

4 Ergebnisse und Diskussion

4.1 Identifizierung von LRC-homologen Sequenzen bei verschiedenen Spezies

4.1.1 Radioaktive Hybridisierung mit einer LILR-Sonde

Im Rahmen der Analyse des LRC sollte auch nach homologen Genen von LILRs bei verschiedenen Tierarten gesucht werden. Hierzu wurden Southern-Blot-Membranen von EcoRI-verdauter genomischer DNA verschiedener Spezies („Multispezies-Membranen“) hergestellt. Eine humane LILRA2 (ILT1)-cDNA wurde radioaktiv markiert und als Sonde zur Hybridisierung einer dieser Multispezies-Membranen eingesetzt. Das resultierende Signal wurde mit Hilfe eines Röntgenfilmes detektiert (Abb. 6). Wie erwartet zeigten sich besonders starke Hybridisierungssignale bereits nach 24 Stunden beim Menschen und verschiedenen Primaten. Darüber hinaus konnten mehrere schwächere Signale durch lange Expositionsdauer (14 Tage) auch bei Rind, Schwein, Hamster, Hund und Huhn festgestellt werden. Im Gegensatz dazu zeigten Kaninchen, Ratte und Maus nur wenige hybridisierte Fragmente.

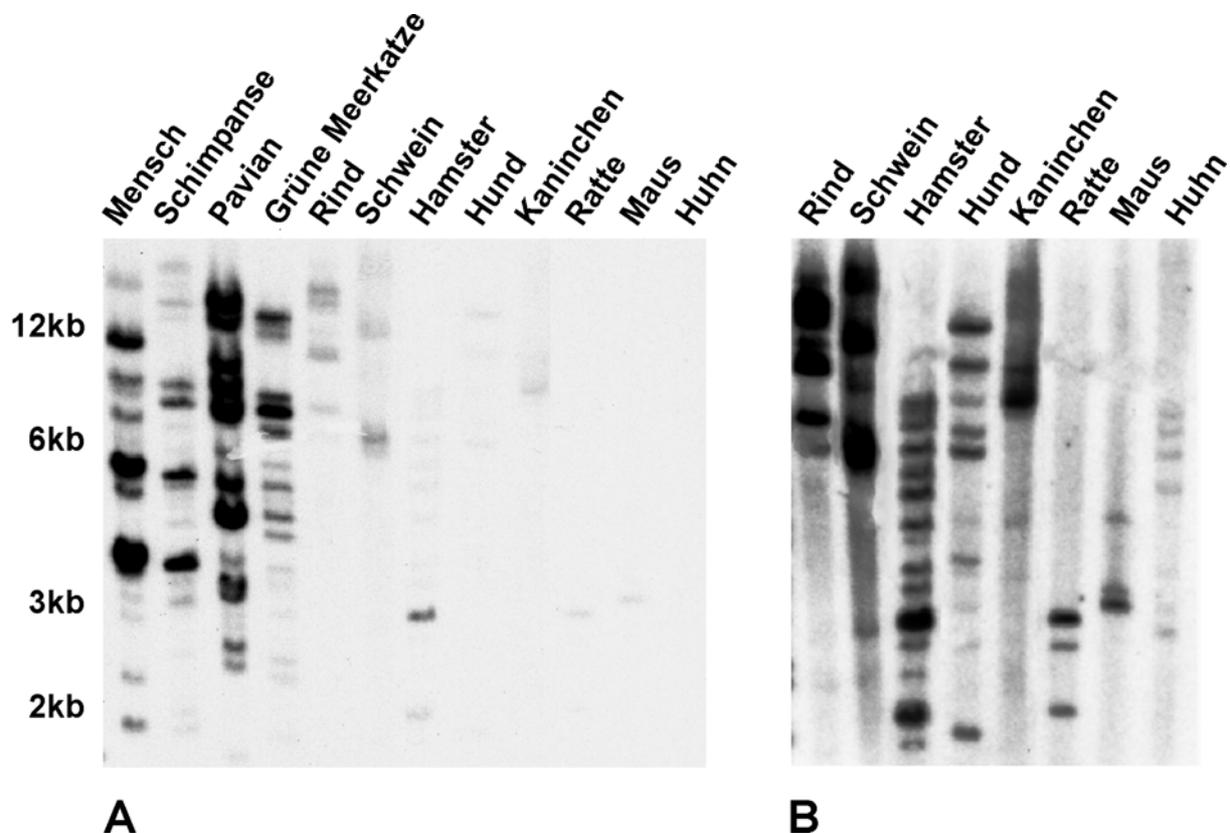


Abb. 6: Hybridisierung einer „Multispezies-Membran“ mit der Sonde ILT1 (LILRA2)
Bild „A“ zeigt das Ergebnis der Hybridisierung nach 24 Stunden, Bild „B“ nach 14 Tagen.

Das Ergebnis deutet darauf hin, daß neben anderen Spezies auch beim Huhn LILR-homologe Gene zu finden sind. Dies ist umso erstaunlicher, da es den letzten gemeinsamen Vorfahren von Mensch und Huhn vor ca. 300 Millionen Jahren gab. Die Hybridisierung zeigte zudem, daß es sich beim Huhn vermutlich wie beim Menschen um eine Multigenfamilie handelt, da fünf starke und zwei schwache Banden eindeutig zugeordnet werden konnten. Neben den LILRs gibt es im LRC weitere Gene, von denen bereits bekannt war, daß sie evolutionär sehr konserviert sind. Im Folgenden wurden deshalb weitere Hybridisierungsexperimente gegen „Multispezies-Membranen“ mit DNA-Sonden einiger dieser Rezeptoren durchgeführt.

4.1.2 Radioaktive Hybridisierungen mit Sonden anderer Gene des LRC

NKp46 ist, wie bereits in der Einleitung erwähnt, in verschiedenen Spezies nachgewiesen und kloniert worden. Neben einer humanen Sonde sollte deshalb auch eine spezifische Sonde für NKp46 der Maus zur Hybridisierung eingesetzt werden, um die Wahrscheinlichkeit zu erhöhen, NKp46 beim Huhn zu finden. Beide NKp46-Sonden wurden jeweils mit einer „Multispezies-Membran“ hybridisiert (Abb. 7).

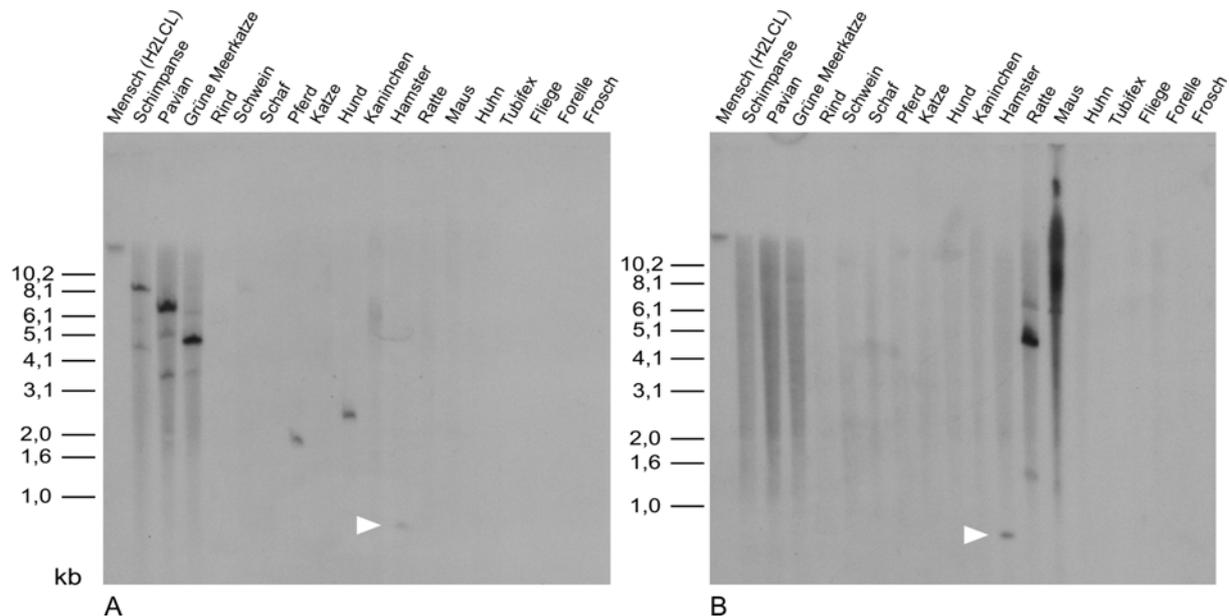


Abb. 7: Multispezies-Membranen, welche mit zwei unterschiedlichen NKp46-Sonden hybridisiert wurden. Hybridisierung mit einer (A) humanen und (B) murinen NKp46-Sonde. Die Expositionsdauer betrug bei beiden Hybridisierungen 14 Tage. Der Röntgenfilm in B zeigt ein stark überexponiertes Signal bei der Maus, so daß einzelne Banden kaum noch zu erkennen sind. Diese Vorgehensweise ermöglicht jedoch auch, schwächere Kreuzhybridisierungen mit anderen Spezies zu erkennen. Die den Membranen zugrunde liegenden Gele unterscheiden sich leicht in der Laufweite, weshalb ein direkter Vergleich der Hybridisierungssignale nicht möglich ist. Es kann jedoch vermutet werden, daß die mit der Pfeilspitze markierten Banden beim Hamster miteinander korrespondieren. Auf dem Röntgenfilm der Hybridisierung mit der humanen NKp46-Sonde ist zusätzlich eine sehr schwache Bande beim Schwein zu erkennen. Bei dem Signal in A in der Spur mit genomischer DNA des Pferdes handelt es sich vermutlich um einen Artefakt der Filmentwicklung.

Die humane NKp46-Sonde zeigte wie erwartet ein Signal bei der menschliche Zelllinie H2LCL und kreuzhybridisierte mit der genomischen DNA verschiedener Primaten (Schimpanse, Pavian und Grüne Meerkatze). Auch bei Pferd, Hund, Kaninchen (sehr schwach) und

Hamster konnten NKp46-homologe DNA-Fragmente detektiert werden. Die murine NKp46-Sonde zeigte die stärksten Hybridisierungssignale bei Maus und der eng verwandten Ratte. Zusätzlich ist ein moderates Signal beim Hamster zu erkennen, welches vermutlich mit der Bande der humanen Sonde korrespondiert (Pfeilspitze in Abb. 7). Eine schwache Kreuzhybridisierung zeigt darüber hinaus die genomische DNA des Menschen (H2LCL), und auch hier korrespondieren vermutlich die Banden der beiden Sonden miteinander. Ob es sich bei den Hybridisierungssignalen der verschiedenen Primaten-DNAs mit der NKp46-Sonde der Maus um spezifische Signale handelt, kann nicht eindeutig geklärt werden, da sie von einem relativ starken Hintergrund überdeckt sind. Es wäre jedoch zu erwarten, daß auch hier eine Kreuzhybridisierung stattfindet, wenn beim Menschen ein DNA-Fragment nachweisbar ist. Beide

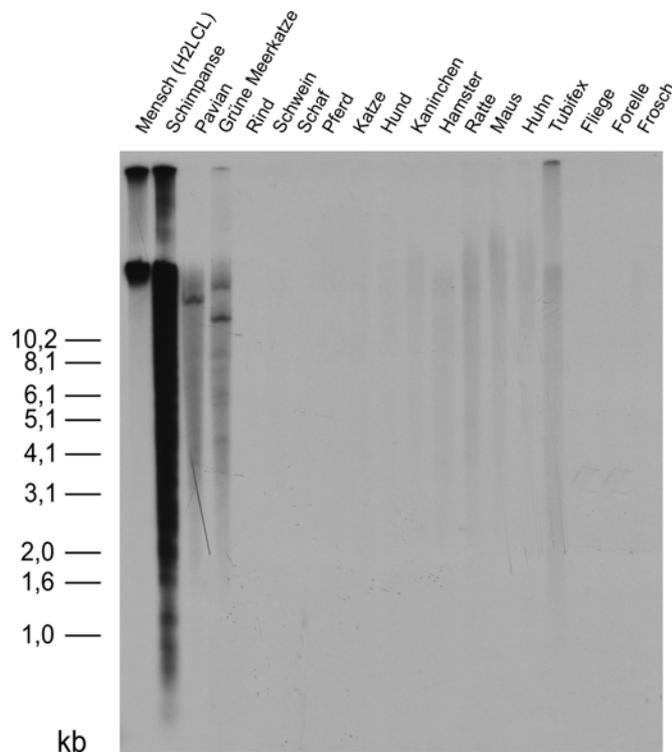


Abb. 8: Hybridisierung einer Multispezies-Membran mit einer humanen $Fc\alpha R$ -Sonde. Bei den Primaten sind sehr starke Hybridisierungssignale nachweisbar. Die Expositionsdauer betrug einen Tag. Da bei verschiedenen Spezies bereits Hintergrundsignale erkennbar wurden, ohne daß sich diskrete Banden abzeichneten, wurde von einer längeren Exposition abgesehen.

rechnet werden können. Da dies vermutlich aufgrund mangelnder Homologie nicht der Fall ist, verwundert es nicht, daß bei dem stammesgeschichtlich weiter entfernten Huhn auch keine Hybridisierung festgestellt werden konnte. Ob also der gemeinsame Vorläufer von Säugern und Vögeln ein *FCAR*-Gen besaß oder nicht, kann auf diese Weise nicht geklärt werden.

Neben den oben erwähnten konservierten LRC-Genen wurde auch eine Hybridisierung mit KIR-Sonden durchgeführt. Dazu wurden mehrere humane KIR-cDNAs (*KIR2DL1*, *KIR2DL3*,

NKp46-Sonden zeigten kein Hybridisierungssignal bei der genomischen DNA des Huhns, was entweder bedeutet, daß das Huhn kein entsprechendes Gen besitzt, oder daß die Sequenzunterschiede so groß sind, daß eine Kreuzhybridisierung nicht mehr erfolgen kann.

Eine weitere Hybridisierung wurde mit einer Sonde des $Fc\alpha$ -Rezeptors (*FCAR*) durchgeführt. Die humane DNA und die DNA der Primaten zeigten deutliche Hybridisierungssignale (Abb. 8). Im Gegensatz dazu waren bei den anderen aufgetragenen Spezies keine Signale zu erkennen, die sich vom Hintergrund abhoben. Da bekannt ist, daß die Ratte ein *FCAR*-Gen besitzt, hätte zumindest dort mit einer Kreuzhybridisierung ge-

KIR3DL1 und *KIR3DL2*) verwendet. Wie erwartet zeigten die Proben der Primaten bereits nach 24 Stunden Exposition positive Signale (Abb. 9). Nach 14 Tagen lassen sich zusätzlich Hybridisierungsergebnisse für Rind, Schwein, Hamster und sehr schwach für die Ratte nachweisen. Maus und Huhn hybridisierten nicht mit den KIR-Sonden. Das Ergebnis bestätigt, daß es sich bei KIRs um eine evolutionär junge Genfamilie handelt, deren Genzahl sich erst in den Primatenlinien deutlich erhöht hat, was an der Vielzahl der Hybridisierungssignale erkennbar ist. Das Huhn besitzt vermutlich keine KIR-homologen Gene.

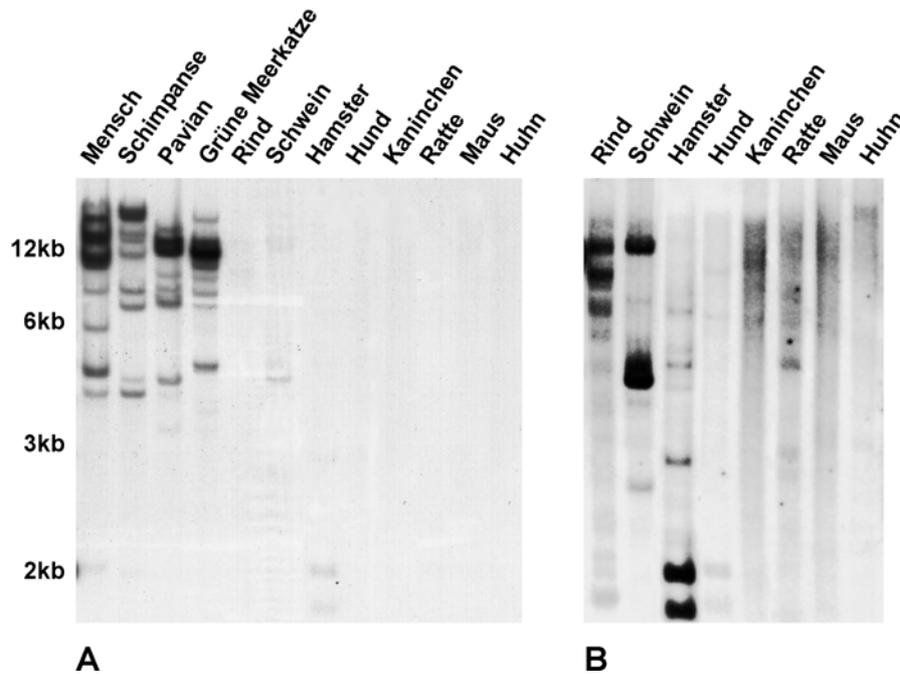


Abb. 9: Hybridisierung einer „Multispezies-Membran“ mit KIR-Sequenzen
Eine Multispezies-Membran wurde mit einem Gemisch der Sonden für *KIR2DL1*, *KIR2DL3*, *KIR3DL1* und *KIR3DL2* hybridisiert. Die Abbildung zeigt das Ergebnis der Hybridisierung nach Exposition für 24 Stunden (A) und einen Ausschnitt des Röntgenfilms nach einer Exposition von 14 Tagen (B).

4.1.3 Identifikation von LILR-Homologen beim Huhn

Die durchgeführten Hybridisierungsexperimente zeigten, daß es eine Multigenfamilie von LILR-homologen Genen beim Huhn gibt. Die evolutionäre Verwandtschaft von CHIRs (Dennis, Jr. et. al. 2000) zu den PIRs von Ratte und Maus sowie zu den LILRs legte nahe, daß es sich bei den hier identifizierten LILR-homologen Genen um CHIRs handelt. Aufgrund der bekannten Duplikationsereignisse innerhalb des LRC von Maus und Mensch wäre es jedoch auch möglich, daß es sich um eine mit den CHIRs verwandte Genfamilie handelt. Um diese Hypothesen zu prüfen, wurde eine Bursa-cDNA Bibliothek mit einer LILR-Sonde hybridisiert (siehe Abschnitt 4.3.1). Die dabei identifizierten Klone enthielten Sequenzen, die zu den publizierten CHIR-Sequenzen homolog waren. CHIRs sind also die LILR-homologen Gene des Huhns.

4.1.4 Diskussion der Hybridisierungsergebnisse

Von allen verwendeten Sonden wurde nur mit der LILRA2-cDNA ein Hybridisierungssignal beim Huhn erhalten. Die LILRA2-Sonde hybridisierte mit einer Vielzahl von Fragmenten genomischer DNA des Huhns, was darauf hindeutet, daß es sich um Mitglieder einer Multigenfamilie handelt. Neben LILRs können nur noch NKp46 und GPVI im LRC der Maus gefunden werden (und FCAR bei der Ratte). Die Konservierung von NKp46 und FCAR reichte aber nicht aus, um diese Gene mit humanen Sequenzen bei der Maus bzw. Ratte nachzuweisen. Damit ist es sehr unwahrscheinlich, daß mit Hilfe einer Hybridisierung unter ähnlichen Bedingungen im Stammbaum weiter entfernte und damit divergentere Sequenzen, wie NKp46-homologe beim Huhn, nachgewiesen werden können.

Bisher war man davon ausgegangen, daß Nager keine KIR-Gene besitzen und stattdessen Ly49-Gene für KIR-Funktionen verwenden. Erstaunlicherweise zeigte der Hamster jedoch einige starke Hybridisierungssignale mit der humanen KIR-Sonde. Möglicherweise handelt es sich hierbei um mehrere KIR-homologe Gene, was einen Hinweis auf funktionelle KIR-Gene bei diesem Nager darstellen würde. Das sehr schwache Hybridisierungssignal der humanen KIR-cDNA mit der genomischen DNA der Ratte spiegelt vermutlich ebenfalls eine spezifische Hybridisierung mit KIR-Sequenzen wider, da kürzlich KIR cDNAs aus Maus und Ratte kloniert werden konnten (Hoelsbrekken et. al. 2003). Die Aminosäuresequenz-Übereinstimmung gegenüber humanen KIRs beträgt bei der Maus nur 37,5 %, bei der Ratte jedoch 45,4 %. Die etwas höhere Homologie der humanen Sequenz zur Ratte könnte auch erklären, warum bei der genomischen DNA der Maus kein Hybridisierungssignal zu erkennen ist (siehe Abb. 9). Das KIR-Gen der Ratte ist im LRC der Ratte kodiert, während das KIR-Gen der Maus auf Chromosom X lokalisiert ist. Da es auch Reste eines Ly49-Gens beim Menschen gibt, haben die Vorläufer beider Genfamilien offenbar schon vor der Teilung der Entwicklungslinien von Nagern und Primaten existiert. Interessanterweise haben bei diesen Spezies unterschiedliche Genfamilien näherungsweise gleiche Funktionen übernommen.

Neben der Suche nach LRC-homologen Genen beim Huhn sollten die Hybridisierungsexperimente auch dazu dienen, die LILR-homologen Gene in den Kontext weiterer LRC-Gene zu setzen. Da dies nicht gelungen ist, bleibt unklar, ob es beim Huhn einen dem LRC vergleichbaren Genkomplex gibt, und ob die LILR-homologen CHIRs Teil eines solchen Komplexes sind.

4.2 Analyse der genomischen Organisation der CHIR-Gene des Huhns

4.2.1 Durchmusterung einer Bibliothek mit Klonen genomischer Hühner-DNA

Die oben beschriebenen Hybridisierungsergebnisse deuteten darauf hin, daß es sich bei den CHIRs um eine Multigenfamilie handelt. Um möglichst alle Gene dieser Familie zu identifizieren und deren chromosomale Organisation aufzuklären, wurde eine Bibliothek genomischer Klone des Haushuhns nach CHIR-Sequenzen durchmustert. Bei der Analyse großer Genomabschnitte haben sich künstliche Bakterienchromosomen (‚bacterial artificial chromosomes‘, BACs) sehr bewährt. Hierbei handelt es sich um Plasmide, in die große DNA-Fragmente (Inserts) kloniert und die mit Hilfe von Bakterien vermehrt werden können. In einer BAC-Bibliothek genomischer DNA ist das gesamte Genom eines Organismus zufällig in den BAC-Vektor kloniert. Die Summe der Klone repräsentiert dann das Genom des gewählten Organismus in der Regel mehrfach. Mir stand eine genomische Bibliothek des Haushuhns (*Gallus gallus domesticus*) zur Verfügung. Die durchschnittliche Größe der Inserts der BACs dieser Bibliothek beträgt 130kb. In der Bibliothek sind ca. 50000 Klone enthalten, die in Form von „Makro-Arrays“ zur Verfügung standen. Die Bibliothek deckt das Hühner-Genom ca. 5,4-fach ab.

Um nach Klonen mit CHIR-Genen zu suchen, mußte eine passende Sonde hergestellt werden. Experimente beim Menschen hatten gezeigt, daß einzelne LILR-cDNAs oder KIR-cDNAs aufgrund ihrer Sequenzähnlichkeit mit allen Genen der Familie kreuzhybridisieren. Da die LILRA2-cDNA mit mehreren CHIR-Genen hybridisierte, kann davon ausgegangen werden, daß eine CHIR-cDNA, die alle Exone umfaßt, mit sämtlichen CHIR-Genen kreuzhybridisiert. Aus diesem Grund habe ich zwei Sonden mittels RT-PCR hergestellt. Die PCR wurde mit spezifischen Primern für die bereits publizierten Sequenzen CHIR-A und CHIR-B (Dennis, Jr. et. al. 2000) durchgeführt, und die erhaltenen PCR-Produkte wurden mit dem TOPO-TA-Kloning Kit kloniert. Die erhaltenen Plasmide wurden sequenziert und das Insert anschließend mittels Restriktionsverdau freigesetzt, isoliert und in den folgenden Hybridisierungsexperimenten eingesetzt. Bei den Klonen handelt es sich um die in Abschnitt 4.3.2 beschriebenen CHIR-cDNAs-32 und -43.

Die CHIR-A und -B-Sonden wurden mittels des Ambion-StripeEZ-Protokolls radioaktiv markiert und anschließend miteinander vermischt. Diese Sondenmischung wurde über Nacht mit einer Nylon-Membran hybridisiert, an die genomische DNA des Huhns gebunden war. Auf diese Weise können unspezifische Bindungen des Sondengemisches an die Filter der an-

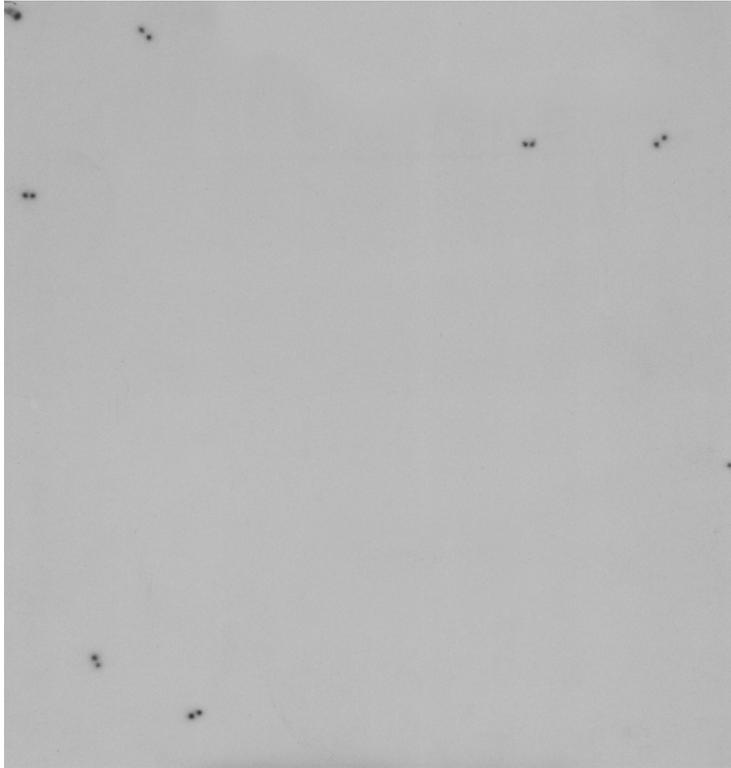


Abb. 10 Hybridisierungsergebnis der „Makro-Array“-Membran 2 der Hühner-BAC-Bibliothek. Auf diesem Filter konnten acht CHIR-positive Klone identifiziert werden. Jeder Klon ist im Duplikat auf der Membran aufgetragen. Aus der Lage der Punkte zueinander und ihrer Position auf der Membran läßt sich ableiten, um welchen Klon es sich handelt.

schließend verwendeten BAC-Bibliothek verhindert werden. Am nächsten Tag wurden dann die Membranen der Hühner-BAC-Bibliothek mit dem CHIR-Sonden-Gemisch hybridisiert. Nach einer Inkubation über Nacht wurden diese gewaschen und anschließend für einen Tag auf einen Röntgenfilm aufgelegt.

Das Ergebnis der Hybridisierung einer dieser Membranen ist in Abb. 10 dargestellt. Ein positives Hybridisierungssignal ist an zwei nahe beieinander liegenden runden Punkten zu erkennen. Aus der Lage der Punkte zueinander und ihrer Position auf dem Filter läßt sich der zugehörige Klon identifizieren. Es wurden insgesamt 19 positive Klone nachgewiesen (Membran 1: sechs Klone; Membran 2: acht Klone; Membran 3: fünf Klone). Alle markierten Klone wurden vom MRC-Geneservice (Cambridge, GB) bezogen. Die 19 Klone wurden anschließend in Kultur genommen und eine Plasmid-Präparation durchgeführt. Um zu überprüfen, ob alle Klone die erwarteten Sequenzen enthalten, wurden die isolierten Plasmide mit der Restriktionsendonuklease EcoRI geschnitten und gelelektrophoretisch aufgetrennt. Die Fragmente wurden auf eine Nylon-Membran transferiert und mit den bei der Durchmusterung der BAC-Bibliothek verwendeten Sonden hybridisiert (Abb. 11).

Alle Klone bis auf BAC 64A24 hybridisierten mit den CHIR-Sonden, d.h. sie enthalten CHIR-homologe Sequenzen. Der Grund für das Fehlen von Hybridisierungssignalen in BAC 64A24 könnte möglicherweise mit einem Fehler bei der Auswertung der BAC-Bibliothek-Membranen zusammenhängen. Bei Klon 64A24 handelt es sich um das Hybridisierungssignal ganz oben links in Abb. 10. Bei diesem Klon ist einer der beiden Punkte nicht genau lokalisiert, so daß möglicherweise das Muster nicht korrekt erkannt wurde. Da die Bibliothek jedoch eine fünffache Redundanz aufweist, wurde darauf verzichtet, die restlichen sieben Klone dieser Koordinate auf deren Gehalt an CHIR-Sequenzen zu untersuchen.

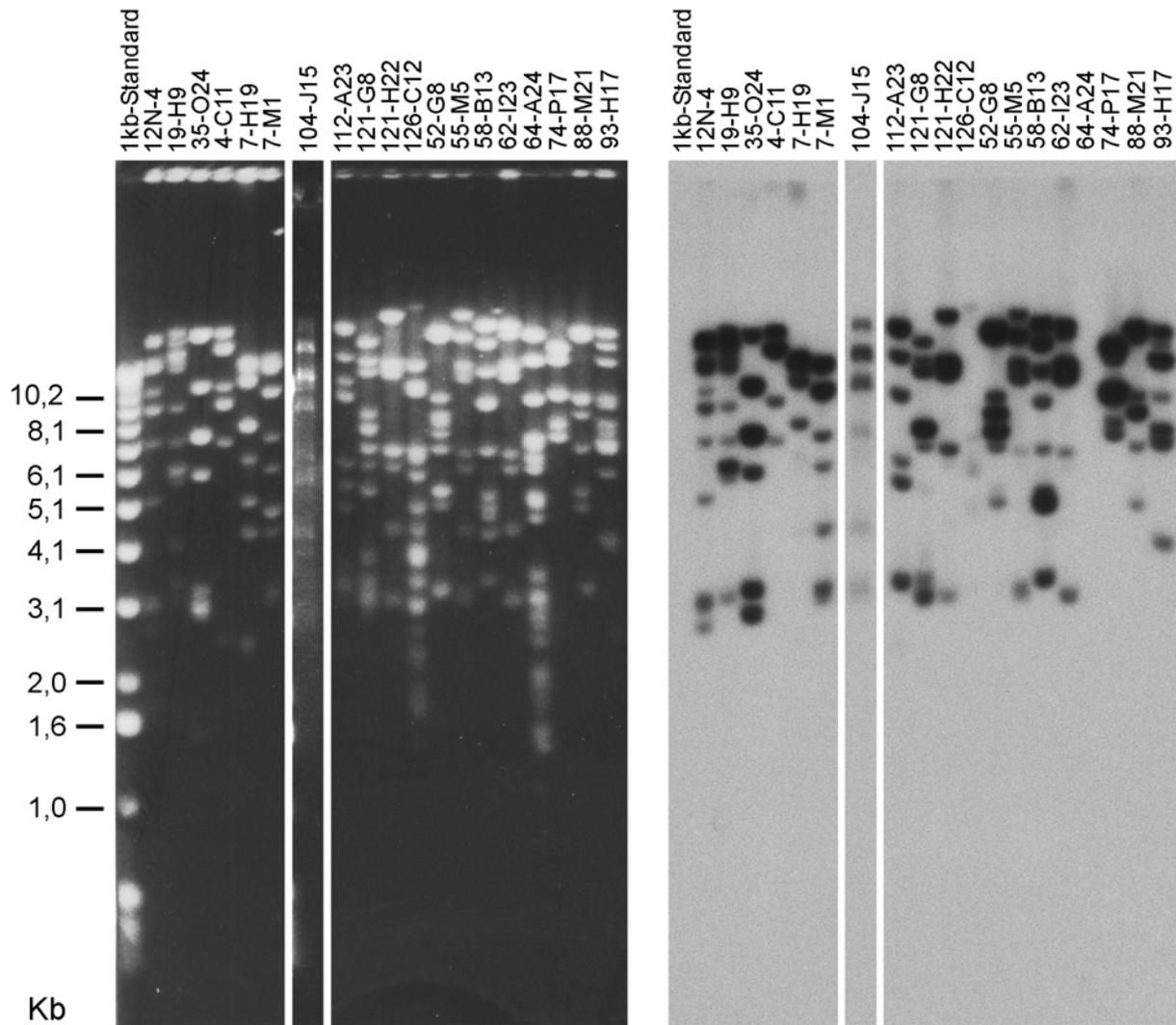


Abb. 11 Gelelektrophoretische Auftrennung der CHIR-BAC-Klone und deren radioaktive Hybridisierung mit CHIR-A und -B-spezifischen Sonden

Das linke Bild zeigt ein Gelfoto der mit EcoRI verdauten 19 BAC-Klone, die bei der Durchmusterung der genomischen BAC-Bibliothek des Huhns gefunden wurden. Die Namen der Klone sind über dem Gel und der Autoradiografie angegeben. Einer der BAC-Klone (104-J15) mußte aufgrund einer zu geringen DNA-Menge in einem separaten Experiment nochmals auf eine Nylonmembran übertragen werden. Die Ergebnisse dieses Experiments wurden an den entsprechenden Stellen des ursprünglichen Gels eingefügt. Bis auf einen Klon (64-A24) enthalten alle BAC-Klone CHIR-Sequenzen.

Die Hybridisierung der CHIR-BAC-Klone in Abb. 11 zeigt neben dem Vorhandensein von CHIR-Genen, daß in allen Klonen mehrere Banden mit den verwendeten Sonden hybridisiert. Daraus kann geschlossen werden, daß jeder Klon mehrere CHIR-ähnliche Sequenzen enthält oder daß die CHIR-Gene sehr groß sind und sich deshalb über mehrere Restriktionsfragmente erstrecken. Um wie viele Gene es sich handelt, kann aus diesem Ergebnis nicht abgeleitet werden. Eine weitere interessante Information ergibt sich aus der Anzahl der gefundenen Klone (18): Wenn man eine ca. fünffache Abdeckung des Genoms des Huhns zugrunde legt und in Betracht zieht, daß ein BAC-Klon im Durchschnitt 130 kb groß ist, folgt daraus, daß sich die CHIR-enthaltenden chromosomalen Abschnitte über etwa drei BAC-Klone und damit ca. 450 kb erstrecken. Das ist allerdings nur eine grobe Abschätzung, da verschiedene

genomische Bereiche zu unterschiedlichen Anteilen in der BAC-Bibliothek repräsentiert sein können.

4.2.2 Sequenzierung ausgewählter BAC-Klone

Im Rahmen einer Kooperation mit Dr. Stefan Beck wurden die 18 CHIR-positiven BAC-Klone zur Sequenzierung dem „Welcome Trust Sanger Center“ (Hinxton, GB) übergeben. Um den Sequenzieraufwand und damit die Kosten zu minimieren, wurden die Klone dort zunächst einer ‚fingerprint‘-Analyse unterzogen. Mit Hilfe eines solchen Experiments läßt sich herausfinden, welche Klone in wie großem Maße miteinander überlappen. Auf der Grundlage dieser Ergebnisse wurden die Klone in Contigs arrangiert. Zur Sequenzierung wurden dann Klone ausgewählt, die zwar miteinander überlappen, aber gleichzeitig nur wenig redundante Sequenzabschnitte enthalten. Die Wahl fiel auf Klone aus zwei Contigs (bestehend aus fünf bzw. zwei Klonen), die den Großteil der CHIR-Genabschnitte abdecken sollten. Insgesamt wurden sieben der 18 Klone zur Sequenzierung ausgewählt. Dabei handelte es sich um 112A23 und 58B13 (Contig 1), sowie 19H9, 52G8, 88M21, 93H17 und 4C11 (Contig 2).

4.2.3 Analyse der Sequenzen der CHIR-enthaltenden BAC-Klone

4.2.3.1 Vergleich der BAC-Sequenzen untereinander

Am Anfang der Sequenzauswertung sollte festgestellt werden, ob und wie die sequenzierten BACs tatsächlich miteinander in Zusammenhang stehen. Die vor der Sequenzierung durchgeführten ‚fingerprint‘-Analysen geben zwar einen Hinweis auf den Grad der Überlappung der Klone, sind aber zu ungenau, um diesen exakt zu bestimmen. Aus diesem Grund wurden die Sequenzen aller BACs mit Hilfe einer Dot-Matrix-Analyse miteinander verglichen. Mit dieser Methode konnte bestätigt werden, daß die Klone 112A23 und 58B13 ein Contig bilden und 2092 bp miteinander überlappen. Beide Klone decken zusammen 174 kbp genomischer DNA ab. Die Überlappung der Klone ist im Dot-Matrix-Plot als diagonale Linie erkennbar (Abb. 12), die eine Homologie zwischen dem Ende von Klon 112A23 und dem Anfang von Klon 58B13 anzeigt. Darüber hinaus ist eine Vielzahl weiterer Diagonalen zu erkennen, die Homologien zwischen den Klonen repräsentieren. Da letztere aber nicht am Anfang oder Ende der beiden Klone liegen, sind sie nicht Teil einer Überlappung (zu ihrer Bedeutung siehe unten). Zur Absicherung dieses Ergebnisses wurde zusätzlich ein Sequenzvergleich durchgeführt, der im Bereich der Überlappung eine hundertprozentige Sequenzübereinstimmung beider Klone ergab.

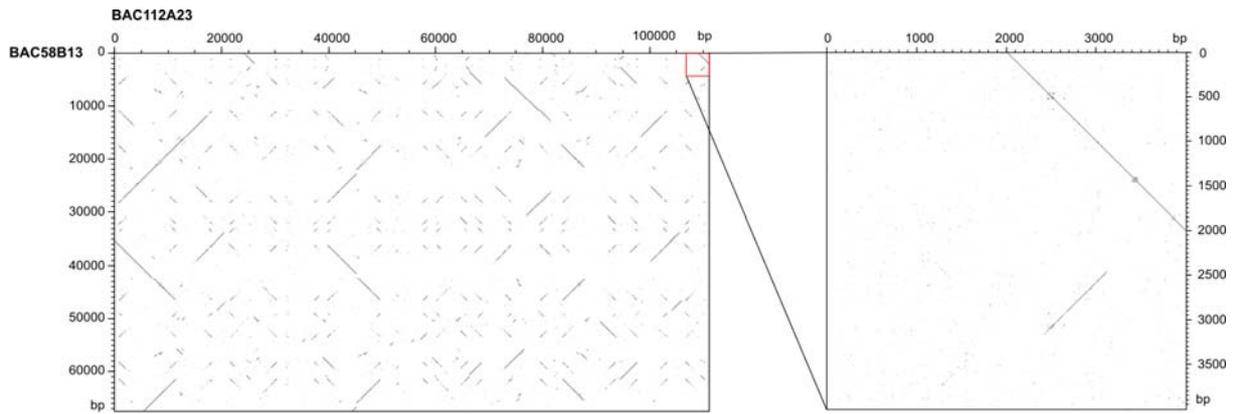


Abb. 12 Überlappung der Klone 112A23 und 58B13

Die linke Grafik zeigt das mit Hilfe der Dot-Matrix-Analyse erstellte Bild des Vergleichs beider Sequenzen. In der oberen rechten Ecke befindet sich eine diagonale Linie, die die Überlappung der beiden Klone anzeigt. Der rot eingerahmte Bereich ist in der rechten Abbildung vergrößert dargestellt. Die Analyse zeigt, daß das Ende von BAC 112A23 mit dem Anfang von BAC 58B13 2092 bp überlappt. (auf die weiteren Homologien im linken Bild wird später näher eingegangen).

Für das zweite postulierte Contig (bestehend aus fünf Klonen) ergab sich ein von den ursprünglichen ‚fingerprint‘-Daten abweichendes Ergebnis: BAC 19H9 überlappt mit keinem der anderen Klone, er hat eine Länge von 133 kbp. BAC 88M21 ist vollständig in BAC 52G8 enthalten, letzterer überlappt ebenfalls mit keinem anderen Klon; seine Länge beträgt ca. 123 kbp. Im Gegensatz dazu bilden die Klone 93H17 und 4C11 ein Contig; 93H17 wurde nicht vollständig sequenziert, da die Analyse der unvollständigen Sequenz des Klones bereits zeigte, daß dieser zum großen Teil mit 4C11 überlappte. Aus diesem Grund wurde die Assemblierung der Sequenzrohdaten von 93H17 nur für die Sequenzen beendet, die nicht in Klon 4C11 enthalten sind. Insgesamt decken die beiden Klone eine Länge von ca. 134 kbp ab. Eine Zusammenfassung der Längen und Überlappungen der sequenzierten Klone gibt Tab. 1.

Name	Länge der Sequenz (bp)	Contig	Länge des Contigs (bp)	Bemerkungen
112A23	111.067	1	176.030	Beide Klone überlappen 2092 bp miteinander.
58B13	67.055			
19H9	133.037	2	133.037	
52G8	123.203	3	123.203	Klon 88M21 ist vollständig in 52G8 enthalten.
88M21	107.816			
93H17	50.749	4	134.753	Die Sequenz von Klon 93H17 wurde nicht vollständig assembliert, der Klon ist also länger als die angegebene Sequenz.
4C11	86.004			
Gesamtlänge der sequenzierten genomischen DNA			567.023	

Tab. 1 Längen, Contigs und Eigenschaften der sequenzierten CHIR-BAC-Klone.

4.2.3.2 Vergleich der Fingerprint-Daten mit den Dot-Matrix-Ergebnissen

Die Fingerprint-Daten der BAC-Klone legten nahe, daß die Klone 19H9, 52G8, 88M21, 93H17 und 4C11 ein zusammenhängendes Contig bilden. Die Sequenzierung dieser Klone zeigte jedoch, daß sie drei nicht miteinander überlappende Sequenzabschnitte überdecken.

Wie kann man sich diese Diskrepanz erklären? Neben den Überlappungen zeigten die Dot-Matrix-Plots zwischen allen Klonen eine Vielzahl von zum Teil weitreichenden Homologien auf, wie sie zum Beispiel in Abb. 12 zu sehen sind. Diese Information zeigt die wahrscheinlichste Fehlerquelle des Fingerprint-Verfahrens auf. So ist es mit dieser Methode möglich, gleiche (oder sehr ähnliche) Sequenzabschnitte aufgrund ihres Restriktionsmusters zu identifizieren. Enthalten nun zwei oder mehrere Klone sehr ähnliche Sequenzen, würden die daraus abgeleiteten Restriktionsfragmente des Fingerprints fälschlicherweise als identische Fragmente identifiziert. Daraus könnte abgeleitet werden, daß die entsprechenden Klone das gleiche Restriktionsfragment enthalten, also miteinander überlappen. Es zeigt sich also, daß die fingerprint-Daten einen Hinweis, aber keinen Beweis für eine Überlappung darstellen. Das Verfahren des Fingerprintings versagt offenbar bei DNA-Sequenzen, die sehr ähnlich zueinander sind, wie es zum Beispiel bei Sequenzen der Fall ist, die durch Duplikation auseinander hervorgegangen sind (siehe unten).

4.2.3.3 Detaillierte Sequenzanalyse der BAC-Klone

Um die einzelnen CHIR-Gene zu lokalisieren und deren Orientierung und Exon-Intron-Struktur innerhalb der BACs zu identifizieren, wurden die Sequenzen der einzelnen BAC-Klone mit den bereits bekannten cDNA-Sequenzen von CHIR-A und -B mittels Dot-Matrix-Analyse verglichen (Abb. 13). Es zeigte sich, daß alle sequenzierten BACs mehrere CHIR-Gene und -Genfragmente enthielten, wie es bereits die Hybridisierungsdaten nahe legten (siehe Abschnitt 4.2.1).

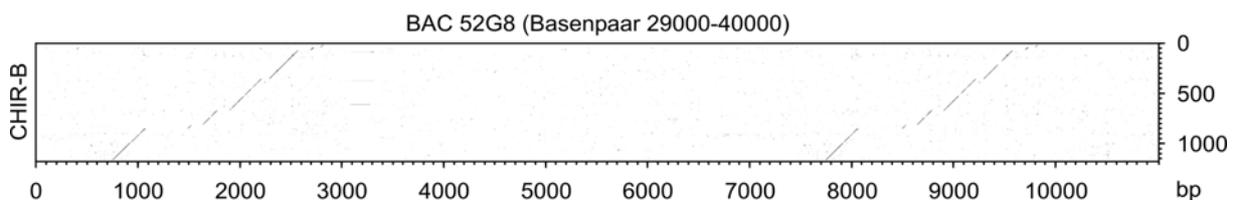


Abb. 13 Dot-Matrix-Sequenzvergleich eines Teils von BAC 52G8 mit der CHIR-B-cDNA
Der Dot-Plot zeigt Homologien zwischen beiden Sequenzen auf. Die diagonalen Linien entsprechen den Exonen der CHIR-B-cDNA, die von den Intronen in 52G8 unterbrochen sind. In diesem Sequenzabschnitt sind zwei CHIR-Gene enthalten.

Um jede der gefundenen Homologien im Detail auf Sequenzebene zu analysieren, wurden die CHIR-A- und CHIR-B-cDNA-Sequenzen mit Hilfe des Programms ‚Multalin‘ an den entsprechenden Sequenzabschnitten der BACs ausgerichtet (Sequenz-Alignment). Dieses Verfahren erlaubt einen basengenauen Sequenzabgleich und ermöglicht es so, die Exon-Intron-Struktur der Gene zu entschlüsseln. Nicht in jedem Fall war es jedoch möglich, einen solchen Sequenzabgleich mit Hilfe der bekannten cDNAs durchzuführen, weil die Homologie zwischen den cDNAs und den Genen zu gering war. In solchen Fällen wurde die genomische Sequenz des Gens verwendet, um Sequenzdatenbanken nach transkribierten Sequenzen, so-

nannten ‚expressed sequence tags‘ (ESTs) zu durchsuchen. Die erhaltenen ESTs wurden dann wiederum einem Sequenz-*Alignment* mit der Gensequenz unterzogen, um die Exon-Intron-Grenzen aufzuspüren.

Insgesamt konnten zwei verschiedene Exon-Intron-Konfigurationen identifiziert werden: Zum einen Gene mit fünf Exonen und einer Länge von etwa 1,4 kbp, die in ihrer Struktur CHIR-A-ähnliche Gene darstellen und zum anderen Gene mit sieben Exonen und einer Länge von etwa 2 kbp, die mit CHIR-B vergleichbar sind (Abb. 14). Die Exon-Intron-Struktur ist zwischen den CHIR-Genen erstaunlich konserviert. Die Konservierung ist jedoch besonders am 5'-Ende ausgeprägt und umfaßt dort sogar die nicht-kodierenden Gensequenzen. So ist das Intron 1 aller CHIR-Gene 73 bp lang. Das zweite Intron zeigt verstärkte Variabilität und die Konservierung zwischen den Genen nimmt zum 3'-Ende hin weiter ab. So besitzen nur noch etwa die Hälfte der Gene Exon 6 und Exon 7.

Neben Genen mit fünf oder sieben Exonen wurden auch Gene gefunden, denen eine oder mehrere Exone oder Teile davon fehlen; bei diesen Genen handelt es sich vermutlich um Pseudogene (siehe Abschnitt 4.2.3.5). In allen BACs sind zudem Fragmente von CHIR-Genen verteilt. Welche der gefundenen Gene und der Gene mit fehlenden Exonen für funktionelle Proteine kodieren, wird in Abschnitt 4.2.3.5 diskutiert.

Eine detaillierte genomische Karte der vier CHIR-enthaltenden chromosomalen Bereiche ist in Abb. 15 dargestellt. Hier sind nur die Gene als funktionell dargestellt, bei denen korrekte Exon-Intron-Grenzen gefunden wurden und die in weiteren Analysen zeigten, daß ein sinnvolles hypothetisches Transkript abgeleitet werden kann. Mit Hilfe des Internet-Programms „Emboss-Isochore“ wurde zudem der Verlauf des G-C-Gehaltes innerhalb der BAC-

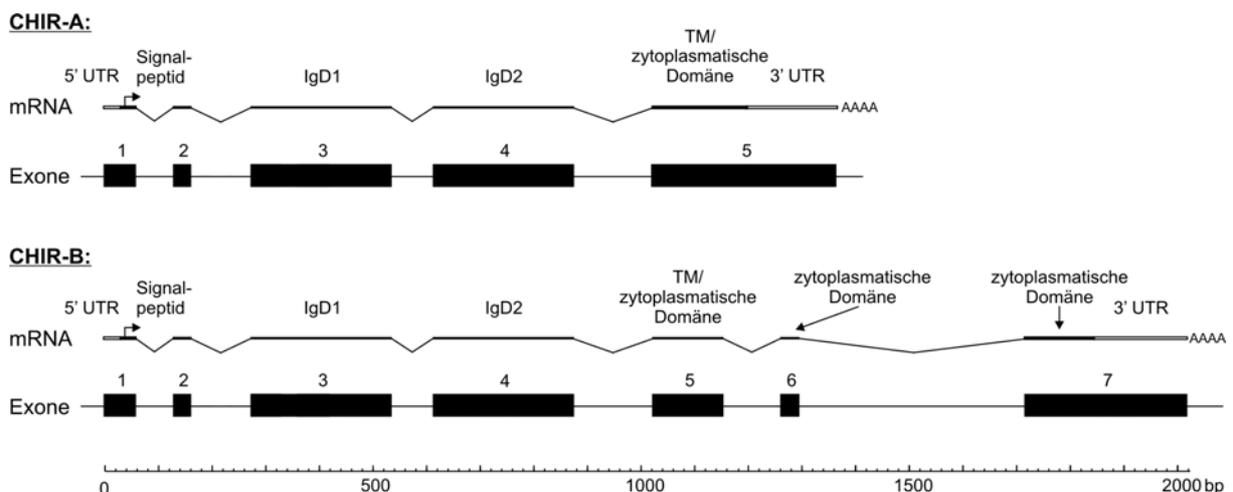


Abb. 14 Exon-Intron-Struktur der CHIR-Gene.

Bei beiden Genen ist jeweils im oberen Teil die sich aus den gespleißten Exonen ergebende mRNA dargestellt. Zusätzlich sind die Proteindomänen angegeben, für die die einzelnen mRNA-Abschnitte kodieren. Unter der mRNA sind die Gene schematisch dargestellt, die aus fünf (CHIR-A) oder sieben (CHIR-B) Exonen aufgebaut sind.

Sequenzen bestimmt. Isochore sind große DNA-Bereiche, die relativ homogen in ihrer Basenzusammensetzung sind und welche mit den darin enthaltenen kodierenden Sequenzen korreliert. Die Grenzen einer Isochores können deshalb als Indiz für das Ende eines Genclusters herangezogen werden. Die Analyse der Isochore ergab keine auffälligen Ergebnisse. Zusätzlich wurden die Sequenzen mit Hilfe des Programms „Repeatmasker“ auf deren Gehalt an repetitiven Elementen untersucht (Abb. 15 unterhalb der Genorganisation dargestellt). Die BAC-Sequenzen wurden auf das Vorhandensein von CpG-Inseln mit Hilfe des „CpG island-searcher“-Programms (<http://cpgislands.usc.edu/>) untersucht. CpG-Inseln findet man häufig im Promotorbereich von Genen, da deren Methylierung zur Transkriptionskontrolle verwendet wird. Insgesamt konnten 19 CpG-Inseln in den BAC-Sequenzen gefunden werden. Sie zeigen aber keine offensichtliche Assoziation mit den CHIR-Genen.

Offenbar ist die Vielzahl der CHIR-Gene durch wiederholte Duplikation eines oder mehrerer Gene entstanden. Um zu prüfen, ob diese Duplikationsereignisse auch die Intergenregionen beinhalteten, wurden verschiedene dieser Regionen mit allen BAC-Sequenzen verglichen. Es zeigte sich, daß auch die Intergenregionen Teil von Duplikationen waren, da sich homologe Sequenzen in mehreren BACs an verschiedenen Stellen finden. Da Intergenregionen dazu neigen, verstärkt Mutationen anzusammeln, deutet dieses Ergebnis darauf hin, daß die Duplikationen evolutionär gesehen in der jüngeren Vergangenheit stattgefunden haben und eventuell bis heute anhalten. Homologe Intergenregionen sind in Abb. 15 in gleichen Farben dargestellt.

Die Intergenregionen sollten zusätzlich auf das Vorhandensein anderer Gene untersucht werden. Deshalb wurden die BAC-Sequenzen mit Datenbanken verglichen, um alle homologen Sequenzen zu identifizieren. Dabei stellte sich heraus, daß innerhalb der BACs mehrere Sequenzen mit Homologie zu G-Protein-gekoppelten Rezeptoren (GPR41-43) vorhanden sind (siehe Abb. 15). Alle GPR-homologen Gene liegen in einer mehrmals duplizierten Intergenregion (gelb Abb. 15), so daß davon ausgegangen werden kann, daß es sich um Nachkommen eines Gens handelt, welches vermutlich zufällig zusammen mit den CHIR-Genen dupliziert wurde. Die Sequenzen sind jedoch unvollständig und es kann kein offenes Leseraster abgeleitet werden. Vermutlich handelt es sich also um Pseudogene. Interessanterweise wurden von Trowsdale und Kollegen auch im „extended LRC“ des Menschen GPR41-43-homologe Gene gefunden (Trowsdale et. al. 2001). Dies ist ein Hinweis darauf, daß es sich bei dem CHIR-Cluster tatsächlich um eine LRC-orthologe Region handelt.

Neben diesen Genen konnten verschiedene ESTs den Intergenregionen zugeordnet werden. Die Homologie war aber jeweils nur schwach und deckte nur einen Teil der ESTs ab. Weitere

im CHIR-Cluster kodierte Transkripte lassen sich daraus nicht ableiten. Alle diese Homologien sind in Abb. 15 unter den Inter-CHIR-Regionen dargestellt.

4.2.3.4 Genomische Organisation der CHIR-Gene

Die in dieser Arbeit analysierten BAC-Sequenzen haben eine Gesamtlänge von etwa 550 kbp. Die BACs enthalten 60 CHIR-Gene sowie 43 CHIR-Pseudogene oder Genfragmente, die in Abb. 15 maßstabsgetreu dargestellt sind. CHIR-Gene zeigen im Gegensatz zu den humanen LRC-Genen (Wende et. al. 2000) keine Cluster, in denen alle Gene eine gemeinsame Transkriptionsrichtung aufweisen. Die Transkriptionsrichtung scheint vielmehr vollkommen zufällig zu sein. Die Gendichte der CHIR-Cluster ist mit einem Gen/Pseudogen pro 5 kbp extrem hoch und entspricht etwa der Gendichte des Haupthistokompatibilitätskomplexes des Huhns (Shiina et. al. 2004). Die CHIR-Gene bzw. -Pseudogene/-Fragmente sind gleichmäßig über alle vier Regionen verteilt. Auch funktionell unterschiedliche Typen von Rezeptoren (siehe Abschnitt 4.2.3.5) kommen nicht auffällig gehäuft in bestimmten Regionen vor, wie man es zum Beispiel bei Tandem-Duplikationen erwarten würde, die unter anderem durch ungleiches crossing-over entstehen können.

Um weitere Informationen über die Verbindung zwischen den einzelnen Contigs zu erhalten, wurden von allen BACs, die in der ursprünglichen Suche identifiziert worden waren, die Sequenzen der Insert-Enden bestimmt. Mit Hilfe dieser Sequenzen war es dann möglich, in den sequenzierten Klonen nach einer Sequenzübereinstimmung zu suchen. Wenn man auf diese Weise mit den Enden eines Klons Sequenzübereinstimmungen in zwei verschiedenen Contigs finden würde, ließe sich daraus ableiten, daß beide Contigs benachbart sind. Über die Größe des überbrückenden Klons und den Grad der Überlappung ließe sich zudem der Abstand der Contigs bestimmen. Die Ergebnisse dieser Untersuchung zeigten, daß es zu einigen BAC-Enden übereinstimmende Bereiche innerhalb der Contigs gibt (in Abb. 15 unterhalb der BAC-Symbole dargestellt). Allerdings überbrückte keiner der Klone die Lücke zwischen zwei Contigs. Aufgrund der hohen Homologie der BAC-Enden zu verschiedenen Bereichen der sequenzierten BACs kann nicht einmal zu 100% sichergestellt werden, daß wirklich eine Überlappung der entsprechenden Klone vorhanden ist. Die Ausdehnung des CHIR-Clusters ist

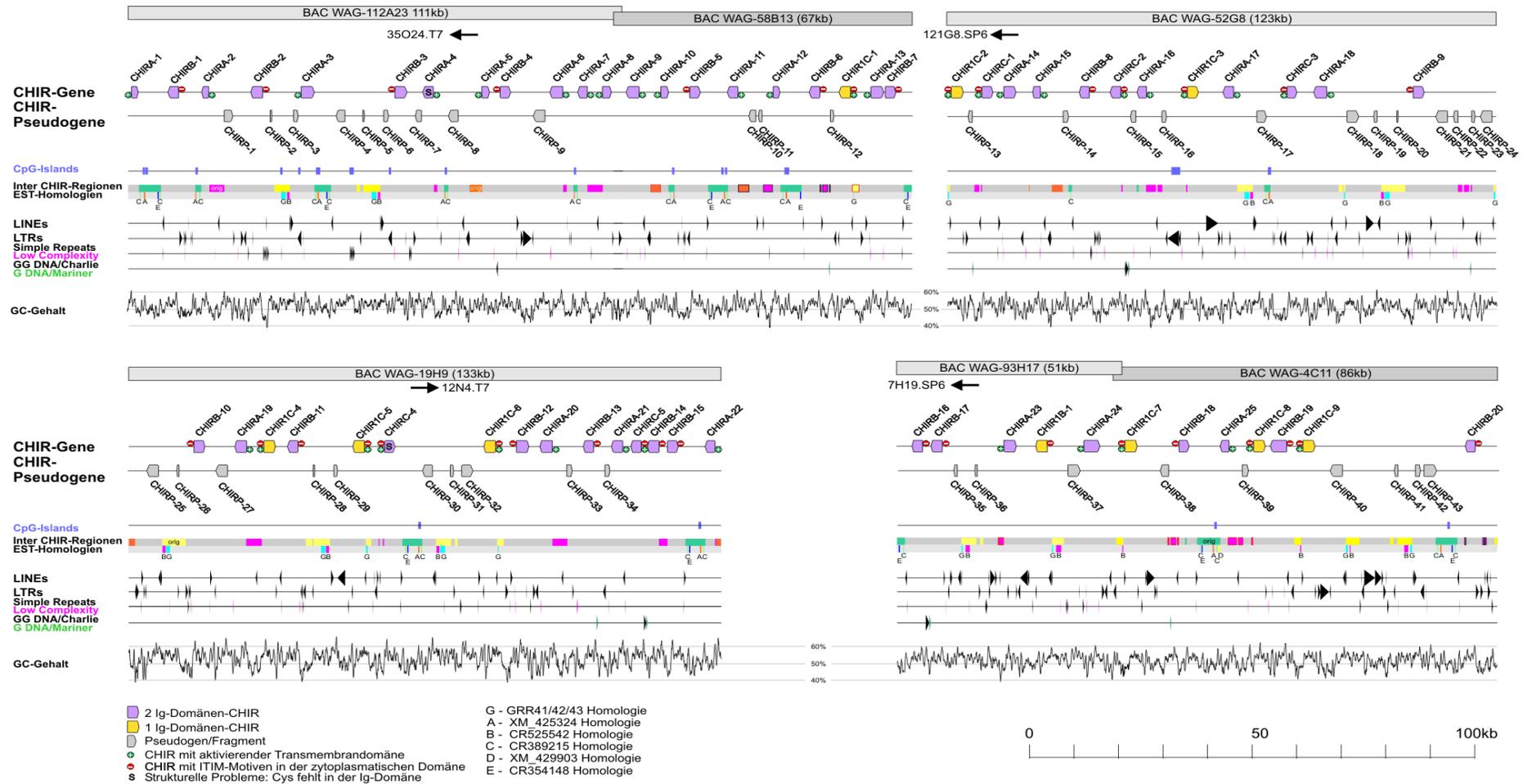


Abb. 15 Grafische Darstellung der Organisation der CHIR-Gene innerhalb der BAC-Contigs.

Die Abfolge der CHIR-Gene und deren Größe sind durch verschiedenfarbige Symbole unterhalb der zugehörigen BAC-Abschnitte maßstabsgerecht dargestellt, Die Spitze der Gensymbole gibt die Orientierung des jeweiligen Gens innerhalb des BAC-Inserts an und zeigt in Richtung des 3'-Endes. Der Rezeptortyp wird durch ein grünes Plus (aktivierender Rezeptor), rotes Minus (inhibitorischer Rezeptor) oder beides (bifunktionaler Rezeptor) symbolisiert (siehe Abschnitt 4.2.3.5). Die Pseudogene sind unter den Genen grau dargestellt. Die grauen Rechtecke über den Gensymbolen stellen die BACs dar, aus denen die jeweiligen Contigs bestehen. Um nach Duplikationen von Intergenregionen zu suchen, wurden verschiedene Bereiche verwendet (mit „Orig“ bezeichnet) und damit ein Sequenzvergleich mit allen BAC-Sequenzen durchgeführt. Homologe Bereiche sind in der gleichen Farbe dargestellt. Alle weiteren Kartenelemente sind links beschriftet, bzw. können der Legende entnommen werden.

also vermutlich deutlich größer als die Sequenz, die von den hier sequenzierten Klonen abgedeckt wird.

Die hohe Sequenzübereinstimmung zwischen den Genen, aber auch zwischen den Inter-genregionen legt nahe, daß das Gencluster durch wiederholte Duplikation von Genen und Genabschnitten entstanden ist. Das Ausmaß dieser Duplikationsereignisse übersteigt jedoch bei weitem das der evolutionär verwandten LILR- und KIR-Gene im LRC der Säugetiere (Wagtmann et. al. 1997; Wende et. al. 2000; Volz et. al. 2001; Trowsdale et. al. 2001). Die große Zahl von Pseudogenen und Genfragmenten stellt hierbei offenbar ein Überbleibsel der dynamischen Entwicklung des Clusters dar, bei der schnell neue Gene entstehen, die sich verändern und bei Verlust der Funktionalität wieder beseitigt werden.

4.2.3.5 *Detailanalyse der CHIR-Gene*

Während der Sequenzanalyse der einzelnen Contigs wurde die Exon-Intron-Struktur der einzelnen Gene bestimmt. Anschließend wurden die einzelnen Exonsequenzen *in silico* aneinandergereiht, um eine virtuelle cDNA zu erhalten. Alle sich ableitenden cDNAs wurden in Aminosäuresequenzen übersetzt. Diese wurden auf ihre Domänenstruktur hin analysiert und es wurde eine Suche nach Signaltransduktionsmotiven (ITIMs, geladene Aminosäure in der Transmembrandomäne [TM]) durchgeführt. In Abb. 15 wurden nur die Gene farbig dargestellt, aus denen ein funktionelles Protein abgeleitet werden konnte. Alle anderen Gene wurden als Pseudogene definiert und sind grau dargestellt (siehe unten). Im Folgenden sind die dazu durchgeführten Analysen im Detail beschrieben.

Die aus den cDNAs abgeleiteten Proteine weisen ein oder zwei Ig-Domänen, eine Transmembrandomäne und eine kurze oder lange zytoplasmatische Domäne auf. Innerhalb der Ig-Domänen befinden sich jeweils zwei konservierte Cysteine, die für die Faltung der Ig-Domäne essentiell sind. In der Transmembrandomäne findet sich bei einigen Proteinen eine für aktivierende Rezeptoren typische geladene Aminosäure. Zusätzlich weisen 46 von 60 CHIRs ein konserviertes Cystein in der Transmembrandomäne auf. Dieses Cystein könnte zur Dimerisierung der Rezeptoren dienen, wie es zum Beispiel für KIR2DL1 gezeigt wurde (Fan et. al. 2000).

Beim Aminosäuresequenzvergleich aller Proteine (Abb. 16) fällt zunächst die hohe Homologie der einzelnen Sequenzen untereinander auf, die im Bereich des Signalpeptides und des vorderen Teils der ersten Ig-Domäne besonders ausgeprägt ist. Die Transmembrandomäne kann zur Unterscheidung von aktivierenden und inhibitorischen Rezeptoren herangezogen

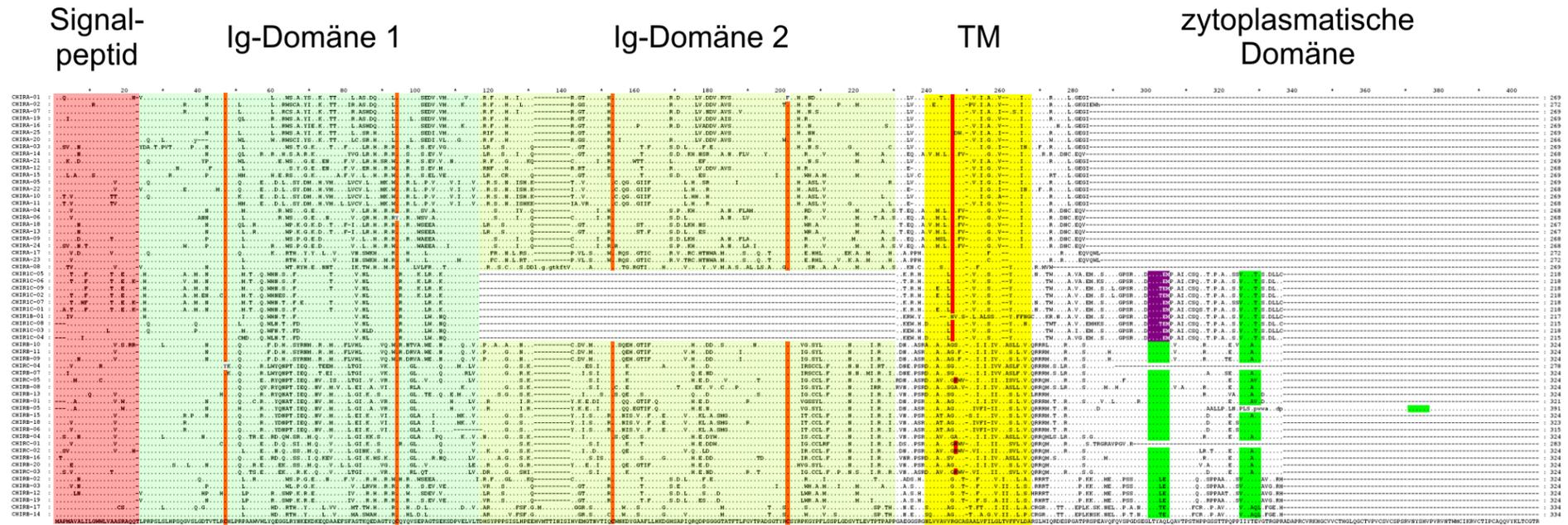


Abb. 16: Aminosäuresequenzvergleich aller 60 funktionellen CHIR-Rezeptoren. Aufgrund struktureller Unterschiede lassen sich die Sequenzen in zwei große Gruppen unterteilen. Die eine Gruppe läßt sich aufgrund der geladenen Aminosäure in der TM den aktivierenden Rezeptoren zuordnen, die andere den inhibitorischen Rezeptoren. Eine Besonderheit stellen zwei Gene dar, die strukturell zum Typ der inhibierenden Rezeptoren gehören, aber zusätzlich eine geladene Aminosäure in der TM besitzen. Die einzelnen Domänen wurden mit unterschiedlichen Farben hinterlegt. Die konservierten Cysteinreste der Ig-Domänen wurden orange hinterlegt und die geladenen Argininreste in der Transmembrandomäne wurden rot unterlegt. Im zytoplasmatischen Teil wurden ITIMs grün und YxxM-Motive lila hinterlegt.

werden, da nur aktivierende Rezeptoren eine geladene Aminosäure in der TM besitzen. Aus diesem Grund wurde besonderes Augenmerk auf diese Domäne gerichtet, um die einzelnen Rezeptoren zu klassifizieren. Der Sequenzvergleich zeigt, daß sich die Rezeptoren aufgrund von Ähnlichkeiten der TM in zwei Gruppen einteilen lassen. Gruppe I zeigt hohe Ähnlichkeit zur TM von CHIR-A, einem aktivierenden Rezeptor, während Gruppe II Ähnlichkeit zur TM von CHIR-B, einem inhibitorischen Rezeptor zeigt. Innerhalb von Gruppe II finden sich vier Gene, die zwar Sequenzähnlichkeit zu CHIR-B aufweisen, aber wie aktivierende Rezeptoren ein Arginin in der Transmembrandomäne tragen. Dieses Arginin ist jedoch im Vergleich zu den aktivierenden Rezeptoren um eine Position versetzt (Abb. 16) und wird von komplett anderen Basen kodiert. Daraus folgt, dass dieses Arginin sehr wahrscheinlich unabhängig von dem inaktiven Rezeptor entstanden ist.

Nach der groben Einteilung der Gene in die oben genannten zwei Gruppen wurden weitere typische Strukturmerkmale der bereits bekannten CHIRs herangezogen, um die Gene weiteren Untergruppen zuzuordnen und die möglichen Signaltransduktionskomponenten und damit die Funktion dieser Rezeptoren zu bestimmen.

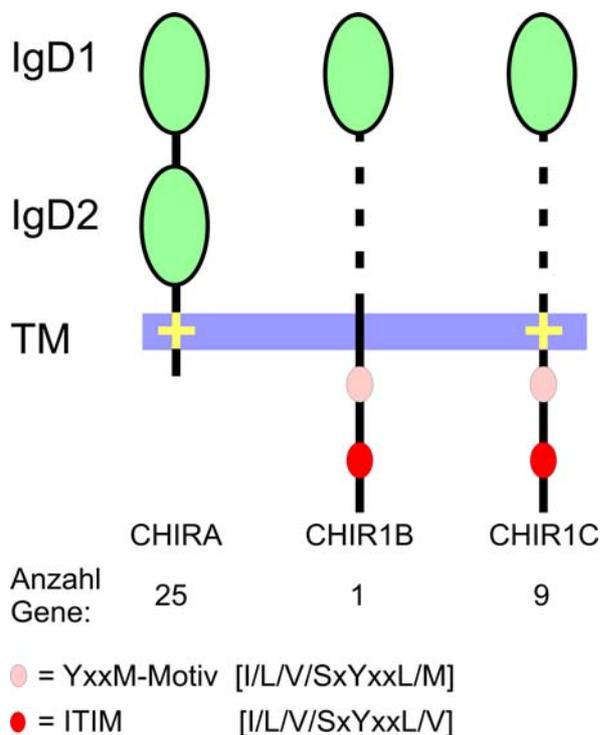


Abb. 17: Schematische Darstellung der Rezeptortypen CHIRA, CHIR1B und CHIR1C.

Rezeptorgruppe I

Ein Aminosäuresequenzvergleich aller Rezeptoren dieser Gruppe zeigt, daß sich aufgrund der Domänenstruktur zwei Rezeptortypen ableiten lassen (Abb. 17). Der Großteil der Gene entspricht strukturell dem bereits bekannten CHIR-A (25 Gene). Diese Gene besitzen zwei Ig-Domänen im extrazellulären Bereich, eine geladene Aminosäure in der TM und eine kurze zytoplasmatische Domäne. Weitere neun Gene bestehen aus sechs Exonen. Diesen Genen fehlt Exon 4, das für die zweite Ig-Domäne kodiert. Bei allen diesen Genen können zwar noch Sequenzhomologien zu Exon 4 gefunden werden, die Spleißstellen sind jedoch mutiert; es handelt sich also um ein Pseudoexon (Abb. 18). Die Gene dieser Gruppe

werden als „CHIR1C“ bezeichnet, da sie nur die äußere Ig-Domäne besitzen und gleichzeitig Charakteristika aktivierender und inhibitorischer Rezeptoren aufweisen. So besitzen sie eine

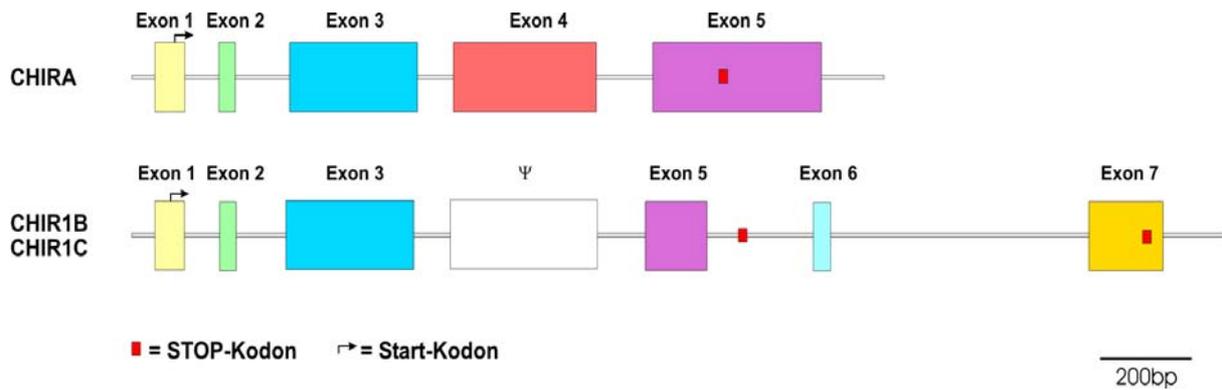


Abb. 18: Schematische Darstellung der Intron-Exon-Struktur der CHIRA- und CHIR1B/CHIR1C-Gene.

Die dünnen grauen Linien stellen die Introne, die farbigen Kästen die Exone dar. CHIRA besteht aus 5 Exonen. Zu CHIR1B/CHIR1C gehörende Gene haben sechs Exone, da die zu Exon 4 homologe Sequenz ein Pseudoexon darstellt.

geladene Aminosäure in der Transmembrandomäne, haben aber auch einen langen zytoplasmatischen Schwanz mit einem ITIM (VIYTTV) und einem ITIM-ähnlichen Motiv (LTY[A/T]EM), das als YxxM-Motiv bezeichnet wird (Abb. 17). Bei DAP10 konnte gezeigt werden, daß das YxxM-Motiv die p85-Untereinheit der PI3K rekrutiert (Wu et. al. 2000). Dieses Motiv könnte also aktivierende Signale weiterleiten. Bei einem der bifunktionellen Gene ist die Transmembrandomäne so mutiert, daß sie keine geladene Aminosäure trägt. Dieser als CHIR1B bezeichnete Rezeptor hat aber ein ITIM und ein YxxM-Motiv behalten. Interessanterweise besitzen alle diese Gene im Intron 5 (zwischen Exon 5 und Exon 6) an der gleichen Position ein im Leseraster befindliches Stopkodon (Abb. 18). Sollte Intron 5 also nicht herausgespleißt werden, würde die mRNA in einen aktivierenden Rezeptor mit einer Ig-Domäne übersetzt werden (man könnte diesen Typ als CHIR1A bezeichnen). Es wäre zum Beispiel denkbar, daß die zwei Varianten durch alternatives Spleißen entstehen. Ob diese Varianten in vivo translatiert werden, kann nur über Proteinanalysen nachgewiesen werden, da ESTs und cDNAs in der Regel aus Gesamt-RNA der Zelle hergestellt werden, die viele unvollständig gespleißte Transkripte enthält.

Der Sequenzvergleich aller Gruppe-I-Rezeptoren zeigt, daß die CHIR1B/CHIR1C-Rezeptoren untereinander stärker konserviert sind, als die CHIRA-Rezeptoren (siehe Abb. 16). Eine mögliche Erklärung hierfür ist, daß diese Rezeptorgruppe erst in der jüngeren Vergangenheit expandierte und deshalb noch keine so große Sequenzdiversität wie die CHIRA-Rezeptorgruppe zeigt. Eine andere Möglichkeit besteht darin, daß diese Gene stärker konserviert sind, da sie zum Beispiel mit einem nicht-polymorphen Liganden interagieren, oder weil schon wenige Sequenzveränderungen zu einem Funktionsverlust führen würden.

In Tab. 2 sind zusammenfassend alle Gene der Typen CHIRA, CHIR1B und CHIR1C und ihre Contig-Zugehörigkeit aufgelistet. Auffällig ist, daß CHIRA-Gene gehäuft in Contig 1 kodiert sind.

CHIRA-Gene				CHIR1B/CHIR1C-Gene			
Contig 1	Contig 2	Contig 3	Contig 4	Contig 1	Contig 2	Contig 3	Contig 4
CHIRA-01	CHIRA-14	CHIRA-19	CHIRA-23	CHIR1C-01	CHIR1C-02	CHIR1C-04	CHIR1C-07
CHIRA-02	CHIRA-15	CHIRA-20	CHIRA-24		CHIR1C-03	CHIR1C-05	CHIR1C-08
CHIRA-03	CHIRA-16	CHIRA-21	CHIRA-25			CHIR1C-06	CHIR1C-09
CHIRA-04	CHIRA-17	CHIRA-22					CHIR1B-01
CHIRA-05	CHIRA-18						
CHIRA-06							
CHIRA-07							
CHIRA-08							
CHIRA-09							
CHIRA-10							
CHIRA-11							
CHIRA-12							
CHIRA-13							

Tab. 2 Liste aller Gene von Gruppe-I-Rezeptoren.

Die einzelnen Gene sind den Typen CHIRA und CHIR1C zugeordnet und nach ihrer Zugehörigkeit zu den Contigs sortiert.

Die Exon-Intron-Grenzen aller Gene folgen der AG-GT-Regel. In Abb. 19 sind exemplarisch für ein CHIRA-Gen die Sequenzen aller Exon-Intron-Grenzen dargestellt.

	Exon 1 (65 bp) 5' UTR/„leader“	CATCCTTGgtgagtga	Intron 1 73 bp	...
...cettgcagGTTGGTGG	Exon 2 (36 bp) „leader“	ACAGCACGgtgagtat	Intron 2 116 bp	...
...cgtcctagTGCCCCGA	Exon 3 (279 bp) Ig-Domäne 1	GCTGACAGgtgaggac	Intron 3 79 bp	...
...tcatgcagATCGCAGT	Exon 4 (312 bp) Ig-Domäne 2	ACCCCCAGgtaggtac	Intron 4 121 bp	...
...ccccacagGTGCTGAG	Exon 5 (366 bp) Stamm/Tm/zytoplasmatische Domäne/ 3' UTR	aaaaaaaaaaaaa		...

Abb. 19 Exon-Intron-Struktur für ein CHIRA-Typ-Gen (CHIRA-01)

In Abb. 20 ist ein Aminosäure-Sequenzvergleich von je zwei CHIRA- und CHIR1C-Rezeptoren dargestellt. In den Ig-Domänen sind bei den Proteinen an den gleichen vier Positionen Cysteine konserviert, die für die Stabilität der Ig-Domänen von entscheidender Bedeutung sind. Die CHIR1C-Rezeptoren haben ein ITIM und ein YxxM-Motiv in der zytoplasmatischen Domäne. Alle Gruppe-I-Rezeptoren tragen darüber hinaus eine geladene Aminosäure in der Transmembrandomäne (siehe auch Abb. 16). Den Rezeptoren des Typs CHIR1C fehlt die Ig-Domäne 2.

Innerhalb der Gruppe der CHIRA-Rezeptoren zeigen einige Gene Besonderheiten, auf die im Folgenden eingegangen werden soll. Bei zwei Genen, *CHIRA-01* und *CHIRA-04*, ist jeweils eines der zwei konservierten und für die Faltung der Ig-Domäne wichtigen Cysteine

mutiert. Bei *CHIRA-01* ist das zweite Cystein der zweiten Ig-Domäne zu Phenylalanin mutiert, was durch eine Mutation der zweiten Base (G→T) des zugehörigen Triplets hervorgerufen wurde. Bei *CHIRA-04* ist das zweite Cystein der ersten Ig-Domäne zu Tyrosin mutiert, was ebenfalls durch eine Mutation der zweiten Base (G→A) des zugehörigen Triplets verursacht wurde. Ob diese Proteine korrekt oder zumindest funktionell gefaltet werden, kann nur durch Untersuchung der Proteine bestimmt werden. Vermutlich ist aber die Stabilität der Ig-Domänen so stark beeinträchtigt, daß diese Proteine nicht exprimiert werden. Eine weitere Besonderheit zeigt das Gen *CHIRA-08*, dessen abgeleitetes Protein zehn zusätzliche Aminosäuren am Anfang der Ig-Domäne 2 trägt (siehe Abb. 20). Möglicherweise hätte dies Konsequenzen für die Ligandenbindung. Darüber hinaus zeigt sich bei einer Reihe von Rezeptoren eine gewisse Variabilität der Anzahl der Aminosäuren in der Stamm-Region, direkt distal der TM und bei der Anzahl der Aminosäuren in der zytoplasmatischen Domäne, was durch unterschiedliche Stop-Kodons hervorgerufen wird. Beides hat vermutlich keine Konsequenzen für die Funktion der betroffenen Proteine.

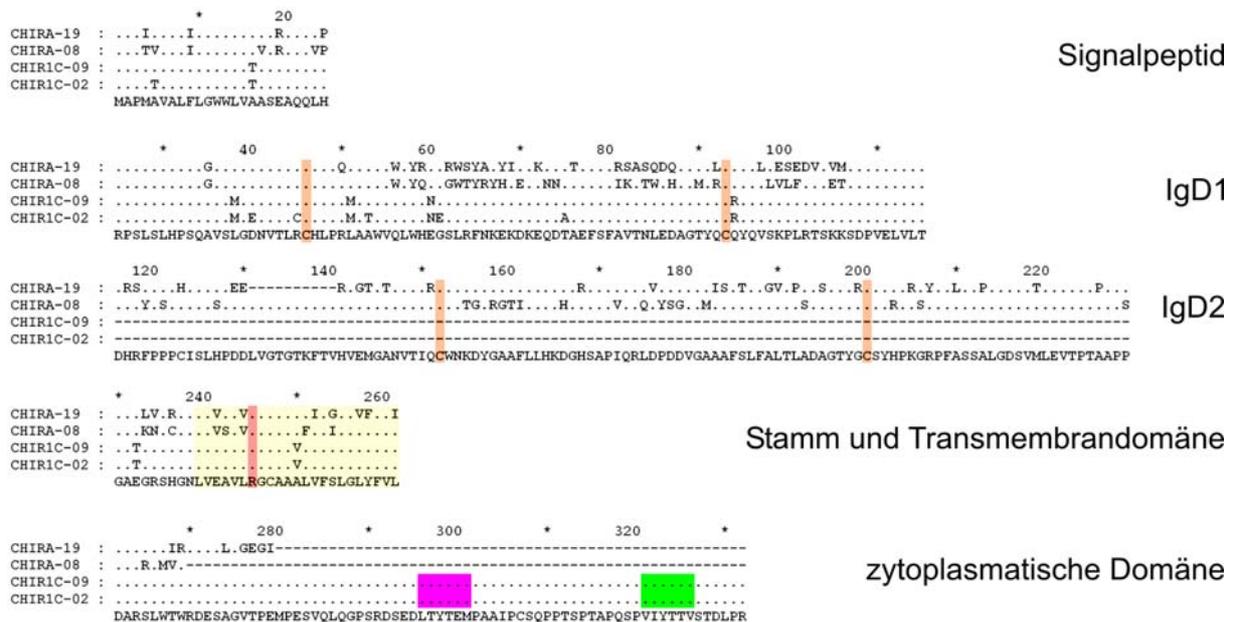
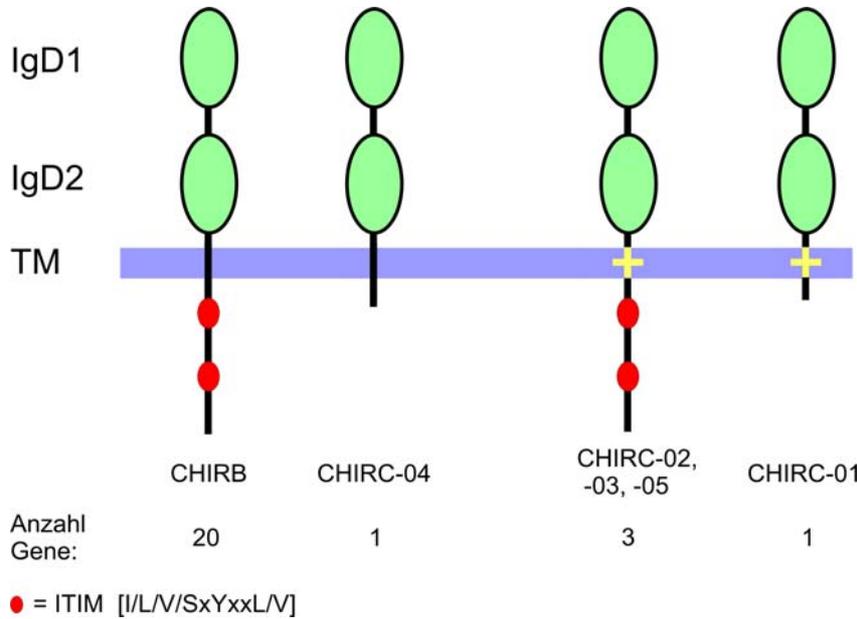


Abb. 20: Aminosäure-Sequenzvergleich von je zwei Genen pro aktivierendem Rezeptor-Typ. Der Sequenzvergleich ist in die einzelnen Proteindomänen aufgeteilt. *CHIRA-19* und *CHIRA-08* sind *CHIRA*-Rezeptoren, wobei *CHIRA-08* in Ig-Domäne 2 zehn zusätzliche Aminosäuren besitzt. Die Gene *CHIRIC-09* und *CHIRIC-02* sind Rezeptorgene des Typs *CHIRIC*. Die konservierten Cysteine der Ig-Domänen sind orange unterlegt, die Transmembrandomäne gelb, die geladene Aminosäure Arginin rot, das YxxM-Motiv rosa und das ITIM grün.

Rezeptorgruppe II

Auch die Gene der Rezeptorgruppe II lassen sich aufgrund ihrer Domänenstruktur in verschiedene Gruppen unterteilen. Die größte Gruppe bilden CHIRB-ähnliche Rezeptoren, die auch eine Sonderform beinhaltet (*CHIRC-04*), bei der der zytoplasmatische Schwanz verkürzt ist.



ist. Auf die zweite Gruppe wurde bereits oben kurz eingegangen; ihre Mitglieder weisen eine geladene Aminosäure in der Transmembrandomäne auf. Die TM unterscheidet sich jedoch in ihrer Aminosäuresequenz von denen der CHIR-A-ähnlichen Rezeptoren, und auch das Arginin ist um eine Aminosäure verschoben. Aufgrund der Länge der zytoplasmatischen Domäne lassen sich die Rezeptoren dieser Gruppe noch weiter unterteilen. Insgesamt läßt sich die Rezeptorgruppe II also in vier Untergruppen teilen (Abb. 21).

Abb. 21: Mögliche Mitglieder der Rezeptor-Gruppen II, die Homologie zu inhibitorischen CHIRs zeigen.

Von den insgesamt 25 Genen der Gruppe II läßt sich der überwiegende Teil (20 Gene) dem

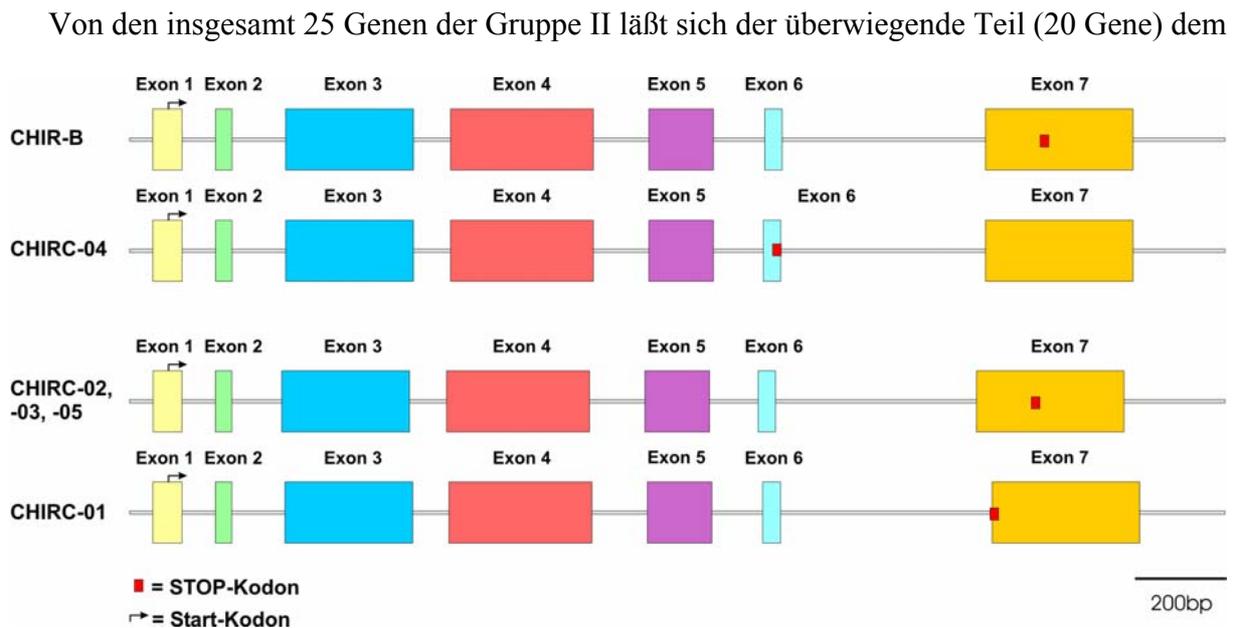


Abb. 22: Schematische Darstellung der Exon-Intron-Struktur der verschiedenen inhibitorischen CHIR-Gen-Typen.

Typ CHIRB zuordnen. Diese Gene haben sieben Exone (Abb. 22), die für ein Protein mit zwei Ig-Domänen, einer ungeladenen Transmembrandomäne und einer langen zytoplasmatischen Domäne mit zwei ITIMs kodieren. Das Gen *CHIRB-05* besitzt zwei zusätzliche Nukleotide in Exon 7, wodurch es nach dem ersten ITIM zu einer Verschiebung des Leserahfers kommt. Daraus resultiert im Vergleich zu anderen CHIRB-Rezeptoren ein längerer zytoplasmatischer Schwanz, in dem ein zweites ITIM folgt. Der Abstand zwischen beiden ITIMs ist jedoch, im Vergleich zu den anderen CHIRB-Rezeptoren, vergrößert. Trotz dieser Abweichung wird dieses Gen der Gruppe der CHIR-B-Rezeptoren zugerechnet, da es alle für diese Gruppe notwendigen Strukturmerkmale aufweist. Im Gen *CHIRC-04* bilden die letzten drei Basen des Exon 6 ein STOP-Codon. Dadurch terminiert das Protein frühzeitig und die zytoplasmatische Domäne enthält keine ITIMs (Abb. 21). Vier weitere Gene folgen strukturell den CHIRB-Rezeptoren; interessant ist jedoch, daß sie in der Transmembrandomäne eine geladene Aminosäure (Arginin) besitzen. Dieses für aktivierende Rezeptoren der Gruppe I typische Merkmal ist in diesem Fall mit einem langen zytoplasmatischen Schwanz, der ITIMs trägt, kombiniert. Es handelt sich also um bifunktionelle Rezeptoren, die sowohl zur Signaltransduktion von aktivierenden als auch von inhibierenden Signalen befähigt sein sollten. Diese Gene wurden als CHIRC klassifiziert, wobei eines dieser vier Gene wiederum eine Ausnahme bildet. Das Gen *CHIRC-01* besitzt ein zusätzliches Nukleotid in Exon 6, wodurch es zu einer Leserasterverschiebung kommt. Dadurch terminiert das Protein frühzeitig in Exon 7 und besitzt nur einen kurzen zytoplasmatischen Schwanz (Abb. 22). Dieser Rezeptor entspricht strukturell Rezeptoren der CHIRA-Gruppe. Es sollte hier jedoch nochmals darauf hingewiesen werden, daß Sequenzhomologien zeigen, daß dieser Rezeptor zur Familie der CHIRB-Rezeptoren gehört.

Die Rezeptorgene der Gruppe II sind im Gegensatz zu den Genen der Gruppe I relativ gleichmäßig über die Contigs verteilt. Relativ wenige CHIR-B-Gene sind im Contig 2 vorhanden; dieses Contig enthält aber auch relativ viele Pseudogene (siehe unten). Tab. 3 gibt einen Überblick über die Verteilung der Gruppe-II-Rezeptorgene über die Contigs.

CHIRB-Gene				CHIRC-Gene	
Contig 1	Contig 2	Contig 3	Contig 4	Contig 2	Contig 3
CHIRB-01	CHIRB-08	CHIRB-10	CHIRB-16	CHIRC-01	CHIRC-05
CHIRB-02	CHIRB-09	CHIRB-11	CHIRB-17	CHIRC-02	
CHIRB-03		CHIRB-12	CHIRB-18	CHIRC-03	
CHIRB-04		CHIRB-13	CHIRB-19		
CHIRB-05		CHIRB-14	CHIRB-20		
CHIRB-06		CHIRB-15			
CHIRB-07		CHIRC-04			

Tab. 3 Gene der Rezeptorgruppe II und ihre Zuordnung zu den einzelnen Contigs.

Bei der Charakterisierung der Gene wurden nur die Exone als funktionell betrachtet, deren Exon-Intron-Grenzen der AG-GT-Regel folgen. Exemplarisch sind in Abb. 23 alle Exon-Intron-Grenzen eines CHIRB-Rezeptorgens dargestellt.

	Exon 1 (64 bp) 5' UTR/„leader“	TGCAGTCTgtgagtga...	Intron 1 73 bp	...
...ccttgcag	Exon 2 (36 bp) „leader“	ACAGCAACgtgagtat...	Intron 2 117 bp	...
...tgctcag	Exon 3 (279 bp) Ig-Domäne 1	GGTGACAGgtgagggc...	Intron 3 81 bp	...
...gtgctcag	Exon 4 (312 bp) Ig-Domäne 2	ACCCCCAGgttaggtct...	Intron 4 121 bp	...
...cccctcag	Exon 5 (141 bp) Stamm/Tm/zytoplasmatische Domäne	GAGGCCTGgtggggaa...	Intron 5 112 bp	...
...tgactcag	Exon 6 (38 bp) Zytoplasmatische Domäne	AGTTGCAGgtcagtg...	Intron 6 444 bp	...
...gtccccag	Exon 7 (322 bp) Zytoplasmatische Domäne /3' UTR	aaaaaaaaaaaaa		

Abb. 23 Exon-Intron-Struktur eines CHIRB-Typ-Gens (*CHIRB-01*)
Das CHIRB-Gen besteht aus sieben Exonen. Die Exon-Sequenzen sind in Großbuchstaben dargestellt, Intronsequenzen in kleinen Buchstaben. Alle Exon-Intron-Grenzen folgen der AG/GT-Regel.

Ein Aminosäuresequenzvergleich einiger ausgewählter Gruppe-II-Rezeptoren (Abb. 24) zeigt wie bei den Rezeptoren der Gruppe I, daß die größte Variabilität zwischen den verschiedenen Rezeptoren innerhalb der Ig-Domänen und in der Stammregion beobachtet wird. Die Sequenz des Signalpeptids, die Grenzen der Ig-Domänen, die TM und der zytoplasmatische Schwanz ist im Gegensatz dazu relativ konserviert. Bei allen Genen sind bis auf eine Ausnahme (*CHIRC-04*, Abb. 16) die vier Cysteine in den zwei Ig-Domänen vorhanden. Hiervon ist vermutlich die Stabilität der Ig-Domäne betroffen, wie es bei den zwei CHIRA-Genen ebenfalls beobachtet wurde. Die bei den Genen der Gruppe CHIRC auftretenden positiv geladenen Aminosäuren (Arginin) in der Transmembrandomäne und die ITIMs im zytoplasmatischen Teil sind farbig unterlegt. Wie bereits weiter oben beschrieben, ist die Position des Arginins im Vergleich zu den klassischen aktivierenden Rezeptoren um eine Position zum C-Terminus hin versetzt; die ITIMs der drei bifunktionellen Rezeptoren (CHIRC) befinden sich jedoch an der gleichen Position wie bei den CHIRB-Rezeptoren. Bei CHIRC-01 ist die zytoplasmatische Domäne verkürzt und enthält keine ITIMs (Abb. 24).

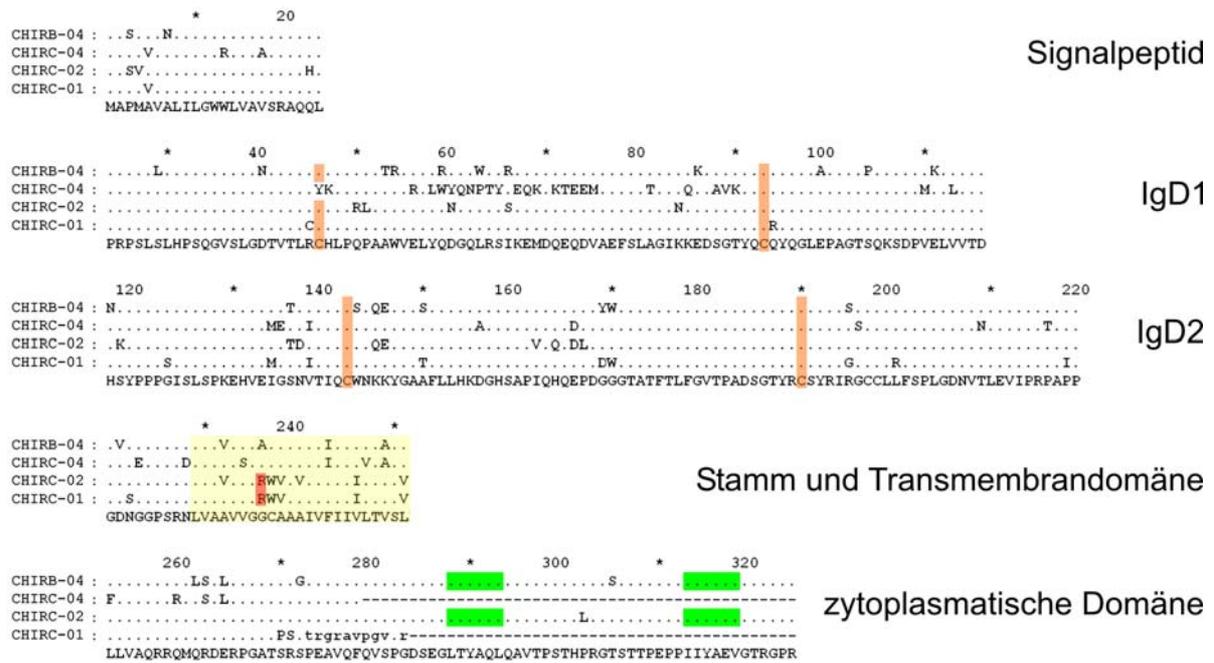


Abb. 24: Aminosäuren-Sequenzvergleich von je einem Gen pro inhibierenden Rezeptor-Typ CHIRB-04 ist ein Vertreter für den Typ CHIRB, CHIRC-04 ist ein CHIRB-Rezeptor mit zytoplasmatischem Schwanz ohne ITIMs, CHIRC-02 ist ein bifunktionaler Rezeptor und CHIRC-01 ist ein aktivierender Rezeptor mit Homologie zu inhibitorischen CHIRs. Orange unterlegt sind die konservierten Cysteine, gelb die Transmembrandomäne, rot die Arginine und grün die ITIMs.

Pseudogene/Genfragmente: CHIR-P

Neben Genen, aus denen sich funktionelle Transkripte ableiten ließen, fanden sich in den BAC-Sequenzen auch 43 nicht-funktionelle Gene (Pseudogene) oder Genfragmente. Diese wiesen entweder nur kurze Homologien zu CHIR-A oder -B auf, besaßen mutierte Spleißdonor- und/oder Spleißakzeptormotive, enthielten Deletionen, Leserasterverschiebungen oder vorzeitige STOP-Kodons. Tab. 4 gibt einen Überblick über die Charakteristika der Pseudogene und Genfragmente.

Im Zusammenhang mit den später durchgeführten Transkriptanalysen sind einige als Pseudogen charakterisierte CHIR-Gene von Interesse, weil sich aus ihnen bis auf ein vorzeitiges STOP-Kodon in Exon 3 oder 4 ein funktionelles Transkript ableiten läßt (CHIRP-01, CHIRP-04, CHIRP-17, CHIRP-18, CHIRP-30 und CHIRP-32). Wird Exon 3 oder Exon 4 mit den flankierenden Intronen aus den Transkripten herausgespleißt, ergäbe sich eine mRNA, die für einen funktionellen Rezeptor mit nur einer Ig-Domäne kodiert. Die durchgeführten Transkriptanalysen zeigen, daß es Transkripte zumindest mit Sequenzen für nur eine Ig-Domäne gibt. Das bedeutet, daß es prinzipiell CHIR-Rezeptoren mit nur einer Ig-Domäne geben könnte. Ob entsprechende Transkripte von den hier erwähnten Pseudogenen abgelesen und gespleißt werden können, ist jedoch unklar, weil kein Transkript mit vollständiger Sequenzübereinstimmung mit diesen Genen gefunden wurde (siehe Abschnitt 4.3.4).

	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7
CHIRP-37	→ + ←	→ + ←	→ + ←	→ + ←	+ ←	→ + ←	→ + ←
CHIRP-4	→ + ←	→ + ←	→ +S←	→ + ←	→ + ←	→ + ←	→ + ←
CHIRP-32	→ + ←	→ + ←	→ +S←	→ + ←	→ + ←	→ + ←	→ + ←
CHIRP-18	→ + ←	→ + ←	→ + ←	→ +S←	→ + ←	→ + ←	→ +
CHIRP-24	→ + ←	→ + ←	→ + ←	→ +S	+F	→ + ←	→ +
CHIRP-9	→ + ←	→ + ←	→ + ←	→ +	→ + ←	→ + ←	→ + ←
CHIRP-43	→ + ←	→ + ←	→ +S←	→ + ←	→ + ←	+	-
CHIRP-25	→ + ←	→ + ←	→ + ←	+	+ ←	-	+ ←
CHIRP-27	-	→ + ←	→ + ←	+	→ + ←	+	→ + ←
CHIRP-30	→ + ←	→ + ←	→ +S←	→ + ←	→ +	-	+
CHIRP-1	→ + ←	→ + ←	→ + ←	→ +S←	→ + ←	-	-
CHIRP-17	→ + ←	→ + ←	→ +S←	→ + ←	→ +	-	-
CHIRP-14	→ + ←	→ + ←	→ + ←	→ + ←	+	-	-
CHIRP-6	→ + ←	→ +FS←	→ + ←	→ +	-	-	-
CHIRP-8	→ + ←	→ + ←	→ +FS←	→ + ←	-	-	-
CHIRP-38	→ + ←	→ + ←	→ +S←	→ +	-	-	-
CHIRP-21	→ + ←	→ + ←	→ + ←	→ +S	-	-	-
CHIRP-3	→ + ←	→ +	→ + ←	+	-	-	-
CHIRP-7	→ + ←	→ + ←	→ + ←	+	-	-	-
CHIRP-10	→ + ←	→ + ←	→ + ←	+	-	-	-
CHIRP-19	→ + ←	→ + ←	→ +S←	+	-	-	-
CHIRP-29	→ + ←	→ + ←	→ + ←	-	+	-	+
CHIRP-11	→ + ←	→ + ←	→ + ←	-	-	-	+
CHIRP-12	→ + ←	→ + ←	→ + ←	-	-	-	-
CHIRP-13	→ + ←	→ + ←	→ + ←	-	-	-	-
CHIRP-35	→ + ←	→ + ←	→ +S←	-	-	-	-
CHIRP-23	→ + ←	→ + ←	→ + ←	-	-	-	-
CHIRP-16	→ + ←	→ + ←	→ +	-	-	-	-
CHIRP-22	→ + ←	→ + ←	→ +	-	-	-	-
CHIRP-39	→ + ←	→ + ←	→ +	-	-	-	-
CHIRP-42	→ + ←	→ +	→ + ←	-	-	-	-
CHIRP-40	→ + ←	+	→ + ←	-	-	-	-
CHIRP-2	+	-	-	+	-	-	-
CHIRP-20	-	-	+	-	-	-	-
CHIRP-36	-	-	+	-	-	-	-
CHIRP-41	-	-	+	-	-	-	-
CHIRP-5	-	-	-	→ +	-	-	-
CHIRP-26	-	-	-	+	-	-	-
CHIRP-28	-	-	-	+	-	-	-
CHIRP-31	-	-	-	+	-	-	-
CHIRP-33	-	-	-	-	-	-	+
CHIRP-34	-	-	-	-	-	-	+
CHIRP-15	-	-	-	-	-	-	+

Tab. 4 Charakteristika der Pseudogene und Genfragmente

Die Tabelle gibt Auskunft über die Fehler, die dazu führen, daß die entsprechenden Gene als Pseudogene klassifiziert wurden. Die Pseudogene sind nach der Schwere der Fehler sortiert. Pfeile geben das Vorhandensein der Spleiß-Donor bzw. -Akzeptorstellen an; + bedeutet, daß Homologie zu der entsprechenden Sequenz im Gen vorhanden ist; S markiert ein frühzeitiges STOP-Kodon; F zeigt Verschiebungen des Leserasters an. - bedeutet, daß keinerlei Sequenzübereinstimmung vorhanden ist.

4.3 Charakterisierung von Transkripten der CHIR-Gene

Da aus der Analyse der DNA-Sequenzen der CHIR-BACs nicht abgeleitet werden kann, ob ein Gen tatsächlich abgelesen wird, habe ich neben der Charakterisierung der CHIR-Gene auch Experimente und Analysen durchgeführt, um die zugehörigen Transkripte zu identifizieren. Dies sollte Rückschlüsse auf die Zahl und Identität der abgelesenen Gene zulassen. Zu

Beginn meiner Arbeit waren nur Transkripte von zwei Genen beschrieben worden (CHIR-A und CHIR-B) (Dennis, Jr. et. al. 2000). Zusätzlich wurden kürzlich vier weitere Transkripte beschrieben, bei denen es sich um CHIRB-Rezeptoren handelt. Sie wurden als CHIR-B1, -B2, -B3 und -B4 bezeichnet, keines der beschriebenen Transkripte zeigt allerdings eine vollständige Übereinstimmung mit einem der hier beschriebenen Gene, weshalb eine Zuordnung nicht möglich war. Einer dieser Rezeptor (CHIR-B2) wurde zudem funktionell charakterisiert (Viertlboeck et. al. 2004).

Die Transkriptanalyse erfolgte auf drei Wegen: Zum einen habe ich eine im Labor verfügbare cDNA-Bibliothek aus der Bursa fabricii eines Huhnes nach LILR-homologen Klonen durchmustert, zum anderen wurden mit verschiedenen spezifischen Primern CHIR-Transkripte aus selbst hergestellter cDNA aus peripheren Leukozyten des Huhns amplifiziert. Zusätzlich wurde eine „Expressed Sequence Tag“ (EST)-Datenbank mit Hilfe einer CHIR-Sequenz durchmustert und die gefundenen Sequenzen analysiert.

4.3.1 Durchmusterung einer Hühner-cDNA-Bibliothek

Eine cDNA-Bibliothek aus der Bursa fabricii eines Huhnes, deren Klone auf „Makro-Array“-Membranen aufgetragen waren, wurde mit der LILRA2 (ILT1)-Sonde hybridisiert. Zehn LILR-homologe Klone konnten identifiziert werden. Diese wurden beim Deutschen Ressourcenzentrum für Genomforschung (RZPD) bestellt, und nach einer Plasmidpräparation sequenziert. Bei allen Klonen handelte es sich um Transkripte von CHIR-Genen. Die sequenzierten cDNAs wurden mit der Sequenz eines CHIR-A- und eines CHIR-B-Rezeptorgens verglichen. Es stellte sich heraus, daß keines der Transkripte einer korrekt maturierten mRNA dieser Gene entsprach. Entweder fehlten ihnen am 5'-Ende Basenpaare, sie enthielten noch Introne, oder es fehlten Exone. Bei mRNAs mit fehlenden Exonen kann es sich zwar auch um Produkte handeln, die durch alternatives Spleißen entstanden sind; wenn jedoch essentielle Exone fehlen, dann ist die korrekte Prozessierung des resultierenden Proteins fraglich. Weiterhin kann das Fehlen von Exonen zu einer Verschiebung des Leserasters und damit zu einem frühzeitigen Abbruch der Translation führen. Die Auswertung und schematische Darstellung dieser Sequenzen ist in Abb. 25 zusammenfassend dargestellt.

Aufgrund ihrer Struktur lassen sich die cDNAs in drei Gruppen einteilen. Die Klone L2437 und K0466 sind an ihrem 5'-Ende verkürzte mRNAs; es fehlen die Exone E1 und E2, bzw. E1 bis E4, wobei letzteres nur unvollständig vorhanden ist. Ursächlich hierfür sind vermutlich Schäden an der mRNA, die bei der Präparation auftreten können (Abbrechen der Enden), oder es handelt sich um teilweise degradierte mRNAs. F0410 fehlt Exon 2, was durch falsches

Spleißen entstanden sein kann, und die Sequenz von Exon 6 ist unvollständig. Aus keinem der drei Klone läßt sich ein Protein ableiten, das einen membranständigen Rezeptor mit funktionellem Signalpeptid repräsentiert.

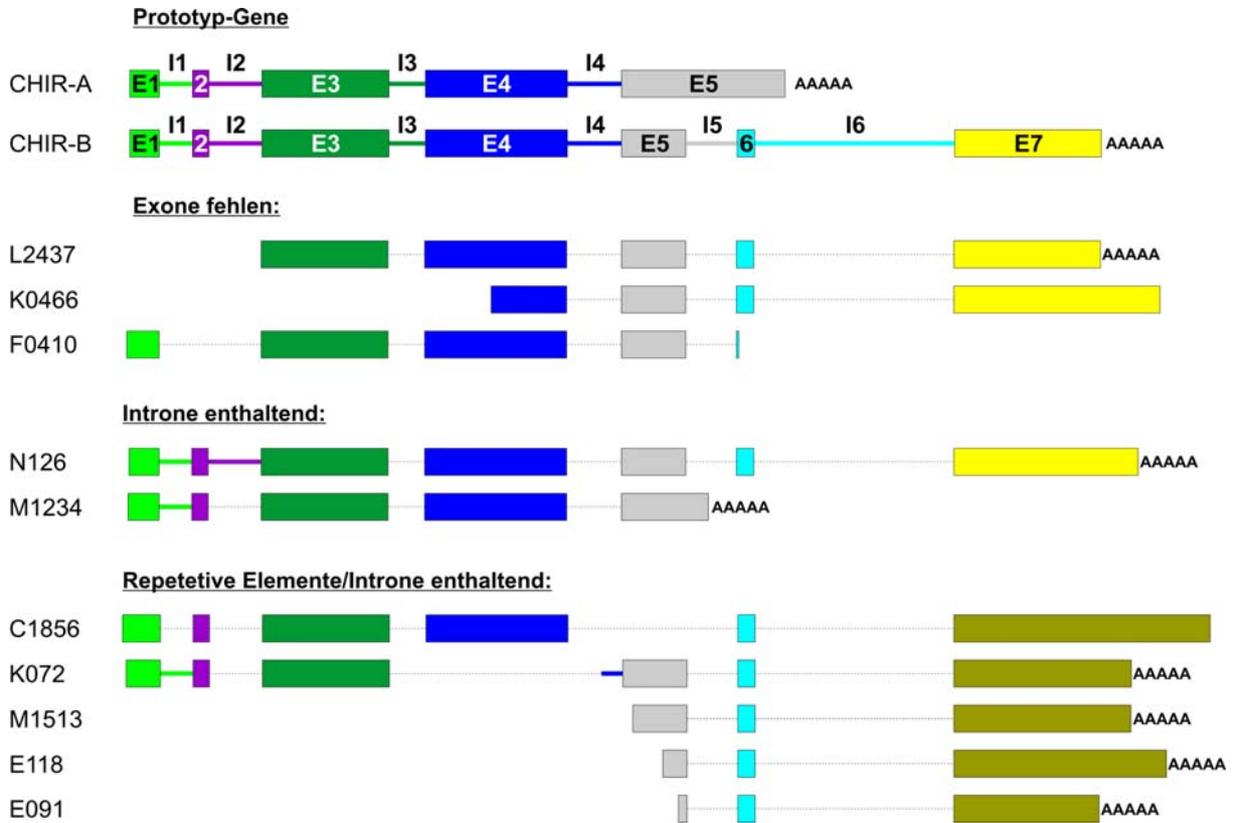


Abb. 25: Schematische Darstellung der Exon-Intron-Struktur der sequenzierten Bursa-cDNA-Klone.

Die in den cDNAs vorhandenen Sequenzen sind entsprechend der oben angegebenen Prototyp-Sequenzen von CHIR-A und CHIR-B farblich markiert, fehlen die Sequenzen, wurde nur eine graue Verbindungslinie eingezeichnet. Die Sequenzen wurden entsprechend der Fehlerhaftigkeit gruppiert. Exone mit Homologie zu Exon 7, die repetitive Elemente enthalten, sind olivfarben statt gelb markiert, um diesen Unterschied zu verdeutlichen. E=Exon; I=Intron

Neben verkürzten Sequenzen beziehungsweise fehlenden Exonen gibt es auch Transkripte, die durch fehlerhafte Spleißung noch Introne enthalten. Solche Transkripte wurden der zweiten Gruppe zugeordnet. Dieses trifft für den Klon N126 (Intron 1 und 2 enthaltend) und den Klon M1234 (Intron 1 enthaltend) zu. Entfernt man die Intron-Sequenzen, so ergibt sich für N126 aus der Aminosäuresequenz ein Rezeptor vom Typ CHIRB, für M1234 ein Rezeptor vom Typ CHIRB mit verkürztem zytoplasmatischem Schwanz (vergleichbar zu CHIRC-04).

Die dritte Gruppe enthält Sequenzen, bei denen zum einen unterschiedlich lange Abschnitte am 5'-Ende fehlen, und zum anderen noch Reste von Intronen enthalten sind. Gemeinsam ist dieser Gruppe, daß in Exon 7 repetitive Elemente (Mariner-Transposon) enthalten sind. Dieses Erscheinung ist nicht ungewöhnlich und kommt bei verschiedensten Genen/Transkripten insbesondere im 3'-Bereich vor. Die repetitiven Elemente haben keine Auswirkung auf die Funktion des Rezeptors; sie liegen meist in einem Sequenzbereich, der

nicht translatiert wird und somit für die Funktion eines sich daraus ableitenden Proteins unbedeutend ist (3' des STOP-Kodons). Die Klone M1513, E118 und E091 enthalten zu wenig Sequenz, als daß man ein Protein aus der cDNA ableiten könnte. C1856 fehlt Exon 5. Die Sequenz von K072 läßt auf einen Rezeptor schließen, der aus einer Ig-Domäne, einer ungeladenen Transmembrandomäne und einem langen zytoplasmatischen Schwanz besteht. Die zytoplasmatische Domäne enthält aber keine ITIMs.

Durch einen Sequenzvergleich aller Bursa-cDNAs miteinander konnte festgestellt werden, daß einige der Klone Transkripte des gleichen Gens darstellen, da sie eine vollständige Sequenzübereinstimmung zeigen. Insgesamt war von zwei Genen oder Allelen mehr als ein Transkript in der Bursa-cDNA-Bibliothek enthalten. Bei L2437 und N126 handelt es sich um Transkripte des einen Gens, sowie bei K072, M1513 und E118 um Transkripte des anderen Gens. Aus der Bursa-Bibliothek konnten also Transkripte von sieben Genen/Allelen isoliert und charakterisiert werden. Interessanterweise konnte zu keinem der hier sequenzierten Klone ein Gen in den BAC-Sequenzen gefunden werden, dessen Sequenz vollständige Übereinstimmung zeigte. Hiefür kommen zwei Gründe in Frage: Zum einen ist es, da die einzelnen Contigs nicht überlappen und zahlreiche weitere CHIR-positive BACs gefunden wurden, wahrscheinlich, daß es mehr als die hier gefundenen CHIR-Gene gibt. Zum anderen könnte es bedeuten, daß CHIRs polymorph sind, und sich die Allele eines Gens bei verschiedenen Individuen (die BAC-Bibliothek und Bursa-cDNA-Bibliothek stammen nicht von demselben Tier) unterscheiden. Für letztere Überlegung sprechen vor allem die Ergebnisse der Auswertung weiterer Transkripte in den folgenden Abschnitten. Die mögliche Zugehörigkeit der Transkripte zu einzelnen CHIR-Genen wird in Abschnitt 4.3.4 zusammen mit den im Folgenden analysierten Transkripten diskutiert.

4.3.2 Herstellung von Hühner-cDNA und deren Verwendung für CHIR-spezifische PCRs

Neben der Durchmusterung einer bereits vorhandenen cDNA-Bibliothek habe ich auch CHIR-Transkripte aus selbst hergestellter cDNA amplifiziert. Dazu wurden aus dem Blut eines Huhns Leukozyten isoliert. Aus den Leukozyten wurde dann die Gesamt-RNA isoliert und mittels Reverser-Transkriptase-(RT)-Reaktion Gesamt-cDNA synthetisiert. Die cDNA diente dann als Matrize für die folgenden PCRs mit drei verschiedenen Primerkombinationen (Abb. 26). Die Primer wurden, basierend auf den Sequenzen der Bursa-cDNA-Klone und der CHIR-A- und CHIR-B-Sequenzen so gewählt, daß sie an möglichst konservierte Sequenzen banden, so daß die Amplifikation aller CHIR-Transkripte zumindest mit einer der Primer-

kombinationen möglich sein sollte. Der Primer Chir all Fw 01 wurde von Exon 1 abgeleitet, CHIR all Fw 02 aus dem Anfang von Exon 3. Als Gegenprimer wurden zum einen die von Dennis und Kollegen (Dennis, Jr. et. al. 2000) entworfenen Primer verwendet, die an CHIR-A-Transkripte in Exon 5 bzw. an CHIR-B-Transkripte in Exon 7 binden können. Zum anderen wurde der Primer SMART_Tag_1 (Adapter-Sequenz) benutzt, der an das 3'-Ende aller cDNAs binden kann.

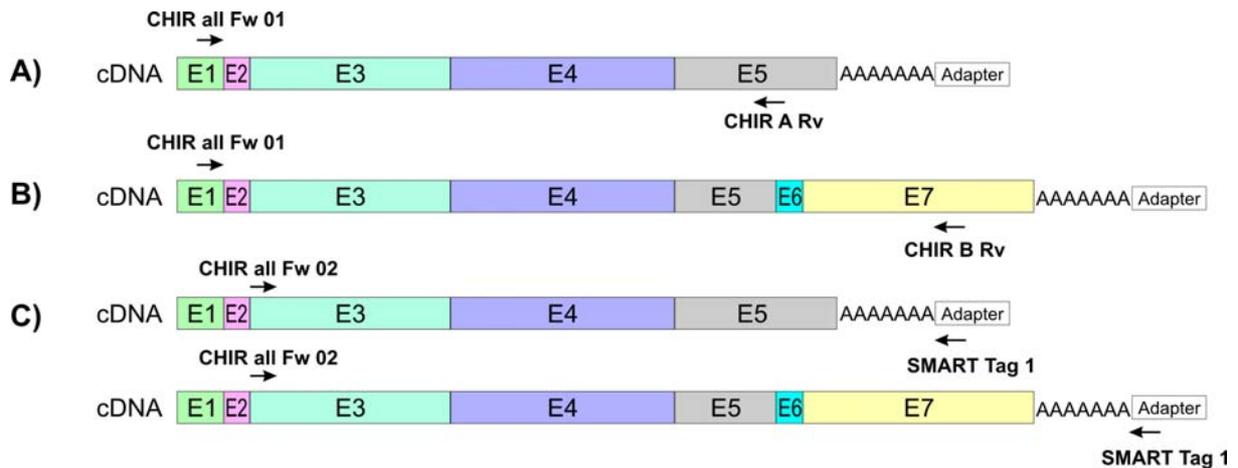


Abb. 26: Schematische Darstellung der Lage der zur CHIR-Amplifikation verwendeten Primer. Primerkombination (A) bestand aus den Primern Chir all Fw 01 und Chir A Rv und diente zur Amplifikation von CHIR-A ähnlichen Rezeptoren. Primerkombination (B) sollte zur Amplifikation CHIR-B ähnlichen Rezeptoren dienen. Hierzu wurden die Primer Chir all Fw 01 und Chir B Rv eingesetzt. Die Primersequenzen von CHIR A Rv und CHIR B Rv wurden der Publikation von Dennis und Kollegen (Dennis, Jr. et. al. 2000) entnommen. C) war eine Kombination aus Chir all Fw 02 und dem SMART Tag 1 Primer, der an das 3'-Ende des zur cDNA-Synthese verwendeten Poly-T-Tag Primers binden kann.

Da erwartet wurde, daß die PCR ein Gemisch verschiedener Transkripte ergeben würde, wurden die Ansätze kloniert und mehrere Klone von jedem Ansatz sequenziert. Insgesamt wurden die Sequenzen von 54 derartigen Klonen untersucht. Die Auswertung ist in Abb. 27 schematisch für einige cDNAs dargestellt.

Wie bei den Bursa-cDNA-Sequenzen lassen sich die Transkripte zu Gruppen zusammenfassen. Die erste Gruppe beinhaltet zwei Klone, CHIR cDNA-43 und -44. Diese Klone zeigen als einzige ein korrektes Spleißmuster und enthalten die Exone eines Rezeptors der CHIRA-Gruppe. Beiden Klonen fehlen zwar die ersten beiden Exone, dies ist jedoch durch die Lage des verwendeten Vorwärts-Primers (CHIR all Fw 02) bedingt. Die *in silico* Translation der beiden Transkripte zeigt jedoch, daß sie eine Verschiebung des Leserasters in der zweiten Ig-Domäne enthalten. Somit läßt sich kein funktionelles Protein aus ihnen ableiten. Es handelt sich also um Transkripte von Pseudogenen.

Die zweite Gruppe enthält Sequenzen, bei denen ein oder mehrere Exone fehlen. Bei einigen Transkripten ist das Exon 3 abwesend. Diese Transkripte könnten fehlerhaft gespleißt sein. Übersetzt man die Sequenzen in Proteine, so ergeben sich jedoch Rezeptoren mit einer

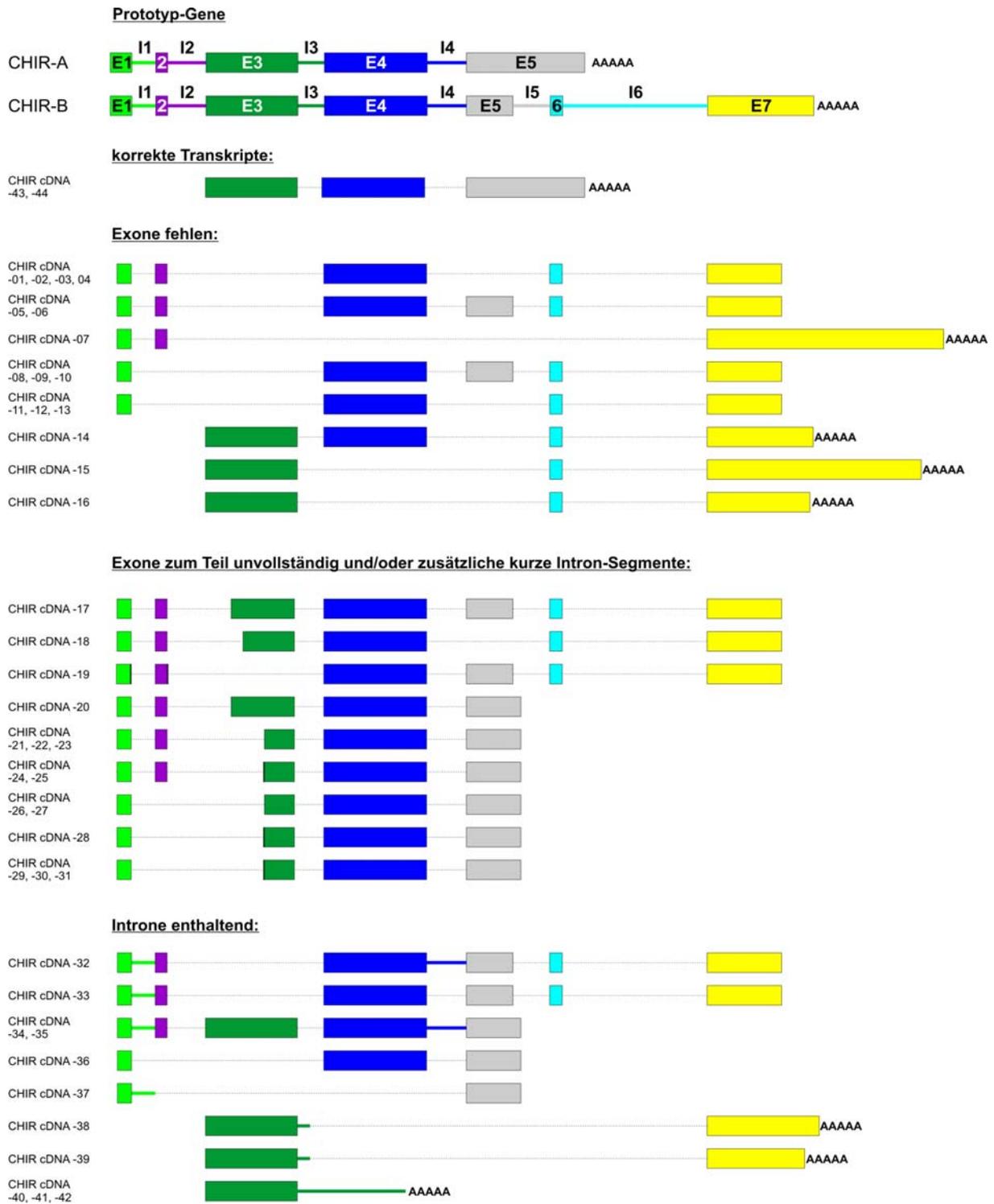


Abb. 27: Schematische Darstellung aller amplifizierten CHIR-cDNA-Sequenzen. Die in den cDNAs vorhandenen Sequenzen sind entsprechend der oben angegebenen Prototyp-Sequenzen von CHIR-A und CHIR-B farblich markiert, fehlen die Sequenzen, wurde nur eine graue Verbindungslinie eingezeichnet. Mehrere cDNA-Namen vor dem jeweiligen Kästchen-Schema bedeuten, daß diese Sequenzen dem gleichen Spleißmuster folgen. Dies bedeutet aber nicht unbedingt, daß sie die gleiche Sequenz haben. E=Exon; I=Intron

Ig-Domäne (Ig-Domäne 2), einer ungeladenen Transmembrandomäne und einem langen zytoplasmatischen Schwanz, der zwei ITIMs enthält. Einen solchen Rezeptor könnte man als CHIR2B klassifizieren. Einigen dieser Transkripte fehlt jedoch auch Exon 2 und damit ein Teil des Signalpeptids, das für die korrekte Prozessierung des Proteins notwendig ist. Die ab-

geleiteten Proteine unterlägen vermutlich nicht dem cotranslationalen Transport in das ER Lumen und könnten deshalb nicht an die Zelloberfläche transportiert werden. Bei den meisten weiteren Sequenzen dieser Gruppe fehlen mehrere Exone, so daß sich keine sinnvollen Transkripte ableiten lassen.

Der dritten Transkriptgruppe wurden cDNAs zugeordnet, bei denen einige Exone nur partiell vorhanden sind. So ist in fast allen Transkripten das Exon 3 nur bruchstückhaft vorhanden. In allen Fällen findet sich auf genomischer Ebene eine Spleißakzeptorstelle (AG) direkt am 5'-Ende des verkürzten Exons 3, so daß es sich bei diesen Transkripten vermutlich um fehlerhaft gespleißte Varianten handelt. Bei den cDNA-Klonen 17, -18, -19 und -20 ergibt sich trotz der fehlenden Sequenzen zwar keine Verschiebung des Leserasters, allerdings fehlen die von den entsprechenden Sequenzen kodierten Aminosäuren. Anders ist dies bei den Transkripten CHIR-cDNA-21 bis -31. Hier resultiert das alternative Spleißen in einer Verschiebung des Leserasters, das zur vorzeitigen Termination der Translation führt.

Die vierte Gruppe beinhaltet CHIR-Sequenzen, bei denen - vermutlich durch unvollständiges Spleißen - ein oder mehrere Introne oder Teile von Intronen im Transkript enthalten sind. Entfernt man diese bei den cDNA-Sequenzen-32 und -33, so ergibt sich wie bereits oben beschrieben ein Rezeptor mit einer Ig-Domäne, TM und einem langen zytoplasmatischen Schwanz. In der zytoplasmatischen Domäne ist allerdings nur ein ITIM enthalten, da die erste Aminosäure des zweiten ITIMs durch eine Mutation der zweiten Base des Triplets mutiert ist (I→T). Bei den Sequenzen 34 bis 37 sind nur Sequenzen für 5 Exone vorhanden, es sollte sich also um Mitglieder der CHIR-A-Familie handeln. Alle Sequenzen besitzen jedoch keine geladene Aminosäure in der Transmembrandomäne. Eine Analyse der CHIR-Gene aus Abschnitt 4.2.3.5 zeigt, daß der zur Amplifikation dieser Transkripte verwendete Primer (CHIR A Rv) auch im Intron 5 einiger inhibitorischer Rezeptoren hätte binden können, so daß es sich auch um Amplifikate von inhibitorischen Rezeptoren handeln könnte. Nach Entfernung der Introne bei den Sequenzen 34 und 35 ergaben sich keinerlei sinnvolle Produkte, da beide cDNAs Verschiebungen des Leserasters in Exon 4 (Ig-Domäne 2) besitzen, was in beiden Fällen zu einer frühzeitigen Termination der Translation führen würde. Es kann allerdings nicht ausgeschlossen werden, daß Amplifikationsartefakte oder Sequenzierfehler zu fehlenden Basen in den Transkripten beitragen. Den restlichen Transkripten fehlen mehrere Exone, einige enthalten auch Fragmente von Intronen und drei Transkripte sind vorzeitig polyadenyliert. Der physiologische Sinn dieser Transkripte ist unklar.

Neben der Analyse ihrer Organisation wurden alle cDNAs miteinander verglichen. Wie bei den Bursa-cDNA-Sequenzen sind bei den hier amplifizierten cDNAs redundante Sequenzen

vorhanden, die von einem Gen transkribiert wurden. Darüber hinaus unterscheiden sich einige cDNAs nur in einzelnen Basenpaaren. Bei diesen Sequenzen handelt es sich entweder um Transkripte von Allelen oder sehr ähnlichen Genen. Auch hier kann jedoch nicht ausgeschlossen werden, daß die Unterschiede durch fehlerhafte Amplifikation und/oder Sequenzierfehler hervorgerufen wurden. Von den 54 untersuchten Sequenzen kann man auf 21 unterschiedliche Gene/Allele schließen. Die größte Redundanz zeigen drei Transkripte, von denen 11, 10 bzw. sechs Klone isoliert wurden. Bei den redundanten Sequenzen handelt es sich um Spleißvarianten desselben Gens, die jedoch wie oben beschrieben meist keine funktionellen Transkripte darstellen.

Insgesamt zeigte sich, daß die Transkripte sehr heterogen sind. Teilweise mag dies von methodischen Problemen bei der Amplifikation der Transkripte mittels PCR herrühren. So könnten zum Beispiel Punktmutationen, der Verlust einzelner Basen, die Überrepräsentation von besonders leicht amplifizierbaren Produkten und sogar Chimärenbildung erklärt werden. Das Ausmaß der beobachteten nicht klassischen Varianten ist jedoch viel zu hoch, um auf diese Weise erklärt werden zu können. Es wurde deshalb auf die Analyse von ESTs, die auf andere Weise hergestellt wurden, zurückgegriffen.

4.3.3 „Expressed Sequence Tags“ (ESTs)

Die von mir durchsuchte NCBI-EST-Datenbank enthält Sequenzen von willkürlich sequenzierten Klonen aus cDNA-Bibliotheken verschiedenster Gewebe. Mit Hilfe des Blast-Programms (Altschul et. al. 1990) habe ich alle ESTs des Huhns (*Gallus gallus*) nach CHIR-Sequenzen durchsucht. Als Suchsequenz wurde das CHIR-Gen CHIRB-01 (ohne Introne) verwendet. Insgesamt konnten damit 121 CHIR-homologe ESTs identifiziert werden, deren Sequenz durch einen Vergleich mit CHIR-A, -B und -C-Sequenzen analysiert und klassifiziert wurde. In Abb. 28 sind diese Gruppen und einige darin enthaltene repräsentative ESTs schematisch dargestellt.

Aufgrund der großen Menge an Sequenzen soll hier nicht im Detail auf die Struktur der einzelnen ESTs eingegangen werden. Die Ergebnisse ähneln jedoch sehr stark den bereits beschriebenen: Nur wenige ESTs sind vollständig und korrekt gespleißt; der Mehrzahl fehlen Exone, sie enthalten noch Introne, oder sie weisen andere Fehler bei der Prozessierung auf.

Bei neun EST-Sequenzen wurden zusätzliche Exone (A / B / C) gefunden, die bei einigen Genen vor Exon 1 (A, B) oder nach Exon 5 (C) lokalisiert sind. Zu diesen Exonen (besonders Exon A) homologe Sequenzen sind relativ verbreitet innerhalb der BAC-Sequenzen. Sie liegen aber häufig nicht in der Nähe von CHIR-Genen und sind oft nur partiell oder in der

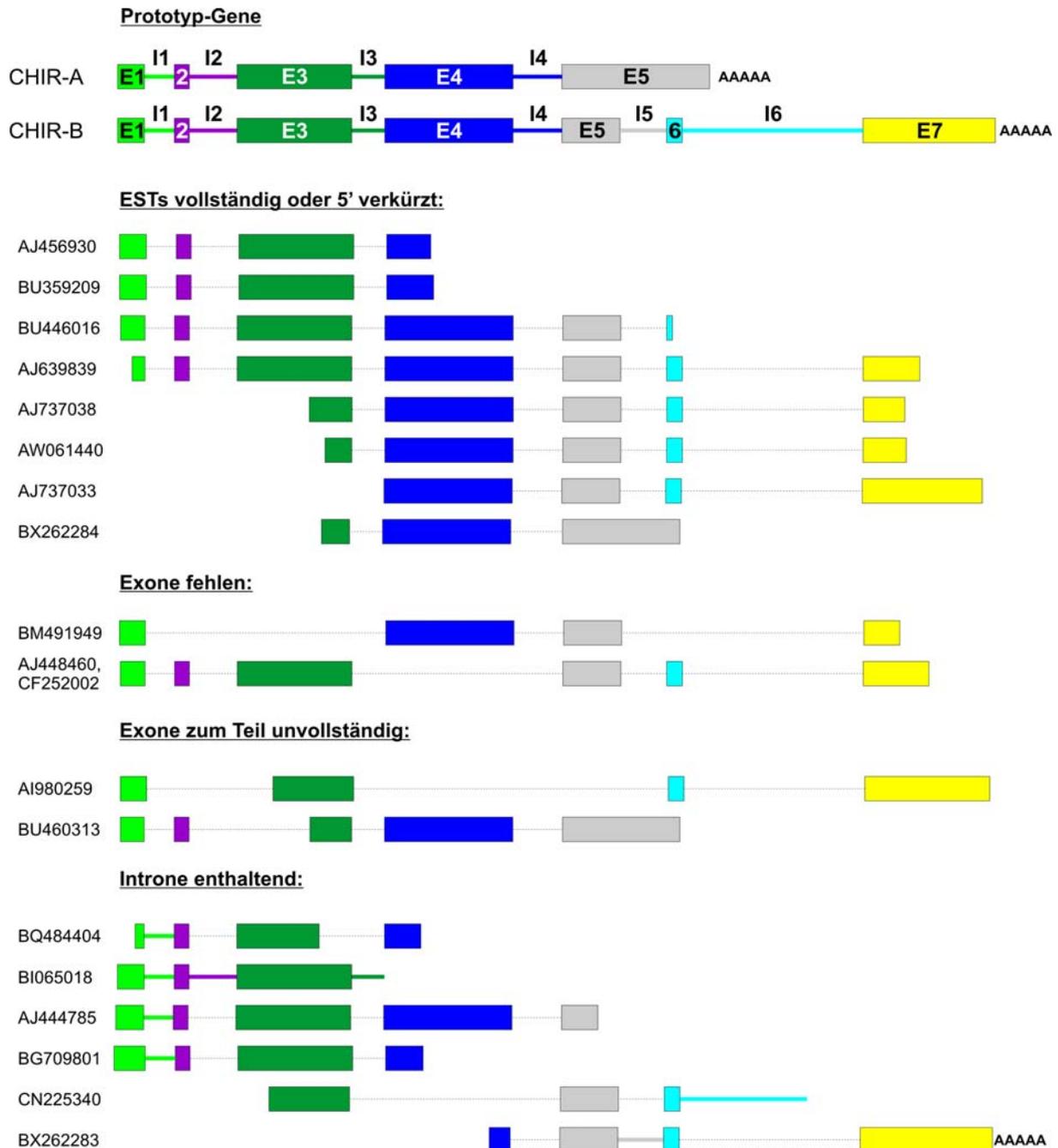


Abb. 28: Schematische Darstellung einiger ausgewählter ESTs.

Die in den cDNAs vorhandenen Sequenzen sind entsprechend der oben angegebenen Prototyp-Sequenzen von CHIR-A und CHIR-B farblich markiert, fehlen die Sequenzen, wurde nur eine graue Verbindungslinie eingezeichnet. Die Gruppierung erfolgte wie in den vorangegangenen Abschnitten aufgrund der beobachteten Spleißvariationen bzw. Fehlerhaftigkeiten der ESTs. Mehrere cDNA-Namen vor dem jeweiligen Kästchen-Schema bedeuten, daß diese Sequenzen dem gleichen Spleißmuster folgen. Dies bedeutet aber nicht unbedingt, daß sie die gleiche Sequenz haben. E=Exon; I=Intron

falschen Orientierung vorhanden. Welche physiologische Relevanz Transkripte mit diesen Exonen haben, ist unklar.

Ein Sequenzvergleich aller ESTs untereinander zeigte, daß auch hier mehrere sequenzidentische Transkripte vorkommen, die wohl von demselben Gen herrühren. Auf dieser Grundlage ließen sich einzelne Sequenzen zu Gruppen zusammenfassen, die sich entweder gar nicht oder nur durch wenige Basenaustausche unterschieden. Bei letzteren handelt es sich möglicherwei-

se um Allele desselben Gens. Da die Transkripte zum Teil sehr unterschiedliche Sequenzbereiche abdeckten, war es nicht in jedem Fall möglich, einen Vergleich mit allen anderen Sequenzen durchzuführen. Aus diesem Grund ist die Redundanz innerhalb der ESTs eventuell etwas höher, als es hier festgestellt werden konnte. Die 121 gefundenen ESTs ließen sich in insgesamt 64 unterschiedliche Gruppen mit großer oder vollständiger Sequenzhomologie einteilen, wobei jede Gruppe vermutlich Sequenzen eines Gens enthält. Die Anzahl der ESTs pro Gruppe schwankte dabei zwischen 1 und 5, nur bei einem Transkript war die Redundanz mit 39 ESTs sehr hoch.

Neben den Sequenzinformationen lassen sich aus den EST-Datenbanken auch Informationen über die gewebespezifische Expression der Transkripte ableiten. Am häufigsten wurden

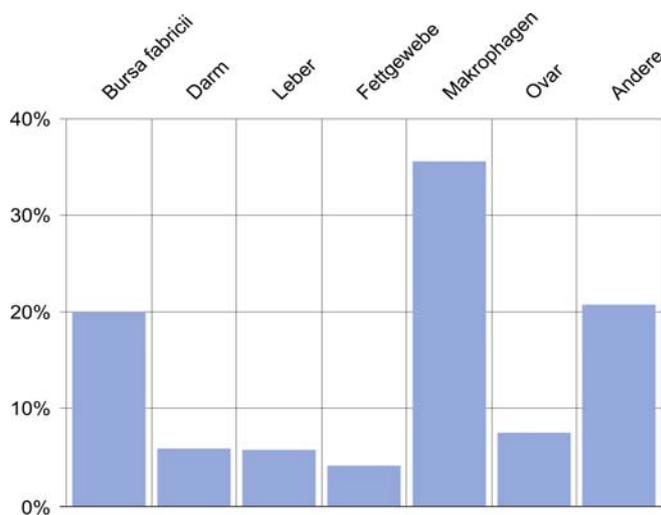


Abb. 29 Gewebeverteilung der analysierten CHIR-ESTs
Der prozentuale Anteil einzelner Gewebe oder Zellpopulationen an der Gesamtzahl der analysierten CHIR-ESTs ist angegeben.

CHIR-Transkripte in Makrophagen-cDNA-Bibliotheken gefunden, aber auch in Bibliotheken von Lymphozyten und Blut kommen sie gehäuft vor. Einige CHIR-Transkripte konnten jedoch auch unterschiedlichsten Geweben (Leber, Milz, Ovar u.a.) zugeordnet werden (Abb. 29). Da CHIRs offenbar vermehrt in zirkulierenden Blutzellen exprimiert werden, ist eine definitive Zuordnung zu verschiedenen Gewebetypen jedoch schwierig, da die CHIR-Transkripte aus den darin enthaltenen Blutzellen stammen könnten.

4.3.4 Zuordnung der Transkripte zu CHIR-Genen

Neben der in den vorangegangenen Abschnitten beschriebenen Analyse der Organisation und Sequenz der Transkripte sollte herausgefunden werden, welche der identifizierten CHIR-Gene abgelesen werden. Hierzu wurden alle analysierten cDNA-Sequenzen der verschiedenen Quellen miteinander verglichen. Dabei zeigte sich, daß auch bei den CHIR-Transkripten aus unterschiedlichen Quellen redundante Sequenzen vorhanden sind. Die Summe aller Transkripte betrug 185. Diese Transkripte lassen sich auf mindestens 85 verschiedene CHIR-Gene bzw. -Allele zurückführen. Einige dieser 85 Sequenzen zeigten jedoch sehr hohe Ähnlichkeit zu anderen CHIR-Transkripten, weshalb es sich hier auch um Allele des jeweils sel-

ben Gens handeln könnte. Die Anzahl von Transkripten pro Gen/Allel reicht dabei von 1 bis 39. Ein Sequenzvergleich dieser 85 einzigartigen Transkriptsequenzen mit den sequenzierten BAC-Genen zeigte, daß nur fünf Gene (CHIRA-13, CHIRA-21, CHIRB-13, CHIRC-04 und CHIRP-09) eine vollständige Sequenzübereinstimmung mit einem oder mehreren Transkripten aufwiesen. An dieser Stelle sollte noch erwähnt werden, daß nur zwei Gene auf den genomischen Kontext bezogen die Transkription der Bursa-Transkripte mit dem Mariner-Transposon in Exon 7 zulassen. Hierbei handelt es sich um CHIRB-14 und CHIRB-17. Tatsächlich zeigen die entsprechenden Transkripte höchste Homologie zu CHIRB-14. Die Sequenzübereinstimmung liegt jedoch nur bei 97%, was darauf hindeutet, daß die Sequenzvariabilität zwischen einzelnen Allelen sehr hoch sein kann.

Bei 80 Transkriptsequenzen ließ sich keine vollständige Übereinstimmung mit einem der BAC-Gene feststellen. Aus diesem Befund läßt sich wahrscheinlich ableiten, daß die CHIR-Gene polymorph sein müssen. Das bedeutet, daß sich die Sequenz der Gene zwischen den Individuen, deren Gewebe zur Herstellung der cDNA-Bibliotheken verwendet wurde, unterscheidet. Für diese Überlegung spricht auch, daß sich einige der Transkripte - wie oben beschrieben - nur durch sehr wenige Basenpaare unterscheiden. Bei solchen Sequenzen handelt es sich vermutlich um Transkripte von Allelen eines Gens. Eine eindeutige Zuordnung zu einem Gen gestaltet sich jedoch bei solchen - aber auch anderen - Sequenzen schwierig, wenn die Sequenzen ähnlich hohe Sequenzübereinstimmung zu zwei verschiedenen Genen aufweisen. Einzelne Gene zeigen offenbar eine ähnlich hohe Sequenzübereinstimmung, wie verschiedene Allele eines Gens untereinander (Abb. 30).

4.3.5 Diskussion der CHIR-Transkript-Analyse

In den vorangegangenen Abschnitten konnte ich zeigen, daß eine Vielzahl von CHIR-Genen *in vivo* transkribiert wird. Diese Transkripte werden von fünf oder sieben Exonen abgelesen und haben eine Länge von etwa 1000 bp bzw. 1300 bp. Mit einigen Ausnahmen lassen sich Transkripte, die aus fünf Exonen bestehen, aktivierenden Rezeptoren zuordnen, während Transkripte von Genen mit sieben Exonen für inhibitorische Rezeptoren kodieren. Die überwiegende Zahl der analysierten Sequenzen war jedoch fehlerhaft oder unvollständig prozessiert. Diese Beobachtung läßt sich zwanglos auf die zur Isolierung und Herstellung der cDNAs verwendeten Methoden zurückführen. So handelt es sich bei den RNA-Präparationen in aller Regel um Lysate der gesamten Zelle, also auch des Zellkerns. Letzterer enthält die Vorstufen der vollständig gereiften mRNAs. Somit lassen sich Transkripte erklären, die noch Introne enthalten. Aber auch falsch gespleißte Transkripte unterliegen Kontrollmechanismen,

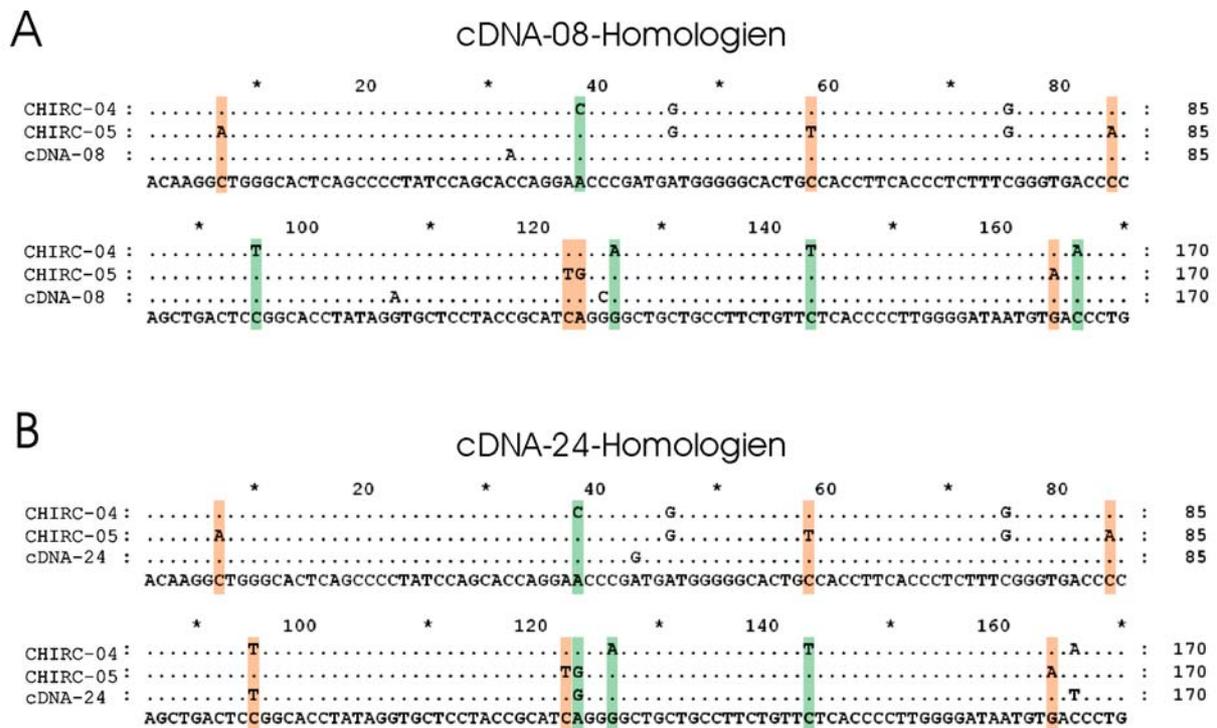


Abb. 30 Sequenzvergleich eines Teils der Gene *CHIRC-04* und *CHIRC-05* mit den cDNAs-08 und -24.

Bei den cDNAs-08 und -24 handelt es sich vermutlich um Allele eines Gens, die Zuordnung zu einem Gen der sequenzierten BACs war jedoch trotz weitreichender Sequenzübereinstimmung nicht möglich, da sowohl *CHIRC-04* als auch *CHIRC-05* ähnlich hohe Übereinstimmung mit beiden cDNAs zeigen. In (A) sind alle Sequenzübereinstimmungen der cDNA-08, die auf eine Abstammung von *CHIRC-04* schließen lassen, orange unterlegt. Übereinstimmungen mit *CHIRC-05* wurden grün unterlegt. Analog wurden in (B) alle Sequenzübereinstimmungen der CHIR-cDNA-24 mit *CHIRC-04* orange und mit *CHIRC-05* grün unterlegt. Es zeigt sich, daß ähnlich viele Übereinstimmungen eine Zuordnung zu dem einen bzw. das anderen Gen zulassen, mit einer leichten Tendenz beider cDNAs zu *CHIRC-04*. Eine eindeutige Zuordnung ist jedoch nicht möglich.

die die korrekte Prozessierung sicherstellen und für die Eliminierung entsprechender mRNAs sorgen (zur Übersicht Baker und Parker 2004). Neben technischen Problemen hängt die Vielfalt der analysierten Transkripte vermutlich auch mit der Entwicklung des CHIR-Clusters zusammen. Die große Sequenzübereinstimmung zwischen den Genen legt nahe, daß Duplikationen auch in jüngerer Vergangenheit stattgefunden haben. Vermutlich hält dieser Prozess bis heute an. Somit entstehen neue Gene durch Duplikation, während andere durch Mutationen ihre Funktionalität einbüßen. Wie weiter unten beschrieben, können Transkripte von defekten Genen durchaus zu einem erheblichen Teil in RNA-Präparationen enthalten sein, solange deren Promotoren weiterhin aktiv sind.

Die Zuordnung der Transkripte zu den entsprechenden Genen in den BACs gelang nur in wenigen Fällen, was vermutlich auf den außergewöhnlich hohen Grad an Polymorphismus dieser Gene und die hohe Sequenzübereinstimmung zwischen einigen Genen zurückzuführen ist. Erstaunlicherweise wurden relativ viele cDNAs gefunden, die hohe Sequenzübereinstimmungen mit Pseudogenen in den BACs zeigen. Da es sich nur in einem Fall um eine vollständige Übereinstimmung handelt, ist es zwar möglich, daß die Gene der meisten Transkripte nicht in den BACs enthalten sind, es erscheint jedoch plausibel, daß zumindest einige dieser Transkripte Genen in den BACs zuzuordnen sind. Es handelt sich in diesen Fällen um

Transkripte von Allelen dieser Gene, aber auch um Transkripte von bisher unbekanten CHIR-Genen. Es erscheint zunächst nicht sinnvoll, daß Pseudogene abgelesen werden; als mögliche Erklärung des Vorhandenseins solcher Transkripte ergeben sich zwei Möglichkeiten: 1) bei den analysierten cDNAs handelt es sich um Transkripte von funktionellen Allelen des gleichen Gens, so daß nur einige Allele Pseudogene darstellen. Ähnliches konnte bereits bei der Analyse eines Clusters olfaktorischer Rezeptoren beim Menschen gezeigt werden (Ehlers et. al. 2000). Zur Überprüfung dieser Hypothese müßten die entsprechenden Gene bei verschiedenen Individuen sequenziert werden, um so verschiedene Allele identifizieren zu können und zu prüfen, ob sich sowohl funktionelle als auch nicht-funktionelle Allele eines Gens finden. 2) Es ist jedoch auch möglich, daß in den CHIR-Pseudogenen Mutationen die korrekte Translation verhindern, ohne daß der Promotor zuvor inaktiviert wurde. Solche Gene sammeln im Laufe der Zeit weitere Mutationen an, weil diese keine funktionellen Konsequenzen für das Individuum haben. Es werden hierbei auch regulatorische Promotorelemente eliminiert, wodurch dann auch die Transkription des Gens verloren gehen kann. Alternativ können aber auch mutierte Sequenzen durch *crossing over* oder *gene conversion* repariert werden, so daß die Funktionalität eines Gens prinzipiell auch wiederhergestellt werden kann (Nei und Rooney 2005). Gerade *gene conversion* stellt beim Huhn in der Entwicklung der B-Zellen eine herausragende Rolle bei der Diversifizierung der leichten und schweren Immunglobulinketten dar (Litman et. al. 1999; Arakawa und Buerstedde 2004) und könnte möglicherweise bei CHIRs auch zu somatischen Rearrangements von Pseudogenen beitragen.

Die Transkriptanalyse ergab weiterhin, daß einige cDNAs kein Exon 3 enthalten. Aus solchen Transkripten ergäbe sich ein Rezeptor mit nur einer Ig-Domäne. Prinzipiell können solche Transkripte sogar aus allen CHIR-Genen durch alternatives Spleißen hervorgehen. Zusätzlich gibt es einige Gene, die ein STOP-Kodon in Exon 3 enthalten. Diese Pseudogene könnten funktionell sein, wenn das Exon 3 und die angrenzenden Introne herausgespleißt würden. Interessant ist in diesem Zusammenhang, daß bei humanen KIR-Proteinen mit 2 Ig-Domänen auf genomischer Ebene auch 3 Exone vorhanden sind, die für je eine Ig-Domäne kodieren können. Einige dieser Sequenzen weisen dabei keinerlei Anomalien auf, andere haben mutierte Spleißstellen oder besitzen Mutationen (Vilches et. al. 2000). KIRs mit 2 Ig-Domänen sind also vermutlich aus einer Duplikation eines Rezeptorgens mit drei Ig-Domänen hervorgegangen, bei dem die erste Ig-Domäne dann inaktiviert wurde und während der Prozessierung herausgespleißt wird. Möglicherweise führen alternatives Spleißen und/oder Mutationen in Exon 3 oder Exon 4 auch bei den CHIR-Genen zu einer zusätzlichen Diversifizierung des Rezeptorrepertoires. Ob tatsächlich CHIRs mit nur einer Ig-Domäne ex-

primiert werden, könnte mit Hilfe von Immunpräzipitationen von CHIR-Proteinen aus geeigneten Zelllinien oder Blutzellen nachgewiesen werden, sobald Antikörper zur Verfügung stehen, die die erste und zweite Ig-Domäne der CHIRs erkennen.

Erstaunlicherweise wiesen eine Reihe von cDNAs partielle Sequenzen von Exon 3, aber auch Exon 4 auf. Auf genomischer Ebene werden diese untypischen Spleißstellen in der Regel von korrekten Spleißakzeptorsequenzen flankiert. Interessant ist in diesem Zusammenhang, daß diese Spleißakzeptorstellen in hypervariablen Regionen von Exon 3 liegen. Möglicherweise sind diese kryptischen Spleißstellen eine Folge der Variabilität der CHIR-Rezeptoren, die wohl in einigen Fällen auch zur Entstehung von Spleißstellen innerhalb des Exons geführt hat.

Die Analyse der ESTs zeigte, daß sich die Expression der CHIRs in verschiedenen Geweben nachweisen läßt (Abb. 29). Eine klare Häufung findet sich jedoch in Geweben und Zelltypen des Immunsystems. Diese Expression deutet darauf hin, daß CHIRs eine Funktion im Immunsystem zukommt. Entsprechendes wäre auch aus der Homologie zu den Rezeptoren des LRC der Säuger zu vermuten, da alle LRC-Rezeptoren auf Zellen des Blutes exprimiert werden (zur Übersicht McQueen und Parham 2002).

Insgesamt zeigt das CHIR-Transkriptom alle von einer dynamischen und polymorphen Genfamilie erwarteten Eigenschaften. Die Transkripte sind sowohl hochvariabel in ihrer Sequenz, als auch ihrem Spleißmuster; es existiert vermutlich eine Vielzahl von Allelen, die zum Teil Pseudogene sind. Auch wenn hier nur eine begrenzte Zahl von Transkripten analysiert werden konnte, ist es doch beachtlich, daß nur fünf Transkripte ihren jeweiligen Loci zugeordnet werden konnten und daß eines dieser Gene ein Pseudogen ist.

4.4 Evolutionäre Betrachtungen zur Entwicklung der CHIR-Gene

Die in den zurückliegenden Abschnitten beschriebenen Analysen zeigen, daß es sich bei der CHIR-Genfamilie um eine vermutlich hochpolymorphe Multigenfamilie handelt. Die Diversität der Rezeptoren wird zusätzlich durch eine Vielzahl von Rezeptortypen ergänzt, die zum Teil durch alternatives Spleißen entstehen. Die hohe Sequenzhomologie zwischen den einzelnen Rezeptorgenen läßt den Schluß zu, daß sich die CHIR-Gene aus einem gemeinsamen Vorläufer entwickelt haben. Um die evolutionären Beziehungen zwischen den verschiedenen Rezeptorgenen zu untersuchen, habe ich deshalb Stammbäume aller jener Exone erstellt, die in den meisten Rezeptorgenen vorhanden sind. Abb. 31 zeigt einen Stammbaum vom Exon 5 aller vermutlich funktionellen CHIR-Gene. Exon 5 enthält u.a. die Sequenz, die für die Transmembrandomäne kodiert. Somit spiegelt der

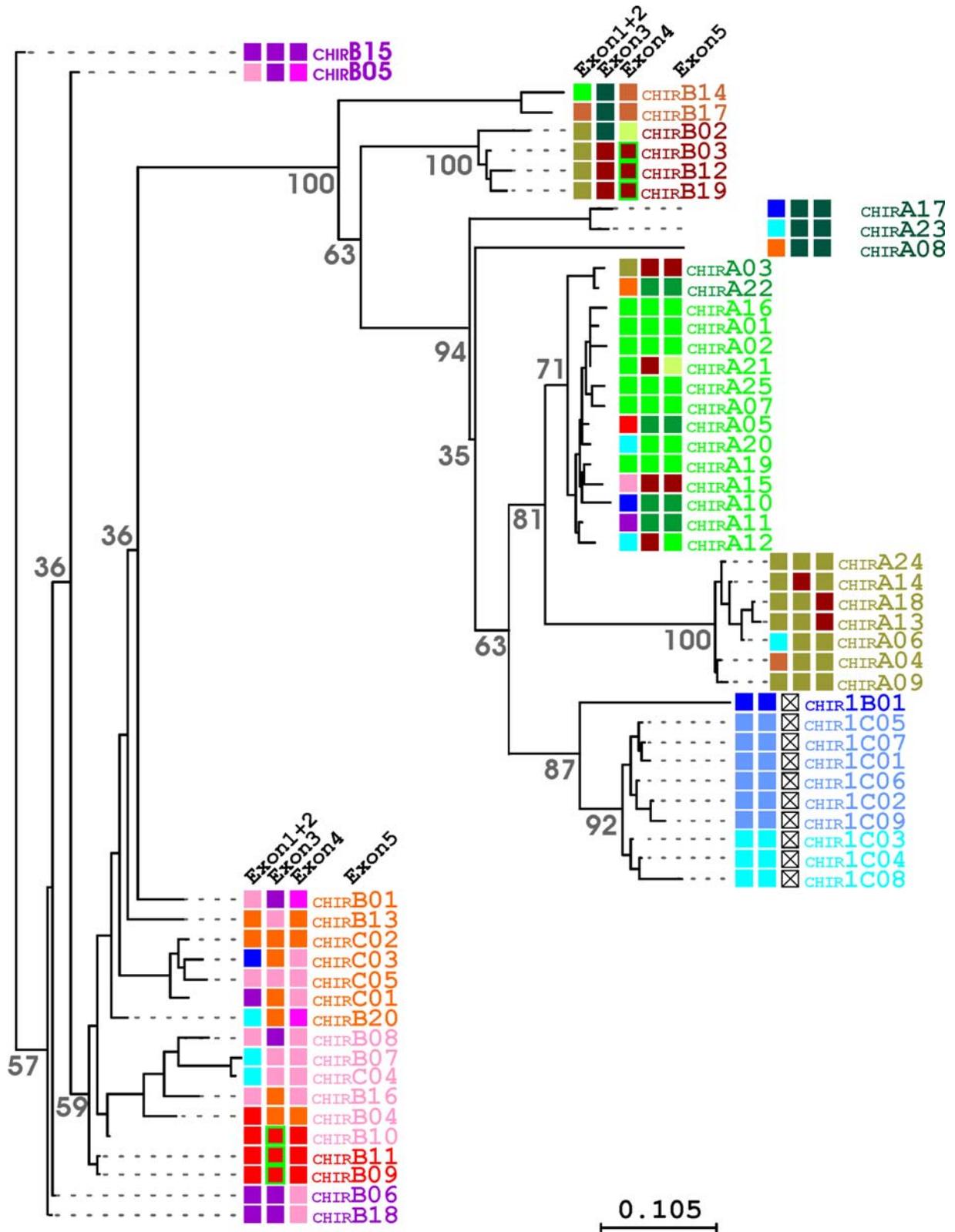


Abb. 31 Stammbaum von Exon 5 aller funktionellen CHIR-Gene
 Der Stammbaum von Exon 5 aller 60 funktionellen Gene wurde mit der „Neighbor joining with pairwise gap removal“-Methode, mit Kimura Abstand und 1000 *bootstrap* Replikaten erstellt. Die *Bootstrap*-Werte sind in grau an den Hauptverzweigungen angegeben. Die Namen der Gene sind entsprechend der Gruppenzugehörigkeit gefärbt (siehe Text). Die Kästchen hinter den Namen zeigen an, welcher Gruppe die entsprechenden Gene in den Stammbäumen von Exon 4, Exon 3 und Exon 1+2 angehören. Kästchen mit Kreuz bedeuten, daß keine Daten für dieses Exon vorliegen, da es in den Genen nicht enthalten ist.

Stammbaum auch die in Abschnitt 4.2.3.5 vorgenommene Gruppierung der CHIRs wider. Die CHIR-Rezeptoren fallen in sechs klare Gruppen, die in Abb. 31 mit unterschiedlich gefärbten Namen gekennzeichnet sind. Bei diesen Gruppen handelt es sich um: 1) inhibitorische Rezeptoren und bifunktionelle Rezeptoren, deren Transmembrandomäne Homologie zu inhibitorischen Rezeptoren aufweist (Gennamen in Rottönen); 2) inhibitorische CHIRs mit erhöhter Homologie zu aktivierenden CHIRs (braun); 3) bifunktionelle CHIRs mit nur einer Ig-Domäne (Gennamen in Blautönen); 4-6) drei Gruppen aktivierender Rezeptoren (Gennamen in Grüntönen). Die Kästchen neben den Namen zeigen an, welcher Gruppe das entsprechende Gen in den Stammbäumen von Exon 1+2, Exon 3 bzw. Exon 4 angehört, wobei analoge Farben verwendet wurden. Hierbei zeigt sich, daß die Zuordnung zu einer der Gruppen im Stammbaum von Exon 1+2 wenig stringent ist, was vermutlich mit der hohen Konservierung der für das Signalpeptid kodierenden Sequenzen zu erklären ist. Hier können einzelne Basenaustausche dazu führen, daß eine Sequenz in eine andere Gruppe fällt. Die Gruppenzuordnung wird jedoch eindeutiger bei den Stammbäumen der Exone 3 und 4, wo die Längen der Zweige innerhalb der Gruppen sehr kurz, zwischen den Gruppen jedoch relativ lang sind. Die kurzen Zweige weisen darauf hin, daß die Expansion der CHIR-Gene innerhalb einer der Gruppen ein evolutionär junges Ereignis ist und daß vor dieser Expansion die Vorläufergene der Gruppen längere Zeit stabil waren und sich deshalb stärker voneinander unterscheiden.

Aus dem Vergleich der verschiedenen Stammbäume läßt sich auch ableiten, daß die meisten CHIR-Gene durch Duplikation ganzer Gene entstanden sind, da die Gene in den verschiedenen Stammbäumen in die gleichen Gruppen fallen (gleiche Farben der Namen und Kästchen). Dies steht im Gegensatz zu evolutionär jungen LRC-Genen wie den KIRs, bei denen eine starke Durchmischung der Domänen stattgefunden hat, die zur Diversifizierung des Rezeptorrepertoires beiträgt (Rajalingam et. al. 2004).

4.5 Variabilität der CHIRs

Der Aminosäuresequenzvergleich aller CHIR-Gene (Abb. 16 auf Seite 54) zeigte, daß sich die Variabilität zwischen den einzelnen CHIR-Rezeptoren nicht gleichmäßig über die gesamten Sequenzen verteilt. Vielmehr scheint der Grad der Sequenzvariabilität von der Aminosäureposition innerhalb des Proteins abzuhängen. So ist die Variabilität der CHIRs im Bereich des Signalpeptids und am Anfang von Ig-Domäne 1 - wie oben beschrieben - klar geringer als in anderen Teilen der CHIRs. Um die Sequenzvariabilität näher zu untersuchen, wurden die CHIR-Sequenzen auf synonyme und nicht-synonyme Basenaustausche hin analysiert (Nei und Gojobori 1986; Ota und Nei 1994). Synonyme Basenaustausche sind durch Mutationen

hervorgerufene Sequenzveränderungen auf DNA-Ebene, die zu keiner Veränderung der von dem betroffenen Triplet kodierten Aminosäure führt. Im Gegensatz dazu ändert sich durch einen nicht-synonymen Basenaustausch die vom betroffenen Triplet kodierte Aminosäure. Synonyme Substitutionsraten sind meist höher als nicht-synonyme Substitutionsraten, weil sie für die Funktion des Proteins bedeutungslos sind. Die meisten nicht-synonymen Substitutionen beeinträchtigen hingegen die Funktion des Proteins und werden durch einen Prozeß, den man als „reinigende Selektion“ bezeichnet, beseitigt. Ist die Anzahl nicht-synonymer Austausche erhöht, läßt sich daraus ableiten, daß eine positive Selektion vorliegt, die die veränderten Eigenschaften des Proteins erhält. Bei der Analyse der CHIR-Sequenzen konnte ich feststellen, daß die Anzahl der nicht-synonymen Austausche an einigen Stellen wesentlich höher ist als die der synonymen Austausche. Es läßt sich also folgern, dass es eine gerichtete Selektion zur Diversität der CHIRs in diesen Regionen gegeben hat. Innerhalb der ersten Ig-Domäne gibt es drei Regionen mit erhöhter Variabilität, in der zweiten Ig-Domäne vier (Abb. 32).

Hypervariable Regionen werden bei Molekülen der angepassten Immunität üblicherweise mit Hilfe eines Variabilitätsplots dargestellt (Wu und Kabat 1970). Diese Darstellung beschreibt das Verhältnis zwischen der Anzahl der verschiedenen Aminosäurereste an einer bestimmten Position im Protein und der Anzahl des häufigsten Aminosäurerestes. Werden in verschiedenen Proteinen an einer bestimmten Position sehr viele verschiedene Aminosäurereste vorgefunden, ergeben sich sehr hohe Wu-Kabat-Werte. Der Wu-Kabat-Graph der potentiell funktionellen 60 CHIR-Gene (Abb. 32) zeigt hohe Variabilität etwa in den gleichen Regionen, wie der Graph der nicht-synonymen Austausche, aber besonders in der Stamm- und Transmembrandomäne zeigt sich eine geringere Variabilität. Diese Beobachtung stimmt auch gut mit den Resultaten des phylogenetischen Baums überein: Dieser zeigt, daß die Gene innerhalb der Gruppen eine geringe Variabilität untereinander zeigen. Somit ergeben sich nur eine geringe Anzahl verschiedener Aminosäuren an einer bestimmten Position, da sich letztlich nur zwischen den Gruppen Unterschiede abzeichnen. Dies führt zu den moderat erhöhten Wu-Kabat-Werten. Auf der anderen Seite führt der Vergleich auf nicht-synonyme Austausche zwischen Mitgliedern unterschiedlicher Gruppen zu hohen Variabilitätswerten, da diese sich recht stark unterscheiden können. Die hohen Ähnlichkeiten innerhalb der Gruppen verringern diesen Wert aber nicht, so wie es bei der Berechnung nach Wu-Kabat der Fall ist und deshalb ergeben sich relativ hohe Werte für nicht-synonyme Austausche in der Stamm- und Transmembrandomäne.

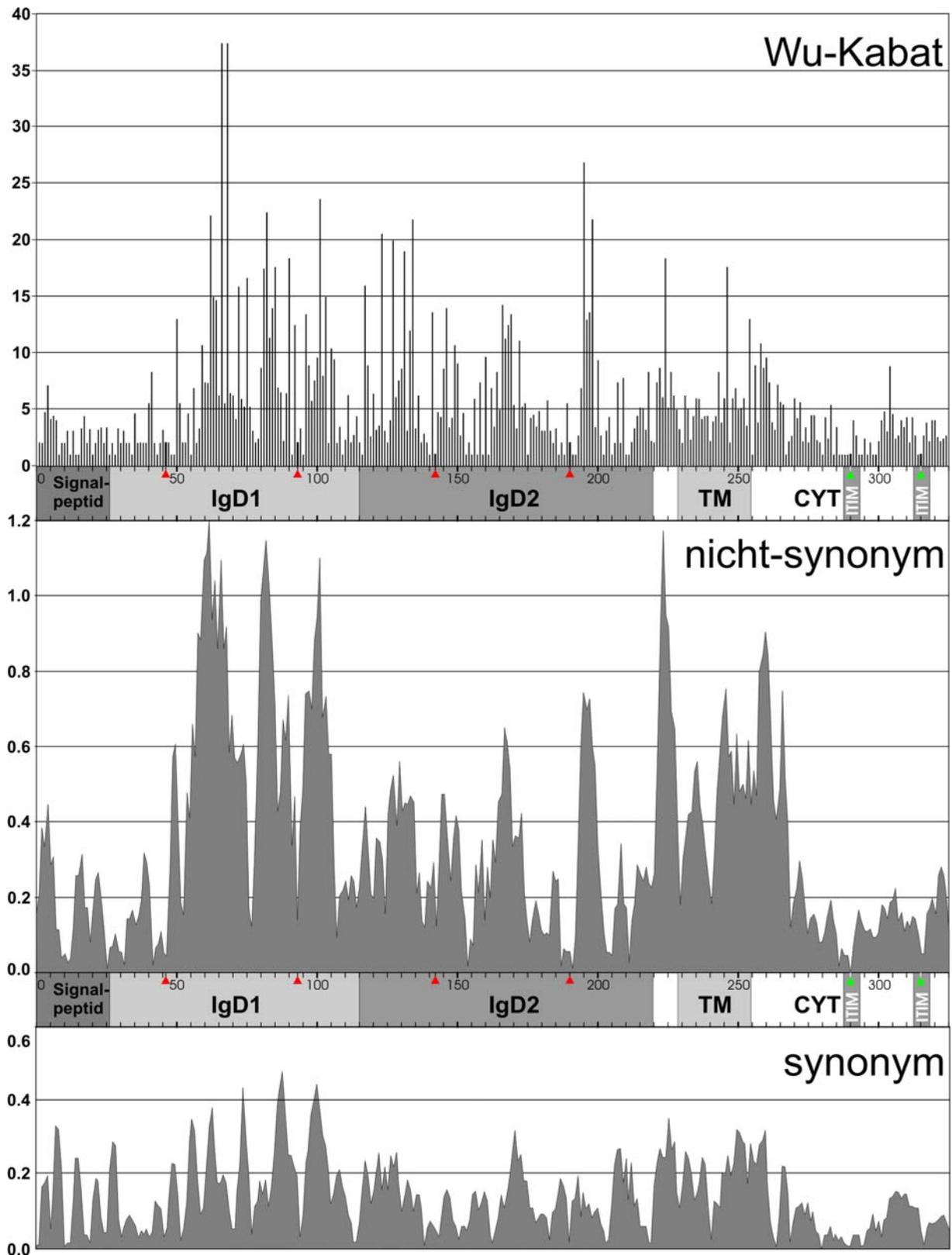


Abb. 32 Sequenzdiversität der CHIRs

Im unteren Teil der Abbildung sind die Kurven der synonymen und nicht-synonymen Substitutionen dargestellt, die für alle 60 potentiell funktionellen CHIR-Gene berechnet wurden. Im oberen Teil der Abbildung beschreiben die schwarzen Balken die zugehörigen Wu-Kabat-Werte. Die Positionen der konservierten Cysteine in den Ig-Domänen sind mit roten Pfeilspitzen markiert und die ITIM-Tyrosine mit grünen Pfeilspitzen. Zwischen den einzelnen Kurven sind die Proteindomänen eingezeichnet und die Nummerierung der Aminosäurereste ist angegeben. Die Ordinaten geben die Wu-Kabat-Werte (Wu-Kabat-Plot) bzw. die durchschnittliche Zahl der synonymen/nicht-synonymen Austausche pro Triplet an.

Die Daten zur Variabilität der CHIRs zeigen, daß es in der Familie der CHIRs eine gerichtete Selektion hin zu Sequenzdiversität gegeben hat und daß die Aminosäurevariabilität in den beiden Ig-Domänen besonders hoch ist. Dies ist insofern interessant, da nur die Ig-Domänen für die Interaktion der CHIRs mit ihren Liganden in Frage kommen. Aufgrund der Variabilität der Ig-Domänen läßt sich also folgern, daß auch die Liganden variabel sein müssen. Vermutlich ist es so, daß die gerichtete Selektion hin zu CHIR-Sequenzdiversität erfolgte, um die Interaktion mit Mitgliedern von variablen Ligandenfamilien zu erlauben.

4.6 Strukturmodelle für CHIRs

Um mehr über die Eigenschaften der CHIR-Rezeptoren zu erfahren, wurde eine Vorhersage der Proteinstruktur mittels SWISS-MODEL (Guex und Peitsch 1997) durchgeführt. Hierzu wurde die extrazelluläre Domäne von CHIRC-02 verwendet, da dessen Sequenz der Konsensus-Sequenz aller CHIRs am ähnlichsten ist. Sequenzvergleiche zeigten die höchsten Ähnlichkeiten zu KIR2DL3/KIR2DS2 (32%) LILRB1 (35%), NKP46 (38%) und KIR2DL1/KIR2DL2 (46%) auf. Trotz der hohen Ähnlichkeit zu KIR2DL1/KIR2DL2 wurde durch den automatischen Modellierungsprozess eine Struktur auf der Basis von LILRB1 und NKp46 entworfen (Abb. 33). Um dennoch ein Strukturmodell auf Basis der KIRs zu erhalten, wurde das Modellierungsprogramm manuell unter Vorgabe der KIR2DL1- und KIR2DL2-Matrizen gestartet (Abb. 33). Beide Modelle wurden dann verwendet, um die Positionen der hoch-variablen Aminosäuren in den Proteinstrukturen darzustellen. Die Modelle zeigen dabei, daß sich die variablen Reste besonders auf der „Vorderseite“ der Moleküle befinden. Auffällig ist vor allem, daß sich viele der variablen Reste in der Berührungsfläche des KIR-Moleküls mit seinem Liganden, einem polymorphen HLA-C-Molekül finden.

Wie oben beschrieben legt die Sequenzvariabilität der CHIR-Proteine nahe, daß sie mit einem polymorphen Liganden interagieren. Diese Hypothese wird durch die Häufung von variablen Aminosäureresten auf einer Seite des Moleküls unterstützt. Es wäre denkbar, daß die variablen Reste Teil einer Oberfläche sind, die geeignet ist, mit variablen Liganden, wie zum Beispiel einem MHC-Molekül des Huhns mit einem gebundenen Peptid, zu interagieren.

4.7 Abschließende Betrachtung und Ausblick

Im Folgenden sollen die Ergebnisse dieser Arbeit, die in den vorangegangenen Abschnitten dargestellt und diskutiert wurden, im Zusammenhang erörtert werden.

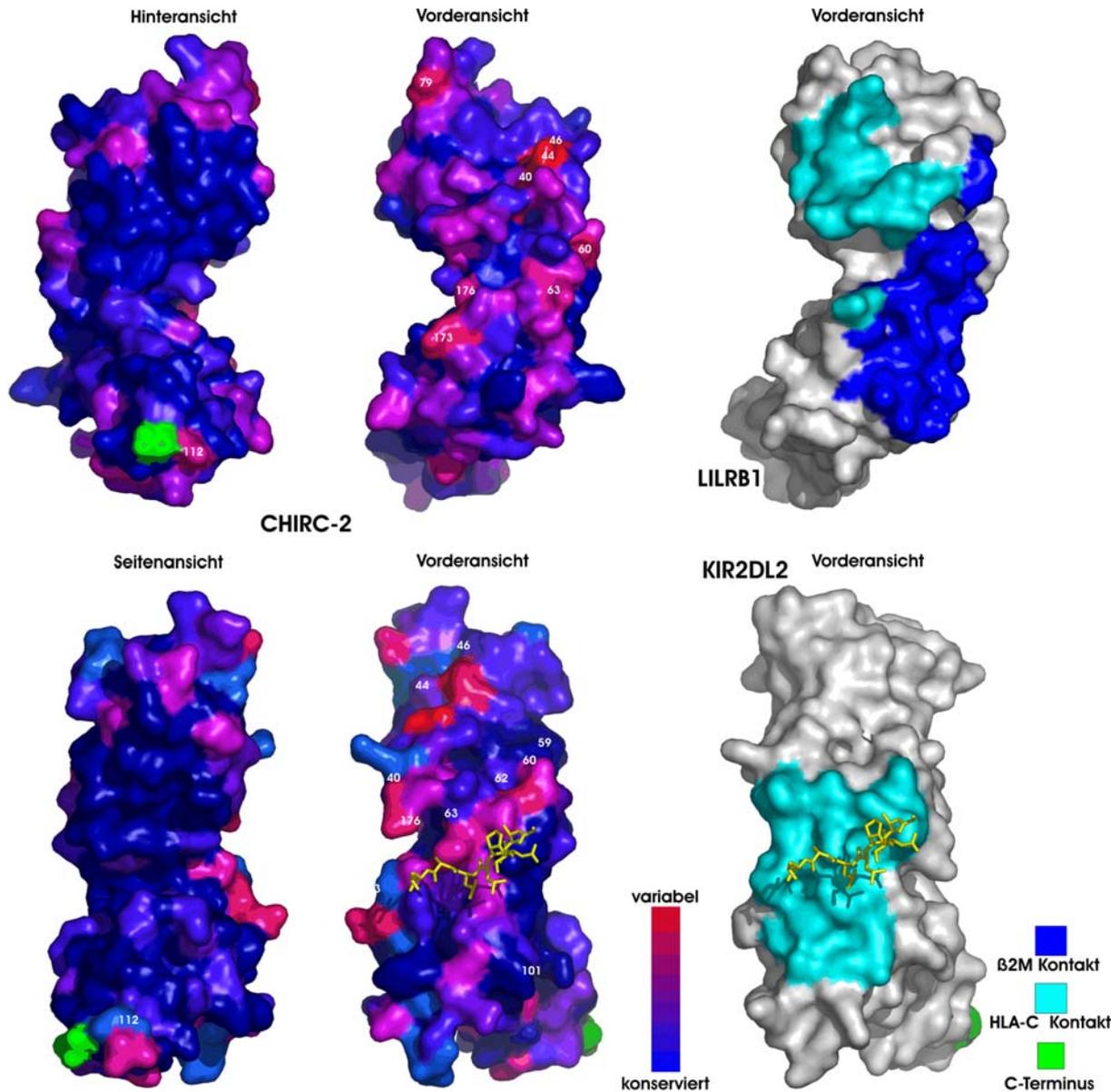


Abb. 33 Strukturmodelle von *CHIRC-02* basierend auf *LILRB1* und *KIR2DL2*.

Die obere Reihe zeigt die Vorder- und Seitenansicht einer Strukturvorhersage für die extrazelluläre Domäne von *CHIRC-02* basierend auf *ILT2* (PDB Codes 1g0xA & 1ufuA), *LIR1* (PDB Codes 1ugnA & 1p7qD) und *Nkp46* (PDB Code 1olla). In der unteren Reihe wurden *KIR2DL1* (PDB Code 1in9D) und *KIR2DL2* (PDB Code 1efxE) für die Vorhersage verwendet. Zur besseren Orientierung wurde das vom MHC-Molekül präsentierte Peptid in der unteren Reihe gelb dargestellt. In beiden Reihen sind rechts die Strukturmodelle einer der Vorlagen dargestellt, bei denen die Berührungsfächen mit dem MHC-Molekül eingefärbt sind. In den beiden linken Spalten wurden die Aminosäurereste entsprechend der Wu-Kabat-Werte (Abb. 32) eingefärbt. Die Nummern der Aminosäuren wurden nur angegeben, wenn deren Wu-Kabat-Werte über 20 lagen (im KIR-Modell sind zusätzlich die Reste 59 und 62 beziffert). Die Nummerierung entspricht dem ausgereiften Protein und beinhaltet die 22 Aminosäuren des Signalpeptids nicht.

4.7.1 Genomische Organisation

Die vier hier charakterisierten CHIR-Contigs haben insgesamt eine Länge von 550.000 Basenpaaren und enthalten 60 CHIR-Gene und 43 CHIR-Pseudogene und Genfragmente. Die CHIR-Gene zeigen eine hochkonservierte Exon-Intron-Struktur. Das Signalpeptid aller Gene wird durch zwei Exone kodiert, während die Ig-Domänen von jeweils einem Exon kodiert sind. Das fünfte Exon kodiert für die Stammregion und die Transmembrandomäne. Die Konservierung der Genstruktur ist bei der zytoplasmatischen Domäne nicht mehr ausgeprägt, da

nur etwa die Hälfte der CHIR-Gene über die Exone 6 und 7 verfügten. Insgesamt zeigt sich, daß die Exon-Intron-Struktur vergleichbar zu Rezeptoren der Ig-SF bei Säugern ist (de Wit et. al. 1995; Wilson et. al. 1997; Martin et. al. 2002). In einer kürzlich erschienen Publikation wurden die gleichen Sequenzen wie in dieser Arbeit auf ihren Gehalt an CHIR-Genen hin untersucht (Nikolaidis et. al. 2005b). Erstaunlicherweise wurden dort 70 funktionelle CHIR-Gene beschrieben. Eine genauere Untersuchung dieser Diskrepanz zeigte, daß sich der Unterschied zumindest bei einigen Genen dadurch erklären läßt, daß Nikolaidis und Kollegen Gene, die frühzeitige STOP-Kodons in nur einer Ig-Domäne tragen, als funktionell betrachteten. Dies trifft allerdings nur dann zu, wenn das entsprechende Exon herausgespleißt wird. Aus diesem Grund habe ich diese Gene den Pseudogenen zugeordnet. Der zahlenmäßige Unterschied läßt sich jedoch mit diesen Genen nicht vollständig erklären. Es zeigte sich, daß einigen als funktionell deklarierten Genen wichtige Exone fehlen (CHIRP-06, CHIRP-09, CHIRP-10) oder sie enthalten Leserasterverschiebungen oder nicht-funktionelle Spleißstellen (CHIRP-24, CHIRP-37, CHIRP-43). Es handelt sich also zumindest bei einem Teil der Gene klar um Pseudogene und nicht um funktionelle Gene. Darüber hinaus wurde in der Arbeit von Nikolaidis und Kollegen die Richtung eines Gens (CHIRA-22) falsch angegeben. Offensichtlich ist die Analyse dieses komplexen Genclusters von Nikolaidis und Kollegen nur sehr oberflächlich durchgeführt worden. Dies hat letztlich dazu geführt, daß die erhaltenen Daten zum Teil falsch interpretiert wurden und daß aufgrunddessen falsche Schlüsse zum Beispiel bezüglich der Evolution des Clusters gezogen wurden.

Neben der Vielfalt an strukturellen Formen sowie der Variabilität der CHIR-Gene kommt auch alternatives Spleißen als mögliche Form der Diversifizierung des Rezeptorrepertoires in Betracht. So zeigten die Transkriptionsanalysen ausgesprochen vielgestaltige Spleißvarianten, deren funktionelle Relevanz jedoch näher untersucht werden muß. Ein Teil der Spleißvarianten läßt sich vermutlich auf fehlerhaftes Spleißen zurückführen, eine Eigenschaft, die im Zusammenhang mit der hohen Variabilität der Gene stehen könnte. So ist es möglich, daß Mutationen in den Ig-Domänen zu kryptischen Spleißdonor- und -akzeptorstellen führen. Die aus diesen Genen abgeleiteten Transkripte stellen letztlich ein Abfallprodukt der schnellen Evolution der CHIRs dar. Neben solchen fehlerhaften Produkten finden sich aber auch Transkripte, bei denen ganze Ig-Domänen oder Teile davon fehlen, ohne daß das Leseraster verschoben wird. Solche Transkripte könnten durchaus funktionell sein. Ähnliche Varianten wurden zum Beispiel für humane KIRs beschrieben (Döhning et. al. 1996). Auch am 3'-Ende erlauben die Sequenzen einiger Gene alternatives Spleißen; so befinden sich unter den 60 CHIR-Genen acht Gene, aus denen sowohl Rezeptoren mit langen als auch mit kurzen zytoplasmatischen

Domänen hervorgehen können. Es würde sich dann um inhibitorische Rezeptoren (CHIRB) oder um Rezeptoren mit ungeladener TM und kurzer zytoplasmatischer Domäne (ähnlich CHIRC-04) handeln. Ob Transkripte eines Gens in dieser Form *in vivo* alternativ gespleißt werden, konnte durch die Transkriptanalysen nicht herausgefunden werden, weil der Polymorphismus der Gene eine genaue Zuordnung verhinderte. Um zu klären, ob alternativ gespleißte Produkte entstehen, müßten die Transkripte der entsprechenden Gene gezielt mit spezifischen Primerkombinationen und RT-PCR analysiert werden. Zudem muß der Nachweis erbracht werden, daß die Produkte solcher Spleißvarianten auch tatsächlich auf der Zelloberfläche exprimiert werden.

Obwohl die Homologie der CHIR-Gene zu den Genen des LRC nahe liegt, war eine eindeutige Feststellung der Orthologie zu bestimmten LRC-Genen nicht möglich. Informativ wären in diesem Zusammenhang Gene, die sich in der Nähe der CHIR-Cluster befinden, da über das gemeinsame Vorkommen der gleichen Gene eine Orthologie unterstützt werden könnte. Das Vorkommen von Sequenzen mit Homologie zu den G-Protein gekoppelten Rezeptoren 40-43 von Säugern ist ein erster Hinweis auf Orthologie zum LRC, da homologen Gene im sogenannten „extended“ LRC lokalisiert sind (Trowsdale et. al. 2001). Die erste Version des Genomprojekts des Huhns enthält keine CHIR-homologen Sequenzen, die einem bestimmten Chromosom zugeordnet sind. Außerdem finden sich in den assemblierten Contigs keine Gene, die Rückschlüsse auf benachbarte Gene des LRC zulassen (Hillier et. al. 2004). Der Abschluß der Sequenzierung des Genoms des Huhns, der auch zur Schließung der Lücken zwischen den Contigs führen wird, sollte in Zukunft die Klärung dieser Frage erlauben. Fluoreszenz-*in situ*-Hybridisierungen an Metaphase-Chromosomen von Hühner-Fibroblasten mit DNA nahezu sämtlichen von mir identifizierten BACs zeigen jedoch, daß die CHIR-Cluster zusammen auf einem Minichromosom des Huhns lokalisiert sind (Laun et. al. 2006 zur Veröffentlichung eingereicht). Die CHIR-Contigs sind also zumindest dicht beieinander lokalisiert und bilden vermutlich ein großes zusammenhängendes Gencluster.

Die Orientierung der CHIR-Gene in den Contigs scheint zufällig zu sein und unterscheidet sich damit von der Tandem-Organisation der meisten LRC-Gene, die vermutlich das Resultat von Duplikationsereignissen sind. Die Tandem-Organisation begünstigt vermutlich auch Mechanismen zur Diversifizierung der Rezeptoren, wie die bei KIRs beobachtete Durchmischung von Domänen (Rajalingam et. al. 2004). Konsistent mit dieser Beobachtung ist, daß die Duplikation von CHIR-Genen vor allem *en bloc* erfolgt sein muß, da die hier durchgeführten phylogenetischen Analysen zeigen, daß eine Durchmischung von Domänen bei CHIRs ein seltenes Ereignis ist (siehe Abschnitt 4.4).

Wie groß die Dynamik der CHIR-Gene ist, läßt sich durch die Sequenzanalyse nur eines Individuums nicht feststellen. Die große Zahl der Pseudogene und Genfragmente deutet jedoch auf eine hohe Dynamik hin (siehe unten). Bei der Untersuchung der KIR-Gene des Menschen bei mehreren Haplotypen konnte festgestellt werden, daß sich die Variabilität der Gene auch in der An- bzw. Abwesenheit einzelner Gene in bestimmten Haplotypen widerspiegeln kann (Trowsdale 2001). Letztlich führt dies zusätzlich zu einer weiteren Diversifizierung des Rezeptorrepertoires. Das Gleiche könnte auch auf die CHIR-Gene zutreffen. Die Untersuchung der Sequenz der CHIR-Gene von verschiedenen Haplotypen würde zusätzlich Licht auf das Ausmaß des Polymorphismus der einzelnen Gene werfen. So wäre denkbar, daß nur bestimmte Gruppen polymorph sind. Vielleicht interagieren verschiedene CHIR-Gruppen auch mit verschiedenen Liganden, was sich in einem unterschiedlichen Grad an Polymorphismus widerspiegeln könnte. Solche Untersuchungen sollten also Licht auf funktionelle Aspekte der CHIR-Gene werfen und Rückschlüsse auf ihre Entwicklung zulassen.

4.7.2 Struktur und Funktion der CHIRs

Die CHIR-Multigenfamilie kodiert für Rezeptoren der Ig-SF. In der Regel besitzen CHIRs zwei Ig-Domänen des C2-Typs im extrazellulären Teil des Rezeptors. Daneben kommen aber auch einige CHIRs vor, denen die zweite Ig-Domäne fehlt. CHIRs kodieren für eine Reihe von funktionell unterschiedlichen Rezeptoren, die sich durch die vorhandenen Signaltransduktionsmotive unterscheiden. Insgesamt konnten sieben verschiedene Varianten identifiziert werden. Von einigen der funktionellen CHIR-Varianten sind keine vergleichbaren Rezeptoren bei Säugern bekannt. Vor allem die bifunktionellen Rezeptoren, die sowohl eine geladene Aminosäure in der Transmembrandomäne tragen und zwei ITIMs bzw. ein ITIM und ein YxxM-Motiv in der zytoplasmatischen Domäne besitzen, sind ungewöhnlich. Vergleichbare bifunktionelle Rezeptoren der Ig-SF bei Säugern sind KIR2DL4 (Faure und Long 2002; Yusa et. al. 2002) und NKp44 (Vitale et. al. 1998); beide besitzen eine geladene Aminosäure in der TM, aber nur ein ITIM. KIR2DL4 kann sowohl aktivierende als auch inhibitorische Signale in das Zellinnere weiterleiten (Rajagopalan et. al. 2001; Faure und Long 2002), während für NKp44 gezeigt wurde, daß das ITIM die aktivierende Funktion von NKp44 nicht blockieren kann (Vitale et. al. 1998; Campbell et. al. 2004). Welche Funktion den entsprechenden CHIR-Varianten zukommt, kann daher nur schwer abgeschätzt werden. Alle bifunktionellen Rezeptoren weisen jedoch eine sehr hohe Sequenzhomologie untereinander auf, was auf eine kürzlich erfolgte Expansion dieser Gene schließen läßt. Es bleibt daher zu vermuten, daß ihnen eine wesentliche Rolle im Immunsystem des Huhns zukommt.

Nikolaidis und Kollegen (Nikolaidis et. al. 2005b) unterscheiden nur 6 CHIR-Varianten, die sich jedoch mit den hier beschriebenen decken (es fehlt die Variante CHIRC-04). Ob es sich um die exakt gleichen Typen handelt, ist jedoch schwierig zu beurteilen, da die Daten sehr knapp beschrieben wurden. Offensichtlich ist jedoch, daß auf Rezeptoren mit YxxM-Motiv im zytoplasmatischen Teil nicht eingegangen, sondern nur auf ITIMs verwiesen wird. In einer Darstellung zur Konservierung des zytoplasmatischen Schwanzes (s. Abbildung 1) ist jedoch erkennbar, daß ein Teil der ITIMs ein Methionin an Position 6 aufweist. Es sollte sich also eher um ein YxxM-Motiv handeln. Da das YxxM-Motiv im Gegensatz zum ITIM vermutlich eher aktivierende Signale vermittelt, würden sich hier abweichende Interpretationen zur Funktion der Rezeptoren ergeben. Darauf wird jedoch von (Nikolaidis et. al. 2005b) nicht eingegangen. Weiterhin wird nicht erwähnt, daß sich die geladene Aminosäure in der TM-Region einiger bifunktionaler Rezeptoren unabhängig von den aktivierenden Rezeptoren entwickelt haben muß. Darüber hinaus wurden fünf Rezeptorgene (CHIRA-6, -18, CHIRC-4, -5 und CHIRB-14) den falschen Varianten (aktivierend/inhibierend/bifunktional) zugeordnet (Nikolaidis et. al. 2005b).

Neben den strukturellen Eigenschaften der Rezeptoren bestimmen vor allem ihre Liganden die Funktion. Bisher war es nicht möglich, einen Liganden für einen der zahlreichen CHIRs zu identifizieren. Die in dieser Arbeit durchgeführten Analysen zur Variabilität der CHIRs (Abschnitt 4.5) sowie die Vorhersagen der Molekülstruktur geben einen Hinweis auf die möglichen Liganden der CHIRs, bei denen es sich sehr wahrscheinlich um variable Moleküle handelt. Ob es sich um die MHC-Moleküle des Huhns handelt ist zwar Spekulation, aber es ist auffällig, daß bei den KIRs des Menschen und bei LY49-Genen von Ratte und Maus, bei denen es sich um auf NK-Zellen exprimierte Rezeptoren von MHC-Klasse-I-Molekülen handelt, ebenfalls eine Expansion der Zahl der Gene festgestellt wurde (Volz et. al. 2001; Trowsdale et. al. 2001; Hao und Nei 2004). Die mögliche Interaktion von CHIRs mit MHC-Klasse-I-Molekülen des Huhns ist besonders interessant, weil der MHC des Huhns sehr viel einfacher, kompakter und kleiner als bei Säugern ist (Kaufman et. al. 1999b). Jeder Haplotyp exprimiert ein dominantes MHC-Klasse-I-Molekül mit unterschiedlichem Expressionslevel. Durch die kleinere Anzahl an MHC-Klasse-I-Genen kommt es zu einem deutlich höheren Selektionsdruck auf die einzelnen Gene, als bei einer Multigenfamilie. Entsprechend konnte beim Huhn gezeigt werden, daß bestimmte MHC-Haplotypen mit der Resistenz bzw. Anfälligkeit gegenüber bestimmten Infektionen assoziiert sind (zur Übersicht Zekarias et. al. 2002). Neben der eingeschränkten Präsentation viraler Peptide durch bestimmte MHC-Haplotypen wäre auch die fehlende Erkennung des Verlusts von MHC-Molekülen durch NK-Zellen eine

mögliche Erklärung des Zusammenhangs bestimmter MHC-Haplotypen mit Anfälligkeiten gegenüber Infektionen. In diesem Kontext wäre eine Rolle von CHIRs denkbar. So wäre es möglich, daß bestimmte MHC-Moleküle nur schlecht oder gar nicht von CHIRs gebunden werden können. Während der Entwicklung würde dies vermutlich zur Selbst-Toleranz führen, wie es zum Beispiel bei MHC-defizienten Mäusen gezeigt wurde (Andersson et. al. 1998). Ein Verlust des entsprechenden MHC-Moleküls würde jedoch der Erkennung durch CHIRs entgehen und somit könnte ein bestimmter MHC-Haplotyp zur verstärkten Anfälligkeit führen.

In Verbindung mit Virusinfektionen ist weiterhin interessant, daß viele Viren Strategien entwickelt haben, einer Immunantwort zu entgehen. So wird eine T-Zell-Antwort unterbunden, indem die MHC-Klasse-I-Expression auf der Oberfläche der infizierten Zelle durch verschiedene Mechanismen blockiert wird (Ambagala et. al. 2005). Gleichzeitig kommt es zum Beispiel beim humanen Cytomegalovirus (hCMV) zur Expression eines MHC-ähnlichen Virusproteins (UL18), welches als Ligand für inhibitorische Rezeptoren von MHC-Klasse-I-Molekülen dient (zur Übersicht Lodoen und Lanier 2005). Letzteres inhibiert eine NK-Zellantwort, die durch den Verlust der Expression von MHC-Klasse-I-Molekülen hervorgerufen werden würde. Es ist erwähnenswert, daß T-Zelllinien des Huhns nach Transfektion mit dem Virus der Marekschen Krankheit resistent gegenüber der Lyse durch NK-Zellen sind, was auf die Expression von virus-kodierten Liganden für inhibitorische Rezeptoren - möglicherweise CHIRs- auf den T-Zellen hindeutet (Sharma und Okazaki 1981).

Zur genaueren Charakterisierung der Funktion der CHIRs wäre es nötig, auch die Gewebeverteilung der einzelnen CHIRs zu bestimmen, da bestimmte Rezeptoren möglicherweise spezielle Funktionen in Subpopulationen von Leukozyten wahrnehmen. In diesem Zusammenhang zeigen die EST-Analysen (siehe Abschnitt 4.3.3 auf Seite 70), daß CHIRs vor allem in Leukozyten exprimiert werden. Das Expressionsmuster der einzelnen CHIR-Gene kann aus diesen Daten aber nicht abgeleitet werden. Zumindest für ein CHIR-Protein konnte jedoch gezeigt werden, daß es präferentiell auf B-Lymphozyten exprimiert wird (Viertlboeck et. al. 2004).

4.7.3 Evolution der CHIR-Gene

Die Daten der Stammbaumanalysen zeigen, daß sich die CHIR-Gene in Gruppen einteilen lassen, wobei Gene mit gleichen strukturellen und funktionellen Eigenschaften in der Regel in eine Gruppe fallen. Erstaunlicherweise spiegelt die genomische Organisation der CHIR-Gene keine derartigen Gruppierungen wider. Vielmehr liegen die verschiedenen Gene durchmischt

in den Contigs vor. Diese Beobachtung läßt den Schluß zu, daß es während und vermutlich auch nach der Expansion des Clusters zu Rearrangements in der CHIR-Region gekommen ist. Außerdem zeigte die Analyse der Intergenregionen, daß Genomabschnitte dupliziert wurden, die mehr als ein Gen enthalten. Auf diese Weise kann sich gleichzeitig die Zahl von strukturell unterschiedlichen Genen erhöhen, weil sie zusammen dupliziert werden. Vermutlich expandierten alle CHIR-Typen etwa zur gleichen Zeit. Dies wird vor allem durch die kurzen Astlängen der Stammbäume innerhalb der Gruppen gestützt. Die kurzen Astlängen zeigen zudem, daß die Expansion des Clusters vermutlich ein evolutionär junges Ereignis ist.

Im Gegensatz zu den CHIRs kann bei Stammbaumanalysen von Rezeptoren des LRC des Menschen keine Gruppierung nach strukturellen Merkmalen festgestellt werden (Martin et. al. 2002). Die verschiedenen strukturellen Varianten der LRC-Gene sind aber wie bei CHIRs zufällig in den Clustern durchmischt. Ein auffälliger Unterschied ist jedoch, daß viele LRC-Gene die gleiche Transkriptionsrichtung haben (Wende et. al. 2000), der Duplikationsmechanismus scheint sich also von dem innerhalb des CHIR-Clusters zu unterscheiden.

Die Stammbaumanalysen, die von Nikolaidis und Kollegen 2005 durchgeführt wurden, führten zu vergleichbaren Ergebnissen mit den Stammbaumanalysen dieser Arbeit. Allerdings wurden von Nikolaidis und Kollegen Stammbäume auf der Grundlage von Aminosäuresequenzen erstellt. Aus diesem Grund sind die Astlängen kürzer, als bei den hier erstellten Stammbäumen. Dies ist darauf zurückzuführen, daß die phylogenetische Differenzierung zwischen den CHIR-Sequenzen feiner ist, wenn die Stammbäume auf DNA-Sequenzen basieren, weil stille Mutationen in aminosäurebasierenden Stammbäumen keine Berücksichtigung finden. Für sehr ähnliche Gene ist deshalb die Verwendung von DNA-Sequenzen sinnvoller. Der genaue Vergleich der Daten gestaltete sich jedoch aus anderen Gründen schwierig: Zum einen sind in den Stammbäumen nur 65 der 70 gefundenen Gene enthalten (Abbildung 2 in Nikolaidis et. al. 2005b), weshalb sich die Frage stellt, ob doch nur 65 Gene funktionell sind. Zum anderen wird ein Gen im Stammbaum gezeigt (*CHIR2DL25*), welches in den Genkarten nicht vorkommt. Zwei Gene sind in einem der Stammbäume doppelt enthalten, allerdings in verschiedenen Gruppen (Abbildung 3 in Nikolaidis et. al. 2005b). Ob es sich bei diesen Diskrepanzen nur um Schreibfehler handelt, kann aus den präsentierten Daten nicht abgelesen werden.

Der LRC der Säuger kodiert für verschiedene Familien von Ig-Rezeptoren wie KIRs, LILRs oder LAIRs (Wende et. al. 2000). Die geringe Sequenzübereinstimmung der CHIRs mit diesen Rezeptorfamilien erlaubt es nicht, einen gemeinsamen Vorfahren zu bestimmen. Die hohe Sequenzvariabilität und das Vorhandensein von 2 Ig-Domänen ähneln jedoch eher

den KIR-Rezeptoren. KIRs werden jedoch fast ausschließlich auf NK-Zellen exprimiert (Colonna und Samaridis 1995), eine Eigenschaft, die nicht auf CHIRs zutrifft. Ihr Expressionsmuster entspricht eher dem der LILR-Familie, deren Mitglieder auf verschiedenen Zellen der myeloischen Reihe exprimiert werden (Colonna et. al. 2000). LILRs wiederum sind weniger variabel und besitzen vier Ig-Domänen. Eine eindeutige Zuordnung von CHIRs zu einer der Familien scheint also nicht möglich, vermutlich hatten Vögel und Säuger den letzten gemeinsamen Vorfahren, als sich die Familien der LRC-Rezeptoren noch nicht separiert hatten (Nikolaidis et. al. 2005a). Dies trifft sicher für KIRs zu, die vermutlich vor rund 35 - 40 Millionen Jahren aus einem LILR-Gen hervorgegangen sind (Wilson et. al. 2000; Wende H. 2004). Eine genauere Abstammung der CHIRs wird möglicherweise im Zuge der Sequenzierung des Genoms des Huhns festgestellt werden können, falls weitere LRC-homologe Gene in der Nähe des CHIR-Clusters gefunden werden.

Die Duplikation von Genen ist ein in der Evolution weit verbreiteter Mechanismus zur Diversifizierung von Genfunktionen (Ohno 1970). Die große Anzahl der gefundenen CHIR-Pseudogene weist hierbei darauf hin, daß sich das CHIR-Cluster nach dem Modell der „birth and death“-Evolution entwickelt hat. Nach diesem Modell werden neue Gene durch Genduplikation erzeugt. Einige dieser Gene werden über lange Zeiten im Genom erhalten, während andere deletiert werden oder ihre Funktion durch schädliche Mutationen verlieren. Dieses Modell trifft auf die meisten Multigenfamilien des Immunsystems zu (Nei und Rooney 2005).

Die bei Duplikationen entstehenden Tochtergene können sich unabhängig voneinander entwickeln. Während eines der Gene die ursprüngliche Funktion erfüllt, kann sich die andere Kopie weiterentwickeln und neue Funktionen annehmen, ohne die Vitalität des Organismus zu beeinflussen. Genduplikationen sind bei Genen des Immunsystems weit verbreitet, man findet sie zum Beispiel bei MHC-Genen (Dawkins et. al. 1999), leichten und schweren Ig-Kettengenen (Hieter et. al. 1981; Honjo et. al. 1981), Fc Rezeptorgenen (Kinet und Launay 2000), lectin-artigen Rezeptoren des NK-Komplexes (Barten et. al. 2001) oder den Genen des LRC (Volz et. al. 2001; Trowsdale et. al. 2001; Martin et. al. 2002). Offensichtlich führen Genduplikationen im Immunsystem zu einem Selektionsvorteil für die Population. Gerade unter dem Aspekt des äußerst kompakten MHC des Huhns (Kaufman et. al. 1999a), bei dem einzelne Haplotypen mit der Anfälligkeit gegenüber bestimmten Infektionen korreliert werden können, ist die extrem große Zahl an CHIR-Genen interessant. Möglicherweise wirkt diese große Zahl an CHIR-Genen kompensatorisch auf die fehlende Komplexität des MHC des Huhns. Es wird spannend sein herauszufinden, welche Funktion dieser Genfamilie zukommt und welchen Beitrag CHIRs zur Immunantwort des Huhns leisten.