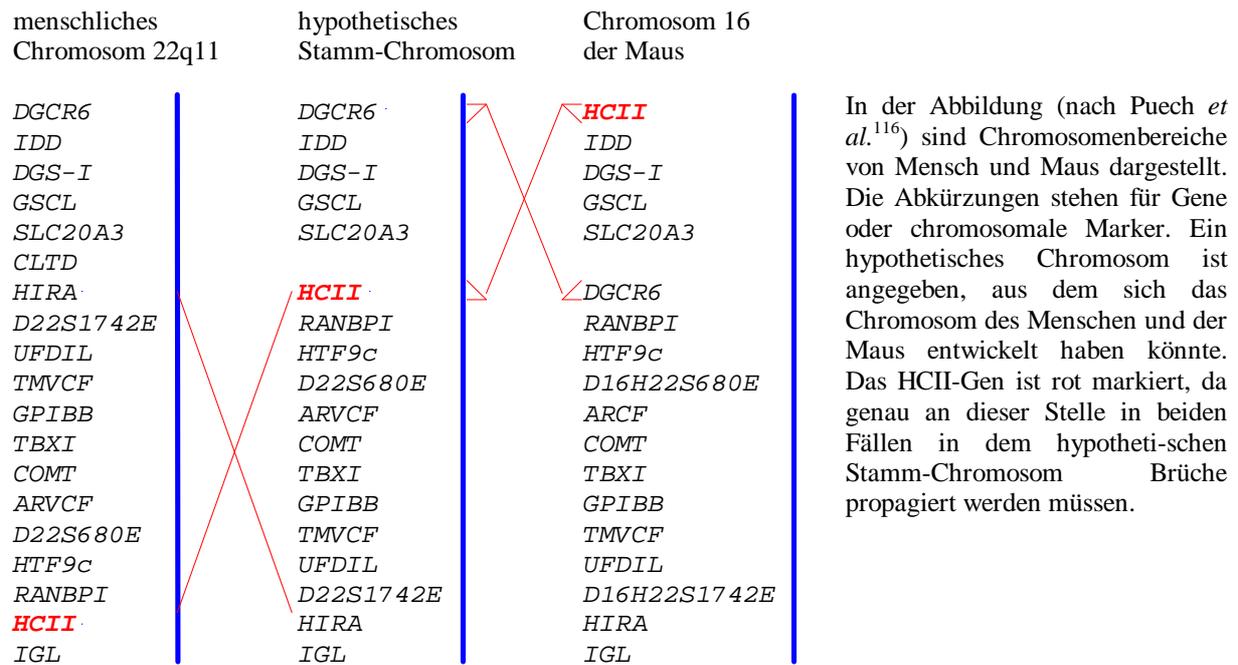
¹¹⁵ Villa *et al.*, 1999

¹¹⁶ Puech *et al.*, 1997

¹¹⁷ Budarf *et al.*, 1996

Einen weiteren Hinweis auf verstärkte evolutive Veränderungen in diesem Bereich geben Vergleiche der Lage homologer Gene auf Maus- und Mensch-Chromosomen. Ein Bereich, der beim Menschen auf Chromosom 22 liegt, verteilt sich in der Maus auf die Chromosomen 6, 10 und 16. Um die veränderte Reihenfolge der Gene um HCII zwischen Maus und Mensch (bei einem hypothetischen gemeinsamen Vorfahren) mit minimalen Mutationen erklären zu können, wurde ein Chromosomenbruch direkt bei dem HCII-Gen vorgeschlagen, mit einer Inversion beim Menschen zwischen den HCII- und HIRA-Genbereichen und direkten Austauschen der Gene von HCII und DGCR6 bei der Maus (siehe Abbildung 16).¹¹⁶

Abbildung 16: Vergleich chromosomaler Bereiche von Mensch und Maus



Im Vergleich zu den umgebenen Sequenzen (Genommutationen in der Nähe des HCII-Gens) ist das Gen dagegen höher konserviert. Im reaktiven Zentrum mit der Schleife von P6 bis P3' sind die Aminosäuren bei Frosch, Huhn und Säugern identisch.^{96,97} Innerhalb der Säuger zeigen auch die Thrombin-Bindungsstelle und die GAG-Bindungsstelle wenig Veränderungen.

Obwohl das Reaktionszentrum in den jeweiligen Serpinen hoch konserviert ist, scheint im Verlauf der Evolution ein positiver Selektionsdruck (bei Nagetieren) auf Genduplikation und Diversifikation der Reaktionszentren zu liegen, da durch das Reaktionszentrum auch die Substratspezifität (durch den Aminosäurerest in der P1 Position) bestimmt wird. So werden neue Enzyme kodiert, die andere Proteinase erkennen und inhibieren können.¹¹⁸

¹¹⁸ Hill *et al.*, 1984; Hill & Hastie, 1987

B.6 Zielsetzung der Arbeit

Die Exon/Intronstruktur vieler Vertebratengenfamilien ist im Blick auf Zahl und Position von Introns stark konserviert (die α - und β -Globingene, die β -Tubulingene und in eingeschränktem Maße auch das Insulingen sind Beispiele^{14,15}). Im Gegensatz dazu bestehen in der Serpingenfamilie hinsichtlich der Zahl und Positionen der Introns erhebliche Unterschiede (siehe B.3). Einige Beobachtungen deuteten daraufhin, daß sogar innerhalb der orthologen HCII-Gene (ein Serpingen) eventuell eine Variabilität der Genstruktur vorliegt:

- 1) Das menschliche HCII-Gen weist eine Struktur mit fünf Exons und vier Introns auf.⁹⁴
- 2) Bei der Analyse der HCII-mRNA aus Rattenleber wurden zwei cDNAs erhalten, die sich durch das Auftreten einer 57 Bp langen Sequenz in der 5'-nicht translatierten Region voneinander unterscheiden.⁹⁷
- 3) Nach Zhang *et al.* enthält das HCII-Gen der Maus nur vier Exons und drei Introns. Es unterscheidet sich damit von der Struktur des menschlichen Gens.⁸⁶

Es lagen also mehrere Beobachtungen vor, die eine Analyse der HCII-Genstruktur der Maus, der Ratte und des Menschen erforderten. Möglicherweise stellen die Ergebnisse ein Modell für Mechanismen dar, die zur Änderung der Organisation von Genen führen. Die Untersuchungen können dementsprechend einen Beitrag zu Kenntnissen über die Evolution von Genen liefern.

Zur Aufklärung der Strukturen sollten daher die 5'-Genbereiche der HCII-Gene von Maus und Ratte sequenziert und zusammen mit der entsprechenden menschlichen Genregion verglichen werden. Darüber hinaus mußte die Untersuchung der 5'-Regionen der HCII-mRNAs erfolgen.

Bisher sind nur wenige Informationen zur Regulation des HCII-Gens vorhanden. Ein zweiter Themenbereich beschäftigt sich dementsprechend mit der regulierten Transkription des Gens. Die Synthese von HCII findet hauptsächlich in der Leber statt. Sehr wenig ist dagegen über die Expression des HCII in anderen Geweben und über die Regulation des Gens bekannt.

Es sollten daher Untersuchungen zur gewebsspezifischen Expression und zu den regulatorischen Elementen des HCII-Gens durchgeführt werden. Dazu sollte u. a. der 5'-Bereich des Gens durch Bestimmung des Transkriptionsstartpunktes und mittels Reportergenkonstrukten analysiert werden.

C. Material und Methoden

C.1 Material

C.1.1 Geräte

Gerät	Hersteller
Heizblock, Thermostat 5320	Eppendorf
Heizblock, TB1	Biometra
Hybridisierungssofen OV2	Biometra
Luminometer CB 9506	EG & G Berthold
Mikrowelle	Siemens
Milli-Q®-Anlage	Millipore
Mikropipetten	Biozyme, Eppendorf, Gilson
Pipetten	Biozyme, Eppendorf, Gilson
1500 DNA Sequencer	GATC
Spectrophotometer UV-1602	Shimadzu
Thermocycler Progene	Thermo-Dux
Thermocycler DNA Thermal Cycler 480	Perkin Elmer
Thermostat PC /4 für Wasserbäder	Julabo
UV-Stratalinker 2400	Stratagene
Vortex Genie 2, G-560 E	Scientific Industries
Waagen:	
Analytische Waage A120S	Sartorius
Universal Waage U4800 P	Sartorius
Zentrifugen:	
Biofuge 28 RS	Heraeus
RC5C Sorvall-Zentrifuge, Rotoren: SS-34; GS-3	Sorvall Instruments, Du Pont
Tisch-Zentrifuge 5415 C	Eppendorf
Tisch-Zentrifuge Biofuge pico	Heraeus
Vakuum-Zentrifuge BA-VC-300H, beheizbar	Bachhofer

C.1.2 Kits

Tabelle 4: Verwendete Systeme

Kit	Kit-Nr.	Hersteller
AmpliAq PCR Kit	N 808-0193	Perkin Elmer
BioCycle Sequencing Kit	SR 100.30	GATC
C. therm. Polymerase 2 Schritt reverse Transkriptase Kit	2 016 311	Boehringer Mannheim
DIG-Nucleic-Acid-Detection Kit	1 175 041	Boehringer Mannheim
Dual-Luciferase™ Reporter Assay System	E 1910	Promega
PCR DIG Probe Synthesis Kit	1 636 090	Boehringer Mannheim
Omniscript reverse Transkriptase-Kit	205111	Qiagen
QIAquick PCR Purification Kit	28104	Qiagen
RNeasy Mini Kit	74104	Qiagen
reverse Transkriptase-PCR Kit	N 808-0017	Perkin Elmer

C.1.3 Dokumentation

Dokumentationssystem	Hersteller
Agarosegele	
Transilluminator (312 nm)	Biometra
CCD-Videokamera	Biometra
Photo Finish Zsoft 3.0 Software	WordStar Atlanta Technology Center®
Southern Blot	
BioMax Light-1 Filme	Kodak
Entwickler D-9	Kodak
Fixierer Unifix	Kodak

C.1.4 Chemikalien und Verbrauchsmaterialien

Chemikalien und Verbrauchsmaterialien	Hersteller
Agarose zur Elektrophorese, Typ 15510-027	Life Technologies
Ammoniumpersulfat (APS)	SERVA
Ampicillin	SIGMA
ATP	Biometra
BCIP/NBT-Lösung	Boehringer Mannheim
Betain Monohydrat	Fluka
Blotpapier 1,2 mm GB004	Schleicher&Schüll
Blot-Membran: Hybond N	Amersham Buchler
Bovine Serum Albumin Fraktion V (BSA) standard grade	SERVA
5-Bromo-4-chloro-3-indolyphosphat (BCIP)	Biometra
Bromphenolblau	SIGMA
CDP Chemolumineszenzsubstrat	Ambion
dATP	Genecraft
dCTP	Genecraft
DEPC	SIGMA
dGTP	Genecraft
dTTP	Genecraft
Ethylendiaminotetraessigsäure, Di-Natrium-Salz (EDTA)	Merck
Eosin	Merck
Ethidiumbromid	SIGMA
Formamid	SIGMA
Glukose	Merck
Glycerin	SIGMA
Guanosin	Fluka
Hefeextrakt	DIFCO
Hexadecyltrimethylammoniumbromid (CTAB)	Fluka
Isopropyl- β -D-thiogalactopyranosid (IPTG)	SIGMA
Laktose	SIGMA
'100 bp-Leiter', 'Perfect-1kb-DNA-ladder', 'kb-DNA-Leiter'	Life Technologies, Genecraft
Lennox L Agar	Life Technologies
Lennox L Broth Base	Life Technologies
β -Mercaptoethanol	SIGMA
Methylenblau	SIGMA
Natrium-Dodecylsulfat (SDS)	Fluka
Nitroblau-Tetrazoliumchlorid (NBT)	SIGMA
PAA-Gellösung, 19:1, 30 % (GATC) oder 40 % (AccuGel)	GATC, national diagnostics
Pipettenspitzen	Alpha, Eppendorf
Roti-Phenol	Roth
Sequenator	GATC
Spectinomycin	Fluka
Streptavidin-alkalische Phosphatase-Konjugat (SA-AP)	GATC
Tetramethylethylendiamin (TEMED)	SERVA
Tris[hydroxymethyl]aminomethan (Tris-Base)	USB
Triton X-100	SERVA
Xylencyanol	SIGMA

Alle anderen Chemikalien sind von z.A. Qualität und stammen von Merck, Darmstadt oder Boehringer Mannheim. Für alle Versuche werden ausschließlich sterile, autoklavierte Gefäße und sterile, autoklavierte Pipettenspitzen verwendet. Das jeweils zur Herstellung von Medien und Puffern eingesetzte Reinstwasser wird über eine Milli-Q[®]-Anlage gewonnen. Für Ansätze in der Polymerase-Kettenreaktion (PCR) und Restriktionsspaltung wird es zusätzlich autoklaviert und in Aliquots bei $-20\text{ }^{\circ}\text{C}$ gelagert. Im weiteren Text wird dieses Reinstwasser mit dH₂O abgekürzt.

C.1.5 Software und Internet-Datenbanken

Software	Verwendungszweck	Hersteller / Internet-Adresse
BLAST 2.0	Suche in DNA-Datenbanken nach ähnlichen Sequenzen	http://www.ncbi.nlm.nih.gov/BLAST/
Censor Web Server	Suche in Sequenzen nach repetitiven Elementen	http://www.girinst.org/~server/censor.html
ClustalW 1.6	Alignment von DNA-Sequenzen	Thompson <i>et al.</i> , 1994
Enhance V1.03	Erstellung von Plasmidkarten	Scientific & Educational Software, 1991
MatInspector V2.2	Suche in Sequenzen nach Bindungsstellen von Transkriptionsfaktoren	http://transfac.gbf.de/cgi-bin/matSearch/ Quand <i>et al.</i> , 1995
ModelInspector V1.1	Suche in Sequenzen nach 2 kombinierten Bindungsstellen von Transkriptionsfaktoren	http://transfac.gbf.de/cgi-bin/matSearch/ Frech <i>et al.</i> , 1997
Oligo 4.0 - 2004	Primerdesign	Wojciech Rychlik
ORF Finder	Suche in Sequenzen nach offenen Leserastern	http://www.ncbi.nlm.nih.gov/gorf/gorf.html
RepeatMasker Web Server	Suche in Sequenzen nach repetitiven Elementen	http://ftp.genome.washington.edu/cgi-bin/RepeatMasker
Rev-Seq	Erzeugung einer komplementären DNA-Sequenz	diese Arbeit (siehe C.2.1 Basic Programme)
WinCam 2.2	Auswertung von Gelelektrophoresen	Cybertech Berlin

C.1.6 Enzyme

Tabelle 5: Restriktionsenzyme

Enzym	Erkennungssequenz und Spaltstelle	Verwendeter Puffer	Hersteller
Restriktionsendonuklease <i>AluI</i>	5'AG ∇ CT ^{3'}	Puffer A	Genecraft
Restriktionsendonuklease <i>Asp718</i>	5'G ∇ GATCC ^{3'}	Puffer A	Boehringer Mannheim
Restriktionsendonuklease <i>BamHI</i>	5'G ∇ GATCC ^{3'}	Puffer B	New England Biolabs
Restriktionsendonuklease <i>DraI</i>	5'TTT ∇ AAA ^{3'}	Puffer M	New England Biolabs
Restriktionsendonuklease <i>EcoRI</i>	5'G ∇ AATTC ^{3'}	Puffer H	Genecraft
Restriktionsendonuklease <i>KpnI</i>	5'GGATC ∇ C ^{3'}	Puffer L	Genecraft
Restriktionsendonuklease <i>NcoI</i>	5'C ∇ CATGG ^{3'}	Puffer H	New England Biolabs
Restriktionsendonuklease <i>NdeI</i>	5'CA ∇ TATG ^{3'}	Puffer H	New England Biolabs
Restriktionsendonuklease <i>PstI</i>	5'CTGCA ∇ G ^{3'}	Puffer H	New England Biolabs
Restriktionsendonuklease <i>PvuII</i>	5'CAG ∇ CTG ^{3'}	Puffer M	New England Biolabs
Restriktionsendonuklease <i>RcaI</i>	5'T ∇ CATGA ^{3'}	Puffer B	Boehringer Mannheim
Restriktionsendonuklease <i>RsaI</i>	5'GT ∇ AC ^{3'}	Puffer L	Boehringer Mannheim
Restriktionsendonuklease <i>SallI</i>	5'G ∇ TCGAC ^{3'}	Puffer H	New England Biolabs
Restriktionsendonuklease <i>ScaI</i>	5'AGT ∇ ACT ^{3'}	mitgeliefert	Promega
Restriktionsendonuklease <i>XhoI</i>	5'C ∇ TCGAG ^{3'}	Puffer H	New England Biolabs

Die Spaltstelle der Enzyme ist durch einen Pfeil gekennzeichnet. Die Angaben für Puffer beziehen sich auf 10x Puffer für Restriktionsenzyme von Boehringer.

Tabelle 6: Weitere Enzyme

Enzym	Verwendeter Puffer	Hersteller
alkalische Phosphatase aus Kälberdarm	Puffer 3	Boehringer Mannheim
Genescript reverse Transkriptase	mitgeliefert	Genecraft
Klonierte <i>Pfu</i> DNA-Polymerase	-	Stratagene
Klen-Therm DNA-Polymerase	mitgeliefert	Genecraft
KlenThermase Sequenase	mitgeliefert	Genecraft
RNase A	-	Amresco
RNase-Inhibitor	-	Amresco
alkalische Phosphatase aus Krabben	mitgeliefert	Boehringer Mannheim
Streptavidin-alkalische Phosphatase-Konjugat (SA-AP)	Puffer 3	GATC
T4 DNA Ligase	mitgeliefert	Genecraft
T4 RNA Ligase	mitgeliefert	New England Biolabs
terminale Nukleotidyl-Transferase	mitgeliefert	Promega

C.1.7 Primer

Die Primer wurden von MWG Biotech und Merck bezogen. Die Angabe der Bindungsstellen von Primern im menschlichen HCII-Gen wird nach der Numerierung der Nukleotide nach Herzog¹¹⁹ durchgeführt. Die Bindungsstellen der HCII-Primer bei Maus und Ratte sind nach der ersten Position im jeweiligen Intron 1 angegeben¹²⁰ (ebenfalls die verwendeten Primer, die im Exon 2 binden). Manche Primer aus dem Intron 1 können sowohl bei Maus und bei Ratte binden, sind jedoch nur als Maus- oder Ratten-Primer aufgeführt (Die Intron-Primer sind im Alignment im Ergebnisteil eingezeichnet).

Die Positionsangabe bezieht sich jeweils auf die erste Base am 5'-Ende des Primers. In den Primer-Tabellen sind Restriktionsschnittstellen durch kursive Buchstaben und Veränderungen gegenüber der chromosomalen Sequenz (u.a. zur Erzeugung von Restriktionsschnittstellen) durch fette Buchstaben gekennzeichnet.

Tabelle 7: Verwendete Primer für die menschliche HCII-Promotor-Region

Name	Organismus	Sinn (+/-)	Position	Sequenz
Mst90	Mensch	+	Position -4957	HO-CGT GGC TGG <i>GAG GTA</i> CCA AGT GGT TTT GAC ACA GA
Mst91	Mensch	+	Position -4402	HO-AGG GGG GAA GTG <i>GGT ACC</i> CTG AAA TAG ATA AAG
Mst92	Mensch	+	Position -2262	HO-TGG CTT TAA <i>AGG TAC</i> CAA AAG TAA TAC ATA CA
Mst93	Mensch	+	Position -688	HO-CAC CCA TAA ACA CAA <i>AAG GTA</i> CCA CAG CCT CAG A
Mst94	Mensch	-	Position 10	HO-GCA GCC AAG GTC <i>TCG AGT</i> GGG ATT TTG ATG TGA A
Mst95	Mensch	-	Position 5226	HO-TGA GTG TTC <i>CAT GGT</i> GGC GGA GCT AAA G
Mst96	Mensch	+	Position 5102	HO-TCC TGG GTC AAA GCC <i>ACA</i> GGG AAC CT
Mst97	Mensch	-	Position 155	HO-TGT AGA <i>AAG CTT</i> CTG <i>AAT TCT</i> TAG TT
Luc+	-	(+)	stromaufw. der mcs auf pGL3	HO-CCA GTG CAA GTG CAG GTG CCA GAA
Luc-	-	(-)	stromabw. der mcs auf pGL3	HO-TAC CGG AAT GCC AAG CTT ACT TA

¹¹⁹ Herzog *et al.*, 1991

¹²⁰ Kamp & Ragg, 1999

Tabelle 8: Verwendete Primer zur Charakterisierung des HCII-Intron 1 aus der Maus

Name	Organismus	Sinn (+/-)	Position	Sequenz
Mst2	Maus	-	Position 4838	HO-GAA AAA CAG ACA ACA AGG AAT GGC AGG GAG CCT
Mst3	Maus	-	Position 4894	HO-GAG AGA AAG AAG AGT GCA GAG TGG ATG TTT C
Mst12	Maus	-	Position -32	HO-TCT CTG CCA GAC CAC TGC TCA GAG
Mst12 H	Maus	-	Position -32	HO-ATA ATA AGC TTC TCT GCC AGA CCA CTG CTC AGA G
Mst13	Maus	+	Position 4726	HO-TGA GCA TTG GTG CTG TGA AGG CTT C
Mst14	Maus	-	Position 4750	HO-GAA GCC TTC ACA GCA CCA ATG CTC A
Mst14 B	Maus	-	Position 4750	HO-AAT GGA TCC GAA GCC TTC ACA GCA CCA ATG CTC A
Mst36	Maus	-	Position 4927	HO-ATT AGT TAG CTG CTC AGC AAG GC
Mst37	Maus	+	Position 65	HO-CAA TTG GTT CCT TTA TTG AGA A
Mst38	Maus	-	Position 4615	HO-TAT GAG CCC CTG GAG CAG TGG TT
Mst39	Maus	-	Position 4538	HO-GCC GGA TCC CCA ACC ATA AGA TTA GTT C
Mst40	Maus	+	Position 158	HO-ACC GGA TCC ATT GGT ACA ACT GAC AAA AAT G
Mst43	Maus	+	Position 2997	HO-TGG TCA CAG GCA CAA AAT AGA CTT C
Mst44	Maus	+	Position 2642	HO-AGG CAA GCA GCA GTG AGA CTA GAG
Mst45	Maus	+	Position 3585	HO-TGG ACA CCT CTC CCT TTC
Mst46	Maus	+	Position 1980	HO-GTC CCT AAG TTA TGA TTA AAA
Mst47	Maus	-	Position 2930	HO-AAG GGT TGC TGG CTC CTG ATA AA
Mst48	Maus	-	Position 1992	HO-TGG CCA CAG GAT TGT TTT AAT CA
Mst49	Maus	+	Position 2402	HO-TCA ATT CCC AAC AAC CAC ATG
Mst50	Maus	-	Position 3833	HO-GGG CGA TTC CTA GAG CCT CTT T
Mst51	Maus	-	Position 3680	HO-TGG GGG CAG GGA AAG GGA GAG
Mst52	Maus	-	Position 1578	HO-AGG GGG CTC AGC AGT TAG TAC
Mst53	Maus	-	Position 3764	HO-TTG CCT GCT GAA GGA CAC AGA AT
Mst54	Maus	+	Position 3281	HO-CCC TGT TTG CAC AGT GAG ATT GA
Mst55	Maus	-	Position 1210	HO-CTG AAG GGT AAG CAT CAT ATA
Mst56	Maus	-	Position 896	HO-TGG AGT TGG CCT TTT TGT ATA TG
Mst70	Maus	+	Position 3314	HO-TAG AAA AGA TTT TCT TTC TCC TGA C
Mst71	Maus	-	Position 3990	HO-TGC ACA CAC ATG TAA ACA TGT A
Mst72	Maus	-	Position 478	HO-GGG GGG TAC ATG TAT ATG TAT CA
Mst73	Maus	+	Position 475	HO-CCC CAT GTG CAC CAC ATC ACA TC
Mst74	Maus	+	Position 808	HO-TTC CCT TCT TGC TTG TTC TGT TT
Mst75	Maus	+	Position 1058	HO-GAG CAA TAT AGA AAG ATT TAT CTG
Mst76	Maus	-	Position 2571	HO-TCC CTA TAA TGT GGT TAA AAT TA
Mst77	Maus	-	Position 2836	HO-TGG CCT TCC TTT ACC ACT ACT AA
Mst78	Maus	-	Position 4254	HO-CAC TGC CCT AGA TGA AAT TTA TT
Mst79	Maus	+	Position 4245	HO-AGG GCA GTG GTT CTC AAC CTG T
Mst80	Maus	-	Position 2354	HO-GGC CCA AAG ATT TAT TTA TTA TTA TAC
Mst81	Maus	-	Position 4192	HO-ATG GAG GTG GAC AGG TTT AC
Mst82	Maus	+	Position 3835	HO-GGC CCC TCT TTT ACA TCA GTA AC
Mst83	Maus	-	Position 4443	HO-AGG CCA GCC TAG CCT ACA AAG
Mst84	Maus	+	Position 3127	HO-CTG GGC AGA AAT CTG GAT TTA ATA
Mst85	Maus	-	Position 2304	HO-GAG GGC ATC AGA TCT CAT TAC A
Mst86	Maus	+	Position 4606	HO-GGG GCT CAT AGA TTC CAG GAG AT

Tabelle 9: Verwendete Primer zur Charakterisierung des HCII-Intron 1 aus der Ratte

Name	Organismus	Sinn (+/-)	Position	Sequenz
Mst4	Ratte	+	Position -21	HO-CCA CTG CTC AGA GGC ACA GA
Mst5	Ratte	-	Position 4622	HO-TGG GGA AAC TGC ATC GAC AAC
Mst41	Ratte	-	Position 4019	HO-TCG GGA TCC AGA CAG CCA GGC TAC ACA G
Mst42	Ratte	+	Position 3681	HO-GCC CAG CAG GCG AAG GCGCTA CT
Mst57	Ratte	+	Position 838	HO-TGG CTC ACA ACC ATC TGT AAT GA

Fortsetzung der Tabelle 9

Name	Organismus	Sinn (+/-)	Position	Sequenz
Mst61	Ratte	+	Position 2062	HO-GGT AAA AGG GTC AGA AAC TTA AG
Mst62	Ratte	-	Position 2997	HO-CTT GCC TTG TTT TTG AGG AGA ACT
Mst63	Ratte	-	Position 1312	HO-GGC TTG GGA GGG TAA GCA TCA G
Mst64	Ratte	+	Position 1380	HO-TGG GCT CAG ACT AGT CCC CTT GA
Mst65	Ratte	-	Position 2763	HO-TGC TGG CTC CTA ATA AAG TAT TC
Mst66	Ratte	+	Position 1119	HO-TTC CCT TCC TCT AAG AAG AGA ATA
Mst67	Ratte	-	Position 1781	HO-GGG GCT TCA GAA GTG CTT ATT TT
Mst68	Ratte	-	Position 2123	HO-AGC CTT GAC TGT TCC TGG AAC T
Mst69	Ratte	+	Position 2707	HO-CCC TCA ATC CCT ACA TTT TAG AGC T

Tabelle 10: Verwendete Primer für mRNA-Analysen und Klonierungen

Name	Organismus	Sinn (+/-)	Position im HCII-Gen	Sequenz
Dag21	Mensch	-	Position 5470	HO-GC GAA TTC ATA TCT TCT CCA GGT CCA GAT AGT
Dag31	Mensch	-	Position 5855	HO-GC GAA TTC AAA TTC CTC CTG AAG AGG CGA TG
Dag50	Mensch	+	Position 5391	HO-GC GAA TTC CCA CAA GGA AAA CAC CGT CAC
Dag70	Mensch	-	Position 5958	HO-GGC CTC AGC AAA GTA ATA CTC
USC1	Mensch	+	Position 4739	HO-CCC CCA TGT GCT CTG GCA GCA TTA T
USC2	Mensch	+	Position 4977	HO-GGG GCC CAC AGA TCC TTC ATT GAG GTT TAT
USC3	Mensch	-	Position 12891	HO-GGG GCG GTC GAC AGT GAA GCG GAC TT
USC4	Mensch	+	Position 10107	HO-CCC CGG GTG GTG GAG AGA TGG CAA AA
USC8	Mensch	-	Position 5336	HO-TGG GGA TCT GCA GAC TGA GCA GTT
USC9	Mensch	-	Position 4781	HO- AAA GCG AAG GCT CTC AAA ATA ATG
USC11	Mensch	+	Position 23	HO-AGC GCC ACT TCT CAG AAA CAC AG
USC12	Mensch	-	Position 4780	HO-AAG CGA AGG CTC TCA AAA TAA TG
USC13	Mensch	-	Position 10118	HO-CAC CAC CCG GGG TGT CAG TTG
USC14	Mensch	-	Position 10046	HO-GCT GAT GCC CCC CAC GTA TTC
USC15	Mensch	+	Position 5142	HO-GCA GCG GGG TGT GGA TCA GC
USC16	Mensch	+	Position 5164	HO-GGC CGC CTT TCA CTG TGT TCT GTT
USC19	Mensch	+	Position 0	HO-CCT TGG CTG CTT TCA TCT CTG AAG
USC20	Mensch	-	Position 4643	p-TGG GGC TGG TGA TTG GAG AGT AC
USC22	Mensch	+	Position 4521	HO-CTG GCC CAT GGA CAT TTT TCA AC
USC23	Mensch	-	Position 5374	HO-GAG GCA TGC TCA GGT TTT TGT TAT
USC24	Mensch	-	Position 4339	HO-CGG CCT CCT GGA CTC AAG CAA TTC T
USC25	Mensch	+	Position 3661	HO-GCT CAT CTA ACC AGA GTG CAG TTA A
USC26	Mensch	-	Position 4039	HO-AGC CCC AGA GGT GAT ATC ACT TG
USC27	Mensch	+	Position 4582	HO-CAG GCA TGG CTG AGT CCC ACT A
HCII5'	Mensch	-	Position 5859	HO-CCC AAA ATT CCT CCT
ATIII12	Mensch	+	(Position 176) ¹²¹	HO-CCG CGG GAC ATT CCC ATG AAT C
ATIII148	Mensch	-	(Position 584) ¹²¹	HO-CAA AAA GGC GAT TGG CTG ATA CTA
Ap ^r -1	(<i>E. coli</i>) pUC	+	Ap ^r -1/2 binden jeweils ca. 70 Bp stromabwärts bzw. -aufwärts der	Bio-TGG CAG CAC TGC ATA ATT CTC TT
Ap ^r -2	(<i>E. coli</i>) pUC	-	<i>ScaI</i> -Schnittstelle im Ap ^r -Gen in pUC-Derivaten	Bio-GGC GCG GTA TTA TCC CGT ATT G
A5'-NV	Ankerprimer mit ddC am Ende für Anchor-PCR			P-CTG CAT CTA TCT AAT GCT CCT CTC GCT ACC TGC TCA CTC TC-STOP
A5'-1	1. Primer zu A5'-NV, für A-PCR			HO-GAG AGT GAG CAG GTA GCG AGA
A5'-2	2. Primer zu A5'-NV, für A-PCR			HO-AGA GTG AGC AGG TAG CGA GAG GAG
A5'-3	3. Primer zu A5'-NV, für A-PCR			p-CGA GAG GAG CAT TAG ATA GAT GCA G
A5'-Tail	Primer für (pC) _n -Template, kann mit A5'-3 amplifiziert werden			p-CGA GAG GAG CAT TAG ATA GAT GCA GCC CCC CCC CCC CCC CC

p-: Primer ist am 5'-Ende phosphoryliert; Bio-: Primer ist am 5'-Ende biotinyliert; -STOP: ddCMP am 3'-Ende

¹²¹ Bock *et al.*, 1982

C.1.8 Plasmide

Tabelle 11: Plasmide

Plasmid	Genotyp / Eigenschaften	ori	Herkunft
p4R9	pUC13 mit ca. 5 kB <i>XbaI</i> -Fragment aus Comid p4R	pUC8	H. Ragg
pGEM-5Zf(+)	<i>lacZ</i> mit spezieller mcs, Ap ^r , Größe: 3 kB	pUC8	Promega
pGEM-7Zf(+)	<i>lacZ</i> mit spezieller mcs, Ap ^r , Größe: 3 kB	pUC8	Promega
pGL3Basis	Plasmid zur Promotoranalyse, Ap ^r , Größe: 4,8 kB	pUC8	Promega
pGL3Control	Plasmid zur Promotoranalyse, Ap ^r , Größe: 5,3 kB	pUC8	Promega
pPBK14	<i>HindIII</i> -2kB-Spc ^r - <i>HindIII</i> -Kassette aus pHP45Spc in pGEM7Zf, Größe: 5 kB	pUC8	P.B. Kamp ¹²²
pPBK30-x	<i>AluI</i> -Fragmente des PCR-Produkts (Primer Mst12H-Mst14B) von genomischer Maus DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK31-x	<i>AluI</i> -Fragmente des PCR-Produkts (Primer Mst12H-Mst5) von genomischer Ratten DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK32-x	<i>AluI</i> -Fragmente des PCR-Produkts (Primer Mst4-Mill54) von genomischer Ratten DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK33-x	<i>DraI/HaeIII/RsaI</i> -Fragmente des PCR-Produkts (Primer Mst40-Mst52) von genomischer Maus DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK34-x	<i>PvuII/XbaI</i> -Fragmente des PCR-Produkts (Primer Mst12H-Mst14B) von genomischer Maus DNA in pPBK14 (geschnitten mit <i>ScaI/XbaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK35-x	<i>HaeIII</i> -Fragmente des PCR-Produkts (Primer Mst4-Mill54) von genomischer Ratten DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK36-x	<i>RsaI</i> -Fragmente des PCR-Produkts (Primer Mst4-Mill54) von genomischer Ratten DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK37-x	<i>AluI</i> -Fragmente des PCR-Produkts (Primer Mst52-Mst57) von genomischer Ratten DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK38-x	<i>AluI/HaeIII</i> -Fragmente des PCR-Produkts (Primer Mst39-Mst42) von genomischer Ratten DNA in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: 5,5-7,5 kB	pUC8	diese Arbeit
pPBK50-x	<i>HaeIII</i> -Fragment des PCR-Produkts (Primer A5'-3-Usc8) von poly(A) ⁺ -RNA aus menschlicher Leber in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: ca. 5,5 kB	pUC8	diese Arbeit
pPBK51-A	PCR-Produkt (Primer A5'-3-Usc20) von Gesamt-RNA (über A-PCR) aus A549- Zellen in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: ca. 5,5 kB	pUC8	diese Arbeit
pPBK51-B	PCR-Produkt (Primer A5'-3-Usc20) von Gesamt-RNA (über T-PCR) aus A549- Zellen in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: ca. 5,5 kB	pUC8	diese Arbeit
pPBK51-C	kleines PCR-Produkt (Primer Mst4-Mst36) von Maus poly(A) ⁺ -RNA (über A-PCR) aus Leber in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: ca. 5,5 kB	pUC8	diese Arbeit
pPBK60	<i>Asp718-XhoI</i> -2,2 kB-Fragment von Plasmid pPBK61 in pGL3Basis (geschnitten mit <i>Asp718-XhoI</i>), Größe: ca. 7,0 kB	pUC8	diese Arbeit
pPBK61	PCR-Produkt (Primer Mst92-Mst94) von Plasmid p4R9 in pPBK14 (geschnitten mit <i>ScaI</i>), Größe: ca. 7,2 kB	pUC8	diese Arbeit
pPBK62	PCR-Produkt (Primer Mst92-Mst94) von Plasmid p4R9 in pGL3Basis (geschnitten mit <i>SmaI</i>), Größe: ca. 7,0 kB, entgegengestzt zum <i>Luc</i> -Gen	pUC8	diese Arbeit
pUC8	<i>lacZ'</i> mit mcs, Ap ^r , Größe: 2,7 kB	-	
pWTBi1	pCDM8 mit <i>HCII</i> -Gen-Insert, Größe: 6,6 kB	pCDM	H. Ragg

Die jeweiligen Inserts spezieller Klone der Plasmidreihen können der Sequenzierungsaufstellung der Intron 1-Sequenzierungen des *HCII*-Gens im Anhang G.2 entnommen werden

¹²² Kamp, 1995

C.1.9 Nukleinsäuren

Tabelle 12: Verwendete Nukleinsäuren

Organismus/Zelllinien	Herkunft		Beschreibung
A549 Lungen - Zelllinie (menschlich)	ECACC; 86012804	RNA	1 µg/µl; Isolierung mit RNeasy Kit
HepG2 Leber - Zelllinie (menschlich)	ECACC; 85011430	RNA	1 µg/µl; Isolierung mit RNeasy Kit
Maus (<i>Mus musculus</i>)	Clontech	genom. DNA	0,1 µg/µl
Maus (<i>Mus musculus</i>), Leber	H. Ragg	poly(A) ⁺ -RNA	aus Lebergewebe, 1 µg/µl
Mensch (<i>Homo sapiens</i>), Leber	H. Ragg	poly(A) ⁺ -RNA	aus Lebergewebe, 1 µg/µl (Ch4 +5; 2. Charge 26.9.83)
Mensch (<i>Homo sapiens</i>), Lunge	Clontech	RNA	aus Lungengewebe, 1 µg/µl
Mensch (<i>Homo sapiens</i>), Lunge	Clontech; Lot 7050104	poly(A) ⁺ -RNA	aus Lungengewebe, 1 µg/µl
Mensch (<i>Homo sapiens</i>), Herz	Clontech; Lot 6120284	poly(A) ⁺ -RNA	aus Herzgewebe, 1 µg/µl
Mensch (<i>Homo sapiens</i>), Niere	Clontech; Lot 6120304	poly(A) ⁺ -RNA	aus Nierengewebe, 1 µg/µl
Mensch (<i>Homo sapiens</i>), Plazenta	Clontech; Lot 6120283	poly(A) ⁺ -RNA	aus Plazentagewebe, 1 µg/µl
Mensch (<i>Homo sapiens</i>), Gehirn	Clontech; Lot 8090440	poly(A) ⁺ -RNA	aus Gehirngewebe, 1 µg/µl
Ratte (<i>Rattus norvegicus</i>)	Clontech	genom. DNA	0,1 µg/µl

Konzentrationsangaben beziehen sich auf die Konzentrationen der in den jeweiligen Versuchen verwendeten Aliquots (gelöst in DEPC-dH₂O)

C.1.10 Bakterienstämme

Tabelle 13: *E. coli* Bakterienstämme

Stamm	Genotyp / relevante Eigenschaften	Referenz
<i>E. coli</i> JM83	<i>supE44 hsdR17 recA1 endA1 gyrA96 thi relA1 lac</i> ⁻ F'[<i>proAB</i> ⁺ <i>lacI</i> ^q <i>lacZ</i> ΔM15 Tn10 (<i>tet</i> ^r)]	Vieira & Messing, 1982
<i>Epicurian Coli</i> ⁰ XL1-Blue MRF' Kan	Δ(<i>mcrA</i>)183 Δ(<i>mcrCB-hsdSMR-mrr</i>) 173 <i>endA1 supE44 thi</i> 1 <i>recA1 gyrA96 relA1 lac</i> [F' <i>proAB lacI</i> ^q ZΔM15 Tn5 (Kan ^r)]	Stratagene

C.1.11 Antibiotika

Tabelle 14: Antibiotika

Antibiotikum	Abkürzung	Endkonzentration	Stammlösung
Ampicillin	Ap	150 µg/ml	150 mg/ml, bei 4°C mehrere Wochen haltbar
Kanamycin	Km	25 µg/ml	25 mg/ml, bei 4°C mehrere Wochen haltbar
Spectinomycin	Spc	100 µg/ml	10 mg/ml, bei 4°C mehrere Wochen haltbar

C.1.12 Medien

Tabelle 15: Verwendete Medien für *E. coli*-Kulturen

Grundmedium	Anmerkung	Zusammensetzung
Lennox L Powder:	<i>E. coli</i> -Vollmedium	50 % SELECT Peptone 140 25 % SELECT Hefe Extrakt 25 % Natriumchlorid
LB-Medium (Lennox L):	<i>E. coli</i> -Vollmedium	20 g Lennox L Powder ad 1 l dH ₂ O
LBG:	<i>E. coli</i> -Vollmedium	2 g Glukose ad 1 l LB-Medium
Lennox L Agar Powder:	<i>E. coli</i> -Vollmedium für Platten	33,3 % SELECT Peptone 140 16,6 % SELECT Hefe Extrakt 16,6 % Natriumchlorid 37,5 % SELECT Agar
LB-Agar (Lennox L Agar):	<i>E. coli</i> -Vollmedium	32 g Lennox L Powder ad 1 l dH ₂ O
LBG-Agar:	<i>E. coli</i> -Vollmedium	2 g Glukose ad 1 l LB-Agar
EMB-LB-Agar:	<i>E. coli</i> -Vollmedium für Platten	32 g Lennox L Agar Powder 10 ml EMB-Stammlösung ad 1 l dH ₂ O
	Zusatz nach dem Autoklavieren:	50 ml Laktose-IPTG-Lösung
Laktose-IPTG-Lösung:		0,585 M Laktose 6,3 mM IPTG in dH ₂ O
EMB-Stamm-Lösung:		3 M KH ₂ PO ₄ 90 mM Eosin 25 mM Methylenblau
SOB:	<i>E. coli</i> -Vollmedium: SOC ohne Glukose	
SOC:	<i>E. coli</i> -Vollmedium	5 g Hefeextrakt (w/v) 20 g Bacto-Tryptone 10 mM NaCl 2,5 mM KCl 10 mM MgCl ₂ 10 mM MgSO ₄ 2 mM Glukose ad 1 l dH ₂ O pH 6,9 eingestellt mit 1 N HCl

C.1.13 PufferTabelle 16: Puffer

Puffer	Anmerkung	Zusammensetzung
Assay-Puffer	Puffer für Chemolumineszenznachweis mit CDP und alkalischer Phosphatase	Fertiglösung von Ambion
10x BPB-Ladepuffer	Verwendung bei Agarosegelelektrophorese	50 % Glycerin (v/v) 0,1 % Bromphenolblau (w/v) 1 % SDS (w/v) 0,1 M EDTA, pH 7,5
2x BPB-Lade/Lyse-puffer	Verwendung bei Agarosegelelektrophorese	33 % Phenol (v/v) 33 % Chloroform (v/v) 7 % Ficoll 400 (w/v) 0,3 % SDS (w/v) 0,2 mg/ml RNase A 0,1 % BPB (w/v) 33 mM EDTA, pH 8,0
PBS	Puffer zum Waschen von eukaryotischen Zellen (autoklaviert)	1,44 g Na ₂ HPO ₄ ·7 H ₂ O 0,24 g KH ₂ PO ₄ 0,20 g KCl 8,00 g NaCl ad 1 l dH ₂ O pH 7,4 eingestellt mit 1 N HCl
Puffer 1	Puffer für Farbnachweis der Aktivität von Alkalischer Phosphatase (autoklaviert)	0,15 M NaCl 0,1 M Maleinsäure, pH 7,5
Puffer 2	Puffer für Farbnachweis der Aktivität von Alkalischer Phosphatase (autoklaviert)	10 % Blockierungs-Lösung (v/v) 90 % Puffer 1
Puffer 3	Puffer für Farbnachweis der Aktivität von Alkalischer Phosphatase (autoklaviert)	50 mM MgCl ₂ 0,1 M NaCl 0,1 M Tris-HCl, pH 9,5
Puffer A	Puffer für Restriktionsspaltung (10x konzentriert)	5 mM DTT 0,1 M Mg-Acetat 0,66 M K-Acetat 0,33 M Tris-Acetat, pH 7,9
Puffer B	Puffer für Restriktionsspaltung (10x konzentriert)	0,01 M β-Mercaptoethanol 0,05 M MgCl ₂ 1,0 M NaCl 0,1 M Tris-HCl, pH 8,0
Puffer L	Puffer für Restriktionsspaltung (10x konzentriert)	0,01 M DTE 0,1 M MgCl ₂ 0,1 M Tris-HCl, pH 7,5

Fortsetzung von Tabelle 16: Puffer

Puffer	Anmerkung	Zusammensetzung
Puffer M	Puffer für Restriktionsspaltung (10x konzentriert)	0,01 M DTE 0,1 M MgCl ₂ 0,5 M NaCl 0,1 M Tris-HCl, pH 7,5
Puffer H	Puffer für Restriktionsspaltung (10x konzentriert)	0,01 M DTE 0,1 M MgCl ₂ 1,0 M NaCl 0,5 M Tris-HCl, pH 7,5
5x SSC	Verwendung im Southernblot (autoklaviert)	250 ml 20x TBE ad 1 l dH ₂ O
20x SSC	Verwendung im Southernblot (autoklaviert)	3 M NaCl 0,3 M NaCitrat, pH 7,0
STE	Verwendung bei der 'S-Lyse' (autoklaviert)	50 mM EDTA 8 % Saccharose (w/v) 50 mM Tris pH 8,0 eingestellt mit 1 N HCl
TAE	Verwendung bei Agarosegelelektrophorese (autoklaviert)	1 mM EDTA 10 mM Na-Acetat 40 mM Tris pH 8,0 eingestellt mit Eisessig
1x TBE	Verwendung bei Sequenzierungen (autoklaviert)	200 ml 5x TBE 800 ml dH ₂ O
5x TBE	Verwendung bei Sequenzierungen (autoklaviert)	54 g Tris 27,5 g Borsäure 4,0 g EDTA ad 2 l dH ₂ O
TCM	Verwendung bei der Transformation von <i>E. coli</i> (autoklaviert)	10 mM CaCl ₂ 10 mM MgCl ₂ 10 mM Tris-HCl, pH 7,5
Waschpuffer	Puffer für Farbnachweis der Aktivität von Alkalischer Phosphatase (autoklaviert)	0,3 % Tween 20 (v/v) 99,7 % Puffer 1

C.1.14 LösungenTabelle 17: Lösungen

Lösung	Anmerkung	Zusammensetzung
ASL-Lösung 1	alkalische Lyse von Bakterien (autoklaviert)	1 % Glukose (w/v) 10 mM EDTA 25 mM Tris-HCl, pH 7,5

Fortsetzung von Tabelle 17: Lösungen

Lösung	Anmerkung	Zusammensetzung
ASL-Lösung 1*	Verwendung für 'AS-Lyse' von Bakterien (jeweils frisch ansetzen)	10 ml ASL-Lösung 1 1 mg RNase A 1 mg Lysozym
ASL-Lösung 2a	Verwendung für 'AS-Lyse' von Bakterien	0,4 N NaOH
ASL-Lösung 2b	Verwendung für 'AS-Lyse' von Bakterien (autoklaviert)	2 % SDS (w/v)
ASL-Lösung 2*	Verwendung für 'AS-Lyse' von Bakterien (jeweils frisch ansetzen)	50 % ASL-Lösung 2a (v/v) 50 % ASL-Lösung 2b (v/v)
ASL-Lösung 3	Verwendung für 'AS-Lyse' von Bakterien	3 M Na-Acetat, pH 5,2
ASL-Lösung 4	Verwendung für 'AS-Lyse' von Bakterien	35 % PEG 4000 (w/v)
APS-Lösung	(Lagerung in 150 µl Aliquots bei -20°C)	1 % Amoniumpersulfat (w/v)
ATP-Lösung		10 mM ATP
BCIP/NBT-Färbelösung	(jeweils frisch ansetzen)	0,2 ml BCIP/NBT 50 ml Puffer 3
Blockierungs-Lösung	Puffer für BCIP/NBT-Farbnachweis bei Sequenzierungen (autoklaviert)	10 g Block-Reagenz (w/v) ad 100 ml Puffer 1
EtBr-Färbebad	Färbung von DNA-Banden in Gelen	20 µg/l EtBr in dH ₂ O
PCR-Master-Mix	für PCR mit KlenTherm- oder KlenTherm/ <i>Pfu</i> -DNA-Polymerase: (in 85 µl Aliquots bei -20°C gelagert)	25 µl 100 mM dATP 25 µl 100 mM dCTP 25 µl 100 mM dGTP 25 µl 100 mM dTTP 1 ml 10x Klen-Reaktions-Puffer 7,4 ml dH ₂ O
RNase-Lösung	10 Minuten bei 100°C inkubiert	10 g/l RNase A
4 % -PAA-Gel	Verwendung als Sequenziergel	15 ml 40 % 19:1 PAA-Gellösung (Acrylamid:Bisacrylamid) 10 ml dH ₂ O 125 ml 10x Verdünnungs-Lösung
Silan-Lösung		50 µl Silan 30 µl 100 % Essigsäure 300 µl dH ₂ O 9,6 ml 96 % Ethanol
Triton X-Lösung	Verwendung für 'S-Lyse' (jeweils frisch ansetzen)	2 % Triton X-100 (v/v) 98 % STE (v/v)
10x Verdünnungs-Lösung	Verwendung bei Sequenzierungen (sterilfiltriert)	252 g Harnstoff 112 ml 2x TBE 170 ml dH ₂ O

C.1.15 DNA-Längenstandards

Die folgenden Längenmarker werden in dieser Arbeit verwendet:

- ‘Perfect 1 kb DNA Ladder’ (Genecraft)
Dieser Marker wird in der Arbeit als ‘kB-Leiter’ bezeichnet.
- ‘100bp Ladder’ (Genecraft)
Dieser Marker wird in der Arbeit als ‘100 Bp-Leiter’ bezeichnet.
- ‘DNA Molecular Weight Marker VIII’ (Boehringer Mannheim)
Dieser Marker wird in der Arbeit als ‘Marker VIII’ bezeichnet.

Längenstandard in Auftragspuffer:

‘kB-Leiter’, ‘100 Bp-Leiter’	DNA (1 µg/µl)	2,5 % (v/v)
(Lagerung bei -20 °C)	10x BPB-Ladepuffer	10 % (v/v)
	dH ₂ O	87,5 % (v/v)

Längenstandard in Auftragspuffer:

‘Marker VIII’	DNA (0,25 µg/µl)	10 % (v/v)
(Lagerung bei -20 °C)	10x BPB-Ladepuffer	10 % (v/v)
	dH ₂ O	80 % (v/v)

Die ‘kB-Leiter’ und der ‘Marker VIII’ werden zur Fragmentlängenbestimmung von linearen doppelsträngigen DNA-Fragmenten von ca. 300 Bp bis zu 10 kB genutzt.

Nach einer Agarosegelelektrophorese in einem 1 % Agarosegel sind bei der ‘kB-Leiter’ folgende Banden zu sehen: 10000 Bp, 8000 Bp, 6000 Bp, 5000 Bp, 4000 Bp, 3000 Bp, 2500 Bp, 2000 Bp, 1500 Bp, 1000 Bp, 750 Bp, 500 Bp, 250 Bp. Die 1000 Bp und die 3000 Bp Bande haben zur besseren Unterscheidung die dreifache Konzentration gegenüber den anderen Banden.

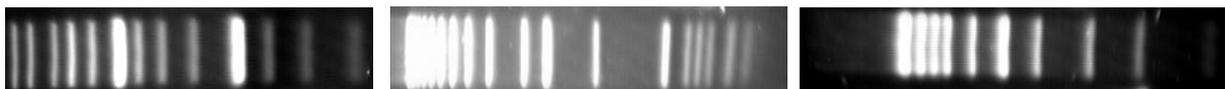
Nach einer Agarosegelelektrophorese in einem 1 % Agarosegel sind bei dem ‘Marker VIII’ folgende Banden zu sehen: 10000 Bp, 9000 Bp, 8000 Bp, 7000 Bp, 6000 Bp, 5000 Bp, 4000 Bp, 3000 Bp, 2000 Bp, 1600 Bp, 1000 Bp, 500 Bp, ca. 400 Bp, ca. 350 Bp, ca. 300 Bp, ca. 220 Bp (und kleinere Banden).

Die ‘100 Bp-Leiter’ wird zur Fragmentlängenbestimmung von linearen doppelsträngigen DNA-Fragmenten von ca. 100 Bp bis zu 1000 Bp genutzt.

Der Standard enthält DNA-Fragmente von 100 Bp bis 1000 Bp in 100 Bp Schritten. Nach der Elektrophorese von 0,25 µg des Fragmentgemisches in einem 2,5 % Agarose Gel sind zehn Banden sichtbar, wobei die 100 Bp Bande sehr schwach ist (Abbildung 17: Längenmarker).

Zur Fragmentlängen- und Konzentrationsbestimmung werden 10 µl der Längenstandards [0,025 µg/µl] in Auftragspuffer in einem Agarosegel parallel zu den zu analysierenden Proben aufgetragen.

Abbildung 17: Längenmarker



Gezeigt ist die Auftrennung von DNA-Längenstandards in Agarosegelen (Laufrichtung von links nach rechts). 0,25 µg ‘kB-Leiter’ (1 % Gel; links), 0,25 µg ‘Marker VIII’ (1 % Gel; Mitte) und 0,25 µg ‘100 Bp-Leiter’ (2,5 % Gel; rechts; die 100 Bp-Bande ist hier nicht zu erkennen)

C.2 Methoden

C.2.1 Basic-Programme

Programm: 'Rev-Seq.bas', zur Darstellung des komplementären DNA-Stranges

Das Programm öffnet die Datei 'seq.txt' und schreibt die komplementären Nukleotid-Kürzel in umgekehrter Reihenfolge (z.B.: AGTC -> GACT) in die Datei 'rev-seq.txt'. Die Sequenzdatei 'seq.txt' darf keine Zeilenumbrüche, Kommas, Punkte oder Semikolons enthalten, da ansonsten die Sequenz an dieser Stelle abbricht. Groß- und Kleinschreibung bleibt erhalten, andere Zeichen als A/C/G/T bzw. a/c/g/t bleiben unverändert erhalten.

```

OPEN "seq.txt" FOR INPUT AS #1 LEN = 5000
INPUT #1, a$
CLOSE
a = LEN(a$): b = a
FOR i = 1 TO a
  b$ = MID$(a$, b, 1)
  c$ = b$
  IF b$ = "a" THEN c$ = "t"
  IF b$ = "A" THEN c$ = "T"
  IF b$ = "c" THEN c$ = "g"
  IF b$ = "C" THEN c$ = "G"
  IF b$ = "g" THEN c$ = "c"
  IF b$ = "G" THEN c$ = "C"
  IF b$ = "t" THEN c$ = "a"
  IF b$ = "T" THEN c$ = "A"
  d$ = d$ + c$
  b = b - 1
NEXT i
OPEN "rev-seq.txt" FOR OUTPUT AS #1
PRINT #1, d$
CLOSE

```

Programm: '%GC.bas', zur Angabe des Gehalts an Nukleotiden in einer Sequenz

Das Programm öffnet die Datei 'seq.txt' und berechnet die Anzahl und den prozentualen Anteil der verschiedenen Nukleotide in der gespeicherten Sequenz. Die Sequenzdatei 'seq.txt' darf keine Zeilenumbrüche, Kommas, Punkte oder Semikolons enthalten, da ansonsten die Berechnung an dieser Stelle abbricht.

```

OPEN "seq.txt" FOR INPUT AS #1 LEN = 6000
INPUT #1, a$
CLOSE
a = LEN(a$)
aa = 0: cc = 0: gg = 0: tt = 0
FOR i = 1 TO a
  b$ = MID$(a$, i, 1)
  IF b$ = "a" or b$ = "A" THEN aa = aa + 1
  IF b$ = "c" or b$ = "C" THEN cc = cc + 1
  IF b$ = "g" or b$ = "G" THEN gg = gg + 1
  IF b$ = "t" or b$ = "T" THEN tt = tt + 1
NEXT i

```

```

PRINT "Gesamte Zeichen: "; a
PRINT "Andere Zeichen: "; a - aa - cc - gg - tt
PRINT "ACGT Zeichen: "; aa + cc + gg + tt
PRINT "A: "; aa; "- "; INT(aa * 100 / a); "%"
PRINT "C: "; cc; "- "; INT(cc * 100 / a); "%"
PRINT "G: "; gg; "- "; INT(gg * 100 / a); "%"
PRINT "T: "; tt; "- "; INT(tt * 100 / a); "%"
PRINT "G+C: "; INT((gg + cc) * 100 / a); "%"

```

Programm: '3'Intron.bas', zur Suche nach potentiellen 3'-Spleißstellen in Sequenzen

Das Programm öffnet die Datei 'seq.txt' und sucht nach möglichen 3'-Spleißstellen in einer Sequenz. Die angewendete Matrix bezieht sich auf die bekannten Strukturen,¹²³ wurde jedoch etwas ergänzt und vereinfacht. Die Auswertung wird in die Datei 'seq2.txt' geschrieben. Zuerst wird die Sequenz der möglichen 3'-Spleißstelle angegeben und dann die prozentuale Übereinstimmung mit der Matrix. Als nächstes erfolgt eine identische Angabe zu einem eventuellen Verzweigungspunkt. Zuletzt wird eine Bewertungszahl angegeben, die sich aus der Summe der beiden Übereinstimmungen mit den Matrizen zusammensetzt.

Eine mögliche Spleißstelle wird ausgegeben, falls die Übereinstimmung mit der Matrize für die 3'-Spleißstelle größer als 70 % (Variable: b), die Übereinstimmung der variablen Positionen an dem Verzweigungspunkt zur Matrix mindestens 37 % (Variable: g) beträgt und die Summe aus beiden Werten größer als 112 (Variable: d) ist. Damit dieses Programm auch andere Spleißstellen erkennt, die mit den gesetzten Parametern nicht erkannt werden, können die Minimalwerte in dem rotgedruckten Bereich herabgesetzt werden.

Dieses Programm erkennt alle bekannten 3'-Spleißstellen von Introns des HCII- und des PAI-1-Gen (Bereich Exon 2-9) des Menschen, alle 3'-Introngrenzen vom Ovalbumingen, sowie die experimentell ermittelten 3'-Introngrenzen des HCII-Gens bei Maus und Ratte innerhalb des 5'-Bereichs der Gene. Es werden allerdings ebenfalls viele weitere, nicht verwendete mögliche Grenzen ausgegeben. Damit dieses Programm auch andere Spleißstellen erkennt, die bisher eventuell nicht erkannt werden, können die Minimalwerte in dem grau hinterlegten Bereich herabgesetzt werden. Andererseits können die Werte für eine stringenter Suche angehoben werden, allerdings werden in diesem Fall nicht mehr alle der getesteten Introns erkannt. Die Sequenzdatei 'seq.txt' darf keine Zeilenumbrüche, Kommas, Punkte oder Semikolons enthalten, da ansonsten die Sequenz an dieser Stelle abbricht.

```

OPEN "seq.txt" FOR INPUT AS #1 LEN = 8000
INPUT #1, a$
CLOSE
a = LEN(a$): c$ = "": PRINT
FOR i = 1 TO a
  b$ = MID$(a$, i, 1)
  d$ = ""
  IF b$ = "a" OR b$ = "A" THEN d$ = "A"
  IF b$ = "c" OR b$ = "C" THEN d$ = "C"
  IF b$ = "g" OR b$ = "G" THEN d$ = "G"
  IF b$ = "t" OR b$ = "T" THEN d$ = "T"
  c$ = c$ + d$
NEXT i
OPEN "seq2.txt" FOR OUTPUT AS #1
a = LEN(c$)
PRINT #1, "Sequenzlaenge: "; a: PRINT #1,

```

¹²³ Lewin, 1991; Krawczak *et al.*, 1992

```

FOR i = 50 TO a - 1
  b$ = MID$(c$, i, 2)
  IF b$ <> "AG" THEN 100
  b$ = MID$(c$, i - 1, 1)
  IF b$ <> "C" AND b$ <> "T" AND MID$(c$, i + 2, 1) <> "G" THEN 100
  b = 0: t = 0
  FOR ii = i - 14 TO i - 3
    IF MID$(c$, ii, 1) = "C" THEN b = b + 90
    IF MID$(c$, ii, 1) = "T" THEN b = b + 90: t = t + 1
  NEXT ii
  IF MID$(c$, i - 1, 1) = "C" THEN b = b + 78
  IF MID$(c$, i + 2, 1) = "G" THEN b = b + 55
  IF t < 4 THEN 100
  f = 0
  FOR ii = i - 45 TO i - 21
    c = 0
    IF MID$(c$, ii + 4, 2) <> "AC" AND MID$(c$, ii + 4, 2) <> "AT" THEN 50
    IF MID$(c$, ii - 1, 1) = "C" THEN c = 40
    IF MID$(c$, ii - 1, 1) = "T" THEN c = 40
    IF MID$(c$, ii + 1, 1) = "C" THEN c = c + 40
    IF MID$(c$, ii + 1, 1) = "T" THEN c = c + 40
    IF MID$(c$, ii + 2, 1) = "T" THEN c = c + 44
    IF MID$(c$, ii + 2, 1) = "C" THEN c = c + 44
    IF MID$(c$, ii + 3, 1) = "A" THEN c = c + 38
    IF MID$(c$, ii + 3, 1) = "G" THEN c = c + 38
    IF MID$(c$, ii + 6, 1) = "T" THEN c = c + 40
    IF MID$(c$, ii + 6, 1) = "C" THEN c = c + 40
    IF c > g THEN g = c: f = ii
  50
  NEXT ii
  IF f = 0 THEN 100
  g = INT(g * 100 / 202)
  b = INT(b * 100 / 1213)
  d = b + g
  IF b > 70 AND g > 37 AND d > 112 THEN PRINT #1, "Pos.:"; i; MID$(c$, i - 14, 16); b;
  IF b > 70 AND g > 37 AND d > 112 THEN PRINT #1; "%, "; MID$(c$, f - 1, 8); g; "%, Gesamt:"; d;
  IF b > 70 AND g > 37 AND d > 112 THEN PRINT #1,
  100
NEXT i
CLOSE

```

Programm: '5'Intron.bas', Suche nach potentiellen 5'-Spleißstellen in Sequenzen

Das Programm öffnet die Datei 'seq.txt' und sucht nach möglichen 5'-Spleißstellen in einer Sequenz. Die angewendete Matrix bezieht sich auf die bekannten Strukturen,¹²⁴ wurde jedoch etwas ergänzt und vereinfacht. Die Auswertung wird in die Datei 'seq2.txt' geschrieben. Zuerst wird die Sequenz der möglichen 5'-Spleißstelle angegeben und dann die prozentuale Übereinstimmung mit der Matrix. Eine mögliche Spleißstelle wird ausgegeben, falls die Übereinstimmung mit der Matrize für die 5'-Spleißstelle größer als 48 % (Variable: b) ist. Damit dieses Programm auch andere Spleißstellen erkennt, die bisher eventuell nicht erkannt werden, kann der Minimalwert in dem rotgedruckten Bereich herabgesetzt werden.

Das Programm erkennt alle bekannten 5'-Spleißstellen von Introns des HCII- und des PAI-1-Gen (Bereich Exon 2-9) des Menschen, alle 5'-Introngrenzen vom Ovalbumingen, sowie die experimentell ermittelten 5'-Introngrenzen des HCII-Gens bei Maus und Ratte innerhalb des 5'-Bereichs der Gene.

¹²⁴ Lewin, 1991; Krawczak *et al.*, 1992

Es werden allerdings ebenfalls viele weitere, nicht verwendete mögliche Grenzen ausgegeben. Damit dieses Programm auch andere Spleißstellen erkennt, die bisher eventuell nicht erkannt werden, kann die Übereinstimmung der Variable 'b' mit der Matrize in dem blau hinterlegten Bereich herabgesetzt werden. Andererseits können die Werte für eine stringendere Suche angehoben werden, allerdings werden in diesem Fall nicht mehr alle der getesteten Introns erkannt. Die Sequenzdatei 'seq.txt' darf keine Zeilenumbrüche, Kommas, Punkte oder Semikolons enthalten, da ansonsten die Sequenz an dieser Stelle abbricht.

```

OPEN "seq.txt" FOR INPUT AS #1 LEN = 8000
INPUT #1, a$
CLOSE
a = LEN(a$): c$ = ""
FOR i = 1 TO a
  b$ = MID$(a$, i, 1)
  d$ = ""
  IF b$ = "a" OR b$ = "A" THEN d$ = "A"
  IF b$ = "c" OR b$ = "C" THEN d$ = "C"
  IF b$ = "g" OR b$ = "G" THEN d$ = "G"
  IF b$ = "t" OR b$ = "T" THEN d$ = "T"
  c$ = c$ + d$
NEXT i
OPEN "seq2.txt" FOR OUTPUT AS #1
a = LEN(c$): PRINT #1, "Sequenzlaenge:"; a: PRINT #1,
FOR i = 5 TO a - 5
  b$ = MID$(c$, i, 2)
  IF b$ <> "GT" THEN 100
  b = 0
  IF MID$(c$, i - 4, 1) = "C" THEN b = b + 29
  IF MID$(c$, i - 4, 1) = "A" THEN b = b + 34
  IF MID$(c$, i - 3, 1) = "C" THEN b = b + 38
  IF MID$(c$, i - 3, 1) = "A" THEN b = b + 35
  IF MID$(c$, i - 2, 1) = "A" THEN b = b + 62
  IF MID$(c$, i - 1, 1) = "G" THEN b = b + 77
  IF MID$(c$, i + 2, 1) = "A" THEN b = b + 60
  IF MID$(c$, i + 3, 1) = "A" THEN b = b + 74
  IF MID$(c$, i + 4, 1) = "G" THEN b = b + 84
  IF MID$(c$, i + 5, 1) = "T" THEN b = b + 50
  b = b * 100 / 479
  IF b > 48 THEN PRINT #1, i, MID$(c$, i - 4, 10), INT(b); "%": PRINT #1,
  100
NEXT i
CLOSE

```

C.2.2 PCR

Die Polymerase-Kettenreaktion (Polymerase chain reaction; PCR), die 1983 von Kary B. Mullis¹²⁵ erfunden wurde, wird zur Amplifizierung von spezifischen DNA-Sequenzen *in vitro* genutzt. Eine DNA-Polymerase synthetisiert anhand einer Nukleinsäure-Matrize den dazu komplementären Gegenstrang. *In vitro* werden dazu eine thermostabile DNA-Polymerase, ein zu amplifizierender DNA-Einzel- oder Doppel-Strang (DNA-Matrize), Oligonukleotidprimer, Desoxyribonukleotide und ein geeignetes Puffersystem benötigt.

¹²⁵ Mullis, 1987; Mullis, 1990

Die Oligonukleotidprimer werden synthetisch hergestellt (MWG Biotech, Gibco/BRL). Sie leiten sich aus den flankierenden Bereichen der zu amplifizierenden Sequenz ab, so daß das vervielfältigte Fragment zwischen ihnen liegt. Einer der beiden Oligonukleotidprimer ist komplementär zum Plus-Strang, der zweite komplementär zum Minus-Strang der DNA-Zielsequenz.

Die PCR-Reaktion läßt sich in drei Schritte unterteilen, die 25-45mal wiederholt werden. Standardmäßig werden in dieser Arbeit, sofern nicht anders beschrieben, jeweils 35 Zyklen verwendet. Als erstes wird der DNA-Doppelstrang durch hohe Temperaturen (94 °C) denaturiert. Anschließendes Abkühlen auf 37-68 °C ermöglicht die Hybridisierung der im großen Überschuß (um den Faktor 10^6 gegenüber der Zielsequenz) eingesetzten synthetischen Oligonukleotidprimer mit der komplementären Matrizen-DNA. Dieser Vorgang wird als Annealing bezeichnet. Die dafür verwendete Annealingtemperatur ist u.a. von der Sequenz der Oligonukleotidprimer abhängig. Die Schmelztemperatur eines DNA-Doppelstranges und damit auch die optimale Anlagerungstemperatur der Oligonukleotidprimer an die DNA läßt sich in erster Näherung aus der Anzahl der Guanin-Cytosin und Adenin-Thymin Basenpaare und ihrer Abfolge vorhersagen. Die Annealing-Temperatur bei einer PCR in dieser Arbeit wird nach folgender, vereinfachten Formel berechnet:

$$T_{\text{Annealing}} = 2 \text{ °C} * n_{\text{A+T}} + 4 \text{ °C} * n_{\text{G+C}}; \quad n_{\text{A+T}}: \text{Anzahl von A- und T-Nukleotiden}$$

$$n_{\text{G+C}}: \text{Anzahl von G- und C-Nukleotiden}$$

Bei Verwendung von Primerpaaren mit unterschiedlicher Annealingtemperatur wird die niedrigere Temperatur verwendet.

Für eine PCR wäre es optimal, zwei Oligonukleotidprimer mit gleichen oder sehr ähnlichen Schmelztemperaturen zu verwenden. Im dritten Schritt, der sogenannten Elongationsphase, synthetisiert die DNA-Polymerase von dem 3'terminale Ende dieser Oligonukleotidprimer ausgehend jeweils in 5' → 3'-Richtung bei 68 °C bis 72 °C den komplementären DNA-Gegenstrang. Im ersten PCR-Zyklus werden DNA-Einzelstränge synthetisiert, deren 5'-Ende zwar durch den Oligonukleotidprimer begrenzt ist, deren 3'-Ende sich aber durch einen undefinierten Polymerisationsabbruch an beliebiger Stelle befinden kann. Ab dem zweiten Zyklus wird die von beiden Oligonukleotidprimern definiert eingegrenzte Sequenz amplifiziert, da auch die Produkte des ersten Reaktionszyklus als Matrize dienen, so daß ein Produkt definierter Größe gebildet wird. Die Zahl dieser PCR-Produkte wächst exponentiell mit der Zykluszahl, da sich ihre Konzentration in jedem Zyklus verdoppelt. Nach 20 PCR-Zyklen ergibt sich eine theoretische Anreicherung um den Faktor 10^6 . Es wird allerdings meist nur eine Ausbeute von ca. 85 % erreicht, da die Ausbeuten von Zyklus zu Zyklus variieren, so daß man einen durchschnittlichen Amplifikationsfaktor von ungefähr 10^5 erhält.¹²⁶

C.2.2.1 DNA-Polymerasen

KlenTherm DNA-Polymerase

Die KlenTherm DNA-Polymerase ist ein Fragment der *Taq* DNA-Polymerase, der die 5' → 3'-Exonuklease-Aktivität fehlt und die hitzestabiler als die *Taq* DNA-Polymerase ist. Die KlenTherm DNA-Polymerase ist bis zu einer Temperatur von 99 °C stabil. Außerdem ist der Einbau fehlerhafter Nukleotide während der Polymerisation um den Faktor 2 geringer als bei der *Taq* DNA-Polymerase.¹²⁷

¹²⁶ Saiki *et al.*, 1985; Saiki *et al.* 1988; White *et al.* 1989; Linz, 1990

¹²⁷ Barnes, 1992; Barnes, 1994

KlenTherm DNA-Polymerase:	10 U/μl	
Lagerung bei -20 °C in:	100 mM	NaCl
	0,5 mM	EDTA
	1,0 mM	DTT
	0,01 % (v/v)	Tween 20
	50 % (v/v)	Glycerin
	10 mM	Kaliumphosphat-Puffer, pH 7,0

10x PCR-Reaktions-Puffer der KlenTherm DNA-Polymerase:	160 mM	(NH ₄) ₂ SO ₄
	35 mM	MgCl ₂
	0,15 % (w/v)	BSA
	670 mM	Tris-HCl, pH 9,1 bei 25 °C

Klonierte *Pfu* DNA-Polymerase

Die klonierte *Pfu* DNA-Polymerase wurde ursprünglich aus dem hyperthermophilen, marinen Archaeobakterium *Pyrococcus furiosus* isoliert. Dieses multifunktionale, thermostabile Enzym besitzt sowohl 5'→3'-DNA-Polymerase-Aktivität als auch 3'→5'-Exonuklease-Aktivität, was ihre DNA-Synthese-Genauigkeit um das 12fache gegenüber der *Taq* DNA-Polymerase erhöht. Das Temperaturoptimum liegt zwischen 72 °C und 78 °C. Die Aktivität bleibt auch nach einer einstündigen Inkubation bei 95 °C erhalten.¹²⁸

Klonierte <i>Pfu</i> DNA-Polymerase:	2,5 U/μl		
Lagerung bei -20 °C in:	50 % (v/v)	Glycerin	0,01 % (v/v) Tween 20
	0,1 mM	EDTA	0,1 % (v/v) Nonidet P-40
	1,0 mM	DTT	50 mM Tris-HCl, pH 8,2

DNA-Polymerase-Mix aus KlenTherm- und *Pfu* DNA-Polymerase

Um eine optimale Polymerase-Aktivität auch für längere Matrizen bei gleichzeitiger niedriger Fehlerrate und geringen Kosten zu erhalten, wird ein Polymerase-Mix aus Klen-Therm- und *Pfu* DNA-Polymerase (80 U KlenTherm / 1 U *Pfu*; *gesamt*: 0,1 U/μl im PCR-Ansatz) verwendet.¹²⁹

C.2.2.2 PCR-Ansatz

Wenn es nicht anders vermerkt ist, wird eine PCR nach folgendem Schema durchgeführt. Die Ansätze werden auf Eis pipettiert und anschließend sofort in den auf 94°C vorgeheizten Thermocycler gestellt ('Hotstart' zur Vermeidung von Primer-Fehlpaarung bei niedrigeren Temperaturen). Als Negativkontrolle dienen Ansätze ohne DNA-Matrize und Positivkontrollen werden mit Matrizen durchgeführt, die zuvor für die PCR unter diesen Bedingungen getestet wurden.

<u>je Reaktionsansatz:</u>	
DNA-Matrize + dH ₂ O	2,5 μl
PCR-Master-Mix (siehe C.1.14)	21,25 μl
2 Primer (je 20 pmol/μl)	2x 0,5 μl
<u>Klen/<i>Pfu</i>-Polymerase</u>	<u>0,25 μl</u>
	25,0 μl

¹²⁸ Stratagene Produktbeschreibung, 1998

¹²⁹ Cline *et al.*, 1996

PCR Temperaturzyklen:

1 Zyklus:	94 °C	2 Minuten	Denaturierung
35 Zyklen:	94 °C	25 Sekunden	Denaturierung
	40-68 °C	10 Sekunden	Annealing
	68 °C	2 Minuten je kB	Elongation
1 Zyklus:	68 °C	2 Minuten je kB	terminale Elongation
Endlos-Schleife:	10 °C		

C.2.2.3 Nested PCR

In der nested PCR werden Primer verwendet, die innerhalb des in der ersten PCR amplifizierten PCR-Fragmentes liegen. In eine neue PCR werden 1/100 oder 1/1000 Volumen des vorher durchgeführten PCR-Ansatzes als Matrize gegeben.

Eine nested PCR wird durchgeführt, wenn die Matrizen-Konzentration in der PCR sehr gering ist und wenn geringe Konzentrationen an einer zu amplifizierenden Matrize im Gemisch mit anderen Matrizen vorliegen. Unspezifisch amplifizierte Fragmente aus der ersten PCR können nicht erneut vermehrt werden.

C.2.2.4 Kolonie-PCR

Zur Überprüfung von Bakterien-Kolonien auf ein spezielles Plasmid kann eine PCR mit Plasmid-spezifischen (bzw. Insert-spezifischen) Primern direkt von den Bakterien-Kolonien durchgeführt werden.¹³⁰ Hierzu wird eine Einzelkolonie mit einem sterilen Zahnstocher in 30 µl dH₂O überführt und fünf Minuten bei 95 °C inkubiert. 0,5 µl des abzentrifugierten Überstandes werden als PCR-Matrize in eine PCR (C.2.2.2) eingesetzt.

C.2.3 Reverse Transkriptase gekoppelte PCR

Die reverse Transkriptase gekoppelte PCR (RT-PCR) beruht auf der Umschreibung einer RNA-Matrize in eine cDNA und deren anschließender Amplifikation mit Hilfe der PCR. An die (einzelsträngige) RNA kann ein spezifischer Primer (bzw. ein unspezifischer poly(T)-Primer für poly(A)⁺-RNAs) binden (siehe PCR-Beschreibung C.2.2). Mit Hilfe von Reversen Transkriptasen, RNA-abhängigen DNA-Polymerasen, wird unter geeigneten Bedingungen ein komplementärer DNA-Einzelstrang synthetisiert.

Es gibt verschiedene reverse Transkriptasen, die jeweils für unterschiedliche Problemstellungen geeignet sind. In dieser Arbeit werden Enzyme verwendet, die vor allem eine hohe Effizienz der Umschreibung gewährleisten, mit denen also auch bei extrem niedriger Matrizen-Konzentration eine DNA-Synthese möglich ist. Unter anderem wird dies über gentechnisch modifizierte Enzyme erreicht, denen die üblicherweise vorhandene RNase H-Aktivität fehlt. Nach der Reaktion wird die cDNA in eine PCR eingesetzt (1/20 Vol.), wobei bei der PCR Primerpaare verwendet werden, die stromaufwärts zu dem in der reversen Transkriptase-Reaktion verwendeten Primer liegen. Für die Bestimmung von 5'-Enden von mRNAs werden vor der PCR weitere enzymatische Reaktionen mit der cDNA durchgeführt (siehe A-PCR C.2.4; T-PCR C.2.5).

¹³⁰ Dallas-Yang *et al.*, 1998

C.2.3.1 Genescript reverse Transkriptase

Die Genescript reverse Transkriptase ist eine modifizierte reverse Transkriptase aus dem Moloney Murine Leukemia Virus (MMLV-reverse Transkriptase), bei der der für die RNase H kodierende Teil deletiert ist. Das Enzym kann einen komplementären DNA-Strang zu einer DNA- oder RNA-Matrize synthetisieren.

Sofern es im folgenden nicht weiter erwähnt ist, wird als Matrize je Reaktion 5 µg Gesamt-RNA oder 0,1 µg poly(A)⁺-RNA eingesetzt.

Genescript reverse Transkriptase:	100 U/µl	
Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	1 mM	EDTA
	0,1 mM	DTT
	0,1 mM	NaCl
	0,1 % (v/v)	Triton X-100
	50 mM	Tris-HCl, pH 8,3

1x PCR-Reaktions-Puffer der Genescript reverse Transkriptase:		
	100 mM	KCl
	5 mM	MgCl ₂
	10 mM	DTT
	50 mM	Tris-HCl, pH 8,3

Wenn es nicht anders vermerkt ist, wird eine RT-PCR nach folgendem Schema durchgeführt:

je Reaktionsansatz:

RNA-Matrize + DEPC-dH ₂ O	26,2 µl
5x Reaktionspuffer für Genescript	10,0 µl
50 mM MnCl ₂	3,0 µl
dNTP-Mix (je dNTP: 2,5 mM)	10,0 µl
Primer (20 µM)	0,3 µl
RNase-Inhibitor (20 U/µl)	0,25 µl
<u>Genescript Polymerase (100 U/µl)</u>	<u>0,25 µl</u>
	50,0 µl

Als Negativkontrollen dienen Ansätze ohne RNA-Matrize. Eine zusätzliche Negativkontrolle wird mit einem hitzeinaktivierten Gesamtansatz durchgeführt. Positivkontrollen werden mit Matrizen durchgeführt, die zuvor für die RT-PCR unter diesen Bedingungen getestet wurden. Bevor die reverse Transkriptase zu dem Ansatz gegeben wird, kann eine einminütige Inkubation bei 94 °C mit anschließendem raschem Abkühlen in Eiswasser durchgeführt werden, um doppelsträngige RNA-Bereiche zu denaturieren. Der Ansatz wird 45 Minuten bei 42 °C inkubiert. Anschließend wird die Reaktion durch fünfminütige Inkubation bei 94 °C (Inaktivierung der reversen Transkriptase) gestoppt.

C.2.3.2 Omniscrypt reverse Transkriptase

Die Omniscrypt reverse Transkriptase ist ein rekombinantes, heterodimeres Enzym, das in *E. coli* exprimiert wird. Das Enzym leitet sich nicht von der MMLV-, AMV-, oder HIV-1-reversen Transkriptase ab. Eine genauere Angabe des Herstellers liegt nicht vor.

Die Vorteile des Enzyms sind: hohe Effizienz, hohe Sensitivität auch für geringe RNA-Matrizen-Konzentrationen und ein weiter dynamischer Bereich der RNA-Matrizen Konzentration, in welchem mit der anschließenden PCR spezifische DNA-Fragmente gebildet werden können. Sofern es nicht anders beschrieben ist, wird als Matrize je Reaktion 0,1 µg poly(A)⁺-RNA eingesetzt.

Omniscript reverse Transkriptase: 4 U/µl

Lagerung bei -20 °C

Lagerungs- und Reaktionspuffer: nicht angegeben

Reaktionsansatz mit der Omniscript reversen Transkriptase:

je Reaktionsansatz:

RNA-Matrize + DEPC-dH ₂ O	13,0 µl
10x Reaktionspuffer für Omniscript rev. T.	2,0 µl
dNTP-Mix (je dNTP: 2,5 mM)	2,0 µl
Primer (20 µM)	1,0 µl
RNase-Inhibitor (20 U/µl)	1,0 µl
<u>Omniscript Polymerase (4 U/µl)</u>	<u>1,0 µl</u>
	20,0 µl

Als Negativkontrollen dienen Ansätze ohne RNA-Matrize. Eine zusätzliche Negativkontrolle wird mit einem hitzeinaktivierten Gesamtansatz durchgeführt. Positivkontrollen werden mit Matrizen durchgeführt, die zuvor für die RT-PCR unter diesen Bedingungen getestet wurden. Bevor die reverse Transkriptase zu dem Ansatz gegeben wird, kann eine einminütige Inkubation bei 94 °C mit anschließendem raschem Abkühlen in Eiswasser durchgeführt werden, um RNA, die doppelsträngige Bereiche ausbildet, zu denaturieren. Der Ansatz wird 45 Minuten bei 42 °C inkubiert. Anschließend wird die Reaktion durch fünfminütige Inkubation bei 94 °C (Inaktivierung der reversen Transkriptase) gestoppt.

C.2.3.3 'C. therm. Polymerase'

Zu der Herkunft dieser Reversen Transkriptase wird vom Hersteller keine Angabe gemacht. Die Vorteile des Enzyms sind spezifische Bindung der Primer und wenig Probleme mit Sekundärstrukturen der Matrize durch hohe Reaktionstemperatur, keine negative Beeinflussung einer anschließenden PCR durch Manganionen und keine Inhibierung des Enzyms durch hohe Konzentrationen von RNA.

Sofern es nicht im besonderen beschrieben ist, wird als Matrize je Reaktion 5 µg Gesamt-RNA oder 0,1 µg poly(A)⁺-RNA eingesetzt.

'C. therm. Polymerase':	4 U/µl	
Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	0,1 mM	EDTA
	10 mM	β-Mercaptoethanol
	50 mM	KCl
	0,5 % (w/v)	Thesit
	0,5 % (v/v)	Nonidet P-40 (NP-40)
	10 mM	Tris-HCl, pH 8,5 (4°C)

Reaktionsansatz mit dem 'C. therm. Polymerase for reverse transkription / Two-Step Kit':

<u>je Reaktionsansatz:</u>	
RNA-Matrize + DEPC-dH ₂ O	10,3 µl
5x reverse Transkriptase-Puffer (mit 25 mM MgCl ₂ und 10 % DMSO)	4,0 µl
dNTP-Mix (je dNTP: 2,5 mM)	1,6 µl
Primer (20 µM)	1,0 µl
DTT	1,0 µl
DMSO	0,6 µl
'C. therm. Polymerase' (4 U/µl)	1,5 µl
	20,0 µl

Als Negativ-Kontrollen dienen Ansätze ohne RNA-Matrize. Eine zusätzliche Negativkontrolle wird mit RNA-Matrize, aber ohne reverse Transkriptase angesetzt. Positivkontrollen werden gegebenenfalls mit Matrizen durchgeführt, die zuvor für die RT-PCR unter diesen Bedingungen getestet wurden. Der Ansatz wird fünf Minuten bei 70 °C und 45 Minuten bei 65 °C inkubiert. Danach wird das Enzym durch zweiminütige Inkubation bei 95 °C inaktiviert.

C.2.3.4 MuLV reverse Transkriptase

In dem RT-PCR-Kit von Perkin Elmer wird mit der Reversen Transkriptase aus dem Moloney Murine Leukemia Virus (MMLV-reverse Transkriptase) gearbeitet. Die Reaktion erfolgt, sofern nicht abweichend beschrieben, wie im Protokoll des Herstellers angegeben. Die cDNA-Synthese und die PCR werden in zwei Schritten im gleichen Reaktionsgefäß durchgeführt. Der PCR-Mix wird nach der reversen Transkriptase-Reaktion zu dem gesamten Ansatz zugegeben und die Inkubation durchgeführt.

Sofern es nicht anders beschrieben ist, wird als Matrize je Reaktion 5 µg Gesamt-RNA oder 0,1 µg poly(A)⁺-RNA eingesetzt.

MMLV-reverse Transkriptase:	4 U/µl	
Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	0,1 mM	EDTA
	1 mM	DTT
	0,01 % (v/v)	Nonidet P-40 (NP-40)
	20 mM	Tris-HCl, pH 7,5 (bei RT)

Reaktionsansatz für die Reverse-Transkriptase-Reaktion:

<u>je Reaktionsansatz:</u>	
RNA-Matrize + DEPC-dH ₂ O	1,5 µl
10x PCR Puffer II	1,0 µl
25 mM MgCl ₂	2,0 µl
dNTP-Mix (je dNTP: 2,5 mM)	4,0 µl
Stromabwärts-Primer (20 µM)	0,5 µl
RNase Inhibitor (20 U/µl)	0,5 µl
MuLV reverse Transkriptase (50 U/µl)	0,5 µl
	10,0 µl

Die Inkubation wird bei 42 °C über 15 Minuten durchgeführt mit anschließender Inaktivierung der Reversen Transkriptase durch zehnminütige Inkubation bei 95 °C.

AmpliTaQ DNA-Polymerase:	5 U/μl	
Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	0,1 mM	EDTA
	100 mM	KCl
	1 mM	DTT
	0,5 % (v/v)	Nonidet P-40 (NP-40)
	0,5 % (v/v)	Tween 20
	20 mM	Tris-HCl, pH 8,0 (bei RT)

Reaktionsansatz für die anschließende PCR-Reaktion:

je Reaktionsansatz:

(reverse Transkriptase-Ansatz	10 μl)
DEPC-dH ₂ O	33,25 μl
10x PCR Puffer II	4,0 μl
25 mM MgCl ₂	2,0 μl
Stromaufwärts-Primer (20 μM)	0,5 μl
AmpliTaQ DNA Polymerase (5 U/μl)	0,25 μl
	50,0 μl

Als Negativkontrollen dienen Ansätze ohne RNA-Matrize. Positivkontrollen werden gegebenenfalls mit RNAs durchgeführt, die zuvor für die reverse Transkriptase-PCR unter diesen Bedingungen getestet wurden. Die PCR Bedingungen werden durchgeführt wie unter C.2.2.2 beschrieben.

C.2.4 Anker-PCR (SLIC-Strategie)

Für die Analyse des 5'-Endes der mRNA muß die Sequenz mit Hilfe der PCR amplifiziert werden. Allerdings kann eine PCR nur durchgeführt werden, wenn beide Sequenzenden um das zu amplifizierende Stück bekannt sind. Die 5'-gelegene Sequenz ist jedoch unbekannt und somit kann ein komplementärer Primer nicht gefunden werden. Die Anker-PCR (anchor-PCR, A-PCR) wird verwendet, um das unbekannte 5'- oder 3'-Ende einer mRNA zu ermitteln. In dieser Arbeit wird die SLIC-Strategie¹³¹ für diese Methode angewandt (Abbildung 30). Die A-PCR umgeht das Problem, indem mit Hilfe von T4 RNA Ligase ein spezifischen Oligonukleotids an das 5'-Ende einer zuvor über RT-PCR synthetisierten cDNA ligiert wird. Anschließend kann das 5'-Ende der mRNA mit PCR über einen spezifischen Antisinn-Primer und einen Sinn-Primer, der komplementär zum Anker-Primer ist, durch Sequenzierung ermittelt werden. Bei sehr seltenen mRNAs ist eine nested PCR und eventuell eine zweite nested PCR erforderlich. Zu diesem Zweck werden drei Primer verwendet, die jeweils stückweise komplementär zum Anker-Primer sind.

Die cDNA-Synthese wird mit 1 μg poly(A)⁺-RNA und Genescript reverse Transkriptase (C.2.3.1) oder 'C. therm. Polymerase' (C.2.3.3) durchgeführt. In weiteren Ansätzen mit doppeltem Volumen und Genescript wurden 100 μg A549-Gesamt-RNA revers transkribiert. Anschließend wird die cDNA über CTAB-Fällung gereinigt (C.2.6.2). Als Antisinn-Primer werden bei der reversen Transkriptase-Reaktion mit Genescript der Primer HCII5'- und bei Verwendung von 'C. therm. Polymerase' der Primer Dag21 eingesetzt.

¹³¹ Edwards *et al.*, 1997

In den anschließenden PCRs werden die Oligonukleotide USC23, USC8, USC9 und USC12 als Antisinn-Primer verwendet. Die zum Anker-Primer komplementären Sinn-Primer tragen die Bezeichnung A5'-1, A5'-2 und A5'-3.

T4 RNA Ligase:	20 U/μl	
Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	0,1 mM	EDTA
	10 mM	β-Mercaptoethanol
	50 mM	KCl
	1 mM	DTT
	10 mM	Tris-HCl, pH 7,4

1x T4 RNA Ligase-Reakt.-Puffer:	1 mM	ATP
	10 mM	MgCl ₂
	10 mM	DTT
	50 mM	Tris-HCl, pH 7,8

je Reaktionsansatz:

ssDNA	10,0 μl
Primer A5'-NV (1 μM)	0,5 μl
10x T4 RNA Ligase Puffer	1,7 μl
30 % PEG 4000	3,8 μl
T4 RNA Ligase (20 U/μl)	1,0 μl
	17,0 μl

Als Negativkontrollen dienen reverse Transkriptase-Ansätze, die ohne RNA-Matrize durchgeführt werden. Der Reaktionsansatz wird mit Öl überschichtet und 48 Stunden bei Raumtemperatur (RT) inkubiert. Anschließend erfolgt die Inaktivierung der T4-RNA-Ligase durch fünf Minuten Inkubation bei 94 °C. Die nachfolgende PCR und die nested PCR werden unter den Standardbedingungen durchgeführt (C.2.2.2 und C.2.2.3). In die erste PCR (50 μl Gesamt-Volumen) werden 1/10 Volumenanteil des Ligase-Ansatzes als Matrize gegeben.

C.2.5 Tailing-PCR

Die Tailing-PCR¹³² (T-PCR) wird ebenfalls verwendet, um das unbekanntende Ende einer mRNA zu ermitteln. Das Problem der notwendigen Amplifikation mit einer PCR wird bei der A-PCR (C.2.4) beschrieben. Die T-PCR umgeht das Problem, indem mit Hilfe von terminaler Desoxynukleotidyl-Transferase eine Vielzahl von identischen Nukleotiden an das 5'-Ende einer zuvor über RT-PCR synthetisierten cDNA angehängt werden. Anschließend kann das 5'-Ende der mRNA mit PCR über einen spezifischen Antisinn-Primer und einen Sinn-Primer, der komplementäre Nukleotide zu den angehängten Nukleotiden am 3'-Ende enthält, ermittelt werden. Bei sehr seltenen mRNA-Transkripten ist eine (semi-)nested PCR und eventuell eine zweite (semi-)nested PCR erforderlich.

Die cDNA Synthese wird mit 0,7 μg poly(A)⁺-RNA und Omniscript reverse Transkriptase (C.2.3.2) oder C. therm. Polymerase (C.2.3.3) durchgeführt (Matrizen in diesem Fall: 20 μg Gesamt-RNA aus HepG2-Zellen, 30 μg Gesamt-RNA aus A549-Zellen, 42 μg Gesamt-RNA aus Lunge, oder 1 μg Leber-poly(A)⁺-RNA).

¹³² Bertoli, 1997

Anschließend wird die cDNA über CTAB-Fällung gereinigt (C.2.6.2) und eine Reihe von G-Nukleotiden angehängt ('tailing'). Als Antisinn-Primer werden bei der reversen Transkriptase-Reaktion mit 'C. therm. Polymerase' Dag21 oder USC23 verwendet, und USC12 wird zusammen mit der Omniscript Polymerase eingesetzt. In den anschließenden PCRs werden die Primer USC8, USC9, USC12, USC20, USC24 und USC26 verwendet. Der zum poly(G) komplementäre Sinn-Primer, mit dem die anschließende PCR durchgeführt wird, trägt die Bezeichnung A5'-Tail.

Terminale Desoxyribonukleotidyl Transferase (TdTase): 20 U/μl

Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	1 mM	β-Mercaptoethanol
	50 mM	Kaliumphosphatpuffer, pH 7,4

5x TdTase-Reaktionspuffer:	0,5 mg/ml	BSA
	5 mM	CoCl ₂
	0,5 mM	DTT
	500 mM	Kakodylatpuffer, pH 6,8

Reaktionsansatz für die terminale Transferase-Reaktion:

<u>je Reaktionsansatz:</u>	
ss cDNA + dH ₂ O	15,25 μl
dGTP (100 μM)	3,75 μl
5x TdTase Puffer	5,0 μl
terminale Transferase (20 U/μl)	1,0 μl
	<u>25,0 μl</u>

Als Negativkontrollen dienen hitzeinaktivierte Ansätze aus der reversen Transkriptase-Reaktion. Ebenfalls werden Ansätze, denen keine Matrize zugegeben wird, als Negativkontrollen verwendet. Der Reaktionsansatz wird eine Stunde bei 37 °C inkubiert. Anschließend erfolgt die Inaktivierung der Terminalen Transferase durch zweiminütige Inkubation bei 94 °C. Die nachfolgende PCR und nested PCR wird unter den Standardbedingungen durchgeführt (C.2.2.2 und C.2.2.3). In die erste PCR (50 μl Gesamtvolumen) werden 1/50 Volumenanteil des Ansatzes eingesetzt.

C.2.6 Techniken zur Analyse von DNA und RNA

C.2.6.1 Agarosegelelektrophorese von DNA¹³³

Die Agarosegelelektrophorese trennt die aufgrund ihrer Phosphatgruppen bei neutralem pH-Wert negativ geladenen DNA-Moleküle in einem elektrischen Feld auf. DNA-Fragmente wandern je nach Größe und Konformation mit unterschiedlicher Geschwindigkeit durch die Agarosematrix zur Anode. Durch Vergleich mit einem DNA-Längenstandard können Konzentration und Größe bestimmt werden. Agarose wird aus Rotalgen isoliert und ist ein lineares Polymer aus dem Dimer D-Galaktose und 3,6-Anhydro-L-Galaktose.

¹³³ Southern, 1979

Trennbereiche

2 %- Agarose-Gel: 2 % (w/v) Agarose in 1x TAE-Puffer; Trennung von 0,1-2 kB

1 %- Agarose-Gel: 1 % (w/v) Agarose in 1x TAE-Puffer; Trennung von 0,4-7 kB

Elektrophorese

Die Agarose wird mit TAE in einem Mikrowellenherd aufgeköcht und heiß in eine Elektrophoresekammer mit einem Probenkamm gegossen. Das Gel erhärtet innerhalb einer halben Stunde. Anschließend wird das Gel in der Elektrophoresekammer mit TAE-Puffer überschichtet (submarine Technik).

Zum Nachweis von Plasmid-DNA, Restriktionsfragmenten oder PCR-Produkten werden 10 µl Probe mit 1 µl 10x Elektrophorese-Auftragspuffer vermischt und in die Taschen des Agarosegels aufgetragen (Geldicke ca. 0,5 cm). Zur Größenbestimmung werden mindestens in einer Spur 0,25 µg DNA-Längenstandard aufgetragen. Die Elektrophorese läuft für ca. eine halbe Stunde bei einer Feldstärke von ca. 20 V/cm in TAE-Puffer (120-150 V, 250-400 mA). Anschließend wird das Gel für zehn Minuten in ein Färbebad mit Ethidiumbromid (EtBr) gelegt. Der Farbstoff interkaliert in DNA und RNA. Da EtBr-DNA-Komplexe bei UV-Bestrahlung mit einer Wellenlänge von 312 nm fluoreszieren, können so die einzelnen DNA-Fragmente sichtbar gemacht werden. Das Gel wird anschließend für fünf Minuten in dH₂O entfärbt. Zur Dokumentation wird das Agarosegel auf einer UV-Durchlichtplatte bei 312 nm fotografiert.

Die DNA-Fragment-Größen werden durch Vergleich mit dem DNA-Längenstandard ('kB-Leiter', 'Marker VIII' oder '100 Bp-Leiter') bestimmt. Mit Hilfe der WinCam Version 2.2 Software kann die Konzentration densitometrisch abgeschätzt werden.

C.2.6.2 Reinigung von DNA

Fällung mit Hexadecyltrimethylammoniumbromid¹³²

(Reinigung von einzelsträngiger und doppelsträngiger DNA und RNA von Proteinen und kurzen Oligomeren; Umpufferung von DNA-Lösungen; Konzentrierung von DNA-Lösungen) Zu der Nukleinsäure-Lösung werden 1/10 Vol. 3,5 M NaCl und 1/10 Vol. 5 % (v/v) CTAB gegeben und jeweils gut gemischt. Nach einer 30minütigen Zentrifugation bei 13000 UpM wird das Pellet in 100 µl 1,2 M NaCl und 270 µl 96 % EtOH aufgenommen und erneut 30 Minuten lang zentrifugiert. Anschließend wird das Pellet mit 70 % EtOH gewaschen. Das Pellet wird in 10 µl dH₂O aufgenommen und 5 Minuten bei 75 °C inkubiert.

Ethanol- / Isopropanolfällung

(Reinigung von einzelsträngiger und doppelsträngiger DNA und RNA von Proteinen, Phenol und kurzen Oligomeren; Umpufferung von DNA-Lösungen; Konzentrierung von DNA-Lösungen)

Um die Konzentration von DNA-Lösungen zu erhöhen oder um einen Pufferwechsel durchzuführen wird die DNA mit EtOH präzipitiert. Dazu wird die Salzkonzentration der wässrigen DNA-Lösung durch Zugabe von 1/10 Vol. 5 M NaCl oder 3 M Na-Acetat-Lösung (pH 5,2) stark erhöht. Dann erfolgt die Zugabe und Mischen mit zwei Vol. 96 % EtOH oder ein Vol. Isopropanol. Die DNA wird durch Inkubation bei -20 °C über 30-60 Minuten gefällt. Danach wird die Probe bei 13000 UpM abzentrifugiert, das Pellet mit 70 % EtOH gewaschen und für zehn Minuten bei 60 °C getrocknet. Dann kann das DNA-Pellet in der gewünschten Menge dH₂O oder Puffer aufgenommen werden.

Phenolextraktion

(Reinigung von einzelsträngiger und doppelsträngiger DNA und RNA von Proteinen)

Die Nukleinsäurelösung wird mit 1/3 Vol. Roti-Phenol versetzt, gevortext und fünf Minuten bei 13000 UpM abzentrifugiert. Der DNA-haltige Überstand (wäßrige Phase) wird abgezogen und mit Ethanol gefällt. Soll die gereinigte DNA für Sequenzierungen verwendet werden, findet anschließend eine weitere Reinigung (Entfernung minimaler Phenolreste) über Microcon-50 statt (s. u.).

Reinigung von DNA mit dem PCR-Purification-Kit (QIAquick)

(Reinigung von einzelsträngiger und doppelsträngiger DNA und RNA von Proteinen, Phenol und kurzen Oligomeren)

Die Reinigung mit dem PCR-Purification-Kit wird entsprechend dem Protokoll des Herstellers durchgeführt¹³⁴. Die Aufnahme der DNA erfolgt immer in 35 µl Elutions-Puffer, mit dem die Säule vor der Zentrifugation eine Minute bei RT inkubiert wird.

Reinigung mit Microcon-50 (Amicon)

(Reinigung von einzelsträngiger und doppelsträngiger DNA und RNA von Phenol und kurzen Oligonukleotiden)

Die Nukleinsäure-Lösung wird nach Angaben des Herstellers auf die Säule gegeben und bei 13000 UpM 1-5 Minuten zentrifugiert, bis nur noch ca. 10 µl Flüssigkeit oberhalb der Membran vorhanden sind. Dann wird zweimal mit 200 µl dH₂O gewaschen, wobei ebenfalls ein geringer Flüssigkeitsfilm erhalten bleiben muß, um die Membran nicht zu zerstören. Anschließend werden 15 µl dH₂O direkt auf die Membran pipettiert und die Säule eine Minute bei RT inkubiert. Abschließend wird die Säule in einem sauberen Reaktionsgefäß invertiert und zwei Minuten lang abzentrifugiert.

C.2.6.3 Konzentrationsbestimmung von DNA und RNA

Konzentrationsbestimmung über Agarosegele

Die Konzentration einer DNA-Lösung kann durch Vergleich mit einer DNA-Lösung bekannter Konzentration im Agarosegel abgeschätzt werden. Es werden Verdünnungsreihen angelegt und die Proben auf ein Minigel aufgetragen. Nach dem Lauf wird das Gel mit Ethidiumbromid gefärbt und fotografiert. Je nach Intensität der Banden im Vergleich zu den Banden mit bekannter Konzentration kann nun die Konzentration der DNA-Lösung abgeschätzt werden.

Eine genauere densitometrische Konzentrationsbestimmung ermöglicht die WinCam (Version 2.2) Software. Die Bestimmung verläuft nach Angaben des Herstellers.

Konzentrationsbestimmung über Messung der Absorption bei einer Wellenlänge von 260 nm

Eine weitere Möglichkeit die Konzentration von DNA oder RNA in einer Lösung zu bestimmen, ist die Messung der optischen Dichte. Nukleinsäuren haben ein Absorptionsmaximum bei 260 nm. Zusätzlich kann die relative Verunreinigung der Lösung mit Proteinen bestimmt werden. Proteine absorbieren UV-Licht maximal bei einer Wellenlänge von 280 nm. Reine DNA weist ein oD_{260}/oD_{280} -Verhältnis von bis zu 2 auf. Je mehr Proteine enthalten sind, desto niedriger ist das Verhältnis. Bei reinen Proteinen entspricht 1 oD_{280} etwa 1,8 µg/µl Protein.

¹³⁴ QIAgen, 1997

Für reine Nukleinsäuren gilt: 1 oD₂₆₀ entspricht einer Konzentration bei:

dsDNA:	50 µg/ml	Oligonukleotide:	20-30 µg/ml
ssDNA:	33 µg/ml	RNA:	40 µg/ml

C.2.6.4 Spaltung von DNA mit Restriktionsenzymen¹³⁵

Restriktionsendonukleasen vom Typ II erkennen eine definierte, palindrome DNA-Zielsequenz von 4 bis 8 Nukleotiden. Innerhalb dieser Sequenz wird spezifisch eine Phosphodiester-Bindung hydrolysiert, so daß ein Ende mit einer 5'-Phosphat-Gruppe und das andere Ende mit einer 3'-Hydroxyl-Gruppe an der Desoxyribose entstehen. Die Restriktion kann stattfinden, sofern die DNA nicht durch ein entsprechendes Modifikationsenzym methyliert ist. Manche Enzyme schneiden den DNA-Doppelstrang versetzt, so daß überstehende Enden entstehen ('sticky ends'), während andere an der gleichen Stelle in den komplementären Strängen schneiden (ergibt 'blunt ends'). Um ihre spezifische, selektive Aktivität zu erhalten, müssen sie in dem jeweils geeigneten Puffer mit entsprechend hoher Salzkonzentration eingesetzt werden. Diese Bedingungen werden mit den 10x Restriktionspuffern von Boehringer erreicht. Bei Mehrfachspaltungen wird normalerweise der Puffer mit der niedrigeren Salzkonzentration verwendet, es muß aber überprüft werden, ob die Restriktionsenzyme eine unspezifische Spaltungsaktivität ('Star-Aktivität') unter diesen Bedingungen aufweisen.

Reaktionsansatz einer Restriktionsspaltung von DNA:

<u>je Reaktionsansatz:</u>	
DNA + dH ₂ O	X µl
10x Spaltungs-Puffer (Tabelle 5)	1/10 Vol.
RNase-Lösung (10 mg/ml)	1/10 Vol.
<u>Restriktionsenzym</u>	Y µl
	1 Vol. Gesamtansatz

Der Ansatz wird gemischt, abzentrifugiert und für 1-1,5 h bei dem Temperaturoptimum des Restriktionsenzym im Wasserbad inkubiert. Anschließend werden die Restriktionsenzyme durch Hitze (15minütige Inkubation bei 65 °C und Einfrieren), Phenolextraktion, oder durch Zugabe von komplexierenden Agenzien (SDS und EDTA im Gel-Ladepuffer) inaktiviert.

C.2.6.5 Ligation von DNA-Restriktionsfragmenten

Die Spaltungsansätze mit der in Restriktionsspaltungen geschnittenen Vektor- und Insert-DNA werden zur Enzyminaktivierung und Lagerung eingefroren. Nach dem Auftauen werden sie im geeigneten Verhältnis gemischt, wobei die Ligase-Reaktion am besten in einem kleinen Reaktionsvolumen mit hoher DNA-Konzentration abläuft (0,1-1 µg/µl). Die Insert-DNA sollte möglichst im Überschuß vorliegen. Dies unterdrückt auch eine mögliche Religierung des Vektors. Sofern die Enden religieren können, oder schwierig zu klonierende Inserts vorliegen, wird eine Dephosphorylierung des Vektors durchgeführt (C.2.6.7). Die Ligase verbindet die Enden zu kovalent-gebundenen DNA-Strängen. Zuerst erfolgt eine Inkubation der DNA-Lösung für zehn Minuten bei 65 °C im Wasserbad, um 'sticky ends' zu lösen und zur Deaktivierung einiger Restriktionsenzyme (evtl. muß die DNA mit einer Phenolextraktion von den Enzymen gereinigt werden, z.B. bei *Bam*HI).

¹³⁵ Smith & Birnstiel, 1976

Eine Ligation wird üblicherweise in einem Ansatz mit 25-40 µl Volumen durchgeführt. Der Ansatz wird über Nacht in einem Thermocycler jeweils 30 Sekunden abwechselnd bei 10 °C und 27 °C inkubiert.¹³⁶ Anschließend wird eine Hitzeinaktivierung (15 Minuten bei 65 °C) durchgeführt. Danach kann die Transformation durchgeführt werden.

Je Reaktion für eine 'sticky-end'-Ligation:			je Reaktion für eine 'blunt-end'-Ligation:		
DNA-Spaltungsansätze	X	µl	DNA-Spaltungsansätze	X	µl
ATP-Lösung (10 mM)	1/10	Vol.	ATP-Lösung (10 mM)	1/100	Vol.
DTT-Lösung (100 mM)	1/10	Vol.	DTT-Lösung (100 mM)	1/10	Vol.
BSA-Lösung (10 mg/ml)	1/20	Vol.	BSA-Lösung (10 mg/ml)	1/20	Vol.
T4-DNA-Ligase	0,01	U	T4-DNA-Ligase	1	U
dH ₂ O	Y	Vol.	dH ₂ O	Y	Vol.
Gesamtansatz	1	Vol.	Gesamtansatz	1	Vol.

C.2.6.6 Isolierung von Restriktionsfragmenten

In einigen Experimenten werden nur bestimmte, definierte DNA-Fragmente aus einer Restriktionsspaltung von DNA benötigt. Dazu werden die Fragmente aus einem Spaltungsansatz in einem Agarosegel aufgetrennt. Eine optimale Auftrennung der gewünschten Fragmente kann durch Veränderung der Agarosekonzentration im Gel erreicht werden. Das Gel wird mit EtBr gefärbt und die gewünschten Restriktionsfragmente werden unter UV-Licht mit einer abgeschnittenen blauen Eppendorf-Spitze aus dem Gel ausgestochen.¹³⁷

Fragmente für Klonierungen

Die Isolierung der DNA-Fragmente aus dem Agarosegel erfolgt über einen Grobfilter (Zurückhaltung der Gelmatrix) und ein Molekularsieb (Microcon 50, DNA-Reinigung von kleinen Molekülen, C.2.6.2) nach Angaben des Herstellers (Gel-Extraktions-System von Amicon).

Fragmente als Matrizen für PCR

Der Gelblock wird in 500 µl dH₂O für fünf Minuten bei 95 °C inkubiert. 0,5 µl der gevortexten Lösung wird anschließend als Matrize in eine PCR-Reaktion eingesetzt.

C.2.6.7 Dephosphorylierung von DNA

Dephosphorylierung mit Alkalischer Phosphatase aus *P. borealis*

Die alkalische Phosphatase aus der arktischen Garnele *P. borealis* (shrimp alkaline phosphatase, SAP) ist eine Orthophosphorsäuremonoesterphosphohydrolase, EC 3.1.3.1. SAP erkennt DNA und RNA als Substrat und dephosphoryliert das 5'-Ende von RNA und DNA an 5'-überhängenden, 5'-rezessiven und an glatten Enden. Eine vollständige und irreversible Inaktivierung wird durch 15 Minuten Inkubation bei 65 °C bewirkt.¹³⁸ Der Ansatz wird 10 Minuten bei 37 °C inkubiert und anschließend zur Inaktivierung 15 Minuten bei 65 °C erhitzt.

¹³⁶ Lund *et al.*, 1996

¹³⁷ Thomson & Compton, 1998

¹³⁸ Maxam & Gilbert, 1980; Moessner *et al.*, 1980

Lagerung bei -20 °C in:	50 % (v/v)	Glycerin
	1 mM	MgCl ₂
	0,1 mM	ZnCl ₂
	25 mM	Tris-HCl, pH 7,6

Dephosphorylierungspuffer (10x konz.):	1 mM	EDTA
	0,5 M	Tris-HCl, pH 8,5

Reaktionsansatz einer Dephosphorylierung:

<u>je Reaktion:</u>		
DNA + dH ₂ O		24,0 µl
SAP (1 U/µl)		3,0 µl
10x Dephosphorylierungspuffer		3,0 µl
		30,0 µl

Dephosphorylierung mit Alkalischer Phosphatase aus Kälberdarm

Die Reaktion wird ebenso durchgeführt, wie für die SAP, allerdings wird 10x Puffer 3 (Boehringer) als Reaktionspuffer verwendet. Eine Inaktivierung des Enzyms durch Hitze ist nur mit anschließender Phenolextraktion möglich. Aus diesem Grund wird eine Reinigung des Produkts über Phenolextraktion oder QIAquick-PCR-Purification durchgeführt.

C.2.6.8 DEPC-Behandlung von Puffern und Gefäßen

Um RNAsen auf Geräten und in dH₂O zu inaktivieren, wird eine DEPC-Behandlung durchgeführt. dH₂O wird 1/1000 Volumen DEPC zugesetzt und zwei Stunden gerührt. Danach wird die Lösung autoklaviert und das DEPC damit zerstört. Oberflächen von Gelkammern können mit 2 %-SDS-Lösung zur Inaktivierung von RNAsen behandelt werden.

C.2.6.9 Isolierung von Gesamt-RNA aus tierischen Zellen

Gesamt-RNA-Isolierungen aus Zellkulturen von HepG2-Zellen und A549-Zellen werden mit dem QIAGEN RNeasy Mini Kit nach den Angaben des Herstellers durchgeführt. Die RNA wird in dH₂O aufgenommen und auf eine Konzentration von 1 µg/µl eingestellt.

C.2.7 Anzucht und Lagerung von Zellen

C.2.7.1 Anzucht und Lagerung von *E. coli*-Stämmen

Herstellung der Kulturmedien

Die Medien werden bei 121 °C für 30 Minuten autoklaviert. Für Festmedien können Antibiotika (Endkonzentration: siehe Tabelle 14) hinzugegeben werden, sobald die Temperatur auf weniger als 60 °C gesunken ist. Das Medium wird anschließend unter sterilen Bedingungen in der Reinraumwerkbank in Petrischalen gegossen und kann mehrere Monate bei 4 °C gelagert werden.

Anzucht und Lagerung

Für die Anzucht der Bakterien werden 5 ml LB Medium mit einer Einzelkolonie beimpft und bei 37 °C und 150 UpM ü.N. inkubiert. Zur Gewinnung von größeren Mengen von Bakterien werden 50-100 ml LB Medium in einer 1000 ml-Flasche angeimpft und im Luftschüttler bei 37 °C bis zu einer OD_{600} von 0,4 oder 0,9 inkubiert. Die Bebrütung von Bakterienausstrichen auf Festmedien (LBG) erfolgt bei 37 °C ü.N. im Wärmeschrank. Die bewachsenen Platten können mehrere Monate bei 4 °C gelagert werden.

Messung der optischen Dichte

Zur Überprüfung des Bakterienwachstums in Flüssigkultur wird die optische Dichte mit einem Spektralphotometer bestimmt. Dabei wird die optische Dichte von *E. coli* bei einer Wellenlänge von 600 nm gemessen. Als Referenz dient das unbeimpfte Medium.

C.2.7.2 Anzucht und Lagerung von eukaryotischen Zelllinien

In dieser Arbeit werden HepG2-Zellen¹³⁹ zur RNA-Isolierung, Transfektion und Promotoranalyse und A549-Zellen zur RNA-Isolierung verwendet.

HepG2-Zellen leiten sich von einem Leberhepatom ab. Die HepG2-Zellen der Passage 89 wurden im Oktober 1997 von der Firma ECACC (ECACC Nr.: 85011430, CB NR.: 2551) bezogen. In dieser Arbeit wurden Zellen ab der 99. Passage für die Analysen verwendet.

Die A549-Zellen¹⁴⁰ der Passage 86 wurden im August 1997 von der Firma ECACC (ECACC Nr.: 86012804, CB NR.: 2493) bezogen. A549-Zellen sind menschliche Lungenkarzinomzellen. Die RNA wurde aus Zellen unterschiedlicher Passagen gewonnen.

100x Penicillin-Streptomycin:	0,63 g	Penicillin G (Sigma P3032)
	1,0 g	Streptomycinsulfat (Merck 722K779117)
	100 ml	dH ₂ O
		Die Lösung wird sterilfiltriert und bei -20 °C gelagert.
Medium für HepG2-Zellen:	10 g	DMEM (Dulbecco's modified Eagle medium; Gibco/BRL Nr. 074-01600P; Lagerung bei -20°C)
	3,7 g	NaHCO ₃ (bei 10 % CO ₂)
	100 ml	Calf serum (inaktiviert durch 30minütige Inkubation bei 56 °C) (NCS; Flow 29-121-54 Lot 29121109)
	10 ml	100x Penicillin-Streptomycin
	0,47 g	L-Glutamin
	890 ml	dH ₂ O
		Das Medium wird sterilfiltriert und bei 4 °C gelagert.
Medium für A549-Zellen:	10 g	DMEM (Dulbecco's modified Eagle medium; Gibco/BRL Nr. 074-01600P; Lagerung bei -20°C)
	3,7 g	NaHCO ₃ (bei 10 % CO ₂)
	100 ml	Calf serum (inaktiviert durch 30minütige Inkubation bei 56 °C) (NCS; Flow 29-121-54 Lot 29121109)
	10 ml	100x Penicillin-Streptomycin
	0,47 g	L-Glutamin
	3,0 g	D-Glukose
	890 ml	dH ₂ O
		Das Medium wird sterilfiltriert und bei 4 °C gelagert.

¹³⁹ Aden *et al.*, 1979; Knowles *et al.*, 1980

¹⁴⁰ Giard *et al.*, 1973; Lieber *et al.*, 1976

Einfrieren und Lagerung von adhärennten eukaryotischen Zellen

Es werden $5-6 \times 10^6$ Zellen (von je einer 10 cm Platte) eingefroren. Dazu wird von einer konfluent bewachsenen Kulturschale oder T-Flasche das Medium abgezogen, die Zellen mit PBS gewaschen und trypsiniert (s.u.). Nach dem Ablösen der Zellen wird die Reaktion mit 10 ml Medium gestoppt, und die Zellen werden bei 800 UpM fünf Minuten abzentrifugiert. Nach vollständigem Absaugen des Mediums werden die sedimentierten Zellen in 1,5 ml Vollmedium mit 10 % DMSO resuspendiert und mit einer sterilen 1 ml Pipette in Kryoröhrchen überführt. Die Zellen werden möglichst schonend eingefroren, indem sie zunächst zwei Stunden auf -20 °C abgekühlt werden, dann über Nacht bei -80 °C inkubiert und anschließend in der Gasphase von flüssigem Stickstoff gelagert werden (-196 °C).

Trypsinieren und Passagieren von adhärennten eukaryotischen Zellen

Die verwendeten Lösungen werden auf 37 °C temperiert.

Von einer konfluenten Monolayerkultur wird das alte Medium abgesaugt. Dann wird die Kultur mit der gleichen Menge an PBS gewaschen und danach soviel Trypsin-Lösung (Sigma) zugegeben, daß die Zellen gerade bedeckt sind. Die Zellen sollten nach zehn Minuten Inkubation bei 37 °C unter mikroskopischer Kontrolle und leichtem Schwenken abgelöst und abgerundet sein. Anschließend wird frisches Medium (je nach Zellen, s.u.) zugegeben. Die Kultur wird 1:2 bis 1:3 in neue Kulturschalen ausverdünt.

C.2.8 Plasmidisolierung

C.2.8.1 'S-Lyse'¹⁴¹ (Plasmid-DNA für Restriktionsspaltungen und Klonierungen)

Diese Methode dient dazu, in kurzer Zeit kleine Mengen an Plasmid-DNA mit relativ hohem Reinheitsgrad zu isolieren. Durch Lysozym- und Triton X-100-Behandlung wird die Bakterienzellwand und Membran teilweise durchlässig. Die chromosomale DNA verbleibt mit den restlichen größeren Zellbestandteilen in der Zelle.

1,5 ml der ü.N.-Kultur werden 1 Minute lang in einem Eppendorfgefäß abzentrifugiert und der Überstand quantitativ abgezogen. Das Pellet wird gevortext, 50 µl Lysozym-Lösung zugegeben und wiederum kurz gevortext und fünf Minuten bei RT inkubiert. Nach Zugabe von 0,8 ml Triton X-Lösung und einminütiger Inkubation bei 100 °C im Wasserbad wird der Ansatz für fünf Minuten auf Eis gestellt. Dann erfolgt eine Zentrifugation für 30 Minuten bei 13000 UpM. Der Überstand wird zweimal mit 300 µl Phenol extrahiert und fünf Minuten bei 13000 UpM zentrifugiert. 700 µl des Überstandes werden in ein neues Eppendorf-Gefäß gegeben. Anschließend wird die Plasmid-DNA durch Fällung mit Isopropanol gereinigt. Das Pellet wird 10 Minuten bei 60 °C getrocknet und in 25-100 µl (je nach Kopienzahl des Replikons) dH₂O (oder TE) aufgenommen und zehn Minuten bei 65 °C gelöst. Soll die DNA für Sequenzierungen verwendet werden, müssen zuerst Phenolreste durch Aufreinigung über Microcon-50 (C.2.6.2) abgetrennt werden.

¹⁴¹ Priefer, 1984

C.2.8.2 'AS-Lyse' (Plasmid-DNA für Sequenzierungen)¹⁴²

Eine *E. coli*-Kultur wird flächig auf eine Agarplatte mit Selektivmedium ausgestrichen und ü.N. inkubiert. 1/3 der Kultur wird in ein Eppendorf-Reaktionsgefäß überführt und in 300 µl ASL-Lösung 1* vollständig resuspendiert. Nach fünf Minuten Inkubation wird ASL-Lösung 2* zugegeben und beide Lösungen bis zum Aufklaren gut gemischt. Danach werden 350 µl ASL-Lösung 3 zugegeben und der Ansatz fünf Minuten auf Eis inkubiert. Nach einer fünfminütigen Zentrifugation bei 13000 UpM wird der Überstand in ein neues Gefäß überführt und die DNA mit 300 µl ASL-Lösung 4 für 30 Minuten auf Eis gefällt. Anschließend wird das nach 10 Minuten Zentrifugation bei 13000 UpM erhaltene Pellet mit 70 % und 100 % Ethanol gewaschen, bei RT getrocknet und in 100 µl dH₂O aufgenommen. Enthält die DNA noch zu viele Verunreinigungen, oder soll die DNA für Restriktionsspaltung oder Ligation verwendet werden, müssen zuerst restliche enthaltene Proteine durch Phenol-Extraktion (C.2.6.2) entfernt werden.

C.2.9 Klonierung von Plasmid-DNA

Spaltungsansätze mit Plasmid-DNA, die mit den gleichen Restriktionsenzyme geschnitten sind, können nach Enzyminaktivierung (Einfrieren und 10 Minuten erhitzen auf 65 °C, oder durch Phenolextraktion) in einem geeigneten Verhältnis vereinigt und gemischt werden. Die nach einer Ligation entstehenden Plasmide werden in *E. coli* transformiert. Die Bakterien mit den richtigen Plasmiden werden über unterschiedliche Selektionsverfahren identifiziert. Dabei gibt es die unten aufgeführten Strategien, um die gewünschten Klone zu erhalten und identifizieren zu können. Anschließend läßt sich über Plasmidisolierung, eine geeignete Restriktionsspaltung und Gelelektrophorese nachweisen, ob das Plasmid die gewünschte Insertion enthält.

C.2.9.1 Insertionsinaktivierung eines Antibiotika-Resistenzgens

Bei dieser Methode wird die Antibiotikaresistenz dadurch zerstört, indem in die (geschnittene) Sequenz des Antibiotika-Resistenzgens die Fremd-DNA inseriert wird. *E. coli*-Bakterien werden mit der Plasmid-DNA transformiert und auf einem antibiotikahaltigen Medium, das nach einer intakten Resistenz auf dem Plasmid gewählt wird, angezogen. Anschließend werden Einzelkolonien parallel auf beide Antibiotikaresistenzen getestet. Plasmid-DNA wird aus den Zellen isoliert, die sensitiv gegen das inaktivierte Antibiotikum sind.

C.2.9.2 Insertionsinaktivierung des lacZ'-Gens

Hier wird die Fähigkeit zur α -Komplementation bestimmter *lacZ'* deletierter *E. coli*-Stämme ausgeschaltet, indem in die geschnittene Sequenz dieses Gens (meist in eine singuläre Schnittstelle) ein anderes Fragment inseriert wird. Nach der Transformation und Anzucht der Bakterien auf antibiotikahaltigem EMB-Medium der Transformanten wird die Plasmid-DNA aus den Zellen isoliert, die hell auf dem Agar erscheinen. Dunkle Kolonien zeigen durch Säureproduktion eine intakte β -Galaktosidase an und haben somit keine Inserts in den Plasmiden.¹⁴³

¹⁴² Good & Nazar, 1997

¹⁴³ Schlegel, 1992

C.2.9.3 Plasmid-Schnell-Charakterisierung anhand der Größe der ccc-Form¹⁴⁴

Einzelne, auf Platten mit Selektionsmedium gewachsene Kolonien werden mit einem sterilen Zahnstocher in 30 µl LBG Medium resuspendiert. Dann werden 30 µl 2x BPB-Lade/Lysepuffer zugegeben, fünf Sekunden gevortext und abzentrifugiert. 15 µl des Überstandes werden vorsichtig direkt in einer Agarose-Gelelektrophorese aufgetrennt. Der Vergleich mit einer Kolonie von Bakterien, die das Ausgangsplasmid enthalten, erlaubt Rückschlüsse auf die Größe eines eventuellen Inserts.

C.2.9.4 Kolonie-PCR

Plasmide und Inserts werden direkt von einer Bakterien-Einzelkolonie mit PCR überprüft (siehe C.2.2.4).

C.2.10 Transformation von Bakterienzellen¹⁴⁵

Das Einbringen von DNA in Zellen ist eine der wichtigsten Voraussetzungen in der molekularen Genetik bzw. Gentechnik. Die Methode, nackte DNA in Bakterien einzuschleusen, wird Transformation genannt.

Das Bakterium *E. coli* besitzt kein natürliches System zur Aufnahme von DNA. Durch die CaCl₂-Behandlung wird allerdings erreicht, daß diese Bakterien sehr effektiv Plasmid-DNA aufnehmen. Diese sogenannten kompetenten Zellen erreichen Transformationsraten von 10⁶ bis 10⁸ je nach transformiertem *E. coli*-Stamm, Größe und Reinheit der Plasmid-DNA. Um möglichst hohe Transformationsraten zu erreichen, muß mit den Zellen jeweils frisch die CaCl₂-Behandlung durchgeführt werden.

C.2.10.1 Herstellung der kompetenten Zellen

1 ml der ü.N. Kultur wird mit 50 ml LB versetzt und bis zu einer oD₆₀₀ von 0,4 oder 0,9 bei 37 °C inkubiert, dann in Zentrifugenröhrchen überführt und 10 Minuten bei 6000 UpM und 4 °C abzentrifugiert. Das Pellet wird in 25 ml kaltem 50 mM CaCl₂ gelöst. Dann erfolgt eine 30minütige Inkubation auf Eis, und danach werden durch 10minütige Zentrifugation bei 6000 UpM und 4 °C die Bakterien wieder pelletiert. Das Pellet wird in 5 ml kaltem 50 mM CaCl₂ aufgenommen und bei 0 °C gelagert.

C.2.10.2 Transformation

Bei einer Transformation werden bis zu 0,1 µg DNA an intakter Plasmid-DNA eingesetzt. Der Reaktionsansatz einer Transformation wird bei 0 °C wie folgt pipettiert:

je Reaktionsansatz:		
DNA	X	µl
β-Mercaptoethanol	1,0	µl
TCM	0,1	ml
kompetente Zellen	0,2	ml
	~ 0,3	ml

¹⁴⁴ Davis *et. al.*, 1999

¹⁴⁵ Mandel & Higa, 1970

Die Ansätze werden durch Invertieren gemischt und 30 Minuten auf Eis inkubiert. Danach erfolgt ein Hitzeschock der Zellen für drei Minuten bei 42,5 °C. Es werden 0,7 ml SOC zugegeben, invertiert und 30 Minuten bis zwei Stunden bei 37 °C zur Expression der Antibiotika-Resistenz inkubiert. Anschließend erfolgt das Ausplattieren der Ansätze auf Selektivmedium mit verschiedenen Verdünnungsstufen.

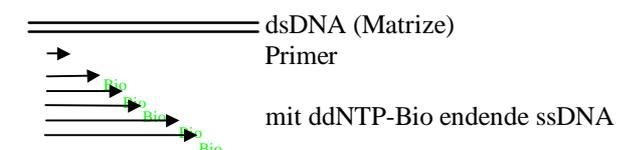
C.2.11 Sequenzierung

Die Didesoxymethode der DNA-Sequenzierung nach F. Sanger benutzt die vier 2',3'-Didesoxynukleosid-5'-Triphosphate ddGTP, ddATP, ddTTP und ddCTP, die in vier getrennten, DNA-Polymerase-katalysierten Reaktionen zur Termination der DNA-Synthese führen. Der Abbruch erfolgt, weil aufgrund der fehlenden 3'-Hydroxy-Gruppe eines eingebauten 2',3'-Didesoxynukleosid-Restes die Polymerisation nicht fortgesetzt werden kann.

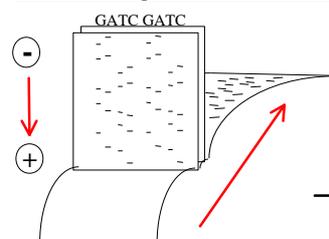
Die einzelsträngigen DNA-Fragmente werden mit einer thermostabilen DNA-Polymerase durch zyklische Denaturierung und erneute Anlagerung der Primer (Cycle Sequencing) synthetisiert. Das Verhältnis der ddNTPs zu den dNTPs ist so abgestimmt, daß je DNA-Strang statistisch ein ddNTP an jeder komplementären Basenposition der Sequenz eingebaut wird und es somit an jeder Position zum Abbruch kommen kann. Die so entstandenen DNA-Fragmente unterschiedlicher Länge werden bei 95 °C denaturiert und in Eiswasser abgekühlt, um ein Reannealing der DNA-Stränge zu verhindern. Im Anschluß werden die DNA-Fragmente parallel auf ein Polyacrylamidgel aufgetragen, elektrophoretisch getrennt und beim Austritt aus dem Gel auf eine Blotting-Membran übertragen.

Die verwendeten ddNTPs sind biotinmarkiert. Dadurch können die einzelnen Fragmente, die alle mit einem der ddNTPs enden, über ein (biotinbindendes) Streptavidin-Alkalische-Phosphatase-Konjugat detektiert werden. Zur kolorimetrischen Detektion wird das NBT/BCIP Färbesystem verwendet. Die alkalische Phosphatase spaltet die Phosphat-Gruppe von BCIP (5-Bromo-4-chloro-3-indolylphosphat) ab. Nach einer Oxidation fällt ein blauer Indigo-Farbstoff am Entstehungsort aus. NBT (Nitroblau-Tetrazoliumchlorid) verstärkt die Farbreaktion, indem es durch Aufnahme des entstandenen Protons zu einem blauen Diformazan reduziert wird. In dem nachfolgenden Schema sind die Schritte dargestellt:

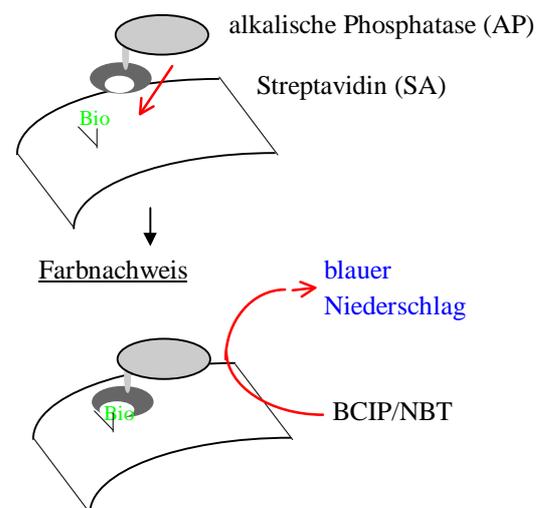
Sequenzierreaktionen



Auftrennung im PAA-Gel & Blotten



Hybridisierung mit SA-AP



C.2.11.1 Sequenzierreaktionen

Für die Sequenzierungsreaktionen wird das BioCycle Sequencing Kit von GATC verwendet. Hierbei wird anstelle eines dNTP-Mixes, der Desoxy-Guanosintri-phosphat enthält, ein dITP-Mix verwendet, der stattdessen Desoxy-Inosintri-phosphat enthält. Der Vorteil dieses Austausches ist ein stärkeres Signal und reduzierte Kompressionen, die durch Sekundärstrukturen entstehen.¹⁴⁶ Der Sequenator, ein Enzymkomplex-Puffergemisch, dient als Zusatz zur Verbesserung der Sequenzierergebnisse. Als DNA-Matrizen werden doppelsträngige Plasmid-DNA (isoliert mit AS-Lyse, C.2.8.2), oder direkt PCR-Produkte (gereinigt mit QIAGEN PCR Purification, C.2.6.2) verwendet.

Reaktionsansatz für eine Sequenzierung:

<u>je Reaktionsansatz:</u>	
DNA (1-5 µg) + dH ₂ O	10,6 µl
10x KlenTherm-Puffer (C.2.2.1)	2,1 µl
dITP-Mix	2,0 µl
Primer (20 µM)	0,5 µl
Sequenator	5,0 µl
<u>KlenThermase DNA-Polymerase (10 U/µl)</u>	<u>0,8 µl</u>
	21,0 µl

Das Reaktionsgemisch wird auf vier Ansätze zu je 5 µl verteilt und jeweils 1 µl von einer der ddNTP-Lösungen (aus dem BioCycle Sequencing Kit; vom Hersteller werden keine Konzentrationen angegeben) zugegeben.

Temperaturzyklen:

1 Zyklus:	94 °C	2 Minuten	Denaturierung
55 Zyklen:	94 °C	20 Sekunden	Denaturierung
	60 °C	10 Sekunden	Annealing
	65 °C	30 Sekunden	Elongation
1 Zyklus:	65 °C	4 Minuten	Elongation
Endlos-Schleife:	10 °C		

Die Sequenzierreaktionen werden durch Zugabe von 4 µl Stop-Mix beendet. Dieser Stop-Mix enthält Formamid, das Basenpaarungen zwischen den DNA-Strängen schwächt.

C.2.11.2 Gellauf für GATC 1500 DNA Sequencer ('Long Run System')

Die Vorbereitung der Glasplatten erfolgt analog zum GATC-Protokoll. Die Sequenzierreaktionen werden für drei Minuten bei 95 °C denaturiert und sofort auf Eiswasser abgekühlt. 3,5 µl der Ansätze werden vorsichtig auf das Sequenziergel in der Reihenfolge G-A-T-C aufgetragen und die Polyacrylamidgelelektrophorese wird gemäß des GATC-Protokolls durchgeführt. Abweichend davon werden 30 ml sterilfiltrierte und mittels Ultraschall entgaste Gellösung mit 135 µl 10 % APS-Lösung und 30 µl TEMED für das Polyacrylamidgel vermischt.

¹⁴⁶ Handbuch des GATC Sequencing Kits

Zu beachten ist, daß die untere Gelkante und die Geltaschen vollständig frei von Gelresten sind. Die Membran (ungeladene Nylonmembran von GATC) muß faltenfrei auf dem Förderband aufliegen.

Die Membran wird nach der Elektrophorese kurz getrocknet, mit UV im Stratalinker bestrahlt (254 nm, 100 mJ/cm², ca. 30 Sekunden) und anschließend 20 Minuten bei 80 °C inkubiert.

Die Detektion wird nach der Vorschrift des Boehringer DIG-Nucleic-Acid-Detection-Kit mit einem SA-AP-Konjugat durchgeführt. Abweichend von der Vorschrift werden 2,5 µl des SA-AP-Konjugates (0,7-1 U/µl) in 20 ml Puffer 2 eingesetzt. Die Equilibrierung vor der Färbung und die Färbung selbst werden bei 37 °C durchgeführt.

C.2.12 Southernblot

Das Ziel der Southernblot-Methode ist ein spezifischer Nachweis von DNA-Fragmenten. Die DNA wird in einer Agarose-Gelelektrophorese aufgetrennt und anschließend auf eine Membran übertragen. Der Nachweis findet über markierte Sonden statt. Diese Sonden bestehen aus komplementärer einzel- oder doppelsträngiger DNA, die radioaktiv- oder mit einem Antigen markiert sind. Letztere können immunologisch nachgewiesen werden.

Bei dem hier verwendeten System ist alkalische Phosphatase an einen Antikörper gegen Digoxigenin gekoppelt und bindet an die Digoxigenin-markierte Sonde (Boehringer).

Der Nachweis findet kolorimetrisch oder luminometrisch statt, indem die Membran, auf der die vom Gel übertragene und mit der Sonde markierte DNA mit einem geeigneten Substrat für die alkalische Phosphatase inkubiert wird und im Falle des Lumineszenz-Nachweises ein Röntgenfilm belichtet wird.

Sollen bestimmte, nur einfach vorhandene, kurze, chromosomale DNA-Fragmente von höheren Säugetieren detektiert werden, müssen mindestens 40 µg Gesamt-DNA auf ein Agarosegel aufgetragen werden.

C.2.12.1 Vorbereitung des Agarosegels

Das Agarosegel und die Restriktionsspaltungen werden wie unter C.2.6.1 und C.2.6.4 beschrieben durchgeführt. In die Restriktionsspaltung werden je Spaltungsansatz 40 µg chromosomale Maus-DNA (Clontech) und 5 µl Restriktionsenzym (100 U von *EcoRI*, *PvuII* und *XbaI*) eingesetzt und zwei Stunden inkubiert.

DNA-Fragmentierung

Nach der Elektrophorese wird die DNA im Gel unter leichtem Schwenken 30 Minuten in einer Mischung aus 0,5 N NaOH und 1,5 M NaCl inkubiert, um die DNA zum besseren Blotten zu fragmentieren. Danach wird das Gel mit dH₂O gespült und 30 Minuten in einer Lösung aus 1 M Tris-HCl und 1,5 M NaCl (pH 8,0) äquilibriert.

C.2.12.2 DNA-Transfer

Der Transfer erfolgt mittels eines inversen Kapillarblots von oben nach unten.¹⁴⁷ Dazu wird ein etwa 15 cm hoher Stapel aus Papiertüchern aufgeschichtet. Als oberstes werden drei dicke Whatman-Papiere (1 mm) und blasenfrei zwei dünne, in Transferpuffer getauchte Whatman-Papiere der gleichen Größe aufgelegt.

¹⁴⁷ Chomczynski, 1992

Die ungeladene Nylon-Membran Hybond N in der Größe des Gels wird zunächst eine Minute in DEPC-Wasser, danach eine Minute in 5x SSC getränkt und blasenfrei aufgelegt. Obenauf wird vorsichtig das Agarosegel gelegt. Mit einem weichen Bleistift werden die Geltaschen und Gelränder vorsichtig markiert. Wiederum werden drei dünne angefeuchtete Whatman-Papiere blasenfrei aufgelegt. Parafilm wird an die Ränder vom Gel angelegt, um eine Flüssigkeitsübertragung neben dem Gel zu verhindern. Als Brücke zwischen Gelstapel und erhöhtem Pufferreservoir werden zwei dünne, angefeuchtete Whatman-Papiere verwendet, auf die eine Glasplatte und 150-200 g Gewicht gelegt wird. Der Transfer wird für fünf Stunden mit 20x SSC als Transferpuffer durchgeführt.

C.2.12.3 Herstellung der Sonden

Herstellung der Sonde für den 3'-Bereich des Intron 1 im HCII-Gen der Maus

Eine PCR wird mit den Primern Mst2-Mst13 gemäß C.2.2.2 durchgeführt (Matrize: PCR-Fragment mit Primern Mst2-Mst86). Die Aufreinigung erfolgt mit dem QIAquick PCR Purification Kit (C.2.6.2) und die Konzentration der DNA wird für die Sonden-Synthese auf 10 ng/µl eingestellt.

Reaktionsansatz für die Herstellung der Sonde:

Reaktionsansatz für eine Sonde mit interner Markierung:

DNA (10 ng)	1,0	µl
Primer Mst2 (20 µM)	2,5	µl
Primer Mst13 (20 µM)	2,5	µl
10x KlenTherm-Puffer (C.2.2.1)	5,0	µl
10x PCR DIG Probe Labeling Mix (Boehringer)	5,0	µl
dH ₂ O	33,0	µl
<u>KlenTherm DNA-Polymerase (10 U/µl)</u>	<u>1,0</u>	<u>µl</u>
	50,0	µl

Temperaturzyklen:

1 Zyklus:	94 °C	2 Min	Denaturierung
35 Zyklen:	94 °C	25 Sek	Denaturierung
	62 °C	10 Sek	Annealing
	68 °C	1 Min	Elongation
1 Zyklus:	68 °C	2 Min	Elongation
Endlos-Schleife:	10 °C		

Nach der PCR erfolgt die Reinigung über das QIAquick PCR Purification Kit (C.2.6.2). Das PCR-Fragment hat eine Länge von 112 Bp. Eine erfolgreiche interne Markierung kann bei kleinen Sonden (100-200 Bp) über ein Fragment-Längenvergleich im Agarosegel festgestellt werden, da die markierten Produkte nach der Gelelektrophorese länger zu sein scheinen.

Herstellung der Sonde für den HCII-Exon 2-Bereich in Maus und Ratte

Eine PCR wird mit den Primern Dag50-Dag70 (Produkt: 567 Bp) gemäß C.2.2.2 durchgeführt (Matrize: Plasmid pWTBi1). Die Aufreinigung erfolgt mit dem QIAquick PCR Purification Kit (C.2.6.2) und die Konzentration der DNA wird für die Sonden-Synthese auf 10 ng/µl eingestellt. Die Synthese der Sonde wird in diesem Fall mit den gleichen Primern (Dag50-Dag70) und unter den Bedingungen, die für die Intron-Sonde verwendet werden, durchgeführt.

C.2.12.4 Fixierung und Hybridisierung

Fixierung der DNA

Nachdem Transfer werden die oberen Papierlagen entfernt und die Membran fünf Minuten in 5x SSC gewaschen. Das Gel wird im EtBr-Bad gefärbt und unter UV-Licht der eventuelle Verbleib von DNA-Fragmenten überprüft. Die Membran wird bei Raumtemperatur zwei Stunden auf Whatman-Papier getrocknet. Anschließend erfolgt eine UV-Bestrahlung im Stratalinker (254 nm, 10 mJ/cm², ca. 3 Sekunden) zur kovalenten Bindung der DNA an die Membran. Die Membran kann im Dunklen trocken bei RT gelagert werden.

Hybridisierung mit der Sonde

Die trockene Membran wird in eine Glasröhre überführt, und die folgenden Schritte werden in den Röhren durchgeführt. Zuerst wird kurz mit sterilem dH₂O gespült und dann der Filter bei 42 °C in 20 ml Hybridisierungspuffer (5x SSC, 10 % Blockierungs-Stocklösung, 0,1 % Sarkosyl, 0,02 % SDS, 50 % deionisiertes Formamid, 100 µg/ml gescherte Heringssperma DNA) 1,5 Stunden unter Bewegung prähybridisiert. Danach erfolgt die Hybridisierung mit der Sonde (End-Konzentration: 0,1 µg/ml im Hybridisierungspuffer) bei 42 °C ü.N.

Abschließend wird die Membran 2x für 15 Minuten bei RT mit 2x SSC, 0,1 % SDS und 2x 15 Minuten bei 68 °C mit 0,1x SSC, 0,1 % SDS gewaschen. Die Membran wird luftgetrocknet und die Sonde wird über UV-Licht im Stratalinker (254 nm, 10 mJ/cm², ca. 3 Sekunden) fixiert.

C.2.12.5 Chemolumineszenz-Nachweisreaktion

Die Detektion wird bei RT im Hybridisierungsofen in Glasröhren durchgeführt. Zuerst wird die Membran fünf Minuten mit 15 ml Waschpuffer und fünf Minuten mit 10 ml Blockierungspuffer gewaschen. Danach wird die Membran 30 Minuten in 20 ml Puffer 2 unter Bewegung inkubiert. Anschließend findet unter den gleichen Bedingungen die Bindung des Antikörpers an die mit Digoxigenin markierte Sonde statt (25 ml Puffer 2 + 1 µl Anti-DIG-AP).

Überschüssiges alkalische Phosphatase-Konjugat wird durch dreimalige Waschung für jeweils 15 Minuten mit je 25 ml Waschpuffer entfernt. Die Membran wird zwei Minuten lang in Assaypuffer äquilibriert und feucht auf eine Glasscheibe gelegt. CDP-Lösung (Boehringer) wird aufgetropft und die Membran für drei Minuten bei RT inkubiert. Danach wird eine Frapan-Folie blasenfrei aufgelegt und die Lösung gründlich ausgestrichen. Mit Tesa-Film werden die Ränder versiegelt, um zu verhindern, daß die Membran austrocknet.

Im Dunkeln wird in einer Filmkassette ein Röntgenfilm auf die Membran gelegt und je nach Intensität der Signale für zehn Sekunden bis vier Stunden belichtet. Der Film wird durch dreiminütige Inkubation im Entwickler, eine Minute Lagerung in dH₂O und dreiminütige Inkubation im Fixierbad entwickelt und fixiert (jeweils im Dunkeln). Der Film wird anschließend für mindestens zehn Minuten in dH₂O gewässert und dann getrocknet.

C.2.13 Promotoranalyse

C.2.13.1 Transfektion von HepG2-Zellen

Im folgenden wird die Methode zur Transfektion von adherenten eukaryotischen Zellen unter Verwendung von Lipofectamin Plus Reagenz (Life Technologies) beschrieben. Für die Promotoranalyse werden in dieser Arbeit nur HepG2-Zellen verwendet.

Die zur Transfektion optimale DNA-Menge muß je nach Plasmid ausgetestet werden, als Startkonzentration kann 80 fmol je Testplasmid und 16 fmol je Referenzplasmid (gereinigt nach C.2.6.2, DNA in Sequenzierqualität, möglichst nur ccc-Form, ohne Endotoxine) je Ansatz verwendet werden (berechnet für 12-well-Platten). Das zu testende und das Referenzplasmid werden vor der Transfektion vermischt. Für jeden gleichen Meßansatz werden jeweils drei parallele Platten unter den gleichen Bedingungen transfiziert.

Die Transfektion und Anzucht der Zellen findet auf 12-well-Platten mit Vertiefungsdurchmessern von je 2 cm statt. Die Zellen werden einen Tag vor der Transfektion abtrypsinisiert und in Vollmedium ausverdünnt.

Am nächsten Tag sollen die Zellen 50-60 % konfluent und gut verteilt sein. Die Plasmid-DNA wird in 40 µl DMEM + Gln verdünnt, 2,4 µl PLUS-Reagenz zugeben und 15 Minuten bei RT inkubiert. Währenddessen werden 1,6 µl Lipofectamin-Reagenz in 40 µl DMEM + Gln verdünnt, beide Ansätze vermischt und anschließend für 15 Minuten bei RT inkubiert. Danach wird das Medium von den Zellen abgezogen und zweimal mit DMEM + Gln gewaschen. Nach Zugabe von 0,32 ml DMEM + Gln auf die Zellen wird der DNA-Lipofectamin-Plus-Komplex unter Schwenken zugegeben und der Ansatz wird drei Stunden bei 37 °C inkubiert. Abschließend erfolgt die Zugabe von 1,6 ml DMEM + Gln + FCS. Die Auswertung wird nach zwei Tagen durchgeführt.

DMEM + Gln:	10,0 g	DMEM (Dulbeccos modified Eagle medium; Gibco/BRL) Nr. 074-01600P; Lagerung bei -20°C)
	3,7 g	NaHCO ₃ (bei 10 % CO ₂)
	0,47 g	L-Glutamin
	auf 1000 ml dH ₂ O	
	Das Medium wird sterilfiltriert und bei 4 °C gelagert.	

DMEM + Gln + FCS:	1,0 g	DMEM (Dulbeccos modified Eagle medium; Gibco/BRL) Nr. 074-01600P; Lagerung bei -20°C)
	0,37 g	NaHCO ₃ (bei 10 % CO ₂)
	47 mg	L-Glutamin
	20 ml	Calf serum (inaktiviert durch 30minütige Inkubation bei 56 °C) (NCS; Flow 29-121-54 Lot 29121109)
	1,0 ml	100x Penicillin-Streptomycin
	auf 100 ml dH ₂ O	
	Das Medium wird sterilfiltriert und bei 4 °C gelagert.	

C.2.13.2 Bestimmung der Promotoraktivität

Die hier beschriebene Methode beruht auf dem Dual-LuciferaseTM Reporter Assay System der Firma Promega (DLR-Assay).¹⁴⁸ Das Ziel dieser Experimente ist die Promotoraktivität eines bestimmten DNA-Abschnittes in eukaryotischen Zellen zu testen. Zu diesem Zweck wird ein Plasmid verwendet, das als Reporter das Luciferasegen (*luc*) des Leuchtkäfers (*Photinus pyralis*) beinhaltet.

¹⁴⁸ Promega, 1998

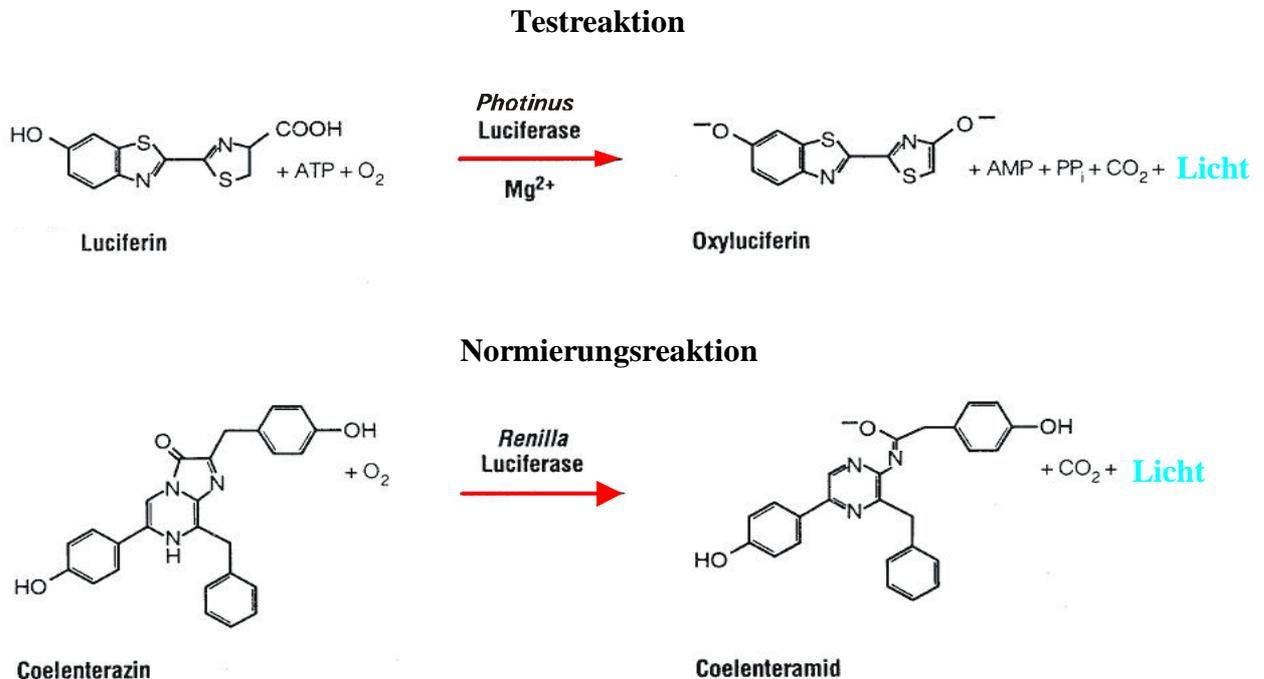
Das Plasmid (pGL3Basis) selbst besitzt keinen funktionellen Enhancer oder Promotor und wird als Negativkontrolle eingesetzt. Das zu testende DNA-Segment wird vor das *luc*-Gen kloniert (C.2.9).

Als Positivkontrolle wird ein Vektor verwendet (pGL3Control), der im Vergleich zum pGL3Basis-Plasmid zusätzlich einen Promotor und einen Enhancer (jeweils aus SV40) enthält. Die zu testenden eukaryotischen Zellen werden mit den Plasmiden transfiziert und nach Expression der Luciferase wird ein Zellaufschluß durchgeführt. Die lysierten Zellen werden zusammen mit dem Luciferase-Substrat in ein Luminometer gebracht und die Lichtausbeute bestimmt.

Zur Normierung der unterschiedlichen Transfektionseffizienz werden die Zellen mit einem zweiten Plasmid kotransfiziert. Dieses Plasmid (pRL-TK) besitzt ebenfalls ein *luc*-Gen unter der Kontrolle eines konstitutiven Promotors. Dieses *luc*-Gen stammt aus *Renilla reniformis*. Das *Renilla-luc*-Genprodukt weist eine andere Substratspezifität auf. Nachdem Auslöschung der Lumineszenz der *Photinus*-Luciferase-Reaktion durch eine spezielle Lösung und Zugabe eines entsprechenden Substrats für die *Renilla* Luciferase können in einem Assay nacheinander die Lichtausbeuten gemessen werden. Somit kann die Promotoraktivität des Inserts im pGL3-Basis-Plasmid ohne Schwankungen durch die jeweilige Transfektionseffizienz bestimmt werden. Die von Promega gelieferte 'Stop&Glo'-Lösung senkt das erste Lumineszenz-Signal um den Faktor 10^5 innerhalb von einer Sekunde und aktiviert die folgende Reaktion im gleichen Zeitraum.

Eine Meßreihe wird zur Bestimmung des Hintergrunds mit nichttransfizierten Zellen durchgeführt. In der nachfolgenden Abbildung sind die enzymatischen Reaktionen der beiden verwendeten Luciferasen dargestellt.

Abbildung 18: Luciferase-Reaktionen



Herstellung des Zell-Lysats von adherenten Zellen in Zellkulturschalen

Die Zellen (HepG2) werden in 12-well-Platten mit Vertiefungsdurchmessern von je 2 cm angezogen und transfiziert (C.2.13.1). Es werden jeweils drei identisch behandelte Ansätze gemessen, um die Standardabweichung bestimmen zu können.

Zuerst wird das Medium von den Zellen abgenommen und einmal mit PBS gewaschen. Das PBS wird quantitativ entfernt. Dann werden die Zellen unter Zugabe von 80 µl 1x PLB ('passive lysis buffer': 1x PLB wird jeweils frisch aus mitgeliefertem 5x Konzentrat mit dH₂O angesetzt) mit einem Greiner 'Disposal cell scraper' lysiert. Das Lysat wird mit einer sterilen Spitze einige Male hochgezogen, um es zu homogenisieren und kann direkt in den DLR-Nachweis eingesetzt werden. Die lysierten Zellen sind bis zu 16 Stunden auf Eis lagerbar.

Bestimmung der Lumineszenz (Nachweis mit dem DLR-System)

Je Ansatz werden 100 µl LAR-II-Puffer in ein Luminometer-Röhrchen gegeben. Dann werden je Röhrchen 20 µl Lysat zugeben und mehrfach zum Mischen der Probe pipettiert. Das Luminometer (Minilumat CB9506) mißt anschließend die Lichtemission nach zwei Sekunden für einen Zeitraum von zehn Sekunden. Danach werden 100 µl von 1x Stop&Glo Reagenz zugegeben und mit dem Ansatz vermischt. Wiederum wird wie zuvor die Lichtemission bestimmt.

D. Ergebnisse

D.1 Analyse des 5'-Bereichs vom Heparinkofaktor II-Gen der Ratte

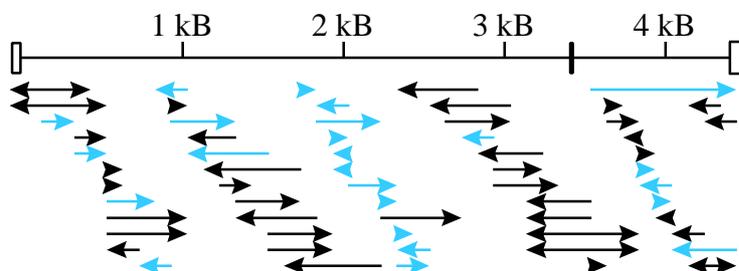
D.1.1 Sequenzierung des 5'-Bereichs vom HCII-Gen der Ratte

Die Exon/Intronstruktur des HCII-Gens ist zwischen verschiedenen Arten hoch konserviert. Auch die Struktur der Gene von Ratte, Maus und Mensch ist sehr ähnlich, da die Positionen der Introns 2, 3 und 4 übereinstimmen (siehe Abbildung 52).

Allerdings gibt es gerade zwischen den sehr nahe verwandten Arten Maus und Ratte einen deutlichen Unterschied. Vor Beginn dieser Arbeit war bekannt, daß in einem Teil der HCII-Transkripte der Ratte¹⁴⁹ eine kurze zusätzliche Sequenz vorkommt (57 Bp), die bei den anderen mRNAs fehlt (siehe Abbildung 41). Bei einem Vergleich mit menschlichen mRNAs und mRNAs der Maus läßt sich kein identischer Bereich zu dieser Sequenz erkennen. Southernblotexperimente zeigten, daß die unterschiedlichen Transkripte von Ratte nicht auf zwei HCII-Genen beruhen können. Zur Analyse der molekularen Ursache für die unterschiedlichen Transkripte wurde in dieser Arbeit der 5'-Bereich des HCII-Gens der Ratte (rHCII) sequenziert.

Die Sequenzierungen wurden mit unterschiedlichen Methoden und hauptsächlich über das 'direct-blotting' Verfahren durchgeführt. Da für dieses Verfahren große Mengen an zu sequenzierender DNA verwendet werden müssen, wurden PCR-Produkte der Intron 1-Sequenz subkloniert. Größere Bereiche konnten nicht kloniert werden, da die Sequenz viele Bereiche enthält (repetitive Elemente, CA-Repeats), die offensichtlich eine Klonierung in Bakterien ausschließen. Erst nach einer Fragmentierung der Sequenz in sehr kleine Bruchstücke (Restriktionsspaltung mit *AluI* o. ä.) konnten diese kloniert werden. Für die Sequenzierung der Inserts wurden Primer verwendet, die stromaufwärts des Inserts innerhalb der Plasmidsequenz hybridisieren. Zusätzlich wurden Sequenzbereiche über direktes Sequenzieren von PCR-Produkten analysiert. Für die Sequenzierung wurden in diesem Fall die gleichen Primer verwendet, wie für die Amplifikation der PCR-Produkte. In beiden Fällen ergaben sich Schwierigkeiten, die Sequenzen langer CA-Repeats eindeutig zu lesen und jeweils beide DNA-Stränge zu sequenzieren. Aus diesem Grund konnte die Sequenz in einem Teil der Bereiche nur einzelsträngig bestimmt werden. Hier liegt auch der Grund für die teilweise nur sehr geringe Leseweite der Einzelsequenzierungen. Nachfolgend ist ein Schema der durchgeführten Sequenzierungen in Abbildung 19 gegeben.

Abbildung 19: Schema der Sequenzierungen (HCII-Gen der Ratte, Intron 1)



Das Schema stellt die durchgeführten Sequenzierungen für den Intron 1-Bereich des HCII-Gens der Ratte dar.

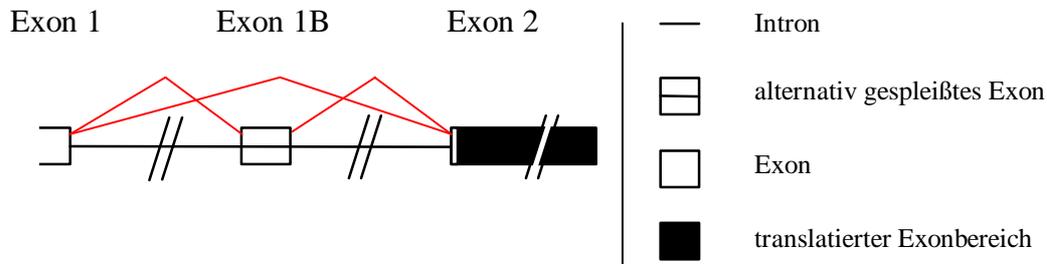
Die Pfeilrichtung deutet die Sequenzierung des Plus- oder Minusstranges an. Doppelpfeile bezeichnen mehrfache Sequenzierungen beider DNA-Stränge. Blaue Pfeile symbolisieren die Sequenzierung von klonierten Fragmenten der Sequenz, während schwarze Pfeile Direktsequenzierungen darstellen.

Die Kästen deuten Exonbereiche an, wobei der linke Kasten für das Exon 1 und der rechte offene Kasten für den Beginn von Exon 2 steht.

¹⁴⁹ Westrup & Ragg, 1994

Die Sequenzierungsdaten der Einzelsequenzen, die zum Teil wiederum aus verschiedenen Sequenzierungen zusammengestellt wurden, sind zusammen mit den Plasmiden, in denen die entsprechenden Inserts vorliegen, im Anhang (G.2.2) abgebildet. Die Sequenz wurde von beiden DNA-Strängen bestimmt und mindestens jeweils zweimal eindeutig sequenziert. Ein Vergleich der ermittelten genomischen Sequenz mit der cDNA-Sequenz zeigt ein alternatives Spleißen der rHCII-Transkripte (siehe Abbildung 20).

Abbildung 20: Alternatives Spleißen von rHCII-Transkripten



Die roten Linien stellen das Spleißen der 5'-gelegenen Exon/Intronbereiche des rHCII-Transkripts dar.

Das erste Intron (incl. Exon 1B) des rHCII ist 4348 Bp lang. Die Sequenz wurde am 02.10.1998 unter der Nummer 227626 an 'Genbank' übermittelt. In der anschließenden Tabelle 18 ist die Sequenz mit den umgebenen Exonbereichen dargestellt.

Tabelle 18: Sequenzen des Introns 1A und 1B des HCII-Gens der Ratte

Bereich	Position	Sequenz
Exon 1*	0001-0047	GCCCCTCTGGTCTCATCTCTGCCACGCCACTGCTCAGAGGCACAGAG
Intron 1A	0001-0050	gtaagtttgggtactaacatTTTTTgccaatataaatttttagttatatt
	0051-0100	taaagctagataaattgggttcctttattaaagtacaaattgagaaaaagaa
	0101-0150	tatatcattaatttttaatttttaatttaatttttttaaacttagttaa
	0151-0200	atatttagtttaataatattttggcataactgacaaaactggaacttga
	0201-0250	atTTtaagagacaagagtattgtacctttgatactttaaactggagactg
	0251-0300	gcatggtagcacaaggctgaaattccagcatgggtagcgtgggtatcca
	0301-0350	tcacctcagggcacttagatgactgtctagtgtttagtatgacaatatt
	0351-0400	cctgttctcaggaagtttccctaagcatcactattaaaggctcggaggaat
	0401-0450	gtaagagttctaacctcaaaggctcagaacaatgtacatatgtagacac
	0451-0500	atatacatgtacacacacatgcacatatcacaccacacacacacacac
	0501-0550	acaaaggagacacagcggctgcactgctattctattctttcaattcttct
	0551-0600	gtaagattgtaacataaaaagggtttccatgtttcccccgcttcttgcttg
	0601-0650	ctctgtgtgaatctggagacatcatggcaggagatggagttctctatgac
	0651-0700	acacatacaaaaaggccaactcaatccatgcaactctcaagaatacacta
	0701-0750	tatgggtgttccctgaaaatacagcttggatgatattactttaaaagcacag
	0751-0800	cacttcgggctggagagatggctcagtggttaagagcactgactgctctt
	0801-0850	cctgaggtcctgagttcaaattcccagcaaccacatgggtggctcacaacca
	0851-0900	tctgtaatgaaatctgatgccctcttctgggtgtgctgaaggcagctaca
	0901-0950	gtgtactcatacaataaaaataaaaatcttaaaaaaaaaaaaaaaaaagcac
	0951-1000	agcattcaattcaagacagtaacttgaatgcaaacacaaagttaacaggt
1001-1050	gtctttaaatgccttaattgttttatgaaactgatgagcaaaatagaaag	
1051-1100	attacctgtcagatctctaatTTTcagtgcataaaaataactttatattt	
1101-1150	ttcccttctctaaagaagagaataatgaaactatggcttatggcttttgg	
1151-1200	atatagaataatgaactatggctcatggcttttggatatagaataatgaa	
1201-1250	actatggctcatggcttttggatatagaataatgaactatggctcatggc	
1251-1300	tttggatatagaataatgaactatggctcatggcttttggctgatgctta	
1301-1350	ccctccaagccaccggagcccaaccggaatgtgtctgtgtctgcagttc	
1351-1400	caggataagacgaagaatagctttaccactgggctcagactagtccctt	
1401-1450	gagcacacattaactgaagttggcttttcttctgttctttaataattactg	
1451-1500	ttactaattgctgttttagccaagggtacttttaattgtaatacaaaaa	
1501-1550	cacttccattaagttgcttctgtcctccctttcttttagagagatgtat	
1551-1600	tttctttttatttgtgaatatgcacttgtctgttcaagtgtgcaccacat	

Intron 1A	1601-1650	gtgtgcagattgccagacgagtggtcaaaacccttggaaactggagttaca
	1651-1700	gatggctgagccacctacctggagggctgggacctgaactcttgctcctc
	1701-1750	tgaaaagtaggaagtactgagtgctgagccccctctctagcccgttact
	1751-1800	gtcagtgcacccatgctgattctgtcttgaatagctctgggtgtagct
	1801-1850	gcatacattaggtactgttcctagaaaactgttccacatgtgcagattatg
	1851-1900	cagcaccggtctcagcaaagtgtgaggaggctcaggcatattcagaatc
	1901-1950	agaaaacctgtgagctctttttccaaaataagcacttctgaagccccgcag
	1951-2000	ccagctgtagatatagctaggagagcgcctgctagcgtgagagccctg
	2001-2050	tctcaccccagcactgtacagaatagttacggtgatacagttctacaact
	2051-2100	gaggaggtataggtaaaagggtcagaaaacttaaggctgaccttgggatac
	2101-2150	aagattttcagctggcctgagatgctcaatgtcttttaaaaagtcctc
	2151-2200	agaaaataaacatggttagggccaggagggtggaaataaccacgcctttaat
	2201-2250	cccagtgcttgggaagcagaagaaggctgatctctacaagttcaaccagc
	2251-2300	ctggctaaagagccagttccaggacagtcaaggctgttgagaaaccttg
	2301-2350	tctctaaaaaatcaaatcaaaacaaacaaaaaaacaaagctgggaatgcta
	2351-2400	atthtagccacattgtaggggataggccctaaggacaaagattgacaaca
	2401-2450	gctgctcccgaagcgtgagcagttctgagtgctcctggcactctacggcaa
	2451-2500	gcaaggcaggcctgtgagaccaggacctggtttaaccaacctggttca
	2501-2550	aatctagtcattatctaccatccctcaatccctacatttttagagctaggg
	2551-2600	actgaacctgggtcccacatgcaatctgtgtgagagtagttaataattatct
	2601-2650	atattcctggctcaaaccactgccatctcttagtagctgcaaaggaaggc
	2651-2700	tacatctcatgatagggttaaaatcgtggcagaatttcaaatagagcatct
	2701-2750	gcaaagctagtcctaactgtcttcagggtcaaacaggctctgaatacttta
	2751-2800	ttaggagccagcaacccttaataatagactggcaacagatatattctaa
	2801-2850	accaagaaatgggtcacaggcacaacacagacttctctttttcctagaaga
	2851-2900	tatgcctagctgctgagacagggaaagcaataagagaggccatctacctct
	2901-2950	ttagtaactgagagaccgactggaagggaaatcaaagccactgggcagaa
	2951-3000	aagctggatctgataagtcatcaagtctcctcaaaaacaaggcaagggga
	3001-3050	tgagcagataaagaacataaaaataccccacctgtccacctgtccacctgt
	3051-3100	ctgcctgtctgtctgtctacagacctctacaaagagtggggcctgacagcg
	3101-3150	ccctcccgcagacctgtgacagtgagactgcaaacaggaggtagaaaa
	3151-3200	gattttctttctcctgacagtaaatctctccctacagcaactcacattaa
3201-3250	cccaatagcccttgaaactacaggagagacacgggccccagccaaactc	
3251-3300	cttcagtgggctcctaagagagctcttgccaaattccaagacttagacaaa	
3301-3350	taaccaaatgactgcacagagcacagcaatctttatgatagtttcttag	
3351-3394	aaggctggttcattggtcattcacaatgtcctcctctccctcag	
Exon 1B	3395-3444	<i>TGAGCAACAGTTTGGTAGCGCTGAGTCCACTGGAAATTCTCCTCCT</i>
	3445-3451	<i>GTCTCAG</i>
Intron 1B	3452-3501	gtatcactctagctcaactcttatgaaacctctttgactctctgggcacc
	3502-3551	ccaccctctccctacttgtcccagtgagtgctgccccttcagcaggcagac
	3552-3601	gtcaatgtctctgggtgccagccctgttttctggagacagatgatgatgtt
	3602-3651	ggtagaaaagagactctggggctctcctgggttccctcttttccaccatt
	3652-3701	aaagaagatggtcttacatcagtgaccagcccagcaggcgaaggcgcta
	3702-3751	cttgtagcaagcacggctacctgagcttcctcaccacccacataacgg
	3752-3801	tggagaaccaaccatctgactgcggcatacatgtttacatgtgtgtgc
	3802-3851	acaccac
	3852-3901	accac
	3902-3951	acacacacacacacacataggggtaggggtgggtgggagtgggggtgggtgg
	3952-4001	gagggttggaaatcatgtcctcctgggtctctctgctctctgtgtagccct
	4002-4051	ggctgtcctggaactcgatttgtagaacaggctggcctcgaactcacaga
	4052-4101	gatctgcctgcctctgcctcccaagcactgggattaagggtgtgcaccac
	4102-4151	tgcccggctgcagtgaataaactttctggatgggggtcaccacaacacga
	4152-4201	gaaactgtattaaagggttgcagcatgaggaagggtgagaaagcactgctc
	4202-4251	taggggtcatagatccaggagatgtatgaaacttacacacactcacact
4252-4301	ggtgtgagccggtgtctatcaagttagacctgagagagcccactgcagg	
4302-4348	gtggcagcagccaggctgcctgggtgctcactggtgtctggttttacag	
Exon 2*	0001-0022	CGGCTTTAGCTCTGCAGAAATG

*Die Exonbereiche sind nur soweit bekannt und bis zum ATG-Startkodon (grün) dargestellt. Die Exon/Intron-grenzen wurden von den bekannten mRNA-Sequenzen¹⁵⁰ abgeleitet. Aus dem Vergleich der genomischen Sequenz mit der cDNA geht weiterhin hervor, daß die beiden ermittelten mRNAs durch alternatives Spleißen gebildet werden (unter Annahme nur eines Gens für HCII, siehe D.1.1, E.1.1).

¹⁵⁰ Westrup & Ragg, 1994

D.1.2 Theoretische Exon/Introngrenzen im 5'-Bereich des rHCII-Gens

Um Aussagen treffen zu können, warum im HCII-Gen der Ratte ein zusätzliches Exon erkannt wird, müssen Daten über die Signale für Exon/Introngrenzen vorhanden sein. Demzufolge ergab sich die Fragestellung nach den Positionen im Intron 1, an denen nach bisherigen Kenntnissen Exon/Introngrenzen erkannt werden können. Ein Vergleich der potentiell erkannten Exon/Introngrenzen von Ratte und Maus kann die Frage nach dem zusätzlichen Exon 1B von Ratte gegenüber der Maus (und anderen Arten) beantworten.

Für den Vergleich wurden zwei Programme (5'Intron.bas, 3'Intron.bas) geschrieben, die nach Klasse I-Introngrenzen (nach den bisher bekannten Kriterien) suchen.

Mit Hilfe dieser Programme wurde der 5'-Bereich des HCII-Gens der Ratte sowie der entsprechende Bereich des Maus- und des menschlichen HCII-Gens untersucht (siehe D.3.3). In den folgenden Tabellen sind die gefundenen Introngrenzen für das 5'- und das 3'-Ende aufgelistet, wobei die tatsächlichen Exon/Introngrenzen fett gedruckt sind. Die invarianten GT- und AG-Dinukleotide sind in Rot und potentielle Intronbereiche sind kursiv gedruckt. Für die Suche nach eventuellen 3'-Introngrenzen wurde nur die Intron 1-Sequenz überprüft, während für die Suche nach 5'-Introngrenzen der in dieser Arbeit angegebene Exon 1-Bereich mit einbezogen wurde (siehe D.1.1, Exon 1: 47 Bp). Der Vielzahl der gefundenen möglichen Positionen von Exon/Introngrenzen ist folgendes zu entnehmen. Zum einen sind die bisher bekannten Kriterien für Grenzen nicht vollständig, so daß viel mehr potentielle Grenzen von den Programmen erkannt werden, als tatsächlich durch analysierte mRNAs nachgewiesen werden können. Zum anderen ist festzustellen, daß sich bezüglich der Erkennung von Exon 1B nur durch ein Vergleich mit den entsprechenden Daten der Maus (und des Menschen) und durch einen Vergleich größerer Sequenzbereiche des Introns eine Aussage treffen läßt (siehe E.1.1, E.2.2).

Tabelle 19: Potentielle Signale für 5'-Intron 1-Grenzen im HCII-Gen der Ratte

Pos.	5'-Spleißstelle	Pos.	5'-Spleißstelle	Pos.	5'-Spleißstelle
48	AGAGGTAAAGT 92 %	437	AAAGGT ^{<i>CGGA</i>} 60 %	448	GAATGT ^{<i>AAGA</i>} 65 %
482	CAATGT ^{<i>ACAT</i>} 49 %	927	TCTGTGT ^{<i>GTGT</i>} 51 %	1046	ACAGGT ^{<i>GTCT</i>} 54 %
1106	ACCTGT ^{<i>CAGA</i>} 48 %	1523	AAGGT ^{<i>ACTT</i>} 53 %	1754	AAAAGT ^{<i>AGGA</i>} 57 %
1798	TACTGT ^{<i>CAGT</i>} 50 %	1957	ACCTGT ^{<i>GAGT</i>} 58 %	2110	ATAGGT ^{<i>AAAA</i>} 64 %
2118	AAGGT ^{<i>CAGA</i>} 63 %	2212	CCATGT ^{<i>TAGG</i>} 59 %	2713	ATAGGT ^{<i>TAAA</i>} 51 %
2775	TCAGGT ^{<i>CAAA</i>} 52 %	2784	ACAGGT ^{<i>CTGA</i>} 61 %	3499	TCAGGTATCA 49 %
3841	ACATGT ^{<i>GTGT</i>} 55 %	4138	AAAGGT ^{<i>GTGC</i>} 60 %	4273	AGATGT ^{<i>ATGA</i>} 50 %
4300	ACTGTGT ^{<i>GTGA</i>} 48 %	4322	TCAA ^{<i>GT</i>} TAGA 53 %		

Die angegebene Position (Pos.) bezieht sich auf die Position der ersten Base des invarianten 'GT' (rot). Die tatsächlichen Spleißstellen sind fett gedruckt. Potentielle Intronbereiche sind kursiv geschrieben. Die Positionsangabe weicht von der Numerierung der beschriebenen Sequenz (D.1.1) ab, da die 47 Bp vom Exon 1 in die Sequenzanalyse mit einbezogen wurden. Der Prozentwert gibt die prozentuale Übereinstimmung mit der verwendeten Matrix für die Spleißstelle an (siehe C.2.1).

Tabelle 20: Potentielle Signale für 3'-Intron 1-Grenzen im HCII-Gen der Ratte

Position	Verzweigungsbereich	3'-Spleißstelle	Gesamtbewertung
360	<i>AGATG</i> ACT 60 %	<i>ATATTCCTGTTCTC</i> AG 77 %	137
1539	<i>CTTCC</i> ATT 81 %	<i>CTCCCTTTCTTTTT</i> AG 89 %	170
2726	<i>TTCAA</i> ATA 58 %	<i>CCCTAACTGTCTTC</i> AG 77 %	135
2845	<i>TGGTC</i> ACA 41 %	<i>CTTCTCTTTTTCTC</i> AG 89 %	130
2903	<i>AAGCA</i> ATA 40 %	<i>CCATCTACCTCTTT</i> AG 74 %	114
3071	<i>TGTCC</i> ACC 81 %	<i>GTCTGTCTGTCTAC</i> AG 73 %	154
3169	<i>TGCAA</i> ACA 58 %	<i>TTCTTTCTCCTGAC</i> AG 88 %	146
3186	<i>TCCTG</i> ACA 80 %	<i>AAATCTCTCCCTAC</i> AG 73 %	153
3393	<i>TGTTCA</i>TTT 81 %	<i>CCTCCTTCTCCCTC</i>AG 95 %	176
3450	<i>CATCC</i> ACT 81 %	<i>TCTCCTCCTGTCTC</i> AG 92 %	173
3544	<i>TCCCT</i> ACT 81 %	<i>CTGTGCCCTTCAGC</i> AG 77 %	158
3996	<i>TTGGA</i> ATC 58 %	<i>TCTGTCTCTGTGT</i> AG 74 %	132
4347	<i>TGCTC</i>ACT 81 %	<i>GTTGTCTGTTTTAC</i>AG 73 %	154

Die angegebene Position (Pos.) bezieht sich auf die Position der ersten Base des invarianten 'AG' (rot). Die tatsächlichen Spleißstellen sind fett gedruckt. Potentielle Intronbereiche sind kursiv geschrieben. Das Adenin-Nukleotid, an welches möglicherweise das 5'-Ende des Introns ligiert wird (Lariatbildung) ist ebenfalls rot hervorgehoben. Der Prozentwert gibt die prozentuale Übereinstimmung mit der verwendeten Matrix für den Verzweigungsbereich und die Spleißstelle an. Die Gesamtbewertung errechnet sich aus der Summe der beiden Einzelprozentwerte (siehe C.2.1).

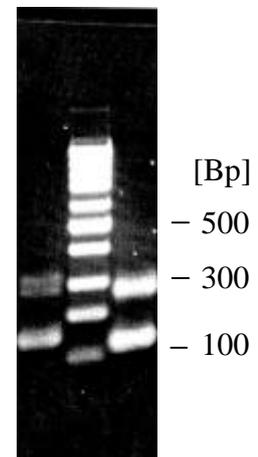
D.2 Identifizierung von zwei verschiedenen HCII-mRNAs in Mausleber

Aufgrund des bei der Ratte zusätzlich vorhandenen Exons im HCII-Gen wurde zum Vergleich der 5'-Bereich der Maus-HCII-mRNA auf zusätzliche Exonsequenzen hin analysiert. Primer-extension- und PCR-Ergebnissen von Ragg¹⁵¹ können entnommen werden, daß auch das HCII-Gen der Maus ein Exon 1 enthält, welches mit den entsprechenden Exons von Mensch und Ratte vergleichbar ist.

Zur Überprüfung der mRNAs wurde die RT-PCR Methode verwendet. Die cDNA Synthese wurde mit der Genescript Reversen Transkriptase, Maus poly(A)⁺-RNA aus Leber und dem Primer Dag5 (1 Fehlpaarung) durchgeführt. In der darauffolgenden PCR mit den Primern Mst12/Mst36 und in der nested PCR (Mst4/Mst36) wurden jeweils zwei Fragmente erhalten (Abbildung 21; Lage der Primer: siehe D.3.2, Abbildung 23). Die Fragmente haben eine Größe von ca. 120 Bp und 260-300 Bp. Das kleine Fragment wurde über Gelextraktion isoliert und kloniert, während das große Fragment über eine PCR mit den Primern Mst4 und Mst14 erhalten wurde. In einer nested PCR wurden 1/10 Volumen des Produkts als Megaprimer zusammen mit dem Primer Mst36 (Matrize: 1/300 Volumen RT-PCR-Produkt) verlängert und amplifiziert (Zugabe des Primers Mst4 vor den letzten fünf Zyklen). Danach erfolgte eine direkte Sequenzierung der Antisinn-DNA-Stränge der PCR-Produkte. Das kleine Fragment wurde mit dem Primer Ap^r1 und das große Fragment mit dem Primer Mst2 sequenziert.

Abbildung 21: Analyse des 5'-Bereichs der mHCII-mRNA mittels RT-PCR

Die Abbildung zeigt ein 2 % Agarosegel: In der linken Spur sind die Fragmente der PCR mit den Primern Mst12-Mst36 aufgetragen. Die rechte Spur zeigt die aufgetrennten Fragmente der nested PCR mit den Primern Mst4-Mst36. In der mittleren Spur sind 0,25 µg der '100 Bp-Leiter' aufgetragen.



¹⁵¹ in: Kamp & Ragg, 1999

In der jeweiligen komplementären Notation ergibt sich für die Sinnsequenzen der mRNAs:

Exon 1 / Exon 2

kleines Transkript: $5'$ TCAGAGGNAC AGAG/CAGCT TTAGCTCTGC AGAAATGAAA $3'$...
großes Transkript: $5'$ AGACCACTGC TCAGAGGCAC AGAG/ATAGC TGGTGTGAGC ATTGGTGTCTG $3'$...

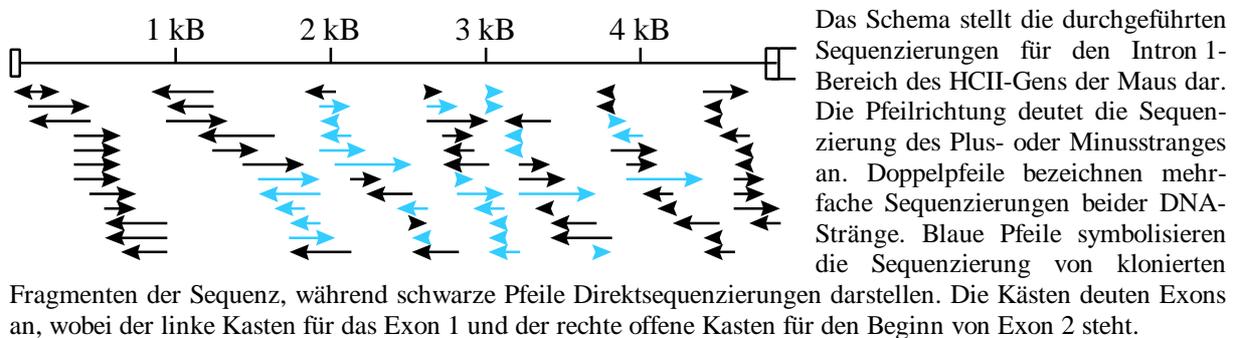
Dem Ergebnis läßt sich entnehmen, daß ein etwa 120 Bp großes Fragment in einem Teil der Mausleber-cDNAs im Vergleich zu den cDNAs der Ratte eingeschoben ist (siehe D.3.2, Abbildung 23). Die verschiedenen mRNAs können aufgrund von zwei unterschiedlichen HCII-Allelen oder aufgrund von alternativen Spleißvorgängen eines einzigen Gens zustande kommen. Durch Analyse der genomischen Sequenz kann die Möglichkeit alternativ gespleißter HCII-mRNAs geprüft werden.

D.3 Analyse des 5'-Bereichs vom Heparinkofaktor II-Gen der Maus

D.3.1 Sequenzierung des 5'-Bereichs vom HCII-Gen der Maus

Nachdem zwei unterschiedliche HCII-Transkripte in der Maus identifiziert wurden, kann über die genomische Sequenz geklärt werden, wie diese mRNAs zustande kommen. Zusätzlich können eventuell die Fragen beantwortet werden, auf welcher molekularen Ursache das zusätzliche Exon 1B des HCII-Gens der Ratte beruht und wie die vergleichbare genomische Sequenz der Maus beschaffen ist: Warum gibt es kein Exon 1B bei dem HCII-Gen der Maus (mHCII)? Die Frage läßt sich beantworten, wenn die gesamten 5'-Regionen des HCII-Gens des Menschen, der Maus und der Ratte verglichen werden (Teil D.4). Die menschliche Sequenz dieses Bereichs ist schon seit längerer Zeit bekannt.¹⁵² Um die drei Intronregionen vergleichen zu können, wurde also in dieser Arbeit nicht nur die Sequenz des 5'-Bereichs des HCII-Gens der Ratte, sondern auch der Maus bestimmt. Für die Sequenzierung des Intron 1 wurden die Daten aus vielen Einzelsequenzierungen des Plus- und des Minusstrangs zusammengestellt (es gelten die gleichen Anmerkungen zu den Sequenzierungen dieses Intronbereichs, wie sie bereits für die Sequenzierungen des rHCII-Introns unter D.1.1 beschrieben wurden). Die Daten, die zum Teil wiederum aus verschiedenen Sequenzierungen zusammengestellt wurden, sind zusammen mit den Bezeichnungen der sequenzierten Plasmide im Anhang (G.2.2) dargestellt. In der folgenden Abbildung ist eine schematische Übersicht über die einzelnen durchgeführten Sequenzierungen gegeben.

Abbildung 22: Schema der Sequenzierungen (HCII-Gen der Maus, Intron 1)



¹⁵² Herzog *et al.*, 1991

Die Sequenz des mHCII-Intron 1 ist 4842 Bp lang und wurde unter der Nummer 228451 am 06.10.1998 an 'Genbank' übermittelt. In der anschließenden Tabelle 21 ist die Sequenz mit den umgebenen Exonbereichen dargestellt.

Tabelle 21: HCII-Intron 1 Sequenz der Maus

Bereich	Position	Sequenz
Exon 1*	0001-0029	CTGCCAGACCACTGCTCAGAGGTACAGAG
Intron 1	0001-0050	gtaagtttggttttctaacttttttgccaattataaattttttatttttaa
	0051-0100	tgcttagagctaaacaattggttcctttattgagaaaagaatatgttatt
	0101-0150	tttaaaatttagtttaattttcttttaaaatttagtttaagtgttagttta
	0151-0200	atattaatattggtacaactgacaaaaatggaatttgaattttaagagac
	0201-0250	aagattattatacctttgatactttataactggaactgggcatggcagca
	0251-0300	caagattgaaattccagcacttgtagcccctgggtatctatcacccatgg
	0301-0350	gcacaaaatcaattttcttttcaaattagataattgtctagtgattaagaa
	0351-0400	tgagaatattcttggtcttagaggaagtttccacctaagcattactattaa
	0401-0450	aggttagaggaatgtaagagatctaactcaaaatgccagaaaaacgtgt
	0451-0500	acatatgttgatacatatacatgtaccccccatgtgcaccacatcacat
	0501-0550	catacacacacaccacacatacacacgcacatacaccataccacacgcaca
	0551-0600	cacacaccaacacacccccacacacatacaccacacataccacaccacacca
	0601-0650	cacacatatcataccacaccacacatacaccacacacacacacacacacacc
	0651-0700	acaccacacacatacacacacaccacacacacacacataccacacccgtgacaca
	0701-0750	cacacaccacatcacaccacacacgcacacaacagtaaaagggaaacacagt
	0751-0800	ggctgctctgctactctattctttcaattcttctgtgaagtttgaagttat
	0801-0850	tttaacataaaaaggtttctctcttgcttctgttctgtttggatctggagag
	0851-0900	atcacagtaggaaatgggtgttctttatgacacatacaaaaaggccaac
	0901-0950	tccaatccatgcaactctcaagaatacactatatgatgttccatgaaaa
	0951-1000	caaaattttgatgaatatcttaaaattatagcactttaattcaaaagacag
	1001-1050	caatttgaacgcatacacaacttacaggatTTTTTaaatgccttaatttt
	1051-1100	TTTTTaatgaaactgatgagcaatatagaaagatttatctgtcagatcgt
	1101-1150	taattttcaatgcataaaaataactacattttttcctttctctaagaagag
	1151-1200	aataatgaaaattatggcttatggcttttgacataccttcaggaacacagg
	1201-1250	gtatatgatgcttacccttcagaaccaaccaaccaaccaacctaacc
	1251-1300	aaaccagaatgtgttctgtagttccaggataagatgaagaacatcttttct
	1301-1350	atgtgggctcagactagtctccttgaacctacattaattcaagttggctt
	1351-1400	ttcttttggtcttttaacattattgtaactactgttttagccaaggaaatt
	1401-1450	tttaactgtaatcataaaaagggttccattaagtttttttatTTTgtgcat
	1451-1500	atgcacttgtctgttcaagtatgcaccacattgtgcagactgccagaagg
	1501-1550	gtacaaaactccttgcagctggagttaacagatggctgagccaccacctg
	1551-1600	gagggtgggacctgaactctcgtaactgctgagcccctctctagc
	1601-1650	ccacttaatttcaatgatacccatgcctgattttatcttgaaactagtct
	1651-1700	tatttacattagatactgtttccagtaattgttccacatgtgtagattat
	1701-1750	gtagccatgagtctcagcaaagtgtgagatggatcgggcttattcacag
	1751-1800	tcagaaagcatgtggctctttttccaaaataagcacttcagaagccccc
	1801-1850	gccagctgtagatgtacctcagctgggagagcatactcctagcactgac
	1851-1900	gcctgactcattgcagcactgtacagataggtatggtgatctgggtct
	1901-1950	acaaaataaggaggtataatcaaaagggtcagaaacgtaaagtcagctgg
	1951-2000	tacacaagattttcagctagcctgagatgcacagtgcttttttaaaaagt
	2001-2050	ccctaagttatgattaaaacaatcctgtggccaggaggtgggtggtacaca
	2051-2100	cctttaatcccagcacttgaggaggtagaggaagggcggatttctgagttca
	2101-2150	aggccaacctggtctacagagtgagttccaggacagccagggctacacag
	2151-2200	agaaacctgtctgaaaaaaccaaaaaccataaaacaaaaaacaaaaaac
	2201-2250	aaaaaaaaaaaaaaaaaatacaaaacaaagggcgggtgagatggctcagtg
	2251-2300	gtaagagcactgactgctcttccgaaggtccaagttcaaatcccagcaa
	2301-2350	ccacctgtaatgagatctgatgcctcttctgggtgcatctgaagacatct
	2351-2400	actgtgtataataataataataatctttgggcctgagcgagcagggttgac
	2401-2450	cagagtttagcagaggtcgtaaaaattcaattcccacaaccacatggagg
	2451-2500	ctcacaaccatctgtacagctatggtatactcacatacattaaattaaaa
	2501-2550	aaatttttttaaatcttttaaaaaaacaaaaacaaacaaaaataaacaa
	2551-2600	acaaaaacaaaaggctgggaatgctaattttaaccacattatagggacagt
	2601-2650	ccctaaggacaaaaggctggtaaaagctgttcccagaagtgtgagcagtttg
	2651-2700	agtgtcctggcattctaaaggcaagcaacatgagactagagatctggggac
	2701-2750	ttaaccaacctactactcatcatctaccatccctcaatccctacatt
	2751-2800	ttagtgtatggattgaacctgatcccatgcaagctgggtgagtgtagt
	2801-2850	taatatttatctatattcctggatcaaacactgccatctcttagtagtg

Intron 1	2851-2900	gtaaaggaaggccacatcttatgataggttaaaatggcagtaaatTTTTca
	2901-2950	aatagagcatctgcaaagctagtcacctactgtttataggcctaacacgt
	2951-3000	cgaatactttatcaggagccagcaacccttactgaatatagatagactg
	3001-3050	gcaacagatattctaaatcaagaaatgggtcacaggcacaaaatagacttc
	3051-3100	TTTTTTTTTctctagaagatataacctagctgctgagacaggggaagcaac
	3101-3150	aagagagggtcatctacctctttagaaattgagagactcaagagcaaatc
	3151-3200	aaagccactgggcagaaatctggatttaataagctatcaaggctcctca
	3201-3250	gaaacaaggcaaggataagcagataaagacaagaacaggggtctacacaa
	3251-3300	aataccttgcttgtctacagacctctacaaagggttaagcctctgacatca
	3301-3350	ctctcctgcagaccctgtttgcacagtgagattgaaagcaggaagtagaa
	3351-3400	aagatTTTCTTTCTCCTGACAGCAAAGTTCTCTCTACAGTCCCTAATAGC
	3401-3450	TCTTGGGACTGCAGGAGAGACACAGGACCCAGCCATACTCCTTTCTCAG
	3451-3500	TGAGTCCTAAGGAAGTCTTATCAAGTTCAAATCTTAAACAAATTAATCC
	3501-3550	AAATGACTGCAGAGAGTACAGCAGTCTTACAATATGTTTCTAGAAGGCT
	3551-3600	GGCTCATGTTCTGTTATAACACCTCCTTCCACAGGTGAGCAGGAGTTGG
	3601-3650	TAAGGCTTGGTCCATCTACTGGACACTCCTCCTGCCTCACTCTAGGCATT
	3651-3700	CACTCTAGCCAGCTCTTAGAAAACCTCTGTAACCTCTCTGGACACCTCTC
	3701-3750	CCTTCCCTGCCCCAAATCATGGCAGCACTGTTAGAAAGCTCTTGATC
	3751-3800	CGGAAAGTCTCCCTGCTTGTCCCATGATTCTGTGCTCTCAGCAGGCA
	3801-3850	AACCTCAGTGCCAGAGACAGGTGATGATGTTGGTGAATGAATGAA
	3851-3900	AAGAGGCTCTAGGAATCGCCAGGCCCTCTTTACATCAGTAACGGAGG
	3901-3950	CAGTGCAGATCAGTGAAGTGGCTCACCAGGTGCAAACCTGAGTTCTGTC
	3951-4000	TCCACAACTCATATAAAGGTGGAATAAGCAACTTCACCCAACCGACTGAC
	4001-4050	TGTGGCATGCATGTTTACATGTGTGTGCACAAACTACACACACACACACA
	4051-4100	CACACACACACACACACACACACACACACACACAGAGACACGTAATAAA
	4101-4150	TAATAAAAAAATTGCGTGCTGAGATGGCTCAGCAGGTACAGGTGCTG
4151-4200	GGACTCACATGGTAGGAGAGTGAGTTCTACCAGTTGTCTATGACTTCCA	
4201-4250	CACACACAGCATGGTAAACCTGTCCACCTCCATAATAATAATAACAAA	
4251-4300	AAATCTAAAACATTTTTTAAAGAATAAATTTTTCATCTAGGGCAGNGGTTCT	
4301-4350	TCAACCTGTGGGTCAACCAGATAACCTGAATATCAGATATTTACAATTTA	
4351-4400	TAACAGCAAATTTATAGTTATAAGGTAGCAACAAAATATATATATGGGGT	
4401-4450	GGTTTGGAATTGTGTCCCCCTCTCACCAGGGTCTCTTGTGTAGCCCTG	
4451-4500	GCTGCCCTGGAACCTCGCTTGTAGGCTAGGCTGGCCTCGAATTCACAGAG	
4501-4550	ATCCACCTGCCTCTGCTTCCCGAATGCTGGGATTAAAGGAGCACAAACCC	
4551-4600	TGCCTGGGTGCAATGAAC TAATCTTATGGTTGGGGTCAACACAGCATGAA	
4601-4650	GAAC TATATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCTCC	
4651-4700	AGGGGCTCATAGATTCAGGAGATGTATGAAACTTACAAAACACTATTTA	
4701-4715	TTATTTTTTATCAG	
Intron 1 / Exon 2	4716-4765	ATAGCTGGTGTGAGCATTGGTGTCTGTGAAGGCTTCTTGTCTGTGTTAAGAG
	4766-4815	AATCTGTCTGTGACATTCCAGTGGGGGTGGTGTCTAGCCAGGCTCCCTG
	4816-4842	CCATTCTTGTGTCTGTTTTTTCACAG
Exon 2*	0001-0022	CAGCTTTAGCTCTGCAGAAATG

*Die Exonbereiche sind nur soweit bekannt und bis zum ATG-Startkodon (grün) dargestellt. Die Exon/Intron-grenzen wurden von den bekannten mRNA-Sequenzen¹⁵³ abgeleitet. Aus dem Vergleich der genomischen Sequenz mit der cDNA geht hervor, daß die beiden ermittelten mRNAs (siehe D.2, D.3.2) durch alternatives Spleißen gebildet werden.

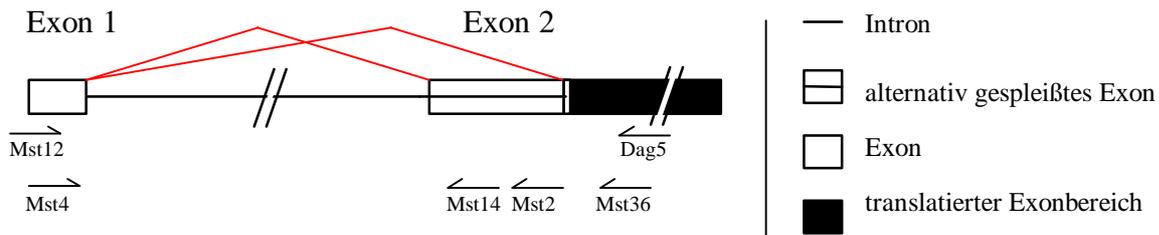
D.3.2 Alternatives Spleißen von HCII-mRNAs der Maus

Bei einem Vergleich der genomischen HCII-Sequenz (D.3.1, Tabelle 21) mit den beiden ermittelten cDNA-Sequenzen (D.2) ist zu erkennen, daß die zusätzliche Sequenz, die nur in einem Teil der mRNAs enthalten ist, direkt vor dem Exon 2 gelegen ist.

Somit gibt es (unter Annahme nur eines Gens für HCII; vergleiche: D.5) auch in der Leber der Maus alternativ gespleißte HCII-Transkripte. Den Sequenzierungen der cDNA und des ersten Introns der Maus ist weiterhin zu entnehmen, daß es keinen vergleichbaren Bereich zum Exon 1B der Ratte gibt, jedoch eine alternative Verlängerung des 5'-Bereichs von Exon 2, der in den mRNAs von Mensch und Ratte nicht vorhanden ist. In dem folgenden Schema ist das Ergebnis dargestellt.

¹⁵³Zhang *et al.*, 1994

Abbildung 23: Mögliches alternatives Spleißen von mHCII-Transkripten



Die roten Linien stellen das Spleißen der 5'-gelegenen Exon/Intronbereiche des mHCII-Transkripts dar. In das Schema sind die Bezeichnungen für die in der Analyse verwendeten Primer eingetragen. Die genomische Sequenz vor dem Exon 2 stimmt mit dem in einem Teil der cDNAs inserierten Sequenz überein (D.2; Tabelle 21).

D.3.3 Theoretische Exon/Introngrenzen in dem sequenzierten Bereich

Nachdem die Intron 1-Sequenz des HCII-Gens der Ratte auf mögliche Exon/Introngrenzen hin untersucht wurde (siehe D.1.2), werden für einen Vergleich die entsprechenden Daten von der Maus (als nahe verwandte Art) und dem Menschen (als ferner verwandte Referenz) benötigt. Darüber hinaus wurden mit Hilfe von RT-PCR in Mausleber mRNAs gefunden (D.2), die ein am 5'-Ende verlängertes Exon 2 des HCII-Gens aufweisen. Eine Erklärung für diese Veränderung kann ebenfalls über die Kenntnis von Spleißsignalen gegeben werden. Die HCII-Intron 1-Sequenzen der Maus (siehe D.3.1) und des Menschen (Verwendung der von Herzog¹⁵² ermittelten HCII-Gen-Sequenz) wurden mit den Programmen '5'Intron.bas' und '3'Intron.bas' auf mögliche Exon/Introngrenzen hin überprüft. Für die Überprüfungen gelten die gleichen Anmerkungen, die zuvor unter Punkt D.1.2 beschrieben wurden.

Nachfolgend sind die gefundenen Introngrenzen für das 5'- und das 3'-Ende tabellarisch aufgelistet. Für die Suche nach eventuellen 3'-Introngrenzen wurde nur die jeweilige Intron 1-Sequenz überprüft, während für die Suche nach 5'-Introngrenzen die in dieser Arbeit angegebenen Exon 1-Bereiche mit einbezogen wurden (D.3.1, Maus: 29 Bp; D.8.2.2, Mensch: 65 Bp). Den Ergebnissen ist zu entnehmen (entsprechend zu D.1.2), daß Aussagen über den Grund für unterschiedliche artspezifische Exon/Introngrenzen nur durch einen Vergleich mit den entsprechenden Daten der Ratte (und des Menschen) gemacht werden können (siehe E.1.2, E.2.2), da in den analysierten Intron 1-Bereichen eine Reihe von potentiellen Spleißsignalen vorhanden sind.

Tabelle 22: Potentielle Signale für 5'-Intron 1-Grenzen im HCII-Gen der Maus

Pos.	5'-Spleißstelle	Pos.	5'-Spleißstelle	Pos.	5'-Spleißstelle
22	AGAG GT ACAG 48 %	30	AGAG GT AAGT 92 %	432	AAAG GT TAGA 76 %
443	GAAT GT AAGA 65 %	814	TTCT GT AAGT 55 %	1057	ACAG GT ATTT 67 %
1230	CAGG GT ATAT 52 %	1498	TCAA GT ATGC 50 %	1912	ATAG GT ATGG 66 %
1955	AAAG GT CAGA 76 %	1964	AAAC GT AAGG 72 %	1969	TAAG GT CAGC 69 %
2150	CAG GT GAGT 56 %	2263	GCGG GT GAGA 56 %	2280	GTGG GT AAGA 61 %
2648	GCTG GT AAAG 51 %	2819	CTGG GT GAGT 65 %	2880	AGTG GT AAAG 51 %
2907	ATAG GT TAAA 51 %	3137	AGAG GT CATC 51 %	3312	AAAG GT TAAAG 58 %
3479	CTCAG GT GAGT 49 %	3587	TCAT GT TCGT 48 %	3614	ACAG GT GAGC 77 %
3629	TTTG GT AAGG 61 %	3850	ACAG GT GATG 59 %	4050	ACAT GT GTGT 55 %
4121	ACAC GT AAAT 66 %	4167	GCAG GT ACAG 49 %	4173	ACAG GT GCCT 54 %
4191	CATG GT AGGA 59 %	4199	GAG GT GAGT 50 %	4243	CATG GT AAAC 57 %
4403	TAAG GT AGCA 48 %	4703	AGAT GT ATGA 50 %		

Die angegebene Position (Pos.) bezieht sich auf die Position der ersten Base des invarianten 'GT' (rot). Die tatsächlichen Spleißstellen sind fett gedruckt. Potentielle Intronbereiche sind kursiv geschrieben. Die Positionsangabe weicht von der Numerierung der beschriebenen Sequenz (D.3.1) ab, da die 29 Bp vom Exon 1 in die Sequenzanalyse mit einbezogen wurden. Der Prozentwert ergibt die prozentuale Übereinstimmung mit der verwendeten Matrix für die Spleißstelle (siehe C.2.1).

Tabelle 23: Potentielle Signale für 3'-Intron 1-Grenzen im HCII-Gen der Maus

Position	Verzweigungsbereich		3'-Spleißstelle		Gesamtbewertung
371	<i>TAGTGATT</i>	80 %	<i>TATTCTTGTTCTAGAG</i>	78 %	158
1866	<i>CACTGACA</i>	80 %	<i>CCCTGACTCATTGCAG</i>	73 %	153
2938	<i>TTCAAATA</i>	58 %	<i>CCCTAACTGTTTATAG</i>	71 %	129
3065	<i>AATAGACT</i>	58 %	<i>TTTTTTTTTTCTCTAG</i>	89 %	147
3123	<i>AAGCAACA</i>	40 %	<i>TCATCTACCTCTTTAG</i>	74 %	114
3269	<i>GGTCTACA</i>	41 %	<i>CCTTGCTTGTCTACAG</i>	80 %	121
3310	<i>TACAGACC</i>	78 %	<i>CATCACTCTCCTGCAG</i>	80 %	158
3449	<i>CTGGGACT</i>	58 %	<i>CATACTCCTTTCTCAG</i>	80 %	138
3644	<i>CATCTACT</i>	81 %	<i>TCCTGCCTCACTCTAG</i>	78 %	159
3793	<i>GATCCATG</i>	41 %	<i>ATTCTGTGTCCTTCAG</i>	73 %	114
3796	<i>CCATGATT</i>	80 %	<i>CTGTGTCTTTCAGCAG</i>	77 %	157
3890	<i>TAGGAATC</i>	58 %	<i>CCCTCTTTTACATCAG</i>	80 %	138
4713	<i>AACTTACA</i>	41 %	<i>ATTTATTTTTTATCAG</i>	73 %	114
4840	<i>CTGCCATT</i>	61 %	<i>TGTCTGTTTTTCACAG</i>	80 %	141
4843	<i>CTGCCATT</i>	61 %	<i>CTGTTTTTCACAGCAG</i>	73 %	134

Die angegebene Position (Pos.) bezieht sich auf die Position der ersten Base des invarianten 'AG' (rot). Die tatsächlichen Spleißstellen sind fett gedruckt. Potentielle Intronbereiche sind kursiv geschrieben. Das Adenin-Nukleotid, an welches möglicherweise das 5'-Ende des Introns ligiert wird (Lariatausbildung) ist ebenfalls rot hervorgehoben. Der Prozentwert ergibt die prozentuale Übereinstimmung mit der verwendeten Matrix für den Verzweigungsbereich und die Spleißstelle. Die Gesamtbewertung errechnet sich aus der Summe der beiden Einzelprozentwerte (siehe C.2.1).

Tabelle 24: Potentielle Signale für 5'-Intron 1-Grenzen im HCII-Gen des Menschen

Pos.	5'-Spleißstelle		Pos.	5'-Spleißstelle		Pos.	5'-Spleißstelle	
66	<i>AGAGGTAAGT</i>	92 %	393	<i>AAAGGTTAGA</i>	76 %	468	<i>CGAGGTCAGG</i>	68 %
714	<i>AAATGTTAGA</i>	60 %	798	<i>ACACGTACGT</i>	68 %	855	<i>CCTGGTAAAT</i>	68 %
865	<i>CTCGGTACGG</i>	52 %	1234	<i>CAATGTTTGT</i>	54 %	1331	<i>TCTGGTAAGG</i>	69 %
1489	<i>TCAAAGTTGGT</i>	48 %	1688	<i>ACAGGTGTGA</i>	61 %	1828	<i>AAAAGTGTGT</i>	55 %
1861	<i>ATATGTAAAA</i>	48 %	1971	<i>CATGTA AACT</i>	67 %	1984	<i>ACCTGTAATT</i>	53 %
2061	<i>CATGGTAAAA</i>	57 %	2377	<i>CAATGTGACT</i>	52 %	3015	<i>CCAGGTTGGA</i>	60 %
3196	<i>AATGGTGCGA</i>	48 %	3552	<i>AGGCGTAAGC</i>	52 %	3646	<i>TATGGTCACT</i>	49 %
3714	<i>GGAAGT AAGC</i>	58 %	3768	<i>CAAAGTCAGA</i>	59 %	3789	<i>ACGAGTCAGA</i>	48 %
4140	<i>CCAGGTCAGG</i>	75 %	4190	<i>TGAGGTGGGT</i>	56 %	4208	<i>TGAGGTCAGG</i>	62 %
4528	<i>AATGGTGTGA</i>	48 %	4807	<i>GCACGTCTGT</i>	48 %	4865	<i>TCTGGTA ACT</i>	62 %
4898	<i>CAAGGTGTCT</i>	52 %	4956	<i>GAAGGTTAGT</i>	79 %	5041	<i>AAAGGTTGGG</i>	60 %

Anmerkungen zu der Tabelle: siehe Tabelle 22

Tabelle 25: Potentielle Signale für 3'-Intron 1-Grenzen im HCII-Gen des Menschen

Position	Verzweigungsbereich		3'-Spleißstelle		Gesamtbewertung
142	<i>TTTCTACA</i>	61 %	<i>TGGTTTTCATTTTTCAG</i>	73 %	134
1115	<i>GATTTATC</i>	61 %	<i>GTTTCTCTTCAAACAG</i>	73 %	134
1269	<i>AATCCATA</i>	41 %	<i>TCCTTCTTCTGGTAAG</i>	78 %	119
1303	<i>CGAAA ACT</i>	58 %	<i>TTTGGCTTTTTATCAG</i>	73 %	131
1401	<i>TCTTTACT</i>	81 %	<i>CTCTATCTGGCCTCAG</i>	73 %	154
1662	<i>CCCCTATG</i>	61 %	<i>TTCTTCTTGATCTTAG</i>	74 %	135
1866	<i>CAGAACT</i>	58 %	<i>CCAGTTCTATTTGCAG</i>	73 %	131
2394	<i>TTGTGACT</i>	80 %	<i>CTTACTATTTCAACAG</i>	73 %	153
2746	<i>TCCGAACA</i>	58 %	<i>GTTACCCCATCTTCAG</i>	77 %	135
3011	<i>TGCAGACT</i>	78 %	<i>GATCCTCCTACTTCAG</i>	73 %	151
3256	<i>GGCTAATT</i>	80 %	<i>TTTGTATTTTTTGTAG</i>	74 %	154
3392	<i>TACTGACC</i>	100 %	<i>ATTTAATTTTTTCACAG</i>	73 %	173
3469	<i>AAGTGATC</i>	60 %	<i>TGCCTTGGTCTCCCAG</i>	73 %	133
4513	<i>TTTCAACA</i>	80 %	<i>CATCTCCATTGCAAAG</i>	71 %	151
4572	<i>GTCCC ACT</i>	61 %	<i>TGCACATCCCTTCTAG</i>	71 %	132
5150	<i>CTTTC ACT</i>	81 %	<i>CTGTTTTCCCTCCCAG</i>	88 %	169

Anmerkungen zu der Tabelle: siehe Tabelle 23

D.4 Alignment der Sequenzen von Intron 1 des Heparinkofaktor II-Gens

Im HCII-Gen der Ratte (siehe D.1.1) und der Maus (siehe D.2, D.3.1) wurden alternativ gespleißte Exonbereiche gefunden, die bei der jeweils anderen Art nicht vorkommen.

Das Exon 1B des HCII-Gens der Ratte ist ebenfalls nicht in anderen Arten und bei anderen Serpinen vorhanden. Durch einen Sequenzvergleich der HCII-Intron 1-Sequenz zwischen Maus und Ratte werden Hinweise für das Auftreten des (neuen) Exons gesucht. Parallel können Hinweise zu der veränderten Exon 2-Grenze im Maus-HCII-Gen erhalten werden. Um die Veränderungen besser vergleichen zu können und um Aussagen treffen zu können, welche Positionen direkt durch artbedingte Mutationen verändert sind, wurde als Referenz die menschliche Sequenz (als entfernt verwandte Art) in das Alignment miteinbezogen. Bisher ist kein alternatives Spleißen im menschlichen HCII-Gen bekannt.

D.4.1 Alignment

Mit dem Programm 'ClustalW'¹⁵⁴ wurde mit Standardeinstellungen ein Alignment durchgeführt. Da das Programm allerdings keine Sequenzen von etwa 5 kB Länge verarbeiten kann und sich außerdem bei einer Durchsicht zeigte, daß das Alignment nicht optimal durchgeführt wurde (d. h. es erreichte nicht die maximale Anzahl an übereinstimmenden Positionen), erfolgte eine manuelle Bearbeitung der Sequenzen in mehreren Schritten.

Zu diesem Zweck wurden Sequenzen, deren Enden sich durch bestätigte Ähnlichkeiten aus einer Datenbanksuche mit dem Programm BLAST2¹⁵⁵ direkt zuordnen lassen, zuerst in etwa 1 kB große Blöcke unterteilt. Danach ergab manuelle Bearbeitung ein Alignment dieser Sequenzen. Anschließend wurden die bearbeiteten Stücke in das Programm 'ClustalW' eingegeben. Nach dem Alignment erfolgte die Zusammensetzung und Nachbearbeitung zu dem gesamten Intron 1.

Das Alignment der Intron 1-Sequenzen wurde in die nachfolgende Tabelle zusammen mit den in dieser Arbeit verwendeten Primern eingetragen. Zur besseren Übersicht wurden größere Insertionen in der menschlichen Sequenz entfernt, die von repetitiven Elementen herrühren. Konservierte Positionen und Exonbereiche sind hervorgehoben (siehe Zeichenerklärung zur Tabelle 26). Die Intronstruktur (Insertionen, Deletionen und repetitive Elemente) ist zusätzlich schematisch in der Abbildung 45 dargestellt.

Tabelle 26: Alignment der HCII-Intron 1-Sequenzen

0001- 0060	Ratte Maus Mensch	gtaagtttggttactaacatTTTTTgcaa-tataaatTTTTtagtt--atatttaaagc gtaagtttggtttctaacaTTTTTgcaaattataaatTTTTtatttttaagccttagagc gtaagttgggtttctaagtTTT-ctgctgattataaattatttttg--gtgtttacgga ***** ** ** ** ** * ** * ** * ** * ** * ** * ** * ** *
0061- 0120	Ratte Maus Mensch	caattggttccttt----att-----gagaa Mst37 -> tagataattggttccttt----attaaagtacaaattgagaaaaagaatatcattaat taacaattggttccttt----att-----gagaaaa-gaatat---gttatt taggcaactggttcatttttctagcaaaactaagaattcagaag-----ctttct ** ** ***** ** * ** * ** * ** * ** * ** *

Fortsetzung der Tabelle auf der nächsten Seite

¹⁵⁴ Thompson *et al.*, 1994

¹⁵⁵ Altschul *et al.*, 1990; Altschul *et al.*, 1994; Claverie *et al.*, 1993; Karlin & Altschul, 1993. Für die Suche wurden Standardeinstellungen verwendet. Das Programm sucht in mehreren DNA-Datenbanken nach Sequenzen, die Ähnlichkeit zu einer ausgewählten Sequenz aufweisen.

Zeichenerklärung zu der Tabelle 26

Größere Insertionen in der menschlichen Sequenz, die von eingefügten repetitiven Elementen stammen (wie z.B. *Alu*) wurden entfernt und durch das Symbol 'E' markiert. Übereinstimmung von Positionen bei allen drei Arten sind durch ein Sternchen hervorgehoben. In den Bereichen der Intron 1-Sequenz, die eine starke Konservierung bei der Datenbanksuche zeigen (Erkennung einer kurzen HCII-Intron 1-Sequenz mittels BLAST2, die Ähnlichkeit mit den entsprechenden Bereichen der anderen Arten aufweist), sind die Sternchen rot gedruckt. Unter dem Alignment sind gegebenenfalls repetitive Elemente (ermittelt mit dem Programm 'RepeatMasker2') eingetragten. Einige kurze Bereiche sind durch Unterstreichen gekennzeichnet. Hier liegen direkte oder inverse Sequenzwiederholungen vor. Exonbereiche sind gelb hinterlegt. Fett gedruckte Buchstaben innerhalb des gelb markierten Bereichs deuten Exonbereiche an, die in allen cDNAs vorkommen. Die anderen Bereiche (kursiv gedruckt) zeigen alternativ gespleißte Exons an. Die Numerierung bezieht sich nicht auf die HCII-Sequenz einer einzelnen Art, sondern auf das hier durchgeführte Alignment.

Dem Alignment ist zu entnehmen, daß viele Sequenzbereiche des Intron 1 vom HCII-Gen innerhalb der Arten Ratte, Maus und Mensch konserviert sind. In den drei Arten sind jeweils auch die homologen Intronsequenzen vorhanden, die den alternativ gespleißten Exonsequenzen in den HCII-mRNAs der Maus und der Ratte entsprechen.

Größere Unterschiede zwischen den Sequenzen werden hauptsächlich durch inserierte repetitive Elemente bewirkt. Im Teil E.2.3 wird die Struktur der drei Introns verglichen (Abbildung 45). Eine ausführliche Diskussion der Ergebnisse findet im gesamten Teil E.2 statt. Die HCII-Genstrukturen von Ratte, Maus und Mensch werden in Abbildung 52 (Abschnitt E.5.1) vergleichend dargestellt.

D.4.2 Suche nach repetitiven Elementen in den Sequenzen

Voraussetzung für das durchgeführte Alignment (D.4.1) war, eventuell vorhandene größere Insertionen bzw. Deletionen in den einzelnen Sequenzen zu kennen. Aus diesem Grund wurden mit dem Programm 'RepeatMasker2' die Intron 1-Bereiche des HCII-Gens von Mensch,¹⁵⁶ Maus und Ratte auf repetitive Elemente hin durchsucht. Die Analyse einer Sequenz mit Repeatmasker2 beruht auf dem Programm 'cross_match' von Phil Green, welches einen Smith-Waterman-Gotoh-Algorithmus verwendet.¹⁵⁷ Die essentiellen Teile der Auswertung wurden gekürzt und in den beiden folgenden Tabellen abgebildet. Für die Durchsuchung der Sequenzen bei der Maus und der Ratte mit 'RepeatMasker2' wurde die Version vom 19.9.1997 mit 'ProcessRepeats'-Version vom 15.12.1997 verwendet. Die Suche beim Menschen wurde mit den entsprechenden Versionen vom 21.4.1999 durchgeführt. In allen Fällen wurde die 'sensitive' Suche verwendet. Diese Einstellung bewirkt eventuell ein besseres Alignment des Sequenzbereiches mit dem gefundenen Element und die Erkennung von bis zu 5 % mehr repetitiven Elementen in der Sequenz.¹⁵⁷

Die meisten gefundenen repetitiven Elemente gehören zu der SINE- und LINE-Gruppe. Auf diese Elemente wird kurz in der Einleitung dieser Arbeit eingegangen (B.2.1).

Die gefundenen Insertionen von repetitiven Elementen wurden zum einen für das Alignment der Intron 1-Sequenzen zur Hilfe genommen. Zum anderen werden die Ergebnisse dieser Analyse ebenfalls in die Diskussion zu veränderten Exon/Introngrenzen (E.2.2, E.2.3) und in die Fragestellung zu potentiellen Promotorbereichen miteinbezogen (E.4.3.2).

¹⁵⁶ Es wurde die von Herzog (Herzog *et al.*, 1991) beschriebene Sequenz verwendet.

¹⁵⁷ Es gibt bisher keine Veröffentlichung des Programms in einer Zeitschrift. Die Autoren schlagen als Referenz die Anleitung des Programms vor: Smit, A. F. A. & Green, F.; RepeatMasker at: <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Tabelle 27: DNA-Elemente im Intron 1 des Heparinkofaktor II-Gens

HCII-Intron 1 der Ratte											
Nr.	SW	%	%	%	Pos.in Seq.		Erkannt.	Typ:	Pos.im Elem.		
	Wert	Div	Del	Ins	Beginn	Ende	+/-	Element	Klasse/Familie	Beginn	Ende
1	51	5.3	0.0	0.0	99	173	+	AT_rich	Low_complexity	1	71
2	337	19.4	1.5	1.5	437	503	+	(CA)n	Simple_repeat	2	68
3	1401	3.7	6.8	0.0	757	947	+	B2	SINE/B2	2	201
4	537	26.2	8.6	1.9	1534	1743	-	B3	SINE/B2	(0)	221
5	224	16.9	6.8	13.6	2157	2274	+	MT2B	LTR/MaLR	155	264
6	626	18.0	3.1	1.6	2189	2316	+	B1_MM	SINE/Alu	17	146
7	234	18.4	0.0	0.0	3031	3068	-	(CAGA)n	Simple_repeat	(0)	31
8	235	26.7	2.5	9.9	3681	3841	+	RSINE1	SINE/B4	17	161
9	402	0.0	3.5	0.0	3801	3916	+	(CA)n	Simple_repeat	1	120
10	817	8.9	1.6	1.6	3987	4109	-	B1_MM	SINE/Alu	(24)	121
11	582	11.5	0.0	0.0	4117	4203	-	URR1B	DNA/MER1_type	(148)	81
HCII-Intron 1 der Maus											
Nr.	SW	%	%	%	Pos.in Seq.		Erkannt.	Typ:	Pos.im Elem.		
	Wert	Div	Del	Ins	Beginn	Ende	+/-	Element	Klasse/Familie	Beginn	Ende
1	23	0.0	0.0	0.0	29	51	+	AT_rich	Low_complexity	1	21
2	36	7.6	0.0	0.0	96	161	+	AT_rich	Low_complexity	1	66
3	207	29.6	2.8	6.3	304	445	+	L1MC4	LINE/L1	7696	7831
4	537	7.4	7.8	3.5	450	731	+	(CA)n	Simple_repeat	1	246
5	290	17.7	0.0	5.1	735	813	+	L1MC4	LINE/L1	7903	7971
6	274	24.6	3.5	8.4	1429	1570	-	B3A	SINE/B2	(19)	191
7	1067	9.0	0.0	0.0	2034	2277	+	B1_MM	SINE/Alu	4	141
8	248	16.7	0.0	0.0	2280	2227	+	A-rich	Low_complexity	1	48
9	836	8.9	16.4	0.0	2228	2473	+	B2	SINE/B2	2	171
10	287	15.8	27.1	4.5	2410	2542	+	B2	SINE/B2	48	210
11	204	4.7	0.0	1.6	4028	4091	+	(CA)n	Simple_repeat	1	61
12	22	0.0	0.0	0.0	4092	4114	+	AT_rich	Low_complexity	1	21
13	317	20.0	8.6	0.0	4147	4216	+	RSINE1	SINE/B4	68	141
14	27	7.8	0.0	0.0	4232	4282	+	AT_rich	Low_complexity	1	51
15	348	12.5	11.2	0.0	4318	4397	-	URR1A	DNA/MER1_type	(74)	161
16	784	13.7	0.8	1.5	4428	4558	-	B1_MM	SINE/Alu	(17)	130
17	577	12.0	2.2	0.0	4560	4651	-	URR1B	DNA/MER1_type	(141)	94
HCII-Intron 1 des Menschen											
Nr.	SW	%	%	%	Pos.in Seq.		Erkannt.	Typ:	Pos.im Elem.		
	Wert	Div	Del	Ins	Beginn	Ende	+/-	Element	Klasse/Familie	Beginn	Ende
1	587	24.2	0.4	0.4	172	335	+	L1MC4	LINE/L1	7635	7771
2	2364	8.1	0.0	0.0	336	645	+	AluSg	SINE/Alu	1	310
3	587	24.2	0.4	0.4	646	707	+	L1MC4	LINE/L1	7779	7844
4	500	21.6	1.4	1.4	738	876	+	L1MC4	LINE/L1	7841	7971
5	810	15.9	0.0	0.0	1512	1643	-	FLAM_C	SINE/Alu	(5)	121
6	262	6.5	0.0	0.0	1652	1682	-	U2	snRNA	(157)	31
7	1642	18.2	0.0	0.0	1894	2200	+	AluJb	SINE/Alu	1	301
8	222	26.7	20.9	20.9	2523	2627	+	MIR	SINE/MIR	28	151
9	1571	15.7	5.2	5.2	2891	3061	-	AluJo	SINE/Alu	(0)	311
10	2253	8.7	0.3	0.3	3062	3370	-	AluSp	SINE/Alu	(5)	301
11	1571	15.7	5.2	5.2	3371	3504	-	AluJo	SINE/Alu	(181)	131
12	2051	13.0	0.0	0.0	4072	4371	+	AluSx	SINE/Alu	1	291

In der Tabelle sind die mit RepeatMasker2 gefundenen DNA-Elemente in den HCII-Intron 1-Sequenzen von Ratte, Maus und Mensch aufgeführt. Nr.: Die gefundenen Elemente sind 5' → 3' durchnummeriert; SW-Wert: Smith-Watermann score, ein Wert für die Übereinstimmung der Sequenz mit einem Element; Div: Prozentualer Unterschied der Sequenz zum Element; Del/Ins: Prozentuale Angabe der Deletionen und Insertionen in der Sequenz; Pos.in Seq./Pos.im Elem.: Angabe der Position der Sequenz (mit Beginn und Ende) in der Gesamtsequenz bzw. Abschnitt innerhalb des Elements. +/-: Sinn- bzw. Antisinnorientierung des Elements. Aufgeführte Typen von DNA-Elementen: Low_complexity: DNA-Sequenz mit niedriger Komplexität, d. h. hohem Anteil von A-Nukleotiden (A-rich), oder von A- und T-Nukleotiden (AT_rich); Simple_repeat: Di-, Tri- oder Tetranukleotidwiederholungen; MER: 'Medium reiteration frequency sequence'; LTR/MaLR: 'Mammalian LTR-transposons'¹⁵⁸; Alu: Nomenklatur der speziellen Alu-Typen nach Batzer¹⁵⁹; MIR: 'Mammalian-wide interspersed repeats'¹⁶⁰. Die HCII-Intron 1-Sequenz des Menschen wurde von Herzog¹⁵⁶ übernommen.

¹⁵⁸ Smit, 1993¹⁵⁹ Batzer *et al.*, 1996¹⁶⁰ Smit & Riggs, 1995

Tabelle 28: Zusammenfassung der Elemente im Intron 1 des HCII-Gens

		Ratte 4348 Bp, 44 % GC-Gehalt			Maus 4842 Bp, 41 % GC-Gehalt			Mensch 5151 45 % GC-Gehalt		
Element-Familie	Typ	Zahl	belegt [Bp]	belegt [%]	Zahl	belegt [Bp]	belegt [%]	Zahl	belegt [Bp]	belegt [%]
SINE	(Gesamt)	5	727	16,72	6	766	15,82	7	1768	34,32
	B1	2	165	3,79	2	275	5,68	-	-	-
	B2	2	401	9,22	3	421	8,69	-	-	-
	B4	1	161	3,70	1	70	1,45	-	-	-
	Alu	-	-	-	-	-	-	6	1663	32,28
	MIR	-	-	-	-	-	-	1	105	2,04
LINE	LINE1	-	-	-	1	221	4,56	1	365	7,09
DNA-Elem.	MER1	1	87	2,00	2	172	3,55	-	-	-
LTR-Elem.	MaLR	1	118	2,71	-	-	-	-	-	-
Σ (Komplexe Wiederhol.)		7	932	21,44	9	1159	23,94	8	2133	41,41
Einfache Wiederholungen		3	180	4,14	3	394	8,14	-	-	-
Kleine RNAs		-	-	-	-	-	-	1	31	0,60
Niedrige Komplexität		1	75	1,72	2	162	3,35	-	-	-
Gesamt		11	1196	27,51	14	1727	35,67	9	2164	42,01

In der Tabelle sind die Daten der mit RepeatMasker2 gefundenen DNA-Elemente aus Tabelle 27 zusammengefaßt. Aufgeführte Typen von DNA-Elementen sind: SINE: 'short interspersed elements'; LINE: 'long interspersed elements'; Niedrige Komplexität: DNA-Sequenz mit einem hohen Anteil von A- und T-Nukleotiden; einfache Wiederholungen: Di-, Tri- oder Tetranukleotidwiederholungen; MaLR: 'Mammalian LTR-trans-posons'¹⁵⁸; MIR: 'Mammalian-wide interspersed Repeat'¹⁶⁰

D.4.3 Überprüfung auf ein Exon 1B im menschlichen Gen

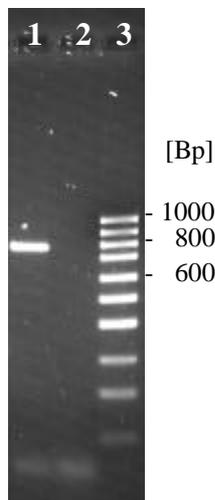
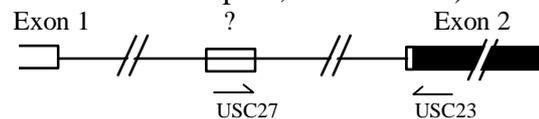


Abbildung 24: Analyse zur Identifizierung eines potentiellen Exons 1B im hHCII

In der Abbildung ist ein 2,5 % Agarosegel nach Elektrophorese von PCR-Produkten zu sehen. In Spur 1 sind 5 µl des RT-PCR-Produkts von hHCII-Leberpoly(A)⁺-RNA und den Primern USC23 und USC27 aufgetragen. Spur 2 zeigt 5 µl der Negativkontrolle (die RT-Reaktion wurde ohne RNA durchgeführt). In Spur 3 sind 0,5 µg '100 Bp-Leiter' aufgetragen.

Nach der Überprüfung mit den Programmen zur Identifizierung potentieller Exon/Introngrenzen besteht die Möglichkeit der Existenz eines Exons 1B im hHCII-Gen analog zum Exon 1B im rHCII-Gen (siehe Tabelle 24, Tabelle 25 und Abbildung 47). Deshalb wurde zum Test eine RT-PCR mit Omniskript-Reverser Transkriptase durchgeführt und die cDNA mit poly(A)⁺-RNA aus menschlicher Leber und einem d(T)₁₆-Primer synthetisiert. Die anschließende PCR wurde mit einem Sinnprimer (USC27) durchgeführt, der in der zum Exon 1B analogen Sequenz bindet. Der verwendete Antisinnprimer (USC23) liegt am Beginn von Exon 2. Sofern kein Intronbereich aus der prä-mRNA (des möglichen Transkripts in Leber) entfernt wird, kann nur ein PCR-Produkt mit einer Länge von 793 Bp amplifiziert werden (basierend auf den Resten von chromosomaler DNA in den RNAs oder langen mRNA-Transkripten; siehe Skizze).



Der Abbildung 24 ist zu entnehmen, daß trotz der Übereinstimmung mit Spleißsignalen kein Exon 1B im menschlichen HCII-Gen vorhanden ist (zumindest ist es nicht in Lebertranskripten nachzuweisen).

D.5 Beteiligung eines repetitiven Elements an einer Introngrenze

Die HCII-Intron 1-Sequenz der Maus zeigt kurz vor dem Exon 2 einen zusätzlichen Bereich von etwa 100 Bp einer auffälligen Insertion im Maus-HCII-Gen gegenüber dem Ratten-HCII-Gen (siehe Abbildung 25, Abbildung 48). Um feststellen zu können, ob die Insertionen bei der Maus von einem (unbekannten) repetitiven Element stammen, wurde ein Southernblot mit genomischer Maus-DNA durchgeführt. Dazu wurde die genomische DNA mit den Restriktionsenzymen *EcoRI*, *PvuII* und *XbaI* gespalten und als Sonde ein mit den Primern Mst2 und Mst13 amplifiziertes PCR-Fragment (112 Bp) verwendet.

Abbildung 25: Schema der Intron 1/Exon 2-Grenze

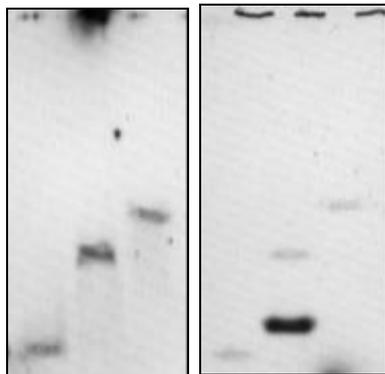
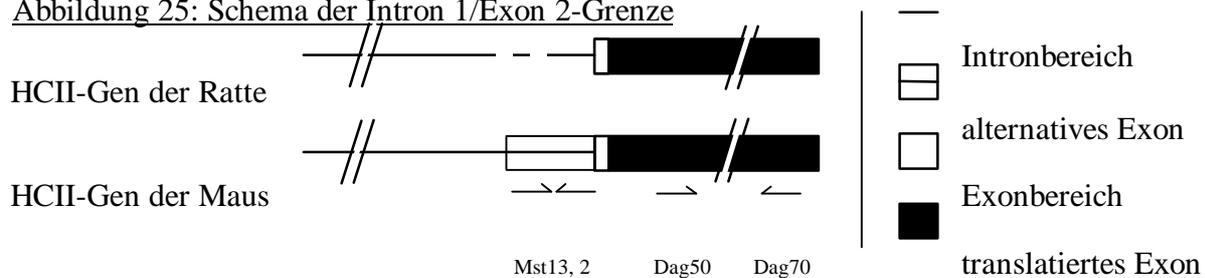


Abbildung 26: Southernblots

Die Abbildung links zeigt einen mit Mst2-Mst13-Sonde hybridisierten Blot. Der Blot auf der rechten Seite wurde mit einer Sonde des PCR-Fragments mit den Primern Dag50-Dag70 hybridisiert. Die chromosomale Maus-DNA (je 40 µg) in den Spuren der einzelnen Blots wurde jeweils mit *EcoRI*, *PvuII* und *XbaI* (von links nach rechts) gespalten.

entnommen werden, daß im Vergleich mit der anderen Sonde die gleichen Fragmente erkannt werden. Somit beinhalten die von der Mst2-Mst13-Sonde erkannten Fragmente das HCII-Exon 2, und das Ergebnis eines fehlenden repetitiven Elements in dieser Region bestätigt sich. In der mittleren Spur des HCII-spezifischen Blots erscheint eine zusätzliche, dicke Bande. Mittels einer PCR konnte nachgewiesen werden, daß eines der Restriktionsenzyme (*PvuII*) mit einer hHCII-cDNA verunreinigt war. Somit läßt sich diese abweichende Bande erklären.

Das Ergebnis des Southernblots ist auf der linken Hälfte von Abbildung 26 zu sehen. Die Sonde hybridisiert in jeder Spur mit nur einem einzigen Fragment, so daß ein repetitives Element ausgeschlossen werden kann.

Um sicher zu gehen, daß die zu erkennenden Hybridisierungssignale von HCII-spezifischer DNA stammen, wurde ein identischer Blot mit einer für das Exon 2 von HCII spezifischen Sonde durchgeführt (siehe Abbildung 25). Die Sonde bestand aus einem 567 Bp großen, mit den Primern Dag50 und Dag70 amplifizierten PCR-Fragment. Als Matrize wurde klonierte menschliche cDNA (pWTBi1) verwendet. Die für die Spaltung der genomischen DNA verwendeten Restriktionsenzyme erkennen keine Schnittstellen zwischen den jeweiligen Sequenzen, mit denen die Sonden hybridisieren können. Die Restriktionsschnittstellen liegen vor dem 3'-Intron 1-Bereich, der bei Maus und Ratte Insertionen bzw. Deletionen enthält (siehe Tabelle 26, ab Position 5541) und hinter Exon 2. Aus diesem Grund erkennen beide Sonden die gleichen (also gleich große) Restriktionsfragmente.

Ein mit der Dag50-Dag70-Sonde hybridisierter Blot ist auf der rechten Seite von Abbildung 26 dargestellt. Dem Blot kann entnommen werden, daß im Vergleich mit der anderen Sonde die gleichen Fragmente erkannt werden. Somit beinhalten die von der Mst2-Mst13-Sonde erkannten Fragmente das HCII-Exon 2, und das Ergebnis eines fehlenden repetitiven Elements in dieser Region bestätigt sich. In der mittleren Spur des HCII-spezifischen Blots erscheint eine zusätzliche, dicke Bande. Mittels einer PCR konnte nachgewiesen werden, daß eines der Restriktionsenzyme (*PvuII*) mit einer hHCII-cDNA verunreinigt war. Somit läßt sich diese abweichende Bande erklären.

D.6 Alignment von Heparinkofaktor II-cDNA-Sequenzen

Bisher wurde in dieser Arbeit der 5'-Bereich der HCII-Genstruktur von Ratte und Maus (und Mensch) verglichen. Da bei den Vergleichen teilweise erhebliche Unterschiede festgestellt werden können, stellen sich die folgende Fragen: Zum einen, ob es weitere Hinweise auf Änderungen der Struktur in weiteren Bereichen des HCII-Gens gibt und zum anderen, ob Unterschiede zwischen verschiedenen Organismen vorhanden sind.

Aus diesem Grund wird in diesem Abschnitt ein Alignment bekannter HCII-cDNA-Sequenzen verschiedener Organismen durchgeführt, um die aufgetretenen Mutationen bzw. Mutationsraten im Intron 1 mit Exonbereichen vergleichen zu können. Die Auswertung der Ergebnisse des Alignments wird in dem Teil E.3 durchgeführt.

Die HCII-cDNA Sequenzen wurden verschiedenen Publikationen entnommen.¹⁶¹ Die bisher bekannten cDNAs von Frosch und Huhn scheinen nicht den 5'-Bereich der jeweiligen mRNA-Sequenz zu enthalten (insbesondere fehlt das ATG-Startkodon vom Huhn). Das Alignment kann dementsprechend nur für die jeweils bekannte cDNA-Sequenz durchgeführt werden. Der 3'-gelegene nicht translatierte Bereich unterscheidet sich nach wenigen Basen sehr stark bei den weniger verwandten Arten. Da die 5'- und 3'-Enden der cDNAs teilweise sehr starke Unterschiede aufweisen, wurde das Alignment für die 5'-Enden manuell durchgeführt. Ein Alignment des 3'-Endes der cDNAs wurde mit Hilfe des Programms 'ClustalW' und anschließender manueller Nachbearbeitung durchgeführt. Die Exon/Intron-grenzen des Frosch- und Huhn-HCII-Gens sind nicht experimentell bewiesen und können aufgrund der Sequenzähnlichkeit nur vermutet werden. Konservierte Positionen in allen Arten sowie bei Säugetieren wurden markiert (siehe Zeichenerklärung zur Tabelle). Im Anschluß ist in Tabelle 29 das Alignment der HCII-cDNAs dargestellt.

Tabelle 29: Alignment von HCII-cDNAs verschiedener Organismen

Exon 1 ¹⁶² 0001- 0060	Maus Ratte Kaninchen Mensch	TCTCTGCCAGACCACTGCTCAGA GCCCCCTCTGGTCTCATCTCTGCCACGCCACTGCTCAGA GACACTGAAAGCCTGCTGGTGTATCTTAGCGGATTTTGTCTCTGAAACGCTACTTCTCTGA TCAAAATCCCCTGATGACCTTGGCTGCTTTTCATCTCTGAAGCGCCACTTCTCAGA
Exon 1 ¹⁶² 0061- 0069	Maus Ratte Kaninchen Mensch	GGTACAGAG__Intron_1_____ GGCACAGAG__Intron_1_____ AGCACAGAG__Intron_1_(?)_____ AACACAGAG__Intron_1_____
Exon 2 0070- 0129 AS: -19 - -6	Maus Ratte Kaninchen Mensch Frosch	CAGCTTTACGTCTGCAGAAATGAAACATCCACTCTGCACT--CCTCTTCTCTCATCAC CGGCTTTAGCTCTGCAGAAATGAAGCATCCAGCCTACACTCTCCTTCTCTCTCATCAT -----CTCTGCA-AAATGCAACATCGACCCACCTGCTTCTCATTCTCTCACCAT ---CTTTAGCTCCGCCAAAATGAAACACTCATTAAACGCACTTCTCATTTCTCTCATCAT AAAATGAAACTT---CTACACTTAGCTACAATAATTTCTTCTTAT ***** * * . * . * * **.. *
Exon 2 0130- 0189 AS: -6 - 15	Maus Ratte Kaninchen Mensch Huhn Frosch	A---TTTATGTGTATTGGGAGCAAAGGCCTTGCTGAGCAGCTA----- G---TCTATGTGTGCTGGGAGCAAAGGCCTTGCTGAGCAGCTA----- AATGTCTGTGTGTGGTGGGAGCAACGGTCTCACAGATCAGCTG----- ACATCTGCGTGGGGTGGGAGCAAAGGCCGCTGGATCAGCTAGAGAAAGGAGGGGAAC GGCACATTCTGTGGAATCAAGGACTTCAGTGACCATTTTGAA----AGCCTCAA-- TCACGCAACACTCGGTGGAGTTAAGGACTTGACGAGCACTTTGAG-----GACC ***. . . * * . * * * * . *

¹⁶¹ HCII-cDNA-Sequenzen:

Frosch und Huhn: Colwell & Tollefsen, 1998; Kaninchen: Sheffield *et al.*, 1994; Maus: Zhang *et al.*, 1994, diese Arbeit; Mensch: Herzog *et al.*, 1991, diese Arbeit; Ratte: Westrup & Ragg, 1994

¹⁶² nach: Kamp & Ragg, 1999

Exon 2 0190- 0249 AS: 15 - 35	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>-----ACTAATGAAACCTGACCACGTCCCT -----ACTAAGGAAACCTGACCAGTGTCCT -----AATAACAAGAACCTGACCATGCCCC T<u>CT</u>CAGT<u>TCTGCAGA</u>-TCC<u>CAGTGGAGCAGTTA</u>AAATAACAAAACCTGAGCATGCCCT -----GATGCACATAC-----ACACGAA-AATGGAACCTACAATATGCCAG TCTACTGGT--ATAAACCCAGGGGAAG----TCAAACCCAAGTTAGAAACCTCCCT</p> <p>.....*.....*</p>
Exon 2 0250- 0309 AS: 35 - 55	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>TCCTGCCTGCTAACTTCACAAAGAAACACAGTCACCAATGACTGGATCCAGAGGGGG TCCTGCCTCCTAACTTCACAAAGAGAACACAGTCACCAATGACTGGATTCAGAGGGAG TGCTGCCAATCGAGTTCACAAAGGAAACACAGTCACCAATGACTGGATTCAGAGGGGG TTCTCCCTGCCGACTTCACAAAGGAAACACCGTCACCAACGACTGGATTCCAGAGGGGG ACCTACCACCTGGAGTTCACAGAGAAACACTATACCAATGACTTGTATCCTGAAGAGG -----GGATGATACGGTTACCAATGACTTGACACTGAAGGAG*.....*</p>
Exon 2 0310- 0369 AS: 55 - 75	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>AAGAGGATGAGGACTACCTGGACCTGGAGAAGCTGCTTGGTGAAGATGACGACTACATT AGGAAGATGACGACTACCTGGACCTGGAGAAGCTGCTCAGTGAAGATGACGACTACATC AGGAGGACGACGACTACCTGGACCTGGAGAAGTTACTCAGTGAAGATGACGACTACATC AGGAGGACGACGACTATCTGGACTGGAGAAGATATCAGTGAAGACGACTACATC AGGAGGAAGAGGACTATCTAGATCTTGACAAGATATGGGTGAAGACGACTAC--AGTC AAGATGAAGAAGACTACCTGGACTTTGATAAAAATATTTGGAGAAGATGAAGATTATATTC *****</p>
Exon 2 0370- 0429 AS: 75 - 95	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>ATATCATCGAT-----GCCGTTTCCCGACAGACTCAGAATCAAGTGTGGCAACATCC ATGTTGTCGAT-----GCAGTTTCCCAACAGACTCAGAATCAAGTGTGGCAACATCC ACATCATCGAC-----GCTGTTTCCCGACCGACTCGGAAGCCAGCGCCGGAAATCC ACATCGTCGACAGTCTGTCAGTTTCCCAGACTCTGATGTGAGTGTGGGAACATCC ACATTATCGAT-----GCTGCTCCACACATAGTTTCTGAAATTCAGCAAGGAAATATTC ATATAATCGATGCT---GCCCTGAAATTAATAATTCGAACCAGCAGGGAACATT *****</p>
Exon 2 0430- 0489 AS: 95 - 115	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>TGCAGCTTTTCCAAGGCAAGAGCCGGATCCAGCGTCTTAATATCCTCAATGCAAAGTTTC TGCAGCTTTTCCAAGGCAAAAGCCGGATCCAGCGCCTTAACATCCTCAATGCAAAGTTTC TGCAGCTCTTCCAAGGCAAGAGCAGGATCCAGCGTCTTAACATCCTCAATGCAAAGTTTC TCCAGCTTTTTCATGGCAAGAGCCGATCCAGCGCTTTAAACATCCTCAACGCCAAGTTTCG TGGAACTTTCCAAGGCAAAACCAGAATCCAGCGTCTTAATATCCTCAATGCAAACTTTC TTGAACTGTTCCATGGGAAAACAAGGTTCAAAGGCTCAATATAATCAATGTAATTTTC *****</p>
Exon 2 0490- 0549 AS:115 - 135	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>CCTTCAACCTTTACCAGTCTCTGAAGGACCAGGCTACCACATCTGATAACCTTTCATAC CCTTCAACCTTTACCAGTCTCTGAAGGACCAGGCTACCCTTCTGATAACATTTTCATAC CTTTCAGCCTTTACCAGACTGAAGGACCAGGCCAACGCTTTTGATAACATCTTTATAC CTTTC AACCTT ACCAGAGTCTGAAAGACCGACTCACACTTTCGATAACATCTTTCATAG GCTTCAATCTTTATCGGAGTGAGCGGACAAAGCCAACCTCCTCAGACAATATTCATAGC GGTTAACCTTTACCCTGCAATAAAACAATAACAGACGATCTGAAAACATTTCTGCTTC *****</p>
Exon 2 0550- 0609 AS:135 - 155	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>CACCTGTTGGCATTTCTACTGCCATGGGGATGATCTCCTTAGGTCTGAGGGGAGAGACC CACCGTTGGCATTTCCACTGCCATGGGGATGATCTCCTTAGGTCTGAGGGGAGAGACC CGCCTGTTGGCATTTCTACTGCCATGGGGATGATCTCCTTAGGTCTGAGGGGAGAGACC CACCGTTGGCATTTCTACTGCGATGGGATGATTTCTTAGGTCTGAAGGGAGAGACC CTCCTGTTGGTATTTCCACTGCAATGGCTATGATTTCCCTGGGTCTGAAGGTTCAAACCT CACCTGTTGGCATTTCCACTGCAATGGCCACAATATCATTAGGTACAAAAGCCAGACT *****</p>
Exon 2 0610- 0669 AS:155 - 175	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>ATGAAGAAGTACACTCAGTTCTACACTTCAGAGACTTTGTCAATGCTAGCAGCAAGTACC ATGAGGAAGTACACTCAGTTCTACACTTCAAAGACTTTGTCAATGCTAGCAGCAAGTACC ATGAACAGGTGCACTCGTTTTGTCATTTAGAGATTTGTTAAACGCCAGCAGCAAGTATC ATGAACAAGTGCACTCGATTTGGCATTTTAAAGACTTTGTTAATGCCAGCAGCAAGTATC AGCAGGAAGTATTATCTGTTCTGGCTTTGAAGACTTCAATTAATGCCAGCGGAAATATC TGGAGCAAGTACTTTTAACACTGGGCTTTAAAGATTTTCTCAATGCTAGCTCAAAGTACC *****</p>
Exon 2 0670- 0729 AS:175 - 195	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>AGGTTACCACCATTCACAATCTCTTTCGAAAGCTGACCCATCGTCTATTTCAGGAGGAAT AGGTGACAACCATTACAACCTCTTTCGAAAGCTGACCCATCGCCTATTTCAGAAGGAAT AGATCCTGACCATTACAACCTCTTTCGGAAGCTGACGCATCGCCTCTTCAGGAGGAAC AAATCACGACCATTACAATCTCTTTCGTAAGCTGACTCATCGCCTTTCAGGAGGAAT AGCTTATGACTGTTCAACCTCTTTCGAAACTCACTCATCGCCTTTCAGGCGCAAT AGATATTAACATTACAATGTCTTTCGGAATGACTCACAGGCTTTCAGGAGGAAT *****</p>
Exon 2 0730- 0789 AS:195 - 215	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>TTGGGTACACACTTCGGTCAGTTAATGGCCTTTATATTCAGAAGCAGTTTCCCATACGGC TTGGGTACACACTTCAGTCAGTTAATGACCTTTATATTCAGAAGCAGTTTCCCATACGGC TTGGGTACACACTTCGGTCAGTCAATGACCTTACGTCAGAAGCAGTTTCCCATACGGC TTGGGTACACTGCAGTCAGTCAATGACCTTACGTCAGAAGCAGTTTCCCATACGGC TTGGGTACACTGCAGTCAGTCAATGACCTTTATTCAGAAGCAGTTTCCCATACGGC TTGGTTATACTCTGAGGTCTGTCAACGATCTTTATATTCGTAAGGACTTTTCTATTCTGT TTGGTACACCTTGAGATCAGTCAATGACATATATGTGAAAAGGGACTTTTAAATTCGAC *****</p>

Fortsetzung der Tabelle 29

Exon 2 0790-0849 AS: 215 - 235	Maus Ratte Kaninchen Mensch Huhn Frosch	AGGACTTCAAAGCTGCTATGAGAGAGTTTACTTTGCTGAGGCCCAGGAGGCTAACTTCC AGGACTTCAAAGCTGCCATGAGAGAGTTTACTTTGCTGAGGCCCAGGAGGCTGACTTC AGGACTTCAAAGCTAAAGTAAGAGAGTATTACTTTGCTGAGGCCCAGGCGGTGACTTCT TTGACTTCAAGACTAAAGTAAGAGAGTATTACTTTGCTGAGGCCCAGATAGCTGACTTCT ATGATTTCAAGAAACAATATGAAAACATACTACTTTCGCTGATGCCCAACACAGCTGATTTCC AGCCTTCAAGAACAACCTGAAAACACTACTACTTTGCAGAGGCACAGACGGTAGATTTTC ...****...*...*****...*...****..
Exon 2 0850-0909 AS: 235 - 255	Maus Ratte Kaninchen Mensch Huhn Frosch	CAGATCCTGCCTTTCATATCAAAGGCCAAACACCACATTTCTGAAACTCACCAAGGGCCTT CAGATCCTGCCTTTCATATCAAAGGCCAACCAGCCACATTTCTGAAGCTCACCAGGGCCTC CGGACCCTGCTTTTATATCAAAGGCCAAACACCACATCTTAAAGGTCACCAAGGGCCTC CAGACCTGCCCTTCATATCAAAGCCAAACACCACATCATGAAGCTCACCAGGGCCTCA CAGATCCTAACTTCAATAACCAAAACCAATGAACGCATCTTGAAGCTGACCAAAAGGATTA GATACAAAGATTTCTACTAAAGCCAAACAGTATCCAGCAGCTCCTAAGGGCTTA ...*...****...*...*****...*...****..
Exon 2 0910-0969 AS: 255 - 275	Maus Ratte Kaninchen Mensch Huhn Frosch	TAAAAGAAGCTCTGGAGAATATAGACTCTGCTACTCAGATGCTGATTCTGAACTGCATCT TAAAAGAAGCTTTGGAGAATACAGACTCTGCTACTCAGATGATGATTCTGAACTGCATCT TAAAAGAAGCGTTGGAGAAGCTAGACCTTCTACTCAGATGATGATTCTGAACTGCATCT TAAAAGATGCTCTGGAGAATATAGACTCTGCTACTCCAGATGATGATTCTCAACTGCATCT TAAAAGAAGCTCTTGTGAATGTAACCCCAACAGCTTATGATGATTCTGAACTGCATCT TAGAAGAAGCAGCTGACAAATGTGACCCAGCATTACTCATGTTGCTGCTGATTTGATCT *...*...****...*...*****...*...****..
Exon 2 0970-0978 AS: 275 - 278	Maus Ratte Kaninchen Mensch Huhn Frosch	ACTTCAAAG_ Intron 2 ACTTCAAAG_ Intron 2 ACTTCAAAG_ Intron 2_(?) ACTTCAAAG_ Intron 2 ACTTCAAAG_ Intron 2_(?) ACTTCAAAG_ Intron 2_(?) *****
Exon 3 0979-1038 AS: 278 - 298	Maus Ratte Kaninchen Mensch Huhn Frosch	GAACTTGGGTGAATAAAATCCAGTAGAAATGACGCACAACCACAACCTCCGGCTGAATC GAGCTTGGATGAATAAAATCCAGTAGAAATGACGCACAACCACAACCTCCGGCTGAATC GAACCTGGGTGAATAAAGTTCCECGTGAAATGACACACAACCACAACCTCCGGCTGAATC GATCCTGGGTGAATAAATCCAGTAGAAATGACGCACAACCACAACCTCCGGCTGAATC GAACGTGGGAGAATAAAGTTCCECGTGAAATGACACAACCACAAGGAGAACTTTTCGACTGAATC GTACTTGGGAGAACAAGTTTCAGTGGAGTATACGCAAAACATGAATTTTCGGCTAAATC *...*...****...*...*****...*...****..
Exon 3 1039-1098 AS: 298 - 318	Maus Ratte Kaninchen Mensch Huhn Frosch	AGAGAGAAGTAGTCAAAGTCTCCATGATGCAAACTAAAGGGAATTCCTTGCAGGCAAAATC AGAGAGAAGTAGTCAAAGTCTCCATGATGCAAACTAAAGGGAATTCCTTGCAGGCAAAATC AAAGGGAGGTGGTCAAGGTTTCTATGATGCAGACCAAGGGTAACCTCCTCGCAGCAAAATC AGAGAGAAGTAGTCAAAGGTTTCCATGATGCAGACCAAGGGTAACCTCCTCGCAGCAAAATC AAAAGCAAAACAATAAAGGTTTCTATGATGCAGACTAAAGGGAATTCCTAGCTGCTGCAC AAAAGGAGCTGGTGAAGTACCAGTATGAGACAAGGGAAACCTCCTTGGTTCGAGCTC *...*...****...*...*****...*...****..
Exon 3 1099-1158 AS: 318 - 338	Maus Ratte Kaninchen Mensch Huhn Frosch	ACCAGGAGCTAGACTGTGACATTTCTCCAGCTGGAGTACGTAGGGGGCATCAGCATGCTT ACCAGGAGCTAGACTGTGACATTTCTCCAGCTGGAGTACGTAGGGGGCATCAGCATGCTTA ACCAGGAGCTGGACTGTGATGTTCTCCAGCTGGAGTACGTAGGGGGCATCAGCATGCTTA ACCAGGAGCTGGACTGTGATGTTCTCCAGCTGGAGTACGTAGGGGGCATCAGCATGCTTA ATCCAGAGCTGGACTGTGGCGTATCCAGCTTCCATTTGTGGGGAACATCAGCATGCTT ACCCGAGTATGATTGTGGGTTTTCAGTACCATATGTTGGGGAACATCAGCATGCTTA *...*...****...*...*****...*...****..
Exon 3 1159-1218 AS: 338 - 358	Maus Ratte Kaninchen Mensch Huhn Frosch	TTGTAGTTCACGGAAGCTATCGGGAATGAAGACTCTTGAAGCAGCTTACACCCAGC TTGTAATTCCACGGAAGCTGTGCGGGAATGAAGACTCTTGAAGCAGCTTACACCCAGC TTGTGGTCCCACATAAGCTGTGCGGGATGAAGACTCTTGAAGCAGCTGACACCCAGC TTGTGGTCCCACACAAAGATGCTGCGGATGAAGACTCTGAAAGCCCTCAGACCCCGG TTGTACTTCCACACAAGCTCTCTGGCATGAAAGCCCTAGAAAAGCAAATAACCCCTCAGC TTGTACTGCCTCATAACTCTCTGGCATGAAACTTCTGGAAAAGCAGATCAGTCCCGAGC ****...*...****...*...*****...*...****..
Exon 3 1219-1252 AS: 358 - 369	Maus Ratte Kaninchen Mensch Huhn Frosch	TGGTGGAGAGATGGCAAAAAGCATGACAAACAG_ Intron 3 TAGTGGAGAGATGGCAAAAAGCATGACAAACAG_ Intron 3 TGGTGGAGAGATGGCAAAAAGCATGACAAACAG_ Intron 3_(?) TGGTGGAGAGATGGCAAAAAGCATGACAAACAG_ Intron 3 TGGTGGAAAATGGCAGAAGAGCATGACGAACAG_ Intron 3_(?) TGGTAGAACGGTGGCAGAATATCATGACAAACAG_ Intron 3_(?) ****...*...****...*...*****...*...****..
Exon 4 1253-1312 AS: 369 - 389	Maus Ratte Kaninchen Mensch Huhn Frosch	AACCTCGAGAGGTACTTCTGCCAAGTTTAAGCTTGAAGAAGTTACAACCTGGTGGAGG AACCCGAGAGGTACTTCTGCCAAGTTTAAGCTTGAAGAAGTTACAACCTGGTGGAGG AACCCGAGAGGTGCTCCTTCTAAATTCAGCTGGAGAAGAACTACAACCTGGTGGAGG AACCTCGAGAAGTCTTCTGCCAAGTTTCAGCTGGAGAAGAACTACAACCTAGTGGAGG AACCCGAGAGGTACTTCTGCCAAGTTTAAGCTTGAAGAAGTTACAACCTGATTGGTT GACAAGGGAGGTCTTCTACTCTGTTTCAACTGGAGAAGAACTACTGACTTACAGAAAG *...*...****...*...*****...*...****..

Fortsetzung der Tabelle 29

Exon 5 1878- 1837	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>ACACTCTCTCCTCAG---GCAGCCTCCTCAGCCTTTATTCGTCAACGGCCTCTGCTCAGC ACACTTTCTCCTCAG---TTAGCCGTACCAGTC-----AAGGTGTCTCCTCAGC CTCCAAGCACACCTGCACAGTACCCCTGGGCACCTCCTTCCAGGCAGCCTGCAGGGTAT ATCCTTGACACCTGA-CTCTGTCACTAAAGCCTTTCTCCACCAGGCCCTCATCTGAAT TAAGAAATATCTCAA---GTAACCTAC---TCTGTGTTTTACTTCACTCTGTAATGT TTCCCACTACCCAAAAAATAAGTTTACTAAAATGTGTGTTAGAAAATTAAGAACAAT</p> <p style="text-align: center;">. MER105-ELEMENT -> .</p>
Exon 5 1838- 1897	Maus Ratte Kaninchen Mensch Huhn Frosch	<p>CAGGCCCTTCCAGACAAAAGGCACTGAGGAACTTGCACCCCTTCCAG----- CAGGCCCTCCATATGGATGGTACTGAGGAACTCTCACCC-TTCCAG----- ACCTACTTCCGTCACTCAAAGCCTTTCTCAACTAGGGCCCTTTCTGAGCACAGAAATGA ACCAAGCACAGAAATGAGTGGTGTGACTAATTCTTACCTCTCCAA----- AATCAATAAA-ATATG-AAGACTATCTGGA(A)_n GCTTAATAAACACATACAAAATTATA(A)_n</p> <p style="text-align: center;">*</p>
Exon 5 1898- 1957	Maus Ratte Kaninchen Mensch	<p>GGAGGGTGTC-CAAGTAGCACCGTTCCTG-----CCCAGTGGGAAAGCAATCCCAACATG GGAGGATGTC-CAAGGAGCACCATTCTTG-----CCTAGTGGAAAAGAACTCAACATG GGAGTGTGACAAAGTGTTCACCATCTGTGATGTTCCCAAGGGAGAAGCAACCTTGAGACA GGAGGGT-ACACAAC TAGCACCATTCTTGATGT--CCAGGGAAGAAGCCACCTCAAGACA </p>
Exon 5 1958- 2017	Maus Ratte Kaninchen Mensch	<p>TCTGAG-----ATGTTCCGACTT-----T TCTGAG-----ACATTTGGACTTAATTC----AAGGCTGAACCTTT GACAGGCGTGGAGTGCCCTGGGCTAACGGTACGGTTTAATTTCTCTCAAGGCCCACTTTC TATGAG---GGGTGCCCTGGGCTAATGTTAGGGCTTAATTTCTCAAAGCCTGACCTT </p>
Exon 5 2018- 2077	Maus Ratte Kaninchen Mensch	<p>TAAATTCATG-TAAATATCTCCTCTTCTGCTG--TGTG-----TACATAGGAC TAAATTCATG-TAAATATCATCTGTTCTGCTT--TGTG-----TCAATAGGGC CAAATCCACG-TGAA-GTCATCTGTCCCTCCTGCTGGTGCCTCGCTGAGACCCAGAGGAC CAAATCCATGATGAATGCCATCAGTCCCTCCTGCTGTTGCCTCCCTGTGACCTGGAGGAC </p>
Exon 5 2078- 2137	Maus Ratte Kaninchen Mensch	<p>TAC-----ATAGTCTATCTAGAGATAAATAAAC--AACCATATTTACCA TACC-----ATAGCCTATTCTAGAGACAAATAAAC--AACCATATTTACTA AGTGCAGCATGTGGCCATGTCTTCCACCCTAAAGATAAATAAATGTAGCCACTTTTCATG AGTG-----TGTG-CCATGCTCCCATACTAGAGATAAATAAATGTAGCCACTTTACTG </p>
Exon 5 2138-	Maus Ratte Kaninchen Mensch	<p>TGA(A)_n AGCA(A)_n TGA(A)_n TGA(A)_n .</p>

Zeichenerklärung zu der Tabelle

Die HCII-cDNA Sequenzen für dieses Alignment wurden verschiedenen Publikationen entnommen.¹⁶¹

Direkte und inverse Sequenzwiederholungen sind fett dargestellt und mit einem Pfeil markiert. Für inverse Wiederholungen wurden zusätzlich kursive Zeichen verwendet. Das ATG-Startkodon ist in grünen Buchstaben eingetragen. Das Kodon für die erste Aminosäure des reifen Proteins und die Stopkodons sind fett dargestellt und gelb hinterlegt. Im 5'-Bereich der cDNAs ist das jeweilige Leseraster bis zum Ende einer Deletion im HCII des Frosches rot und schwarz gedruckt, um die Rasterverschiebungen zu verdeutlichen. Diese Rasterung wurde in der menschlichen cDNA-Sequenz vollständig durchgeführt, um eine Zuordnung der Aminosäurepositionen zu erleichtern, die sich auf die menschliche Sequenz bezieht.

Übereinstimmung von Positionen bei allen Arten sind durch ein Sternchen unter diesen Positionen markiert. Positionen, die nur bei Mammaliern (Maus, Ratte, Kaninchen und Mensch) identisch sind, sind durch einen Punkt gekennzeichnet. Ein mit dem Programm 'RepeatMasker2' in der menschlichen cDNA identifiziertes repetitives Element (MER105, Position 1906-2007) ist gelb hinterlegt.

D.7 Heparinkofaktor II-Transkripte in verschiedenen Gewebetypen

Das HCII-Gen wird in Lebergewebe exprimiert. In weiteren Untersuchungen von Schmidt¹⁶⁶ wurden auch im Lungengewebe HCII-mRNAs mit Hilfe von RT-PCR nachgewiesen. Diese Transkripte scheinen jedoch nur in sehr geringer Konzentration vorzuliegen, da sie mittels Northernblot nicht detektiert werden konnten. Ein falsches positives Ergebnis aufgrund von Amplifikation chromosomaler Fragmente kann nur durch eine PCR über Exon/Introngrenzen hinweg ausgeschlossen werden. Um diese Transkripte definitiv und eventuell auch in anderen Geweben nachweisen zu können, wurden parallel mRNA-Isolate aus Leber, Lunge, Herz, Niere, Plazenta und Gehirn getestet.

Zuerst synthetisierte Schmidt¹⁶³ cDNAs mit Omniscript-reverser Transkriptase und jeweils 0,1 µg mRNA und daran anschließend ein 1207 Bp-Produkt mit den Primern USC13 und USC15. In dieser PCR entstanden viele unspezifische Fragmente.

Aus diesem Grund wurde in dieser Arbeit eine nested PCR mit den Primern USC14 und USC16 sowie dem USC13/15-Produkt als Matrize durchgeführt (erwartetes PCR-Produkt: 1113 Bp). In der Abbildung 27 sind die im Agarosegel aufgetrennten PCR-Produkte und in der folgenden Skizze sind die Positionen der Primer dargestellt. Die Angabe der Positionen erfolgte nach Herzog.¹⁶⁷ Zusätzlich wurden einige Restriktionsschnittstellen in die Skizze eingetragen (B=*Bam*HI, D=*Dra*I und S=*Sal*I).

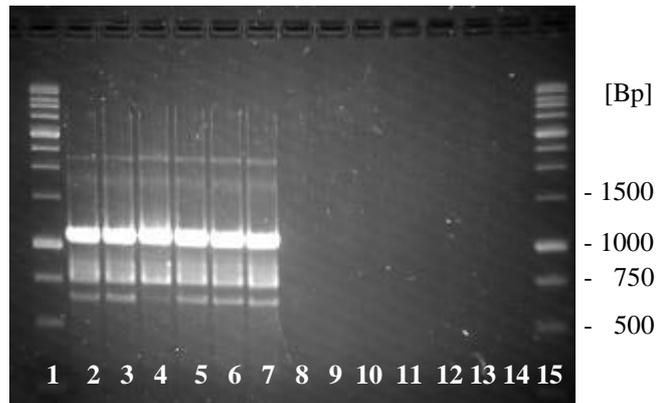


Abbildung 27: HCII-Transkripte in unterschiedlichen Geweben

Gezeigt sind aufgetrennte PCR-Produkte in einem 1,5 % Agarosegel nach einer Elektrophorese. In den Spuren 1 und 15 sind jeweils 0,25 µg 'kB-Leiter' aufgetragen. Die anderen Spuren enthalten jeweils 5 µl PCR-Produkte (USC14-USC16). In den Spuren 2-7 und 8-13 sind jeweils die Ansätze von Leber, Lunge, Herz, Niere, Plazenta und Gehirn aufgetragen. Die Ansätze 8-13 wurden mit hitzeinaktivierten reverse Transkriptase-Ansätzen durchgeführt. Die Spur 14 enthält eine Negativkontrolle, in die zu Beginn keine mRNA gegeben wurde.

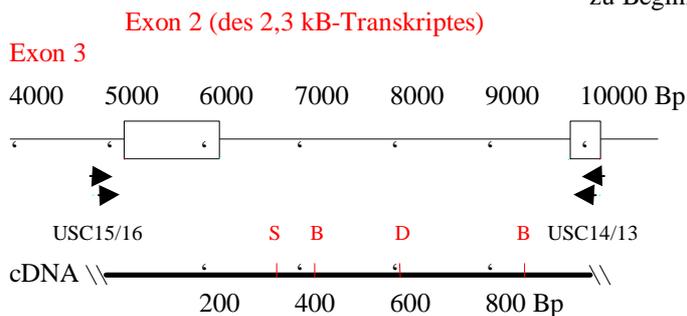
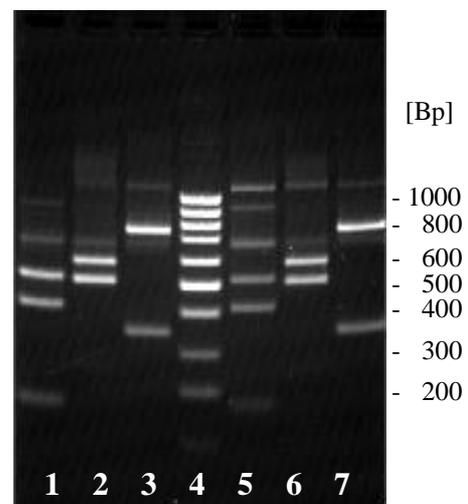


Abbildung 28: Spaltungen von RT-PCR-Produkten

Dargestellt ist ein 2,5 % Agarosegel nach einer Elektrophorese von Restriktionsspaltungsansätzen von RT-PCR-Produkten (durchgeführt mit den Primern USC14 und USC16). In der Spur 4 ist 0,25 µg der '100 Bp-Leiter' aufgetragen. Die Spuren 1-3 zeigen gespaltenes Produkt von Leber-mRNA-Ausgangsmaterial, während in den Spuren 5-7 Spaltungsansätze einer Herz-RT-PCR aufgetrennt sind. Die Spaltungen wurden mit *Bam*HI (Spuren 1+5 [partielle Spaltung]), *Dra*I (Spuren 2+6) und *Sal*I (Spuren 3+7) durchgeführt.



¹⁶³ U. Schmidt, unveröffentlichte Daten

Beispielhaft wurden zwei der erhaltenen Produkte mit unterschiedlichen Restriktionsenzymen gespalten und in einer Gelelektrophorese aufgetrennt. Für die Spaltung der RT-PCR-Produkte von Leber- und Herz-mRNA ließen sich die Enzyme *Bam*HI, *Dra*I und *Sal*I verwenden. In der Abbildung 28 sind die aufgetrennten Spaltprodukte gezeigt. Die erwarteten drei Spaltprodukte für eine *Bam*HI-Spaltung sind 171 Bp, 418 Bp und 524 Bp lang. Bei der Spaltung mit *Dra*I entstehen zwei Fragmente (513 Bp und 600 Bp), ebenso wie bei der Spaltung mit *Sal*I (339 Bp und 774 Bp; siehe obige Skizze). Obwohl die *Bam*HI-Spaltungen teilweise unvollständig sind, kann dem Gelbild (Abbildung 28) entnommen werden, daß Spaltungen der cDNAs aus beiden Organen die berechneten Fragmente ergeben. Damit ist sicher, daß HCII-Transkripte mit einem verlängerten Exon 2 (Exon 1*; die Sinnprimer können nur mit einer Region der cDNA hybridisieren, die eigentlich eine Intron 1-Region ist) in Leber und Herz vorhanden sind. Aufgrund der identischen PCR-Produkte können diese Transkripte auch in den anderen getesteten Gewebetypen erwartet werden.

Die Frage, die sich bei diesem Ergebnis zwangsläufig stellt ist, in welcher Menge HCII-Transkripte in den verschiedenen Gewebetypen vorkommen. Um semiquantitative Aussagen darüber machen zu können, wieviel Transkripte vorliegen, wurden Fragmente von HCII- und Antithrombin-kodierenden Transkripten in den einzelnen Geweben koamplifiziert. Antithrombin wurde als Referenz gewählt, weil beide Proteine als Hauptzielproteinase Thrombin inhibieren. Zur Analyse wurden zunächst die mRNAs der verschiedenen Gewebe mit einem Oligo(dT)₁₆-Primer in Gegenwart von Omniskript-Reverser Transkriptase revers transkribiert. Die Amplifikation der PCR-Produkte erfolgte über Exon/Introngrenzen hinweg. Zum Erhalt eines 300 Bp-HCII-Fragmentes kamen die Primer USC3 und USC4 und für Antithrombin die Primern ATIII12 und ATIII148 zur Verwendung (409 Bp-Fragment). In Abbildung 29 sind die im Agarosegel aufgetrennten PCR-Produkte nach 25, 28, 31 und 34 PCR-Zyklen zu sehen. Die Lage der Primer und der Exongrenzen (rot) ist in die folgende Skizze eingetragen. Auf den Agarosegelen ist erwartungsgemäß zu erkennen, daß die größte Menge an HCII- und Antithrombin-Transkripten in der Leber vorhanden ist. Die Zahl der HCII-mRNAs in anderen Geweben ist um mehrere Größenordnungen niedriger. Eine Bandenstärke, die der Bande von Leber nach 25 Zyklen entspricht, wird mit Lungenmaterial erst nach etwa 9 weiteren Zyklen erreicht. In den anderen Geweben sind noch weniger Transkripte vorhanden. Die Diskussion findet im Teil E.4.2 statt.

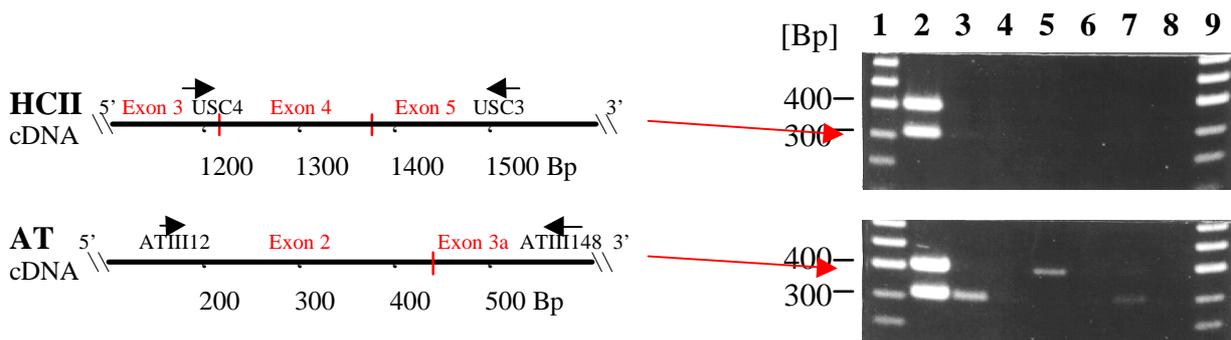


Abbildung 29: Gewebsspezifische HCII/AT-Transkripte

Gezeigt sind 2,5 % Agarosegele nach einer Elektrophorese. In den Spuren 1 und 9 wurden jeweils 0,5 µg '100 Bp-Leiter' aufgetragen. Die anderen Spuren enthalten jeweils 5 µl PCR-Produkte. In den Spuren 2-7 sind jeweils die Ansätze von Leber, Lunge, Herz, Niere, Plazenta und Gehirn aufgetragen. Die Spur 8 enthält jeweils eine Negativkontrolle, in die zu Beginn keine poly(A)⁺-RNA gegeben wurde. Das obere Gel zeigt das PCR-Ergebnis nach 25 Zyklen, das zweite Gel nach 28 Zyklen, im darauffolgenden Gel ist das Ergebnis nach 31 Zyklen zu sehen und im unteren Gel wurden die Produkte nach 34 Zyklen aufgetragen.

D.8 Startpunkte von menschlichen Heparinkofaktor II-Transkripten

Die Analyse der menschlichen HCII-mRNAs wurde zuerst u. a. mittels Northernblot durchgeführt.¹⁶⁴ Mit dieser Methode läßt sich die ungefähre Länge der Transkripte (2,3 kB) bestimmen, der genaue Startpunkt der Transkripte kann jedoch nicht ermittelt werden. Aus diesem Grund untersuchten Ragg und Preibisch die Startpunkte der HCII-mRNA mit Hilfe von S1-Nuklease- und RNaseA/T1-Kartierung. Die Ergebnisse von beiden Methoden stimmen grundsätzlich überein. Es lassen sich mehrere Startpositionen der Transkripte entnehmen, sowie eine Länge des ersten HCII-Exons von 25-115 Bp. Eine Ausnahme bildet der am weitesten stromaufwärts gelegene Transkriptionsstartpunkt (entsprechend einer Länge des Exon 1 von etwa 115 Bp), der nur mit der S1-Kartierung erhalten wird.¹⁶⁵

Für die Analyse der HCII-Genstruktur und des Promotors war es sehr wichtig, die Startpunkte der mRNAs zu kennen. Aus diesem Grund wurden in dieser Arbeit die über Nukleaseverdau erhaltenen Ergebnisse mit weiteren, unabhängigen Methoden (A-PCR- und T-PCR) in Hinblick auf die fragliche Position überprüft (D.8.2).

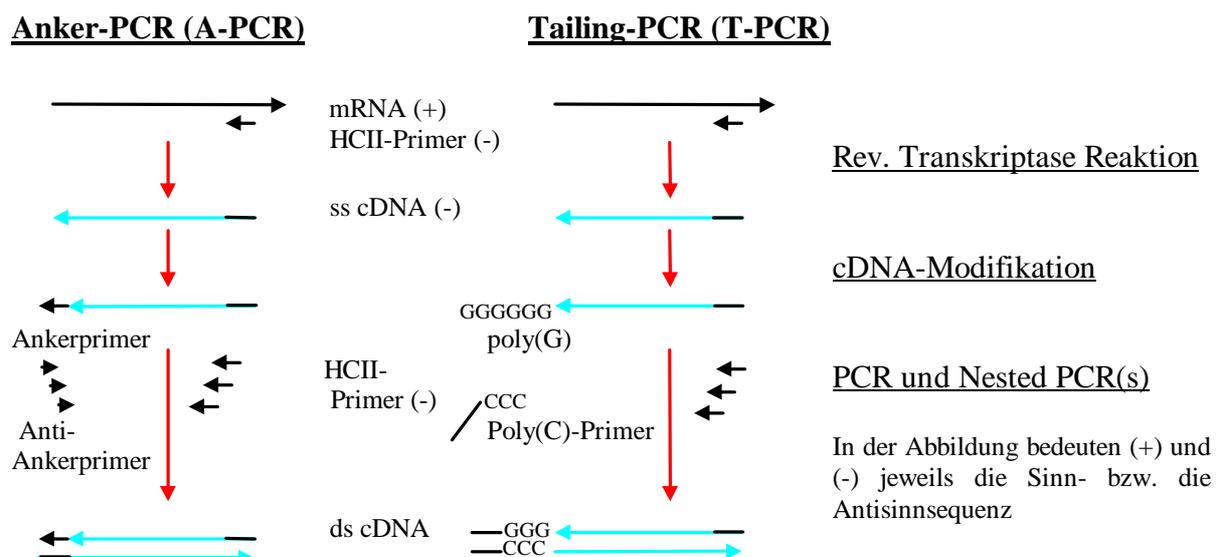
Während der Arbeit entdeckte U. Schmidt¹⁶⁶ HCII-Transkripte im Lungengewebe. Die Überprüfung der Startpunkte dieser Transkripte, die eine im 5'-Bereich veränderte Genstruktur aufweisen, wurde parallel durchgeführt (D.8.3). Mit Hilfe der erhaltenen Daten kann die veränderte Genstruktur der beiden Transkripte verglichen werden, und es werden Hinweise auf eventuell vorhandene unterschiedliche Promotorbereiche erhalten.

Im folgenden wird zuerst kurz die verwendete Methodik der Bestimmung von Startpunkten beschrieben, die sich sehr von den Nuklease-Kartierungsmethoden unterscheidet.

D.8.1 Methodik zur Identifizierung von Transkriptionsstartpunkten

Die Startpunkte von HCII-mRNAs wurden über die Methode der Anker-PCR (A-PCR) und Tailing-PCR (T-PCR) ermittelt. In der folgenden Abbildung ist die Synthese dieser Produkte dargestellt, um den Aufbau der analysierten PCR-Produkte zu verdeutlichen.

Abbildung 30: Schema der Anker- und der Tailing-PCR



¹⁶⁴ Ragg, 1986

¹⁶⁵ Ragg & Preibisch, 1988

¹⁶⁶ Schmidt, 1997

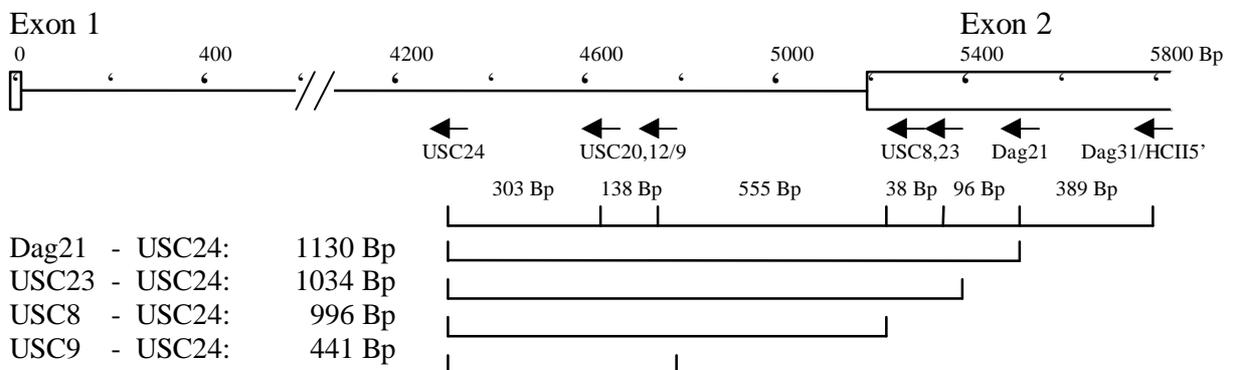
Die Sequenz des menschlichen HCII-Gens im Bereich des Beginns von Exon 2 ist nachfolgend dargestellt, um die Ergebnisse der A- und T-PCR besser verständlich zu machen. Die Numerierung wurde von Herzog¹⁶⁷ übernommen.

Abbildung 31: Menschliche HCII-Sequenz im Bereich des Beginns von Exon 2

```

4311  5' caagagaatt gcttgagtcc aggaggccga agttgcagta agccgagatc gcgccactgc
      <- tcttaa cgaactcagg tctctccggc USC24
4371  actccagcct ggggtgacaga gtgagactcc atttcaaaaa ataaaaacaa caaaagccaa
4431  ttacaacaac aacaacaaaa aaacaacgaa ttaacaacc ccaaagattg cacaaatttc
4491  aagtatcttt agaatatggt ttcagaaagc ctggcccatg gacatttttc aacagcatct
4551  ccattgcaaa ggtggaatgg tgtgagtcac acaggcatgg ctgagtccca ctaatgcaca
4611  tcccttctag gtactctcca atcaccagcc ccagggtgcc actcaagccc agctcttagt
      <- catgagaggt tagtggtcgg ggt USC20
4671  gaggtttccc tgactctctg ggcacttcca ctctaccac acagggtaga gccacacccc
4731  tttccgtacc cccatgtgct ctggcagcat tattttgaga gccttcgctt tactgcacgt
      <- gta ataaaactct cggaagcгаа a USC9/12
4791  ctgtcccata tgtcccctga ctgggtccatg agcccctggg ggaactttg tctctggtaa
4851  ctaaacactg tctggagggtg gtggacaagg tgtctggaga aaaacaaact cctccctggg
4911  atgcttgagc tcccaggatt ctagaaggtt agttttgcaa acctttaaag aagggatttt
4971  catcaagggg cccacagatc cttcattgag gtttatgagt cccacatcaa aggttgggtg
5031  tctatctaca tcagattctc ttaaagtcca tgatcctaaa acagttaaga actaatgctg
5091  tgagggcctc ttcttgggtc aaagccacag ggaacctgcc atgtggatgc tgcagcgggg
5151  tgtggatcag ccaggccgcc tttcactgtg ttctgttttc cctcccaCT TTAGCTCCGC
5211  CAAAATGAAA CACTCATTAA ACGCATTCT CATTTTCCTC ATCATAACAT CTGCGTGGGG
5271  TGGGAGCAAA GGCCCGCTGG ATCAGCTAGA GAAAGGAGGG GAAACTGCTC AGTCTGCAGA
      <- ttgacgag tcagacgtct
5331  TCCCAGTGG GAGCAGTTAA ATAACAAAAA CCTGAGCATG CCTCTTCTCC CTGCCGACTT
      aggggt USC8 <- tattgttttt ggactcgtac ggag USC23
5391  CCACAAGGAA AACACCGTCA CCAACGACTG GATTCCAGAG GGGGAGGAGG ACGACGACTA
      <- tgat
5451  TCTGGACCTG GAGAAGATAT TCAGTGAAGA CGACGACTAC ATCGACATCG TCGACAGTCT
      agacctggac ctcttctata Dag21
5511  GTCAGTTTCC CCGACAGACT CTGATGTGAG TGCTGGGAAC ATCCTCCAGC TTTTTCATGG
5571  CAAGAGCCGG ATCCAGCGTC TTAACATCCT CAACGCCAAG TTCGTTTCA ACCTCTACCG
5631  AGTGCTGAAA GACCAGGTCA ACACTTTCGA TAACATCTTC ATAGCACCCG TTGGCATTTC
5691  TACTGCGATG GGTATGATT CCTTAGGTCT GAAGGGAGAG ACCCATGAAC AAGTGCATTC
5751  GATTTTGCAT TTTAAAGACT TTGTTAATGC CAGCAGCAAG TATGAAATCA CGACCATTCA
5811  TAATCTCTTC CGTAAGCTGA CTCATCGCCT C TTCAGGAGG AATTTTGGGT ACACACTGCG'
      <- gtagcgga gaagtcctcc ttaaa Dag31
      <- tctcc ttaaaaccc HCII5'

```



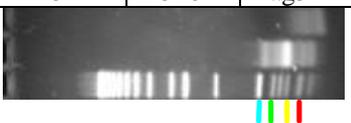
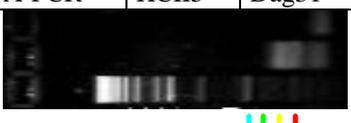
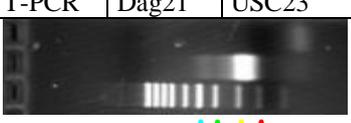
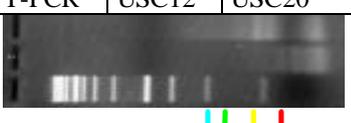
Die verwendeten Antisinnprimer (rot) sind jeweils unter der Sequenz in 3'-5'-Richtung eingetragen und ihre Position ist in der Übersichtsskizze angegeben. Exonbereiche (Exon 2 der bekannten 2,3 kB mRNA) sind mittels großer und fetter Buchstaben dargestellt. Das ATG-Startkodon ist in Grün gedruckt.

D.8.2 Startpunkte der HCII-Transkripte in menschlicher Leber

D.8.2.1 Vergleich von A-PCR- und T-PCR-Produkten

Wie bereits erwähnt, wurden im Verlauf der Arbeit HCII-Transkripte auch in der Lunge entdeckt, die sich im 5'-Bereich von den zuvor bekannten Transkripten unterscheiden. Aus diesem Grund wurde auch die RNA aus Leber auf Startpunkte dieser Transkripte untersucht. Da sich ein Problem mit der Exon- bzw. Intronbezeichnung der 5'-gelegenen Sequenzen der beiden unterschiedlichen Transkripte ergibt, wird das erste Exon des längeren Transkripts in dieser Arbeit als Exon 1* bezeichnet (dieses entspricht Sequenzen vom 3'-Bereich des Intron 1 und dem Exon 2 des kürzeren Transkripts). Zur Bestimmung der mRNA-Startpunkte von hHCII in Leberzellen wurde entweder die A-PCR- oder die T-PCR-Methode verwendet. Darüber hinaus wurde bei den Versuchen unterschiedliches Ausgangsmaterial eingesetzt (zum einen Gesamt-RNA aus HepG2-Zellen und zum anderen poly(A)⁺-RNA aus menschlicher Leber). Aufgrund der unterschiedlich durchgeführten Experimente (Material, Methode, Negativkontrollen) ist eine parallele Auflistung dieser Daten zu den jeweiligen Ergebnissen der Versuche unumgänglich. Die mittels A-PCR und T-PCR sowie mit unterschiedlichen Primern erhaltenen Produkte sind in der Tabelle 30 dokumentiert. Die Amplifikation der PCR-Produkte erfolgte über eine erste oder zweite Nested PCR. Eine zugehörige Legende erläutert die Tabelle näher (s. u.). Die Ergebnisse der ermittelten Bereiche für Transkriptionsstarts sind in der Tabelle 30 zusammengefaßt.

Tabelle 30: 5'-Enden von Leber-spezifischen HCII-mRNAs

Herkunft der RNA	Methode	Primer f. Reverse Transkr.	Primer für die PCR	Primer für 1. Nested PCR	Primer für 2. Nested PCR	Gelstärke, erhaltene Fragmente (ca.)	Negativkontrollen
poly(A) ⁺ RNA aus Leber	A-PCR	HCII5'	Dag31	Dag21	-	1 % Gel 350 Bp, 450 Bp	PCR ohne cDNA
				5 µl Negativkontrolle 5 µl PCR-Produkt 0,25 µg 'Marker VIII'			
poly(A) ⁺ RNA aus Leber	A-PCR	HCII5'	Dag31	Dag21	USC8	1,5 % Gel 320 Bp	PCR ohne cDNA
				5 µl Negativkontrolle 5 µl PCR-Produkt 0,25 µg 'Marker VIII'			
poly(A) ⁺ RNA aus Leber	T-PCR	Dag21	USC23	USC8	-	2 % Gel 250 Bp	PCR ohne cDNA
				5 µl Negativkontrolle 5 µl PCR-Produkt 0,25 µg '100 Bp-Leiter'			
RNA aus HepG2-Zellen	T-PCR	Dag21	USC9	USC20	-	2,5 % Gel PCR-Fragmente nach Amplifikation von DNA aus Gelextraktion	PCR mit Matrize und Anker-Primer, aber ohne USC20 (ebenfalls durchgeführt: PCR ohne Matrize)
				5 µl Negativkontrolle 5 µl '150' Bp Fragment 5 µl '200' Bp Fragment 5 µl '350' Bp Fragment 0,25 µg '100 Bp-Leiter'			
poly(A) ⁺ RNA aus Leber	T-PCR	USC12	USC20	USC24	-	2,5 % Gel > 250 Bp	Durch Hitze inaktivierter Reverse Transkriptase-Ansatz
				5 µl PCR-Produkt 5 µl Negativkontrolle 0,25 µg 'kB-Leiter'			

Die Legende zu der Tabelle folgt auf der nächsten Seite.

Legende zu der Tabelle 30

In der Tabelle sind die in einer PCR, Nested PCR und gegebenenfalls einer zweiten Nested PCR verwendeten Antisinnprimer von links nach rechts aufgelistet. Die auf den Gelbildern zu erkennenden PCR-Produkte wurden also mit dem in der Tabelle am weitesten rechts genannten Primer erhalten.

Die erste Spalte der Tabelle gibt das in der jeweiligen Methode verwendete Ausgangsmaterial an. In der zweiten Spalte ist vermerkt, ob die PCR-Fragmente mittels A-PCR oder T-PCR erhalten wurden. Als nächstes ist der Primer (Primer f. Reverse Transkr.) aufgeführt, mit dem die cDNA-Synthese durchgeführt wurde. In den folgenden Spalten (Primer für die PCR, etc.) sind die verwendeten Antisinnprimer eingetragen. Die PCR-Produkte wurden in einer ersten oder zweiten Nested PCR erhalten und die jeweiligen spezifischen Antisinnprimer sind angegeben. Da die Sinnprimer in den Experimenten jeweils gleich sind (Anti-Ankerprimer bei der A-PCR und poly(C)-Primer bei der T-PCR), wurden für die durchgeführten PCR- bzw. Nested PCR-Reaktionen jeweils nur die spezifischen Antisinnprimer genannt. Falls in der Spalte für die zweite Nested PCR nur ein Strich steht, bedeutet dies, daß keine zweite Nested PCR durchgeführt wurde, und die auf den Gelbildern zu sehenden Fragmente zeigen Amplifikate der ersten Nested PCR. Die siebte Spalte gibt die verwendete Gelstärke und eventuell Bemerkungen zu der Durchführung der Methode an. In der letzten Spalte sind Bemerkungen zu der Durchführung der jeweiligen (teilweise unterschiedlichen) Negativkontrollen aufgeführt.

Die Auftrennung der DNA in den gezeigten Agarosegelen ist von links nach rechts verlaufen. Es wurden unterschiedliche Gelstärken und DNA-Längenmarker verwendet. Um die Fragmentlängen besser abschätzen zu können, ist deshalb unter den Gelen ein Farbcode angegeben. Die Positionen sind am unteren Rand der Agarosegele markiert, an denen die 500 Bp- (blau), 400 Bp- (grün), 300 Bp- (gelb) und 200 Bp-Fragmente (rot) zu sehen sind.

In der sich anschließenden Tabelle 31 sind die nach den Gelen (etwa ± 50 Bp Abweichung) abgeschätzten Startbereiche zusammengefaßt. Zu beachten ist, daß bei der A-PCR durch den Anker-Primer und bei der T-PCR durch die angehängten Nukleotide die PCR-Produkte um 30-50 Bp gegenüber der HCII-Sequenz verlängert sind.

Den Ergebnissen für die Startbereiche der bekannten 2,3 kB-Transkripte, die das Exon 1 enthalten, ist zu entnehmen, daß sie gut mit den von Ragg und Preibisch¹⁶⁵ erhaltenen Daten übereinstimmen (siehe Skizze unter 8.2.2). Es wird in beiden Fällen eine Exon 1-Länge von etwa 30 Bp bzw. 130 Bp erhalten. Der Startbereich der ursprünglich zuerst in Lunge festgestellten längeren HCII-Transkripte im Lebergewebe scheint sich über einen sehr großen Bereich (<700 Bp bis >1000 Bp vor Exon 2) zu erstrecken. Eine ausführliche Diskussion der Ergebnisse wird im Teil E.4.1 und E.4.2 durchgeführt.

Tabelle 31: Startbereich von leberspezifischen HCII-mRNAs

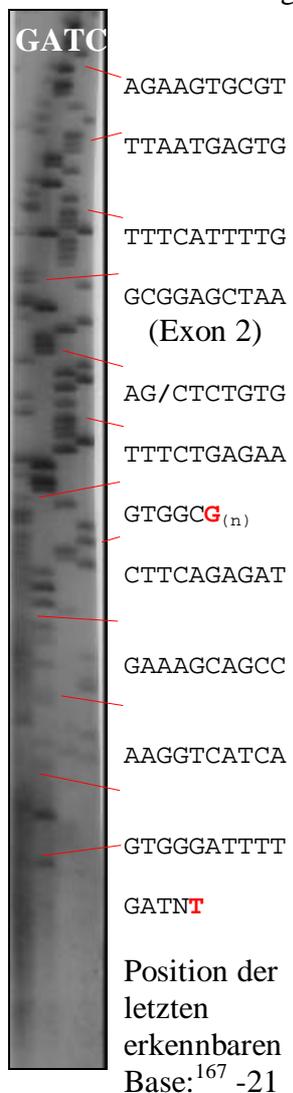
RNA	Antisinnprimer	Fragmentlänge	Geschätzter Start (siehe Abbildung 31)
Leber	Dag21	350 Bp + 450 Bp	Intron 1 ist entfernt (Sequenzierung der nachfolgenden nested PCR). Start: 30 Bp und 130 Bp vor dem 3'-Ende des Exon 1
Leber	USC8	250 Bp + 320 Bp	Intron 1 ist entfernt (Sequenzierungsergebnis unter D.8.2.2). Start: 50 Bp und 120 Bp vor dem 3'-Ende des Exon 1
HepG2	USC20	150 Bp - 350 Bp	650-850 Bp vor Exon 2
Leber	USC24	> 250 Bp	mehr als 1050 Bp vor Exon 2

D.8.2.2 Sequenzierung des 5'-Endes der cDNA aus Leber

Um die Startpunkte der HCII-Transkripte genauer bestimmen zu können, ist es erforderlich einige PCR-Produkte zu sequenzieren. In diesem Teil werden die Sequenzierungen der jeweils längsten PCR-Produkte für das 2,3 kB- und das längere HCII-Transkript beispielhaft gezeigt. Zuerst ist das Ergebnis für die schon zuvor von Ragg und Preibisch¹⁶⁵ charakterisierten kürzeren Transkripte aufgeführt.

Die Sequenzierung fand nach der 'Direct-Blotting'-Methode statt. Parallel zur Auftrennung der Sequenzierprodukte durch das Gel werden die durchgelaufenen DNA-Fragmente direkt auf eine mitlaufende Membran gebロットet. Aus diesem Grund werden zuerst die kurzen Fragmente detektiert (analog zu den moderneren Sequenzierverfahren mit dem ALF oder LICOR) und erscheinen am oberen Rand der Membran. Hier unterscheidet sich die Methode vom konservativen Blotten von Sequenziergelen, bei denen sich die kürzeren Banden im unteren Teil der Membran befinden.

Die Membran (Abbildung 32) zeigt die Antisinnsequenz. Um die Sequenz besser verfolgen zu können, ist nebenstehend die ermittelte Nukleotidabfolge notiert. Rote Striche ermöglichen die bessere Zuordnung zur gezeigten Membran.



Zusätzlich ist parallel die Sinnsequenz (identisch zu der von Herzog *et al.* beschriebenen Sequenz¹⁶⁷) mit dem gefundenen Startbereich dargestellt (Skizze s. u.). Die zuvor mit Nuklease-Kartierung bestimmten Startbereiche (unterstrichene Sequenzabschnitte) wurden von Ragg und Preibisch¹⁶⁵ übernommen und eingetragen.

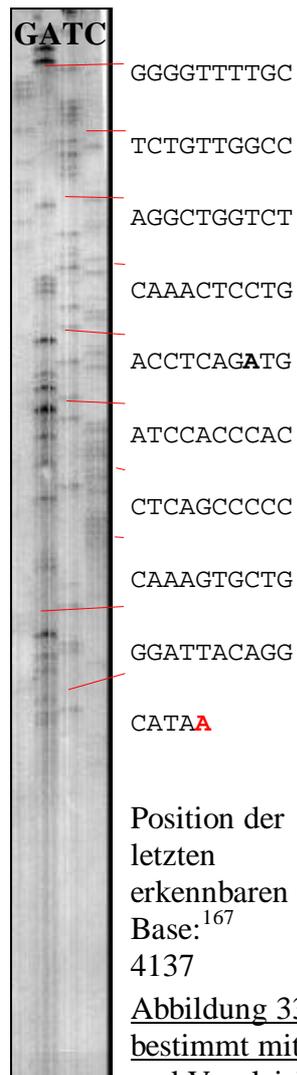
Der Startbereich der 2,3 kB-HCII-mRNAs wurde über Sequenzierung eines T-PCR-Produkts mit dem Primer USC8 bestimmt. Es ist kein festgelegtes Ende der Sequenz zu erkennen, sie kann jedoch eindeutig bis zur Position +25 gelesen werden. Anschließend beginnt in der G-Spur eine Bandenleiter. Die Sequenz läßt sich jedoch anhand der bekannten chromosomalen hHCII-Sequenz bis zur Position -21 verfolgen. In der untenstehenden Skizze ist der Bereich zwischen diesen beiden Positionen in Rot gedruckt. Die 3'-gelegene Position markiert die Stelle in der Sequenz, an der die kürzesten Transkripte beginnen. Es ist gut zu erkennen, daß der gefundene 5'-gelegene Startpunkt (Position -21) mit einer vor Nukleaseverdau geschützten Stelle (unterstrichen) übereinstimmt. Der Bereich der G-Bandenleiter überdeckt fast exakt den mit Nukleaseverdau ermittelten Startbereich ab. Eine Übereinstimmung mit der geschützten Stelle um Position -70 konnte nicht gefunden werden, allerdings setzt sich der Bandenschmier bis über die Position -21 fort, aus diesem Grund müssen auch noch längere Transkripte vorhanden sein.

```

-150 5' TGCAAGCCAC TCTACAGTTA TCAAAGGCAC AGCTGAGGGG
-110 GTTTGTGCTG ACCAAGCTGG TTGCCTGGTG TTTGGATTGG
-70  GACTTATTTA CTTTGAAAA TATGCAGCAA CAGCCCAGCA
-30  CCAAAGTTCA CATCAAAATC CCACTGATGA CCTTGGCTGC
11  TTTCATCTCT GAAGCGCCAC TTCTCAGAAA CACAGAGGTA3'

```

Abbildung 32: 5'-Ende der HCII-mRNA, bestimmt mit dem Primer USC8 und Vergleich mit der genomischen Sequenz



Im Lungengewebe wurden HCII-Transkripte detektiert, deren Exon 1* im zuvor beschriebenen Intron 1-Bereich beginnt.¹⁶⁶

Um festzustellen, ob es auch im Lebergewebe Transkripte gibt, deren Startpunkte ebenfalls im Intron 1 liegen, wurden weitere Untersuchungen mit Hilfe von T-PCR und spezifischen Antisinnprimern, die im Intron 1-Bereich binden, durchgeführt.

Die Nested PCR zur T-PCR und die Sequenzierung wurde mit dem Primer USC24 durchgeführt. Die Sequenz kann anhand der bekannten chromosomalen HCII-Sequenz bis zur Position 4137 verfolgt werden (Abbildung 33), es kann jedoch kein definiertes Ende erkannt werden. Eine Abweichung von der bekannten Sequenz ist an Position 4184 zu erkennen. Statt einer 'G'- (komplementäre Sequenz) ist eine 'A'-Bande (fett) zu sehen. Weiterhin kann der Sequenzierung entnommen werden, daß es auch längere Transkripte geben muß.

Nachfolgend ist ein Sequenzausschnitt (nach Herzog *et al.*,¹⁶⁷ vergleiche: Abbildung 31) um den ermittelten Startpunkt gezeigt. Die Darstellung wurde analog zur Abbildung 32 mit zugehöriger Skizze vorgenommen. Das rote 'T' zeigt die letzte erkennbare Position der Sequenzierung. Dieses Ergebnis und weitere Sequenzierungsergebnisse zur Ermittlung von Transkriptionsstartpunkten im Intron 1-Bereich der 2,3 kB mRNAs werden im Teil E.4 diskutiert.

Position der letzten erkennbaren Base: ¹⁶⁷	4131	5' GGTGGC	TAT	GCCTGTAATC	CCAGCACTTT	GGGGGGCTGA
	4171	GGTGGGTGGA	TCACCTGAGG	TCAGGAGTTT	GAGACCAGCC	
	4211	TGGCCAACAG	AGCAAAACCC	CATCTCTACT	AAACAAAATA ^{3'}	

D.8.3 Startpunkte von HCII-Transkripten in menschlicher Lunge

D.8.3.1 Vergleich von A-PCR- und T-PCR-Produkten

HCII-Transkripte waren bisher nur im Lebergewebe bekannt. Bei einer Überprüfung einer Lungen-cDNA-Bank entdeckte Schmidt jedoch ebenfalls HCII-cDNA-Sequenzen.¹⁶⁶ Die Transkripte unterscheiden sich im 5'-Bereich von den 2,3 kB-Transkripten dadurch, daß das Exon 1 nicht vorhanden ist. Der Startbereich dieser mRNAs liegt vor dem eigentlichen Exon 2 (Bezeichnung aus diesem Grund: Exon 1*). Mit Hilfe von A-PCR und T-PCR wird in dieser Arbeit der Startbereich dieser Transkripte und damit auch der 5'-Bereich der HCII-Genstruktur für Lungengewebe (und zur Kontrolle einer Lungenzelllinie) bestimmt. Bei den Versuchen wurde unterschiedliches Ausgangsmaterial eingesetzt (Gesamt-RNA aus A549-Zellen und Gesamt-RNA und poly(A)⁺-RNA aus menschlicher Lunge), und die Negativkontrollen wurden teilweise abgewandelt durchgeführt. Aus diesem Grund werden die Ergebnisse zusammen mit den Daten der Versuchsdurchführung dargestellt. Die mit den PCR-Methoden und unterschiedlichen Primern (siehe Abbildung 30) erhaltenen PCR-Produkte sind in der Tabelle 32 als Bilder von Agarosegelen zusammengefaßt.

¹⁶⁷ Herzog *et al.*, 1991

Die Ergebnisse sind genau wie die Ergebnisse für HCII-Transkripte in der Leber dargestellt (siehe D.8.2.1). Eine Zusammenfassung der Ergebnisse ist in der Tabelle 33 gegeben.

Tabelle 32: 5'-Enden von HCII-mRNAs aus Lunge

Herkunft der RNA	Methode	Primer f. Reverse Transkr.	Primer für die PCR	Primer für 1. Nested PCR	Primer für 2. Nested PCR	Gelstärke, erhaltene Fragmente (ca.)	Negativkontrollen
RNA aus A549 / poly(A) ⁺ RNA aus Lunge	A-PCR	HCI15'	Dag31	Dag21	USC8	1,5 % Gel A549: 300-1000 Bp Lunge: 300-1000 Bp	PCR ohne cDNA
RNA aus Lunge	T-PCR	USC23	USC9	USC20	-	2,5 % Gel	1) Ansatz ohne Reverse Transkriptase 2) PCR ohne Primer USC20 3) PCR ohne cDNA
				5 µl Negativkontrolle 1 5 µl PCR-Produkt 5 µl Negativkontrolle 2 0,25 µg '100 Bp-Leiter' 5 µl Negativkontrolle 3		120-700 Bp	
				5 µl '200' Bp Fragment 5 µl '300' Bp Fragment 0,25 µg '100 Bp-Leiter' 5 µl '400' Bp Fragment 5 µl '500' Bp Fragment		PCR-Fragmente nach Reamplifikation von DNA aus Gelextraktion	
poly(A) ⁺ RNA aus Lunge	T-PCR	USC23	USC9	USC20	-	2,5 % Gel 120 Bp, 450 Bp, 530 Bp, 850 Bp	Durch Hitze inaktivierter Rev. Transkriptase-Ansatz
poly(A) ⁺ RNA aus Lunge	T-PCR	USC12	USC20	USC24	-	2,5 % Gel 200 Bp, 500 Bp	Durch Hitze inaktivierter Rev. Transkriptase-Ansatz

In der Tabelle sind die in der PCR, Nested PCR und gegebenenfalls der zweiten Nested PCR verwendeten Antisinnprimer von links nach rechts aufgelistet. Die auf den Gelbildern zu erkennenden PCR-Produkte wurden mit dem in der Tabelle am weitesten rechts genannten Primer erhalten.

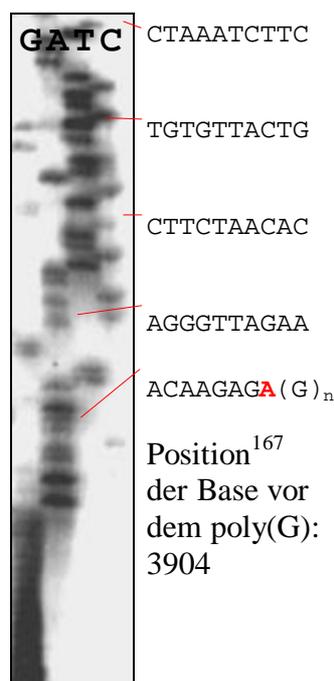
Die erste Spalte der Tabelle gibt das in der jeweiligen Methode verwendete Ausgangsmaterial an. In der zweiten Spalte ist vermerkt, ob die PCR-Fragmente mittels A-PCR oder T-PCR erhalten wurden. Als nächstes ist der Primer (Primer f. Reverse Transkr.) aufgeführt, mit dem die cDNA-Synthese durchgeführt wurde. In den folgenden Spalten (Primer für die PCR, etc.) sind die verwendeten Antisinnprimer eingetragen. Die PCR-Produkte wurden in einer ersten oder zweiten Nested PCR mit den aufgeführten spezifischen Antisinnprimer erhalten. Da die Sinnprimer in den Experimenten jeweils gleich sind (Anti-Ankerprimer bei der A-PCR und poly(C)-Primer bei der T-PCR), wurden für die durchgeführten PCR- bzw. Nested PCR-Reaktionen jeweils nur die spezifischen Antisinnprimer angegeben. Falls in der Spalte für die zweite Nested PCR nur ein Strich steht, bedeutet dies, daß keine zweite Nested PCR durchgeführt wurde und die auf den Gelbildern zu sehenden Fragmente in einer ersten Nested PCR amplifiziert wurden. Die siebte Spalte gibt die verwendete Gelstärke und eventuell Bemerkungen zu der Durchführung der Methode an. In der letzten Spalte sind Bemerkungen zu der Durchführung der jeweiligen (teilweise unterschiedlichen) Negativkontrollen aufgeführt. Die Auftrennung der DNA in den gezeigten Agarosegelen ist von links nach rechts verlaufen. Es wurden unterschiedliche Gelstärken und DNA-Längenmarker verwendet. Um die Fragmentlängen besser abschätzen zu können, ist deshalb unter den Gelen ein Farbcode angegeben. Die Positionen sind am unteren Rand der Agarosegele markiert, an denen die 500 Bp- (blau), 400 Bp- (grün), 300 Bp- (gelb) und 200 Bp-Fragmente (rot) zu sehen sind.

In der sich anschließenden Tabelle 33 sind die nach den Gelen (± 50 Bp Abweichung) abgeschätzten Startbereiche zusammengefaßt (Lage der Primer: Abbildung 31). Zu beachten ist, daß bei der A-PCR durch den Anker-Primer und bei der T-PCR durch die angehängten Nukleotide die PCR-Produkte ca. 30-50 Bp gegenüber der HCII-Sequenz verlängert sind. Den Ergebnissen kann entnommen werden, daß es zumindest für menschliches Lungengewebe eine unterschiedliche HCII-Genstruktur gibt. Im Intron 1 des 2,3 kB-Transkripts ist eine sehr große Region (1 kB) vorhanden, innerhalb der an bestimmten Stellen die Transkription der HCII-mRNA beginnt (Exon 1*). Eine ausführliche Diskussion der Ergebnisse wird im Teil E.4.2 durchgeführt.

Tabelle 33: Startbereich von Lungen-spezifischen HCII-mRNAs

Herkunft der RNA	Antisinnprimer	Fragmentlänge	Geschätzter Start
A549	USC8	300-1000 Bp	300-800 Bp vor Exon 2
Lunge	USC8	300-1000 Bp	300-800 Bp vor Exon 2
Lunge	USC20	100-850 Bp	630-1350 Bp vor Exon 2
Lunge	USC24	200 Bp, 500 Bp	1000-1300 Bp vor Exon 2

D.8.3.2 Sequenzierung des 5'-Endes der cDNA aus Lunge



Die mit Hilfe von A-PCR und T-PCR gefundenen spezifischen Produkte, die die 5'-Enden von HCII-mRNAs aus Lungengewebe enthalten, wurden teilweise sequenziert.

Als Beispiele sind die Sequenzierungen der zwei am weitesten stromaufwärts gefundenen Startpunkte detailliert aufgeführt (Abbildung 34, Abbildung 35 und zugehörige Sinnsequenz s.u.). Weitere, über Sequenzierung ermittelte Transkriptionsstartpunkte werden beispielhaft in Abbildung 36 gezeigt.

Die Sequenzierung wurde mit der 'Direct-Blotting'-Methode identisch zu der Analyse von Startpunkten der 2,3 kB-Transkripte durchgeführt. Es gelten also die unter Punkt D.8.2.2 beschriebenen Anmerkungen. In den folgenden Abbildungen sind einige Ergebnisse der Sequenzierungen von T-PCR-Produkten zu sehen. Das Ende der lesbaren Antisinnsequenz (an welches die G-Nukleotide anschließen) ist aufgeführt, sowie die Position der ersten Base, die mit der chromosomalen Sequenz übereinstimmt (Positionsangabe nach Herzog *et al.*¹⁶⁷).

Abbildung 34: 5'-Ende der HCII-mRNA aus Lunge, bestimmt mit dem Primer USC26 und Vergleich mit der genomischen Sequenz

Neben der Sequenziermembran ist die ermittelte Nukleotidabfolge notiert, wobei die ermittelte Sequenz über rote Striche den Banden zugeordnet ist. Die Startpunkte der HCII-mRNAs aus Lunge (bzw. A549-Zellen) wurden über Sequenzierung mit den Primern USC26 (Abbildung 34), USC24 (Abbildung 35) und USC20 (Abbildung 36) bestimmt. Die Position, an der die poly(G)-Sequenz beginnt (und damit das 5'-mRNA- bzw. 3'-cDNA-Ende anzeigt) ist jeweils deutlich zu erkennen.

Nachfolgend sind die ermittelten Startpunkte der Transkripte in Rot in der unten aufgeführten Sinnsequenz eingetragen (identisch zu der von Herzog *et al.* beschriebenen Sequenz¹⁶⁷).

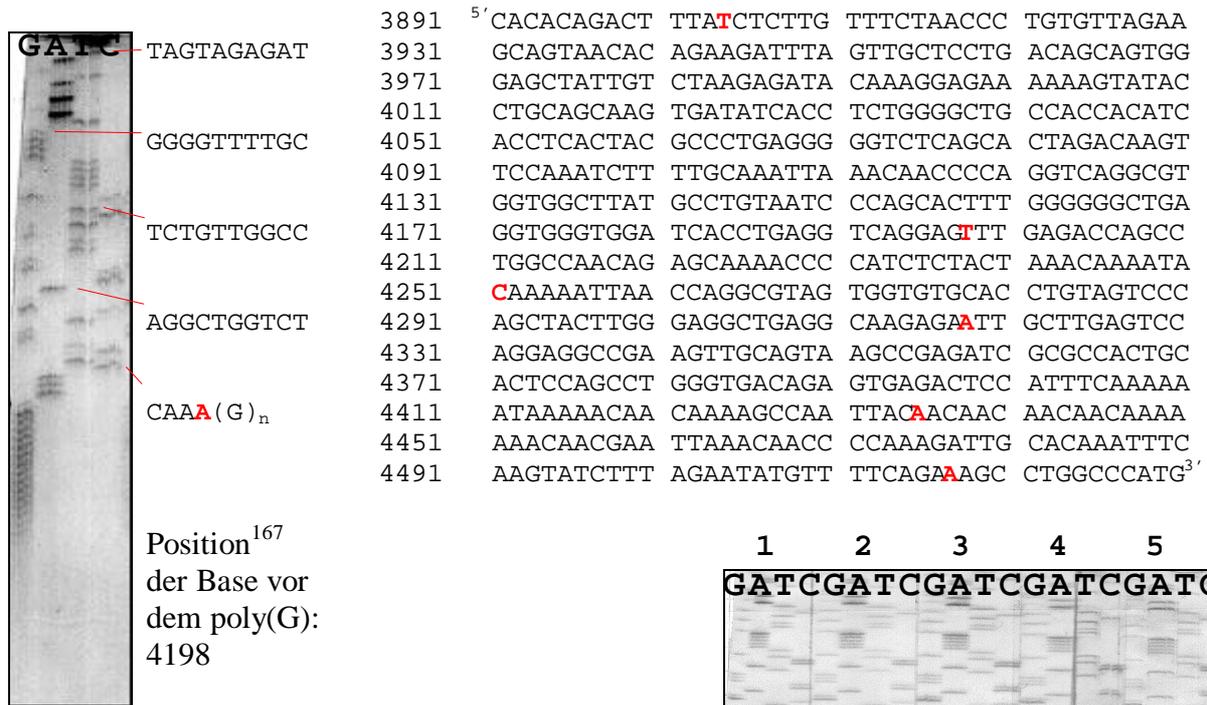


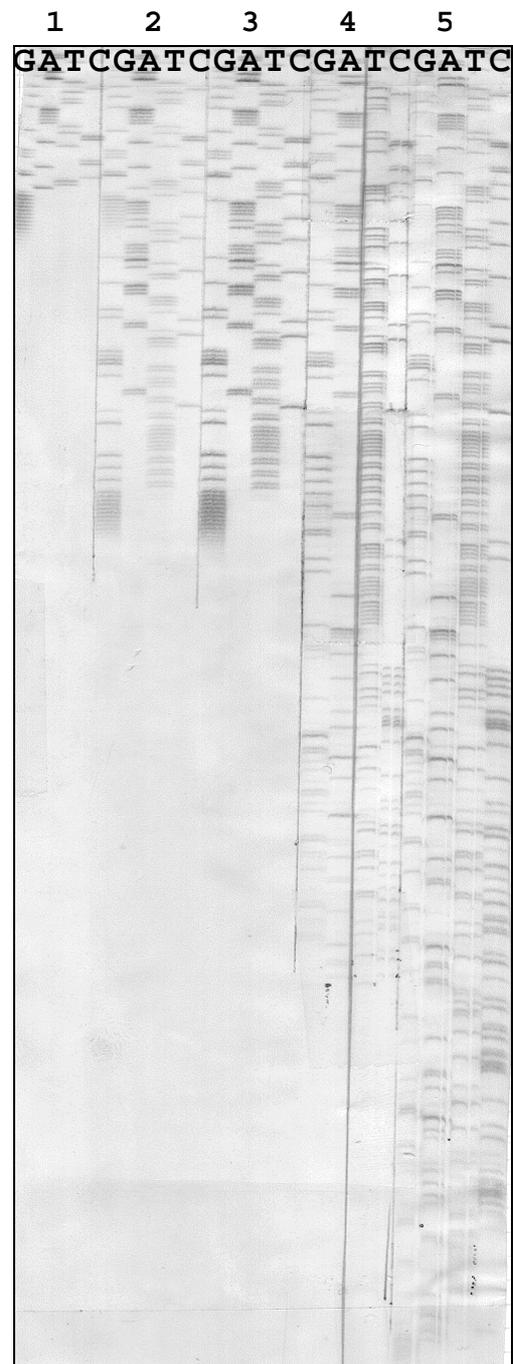
Abbildung 35: 5'-Ende der HCII-mRNA aus Lunge, bestimmt mit dem Primer USC24 und Vergleich mit der genomischen Sequenz

Es wurden keine Sequenzierungen von PCR-Produkten der Zelllinie durchgeführt, da sich, nach der Auftrennung der PCR-Produkte in den Agarosegelen, der Startbereich der HCII-Transkripte von A549-Zellen und Lungengewebe nur wenig unterscheidet. In der oben gezeigten Sequenz sind nicht sämtliche Startpunkte aufgeführt, sondern nur die über Sequenzierung ermittelten cDNA-Enden. Wie anhand der Gelbilder (Tabelle 32, Tabelle 33) zu erkennen ist, muß es weitere Startpunkte stromaufwärts geben. Die Startpunkte, die teilweise INR-Elementen entsprechen, werden in E.4.2 und E.4.3 diskutiert.

Abbildung 36: 5'-Enden der HCII-mRNA aus Lunge, bestimmt mit dem Primer USC20

Nach Gelextraktion von unterschiedlichen PCR-Fragmenten aus der selben T-PCR (durchgeführt mit Primer USC20) wurden fünf Fragmente mit dem gleichen Antisinnprimer sequenziert. Es wurden vier unterschiedliche Enden gefunden, die nachfolgend aufgelistet sind (Antisinnsequenz). Die Position der letzten erkennbaren Base (rot) ist nach Herzog *et al.*¹⁶⁷ angegeben.

- 1) ... GGGCCAGGCTTA(G)_n: Position der Base vor (G)_n: 4517
- 2) ... TTTGTTGTTGT(G)_n: Position der Base vor (G)_n: 4437
- 3) ... GACTCAAGCANT(G)_n: Position der Base vor (G)_n: 4318
- 4) ... GGTAAATTTTGG... (G)_n: Position der letzten Base: 4251



D.8.4 Gegenüberstellung der gefundenen Startpunkte

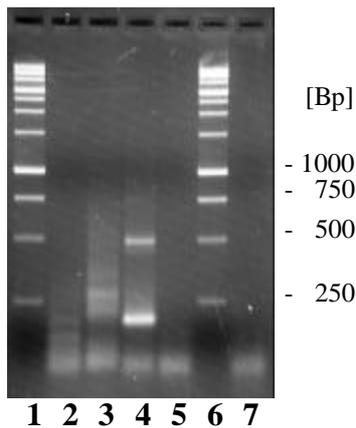


Abbildung 37: Vergleich der 5'-Startbereiche im Exon 1*

Das 2,5 %-Agarosegel zeigt mit USC24 erhaltene, aufgetrennte T-PCR-Produkte. Die Spuren 1+6 enthalten 0,25 µg 'kB-Leiter'. Die Spuren 2+3 enthalten jeweils 5 µl Produkte der Negativkontrolle (RT-Reaktion mit hitzeinaktiviertem Enzym) bzw. des Positivansatzes von Lungen-mRNA. Spuren 4+5 enthalten je 5 µl Produkte des Positivansatzes und der Negativkontrolle von Leber-mRNA. Spur 7 zeigt eine Negativkontrolle, die mit dH₂O als Matrize durchgeführt wurde.

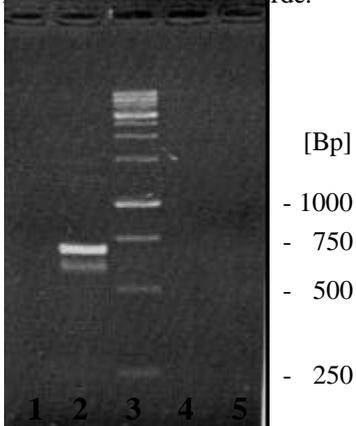
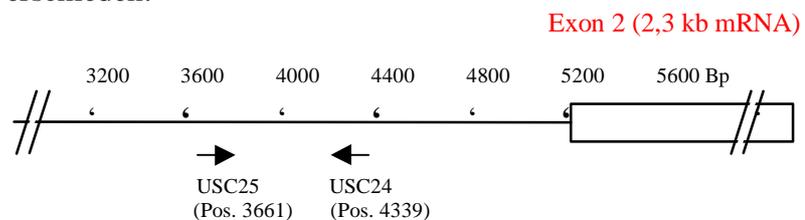


Abbildung 38: Überprüfung der 5'-Startbereiche

In der Abbildung ist ein 2 % Agarosegel abgebildet, in dem PCR-Fragmente nach einer RT-PCR-Reaktion mit Leber- und Lungen-poly(A)⁺-RNA aufgetrennt wurden. Die PCR wurde mit den Primern USC24 und USC25 durchgeführt. In den Spuren 1 und 2 sind die mit Leber poly(A)⁺-RNA erhaltenen Produkte aufgetrennt, während in den Spuren 4 und 5 das Ergebnis von Lungen poly(A)⁺-RNA zu erkennen ist. Die Spuren 1 und 5 enthalten die Negativkontrollen (RT-Reaktion mit hitzeinaktivierter reverser Transkriptase). Die mittlere Spur enthält 0,25 µg 'kB-Leiter'.

Den Ergebnissen für Startbereiche von längeren HCII-Transkripten in Leber- (D.8.2) und Lungengewebe (D.8.3) ist zu entnehmen, daß in beiden Organen mRNAs vorhanden sein müssen, die ihre jeweiligen Startpunkte mehr als 1 kB vor dem Exon 2 haben. Zum Erhalt eines qualitativen und quantitativen direkten Vergleichs der am weitesten stromaufwärts gelegenen Startbereiche von Exon 1* wurden die jeweiligen poly(A)⁺-RNAs mittels T-PCR überprüft. Die PCR-Produkte wurden mit dem Primer USC24 amplifiziert und parallel in einem Agarosegel aufgetrennt (Abbildung 37).

Aus dem Gelbild (vergleiche untenstehende Skizze zur Lage des Antisinnprimers) kann geschlossen werden, daß es in beiden Geweben Transkripte gibt, die in einem Bereich von 1000-1300 Bp vor Exon 2 beginnen. Allerdings scheint die Verteilung der einzelnen Startpunkte innerhalb dieses Bereichs je nach Gewebe unterschiedlich zu sein. Auch die Menge an Transkripten mit einem definierten Startpunkt ist deutlich verschieden.



Um eine eindeutige Aussage bezüglich der am weitesten 5'-gelegenen Startpunkte in Lunge und Leber machen zu können, wurde eine RT-PCR durchgeführt. Als Matrize wurde von poly(A)⁺-RNA aus Lunge und Leber unter identischen Bedingungen gegenüber hitzeinaktivierten Negativkontrollen ein 678 Bp großes Fragment aus dem Intron 1 mit den Primern USC24 und USC25 amplifiziert (Lage der Primer: siehe obige Skizze). Das Fragment beginnt 1538 Bp vor Exon 2. Das Ergebnis der aufgetrennten PCR-Produkte ist in der nebenstehenden Abbildung 38 zu sehen. Während die RNA aus Leber ein klares Signal liefert, wurde kein Fragment mit eingesetzter Lungen-mRNA erhalten.

Dies Ergebnis ist insofern überraschend, als mittels T-PCR und Primer USC24 (Abbildung 37) vermehrt größere Fragmente mit RNA aus Lunge erhalten werden, also weiter stromaufwärts gelegene Startpunkte. Die Diskussion der Ergebnisse findet im Teil E.4 statt.

D.9 Untersuchungen zur Promotorstruktur des Heparinkofaktor II-Gens

Das Ziel dieser Arbeit ist die Analyse der Struktur des HCII-Gens. Bisher sind HCII-mRNAs einiger Arten bekannt, die jeweils hohe Übereinstimmungen zeigen (D.6). Den bisher beschriebenen Ergebnissen ist zu entnehmen, daß dies nicht unbedingt für den 5'-Bereich des HCII-Gens bis zum Start von Exon 2 zutreffend ist. Beim Menschen sind zwei unterschiedliche HCII-Transkripte vorhanden (siehe D.8), die in verschiedenen Sequenzbereichen beginnen und somit anscheinend durch unterschiedliche Promotorregionen reguliert werden. Über die Regulation des HCII-Gens auf Transkriptionsebene ist nahezu nichts bekannt. Daher sollten mit Hilfe von Reportergenkonstrukten die regulatorischen Bereiche des Gens eingegrenzt werden. Die Promotoranalyse des menschlichen HCII-Promotors beschränkte sich jedoch größtenteils auf die theoretische Ebene. Als Vorbereitung für die Klonierung eines möglichen Promotorbereichs wurde zunächst die 5' vor dem HCII-Gen gelegene Sequenz analysiert. Zu diesem Zweck wurden die Sequenzen mit Hilfe der Programme 'MatInspector' und 'RepeatMasker' (Anmerkungen zu dem Programm: siehe D.4.2) auf eventuelle Bindungsstellen von Transkriptionsfaktoren und auf repetitive Elemente hin durchsucht. Dann erfolgte die Klonierung einer 2,2 kB langen 5'-gelegenen Sequenz. Die Überprüfung dieses potentiellen Promotorbereichs des 2,3 kb-Transkripts wurde mit Hilfe eines Luciferase-Reporterplasmids durchgeführt.

D.9.1 Bindungsstellen von Transkriptionsfaktoren

Promotorregionen setzen sich aus vielen einzelnen Elementen zusammen, von denen einige nahe des Transkriptionsstartpunktes liegen, während andere weit (mehrere kB) vor dem Exon 1 eines Gens liegen können.¹⁸⁰ Es gibt auch Beispiele dafür, daß wichtige regulative Sequenzen im 3'-Bereich des Gens liegen können.¹⁶⁸ Da diese Fälle aber eher Ausnahmen zu sein scheinen, wurde dieser Bereich nicht überprüft.

Bei den einzelnen Elementen, aus denen eine Promotorregion besteht, handelt es sich um kurze DNA-Sequenzen bzw. deren Kombination, an die spezielle Transkriptionsfaktoren binden können. Bestimmte Faktoren können nur dann binden, wenn sich in einem mehr oder weniger definierten Abstand ein anderer Faktor an der DNA anlagern kann. Mittlerweile sind viele Transkriptionsfaktoren bekannt, ebenso wie deren Zielsequenzen.¹⁶⁹ Die Zielsequenzen bzw. deren Matrizen sind jedoch größtenteils nur schlecht charakterisiert, so daß bei einer Suche nach Bindungsstellen in einer Sequenz normalerweise viel mehr potentielle Stellen ausgegeben werden, als tatsächlich in der Zelle von Transkriptionsfaktoren erkannt werden.

Für die Analyse von Sequenzen auf eventuelle Bindungsstellen von Transkriptionsfaktoren wurde das Programm 'MatInspector' mit Standardeinstellungen verwendet¹⁶⁹ (Kern-Matrix-Übereinstimmung: 75 %, Matrix-Übereinstimmung: 85 %). Teilweise wurde nach einer Kombination von Bindungsstellen mit dem 'ModelInspector'¹⁷⁰ gesucht, hierbei wurde die minimale Übereinstimmung der jeweiligen Matrix auf 80 % gesenkt. Da 'MatInspector' mit diesen Einstellungen eine sehr große Zahl von Treffern liefert, wurden nur einige Bindungsstellen ausgewählt. Für die detaillierte Betrachtung wurden Faktoren ausgewählt, die leberspezifisch sind, da HCII hauptsächlich in der Leber exprimiert wird. Weiterhin wurden allgemein wichtige Transkriptionsfaktoren wie C/EBP berücksichtigt, oder Faktoren, die aufgrund anderer Beobachtungen an der Regulation der HCII-Expression beteiligt sein könnten (z. B.: Antibabypille führt zu einem erhöhten HCII-Level im Blutplasma).

¹⁶⁸ z. B.: Morgan & Kalsheker, 1997

¹⁶⁹ Quandt *et al.*, 1995. Die Matrizen stammen aus der Transfac-Datenbank: Heinemeyer *et al.*, 1999

¹⁷⁰ Frech *et al.*, 1997

Um die Zahl der möglichen Übereinstimmungen mit den Matrizen zu erniedrigen, wäre es leicht, die minimale Übereinstimmung heraufzusetzen. Allerdings besteht dann die Gefahr, wichtige Stellen zu übersehen. Die hier aufgelisteten Matrizen umfassen Bindungsstellen für einige bekannte allgemeine und leberspezifische Transkriptionsfaktoren für Vertebraten. Bindungsstellen für mutierte, viruskodierte oder vor allem in Tumoren exprimierte Faktoren wurden nicht berücksichtigt. In die Betrachtung wurden gleichermaßen Enhancersequenzen und Promotorelemente aufgenommen. Falls unterschiedliche Matrizen des gleichen Faktors mit der Sequenz übereinstimmen, sind diese nur einmal gezählt. Tritt eine Bindungsstelle in einer Sequenz innerhalb von 20 Bp mehrfach auf, ist nur die erste Position zusammen mit der Zahl der Übereinstimmungen angegeben. Nachfolgend sind in Tabelle 34 die in dieser Arbeit angegebenen Bindungsstellen mit den zugehörigen Transkriptionsfaktoren aufgelistet. Die Einteilung der Faktoren in dieser Tabelle ist grob nach den jeweiligen Funktionen und nicht nach der Ähnlichkeit der Bindungsstellen durchgeführt.

Tabelle 34: Überprüfte Bindungsstellen für Transkriptionsfaktoren

Matrize	Faktor	Funktion	Vorkommen
AP1/AP4	Aktivator Protein (AP1: Leucine-zipper Protein)	Universelle Transkriptionsaktivatoren	
CEBP ¹⁷¹	C/EBP-Proteine (Leucine-zipper Protein; 'CCAAT/enhancer binding protein')	Zellspezifische Transkription	
CHOP ¹⁷² (GADD153)	CHOP, Stress-induzierter Faktor	Inhibition von C/EBP-Proteinen in der 'acute phase', oder bei Anwesenheit von Zellgiften, Transkriptionsaktivator	
ER	Östrogenrezeptor	Regulation ER-responsiver Gene	
GC		Universeller Transkriptionsaktivator	
GKLF	'Gut-enriched Krüppel-like' Faktor	Hinweise auf Interaktion mit SP1 Regulation der Transkription	
HFH ¹⁷³	Hepatozyten-Gabel-Proteine	Zellspezifische Transkription (HNF3: leberspezifisch), Regulierung der Entwicklung	Mesodermzellen
HNF ¹⁷⁴	Hepatozyten-Kernfaktoren (HNF1, HNF3, HNF4, HNF8) (HNF1: mit Homeodomäne, HNF4: mit Zink-Finger)	Regulation des Lebertranskriptions- Netzwerkes, Interaktion mit anderen leberspezifischen Faktoren	meist Leberzellen
IRF ¹⁷⁵	Interferonregulierender Faktor	Regulation des IFN-Systems	
NFY ¹⁷⁶ (CCAAT)	Kernfaktor Y	Universeller Transkriptionsaktivator, evtl. Interaktion mit TATA-Box- bindenden Proteinen	
OCT ¹⁷⁷	Oct-1 (Protein POU-Domäne)	Regulation der Entwicklung	
RORE ¹⁷⁸	Steroidhormon Nuklear Rezeptor ('RAR-related orphan receptor')	Regulation Steroid-responsiver Gene	

Fortsetzung der Tabelle auf der nächsten Seite

¹⁷¹ Akira *et al.*, 1990

¹⁷² Ubeda *et al.*, 1996

¹⁷³ Overdier *et al.*, 1994; Overdier *et al.*, 1997

¹⁷⁴ Tronche & Yaniv, 1992

¹⁷⁵ Tanaka *et al.*, 1993

¹⁷⁶ Mantovani, 1998

¹⁷⁷ Groenen *et al.*, 1992; van Leeuwen *et al.*, 1995; Verrijzer *et al.*, 1992;

¹⁷⁸ Giguere *et al.*, 1994

Fortsetzung der Tabelle

S8 ¹⁷⁹	S8 (Protein mit Homeodomäne)	Universeller Transkriptionsaktivator, Regulation der Entwicklung	Mesodermzellen
SP1	Stimulierendes Protein 1	Universeller Transkriptionsaktivator	
TATA	Teile des Initiationskomplexes	Universelles Promotorelement	

In der Tabelle sind bekannte, wichtige Bindungsstellen für allgemeine und leberspezifische Transkriptionsfaktoren aufgeführt. Soweit bekannt, ist die Anwesenheit des Transkriptionsfaktors in einem bestimmten Gewebe aufgeführt.¹⁸⁰ In dieser Tabelle sind nur die in dieser Arbeit zur Analyse ausgewählten Elemente aufgeführt und einige funktionell ähnliche Elemente sind in einer Rubrik zusammengefaßt.

D.9.1.1 Mögliche Bindungsstellen vor dem 2,3 kB-Transkript

Der eigentliche HCII-Promotorbereich, der die Expression des Proteins in Lebergewebe bewirkt, liegt wahrscheinlich vor dem Exon 1. Dementsprechend wurde vornehmlich der 5'-gelegene Bereich untersucht. Die später entdeckten, zahlenmäßig um Größenordnungen niedriger auch in der Leber vorhandenen, längeren Transkripte mit einem Startpunkt im Intron 1 haben offensichtlich (teilweise) einen eigenen Promotorbereich. Die Untersuchung für Bindungsstellen von Faktoren im Intron 1-Bereich findet nachfolgend im Punkt D.9.1.2 statt. Natürlich ist es möglich bzw. sehr wahrscheinlich, daß beide Transkriptionsstartpunkte von den gleichen weiter entfernt liegenden Enhancer- bzw. Silencersequenzen kontrolliert werden.

Die überprüfte Sequenz umfaßt 5994 Bp stromaufwärts des menschlichen HCII-Gens (Exon 1 eingeschlossen). Die Numerierung beginnt bei dem ersten Nukleotid dieser Sequenz. Bei einer Überprüfung auf gekoppelte Bindungsstellen wurde der Abstand 1-20 Bp gewählt. Die Tabelle 35 zeigt die Positionen in der Sequenz, an denen Übereinstimmungen mit spezifischen Matrizen gefunden wurden. Die Zeilen mit AP1/AP4, CEBP und HFH/HNF sind nur aus Gründen der Vollständigkeit aufgeführt. Eine Auswertung der jeweiligen einzelnen Bindungsstellen dieser Faktoren ist aufgrund der Vielzahl kaum möglich.

Tabelle 35: Position der Matrizen-Übereinstimmungen für den 5'-Bereich

Faktor	Position, Orientierung
AP1 / AP4	2x 2(±), 83(+), 2x 249(±), 270(-), 610(-), 3x 638(±), 2x 723(+), 762(+), 790(+), 1015(+), 1070(+), 2x 1144(±), 1230(-), 1268(+), 1301(-), 1537(-), 1593(-), 1619(-), 2x 1658(±), 1737(+), 1798(±), 2148(-), 2248(+), 2421(-), 2482(-), 2507(+), 2670(-), 2702(+), 2779(+), 2954(+), 2996(-), 3024(+), 3157(+), 3217(-), 3294(-), 3x 3319(±), 3461(-), 3556(-), 3588(+), 3631(-), 3655(+), 3691(-), 3773(-), 3839(-), 4011(-), 4045(-), 4085(+), 2x 4121(±), 2x 4181(-), 4265(+), 4312(+), 2x 4336(+), 4489(-), 4566(-), 3x 4666(±), 4776(-), 2x4820(+), 4856(+), 4981(+), 5177(-), 2x 5211(+), 5372(-), 2x 5412(±), 5529(-), 5634(+), 2x 5692(±), 4x 5748(±), 3x 5822(±), 3x 5840(±), 5943(+)
CEBP	102(+), 184(+), 440(-), 855(-), 2x 931(+), 1193(+), 1470(-), 1513(-), 1664(+), 1817(+), 2x 2300(±), 2414(-), 2520(-), 2594(+), 2x 2613(-), 3089(+), 3225(-), 3374(-), 3878(-), 3908(-), 3981(-), 4424(-), 4849(-), 5048(+), 5414(+), 5444(+), 5500(+), 2x 5886(+)
CHOP	398(-), 4063(-), 4199(-), 4972(+)
CEBP&CHOP	1470(-/-), 2301(+/-), 2594(+/-), 2629(-/-), 3225(-/-), 4424(-/-)
ER	1538(-), 1610(+), 2149(-), 2x 3632(±), 3774(-), 5835(+)
GC	618(+), 3045(+), 3200(+), 3817(-), 4191(+)

Fortsetzung der Tabelle auf der nächsten Seite

¹⁷⁹ de Jong *et al.*, 1993

¹⁸⁰ Darnell *et al.*, 1990

Fortsetzung der Tabelle

HFH / HNF	2x 42(+), 65(-), 4x 84(+), 2x 226(+), 2x 279(-), 327(-), 1044(+), 1360(-), 9x 1519(-), 1625(-), 1764(-), 1822(-), 1903(+), 4x 2047(+), 2207(-), 15x 2348(-), 2620(-), 2714(+), 2896(-), 2935(-), 3116(-), 4x 3274(-), 2x 3390(-), 6x 3408(+), 3592(+), 3743(-), 3780(-), 3840(-), 3916(-), 4632(-), 4770(-), 4937(-), 5075(-), 5x 5235(-), 5349(-), 5678(+), 2x 5863(+), 5980(-)
IRF	401(-)
NFY	25(+), 1881(+), 3361(+), 5119(+), 5867(-)
NFY&TATA	1879(+/-)
Oct	57(+), (-), 226(-), 303(+), 2x 325(±), 1046(+), 1609(-), 1688(+), 1812(-), 1857(-), 2x 1867(±), 2x 2709(±), 3328(-), 3352(-), 3x 3845(±), 4x 3929(±), 2x 3969(-), 4199(-), 4302(+), 2x 4586(±), 3x 4641(±), 5893(+), 5926(-)
RORE	1071(-), 1227(+), 1298(+), 1534(+), 2145(+), 3628(+), 3656(-), 3770(+), 5693(-), 5845(-), 5944(-)
S8	2x 37(±), 186(+), 1362(+), 1626(+), 1870(+), 2208(+), 3590(-), 3916(+), 2x 3943(±), 3x 3945(±), 3976(-), 2x 4626(+), 5077(+), 5338(-), 5457(-)
SP1	3045(+), 3200(+), 3818(-)
SP1&GKLF	470(+/+), 618(+/+), 3200(+/+), 3818(-/-), 4231(+/+), 4291(+/+), 5002(+/+), 5519(-/+)
TATA	298(+), 2x 318(±), 459(-), 2x 1625(+), 1892(-), 1913(-), 2530(-), 3737(+), 2x 4613(+)

Die Tabelle gibt Positionen potentieller Bindungsstellen von Transkriptionsfaktoren an, die mit MatInspector und Modell-Inspector im 5'-Bereich des HCII-Gens gefunden wurden (siehe Text). In einer Zeile sind teilweise Stellen von mehreren Faktoren einer Klasse (Proteinstruktur und Wirkungsweise) zusammen aufgeführt (dargestellt durch ein '/'-Symbol). Außerdem wurde nach kombinierten Bindungsstellen gesucht ('&'-Symbol). In Klammern hinter den Positionsangaben ist die Orientierung der Bindungsstelle angegeben: '+' entspricht Sinn-Orientierung und '-' entspricht Antisinn-Orientierung.

In der Tabelle 36 sind die gefundenen Übereinstimmungen mit den aus der Transfac-Datenbank¹⁶⁹ ausgewählten spezifischen Matrizen innerhalb des überprüften Bereiches zusammengefaßt. Zusätzlich ist der Re-Wert angegeben, der ein Maß darstellt, wie gut die jeweilige Matrize charakterisiert ist.

Tabelle 36: Anzahl der Matrizen-Übereinstimmungen für den 5'-Bereich

Zahl	Matrize	Re-Wert	Zahl	Matrize	Re-Wert	Zahl	Matrize	Re-Wert
40	V\$AP1FJ_Q2	2.45	5	V\$GC_01	2.12	1	V\$OCT1_01	0.04
11	V\$AP1_C	1.08	20	V\$GKLF_01	4.76	8	V\$OCT1_02	2.34
35	V\$AP1_Q2	1.82	6	V\$HFH1_01	0.12	23	V\$OCT1_06	0.97
19	V\$AP1_Q4	2.48	55	V\$HFH2_01	2.03	1	V\$OCT1_07	0.04
1	V\$AP2_Q6	4.78	70	V\$HFH3_01	0.31	1	V\$OCT1_Q6	4.82
1	V\$AP4_01	0.01	23	V\$HFH8_01	0.23	10	V\$RORA1_01	0.19
47	V\$AP4_Q5	0.96	2	V\$HNF1_01	0.51	5	V\$RORA2_01	0.01
18	V\$AP4_Q6	0.50	70	V\$HNF3B_01	1.45	20	V\$S8_01	0.67
32	V\$CEBPB_01	2.07	1	V\$IRF1_01	0.03	3	V\$SP1_Q6	1.70
2	V\$CEBP_C	0.27	1	V\$IRF2_01	<0.01	6	V\$TATA_01	5.33
4	V\$CHOP_01	1.10	3	V\$NFY_01	0.49	10	V\$TATA_C	0.30
7	V\$ER_Q6	1.73	5	V\$NFY_Q6	0.70			

Die Tabelle zeigt die Anzahl potentieller Bindungsstellen von Transkriptionsfaktoren, die mit MatInspector und Modell-Inspector im 5'-Bereich des HCII-Gens gefunden wurden. Der Re-Wert gibt die Anzahl zufälliger Übereinstimmungen der Matrix in einer beliebigen 1000 Bp-Sequenz an (nach Angabe bei MatInspector). Somit ist eine Matrix mit einem niedrigen Wert sehr gut definiert und besitzt eine größere Wahrscheinlichkeit einer Bindungsstelle zu entsprechen, die in der Zelle erkannt wird. Somit können für Promotoranalysen besser Matrizen wie V\$IRF2_01 als V\$GKLF_01 verwendet werden.

Die verschiedenen Matrizen sind in einer MatInspector untergeordneten Datenbank enthalten. Abkürzungen für Matrizen spezieller Bindungsstellen: AP: Aktivatorprotein; CEBP: C/EBP-Proteine; CHOP: Stress-induzierter Faktor; ER: Östrogen-Rezeptor; GC: GC-Box-Bindungsstelle; GKLF: 'Gut-enriched Krüppel-like' Faktor; HFH: Hepatozyten-Gabel-Proteine; HNF: Hepatozyten-Nuklearfaktoren; IRF: Interferon-regulierender Faktor; NFY: Nuklearfaktor Y; OCT: Oktamerbindungsstelle; ROR: Steroidhormon Nuklear Rezeptor; S8: S8-Bindungsstelle; SP1: SP1-Bindungsstelle; TATA: TATA-Box (siehe Tabelle 34).

Den zwei vorhergegangenen Tabellen ist zu entnehmen, daß es eine große Anzahl von AP-, HFH- und HNF-Bindungsstellen gibt. Für ein Protein, das in der Leber stark exprimiert wird, ist dies zu erwarten. Weiterhin bleibt die Frage allerdings, welche Regionen in dem untersuchten Bereich für die Promotoraktivität essentiell notwendig sind. Für eine Abschätzung der wirklich relevanten Stellen wurden deshalb nur die jeweils höchstsignifikanten Matrizen (möglichst geringer Re-Wert) für diese Faktoren (V\$CEBP_C, V\$HFH1_01, V\$HNF1_01) in die anschließenden Betrachtungen mit einbezogen (E.4.3.1, Abbildung 50).

D.9.1.2 Mögliche Bindungsstellen vor den längeren Transkripten

Die entdeckten längeren HCII-mRNAs (siehe D.8.2, D.8.3) mit Startpunkten im Bereich des Intron 1 zeigen, daß es einen zweiten, schwächeren Promotorbereich geben muß. Wahrscheinlich teilen sich beide Bereiche distale Regulationselemente, jedoch muß es weitere Signale geben, die einen zusätzlichen Startbereich im Intron 1 ermöglichen.

Um weitere regulative Elemente auf der DNA identifizieren zu können, wurde das gesamte erste Intron (5151 Bp) des menschlichen HCII-Gens mit MatInspector analysiert (analog zur Analyse des 5'-Bereichs unter Punkt D.9.1.1). Die Numerierung beginnt bei dem ersten Nukleotid dieser Sequenz. Gekoppelte Bindungsstellen (Abstand 1-20 Bp) wurden ebenfalls überprüft.

Tabelle 37: Position der Matrizen-Übereinstimmungen für den Intron 1-Bereich

Faktor	Position, Orientierung
AP1 / AP4	21(-), 159(-), 185(±), 398(-), 422(+), 466(-), 498(+), 606(-), 755(+), 954(+), 2x 986(+), 1078(-), 1126(+), 1273(+), 1424(+), 2x 1552(±), 2x 1726(±), 1749(+), 2x 1822(±), 1863(+), 2x 1609(±), 2060(+), 2155(+), 2257(-), 2x 2295 (±), 2311(+), 2x 2358(±), 2x 2439(±), 2575(+), 2581(+), 2656(-), 2700(+), 2810(+), 2850(-), 2x 3334(±), 3439(+), 3x 3554(±), 3674(-), 2x 3698(-), 2x 3816(±), 2x 3910(+), 4027(+), 4070(-), 2x 4103(±), 4138(-), 4241(+), 4335(+), 4524(±), 4611(+), 4632(+), 4759(+), 5056(-), 5094(+)
CEBP	73(+), 98(-), 768(-), 3x 833(±), 877(-), 1007(+), 3x 1264(+), 1306(+), 1328(-), 1407(-), 1545(-), 2x 1704(+), 2x 1820(±), 1884(-), 1920(-), 1952(-), 2007(-), 2x 2202(±), 2384(-), 2580(+), 2633(-), 2730(-), 2764(-), 3263(-), 2x 3764(-), 3812(+), 3962(+), 4049(+), 2x 4255(±), 4294(+), 4352(-), 4381(-), 4428(-), 2x 4460(±), 4487(-), 4626(-), 4681(-), 4730(-), 4835(+), 4893(+), 4918(-)
CHOP	2331(+), 3121(+), 4426(-)
CEBP&CHOP	1920(-/+), 2328 (+/+), 4352(-/-), 4381(-/-), 4428(-/-), 4487(-/-)
ER	414(+), 2855(-), 3338(+), 3577(-), 4071(-), 5057(-)
GC	2082(+), 3309(-), 4672(-), 4849(-)
HFH / HNF	13(+), 44(+), 82(-), 217(-), 457(-), 613(-), 637(-), 692(-), 868(-), 1049(-), 2x 1095(+), 1163(+), 1210(+), 1243(+), 3x 1417(+), 1457(+), 1473(-), 1822(+), 2x 2018(-), 2123(-), 3x 2182(-), 2300(-), 2366(-), 2635(+), 2683(-), 2759(+), 7x 2884(+), 8x 3057(+), 4x 3233(+), 2x 3368(+), 3565(-), 2x 3730(-), 3777(+), 3786(-), 3859(-), 4058(-), 2x 4189(-), 2x 4357(-), 3x 4395(-), 2x 4801(-)
IRF	129(-), 634(+), 1463(-), 1511(-), 2909(-), 4724(-),
NFY	245(-), 2121(-), 2781(+), 2x 4375(+), 4567(-), 4760(-),
NFY&TATA	245(-/-), 4373(+/+)
Oct	151(-), 2x 214(±), 234(-), 2x 274(±), 1458(+), 1828(+), 2x 2602(-), 2665(-), 2706(-), 3375(-), 3597(-), 3745(-), 4050(+), 4855(+), 4916(-), 4966(-), 2x 4989(±),
RORE	395(+), (±), (423-), 3348(-), 4067(+), 4135(+)
S8	2x 21(±), 2x 234(±), 458(+), 1009(+), 1215(-), 1916(-), 2020(+), 2564(+), 2674(+), 2x 3052(±), 3231(-), 3377(-), 3632(-), 4052 (+), 4201(+), 4375(+), 4405(+)
SP1	2082(+), 2168(-), 3310(-)
SP1&GKLF	333 (+/+), 471 (+/+), 2082(+/+)
TATA	30(+), 151(+), 778(-), 868(+), 1086(+), 1184(-), 1228(+), 1289(-), 1356(+), 1416(-), 1456(-), 1518(±), 1754(±), 2x 2182(+), 3x 2598(±), 3367(-), 3566(+), 4361(+)

Erklärung zu der Tabelle: siehe Tabelle 35

In der Tabelle 38 sind die gefundenen Übereinstimmungen mit den aus der Transfac-Datenbank¹⁶⁹ ausgewählten spezifischen Matrizen (und der jeweilige Re-Wert) innerhalb des Intron 1-Bereiches zusammengefaßt.

Tabelle 38: Anzahl der Matrizen-Übereinstimmungen für den Intron 1-Bereich

Anzahl	Matrize	Re-Wert	Anzahl	Matrize	Re-Wert	Anzahl	Matrize	Re-Wert
34	V\$AP1FJ_Q2	2.45	24	V\$GKLF_01	4.76	6	V\$NFY_Q6	0.70
22	V\$AP1_C	1.08	7	V\$HFH1_01	0.12	2	V\$OCT1_02	2.34
22	V\$AP1_Q4	2.48	47	V\$HFH2_01	2.03	1	V\$OCT1_05	0.03
1	V\$AP2_Q6	4.78	61	V\$HFH3_01	0.31	17	V\$OCT1_06	0.97
24	V\$AP4_Q5	0.96	25	V\$HFH8_01	0.23	3	V\$OCT1_Q6	4.82
12	V\$AP4_Q6	0.50	3	V\$HNF1_01	0.51	5	V\$RORA1_01	0.19
49	V\$CEBPB_01	2.07	59	V\$HNF3B_01	1.45	1	V\$RORA2_01	0.01
4	V\$CEBP_C	0.27	5	V\$IRF1_01	0.03	21	V\$S8_01	0.67
3	V\$CHOP_01	1.10	3	V\$IRF2_01	<0.01	3	V\$SP1_Q6	1.70
6	V\$ER_Q6	1.73	4	V\$NFY_01	0.49	17	V\$TATA_C	0.30
4	V\$GC_01	2.12	1	V\$NFY_C	<0.01			

Erklärungen zu der Tabelle: siehe Tabelle 36

Es ist gut zu erkennen, daß auch in dem Intron1-Bereich eine große Anzahl von AP-, HFH- und HNF-Bindungsstellen (siehe vorhergegangene Tabellen, D.9.1.1) vorhanden ist. Um essentielle Promotorbereiche vorhersagen zu können, müssen Matrizen betrachtet werden, die so gut definiert sind, daß sie rein zufällig nur sehr selten in beliebigen Sequenzen vorkommen (vergleiche: D.9.1.1). Die Position von potentiellen Bindungsstellen (Matrizen mit niedrigem Re-Wert) in diesem Bereich ist unter Punkt E.4.3.2 in Abbildung 51 eingetragen. Die weitere Diskussion der Ergebnisse wird ebenfalls im Abschnitt E.4 durchgeführt.

D.9.2 Repetitive Elemente vor dem Exon 1 des Heparinkofaktor II-Gens

Werden Sequenzen mit Programmen in Hinsicht auf mögliche Bindungsstellen von Transkriptionsfaktoren durchsucht, erhält man normalerweise viel mehr Bindungsstellen als tatsächlich vorhanden sind (vergleiche: D.9.1). In den Zellen wird die Erkennung einer DNA-Sequenz von weiteren spezifischen Bedingungen beeinflusst, wie z. B. durch die Konzentration spezifischer Transkriptionsfaktoren oder durch benachbarte Bindungsstellen anderer regulatorischer Proteine. Um die mit 'MatInspector' gefundenen Startpunkte sinnvoll eingrenzen zu können, wurde eine Suche mit 'RepeatMasker2' nach repetitiven Elementen (Tabelle 39) innerhalb einer 6 kB großen Sequenz vor dem Exon 1 durchgeführt. Die Analyse fand mit der RepeatMasker-Version vom 21.4.1999 mit 'ProcessRepeats'-Version vom 15.12.1998 statt (für Anmerkungen zu RepeatMaker siehe: D.4.2). Die überprüfte Sequenz umfaßt 5994 Bp vor dem menschlichen HCII-Gen einschließlich des ersten Exons (entsprechend der untersuchten Sequenz unter Punkt D.9.1.1).

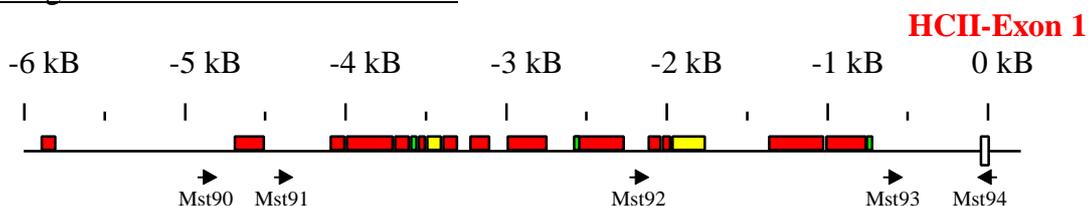
Repetitive Elemente können wichtige Promotorbereiche für ein Gen liefern (siehe B.2.2.3). In Bereichen der Sequenz, die von einem repetitiven Elementen abstammen, ist es allerdings weniger wahrscheinlich HCII-spezifische Promotorbereiche zu finden. Die Sequenzen dieser Elemente kommen sehr häufig im Genom vor und sind mehr oder weniger statistisch verteilt. Natürlich kann dennoch nicht ausgeschlossen werden, daß derartige Elemente (über die individuellen, teilweise geringen Sequenzunterschiede) ebenfalls genspezifische Promotorelemente enthalten.

Tabelle 39: DNA-Elemente vor dem Exon 1 des Heparinkofaktor II-Gens

HCII-Intron 1 des Menschen												
Nr.	SW Wert	% Div	% Del	% Ins	Pos.in Seq.		Erkanntes Element	Typ: Klasse/Fam.	Pos.im Elem.			
					Beginn	Ende			+/-	Beginn	Ende	
1	834	17.2	0.0	0.0	89	222	-	AlJo/FLAM	SINE/Alu	(179)	133	
2	21	7.1	0.0	0.0	299	340	+	AT_rich	Low_complexity	1	42	
3	2307	9.7	0.0	0.0	1239	1536	+	AluSg	SINE/Alu	1	297	
4	1084	16.1	5.2	5.2	1910	2083	-	AluJ	SINE/Alu	(21)	291	
5	2462	8.2	0.3	0.3	2084	2389	+	AluSq	SINE/Alu	1	306	
6	1084	16.1	5.2	5.2	2390	2435	-	AluJo	SINE/Alu	(211)	101	
7	441	0.0	0.0	0.0	2436	2484	+	(CA)n	Simple_repeat	2	50	
8	1084	16.1	5.2	5.2	2485	2506	-	AluJo	SINE/Alu	(260)	52	
9	498	18.7	1.6	1.6	2507	2629	+	L1MC4	LINE/L1	7857	7977	
10	1084	16.1	5.2	5.2	2630	2708	-	AluJo	SINE/Alu	(283)	29	
11	669	20.8	0.0	0.0	2752	2881	+	FLAM_C	SINE/Alu	3	128	
12	2351	7.0	0.3	0.3	3003	3286	+	AluY	SINE/Alu	9	293	
13	288	0.0	0.0	0.0	3418	3449	+	(TTTA)n	Simple_repeat	1	32	
14	2249	10.7	0.0	0.0	3450	3729	-	AluSx	SINE/Alu	(32)	280	
15	197	31.1	0.0	0.0	3876	3965	-	MIR	SINE/MIR	(21)	241	
16	357	25.0	3.8	3.8	3968	4047	-	MIR	SINE/MIR	(61)	201	
17	244	28.5	15.2	15.2	4064	4228	-	L2	LINE/L2	(240)	3073	
18	2033	14.3	0.0	0.0	4659	4944	+	AluJb	SINE/Alu	4	288	
19	2102	12.0	0.0	0.0	4952	5235	+	AluSx	SINE/Alu	1	283	
20	225	0.0	0.0	0.0	5236	5260	+	(CAA)n	Simple_repeat	2	26	

In der Tabelle sind die mit RepeatMasker2 gefundenen DNA-Elemente in einer 6 kB großen Sequenz vor dem Exon 1 des menschlichen HCII-Gens aufgeführt. Nr.: die gefundenen Elemente sind in 5'-3'-Richtung durchnummeriert; SW-Wert: Smith-Watermann score, ein Wert für die Übereinstimmung der Sequenz mit einem Element; Div: prozentualer Unterschied der Sequenz zum Element; Del/Ins: prozentuale Angabe der Deletionen und Insertionen in der Sequenz; Pos.in Seq./Pos.im Elem.: Angabe der Position der Sequenz (mit Beginn und Ende) in der Gesamtsequenz bzw. Abschnitt innerhalb des Elements. +/-: Sinn- bzw. Antisinnorientierung des Elements. Aufgeführte Typen von DNA-Elementen: Low_complexity: DNA-Sequenz mit niedriger Komplexität, d. h. einem hohen Anteil von A- und T-Nukleotiden (AT_rich); Simple_repeat: Di-, Tri- oder Tetranukleotid-wiederholungen; Alu: Nomenklatur der speziellen Alu-Typen nach Batzer¹⁸¹, MIR: 'Mammalian-wide inter-spersed repeats'¹⁸²

Nachfolgend veranschaulicht eine Skizze (Abbildung 39) die Position repetitiver Elemente in der DNA-Sequenz vor dem Exon 1 des HCII-Gens. In der Abbildung 50 im Diskussionsteil sind die Sequenzbereiche einheitlich markiert, die von repetitiven Elementen abstammen.

Abbildung 39: 5'-Bereich des HCII-Gens

In der Skizze sind die mit RepeatMasker identifizierten repetitiven Elemente in einer 6 kB großen Sequenz, die mit dem 5'-Bereich des HCII-Exon 1 endet, schematisch dargestellt. SINE-Elemente sind rot, LINEs sind gelb und einfache Wiederholungen sind grün gezeichnet. Die Position der Elemente wurde der Tabelle 39 entnommen.

Für die Klonierung eines potentiellen Promotorbereichs wurden Primer (Mst90-Mst93) für eine PCR ausgewählt, die in Sequenzbereichen hybridisieren können, die nicht von repetitiven Elementen stammen. Falls die Primer dementsprechend ausgewählt werden, können weniger unspezifische Hybridisierungen auftreten. Außerdem ist es wahrscheinlich, daß die Existenz genspezifischer Promotorelemente in diesen Bereichen erhöht ist.

¹⁸¹ Batzer *et al.*, 1996¹⁸² Smit & Riggs, 1995

Um möglichst HCII-spezifische Sequenzen aus dem 5'-Bereich klonieren zu können, ist es optimal, die Bindungsstellen der Primer für die PCR-Reaktion in nicht repetitiven Bereichen zu wählen. Aus zeitlichen Gründen wurde nur die Promotoraktivität eines einzigen Bereiches experimentell bestimmt. Normalerweise ist damit zu rechnen, daß der Basis-Promotorbereich innerhalb von 1-2 kB vor dem Exon 1 eines Gens vorhanden ist. Daher bot sich an, eine 2,2 kB große Sequenz (PCR-Produkt der Primer Mst92-Mst94, siehe Abbildung 39) incl. eines Teils von Exon 1 zur weiteren Analyse (D.9.3) zu klonieren.

In diesem Abschnitt wurde nur die Analyse des 5'-Bereichs vor dem Exon 1 des HCII beschrieben, da die äquivalente Untersuchung für den Intron 1-Bereich schon unter einer anderen Fragestellung im Punkt D.4.2 beschrieben ist. Die Abbildung 45 unter Punkt E.3.2 zeigt u.a. die Lage der repetitiven Elemente im menschlichen Intron 1-Bereich.

D.9.3 Promotoranalyse der 2,3 kB-Transkripte

Die erste Überprüfung einer Region vor dem HCII-Exon 1 fand mittels des Promega Dual-Luciferase-Reporter-Systems auf regulatorische Elemente hin statt. Zu diesem Zweck wurde ein 2,2 kB großes Fragment, welches einen Teil von Exon 1 mit einschließt, vor das *Photinus Luciferase*-Gen kloniert (PCR-Produkt der Primer Mst92-Mst94, siehe Abbildung 39). Die Luciferase-Aktivität, die dieses Plasmid (pPBK60; gezeigt in G.1.3, Abbildung 55) in einer Leberzelllinie (HepG2) bewirkt, wurde im Vergleich zu Negativkontrollen (nichttransfizierte Zellen, Plasmid ohne Insert: pGL3Basis) bestimmt. Als Positivkontrolle dienten Ansätze mit dem Plasmid pGL3Control. Dieses Plasmid entspricht dem pGL3Basis-Plasmid mit einem zusätzlichen SV40-Enhancer und dem SV40-Promotor vor dem *Luciferase*-Gen.¹⁸³ Um die Ergebnisse bezüglich der Transfektionseffizienz zu normieren, wurde jeweils eine Kotransfektion mit einem Referenzplasmid¹⁸⁴ durchgeführt (C.2.13.2). Da die jeweiligen Plasmide unterschiedliche Größen aufweisen, wurde nicht die gleiche quantitative Menge, sondern die gleiche molare Menge der Plasmide in den Ansätzen verwendet. Die Durchführung erfolgte mit jeweils zwei bis drei parallelen Transfektionsansätzen, um die Standardabweichung ermitteln zu können. Die Messungen der Ansätze ergaben nach zwei unabhängigen Vorversuchen die folgenden Werte (Tabelle 40, Abbildung 40).

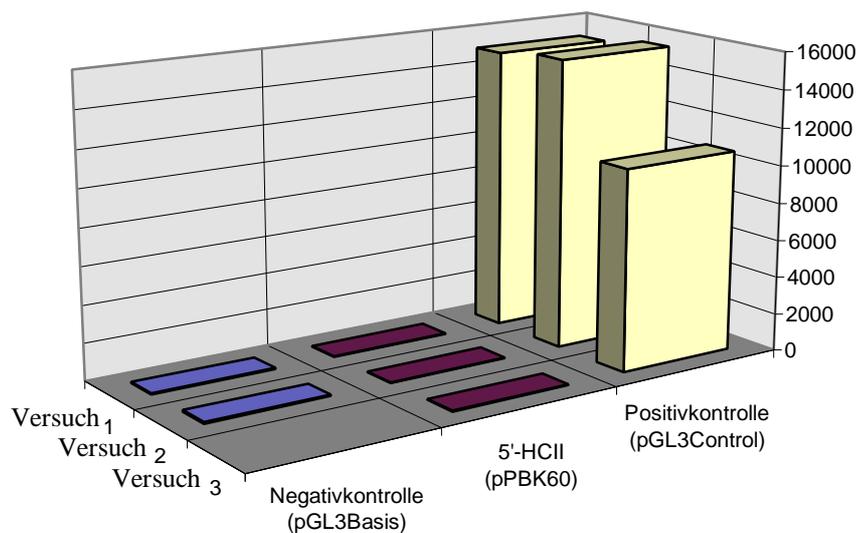
Tabelle 40: Überprüfung einer 2,2 kB-Sequenz auf Funktion als Promotorbereich

Ansatz	Plasmid	RLUs absolut <i>Photinus luc</i>	RLUs absolut <i>Renilla luc</i> (kotransfiziert)	RLUs normiert <i>Photinus luc</i> (* 1000)	Mittelwert ± SD <i>Photinus luc</i>
Negativ	pGL3Basis	689	11587	59	71 ± 11 (71 ± 16 %)
		893	10881	82	
Positiv	pGL3Control	193011	12808	15070	13752 ± 2163 (13752 ± 16 %)
		198564	12824	15484	
		134642	12581	10702	
HCII ⁺	pPBK60	938	22610	41	48 ± 5 (48 ± 10 %)
		1604	30405	53	
		768	15819	49	

Die RLU-Werte (relative light units) für die *Photinus Luciferase* (*Photinus luc*, Testreportergen) wurden normiert ('RLU normiert'), indem sie durch die Werte für die *Renilla Luciferase*aktivität geteilt wurden. Um die Werte anschaulicher zu machen, wurden die erhaltenen Zahlen außerdem mit 1000 multipliziert und gerundet.

¹⁸³ Promega, 1998

¹⁸⁴ pRL-TK: Enthält das *Renilla Luciferase*gen mit einer andere Substratspezifität (G.1.2, Abbildung 54)

Abbildung 40: Überprüfung einer 2,2 kB-HCII-Sequenz auf Funktion als Promotorbereich

In der Abbildung sind die normierten RLU-Daten aus Tabelle 40 graphisch dargestellt. Eine Spalte zeigt jeweils parallele Messungen.

Neben den unten aufgeführten Messungen wurde auch die Luciferaseaktivität von weiteren Negativkontrollen bestimmt (nichttransfizierte Zellen und Plasmid mit einem entsprechenden Insert in umgekehrter Orientierung). Sämtliche Negativkontrollen zeigen allerdings ähnliche Ergebnisse, so daß sie hier nicht aufgeführt sind. Der Mittelwert und die Standardabweichung sind ebenfalls in Tabelle 40 angegeben. Der Tabelle ist zu entnehmen, daß Transfektionen und Messungen von gleichen Ansätzen Werte mit einer Standardabweichung von bis zu 16 % ergeben. Diese Abweichungen entsprechen den normalen Schwankungen bei dieser Methode. Die Werte für die Positivkontrolle (pGL3Control) zeigen, daß ein Plasmid mit dem *Photinus* Luciferasegen und einem intakten Promotor und Enhancer eine starke Luciferaseaktivität in HepG2-Zellen bewirken kann. Demgegenüber zeigt das klonierte HCII-Genfragment (2,2 kB incl. eines Teils von Exon 1) für sich allein keine Promotorfunktion, da die gemessenen Werte mit der Negativkontrolle vergleichbar sind (siehe Abbildung 40).

Wahrscheinlich enthält die klonierte 2,2 kB große Sequenz nicht den vollständigen Promotorbereich des HCII-Gens. Es scheinen wichtige essentielle Regionen (und eventuell ebenfalls weitere Enhancer-Sequenzen) zu fehlen, die für die Transkription des Gens notwendig sind.

Eine weitere Möglichkeit der fehlenden Transkriptionsaktivierung könnte durch einen Silencer bewirkt sein, der ebenfalls innerhalb der klonierten Sequenz liegt. Allerdings sollte ein Silencer innerhalb einer Promotorregion eines in der Leber transkribierten Gens in HepG2-Zellen inaktiv sein. Somit scheidet diese Möglichkeit aus, sofern nicht angenommen wird, daß die HepG2-Zelllinie sich deutlich von primären Leberzellen unterscheidet. Die mögliche Position essentieller Promotorsequenzen wird im Teil E.4.3.1 diskutiert.

E. Diskussion

Die Diskussion der Ergebnisse dieser Arbeit ist in folgende Teile gegliedert:

In dem Abschnitt E.1 werden zuerst die gefundenen und bekannten Exon/Introngrenzen der HCII-Gene von Ratte, Maus und Mensch kurz dargestellt. Im Abschnitt E.2 werden die einzelnen 5'-Bereiche der HCII-Gene verglichen und Aussagen zu den möglichen Mechanismen der Veränderungen gemacht. Eine Diskussion zu den erhaltenen Ergebnissen des cDNA-Alignments wird unter Punkt E.3 durchgeführt und die Ergebnisse der Bestimmung der Transkriptionsstartpunkte vom menschlichen HCII-Gen werden im Abschnitt E.4 diskutiert. Das Ergebnis der Promotoranalyse wird im Punkt E.4.3 analysiert und eine abschließende graphische Zusammenfassung der HCII-Genstrukturen wird im Teil E.5 gegeben.

E.1 Nachweis von Exon/Introngrenzen

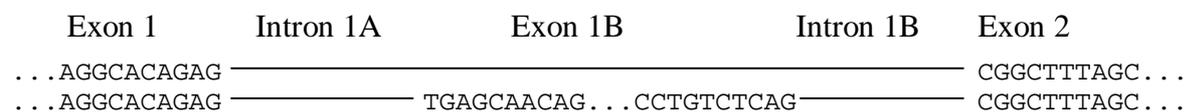
Die Exon/Intronstruktur der α_1 -Antitrypsin-Gruppe innerhalb der Serpin-Genfamilie ist verhältnismäßig stark konserviert. Die in dieser und früheren Arbeiten erhaltenen Daten zeigen jedoch, daß innerhalb dieser Gruppe die Struktur des HCII-Gens variabel ist. Die durchgeführten Untersuchungen bei der Ratte, bei der Maus und beim Menschen lassen einige Schlußfolgerungen hinsichtlich der Mechanismen zu, die diese Veränderungen verursacht haben könnten.

Während im folgenden unter Punkt E.1 die unterschiedlichen 5'-Genstrukturen nur grob dargestellt werden, werden die detaillierten Sequenzen der Grenzbereiche schließlich unter Punkt E.2.4 (Abbildung 46 und folgende) gezeigt und besprochen.

E.1.1 Exon/Introngrenzen im 5'-Bereich des HCII-Gens der Ratte

Die Exon/Introngrenzen im 5'-Bereich des HCII-Gens der Ratte wurden teilweise vor dieser Arbeit von Westrup und Ragg¹⁸⁵ analysiert. Es liegen zwei unterschiedliche HCII-Transkripte vor. Das längere Transkript enthält eine kurze Sequenz, die in den anderen Arten nicht nachgewiesen werden kann und (über einen Vergleich mit der Intron 1-Sequenz) auf ein alternatives Spleißen zurückzuführen ist. Ein Teil der mRNAs enthält also ein zusätzliches Exon (Ergebnis der entsprechenden RT-PCRs für die mHCII-mRNA: D.2, Abbildung 21 und für die hHCII-mRNA: D.4.3, Abbildung 24). In der Abbildung 41 sind die Exon/Introngrenzen der beiden möglichen Transkripte dargestellt.

Abbildung 41: 5'-gelegene Exon/Introngrenzen im Ratten-HCII-Gen



In der Abbildung sind nur jeweils 10 Bp der angrenzenden Exons dargestellt. Die vollständige Intronsequenz kann der Tabelle 18 entnommen werden. Die Sequenzen der angrenzenden Bereiche sind in den Abbildungen unter Punkt E.2.4 genauer gezeigt.

¹⁸⁵ Westrup & Ragg, 1994

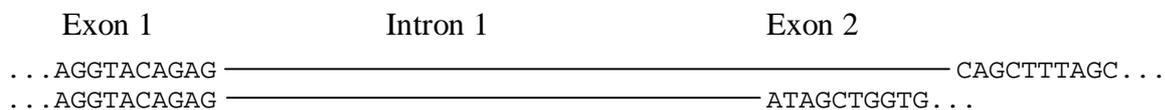
E.1.2 Exon/Introngrenzen im 5'-Bereich des HCII-Gens der Maus

Vor dieser Arbeit war bereits bekannt, daß ein dem Menschen (E.1.3), bzw. der Ratte (E.1.1) vergleichbares mRNA-HCII-Transkript in Maus-Leberzellen vorhanden ist.

Die mögliche Existenz eines zusätzlichen Exons oder Veränderungen der Genstruktur im Verlauf der Evolution lassen sich mittels RT-PCR-Reaktionen nachweisen. Zu diesem Zweck wurden in zwei verschiedenen PCRs mit Primerpaaren, deren Produkt vom ersten Exon bis zum zweiten Exon reicht, der dazwischen liegende Bereich der HCII-cDNAs amplifiziert. Es ist jeweils ein PCR-Produkt mit einer Länge von 117 Bp (Mst12-Mst36) bzw. ein Produkt der Länge 106 Bp (Mst4-Mst36) zu erwarten. Wie der Abbildung 21 zu entnehmen ist, entstanden diese Produkte und auch jeweils ein zweites Produkt, das etwa 140 Bp länger ist. Das kleine Fragment wurde über eine PCR und Klonierung gereinigt während die Reinigung des größeren Fragments durch eine RT-PCR mit einem Antisinn-Primer, der vor dem bekannten Exon 2-Beginn bindet, erfolgte.

Anschließend wurde das erhaltene Produkt in einer weiteren PCR als Megaprimer verwendet, um es für die Sequenzierung zu verlängern. In den nachfolgenden direkten Sequenzierungen von PCR-Produkten beider cDNAs ließen sich zwei verschiedene Exon/Introngrenzen nachweisen. Die Sequenzierung des PCR-Produkts der erwarteten Länge ergab eine Exon/Introngrenze, die an der homologen Stelle zum Exon 2-Beginn in der Ratte liegt. Das zweite PCR-Fragment zeigt nicht einen zum Exon 1B der Ratte entsprechenden Bereich, sondern eine Verlängerung des Exon 2 von 127 Bp (siehe D.2, Abbildung 23 und D.3.1, Tabelle 21) in das Intron 1 hinein. In der Maus kommt es also beim HCII-Gen zu einem alternativen Spleißen am Beginn von Exon 2, das bei den anderen untersuchten Arten nicht stattfindet. In der nachfolgenden Abbildung sind die Exon/Introngrenzen in diesem Bereich schematisch dargestellt.

Abbildung 42: 5'-gelegene Exon/Introngrenzen im Maus-HCII-Gen



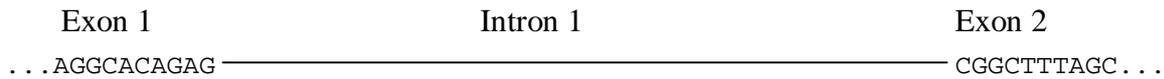
In der Abbildung sind jeweils 10 Bp der angrenzenden Exons gezeigt. Die vollständige Intronsequenz kann der Tabelle 21 entnommen werden. Die Sequenzen angrenzenden Bereiche sind in den Abbildungen unter Punkt E.2.4 genauer dargestellt.

E.1.3 Exon/Introngrenzen im 5'-Bereich des HCII-Gens des Menschen

Von Ragg¹⁸⁶ und Herzog *et al.*¹⁸⁷ wurden vor dieser Arbeit die Exon/Introngrenzen des HCII-Gens des Menschen analysiert. Ein alternatives Spleißen wie bei der Maus oder bei der Ratte konnte bisher nicht gefunden werden (D.4.3). Allerdings konnten Transkripte in sehr geringer Konzentration gefunden werden (Abbildung 27 und Skizze), die nicht das Exon 1 enthalten, stattdessen jedoch ein verlängertes Exon 2 aufweisen (Exon 1*). In der Abbildung 43 sind die Exon/Introngrenzen des Haupttranskripts abgebildet. Da die längeren Transkripte eine Vielzahl von Startpunkten zu haben scheinen, sind sie in der Abbildung nicht dargestellt.

¹⁸⁶ Kamp & Ragg, 1999

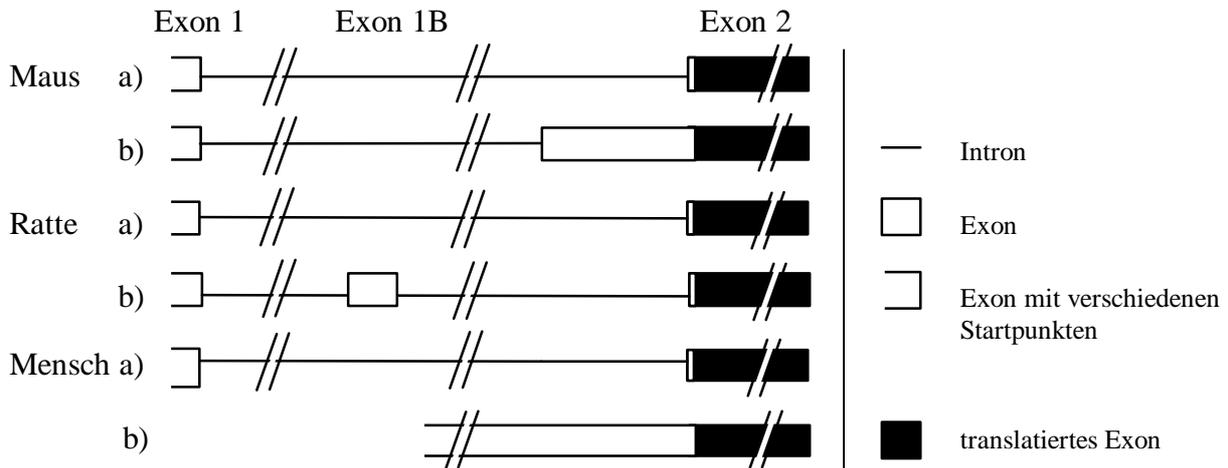
¹⁸⁷ Herzog *et al.*, 1991

Abbildung 43: 5'-gelegene Exon/Introngrenzen im menschlichen HCII-Gen

In der Abbildung ist der 5'-Bereich des vorherrschenden menschlichen HCII-Transkripts (bezüglich der Konzentration an Transkripten in Lebergewebe) schematisch dargestellt. Es sind nur jeweils 10 Bp der angrenzenden Exons eingezeichnet. Die Sequenzen dieser Bereiche sind in den Abbildungen unter Punkt E.2.4 genauer gezeigt.

E.1.4 Vergleich der Exon/Introngrenzen im 5'-Bereich des HCII-Gens

Die unterschiedlichen Exon/Intronstrukturen des HCII-5'-Bereichs von Maus, Ratte und Mensch sind in dem nachfolgendem Schema (Abbildung 44) zusammengefaßt. Eine vergleichende Darstellung der gesamten Genstrukturen ist unter Punkt E.5.1 in Abbildung 52 gegeben.

Abbildung 44: 5'-Genstruktur von HCII

E.2 Alignment von Intron 1-Sequenzen von Heparinkofaktor II

Da die Exon/Intronstruktur bei HCII im Bereich von Intron 1 gerade bei den nahverwandten Arten Maus und Ratte weniger stark konserviert ist, liegt die Frage nahe, durch welche Mechanismen sich im Verlauf einer sehr kurzen evolutiven Zeitspanne (41 Millionen Jahre)¹⁸⁸ die Struktur geändert hat. Daher wird die HCII-Intron 1-Sequenz in dieser Arbeit analysiert.

Für die Analyse der Intron 1-Bereiche und die Veränderungen, die sich im Verlauf der Evolution abgespielt haben, werden nachfolgend die Sequenzen auf verschiedene Weise verglichen und ausgewertet. Zuerst werden die allgemeinen Daten der Introns betrachtet, daran anschließend die Intronstrukturen, und dann werden direkt wichtige Mutationen in den Sequenzen dargestellt. Abschließend wird ein Vorschlag für die unterschiedliche HCII-Exon/Intronstruktur in den einzelnen Arten aufgrund der eingetretenen Veränderungen gegeben.

¹⁸⁸ Kumar & Hedges, 1998

E.2.1 Statistische Daten des Intron 1

Das HCII-Intron 1 der Maus ist 4842 Bp lang. Die entsprechenden Intronsequenzen von Ratte und Mensch haben eine Länge von 4348 Bp bzw. 5151 Bp. Das bedeutet Längenunterschiede der Sequenz der Maus von +10 % gegenüber der Ratte und -6 % gegenüber dem Menschen.

Die Sequenzen zeigen keine auffälligen Abweichungen bezüglich des GC-Gehalts (Maus: 41 %, Ratte 44 % und Mensch 45 %) von eukaryotischen Genomen. In nicht-kodierenden Regionen liegt der GC-Gehalt unter 50 %, da ein hoher Anteil des Genoms repetitive Sequenzen enthält, die reich an AT-Basenpaaren sind.¹⁸⁹

Die Anzahl und der prozentuale Anteil der einzelnen Nukleotide ist in der folgenden Tabelle 41 zusammengefaßt. Wie leicht zu erkennen ist, treten keine größeren Schwankungen in den Nukleotid-Relationen auf. Bemerkenswert ist, daß in der Sinn-Sequenz der DNA bzw. in der prä-mRNA A-Nukleotide häufiger vorkommen als T-Nukleotide. Außerdem kommen C-Nukleotide häufiger (mit Ausnahme der menschlichen Sequenz) als G-Nukleotide vor.

Tabelle 41: Nukleotide im Intron 1 vom Heparinkofaktor II-Gen

	Anzahl der Nukleotide				prozentualer Anteil			
	A	C	G	T	A	C	G	T
Maus	1557	1098	888	1299	32 %	23 %	18 %	27 %
Ratte	1287	1018	903	1140	30 %	23 %	21 %	26 %
Mensch	1454	1147	1148	1402	28 %	22 %	22 %	27 %

Die Intron 1-Sequenzen, die diesen Daten zu Grunde liegen, wurden dieser Arbeit (Maus, Ratte) bzw. der Veröffentlichung von Herzog *et al.*¹⁸⁷ entnommen.

Der Anteil von repetitiver DNA in der Intron 1-Sequenz variiert stärker (siehe Tabelle 28). Bei der Maus bestehen etwa 36 % der Gesamtsequenz aus bekannten Elementen (davon 2/3 komplexe Wiederholungen), der Anteil bei der Ratte liegt hingegen nur bei 28 % (davon etwa 3/4 komplexe Wiederholungen). Beim Menschen ist mit 42 % der weitaus größte Anteil vorhanden (fast ausschließlich komplexe repetitive Elemente). Es gibt Beispiele für eine Zunahme der repetitiven Elemente in einer Sequenz während der Fortentwicklung einer Art.¹⁹⁰ Allerdings ist die Anzahl der erkannten *Alu*-Elemente (beziehungsweise *Alu*-ähnlichen Elemente: B-Elemente bei Maus und Ratte), die jeweils den größten Teil der repetitiven DNA ausmachen (mHCII: 16 %, rHCII: 17 %, hHCII: 32 %), annähernd konstant. Es existieren jeweils sechs *Alu*-Elemente bei Maus und Mensch und fünf solcher Bereiche bei der Ratte.

Während beim Menschen nahezu die gesamten repetitiven DNA-Elemente aus *Alu*-Repeats bestehen, beträgt deren Anteil bei Maus und Ratte nur etwa die Hälfte. Bei diesen Arten kommt ein größerer Teil von 'einfachen Wiederholungen' vor, die beim Menschen im HCII-Intron 1 nicht vorhanden sind.

Abschließend kann gesagt werden, daß sich aus den allgemeinen Daten eine stärkere verwandtschaftliche Beziehung zwischen Maus und Ratte erkennen läßt, aber keine Hinweise auf die veränderte Exon/Intronstruktur erhalten werden.

¹⁸⁹ Knippers *et al.*, 1990

¹⁹⁰ Brookfield, 1995; Kass *et al.*, 1997

E.2.2 Vergleich der theoretischen mit den natürlichen Introngrenzen

Aus den Daten von D.1.2 und D.3.3 läßt sich entnehmen, daß fünf bis sieben mögliche 5'-Intron-Spleißgrenzen pro 1000 Bp erkannt werden (Maus: 7,2, insgesamt: 35 Stellen; Ratte: 5,3, insgesamt: 23 Stellen; Mensch: 6,4, insgesamt: 33 Stellen). Dagegen werden etwa drei mögliche Stellen je 1000 Bp für die 3'-Introngrenze (Maus und Mensch: je 3,1, insgesamt: 15 bzw. 16 Stellen; Ratte: 3,0, insgesamt: 13 Stellen) gefunden. Daraus folgt, daß nach den bekannten Daten¹²³ (siehe C.2.1), auf denen die Programme beruhen, 10 bis 30mal mehr potentielle 5'-Stellen von dem Programm bei Maus und Mensch (nur etwa zehnmal soviel bei Ratte) vorgeschlagen werden, als (experimentell belegt) in der Zelle erkannt werden.

Für die 3'-Seite der Introns verläuft die Erkennung realer Grenzen spezifischer. Bei Maus und Ratte werden nur etwa siebenmal so viele Grenzen erkannt, wie tatsächlich vorhanden sind. Da beim Menschen kein alternatives Spleißen bekannt ist, ist hier das Verhältnis der zuviel erkannten Grenzen etwa doppelt so hoch. Es läßt sich also erkennen, daß nach heutigem Wissen viele Möglichkeiten für zusätzliche Exons bzw. Exonbereiche in dieser Sequenz gegeben sind. Die tatsächlich genutzten Grenzen entsprechen nicht unbedingt den jeweils besten aus der Konsensussequenz abgeleiteten theoretischen Grenzen. Für die bei den drei Arten an gleicher Position liegenden 5'-Introngrenzen liegt die Übereinstimmung mit der Matrix zwar jeweils sehr hoch bei 92 %, das alternativ gespleißte Exon 1B der Ratte hat aber an seinem Ende eine Spleißstelle, die nur zu 49 % mit der Matrix übereinstimmt. Vielleicht liegt hier der Grund dafür, daß nur ein Teil der Transkripte dieses Exon enthält.

Die in etwa an der gleichen Position bei allen Arten gelegene 3'-Spleißstelle weist eine Übereinstimmung mit der Matrix von 73 % bis 88 % auf. Dabei ist in auffälliger Weise die beste Übereinstimmung mit 88 % zur Matrix beim Menschen gegeben (siehe D.3: Tabelle 25, E.2.4.3: Abbildung 48). Hier kommt es zu keinem alternativen Spleißen. In Maus und Ratte, bei denen die Übereinstimmungen schlechter sind, kommen (bei einigen Transkripten) veränderte Grenzen vor. So wird bei der Maus teilweise eine 3'-Spleißstelle verwendet, die stromaufwärts kurz vor der allgemeinen Position liegt. Das zusätzliche Exon 1B der Ratte enthält an seinem 5'-Ende eine fast optimal mit der Matrix übereinstimmende Grenze, die in diesem Bereich ein Exon extrem wahrscheinlich macht.

Um einer Beantwortung der Frage näherzukommen, warum nur die experimentell nachgewiesenen Teile der Sequenzen als Exons erkannt werden, benötigt man also weitere Informationen über Intron- bzw. Exonererkennung. Zuerst wurden die Sequenzen auf potentielle Bindungsstellen von eventuell noch unbekanntem (Spleiß-)Faktoren hin überprüft. Da kein spezielles Programm für diese Suche existiert, wurde stattdessen 'MatInspector' verwendet. Dieses Programm sucht zwar nach Transkriptionsfaktoren, die nicht an einzelsträngige RNA binden, aber bei der großen Zahl der mittlerweile bekannten Faktoren und Matrizen ist es möglich, daß es übereinstimmende Bindungssequenzen gibt. Eventuell gibt es auch gemeinsame Bindungsmotive von Spleiß- und Transkriptionsfaktoren. Anschließend wurde das Muster des Auftretens in den Sequenzen der drei Arten verglichen. Bei einer Übereinstimmung der Matrizenmuster zwischen den drei HCII-Intron 1-Sequenzen wurden zusätzlich andere Serpin-Introns überprüft. Es konnte jedoch kein einheitliches Muster gefunden werden.

Eine andere Möglichkeit für die Nutzung suboptimaler Spleißstellen könnte darin bestehen, daß ihre Sekundärstruktur optimal für den Spleißvorgang ist. Es ist sehr wahrscheinlich, daß die prä-mRNAs Rückfaltungen bilden, die den Zutritt von Spleißfaktoren zur prä-mRNA begünstigen. Für die Überprüfung auf kleine Sekundärstrukturen wurde die Sequenz mit den Programmen 'Oligo 4.0', einem selbstgeschriebenen Programm (nicht aufgeführt) und 'Mfold'¹⁹¹ untersucht. Die beiden erstgenannten Programme ermitteln in kurzen Sequenzen kleine intramolekular komplementäre Bereiche.

Für die Suche wurden 25 Bp und 50 Bp lange Sequenzbereiche mit dem Rest der Sequenz verglichen. Eine Übereinstimmung einer Sequenz wird ausgegeben, falls mindestens die Hälfte der Basen in diesem Bereich komplementär sind, falls es mindestens fünf komplementäre Basen hintereinander gibt und eine Faltungstemperatur (berechnet nach der Annealingtemperatur für Primer: siehe PCR, abzüglich von Fehlpaarungen) von mehr als 30 °C vorhanden ist. Obwohl eine Häufung von kleinen (möglichen) Haarnadelstrukturen an der 5'-Grenze von Introns zu erkennen gewesen ist (die häufig auch das Ende des Exons mit einschloß), konnten keine einheitlichen Strukturen für alle untersuchten Introns gefunden werden. Zusätzlich wurde mit dem Programm 'Mfold'¹⁹¹ die mRNA-Sekundärstruktur in einen Bereich von jeweils 500 Bp über die Exon/Introngrenzen hinweg berechnet. Da jeweils eine große Zahl von möglichen Strukturen mit annähernd gleicher Energie allein für eine Sequenz ausgegeben werden, lassen sich auch mit diesem Programm keine eindeutig übereinstimmenden Strukturen erkennen. Beim Menschen oder bei der Maus wird kein zusätzliches Exon in der zur Ratte homologen Sequenz erkannt (z.B. über die Verwendung von weiter entfernt liegenden möglichen Spleißsignalen). Wahrscheinlich ist hierfür die Sekundärstruktur der gesamten Intronsequenz verantwortlich, die nur in dem speziellen Bereich eine Exonerkenkung erlaubt. Vielleicht besteht nur in dem HCII-Intron der Ratte durch eventuell weiter entfernt liegende inserierte (bzw. fehlende) repetitive Elemente die Möglichkeit für eine Exonerkenkung (s.u.). Die Frage, warum ein zusätzliches Exon in der Sequenz der Ratte erkannt wird und warum bei der Maus das 5'-Ende von Exon 2 in einigen Transkripten verlängert ist, kann also mangels bekannter weiterer Kriterien definitiv nur bezüglich des definierten, lokalen Sequenzbereichs beantwortet werden. Im Punkt E.2.3 wird dennoch versucht Intronbereiche zu identifizieren, die über größere Distanz veränderte Spleißstellen bewirken könnten.

E.2.3 Vergleich der Gesamtstruktur von Intron 1

In diesem Abschnitt wird die Struktur der Intron 1-Sequenzen verglichen, und es wird kurz auf die größeren allgemeinen Mutationen in den Sequenzen eingegangen, die entweder die eine oder die andere Veränderung der Exon/Intronstruktur des Gens für diesen Bereich bewirken könnten. In den nachfolgenden Unterpunkten (E.2.3.1, E.2.3.2) werden spezielle Aussagen zu den Regionen um das Exon 1B und das 3'-Ende des Exons gemacht, die sich auf Veränderungen im Nahbereich dieser Regionen beziehen. In dem Abschnitt E.2.4 werden schließlich detailliert die Sequenzen der jeweiligen Exon/Introngrenzen verglichen. Im folgenden wird das Alignment der Sequenzen von Intron 1 von Ratte, Maus und Mensch unter Punkt D.4.1 diskutiert. Daher wird kurz auf die Signifikanz einer Übereinstimmung von DNA-Sequenzen eingegangen. Werden zwei gleich lange Datenfolgen verglichen, die aus vier verschiedenen Zeichen rein zufällig zusammengesetzt sind, so stimmen sie zu 25 % (4^{-1}) überein. Werden drei dieser Zeichenfolgen verglichen, so liegt die rein zufällige Übereinstimmung bei nur noch 6,25 % (4^{-2}). Zu beachten ist, daß ein Alignment teilweise mit unterschiedlich langen Sequenzen durchgeführt wird. An verschiedenen Positionen werden Lücken eingefügt, die nicht an dieser, stattdessen jedoch an anderen Stellen eine Übereinstimmung ermöglichen. Die Einfügung von Lücken beim Alignment verändert nicht die berechnete Übereinstimmung zwischen den Sequenzen. Für die weiteren Betrachtungen wird also das Alignment mit eingefügten Sequenzlücken verglichen. Bei der menschlichen Sequenz wurden fünf große Insertionen entfernt. Dementsprechend ist die Übereinstimmung zu den anderen Sequenzen erhöht (Alignment-Positionen: 421, 1828, 2258, 3683 und 4237).

¹⁹¹<http://bibiserv.techfak.uni-bielefeld.de/mfold/>; Mathews *et al.*, 1998; Zuker & Stiegler, 1981; Zuker, 1989

Andererseits kann die Insertion eines Elements als nur eine einzige Mutation gerechnet werden, so daß die Übereinstimmung bei Maus und Ratte in diesem Fall wesentlich höher liegt. Aus diesem Grund wird bei den folgenden prozentualen Anteilen ein zweiter, sich darauf beziehender Wert in Klammern angegeben. Dieser Wert wurde um vier große Insertionen korrigiert, die jeweils nur in einer Sequenz vorkommen (Positionen im Alignment: 541->737, 1103->1304, 2709->3074, 4953->5239). Die Summe dieser Positionen wurde von der Gesamtzahl der Alignment-Positionen abgezogen (5745-1052 = 4693) und die Übereinstimmungen auf diesen Wert bezogen. Deshalb gibt der Wert in Klammern ein besseres Bild der Sequenzidentität. Nachfolgend sind die Übereinstimmungen kurz tabellarisch aufgeführt.

Tabelle 42: Verwandtschaft der Intron 1-Sequenzen des HCII-Gens

	Mensch	Maus
Ratte	46 % (57 %)	56 % (68 %)
Maus	33 % (40 %)	

Eine Übereinstimmung liegt bei Maus und Ratte an 3202 Positionen (von 5745) vor. Das entspricht einer Sequenzidentität von 56 % (68 %). Diese Werte liegen deutlich über einer zufälligen Übereinstimmung. Der Wert für den Vergleich der Sequenzen von Ratte und Mensch liegt mit 2667 Positionen bei 46 % (57 %). Bei einem Vergleich der Sequenzen von Maus und Mensch erhält man eine niedrigere Übereinstimmung von 1892 Positionen, dies entspricht 33 % (40 %). Alle Werte zeigen somit eine signifikante Übereinstimmung. Abweichende Werte sind größtenteils durch die großen und kleinen Insertionen in den einzelnen Sequenzen begründet, die sich nicht zuordnen lassen. Eine Ähnlichkeit der drei Sequenzen wird ebenfalls gut durch die Summe der Übereinstimmungen einzelner Positionen aller Sequenzen belegt. Es gibt 1310 in allen Introns identische Positionen, also 23 % (28 %) Übereinstimmung. Dies entspricht einer 3,5fach bzw. 4,5fach höheren Ähnlichkeit, als bei rein zufälligen Sequenzen zu erwarten ist.

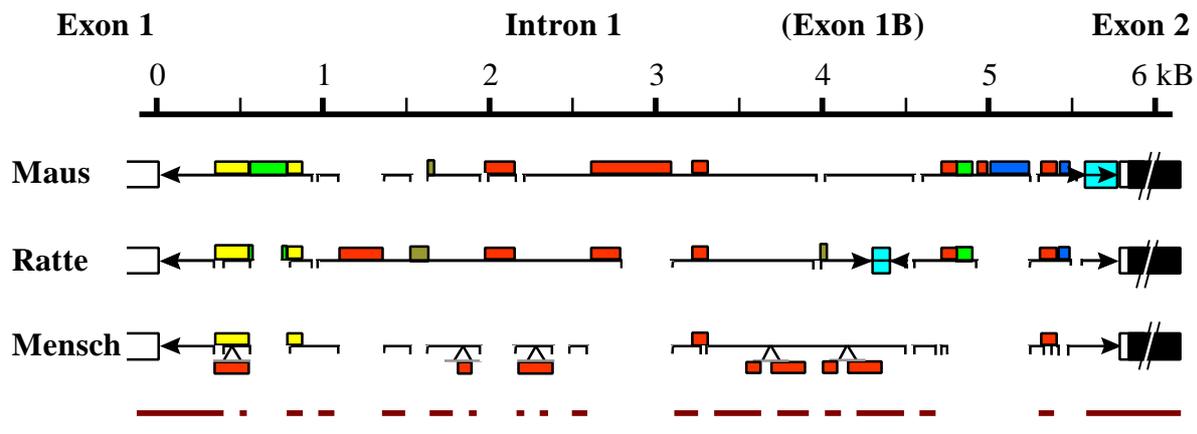
Unter dem Abschnitt E.2.2 wurde bereits erwähnt, daß Exons bzw. Introns wahrscheinlich nicht nur über die lokalen Spleißsignale erkannt werden, sondern eventuell auch über andere Strukturen des Introns. Mögliche Sekundärstrukturen der mRNA werden durch Insertionen bzw. Deletionen von repetitiver DNA und weniger durch Punktmutationen beeinflusst.

Falls eine andere Exon/Intronstruktur in den einzelnen Arten durch die veränderte Intron 1-Struktur bedingt ist, gibt es dafür mehrere Möglichkeiten (siehe Abbildung 45). Obwohl das menschliche Intron 1 Ähnlichkeiten mit den anderen Sequenzen aufweist, wie z. B. anhand des L1-Fragment an Position 0,3-0,8 kB, oder des MIR-Elements im Bereich 3,2 kB zu erkennen ist, so ist die Struktur durch große Insertionen und Deletionen im Vergleich zu Maus und Ratte doch stark verändert.

Ein Beispiel dafür, daß eine größere Insertion Einflüsse (über die mRNA-Struktur) auf die Genstruktur haben kann, wurde von Mühlhardt¹⁹² gegeben. Eine Insertion eines L1-Elementes in das fünfte Intron des *glyrb*-Gens führt in der spastischen Maus zu einem Skipping von Exon 4 und 5, obwohl die Spleißstellen unverändert erhalten sind. Um genauere Aussagen machen zu können, werden im folgenden vor allem die ähnlichen Introns von Maus und Ratte verglichen. Das Alignment der Intron 1-Sequenzen im Teil D.4.1 ermöglicht keinen guten Überblick über die Verteilung von repetitiven Elementen in den Sequenzen, aus diesem Grund ist in der Abbildung 45 ein Schema der Intron 1-Struktur des HCII-Gens der behandelten Arten gegeben.

¹⁹² Mühlhardt *et al.*, 1994

Abbildung 45: Schematische Darstellung repetitiver Elemente im Intron 1 des HCII-Gens



Legende

← →	Intronbereich mit Deletion (>20 Bp)		LINEs, L1-Element-Fragmente
	differenziell gespleißter Exonbereich		SINEs, Alu- bzw. B-Elemente, MIRs
	translatierter Exonbereich		URR1-Elemente
	nicht translatierter Exonbereich		'Simple-Repeats'
	Insertion beim Mensch (>20 Bp)		Direkte Sequenzwiederholungen

Die roten Linien markieren vergleichbare Sequenzabschnitte in den drei untersuchten Arten. Das MIR-Element an Position 3,2 kB wird nur in der menschlichen Sequenz von dem Programm 'RepeatMasker2' erkannt (D.4.). Allerdings ist dem Alignment unter Abschnitt D.4.1 zu entnehmen, daß es deutliche Übereinstimmungen mit der Sequenz in Maus und Ratte gibt. Aus diesem Grund sind auch für Maus und Ratte die Zeichen für SINEs an dieser Position eingetragen.

Fünf größere Unterschiede in der Sequenz fallen bei einem Vergleich der ersten Introns der HCII-Gene von Maus und Ratte auf.

Der diesbezüglich am weitesten 5'gelegene Bereich umfaßt eine große Insertion (246 Bp) eines $(CA)_n$ -Repeats in das L1-Fragment der Maus bei etwa der Position 0,6 kB (bzw. eine entsprechende Deletion in der Sequenz der Ratte). $(CA)_n$ -Wiederholungen können unter bestimmten Bedingungen eine Linksgängigkeit der DNA bewirken und Rekombinationsereignisse verstärken.¹⁹³ Außerdem sind Zusammenlagerungen von zwei Doppelhelices möglich.¹⁹⁴ Ein derartig großes $(CA)_n$ -Repeat wurde in anderen Sequenzen meines Wissens nach noch nicht gefunden. Eine intakte Sequenz dieser Länge würde die DNA sehr stark für Rekombinationsereignisse anfällig machen, allerdings weist dieses $(CA)_n$ -Repeat viele Mutationen auf, so daß die DNA-Konformation weniger beeinflußt wird. Ein weiteres, kürzeres und höher konserviertes $(CA)_n$ -Repeat kommt bei Maus und Ratte im 3'-Bereich des Introns vor (ab Position 4028 bei der Maus mit einer Länge von 63 Bp und ab Position 3801 bei Ratte mit einer Länge von 120 Bp). Für die von dieser Sequenz transkribierte einzelsträngige mRNA mit $(CA)_n$ -Wiederholungen ist allerdings keine Sekundärstrukturbildung bekannt. Sofern also nicht genau das Fehlen einer möglichen Rückfaltung (die bei anderen Sequenzen mit höherer Wahrscheinlichkeit vorkommt) in diesem Bereich eine veränderte Exon/Intronstruktur bewirkt, hat diese Mutation für die mRNA keine Auswirkungen.

¹⁹³ Hamada *et al.*, 1984; Wahls *et al.*, 1990

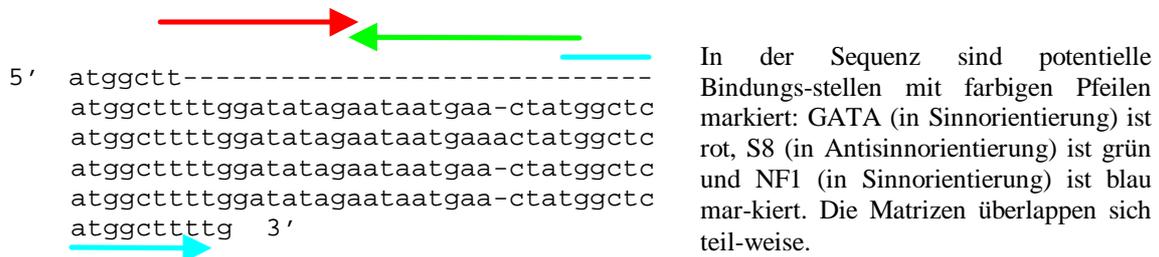
¹⁹⁴ Gaillard & Strauss, 1994

Weiterhin gibt es Insertionen von B-Elementen in das Ratten-Intron 1 (Position 1103 bis 1344) und von Position 2709 bis 3074 in der Maus (bzw. eine Deletion in den jeweiligen Sequenzen der Maus bzw. der Ratte).

Unter dem Gesichtspunkt von unterschiedlichen Genstrukturen zwischen Ratte und Maus ist beachtenswert, daß ein inseriertes B-Element unter Umständen eine artifizielle Spleißstelle bewirken kann.¹⁹⁵ Eventuell bedingen somit inserierte SINES die veränderten Exon/Introngrenzen. Bei beiden Introns besteht etwa 3 kB (siehe Abbildung 45) nach den zusätzlich vorhandenen B-Elementen eine in den anderen Arten nicht erkannte 5'-Exongrenze. Da jedoch mehrere B-Elementfragmente in diesen (wie auch vielen anderen) Introns vorkommen, andererseits jedoch viele Introns keine derartigen Fragmente aufweisen, muß eine Veränderung der Exon/Intronstruktur (allein) durch eine Insertion der Elemente bezweifelt werden.

Eine weitere Insertion (bzw. eine Deletion in der Sequenz der Ratte!) eines repetitiven Elementes, in diesem Fall ein URR1-Element, liegt zwischen den Positionen 4952-5210 in der Intronsequenz der Maus vor. Das zusätzliche Exon in der Ratte kann durch diese Insertion wahrscheinlich nicht bewirkt werden, denn beim Menschen existiert dieses Element sowie das zusätzliche Exon ebenfalls nicht. Allerdings besteht die Möglichkeit, daß einige der zahlreichen anderen Sequenzunterschiede beim Menschen die Erkennung eines Exonbereichs verhindern. Eine Auswirkung auf die etwa 0,5 kB entfernte, alternativ gespleißte 5'-Exongrenze des Maus-HCII-Gens bleibt allerdings möglich.

Eine interessante kleinere Insertion ist in der Ratte ab Position 1449 (bis Position 1648) gegenüber der Maus vorhanden über eine vierfache Wiederholung eines 35 Bp-Sequenz-Motivs. Diese Sequenz zeigt Ähnlichkeiten zu Bindungsstellen von bekannten Transkriptionsfaktoren (GATA, S8 und NF1; siehe Skizze). Nachfolgend ist der Sequenzabschnitt so dargestellt, daß man die wiederholten Bereiche gut erkennen kann.



Der Transkriptionsfaktor, der an die GATA-Sequenz bindet, ist zum Beispiel wichtig für die Regulation von Globin- und Nicht-Globingenen in Erythroid-, Megakaryozyten- und Mastzelllinien.¹⁹⁶ S8 und NF1 (Nuklearfaktor 1) sind universelle Transkriptionsfaktoren (siehe auch Tabelle 34). Die vierfache Sequenzwiederholung könnte eine verstärkte Bindung der entsprechenden Faktoren bewirken und somit zu einer veränderten Transkriptionsrate führen. Ein verändertes Spleißen der mRNA ist nur denkbar, sofern in diesem Fall noch unbekannte Spleißfaktoren ebenfalls diese Sequenz erkennen. Da jedoch schon viele für das Spleißen benötigte Faktoren bekannt sind, sind Bindungsstellen im mittleren Bereich eines Introns weniger wahrscheinlich.

Zusammenfassend kann somit gesagt werden, daß keine größeren Insertionen bzw. Deletionen in unmittelbarer Nähe bzw. über den Bereich der jeweils zusätzlich erkannten Exonregionen gefunden werden, die Unterschiede in der Genstruktur bewirken können. Somit müssen kleinere Mutationen in unmittelbarer Nähe die Veränderungen (hauptsächlich) bewirken. Eine Beteiligung von repetitiven B- und URR1-Elemente sowie des oben gezeigten Sequenzmotivs kann jedoch nicht vollständig ausgeschlossen werden.

¹⁹⁵ Pattanakitsakult *et al.*, 1992

¹⁹⁶ Whyatt *et al.*, 1993

E.2.3.1 Vergleich der Intronstrukturen bezüglich Exon 1B

Bei einem Vergleich der Intron 1-Sequenzen zwischen Maus und Ratte in der Region um das Exon 1B der Ratte fällt auf, daß gerade hier nur wenige größere Veränderungen zu finden sind (Tabelle 26, Abbildung 45). Ungefähr 0,5 kB stromaufwärts des Exons existieren leicht versetzte, kleinere Deletionen (bzw. Insertionen) in beiden Sequenzen und wenige, kurze, direkte Sequenzwiederholungen in der Sequenz der Ratte. Stromabwärts des Exons gibt es ebenfalls kleine Deletionen/Insertionen in beiden Arten. Auffällige Sequenzmotive sind bis auf die genannten Wiederholungen nicht zu erkennen. Eine stark veränderte Sekundärstruktur der prä-mRNA ist also nicht zu erwarten. Weiterhin möglich bleiben andere Bindungsstellen für Spleißfaktoren. Besonders beachtenswert sind hier die genannten kurzen Sequenzwiederholungen in der Sequenz der Ratte (Position 3921-3959). Hier kann der gleiche Faktor jeweils mehrere Bindungsstellen erkennen. Einige Transkriptionsfaktoren könnten theoretisch gut an diese Sequenz binden (Übereinstimmung mit den Matrizen für DeltaEF, E47, LMO2COM, MyoD und USF; Ergebnisse nicht gezeigt). Darüber hinaus existieren weiter entfernt (ab Position 1449, s. o.) die oben besprochenen, direkten Wiederholungen in der Sequenz der Ratte.

Bisher sind die genauen Bindungsstellen für die U1- und U2snRNA bekannt (5'-Introngrenze beziehungsweise Verzweigungsstelle), die anderen beteiligten snRNAs (U4, U5 und U6) lagern sich an das 3'-Ende eines Introns an. Daher kann zur Zeit keine abschließende Aussage über eine eventuelle Funktion kurzer interner Intronbereiche für diese oder bisher noch unbekannte Faktoren gemacht werden.

E.2.3.2 Vergleich der Intronstrukturen bezüglich des 3'-Intronbereichs

Bei einem Vergleich der 3'-Enden der jeweiligen Intron 1-Sequenzen fällt auf, daß bei Maus und Ratte gegenüber der menschlichen Sequenz mehrere kleine Bereiche Deletionen aufweisen (bzw. die menschliche Sequenz zeigt kurze Insertionen). Die Deletionen sind bei der Ratte größer (aus diesem Grund ist nur bei der Ratte in Abbildung 45 das Symbol für eine Deletion eingetragen) und könnten der Grund für eine weitere erkannte Spleißstelle für das 3'-Ende des Introns in der Maus sein. Eine genauere Analyse der Sequenzen wird unter E.2.4 durchgeführt.

E.2.4 Alignment der Exon/Introngrenzen

In den vorausgegangenen Teilen wurde diskutiert, durch welche evolutiven Veränderungen in den Introns in bestimmten Regionen die Möglichkeit für neu erkannte Exonbereiche geschaffen werden konnte. Im folgenden wird nun speziell der zusätzlich auftretende Exonbereich bei Maus und Ratte mit den homologen Bereichen der anderen Introns zur Beantwortung der Frage verglichen, warum nur in der einen Sequenz ein Exon erkannt wird.

E.2.4.1 Alignment der Exon/Introngrenze am 3'-Ende von Exon 1

Dieser Bereich ist nur aus Gründen der Vollständigkeit hier angegeben. Diese Exongrenze ist bei Maus, Ratte und Mensch unverändert erhalten. Wie in Abbildung 46 zu erkennen ist, stimmen alle drei Sequenzen an der Spleißstelle gut überein. Aufgrund der unveränderten Sequenz an diesen Positionen wird von dem Programm '5'Intron.bas' eine Matrix-Übereinstimmung von jeweils 92 % angegeben.

Abbildung 46: Exon/Introngrenze am 3'-Ende von Exon 1 (Alignmentpositionen: -30 - 39)

Ratte	TCTGCCACGCCACTGCTCAGAGGCACAGAG	gt	aagtttggttactaacatTTTTTgcaa-tataaTTTTT	
Maus	CTGCCAGACCCTGCTCAGAGGTACAGAG	gt	aagtttggtttctaacatTTTTTgcaattataaTTTTT	
Mensch	TCTGAAGCGCCACTTCTCAGAAACACAGAG	gt	aagttgggtttctaagtTTT-ctgctgattataaattatt	
Matrix		ccAG	GT	AAGT

In der Abbildung stehen kleine Buchstaben der Matrix für Positionen, in der weniger als 50 % der angegebenen Nukleotide konserviert sind. Die unveränderlichen Positionen an den Exon/Introngrenzen wurden eingerahmt. Die Exon 1-Bereiche (max. sind 30 Bp gezeigt) sind fett gedruckt und mit großen Buchstaben geschrieben. Übereinstimmende Sequenzen mit der Matrix für entsprechende Spleißsignale sind unterstrichen. Abweichungen von der Matrix sind blau hinterlegt.

E.2.4.2 Alignment der Exon 1B-Region

In Abschnitt E.2.3 wurde festgestellt, daß eine unterschiedliche Genstruktur und somit auch das Auftreten des Exon 1B vom HCII-Gen der Ratte (vornehmlich) auf kleinere (Punkt-) Mutationen zurückzuführen sein muß. In der folgenden Abbildung ist die Sequenz der drei Arten (± 50 Bp) im Bereich des Exons 1B der Ratte dargestellt (das Alignment wurde dem Teil D.4.1 entnommen und zur Verdeutlichung etwas abgeändert). Im folgenden wird anhand von Abbildung 47 dargestellt, aus welchem Grund sich die Exon/Intronstrukturen von Maus und Ratte in diesem Bereich unterscheiden.

Wie gut zu erkennen ist, kann ein zur Ratte vergleichbares Exon in diesem Bereich bei der Maus nicht vorhanden sein, da ein zur 5'-Spleißstelle in Ratte homologes Signal infolge einer Punktmutation nicht vorhanden ist (Position 4405 im Alignment, gelb hinterlegt). Das am nächsten gelegene, potentielle Spleißsignal (nach D.3.3, Tabelle 22) befindet sich an Position 3629 (Position 4344 im Alignment) in der Sequenz (statt Position 3642). Die potentielle Exonsequenz ist eventuell zu kurz für einen Spleißvorgang.

Abbildung 47: Vergleich der Sequenzen um den Bereich von Exon 1B der Ratte

	4260	4280	4300	4320
Ratte	atgatatgTTTctagaa-ggctggTTCatgTTCat-tcacaatgtcctccttctccctc	ag		
Maus	acaatatgTTTctagaa-ggctggctcatgTTCgT-tcataa---	ag	g	
Mensch	agaatatgTTTtcagaaagcctggcccacgacattTTTcaacagcattctccattgcaa	ag	gtgg	aatggTg
Matrix		YnYYUAY	YnYYUAY	YYYYYYYYYYYnC
				AG
				G
				ccAGGTA
	4340	4360	4380	4400
Ratte	TGAGCAACAGTTTGGTAGCGCTGAGTCCACTGGA-AATTCTCCT-CCT-GTCT C			AG
Maus	tgagcaggagTTTgtaaggcttggtccatctactggacactcctcctgct-cactc			tag
Mensch	tgagtcacaca--ggcatggctgagtcccactaatgcacatcccttctaggtactctccaatcaccagccccag			
Matrix	AGT	ccAGGTAAGT		cc-----AG
	4420	4440	4460	
Ratte	gt atca-ctctagctcaactctta-tgaaacctcttgactctctgggcacccaccct			
Maus	gc attcactctagcccagctctta-gaaaacctctgtaactctctggacactctcccttccctgccccaaa			
Mensch	gt gcccactcaagcccagctcttagtgaggTTTccctgactctctgggcaett			cca-----
Matrix	GT	AAGT		

In der Abbildung¹⁸⁶ stehen kleine Buchstaben der Matrix für Positionen, in der weniger als 50 % der angegebenen Nukleotide konserviert sind. 'Y' und 'U' stehen für Pyrimidine bzw. für Purine. Die unveränderlichen Positionen an den Exon/Introngrenzen wurden eingerahmt. Das Exon 1B der Ratte ist kursiv, fett und mit großen Buchstaben geschrieben. Übereinstimmungen mit der Matrix für entsprechende Spleißsignale sind unterstrichen. Abweichungen von der Matrix sind blau, gelb oder rot hinterlegt.

Die am nächsten stromabwärts gelegene 5'-Spleißstelle wäre an Position 3850 (im Alignment: 4565), also über 200 Bp entfernt. Eventuell ist ein Spleißen an dieser Stelle durch die prä-mRNA-Struktur nicht mehr möglich.

Abgesehen davon ist auch die homologe Region der Maus an der potentiellen 3'-Spleißstelle für ein Spleißen nicht geeignet. Es liegen z. B. ungünstige Mutationen im Polypyrimidinbereich vor (Alignmentpositionen 4300-4303, 4305). Der ununterbrochene Pyrimidinbereich ist gegenüber der Ratte (15 Bp) um 6 Bp auf 9 Bp verkürzt und verschlechtert damit das Spleißsignal.¹⁹⁷ Diese Mutation sollte aber die Erkennung einer Spleißstelle nicht vollständig ausschließen.¹⁹⁷ Der Hauptgrund für eine ungünstige 3'-Spleißstelle an dieser Position liegt jedoch in einer nicht vorhandenen Verzweigungsstelle. Obwohl die Matrix für eine Verzweigungsstelle nur sehr schwach definiert ist,¹⁹⁸ liegt kein optimales A-Verzweigungsnukleotid vor. Durch eine Punktmutation ist ein 'A' in der Maus gegenüber einem 'G' in der Ratte ausgetauscht (im Alignment: Position 4291, rot hinterlegt). Ein weiterer Bereich (Pos. 4293-4300) für eine eventuelle Verzweigungsstelle (die in der menschlichen Sequenz von dem Programm erkannt wird) wird ebenfalls nicht erkannt, da sie durch eine Deletion von 4 Bp zu nahe an der Spleißstelle liegt. Die stromaufwärts nächstgelegene, berechnete 3'-Verzweigungs- und Spleißstelle ist an Position 3449 (Alignmentposition 4124). Die 135 Bp entfernte Stelle scheint nicht mehr im Bereich einer möglichen Exongrenze zu liegen.

Ein Exon an der homologen Position in der menschlichen Sequenz kommt aus anderen Gründen nicht vor. Ein 3'-Spleiß-AG ist an identischen Positionen vorhanden, obwohl eine (im Vergleich zur Sequenz der Ratte) andere Verzweigungsstelle verwendet werden müßte. Eine gleiche Position der 5'-Spleißstelle wäre hingegen unwahrscheinlich, da die Übereinstimmung mit der Matrix sehr niedrig ist. Ausgeschlossen werden kann die Möglichkeit jedoch nicht vollständig, da das unbedingt benötigte GT-Dinukleotid ab der entsprechenden Position 4404 vorhanden ist. Eine weitere potentielle 5'-Spleißstelle kommt an Alignmentposition 4327 (Pos. 4528 in Tabelle 24) vor, also unmittelbar nach der 3'-Spleißstelle und scheidet somit aus. Die nächste mögliche GT-Spleißstelle wird erst wieder an Position 4807 erkannt, also über 200 Bp weiter stromabwärts und somit ist eine Erkennung in diesem Bereich eher unwahrscheinlich. Im übrigen muß an dieser Stelle noch einmal auf die stark veränderte Intronstruktur vom Menschen gegenüber der Ratte hingewiesen werden (durch andere/fehlende Insertionen von repetitiven Elementen), die eventuell gar kein Exon in diesem Bereich zulassen.

Aufgrund der theoretischen Daten kann ein zusätzliches Exon beim Menschen in diesem Bereich nicht ausgeschlossen werden. Mittels einer RT-PCR (siehe D.4.3, Abbildung 24) konnte jedoch kein PCR-Fragment von einer HCII-mRNA amplifiziert werden, aus dem ein Intronbereich ausgeschnitten ist (es wurde nur das 793 Bp-PCR-Produkt erhalten, das auf dem langen Transkript oder genomischer DNA beruht). Daher kann ein der Ratte entsprechendes Exon 1 B im Mensch ausgeschlossen werden.

In dieser Arbeit kann experimentell gezeigt werden, daß keine Exons bei Mensch und Maus vorkommen, die dem HCII-Exon 1 der Ratte entsprechen. Bisher wurde ebenfalls kein vergleichbarer Exonbereich beim Kaninchen gefunden. Aus diesem Grund kann nach dem Sparsamkeitsprinzip angenommen werden, daß das Exon 1B der Ratte (vor allem) durch wenige Punktmutationen neu entstanden ist und es nicht in den anderen Arten verloren ging.

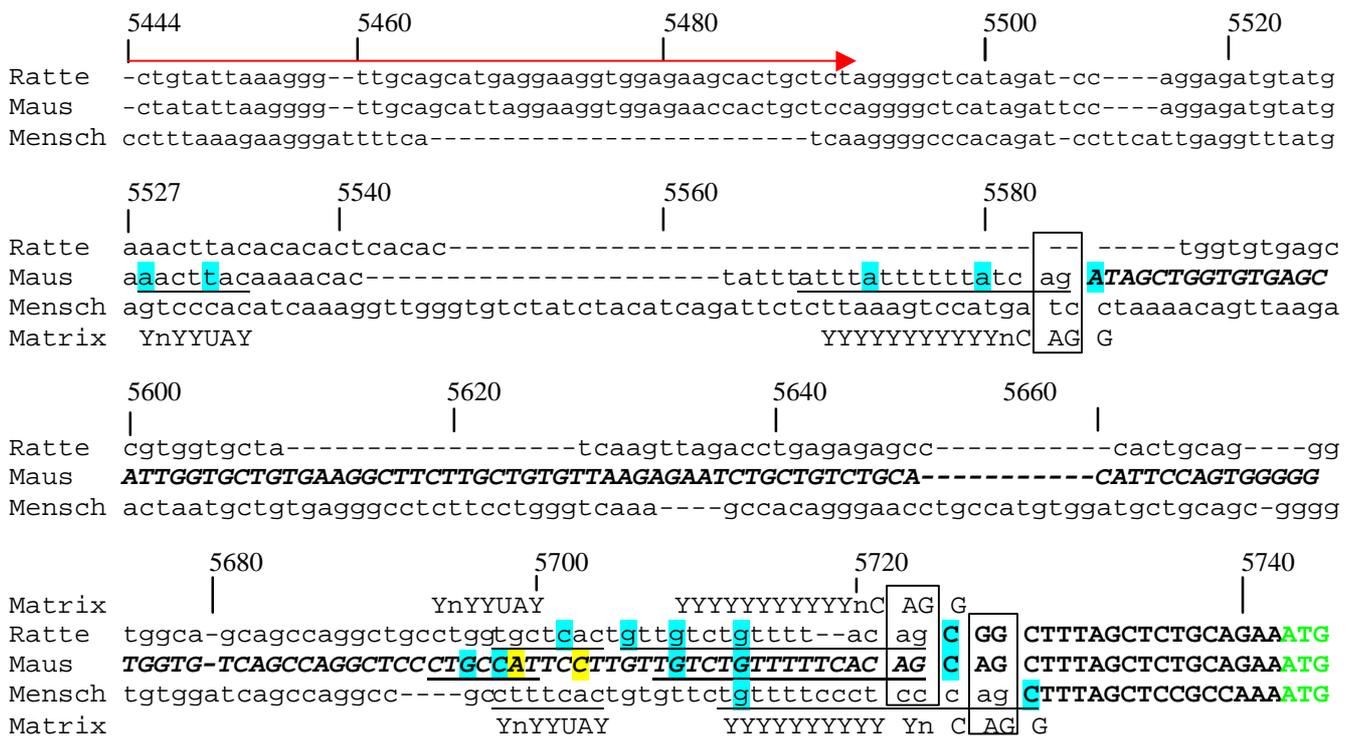
¹⁹⁷ Coolidge *et al.*, 1997

¹⁹⁸ Lewin, 1991

E.2.4.3 Alignment des 5'-Bereichs von Exon 2

Die tatsächlich genutzten 3'-Grenzen von Intron 1 des HCII-Gens unterscheiden sich bei allen drei Arten. Wenn davon ausgegangen wird, daß aufgrund des allgemeinen Aufbaus von Intron 1 in Ratte, Maus und Mensch nur an dieser Region eine Exon/Introngrenze möglich ist, so läßt sich die Variabilität am besten durch das Auftreten von kleineren (Punkt-) Mutationen erklären. Im folgenden werden die Veränderungen in den Sequenzen besprochen, die zu einer unterschiedlichen Genstruktur in dieser Region führen. In Abbildung 48 ist das Alignment dieses Bereiches noch einmal (verändert nach D.4.1) dargestellt.

Abbildung 48: Bereich der 3'-Grenzen des HCII-Intron 1



In der Abbildung steht ein 'Y' für Pyrimidine, ein 'U' steht für Purine. Die unveränderlichen Positionen an den Exon/Introngrenzen sind eingerahmt und Exonbereiche sind fett und mit großen Buchstaben gedruckt. Der alternativ gespleißte Exonteil bei der Maus ist kursiv geschrieben. Das Exon 2 ist nur bis zum Startkodon (grün) abgebildet. Mit der Matrix für entsprechende Spleißsignale übereinstimmende Sequenzen sind unterstrichen. Abweichungen von der Matrix sind blau hinterlegt. Der rote Pfeil markiert das 3'-Ende eines mit RepeatMasker2¹⁵⁷ identifizierten URR1B-Elementes (siehe D.4.1).

Als erstes kann gesagt werden, daß nach der Analyse auf potentielle 3'-Introngrenzen beim Menschen in diesem Bereich eine einzige mögliche Spleißstelle vorliegt (Position 5150 bzw. Positionen 5725 + 5726 des invarianten 'AG' im Alignment), die auch sehr gut mit der Matrix übereinstimmt (88 % bzw. 169 Punkte). Eine andere theoretische Spleißstelle ist erst 578 Bp weiter stromaufwärts (siehe D.3.3, Tabelle 25) lokalisiert. Es ist daher gut zu verstehen, daß beim Menschen nur diese Position als Spleißstelle in der prä-mRNA verwendet wird.

Die nach dem Alignment homologe Exon/Introngrenze bei Maus und Ratte ist um drei Basenpaare weiter zur 5'-Seite hin verlagert. Diese Verschiebung kann durch einfache Punktmutationen erklärt werden. Bei Maus und Ratte ist das für eine 3'-Spleißstelle unbedingt benötigte 'AG'-Dinukleotid an Positionen, an denen beim Menschen ein 'CC'-Duplett vorliegt. Das beim Menschen verwendete 'AG' ist hingegen bei der Ratte durch ein 'GG' ersetzt.

Bei der Maus liegt hier keine Veränderung gegenüber der menschlichen Sequenz vor, und die zur menschlichen Spleißstelle identische Position wird auch von dem Programm erkannt. Die Übereinstimmung mit der Matrix fällt aber niedriger aus (Position 4843 in der Sequenz der Maus, 73 % Übereinstimmung gegenüber 80 % an Position 4840; Tabelle 23). Eventuell wird aus diesem Grund nur das stromaufwärts gelegene Spleiß-AG in der prä-mRNA verwendet. Eine alternative Begründung für die Nutzung einer stromaufwärts gelegenen Spleißstelle anstatt der anderen nahegelegenen Stelle ist ein von 5'- nach 3'-verlaufender Scanning-mechanismus der Spleißfaktoren.¹⁹⁸

Die mutmaßlichen Verzweigungsstellen weichen ebenfalls voneinander ab.¹⁹⁹ Während bei der Ratte und beim Mensch eventuell homologe Stellen verwendet werden (Positionen 5698-5704 im Alignment), sind an zwei wichtigen Stellen der Sequenz der Maus (Position 5699 und 5703 im Alignment, gelb hinterlegt) andere Nukleotide in der DNA-Sequenz enthalten. Das potentielle Verzweigungsnukleotid 'A' in den beiden anderen Sequenzen ist durch ein 'C' ersetzt und stattdessen ist 4 Bp stromaufwärts ein anderes Nukleotid gegen ein 'A' ausgetauscht, so daß eine neue potentielle Verzweigungsstelle entsteht. Diese Stelle ist allerdings von der Bewertung her ungünstiger als die andere Position (nur 61 % gegenüber 81 % bei Mensch und Ratte; Tabelle 20, Tabelle 23 und Tabelle 25). Möglicherweise ist die weniger gute Spleißstelle ein zusätzlicher Grund für die alternative Verwendung der Position 4713 bei der Maus (Alignmentposition: 5583). Es gibt keine gute Übereinstimmung zu der Matrix (73 %, 114 Punkte). Aus diesem Grund kommt die Sequenz wahrscheinlich nur in einem Teil der Maus-HCII-mRNAs vor.

In den beiden anderen Arten kann an dieser Position keine Spleißstelle vorhanden sein, da sich in diesem Bereich die Sequenzen völlig unterscheiden. Im Ratten-Gen liegt hier eine größere Deletion vor, die den gesamten Polypyrimidinbereich und das 'AG' vom Maus-Gen umfaßt. Gegenüber dem Menschen gibt es auch bei der Maus vor dem Polypyrimidinbereich eine Deletion. Etwa 50 Bp vor dieser Deletion endet bei allen Arten das gleiche MER-Element (URR1B: Das Element wird von 'RepeatMasker2' in der menschlichen Sequenz aufgrund größerer Mutationen nicht erkannt). MER-Elemente²⁰⁰ ('Medium reiteration frequency sequences') kommen bei allen Mammaliern vor, sind also vor der Auftrennung in die einzelnen Arten inseriert. Daher sollten sich die Sequenzen sowie ihre Enden ähneln. Allerdings gibt es auch bei anderen Elementen, die bei nahe verwandten Arten vorliegen und evolutiv gesehen vor kurzem inseriert sind, Abweichungen an den Endbereichen. Eine Möglichkeit für den Mechanismus dieser Mutationen ist die nachträgliche Veränderung über homologe Rekombination mit identischen Elementen aus anderen Bereichen des Genoms oder über Rekombination mit revers transkribierten RNAs, die die Sequenz dieser Elemente enthalten. Im letzteren Fall werden häufig 'A'-reiche Sequenzen (bzw. revers: 'T'-reich) ins Genom eingebaut, die von polyadenylierten RNAs abstammen (vergleiche: Abbildung 4).²⁰¹ In Abbildung 48 ist ein 'A/T'-reicher Sequenzabschnitt in der Sequenz der Maus vor der stromaufwärts gelegenen Exon/Introngrenze innerhalb der Positionen 5527-5581 zu erkennen. Dieser Bereich könnte durch einen derartigen Prozeß entstanden sein und den nur in der Maus vorhandenen Polypyrimidin-Abschnitt bewirken, der für die Erkennung der Exon/Introngrenze benötigt wird.

¹⁹⁹ Nach dem Programm: '3'Intron.bas' (C.2.1), basierend auf: Lewin, 1991; Krawczak *et al.*, 1992

²⁰⁰ Smit, 1993

²⁰¹ Darnell *et al.*, 1990; Jurka & Klonowski, 1996

Die Beteiligung eines weiteren repetitiven Elements an der veränderten Genstruktur kann praktisch ausgeschlossen werden, da die Suche in DNA-Datenbanken²⁰² keine Ähnlichkeiten mit bereits bekannten Sequenzen zeigt und ein genomischer Southernblot von Maus-DNA mit einer zu diesem Bereich identischen Sonde nur jeweils eine Bande liefert. Die gleiche Bande kann auch mit einer menschlichen HCII-spezifischen Sonde nachgewiesen werden, also existieren die zusätzlichen Fragmente in der Maus nur einmal im Genom (siehe D.5).

Es gibt eine weitere alternative Erklärungsmöglichkeit für den Mechanismus der aufgetretenen Sequenzveränderungen in diesem Bereich. Die Unterschiede zwischen Ratte und Maus sind eventuell unter Beteiligung der kurzen, in der Sequenz der Ratte vorkommenden $(CA)_n$ - bzw. $(TG)_n$ -Repeats (Positionen 5533-5546 und 5590-5596 im Alignment) entstanden. Die Enden dieses Sequenzbereiches sind schwach konserviert auch in der Sequenz der Maus enthalten. Es ist gut möglich, daß hier nach einem Doppelstrangbruch im $(CA)_n$ -Repeat der DNA eines Vorläufers von Maus oder Ratte der Polypyrimidinbereich in der Maus eingefügt bzw. in der Ratte deletiert wurde (siehe B.2.1).²⁰³

Die besprochenen Mutationsvorgänge (Insertionen von repetitiven Elementen, homologe Rekombination, Mutationen durch $(CA)_n$ -Repeats) ziehen normalerweise über fehlgepaarte Basen und somit einzelsträngige DNA-Bereiche Reparaturvorgänge nach sich (Aktivierung von RecA-ähnlichen und weiteren Reparaturproteinen), die nachfolgend die DNA-Sequenz ebenfalls verändern können.²⁰⁴ Eine genaue Aussage, wie eine Sequenz durch DNA-Reparatur verändert wird (mit der Ausnahme, daß Strangbrüche und Fehlpaarungen behoben sind) läßt sich jedoch nicht grundsätzlich geben.

E.2.5 Zusammenfassung

Aus den Vergleichen der Sequenzen des Exon 1B-Bereichs und des 3'-Bereichs von Intron 1 läßt sich das Folgende zusammenfassen:

Das Exon 1B in Ratte ist mit großer Wahrscheinlichkeit durch Punktmutationen neu entstanden. Ein entsprechendes Exon im Maus HCII-Gen an der Stelle des Exons 1B von Ratte ist durch jeweils eine Punktmutation in der Verzweigungsstelle und in der 5'-Spleißstelle ausgeschlossen. Beim Menschen wird eine Exonerkenung an dieser Stelle durch die ungünstige Umgebung der 5'-Spleißstelle und eventuell der gesamten Intronstruktur behindert.

Die veränderten 5'-Grenzen von Exon 2 sind durch mehrere Punktmutationen bedingt. Dabei scheint eine Verschiebung der Exongrenze um 3 Bp zur 5'-Seite bei Maus und Ratte durch ein in der Sequenz neu vorhandenes 'AG'-Duplett bewirkt zu sein. Eine andere Möglichkeit ist die Verschiebung der menschlichen Exon/Introngrenze. In diesem Fall wäre ein zuvor vorhandenes, der Maus- bzw. Ratte entsprechendes 'AG' mutiert.

Die 5'-gelegene, alternativ gespleißte Exon/Introngrenze vom Maus-HCII-Gen kommt eventuell durch ein verändertes Ende eines repetitiven Elements (URR1B) oder durch Rekombination innerhalb eines $(CA)_n$ -Repeats zustande.

Die aufgetretenen Veränderungen der Exon/Intronstruktur wurden durch Punktmutationen ermöglicht. Weitere Mutationen in der Sequenz wie die Duplikationen von kurzen Sequenzen, Insertionen und Rekombination zwischen früher inserierten repetitiven Elementen können allerdings über eine veränderte RNA-Struktur ein zusätzliches Exon verhindern oder begünstigen.

²⁰² Suche mit 'BLAST2': <http://www.ncbi.nlm.nih.gov/BLAST/>

²⁰³ Hamada *et al.*, 1984; Wahls *et al.*, 1990

²⁰⁴ Knippers *et al.*, 1990

E.3 Vergleich verschiedener cDNAs von Heparinkofaktor II

Bisher wurden die nicht translatierten Genbereiche des HCII-Gens betrachtet. Für Aussagen, ob in diesen Bereichen spezielle und gehäufte Mutationen auftreten, können zum Vergleich die Exonbereiche des HCII-Gens betrachtet werden. Diese Bereiche beinhalten aufgrund der stärkeren Selektion auf eine sinnvolle proteinkodierende Sequenz weniger Mutationen.²⁰⁵ Rastermutationen und Mutationen von Kodons von essentiell für die Proteinfunktion wichtigen Aminosäuren, die zu einem inaktiven HCII-Protein führen, werden normalerweise ausselektiert, da bisher nur wenige Patienten mit einer totalen HCII-Defizienz bekannt sind.²⁰⁶ Um Mutationen festzustellen, die im Verlauf einer längeren Evolutionsspanne in den Exonbereichen des HCII-Gens aufgetreten sind, wurde ein Alignment der bisher bekannten cDNA-Sequenzen durchgeführt. Eine andere Möglichkeit ist ein Vergleich der Proteine auf Aminosäureebene (durchgeführt von Colwell und Tollefsen²⁰⁷). Ein derartiger Vergleich gibt allerdings keinen genauen Aufschluß über einzelne Mutationen in der chromosomalen Sequenz. Für ein Alignment wurden neben den cDNA-Sequenzen der Maus, der Ratte und des Menschen auch die Sequenzen des Kaninchens, des Huhns und des Frosches miteinbezogen. Die letztgenannten cDNAs sind bisher unvollständig, so daß am 5'-Ende kein vollständiger Vergleich möglich ist.

E.3.1 Vergleich der Heparinkofaktor II-cDNAs

E.3.1.1 Evolutiver Vergleich

Mit Hilfe des Programms 'ClustalW' werden die bisher bekannten cDNA-Sequenzen von Maus (2203 Bp), Ratte (2218 Bp), Mensch (2236 Bp), Kaninchen (2240 Bp), Huhn (1834 Bp) und Frosch (1878 Bp) in einem Alignment (D.6; Tabelle 29) einander gegenübergestellt.²⁰⁸ Der kodierende Bereich vom ATG-Startkodon bis zum Stopkodon umfaßt mit eingefügten Leerstellen 1501 Bp. Die jeweils angegebene Gesamtlänge bezieht sich auf die bisher bekannte Sequenz bis zur letzten Base vor der poly(A)-Sequenz der cDNA. In den folgenden Tabellen sind die identisch besetzten Positionen bezüglich des translatierten Bereiches und der nicht translatierten Bereiche angegeben. Die prozentuale Übereinstimmung bezieht sich bei dem translatierten Bereich auf die erwähnten 1501 Bp. Bei den nicht translatierten Regionen werden nur die Positionen bis zu den Enden der jeweils kürzeren Sequenz verglichen.

Tabelle 43: Vergleich der translatierten Bereiche von HCII-cDNAs auf DNA-Ebene

		absolute Zahl der identischen Positionen					
		Ratte	Maus	Kaninchen	Mensch	Huhn	Frosch
% Übereinstimmung	Ratte		1353	1200	1191	958	914
	Maus	90,1		1208	1199	974	918
	Kaninchen	79,9	80,5		1256	969	924
	Mensch	79,3	79,9	83,7		986	929
	Huhn	63,8	64,9	64,6	65,7		939
	Frosch	60,9	61,2	61,6	61,9	62,6	

²⁰⁵ Csihak *et al.*, 1992

²⁰⁶ Villa *et al.*, 1999

²⁰⁷ Colwell & Tollefsen, 1998

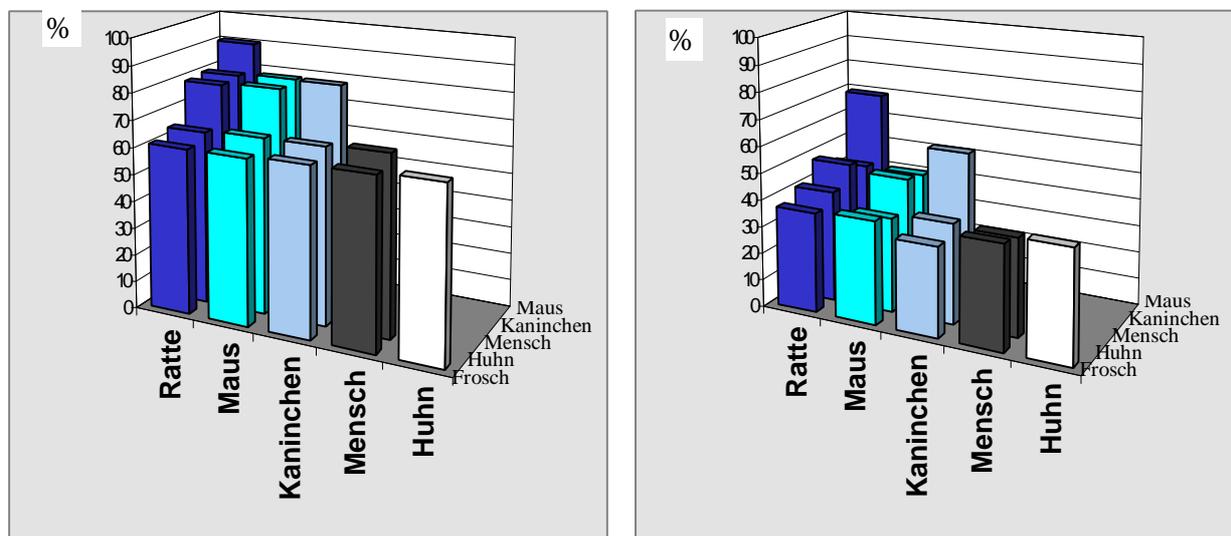
²⁰⁸ HCII-cDNA-Sequenzen: Frosch und Huhn: Colwell & Tollefsen, 1998; Kaninchen: Sheffield *et al.*, 1994; Maus: Zhang *et al.*, 1994, diese Arbeit; Mensch: Herzog *et al.*, 1991, diese Arbeit; Ratte: Westrup & Ragg, 1994

Tabelle 44: Vergleich der 5'- und 3'-nicht translatierten Bereiche von HCII-cDNAs

		absolute Zahl der identischen Positionen					
		Ratte	Maus	Kaninchen	Mensch	Huhn	Frosch
% Übereinstimmung	Ratte		482	315	346	135	139
	Maus	68,7		307	322	118	142
	Kaninchen	43,9	43,7		434	124	123
	Mensch	48,3	45,9	59,0		120	143
	Huhn	40,5	35,4	37,2	36,0		136
	Frosch	36,9	37,7	32,6	37,9	40,8	

Um diese Werte anschaulich darzustellen, wurden sie in die beiden anschließenden Diagramme eingetragen.

Abbildung 49: Übereinstimmungen zwischen den HCII-cDNAs



Übereinstimmungen in den translatierten (links) und nicht translatierten Bereichen (rechts) der HCII-cDNAs verschiedener Organismen

Den Diagrammen ist zu entnehmen, daß die Übereinstimmung in kodierenden Bereichen höher ist als in den nicht kodierenden Regionen. Dies bestätigt alte Befunde. Darüber hinaus ist zu sehen, daß die Übereinstimmungen der nichttranslatierten Bereiche der cDNAs um etwa 20-25 % niedriger sind als diejenigen des kodierenden Bereiches. Austausch in den nicht translatierten Bereichen kommen aufgrund des fehlenden Selektionsdrucks wesentlich häufiger vor. Aus diesem Grund können diese Werte nicht mit bekannten Werten verglichen werden, die auf Vergleichen von Proteinen bzw. proteinkodierenden Sequenzen beruhen.

Anhand des linken Diagramms können folgende Aussagen getroffen werden. Die niedrigste Übereinstimmung der anderen Arten gibt es im Vergleich zum Frosch mit etwa jeweils 60 %. Etwas höher liegt die Ähnlichkeit zum Huhn mit ca. 65 %. Die hier aufgeführten Mammalier zeigen untereinander mehr als 80 % Identität, wobei die höchste Übereinstimmung mit 90 % zwischen Maus und Ratte vorliegt. Diese Werte entsprechen den bekannten verwandtschaftlichen Beziehungen.

Als ursprünglichste Art der hier betrachteten Wirbeltiere zeigt der Frosch als Mitglied der Amphibien die größten Abweichungen gegenüber den anderen Spezies. Amphibien trennten sich vor 350-400 Millionen Jahren von den Reptilien (Evolution der Tetrapoden aus den Quastenflossern im Devon²⁰⁵). Die restlichen Arten gehören zu den Amniota. Neben der Sequenz des Frosches zeigt auch die HCII-Sequenz des Huhns als Mitglied der Vögel/Reptilien (Sauropsiden) weniger Übereinstimmungen zu den anderen Arten. Letzte gemeinsame Vorfahren mit den Mammaliern gab es vor etwa 310 Millionen Jahren. Die Mammalier, die hier besprochen werden, trennten sich vor weniger als 91 Millionen Jahren, und Maus und Ratte hatten die letzten gemeinsamen Vorfahren vor ca. 41 Millionen Jahren.²⁰⁹ Untersuchungen zum Verwandtschaftsgrad von Arten und 'Molekulare Uhren' wurden bereits über eine Reihe von Gegenüberstellungen von Nukleinsäure- oder Proteinsequenzen erstellt (Vergleiche von rRNA oder tRNA-Sequenzen, Vergleiche von Aminosäuresequenzen von Proteinen, wie zum Beispiel von Cytochrom C, oder Vergleiche von Genstrukturen, wie z. B. anhand der Serpine²¹⁰). Bei einem Vergleich von sieben verschiedenen Proteinen errechnet sich ein Unterschied in der Nukleotidsequenz von 0,8 Nukleotiden pro 1 Million Jahren nach Artaufzählung.²⁰⁵

Werden die translatierten Bereiche der HCII-cDNA-Sequenzen verglichen, so ergibt sich folgendes: Die Sequenzabweichung der HCII-cDNAs liegt bei 0,05-0,1 % pro 1 Million Jahren (etwa 0,7-1,5 Nukleotidaustausche in der kodierenden Sequenz pro 1 Million Jahren) und entspricht somit den bekannten Daten.

Stellt man die Werte innerhalb der Mammalier den Werten der restlichen Arten gegenüber, so wird eine zwei- bis vierfach höhere Abweichung je Zeiteinheit erhalten (etwa 20 % gegenüber 5-10 % je 100 Millionen Jahren). Eine erhöhte Mutationsrate wurde bereits bei Säugetieren, d. h. vor allem bei Nagetieren, festgestellt.²¹¹

E.3.1.2 Größere Unterschiede zwischen verschiedenen HCII-cDNA-Sequenzen

Neben den bisher besprochenen Unterschieden in den cDNA-Sequenzen, die auf Mutationen einzelner Basen beruhen, gibt es größere Bereiche in einzelnen Sequenzen, die sich völlig von den anderen unterscheiden. Zuerst wird der translatierte Bereich verglichen und anschließend die 5'- und 3'-Enden. Wie dem cDNA-Alignment (Tabelle 29) zu entnehmen ist, gibt es innerhalb der kodierenden Sequenz nur sechs Bereiche (siehe Tabelle 45), in denen Basen im Verlauf der Evolution inseriert bzw. deletiert wurden.

Tabelle 45: Insertionen/Deletionen in den HCII-cDNAs

Nukleotid-Positionen	Aminos.-Positionen	Maus	Ratte	Kaninchen	Mensch	Huhn	Frosch
99-101	-16						- 3 Bp
111-113	-12	- 3 Bp					
131-133	-5	- 3 Bp	- 3 Bp				
173-271	+10 - +42				+ 51 Bp	+ 27 Bp	+ 37 / - 22 Bp
364-366	+73					- 3 Bp	
381-386	+79 - +80				+ 6 Bp		+ 3 Bp

Die Positionsangaben beziehen sich auf die Numerierung des Alignments in Tabelle 29. Ein '+' steht für die Einfügung von Nukleotiden, während '-' für eine Deletion verwendet wurde.

²⁰⁹ Csihak *et al.*, 1992; Graur *et al.*, 1996; Kumar & Hedges, 1998

²¹⁰ Prochnownik *et al.*, 1985

²¹¹ Workshop on Comparative Genome Organization, 1996

Jeweils drei Basen, die je einem Kodon des Signalpeptids entsprechen, sind an den Positionen 99-101 bei Frosch, 111-113 bei Maus und 131-133 bei Maus und Ratte deletiert. Außerhalb des für das Signalpeptid kodierenden Bereichs sind an den Position 364-366 ebenfalls drei Nukleotide beim Huhn deletiert. Diese Positionen entsprechen genau einem Kodon, so daß das Leseraster unverändert erhalten bleibt und somit keine neuen Aminosäuren in das Protein eingebaut werden. Das gleiche trifft für die Position 381-386 zu. Hier sind beim Mensch 6 Bp (AS 79-80) und beim Frosch drei Nukleotide, die einem Kodon entsprechen, inseriert. Erstaunlicherweise gibt es bei diesen entfernt verwandten Arten eine Insertion an der gleichen Stelle. Entweder liegt ein (minimaler) 'hot spot' (eine Position in der DNA mit punktuell erhöhter Mutationsrate) an dieser Stelle vor, oder sie ist für das Protein nicht essentiell wichtig, so daß hier Mutationen, die nicht das Leseraster beeinflussen, fixiert werden.

Die bisher beschriebenen Mutationen führen nur zu geringen Veränderungen der Proteinsequenzen. Eine deutliche Veränderung des aminoterminalen Endes des HCII-Proteins wird dagegen innerhalb der Evolution durch größere Insertionen bzw. Deletionen am Beginn der für HCII-kodierenden Sequenz entstanden. Nach dem in Tabelle 29 durchgeführten Alignment ist es dabei auch zu Rastermutationen gekommen.

Obwohl nach den bisher aus der Literatur bekannten Daten²¹² und dem Vergleich der HCII-Sequenzen (siehe Abbildung 49) das Kaninchen einen ähnlichen bis näheren Verwandtschaftsgrad zum Menschen besitzt als zur Maus oder Ratte, zeigt das Alignment in diesem Bereich (Positionen 173-271) eine starke Übereinstimmung der Nagetiere untereinander. Der Mensch zeigt dagegen teilweise mehr Ähnlichkeit zu ferner verwandten Arten.

Wie entstanden jene Mutationen? Als erstes kann gesagt werden, daß nach dem Modell für die Thrombin-Inhibierung durch HCII der aminoterminal Bereich keine (sehr stark) festgelegte Primärstruktur besitzt. Er lagert sich bei der Interaktion mit der sauren Domäne an Thrombin an. Die für die sauren Aminosäuren (erste saure AS beim Menschen: 53 Position) kodierenden Sequenzen beginnen allerdings stromabwärts des veränderlichen Bereichs (Alignment-positionen 249-251 für die menschliche 53. AS). Im Anschluß an diese Region gibt es nur beim Frosch eine größere Deletion. Somit wird die Proteinfunktion durch die Mutationen nicht beeinflußt, da sich keine (Raster-)mutation bis in die essentiellen Bereiche erstreckt, und es können größere Veränderungen auch in einem kodierenden Abschnitt des HCII-Gens auftreten.

Die Entstehung der großen Insertion beim Mensch (Positionen 173-223; AS 10-26) gegenüber den anderen Mammaliern läßt sich wie folgt erklären: An den Enden (Positionen 163-172 und 214-223) gibt es eine direkte Sequenzwiederholung von 10 Bp (schwarze Pfeile in Tabelle 29; 2 Bp sind nicht identisch), dazwischen liegt größtenteils eine unvollständige invertierte Sequenzwiederholung (11 Bp). Solche Strukturen können entstehen, wenn in einen DNA-Doppelstrangbruch mit überstehenden Enden (in diesem Fall 10 Bp) eine rückgefaltete kleine Haarnadelstruktur einer Nukleinsäure inseriert. Eine Datenbanksuche erbrachte allerdings keine Hinweise über die Herkunft der Sequenz. Überraschend sind die ebenfalls in diesem Bereich liegenden Insertionen bei dem Huhn und dem Frosch. Obwohl Ähnlichkeiten zu dem inserierten Bereich im Menschen bestehen, ist es einfacher von einer zufälligen Ähnlichkeit auszugehen und somit anderen, parallelen Insertionsereignissen im Verlauf der Evolution in Huhn und Frosch. Andernfalls muß davon ausgegangen werden, daß genau jenes Stück bei Maus, Ratte und Kaninchen deletiert wurde. Es besteht die Möglichkeit, daß bei einem gemeinsamen Vorfahr von Mammaliern, Reptilien und Amphibien an der Stelle ein Chromosomenbruch stattfand, der über unterschiedliche DNA-Reparatur in den Nachkommen behoben wurde.

²¹² Graur *et al.*, 1996; Kumar & Hedges, 1998

Die N-terminalen Aminosäuren bis zur Position 80 zeigen keine Homologie zu anderen Serpinen. Deshalb war es bisher nicht möglich, die gesamte 3D-Struktur von HCII anhand ähnlicher Serpine zu modellieren. Da sämtliche Mutationen in diesem Abschnitt liegen, kann davon ausgegangen werden, daß er sich nur wenig auf die (funktionelle) 3D-Struktur des übrigen Proteins auswirkt.

Die Unterschiede in den hier verglichenen, nicht translatierten 5'-HCII-cDNA-Enden sind vergleichsweise gering (bis auf die bereits erwähnte verschobenene Exon/Introngrenze bei Mammaliern). Die niedrigen Sequenzunterschiede dieses Bereiches zeigen allerdings kein reales Bild, da Teile der 5'-Sequenzen von Frosch und Huhn nicht bekannt sind.

Der 3'-gelegene nicht translatierte Abschnitt der cDNAs zeigt vielerlei Insertionen bzw. Deletionen von 1-37 Bp. Es ist genau zu erkennen, daß hier keine Selektion auf translatierte Kodons stattfindet. Zwei Mutationen sind besonders bemerkenswert: Als Zeichen dafür, wie sehr sich der Selektionsdruck in translatierten und nicht translatierten Bereichen selbst innerhalb der transkribierten DNA-Sequenz unterscheidet, kann die Insertion von zwei Nukleotiden im Kaninchen bzw. die Deletion von 1 Bp im Huhn direkt hinter dem jeweiligen Stopkodon angesehen werden. Die zweite Mutation umfaßt die Insertion eines Fragments eines MER105 Elements in die 3' nicht translatierte Region der cDNA. Dieses Element wird nur in der menschlichen Sequenz vom Programm 'RepeatMasker2' erkannt. Wie allerdings gut zu erkennen ist, kommen mutierte Varianten auch in den anderen Mammaliern vor. MER-Elemente ('medium reiteration frequency sequence') sind wiederholte Sequenzen, die in allen Säugetieren vorkommen können.²¹³ Sie umfassen teilweise noch nicht genauer charakterisierte Wiederholungen mittlerer Häufigkeit im Genom. Die Abstammung der Elemente und ihre eventuelle Funktionen sind bisher nur in einigen Fällen bekannt (z. B.: Zuordnung von MER42 zu L1-Fragmenten, oder MER15/18 zu den LTRs von THE1, einem MaLR-Element).²¹³ Das MER105-Element besitzt 24 Bp LTRs, aber keine verdoppelten Zielstellen (die bekannte Sequenz der LTRs ist 'CAGCCTTTCTCAACCGGGGTTCC'; der rote Pfeil in Tabelle 29 zeigt die entsprechende, etwas abweichende Sequenz in der hHCII-cDNA). Die einzelnen Kopien sind zu etwa 19 % unterschiedlich und wurden vor der Säugetieraufteilung ins Genom inseriert.²¹⁴ Die Insertion eines repetitiven Elements in dem nicht translatierten Abschnitt einer Exonregion der mRNA ist nicht ungewöhnlich. Viele mRNAs enthalten Insertionen von (Fragmenten) von *Alu*-ähnlichen Sequenzen im 5'- oder 3'-Bereich.

Zusammenfassend kann gesagt werden, daß mit Ausnahme des in das 3'-gelegene nicht translatierte Ende inserierte MER-Element im Vergleich zum Intron 1 keine Insertionen von repetitiven Elementen oder Dinukleotidwiederholungen in der HCII-cDNA vorkommen. Durch Selektion auf Individuen mit einem aktiven HCII-Protein scheinen derartige Mutationen in den Exonbereichen nahezu ausgeschlossen zu sein.

²¹³ Smit, 1993

²¹⁴ Arian Smit, persönliche Mitteilung

E.3.2 Vergleich der HCII-Exon/Introngrenzen

Dem Alignment (Tabelle 29) kann entnommen werden, daß die internen Exon/Introngrenzen zwischen den einzelnen Arten identisch sind. Die Exon/Introngrenzen von Kaninchen, Frosch und Huhn werden hier nur postuliert aufgrund der bisher untersuchten, sehr konservierten Exon/Introngrenzen von Ratte, Maus und Mensch. Somit läßt sich zusammenfassen, daß die interne Genstruktur der translatierten Region des HCII-Gens verschiedener Arten stark konserviert ist. Dementsprechend zeigt nur der 5'-Bereich des HCII-Gens (also der Intron 1 Bereich) im Verlauf der Evolution deutliche Veränderungen der Genstruktur. Die unterschiedliche Länge der HCII-cDNAs ist (die noch unbekanntes 5'-Enden von Huhn und Frosch nicht einbezogen) hauptsächlich über die Insertion im 5'-translatierten Bereich und vor allem die Insertion des MER105-Elements in den 3'-Bereich der cDNA in den Mammaliern begründet.

E.3.3 Evolution des Heparinkofaktor II: Vergleich von Intron 1 und cDNA

Bei einem Vergleich der Veränderungen im HCII-Gen von Intron 1 und der cDNA-Sequenz läßt sich folgendes feststellen (Tabelle 42; Tabelle 43): Die nicht translatierten Bereiche der cDNA und die Intron 1-Sequenzen weisen bei Maus, Ratte und Mensch in etwa die gleiche Häufigkeit von Mutationen bzw. die gleiche Sequenzidentität auf (Maus-Ratte, Intron 1: 68 %, cDNA: 69 %; Maus-Mensch, Intron 1: 40 %, cDNA: 46 %; Ratte-Mensch, Intron 1: 57 %, cDNA 48 %). Das Intron 1 enthält also keinen Abschnitt, der im Vergleich zum restlichen Gen (bzgl. cDNA) verstärkt für Mutationen anfällig ist. Daher ergeben sich keine Hinweise auf einen speziellen Mechanismus zur schnellen Veränderung der Gensequenz.

Der translatierte Bereich des HCII-Gens ist wesentlich stärker konserviert. Die geringste Übereinstimmung der bekannten HCII-Sequenzen liegt zwischen Ratte und Frosch mit etwa 61 % vor. Die Mutationsrate innerhalb der Mammalier scheint gegenüber anderen Tiergruppen höher zu liegen. Daher lassen sich mittels PCR HCII-Gene eventuell auch bei entfernten Tierarten, z. B. Fischen, nachweisen. Die Primer müßten in diesem Fall an Positionen gelegt werden, die starke Konservierung aufweisen (Tabelle 29; z. B.: Positionen 283-299, 549-577, 967-978, 1211-1252, etc.). Wenn von etwa 10 % Basenaustausch in der kodierenden Sequenz je 100 Millionen Jahre ausgegangen wird, dann sollte die Übereinstimmung der translatierten HCII-Sequenz von Arten, die sich vor etwa einer halben Milliarde Jahren getrennt haben, immerhin bei etwa 50 % liegen. Konservierte Bereiche, wie z. B. der Bereich des reaktiven Zentrums oder der Exosites, können dabei auch wesentlich höhere Ähnlichkeiten aufweisen. In bestimmten Fällen zeigt jedoch ebenfalls das Reaktionszentrum eine verstärkte Mutationsrate,²¹⁵ so daß nicht unbedingt von essentiellen Bereichen des Proteins zurück auf starke Konservierung geschlossen werden kann. Ein Vergleich mit dem in Tabelle 29 gezeigten Alignment ist notwendig.

Eine Richtung der Evolution, wie z. B. vermehrte spezielle Mutationen, ist nicht festzustellen. Eine Ausnahme bilden repetitive Elemente, deren Insertion häufiger als die Excision dieser Sequenz ist. Diese Elemente können, wie auch am HCII-Gen erkennbar ist, zu einer Verlängerung von Genomabschnitten²¹⁶ (und ebenso der Gene) im Verlauf der Evolution führen und auch deren Expression beeinflussen.²¹⁷ Eine Analyse bezüglich des hHCII-Gens und durch inserierte repetitive Elemente geschaffene Transkriptionsstartpunkte wird in E.4 und E.5 durchgeführt.

²¹⁵ Hill & Hastie, 1987

²¹⁶ Kass *et al.*, 1997

²¹⁷ Britten, 1997

E.4 Transkriptionsstartpunkte des menschlichen Heparinkofaktor II-Gens

In diesem Teil werden die für die unterschiedlichen menschlichen HCII-mRNAs ermittelten Transkriptionsstartpunkte beschrieben, um die Analyse der HCII-Genstruktur bei Maus, Ratte und Mensch zu erweitern. Im folgenden wird zuerst der Startbereich der seit längerem bekannten 2,3 kB-HCII-mRNAs eingegrenzt. Danach wird auf die längeren Transkripte eingegangen. Im Teil E.5 werden anschließend Vorhersagen getroffen zu den eventuell wirksamen Promotorelementen der verschiedenen Transkripte.

HCII-mRNAs kommen in unterschiedlicher Konzentration in unterschiedlichen Geweben vor (Abbildung 29), somit unterliegt die Transkription gewebsspezifischer Kontrolle (siehe E.4.3). Ein quantitativer Vergleich der beiden unterschiedlichen Typen von HCII-Transkripten in den einzelnen Geweben wurde nicht durchgeführt. Erst zum Schluß dieser Arbeit lagen Ergebnisse vor, nach denen beide Transkripte in allen getesteten Geweben vorliegen (unveröffentlichte Daten von U. Schmidt). Die Konzentration der langen Transkripte im Lebergewebe scheint allerdings um mehrere Größenordnungen niedriger zu sein als diejenige der 2,3 kB mRNAs. Der Nachweis der kürzeren Transkripte in anderen Geweben als Leber gestaltete sich ebenfalls schwierig.

E.4.1 Startbereich der 2,3 kB-Transkripte

Die Analyse des Transkriptionsstarts der in der Leber weit überwiegenden 2,3 kb HCII-mRNAs mittels T-PCR-Methodik und Sequenzierung bestätigt, daß die Transkripte nicht sämtlich an einem einzigen Punkt beginnen. Stattdessen kann ein Startbereich eingegrenzt werden, der sich über einen Bereich von mindestens 46 Nukleotiden erstreckt (entsprechend Exon 1-Längen von 23 Bp bis 68 Bp).

Anhand der Sequenziermembran läßt sich sehr gut die 3'-gelegene Position ablesen, die das Ende des Startbereichs markiert (Abbildung 32 in D.8.2.2). Diese Position befindet sich innerhalb der bereits veröffentlichten, aus einem cDNA-Klon abgeleiteten, Exon 1-Sequenz.²¹⁸ Das Ende wird durch eine Reihe von Banden in der 'G'-Spur angezeigt, die 5'-Enden von mRNAs entsprechen. Die komplementäre Sinnsequenz zu dieser Position ist 'CGCCACTTCT' (vergleiche Skizze zur Abbildung 32). Damit ist die Position der ersten Base in der Sequenz an Position +25. Dieser Startpunkt liegt 2 Bp stromabwärts zu der am weitesten 3'-gelegenen Stelle, die mit Nuklease-Kartierung erhalten wurde.²¹⁸ Die Sequenz kann darüber hinaus bis zu der (revers notiert) Sinnsequenz 'ANATCAAAAT' verfolgt werden. Das erste Nukleotid entspricht Position -21 nach Herzog *et al.*,²¹⁸ also 21 Bp vor dem Ende des publizierten cDNA-Klons. Diese Position fällt mit einem potentiellen Startpunkt zusammen, der über Nuklease-Kartierung ermittelt wurde (siehe Skizze zur Abbildung 32).²¹⁸ Die Überprüfung der Startpunkte wurde jedoch vor allem durchgeführt, weil bezüglich der Nuklease-Kartierung und dem am weitesten 5'-gelegenen geschützten Bereich unterschiedliche Ergebnisse mit S1- und RNaseA-Methode erhalten wurden.²¹⁸ Eine endgültige Aussage wurde aber auch mit den hier verwendeten Methoden nicht erhalten. Der Sequenziermembran (Abbildung 32) kann entnommen werden, daß es längere HCII-Transkripte ('G'-Bandenleiter deutet Transkripte an, die vor Position -21 beginnen) geben muß, ohne daß sie eindeutig identifiziert werden können. Dieses Ergebnis spricht für einen nach stromaufwärts verlängerten Startbereich. Außerdem wird das Ergebnis der S1-Kartierung durch die erhaltenen A-PCR- und T-PCR-Produkte bestätigt. Nach Auftrennung der PCR-Produkte in Agarosegelen wird jeweils ein kürzeres Produkt erhalten (Hauptbande) und sehr viel schwächer eine Nebenbande, die ein um etwa 70-100 Bp längeres Produkt zeigt (Tabelle 30, Tabelle 31).

²¹⁸ Ragg & Preibisch, 1988

Dieses Produkt entspricht einer Länge des Exon 1 von etwa 120-130 Bp. Damit stimmt das Ergebnis gut mit der ermittelten 5'-Position der S1-Kartierung überein (berechnete Exon 1-Länge: 116 Bp; vergleiche Skizze zur Abbildung 32).

Die Analysen zeigen einen Startbereich für die 2,3 kB-HCII-mRNAs anstatt eines einzigen Startpunktes. Die möglichen Gründe für das Fehlen eines definierten Startpunktes sind im folgenden aufgeführt. Zum einen kann eventuell von artifiziellen Ergebnissen ausgegangen werden. Ein Grund für die mit Hilfe von A-PCR und T-PCR ermittelten unterschiedlichen cDNA-Enden können Sekundärstrukturen oder Sequenzen in der mRNA sein, an denen die cDNA-Synthese abbricht. Sämtliche Analysemethoden können ähnliche falsche Startpunkte ergeben, falls es im 5'-Bereich der mRNA bevorzugte Spaltstellen von Endonukleasen gibt. Außerdem können die 5'-Enden der mRNAs abhängig von der Isolierungsmethode durch Exonukleasen teilweise abgebaut sein. Es muß jedoch davon ausgegangen werden, daß die ermittelten cDNA-Enden den Startpunkten der HCII-mRNAs entsprechen, da der Vergleich der Ergebnisse von vier unterschiedlichen Methoden eine gute Übereinstimmung für den Transkriptionsstartbereich liefert. Damit verbleibt die Frage, warum die HCII-mRNA im Gegensatz zu vielen anderen mRNAs keinen definierten Startpunkt aufweist. Der Grund könnte darin liegen, daß eine TATA-Box als Startsignal fehlt. Eine mögliche TATA-Box, die einen wahrscheinlichen mRNA-Start stromabwärts anzeigen würde, liegt nach Herzog *et al.*²¹⁸ bei Position -67. Für die beiden beschriebenen Startpunkte ist diese Position allerdings zu weit entfernt, da TATA-Boxen etwa 30 Bp vor dem mRNA-5'-Ende liegen. Andere Startsignale liefern z. B. INR-Elemente. Mögliche Transkriptionssignale werden unter E.4.3 diskutiert.

E.4.2 Startbereich der langen Transkripte (Exon 1*)

Von Schmidt²¹⁹ wurden in einer cDNA-Bank (von Clontech) von menschlichem Lungengewebe HCII-Transkripte entdeckt, die im Gegensatz zu den zuvor aus Leber bekannten Transkripten offensichtlich nicht das Exon 1 enthalten. Der Startbereich liegt in dem bisher als Intron 1 bezeichneten Bereich. Somit ist die zuvor als Exon 2 bezeichnete Sequenz in diesen Transkripten das erste Exon und am 5'-Ende verlängert. Dieses Exon wird in dieser Arbeit als Exon 1* bezeichnet.

Mittels A-PCR und T-PCR wurden von RNAs aus A549-Zellen (einer Lungen-Zelllinie) und von poly(A)⁺-RNA aus menschlicher Lunge der Startbereich mit unterschiedlichen Antisinnprimern bestimmt. Im weiteren Verlauf der Untersuchungen konnten diese Transkripte ebenfalls in Lebergewebe festgestellt und deren Startbereich eingegrenzt werden. Auch bei diesen Transkripten kann kein einheitlicher, definierter Startpunkt ermittelt werden.

Je weiter die Antisinnprimer zur Analyse des 5'-Endes zum 5'-Bereich des Introns verschoben werden, desto längere Transkripte werden gefunden. Bei Beginn der Experimente erschienen diese Ergebnisse sehr unwahrscheinlich. Kontrollexperimente zeigten, daß einige der verwendeten RNAs und mRNAs minimale Reste von chromosomalen DNA-Fragmenten enthalten, die in PCR-Reaktionen amplifiziert werden können und falsche Ergebnisse liefern. Aus diesem Grund wurden bei den A- und T-PCR-Experimenten, auf denen die Daten in dieser Arbeit beruhen, erweiterte Negativkontrollen durchgeführt. Neben den einfachen Negativkontrollen, bei denen die RT-PCR von Beginn an ohne Template angesetzt wurde, wurden Ansätze parallel durchgeführt, in denen als einziger Unterschied zu den Positivansätzen die reverse Transkriptase fehlte oder durch Hitze inaktiviert wurde. Da die Kontrollen keine definierten PCR-Amplifikate zeigten, können chromosomale Verunreinigungen der RNA-Präparationen, die zu falschen, positiven Ergebnissen hätten führen können, ausgeschlossen werden.

²¹⁹ Schmidt, 1997

Darüber hinaus wurde eine RT-PCR des langen Transkripts durchgeführt, in der ein Fragment, das ein Stück von Intron 1 enthält (also Exon 1*) und bis ins Exon 3 reicht, amplifiziert wurde (Abbildung 27). Falls man davon ausgeht, daß aus der prä-mRNA das Intron 2 heraus-geschritten wird, dann zeigt das Fragment die erwartete Größe. Diese RT-PCR wurde neben poly(A)⁺-RNA aus Leber- und Lungengewebe auch parallel mit RNA aus Herz, Niere, Plazenta und Gehirn durchgeführt. Das Ergebnis belegt eindeutig, daß ein verändertes HCII-Transkript in allen Gewebetypen vorkommt. Zur Sicherheit wurden die erhaltenen PCR-Produkte von Leber und Herz mit Restriktionsenzymen gespalten. Teilweise sind aufgrund von nur partieller Spaltung schwach die ungespaltenen Ausgangsfragmente zu erkennen. Die Spaltprodukte selbst zeigen die errechnete Größe (Abbildung 28). Somit läßt sich sagen, daß lange HCII-Transkripte in den Geweben vorliegen und Teile des Intron 1 der 2,3 kB-mRNA enthalten. Außerdem werden durch diese RT-PCR-Ergebnisse die erhaltenen Daten für die 5'-Enden unterstützt. Es kann jedoch nicht ausgeschlossen werden, daß die mit A-PCR und T-PCR erhaltenen cDNAs auf unvollständig prozessierten prä-mRNAs beruhen. Allerdings wurden auch bei anderen Genen alternative Promotorbereiche gefunden, wie z. B. beim Connexin32-, dem Vigilin-Gen oder dem α_1 -Antitrypsin.²²⁰ Für Lungengewebe und -zelllinie wurden Startpunkte in einem Bereich von bis zu 1300 Bp vor dem Exon 2 identifiziert. Zahlreiche definierte Startpunkte konnten über Sequenzierung in dem Bereich festgestellt werden (siehe Abbildung 34 bis Abbildung 36 und zugehörige Skizze der chromosomalen Sequenz). Mit poly(A)⁺-RNA aus Leber wurden PCR-Produkte von 5'-Enden der cDNAs nur bis etwa 1000 Bp vor Exon 2 detektiert. Die Positionen der Transkriptionsstartpunkte sind auch in Übersichtsschemata unter E.4.3 eingetragen (Abbildung 50, Abbildung 51). In beiden Fällen können noch weiter 5'-gelegene Startpunkte nicht ausgeschlossen werden. Um abschätzen zu können, ob eine in der HCII-Intron 1-Sequenz vorhandene potentielle TATA-Box (Position 3566) den 5'-gelegenen Startpunkt des Bereichs definiert, wurde mit Leber- und Lungen-RNA eine RT-PCR durchgeführt. Ein 3'-gelegener Abschnitt zu dieser Position sollte amplifiziert werden, sofern er in den gewebsspezifischen mRNAs vorhanden ist. Das Ergebnis zeigt, daß zwar in beiden Zelltypen eine ganze Reihe von HCII-Startpunkten vorliegen, aber daß mit RNA aus Lungengewebe im Gegensatz zum Leber-Material kein Transkript dieser Länge nachzuweisen ist (D.8.4, Abbildung 38). Somit unterliegt der Transkriptionsstartbereich (zu einem gewissen Maß) gewebsspezifischer Kontrolle. Diese unterschiedliche Verwendung einzelner Startpunkte in dem Bereich in unterschiedlichen Geweben innerhalb eines sehr großen Bereichs von etwa 1500 Bp wird auch von der Gegenüberstellung der erhaltenen 5'-Enden mit dem hier verwendeten, am weitesten 5'-gelegenen Antisinnprimer dokumentiert. Wie in Abbildung 37 zu erkennen ist, scheinen sich in Lunge und Leber die Mengen an Transkripten mit einem Startbereich etwa 1000 Bp vor Exon 2 stark zu unterscheiden.

Zur physiologischen Rolle der langen HCII-Transkripte können folgende Überlegungen gemacht werden: HCII scheint im Gewebe spezielle Funktionen zu besitzen und kann mittels Westernblot in Leber, Lunge, Herz, Niere und Plazenta nachgewiesen werden.²²¹ HCII-Transkripte können jedoch auch im Gehirn gefunden werden, in dem kein (bzw. sehr wenig) Protein nachzuweisen ist. Möglicherweise wird das Protein über die Blutbahn ins Gewebe importiert. Durch die Blut-Hirnschranke läßt sich das Fehlen von HCII im Gehirn erklären. Von den langen Transkripten wird eventuell aus zwei Gründen (so gut wie) kein HCII-Protein translatiert. Zum einen ist die Anzahl aller HCII-Transkripte in anderen Geweben als Leber anscheinend um etwa den Faktor 500 niedriger (siehe D.7, Abbildung 29; E.4.3) und zum anderen ist ein sehr langer 5'-nicht translatierter Bereich vorhanden.

²²⁰ Duga *et al.*, 1999; Kügler *et al.*, 1996; Perlino *et al.*, 1987

²²¹ A. Strathmann, 1999, unveröffentlichte Ergebnisse

Normalerweise ist der 5'-UTR von mRNAs nur kurz. Da die Cap-Struktur der mRNA die Translation beeinflusst,²²² ist eine effiziente Translation dieser Transkripte unwahrscheinlich. Das ATG-Startkodon liegt bei einigen Transkripten mehr als 1 kB stromabwärts. Somit haben diese Transkripte wahrscheinlich keine physiologische Bedeutung, aber bilden eventuell den evolutiven Übergang (über verstärkte Transkription und Bildung einer neuen kodierenden Sequenz mit einem weiter 5'-gelegenen Start-ATG) zu einem neuen, am 5'-Ende verlängerten Protein.

E.4.3 Promotoranalyse

Eine experimentelle Überprüfung auf einen funktionellen Promotorbereich vor dem menschlichen HCII-Gen wurde nur für eine 2,2 kB große Sequenz vor dem Exon 1 mit Hilfe eines Reportergenkonstrukts durchgeführt (siehe D.9.3). Dieser Test ermöglicht die im folgenden besprochenen Aussagen über einen potentiellen transkriptionsaktivierenden Bereich und eventuelle Bindungsstellen von Transkriptionsfaktoren für die seit längerem bekannten 2,3 kB-mRNA-Transkripte (E.4.3.1).

Für die 2,3 kB- und die längeren HCII-Transkripte werden im folgenden Vorschläge für mögliche Sequenzbereiche gegeben, die eine Promotorfunktion besitzen könnten. Es wurde bei beiden Transkriptgruppen ein teilweise umfangreicher Startbereich gefunden.

Die bekanntesten proximalen Elemente, die einen Transkriptionsstartpunkt etwa 30 Bp stromabwärts festlegen können, sind die TATA-Boxen. Die hier analysierten Transkripte scheinen allerdings größtenteils (siehe E.4.3.1 und E.4.3.2) nicht durch TATA-Boxen festgelegt zu sein. Als zur TATA-Box funktionell homologe Sequenzen sind INR-Elemente bekannt. Smale führt an, daß es TATA-Box-freie Gene mit Transkripten gibt, deren Dutzende von Startpunkten sich über einen Bereich von mehreren 100 Bp erstrecken können.²²³

Bisher ist zu den INR-Elementen bekannt, daß sie eine schwach konservierte Konsensussequenz enthalten, die den direkten Transkriptionsstartpunkt umfassen (beim Terminalen-Transferase-Gen der Maus: Position -6 bis +11) und Bindungsstellen für transkriptionsaktivierende Faktoren in der Nähe benötigen (wie G/C-Boxen oder von SP1 erkannte Stellen). Die schwach konservierte Konsensussequenz²²³ ist in der folgenden Skizze dargestellt und die ermittelten Transkriptionsstartpunkte werden mit dieser Sequenz in den Punkten E.4.3.1 und E.4.3.2 verglichen.

-3	-2	-1	+1	+2	+3	+4	+5	+6
y	Y	<u>Y</u>	<u>A</u>	n	<u>V</u>	Y	Y	y

Kleine Buchstaben stehen für nur schwach konservierte Positionen der einzelnen bekannten INR-Elemente. Fette und unterstrichenen Buchstaben stellen stark konservierte Basen dar und ein 'V' steht für ein 'T'- oder 'A'-Nukleotid.

Bezüglich einer gewebsspezifischen Expression läßt sich eine grobe Abschätzung über die relative Menge von HCII-Transkripten in einzelnen Organen über den Vergleich von RT-PCR-Produkten machen, die unter den gleichen Bedingungen amplifiziert wurden. In Abbildung 29 sind über RT-PCR erhaltene HCII-cDNA-Fragmente nach unterschiedlicher Anzahl von PCR-Zyklen zu sehen. Die einzelnen Gele unterscheiden sich jeweils durch drei zusätzliche Zyklen, also rein theoretisch einem Amplifikationsfaktor von bis zu acht (2³). Gut zu erkennen ist die starke HCII-Transkription in Leberzellen, da hier nach 25 PCR-Zyklen nahezu die maximal mögliche Zahl an Amplifikaten in der PCR-Reaktion vorhanden ist.

²²² Darnell *et al.*, 1990

²²³ Smale, 1997

Die gleiche Bandenstärke wird in Lunge und Gehirn, den Geweben mit der offensichtlich nächst stärkeren Transkription, erst nach etwa weiteren neun Zyklen erreicht (unteres Gelbild). Das entspricht rechnerisch einer Anfangskonzentration an HCII-Transkripten, die um den Faktor ≈ 500 (2^9) geringer sein muß als in Leberzellen.

Mit der durchgeführten RT-PCR kann nicht zwischen den beiden verschiedenen HCII-mRNAs unterschieden werden. Beide Produkte lassen sich mit der verwendeten Primerkombination amplifizieren. Daher ist eine Aussage über die relative quantitative Verteilung der beiden Transkripttypen in den einzelnen Geweben nicht möglich.

E.4.3.1 Promotorregion der 2,3 kB-Transkripte

In dieser Arbeit wurde ein Transkriptionsstartbereich für die 2,3 kB mRNAs von HCII ermittelt. Die 3'-Position des Bereichs liegt bei Position -21, die stromaufwärts gelegene Position bei +25 (Positionsangaben nach Herzog *et al.*²¹⁸). Welche regulatorischen Elemente sind für diese Transkripte verantwortlich?

Mit Hilfe eines Reportergens wurde eine 2,2 kB große Sequenz getestet, die stromaufwärts vom Exon 1 liegt. In einer Leberzelllinie (HepG2), in welcher der potentielle Promotor der 2,3 kB-Transkripte aktiv sein sollte, konnte keine Reporterogenaktivität gemessen werden (siehe D.9.3, Abbildung 40). Schwankungen zwischen den Ansätzen lassen sich eventuell durch unterschiedliche Qualität der DNA-Isolationen und durch unterschiedliche Plasmidgröße erklären. Ein großes Plasmid läßt sich schlechter transfizieren als kleinere. Als direkte Negativkontrollen wären also Plasmide mit den gleichen Reporter genen und gleicher Größe ideal. Allerdings ist nach mehreren Versuchen (siehe D.9.3, Tabelle 40) eindeutig zu erkennen, daß der getestete Bereich keine starke Transkription des HCII-Gens in HepG2-Zellen bewirken kann. Entweder fehlen wichtige regulatorische Elemente, oder die Zellen unterscheiden sich von normalen Hepatozyten in ihrem Expressionsmuster von Transkriptionsfaktoren deutlich. Als Folge könnte ein möglicher Silencer in der Sequenz aktiviert werden, der die Transkription des HCII-Gens in anderen Geweben als Leber unterdrückt.

Die Frage stellt sich, an welchen Positionen diese Elemente vorhanden sein können. Die wahrscheinlichste Antwort ist, daß der klonierte Bereich zu klein ausgefallen ist. Im Anschluß an diese Arbeit sollen größere Bereiche vor dem Exon 1 kloniert und überprüft werden. Weiterhin können im gesamten Exon 1, wie z. B. beim Angiotensinogen-Gen²²⁴ (es wurde nur ein Teil des Exons kloniert!) oder auch im Intron 1 wichtige Signale kodiert sein. Darüber hinaus ist vom α_1 -Antitrypsin-Gen bekannt (einem Serpin aus der gleichen Familie, dem auch HCII angehört), daß wichtige Bereiche (Enhancer) nicht nur im 5'-Bereich des Gens, sondern auch stromabwärts des kodierten Bereichs gelegen sind.²²⁵ Die experimentell erhaltenen Daten sind zur Eingrenzung der Promotorregion unzureichend. Sie wurden mit Hilfe von Datenbanken auf mögliche Bindungsstellen von Transkriptionsfaktoren hin durchsucht.

Vor dem HCII-Gen gibt es die nächste TATA-Box etwa 1300 Bp vor Exon 1 (Tabelle 35, Position 4613), also weit außerhalb des funktionellen Bereichs (≈ 30 Bp vor dem mRNA-Beginn). Herzog *et al.* schlagen eine TATA-ähnliche Sequenz bei Position -67 vor.¹⁸⁷ Diese Sequenz zeigte aber bei den hier durchgeführten Untersuchungen keine Übereinstimmung mit der Matrix für TATA-Boxen. Falls diese Sequenz in den Zellen als TATA-Box erkannt wird, so liegt sie allerdings auch für das 5'-gelegene Ende des Startbereichs um etwa 10 Bp zu weit stromaufwärts (s.u., gelb hinterlegt). Eine Überprüfung auf mögliche INR-Elemente erbrachte dagegen ein positives Ergebnis.

²²⁴ Cui *et al.*, 1998

²²⁵ Morgan & Kalsheker, 1997

Obwohl die bekannten Kriterien für INRs nur sehr schwach definiert sind, gibt es die einzigen Übereinstimmungen im Exon 1 jeweils an den beiden ermittelten Endpositionen. Nachfolgend ist zum Vergleich die chromosomale Sequenz mit unterlegten INR-Matrizen dargestellt.

```

- 70  5'  GACTTATTTA CTTTGGAAAA TATGCAGCAA CAGCCAGCA
- 30    CCAAAGTTCA CATCAAATC CCACTGATGA CCTTGGCTGC
        yYYA nVYYy
+ 11    TTTCATCTCT GAAGCGCCAC TTCTCAGAAA CACAGAGGTA 3'
        yYYAn VYYy

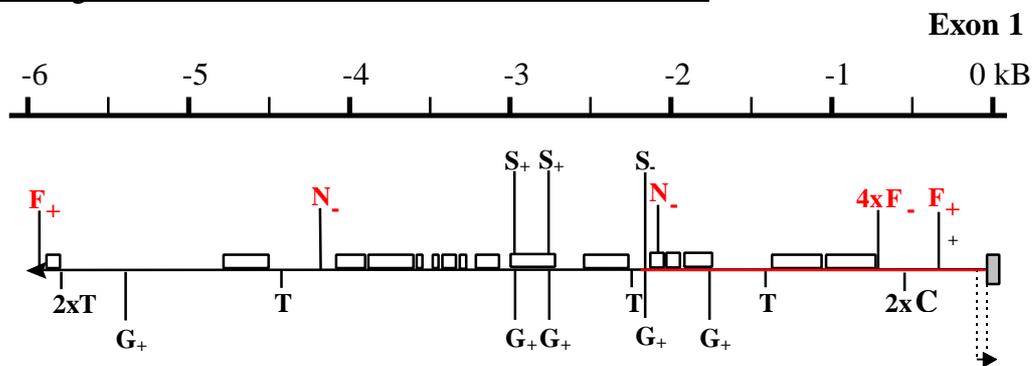
```

In dem Sequenzausschnitt sind die ermittelten 5'- und 3'-Grenzen des Transkriptionsstartbereiches eingetragen (rot), sowie die von Ragg und Preibisch²¹⁸ gefundenen, vor Nukleaseverdau geschützten Bereiche (unterstrichen). Unter der Sequenz sind die INR-Matrizen eingetragen. Kleine Buchstaben innerhalb der Matrize stehen für nur geringe Übereinstimmung der einzelnen bekannten INR-Elemente, während fette Buchstaben hohe Konservierung bedeuten. Ein 'V' steht für ein 'T'- oder 'A'-Nukleotid. Übereinstimmungen mit der INR-Konsensussequenz sind blau hinterlegt.

Die äußeren Startpunkte können aus drei Gründen als gesichert angesehen werden. Erstens stimmen nur die markierten Positionen in diesem Bereich mit der Matrix für INRs überein. Zweitens ist eine sehr hohe Übereinstimmung zu erkennen und drittens erhält man mit mehreren unterschiedlichen Methoden gleiche Ergebnisse.

INR-Elemente benötigen zur Funktion transkriptionsaktivierende Faktoren. In der Literatur sind G/C-Boxen und SP1-Elemente beschrieben.²²³ Daher wurde nach Bindungsstellen für diese Faktoren und für leberspezifische Faktoren gesucht. In der Abbildung 50 sind einige potentielle Bindungsstellen für Transkriptionsfaktoren innerhalb der mutmaßlichen HCII-Promotorregion in ein Schema eingefügt.

Abbildung 50: Schema des Bereichs vor dem HCII-Exon 1



Legende

—	Chromosomale HCII-DNA-Sequenz	T	mögliche TATA-Box
■	Exonbereich	G	mögliche GC-Box
□	Repetitives Element	C	mögliche CEBP-Bindungsstelle
⋮	Ermittelter Transkriptionsstartbereich	F	mögliche HFH1-Bindungsstelle
+	Matrix in Sinnsequenz	N	mögliche HNF1-Bindungsstelle
-	Matrix in Antisinnsequenz	S	mögliche SP1-Bindungsstelle

Erklärungen zu dem Schema folgen auf der nächsten Seite

Erklärungen zur Abbildung 50

In der Abbildung ist der Transkriptionsstartbereich, der mit dem Primer USC8 ermittelt wurde, als Pfeil eingetragen. Die senkrechten gestrichelten Linien markieren die Positionen, an denen nach Sequenzierdaten mRNA-Startpunkte vorhanden sind. Die rote Linie zeigt den vor das Luziferase-Reportergen klonierten Sequenzbereich, der gegenüber Negativkontrollen keine Transkriptionssteigerung bewirkt. Bindungsstellen für Faktoren, die vor allem in Lebergewebe eine Transkription bewirken (CEBP, HFH1 und HNF1) sind in Rot eingezeichnet. Von den CEBP-Matrizen wurde nur die Matrizenübereinstimmungen mit 'CEBP_C' eingetragen, da diese Matrix am besten definiert ist.

Von dem Programm 'MatInspector' wurden neben den eingetragenen Bindungsstellen sehr viele weitere mögliche Bindungsstellen von allgemein die Transkription positiv beeinflussenden Faktoren in der Sequenz gefunden (Tabelle 36: z. B.: AP1-4, HFH1-8, HNF1-3, Octamer1-7). Die Anzahl der möglichen Bindungsstellen dieser Faktoren liegt teilweise sehr stark über der erwarteten Zahl für eine zufällige Sequenz (siehe Re-Wert in der Tabelle). Daher ist zu erwarten, daß diese Faktoren ebenfalls für den Promotor wichtig sind. Für einige Serpine ist die Bedeutung der Faktoren für die transkriptionelle Regulation gezeigt (uPA: AP1,²²⁶ 14q32.1-Serpinkluster: HNF1 + HNF4,²²⁷ α_1 -Antitrypsin: SP1,²²⁸ Angiotensinogen: AP4 + HNF3²²⁹). Außerdem sind Bindungsstellen von Faktoren vorhanden, die unter bestimmten Bedingungen eine Transkription induzieren könnten (CHOP, IRF, RORE). Da die Anzahl der Bindungsstellen für diese Faktoren zu hoch für eine Darstellung ist, wurden sie nicht in die Abbildung eingetragen.

Der Abbildung 50 (und Tabelle 35) ist zu entnehmen, daß eine Reihe von potentiellen G/C-Boxen und SP1-Elemente vorhanden sind. Allerdings liegen diese 1,5 bis 3 kB stromaufwärts des Transkriptionsstartbereiches. Es wurden jedoch auch schon andere essentielle Promotorelemente beschrieben, die mehrere kB stromaufwärts vor dem Startbereich gelegen sind.²²⁶ Die Häufung der Bindungsstellen in dem Bereich von -3 kB bis -1,5 kB liegt über dem Wert für zufällige Übereinstimmung (nach dem 'Re-Wert', Tabelle 36). Damit ist der Bereich potentiell für den HCII-Promotor wichtig. Die Bindungsstellen werden von inserierten SINEs und LINEs zu Verfügung gestellt.

Es bleibt die Frage bestehen, warum genau die beschriebenen INR-Elemente verwendet werden. Es ist wahrscheinlich, daß es weitere wichtige Bindungsstellen für Faktoren gibt, die näher am Transkriptionsstart liegen, um diesen zu definieren. Welche Faktoren können in Frage kommen? Die 2,3 kB-Transkripte werden sehr stark in Leberzellen transkribiert. Es werden also wahrscheinlich leberspezifische Faktoren für die Transkription benötigt. Aus diesem Grund wurde die Sequenz auf bekannte leberspezifische Faktoren hin untersucht.

In der Abbildung 50 ist zu erkennen, daß es eine starke Konzentration von HFH1 und CEBP-Bindungsstellen in dem Bereich von -300 bis -800 gibt. Beide Matrizen sind gut definiert, und eine HFH1-Stelle sollte nur ca. 0,7mal rein zufällig in den 6 kB vorkommen und die CEBP-C-Matrix nur 1,6mal (nach angegebenen Re-Wert, Tabelle 36). Die sehr hohe Dichte der spezifischen Bindungsstellen macht also diese Region als zweites wirksames Promotorelement sehr wahrscheinlich. In diesem Fall sind die Bindungsstellen nicht direkt durch repetitive Elemente bedingt, allerdings wird eine vierfache HFH1-Bindungsstelle über eine kurze wiederholte Sequenz am Ende eines inserierten Elementes gebildet.

²²⁶ Miralles *et al.*, 1999

²²⁷ Rollini & Fournier, 1999

²²⁸ Li *et al.*, 1998

²²⁹ Cui *et al.*, 1998

E.4.3.2 Promotorregion der langen Transkripte mit Start im Intron 1

Der Startbereich für die längeren mRNAs wurde ebenfalls ermittelt. Die Transkripte beginnen in der Intron 1-Region der 2,3 kB-mRNAs und wurden in verschiedenen Geweben gefunden. Die mittels Sequenzierung gefundenen Positionen von 5'-cDNA-Enden (die potentiellen Startpunkten entsprechen) sind 3904, 4198, 4318, 4437, 4517 für Lungengewebe und 4137 für Transkripte aus der Leber (Numerierung nach Herzog *et al.*¹⁸⁷). Den Tabellen (Tabelle 30 - Tabelle 33) ist zu entnehmen, daß wahrscheinlich weitere Startpunkte im 5'-Bereich existieren. Der ermittelte Startbereich ist also noch unvollständig. Im Bereich von Intron 1 gibt es eine Reihe von Positionen mit potentiellen TATA-Boxen (nach MatInspektor, Tabelle 37), möglicherweise sind zwei davon an einer Initiation der Transkription von HCII-mRNAs beteiligt. Die 5'-gelegene Position (Position 3566 in Tabelle 37) kann nur für Transkripte aus Lebergewebe relevant sein, da keine mRNA-Sequenzen in der Lungen-RNA in diesem Bereich nachweisbar waren (Abbildung 38). Falls man die Position der TATA-Boxen mit den Transkriptionsstartpunkten vergleicht, dann beruht ein Transkript aus Lunge sehr wahrscheinlich auf einem TATA-Box-vermittelten Transkriptionsstart. 29 Bp vor dem Start an Position 4437 liegt eine Sequenz, die Übereinstimmung zur TATA-Matrix zeigt (in Tabelle 37 Position 4361, entsprechend 4408 nach der Numerierung von Herzog *et al.*¹⁸⁷). Da dieser Startpunkt nicht vollständig der INR-Konsensussequenz entspricht, ist die Verwendung der TATA-Box als Promotorelement sehr wahrscheinlich.

Die anderen Startpunkte können nicht direkt auf TATA-Boxen zurückgeführt werden. Nach heutigem Wissen müssen demnach INR-Elemente die Startpunkte definieren. Im folgenden ist zum Vergleich die chromosomale Sequenz ausschnittsweise mit unterlegten TATA- und INR-Matrizen dargestellt.

```

3891 5' CACACAGACT TTATCTCTTG TTTCTAACC TGTGTTAGAA
           y YYAnVYYy

3931 GCAGTAACAC AGAAGATTTA GTTGCTCCTG ACAGCAGTGG
3971 GAGCTATTGT CTAAGAGATA CAAAGGAGAA AAAAGTATAC
4011 CTGCAGCAAG TGATATCACC TCTGGGGCTG CCACCACATC
4051 ACCTCACTAC GCCCTGAGGG GGTCTCAGCA CTAGACAAGT
4091 TCCAAATCTT TTGCAAATTA AACAAACCCA GGTCAAGCGT
           yYYA nVYYy

4131 GGTGGCTTAT GCCTGTAATC CCAGCACTTT GGGGGGCTGA
4171 GGTGGGTGGA TCACCTGAGG TCAGGAGTTT GAGACCAGCC
           yYYAnVYY y

4211 TGGCCAACAG AGCAAAACCC CATCTCTACT AAACAAAATA
4251 CAAAAATTAA CCAGGCGTAG TGGTGTGCAC CTGTAGTCCC
4291 AGCTACTTGG GAGGCTGAGG CAAGAGATT GCTTGAGTCC
           yYYAnVYYy yYYAnV Yyy nCT

4331 AGGAGGCCGA AGTTGCAGTA AGCCGAGATC GCGCCACTGC
4371 ACTCCAGCCT GGGTGACAGA GTGAGACTCC ATTTCAAAAA
4411 ATAAAAA CAA CAAAAGCCAA TTACAACAAC AACAAACAAA
           ATAAAAAR y YYAnVYYy

4451 AAACAACGAA TTAAACAACC CCAAAGATTG CACAAATTC
4491 AAGTATCTTT AGAATATGTT TTCAGAAAGC CTGGCCCATG 3'
           yYYAnVYYy

```

In dem Sequenzausschnitt sind die ermittelten 5'-cDNA-Enden (rot) und die INR-Matrizen eingetragen. Kleine Buchstaben innerhalb der Matrize stehen für nur geringe Übereinstimmung mit der Konsensussequenz der einzelnen bekannten INR-Elemente, während fette Buchstaben hohe Konservierung bedeuten. Ein 'V' steht für ein 'T'- oder 'A'-Nukleotid. Übereinstimmungen mit der INR-Matrix sind blau hinterlegt. Eine TATA-Konsensussequenz (TATA_C-Matrize) ist grün markiert. Gelb sind Nukleotide hinterlegt, bei denen die Sequenzierung nicht mit der chromosomalen Sequenz übereinstimmt. In diesem Fall wurde das bekannte chromosomale Nukleotid eingesetzt. Die Nummerierung wurde nach Herzog *et al.* durchgeführt.¹⁸⁷

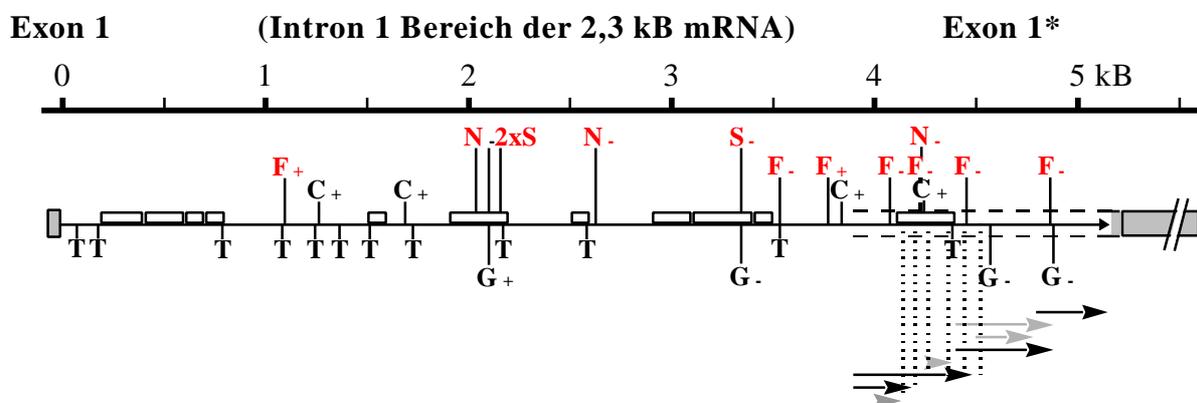
Die meisten Startstellen entsprechen nicht sehr gut der geforderten Matrix für INR-Elemente. Es lassen sich jedoch einige Unterschiede erklären.

Die cDNA-Enden an den Positionen 4137 und 4198 stimmen nicht (oder nur sehr schlecht) mit der INR-Konsensussequenz überein. Wahrscheinlich wurde in diesem Fall das Ende einer trunkeierten mRNA amplifiziert. Eine abgebrochene Sequenz kommt vor allem für die Position 4137 in Frage, da die Sequenz sich nicht bis zum poly(G) lesen läßt (siehe D.8.2.2, Abbildung 33). Die stromaufwärts nächstgelegene potentielle Stelle für ein potentielles INR-Element ist an Position 4108 (s.o.). Der Transkriptionsstart an Position 3904 stimmt mit der INR-Matrix bis auf die dritte Position überein, allerdings ist in der Sequenz das reverse 'T' für das 'A' eines mRNA-5'-Endes nicht zu erkennen. Wahrscheinlich wurde dieses Nukleotid vor dem Anhängen der G-Nukleotide entfernt. Die beiden übrigen sequenzierten 5'-cDNA-Enden (Positionen 4318 und 4517) zeigen teilweise Übereinstimmung zum INR-Element. Jeweils eine der stark konservierten Positionen enthält ein anderes Nukleotid. Wenn davon ausgegangen wird, daß eventuell das 5'-Ende der cDNA mit Start an Position 4517 um zwei Nukleotide verkürzt wurde, ergibt sich eine bessere Übereinstimmung. Für diese Überlegung spricht eine Mutation an der Endposition der cDNA. Eine alternative, stromaufwärts gelegene Position für ein INR-Element, das den natürlichen Transkriptionsstart der mRNA bewirken könnte, auf dem die cDNA mit dem 5'-Ende an Position 4318 beruht, wurde an Position 4292 eingetragen.

In die Abbildung 51 wurden die Positionen der sequenzierten 5'-Enden der cDNAs eingetragen. Zu beachten ist, daß die cDNAs wahrscheinlich auf abgebrochenen mRNAs basieren und die Transkriptionsstartpunkte stromaufwärts gelegen sind. In Zukunft sind Experimente mit S1- oder DNase1-Nuklease notwendig, um die Startpunkte der Transkripte genau festzulegen und die hier erhaltenen Ergebnisse überprüfen zu können. Die hier aufgeführten Transkripte haben wahrscheinlich keine physiologische Rolle (siehe E.4.3).

Welche distalen Promotorelemente vor dem Transkriptionsstart können vorhanden sein? Im Gegensatz zu der 5'-HCII-Genregion gibt es in diesem Bereich keine geklusterten G/C-Boxen oder SP1-Elemente (siehe Abbildung 51). Es existieren jedoch gerade im Transkriptionsstartbereich sehr viele Übereinstimmungen zu Matrizen von HFH1- und CEBP-Bindungsstellen, die eine spezifische Transkription in Leberzellen erklären können. Die in diesem Bereich beginnenden Transkripte können jedoch auch in anderen Geweben nachgewiesen werden, sind also nicht leberspezifisch. Deshalb kann keine Aussage bezüglich potentieller leberspezifische Promotorelemente getroffen werden. In der folgenden Abbildung 51 sind die Ergebnisse bezüglich 5'-Enden der langen mRNAs zusammengefaßt.

Abbildung 51: Startbereich der langen HCII-Transkripte



Erklärungen zur Abbildung : siehe nächste Seite

Legende

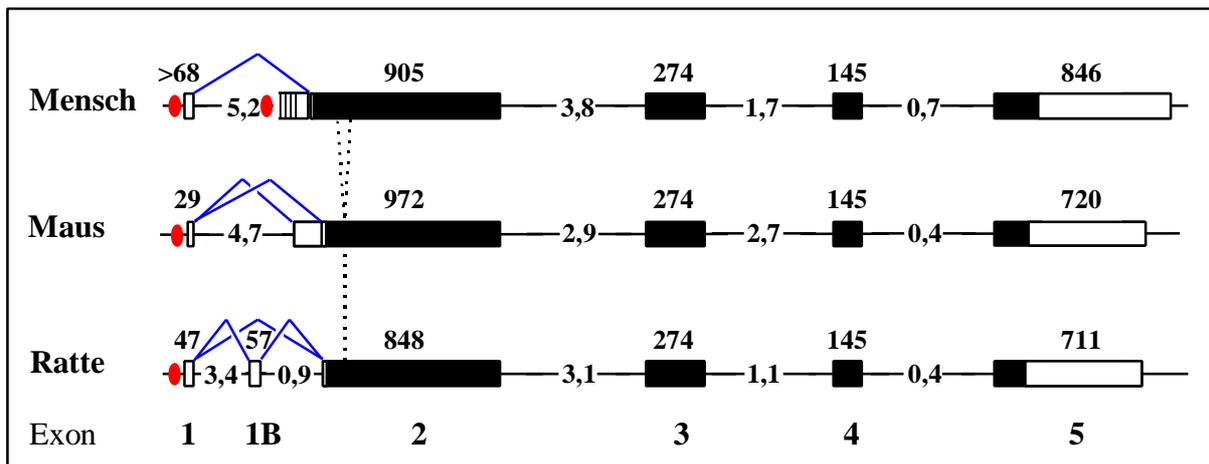
—	Chromosomale HCII-DNA-Sequenz	T	mögliche TATA-Box
■	Exonbereich	G	mögliche GC-Box
□	Repetitives Element	C	mögliche CEBP-Bindungsstelle
→	Ermittelter Transkriptionsstartbereich	F	mögliche HFH1-Bindungsstelle
+	Matrix in Sinnsequenz	N	mögliche HNF1-Bindungsstelle
-	Matrix in Antisinnsequenz	S	mögliche SP1-Bindungsstelle

In der Abbildung 51 sind die mit Hilfe von Gelbildern von A-PCRs und T-PCRs abgeschätzten Transkriptionsstartbereiche als Pfeil eingetragen. Die Länge der Pfeile markiert den jeweils mit einer PCR-Reaktion (und einem spezifischen Antisinnprimer) erhaltenen Startbereich (bei einer PCR wurden mehrere cDNA-Fragmente erhalten). Hellgraue Pfeile zeigen Ergebnisse von A549-Zellen, schwarze Pfeile markieren Bereiche für Lungen-RNA und ein dunkelgrauer Pfeil zeigt das mit Leber poly(A)⁺-RNA erhaltene Ergebnis. Die senkrechten gestrichelten Linien markieren die Positionen, an denen nach Sequenzierdaten mRNA-Startpunkte vorhanden sind. Bindungsstellen für Faktoren, die vor allem in Lebergewebe eine Transkription bewirken (CEBP, HFH1 und HNF1) sind in Rot eingezeichnet. Von den CEBP-Matrizen wurde nur die Matrizenübereinstimmungen mit 'CEBP_C' eingetragen, da diese Matrix am besten definiert ist. Der Exon-Bereich ist länger skizziert als das Ergebnis der Sequenzierung der längsten cDNA zeigt, da eine RT-PCR für Leber poly(A)⁺-RNA ein positives Ergebnis für einen Bereich bis 1538 Bp vor Exon 2 liefert (Position 3661).

E.5 Zusammenfassung der HCII-Genstruktur

Um einen Überblick der in dieser Arbeit erhaltenen Ergebnisse und der bereits bekannten Daten zum HCII-Gen bei Mensch, Maus und Ratte zu geben, werden in der folgenden Abbildung 52 die genomischen Strukturen zusammenfassend dargestellt.

Abbildung 52: Struktur des Heparinkofaktor II-Gens in Mensch, Maus und Ratte



In dem Schema ist die HCII-Genstruktur von Mensch, Maus und Ratte dargestellt. Balken stellen Exon-Regionen dar. Ausgefüllte Bereiche werden in HCII-Protein translatiert. Zahlen über den Balken geben die Exongröße in [Bp] an. Die Zahlen zwischen den Linien beschreiben hingegen die jeweiligen Introngrößen in kB. Die ermittelten Spleißvorgänge bei Maus und Ratte sind durch die blauen, gewinkelten Linien angedeutet. Mutmaßliche Promotorbereiche sind durch ein rotes Oval markiert. Der 5'-Bereich des Exon 2 des menschlichen HCII ist gestrichelt dargestellt, da ein kleiner Teil der HCII-Transkripte in dieser über 1 kB großen Region beginnt.

Die Frage, warum die Struktur in der 5'-terminalen Region des HCII-Gens im Lauf der Evolution stark variiert, kann auf mehrfache Art beantwortet werden. Es scheinen besonders die unterschiedlichen 5'-gelegenen Exon/Introngrenzen durch einige Punktmutationen in den Sequenzen begründet zu sein. Diese Mutationen sind normale, zufällige Sequenzveränderungen innerhalb der Evolution. Es war kein spezieller Mechanismus, der zu einer Veränderung von Exongrenzen führt, auffindbar. Die Mutationsrate scheint auch in diesem Bereich nicht erhöht zu sein. Teilweise ist es jedoch durch Insertionen von repetitiven Elementen, bzw. durch Rekombinationsereignisse zu umfangreicheren Veränderungen der HCII-Sequenz gekommen.

Kurze Bereiche innerhalb der Sequenz begünstigen die Rekombinationsereignisse. Unterschiedliche Enden der gleichen repetitiven Elemente (an identischen Positionen in den untersuchten Arten) spiegeln homologe Rekombinationen wider. Es gibt große Abschnitte mit (CA)_n-Repeats bei Maus und Ratte, sowie einen Bereich in den Sequenzen, der Ähnlichkeit zu einem Bereich aus der V-J-Region von Immunglobulinen besitzt. Die Ähnlichkeit bezieht sich allerdings nur auf dort inserierte repetitive Elemente. Gehäufte Rekombinationsereignisse im chromosomalen Bereich des HCII-Gens werden unter anderem auch über chromosomale Instabilitäten dokumentiert.²³⁰ Die festgestellten größeren Sequenzunterschiede sind nicht unmittelbar in der Nähe von Exongrenzen. Somit haben sie wahrscheinlich keinen direkten Einfluß auf die Genstruktur.

Neben den Möglichkeiten zur Rekombination scheint der Selektionsdruck auf einen unveränderten 5'-Bereich des HCII-Gens gering zu sein. Einzelne Mutationen haben eine geringe Bedeutung, da unterschiedliche Promotorregionen und ein großer Startbereich vorhanden sind. Da das erste Exon nicht translatiert wird, ist ein Verlust für das Genprodukt ohne Bedeutung, sofern ein Transkriptionsstart vor Exon 2 möglich ist. Außerdem scheint selbst der 5'-translatierte Genbereich für Mutationen anfällig zu sein. Ohne Beeinträchtigung der Proteinfunktion kann das Signalpeptid und der aminoternale Bereich des Proteins leicht verändert sein.

In dieser Arbeit wurde bisher nicht überprüft, ob in Maus und Ratte HCII-Transkripte in anderen Geweben als Leber nachzuweisen sind. Es bleibt weiterhin offen, ob das differentielle Spleißen bei Maus und Ratte gewebsspezifisch ist.

Es müssen ebenfalls Untersuchungen auf HCII-Transkripte mit einem Startpunkt vor Exon 2 in Maus und Ratte, die den langen menschlichen Transkripten entsprechen, folgen. Ob die langen Transkripte jedoch in der Maus oder der Ratte vorliegen, ist fraglich, da sich durch inserierte repetitive Elemente der 3'-Bereich des menschlichen HCII-Intron 1 deutlich unterscheidet.

²³⁰ Puech *et al.*, 1997

F. Verzeichnisse

F.1 Literaturverzeichnis

- Aden, D. P., Fogel, A., Plotkin, S., Damjanov, I. & Knowles, B. B. (1979)**
„Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line“
Nature **282**, 615-616
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990)**
„Basic local alignment search tool“
Journal of Molecular Biology **215**, 403-410
- Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994)**
„Issues in searching molecular sequence databases“
Nature Genetics **6**, 119-129
- Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. & Kishimoto T. (1990)**
„A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family“
The EMBO Journal **9**, 1897-1906
- Andersson, T., Lorentzen, B., Hogdahl, H., Clausen, T., Mowinkel, M.-C. & Abildgaard, U. (1996)**
„Thrombin-inhibitor complexes in the blood during and after delivery“
Thrombosis Research **82**, 109-117
- Barnes, W. (1992)**
„The fidelity of Taq polymerase catalyzing PCR is improved by N-terminal deletion“
Gene **112**, 29-35
- Barnes, W. (1994)**
„PCR amplification of up to 35 kb DNA with high fidelity and high yield from λ bacteriophage templates“
Proceedings of the National Academy of Sciences USA **91**, 2216-2220
- Bartuski, A. J., Kamachi, Y., Schick, C., Overhauser, J. & Silverman, G. A. (1997)**
„Cytoplasmatic antiproteinase 2 (PI8) and bomapin (PI10) map to the serpin cluster at 18q21.3“
Genomics **43**, 321-328
- Batzer, M. A., Deininger, P. L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C. M., Schmid, C. W., Zietkiewicz, E. & Zuckerkandl, E. (1996)**
„Standardized nomenclature for Alu repeats“
Journal of Molecular Evolution **42**, 3-6
- Bellart, J., Gilabert, R., Cabero, L., Fontcuberta, J., Monasterio, J. & Miralles, R. (1998)**
„Heparin cofactor II: a new marker for pre-eclampsia“
Blood Coagulation and Fibrinolysis **9**, 205-208
- Berger, P., Kozlov, S. V., Krueger, S. R. & Sonderegger, P. (1998)**
„Structure of the mouse gene for the serine protease inhibitor neuroserpin (PI12)“
Gene **214**, 25-33
- Bertioli, D. (1997)**
„Rapid amplification of cDNA ends“
Methods in Molecular Biology **67**: PCR cloning protocols, 233-238
- Blinder, M. & Tollefsen, D. (1990)**
„Site-directed mutagenesis of arginine 103 and lysine 185 in the proposed glycosaminoglycan-binding site of heparin cofactor II“
Journal of Biological Chemistry **265**, 286-291

Blinder, M., Marasa, J., Reynolds, C., Deaven, L. & Tollefsen, D. M. (1988)

„Heparin cofactor II: cDNA sequence, chromosome localization, restriction fragment length polymorphism, and expression in *Escherichia coli*“
Biochemistry **27**, No. 2, 752-759

Bock, S. C., Wion, K. L., Vehar, G. A. & Lawn, R.M. (1982)

„Cloning and expression of the cDNA for human antithrombin III“
Nucleic Acids Research **10**, 8113-8125

Bock, S. C., Marrinan, J. A. & Radziejewska, E. (1988)

„Antithrombin III utah: proline 407 to leucine mutation in a highly conserved region near the inhibitor reactive site“
Biochemistry **27**, 6171-6178

Boeke, J. D. (1996)

„A little help for my ends“
Nature **383**, 579-581

Bosma, P. J., van den Berg, E. A., Kooistra, T., Siemieniak, D. R. & Slightom, J. L. (1988)

„Human plasminogen activator inhibitor-1 gene. Promotor and structural gene nucleotide sequences“
Journal of Biological Chemistry **263**, 9129-9141

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P. (1978)

„Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries“
Proceedings of the National Academy of Sciences USA **75**, 4853-4857

Breathnach, R. & Chambon, P. (1981)

„Organization and expression of eucaryotic split genes coding for proteins“
Annual Reviews of Biochemistry **50**, 349-83

Britten, R. J. (1997)

„Mobile elements inserted in the distant past have taken on important functions“
Gene **205**, 177-182

Brookfield, J.F. (1995)

„Molecular evolution. Retroposon revivals“
Current Biology **5**, 255-256

Budarf, M. L., Eckman, B., Michaud, D., McDonald, T., Gavigan, S., Buetow, K. H., Tatsumura, Y., Liu, Z., Hilliard, C., Driscoll, D., Goldmuntz, E., Meese, E., Zwarthoff, E. C., Williams, S., McDermid, H., Dumanski, J. P., Biegel, J., Bell, C. J. & Emanuel, B. S. (1996)

„Regional localization of over 300 loci on human chromosome 22 using a somatic cell hybrid mapping panel“
Genomics **35**, 275-288

Carrell, R. & Travis, J. (1985)

„ α_1 -antichymotrypsin and the serpins: variation and counter variation“
Trends in Biochemical Sciences **10**, 20-24

Carter, P. E., Duponchel, C., Tosi, M. & Fothergill, J. E. (1991)

„Complete nucleotide sequence of the gene for human C1 inhibitor with an unusually high density of Alu elements“
European Journal of Biochemistry **197**, 301-308

Chai, K. X., Ward, D. C., Chao, J. & Chao, L. (1994)

„Molecular cloning, sequence analysis, and chromosomal localization of the human protease inhibitor 4 (kallistatin) gene (PI4)“
Genomics **23**, 370-378

Chomczynski, P. (1992)

„One-hour downward alkaline capillary transfer for blotting of DNA and RNA“
Anal Biochemistry **201**, 134-139

Church, F., Noyes, C. & Griffith, M. (1985)

„Inhibition of chymotrypsin by heparin cofactor II“
Proceedings of the National Academy of Sciences USA **82**, 6431-6434

Claverie, J.-M. & States, D. J. (1993)

„Information enhancement methods for large scale sequence analysis“
Computers in Chemistry **17**, 191-201

Cline, J., Braman, J. C. & Hogrefe, H. H. (1996)

„PCR fidelity of *Pfu* DNA polymerases and other thermostable DNA polymerases“
Nucleic Acids Research **24**, 3546-3551

Colwell, N. S., Grupe, M. J. & Tollefsen, D. M. (1999)

„Amino acid residues of heparin cofactor II required for stimulation of thrombin inhibition by sulphated polyanions“

Biochimica et Biophysica Acta **1431**, 148-156

Colwell, N. S. & Tollefsen, D. M. (1998)

„Isolation of frog and chicken cDNAs encoding heparin cofactor II“
Thrombosis and Haemostasis **80**, 784-790

Coolidge, C. J., Seely, R. J. & Patton, J. G. (1997)

„Functional analysis of the polypyrimidine tract in pre-mRNA splicing“
Nucleic Acids Research **25**, 888-896

Coughlin, P., Sun, J., Cerruti, L., Salem, H.H. & Bird, P. (1993)

„Cloning and molecular characterization of a human intracellular serine proteinase inhibitor“
Proceedings of the National Academy of Sciences USA **90**, 9417-9421

Cui, Y., Narayanan, C. S., Zhou, J. & Kumar, A. (1998)

„Exon-I is involved in positive as well as negative regulation of human angiotensinogen gene expression“
Gene **224**, 97-107

Czihak, G., Langer, H. & Ziegler, H. (1992)

„Biologie - Ein Lehrbuch“
Springer Verlag, Berlin, 5. Auflage

Dallas-Yang, Q., et al. (1998)

„Avoiding false positives in colony PCR“
BioTechniques **24**, 580-582

Darnell, J., Lodish, H. & Baltimore, D. (1990)

„Molecular cell biology, second edition“
Scientific American Books, distributed by W. H. Freeman and Company, New York

Davis, A. R., Mascolo, P. L. & Quinn, M. T. (1999)

„Rapid miniprep of plasmid DNA for screening multiple colonies“
BioTechniques **26**, 66-68

de Jong, R., van der Heijden, J. & Meijlink, F. (1993)

„DNA-binding specificity of the S8 homeodomain“
Nucleic Acids Research **21**, 4711-4720

Duga, S., Asselta, R., Giacco, L. D., Malcovati, M., Ronchi, S., Tenchini, M. L. & Simonic, T. (1999)

„A new exon in the 5' untranslated region of the *connexin32* gene“
European Journal of Biochemistry **259**, 188-196

Edwards, J. B. D. M., Valdenaire, O. & Mallet, J. (1997)

„Anchoring a defined sequence to the 5'-ends of mRNAs“
Methods in Molecular Biology **67**: PCR Cloning Protocols, 261-278

Eyre, H. J., Sun, J., Sutherland, G. R. & Bird, P. (1996)

„Chromosomal mapping of the gene (PI9) encoding the intracellular serpin proteinase inhibitor 9 to 6p25 by fluorescence *in situ* hybridisation“
Genomics **37**, 406-408

Evans, E., Cooley, J. & Remold-O'Donnell (1995)

„Characterization and chromosomal localization of the ELANH2, the gene encoding human monocyte/neutrophil elastase inhibitor“
Genomics **28**, 235-240

Fernandez-Rachubinski, F. A., Weiner, J. H. & Blajchman, M. A. (1996)

„Regions flanking exon 1 regulate constitutive expression of the human antithrombin gene“
Journal of Biological Chemistry **271**, 29502-29512

Frech, K., Danescu-Mayer, J., Werner, T. (1997)

„A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter“
Journal of Molecular Biology **270**, 674-687

Fukamizu, A., Takahashi, S., Seo, M. S., Tada, M., Tanimoto, K., Uehara, S. & Murakami, K. (1990)

„Structure and expression of the human angiotensinogen gene“
Journal of Biological Chemistry **265**, 7567-7582

Gaillard, C. & Strauss, F. (1994)

„Association of poly(CA)•poly(TG) DNA fragments into four-stranded complexes bound by HMG1 and 2“
Science **264**, 433-436

Gaillard, I., Clauser, E. & Corvol, P. (1989)

„Structure of human angiotensinogen gene“
DNA **8**, 87-99

Giard, D. J., Aaronson, S. A., Todaro, G. J., Arnstein, P., Kersey, J. H., Dosik, H. & Parks, W. P. (1973)

„*In vitro* cultivation of human tumors: establishment of cell lines derived from a series of solid tumors“
Journal of the National Cancer Institute **51**, 1417-1423

Giguere, V., Tini, M., Flock, G., Ong, E., Evans, R. M. & Otulakowski, G. (1994)

„Isoform-specific amino-terminal domains dictate DNA-binding properties of RORalpha, a novel family of orphan hormone nuclear receptors“
Genes Development **8**, 538-553

Gils, A. & Declerck, P. (1998)

„Structure-function relationships in serpins: current concepts and controversies“
Thrombosis and Haemostasis **80**, 531-541

Good, L. & Nazar, R. N. (1997)

„Plasmid mini-preparations from culture streaks“
BioTechniques **22**, 404-406

Grabowski, P. J. (1998)

„Splicing regulation in neurons: tinkering with cell-specific control“
Cell **92**, 709-712

Graur, D., Duret, L. & Gouy, M. (1996)

„Phylogenetic position of the order *Lagomorpha* (rabbits, hares and allies)“
Nature **379**, 333-335

- Groenen, M. A. M., Dijkhof, R. J. M., van der Poel, J. J., van Diggelen, R. & Verstege, E. (1992)**
„Multiple octamer binding sites in the promoter region of the bovine α_2 -casein gene“
Nucleic Acids Research **20**, 4311-4318
- Gruhl, M., Kao, W.-Y. & Bergtrom, G. (1997)**
"Evolution of orthologous intronless and intron-bearing globin genes in two insect species"
Journal of Molecular Evolution **45**, 499-508
- Hafeez, W., Ciliberto, G. & Perlmutter, D. H. (1992)**
„Constitutive and modulated expression of the human α_1 -antitrypsin gene“
Journal of Clinical Investigation **89**, 1214-1222
- Häger, K.-P., Müller, B., Wind, C., Erbach, S. & Fischer, H. (1996)**
"Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification"
FEBS Letters **387**, 94-98
- Hall, S. L. & Padgett, R. A. (1994)**
„Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites“
Journal of Molecular Biology **239**, 357-365
- Hall, S. L. & Padgett, R. A. (1996)**
„Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns“
Science **271**, 1716-1718
- Hamada, H., Petrino, M. G., Kakunaga, T., Seidman, M. & Stollar, B. D. (1984)**
„Characterization of genomic poly(dT-dG)•poly(dC-dA) sequences: structure, organization, and conformation“
Molecular and Cellular Biology **4**, 2610-2621
- Hammond, G. H., Smith, C. L., Goping, I. S., Underhill, D. A., Harley, M. J., Reventos, J., Musto, N. A., Gunsalus, G. L. & Bardin, C. W. (1987)**
„Primary structure of human corticosteroid binding globulin, deduced from hepatic and pulmonary cDNAs, exhibits homology with serine protease inhibitors“
Proceedings of the National Academy of Sciences USA **84**, 5153-5157
- Hayashi, T., Mori, Y., Janssen, O. E., Sunthornthepvarakul, T., Weiss, R. E., Takeda, K., Weinberg, M., Seo, H., Bell, G. I. & Refetoff, S. (1993)**
„Human thyroxin-binding globulin gene: complete sequence and transcriptional regulation“
Molecular Endocrinology **7**, 1049-1060
- Hayashi, T. & Suzuki, K. (1993)**
„Gene organization of human protein C inhibitor, a member of serpin family proteins encoded in five exons“
International Journal of Hematology **58**, 213-224
- Heilig, R., Muraskowsky, R., Kloepfer, C. & Mandel, J. L. (1982)**
„The ovalbumin gene family: complete sequence and structure of the Y gene“
Nucleic Acids Research **10**, 4363-4382
- Heinemeyer, T., Chen, X., Karas, H., Kel, A. E., Kel, O. V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F., Wingender, E. (1999)**
„Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms“
Nucleic Acids Research **27**, 318-322
- Herzog, R., Lutz, S., Blin, N., Marasa, J. C., Blinder, M. A. & Tollefsen, D. M. (1991)**
„Complete nucleotide sequence of the gene for human heparin cofactor II and mapping to chromosomal band 22q11“
Biochemistry **30**, 1350-1357
- Hill, R. E. & Hastie, N. D. (1987)**
„Accelerated evolution in the reactive centre regions of serine protease inhibitors“
Nature **326**, 96-99

Hill, R. E., Shaw, P. H., Boyd, P. A., Baumann, H. & Hastie, N. D. (1984)

„Plasma protease inhibitors in mouse and man: divergence within the reactive centre regions“
Nature **311**, 175-177

Hirosawa, S., Nakamura, Y., Miura, O., Hirose, S., Sumi, Y., Aoki, N. (1988)

„Organization of the human α_2 -plasmin inhibitor gene“
Proceedings of the National Academy of Sciences USA **85**, 6836-6840

Hsiang, C.-H., Marten, N. W. & Straus, D. S. (1999)

„Upstream region of rat serum albumin gene promoter contributes to promoter activity: presence of functional binding site for hepatocyte nuclear factor-3“
Biochemical Journal **338**, 241-249

Huang, C.-H., Xie, S.-S., Socha, W. & Blumenfeld, O. O. (1995)

"Sequence diversification and exon inactivation in the glycophorin A gene family from chimpanzee to human"
Journal of Molecular Evolution **41**, 478-486

Huber, R. & Carrell, R. (1989)

„Implications of the three-dimensional structure of α_1 -antitrypsin for structure and function of serpins“
Biochemistry **28**, 8951-8965

Hunt, L. & Dayhoff, M. (1980)

„A surprising new protein superfamily containing ovalbumin“
Biochemical and Biophysical Research Communications **95**, 864-871

Jiang, H., Wang, Y., Huang, Y., Mulnix, A. B., Kadel, J., Cole, K. & Kanost M. R. (1996)

„Organisation of serpin gene-1 from *Manduca sexta*“
Journal of Biological Chemistry **271**, 28017-28023

Jurka, J. & Klonowski, P. (1996)

„Integration of retroposable elements in mammals: selection of target sites“
Journal of Molecular Evolution **43**, 685-689

Kalmovarin, N., Friedrichs, W. E., O'Brien, H. V., Linehan, L. A., Bowman, B. H. & Yang, F. (1991)

„Extrahepatic expression of plasma protein genes during inflammation“
Inflammation **15**, 369-379

Kamp, P.-B. (1995)

„Transkriptionale und posttranslationale Regulation der Nitrogenase in *Rhodobacter capsulatus*“
Diplomarbeit der Fakultät für Biologie, Universität Bielefeld

Kamp, P.-B. & Ragg, H. (1999)

„Rapid changes in the exon/intron structure of a mammalian thrombin inhibitor gene“
Gene **229**, 137-144

Karlin, S. & Altschul, S. F. (1993)

„Applications and statistics for multiple high-scoring segments in molecular sequences“
Proceedings of the National Academy of Sciences USA **90**, 5873-5877

Kass, D. H., Kim, J., Rao, A. & Deininger, P. L. (1997)

„Evolution of B2 repeats: the muroid explosion“
Genetica **99**, 1-13

Kim, U., Shizuya, H., Kang, H. (1996)

„A bacterial artificial chromosome-based framework contig map of human chromosome 22q“
Journal of the National Academy of Science of the USA **93**, 6297-6301

Kim, Y., Lee, K. & Linhardt, R. (1988)

„Microheterogeneity of plasma glycoproteins heparin cofactor II and antithrombin III and their carbohydrate analysis“

Thrombosis Research **51**, 97-104

Knippers, R., Philippsen, P., Schäfer, K. P. & Fanning, E. (1990)

„Molekulare Genetik, 5. neubearbeitete und erweiterte Auflage“

Georg Thieme Verlag Stuttgart

Knowles, B. B., Howe, C. C. & Aden, D. P. (1979)

„Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen“

Science **209**, 497-499

Koike, C., Hayakawa, Y., Niiya, K., Sakuragawa, N. & Sasaki, H. (1996)

„The production of heparin cofactor II is not regulated by inflammatory cytokines in human hepatoma cells: comparison with PAI-1“

Thrombosis and Haemostasis **75**, 298-302

Kounnas, M., Church, F., Argraves, W. & Strickland, D. (1996)

„Cellular internalization and degradation of antithrombin III-thrombin, heparin cofactor II-thrombin, and α_1 -antitrypsin complexes is mediated by the low density lipoprotein receptor-related protein“

Journal of Biological Chemistry **271**, 6523-6529

Kramerov, D. A., Tillib, S. V., Shumyatsky, G. P. & Georgiev, G. P. (1990)

„The most abundant nascent poly(A)⁺ RNAs are transcribed by RNA polymerase III in murine tumor cells“

Nucleic Acids Research **18**, 4499-4506

Krawczak, M., Reiss, J. & Cooper, D. N. (1992)

„The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences“

Human Genetics **90**, 41-54

Krepulat, F. (1999)

„Untersuchungen zur Expression von Heparinkofaktor II & Varianten mit Immun-Tags in *E. coli*“

Diplomarbeit der Fakultät für Biologie, Universität Bielefeld

Kress, M., Barra, Y., Seidman, J. G., Khoury, G. & Jay, G. (1984)

„Functional insertion of an Alu type 2 (B2 SINE) repetitive sequence in murine class I genes“

Science **226**, 974-977

Kudo, S. & Fukuda, M. (1989)

„Structural organization of glykophorin A and B genes: glykophorin B gene evolved by homologous recombination at *Alu* repeat sequences“

Proceedings of the National Academy of Sciences USA **86**, 4619-4623

Kügler, S., Plenz, G. & Müller, P. K. (1996)

„Two additional 5' exons in the human *Vigilin* gene distinguish it from the chicken gene and provide the structural basis for differential routes of gene expression“

European Journal of Biochemistry **238**, 410-417

Kumar, S. & Hedges, S. B. (1998)

„A molecular timescale for vertebrate evolution“

Nature **392**, 917-920

Lewin, B. (1991)

„GENE II - Lehrbuch der molekularen Genetik“

VCH Verlagsgesellschaft, Weinheim, 2. Auflage

- Li, Y., Zhou, L., Twining, A. S., Sugar, J. & Yue, B. Y. J. T. (1998)**
„Involvement of SP1 elements in the promotor activity of the α_1 -proteinase inhibitor gene“
Journal of Biological Chemistry **273**, 9959-9965
- Lieber, M., Smith, B., Szakal, A., Nelson-Rees, W. & Todaro, G. (1976)**
„A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells“
International Journal of Cancer **17**, 62-70
- Lindl, T. & Bauer, J. (1994)**
„Zell- und Gewebekultur“
Gustav Fischer Verlag, 3. Auflage
- Linz, U. & Degenhardt, H. (1990)**
„Die Polymerase-Kettenreaktion“
Naturwissenschaften **77**, 515-530
- Long, G. L., Chandra, T., Woo, S. L. C., Davie, E. W. & Kurachi, K. (1984)**
„Complete sequence of the cDNA for human α_1 -antitrypsin and the gene for the S variant“
Biochemistry **23**, 4828-4837
- Loskutoff, D. J., Linders, M., Keijer, J., Veerman H., van Heerikhuizen, H. & Pannekoek, H. (1987)**
„Structure of the human plasminogen activator inhibitor 1 gene: nonrandom distribution of introns“
Biochemistry **26**, 3763-3768
- Lund, A. H., Duch, M. & Pedersen, F. S. (1996)**
„Increased cloning efficiency by temperature-cycle ligation“
Nucleic Acids Research **24**, 800-801
- Mandel, M. & Higa, A. (1970)**
„Calcium dependent bacteriophage DNA infection“
Journal of Molecular Biology **53**, 154-162
- Mantovani, R. (1998)**
„A survey of 178 NF-Y binding CCAAT boxes“
Nucleic Acids Research **26**, 1135-1143
- Marshall, C. J. (1993)**
„Evolutionary relationships among the serpins“
Philosophical Transactions of the Royal Society of London **342**, 101-119
- Mathews, D. H., Andre, T. C., Kim, J., Turner, D. H. & Zuker, M. (1998)**
„An updated recursive algorithm for RNA secondary structure prediction with improved free energy parameters“
American Chemical Society Symposium Series **682**, 246-257
- Maxam, A. & Gilbert, W. (1980)**
Methods in Enzymology **65**, 499
- McCarthy, B. & Warrel, D. (1997)**
„Ovalbumin and serpin inhibitory function“
Journal of Molecular Biology **267**, 561-569
- McGrogan, M., Kennedy, J., Golini, F., Ashton, N., Dunn, F., Bell, K., Tate, E., Scott, R. W. & Simonsen, C. C. (1990)**
„Structure of the human protease nexin gene and expression of recombinant forms of PN-I“
Serine proteases and their serpin inhibitors in the nervous system, Herausgeber: B. Festoff, Elsevier Amsterdam
- Miralles, F., Ibanez-Tallon, I., Parra, M., Crippa, M., Blasi, F., Besser, D., Nagamine, Y. & Munoz-Canoves, P. (1999)**
„Transcriptional regulation of the murine urokinase-type plasminogen activator gene in skeletal myoblasts“
Thrombosis and Haemostasis **81**, 767-774

Moore, J. K. & Haber J. E. (1996)

„Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks“
Nature **383**, 644-646

Morgan, K. & Kalsheker, N. A. (1997)

„Regulation of the serine proteinase inhibitor (SERPIN) gene α_1 -antitrypsin: a paradigm for other SERPINs“
International Journal of Biochem. Cell. Biology **29**, 1501-1511

Mount, S. M. (1982)

„A catalogue of splice junction sequences“
Nucleic Acids Research **10**, 459-472

Müllhard, C., Fischer, M., Gass, P., Simon-Chazottes, D., Guenet, J.-L., Kuhse, J., Betz, H. & Becker C.-M. (1994)

"The spastic mouse: aberrant splicing of glycine receptor β subunit mRNA caused by intronic insertion of L1 element"
Neuron **13**, 1003-1015

Mullis, K. B. (1987)

„Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction“
Methods in Enzymology **155**, 335-350

Mullis, K. B. (1990)

„Eine Nachtfahrt und die Polymerase-Kettenreaktion“
Spektrum der Wissenschaft **6**, 60-67

Naas, T. P., DeBerardinis, R. J., Moran, J. V., Ostertag, E. M., Kingsmore, S. F., Seldin, M. F., Hayashizaki, Y., Martin, S. L. & Kazazian, H. H. (1998)

„An actively retrotransposing, novel subfamily of mouse L1 elements“
The EMBO Journal **17**, 590-597

Okada, K., Lijnen, H. R., Dewerchin, M., Belayew, A., Matsuo, O., Collen, D. & Bernaerts, R. (1997)

„Characterization and targeting of the murine α_2 -antiplasmin gene“
Thrombosis and Haemostasis **78**, 1104-1110

Olds, R. J., Lane, D. A., Ireland, H., Leone, G., De Stefano, V., Wiesel, M. L., Cazenave, J. P. & Thein, S. L. (1991)

„Novel point mutations leading to type 1 antithrombin deficiency and thrombosis“
British Journal of Haematology **78**, 408-413

Olds, R. J., Lane, D. A., Chowdhury, V., De Stefano, V., Leone, G. & Thein, S. L. (1993)

„Complete sequence of the antithrombin gene: evidence for homologous recombination causing thrombophilia“
Biochemistry **32**, 4216-4224

O'Neill, R. J. W., Brennan, F. E., Delbridge, M. L., Crozier, R. H. & Graves, J. A. M. (1998)

"*De novo* insertion of an intron into the mammalian sex determining gene, SRY"
Proceedings of the National Academy of Sciences USA **95**, 1653-1657

Overdier, D. G., Porcella, A. & Costa, R. H. (1994)

„The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino acid residues adjacent to the recognition helix“
Molecular and Cellular Biology **14**, 2755-2766

Overdier D. G., Ye H., Peterson R. S., Clevidence D.E. & Costa R. H. (1997)

„The winged helix transcriptional activator HFH-3 is expressed in the distal tubules of embryonic and adult kidney“
Journal of Biological Chemistry **272**, 13725-13730

- Pattanakitsakul, S., Zheng, J.-H., Natsuume-Sakai, S., Takahashi, M. & Nonaka, M. (1992)**
„Aberrant splicing caused by the insertion of the B2 sequence into an intron of the complement C4 gene is the basis for low C4 production in H-2^k mice“
Journal of Biological Chemistry **267**, 7814-7820
- Perlino, E., Cortese, R. & Ciliberto, G. (1987)**
„The human α_1 -antitrypsin gene is transcribed from two different promoters in macrophages and hepatocytes“
The EMBO Journal **6**, 2767-2771
- Poller, W., Faber, J. P., Weidinger, S., Tief, K., Scholz, S., Fischer, M., Olek, K., Kirchgesser, M. & Heidtmann, H. H. (1993)**
„A leucine-to-proline substitution causes a defective alpha 1-antichymotrypsin allele associated with familial obstructive lung disease“
Genomics **17**, 740-743
- Potempa, J., Korzus, E. & Travis, J. (1994)**
„The serpin superfamily of proteinase inhibitors: structure, function and regulation“
Journal of Biological Chemistry **269**, 15957-15960
- Pratt, C., Whinna, H., Meade, J., Treanor, R. & Church, F. (1989)**
„Physicochemical aspects of heparin cofactor II“
Annals of the New York Academy of Sciences **566**, 104-114
- Priefer, U. (1984)**
„Isolation of plasmid DNA“ in: „Advanced molecular genetics“
Pühler, A., Timmis, K. N. (eds.), *Springer Verlag, Heidelberg*, 14-25
- Prochownik, E. V., Bock, S. C. & Orkin, S. H. (1985)**
„Intron structure of the human antithrombin III gene differs from that of other members of the serine protease inhibitor superfamily“
Journal of Biological Chemistry **260**, 9608-9612
- Promega GmbH, 1998**
Technical Manual: „Dual-Luciferase™ Reporter Assay System“
- Puech, A., Saint-Jore, B., Funke, B., Gilbert, D. J., Sirotkin, H., Copeland, N. G., Jenkins, N. A., Kucherlapati, R., Morrow, B. & Skoultschi, A. I. (1997)**
„Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization“
Proceedings of the National Academy of Sciences USA **94**, 14608-14613
- Qiagen Produktbeschreibung: Qiaquick Spin Handbook (1997)**
„Qiaquick PCR Purification Kit“
- Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995)**
„MatInd and MatInspector - new fast and versatile tools for detection of consensus matches in nucleotide sequence data“
Nucleic Acids Research **23**, 4878-4884
- Ragg, H., Ulshöfer, T. & Gerewitz, J. (1990a)**
„Glycosaminoglycan-mediated leuserpin-2/thrombin interaction structure-function relationships“
Journal of Biological Chemistry **265**, 22386-22391
- Ragg, H., Ulshöfer, T. & Gerewitz, J. (1990b)**
„On the activation of human leuserpin-2, a thrombin inhibitor, by glycosaminoglycans“
Journal of Biological Chemistry **265**, 5211-5218
- Ragg, H. & Preibisch, G. (1988)**
„Structure and expression of the gene coding for human serpin hLS2“
Journal of Biological Chemistry **263**, 12129-12134

Ragg, H. (1986)

„A new member of the plasma protease inhibitor gene family“
Nucleic Acids Research **14**, 1073-1088

Riewald, M. & Schleef, R. R. (1995)

„Molecular cloning of bomapin (protease inhibitor 10), a novel human serpin that is expressed specifically in the bone marrow“
Journal of Biological Chemistry. **270**, 26754-26757

Robins, D. M. & Samuelson, L. C. (1992)

"Retrotransposon and the evolution of mammalian gene expression"
Genetica **86**, 191-201

Rollini, P. & Fournier, R. E. K. (1997)

„A 370-kb cosmid contig of the serpin gene cluster on human chromosome 14q32.1: molecular linkage of the genes encoding α_1 -antichymotrypsin, protein C inhibitor, Kallistatin, α_1 -antitrypsin, and corticosteroid-binding globulin“
Genomics **46**, 409-415

Saiki, R. K., Gelfand, D., Stoffel, S., Higuchi, R., Horn, G., Mullis, K. & Erlich, H. (1988)

„Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase“
Science **239**, 487-491

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H. & Arnheim, N. (1985)

„Enzymatic amplification of β -globin genomic sequence and restriction site analysis for diagnosis of sickle cell anemia“
Science **230**, 1350-1354

Schlegel, H. G. (1992)

„Allgemeine Mikrobiologie, 7., überarbeitete Auflage“
Georg Thieme Verlag Stuttgart

Schmidt, U. (1997)

„Untersuchungen zur Expression von Leuserpin-2-mRNA in menschlichen Geweben“
Diplomarbeit der Fakultät für Chemie, Universität Bielefeld

Schneider, S.S., Schick, C., Fish, K. E., Miller, E., Pena, J. C., Treter, S. D., Hui, S. M. & Silverman, G. A. (1995)

„A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication of the human squamous cell carcinoma antigen gene“
Proceedings of the National Academy of Sciences USA **92**, 3147-3151

Seralini, G. E., Berube, D., Gagne, R. & Hammond, G. L. (1990)

„The human corticosteroid binding globulin gene is located on chromosome 14q31-q32.1 near two other serine protease inhibitor genes“
Human Genetics **86**, 73-75

Sheffield, W. P., Schuyler, P. D. & Blajchman, M. A. (1994)

„Molecular cloning and expression of rabbit heparin cofactor II: a plasma thrombin inhibitor highly conserved between species“
Thrombosis and Haemostasis **71**, 778-782

Simioni, P., Lazzaro, A., Coser, E., Salmistrato, G. & Girolami, A. (1990)

„Hereditary heparin cofactor II deficiency and thrombosis: report of six patients belonging to two separate kindreds“
Blood Coagulation and Fibrinolysis **1**, 351-356

Smale, S. T. (1997)

„Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes“
Biochimica et Biophysica Acta **1351**, 73-88

Smit, A. F. (1993)

„Identification of a new, abundant superfamily of mammalian LTR-transposons“
Journal of Molecular Biology **246**, 401-417

Smit, A. F. & Riggs, A. D. (1995)

„MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation“
Nucleic Acids Research **23**, 98-102

Smit, A. F., Toth, G., Riggs, A. D. & Jurka, J. (1995)

„Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences“
Nucleic Acids Research **21**, 1863-1872

Smith, H. O. and M. L. Birnstiel (1976)

„A simple method for DNA restriction mapping“
Journal of Bacteriology **172**, 2387-2398

Southern, E. (1979)

„Gelelectrophoresis of restriction fragments“
Methods in Enzymology **68**, 152-176

Sprecher, C. A., Morgenstern, K. A., Mathewes, S., Dahlen, J. R., Schrader, S. K., Foster, D. C. & Kisiel, W. (1995)

„Molecular cloning, expression, and partial characterization of two novel members of the ovalbumin family of serine proteinase inhibitors“
Journal of Biological Chemistry. **270**, 29854-29861

Steinmeyer, K., Klocke, R., Ortland, C., Gronemeier, M., Jockusch, H., Gründer, S. & Jentsch, T. J. (1991)

„Inactivation of muscle chloride channel by transposon insertion in myotonic mice“
Nature **354**, 304-308

Stratogene Produktbeschreibung (1996)

„*Pfu* DNA polymerase. Instruction manual“

Sun, J., Rose, J. B. & Bird, P. I. (1995)

„Gene structure, chromosomal localization, and expression of the murine homologue of human proteinase inhibitor 6 (PI-6) suggests divergence of PI-6 from the ovalbumin serpins“
Journal of Biological Chemistry. **270**, 16089-16096

Sun, J., Stephens, R., Mirza, G., Kanai, H., Ragoussis, J. & Bird, P. I. (1998)

„A serpin gene cluster on human chromosome 6p25 contains PI6, PI9 and ELANH2 which have a common structure almost identical to the 18q21 ovalbumin serpin genes“
Cytogenetics and Cell Genetics **82**, 273-277

Takeda, A., Yamamoto, T., Nakamura, Y., Takahashi, T. & Hibino, T. (1995)

„Squamous cell carcinoma antigen is a protein inhibitor of cysteine proteinase cathepsin L“
FEBS Letters **359**, 78-80

Tanaka, N., Kawakami, T. & Taniguchi, T. (1993)

„Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system“
Molecular and Cellular Biology **13**, 4531-4538

Tarn, W.-Y. & Steitz, J. A. (1996)

„Highly diverged U4 and U6 small nuclear RNA required for splicing rare AT-AC introns“
Science **273**, 1824-1832

Tarn, W.-Y. & Steitz, J. A. (1997)

„Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge“
Trends in Biological Sciences **22**, 132-137

Teng, S.-C., Kim, B. & Gabriel, A. (1996)

„Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks“
Nature **383**, 641-644

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994)

„CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice“
Nucleic Acids Research **22**, 4673-4680

Thomson, J. M. & Compton, M. M. (1998)

„Disposable device for the isolation of DNA from agarose gels“
BioTechniques **24**, 942

Tollefsen, D. M. (1997)

„Heparin Cofactor II“
Chemistry and Biology of serpins, 35-44

Tollefsen, D. M. (1995)

„Insight into the mechanism of action of heparin cofactor II“
Thrombosis and Haemostasis **74**, 1209-1214

Tollefsen, D. M. (1988)

„Structure and function of heparin cofactor II“
First Atlas of Science, 19-24

Tollefsen, D. M., Majerus, D. & Blank, M. (1982)

„Heparin cofactor II: purification and properties of a heparin-dependent inhibitor of thrombin in human plasma“
Journal of Biological Chemistry **257**, 2162-2169

Toulon, P., Lamine, M., Ledjev, I., Guez, T., Holleman, M. E., Sereni, D. & Sicard, D. (1993)

„Heparin cofactor II deficiency in patients infected with the human immunodeficiency virus“
Thrombosis and Haemostasis **70**, 730-735

Tran, T., Marbet, G. & Duckert, F. (1985)

„Association of hereditary heparin cofactor II deficiency with thrombosis“
Lancet, 413-414

Tronche, F., Ringeisen, F., Blumenfeld, M., Yaniv, M. & Pontoglio, M. (1997)

„Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome“
Journal of Molecular Biology **266**, 231-245

Tronche, F. & Yaniv, M. (1992)

„HNF1, a homeoprotein member of the hepatic transcription regulatory network“
BioEssays **14**, 579-587

Ubeda, M., Wang, X.-Z., Zinszner, H., Wu, I., Habener, J. F. & Ron, D. (1996)

„Stress-induced binding of the transcription factor CHOP to a novel DNA control element“
Molecular and Cellular Biology **16**, 1479-1489

Underhill, D. A. & Hammond, G. L. (1989)

„Organization of the human corticosteroid binding globulin gene and analysis of its 5'-flanking region“
Molecular Endocrinology **3**, 1448-1454

Van Leeuwen, H. C., Strating, M. J., Cox, M., Kaptein, R. & van der Vliet, P. C. (1995)

„Mutation of the Oct-1 POU-specific recognition helix leads to altered DNA binding and influences enhancement of adenovirus DNA replication“
Nucleic Acids Research **23**, 3189-3197

Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J. J. & van der Vliet P. C. (1992)

„The DNA binding specificity of the bipartite POU domain and its subdomains“
The EMBO Journal **11**, 4993-5003

Villa, P., Aznar, J., Vaya, A., Espana, F., Ferrando, F., Mira, Y. & Estelles, A. (1999)

„Hereditary homozygous heparin cofactor II deficiency and the risk of developing venous thrombosis“
Thrombosis and Haemostasis **82**, 1011-1014

Voss, G. C., Botteri, F., Monard, D., Jockusch, H. (1996)

„Localization of protease nexin-1 gene Spi4 between villin and acetylcholine receptor gamma genes on mouse chromosome 1“
Mammalian Genome **7**, 167-168

Wahls, W. P., Wallace, L. J. & Moore P. D. (1990)

„The Z-DNA motif d(TG)₃₀ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture“
Molecular and Cellular Biology **10**, 785-793

Wang, S. Y. (1992)

„Structure of the gene and its retinoic acid-regulatory region for murine J6 serpin. An F9 teratocarcinoma cell retinoic acid-inducible protein“
Journal of Biological Chemistry **267**, 15362-15366

Westrup, D. & Ragg, H. (1994)

„Secondary thrombin-binding site, glycosaminoglycan binding domain and reactive center region of leuserpin-2 are strongly conserved in mammalian species“
Biochimica et Biophysica Acta **1217**, 93-96

White, T., Arnheim, N. & Erlich, H. (1989)

„The polymerase chain reaction“
Trends in Genetics **6**, 185-189

Whyatt, D. J., deBoer, E. & Grosveld, F. (1993)

„The two zinc finger-like domains of GATA-1 have different DNA binding specificities“
The EMBO Journal **12**, 4993-5005

Workshop on Comparative Genome Organization (1996)

„Comparative genome organization of vertebrates“
Mammalian Genome **7**, 717-734

Woo, S. L., Beattie, W. G., Catterall, J. F., Dugaiczky, A., Staden, R., Brownlee, G. G. & O'Malley, B. W. (1981)

„Complete nucleotide sequence of the chicken chromosomal ovalbumin gene and its biological significance“
Biochemistry **20**, 6437-6446

Wright, H. (1984)

„Ovalbumin is an elastase substrate“
Journal of Biological Chemistry. **259**, 14335-14336

Wu, Q. & Krainer, A. R. (1996)

„U1-mediated exon definition interactions between AT-AC and GT-AG introns“
Science **274**, 1005-1008

Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. (1996)

„Efficient integration of an intron RNA into double-stranded DNA by reverse splicing“
Nature **381**, 332-335

Zeng, W., Silverman, G. A. & Remold-O'Donnell, E. (1998)

„Structure and sequence of human M/NEI (monocyte/neutrophil elastase inhibitor), an ov-serpin family gene“
Gene **213**, 179-187

Zhang, G. S., Mehringer, J. H., Van Deelin, V. M. D., Kozak, C. A. & Tollefsen, D. M. (1994)

„Murine heparin cofactor II: purification, cDNA sequence, expression, and gene structure“
Biochemistry **33**, 3632-3642

Zou, Z., Anisowicz, A., Neveu, M., Rafidi, K., Sheng, S., Hendrix, M. J., Seftor, E. & Thor, A. (1994)

„Maspin, a serpin with tumor suppressing activity in human mammary epithelial cells“
Science **263**, 526-529

Zuker, M. & Stiegler, P. (1981)

„Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information“
Nucleic Acids Research **9**, 133-148

Zuker, M. (1989)

„The use of dynamic programming algorithms in RNA secondary structure prediction“
in: *Mathematical Methods for DNA Sequences*, M. S. Waterman ed. CRC Press, Inc., 159-184

F.2 Abkürzungen

A	Adenosin
A549	Lungenzelllinie
Ab	Antibiotikum
Abb.	Abbildung
Ak	Antikörper
Ala	Alanin
Amp	Ampicillin
Ap ^r	Ampicillin-Resistenz, vermittelt über die Wirkung der β -Laktamase (Gen: <i>bla</i>)
A-PCR	Anchor-PCR (Methoden C.2.4)
APS	Ammoniumsulfat
As	Aminosäure
ATP	Adenosintriphosphat
BCIP	5-Bromo-4-chloro-3-indolylphosphat
<i>bla</i>	Gen der β -Laktamase, die der Zelle Ampicillin-Resistenz verleiht
Bp	basepair(s) (engl.); Basenpaar(e)
'100 Bp- Leiter'	spezieller DNA-Längenstandard (siehe C.1.15): 100bp Ladder
BPB	Bromphenolblau
BSA	Bovine Serum Albumin; Rinder Serum Albumin
c	konstitutive Expression des Gens
C	Cytosin
° C	Grad Celsius
cDNA	complementary DNA (engl.: Eine synthetische DNA, die <i>in vitro</i> durch reverse Transkription von einer bestimmten RNA hergestellt wurde und somit komplementär zu dieser ist. Manche Autoren bezeichnen sie als copy-DNA)
COS	Zelllinie der afrikanischen grünen Meerkatze (CV1 Origin SV40)
CTAB	Hexadecyltrimethylammoniumbromid
CTP	Cytosintriphosphat
(DE3)	Bakterienstamm ist lysogen für den Bakteriophagen DE3
D	Dalton, Masseneinheit; entspricht 1/12 der Masse eines C-Atoms. 1D \approx 1,66 10 ⁻²⁷ kg
d	Desoxy-
dcm	Keine Methylierung von Cytosin in der Sequenz CCWGG
dd	Didesoxy-
Dag	Abkürzung für HCII-spezifische Primer
DEPC	Diethylpyrocarbonat
DEPC-dH ₂ O	Diethylpyrocarbonat-behandeltes dH ₂ O (siehe Methoden C.2.6.8)
dH ₂ O	demineralisiertes, autoklaviertes Reinstwasser (siehe Material C.1.4)
DNA	Deoxyribonucleic Acid (engl.); Desoxyribonukleinsäure
dsDNA	double stranded DNA (engl.); doppelsträngige DNA
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraessigsäure
EMB	Eosin-Methylenblau
<i>et al.</i>	et alii (lat.); und andere
EtBr	2,7-Diamino-10-Ethyl-9-Phenyl-Phenanthridiniumbromid; Ethidiumbromid
EtOH	Ethanol
Exon 1*	Das Exon 1 des längeren HCII-Transkripts, entspricht Teilen von Intron 1 und dem Exon 2 des 2,3 kB-Transkripts
F ⁻	Bakterienstamm enthält kein F-Episom
G	Guanin
g	Gramm
GAG	Glykosaminoglykan
gal	Galactose kann von einem Bakterium nicht verwertet werden
Gly	Glycin
GTP	Guanosintriphosphat
h	hour (engl.); Stunde
HCII	Heparin cofactor II (engl.); Heparinkofaktor II; Leuserpin-2

HepG2	Leberzelllinie
hHCII	human HCII (engl.); menschliches HCII
hnRNA	heterous nuclear RNA (engl.); RNAs unterschiedlicher Länge, prä-mRNA
hds _B (r _B ⁻ m _B ⁻)	Verhindert bei Bakterien die Restriktion und Adenosin-Methylierung an bestimmten Sequenzen TGA(N) ₆ TGCT
I	Inosin
IgG	Immunglobulin G
IPTG	Isopropyl-β-D-Thiogalaktosid
ITP	Inosintriphosphat
k	Kilo-, 10 ³
Kan	Kanamycin
kB	Kilobasen, eine Längeneinheit, die 1000 Basen eines Nucleinsäuremoleküls entspricht. Eine Kilobase doppelsträngiger DNA besitzt eine Länge von 0,34 μm und eine Masse von etwa 660 kD.
'kB-Leiter'	spezieller DNA-Längenstandard (siehe C.1.15): Perfect 1 kb DNA Ladder
<i>Klen</i>	KlenTherm-Polymerase, verkürzte <i>Taq</i> -Polymerase
Km ^r	Kanamycin-Resistenz
l	Liter
<i>lacZ</i>	für die β-Glukosidase kodierendes Gen
LB	Lennox L Broth Base
LBG	Lennox L Broth Base mit Glukose
LM	Längenmarker
LMW	low molecular weight (engl.); niedriges Molekulargewicht
M	Molar
m	Milli-, 10 ⁻³
μ	Mikro-, 10 ⁻⁶
'Marker VIII'	spezieller DNA-Längenstandard (siehe C.1.15): DNA Molecular Weight Marker VIII
mcs	multiple cloning site (engl.); Bereich im Plasmid mit vielen Restriktionsschnittstellen
Met	Methionin
mHCII	murine HCII (engl.); Maus HCII
MOPS	Morpholinopropanesulfonic acid
mRNA	messenger ribonucleic acid (engl.); Boten-RNA
Mst	Abkürzung für spezifische mHCII- oder rHCII-Primer
n	Nano-, 10 ⁻⁹
NBT	Nitroblau-Tetrazoliumchlorid
nested PCR	zweite PCR mit innenliegenden Primern zu einer DNA-Matrize aus einer ersten PCR
NTP	Nukleosidtriphosphat
oD ₆₀₀	optische Dichte einer Lösung, gemessen bei einer Wellenlänge von 600 nm
ORF	open reading frame (engl.); Offenes Leseraster
P _{xyz}	spezifischer Promotor des Gens <i>xyz</i>
p.a.	pro analysis (lat.); zur Analyse
PAA	Polyacrylamid
PBS	phosphate buffered saline (engl.); Phosphatpuffer
PCR	polymerase chain reaction (engl.); Polymerase-Kettenreaktion
<i>Pfu</i>	klonierte <i>Pfu</i> DNA Polymerase aus <i>Pyrococcus furiosus</i>
poly(A) ⁺ -RNA	RNA mit einem 3'-Ende aus einer Vielzahl von A-Nukleotiden
pWTB11 ^r	Shuttle-Vektor, der die HCII-cDNA enthält Resistent
rHCII	rat HCII (engl.); HCII der Ratte
rabHCII	rabbit HCII (engl.); HCII des Kaninchens
<i>RcaI</i>	Restriktionsendonuklease aus <i>Rhodococcus capsulatum</i>
RNA	ribonucleic acid (engl.); Ribonukleinsäure
Rollern	Inkubieren von Membranen in einem Hybridisierungssofen
<i>RsaRI</i>	Restriktionsendonuklease aus <i>Rhodopseudomonas sphaeroides</i>
RT	Raumtemperatur
RT-PCR	reverse Transkriptase Reaktion mit anschließender PCR
^s	Sensitiv
SAP	shrimp alkaline phosphatase (engl.); alkalische Phosphatase aus Krabben
SA-AP	Streptavidin konjugierte alkalische Phosphatase
Ser	Serin

SDS	sodium dodecyl sulfate (engl.); Natriumdodecylsulfat
SOC	sodium carbohydrate (engl.); Natrium-Kohlenhydrat (-Puffer)
Spc	Spectinomycin
Spc ^r	Spectinomycin-Resistenz
ssDNA	single stranded DNA (engl.); einzelsträngige DNA
T	Thymin
TAE	Tris-Acetat-EDTA Puffer
<i>Taq</i>	<i>Thermus aquaticus</i>
TBS	tris buffered saline (engl.); Tris-Puffer
TCM	Tris, CaCl ₂ , MgCl ₂ Puffer
TdTase	Terminale Desoxynukleotidyl Transferase
TETA	Triethylentetramin=3,6-Diazaoctan-1,8-diamin
TEMED	Tetramethylethylendiamin
T-PCR	Tailing-PCR (Methoden C.2.5)
Tris	Tris-(hydroxymethyl)-aminomethan
Triton X-100	Octylphenol-polyethylenglycolether
Tyr	Tyrosin
TTP	Thymidintriphosphat
U	Units
ü.N.	über Nacht
UpM	Umdrehungen pro Minute
USC	Abkürzung für hHCII-spezifische Primer
UTR	untranslated region (engl.); nicht translatierte Region einer Nukleinsäure
v/v	volume per volume (engl.); Volumen je Volumen
Vol.	Volumen-Anteil
w/v	weight per volume (engl.); Gewicht je Volumen
WT	Wildtyp
z.A.	zur Analyse

F.3 Tabellen-Verzeichnis

Tabelle 1: RNA-Polymerasen.....	6
Tabelle 2: Inhibierende und Nichtinhibierende Serpine.....	13
Tabelle 3: Serpingen-Familien.....	15
Tabelle 4: Verwendete Systeme.....	28
Tabelle 5: Restriktionsenzyme.....	30
Tabelle 6: Weitere Enzyme	31
Tabelle 7: Verwendete Primer für die menschliche HCII-Promotor-Region.....	31
Tabelle 8: Verwendete Primer zur Charakterisierung des HCII-Intron 1 aus der Maus.....	32
Tabelle 9: Verwendete Primer zur Charakterisierung des HCII-Intron 1 aus der Ratte.....	32
Tabelle 10: Verwendete Primer für mRNA-Analysen und Klonierungen.....	33
Tabelle 11: Plasmide	34
Tabelle 12: Verwendete Nukleinsäuren	35
Tabelle 13: <i>E. coli</i> Bakterienstämme	35
Tabelle 14: Antibiotika.....	35
Tabelle 15: Verwendete Medien für <i>E. coli</i> -Kulturen.....	36
Tabelle 16: Puffer.....	37
Tabelle 17: Lösungen	38
Tabelle 18: Sequenzen des Introns 1A und 1B des HCII-Gens der Ratte.....	72
Tabelle 19: Potentielle Signale für 5'-Intron 1-Grenzen im HCII-Gen der Ratte.....	74
Tabelle 20: Potentielle Signale für 3'-Intron 1-Grenzen im HCII-Gen der Ratte.....	75
Tabelle 21: HCII-Intron 1 Sequenz der Maus	77
Tabelle 22: Potentielle Signale für 5'-Intron 1-Grenzen im HCII-Gen der Maus.....	79
Tabelle 23: Potentielle Signale für 3'-Intron 1-Grenzen im HCII-Gen der Maus.....	80
Tabelle 24: Potentielle Signale für 5'-Intron 1-Grenzen im HCII-Gen des Menschen	80
Tabelle 25: Potentielle Signale für 3'-Intron 1-Grenzen im HCII-Gen des Menschen	80
Tabelle 26: Alignment der HCII-Intron 1-Sequenzen.....	81
Tabelle 27: DNA-Elemente im Intron 1 des Heparinkofaktor II-Gens.....	90
Tabelle 28: Zusammenfassung der Elemente im Intron 1 des HCII-Gens.....	91
Tabelle 29: Alignment von HCII-cDNAs verschiedener Organismen	93
Tabelle 30: 5'-Enden von Leber-spezifischen HCII-mRNAs	102
Tabelle 31: Startbereich von leberspezifischen HCII-mRNAs.....	103
Tabelle 32: 5'-Enden von HCII-mRNAs aus Lunge.....	106
Tabelle 33: Startbereich von Lungen-spezifischen HCII-mRNAs	107
Tabelle 34: Überprüfte Bindungsstellen für Transkriptionsfaktoren.....	111
Tabelle 35: Position der Matrizen-Übereinstimmungen für den 5'-Bereich.....	112
Tabelle 36: Anzahl der Matrizen-Übereinstimmungen für den 5'-Bereich	113
Tabelle 37: Position der Matrizen-Übereinstimmungen für den Intron 1-Bereich.....	114
Tabelle 38: Anzahl der Matrizen-Übereinstimmungen für den Intron 1-Bereich	115
Tabelle 39: DNA-Elemente vor dem Exon 1 des Heparinkofaktor II-Gens	116
Tabelle 40: Überprüfung einer 2,2 kB-Sequenz auf Funktion als Promotorbereich	117
Tabelle 41: Nukleotide im Intron 1 vom Heparinkofaktor II-Gen	122
Tabelle 42: Verwandtschaft der Intron 1-Sequenzen des HCII-Gens.....	125
Tabelle 43: Vergleich der translatierten Bereiche von HCII-cDNAs auf DNA-Ebene.....	134
Tabelle 44: Vergleich der 5'- und 3'-nicht translatierten Bereiche von HCII-cDNAs.....	135
Tabelle 45: Insertionen/Deletionen in den HCII-cDNAs.....	136

F.4 Abbildungs-Verzeichnis

Abbildung 1: Struktur von eukaryotischen proteinkodierenden Genen.....	6
Abbildung 2: Intronsignale der Klasse I Introns.....	7
Abbildung 3: Intronsignale der Klasse II Introns	8
Abbildung 4: Modell zur Verbreitung von Retroposons.....	11
Abbildung 5: Modell der Inaktivierung von Thrombin durch HCII.....	14
Abbildung 6: Genstruktur der „Ov-serpine“.....	16
Abbildung 7: Genstruktur der α_1 -Antitrypsin-Familie.....	17
Abbildung 8: Genstruktur der α_2 -Antiplasmin-Familie.....	18
Abbildung 9: Antithrombin Genstruktur.....	18
Abbildung 10: Vergleich von Exon/Introngrenzen an homologen Proteinpositionen.....	19
Abbildung 11: Colligin-2 (J6) Genstruktur	19
Abbildung 12: <i>Manduca</i> -Serpin Genstruktur	19
Abbildung 13: Genstruktur der PAI-1-Familie.....	20
Abbildung 14: Schematisches HCII-Modell.....	21
Abbildung 15: Inhibierung von Thrombin durch HCII (Modell) ⁹⁰	22
Abbildung 16: Vergleich chromosomaler Bereiche von Mensch und Maus.....	26
Abbildung 17: Längenmarker.....	40
Abbildung 18: Luciferase-Reaktionen	69
Abbildung 19: Schema der Sequenzierungen (HCII-Gen der Ratte, Intron 1)	71
Abbildung 20: Alternatives Spleißen von rHCII-Transkripten	72
Abbildung 21: Analyse des 5'-Bereichs der mHCII-mRNA mittels RT-PCR.....	75
Abbildung 22: Schema der Sequenzierungen (HCII-Gen der Maus, Intron 1).....	76
Abbildung 23: Mögliches alternatives Spleißen von mHCII-Transkripten	79
Abbildung 24: Analyse zur Identifizierung eines potentiellen Exons 1B im hHCII.....	91
Abbildung 25: Schema der Intron 1/Exon 2-Grenze.....	92
Abbildung 26: Southernblots.....	92
Abbildung 27: HCII-Transkripte in unterschiedlichen Geweben.....	98
Abbildung 28: Spaltungen von RT-PCR-Produkten.....	98
Abbildung 29: Gewebsspezifische HCII/AT-Transkripte.....	99
Abbildung 30: Schema der Anker- und der Tailing-PCR	100
Abbildung 31: Menschliche HCII-Sequenz im Bereich des Beginns von Exon 2	101
Abbildung 32: 5'-Ende der HCII-mRNA, bestimmt mit dem Primer USC8	104
Abbildung 33: 5'-Ende der HCII-mRNA, bestimmt mit dem Primer USC24	105
Abbildung 34: 5'-Ende der HCII-mRNA aus Lunge, bestimmt mit dem Primer USC26	107
Abbildung 35: 5'-Ende der HCII-mRNA aus Lunge, bestimmt mit dem Primer USC24	108
Abbildung 36: 5'-Enden der HCII-mRNA aus Lunge, bestimmt mit dem Primer USC20 ...	108
Abbildung 37: Vergleich der 5'-Startbereiche im Exon 1*.....	109
Abbildung 38: Überprüfung der 5'-Startbereiche.....	109
Abbildung 39: 5'-Bereich des HCII-Gens.....	116
Abbildung 40: Überprüfung einer 2,2 kB-HCII-Sequenz auf Funktion als Promotorbereich	118
Abbildung 41: 5'-gelegene Exon/Introngrenzen im Ratten-HCII-Gen.....	119
Abbildung 42: 5'-gelegene Exon/Introngrenzen im Maus-HCII-Gen	120
Abbildung 43: 5'-gelegene Exon/Introngrenzen im menschlichen HCII-Gen.....	121
Abbildung 44: 5'-Genstruktur von HCII.....	121
Abbildung 45: Schematische Darstellung repetitiver Elemente im Intron 1 des HCII-Gens.	126
Abbildung 46: Exon/Introngrenze am 3'-Ende von Exon 1 (Alignmentpositionen: -30 - 39)	129
Abbildung 47: Vergleich der Sequenzen um den Bereich von Exon 1B der Ratte.....	129
Abbildung 48: Bereich der 3'-Grenzen des HCII-Intron 1	131

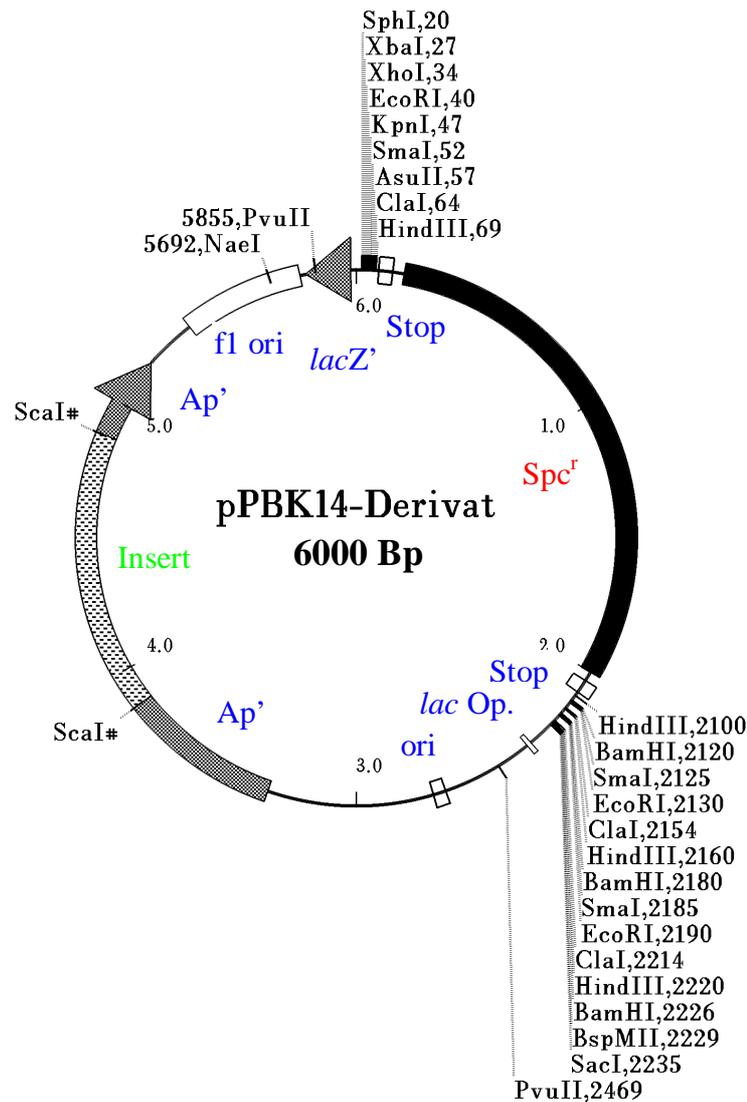
Abbildung 49: Übereinstimmungen zwischen den HCII-cDNAs	135
Abbildung 50: Schema des Bereichs vor dem HCII-Exon 1	145
Abbildung 51: Startbereich der langen HCII-Transkripte.....	148
Abbildung 52: Struktur des Heparinkofaktor II-Gens in Mensch, Maus und Ratte	149
Abbildung 53: pPBK14-Plasmid-Derivate.....	172
Abbildung 54: pRL-TK Kontroll-Vektor	173
Abbildung 55: pPBK60 (Plasmid zur Messung der Promotor-Aktivität des HCII-Gens).....	174
Abbildung 56: Sequenzierungen des mHCII-Intron 1	175
Abbildung 57: Sequenzierungen des rHCII-Intron 1	179

G. Anhang

G.1 Plasmidkarten

G.1.1 pPBK14-Derivate

Abbildung 53: pPBK14-Plasmid-Derivate

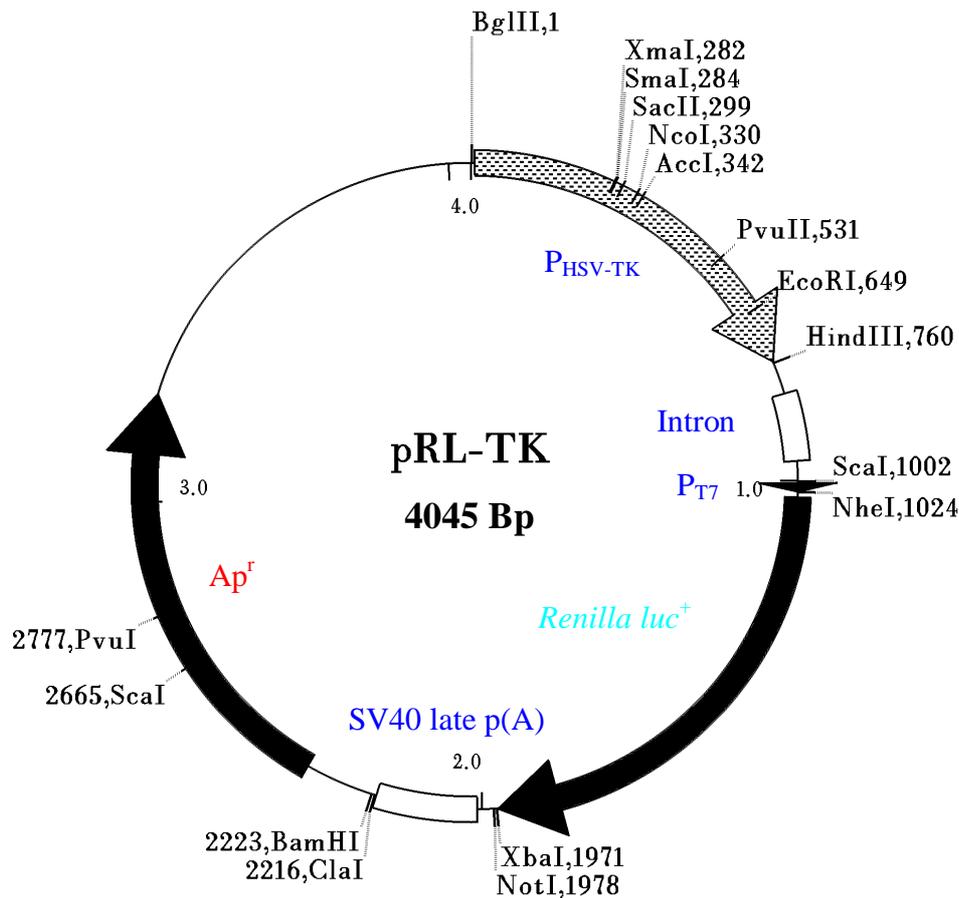


Die Plasmidkarte zeigt Derivate des Plasmids pPBK14 (Kamp, 1995), die durch verschiedenen 'blunt-end'-Insertionen in die *ScaI*-Schnittstelle des Gens für die Ap-Resistenz erhalten wurden. Als Beispiel wurde ein Insert von 1 kB Länge eingezeichnet.

Abkürzungen: *Ap'*: inaktivierte Ampicillin-Resistenz; *fl ori*: Replikationsursprung vom Phagen fl; *Insert*: eingefügte Sequenz; *lac Op.*: Operator vom Lactose-Operon; *lacZ'*: trunkiertes *lacZ*-Gen; *ori*: Replikationsursprung; *Spc^r*: Gen für Spectinomycin-Resistenz; *Stop*: Stop-Signale für Translation (in allen drei Leserastern) und Transkription.

G.1.2 pRL-TK (*Renilla* Luciferase/Kontroll-Vektor)

Abbildung 54: pRL-TK Kontroll-Vektor

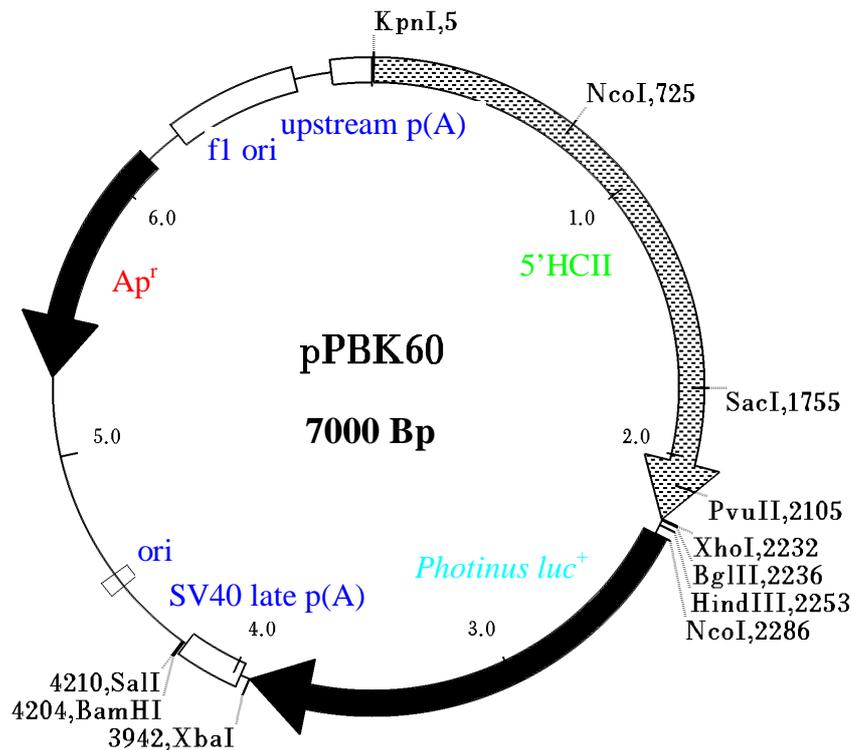


Die Plasmidkarte zeigt den Vektor pRL-TK (Promega, 1998). Das Plasmid dient als Positiv-Kontrolle und Abgleich für Schwankungen in der Transformationsrate.

Abkürzungen: **Ap^r**: Gen für Ampicillin-Resistenz (*bla*); **Intron**: synthetischer Intronbereich zur Transkriptionssteigerung; **Renilla luc⁺**: Luciferase-Gen aus *Renilla reniformis*; **P_{HSV-TK}**: Promotor des Thymidinkinase-Gens aus dem *Herpes simplex*-Virus; **P_{T7}**: Promotor des T7-Bakteriophagen; **SV40 late p(A)**: spätes Polyadenylierungssignal aus dem SV40-Virus.

G.1.3 pPBK60 (HCII-Promotor/Luciferase-Vektor)

Abbildung 55: pPBK60 (Plasmid zur Messung der Promotor-Aktivität des HCII-Gens)



Die Plasmidkarte zeigt den Vektor pPBK60, der auf dem pGL3Basis-Plasmid (Promega, 1998) beruht. Das Plasmid wird zur Ermittlung der Promotoraktivität einer 2,2 kB-Region vor dem hHCII-Exon 1 verwendet. Das Plasmid pPBK62 ist ähnlich. Das Insert wurde jedoch mittels blunt-end-Ligierung in die *Sma*I-Schnittstelle von pGL3Basis kloniert (statt einer Klonierung über die *Kpn*I/*Xho*I-Schnittstellen) und liegt in der umgekehrten Orientierung vor.

Abkürzungen: **Ap^r**: Gen für Ampicillin-Resistenz (*bla*); **fl ori**: Replikationsursprung vom Phagen fl; **5'HCII**: 2,2 Kb-Bereich 5' vom hHCII-Gen gelegen; **Intron**: synthetischer Intronbereich zur Transkriptionssteigerung; **ori**: Replikationsursprung (ColE1-Derivat); **Photinus luc⁺**: Luciferase-Gen aus *Photinus pyralis*; **SV40 late p(A)**: spätes Polyadenylierungssignal aus dem SV40-Virus; **upstream p(A)**: Polyadenylierungssignal verhindert Transkription aus anderen Plasmidbereichen.

G.2 Intron 1-Sequenzierungen des HCII-Gens von Maus und Ratte

G.2.1 Sequenzierungen des Maus HCII-Intron 1

Abbildung 56: Sequenzierungen des mHCII-Intron 1

Mst12 +	TTTATTTTAAATGCTTAGAGCTAAACAAT
Mst72 +	GTACAGGTAAGTTTGGTTTCTAACATTTTTGCCAATTATAAAATTTTATTTTAAATGCTTAGAGCTAAACAAT
Ragg	TCTCTGCCAGACCCTGCTCAGAGGTACAGAGTTAAGTTTGGTTTCTAACATTTTTGCCAATTATAAAATTTTATTTTAAATGCTTAGAGCTAAACAAT
Sequenz	TCTCTGCCAGACCCTGCTCAGAGGTACAGAGTTAAGTTTGGTTTCTAACATTTTTGCCAATTATAAAATTTTATTTTAAATGCTTAGAGCTAAACAAT
Mst12 +	TGTTTCCTTTATTGAGAAAAGAATATGTTATTTTAAAATTTAGTTTAAATATCTTTAAAATTTAGTTTAAATGTTTAAATATTAATATTGGTACA/
Mst72 +	TGTTTCCTTTATTGAGAAAAGAATATGTTATTTTAAAATTTAGTTTAAATATCTTTAAAATTTAGTTTAAATGTTTAAATATTAATATTGGTACA/
Ragg	TGTTTCCTTTATTGAGAAAAGAATATGTTATTTTAAAATTTAGTTTAAATATCTTTAAAATTTAGTTTAAATGTTTAAATATTAATATTGGTACA/
Sequenz	TGTTTCCTTTATTGAGAAAAGAATATGTTATTTTAAAATTTAGTTTAAATATCTTTAAAATTTAGTTTAAATGTTTAAATATTAATATTGGTACA/
Ragg	CTGACAAAAATGGAATTTGAATTTTAAAGAGACAAGATTATTATACCTTTGATACCTTTATACCTGGAACCTGGGCATGGCAGCACAAGATTGAAATCCAG/
Mst12 +	CTGACAAAAATGGAATTTGAATTTTAAAGAGACAAGATTATTATACCTTTGATACCTTTATACCTGGAACCTGGGCATGGCAGCACAAGATTGAAATCCAG/
Mst72 +	CTGACAAAAATGGAATTTGAATTTTAAAGAGACAAGATTATTATACCTTTGATACCTTTATACCTGGAACCTGGGCATGGCAGCACAAGATTGAAATCCAG/
Mst40 +	CTGACAAAAATGGAATTTGAATTTTAAAGAGACAAGATTATTATACCTTTGATACCTTTATACCTGGAACCTGGGCATGGCAGCACAAGATTGAAATCCAG/
Sequenz	CTGACAAAAATGGAATTTGAATTTTAAAGAGACAAGATTATTATACCTTTGATACCTTTATACCTGGAACCTGGGCATGGCAGCACAAGATTGAAATCCAG/
Mst12 +	ACTTGTAGCCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst40 +	ACTTGTAGCCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst40 +	ACTTGTAGCCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst40 +	CCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst40 +	CCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst72 +	ACTTGTAGCCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst60 +	ACTTGTAGCCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Sequenz	ACTTGTAGCCCTGGGTATCTATCACCTAGGGCACAATAATCAATTTCTTTCAAATTAGATAAATGCTAGTGATTAAGAATGAGAATATCTTGTTC/
Mst72 -	AGAGGAAGTTTCCACCTAAGCATTACTATTAAGGTT
Mst12 +	AGAGGAAGTTTCCACCTAAGCATTACTATTAAGGTTAGAGGAATGTAAGAGATCTAACCTCAAATGCCAGAAAACGTGTACATATGTTGATACATA/
Mst40 +	AGAGGAAGTTTCCACCTAAGCATTACTATTAAGGTTAGAGGAATGTAAGAGATCTAACCTCAAATGCCAGAAAACGTGTACATATGTTGATACATA/
Mst60 +	AGAGGAAGTTTCCACCTAAGCATTACTATTAAGGTTAGAGGAATGTAAGAGATCTAACCTCAAATGCCAGAAAACGTGTACATATGTTGATACATA/
Sequenz	AGAGGAAGTTTCCACCTAAGCATTACTATTAAGGTTAGAGGAATGTAAGAGATCTAACCTCAAATGCCAGAAAACGTGTACATATGTTGATACATA/
Mst12 +	ACATGTACCCCCCATGTGCACCACATCACATCATACACACACCACACATACACA
Mst40 +	ACATGTACCCCCCATGTGCACCACATCACATCATACACACACCACACATACACACACCCACATACACCATACCCACACGACACACACCACACACCCCC
Mst60 +	ACATGTACCCCCCATGTGCACCACATCACATCATACACACACCACACATACACACACCCACATACACCATACCCACACGACACACACCACACACCC
Mst73 +	ACATGTACCCCCCATGTGCACCACATCACATCATACACACACCACACATACACACACCCACATACACCATACCCACACGACACACACCACACACCC
Sequenz	ACATGTACCCCCCATGTGCACCACATCACATCATACACACACCACACATACACACACCCACATACACCATACCCACACGACACACACCACACACCC
Mst40 +	ACACATATACCACACCATACCACACCACACACACATATCATACCACACACACATACCACACACACACACACACACACACACACACACACACACAC
Mst60 +	ACACATATACCACACCATACCACACCACACACACACATATCATACCAC
Mst73 +	ACACATATACCACACCATACCACACCACACACACACATATCATACCAC
Mst56 -	CCACACACATATCATACCAC
Mst56 -	CAC
Sequenz	ACACATATACCACACCATACCACACCACACACACATATCATACCAC
Mst73 +	CACCACACACACAC
Mst60 +	CACCACACACACACATACCACACCGTGAC
Mst56 -	CACCACACACACACACACACACCGTGAC
Mst56 -	CACCACACACACACACACACACCGTGAC
Mst55 -	CACACACACACATACCACACCGTGAC
Mst56 -	CACACACACACACACACACCGTGAC
Sequenz	CACCACACACACACATACCACACCGTGAC
Mst56 -	TTCTTCAATTTCTCTGTAAGTTTGAAGTTATTTTAAACATAAAAAGGTTT-CCT
Mst56 -	TTCTTCAATTTCTCTGTAAGTTTGAAGTTATTTTAAACATAAAAAGGTTT-CCTTCTTGCTTGTCTGTTTGGATCTGGAGAGATCACAGTAGGAAATGG
Mst56 -	TTCTTCAATTTCTCTGTAAGTTTGAAGTTATTTTAAACATAAAAAGGTTT-CCTTCTTGCTTGTCTGTTTGGATCTGGAGAGATCACAGTAGGAAATGG
Mst55 -	TTCTTCAATTTCTCTGTAAGTTTGAAGTTATTTTAAACATAAAAAGGTTT-CCTTCTTGCTTGTCTGTTTGGATCTGGAGAGATCACAGTAGGAAATGG
Mst55 -	TTCTTCAATTTCTCTGTAAGTTTGAAGTTATTTTAAACATAAAAAGGTTT-CCTTCTTGCTTGTCTGTTTGGATCTGGAGAGATCACAGTAGGAAATGG
Sequenz	TTCTTCAATTTCTCTGTAAGTTTGAAGTTATTTTAAACATAAAAAGGTTT-CCTTCTTGCTTGTCTGTTTGGATCTGGAGAGATCACAGTAGGAAATGG
Mst56 -	TGTTGT
Mst56 -	TGTTCTTTATGACACA
Mst55 -	TGTTCTTTATGACACATATACAAAAAGGCCAATCCATCCATGCAACTCTCAAGAATACACTATATGATGTTCCCATGAAAACAAAATTTTGTATGAAT/
Mst55 -	TGTTCTTTATGACACATATACAAAAAGGCCAATCCATCCATGCAACTCTCAAGAATACACTATATGATGTTCCCATGAAAACAAAATTTTGTATGAAT/
Mst74 +	AAAGGCCAATCCATCCATGCAACTCTCAAGAATACACTATATGATGTTCCCATGAAAACAAAATTTTGTATGAAT/
Sequenz	TGTTCTTTATGACACATATACAAAAAGGCCAATCCATCCATGCAACTCTCAAGAATACACTATATGATGTTCCCATGAAAACAAAATTTTGTATGAAT/
Mst55 -	TCTTAAAATTTATAGCACTTTAATTTCAAAGACAGCAATTTGAACGCATACACAACCTTACAGGTATTTTAAATGCCTTAATTTTTTTTTTAAATGAAACTGA/
Mst55 -	TCTTAAAATTTATAGCACTTTAATTTCAAAGACAGCAATTTGAACGCATACACAACCTTACAGGTATTTTAAATGCCTTAATTTTTTTTTTAAATGAAACTGA/
Mst74 +	TCTTAAAATTTATAGCACTTTAATTTCAAAGACAGCAATTTGAACGCATACACAACCTTACAGGTATTTTAAATGCCTTAATTTTTTTTTTAAATGAAACTGA/
Mst52 -	TCTTAAAATTTATAGCACTTTAATTTCAAAGACAGCAATTTGAACGCATACACAACCTTACAGGTATTTTAAATGCCTTAATTTTTTTTTTAAATGAAACTGA/
Sequenz	TCTTAAAATTTATAGCACTTTAATTTCAAAGACAGCAATTTGAACGCATACACAACCTTACAGGTATTTTAAATGCCTTAATTTTTTTTTTAAATGAAACTGA/
Mst55 -	GAGCAATATAGAAAAGATTTATCTGTGAGATCGTTAATTTTCAATGCATAAAAATAACTACATTTTTTCTTTCTCTAAG
Mst55 -	GAGCAATATAGAAAAGATTTATCTGTGAGATCGTTAATTTTCAATGCATAAAAATAACTACATTTTTTCTTTCTCTAAG
Mst74 +	GAGCAATATAGAAAAGATTTATCTGTGAGATCGTTAATTTTCAATGCATAAAAATAACTACATTTTTTCTTTCTCTAAG
Mst52 -	GAGCAATATAGAAAAGATTTATCTGTGAGATCGTTAATTTTCAATGCATAAAAATAACTACATTTTTTCTTTCTCTAAG
Mst75 +	GAGCAATATAGAAAAGATTTATCTGTGAGATCGTTAATTTTCAATGCATAAAAATAACTACATTTTTTCTTTCTCTAAG
Sequenz	GAGCAATATAGAAAAGATTTATCTGTGAGATCGTTAATTTTCAATGCATAAAAATAACTACATTTTTTCTTTCTCTAAG

p30/19+2 ATAAATCTTTGGGCCTGAGCGAGCAGGGTTGACCAGAGTTAGCAGAGGTCGTAAAAATTC AATTCCCAACAACCCATGGAGGCTCACAACCATCTGTAT
 Mst76 - ATAAATCTTTGGGCCTGAGCGAGCAGGGTTGACCAGAGTTAGCAGAGGTCGTAAAAATTC AATTCCCAACAACCCATGGAGGCTCACAACCATCTGTAT
 p30/1-2 ATAAATCTTTGGGCCTGAGCGAGCAGGGTTGACCAGAGTTAGCAGAGGTTAAAAA-GTCAAGTCCCAACAACCCATGGAGGCTCACAACCATCTGTAT
 Mst77 - TCTGTAT
 Sequenz ATAAATCTTTGGGCCTGAGCGAGCAGGGTTGACCAGAGTTAGCAGAGGTCGTAAAAATTC AATTCCCAACAACCCATGGAGGCTCACAACCATCTGTAT

Mst76 - AGCTATGGTATACTCACATACATTAAATTA AAAAAAATTTTTTAATCTTTAAAAA
 p30/1-2 AGCTATGGTATACTCACATACATTAAATTA AAAAAAATTTTTTAATCTTTAAAAA
 p30/63+1 AGCTATGGTATACTCACATACATTAAATTA AAAAAAATTTTTTAATCTTTAAAAA
 Mst77 - AGCTATGGTATACTCACATACATTAAATTA AAAAAAATTTTTTAATCTTTAAAAA
 Mst49 + CTCACATACATTAAATTA AAAAAAATTTTTTAATCTTTAAAAA
 p30/19-1 TTTAAAAA ACAAACAACAAAAATAAACAAACAAAAACAAGCTGG
 Mst49 + TTTAAAAA ACAAACAACAAAAATAAACAAACAAAAACAAGCTGG
 Mst77 - TTTAAAAA ACAAACAACAAAAATAAACAAACAAAAACAAGCTGG
 Sequenz AGCTATGGTATACTCACATACATTAAATTA AAAAAAATTTTTTAATCTTTAAAAA ACAAACAACAAAAATAAACAAACAAAAACAAGCTGG

Mst49 + GAATGCTAATTTTAAACCACATTATAGGGACAGTCCCTAAGGAC
 p30/19-1 GAATGCTAATTTTAAACCACATTATAGGGACAGTCCCTAAGGACAAAGGCTGGTAAAGCT
 Mst77 - GAATGCTAATTTTAAACCACATTATAGGGACAGTCCCTAAGGACAAAGGCTGGTAAAGCTGGTCCCAAGTGTGAGCAGTTTGTAGTCTGGCATTCT
 Mst49 + GAATGCTAATTTTAAACCACATTATAGGGACAGTCCCTAAGGACAAAGGCTGGTAAAGCTGGTCCCAAGTGTGAGCAGTTTGTAGTCTGGCATTCT
 Mst49 + GGGACAGTCCCTAAGGACAAAGGCTGGTAAAGCTGGTCCCAAGTGTGAGCAGTTTGTAGTCTGGCATTCT
 p30/8+1 AGCTGTCCCTGTGTGAGCAGTTTGTAGTCTGGCATTCT
 Sequenz GAATGCTAATTTTAAACCACATTATAGGGACAGTCCCTAAGGACAAAGGCTGGTAAAGCTGGTCCCAAGTGTGAGCAGTTTGTAGTCTGGCATTCT

p30/8+1 AGGCAAGCAACATGAGACTAGAGATCTGGGACTTAACCAACCTACTACTCATCATCTACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG
 Mst77 - AGGCAAGCAACATGAGACTAGAGATCTGGGACTTAACCAACCTACTACTCATCATCTACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG
 Mst49 + AGGCAAGCAACATGAGACTAGAGATCTGGGACTTAACCAACCTACTACTCATCATCTACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG
 Mst65 - ACTTAACCAACCTACTACTCATCATCTACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG
 Mst47 - TCATCTACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG
 Mst44 + TACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG
 Sequenz AGGCAAGCAACATGAGACTAGAGATCTGGGACTTAACCAACCTACTACTCATCATCTACCATCCCTCAATCCCTACATTTTAGTGTCTGGATTG

Mst77 - ACCCTGATCCCATGCAAGCTGGGT
 Mst49 + ACCCTGATCCCATGCAAGCTGGGTGAGTGTAGTTAANATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT
 Mst65 - ACCCTGATCCCATGCAAGCTGGGTGAGTGTAGTTAATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT
 Mst47 - ACCCTGATCCCATGCAAGCTGGGTGAGTGTAGTTAATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT
 Mst44 + ACCCTGATCCCATGCAAGCTGGGTGAGTGTAGTTAATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT
 p30/57+1 AGCTGGGTGAGTGTAGTTAATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT
 p30/63+1 AGCTGGGTGAGTGTAGTTAATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT
 p30/6-2 AGTGGTAAAGGAAGGCCACAT
 Sequenz ACCCTGATCCCATGCAAGCTGGGTGAGTGTAGTTAATTTATCTATACTGGATCAAACCACTGCCATCTCTAGTAAATGGTAAAGGAAGGCCACAT

Mst65 - CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCT
 p30/63+1 CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCT
 Mst47 - CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAAGCTGTTTATAGGCCT
 Mst49 + CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAAGCTGTTTATAGGCCT
 p30/57+1 CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG
 p30/6-2 CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG
 p30/1-2 AGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG
 p30/9-2 AGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG
 p30/30+1 AGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG
 p30/62-2 AGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG
 p30/43+1 TAAGGCCTAACACGTCGGAATACTTTATCAG
 p30/7-1 AACACGTCGGAATACTTTATCAG
 Sequenz CTTATGATAGGTTAAATGGCAGTAAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAAGCTGTTTATAGGCCTAACACGTCGGAATACTTTATCAG

p30/57+1 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTT
 p30/6-2 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 p30/1-2 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 p30/9-2 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 p30/30+1 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 p30/62-2 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 p30/43+1 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 p30/7-1 AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG
 Mst70 - TTTTTTTTCTCTAG
 Sequenz AGCCAGCAACCCCTTACTGAATATAGATAGACTGGCAACAGATATTTCTAAATCAAGAAATGGTACAGGCACAAAATAGACTTC-TTTTTTTTCTCTAG

p30/6-2 AAGATATACCTAGCT
 p30/30+1 AAGATATACCTAGCT
 p30/62-2 AAGATATACCTAGCT
 p30/43+1 AAGATATACCTAGCT
 p30/7-1 AAGATATACCTAGCT
 p30/8-2 AGCTGCTGAGACAGGGAAGCAACAAGAGAGGTCATCTACCTTTAGAAATTGAGAG
 p30/8-1 AGCTGCTGAGACAGGGAAGCAACAAGAGAGGTCATCTACCTTTAGAAATTGAGAGACTGGAAGAGCAAA
 p30/1-2 AAGATATACCTAGCTGCTGAGACAGGGAAGCAACAAGAGAGGTCATCTACCTTTAGAAATTGAGAGACTGGAAGAGCAAAATCAAGGCTGGGCAGA
 p30/9-2 AAGATATACCTAGCTGCTGAGACAGGGAAGCAACAAGAGAGGTCATCTACCTTTAGAAATTGAGAGACTGGAAGAGCAAAATCAAGGCTGGGCAGA
 Mst70 - AAGATATACCTAGCTGCTGAGACAGGGAAGCAACAAGAGAGGTCATCTACCTTTAGAAATTGAGAGACTGGAAGAGCAAAATCAAGGCTGGGCAGA
 Mst43 + GGTCATCTACCTTTAGAAATTGAGAGACTGGAAGAGCAAAATCAAGGCTGGGCAGA
 Mst70 - CTGGGCAGA
 Sequenz AAGATATACCTAGCTGCTGAGACAGGGAAGCAACAAGAGAGGTCATCTACCTTTAGAAATTGAGAGACTGGAAGAGCAAAATCAAGGCTGGGCAGA

p30/1-2 AATCTGGATTAA
 p30/9-2 AATCTGGATTAAATAAGCT
 Mst43 + AATCTGGATTAAATAAGCTATCAAGGCTCTCTCAGAAACAAGGCAAGGATAAGCAGATAAAGACAAGAACAGGGTCTACACAAAATACCTTGTCT
 Mst70 - AATCTGGATTAAATAAGCTATCAAGGCTCTCTCAGAAACAAGGCAAGGATAAGCAGATAAAGACAAGAACAGGGTCTACACAAAATACCTTGTCT
 Mst70 - AATCTGGATTAAATAAGCTATCAAGGCTCTCTCAGAAACAAGGCAAGGATAAGCAGATAAAGACAAGAACAGGGTCTACACAAAATACCTTGTCT
 p30/68+1 AGCTATCAAGGCTCTCTCAGAAACAAGGCAAGGATAAGCAGATAAAGACAAGAACAGGGTCTACACAAAATACCTTGTCT
 Mst84 + TATCAAGGCTCTCTCAGAAACAAGGCAAGGATAAGCAGATAAAGACAAGAACAGGGTCTACACAAAATACCTTGTCT
 Sequenz AATCTGGATTAAATAAGCTATCAAGGCTCTCTCAGAAACAAGGCAAGGATAAGCAGATAAAGACAAGAACAGGGTCTACACAAAATACCTTGTCT

Mst39 -	CTTTGTAGG
Mst39 -	CTTTGTAGGCTAGGCTGGCCCTCGAATTCACAGAGATCCACCTGCCTCTGCTTCCCGAATGCTGGG
Mst39 -	CTTTGTAGGCTAGGCTGGCCCTCGAATTCACAGAGATCCACCTGCCTCTGCTTCCCGAATGCTGGG
Mst79 +	CTTTGTAGGCTAGGCTGGCCCTCGAATTCACAGAGATCCACCTGCCTCTGCTTCCCGAATGCTGGGATTAAAGGAGCACAAACCCTGCCTGGCTGCAATG
Mst14 -	CTTTGTAGGCTAGGCTGGCCCTCGAATTCACAGAGATCCACCTGCCTCTGCTTCCCGAATGCTGGGATTAAAGGAGCACAAACCCTGCCTGGCTGCAATG
Mst14 -	TCGAATTCACAGAGATCCACCTGCCTCTGCTTCCCGAATGCTGGGATTAAAGGAGCACAAACCCTGCCTGGCTGCAATG
Mst14 -	CAATG
Sequenz	CTTTGTAGGCTAGGCTGGCCCTCGAATTCACAGAGATCCACCTGCCTCTGCTTCCCGAATGCTGGGATTAAAGGAGCACAAACCCTGCCTGGCTGCAATG

Mst14 -	ACTAATCTTATGGTGGGGTCACCACAGCATGAAGAAGCTATATTAAGGGGTTGCAGCA
Mst79 +	ACTAATCTTATGGTGGGGTCACCACAGCATGAAGAAGCTATATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCCTCAGGGNCTCATAGATNC
Mst14 -	ACTAATCTTATGGTGGGGTCACCACAGCATGAAGAAGCTATATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCCTCAGGGGGCTCATAGATT
Mst14 -	ACTAATCTTATGGTGGGGTCACCACAGCATGAAGAAGCTATATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCCTCAGGGGGCTCATAGATT
Mst22 +	CAGCATGAAGAAGCTATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCCTCAGGGGGCTCATAGATT
Mst22 +	CAGCATGAAGAAGCTATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCCTCAGGGGGCTCATAGATT
Mst22 +	GTGGAGAACCACTGCTCCAGGGGGCTCATAGATT
Mst22 +	GAGAACCCTGCTCCAGGGGGCTCATAGATT
Sequenz	ACTAATCTTATGGTGGGGTCACCACAGCATGAAGAAGCTATATTAAGGGGTTGCAGCATTAGGAAGGTGGAGAACCCTGCCTCAGGGGGCTCATAGATT

Mst14 -	CAGGAGAT
Mst79 +	CAGGAGATGTATG
Mst14 -	CAGGAGATGTATGAAACTTACAAAACACTATTTATTTATTT
Mst22 +	CAGGAGATGTATGAAACTTACAAAACACTATTTATTTATTTTATTTATCAGATAGCTGGTGTGAGCATTGGTGTGTGAAGGCTTC
Mst22 +	CAGGAGATGTATGAAACTTACAAAACACTATTTATTTATTTTATCAGATAGCTGGTGTGAGCATTGGTGTGTGAAGGCTTC
Mst22 +	CAGGAGATGTATGAAACTTACAAAACACTATTTATTTATTTTATCAGATAGCTGGTGTGAGCATTGGTGTGTGAAGGCTTC
Mst22 +	CAGGAGATGTATGAAACTTACAAAACACTATTTATTTATTTTATCAGATAGCTGGTGTGAGCATTGGTGTGTGAAGGCTTC
Mst36 -	TGGTGTGAGCATTGGTGTGTGAAGGCTTC
Mst36 -	TGGTGTGAGCATTGGTGTGTGAAGGCTTC
Mst36 -	TGGTGTGAGCATTGGTGTGTGAAGGCTTC
Sequenz	CAGGAGATGTATGAAACTTACAAAACACTATTTATTTATTTTATCAGATAGCTGGTGTGAGCATTGGTGTGTGAAGGCTTC

Mst22 +	ATCTGCTGTCTGCACATTCAGTGGGGTGGTGTGACAGCAGGCTCCCTGCCATTCCCTGTTGCTGTTTTTACACAGCAGCTTTAGCTCTGCAGAAATGAA
Mst36 -	ATCTGCTGTCTGCACATTCAGTGGGGTGGTGTGACAGCAGGCTCCCTGCCATTCCCTGTTGCTGTTTTTACACAGCAGCTTTAGCTCTGCAGAAATGAA
Sequenz	ATCTGCTGTCTGCACATTCAGTGGGGTGGTGTGACAGCAGGCTCCCTGCCATTCCCTGTTGCTGTTTTTACACAGCAGCTTTAGCTCTGCAGAAATGAA

Basen, die nicht eindeutig zu lesen waren, sind in fett und kursiv dargestellt. Gelb hinterlegt sind Basen in den Sequenzierungen hinterlegt, die nicht mit anderen Sequenzierungen übereinstimmen. Die unterste Zeile zeigt die ermittelte Sequenz. Hier sind die Bereiche gelb hinterlegt, in denen nur 1 Strang (mehrfach) sequenziert wurde. Blau sind Sequenzabschnitte hinterlegt, die die angrenzenden Exonbereiche enthalten. Positionsangaben sind im Ergebnisteil unter D.1 gegeben.

Die erste Spalte gibt die Art der Sequenzierung an: '+' oder '-' zeigt eine Sequenz des Sinn- bzw. des Antisinnstranges. Entweder wurde direkt ein PCR-Produkt sequenziert, in diesem Fall steht der Name des Sequenzierprimers in der ersten Spalte, oder es wurde Insert aus Plasmiden sequenziert. Ein 'p' steht für pPBK-Plasmide (pPBK30-pPBK38). Sequenzierungen von Plasmiden wurden notiert durch die Plasmidbezeichnung mit angehängter Klon-Nr. und einer 1 bzw. 2 für den verwendeten Primer (Ap¹ oder Ap²). Das ATG-Startkodon ist in grünen Buchstaben gedruckt.

G.2.2 Sequenzierungen des Ratten HCII-Intron 1

Abbildung 57: Sequenzierungen des rHCII-Intron 1

Ragg	GCCTCTGGCTCATCTCTGCCACGCCACTGCTCAGAGGCACAGAGTAAGTTTGGTTACTAACATTTTTTGGCAATATAAAATTTTTAGTTATATTTAAAC
Sequenz	GCCTCTGGCTCATCTCTGCCACGCCACTGCTCAGAGGCACAGAGTAAGTTTGGTTACTAACATTTTTTGGCAATATAAAATTTTTAGTTATATTTAAAC

Ragg 2x +/-	CTAGATAATTTGGTTCCTTTATTAAGTACAAATTGAGAAAAAGATATATACATTAATTTTAAATTTTAAATTTTAAATTTTAAACTTAGTTTAAAT
p36/3+1	GTACAAATTGAGAAAAAGAAATATATCATTAAATTTTAAATTTTAAATTTTAAATTTTAAACTTAGTTTAAAT
Sequenz	CTAGATAATTTGGTTCCTTTATTAAGTACAAATTGAGAAAAAGAAATATATCATTAAATTTTAAATTTTAAATTTTAAACTTAGTTTAAAT

p36/3+1	TAGTTTAAATATTAATATTGGCATAACTGACAAAAGTGAACCTTGAATTTTAAAGAGACAAGAATTGTAC
Ragg 2x +/-	TAGTTTAAATATTAATATTGGCATAACTGACAAAAGTGAACCTTGAATTTTAAAGAGACAAGAGATTTGTACCTTTGATACTTTAAACTGGAGACTGGCAT
Mst4	GGCATAAAGTGAACCTTGAATTTTAAAGAGACAAGAGATTTGTACCTTTGATACTTTAAACTGGAGACTGGCAT
p36/11+1	GTACCTTTGATACTTTAAACTGGAGACTGGCAT
Sequenz	TAGTTTAAATATTAATATTGGCATAACTGACAAAAGTGAACCTTGAATTTTAAAGAGACAAGAGATTTGTACCTTTGATACTTTAAACTGGAGACTGGCAT

Ragg 2x +/-	GTAGCACAAGGCTGAAATTCAGCATGGGTAGCGCTGGGTATCCATCACCTCAGGGCACTTAGATGACTGTCTAGTGTCTAAGTATGACAATATTCCTGT
Mst4	GTAGCACAAGGCTGAAATTCAGCATGGGTAGCGCTGGGTATCCATCACCTCAGGGCACTTAGATGACTGTCTAGTGTCTAAGTATGACAATATTCCTGT
p36/11+1	GTAGCACAAGGCTGAAATTCAGCATGGGTAGCGCTGGGTATCCATCACCTCAGGGCACTTAGATGACTGTCTAGTGTCTAAGTATGACAATATTCCTGT
Sequenz	GTAGCACAAGGCTGAAATTCAGCATGGGTAGCGCTGGGTATCCATCACCTCAGGGCACTTAGATGACTGTCTAGTGTCTAAGTATGACAATATTCCTGT

Mst4	TCTCAGGAATTTCCCTAAGCATCACTATTAAGGTCGGAGGAATGTAAGA
p36/11+1	TCTCAGGAAGTTTCCCTAAGCATCACTATTAAGGTCGGAGGAATGTAAGAGTTCTAACCTCAAAGGCTCAGAAACAATGTAC
Ragg +/-	TCTCAGGAAGTTTCCCTAAGCATCACTATTAAGGTCGGAGGAATGTAAGAGTTCTAACCTCAAAGGCTCAGAAACAATGTACATATGTAGACATATA
Mst33 2x +	AAGTTTCCCTAAGCATCACTATTAAGGTCGGAGGAATGTAAGAGTTCTAACCTCAAAGGCTCAGAAACAATGTACATATGTAGACACATATA
Mst33 +	CTAAGCATCACTATTAAGGTCGGAGGAATGTAAGAGTTCTAACCTCAAAGGCTCAGAAACAATGTACATATGTAGACACATATA
Mst33 +	GGTCGGAGGAATGTAAGAGTTCTAACCTCAAAGGCTCAGAAACAATGTACATATGTAGACACATATA
Mst56 -	ACAATGTACATATGTAGACAAATATA
Sequenz	TCTCAGGAAGTTTCCCTAAGCATCACTATTAAGGTCGGAGGAATGTAAGAGTTCTAACCTCAAAGGCTCAGAAACAATGTACATATGTAGACACATATA

Mst33 +	CATGTACACACACAAGCACATATCACACCACACACACACACACACACA
Mst33 +	CATGTACACACACATGCACATATCACACCACACACACACACACACAAGANANA
Mst33 +	CATGTACACACACATGCACATATCACACCACACACACACACACACAAGGAGACACAGCGGCTGCACTGCTGATTTCTATTCTTTCAATTCTTCTG
Mst33 2x +	CATGTACACACACATGCACATATCACACCACACACACACACACACAAGGAGACACAGCGG-CTGCA-CTNTATTCTATTCTTTCAANTCTNCTGNA
Mst56 -	CATGTACACACACATGCACATATCACACCACACACACACACACAAGGAGACACAGCGG-CTGCA-CTGCTATTCTATTCTTTCAATTCTTCTGT
p36/10+2	GTACACACACATGCACATATCACACCACACACACACACACAAGGAGACACAGCGG-CTGCA-CTGCTATTCTATTCTTTCAATTCTTCTGT
Sequenz	CATGTACACACACATGCACATATCACACCACACACACACACAAGGAGACACAGCGG-CTGCA-CTGCTATTCTATTCTTTCAATTCTTCTGT

Mst56 -	AGATTGTAACATAAAAAGGTTTCCATGTTTCCCCCGCTTCTTGCT
Mst33 2x +	AGATTGTAACATAAAAAGNTTCCATGTTTCCCCCGCTTCTTGCTGCTGTGTAATCTGGAGACATCATGGCAGGAGATGGAGTTCTCTATGACAC
p36/10+2	AGATTGTANCATAAAAAGGTTTCCATGTTTCCCCCGCTTCTTGCTGCTGTGTAATCTGGAGACATCATGGCAGGAGATGGAGTTCTCTATGACACA
p36/10-1	TTGCTCTGTGTAATCTGGAGACANTATGGCAGGAGATGGAGTTCTCTATGACACA
Sequenz	AGATTGTAACATAAAAAGGTTTCCATGTTTCCCCCGCTTCTTGCTGCTGTGTAATCTGGAGACATCATGGCAGGAGATGGAGTTCTCTATGACACA

p36/10+2	CATACAAAAAGGCCAA-CTCAATCCATGCAACTCTC
Mst33 2x +	CATACAAAAAGGCCAA-CTCAATCCATGCAACTCTCAAGAATACACTATATGGTGTCCCTGAAAAATACAGCTTGGATGATTTACTTTAAAAAGCACAGCA
p36/10-1	CATACAAAAAGGCCAAAGCTCAATCCATGCAACTCTCAAGAATACACTATATGGTGTCCCTGAAAAATACAGCTTGGATGATTTACTTTAAAAAGCACAGCA
p36/22-1	TTNCTTTAAAGCNACGCN
Sequenz	CATACAAAAAGGCCAA-CTCAATCCATGCAACTCTCAAGAATACACTATATGGTGTCCCTGAAAAATACAGCTTGGATGATTTACTTTAAAAAGCACAGCA

Mst33 2x +	CTTCGGGCTGGAGAGATGGCCAGTGGTTAAGAGCACTGACTGCTTCTCCTGAGGNCTGAGTTCAAATCCCAGCAACCACATGGTGGCTCACACCATC
p36/10-1	CTTCGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCACTGACTGCTTCTCCTGAGTCCCTGAGTTCAAATCCCAGCAACCACATGGTGGCTCACACCATC
p36/22-1	CTTCGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCACTGACTGCTTCTCCTGAGTCCCTGAGTTCAAATCCCAGCAAACATGGTGGCTCACACCATC
Sequenz	CTTCGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCACTGACTGCTTCTCCTGAGTCCCTGAGTTCAAATCCCAGCAACCACATGGTGGCTCACACCATC

p36/10-1	TGTAATGAAATCTGATGCCCTCTTCTGGTGTGTCTGAAGGCAGCTACAGTGTAC
p36/22-1	TGTAATGAAATCTGATGCCCTCTTCTGGTGTGTCTGAAGGCAGCTACAGTGTAC
Mst33 2x +	TGTAATGAAATCTGATGCCCTCTTCTGGTGTGTCTGAAGGCANTACAGTGTACTCATAACAATAAAAATAAAATCTTAAAAAATAAAAAA
Mst57 +	ATGCCCTCTTCTGGTGTGTCTGAAGGCAGCTACAGTGTACTCATAACAATAAAAATAAAATCTTAAAAAATAAAAAA
Mst63 -	AAAAAATAAAAAAGCACAG
p37/1+2	CTACAGTGTACTCATAACAATAAAAATAAAATCTTAAAAAATAAAAAA
Sequenz	TGTAATGAAATCTGATGCCCTCTTCTGGTGTGTCTGAAGGCAGCTACAGTGTACTCATAACAATAAAAATAAAATCTTAAAAAATAAAAAA

Mst63 -	CACCTCAATTCAAAGACAGTAACCTGAATGCAAACACAAGTTACAGGTGTCTTTAAATGCCTTAATGTTTTATGAAACTGATGAGCAAATAGAAAAG
p37/1+2	CACCTCAATTCAAAGACAGTAACCTGAATGCAAACACAAGTTACAGGTGTCTTTAAATGCCTTAATGTTTTATGAAACTGATGAGCAAATAGAAAAG
p37/1-1	TCAATTCAAAGACAGTAACCTGAATGCAAACACAAGTTACAGGTGTCTTTAAATGCCTTAATGTTTTATGAAACTGATGAGCAAATAGAAAAG
MST52 -	AAAAA'
Sequenz	CACCTCAATTCAAAGACAGTAACCTGAATGCAAACACAAGTTACAGGTGTCTTTAAATGCCTTAATGTTTTATGAAACTGATGAGCAAATAGAAAAG

Mst63 -	TTACCTGTCAGATCTCTAATTTTTCAGTGCATAAAAATAACTTTATATTTTTCCCTTCCCTAAGAAGAGAATAATGAAACTATGGCTTATGGCTTTTGGGA
p37/1+2	TTACCTGTCAGATCTCTAATTTTTCAGTGCATAAAAATAACTTTATATTTTTCCCTTCCCTAAGAAGAGAATAATGAAACTATGGCTTATGGCTTTTGGGA
p37/1-1	TTACCTGTCAGATCTCTAATTTTTCAGTGCATAAAAATAACTTTATATTTTTCCCTTCCCTAAGAAGAGAATAATGAAACTATGGCTTATGGCTTTTGGGA
MST52 -	TTACCTGTCAGATCTCTAATTTTTCAGTGCATAAAAATAACTTTATATTTTTCCCTTCCCTAAGAAGAGAATAATGAAACTATGGCTTATGGCTTTTGGGA
MST75 +	AAGAAGAGAATAATGAAACTATGGCTTATGGCTTTTGGAT
Sequenz	TTACCTGTCAGATCTCTAATTTTTCAGTGCATAAAAATAACTTTATATTTTTCCCTTCCCTAAGAAGAGAATAATGAAACTATGGCTTATGGCTTTTGGGA

p37/1+2	ATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAG
Mst63 -	ATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
p37/1-1	ATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST52 -	ATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST75 +	ATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST66 +	AAAATGCTTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
Sequenz	ATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT

p37/1-1	TTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST52 -	TTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST52 -	TTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST75 +	TTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
MST66 +	TTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT
Sequenz	TTGGATATAGAATAATGAACTATGGCTCATGGCTTTTGGATATAGAATAATGAAACTATGGCTCATGGCTTTTGGATATAGAATAATGAACTATGGCTCATGGCT

MST75 +	GGA
p37/1-1	GGATAAGAAGAAGATAG
MST52 -	GGATAAGACGAAGAATAGCTTTACCACCTGGGCTCAGACTAGTCCCTTGAGCACACATTAACCTGAAGTTGGCTTTTCTTCTGTCTTTAATATTACTGT
MST52 -	GGATAAGACGAAGAATAGCTTTACCACCTGGGCTCAGACTAGTCCCTTGAGCACACATTAACCTGAAGTTGGCTTTTCTTCTGTCTTTAATATTACTGT
MST66 +	GGATAAGACGAAGAATAGCTTTACCACCTGGGCTCAGACTAGTCCCTTGAGCACACATTAACCTGAAGTTGGCTTTTCTTCTGTCTTTAATATTACTGT
Sequenz	GGATAAGACGAAGAATAGCTTTACCACCTGGGCTCAGACTAGTCCCTTGAGCACACATTAACCTGAAGTTGGCTTTTCTTCTGTCTTTAATATTACTGT

MST52 -	ACTAATGCTGTTTTAGCCAAGGGTACTTTAAATGTAATCATAAAAAACACTTCCATTAAGTTGCTTCTGCTCCTCTTCTTTTATAGAGAGATGTATT
MST52 -	ACTAATGCTGTTTTAGCCAAGGGTACTTTAAATGTAATCATAAAAAACACTTCCATTAAGTTGCTTCTGCTCCTCTTCTTTTATAGAGAGATGTATT
MST66 +	ACTAATGCTGTTTTAGCCAAGGGTACTTTAAATGTAATCATAAAAAACACTTCCATTAAGTTGCTTCTGCTCCTCTTCTTTTATAGAGAGATGTATT
Mst64 +	ACTAATGCTGTTTTAGCCAAGGGTACTTTAAATGTAATCATAAAAAACACTTCCATTAAGTTGCTTCTGCTCCTCTTCTTTTATAGAGAGATGTATT
Mst64 +	TAATGCTGTTTTAGCCAAGGGTACTTTAAATGTAATCATAAAAAACACTTCCATTAAGTTGCTTCTGCTCCTCTTCTTTTATAGAGAGATGTATT
Sequenz	ACTAATGCTGTTTTAGCCAAGGGTACTTTAAATGTAATCATAAAAAACACTTCCATTAAGTTGCTTCTGCTCCTCTTCTTTTATAGAGAGATGTATT

MST66 +	NCTTTTTATTGTG
MST52 -	TCTTTTTATTGTGAATATGCACTTGCTGTCTCAAGTGTGCACCACATGTGTGCAGATTGCCAGACAGTGTCAAACCCCTTGGAACTGGAGTTACAG
MST52 -	TCTTTTTATTGTGAATATGCTGTGTCTGTCAAGTGTGCACCACATGTGTGCAGATTGCCAGACAGTNTCAAACCCCTTGGAACTGGAGTTACAG
Mst64 +	TCTTTTTATTGTGAATATGCACTTGCTGTCTCAAGTGTGCACCACATGTGTGCAGATTGCCAGACAGTGTCAAACCCCTTGGAACTGGAGTTAAGA
Mst64 +	TCTTTTTATTGTGAATATGCACTTGCTGTCTCAAGTGTGCACCACATGTGTGCAGATTGCCAGACAGTGTCAAACCCCTTGGAACTGGAGTTACAG
Mst68 -	GTGTGCACCACATGTGTGCAGATTGCCAGACAGTGTCAAACCCCTTGGAACTGGAGTTACAG
Sequenz	TCTTTTTATTGTGAATATGCACTTGCTGTCTCAAGTGTGCACCACATGTGTGCAGATTGCCAGACAGTGTCAAACCCCTTGGAACTGGAGTTACAG

Mst52 - TGGCTGAGCCACCTACCTGGAGGGCTGGGACCT-GAACTCTTGT
 Mst64 + TGGCTGAGCCACCTACCTGGAGGGCTGGGACCT-GAACTCTTGTTCCTGAAAAGTAGGAAGTACTGAGTCTGAGCCCCCTCTCTAGCCCGTTACT
 Mst64 + TGGCTGAGCCACCTACCTGGAGGGCTGGGACCT-GAACTCTTGTTCCTGAAAAGTAGGAAGTACTGAGTCTGAGCCCCCTCTCTAGCCCGTTACT
 Mst68 - TGGCTGAGCCACCTACCTGGAGGGCTGGGACCT-GAACTCTTGTTCCTGAAAAGTAGGAAGTACTGAGTCTGAGCCCCCTCTCTAGCCCGTTACT
 p36/22+1 TGGCTGAGCCACCTACCTGGAGGGCTGGGACCT-GAACTCTTGTTCCTGAAAAGTAGGAAGTACTGAGTCTGAGCCCCCTCTCTAGCCCGTTACT
 Sequenz TGGCTGAGCCACCTACCTGGAGGGCTGGGACCT-GAACTCTTGTTCCTGAAAAGTAGGAAGTACTGAGTCTGAGCCCCCTCTCTAGCCCGTTACT

p36/22+1 TCAGTGACACCCATGCTGATTCTGTTGAAATAGCTCTGGTTGAGCTGCATACATTAGGT
 Mst64 + TCAGTGACACCCATGCTGATTCTGTTGAAATAGCTCTGGTTGAGCTGCATACATTAGGTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 Mst64 + TCAGTGACACCCATGCTGATTCTGTTGAAATAGCTCTGGTTGAGCTGCATACATTAGGTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 Mst68 - TCAGTGACACCCATGCTGATTCTGTTGAAATAGCTCTGGTTGAGCTGCATACATTAGGTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 p32/14-2 CTCTGGTTGAGCTGCATACATTAGGTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 p32/11+1 CTCTGGTTGAGCTGCATACATTAGGTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 p36/13-1 GTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 p36/22+2 GTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 p32/11-2 CACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG
 Sequenz TCAGTGACACCCATGCTGATTCTGTTGAAATAGCTCTGGTTGAGCTGCATACATTAGGTACTGTTTCCTAGAAAACCTGTTCCACATGTGCAGATTATG

Mst64 + AGCACCGG
 Mst68 - AGCACCGGTCAGCAAAGTCTGAGGAGG
 p32/14-2 AGCACCGGTCAGCAAAGTCTGAGGAGGCTCAGGCATATTCAGAATCAGAAAACCTGTGAGTCTTTTTCCAAAATAAGCACTTCTGAAGCCCCGAG
 p32/11+1 AGCACCGGTCAGCAAAGTCTGAGGAGGCTCAGGCATATTCAGAATCAGAAAACCTGTGAGTCTTTTTCCAAAATAAGCACTTCTGAAGCCCCGAG
 p36/13-1 AGCACCGGTCAGCAAAGTCTGAGGAGGCTCAGGCATATTCAGAATCAGAAAACCTGTGAGTCTTTTTCCAAAATAAGCACTTCTGAAGCCCCGAG
 p36/22+2 AGCACCGGTCAGCAAAGTCTGAGGAGGCTCAGGCATATTCAGAATCAGAAAACCTGTGAGTCTTTTTCCAAAATAAGCACTTCTGAAGCCCCGAG
 p32/11-2 AGCACCGGTCAGCAAAGTCTGAGGAGGCTCAGGCATATTCAGAATCAGAAAACCTGTGAGTCTTTTTCCAAAATAAGCACTTCTGAAGCCCCGAG
 Sequenz AGCACCGGTCAGCAAAGTCTGAGGAGGCTCAGGCATATTCAGAATCAGAAAACCTGTGAGTCTTTTTCCAAAATAAGCACTTCTGAAGCCCCGAG

p32/11-2 CAGCTGTAGATATAGC
 p32/14-2 CAGCTGTAGATATAGCTAGGAGAGCGCCCTGCTAGCGCTGAGAGCC
 p36/13-1 CAGCTGTAGATATAGCTAGGAGAGCGCCCTGCTAGCGCTGAGAGCCCTGCTCTCAC
 p36/22+2 CAGCTGTAGATATAGCTAGGAGAGCGCCCTGCTAGCGCTGAGAGCCCTGCTCAGCACTGT
 Mst68 - TCTCACCCAGCACTGTACAGAATAGTTACCTGATACAGTTCTACAAGT
 p36/9+1 GTACAGAATAGTTACCGTGATACAGTTCTACAAGT
 p32/11+1 CAGCTGTAGATATAGCTAGGAGAGCGCCCTGCTAGCGCTGAGAGCCCTGCTCACCCAGCACTGTACAGAATAGTTACCGTGATACAGTTCTACAAGT
 Sequenz CAGCTGTAGATATAGCTAGGAGAGCGCCCTGCTAGCGCTGAGAGCCCTGCTCACCCAGCACTGTACAGAATAGTTACCGTGATACAGTTCTACAAGT

p32/11+1 AGGAGGTATAGTAAAAGGGTCAGAACTTAAGGTCAGCCTTGGGATACAAGATTTTCAGCTGGCTGAGATGCTCAATGCTTTTTAA
 Mst68 - AGGAGGTATAGTAAAAGGGTCAGAACTTAAGGTCAGCCTTGGGATACAAGATTTTCAGCTGGCTGAGATGCTCAATGCTTTTTAA
 p36/9+1 AGGAGGTATAGTAAAAGGGTCAGAACTTAAGGTCAGCCTTGGGATACAAGATTTTCAGCTGGCTGAGATGCTCAATGCTTTTTAA
 p35/6+1 AGGAGGTATAGTAAAAGGGTCAGAACTTAAGGTCAGCCTTGGGATACAAGATTTTCAGCTGGCTGAGATGCTCAATGCTTTTTAA
 Mst61 + TTTTAAAAGTCCCTC
 Sequenz AGGAGGTATAGTAAAAGGGTCAGAACTTAAGGTCAGCCTTGGGATACAAGATTTTCAGCTGGCTGAGATGCTCAATGCTTTTTAAAGTCCCTC

Mst68 - GAAATAAAACCATTGTTAGGGCCAGGAGGTGGAATACCACGCCCTTAATCCCAGTCTTGGGAAAGAGA
 p36/9+1 GAAATAAAACCATTGTTAGGGCC
 p35/6+1 GAAATAAAACCATTGTTAGGGCC
 p35/7+2 GGCAGGAGGTGGAAGACCACGCCCTTAATCCCAGTCTTGGGAAAGTAGAAGAGCTGATCT
 p35/6-2 GGCAGGAGGTGGAAGACCACGCCCTTAATCCCAGTCTTGGGAAAGTAGAAGAGCTGATCTTACAAGTTCAACCAGCC
 p35/7+1 GGCAGGAGGTGGAATACCACGCCCTTAATCCCAGTCTTGGGAAAGTAGAAGAGCTGATCTTACAAGTTCAACCAGCC
 Mst61 + GAAATAAAACCATTGTTAGGGCCAGGAGGTGGAATACCACGCCCTTAATCCCAGTCTTGGGAAAGTAGAAGAGCTGATCTTACAAGTTCAACCAGCC
 Sequenz GAAATAAAACCATTGTTAGGGCCAGGAGGTGGAATACCACGCCCTTAATCCCAGTCTTGGGAAAGTAGAAGAGCTGATCTTACAAGTTCAACCAGCC

p35/6-2 TGGTCTAAAGAGCCAGTCCAGGACAGTCAAGGCTGTTGAGAAAACCTTGTCTCTAAAAAATCAAATCAAACAAAACAAAACAAAAGCTGGGAATGCTAA
 p35/7+1 TGGTCTAAAGAGCCAGTCCAGGACAGTCAAGGCTGTTGAGAAAACCTTGTCTCTAAAAAATCAAATCAAACAAAACAAAACAAAAGCTGGGAATGCTAA
 Mst61 + TGGTCTAAAGAGCCAGTCCAGGACAGTCAAGGCTGTTGAGAAAACCTTGTCTCTAAAAAATCAAATCAAACAAAACAAAACAAAAGCTGGGAATGCTAA
 Mst65 - TGGTCTAAAGAGCCAGTCCAGGACAGTCAAGGCTGTTGAGAAAACCTTGTCTCTAAAAAATCAAATCAAACAAAACAAAACAAAAGCTGGGAATGCTAA
 Sequenz TGGTCTAAAGAGCCAGTCCAGGACAGTCAAGGCTGTTGAGAAAACCTTGTCTCTAAAAAATCAAATCAAACAAAACAAAACAAAAGCTGGGAATGCTAA

p35/6-2 TTTTAGCCACATTGTTAGGGGATAGG
 p35/7+1 TTTTAGCCACATTGTTAGGGGATAGG
 Mst61 + TTTTAGCCACATTGTTAGGGGATAGGCTTAAGGACAAAGATTGACAACAGCTGCTCCCAAGCGTGAGCAGTTCTGAGTGTCTGGCACTCTACGGCAAG
 Mst65 - TTTTAGCCACATTGTTAGGGGATAGGCTTAAGGACAAAGATTGACAACAGCTGCTCCCAAGCGTGAGCAGTTCTGAGTGTCTGGCACTCTACGGCAAG
 Mst62 - ACAGTCTGAGTGTCTTGCCTCTACGGCAAG
 Sequenz TTTTAGCCACATTGTTAGGGGATAGGCTTAAGGACAAAGATTGACAACAGCTGCTCCCAAGCGTGAGCAGTTCTGAGTGTCTGGCACTCTACGGCAAG

Mst61 + CAAGGCAGCCCTTGTGAGACCAGGACCTGGTTAAACCAACCTGGTTCAAATCTAGTATTATCTACCATCCCTCAATCCCTACATTTTAGAGCTAGGGA
 Mst65 - CAAGGCAGCCCTTGTGAGACCAGGACCTGGTTAAACCAACCTGGTTCAAATCTAGTATTATCTACCATCCCTCAATCCCTACATTTTAGAGCTAGGGA
 Mst62 - CAAGGCAGCCCTTGTGAGACCAGGACCTGGTTAAACCAACCTGGTTCAAATCTAGTATTATCTACCATCCCTCAATCCCTACATTTTAGAGCTAGGGA
 Sequenz CAAGGCAGCCCTTGTGAGACCAGGACCTGGTTAAACCAACCTGGTTCAAATCTAGTATTATCTACCATCCCTCAATCCCTACATTTTAGAGCTAGGGA

Mst61 + CTGAACCCTGGTCCCATGCAATCTGTGTGAGAGTAGTTAATATTATCTATATTCTGGCTCAA-C-ACTGCCATCTCTTAGTAGCT
 Mst65 - CTGAACCCTGGTCCCATGCAATCTGTGTGAGAGTAGTTAATATTATCTATATTCTGGCTCAAACCCTGCTCTTAGTAGCTGCAAAAGGAAGGCT
 Mst62 - CTGAACCCTGGTCCCATGCAATCTGTGTGAGAGTAGTTAATATTATCTATATTCTGGCTCAAACCCTGCTCTTAGTAGCTGCAAAAGGAAGGCT
 Mst69 + TAATATTATCTATATTCTGGCTCAAACCCTGCTCTTAGTAGCTGCAAAAGGAAGGCT
 p35/11-1 TCTCTTAGTAGCTGCAAAAGGAAGGCT
 Sequenz CTGAACCCTGGTCCCATGCAATCTGTGTGAGAGTAGTTAATATTATCTATATTCTGGCTCAAACCCTGCTCTTAGTAGCTGCAAAAGGAAGGCT

Mst65 - ACATCTCATGATAGGTTAAAATCGTGGCAGAATTTCAA
 Mst62 - ACATCTCATGATAGGTTAAAATCGTGGCAGAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAACTGCTCTCAGGTCAAACAGGCTGTAATACTTTA
 Mst69 + ACATCTCATGATAGGTTAAAATCGTGGCAGAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAACTGCTCTCAGGTCAAACAGGCTGTAATACTTTA
 p35/11-1 ACATCTCATGATAGGTTAAAATCGTGGCAGAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAACTGCTCTCAGGTCAAACAGGCTGTAATACTTTA
 Sequenz ACATCTCATGATAGGTTAAAATCGTGGCAGAATTTCAAATAGAGCATCTGCAAAGCTAGTCCCTAACTGCTCTCAGGTCAAACAGGCTGTAATACTTTA

Mst62 - TAGGAGCCAGCAACCCCTTAATGAATATAGACTGGCAACAGATATTCTAAACCAAGAAATGGTCCACAGGCACAAAACAGACTTCTCTTTTTCCTAGAAGA
 Mst69 + TAGGAGCCAGCAACCCCTTAATGAATATAGACTGGCAACAGATATTCTAAACCAAGAAATGGTCCACAGGCACAAAACAGACTTCTCTTTTTCCTAGAAGA
 p35/11-1 TAGGAGCCAGCAACCCCTTAATGAATATAGACTGGCAACAGATATTCTAAACCAAGAAATGGTCCACAGGCACAAAACAGACTTCTCTTTTTCCTAGAAGA
 Mst70 - TTTCTTTTTCCTAGAAGA
 Sequenz TAGGAGCCAGCAACCCCTTAATGAATATAGACTGGCAACAGATATTCTAAACCAAGAAATGGTCCACAGGCACAAAACAGACTTCTCTTTTTCCTAGAAGA

p35/11-1	ATGCCTAGCTGCTGAGACAGGGAAGCA
Mst62 -	ATGCCTAGCTGCTGAGACAGGGAAGCAATAAGAGAGGCCATCTACCTCTTTAGTAAGTGAAGACCCGACTGGAA
Mst69 +	ATGCCTAGCTGCTGAGACAGGGAAGCAATAAGAGAGGCCATCTACCTCTTTAGTAAGTGAAGACCCGACTGGAAAGGAAATCAAAGCC
Mst70 -	ATGCCTAGCTGCTGAGACAGGGAAGCAATAAGAGAGGCCATCTACCTCTTTAGTAAGTGAAGACCCGACTGGAAAGGAAATCAAAGCCACTGGGCAGAAA
Mst43 + 2x	CTAGCTGCTGAGACAGGGAAGCAATAAGAGAGGCCATCTACCTCTTTAGTAAGTGAAGACCCGACTGGAAAGGAAATCAAAGCCACTGGGCAGAAA
Sequenz	ATGCCTAGCTGCTGAGACAGGGAAGCAATAAGAGAGGCCATCTACCTCTTTAGTAAGTGAAGACCCGACTGGAAAGGAAATCAAAGCCACTGGGCAGAAA

Mst70 -	AGCTGGATCTGATAAGTCATCAAGTTCTCTCAA AAAACAAGGCAAGGGATGAGCAGATAAAGAACATAAAATACCCCACTGTCCACTGTCCACTGT
Mst43 + 2x	AGCTGGATCTGATAAGTCATCAAGTTCTCTCAA AAAACAAGGCAAGGGATGAGCAGATAAAGAACATAAAATACCCCACTGTCCACTGTCCACTGT
Sequenz	AGCTGGATCTGATAAGTCATCAAGTTCTCTCAA AAAACAAGGCAAGGGATGAGCAGATAAAGAACATAAAATACCCCACTGTCCACTGTCCACTGT

Mst70 -	TACCTGTCTGTCTGTCTACA-GACCTCTACAAA-GAGTGGGCTGACAGC
Mst43 + 2x	TGCCTGTCTGTCTGTCTACA-GACCTCTACAAA-GAGTGGGCTGACAGCACCCTCCCGCAGACCCCTGTGACAGTGAAGTGAAGAGGAGGTAAGAAA
Ragg	TACAAGACCTCTACAAAAGAGTGGGCTGACAGCACCCTCCCGCAGACCCCTGTGACACAGTGAAGTGAAGAGGAGGTAAGAAA
Mill54 - 2x	TACA-GACCTCTACAAA-GAGTGGGCTGACAGCACCCTCCCGCAGACCCCTGTGACAGTGAAGTGAAGAGGAGGTAAGAAA
Sequenz	TGCCTGTCTGTCTGTCTACA-GACCTCTACAAA-GAGTGGGCTGACAGCACCCTCCCGCAGACCCCTGTGACACAGTGAAGTGAAGAGGAGGTAAGAAA

Mst43 +	AGATTTTCTTTCTCTGACAGTAAATCTCTCCCTACAGCAA
Ragg 2/3x	AGATTTTCTTTCTCTGACAGTAAATCTCTCCCTACAGCAAACATTAAACCAATAGCCCTTGAACACAGGAGAGACACGGGCCCCAGCCAAAC
Mill54 - 2x	AGATTTTCTTTCTCTGACAGTAAATCTCTCCCTACAGCAAACATTAAACCAATAGCCCTTGAACACAGGAGAGACACGGGCCCCAGCCAAAC
Sequenz	AGATTTTCTTTCTCTGACAGTAAATCTCTCCCTACAGCAAACATTAAACCAATAGCCCTTGAACACAGGAGAGACACGGGCCCCAGCCAAAC

Ragg 2/3x	CCTTCCAGTGGGCTAAGAGAGTCTTGCCAAATCCAAGACTTAGACAATAACCAAAATGACTGCACAGAGCAGCAATCTTTATGATATGTTTCT
Mill54 - 2x	CCTTCCAGTGGGCTAAGAGAGTCTTGCCAAATCCAAGACTTAGACAATAACCAAAATGACTGCACAGAGCAGCAATCTTTATGATATGTTTCT
Sequenz	CCTTCCAGTGGGCTAAGAGAGTCTTGCCAAATCCAAGACTTAGACAATAACCAAAATGACTGCACAGAGCAGCAATCTTTATGATATGTTTCT

Mill54 - 2x	GAAGGCTGGTTCATGTTTCATTCACAATGTCCTCTCTCCCT
Ragg 2/3x	GAAGGCTGGTTCATGTTTCATTCACAATGTCCTCTCTCCCTCAGTGAACAGTCTTGCTGAGCAGGAGAGACACGGGCCCCAGCCAAAC
Sequenz	GAAGGCTGGTTCATGTTTCATTCACAATGTCCTCTCTCCCTCAGTGAACAGTCTTGCTGAGCAGGAGAGACACGGGCCCCAGCCAAAC

Ragg 2/3x	AGGTATCACTTAGTCAACTCTTATGAAACCTCTTGGACTCTCTGGGCACCCACCCTCTCCCTACTTGTCCATGAGTCTGTGCCCTTCAGCAGGCAG
Dag83a +	TAGTCAACTCTTATGAAACCTCTTGGACTCTCTGGGCACCCACCCTCTCCCTACTTGTCCATGAGTCTGTGCCCTTCAGCAGGCAG
p31/16+1	AGTCAACTCTTATGAAACCTCTTGGACTCTCTGGGCACCCACCCTCTCCCTACTTGTCCATGAGTCTGTGCCCTTCAGCAGGCAG
Dag83a +	CCCATGAGTCTGTGCCCTTCAGCAGGCAG
Sequenz	AGGTATCACTTAGTCAACTCTTATGAAACCTCTTGGACTCTCTGGGCACCCACCCTCTCCCTACTTGTCCATGAGTCTGTGCCCTTCAGCAGGCAG

Dag83a +	ACGTCATGTCCTCTGGTCCAGCCCTGTTTTCTGGAGACAGA
Dag83a +	ACGTCAATGTCCTCTGGTCCAGCCCTGTTTTCTGGAGACAGATGATGATGTTGGTGA AAAAGAGACTCTGGGCTCTCCTGGGTCCCTCTTTTCCACCA
Dag83a +	ACGTCAATGTCCTCTGGTCCAGCCCTGTTTTCTGGAGACAGATGATGATGTTGGTGA AAAAGAGACTCTGGGCTCTCCTGGGTCCCTCTTTTCCACC
Ragg 2/3x	ACGTCATGTCCTCTGGTCCAGCCCTGTTTTCTGGAGACAGATGATGATGTTGGTGA AAAAGAGACTCTGGGCTCTCCTGGGTCCCTCTTTTCCACC
p31/16+1	ACGTCATGTCCTCTGGTCCAGCCCTGTTTTCTGGAGACAGATGATGATGTTGGTGA AAAAGAGACTCTGGGCTCTCCTGGGTCCCTCTTTTCCACC
Mst71 -	TCTCTGGTCCAGCCCTGTTTTCTGGAGACAGATGATGATGTTGGTGA AAAAGAGACTCTGGGCTCTCCTGGGTCCCTCTTTTCCACC
Sequenz	ACGTCATGTCCTCTGGTCCAGCCCTGTTTTCTGGAGACAGATGATGATGTTGGTGA AAAAGAGACTCTGGGCTCTCCTGGGTCCCTCTTTTCCACC

Dag83 -	TTAAA
Mst71 -	TTAAAAGAGATGTTCTTACATCAGTGAACAGCCAGCAGCGAAGGCGCTACTTGTGCAAGCAGCGCTACCTGAG
Ragg 2x	TTAAAAGAGATGTTCTTACATCAGTGAACAGCCAGCAGCGAAGGCGCTACTTGTGCAAGCAGCGCTACCTGAGCTTATCTCCACCACCCATAAC
p31/16+1	TTAAAAGAGATGTTCTTACATCAGTGAACAGCCAGCAGCGAAGGCGCTACTTGTGCAAGCAGCGCTACCTGAGCTTATCTCCACCACCCATAAC
Mst42 +	CTTCTATCTCCACCACCCACATAA
Sequenz	TTAAAAGAGATGTTCTTACATCAGTGAACAGCCAGCAGCGAAGGCGCTACTTGTGCAAGCAGCGCTACCTGAGCTTATCTCCACCACCCACATAA

Ragg 2x	GGTGAAG-AACCAACCCATCTGACTGTGGCATAATGTTTACATGTTGTGTG
p38/24+1	GGTGAAG-AACCAACCCATCTGACTGTGGCATAATGTTTACATGTTGTGTGCAC--ACCACACACACACACACACACACACACACACAC
Mst42 +	GGTGAAG-AACCAACCCATCTGACTGTGGCATAATGTTTACATGTTGTGTGCAC--ACCACACACACACACACACACACACACACACACACAC
p31/16+1	GGTGAAG-AACCAACCCATCTGACTGTGGCATAATGTTTACATGTTGTGTGCAC--ACCACACACACACACACACACACACACACACACACAC
p38/20-1	GGTGAAG-AACCAACCCATCTGAGTGTGGCATAATGTTTACATGTTGTGTGCAC--ACCACACACACACACACACACACACACACACACACAC
p38/1+2	AC
Sequenz	GGTGAAG-AACCAACCCATCTGACTGTGGCATAATGTTTACATGTTGTGTGCAC--ACCACACACACACACACACACACACACACACACACACACAC

p38/1+2	CACACACCAC
p38/20-1	CACACACCAC
Mst41 -	ANANANACATATGGGGTAGGGGT
Mst41 -	CATATGGGGTAGGGGTGGTGGAGTGGGG
p31/16+1	CACACACCAC
Dag88 -	CATATGGGGTAGGGGTGGTGGAGTGGGG
Mst39 -	ACATATGGGGTAGGGGTGGTGGAGTGGGG
p31/18-1	CACACACACACACACATATGGGGTAGGGGTGGTGGAGTGGGG
Sequenz	CACACACCAC

p31/16+1	GTGGTGGGAGGGTTGGAATCATGTCCTCCTGGGTCTCTGCTCTCTGTGTAGCCCTGGCTGTCTGGAACCTCGATTTGTAGAACAGGCTGGCCTCGAAC
Dag88 -	GTGGTGGGAGGGTTGGAATCATGTCCTCCTGGGTCTCTGCTCTCTGTGTAGCCCTGGCTGTCTGGAACCTCGATTTGTAGAACAGGCTGGCCTCGAAC
Mst39 -	GTGGTGGGAGGGTTGGAATCATGTCCTCCTGGGTCTCTGCTCTCTGTGTAGCCCTGGCTGTCTGGAACCTCGAATTTGTAGAACAGGCTGGCCTCGAAC
p31/18-1	GTGGTGGGAGGGTTGGAATCATGTCCTCCTGGGTCTCTGCTCTCTGTGTAGCCCTGGCTGTCTGGAACCTCGATTTGTAGAACAGGCTGGCCTCGAAC
Ragg/Mst39-	TT-GTAGAACAGGCTGGCTGGCCTCGAAC
Dag88 -	GTAGAACAGGCTGGCTGGCCTCGAAC
Sequenz	GTGGTGGGAGGGTTGGAATCATGTCCTCCTGGGTCTCTGCTCTCTGTGTAGCCCTGGCTGTCTGGAACCTCGATTTGTAGAACAGGCTGGCCTCGAAC

Dag88 -	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA
p31/16+1	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA
Dag88 -	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA
Dag88 -	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA
p31/18-1	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA
Ragg/Mst39-	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA
Dag88 -	AGGGTCAAC
Sequenz	TCACAGAGATCTGCTGCTCCTCCCAAGCACTGGGATTAAGGTTGTGACCACTGCCCGGCTGCAGTGAATAAATTTCTGGATGGGGGTACCA

Dag88 -	AACACGAGAAACTGTATTAAAGGGTTGCAGCATGAGGAAGGTGGAGAAGCACTGCTCTAGGGGCTCATAGATCC
p31/16+1	AACACGAGAAACTGTATTAAAGGGTTGCAGCATGAGGAAGGTGGAGAAGCACTGCTCTAGGGGCTCATAGATCCAGGAGATGTATGAAACTTACACACA
Dag88 -	AACACGAGAAACTGTATTAAAGGGTTGCAGCATGAGGAAGGTGGAGAAGCACTGCTCTAGGGGCTCATAGATCCAGGAGATGTATGAAACTTACACACA
p31/18-1	AACACGAGAAACTGTATTAAAGGGTTGCAGCATGAGGAAGGTGGAGAAGCACTGCTCTAGGGGCTCATAGATCCAGGAGATGTATGAAACTTACACACA
Ragg/Mst39-	AACACGAGAAACTGTATTAAAGGGTTGCAGCAGGAGGAAGGTGGAGAAGCTCTGCTCTAGGGGCTCATAGATCCAGGAGATGTATGAAACTTACACACAC
Dag88 -	AACAAGAGAAACTGTATTAAAGGGTTGCAGCATGAGGAAGGTGGAGAAGCACTGCTCTAGGGGCTCATAGATCCAGGAGATGTATGAAACTTACACACAC
Sequenz	AACACGAGAAACTGTATTAAAGGGTTGCAGCATGAGGAAGGTGGAGAAGCACTGCTCTAGGGGCTCATAGATCCAGGAGATGTATGAAACTTACACACA

Dag88 -	TCACACT
Dag88 -	TCACACTGGTGTGAGCCGTGGTGTATCAAGTTAGACCTGAGAGAGCCCACTGCAGGGTGCAGCAG-CCA
p31/16+1	TCACACTGGTGTGAGCCGTGGTGTATCAAGTTAGACCTGAGAGAGCCCACTGCAGGGTGGCAGCAG-CCAGGCTGCTGGTGCTC-ACTGTTGTCTGTT
p31/18-1	TCACACTGGTGTGAGCCGTGGTGTATCAAGTTAGACCTGAGAGAGCCCACTGCAGGGTGGCAGCAG-CCAGGCTGCTGGTGCTC-ACTGTTGTCTGTT
Ragg	NCACNNTTTGTGAGCCGTGGTGTATCAAGTTAGACCTGAGAGAGCCCACTGCAGGGTCA-CAGCCAGGCTGCCTGGTGCTCCTACTGTTGTCTGTT
Sequenz	TCACACTGGTGTGAGCCGTGGTGTATCAAGTTAGACCTGAGAGAGCCCACTGCAGGGTGGCAGCAG-CCAGGCTGCCTGGTGCTC-ACTGTTGTCTGTT

p31/16+1	TTACAGCGGCTTTAGCTCTGCAGAAATG
p31/18-1	TTACAGCGGCTTTAGCTCTGCAGAAATG
Ragg	TTACAGCGGCTTTAGCTCTGCAGAAATG
Sequenz	TTACAGCGGCTTTAGCTCTGCAGAAATG

Basen, die nicht eindeutig zu lesen waren, sind in fett und kursiv dargestellt. Gelb hinterlegt sind Basen in den Sequenzierungen hinterlegt, die nicht mit anderen Sequenzierungen übereinstimmen. Die unterste Zeile zeigt die ermittelte Sequenz. Hier sind die Bereiche gelb hinterlegt, in denen nur 1 Strang (mehrfach) sequenziert wurde. Blau sind Sequenzabschnitte hinterlegt, die die angrenzenden Exonbereiche enthalten. Positionsangaben sind im Ergebnisteil unter D.1 gegeben.

Die erste Spalte gibt die Art der Sequenzierung an: '+' oder '-' zeigt eine Sequenz des Sinn- bzw. des Antisinnstranges. Entweder wurde direkt ein PCR-Produkt sequenziert, in diesem Fall steht der Name des Sequenzierprimers in der ersten Spalte, oder es wurde Insert aus Plasmiden sequenziert. Ein 'p' steht für pPBK-Plasmide (pPBK30-pPBK38). Sequenzierungen von Plasmiden wurden notiert durch die Plasmidbezeichnung mit angehängter Klon-Nr. und einer 1 bzw. 2 für den verwendeten Primer (Ap^f1 oder Ap^f2).

Bildungswegdegang

Persönliche Daten

Name: Paul-Bertram Kamp
Anschrift: Nansenstr. 13, 33790 Halle
Geburtsdatum/-ort: 13.03.1969 in Halle
Staatsangehörigkeit: deutsch
Familienstand: ledig

Schulausbildung

Grundschule Halle-Ost: 1975 - 1979
Kreisgymnasium Halle: 1979 - 1988
Abitur: 25.6.1988

Zivildienst und Praktikum

Praktikum im Herzkatheterlabor
der Schüchtermannklinik: 15.07. - 31.11.1988
Zivildienst im Herzkatheterlabor
der Schüchtermannklinik: 01.12.1988 - 31.07.1990

Hochschulstudium

Studium an der Universität Bielefeld,
Fakultät für Biologie: Oktober 1990 - Oktober 1995

Diplomarbeit „Transkriptionale und posttranslationale Regulation der Nitrogenase in *Rhodobacter capsulatus*“ in der Arbeitsgruppe von Prof. Dr. W. Klipp

Diplom der Fakultät für Biologie,
Universität Bielefeld: 06.10.1995
Promotion an der Universität Bielefeld
(Technischen Fakultät,
AG Zellkulturtechnik) unter der Leitung
von Prof. Dr. H. Ragg: Oktober 1995 - Dezember 1999

Praktika während des Studiums

Praktikum in der Arbeitsgruppe
von Prof. Dr. W. Klipp: 16.8.1993 - 8.10.1993
Praktikum bei der
„ASTA Medica“ in Halle: 15.11.1993 - 28.02.1994