

**Die Genstrukturen von *Cobra Venom Factor*
und homologen Komplementgenen
der Kobra *Naja kaouthia***

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
im Fachbereich Chemie der Universität Hamburg

vorgelegt von

Holger Bammert

aus Eltville

Hamburg 2002

Der experimentelle Teil der vorliegenden Arbeit wurde von Januar 1998 bis Juli 2002 am Institut für Biochemie und Lebensmittelchemie der Universität Hamburg, Abteilung für Biochemie und Molekularbiologie, durchgeführt. Die Arbeit wurde im Arbeitskreis von Prof. Dr. Dr. C.-W. Vogel begonnen und im Arbeitskreis von Prof. Dr. R. Bredehorst beendet.

1. Gutachter: Prof. Dr. R. Bredehorst
2. Gutachter: Prof. Dr. Dr. C.-W. Vogel

Tag der Disputation: 31.01.2003

Meinen Eltern

INHALT

| | |
|---|-----------|
| Abkürzungsverzeichnis | III |
| 1 Einleitung | 1 |
| 1.1 Das Komplementsystem | 1 |
| 1.2 Der Cobra Venom Factor | 6 |
| 1.3 Die Struktur des humanen C3-Gens und homologer Komplementgene | 9 |
| 1.4 Herstellung genomischer Bibliotheken | 10 |
| 1.5 Ziel dieser Arbeit | 12 |
| 2 Material und Methoden | 13 |
| 2.1 Material | 13 |
| 2.1.1 Chemikalien und Bioreagenzien | 13 |
| 2.1.2 Enzyme | 13 |
| 2.1.3 Oligonukleotide | 13 |
| 2.1.4 Kits | 14 |
| 2.1.5 Genotypen der <i>E. coli</i> -Bakterienstämme | 14 |
| 2.1.6 Vektoren | 14 |
| 2.1.7 Kobragewebe | 14 |
| 2.1.8 Cosmid-Bibliothek | 15 |
| 2.1.9 Spezielle Materialien zum Screening | 15 |
| 2.1.10 Geräte | 15 |
| 2.1.11 Puffer und Lösungen | 16 |
| 2.2 Methoden | 19 |
| 2.2.1 Isolierung genomischer DNA aus Kobragewebe | 20 |
| 2.2.2 Isolierung von Plasmid- oder Cosmid-DNA | 20 |
| 2.2.3 Quantifizierung der DNA | 22 |
| 2.2.4 Restriktionsverdau | 22 |
| 2.2.5 Agarose-Gelelektrophorese | 23 |
| 2.2.6 DNA-Extraktion aus Agarose-Gelen | 23 |
| 2.2.7 Dephosphorylierung und Ligation von DNA-Fragmenten | 23 |
| 2.2.8 Polymerase-Kettenreaktion (PCR) | 24 |
| 2.2.9 ARMS-Methode | 25 |
| 2.2.10 Klonierung von PCR-Produkten | 26 |
| 2.2.11 Herstellung kompetenter Zellen | 27 |
| 2.2.12 Transformation | 28 |
| 2.2.13 Analyse von positiven Klonen | 29 |
| 2.2.14 Wachstum und Lagerung von Bakterienklonen | 29 |
| 2.2.15 DNA-Sequenzierung | 30 |
| 2.2.16 Screening mit dem Digoxigenin-System | 31 |
| 2.2.17 Southern-Blot | 36 |
| 2.2.18 Genome Walking | 37 |
| 2.2.19 Identifizierung regulatorischer Elemente | 40 |
| 2.2.20 Expression in <i>E. coli</i> | 41 |
| 2.2.21 Denaturierender Zellaufschluß | 41 |
| 2.2.22 Proteinreinigung mittels Affinitätschromatographie | 41 |
| 2.2.23 SDS-PAGE | 42 |

| | | |
|----------|---|------------|
| 3 | Ergebnisse | 45 |
| 3.1 | Struktur des CVF1-Gens | 45 |
| 3.1.1 | Charakterisierung eines CVF1-Cosmid-Klons | 45 |
| 3.1.2 | Identifizierung und Charakterisierung des Cosmid-Klons M26 | 55 |
| 3.1.3 | Aufklärung der CVF1-Genstruktur durch genomische PCR | 62 |
| 3.1.4 | Aufklärung der CVF1-Genstruktur durch Genome Walking | 80 |
| 3.2 | Struktur des CVF2-Gens | 101 |
| 3.2.1 | Aufklärung der CVF2-Genstruktur durch genomische PCR | 101 |
| 3.2.2 | Aufklärung der CVF2-Genstruktur durch Genome Walking | 105 |
| 3.2.3 | Ansatz zum proteinbiochemischen Nachweis von CVF2 | 112 |
| 3.3 | Struktur des Cobra C3-Gens | 115 |
| 3.3.1 | Aufklärung der Cobra C3-Genstruktur durch genomische PCR | 115 |
| 3.3.2 | Aufklärung der Cobra C3-Genstruktur durch Genome Walking | 119 |
| 3.4 | Hinweise auf die Existenz weiterer Gene der CVF/C3-Familie | 122 |
| 4 | Diskussion | 133 |
| 4.1 | Aspekte der angewandten Methoden | 133 |
| 4.2 | Die Struktur des CVF1-Gens | 137 |
| 4.3 | Die Struktur des CVF1-Gens im Vergleich zu verwandten Komplementgenen | 139 |
| 4.4 | Vergleich der Genstrukturen von CVF1, CVF2 und coC3 | 143 |
| 4.5 | Vergleich von Intronsequenzen | 144 |
| 4.6 | Regulation von CVF1, CVF2 und coC3 | 148 |
| 4.7 | Hinweise auf weitere Mitglieder der CVF/C3-Genfamilie | 155 |
| 4.8 | Evolution und Phylogenie der CVF/C3-Genfamilie | 155 |
| 4.9 | Ansatz zum Nachweis des CVF2-Proteins | 159 |
| 4.10 | Ausblick | 161 |
| 5 | Literatur | 163 |
| 6 | Zusammenfassung/Summary | 177 |
| 7 | Anhang | 181 |
| | Anhang 7.1 Sequenzen des CVF1-Gens | 181 |
| | Anhang 7.2 Sequenzen des CVF2-Gens | 203 |
| | Anhang 7.3 Sequenzen des coC3-Gens | 206 |
| | Anhang 7.4 Sequenzen weiterer putativer Gene der CVF/C3-Familie | 210 |
| | Anhang 7.5 Primersequenzen | 214 |
| | Anhang 7.6 PCR-Programme | 217 |
| | Veröffentlichungen | 223 |
| | Danksagung | |
| | Lebenslauf | |

Abkürzungsverzeichnis

| | |
|-------|---|
| amp | Ampicillin-Resistenz |
| AP | Alkalische Phosphatase |
| APS | Ammoniumpersulfat |
| ARMS | <i>amplification refractory mutation system</i> , System der amplifizierungs-resistenten Mutation |
| BAC | <i>bacterial artificial chromosome</i> , künstliches Bakterienchromosom |
| BCIP | 5-Brom-4-chlor-3-indoyl-phosphat |
| bp | Basenpaare |
| BSA | <i>bovine serum albumin</i> , Rinderserumalbumin |
| C3 | Dritte Komplementkomponente |
| CIAP | <i>calf intestinal alkaline phosphatase</i> , Alkalische Phosphatase aus Kälberdarm |
| cDNA | komplementäre DNA |
| co | cobra |
| CVF | <i>Cobra Venom Factor</i> |
| Cy5 | Carboindocyaninfarbstoff mit 5 C-Atomen in der Polymethinkette |
| Da | Dalton |
| DIG | Digoxigenin |
| DMSO | Dimethylsulfoxid |
| DNA | Desoxyribonukleinsäure (<i>deoxyribonucleic acid</i>) |
| dATP | Desoxyadenosintriphosphat |
| dCTP | Desoxycytosintriphosphat |
| dGTP | Desoxyguanidintriphosphat |
| dNTP | Desoxynukleosidtriphosphat |
| DTT | Dithiothreitol |
| dTTP | Desoxythymidintriphosphat |
| dUTP | Desoxyuridintriphosphat |
| EDTA | Ethylendiamintetraacetat |
| Fab | Antigen-bindendes Fragment |
| hu | human |
| kb | Kilobasenpaare |
| IPTG | Isopropyl- β -D-thiogalactosid |
| LacZ | β -Galactosidase-Gen |
| LB | Luria-Bertani |
| MAC | Membranangriffskomplex (<i>membrane attack complex</i>) |
| MHC | Haupt-Histokompatibilitätskomplex (<i>major histocompatibility complex</i>) |
| MCS | <i>Multiple Cloning Site</i> |
| NBT | Nitroblau-Tetrazoliumchlorid |
| OD | Optische Dichte |
| OPA | Oligonukleotid-Primer in antisense-Richtung |
| OPS | Oligonukleotid-Primer in sense-Richtung |
| ori | Replikationsursprung (<i>origin</i>) |
| PAC | <i>P1 Artificial Chromosome</i> = künstliches P1-Chromosom |
| PAGE | Polyacrylamidgelelektrophorese |
| PBS | <i>phosphate buffered saline</i> |
| PCR | Polymerase-Kettenreaktion (<i>polymerase chain reaction</i>) |
| PEG | Polyethylenglykol |
| pUC | <i>plasmid University of California</i> |
| RNA | Ribonukleinsäure |
| RNase | Ribonuklease |

| | |
|----------|--|
| rpm | Umdrehungen pro Minute (<i>revolutions per minute</i>) |
| RT | Raumtemperatur |
| SDS | Natriumdodecylsulfat (<i>sodium dodecyl sulfate</i>) |
| SDS-PAGE | SDS-Polyacrylamid-Gelelektrophorese |
| SSC | <i>Standard Saline Citrate</i> |
| STE | <i>Sodium-Tris-EDTA</i> |
| T3 | T3 RNA-Polymerase-Promotor |
| T7 | T7 RNA-Polymerase-Promotor |
| TAE | Tris-Acetat-EDTA |
| TBE | Tris-Borsäure-EDTA |
| TE | Tris-EDTA |
| TEMED | N,N,N',N'-Tetramethylethyldiamin |
| Tris | Tris-(hydroxymethyl)aminomethan |
| TSS | <i>Transformation and Storage Solution</i> |
| u | Einheit (unit) |
| UTR | untranslatierte Region |
| UV | Ultraviolett |
| V | Volt |
| xg | Mehrfaches der Gravitation |
| X-Gal | 5-Brom-4-chlor-3-indoyl- β -D-galactosid |
| YAC | <i>yeast artificial chromosome</i> |

Ein-Buchstaben-Code für Nukleotide:

| | |
|---|--------------|
| A | Adenosin |
| C | Cytidin |
| G | Guanosin |
| T | Thymidin |
| U | Uridin |
| R | A oder G |
| Y | C oder T |
| W | A oder T |
| M | A oder C |
| K | G oder T |
| S | C oder G |
| N | A,C,G oder T |

1 Einleitung

Gegenstand der vorliegenden Arbeit ist die Aufklärung der Genstruktur des Cobra Venom Factor (CVF), eines nicht-toxischen Proteins aus dem Gift der Kobra-Spezies *Naja*, *Ophiophagus* und *Hemachatus* der *Elapidae*-Familie (Vogel 1996). CVF weist strukturelle und funktionelle Ähnlichkeiten mit der zentralen Komplement-Komponente C3 auf und ist in der Lage, den Alternativen Weg des Komplementsystems zu aktivieren. Nachfolgend soll daher zunächst die Funktion des Komplementsystems und die Bedeutung von C3 sowie die wichtigsten strukturellen und funktionellen Eigenschaften von CVF dargestellt werden.

1.1 Das Komplementsystem

Das Komplementsystem ist ein Teil des Immunsystems höherer Organismen und stellt den wichtigsten humoralen Effektormechanismus dar (Müller-Eberhard 1988, Kinoshita 1991). Es ist ein komplexes System aus mehr als 30 verschiedenen löslichen und membranständigen Proteinen (Komplementfaktoren), die etwa 4% der Proteinmenge im Serum ausmachen. Der Ausdruck *Komplement* wurde vor mehr als 100 Jahren geprägt, um damit die hitzeempfindliche Aktivität eines Serums zu beschreiben, das bei der Lyse von Bakterien die Wirkung von spezifischen Antikörpern komplementiert (Ehrlich und Morgenroth 1899). Das Komplementsystem verrichtet im wesentlichen drei biologische Funktionen: neben der Lyse von Zielzellen sowie der Aktivierung von basophilen Granulozyten und Makrophagen (Chemotaxis) bewirkt es die Opsonisierung, d.h. die Markierung von Pathogenen für die Phagozytose, indem diese mit der Komplement-Komponente C3b besetzt werden (Frank und Fries 1991).

Die Aktivierungsmechanismen des Komplementsystems lassen sich in drei grundlegende Wege einteilen: den Klassischen Weg, den Alternativen Weg sowie den Lektin-Weg (Abb. 1.1). Diese drei Reaktionswege gleichen sich in ihrer Organisation und stellen einen Kaskadenmechanismus dar. Einige der beteiligten Komplementproteine sind Proenzyme (Zymogene), die wiederum durch andere Enzyme aktiviert werden. Zentrales Ereignis der Komplementaktivierung generell ist die Bildung einer C3-Konvertase, die C3 in C3a und C3b spaltet. Nachfolgend wird eine C5-Konvertase gebildet, die die Spaltung von C5 in C5a und C5b bewirkt. Die Anlagerung von C5b an Pathogene initiiert schließlich den terminalen Weg, der zur Bildung des Membranangriffskomplexes (MAC) und somit zur Lyse der Fremdzelle führt.

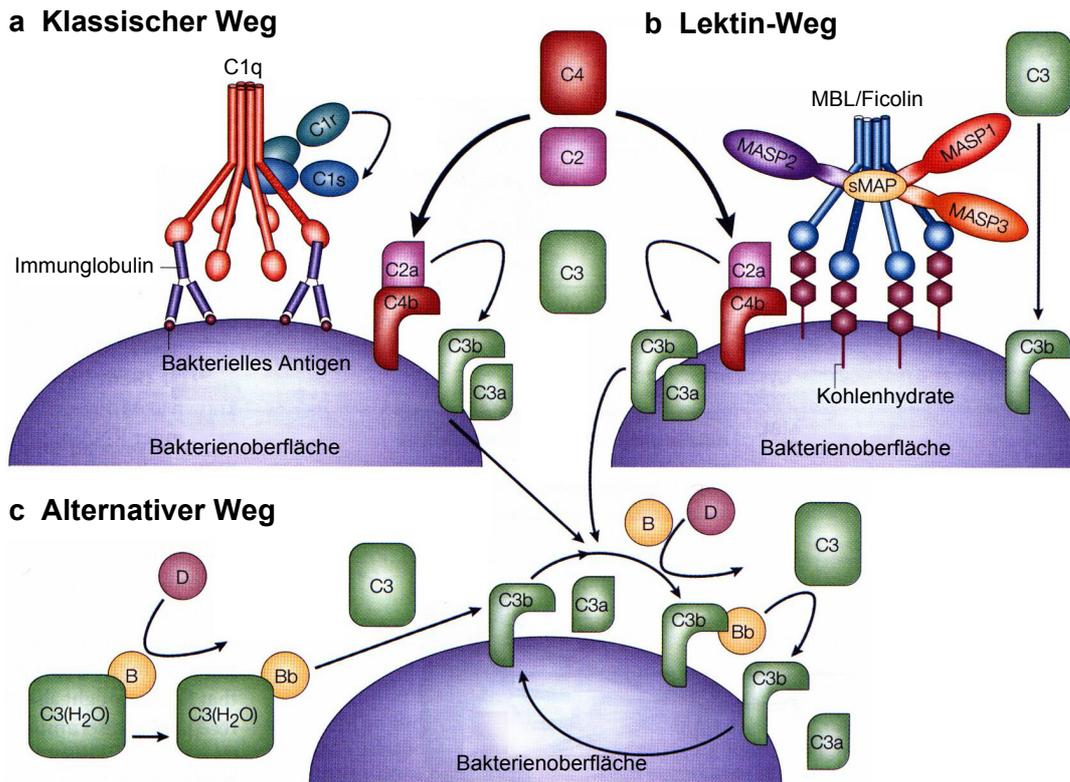


Abb. 1.1: Aktivierungswege des Komplementsystems. Das Komplementsystem kann durch den Klassischen Weg (a), den Lektin-Weg (b) und den Alternativen Weg (c) aktiviert werden (aus: Fujita 2002).

Klassischer Aktivierungsweg

Der Klassische Weg ist antikörperabhängig und wird aktiviert, nachdem zwei Fc-Regionen von Immunglobulinen der Klasse IgG oder IgM durch die Erkennungseinheit C1q des C1-Hexamers verknüpft werden (Schumaker *et al.* 1987, Sim und Reid 1991). Dadurch wird eine Autoaktivierung der Serinprotease C1r ausgelöst. C1r spaltet und aktiviert C1s, das wiederum die beiden Substrate C2 und C4 spaltet. Das entstandene C4b-Fragment assoziiert über einen reaktiven Thioester mit der Zellmembran und bildet zusammen mit C2a den Enzym-Komplex C4bC2a, die C3-Konvertase des Klassischen Wegs.

Lektin-Weg

Die Aktivierung des Lektin-Wegs erfolgt durch die Anlagerung von Mannose-bindendem Lektin (MBL) oder Ficolin an High-Mannose-Polysaccharide auf der Oberfläche von Hefen, Viren und Bakterienzellen (Thompson 1995). MBL und Ficolin sind dabei mit den MBL-assoziierten Serinproteasen 1 (MASP1), MASP2, MASP3 sowie sMAP (*small MBL-associated protein*) assoziiert. Ähnlich wie bei C1s wird durch MASP2 die Spaltung von C4 und C2 aktiviert, wodurch dieselbe C3-Konvertase (C4bC2a) wie beim Klassischen

Weg gebildet wird (Vorup-Jensen *et al.* 1998). Durch MASP1 kann C3 direkt gespalten werden (Rossi *et al.* 2001). Die Funktionen von sMAP und MASP3 sind noch nicht vollständig geklärt. MASP1, MASP2, MASP3 und sMAP werden durch zwei Gene codiert (Schwaeble *et al.* 2002). Bei sMAP handelt es sich um eine verkürzte Form von MASP2 (Thiel *et al.* 2000), und MASP3 wird durch alternatives Spleissen aus dem MASP1-Genprodukt gebildet (Dahl *et al.* 2001).

Alternativer Weg

Der Alternative Weg wird durch direkte Markierung von fremden Zelloberflächen durch aktiviertes C3 ausgelöst (Pangburn und Müller-Eberhard 1984). Dies wird dadurch ermöglicht, dass in der wässrigen Phase stets ein geringer Anteil von C3-Molekülen durch einen nukleophilen Angriff von Wasser auf den reaktiven Thioester hydrolysiert wird (Tack *et al.* 1980, Isenman *et al.* 1981). Dabei kommt es zu einer Konformationsänderung von C3 zu C3(H₂O), das auch als C3b-ähnlich bezeichnet wird. Ähnlich wie bei C3b wird durch Anlagerung von Faktor B und anschließende Spaltung durch Faktor D eine C3-Konvertase gebildet (C3(H₂O)Bb), die C3 in C3b und C3a spaltet. Der Alternative Weg wurde auch in Invertebraten wie dem Seeigel nachgewiesen (Leonard *et al.* 1990) und gilt daher als der evolutionär älteste Aktivierungsweg.

Amplifikationsschleife

Alle drei Aktivierungswege führen letztlich zur Bildung von C3-Konvertasen, die C3 aus dem Serum in C3b und das Anaphylatoxin C3a spalten. Durch Anlagerung von Faktor B an C3b und nachfolgende Spaltung durch Faktor D entstehen weitere C3-Konvertasen (C3bBb) des Alternativen Wegs (Abb. 1.1). Dieser Mechanismus führt zu einer deutlichen Verstärkung des ursprünglichen Signals.

Durch Anlagerung von C3b an die C3-Konvertasen entstehen C5-Konvertasen, die C5 in C5b und das Anaphylatoxin C5a spalten. Diese Reaktion mündet in den terminalen, lytischen Weg (Abb. 1.2).

Terminaler Weg

Die Funktion des terminalen Weges liegt in der Bildung des Membranangriffskomplexes (*membrane attack complex*, MAC), welcher zur Zerstörung der detektierten Pathogene führt.

Das in allen Aktivierungswegen gebildete C5a-Peptid diffundiert in die flüssige Phase, während das C5b-Fragment an der Membran gebunden bleibt und dort als Bindungsstelle für die schrittweise Anlagerung der löslichen Komplement-Komponenten C6-C9 dient. Durch die Bindung an C5b erfährt C6 eine Konformationsänderung, welche die weitere Bindung von C7 an den Komplex C5bC6 ermöglicht. Eine Konformationsänderung des gebundenen C7 legt dann die hydrophobe Region frei, über die sich C7 in die Lipid-Doppelschicht der Plasmamembran integrieren kann. Der Komplex C5b-7 bindet C8, welches durch eine Änderung vom hydrophilen zum amphiphilen Charakter die Zellwand durchdringen kann. Der Komplex C5b678 katalysiert die Polymerisation von C9. Die Anlagerung von bis zu 18 C9-Molekülen führt zur Bildung des Membranangriffskomplexes (MAC). Dabei entsteht eine Pore in der Membran mit einem Durchmesser von ca. 100 Å, durch die gelöste Stoffe unkontrolliert diffundieren können, was schließlich die Lyse der Zielzelle bewirkt.

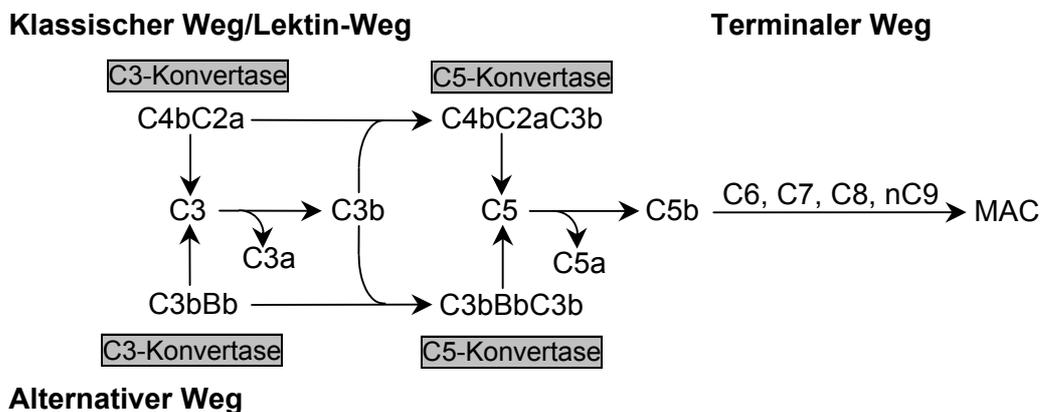


Abb. 1.2: Bildung der C5-Konvertasen und Terminaler Weg.

Regulation des Komplementsystems

Das Komplementsystem unterliegt einer strengen Kontrolle durch regulatorische Proteine (Inhibitoren), die körpereigene Zellen und Gewebe vor einem fehlgeleiteten Angriff durch Komplement schützen. Die Kontrollproteine beeinflussen drei Stadien der Komplementaktivierung durch folgende Mechanismen:

- (a) direkte Inhibierung von Serinproteasen
- (b) Zerfall und Zerstörung von Konvertasen
- (c) Kontrolle des Membranangriffskomplexes (MAC)

Tabelle 1 gibt eine Übersicht über die regulatorischen Proteine des Komplementsystems.

Tab. 1: Lösliche und membrangebundene Kontrollfaktoren der Komplementaktivierung.

I. Serumproteine

| Faktor | Liganden | Funktion |
|--------------|----------|---|
| C1-Inhibitor | C1r | Stabilisierung des C1q-r2-s2-Komplexes, Schutz der katalytischen Domäne |
| C4BP | C4b | Bindung an C4b, Zerfall des C4b2a-Komplexes |
| Faktor H | C3b | Bindung von aktiviertem Serum-C3b, Cofaktor-Aktivität für Faktor I-vermittelte Spaltung |
| Faktor I | C3b/C4b | Spaltung von C3b und C4b |
| Properdin | C3bBb | Stabilisierung der C3-Konvertase des alternativen Weges |
| S Protein | C5b-7 | Bindung von löslichen C5b-7-Komplexen zur Verhinderung von spontaner Zelllyse |

II. Membranproteine und Rezeptoren

| Faktor | Liganden | Funktion |
|------------|-----------------|--|
| CR1 (CD35) | C3b/4b, iC3b/4b | Zerfall der C3- und C5-Konvertasen, Cofaktor-Aktivität für Faktor I-vermittelte Spaltung |
| MCP (CD46) | C3b/4b | Cofaktor-Aktivität für Faktor I-vermittelte Spaltung |
| DAF (CD55) | C4b2a, C3bBb | Zerfall der klassischen und alternativen C3-Konvertasen |
| C8BP | C5b-8, C9 | Verhinderung der MAC-Bildung |
| CD59 | C5b-8, C9 | Verhinderung der MAC-Bildung |

Legende: C4BP - C4 binding protein, CR1 - complement receptor type 1, MCP - membrane cofactor protein, DAF - decay accelerating factor, C8BP - C8 binding protein.

Die regulatorischen Proteine Faktor H, C4BP, MCP, DAF und CR1 sind eng verwandte homologe Proteine, die aus kleinen Domänen (*short consensus repeats, SCR*) aufgebaut sind, von denen jede aus 60 Aminosäuren besteht (Sim *et al.* 1993). Diese Proteine sind zusammen mit Komplement-Rezeptor Typ 2 (CR2) alle in derselben Region des humanen Genoms, dem langen Arm von Chromosom 1, codiert, die als RCA (*regulation of complement activation*)-Gencluster bezeichnet wird (Hourcade *et al.* 1989).

1.2 Der Cobra Venom Factor

Der Cobra Venom Factor (CVF) ist ein Komplement-aktivierendes Protein aus dem Gift der Kobra-Spezies *Naja*, *Ophiophagus* und *Hemachatus* (Vogel *et al.* 1996). Die besondere Eigenschaft der Komplementaktivierung von Kobragift war schon seit Beginn des letzten Jahrhunderts bekannt (Flexner und Noguchi 1903) und führte 1964 zur erstmaligen Identifizierung und Charakterisierung von CVF (Vogt und Schmidt 1964). In der Folge wurden die molekularen Wechselwirkungen mit dem Komplementsystem aufgeklärt, wobei die funktionelle und strukturelle Analogie von CVF und C3b, der aktivierten Form von C3, entdeckt wurde (Alper und Balavitch 1976). Inzwischen konnte die Primärstruktur von CVF aufgeklärt werden (Fritzinger *et al.* 1994b).

CVF aus dem Gift der indischen Kobra *Naja kaouthia*¹ (Monokel-Kobra) ist ein dreikettiges Glykoprotein mit einer Molmasse von 149 kDa (Vogel *et al.* 1996). Es besteht aus drei über Disulfidbrücken verknüpften Ketten mit Molmassen von 68,5 kDa (α -Kette), 48,5 kDa (β -Kette) und 32 kDa (γ -Kette). Die γ -Kette tritt in zwei Größen auf. Diese Größen-Heterogenität ist vermutlich auf unterschiedliche Prozessierung am C-Terminus zurückzuführen (Eggertsen *et al.* 1981, Vogel und Müller-Eberhardt 1984).

Der Kohlenhydratanteil von CVF beträgt insgesamt 7,4% und setzt sich aus drei N-verknüpften Oligosaccharidketten vom komplexen Typ zusammen, von denen zwei N-Glykosylierungsstellen auf der α -Kette und eine auf der β -Kette liegen. Ungefähr 80% der Oligosaccharidketten bestehen aus symmetrisch fucosylierten biantennären Komplexen mit ungewöhnlichen α -Galactosyl-Resten an den nicht-reduzierten Enden, 12% bestehen aus tri- und tetraantennären Strukturen, 8% weisen Oligomannosestrukturen auf (Gowda *et al.* 2001).

Die Oligosaccharidketten von CVF sind allerdings vermutlich nicht für die Komplementaktivierung erforderlich, da auch deglykosyliertes CVF Komplement-aktivierend wirkt (Gowda *et al.* 1994).

Struktur und Funktion von CVF

Die CVF-mRNA hat eine Länge von mehr als 5928 bp und enthält einen offenen Leserahmen von 4926 Nukleotiden, der für ein Prä-pro-Protein aus 1642 Aminosäuren codiert (Fritzinger *et al.* 1994b). Die dreikettige Struktur von CVF entsteht durch posttranslationale Prozessierung, bei der die Argininreste 628-631 am C-Terminus der

¹ bis 1996 übliche taxonomische Bezeichnung: *Naja naja kaouthia* (Wüster 1996).

α -Kette sowie die zu C3 homologen Bereiche C3a und C3d entfernt werden (Fritzinger *et al.* 1994b, Abb. 1.3)

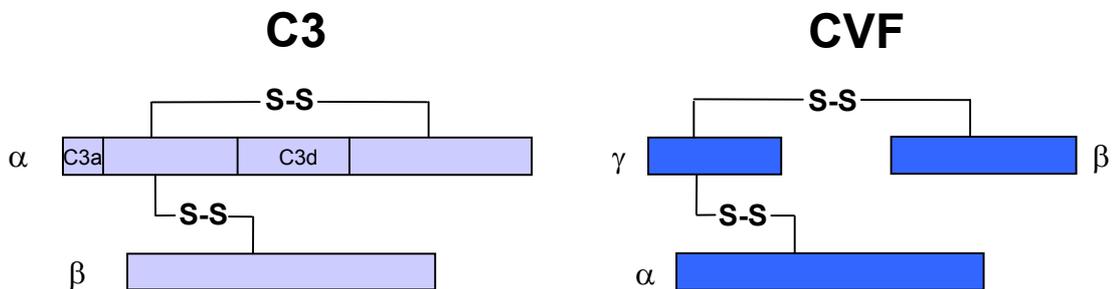


Abb. 1.3: Vergleich der Kettenstrukturen von C3 und CVF. Die zu C3a und C3d homologen Regionen sind improzessierten CVF nicht vorhanden.

Durch Zugabe von CVF zu Serum von Vertebraten erfolgt eine permanente Aktivierung des Alternativen Weges des Komplementsystems. Dabei bindet CVF ebenso wie C3b, die aktivierte Form von C3, an Faktor B.

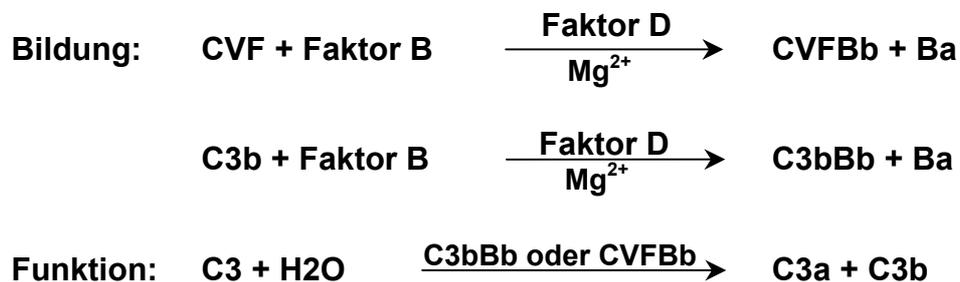


Abb. 1.4: Bildung und Funktion der CVF- bzw. C3b-abhängigen C3-Konvertasen (nach Vogel *et al.* 1996).

Als entscheidender Faktor für die Aktivität von CVF als Komplement-aktivierendes Protein ist seine Befähigung darin zu sehen, in Analogie zu C3 Konvertasen auszubilden, die eine weitere Prozessierung von Komplement-Proteinen nach sich ziehen. Dies wird ermöglicht durch die signifikante Homologie von CVF zu humanem C3 auf cDNA-Ebene von 56,3%. Sowohl CVF als auch C3b binden Faktor B, der in Anwesenheit von Mg^{2+} -Ionen durch Faktor D gespalten wird (Abb. 1.4). Die dabei entstehenden Komplexe CVFBb und C3bBb sind C3-Konvertasen des Alternativen Wegs.

Trotz aller Homologie zeigt CVF signifikante Unterschiede gegenüber C3, die sich in erhöhter funktioneller Aktivität äußern. So zeichnet sich die CVF-abhängige C3-

Konvertase durch eine besonders hohe Stabilität aus. Die Halbwertszeit bei 37°C beträgt bei C3bBb 1,5 min (Medicus *et al.* 1976), wohingegen CVFBb mit 7 Stunden (Vogel und Müller-Eberhard 1982) eine drastisch gesteigerte Halbwertszeit aufweist. Zudem ist die CVFBb-Konvertase resistent gegenüber den Faktoren H und I, während die C3bBb-Konvertase durch die Faktoren H und I inaktiviert wird. Darüber hinaus verfügt CVFBb über eine C5-Konvertase-Aktivität, die bei C3Bb erst durch Anlagerung weiterer C3b-Moleküle erlangt wird (Daha *et al.* 1976).

Die biologische Funktion des nicht-toxischen Kobrafaktors besteht vermutlich in der Verstärkung der Giftwirkung anderer Komponenten des Kobragifts. Bei der durch CVF ausgelösten Dekomplementierung werden die Anaphylatoxine C3a und C5a in grossen Mengen freigesetzt. Diese verursachen die Degranulation von Mastzellen und Basophilen und so nachfolgend einen lokalen Anstieg der Vaskularpermeabilität. Die Erhöhung der Vaskularpermeabilität an der Stelle der Giftinjektion erleichtert den Toxinen den Eintritt in den Blutkreislauf und ermöglicht eine schnellere Wirkung.

Zusätzlich zu CVF wurde beim Screening einer cDNA-Bibliothek aus Giftdrüsen-mRNA die Sequenz eines weiteren Cobra Venom Factor (CVF2) identifiziert. Die CVF2-cDNA-Sequenz wurde partiell aufgeklärt und zeigt eine sehr hohe Homologie zu CVF1 und coC3 (Fritzinger *et al.* 1992a, 1994a). Die cDNAs von CVF1, CVF2 und coC3 sind zu jeweils mehr als 91% identisch. Die biologische Relevanz von CVF2 ist nach wie vor unklar, zumal das korrespondierende Protein noch nicht nachgewiesen wurde.

Mögliche Anwendungen von CVF

Die hohe Physikostabilität der CVF-abhängigen C3/C5-Konvertase und die daraus folgende kontinuierliche Proteolyse von C3 und C5 führt zu einer vollständigen Dekomplementierung des Serums abwärts der Faktoren C3 und C5. Diese Eigenschaft von CVF könnte die unterschiedlichsten Anwendungen in therapeutischen und analytischen Fragestellungen ermöglichen.

Vollständige Dekomplementierung konnte für einen Zeitraum von 7-10 Tagen in Versuchstieren gezeigt werden. Dieser ausgedehnte Zeitraum ermöglicht *in vivo*-Untersuchungen zum Komplementsystem in der Pathogenese unterschiedlichster Erkrankungen und in der Immunabwehr (Cochrane *et al.* 1970, Ryan *et al.* 1986). Da auch bei der Gewebeabstoßung von Xenotransplantaten das Komplementsystem eine wichtige

Rolle spielt, bietet sich hier ein Einsatz von CVF als Immunsuppressivum an. So konnte die Abstoßungszeit von Xenotransplantaten durch Dekomplementierung mit CVF von 90 min auf 92 h verlängert werden (Leventhal *et al.* 1993). Kürzlich generierte CVF-transgene Mäuse, die permanente Dekomplementierung zeigen, können hier und in anderen Bereichen als Modellsystem für Untersuchungen von Komplement-assoziierten Erkrankungen dienen (Andrä *et al.* 2002).

Eine Umgehung der Komplement-Resistenz vieler Tumoren durch Nutzung der Konvertase-Stabilität gegenüber Regulator-Proteinen konnte durch den Einsatz von CVF-gekoppelten Antikörpern (Vogel 1987, Petrella *et al.* 1987) gezeigt werden. Daraus ergeben sich mögliche Anwendungen von Antikörper-CVF-Konjugaten in der Tumorthherapie (Vogel *et al.* 1985, Juhl *et al.* 1995).

Die rekombinante Expression von CVF in geeigneten zellulären Systemen wie dem Baculovirus-System zeigte zwar keine vollständige proteolytische Spaltung von pro-CVF zu einem dreikettigen Molekül, dennoch wies das rekombinante Protein vergleichbare Aktivität auf (Kock 1996). Prinzipiell könnten mittels rekombinanter Expression mittelfristig Analysen der Struktur und Funktion des CVF durchgeführt werden. Diese Informationen können so dazu beitragen, die weitgehende Humanisierung des Moleküles zu ermöglichen, das dann unter Vermeidung seiner immunogenen Eigenschaften für wiederholte therapeutische Anwendungen genutzt werden könnte.

1.3 Die Struktur des humanen C3-Gens und homologer Komplementgene

Das humane C3-Gen ist ein single copy-Gen, das auf dem kurzen Arm von Chromosom 19 (19p13.3-p13.2) lokalisiert ist (Schneider und Rittner 1997). Es weist eine Gesamtlänge von 42 kb auf und ist aus 41 Exons und 40 Introns zusammengesetzt (Vik *et al.* 1991). In seinem molekularen Aufbau ist es hochkonserviert und zeigt Ähnlichkeiten mit den Genstrukturen von C4 (human C4A: Yu 1991, murin C4: Ogata *et al.* 1989) und C5 (human C5: Carney *et al.* 1991, murin C5: Haviland *et al.* 1991b), die allesamt einer Genfamilie zugehörig sind und vermutlich aus einem gemeinsamen Vorläufergen entstanden (Hughes 1994; Sottrup-Jensen *et al.* 1985). Das α 2-Macroglobulin gehört ebenfalls der selben Familie an, obwohl es nur von 36 Exons codiert wird (Matthijs *et al.* 1992). Die Komplement-Komponenten C3, C4 sowie α 2-Macroglobulin zeichnen sich durch die Existenz einer aktiven Thioester-Region aus, die durch Exon 24 codiert wird.

Die humane Komplementkomponente C4 stellt insofern eine Besonderheit dar, da sie durch die Gene C4A und C4B codiert wird, die im MHC III-Komplex auf Chromosom 6 lokalisiert sind. Obwohl die Genprodukte bis auf 4 Aminosäuren identisch sind, unterscheiden sich die Proteine erheblich in ihrer Aktivität: C4A ist 100fach reaktiver mit Zielzellen, die freie Aminogruppen enthalten, und 10fach weniger reaktiv mit Hydroxylgruppen als C4B (Isenman und Young 1984, Yu *et al.* 1986). Beide C4-Gene können Längen von 14,6 kb oder 21 kb aufweisen. Diese Größen-Dichotomie ist auf das Vorhandensein des endogenen Retrovirus HERV-K (C4) (*human endogenous retrovirus*) in Intron 9 zurückzuführen (Tassabehji *et al.* 1994, Dangel *et al.* 1994).

1.4 Herstellung genomischer Bibliotheken

Für die Charakterisierung von genomischen Strukturen und Chromosomenabschnitten werden zwei prinzipiell unterschiedliche Strategien verfolgt. Der Transfer der genomischen Sequenzabschnitte in eine replizierbare Form wird unter Verwendung von entsprechenden Vektoren sichergestellt. Alternativ werden nicht-replizierbare Fragmente der Genstrukturen durch molekularbiologische Modifikationen fragmentierter DNA erhalten und können damit einer detaillierten Analyse zugänglich gemacht werden.

Vektor-Systeme zur Analyse von Genabschnitten

Für die Herstellung von Bibliotheken eukaryontischer Genome werden Klonierungsvektoren benötigt, die für die Aufnahme besonders umfangreicher DNA-Fragmente geeignet sind. Vektoren wie der Phage Lambda und Cosmide (Collins und Hohn 1978, DiLella und Woo 1987) mit Aufnahmekapazitäten von 20 bzw. 45 kb sind für die Analyse einzelner Gene geeignet. Zu den Klonierungssystemen, die ebenfalls *E. coli* als Wirt verwenden, zählen weiterhin der Phage P1 (Pierce *et al.* 1992), PACs (*P1-derived artificial chromosomes*; Ioannou *et al.* 1994), Fosmide (Kim *et al.* 1992) und BACs (*bacterial artificial chromosomes*; Shizuya *et al.* 1992). Diese Vektoren haben Aufnahmekapazitäten von 100 kb (P1) bis zu 300 kb (BACs) und werden für die Sequenzanalyse ganzer Chromosomen oder eines Gesamtgenoms eingesetzt.

Die Entwicklung von YACs (*yeast artificial chromosomes*), die in der Hefe *Saccharomyces cerevisiae* als künstliche Hefechromosomen repliziert werden können, erlaubt die Klonierung von DNA-Fragmenten mit 1 Mb Länge (Burke *et al.* 1987). Allerdings haben YACs den Nachteil, dass die Bildung chimärer Klone bzw. Rekombinationen häufig vorkommen.

Cosmide als Klonierungsvektoren

Cosmide sind eine Weiterentwicklung der Phagenvektoren und sind im wesentlichen Plasmide, die neben den Sequenzen für Replikation und Selektion in bakteriellen Wirten noch eine für den Phagen Lambda relevante Sequenz, die *cos*-Erkennungsstelle (*cohesive*) mit einer Länge von 12 bp, tragen (Furth und Wickner 1983).

Während der Replikation des Phagen Lambda in *E. coli* entstehen DNA-Concatamere mit *cos*-Signalen, die durch das Lambda-Genom (~45 kb) getrennt sind. Die Verpackungsmaschinerie schneidet die *cos*-Stellen und verpackt interne DNA in einen Phagenkopf. Cosmide nutzen dieses Prinzip, wobei Fremd-DNA mit einer Größe von 35-45 kb und zwei eingeführten *cos*-Enden in Phagenhüllen verpackt wird. Diese Partikel sind infektiös wie Phagen, doch sobald die DNA in ein Bakterium eingebracht ist, kann sie wie Plasmide repliziert und permanent kultiviert werden.

Obwohl ursprünglich für die Isolation singulärer Gene entwickelt, konnten Cosmide auch erfolgreich zur Erstellung physikalischer Karten kompletter Genome wie *E. coli* (Kohara *et al.* 1987), *C. elegans* (Waterston und Sulston 1995), *S. cerevisiae* (Riles *et al.* 1993) oder *S. pombe* (Hoheisel *et al.* 1993) eingesetzt werden.

Genome Walking

Neben der Klonierung genomischer DNA-Fragmente in zirkuläre Vektoren bietet sich auch die Ligation kurzer linearer Adapter bekannter Sequenz an (Siebert *et al.* 1995). Diese Methode eignet sich zur Analyse von Sequenzen, die in direkter Nachbarschaft zu einer bekannten Sequenz liegen. Durch verschachtelte Polymerase-Kettenreaktionen (*polymerase chain reactions*, PCRs) mit genspezifischen Primern und Adapter-Primern kann so die gesuchte Sequenz gezielt amplifiziert und anschließend analysiert werden. Diese als Genome Walking bezeichnete Methode eignet sich besonders zur Analyse von Exon/Intron-Übergängen sowie von 5'- und 3'-flankierenden Sequenzen eines Gens.

1.5 Ziel dieser Arbeit

Ziel dieser Arbeit war die Aufklärung der Genstruktur von Cobra Venom Factor 1 (CVF1) aus der Monokel-Kobra *Naja kaouthia*. Neben der Nutzung einer Cosmid-Bibliothek sollten Methoden zur Analyse genomischer Strukturen etabliert werden. Darüber hinaus sollte der strukturelle Aufbau der Gene von CVF1, CVF2, coC3 und humanem C3 miteinander verglichen sowie die Promotorsequenzen und putative regulatorische Elemente identifiziert werden.

2 Material und Methoden

2.1 Material

2.1.1 Chemikalien und Bioreagenzien

Gängige Chemikalien und Bioreagenzien wurden von den Merck (Darmstadt), Sigma-Aldrich (Deisenhofen), Roth (Karlsruhe) oder Peqlab (Erlangen) bezogen. Speziellere Reagenzien sind jeweils gesondert aufgeführt.

2.1.2 Enzyme

Restriktionsenzyme

Restriktionsenzyme sowie die entsprechenden Puffer wurden von MBI Fermentas oder von NEB bezogen.

DNA-Polymerasen

AGS GoldTM DNA-Polymerase (Hybaid AGS)

FastStart Taq DNA Polymerase (Roche)

Sawady "Mid Range PCR"-System (Peqlab)

AGS Taq/Pwo (Proof Sprinter) (Hybaid AGS)

Expand Long Template PCR System (Roche) (Taq, Tgo)

Expand 20 kbPLUS (Roche)

Andere Enzyme

RNAse, Proteinase K, Polynukleotidkinase, T4-DNA-Ligase, CIAP (*Calf intestinal alkaline phosphatase*), T7-Polymerase, Mungbean-Nuklease wurden von MBI Fermentas, NEB oder Sigma bezogen.

2.1.3 Oligonukleotide

Oligonukleotide wurden von den Firmen Pharmacia Biotech (Freiburg), MWG Biotech (Ebersberg) oder Metabion (Martinsried) bezogen.

2.1.4 Kits

E.Z.N.A. Plasmid Miniprep Kit (Peqlab, Erlangen)

Genome Walker Kit (Clontech)

QIAEX II Gel Extraction Kit (Qiagen, Hilden)

Sure Clone Ligation Kit (Pharmacia)

Dye Ex Spin Kit (Qiagen, Hilden)

TOPO-TA Cloning Kit (Invitrogen, Groningen, NL)

BigDye Terminator Ready Reaction Kit (Applied Biosystems, Langen)

2.1.5 Genotypen der *E. coli*-Bakterienstämme

TOP 10

F⁻, *mcrA*, $\Delta(mrr-hsdRMS-mcrBC)$, $\Phi 80lacZ\Delta M15$, $\Delta lacX74$, *recA1*, *deoR*, *araD139*, $\Delta(ara-leu)7697$, *galU*, *galK*, *rpsL*, (Str^R) *endA1*, *nupG*

DH5 α

F⁻, $\phi 80dlacZ\Delta M15$, $\Delta(lacZYA-argF)U169$, *deoR*, *recA1*, *endA1*, *hsdR17*(rk⁻, mk⁺), *phoA*, *supE44*, λ^- , *thi-1*, *gyrA96*, *relA1*

BL21(DE3)

F⁻, *ompT*, *hsdS_B* (r_B⁻, m_B⁻), *dcm*, *gal*, λ (DE3), Cm^r

DH10b

F⁻, *mcrA*, $\Delta(mrr-hsdRMS-mcrBC)$, $\phi 80dlacZ\Delta M15$, $\Delta lacX74$, *deoR*, *recA1*, *endA1*, *araD139*, $\Delta(ara,leu)7697$, *galU*, *galK*, $\Delta rpsL$, *nupG*

2.1.6 Vektoren

Die Vektoren pCR2.1-TOPO und pCR-TOPO-XL waren Bestandteile der entsprechenden *TOPO TA Cloning Kits* (Invitrogen, Groningen, NL). Der Vektor pUC18 war Bestandteil des *Sure Clone Ligation Kits* (Pharmacia Biotech). Der Cosmid-Vektor sCos-1 wurde von Stratagene erworben. Der Expressionsvektor pET-26b(+) wurde von Novagen (Schwalbach) bezogen.

2.1.7 Kobragewebe

Gewebe der Kobra *Naja kaouthia* war im Arbeitskreis vorhanden und wurde bei -80 °C gelagert. Die Opferung der Kobra und die Bedingungen der Gewebe-Entnahme sind an anderer Stelle beschrieben (Fritzinger *et al.* 1992b).

2.1.8 Cosmid-Bibliothek

Eine Cosmid-Bibliothek der Kobra *Naja kaouthia* (Vektor sCos-1, Wirt: *E. coli* DH10b) mit einer Diversität von $4-5 \cdot 10^5$ Klonen wurde 1992 von Kunze erstellt und war im Arbeitskreis vorhanden.

2.1.9 Spezielle Materialien zum Screening

Amersham-Pharmacia Biotech (Freiburg):

Nylonmembranen Hybond NX (neutral) und Hybond N⁺ (positiv geladen)

Roche (Mannheim):

DIG-11-dUTP

Blockierungsreagenz

DIG Easy Hyb (Hybridisierungslösung)

Anti-DIG-AP-Konjugat, Fab-Fragmente

CSPD[®]-Substratlösung

Sigma (Deisenhofen):

Röntgenfilme (X-OMAT, Kodak, 18 x 24 cm und 30 x 40 cm)

Entwickler und Fixierer (Kodak GBX, Developer und Fixer Twin Pack)

2.1.10 Geräte

| | |
|---------------------------------------|---|
| Sepatech (Hannover) | Heraeus Megafuge 1.0R |
| Eppendorf (Hamburg) | Tischzentrifugen 5415 C, 5417C Mastercycler gradient |
| Amersham-Pharmacia Biotech (Freiburg) | UV/VIS-Spektrometer Ultrospec 3000 |
| Hybaid-AGS (Heidelberg) | PCR-Sprint-Thermoblock PCR-OmniGene-HB TR3-Block Hybridisierungsofen Shake'n Stack UV-Crosslinker BIO-LINK [®] BLX 254 |
| IBI, New Haven (USA) | System für Agarosegelelektrophorese |
| New Brunswick Scientific (Nürtingen) | Inkubationsschüttler innova [™] 4300 |

| | |
|--|--|
| Intas (Göttingen) | Video-Image-System |
| Steiner (Hamburg) | UV-Transilluminator (302 nm, 90 W) |
| Hoefler Scientific Instruments (San Francisco, CA, USA) | Minigel-Apparatur LKB 2050 MIDGET für SDS- PAGE |
| Gössner GmbH & Co. (Hamburg) | Laborautoklav GLA; Fabr. Nr. 940306 |

2.1.11 Puffer und Lösungen

| | |
|----------------------------|---|
| Acrylamid-Fertiglösung | 30% (w/v) Acrylamid 0,8 % (w/v) N,N-Methylen-bisacrylamid |
| Ampicillin Stamm-Lösung | 50 mg/ml (1000x) Ampicillin in sterilem Wasser lösen, Aliquots bei -20 °C lagern. |
| Blockierungslösung | 10% Blockierungsvorratslösung im Verhältnis 1:10 mit Maleinsäure-Puffer verdünnen |
| Chloramphenicol Stamm-Lsg. | 34 mg/ml (200x) Chloramphenicol in Ethanol lösen, Aliquots bei -20 °C lagern. |
| Church-Puffer | 7% SDS 50% Formamid 5x SSC 2% Blockierungsreagenz (Roche) 50 mM Natriumphosphat pH 7.0 0,1% (w/v) N-Lauroylsarkosin |
| Coomassie-Färbelösung | 0,25% (w/v) Coomassie Brilliant Blau R-250 10% (v/v) Eisessig 45% (v/v) Methanol sedimentieren |
| Coomassie-Entfärber | 45% (v/v) Methanol 10% (v/v) Eisessig |
| Denaturierungslösung | 0,5 M NaOH 1,5 M NaCl |
| Detektionspuffer (für AP) | 100 mM Tris-HCl, pH 9.5 100 mM MgCl ₂ 100 mM NaCl |

| | |
|--------------------------|---|
| Digestion-Puffer | 100 mM NaCl 10 mM Tris-HCl, pH 8 25 mM EDTA, pH 8 0,5% SDS 0,1 mg/ml Proteinase K Proteinase K ist labil und wurde vor jeder Benutzung frisch zugegeben. |
| DNA-Beladungspuffer (5x) | 20% (w/v) Ficoll 400 100 mM EDTA 0,025% (w/v) Bromphenolblau 0,025% (w/v) Xylencyanol FF |
| Ethidiumbromid-Stammlsg. | 10 mg EtBr/ml H ₂ O |
| Geltrocknungs-Lösung | 175 ml Ethanol (96%) 10 ml Glycerol (87%) ddH ₂ O ad 500 ml |
| Kanamycin-Stammlösung | 50 mg Kanamycin/ml H ₂ O (1000x) |
| LB-Medium | 10 g Trypton 5 g Hefe-Extrakt 10 g NaCl ad 1000 ml dH ₂ O, autoklavieren |
| LB-Agarplatten | wie LB-Medium 15 g Agar ad 1000 ml dH ₂ O, autoklavieren |
| Lösung I | 50 mM Glucose 25 mM Tris-HCl, pH 8.0 10 mM EDTA autoklavieren bei 4 °C lagern |
| Lösung II | 0,2 N NaOH 1% SDS frisch ansetzen |
| Lösung III | 3 M Kaliumacetat 11,5 ml Essigsäure ad 100 ml dH ₂ O bei 4 °C lagern |
| Maleinsäurepuffer | 0,1 M Maleinsäure 0,15 M NaCl pH 7.5, autoklavieren. |

| | |
|---------------------------|--|
| Neutralisierungslösung | 1,5 M NaCl 1,0 M Tris-HCl pH 7,4 (21 °C) |
| PBS | 10 mM Na-Phosphat, pH 7.2 150 mM NaCl |
| Protein-Probenpuffer (4x) | 250 mM Tris-HCl, pH 6,8 8% (w/v) SDS 40% (v/v) Glycerin 0,004% (w/v) Bromphenolblau reduzierend: 100 mM DTT zufügen |
| Sammelgelpuffer (4x) | 1,5 M Tris-HCl 0,4% (w/v) SDS pH 6,8 |
| SOB-Medium | 2% (w/v) Trypton 0,5% (w/v) Hefe-Extrakt 10 mM NaCl 2,5 mM KCl 10 mM MgCl ₂ Die Mischung (1000 ml) ohne MgCl ₂ auf pH7.0 einstellen und autoklavieren. Anschließend erfolgt die Zugabe von 10 ml steriler MgCl ₂ -Lösung (1M). |
| SOC-Medium | 2% (w/v) Trypton 0,5% (w/v) Hefe-Extrakt 10 mM NaCl 2,5 mM KCl 10 mM MgCl ₂ 20 mM Glucose Zugabe von 20 ml steriler Glucose-Lösung (1 M) zu 1000 ml SOB-Medium. |
| SSC (20x) | 3 M NaCl 0,3 M Tri-Natriumcitrat pH 7,0 |
| TAE-Puffer (50x) | 2 M Tris 5,71% Eisessig 50 mM EDTA pH 8,5 einstellen |
| Tankpuffer (5x) | 0,125 M Tris-HCl 0,96 M Glycin 0,5% (w/v) SDS pH 8,3 |

| | |
|--|--|
| TE-Puffer | 10 mM Tris-HCl, pH 8.0 1 mM EDTA autoklavieren |
| TES-Puffer | 0.2 M Tris-HCl, pH 8.0 0,5 mM EDTA 0,5 M Saccharose |
| Trenngelpuffer (4x) | 1,5 M Tris-Base 0,4% (w/v) SDS pH 8,8 |
| TSS (<i>transformation and storage solution</i>) | 85% (v/v) LB-Medium 10% (w/v) PEG-8000 5% DMSO |
| Waschpuffer | Maleinsäurepuffer mit 0,3% Tween 20 (v/v) |
| 2YT-Medium | 16 g Trypton 10 g Hefe-Extrakt 5 g NaCl ad 1000 ml dH ₂ O, autoklavieren |
| 2YT-Agarplatten | wie 2YT-Medium 15 g Agar ad 1000 ml dH ₂ O, autoklavieren |

2.2 Methoden

Molekularbiologische Methoden

Die molekularbiologischen Methoden wurden, soweit nicht anders beschrieben, nach Sambrook *et al.* (1989) durchgeführt. Verwendete Lösungen und Arbeitsgeräte wurden autoklaviert bzw. mit Ethanol sterilisiert. Durch das Verwenden einer Bunsenbrennerflamme wurden Kontaminationen aus der Luft verhindert. Für alle wässrigen Puffer, Lösungen und Medien wurde entweder einfach deionisiertes Wasser oder hochreines Wasser aus der Filteranlage PURELAB Plus PL5121 02 (USF, Ransbach-Baumbach). Einige Puffer und Lösungen wurden zusätzlich mit einer sterilen Einwegspritze BRAUN Inject (B. Braun Melsungen AG, Melsungen) durch einen sterilen Rotilabo-Spritzenfilter (0,22 µm) der Firma Roth (Karlsruhe) sterilfiltriert.

Die Durchführung dieser Arbeit fand in einem S1-Labor (Sicherheitsstufe 1: Keine Gefährdung von Beschäftigten oder der Umwelt) gemäß des Gentechnikgesetzes statt. Sämtliche Organismen wurden vor der Entsorgung durch Sterilisation abgetötet.

2.2.1 Isolierung genomischer DNA aus Kobragewebe

Bei der Isolierung hochmolekularer genomischer DNA ist es wichtig, die DNA nicht durch Scherkräfte zu fragmentieren. Die benutzten Pipettenspitzen wurden daher durch ein Skalpell abgeschnitten, damit eine größere Öffnung entsteht. Jedes Vermischen oder Pipettieren wurde vorsichtig durchgeführt. Mehrfaches Einfrieren und Auftauen wirkt sich ebenfalls ungünstig aus, so dass ein Aliquot der DNA (Größe >20 kb) bei 4°C gelagert wurde.

Die Isolierung genomischer DNA erfolgte nach einer modifizierten Vorschrift von Gros-Bellard *et al.* (1973). Es wurden ca. 1 g tiefgefrorene Kobraleber in einem vorgekühlten sterilen Mörser mit flüssigem Stickstoff versetzt und pulverisiert. Anschließend wurde das pulverisierte Gewebe in ein 2 ml-Gefäß überführt und in je 1.2 ml Digestion-Puffer pro 100 mg Gewebe suspendiert. Die Probe wurde unter Schütteln für 18 Std. bei 50°C inkubiert. Danach wurde mit einem Äquivolumen Phenol/Chloroform/Isoamylalkohol (25:24:1) extrahiert und für 10 min bei 1700 xg zentrifugiert. Der wässrige Überstand wurde in ein neues Gefäß überführt und die Extraktion wiederholt, bis kein weißes Material an der Interphase mehr sichtbar war. Die wässrige Phase wurde mit einem ½ Volumen 7.5 M Ammoniumacetat und dem doppelten Volumen eiskaltem Ethanol (96%) versetzt, wobei sich ein fadenförmiger DNA-Niederschlag bildete. Die DNA wurde für 2 min bei 1700 xg zentrifugiert. Der Überstand wurde abgesaugt und das Pellet mit 70% Ethanol gewaschen. Nach erneutem Absaugen des Überstands wurde das Pellet an der Luft getrocknet. Die DNA wurde in TE-Puffer (pH 8) resuspendiert und unter vorsichtigem Schütteln bei 65°C über Nacht in Lösung gebracht. Danach wurde mit 1 µg/ml DNase-freier RNase für 1 h bei 37°C inkubiert und die Extraktion mit Phenol/Chloroform/Isoamylalkohol und Ethanol-Fällung wie oben beschrieben wiederholt. Die DNA wurde in einer finalen Konzentration von 1 mg/ml in Wasser aufgenommen. Die Lagerung erfolgte bei 4°C, ein Aliquot wurde bei -20°C gelagert.

2.2.2 Isolierung von Plasmid- oder Cosmid-DNA

Die Plasmid- oder Cosmid-DNA-Isolation erfolgte mittels alkalischer Lyse (Birnboim und Doly 1979).

Minipräparation von Plasmid-DNA

LB-Medium (3 ml) mit 50 µg/ml Ampicillin bzw. Kanamycin wurden mit einer Kolonie des plasmidtragenden Bakterienstammes inokuliert und über Nacht in einem Inkubationsschüttler bei 37°C und 220 rpm inkubiert. Von dieser Kultur wurden 2 ml in ein Eppendorfgefäß überführt und zentrifugiert (14.000 rpm, 2 min, 4°C). Der Überstand wurde abgesaugt, und das Bakterienpellet wurde in 100 µl Lösung I durch Vortexen suspendiert. Danach wurden 200 µl frisch hergestellter Lösung II zugegeben, vorsichtig durch Invertieren gemischt und für 5 min bei RT inkubiert. Durch Zugabe von 150 µl eiskalter Lösung III wurde das Lysat neutralisiert und durch mehrfaches Invertieren vorsichtig gemischt, wobei ein weißer Niederschlag aus Protein/SDS/Membran-Komplexen und chromosomaler DNA ausfällt. Nach einer Inkubation für 10 min auf Eis wurde der Ansatz zentrifugiert (14.000 rpm, 10 min, 4°C) und der klare Überstand in ein neues Eppendorf-Gefäß überführt. Die DNA wurde durch Zugabe von 1 ml eiskaltem Ethanol (96% v/v) gefällt und für 10 min auf Eis gelagert. Nach Pelletierung durch Zentrifugation (14.000 rpm, 10 min, 4°C) wurde der Überstand abgesaugt. Das Pellet wurde mit 500 µl eiskaltem Ethanol (70%) gewaschen. Nach erneuter Zentrifugation wurde der Überstand abgesaugt. Schließlich wurde die DNA für 30 min an der Luft getrocknet und in 20 µl ddH₂O aufgenommen und mit 0,5 µl RNase (10 mg/ml) versetzt.

Alternativ können Proteine durch eine Phenol/Chloroform-Extraktion entfernt werden (siehe Midipräparation).

Midipräparation von Cosmid-DNA

LB-Medium (100 ml) mit Kanamycin (50 µg/ml) wurde mit einer einzelnen Bakterienkolonie angeimpft. Die Kultur wurde in einem Erlenmeyerkolben bei 37°C unter Schütteln (220 rpm) inkubiert. Bei einer optischen Dichte (OD₆₀₀) von 0,6-0,8 wurde Chloramphenicol in einer finalen Konzentration von 170 µg/ml zugegeben. Die Kultur wurde über Nacht inkubiert und danach zentrifugiert (10.000 xg, 10 min, 4°C). Der Überstand wurde abgossen, und das Bakterienpellet wurde in 2 ml eiskalter Lösung I durch Vortexen resuspendiert. Danach wurden 4 ml frisch hergestellte Lösung II zugegeben, vorsichtig unter Invertieren gemischt und für 5 min bei RT inkubiert. Danach wurden 3 ml eiskalte Lösung III zugegeben und durch Invertieren gemischt. Der Ansatz wurde für 10 min auf Eis inkubiert und anschließend zentrifugiert (10.000 xg, 10 min, 4°C). Der Überstand wurde in ein neues Eppendorf-Reaktionsgefäß überführt und mit einem Äquivolumen Phenol : Chloroform : Isoamylalkohol (25:24:1) extrahiert. Nach der

Zentrifugation (10.000 xg, 1 min, 4°C) wurde die obere, wässrige Phase in ein neues Gefäß überführt. Zur Fällung der DNA wurde das doppelte Volumen an eiskaltem Ethanol (96%) zugegeben und zentrifugiert (10.000 xg, 20 min, 4°C). Der Überstand wurde vorsichtig abgesaugt. Das Pellet wurde einmal mit 500 µl eiskaltem Ethanol (70%) gewaschen und erneut zentrifugiert (10.000 xg, 10 min, 4°C). Der Überstand wurde vorsichtig abgesaugt und das DNA-Pellet bei RT an der Luft getrocknet. Anschließend wurde die DNA in 20 µl ddH₂O aufgenommen und mit 0,5 µl RNase (10 mg/ml) versetzt.

DNA-Isolation mittels Kit

Alternativ zur klassischen DNA-Präparation wurde Plasmid-DNA auch mit Hilfe des *E.Z.N.A. PeqLab Mini Kit* (PeqLab, Erlangen) isoliert. Das Kit beruht auf der Methode der alkalischen Lyse, wobei die Reinigung der DNA nicht durch Alkohol-Präzipitation erfolgt, sondern mit Hilfe von Zentrifugationssäulen, die mit einer DNA-bindenden Matrix ausgestattet sind. Nach Adsorption der DNA werden mehrere Waschschrte mit chaotropen Salzlösungen und ethanolischer Lösung durchgeführt. Die Elution von der Matrix erfolgt mit ddH₂O und liefert sehr saubere DNA. Die Durchführung erfolgte nach Herstellerangaben.

2.2.3 Quantifizierung der DNA

Die DNA-Konzentration wurde durch Messung der Absorption bei 260 nm bestimmt. Proteinverunreinigungen wurden durch den Vergleich der Messung bei 280 nm ermittelt. Als Reinheitskriterium für Nukleinsäuren dient der Quotient OD₂₆₀/OD₂₈₀, der zwischen 1.8 und 2.0 liegen sollte. Mischungen aus 50% Proteine/50% DNA ergeben ein OD₂₆₀/OD₂₈₀Verhältnis von ungefähr 1.5.

Als Richtwert für die Konzentrationsbestimmung wurde folgende Formel verwendet:

Doppelsträngige DNA = 1 OD₂₆₀ -E/ml = 50 µg/ml.

2.2.4 Restriktionsverdau

Verwendete Restriktionsendonukleasen gehörten dem Typ II an, deren symmetrische DNA-Erkennungseinheit 6-8 Basen umfaßt. Bei der Spaltung doppelsträngiger DNA entstehen entweder glatte Enden (*blunt ends*) oder zueinander komplementäre Ende (*sticky ends*). Jedes Restriktionsenzym benötigt ein spezielles pH-, Temperatur- und Salz-Optimum. Daher wurde die Reaktion in einem auf das jeweilige Enzym abgestimmten Puffer nach

Angaben des Herstellers durchgeführt. Das Reaktionsvolumen betrug typischerweise 10-20 μ l.

2.2.5 Agarose-Gelelektrophorese

Die Agarose-Gelelektrophorese diente zur Auftrennung, Identifizierung und Reinigung von DNA-Molekülen. Die Methode wurde sowohl für analytische als auch für präparative Zwecke angewendet. Agarose ist ein Polysaccharid aus roten Meeresalgen. Zur Herstellung eines Gels wurde die entsprechende Menge Agarose in TAE-Puffer (1%) in der Mikrowelle erhitzt, bis sie vollständig gelöst war. Es wurden ausschließlich Gele mit einem Agarosegehalt von 0,8-2% verwendet. Nach Abkühlen der Lösung auf etwa 60°C wurde Ethidiumbromid in einer Endkonzentration von 1 μ g/ml zugegeben und das Gel in einen mit einem Kamm bestückten Schlitten gegossen.

Durch das Erkalten erhärtete Gel wurde danach in eine mit TAE-Puffer gefüllte Laufkammer (IBI, New Haven, USA) überführt. Die DNA-Proben wurden in 1x DNA-Auftragspuffer in die Probenaschen pipettiert und bei einer Spannung von 100-125 V elektrophoretisch getrennt.

Anhand des verwendeten Farbmärkers (Bromphenolblau, Xylencyanol) zur Markierung der Elektrophoresefront kann der Verlauf der Elektrophorese kontrolliert und die Auftrennung nach Bedarf beendet werden. Die DNA-Banden konnten anschließend durch Fluoreszenz des DNA-Interkalators Ethidiumbromid auf einem UV-Transilluminator bei 302 nm sichtbar gemacht und mit einem Video-Image-System dokumentiert werden. Zur Analyse diente ein mitlaufender DNA-Standard. Üblicherweise kamen hier MspI-geschnittener pUC19-Vektor und Eco130I-geschnittene λ -Phagen-DNA (beides MBI Fermentas, St. Leon-Rot) zum Einsatz.

2.2.6 DNA-Extraktion aus Agarose-Gelen

Zur Isolation von DNA aus Agarosegelen wurde das *QiaexII DNA Extraction Kit* (Qiagen, Hilden) entsprechend der Herstellerangaben verwendet. Die Reinigung basiert auf einem modifizierten Protokoll von Vogelstein und Gillespie (1979). Die Agarose wird in einem speziellen Puffer aufgelöst, und in Gegenwart einer hohen Salz-Konzentration bindet die DNA an speziell gefertigte Silica-Partikel. Die Silica-Partikel werden danach abzentrifugiert, mehrmals gewaschen und schließlich mit reinem Wasser oder TE-Puffer in hoher Reinheit eluiert.

2.2.7 Dephosphorylierung und Ligation von DNA-Fragmenten

Freie 5'-Phosphat- und 3'-Hydroxygruppen doppelsträngiger DNA konnten mit DNA-Ligase kovalent verknüpft werden. Dazu wurde die T4-DNA-Ligase (MBI Fermentas) aus dem *E.coli*-Phagen T4 verwendet, da dieses Enzym sowohl die Verknüpfung kohäsiver als auch glatter Enden katalysiert. Die T4-Ligase benötigt ATP zur Katalyse. Die Ligation erfolgte bei 16°C über Nacht. Um ein DNA-Fragment in einen Vektor zu klonieren, wurde der linearisierte Vektor mit CIAP (*calf intestinal alkaline phosphatase*, New England Biolabs) dephosphoryliert, um eine intramolekulare Reaktion, d.h. eine Recyclisierung, zu verhindern.

Nach der Dephosphorylierung wurde die CIAP entweder durch Hitze inaktiviert (65°C, 10 min) oder durch eine Phenol/Chloroform-Extraktion entfernt. Alternativ konnte die DNA durch QIAEXII aufgereinigt werden.

Anschließend wurde das gereinigte Insert und der dephosphorylierte Vektor in einem stöchiometrischen Verhältnis von 3 :1 eingesetzt und in entsprechendem Ligationspuffer in einem Volumen von 20 µl ligiert.

2.2.8 Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion (*polymerase chain reaction*, PCR) ermöglicht die spezifische *in vitro* Vervielfältigung von DNA-Abschnitten kleinster Mengen unter Verwendung von Oligonukleotiden bekannter Sequenz (Saiki *et al.* 1986, 1988). Die Oligonukleotide werden so gewählt, dass sie das zu amplifizierende DNA-Segment flankieren.

Das Prinzip der PCR beruht auf drei Reaktionsschritten: die zu amplifizierende DNA wird zunächst denaturiert. Bei gemäßigten Temperaturen binden die Oligonukleotide spezifisch an den ihnen komplementären DNA-Strang (Annealing), und es erfolgt die oligonukleotid-abhängige Verlängerung der Komplementärstränge durch hitzestabile DNA-Polymerasen (Elongation) in Gegenwart von freien dNTPs. Dieser Prozeß wird zyklisch wiederholt, was in einer exponentiellen Amplifikation der gewünschten DNA-Fragmente resultiert.

Die PCR-Experimente wurden mit drei Arten von thermostabilen DNA-Polymerasen durchgeführt. Als Standard-Polymerase hat sich insbesondere die thermostabile Polymerase aus *Thermus aquaticus* (Taq) bewährt, die ihr Temperatur-Optimum bei 72°C besitzt. Sie besitzt eine hohe 5'-3'-Polymeraseaktivität, aber keine 3'-5'-Exonukleaseaktivität.

Bei einer Weiterentwicklung wurden Taq-Polymerasen mit einem inhibierenden Antikörper (Fast Start Taq-DNA-Polymerase) in sogenannten Hot Start-PCRs eingesetzt (Sharkey *et al.* 1994). Dazu wird dem PCR-Programm ein 4 minütiger Denaturierungsschritt bei 95°C vorangestellt, um eine Denaturierung des inhibierenden Antikörpers und eine Freisetzung der Polymerase zu bewirken (D'Aquila *et al.* 1991). Hot Start-Polymerasen bieten den Vorteil, dass der Anteil der entstehenden Nebenprodukte meist geringer als bei Verwendung von Taq ist. Taq-Polymerasen und die modifizierten Hot Start-Taq-Polymerasen fügen der amplifizierten DNA am 3'-Ende meist ein dATP an, sodass ein A-Überhang gebildet wird. Für die Amplifikation langer und GC-reicher Sequenzen wurden Polymerase-Mischungen entwickelt, die über Korrektur-Aktivität verfügen (Proofreading-Polymerasen). Proofreading-Polymerasen werden eingesetzt, um die Lesegenauigkeit zu erhöhen und damit das Risiko von Mutationen zu minimieren. Die Mischungen enthalten DNA-Polymerasen aus *Thermus aquaticus* (Taq) und *Thermococcus gorgonarius* (Tgo) oder *Pyrococcus woesei* (Pwo) im Verhältnis 11 (Taq) und 1 (Pwo oder Tgo). Die Pwo- und Tgo-Polymerasen zeichnen sich durch eine 3'-5'-Exonuklease-Aktivität aus, die im Falle einer Fehlpaarung des entstehenden DNA-Stranges zu einer Korrektur des falsch eingebauten Nukleotids führt. Die Verwendung solcher Proofreading-Polymerasen wurde zur Amplifikation von bis zu 35 kb genutzt (Barnes 1994, Cheng *et al.* 1994).

Auswahl von Oligonukleotiden

Für den Erfolg der PCR auf genomischer bzw. Cosmid-DNA war es notwendig, spezifische Polymerisationsstarter (Primer) einzusetzen, die 5' und 3' zu einem Intron liegen. Ausgehend von der CVF cDNA-Sequenz wurden Exon-spezifische Primer so gewählt, dass sie nicht im Bereich einer Exon-/Introngrenze liegen. Die Primer sollten nach Möglichkeit nur ein Exon abdecken, da einerseits die Amplifikation mit zunehmender Distanz der Primer immer schwieriger werden würde, und andererseits mit zunehmender Länge der amplifizierten DNA deren Grössenbestimmung immer ungenauer werden würde.

Vor der Primersynthese wurde die Oligonukleotidsequenz durch das Programm *Amplify* (Version 1.2, Universität Wisconsin, Madison, WI, USA) getestet.

2.2.9 Die ARMS-Methode

Die ARMS-Methode (*amplification refractory mutational system*) ist eine PCR-Anwendung, bei der die Primer so gewählt sind, dass sie zwischen Templates differenzieren können, die sich in einem einzigen Nukleotid unterscheiden (Newton *et al.*

1989). ARMS beruht auf dem Prinzip, dass die Primer am 3'-Ende mit der Base enden, um die sich die beiden Sequenzen unterscheiden. Da der *Taq* DNA-Polymerase eine 3'-5'-Exonuklease-Aktivität fehlt, führt eine Fehlpaarung zwischen dem 3'-Ende eines PCR-Primers und dem Template zu keiner bzw. nur zu einer sehr eingeschränkten Amplifikation. Die Spezifität der Methode kann dadurch erhöht werden, dass die Annealingtemperatur während der PCR hoch gewählt wird, dass darüber hinaus zwei ARMS-Primer verwendet werden, und dass zusätzliche Fehlpaarungen der Primersequenz zu den homologen Templates genutzt werden, welche zu einer verminderten Hybridisierungseffizienz führen. Die Methode wurde ursprünglich zum Nachweis von bekannten Punktmutationen entwickelt und ist auch auf andere Anwendungen wie die gezielte Amplifikation bei mehreren homologen Gensequenzen übertragbar.

2.2.10 Klonierung von PCR-Produkten

PCR-Produkte wurden nach der präparativen Aufreinigung über ein Agarose-Gel mit einem Skalpell ausgeschnitten, mit dem *Qiaex II Gel Extraction Kit* (Qiagen) isoliert und anschliessend nach zwei Methoden in entsprechende Vektoren kloniert. Bei der blunt-end Methode wurden singuläre Nukleotid-Überhänge am 3'-Ende des PCR-Produkts mit Klenow-Fragment verdaut und anschliessend das PCR-Produkt mit den glatten Enden mit Hilfe des *Sure Clone Ligation Kit* (Pharmacia) in pUC18 kloniert.

Ein anderer Ansatz beruht auf der Entdeckung, dass die *Taq*-Polymerase meist Amplifikate mit einem unspezifischen Überhang von einer Base erzeugt (Clark 1988). Diese Base ist in der Mehrzahl der Fälle ein Adenosin, allerdings nicht immer (Hu 1993). Durch die Verwendung eines Vektors mit T-Überhang (Marchuck *et al.* 1991) lassen sich die PCR-Produkte leichter klonieren als in Vektoren mit glatten Enden. Diese Methode wird als TA-Klonierung bezeichnet. Eine Variante davon ist das *TOPO TA Cloning Kit* (Invitrogen), wobei die Ligation durch die Topoisomerase I erfolgt, die kovalent mit dem Vektor verbunden ist (Shuman 1994).

TOPO TA Cloning Kit

Bei der T/A-Klonierung macht man sich zunutze, dass PCR-Produkte mit einem 3'-A-Überhang effizient in einen Vektor mit einem 3'-T-Überhang ligiert werden können. Eine Adenylierung des PCR-Produkts durch die *Taq* DNA-Polymerase nutzt die Template-unabhängige terminale Transferase-Aktivität von DNA-Polymerasen aus. Die gereinigten PCR-Produkte wurden zunächst wie nachfolgend beschrieben adenyliert:

7 µl PCR-Produkt
 1 µl Puffer für AGS Gold DNA-Polymerase (10x, 1.75 mM MgCl₂)
 1 µl dATP (10 mM)
1 µl AGS Gold DNA-Polymerase (1u)
 10 µl

Der Ansatz wurde in einem Thermocycler für 10 min bei 72°C inkubiert.

Die Ligation in einen Vektor mit 3'-T-Überhang erfolgte mit Hilfe des *TOPO-TA Cloning Kits* (Invitrogen). PCR-Produkte mit einer Länge bis zu 3 kb wurden in den Vektor pCR2.1-TOPO kloniert, für größere PCR-Produkte stand der Vektor pCR-XL-TOPO zur Verfügung.

Der Ligationsansatz hatte folgende Zusammensetzung:

4 µl adenyliertes PCR-Produkt
 1 µl Salt Solution
1 µl Vektor
 6 µl

Für die Ligation durch die Topoisomerase I wurde der Ansatz für 5-30 min bei RT inkubiert und danach bis zur Transformation auf Eis gestellt.

2.2.11 Herstellung kompetenter Zellen

Chemisch kompetente Zellen

3 ml LB-Medium wurden mit einer Kolonie des entsprechenden *E. coli*-Stammes (z.B. DH5α) angeimpft und über Nacht bei 37°C und 220 rpm inkubiert. 1 ml der stationär gewachsenen Kultur wurden in 100 ml LB-Medium in einem Erlenmeyerkolben überführt. Die Zellen wurden dann 2,5-3 Stunden wachsen lassen, bis sie eine OD₆₀₀ von 0,5 erreicht hatten. Die Zellen wurden für 20 min im Eisbad gekühlt. Alle folgenden Schritte erfolgten im Kühlraum bei 4°C mit vorgekühlten Geräten. Die Zellen-Suspension wurde in 50 ml-Gefäßen zentrifugiert (4000 rpm, 10 min, 0°C), der Überstand dekantiert und jedes Pellett in 5 ml eiskaltem TSS-Puffer aufgenommen. Die kompetenten Zellen konnten entweder sofort verwendet oder in Aliquots von je 100 µl in flüssigem Stickstoff schockgefroren und bei -80°C gelagert werden.

Elektrokompetente Zellen

Für die Herstellung salzfreier, elektrokompetenter Zellen beimpft und inkubiert man 500 ml LB-Medium mit dem gewünschten Bakterienstamm und pelletiert die Bakterien, sobald die

Kultur eine OD_{600} von 0.5 aufweist. Das Pellet wird in 100 ml kaltem H_2O resuspendiert. Pelletieren und Waschen der Bakterien wiederholt man drei weitere Male bei $4^\circ C$ und resuspendiert dann das Bakterienpellet in 2-5 ml einer kalten 10% (v/v) Glycerinlösung. Es werden 100 μl Aliquots gebildet, die entweder für die anschließende Elektroporation auf Eis aufbewahrt werden oder mit flüssigem Stickstoff schockgefroren und bei $-80^\circ C$ gelagert werden.

2.2.12 Transformation

Hitzeshock-Methode

Die Hitzeshock-Methode zur Transformation kompetenter *E. coli* nach Chung *et al.* (1989) beruht auf der Beobachtung, dass Bakterien, die mit einer eiskalten Lösung von $CaCl_2$ versetzt und anschließend kurz erhitzt werden, mit Bakteriophagen λ transfiziert werden können (Mandel und Higa 1970). Es wurden 3-5 μl des Ligationsansatzes mit 100 μl kompetenten *E. coli*-Zellen (Stamm DH5 α bzw. TOP10) für 30 min auf Eis inkubiert. Anschliessend erfolgte der Hitzeshock für 60 s bei $42^\circ C$. Nach einer Inkubation für 2 min auf Eis wurden 250 μl SOC-Medium zugegeben und für 1 h bei $37^\circ C$ und 220 rpm inkubiert.

Schließlich wurden je 75 μl und 150 μl des Transformationsansatzes auf LB-Agar-Platten mit dem entsprechenden Antibiotikum ausplattiert und bei $37^\circ C$ über Nacht inkubiert.

Elektroporation

Im Vergleich zur chemischen Transformation liefert die Elektroporation eine höhere Ausbeute an Kolonien, und zudem ist die Präparation der Bakterien einfach. Ein weiterer positiver Aspekt ist, dass die Bakterien leichter mit grösseren Plasmiden transformiert werden können (Dower *et al.* 1988).

Für die Transformation wurden kompetente Bakterien auf Eis aufgetaut, 100 μl davon mit 1 μl des Ligationsansatzes gemischt und für 1 min auf Eis gekühlt. Anschließend wurde der gesamte Ansatz in eine vorgekühlte Elektroporationsküvette mit 2 mm Elektrodenabstand pipettiert. Die Elektroporation wurde in einem Easyject-Gerät (Equibio) durchgeführt. Die Einstellung betrug 2500 V/25 μF /200-400 Ω . Direkt nach dem Puls wurden 900 μl SOC-Medium, vorgewärmt auf $37^\circ C$, zupipettiert und der Ansatz für 1 h bei $37^\circ C$ und 220 rpm inkubiert. Schließlich wurden je 75 μl und 150 μl des Transformationsansatzes auf LB-Agar-Platten mit dem entsprechenden Antibiotikum ausplattiert und die Platten bei $37^\circ C$ über Nacht inkubiert.

2.2.13 Analyse von positiven Klonen

Blau-Weiß-Screening

Die *Multiple Cloning Site* der Klonierungsvektoren pUC18, pCR2.1-TOPO und pCR-TOPO-XL ist Teil des lacZ-Gens von *E. coli*, welches für die C-terminale Domäne von β -Galactosidase codiert. Die Induktion mit IPTG führt zur Expression der C-terminalen Domäne von β -Galactosidase. Die Transformation von Plasmiden in *E. coli*-Stämme, welche die N-terminale Domäne von β -Galactosidase exprimieren, führt zur Formation des aktiven Enzyms (α -Komplementierung). Das Wachstum auf Agarplatten, die das β -Galactosidase-Substrat X-Gal enthalten, führt zur Spaltung des Substrats durch das aktive Enzym, wobei ein blauer Farbstoff entsteht. Die Insertion von Fremd-DNA in die multiple Klonierungsstelle (MCS) des Vektors verhindert die Expression der aktiven C-terminalen Domäne und damit die Bildung des blauen Farbstoffs. Daher enthalten blaue Kolonien Plasmid-DNA ohne Insert, wohingegen weisse Kolonien rekombinante Plasmide enthalten und für weitere Analysen gepickt wurden.

Von den erhaltenen Klonen wurden vier bis sechs gepickt und zum Animpfen einer LB-Flüssigkultur verwendet. Die DNA wurde isoliert, und positive Klone wurden durch EcoRI-Restriktionsanalyse bzw. durch PCR identifiziert.

PCR-Screening

Alternativ zur DNA-Präparation von Übernacht-Kulturen wurde eine Zahnstocher-Spitze mit Bakterien in 50 μ l sterilem Wasser durch Vortexen suspendiert, für 5 min gekocht und das Zell-Debris anschliessend für 1 min in einer Eppendorf-Zentrifuge sedimentiert. Vom Überstand wurden dann 1 μ l als Template für eine PCR eingesetzt.

2.2.14 Wachstum und Lagerung von Bakterienklonen

Eine flüssige *E. coli* Bakterienkultur kann bei 4°C bis zu drei Tagen aufbewahrt werden. LB-Agar-Platten mit Bakterienkulturen wurden mit Parafilm verschlossen und bei 4°C aufbewahrt. Die Kultur bleibt dann bis zu 4 Wochen lebensfähig. Für eine Langzeitlagerung wurden 1 ml einer Übernacht-Kultur des aufzubewahrenden Klons mit 250 μ l 75%igem Glycerol in einem Kryogefäß (Greiner) gemischt, in flüssigem Stickstoff schockgefroren und bei -80°C aufbewahrt.

2.2.15 DNA-Sequenzierung

Die DNA-Sequenzierung erfolgte nach dem Kettenabbruch- oder Didesoxynukleotid-verfahren (Sanger 1977). Dabei wurde entweder mit einer T7-DNA-Polymerase und dem *AutoRead Sequencing Kit* (Amersham-Pharmacia, Freiburg) oder mit einer thermostabilen DNA-Polymerase und dem *BigDye Terminator Ready Reaction Kit* (Applied Biosystems, Perkin Elmer Corporation, Weiterstadt) gearbeitet.

DNA-Sequenzierung mit T7-DNA-Polymerase

Die Sequenzierung einzelsträngiger DNA mit der T7-DNA-Polymerase wurde nach den Protokollen *Annealing of Primer to Single-Stranded Template* und *Sequencing Reactions* des *AutoRead Sequencing Kit* (Amersham Pharmacia Biotech, Freiburg) durchgeführt. Die DNA-Sequenzen wurden unter Verwendung des ALF Express-Sequenziergerätes (Amersham Pharmacia Biotech, Freiburg) ermittelt. Durch die Auftrennung im Polyacrylamid-Sequenziergel nach ihrer Größe konnten die fluoreszenzmarkierten DNA-Fragmente mit einem Laser erfaßt und in ihrer Reihenfolge automatisch in die DNA-Sequenz umgeschrieben werden. Die weitere Analyse und Bearbeitung erfolgte am Computer mittels der ALF Manager 3.01 Software.

Sequenzierung nach der Dye-Terminator-Methode

Die Sequenzierung nach der Dye-Terminator-Methode (Cycle Sequencing) beruht auf dem Prinzip der Didesoxymethode nach Sanger. Die zu sequenzierende Probe wird in Gegenwart von fluoreszenzmarkierten 2',3'-Didesoxynukleotiden durch eine PCR mit einer hitzestabilen DNA-Polymerase amplifiziert. Der Vorteil dieser Methode besteht darin, dass die Primer frei wählbar sind. Jedes der vier Didesoxynukleotiden ist mit einem bestimmten Fluoreszenzfarbstoff markiert und liegt neben den Desoxynukleotiden in einer Konzentration vor, dass es statistisch nur einmal pro wachsender Kette eingebaut wird. Durch den Einbau eines Didesoxynukleotids wird die Kettenverlängerung abgebrochen, und das PCR-Produkt erhält gleichzeitig eine spezifische Markierung. Die Sequenzierreaktion wird anschließend gelelektrophoretisch aufgetrennt.

Für die Sequenzierung wurde das *BigDye Terminator Ready Reaction Kit* (Applied Biosystems, Perkin Elmer Corporation, Weiterstadt) verwendet und wie folgt verfahren: Für eine Reaktion wurden 500 ng DNA mit 15 pmol Primer vermischt und auf 12 µl mit ddH₂O aufgefüllt. Dazu wurden 6 µl Half Term-Puffer und 2 µl BigDye Terminator Ready

Reaction Mix zugesetzt und der Ansatz im Thermocycler nach folgendem Programm sequenziert: Denaturierung 10 s bei 96°C, Hybridisierung 10 s bei 50°C, Elongation 4 min bei 60°C. Es wurden 27 Zyklen durchgeführt. Für eine DNA-Fällung wurden 50 µl eiskaltes Ethanol (96%) zugesetzt, zentrifugiert (15800 xg, 30 min, 4°C) und der Überstand verworfen. Das Pellet wurde mit 50 µl eiskaltem Ethanol (70%) gewaschen und zentrifugiert (15800 xg, 10 min, 4°C). Der Überstand wurde verworfen, und das Pellet wurde 30-60 min an der Luft getrocknet.

Die Analyse der Sequenzierreaktion wurde am Institut für Zellbiochemie (Universitätskrankenhaus Eppendorf, Hamburg) vom Arbeitskreis Richter auf einem ABI Sequenator 377 durchgeführt. Die Auswertung der Daten erfolgte mit Hilfe der ABI PRISM 3.3 Software.

2.2.16 Screening mit dem Digoxigenin-System

Das Digoxigenin-System ist ein nicht-radioaktives Markierungs- und Detektionssystem mit einer vergleichbaren Sensitivität zur radioaktiven Detektion. Es basiert auf dem Steroid-Hapten Digoxigenin (DIG) zur Markierung von DNA, RNA oder Oligonukleotiden. Digoxigenin ist ein chemisch abgeleitetes Aglykon von Digoxin aus *Digitalis purpurea*. Das Nukleotid-Derivat DIG-dUTP wird durch eine Kopplung von dUTP und Digoxigenin über einen Spacer hergestellt. Für die Herstellung einer Sonde kann die DNA durch Inkorporation von DIG-dUTP markiert werden (Abb. 2.1). Die nicht-radioaktiv markierten Sonden werden zur Hybridisierung eingesetzt und anschließend mit einer Farb- oder Chemilumineszenz-Reaktion nachgewiesen. Für die Markierung von DNA ist Digoxigenin über eine Alkali-labile Esterbindung an dUTP gekoppelt. Diese Alkali-labile Form des DIG-11-dUTP ermöglicht ein schnelles und effektives Entfernen des Hybridisierungssignals und erlaubt eine zweite Hybridisierung mit einer anderen DNA-Sonde.

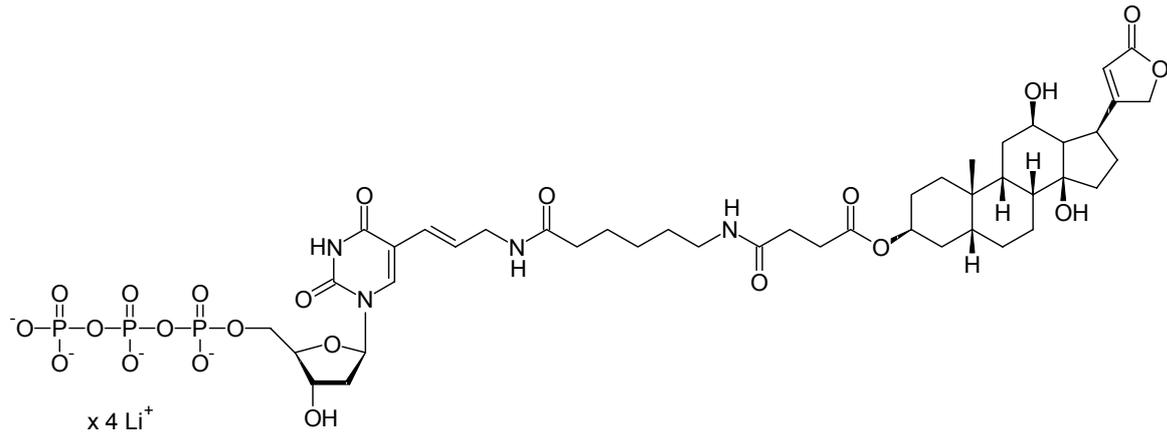


Abb. 2.1: Struktur von Digoxigenin-11-dUTP

Herstellung von DNA-Sonden

Die Herstellung von DIG-markierten DNA-Sonden erfolgte durch Inkorporation von DIG-11-dUTP in der PCR. Auf diese Weise wurden Sonden definierter Länge erhalten, bei denen statistisch jedes 20. Nukleotid DIG-11-dUTP ist. Das optimale Verhältnis von DIG-11-dUTP : dTTP ist 1:2. Für die Sondenherstellung wurde daher eine dNTP-Spezialmischung der Zusammensetzung 2 mM dATP, 2 mM dCTP, 2 mM dGTP, 1.3 mM dTTP und 0.7 mM DIG-11-dUTP eingesetzt.

Die PCR-Produkte wurden mittels Agarose-Gelelektrophorese überprüft, und die markierte DNA-Sonde wurde mittels *QIAEX II Gel Extraction Kit* (Qiagen) gereinigt. Die Sondenkonzentration und die Effizienz der Markierung wurden photometrisch oder durch einen Spot-Test ermittelt.

Screening der Cosmid-Genbibliothek

Die generellen Methoden zur Charakterisierung der Cosmid-Bibliothek wurden, soweit nicht anders beschrieben, nach Birren *et al.* (1997) durchgeführt. Zum Ausplattieren der Genbibliothek wurde mit einer Pipettenspitze aus einem tiefgefrorenen Aliquot ca. 2 µl Zellsuspension abgekratzt und diese in LB-Medium so verdünnt, dass beim Ausplattieren auf einer grossen Agarplatte (24 x 24 cm) eine Koloniedichte von 2000 Klonen pro Platte erhalten wurde.

Kolonie-Lifting

Nach einer Inkubationszeit von 30 Std. wurden die mit Kolonien bewachsenen Platten zunächst für mindestens 30 min bei 4°C aufbewahrt. Danach wurde der erste Membranabdruck hergestellt. Mit einem weichen Bleistift wurde die Nummer der entsprechenden Platte auf eine Seite der Membran geschrieben. Die Membran wurde mit der beschrifteten Seite nach oben luftblasenfrei auf die Platte gelegt. Zur Orientierung nach der Detektion wurde die Membran mit einer dünnen Nadel durch Einstiche an asymmetrischen Stellen markiert. Die Positionen der Nadeleinstiche wurden dann mit einem wasserfesten Markierungsstift auf dem Boden der Schale markiert. Danach wurde die Membran mit einer Flachpinzette von der Platte abgezogen und mit der Koloniseite nach oben luftblasenfrei auf eine neue LB-Kanamycin-Platte gelegt.

Zum Auftüpfeln einer Positiv-Kontrolle wurde am Rand der Membran eine freie Stelle mit Bleistift markiert. Als Positiv-Kontrolle wurde eine kleine Menge einer geeigneten Bakterienkultur auf LB-Agar-Platte mit einem Zahnstocher gepickt und auf die markierte Stelle aufgetüpfelt. Die Platte mit der Membran wurde gemeinsam mit der Originalplatte bei 37°C über Nacht inkubiert. Die Inkubation diente zur Regeneration der Klone.

Nach der Über-Nacht-Inkubation wurde ein zweiter Membranabdruck entweder von der Original-Platte oder dem ersten Membranabdruck hergestellt. Dieser zweite Membranabdruck wurde ebenfalls mit der Koloniseite nach oben luftblasenfrei auf eine neue LB-Kanamycin-Platte gelegt und bei 37°C über Nacht inkubiert.

Danach wurden zur Amplifikation der Cosmid-DNA die zwei Membranabdrücke jeweils mit der Koloniseite nach oben luftblasenfrei auf Chloramphenicol-Platten gelegt und bei 37°C über Nacht inkubiert.

Nach der Über-Nacht-Inkubation auf Chloramphenicol-Platten wurden die Membranen auf Filterpapier gelegt. Am Rand der Membran wurden freie Stellen mit Bleistift markiert und dort jeweils 1 µl einer Positiv- bzw. Negativ-Kontroll-DNA aufgetüpfelt. Nach dem Auftüpfeln wurden die Membranen kurz bei Raumtemperatur getrocknet.

Anschließend wurde die Membran mit der Koloniseite nach oben für 15 min auf mit Denaturierungslösung getränktem Filterpapier (drei Schichten) inkubiert. Danach wurde die Membran kurz auf ein trockenes Filterpapier gelegt.

Zur Neutralisierung wurde die Membran für 15 min auf mit Neutralisierungslösung getränktem Filterpapier (drei Schichten) inkubiert. Danach wurde die Membran kurz auf ein trockenes Filterpapier gelegt.

Im Anschluß wurde die Membran für 10 min auf mit 2x SSC getränktem Filterpapier (drei Schichten) äquilibriert und danach kurz auf ein trockenes Filterpapier gelegt.

Zur Fixierung der DNA an die Membran wurde die Membran mit der Kolonienseite nach oben in den UV-Crosslinker (BIO-LINK[®] BLX, AGS, Heidelberg) gelegt. Durch UV-Strahlung mit einer Energie von 0,07 J/cm² wurde die DNA an die Membran immobilisiert.

Proteine wurden durch Inkubation mit Proteinase K verdaut. Dazu wurden 100 µl Proteinase K (20 mg/ml) im Verhältnis 1:20 mit 2x SSC verdünnt und die Membran für 60 min bei 37°C darin inkubiert. Anschließend wurde die Membran zwischen zwei Lagen Blotpapier in 2x SSC eingetaucht. Um Zellrückstände vollständig zu entfernen, wurde mit sauberen Handschuhen sanfter Druck ausgeübt und die Zellrückstände mit dem Blotpapier entfernt. Schließlich wurde die Membran sorgfältig mit 2x SSC gespült.

Hybridisierung mit DIG-DNA-Sonden

Die Membran wurde mit einem Gaze-Netz, welches ein Zusammenkleben der Membran verhindert, zusammengerollt und in eine Hybridisierungsflasche geschoben. Es wurden 10 ml Prähybridisierungslösung (DIG Easy Hyb oder Church-Puffer) zugegeben und in einem Hybridisierungssofen (HYBAID, Heidelberg) bei 42°C für 60 min inkubiert.

Die Sonde (Endkonzentration 10-15 ng/ml) wurde in 100 µl Hybridisierungslösung (DIG Easy Hyb oder Church buffer) aufgenommen und für 2 min bei 95°C denaturiert. Anschließend wurde das Eppendorfgefäß sofort auf Eis gestellt, vor der Hybridisierung auf 42°C erwärmt und die denaturierte Sonde direkt in die Prähybridisierungslösung pipettiert.

Bei Mehrfachverwendung der Sonde wurde die Hybridisierungslösung für 10 min bei 68°C denaturiert und anschließend auf 42°C temperiert. Die Prähybridisierungslösung wurde in diesem Fall aus der Flasche gegossen und durch die Hybridisierungslösung ersetzt.

Die Hybridisierung erfolgte über Nacht bei 42°C in einem Hybridisierungssofen.

Nach der Hybridisierung wurde die Membran in folgenden Schritten gewaschen:

- (1) zweimal mit 20 ml 2x SSC, 0,1% SDS, für 5 min bei RT,
- (2) zweimal mit 20 ml 0,5x SSC, 0,1% SDS, für 15 min bei 68°C,
oder zweimal mit 20 ml 0,1x SSC, 0,1% SDS, für 15 min bei 68°C (für sehr stringentes Waschen). Je niedriger die Salzkonzentration des Waschpuffers, desto stringenter ist der Waschschrift.

Immunoprinting

Die Membran wurde einmal für 5 min mit 10 ml Waschpuffer gewaschen und mit 10 ml Blockierungslösung für 30 min bei RT blockiert. Das Gefäß mit Anti-DIG-AP-Konjugat wurde zentrifugiert (14.000 rpm, 4°C, 2 min), um Antikörper-Aggregate zu sedimentieren. Danach wurde 1 µl Anti-DIG-AP (Fab-Fragmente) zu 10 ml Blockierungslösung pipettiert (Verdünnung 1:10.000) und die Membran darin für 60 min bei RT inkubiert.

Detektion

Nach der Inkubation mit dem Antikörper-Konjugat wurde die Membran zweimal für 15 min mit 10 ml Waschpuffer bei RT gewaschen. Dann wurde die Membran in 10 ml Detektionspuffer für 5 min äquilibriert. Die Membran wurde dann aus der Flasche genommen und auf eine Plastikfolie gelegt.

Die Detektion mit CSPD[®] ist eine Chemilumineszenz-Reaktion, die durch Alkalische Phosphatase katalysiert wird. Das erzeugte Lichtsignal wird auf einem Röntgenfilm dokumentiert. Dazu wurden 5-10 Tropfen CSPD[®]-Substratlösung auf die Membran gegeben und gleichmäßig verteilt. Die Membran wurde mit der Plastikfolie umhüllt und bei RT für 5 min und danach bei 37°C für weitere 15 min inkubiert. Danach wurde die überschüssige Flüssigkeit von der Membran entfernt. Die Plastikfolie mit der Membran wurde in eine Filmkassette mit Röntgenfilm gelegt und mit Klebestreifen fixiert. Die Position der Membran sowie die Markierungen auf der Membran wurden mit einem wasserfesten Stift auf den Film gezeichnet. Die Exposition erfolgte für 1 bis 60 min bei RT. Danach wurde der Film für 2-5 min in Entwickler-Lösung gelegt, kurz in Wasser gespült und dann für 5-10 min in Fixier-Lösung gelegt. Schließlich wurde der Film mit Wasser gewaschen und getrocknet.

Für die Auswertung wurde der Film anhand der Markierungen mit der Original-Platte verglichen. Nach Zuordnung der Signale zu den entsprechenden Klonen wurden die Klone gepickt und zum Animpfen von Übernachtskulturen eingesetzt.

Stripping

Derselbe Membranabdruck konnte nacheinander mit verschiedenen Sonden hybridisiert werden, wenn die Digoxigenin-Gruppe der zuerst gebundenen Sonde durch Behandlung mit Alkali entfernt wurde. Dazu wurde die Membran nach der Detektion mit CSPD gründlich mit Wasser gewaschen und anschließend zweimal mit 10 ml 0,2 M NaOH, 0,1% SDS (w/v) für 15 min bei 37°C inkubiert. Danach wurde die Membran mit 2x SSC gewaschen und direkt für eine erneute Hybridisierung eingesetzt.

2.2.17 Southern-Blot

Auf der Grundlage des Verfahrens von E.M. Southern (1975) wurde das einseitig gerichtete Kapillarblotting von DNA auf eine Membran durchgeführt, um sie dort mit einer DNA-Sonde zu detektieren. Die zu untersuchende DNA-Probe wurde zunächst mit geeigneten Restriktionsenzymen verdaut und durch Agarose-Gelelektrophorese analysiert. Danach wurde das Agarose-Gelstück für 30 min in 0,4 M NaOH inkubiert, um die DNA zu denaturieren. Die Nylonmembran wurde ebenfalls für 15 min in 0,4 M NaOH äquilibriert. Eine Plastikschaale wurde zur Hälfte mit 0,4 M NaOH gefüllt. Ein Haushaltsschwamm wurde mit der grünen Seite nach unten in die Schale gelegt und mit 0,4 M NaOH getränkt. Mit Natronlauge getränktes Blotpapier wurde in vier Lagen auf den Schwamm gelegt. Dann wurde das Agarosegel mit der Unterseite nach oben auf das Blotpapier gelegt. Die Nylonmembran wurde luftblasenfrei auf dem Gel positioniert und mit einem Bleistift markiert. Mit Natronlauge getränktes Blotpapier wurde in vier Lagen auf die Membran gelegt. Um Kurzschlüsse zu vermeiden, wurde Frischhaltefolie an den Kanten um das Gel und die Membran gelegt. Auf das Blotpapier wurde ein Stapel Papierhandtücher gelegt und mit einem Gewicht beschwert. Es ist möglich, mehrere Membranabdrücke desselben Gels durchzuführen. Für den ersten Blot wurde 15 min geblottet, das Blotting für den zweiten Abdruck erfolgte über Nacht.

Nach dem Blotting wurde die Membran für 15 min mit 2x SSC bei RT gewaschen. Anschließend wurde die DNA durch UV-Crosslinking an die Membran fixiert. Die nachfolgenden Schritte Hybridisierung, Immunprinting und Detektion wurden wie oben beschrieben durchgeführt.

2.2.18 Genome Walking

Die Genome Walking-Methode ist dazu geeignet, unbekannte genomische DNA-Sequenzen in Nachbarschaft zu einer bekannten Sequenz wie z.B. einer cDNA mittels PCR aufzufinden (Siebert *et al.* 1995). Bei dem *GenomeWalkerTM Universal Kit* (Clontech) wird die zu untersuchende genomische DNA in vier Ansätzen mit jeweils einem der Restriktionsenzyme (blunt end) *DraI*, *EcoRV*, *PvuII* und *StuI* komplett verdaut, wobei Fragmente mit glatten DNA-Enden entstehen, die anschließend mit Adapter-Molekülen ligiert werden. Die vier erhaltenen Ansätze mit den Adapter-ligierten DNA-Fragmenten werden nachfolgend als GenomeWalker-Bibliotheken bezeichnet. Die Ausgangs-DNA muß sehr sauber sein und ein großes Molekulargewicht haben. Nach der Herstellung der Bibliotheken werden mit jeder Bibliothek jeweils zwei PCR-Amplifikationen durchgeführt. Bei der Primär-PCR wird der äußere Adapter-Primer (AP1) zusammen mit einem äußeren genspezifischen Primer (GSP1) verwendet. Nach Durchführung der Primär-PCR dient ein Aliquot der Primär-PCR-Mischung als Template für eine verschachtelte Sekundär-PCR mit einem inneren Adapter-spezifischen Primer (AP2) und einem inneren genspezifischen Primer (GSP2). Die PCR-Produkte wurden anschließend über ein Agarose-Gel analysiert. Bei *single copy*-Genen werden mit allen 4 Bibliotheken singuläre Banden erwartet. Das Fehlen von PCR-Produkten ist allerdings nicht ungewöhnlich, da die Distanz zwischen dem genspezifischen Primer und der entsprechenden Restriktions-Schnittstelle so groß sein kann, dass unter den Bedingungen des Systems keine Amplifikate entstehen. Jedes der erhaltenen PCR-Produkte beginnt mit dem 5'-Ende des genspezifischen Primers 2 (GSP2) und reicht in unbekannte benachbarte genomische DNA hinein. Die größten Amplifikate können dann kloniert und sequenziert werden. Diese Methode liefert zugleich eine Restriktionskartierung des amplifizierten genomischen Abschnitts: wird bei einem *single copy*-Gen mit allen vier Bibliotheken eine singuläre spezifische Bande erhalten, so müssen im schwersten Amplifikat die Schnittstellen derjenigen Enzyme vorhanden sein, mit denen die anderen Bibliotheken erstellt wurden.

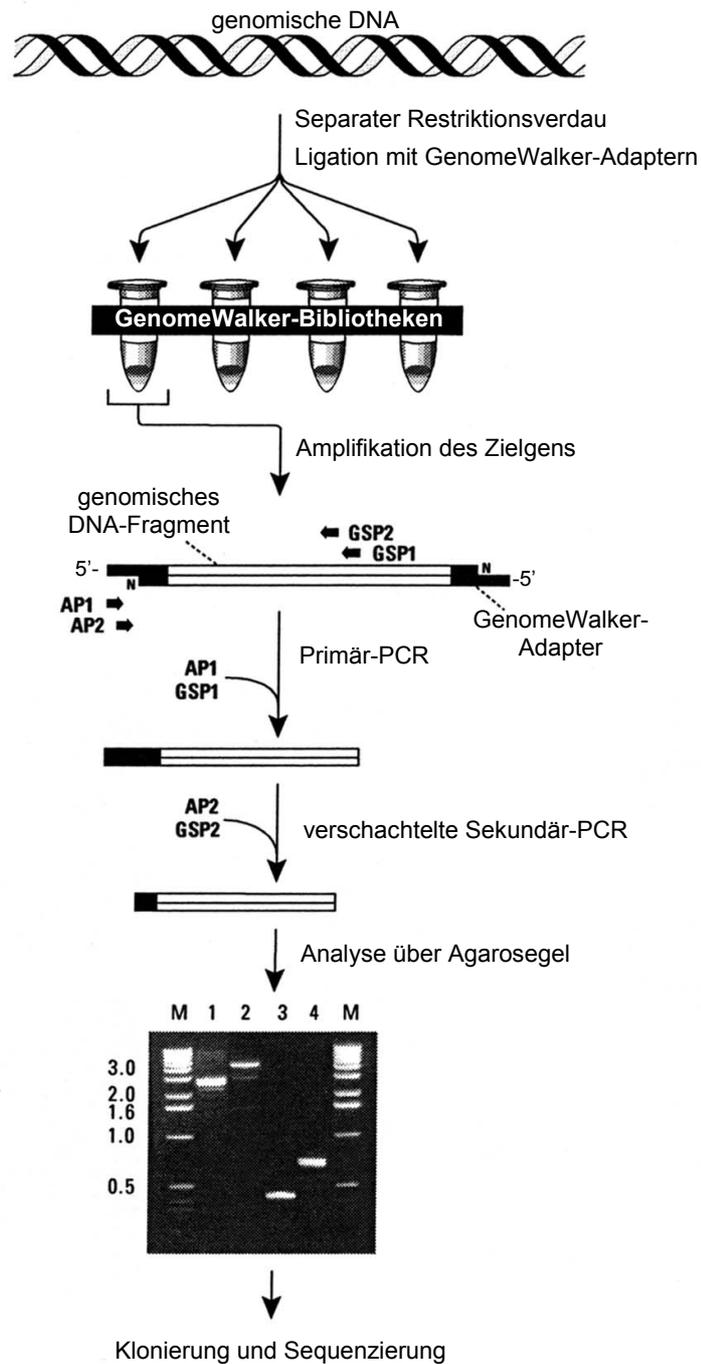


Abb. 2.2: Flußdiagramm für das GenomeWalker-Protokoll. Das Gel zeigt exemplarisch die erhaltenen Amplifikate, die durch PCR mit den vier verschiedenen Bibliotheken erhalten wurden. Bahn 1: *DraI*-Bibliothek, Bahn 2: *EcoRV*-Bibliothek, Bahn 3: *PvuII*-Bibliothek, Bahn 4: *StuI*-Bibliothek, M: DNA-Größenmarker.

AP: Adapter-Primer, GSP: Genspezifischer Primer, N: Aminogruppe, welche die Elongation am 3'-Ende des Adapters verhindert (nach: Anwendungsprotokoll *Universal GenomeWalker Kit*, Clontech).

Die Bindungsstellen der Adapter-Primer sind in Abbildung 2.3 dargestellt. Damit bei der PCR keine Hintergrundamplifikate durch Binden der Adapter-Primer in sense- und antisense-Richtung entstehen, liegt der Adapter nicht komplett als doppelsträngige DNA vor und enthält am 3'-Ende des Gegenstrangs statt einer Hydroxylgruppe eine Aminogruppe, die eine Verlängerung und somit ein unspezifisches Annealing der Adapter-Primer verhindert. Erst nachdem im ersten Zyklus der PCR durch die Anlagerung und Verlängerung des genspezifischen Primers der Adapter als doppelsträngige DNA vorliegt, kann der Adapter-Primer an die neusynthetisierte komplementäre DNA binden.

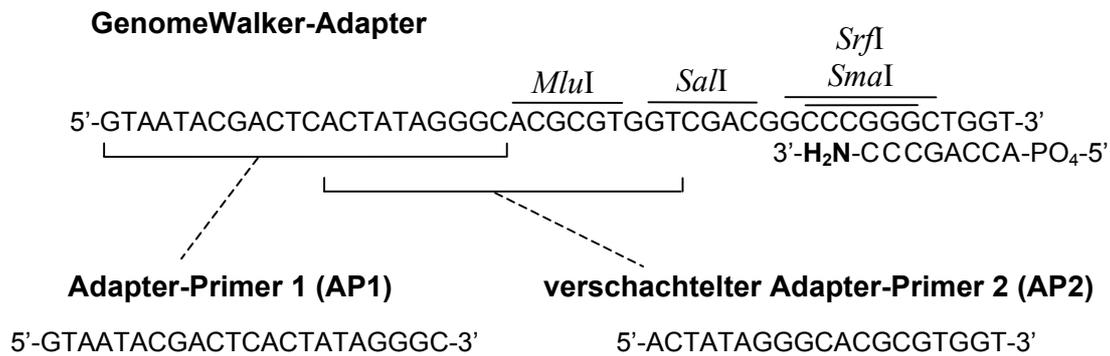


Abb. 2.3: Struktur des GenomeWalker-Adapters und der Adapter-Primer. Der Adapter wird an beide Seiten des genomischen DNA-Fragments ligiert. Die Aminogruppe am Gegenstrang des Adapters blockiert eine Verlängerung in 3'-Richtung und verhindert dadurch die Bildung einer AP1-Bindungsstelle in den Bibliotheken der Adapter-ligierten DNA-Fragmente. Eine AP1-Bindungsstelle entsteht spezifisch erst durch die Verlängerung eines genspezifischen Primers. Die Schmelzpunkte von AP1 und AP2 betragen 59°C und 71°C.

Die Synthese der adapter-ligierten Genome Walker-Bibliotheken erfolgte nach Herstellerangaben. Dazu wurde hochmolekulare Kobra-DNA in vier Ansätzen mit den Restriktionsenzymen DraI, EcoRV, PvuII und StuI vollständig verdaut und die erhaltenen Fragmente mit den Genome Walker-Adaptoren ligiert. Die erhaltenen vier Bibliotheken wurden bei 4°C aufbewahrt und für PCR-Analysen eingesetzt. Ein Aliquot der Bibliotheken wurde bei -20°C gelagert.

Die Analyse der Bibliotheken erfolgte mittels verschachtelter PCR nach einem Touchdown-Programm unter Verwendung der Hot Start-Polymerase *Fast Start Taq DNA-Polymerase* (Roche). Es wurde ein zweistufiges PCR-Programm mit folgenden Parametern durchgeführt:

| | | | |
|-----------|---------------------|------|-------|
| 1 Zyklus | Primärdenaturierung | 95°C | 4 min |
| 7 Zyklen | Denaturierung | 95°C | 20 s |
| | Annealing | 72°C | 3 min |
| 32 Zyklen | Denaturierung | 95°C | 20 s |
| | Annealing | 67°C | 3 min |
| 1 Zyklus | Finale Elongation | 67°C | 7 min |

Ein Aliquot (1 µl) der ersten Amplifikationsrunde wurde dann als Template für die verschachtelte Sekundär-PCR eingesetzt:

| | | | |
|-----------|---------------------|------|-------|
| 1 Zyklus | Primärdenaturierung | 95°C | 4 min |
| 5 Zyklen | Denaturierung | 95°C | 20 s |
| | Annealing | 72°C | 3 min |
| 20 Zyklen | Denaturierung | 95°C | 20 s |
| | Annealing | 67°C | 3 min |
| 1 Zyklus | Finale Elongation | 67°C | 7 min |

Bei den zweistufigen PCR-Programmen finden Annealing der Primer und Elongation durch die Polymerase bei 67°C statt. Durch die Vermeidung niedrigerer Temperaturen soll die Amplifikation unspezifischer Banden verhindert werden. Entsprechend der hohen Annealing-Temperatur wurden die genspezifischen Primer mit einer Länge von 25-35 bp gewählt.

2.2.19 Identifizierung regulatorischer Elemente

Die Datenbanksuche zur Identifizierung regulatorischer Elemente in den 5'-flankierenden Genregionen wurde mit folgenden Programmen durchgeführt:

TFSEARCH (<http://www.cbrc.jp/research/db/TFSEARCH.html>),

Proscan (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html) und

Signal Scan (<http://bimas.dcrn.nih.gov/molbio/signal>).

Für weitere Vergleiche diente eine Liste von Transkriptionsfaktoren der Transcription Factor Database TFD (<http://www.cgr.ki.se/kisac/databases/tfd.html>).

Proteinbiochemische Methoden

Die proteinbiochemischen Methoden wurden, soweit nicht anders beschrieben, nach Sambrook *et al.* (1989) durchgeführt.

2.2.20 Expression in *E. coli*

Kulturvolumen von 3 ml 2YT-Medium wurden mit einer Kolonie des entsprechenden Klones inokuliert und über Nacht bei 37°C und 220 rpm kultiviert. Mit dieser Vorkultur wurden 100 ml 2YT-Medium mit 0.1% Glucose und einer entsprechenden Antibiotikakonzentration inokuliert und die Zellen bei 37°C und 220 rpm bis zur OD₆₀₀ von 1 kultiviert. Daraufhin wurde durch Zugabe von IPTG in einer finalen Konzentration von 1 mM die Proteinexpression induziert und die Kultur für 3 h bei 25-37°C und 220 rpm inkubiert.

2.2.21 Denaturierender Zellaufschluß

Proteine, die im Zielkompartiment als Einschlusskörper (*inclusion bodies*) vorliegen, können in hohen Ausbeuten unter denaturierenden Bedingungen in Lösung gebracht und dann einer Protein-Reinigung zugeführt werden. Dazu werden chaotrope Agentien wie Guanidinium-Chlorid oder Harnstoff eingesetzt.

Das Zellpellet eines 100 ml Expressionsansatzes wurde mit 2 ml 8 M Harnstoff in PBS, pH 7.2, versetzt, inkubiert, bis sich die zelluläre Substanz sichtbar gelöst hatte, und dreimal für je eine Minute sonifiziert. Anschließend wurde die Zellsuspension über Nacht bei RT auf dem Rollbrett inkubiert. Der zelluläre Restbestandteil wurde dann durch Zentrifugation bei 20.000 xg für einen Zeitraum von 20 min bei RT abgetrennt. Der Überstand wurde abgenommen, in ein frisches Reaktionsgefäß überführt und umgehend für die Protein-Reinigung eingesetzt.

2.2.22 Proteinreinigung mittels Affinitätschromatographie

Die Proteinreinigung des denaturierend gelösten Proteinsolubilisats erfolgte mittels Nickelchelat-Affinitätschromatographie (IMAC, *immobilized metal ion affinity chromatography*). Alle Puffer enthielten mit 8 M Harnstoff die Konzentration des Solubilisierungsreagenzes.

Der Überstand wurde mit Ni-NTA-Matrix aus einem Volumen von 1 ml und Imidazol in einer finalen Konzentration von 20 mM versetzt und über Nacht bei 4°C auf dem Rollbrett inkubiert. Die Matrix wurde dann bei 700 xg für einen Zeitraum von 10 min bei 4°C

abzentrifugiert, der Überstand entfernt und die Matrix vorsichtig in 15 ml PBS mit 20 mM Imidazol resuspendiert, mehrfach invertiert und erneut bei 700 xg für einen Zeitraum von 10 min bei 4°C zentrifugiert. Die Matrix wurde erneut in 2 ml PBS mit 20 mM Imidazol aufgenommen und in eine kleine Durchflußsäule gegeben. Der Durchlauf wurde mit der Waschfraktion vereint und anschließend die gebundenen Proteine sukzessiv mit PBS einer Konzentration von 300 mM Imidazol von der Matrix eluiert. Dabei wurden Fraktionen zu je 500 µl in Reaktionsgefäßen aufgefangen und durch Zugabe von jeweils 50 µl 1 M Tris-HCl, pH 7.5, neutralisiert. Anschließend wurden die Fraktionen mittels eines 17,5%igen Polyacrylamid-Gels auf ihren Proteingehalt hin überprüft. Die geeigneten Fraktionen wurden vereint, gegen PBS, pH 7.4, dialysiert und abschließend durch eine analytische SDS-Gelelektrophorese auf Proteingehalt und Reinheitsgrad hin analysiert.

2.2.23 SDS-PAGE

Die SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE) diente zur Auftrennung von Proteinen nach ihrem Molekulargewicht und wurde in einem diskontinuierlichen System aus Sammelgel und Trenngel gemäss der Methode von Laemmli (Laemmli 1970) in einer Minigel-Apparatur (Mighty Small, Hoefer) durchgeführt.

Giessen von SDS-Polyacrylamid-Gelen

Die in einem senkrechten Gießstand (Hoefer) gegossenen Gele hatten eine Dicke von 0,75 mm. Je nach Trennproblem variiert die Polyacrylamid-Konzentration der Trenngele zwischen 7,5% und 20%. Für die Analyse eines Peptids wurden Trenngele mit einer Konzentration von 17,5% verwendet. Die Konzentration des Sammelgels betrug 4%.

Die Trenngel- und Sammelgel-Lösungen setzten sich aus den folgenden Komponenten zusammen (Mengen reichen für 4 Gele):

| | Trenngel (17,5%) | Sammelgel (4%) |
|---|-----------------------------|---------------------------|
| ddH ₂ O | 4,05 ml | 6,125 ml |
| Trenngel- bzw. Sammelgel-Puffer (4x) | 6,25 ml | 2,5 ml |
| Acrylamid-Lösung | 14,6 ml | 1,3 ml |
| Ammoniumpersulfat- Lösung (10%) | 75 µl | 40 µl |
| TEMED | 20 µl | 8 µl |
| gesamt | 25 ml | 10 ml |

Die Trenngel-Lösungen wurden vorsichtig mit Isopropanol überschichtet, um eine ebene Gelkante zu erhalten. Nach der Polymerisationszeit von 1 Stunde wurde die überstehende Flüssigkeit dekantiert. Die polymerisierten Trenngele wurden mit der Sammelgel-Lösung (4%) überschichtet, in welche Teflon-Kämme zur Erzeugung von Probenfächern eingefügt wurden. Die SDS-Polyacrylamidgelelektrophorese wurde in einer senkrechten, mit Tankpuffer befüllten Minigel-Apparatur durchgeführt. Das Auftragen der Proben erfolgte, nachdem diese zuvor mit Probenpuffer (4x) versetzt und dann für 3-5 min in einem Wasserbad aufgekocht wurden. Zur Molekulargewichtsbestimmung der Proteinproben wurden die Gele zusätzlich mit dem Marker Precision Plus Protein Standard (Biorad, München) beladen. Zu Beginn der Elektrophorese wurde eine Spannung von 140 V angelegt, bis die Proben das Trenngel erreicht hatten. Danach wurde mit einer konstanten Spannung von 160 V elektrophoretisch getrennt, bis der als Referenz zugesetzte Farbstoff im Probenpuffer (Bromphenolblau) das untere Ende des Gels erreichte. Die Elektrophorese wurde unter Wasserkühlung durchgeführt.

Färbung von Proteinbanden mit Coomassie Brilliant Blau

Eine Möglichkeit zur Anfärbung von Proteinbanden ist die Coomassie-Färbung. Die Proteinnachweisgrenze liegt bei der Detektion mit diesem Farbstoff bei 0.1-1 µg. Zur Färbung wurden die Trenngele nach beendeter Elektrophorese für 30 min bei RT in eine Coomassie-Färbelösung gelegt. Zur anschließenden Reduktion des Hintergrunds wurden die Gele in eine Entfärber-Lösung überführt und für ca. 2 h inkubiert.

Trocknen und Aufbewahren von SDS-Polyacrylamidgelen

Vor dem Trocknen wurden die Polyacrylamid-Gele für 30 min in Glycerol-haltiger Geltrocknungs-Lösung inkubiert, um ein Reißen beim Trocknungsprozess zu verhindern. Anschliessend wurden die Gele in dem Trocknungsrahmen *Easy Breeze Air Gel dryer SE 1200* (Hoefer Scientific Instruments, San Francisco, CA) zwischen zwei Cellophan-Folien eingespannt und über Nacht an der Luft getrocknet.

3 Ergebnisse

3.1 Struktur des CVF1-Gens

In vorangegangenen Arbeiten wurde zum Zwecke der Aufklärung der CVF1-Genstruktur eine genomische Cosmid-Bibliothek der Monokel-Kobra *Naja kaouthia* im Vektor sCos1 erstellt, die aus $4-5 \cdot 10^5$ verschiedenen Klonen besteht. Durch radioaktives Screening dieser Bibliothek mit einer P-32 markierten cDNA-Sonde konnte ein CVF1-Klon identifiziert, monoklonalisiert und partiell aufgeklärt werden (B. Kunze, unveröffentlichte Ergebnisse). Die Sonde (204 bp Länge, Position 2903-3106 der CVF1-cDNA) war auf der Basis der humanen C3-Genstruktur so gewählt worden, dass sie Exon 24 des CVF1-Gens entspricht. Exon 24 codiert für die Thioester-Region, die in den cDNA-Sequenzen von humanem C3 und CVF1 relativ stark konserviert ist (C3d-Bereich). Daher war zu vermuten, dass diese Region auch in der CVF1-Sequenz durch ein einziges Exon codiert wird. Mit dieser Sonde wurde der CVF1-Klon 1R67 identifiziert.

3.1.1 Charakterisierung eines CVF1-Cosmid-Klons

Im folgenden ist die Aufklärung der kompletten Exon/Intron-Struktur des CVF1-Klons 1R67 beschrieben. Zunächst wurde der Klon durch Restriktionsanalysen mit NotI, NdeI und EcoRI (Abb. 3.1A) charakterisiert. Durch NotI wird das genomische Insert aus dem Vektor (6,7 kb) herausgeschnitten und ist als singuläre Bande (>20 kb) sichtbar. Beim NdeI-Verdau entstehen 3 schwere Banden mit Molekulargewichten von ca. 15 kb, 11 kb und 9 kb sowie Banden bei 3,2 kb und 2,0 kb. Durch Addition der Fragmentgrößen ergibt sich für das CVF1-Insert eine Größe von ca. 33 kb. Beim EcoRI-Verdau entstehen mehrere Banden mittleren Molekulargewichts (8.5 kb, 6.7 kb, 4.5 kb (Doppelbande), 3.7 kb (Doppelbande), 2.4 kb, 1,0 kb) sowie einige kleinere Fragmente. Aufgrund der Insertgröße von ca. 33 kb ergibt sich, dass einige Banden des EcoRI-Verdaus als Überlagerung von Fragmenten gleicher Größe entstanden sind.

Ein Southern-Blot mit verdauter 1R67-DNA wurde anstelle der P-32 markierten Sonde mit einer Digoxigenin-markierten Sonde hybridisiert, die Exon 24 (204 bp) entspricht. In Abbildung 3.1 ist zu erkennen, dass das schwerste EcoRI-Fragment bei 8 kb mit der Sonde hybridisierte und ein Signal ergab. Durch Klonierung und Sequenzierung der EcoRI-Fragmente sowie durch Sequenzierung von Subklonen konnte von Kunze gezeigt werden, dass der Klon 1R67 den CVF1-Genabschnitt mit den Exons 16-27 enthält.

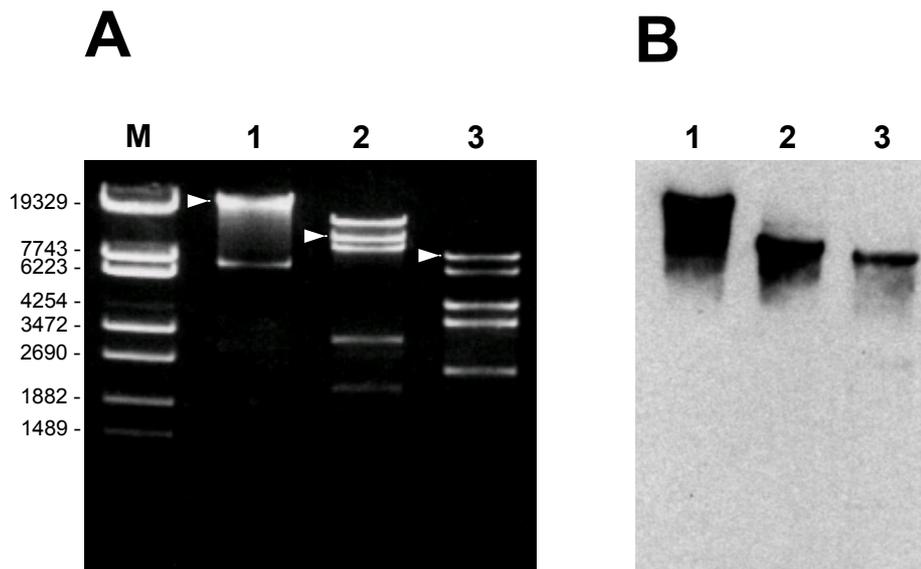


Abb. 3.1: Charakterisierung des CVF1-Klons 1R67. A: Restriktionsanalyse von 1R67-DNA, aufgetrennt über ein 1%iges Agarosegel, M: DNA-Marker λ /Eco130I, Bahn 1: 1R67-DNA/NotI, Bahn 2: 1R67-DNA/NdeI, Bahn 3: 1R67-DNA/EcoRI. Die mit Pfeilen markierten Banden sind im Southern-Blot (B) als Signale zu erkennen. B: Southern-Blot des Agarosegels aus A. Geblottet wurde über Nacht auf eine neutrale Nylonmembran, die mit der DIG-Sonde „Exon 24“ hybridisiert wurde.

Für die exakte Bestimmung der CVF1-Genstruktur von Exons 16-27 wurde die Exonstruktur dieses Genabschnitts in Analogie zur humanen C3-Genstruktur (Vik *et al.* 1991) vorhergesagt (Abb. 3.2). Daraufhin konnten exonspezifische Primer generiert werden, mit denen die Introns des Klons 1R67 durch Verwendung von flankierenden Exon-Primern amplifiziert wurden.

Bei allen nachfolgend beschriebenen PCR-Amplifikationen zur Analyse der CVF1-Genstruktur wurden die in Anhang 7.6 aufgeführten PCR-Programme verwendet. Alle PCR-Produkte wurden nach Klonierung in entsprechende Vektoren (pUC18, pCR2.1-TOPO, pCR-XL-TOPO) mit M13-Primern sequenziert. Die Sequenzdaten sind in Anhang 7.1 aufgeführt. Darüber hinaus wurde für jedes Amplifikat eine EcoRI-Restriktionsanalyse durchgeführt (Fingerprint); die entsprechenden Fragmente und identifizierten Schnittstellen sind im Anhang aufgelistet.

Exon 16

CVF1 ..CAGCTGCAAAGTGTCTCAGCCTGCAAATCGGAGGCGTCGAGTTCTGTTTTGCTGCT 1979
 CVF2 ..CAGCTGCAAAGTGTCTCAGCCTGCAAATCGGAGGCGTCGAGTTCTGTTTTGCTGCT 1027
 coC3 ..CAGCTGCAAAGTGTCTCAGCCTGCAAATCGGAGGCGTCGAGTTCTGTTTTGCTGCT 2002

Exon 17

CVF1 TGACAGCAACGCAAGCAAAGCGGCAGAATTTTCAGGATCAAGACCTGCGTAAATGCTGTGA 2039
 CVF2 TGACAGCAACGCAAGCAAAGCGGCACAGTTTCAGGATCAAGACCTGCGTAAATGCTGTGA 1087
 coC3 TGACAGCAAAGCAAGCAAAGCGGCACAGTTTCAGGATCAAGCCCTGCGTAAATGCTGTGA 2062

CVF1 AGATGTCATGCATGAGAACCCCATGGGGTACACTTGTGAAAAGCGTGCAAATAACATCCA 2099
 CVF2 AGATGTCATGCATGAGAACCCCATGGGGTACACTTGTGAAAAGCGTGAAAATAACATCCA 1147
 coC3 AGATGTCATGCATGAGAACCCCATGGGGTACACTTGTGAAAAGCGTGCAAATAACATCCA 2122

CVF1 GGAGGGAGATGCTTGTAAAGGCTGCCTTCTTGAATGCTGTGCTACATCAAGGGGGTCCG 2159
 CVF2 GGAGGGAGATGCTTGTAAAGGCTGCCTTCTTGAATGCTGTGCTACTACATCAAAGGGAATCCA 1207
 coC3 GGAGGGAGATGCTTGTAAAGGCTGCCTTCTTGAATGCTGTGCTACTACATCAAAGGGAATCCG 2182

Exon 18

CVF1 AGATGAAAACCAACGGGAGAGCGAGTTGTTTCTGGCAAGAGATGATAATGAAGATGGTTT 2219
 CVF2 AGATGAAACCAACGGGAGAGCGAGTTGTTTCTGGCAAGAGATGATAATGAAGATGATTT 1267
 coC3 AGATGAAAACCAACGGGAGAGCGAGTTGTTTCTGGCAAGAGATGATAATGAAGATGAAC 2242

CVF1 CATAGCAGATAGTGATATCATCTCAAGGTCTGATTTCCCAAGAGTTGGTTGTGGCTAAC 2279
 CVF2 CATAGCAGATAGTGATATCATCTCAAGGTCTGATTTCCCTGAGAGTTGGTTGTGGCTAAT 1327
 coC3 CATAGCAGATAGTGATATCATCTCAAGGTCTGATTTCCCAAGAGTTGGTTGTGGCTAAC 2302

Exon 19

CVF1 AAAGGACTTGACCGAGGAGCCTAACAGTCAAGGGATTTCAAGCAAGACAATGTCTTTTTA 2339
 CVF2 AAAGGACTTGACCGAGGAGCCTAACAGTCAAGGGATTTCAAGCAAGACAATGTCTTTTTA 1387
 coC3 AAAGGACTTGACCGAGGAGCCTAACAGTCAAGGGATTTCAAGCAAGACAATGTCTTTTTA 2362

CVF1 TCTGAGGGATTCCATCACAACCTGGGTGGTGCTGGCTGTAAGCTTTACCCACCAAAGG 2399
 CVF2 TCTGAGGGATTCCATCACAACCTGGGTGGTGCTGGCTGTAAGCTTTACCCACCAAAGG 1447
 coC3 TCTGAGGGATTCCATCACAACCTGGGTGGTGCTGGCTGTAAGCTTTACCCACCAAAGG 2422

Exon 20

CVF1 GATCTGTGTGGCTGAACCTTATGAAATAAGAGTCATGAAAGTCTTCTTCATTGATCTTCA 2459
 CVF2 GATCTGTGTGGCTGAACCTTATGAAATAAGAGTCATGAAAGTCTTCTTCATTGATCTTCA 1507
 coC3 GATCTGTGTGGCTGAACCTTATGAAATAAGAGTCATGAAAGTCTTCTTCATTGATCTTCA 2482

CVF1 AATGCCATATTCAGTAGTGAAGAATGAGCAGGTGGAGATTCGAGCTATTCTGCACAAC 2519
 CVF2 AATGCCATATTCAGTAGTGAAGAATGAGCAGGTGGAGATTCGAGCTATTCTGCACAAC 1567
 coC3 AATGCCATATTCAGTAGTGAAGAATGAGCAGGTGGAGATTCGAGCTATTCTGCACAAC 2542

Exon 21

CVF1 CGTTAACGAGGATATTTATGTGCGAGTGGAACTGTTATACAACCCAGCCTTCTGCAGTGC 2579
 CVF2 CGTTAACGAGGATATTTATGTGCGAGTGGAACTGTTATACAACCCAGCCTTCTGCAGTGC 1627
 coC3 CGTTAACGAGGATATTTATGTGCGAGTGGAACTGTTATACAACCCAGCCTTCTGCAGTGC 2602

CVF1 TTCCACAAAAGGACAAAGATAACCGACAGCAGTTCCCAATTAAGCCCTGTCCTCCAGAGC 2639
 CVF2 TTCCACAAAAGGACAAAGATAACCGACAGCAGTTCCCAATTAAGCCCTGTCCTCCAGAGC 1687
 coC3 TTCCACAAAAGGACAAAGATAACCGACAGCAGTTCCCAATTAAGCCCTGTCCTCCAGAGC 2662

CVF1 AGTACCGTTTGTGATAGTCCATTAGAGCAAGGATTGCATGATGTTGAGATTAAAGCAAG 2699
 CVF2 AGTACCGTTTGTGATAGTCCATTAGAGCAAGGATTGCATGATGTTGAGATTACAGCAAG 1747
 coC3 AGTACCGTTTGTGATAGTCCATTAGAGCAAGGATTGCATGATGTTGAGATTACAGCAAG 2722

CVF1 TGTCAGGAAGCGTTGTGGTCAGACGGTGTGAGGAAGAACTGAAAGTTGTACCTGAAGG 2759
 CVF2 TGTCAGGAAGCGTTGTGGTCAGACGGTGTGAGGAAGAACTGAAAGTTGTACCTGAAGG 1807
 coC3 TGTCAGGAAGCGTTGTGGTCAGACGGTGTGAGGAAGAACTGAAAGTTGTACCTGAAGG 2782

Fortsetzung

| | | |
|----------------|---|------|
| Exon 22 | | |
| CVF1 | GGTACAGAAATCCATTGTGACTATTGTTAAACTGGACCCAAGGGCAAAGGAGTTGGTGG | 2819 |
| CVF2 | GGAAATGGAAAAGTATTGTTACTATTATTGAACTGGACCCAATAAAAAGGAATTGGTGG | 1867 |
| coC3 | GGAAACGGAAAATATTGTGACTATTATTGAACTGGACCCAAGTGTAAAAGGAGTTGGTGG | 2842 |
| Exon 23 | | |
| CVF1 | AACACAGCTAGAAGTGATCAAAGCCCGCAAATTAGATGACAGAGTGCCTGACACAGAAAT | 2879 |
| CVF2 | AACACAGCTAGAATTGGTCAAAGCCAAATAAATAATGACAGCGTTCCTGATACCGAAAT | 1927 |
| coC3 | AACCCAGGAACTAACGGTCAATAGCCAAATAAATAATGATGACAAAGTGCCTGATACAGAAAT | 2902 |
| Exon 24 | | |
| CVF1 | TGAAACCAAGATTATCATCCAAGGTGACCCCTGTGGCTCAGATTATTGAAAACCTCAATTGA | 2939 |
| CVF2 | AGAAACCAAGATTACTATTCAAAGGTGATCCTGTGGCTCAGACTATTGAAAACCTCAATTGA | 1987 |
| coC3 | TGAGACCAAGATTCTGTCTAGGTGACCCCTGTGGCTCAGATTATTGAAAACCTCAATTGA | 2962 |
| CVF1 | TGGAAGTAAACTCAACCATCTCATTATCACTCCTTCTGGCTGTGGGGAGCAAAATATGAT | 2999 |
| CVF2 | TGGAAGTAAACTCAACCATCTCATTATCACTCCTTTGGCTGTGGGGAGCAAAATATGAT | 2047 |
| coC3 | TGGAAGTAAACTCAATTCATCTCATTATCACTCCTTCTGGCTGTGGGGAGCAAAATATGAT | 3022 |
| CVF1 | CCGCATGGCCGACCCAGTTATTGCCACCTACTACCTGGACACCACAGAGCAGTGGGAGAC | 3059 |
| CVF2 | CCGCATGACTGCACCCAGTTATTGCCACCTACTACCTGGACACCACAAGCAGTGGGAGAC | 2107 |
| coC3 | CAACCATGACTCATTCGGTCATTGCCACCTACTACTGGACCCAAACAGGGCAGTGGGAGAA | 3082 |
| Exon 25 | | |
| CVF1 | TCTCGGCATAAAATCGCAGGACTGAAGCTGTCAATCAGATCGTGACTGTTATGCCAGCA | 3119 |
| CVF2 | TCTCGGCATAAAATCGCAGGACTGAAGCTGTCAATCAGATCATGACTGTTATGCCAGCA | 2167 |
| coC3 | TCTTGGTCTCATGCGCAGGACTGAAGCTATCAAACAGATCATGACTGTTATGCCAGCA | 3142 |
| CVF1 | GATGGTGTACAAGAAAGCAGATCATTCCCTATGCAGCATTTACAAACCGTGCATCTAGTTC | 3179 |
| CVF2 | GTGGTGTACAAGAAAGCAGACATTCCCTATGCAGCATTTACAAACAGTGCATCTAGTTC | 2227 |
| coC3 | GATGGTGTACAAGAAAGCAGATCATTCCCTATGCAGCATTTACAAACCGTGCATCTAGTTC | 3202 |
| Exon 26 | | |
| CVF1 | TTGGCTAACAGCATATGTCGTAAGTCTTTGCCATGGCTGCCAAAATGGTAGCAGGCAT | 3239 |
| CVF2 | TTGGCTAACAGCATATGTTGTAAGTCTTTGCCATGGCTGCCAAAATGGTAGAAGACAT | 2287 |
| coC3 | TTGGCTAACAGCATATGTTGTAAGTCTTAGCCATGGCTTCCAAACATGGTAGAAGACAT | 3262 |
| CVF1 | TAGTCATGAAATCATTGTGGAGGTGTGAGGTGGCTGATTCTGAACAGGCAACAACCAGA | 3299 |
| CVF2 | TAACCATGAAATCATTGTGGAGGTATGAGGTGCTGATTCTGAACAGGCAACAACCAGA | 2347 |
| coC3 | TAGCCATGAAATATTATTGTGGAGGTGTGAATGGCTCATTCTGAACAGGCAACAACCAGA | 3322 |
| Exon 27 | | |
| CVF1 | TGGAGCGTTCAAAAGAAAATGCCCTGTACTTTCTGGAACAATGCAGGGAGGAATTCAAGG | 3359 |
| CVF2 | TGGAGTGTTCAGAAAAAGCCCCTGTACTTTCTGGAACAATGCAGGGAGGCATCAAGG | 2407 |
| coC3 | TGGAGTGTTCAAAAGAAAATGCCCTGTGATCCATGGAGAATGCTGGAGGAATCAAGG | 3382 |
| CVF1 | TGCTGAAGAAGAAGTATATTTAACAGCTTTTATTCTGGTTGCGTTGTTGGAATCCAAAAC | 3419 |
| CVF2 | TGCTGAAACGAGAAGCATCTTTAACAGCTTTTATTCTGGTTGCGTTGTTGGAATCCAGATC | 2467 |
| coC3 | TGCTGAAACGAGAAGCATCTTTAACAGCAATTCATTGTGACTGCATTTATTGGAATCCAGATC | 3442 |
| CVF1 | AATCTGCAATGACTATGTCAATAGT... | 3444 |
| CVF2 | AATCTGCAATGCATATATCAATATT... | 2492 |
| coC3 | AGTCTGCAAAAGAACAAATCAATATT... | 3467 |

Abb. 3.2: Darstellung der Exonstrukturen von CVF1, CVF2 und coC3. Auszug aus den cDNA-Sequenzen (Exons 16-27). Die Exongrenzen wurden in Analogie zur humanen C3-Genstruktur vorhergesagt und für CVF1 vollständig sowie für CVF2 und coC3 exemplarisch bestätigt.

Introns 16-26:

Die Introns 16-26, welche im Cosmid-Klon 1R67 vollständig enthalten sind, wurden mittels PCR unter Verwendung von exonspezifischen Primern amplifiziert. Für die Generierung dieser Primer wurde die Struktur des CVF1-Gens in Analogie zur humanen C3-Genstruktur (Vik *et al.* 1991) vorhergesagt. In Abbildung 3.2 ist der entsprechende Ausschnitt aus den Exonstrukturen von CVF1, CVF2 und coC3 gezeigt. Die erhaltenen PCR-Amplifikate mit Introns 16-26 des CVF1-Gens sind in Abbildung 3.3 dargestellt.

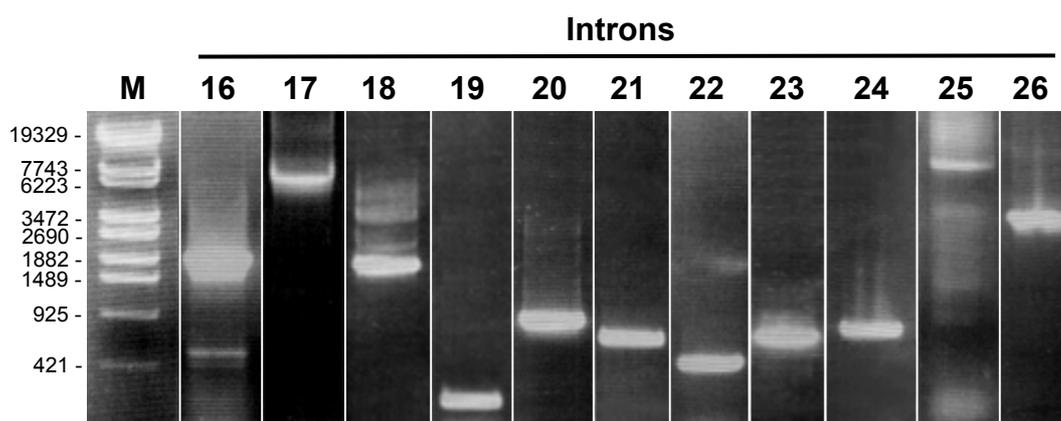


Abb. 3.3: Gelelektrophoretische Analyse der PCR-Amplifikate mit Introns 16-26 des CVF1-Gens (1%iges Agarosegel). M: Marker λ -DNA/Eco130I, Intron 16: 1500 bp, Intron 17: 7000 bp, Intron 18: 1531 bp, Intron 19: 274 bp, Intron 20: 968 bp, Intron 21: 722 bp, Intron 22: 444 bp, Intron 23: 756 bp, Intron 24: 827 bp, Intron 25: 8000 bp, Intron 26: 3000 bp.

Die für die Intron-Amplifikationen eingesetzten Primerpaare sind in Tabelle 3.1 aufgeführt.

Tab. 3.1: Übersicht der PCR-Amplifikate des CVF1-Cosmidklons. Größenangaben in Klammern sind Schätzungen. Exonsequenzen sind in fetten Großbuchstaben und Intronsequenzen in Kleinbuchstaben dargestellt. * = partiell im Cosmidklon 1R67 enthalten.

| Intron | Primerpaar | Größe PCR-Produkt (bp) | Exon-Intron-Übergangssequenzen |
|--------|--------------------------|------------------------|---|
| 15* | Con2eco / OPA72 | (3500) | ...cctccttcag CTG |
| 16 | OPS72 / OPA24 | (1600) | AAG gtatgtggca...catatcttag CGG |
| 17 | OPS24 / OPA22 | (7.000) | GAG gtaagtagca...tatttttcag ATG |
| 18 | OPS22 / OPA21 | 1531 | AGG gtattggagt...ttgttctag GAT |
| 19 | OPS21 / OPA16 | 274 | AAG gtaaagaatt...ttccttgag GGA |
| 20 | OPS16 / OPA17 | 968 | TAT gtaagtgttg...cctccactag GTG |
| 21 | OPS17 / OPA25 | 722 | GTA gtaagtaaca...tctataacag CCT |
| 22 | OPS25 / OPA14 | 444 | TTG gtgagtaact...catttttcag GTG |
| 23 | HB23s / HB24as | 756 | AAG gtaggaacta...tcacattcag GTG |
| 24 | SD24s / HB25as | 827 | CTG gtaaggcaaa...cccttttcag GTT |
| 25 | OPS15 / OPA26 | (8.000) | TTG gtagggacaa...ttctttacag GCT |
| 26 | OPS20 / HB27as | (3.000) | CAG gtatggtttc...tttcttctag GGA |
| 27* | HB1R67-F5-s / sCos1-3'as | (1000) | AAT gtacgtgatg... |

Anhand der aufgeführten Exon/Intron-Übergangssequenzen ist ersichtlich, dass alle Introns im Cosmid-Klon der gt/ag-Konsensussequenz (Mount 1982) folgen.

Am Beispiel von Intron 16 werden exemplarisch die Primerbindungsstellen im Exon sowie die erhaltene genomische Sequenz des PCR-Produkts im Vergleich zu den korrespondierenden cDNA-Sequenzen von CVF1, CVF2 und coC3 gezeigt.

Intron 16

Unter Verwendung des Primerpaars OPS72 / OPA24 wurde ein Amplifikat von ca. 1500 bp erhalten. Die Randsequenzen des klonierten Amplifikats mit den Exon/Intron-Übergangssequenzen sind in Abbildung 3.4 dargestellt.

| | | | |
|------------------|------|---|---------------|
| CVF1 cDNA | 1943 | CTGCAAATCG GAGGCGTCGC AGTTCTGTTT TGCTGCTTGA CAGCAACGCA | 1992 |
| CVF2 cDNA | 991 | CTGCAAATCG GAGGCGTCGC AGTTCTGTTT TGCTGCTTGA CAGCAACGCA | 1040 |
| coC3 cDNA | 1966 | CTGCAAATCG GAGGCGTCGC AGTTCTGTTT TGCTGCTTGA CAGCAAAGCA | 2015 |
| CVF1 Gen | | CTGCAAATCG GAGGCGTCGC AGTTCTGTTT TGCTGCTTGA CAGCAACGCA | |
| | | ↓ | |
| CVF1 cDNA | 1993 | AGCAAAG | 1999 |
| CVF2 cDNA | 1041 | AGCAAAG | 1047 |
| coC3 cDNA | 2016 | AGCAAAG | 2022 |
| CVF1 Gen | | AGCAAAG gta tgtggcaaat aatgaagttt ctgtttactc tgagataaaa | |
| CVF1 Gen | | gaataaaact acagtccact tgcattgatt gaatcaatca ggaaatggag | |
| CVF1 Gen | | gaaggaatgg | |
| | | (ca. 1200 bp) | |
| CVF1 Gen | | aactcatttg | |
| CVF1 Gen | | attggtttat gtagatactc gttgtctctt gtttoccttt atgcacatga | |
| CVF1 cDNA | 2000 | | C 2000 |
| CVF2 cDNA | 1048 | | C 1048 |
| coC3 cDNA | 2023 | | C 2023 |
| CVF1 Gen | | ctatgagaaa taggattctc taatgtttca cgcattttgc atatttttag C | |
| CVF1 cDNA | 2001 | GGCAGAATTT CAGGATCAAG ACCTGCGTAA ATGCTGTGAA GATGTCATGC | 2050 |
| CVF2 cDNA | 1049 | GGCAGAGTTT CAGGATCAAG ACCTGCGTAA ATGCTGTGAA GATGTCATGC | 1098 |
| coC3 cDNA | 2024 | GGCAGAGTTT CAGGATCAAG ACCTGCGTAA ATGCTGTGAA GATGTCATGC | 2073 |
| CVF1 Gen | | GGCAGAATTT CAGGATCAAG ACCTGCGTAA ATGCTGTGAA GATGTCATGC | |
| | | ← | |

Abb. 3.4: Partielle Sequenz des PCR-Produkts mit Intron 16 des CVF1-Gens. Exonsequenzen sind in Großbuchstaben und fett dargestellt. Intronsequenzen sind in Kleinbuchstaben dargestellt. Die entsprechenden cDNA-Sequenzen von CVF1, CVF2 und coC3 sind für Vergleichszwecke oberhalb der Exonsequenzen angeordnet. Die Primer-Sequenzen sind mit Pfeilen unterlegt. Exon 16 hat eine Größe von 75 bp. Intron 16 hat eine Größe von ca. 1400 bp.

Die Exonsequenzen des PCR-Produkts zeigen Übereinstimmung mit der cDNA von CVF1 und identifizieren den Cosmidklon 1R67 als CVF1-Klon. Die postulierten Exongrenzen wurden bestätigt. Die Exon/Intron-Übergänge folgen den gt/ag-Konsensus-Sequenzen.

Eine vollständige Übersicht über die Exon/Intron-Strukturen im Cosmid-Klon 1R67 ist in Tabelle 3.1 dargestellt.

Die Exongrenzen des CVF1-Genabschnitts im Cosmid-Klon 1R67 stimmen mit den Vorhersagen aus der huC3-Genstruktur weitgehend überein. Als einzige Ausnahme von den stark konservierten Exongrenzen in beiden Genen ist die Position von Intron 19 um ein Nukleotid verschoben (Abb. 3.5).

| | | Exon 19 | | | |
|------|------|------------------|---|--------------------------|------|
| CVF1 | 2313 | ...tag | GATTTC AAGCAAGACAATGTCTTTTATCTGAGGGATTCCATCACAACTGGGT | | 2366 |
| huC3 | 2415 | ...cag | AATCTCTACGAAGCTCATGAATATATTTTGAAAGACTCCATCACACCTGGGA | | 2468 |
| | | Intron 19 | | Exon 20 | |
| CVF1 | 2367 | | GGTGCTGGCTGTAAGCTTTACACCCACCAAG gta....cag | GGATCTGTGTGGCTGAA | 2415 |
| huC3 | 2469 | | GATCTCGGCTGTGAGCATCTCGACAAGAAAGG gtg....cag | GATCTGTGTGGCAGAC | 2517 |
| CVF1 | 2416 | | CCTTATGAAATAAGAGTCATGAAAGTCTTCTTCATTGATCTTCAAATGCCATATTAGTA | | 2475 |
| huC3 | 2518 | | CCCTTCGAGGTCACAGTAATGCAGGACTTCTTCATCGACCTCCGGCTACCTACTCTGTT | | 2577 |
| CVF1 | 2476 | | GTGAAGAATGAGCAGGTGGAGATTCGAGCTATTCTGCACAACACTACGTT---AACGAGGAT | | 2532 |
| huC3 | 2578 | | GTTCGAAAACGAGCAGGTGGAAATCCGAGCCCTTCTCTACAAATACCGGCAGAACCAAGAC | | 2637 |
| CVF1 | 2533 | | ATTTAT gta... 2538 | | |
| huC3 | 2638 | | CTCAAG gtg... 2643 | | |

Abb. 3.5: Darstellung der Phasenverschiebung von Intron 19. Die Position von Intron 19 befindet sich im huC3-Gen ein Nukleotid weiter stromabwärts als im CVF1-Gen. Die Nummerierung bezieht sich auf die entsprechenden cDNA-Sequenzen.

Die Analyse der flankierenden 5'- und 3'-Regionen des Cosmidklons 1R67 erfolgte wie nachfolgend beschrieben durch überlappende PCR und Klonierung von EcoRI-Fragmenten.

5'-Ende des CVF1-Klons 1R67

Der 5'-Bereich des CVF1-Klons 1R67 enthält ein ca. 7,5 kb langes Teilstück von Intron 15 (Abb. 3.6). Da eine Amplifizierung des gesamten Intronabschnittes mittels PCR nicht möglich war, wurde dieser Abschnitt durch Klonierung und Sequenzierung der entsprechenden EcoRI-Fragmente und überlappenden PCR-Amplifikate aufgeklärt. Die Positionen der 4,5 kb und 3,6 kb schweren EcoRI-Fragmente (Abb. 3.1) im Cosmid-Klon wurden durch Identifizierung der Randsequenzen bestimmt. Die erhaltenen Intron-Sequenzen ermöglichten die Generierung von Primern für überlappende PCR.

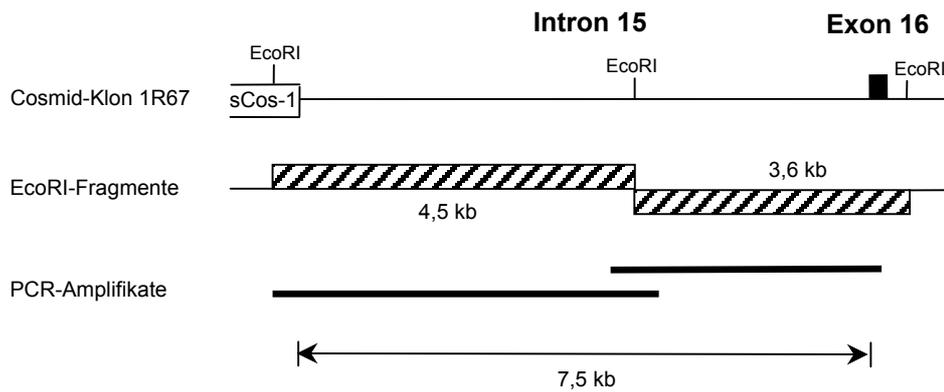


Abb. 3.6: Schematische Darstellung zur Aufklärung der 5'-Region des Cosmid-Klons 1R67. Der 5'-Bereich enthält ein ca. 7,5 kb langes Teilstück von Intron 15 des CVF1-Gens.

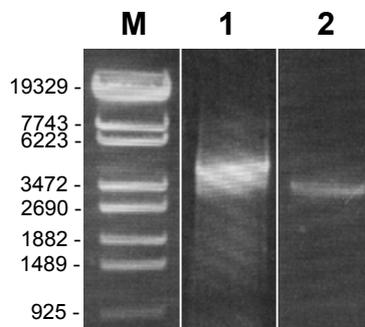


Abb. 3.7: Gelelektrophoretische Analyse der PCR-Amplifikate mit Intron 15-Teilstücken des CVF1-Gens (1%iges Agarosegel). M: Marker λ -DNA/Eco130I, Bahn 1: Intron 15-Teilstück (5'-Ende), 4500 bp Primerpaar sCos1-T3s/HB-Con2eco-as; Bahn 2: Intron 15-Teilstück (3'-Ende), 3500 bp, Primerpaar Con2eco/OPA72.

Mit dem Primerpaar sCos1-T3s / HB-Con2eco-as wurde ein 4,5 kb-Produkt amplifiziert, welches sich vom Vektor sCos-1 bis zur ersten EcoRI-Schnittstelle im CVF1-Insert erstreckt und mit seiner Länge der Größe eines der EcoRI-Fragmente entspricht (siehe Abb. 3.6 und Abb 3.7). Die Sequenz dieses Amplifikats ist in Abbildung 3.8 dargestellt.

```

Vektor sCos-1      CGCAATTAAC CCTCACTAAA GG
CVF1 (Gen)         CGCAATTAAC CCTCACTAAA GGgatcttaa aagagctggt tttataaac
                   |
CVF1 Gen           tcagagagaa cttgggaaag attccatgga tgaaaatcct aaaggggaaa
CVF1 Gen           acaacttaag .....
                   ..... (ca. 4300 bp).....
CVF1 Gen           ..... ctcaagtccc
CVF1 Gen           agcaaacttg ggcgatcctt gtccacgatg gtcagtggca aacgtttggt
CVF1 Gen           gtggccttta aactgcacgg ttgaattccc tcgctccaac aactggaatc
                   |

```

Abb. 3.8: Teilsequenz von Intron 15 des CVF1-Gens. Als Matrize diente der Cosmid-Klon 1R67. Das PCR-Amplifikat mit der 5'-Teilsequenz von Intron 15 im Cosmid-Klon 1R67 hat eine Länge von ca. 4500 bp. Die MboI-Schnittstelle (gatc) am 5'-Ende sowie die EcoRI-Schnittstelle (gaattc) am 3'-Ende des Amplifikats sind unterstrichen dargestellt. Der Bereich der Überlappung mit nachfolgender Sequenz ist kursiv dargestellt.

Mit dem Primerpaar Con2eco/OPA72 wurde mittels PCR ein 3.6 kb-Amplifikat (Abb. 3.7) erhalten, dessen Randsequenz in Abbildung 3.9 gezeigt ist.

```

CVF1 Gen           tcagtggcaa acgtttggtg tggcctttaa actgcacgtt gaattccctc
CVF1 Gen           gtcccaacaa ctggaatcctt gttcccttgg tagtctttga gggtcagtgc
CVF1 Gen           ttgcttctg .....
                   ..... (ca. 3300 bp).....
CVF1 Gen           ..... ctagtgtga
CVF1 Gen           atgttacctt cttactgtaa agaataattg ttccacctc ttctcccctg

CVF1 cDNA 1925      CTGCAAAGTG TCCTCAGCCT GCAAATCGGA GCGTCGCAG 1964
CVF1 Gen           cctccttcag CTGCAAAGTG TCCTCAGCCT GCAAATCGGA GCGTCGCAG
                   |

```

Abb. 3.9: Teilsequenz von Intron 15 des CVF1-Gens. Als Matrize diente der Cosmid-Klon 1R67. Das PCR-Amplifikat mit der 3'-Teilsequenz von Intron 15 im Cosmid-Klon 1R67 hat eine Länge von ca. 3500 bp. Die EcoRI-Schnittstelle (gaattc) am 5'-Ende des Amplifikats ist unterstrichen dargestellt. Der Bereich der Überlappung mit obiger Sequenz ist kursiv dargestellt.

Durch Sequenzanalyse der PCR-Amplifikate konnte gezeigt werden, dass die 4,5 kb- und 3,6 kb-Fragmente direkt benachbart sind und dieser Genabschnitt keine weiteren EcoRI-Schnittstellen und Exons enthält. Die 5'-Region des Klons 1R67 mit dem ca. 7,5 kb Teilabschnitt von Intron 15 ist in Abbildung 3.6 schematisch dargestellt.

3'-Ende des CVF1-Klons 1R67

Die Sequenzinformation des 3'-Bereichs des CVF1-Klons 1R67 ist in dem EcoRI-Fragment von ca. 1000 bp Länge enthalten (Abb. 3.1), welches kloniert und sequenziert wurde. Dieses Fragment enthält einen Teil von Exon 27, ferner 938 bp der Intron 27-Sequenz sowie einen Teil des Cosmidvektors sCos-1. Die Sequenz ist in Abbildung 3.10 dargestellt.

| | | | |
|------------------|------|---|------|
| CVF1 cDNA | 3350 | <u>GAATTCAAGG</u> TGCTGAAGAA GAAGTATATT TAACAGCTTT CATTCTGGTT | 3399 |
| CVF1 Gen | | <u>GAATTCAAGG</u> TGCTGAAGAA GAAGTATATT TAACAGCTTT CATTCTGGTT | |
| CVF1 cDNA | 3400 | GCGTTGTTGG AATCCAAAAC AATCTGCAAT GACTATGTCA AT | 3441 |
| CVF1 Gen | | GCGTTGTTGG AATCCAAAAC AATCTGCAAT GACTATGTCA ATgtacgtga | |
| CVF1 Gen | | tgttgtgatg ggaaatggta tgagggggcc aaaataagat tacatctctc | |
| CVF1 Gen | | cacatagtat | |
| | | (794 bp) | |
| CVF1 Gen | | atctttaa | |
| CVF1 Gen | | atcaaagagc ttttcactgc tgcagtataa tttttcatct gcctcaaagg | |
| Vektor sCos1 | | CCTA TAGTGAGTCG TATTATGCGG CCGCGAATTC | |
| CVF1 (Gen) | | cagatttaga aagatcCCTA TAGTGAGTCG TATTATGCGG CCGCGAATTC | |

Abb. 3.10: Teilsequenz von Intron 27 des CVF1-Gens. Exon 27 hat eine Länge von 96 bp. Die Sequenz von Intron 27 im Cosmid-Klon 1R67 umfasst 938 bp. Die MboI-Klonierungsstelle (gac) ist unterstrichen dargestellt.

Die gezeigten Daten wurden zusätzlich durch Sequenzierung des Cosmidklons mit den Vektorprimern sCos1-T7as sowie sCos1-3'as vom 3'-Ende partiell bestätigt. Somit enthält der Cosmidklon 1R67 zwischen Exon 27 und dem Vektor am 3'-Ende 938 bp von Intron 27. Dieser Intronbereich enthält neben der MboI-Schnittstelle (Sequenz gac) am 3'-Ende des CVF1-Inserts noch vier weitere MboI-Schnittstellen, die beim partiellen Verdau genomischer Kobra-DNA für die Herstellung der Cosmid-Bibliothek intakt blieben (siehe Anhang 7.1).

Schematische Übersicht über das CVF1-Insert im Cosmidklon 1R67

Der Cosmidklon 1R67 hat eine Insertgröße von ca. 33 kb und enthält den Bereich von Exon 16-27 des CVF1-Gens. An den flankierenden Regionen enthält der Klon ca. 7,5 kb von Intron 15 am 5'-Ende sowie 938 bp von Intron 27 am 3'-Ende. Eine graphische Übersicht über die Positionen der Exons und Intronlängen befindet sich in Abbildung 3.11.

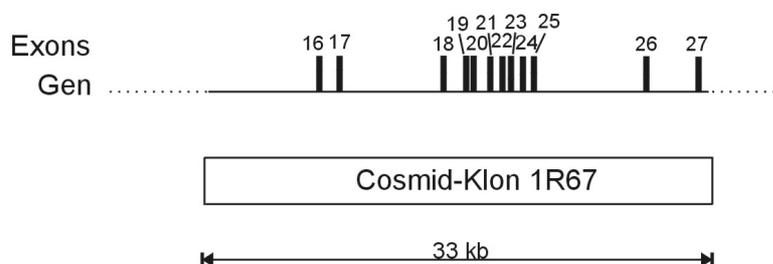


Abb. 3.11: Graphische Darstellung der CVF1-Genstruktur im Cosmid-Klon 1R67.

3.1.2 Identifizierung und Charakterisierung des Cosmid-Klons M26

Screening der Cosmid-Bibliothek

Für die Identifizierung weiterer CVF1-Klone aus der Cosmid-Bibliothek wurden Digoxigenin-markierte Sonden hergestellt. Die beiden neuen Sonden („Exon 12“ und „Intr.26/Ex.27“) waren so gewählt, dass damit die Isolierung von Klonen möglich sein sollte, welche die CVF1-Genabschnitte stromaufwärts und stromabwärts des Klons 1R67 tragen. Der Hybridisierungsbereich der Klone ist in Abbildung 3.12 dargestellt.

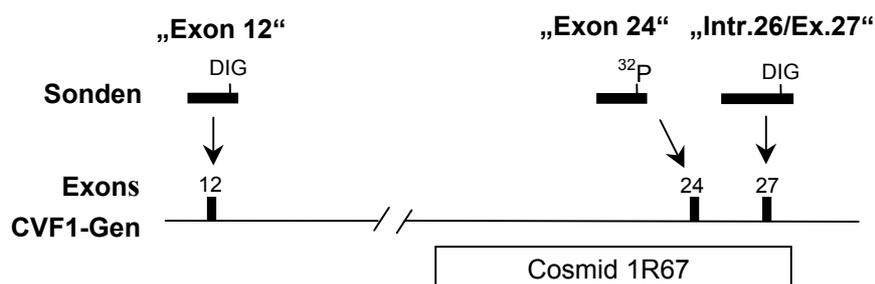


Abb. 3.12: Auswahl der Sonden zum Screening der Cosmid-Bibliothek nach CVF1-Klonen. Länge der Sonden und Art der Markierung: „Exon 12“ (210 bp, Digoxigenin-markiert), „Exon 24“ (204 bp, ^{32}P -markiert), „Intr.26/Ex.27“ (389 bp, Digoxigenin-markiert).

Nachfolgend ist exemplarisch die Sonde „Intr.26/Ex.27“, die später zum Auffinden eines positiven Signals führte, näher charakterisiert.

Auswahl und Herstellung der DIG-markierten Sonde „Intr.26/Ex.27“

Die Sonde „Intr.26/Ex.27“ wurde so ausgewählt, dass sie mit dem 3'-Bereich von 1R67 hybridisieren und somit zur Auffindung eines überlappenden Klons dienen sollte. Auf diese Weise konnte der Klon 1R67 beim Screening der Cosmid-Bibliothek als Positiv-Kontrolle eingesetzt werden. Zunächst wurde der entsprechende Bereich in einer PCR mit den Primern YLIntr.26s-CVF und YL27as-CVF sowie mit 1R67-DNA als Template amplifiziert. Das Produkt wurde gereinigt und als Template für die Sondenherstellung eingesetzt.

Die DNA-Sonde „Intr.26/Ex.27“ wurde mit oben genannten Primern unter Verwendung von DIG-dUTP hergestellt. In einer Kontrollreaktion wurde ein Amplifikat ohne DIG-Markierung erzeugt. Die Sonde mit einer Länge von 389 bp enthält 288 bp der Intron 26-Sequenz, die komplette Exon 27-Sequenz mit 96 bp sowie 5 bp von Intron 27. Diese Sonde wurde fast doppelt so groß gewählt wie die P-32 markierte Sonde „Exon 24“, um durch den Einbau einer größeren Anzahl von Digoxigenin-Moleküle ein verstärktes Signal zu erreichen. Zudem enthält diese Sonde zu mehr als 2/3 Intronsequenzen, um spezifisch weitere CVF1-Klone zu detektieren. Bei einer Sonde mit alleiniger Exonsequenz wäre aufgrund der hohen Homologie der cDNAs von CVF1, CVF2 und coC3 die Wahrscheinlichkeit gegeben, Cosmid-Klone der verwandten Gene zu isolieren.

Die DIG-markierte Sonde „Intr.26/Ex.27“ zeigte durch den Einbau von Digoxigenin ein anderes Laufverhalten als das Kontroll-PCR-Produkt (ohne Abbildung). Die Sonden-Konzentration wurde photometrisch und im Spot-Test durch Vergleich mit einem DNA-Standard ermittelt. Für die Hybridisierung wurde die Sonde in einer Konzentration von ca. 15 ng/ml eingesetzt.

Identifizierung des Klons M26

Die Cosmid-Bibliothek der Kobra *Naja kaouthia*, für die nach der Herstellung eine Diversität von $4-5 \cdot 10^5$ Klonen bestimmt worden war (Kunze, unveröffentlichte Ergebnisse), wurde mit der DIG-Sonde „Intr.26/Ex.27“ nach CVF-Klonen durchmustert (Li 2000). Dazu wurde die Cosmid-Bibliothek verdünnt ausplattiert, so dass eine Dichte von 2000 Klonen pro Agarplatte (24 x 24 cm) erhalten wurde. Nach Inkubation der Klone wurden Abdrücke mit Nylonmembranen (*colony lift*) genommen und diese mit der Sonde hybridisiert. Nach dem Screening von ca. $2,5 \cdot 10^5$ Klonen wurden auf einer Platte positive Signale erhalten (Abb. 3.13).

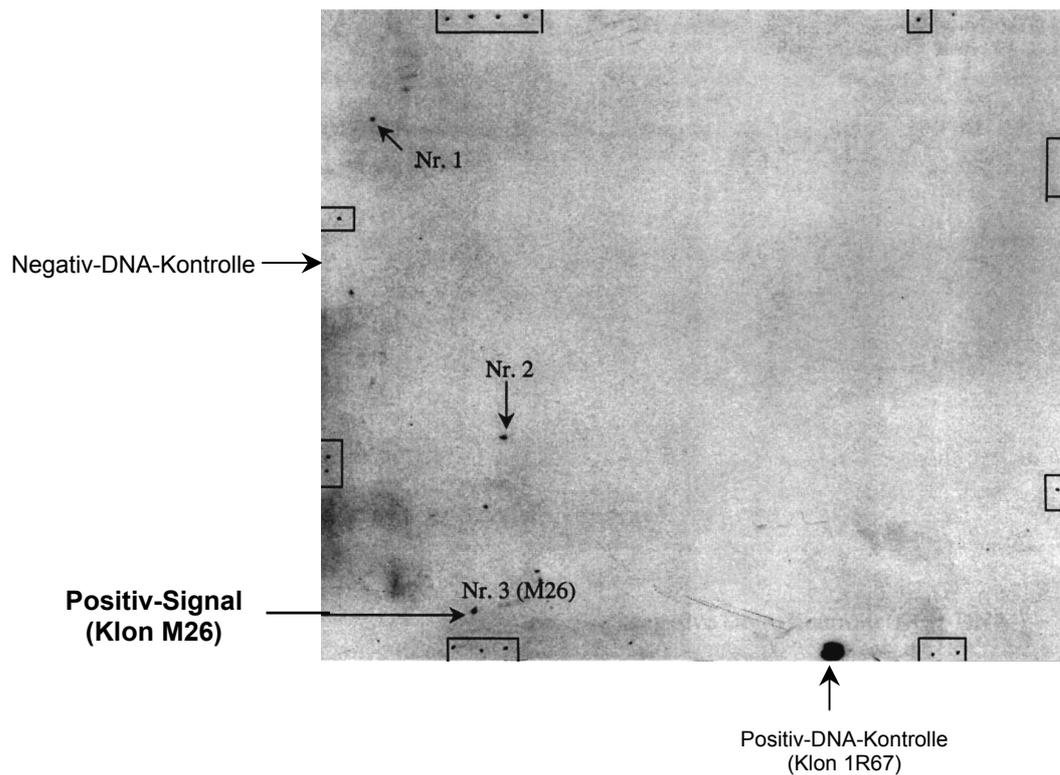


Abb. 3.13: Screening der Cosmid-Bibliothek und Detektion positiver Signale. Film der Hybond-NX-Membran, die mit der Sonde „Intr.26/Ex.27“ hybridisiert wurde. Der Film wurde für 60 min belichtet. Die positiven Signale der Klone Nr. 1, Nr. 2 und Nr. 3 (Klon M26) sind mit Pfeilen gekennzeichnet. Die Original-Platte hatte eine Größe von 24 x 24 cm. Orientierungsmarken sind umrahmt dargestellt.

Nach der Auswertung des Films wurden die Signal-gebenden Klone auf der Platte gepickt und weiterführend mit den Methoden Dot-Blot, Subscreening und PCR analysiert.

Dot-Blot

Für den Dot-Blot wurde DNA der Klone Nr. 1, Nr. 2 und Nr. 3 (vgl. Abb. 3.13) aus Übernachtskulturen isoliert, auf eine Hybond NX-Membran aufgetüpfelt und mit der Sonde „Intr.26/Ex.27“ hybridisiert (Abb. 3.14). Als Positiv-Kontrolle diente 1R67-DNA.

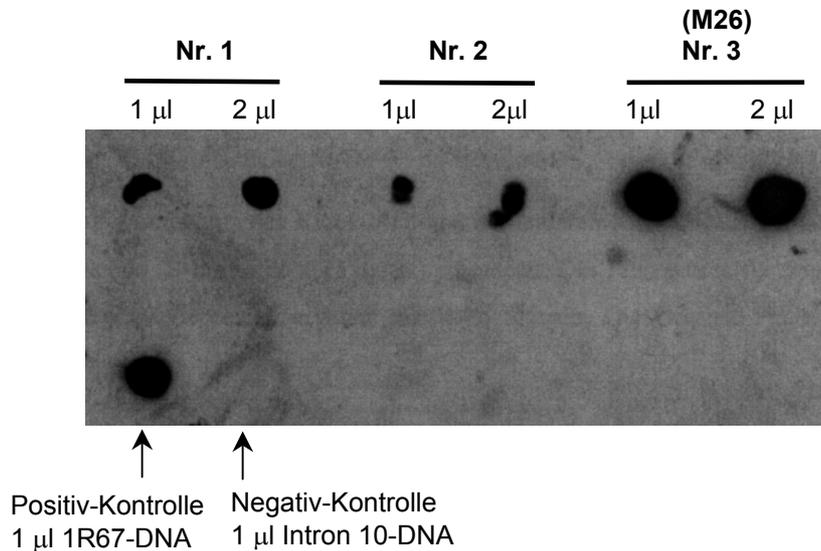


Abb. 3.14: Dot-Blot zur Identifizierung von CVF-Klonen. Auf einer Hybond NX-Membran wurden jeweils 1 µl bzw. 2 µl Cosmid-DNA der Klone Nr. 1, Nr. 2 und Nr. 3 sowie jeweils 1 µl Positiv- bzw. Negativ-Kontroll-DNA aufgetüpfelt und mit der Sonde „Intr.26/Ex.27“ hybridisiert. Der Film wurde für 15 min belichtet. Die Probe Nr. 3 (M26) zeigt die stärksten Signale.

Subscreening

Es wurden Übernachtskulturen der Klone Nr. 1, Nr. 2 und Nr. 3 im Verhältnis 1:10⁶ in LB-Medium verdünnt, und 40 µl davon wurden in einem Drittel einer Agarplatte ausplattiert. Nach Inkubation der Platte bei 37°C über Nacht wurden 30-40 Einzelkolonien erhalten. Ein Membranabdruck dieser Platte wurde erneut bei 37°C inkubiert und wie unter Kapitel 2 beschrieben behandelt. Nach Auftüpfeln von 1R67-DNA als Positiv-Kontrolle wurde die Membran mit der Sonde „Intr.26/Ex.27“ hybridisiert. Auf dem belichteten Film in Abbildung 3.15 sind deutlich die Signale der Kolonien von M26 zu erkennen, während für die Kolonien von Nr. 1 und Nr. 2 nur schwache Signale erhalten wurden.

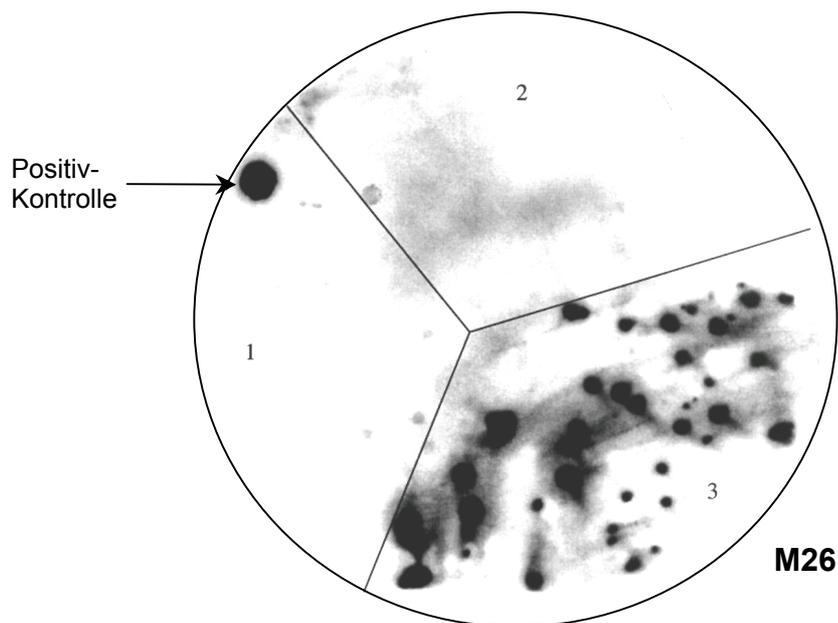


Abb. 3.15: Subscreening der Klone Nr. 1, Nr. 2 und Nr. 3 (M26). Für das Screening der Kolonien wurde die Sonde „Intr.26/Ex.27“ eingesetzt. Der Film wurde für 5 min belichtet. Die Kolonien von M26 im Bereich 3 ergaben deutliche Signale.

Die Ergebnisse des Dot-Blots und des Subscreenings deuteten darauf hin, dass M26 ein positiver CVF-Klon ist, während die beiden Klone Nr. 1 und Nr. 2 offenbar falsch-positiv sind, obwohl sie im Dot-Blot signifikante Signale zeigten.

Charakterisierung des Klons M26

Die detailliertere Charakterisierung des Klons M26 sollte mit verschiedenen Ansätzen vorgenommen werden. Dazu wurden die Analysemethoden PCR, Restriktionsanalyse, Subscreening sowie Southern-Blot eingesetzt.

PCR-Analyse

Zunächst sollte überprüft werden, ob der zur Sonde „Intr.26/Ex.27“ komplementäre Bereich tatsächlich im Klon M26 vorhanden ist. Die Analyse erfolgte mittels PCR unter Verwendung der Primer YLIntr.26s-CVF und YL27as-CVF, die für die Sonden-Herstellung verwendet worden waren, wobei Kontroll-Reaktionen mit 1R67-DNA als Template (Positiv-Kontrolle) sowie ohne Template (Negativ-Kontrolle) durchgeführt wurden. Für beide Cosmidklone wurde ein deutliches Amplifikat in der erwarteten Größe (389 bp) erhalten (ohne Abbildung). Auch die Amplifikation von weiteren Introns führte in einigen Fällen zu Produkten, welche in ihrer Größe und Sequenz mit denen der Positivkontrolle

1R67 identisch waren. Demzufolge schien es sich bei dem Klon M26 um einen CVF1-Klon zu handeln, der partiell mit dem Klon 1R67 überlappt.

Restriktionsanalyse

Die Restriktionsanalyse wurde genutzt, um durch Vergleich der Restriktionsmuster beider Klone mögliche Überlappungen bzw. Unterschiede nachweisen sowie die Insertgröße des Klons M26 bestimmen zu können. Dazu wurden die DNAs von M26 und 1R67 mit den Restriktionsenzymen EcoRI, NotI, NdeI und HindIII verdaut (Abb. 3.16 und 3.17A) und die Restriktionsmuster beider Klone miteinander verglichen.

Insbesondere das EcoRI-Restriktionsmuster (Abb. 3.16) zeigt für beide Klone große Übereinstimmung. Neben der Vektorbande (6,7 kb) scheinen noch die Fragmente bei 4,5 kb, 3,9 kb und 2,6 kb von gleicher Größe zu sein, was auf eine Überlappung beider Klone hinwies.

Eine Addition aller EcoRI-Fragmentlängen von M26 ergibt eine Insertgröße von ca. 42 kb.

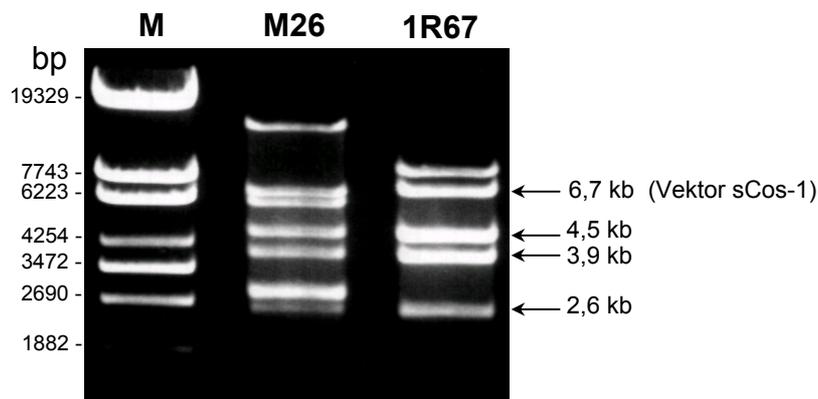


Abb. 3.16: Restriktionsanalyse der Klone M26 und 1R67. Die DNA der Klone M26 und 1R67 wurde mit EcoRI verdaut und über ein 1%iges Agarosegel gelelektrophoretisch aufgetrennt. Die mit Pfeilen markierten Fragmente scheinen jeweils identische Größen zu haben. M: DNA-Marker λ /Eco130I.

Beim NotI-Verdau (Abb. 3.17) zeigen beide Klone ein identisches Muster, wobei das Insert (>20 kb) als singuläre Bande aus dem Vektor (6.7 kb) herausgeschnitten wurde. Beim NdeI- und HindIII-Verdau (Abb. 3.17) sind deutlich unterschiedliche Muster erkennbar. Dass hier keine Banden gleicher Größe erkennbar waren, scheint im Widerspruch zum EcoRI-Muster zu stehen.

Southern-Blot

Die Southern-Blot-Analyse bot die Möglichkeit, die Fragmente zu identifizieren, mit denen die Sonde hybridisierte, sowie die Hybridisierungsmuster der Klone M26 und 1R67 miteinander zu vergleichen. Dazu wurde die DNA beider Klone mit den Restriktionsenzymen NotI, NdeI und HindIII verdaut und die Fragmente im Anschluss an eine Größenauftrennung und Immobilisierung auf einer Membran mit der Sonde „Intr.26/Ex.27“ hybridisiert (Abb. 3.17).

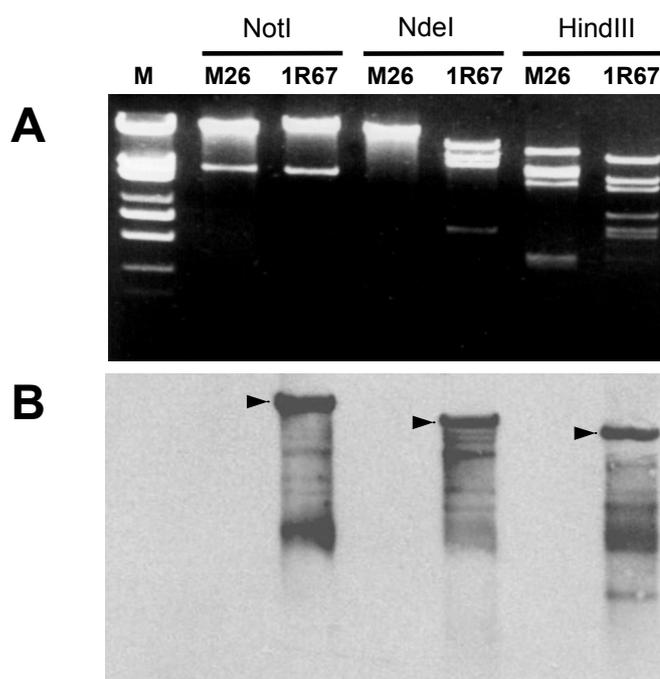


Abb. 3.17: Charakterisierung der Klone M26 und 1R67. **A:** Die DNA der Klone M26 und 1R67 wurde mit NotI, NdeI und HindIII verdaut und über ein 1%iges Agarosegel elektrophoretisch aufgetrennt. M: DNA-Marker λ /Eco130I. **B:** Southern-Blot des unter A abgebildeten Gels. Für die Hybridisierung wurde die DIG-markierte Sonde „Intr.26/Ex.27“ eingesetzt. Trotz Waschschritte mit geringer Stringenz wurden nur für den Klon 1R67 Hybridisierungs-Signale detektiert. Die spezifischen Hybridisierungs-Signale sind mit Pfeilen markiert.

Hybridisierungssignale wurden nur für den Klon 1R67 erhalten, obwohl die Waschschritte mit geringer Stringenz durchgeführt wurden, um dadurch auch die Hybridisierung mit solchen Genabschnitten zu ermöglichen, die mit dem CVF1-Abschnitt der Sonde homolog sind.

Eine Homologie der M26-Sequenzen zu den bisher bekannten genomischen und cDNA-Sequenzen der CVF/C3-Familie konnte nicht festgestellt werden. Es ist insbesondere aufgrund fehlender Signale im Southern-Blot davon auszugehen, dass es sich bei M26 um

einen falsch-positiven Klon handelt. Die PCR-Amplifikate sind möglicherweise auf Verunreinigungen mit 1R67-DNA zurückzuführen.

Da das Screening der Cosmid-Bibliothek nicht zur Identifizierung neuer CVF1-Klone führte, wurde die Struktur des CVF1-Gens mittels alternativer Methoden (genomische PCR und Genome Walking) aufgeklärt.

3.1.3 Aufklärung der CVF1-Genstruktur durch genomische PCR

Als alternative Methode zur Aufklärung der CVF1-Genstruktur wurde genomische PCR eingesetzt. Um die Anwendbarkeit dieser Methode auf die konkrete Fragestellung zu überprüfen, wurde zunächst exemplarisch die Amplifikation einiger Introns untersucht, deren Größe bereits durch Aufklärung des Klons 1R67 bekannt war.

Dazu wurden in jeweils 2 Ansätzen PCR-Reaktionen mit 1R67-DNA bzw. genomischer DNA als Template durchgeführt. Als Primer für die Amplifikation der Introns 16, 21, 22 und 24 wurden die bereits in Kapitel 3.1.1 beschriebenen Primerpaare eingesetzt. In Abbildung 3.18 ist nach gelelektrophoretischer Analyse der PCR-Produkte zu erkennen, dass in allen Ansätzen die erwartete CVF1-Bande vorhanden ist. Bei der genomischen Amplifikation von Intron 16 wurde zusätzlich zum CVF1-Amplifikat (1,5 kb) ein weiteres Produkt von 903 bp gebildet, welches das entsprechende coC3-Intron enthält (vgl. Kapitel 3.3.1). Das Entstehen eines spezifischen Co-Amplifikats anderer Größe ist ein Hinweis auf unterschiedliche Introngrößen innerhalb der CVF/C3-Genfamilie.

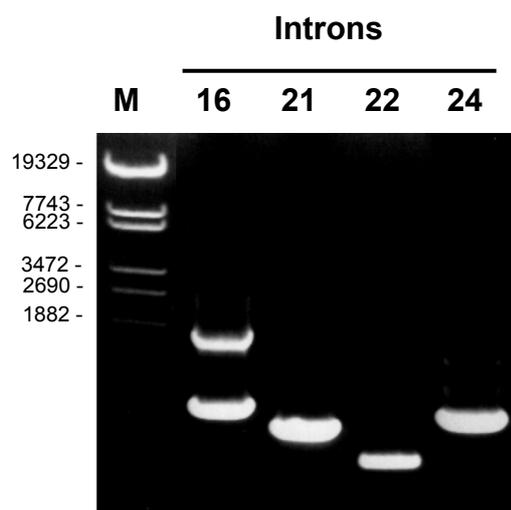


Abb. 3.18: Gelelektrophoretische Analyse ausgewählter Intron-Amplifikate nach genomischer PCR (1%iges Agarosegel). M: DNA-Marker λ /Eco130I, Intron 16: 1500 bp (CVF1), 903 bp (coC3); Intron 21: 722 bp (CVF1); Intron 22: 444 bp (CVF1); Intron 24: 827 bp (CVF1).

Diese Ergebnisse zeigen, dass die Methode für die oben beschriebene Fragestellung grundsätzlich geeignet ist.

Für die Aufklärung der Exon/Intron-Übergänge und zur Bestimmung der Intron-Größen des CVF1-Gens sollte daher genomische PCR durchgeführt werden. Da außer dem CVF1-Gen noch die homologen Gene CVF2 und Cobra C3 im Genom der Kobra *Naja kaouthia* vorhanden sind, wurden für die gezielte Amplifikation von CVF1-spezifischen Produkten die Primer so gewählt, dass sie entweder CVF1-spezifische Intronsequenz enthalten, oder dass die Exon-Primer CVF1-spezifisch sind.

Die Exonstrukturen der Gene CVF1, CVF2 und coC3 wurden erneut in Analogie zur humanen C3-Genstruktur vorhergesagt. In Abb. 3.19 ist der Ausschnitt aus den cDNA-Sequenzen vom Startcodon (Exon 1) bis Exon 15 mit den jeweiligen Exongrenzen dargestellt. Für CVF2 liegt die cDNA-Sequenz erst ab Exon 10 vor. Die exonspezifischen Primer wurden meist so gewählt, dass sie CVF1-spezifisch binden und Fehlpaarungen zu den homologen Sequenzen CVF2 und coC3 insbesondere am 3'-Ende aufweisen (ARMS-Methode, *amplification refractory mutation system*). Auf diese Weise sollte gewährleistet sein, dass CVF1-Amplifikate bevorzugt gebildet werden. Die PCR-Produkte wurden gereinigt, kloniert und sequenziert. Um die Spezifität der sequenzierten PCR-Amplifikate zu demonstrieren, sind in den entsprechenden Abbildungen die homologen cDNA-Sequenzen der entsprechenden Gene zum Vergleich mit angeordnet.

| | | |
|----------------|---|------|
| Exon 1 | | |
| CVF1 | ...ATGGAGAGGATGGCTCTCTATCTGGTGGCTGCTCTATTGATTGGTTTTCCAGGGTCT | 60 |
| coC3 | ...ATGGAGGGATGGCTCTCTATCTGGTGGCTGCTCTATTGATTGGTTTTCCAGGGTCT | 65 |
| Exon 2 | | |
| CVF1 | TCTCATGGGGCTCTCTACACCCCTCATCACCCCTGCTGTTTTGCGAACAGACACAGAAGAG | 120 |
| coC3 | TCACAGGGGGCTCTCTACACCCCTCATCACCCCTGCTGTTTTGCGAACAGACACAGAAGAG | 125 |
| CVF1 | CAAATTTTGGTGGAGGCCCATGGAGACAGTACTCCAAAACAGCTTGACATCTTTGTTTCAT | 180 |
| coC3 | CAAATTTTGGTGGAGGCCCATGGAGACAGTACTCCAAAATCGCTTGACATCTTTGTTTCAT | 185 |
| CVF1 | GATTTTCCACGGAAGCAGAAAACCTTGTTCAAAACCAGAGTAGATATGAATCCAGCAGGA | 240 |
| coC3 | GATTTTCCACGGAAGCAGAAAACCTTGTTCAAAACGAGAGTAGATATGAATCCAGCAGGA | 245 |
| Exon 3 | | |
| CVF1 | GGCATGCTTGTCACTCCAACCTATAGAGATTCCAGCAAAAAGAAGTGAGTACGGACTCCAGG | 300 |
| coC3 | AGCATGTTTTGTCACTCCAACCTATAAAGGTTCCCGCAAAAAGAACTGAATAAGGACTCCAAG | 305 |
| CVF1 | CAAAATCAATATGTGGTTGTGCAAGTAACGGTCCTCAAGTGAGATTGGAAAAGGTGGTT | 360 |
| coC3 | CAAAATCAATATGTGGTTGTGCAAGTAACGGTCCTCAAGTGCCATTGGAAAAGGTGGTT | 365 |
| CVF1 | CTCCTTTCTTACCAGAGTAGCTTTCTGTTTTATCCAGACAGATAAAAGGCATCTATACACCA | 420 |
| coC3 | CTCCTTTCTTACCAGAGTGGCTTTCTGTTTCATCCAGACAGATAAAAGGCATCTATACACCA | 425 |
| Exon 4 | | |
| CVF1 | GGGTCTCCAGTACTCTATCGTGTTTTTTCTATGGATCACAACACAAGCAAGATGAACAAA | 480 |
| coC3 | GGCTCTCCAGTGGCTTATCGTGTCTTTTTCTGGATCACAACATGCACAGGATGCACAAA | 485 |
| Exon 5 | | |
| CVF1 | ACTGTGATTGTTGAGTTTCAGACTCCAGAAGGCATTCTTGTGAGTTCTAATTCAGTTGAC | 540 |
| coC3 | ACTGTGATTGTTGAGTTTCAGACTCCAGAAGGCATTCTTGTGAGTTCTAATTCAGTTCAAT | 545 |
| Exon 6 | | |
| CVF1 | CTA---AACTTCTTCTGGCCTTACAATTTACCAGACCTTGTGAGTTTGGGGACTTGGAGG | 597 |
| coC3 | CCATCAAGGCTCCATCCGGCCTTACAATTTACCAGAGCTTGTGAGTTTGGGACATGGAAAG | 605 |
| CVF1 | ATTGTGGCCAAATATGAACATTCCCCAGA---GAATTATACTGCATATTTTGTGTCAGG | 654 |
| coC3 | GGTGTGGCCAAATATGAACATTCCACAGAAGAGAAGCTACACTGCATATTTTGTGTCAGG | 665 |
| Exon 7 | | |
| CVF1 | AAATATGTTGTTGCCAAGCTTTGAAGTCCGTCTGCAACCATCAGAGAAGTTTTTTTACATT | 714 |
| coC3 | CAATATGTTGTTGCCAAGCTTTGAAGTCCGTCTGCAACCATCAGATAAAGTTTTCTTACATT | 725 |
| Exon 8 | | |
| CVF1 | GACGGCAATGAAAATTTCCACGTGTCTATCACTGCAAGGTACTTGTATGGAGAGGAAGTG | 774 |
| coC3 | GATGGCAATGAAAATTTCCACGTGTCTATCACTGCAAGGTACTTGTATGGAAAGGAAGTG | 785 |
| CVF1 | GAAGGTGTGGCCTTTGTCCCTTTGGAGTGAAAATAGATGATGCTAAAAAGAGTATTCCA | 834 |
| coC3 | GAAGGTGTGGCCTTTGTCCCTTTGGAGTCAAAAATAGATGATGCTAAAAAGAGTATTCCA | 845 |
| Exon 9 | | |
| CVF1 | GACTCACTCACGAGAATTCGATTATTGATGGAGATGGGAAAGCAACACTAAAAAGAGAT | 894 |
| coC3 | GACTCACTCACGAGAATTCGATTATTGATGGAGATGGGCAAGCAACACTAAAAAGAGAT | 905 |
| CVF1 | ACATTCCGTTCTCGATTTCCAAATCTCAATGAGCTTGTGGGCATACTCTGTATGCATCT | 954 |
| coC3 | ACACTACGTTCCCGATTTCAACATCTCAATCAGCTTGTGGTCATACTCTGTATGTTATCT | 965 |
| Exon 10 | | |
| CVF1 | GTAACAGTCATGACAGAATCAGGCAGTGATATGGTAGTACTGAGCAAAGCGGCATTTCAT | 1014 |
| CVF2 |ATCAGGAGGTGATATGGTAAATGACTGAGCAAAGTGGCATTTCAT | 50 |
| coC3 | GTAACAGTCATGACAGAATCAGGCAGTGATATGGTAGTACTGAGCAAAGCGGCATTTCAT | 1025 |

Fortsetzung

CVF1 ATTGTGGCATCTCCCTATCAGATCCACTTCACAAAAACCCCAAAATATTTCAAGCCAGGA 1074
 CVF2 ATTGTGACATCTCCCTATCAGATCTACTTCACAAAAACCCCAAAATATTTCAAGCCAGGA 110
 coC3 ATTGTGACATCTCCCTATCAGATCTACTTCACAAAAACCCCAAAATATTTCAAGCCAGGA 1085

Exon 11

CVF1 ATGCCATATGAAC TGACG GTGTATGTTACCAACCTGATGGCTCACCAGCTGCCCATGTG 1134
 CVF2 ATGCCATATGAAC TGACG GTGTATGTTACCAACCTGATGGCTCACCAGCTGCCCATGTG 170
 coC3 ATGCCATATGAAC TGACG GTGTATGTTACCAACCTGATGGCTCACCAGCTGCCCATGTG 1145

CVF1 CCAGTGGTATCAGAGGCC TTTTCATTCTATGGGAACCACTTTGAGTGATGGGACTGCTAAG 1194
 CVF2 CCAGTGGTATCAGAGGCC TTTTCATTCTATGGGAACCACTTTGAGTGATGGGACTGCTAAG 230
 coC3 CCAGTGGTATCAGAGGCC TTTTCATTCTATGGGAACCACTTTGAGTGATGGGACTGCTAAG 1205

Exon 12

CVF1 CTCATCCTGAACATACCATTGAATGCTCAAAGCCTACCAATCACT GTTAGAACTAACCAT 1254
 CVF2 CTCATCCTGAACATACCATTGAATGCTCAAAGCCTACCAATCACT GTTAGAACTAACCAT 290
 coC3 CTCATCCTGAACATACCATTGAATGCTCAAAGCCTACCAATCACT GTTAGAACTAACCAT 1265

CVF1 GGAGACCTCCCAAGAGAACGCCAGGCAACAAAGTCCATGACAGCCATAGCCTACCAAACC 1314
 CVF2 GGAGACCTCCCAAGAGAACGCCAGGCAACAAAGTCCATGACAGCCATAGCCTACCAAACC 350
 coC3 GGAGACCTCCCAAGAGAACGCCAGGCAACAAAGTCCATGACAGCCATAGCCTACCAAACC 1325

CVF1 CAGGGAGGATCTGGAAACTATCTTCATGTAGCCATTACATCTACAGAGATTAAGCCCGGA 1374
 CVF2 CAGGGAGGATCTGGAAACTATCTTCATGTAGCCATTACATCTACAGAGATTAAGCCCGGA 410
 coC3 CAGGGAGGATCTGGAAACTATCTTCATGTAGCCATTACATCTACAGAGATTAAGCCCGGA 1385

CVF1 GATAACTTACCTGTCAATTTCAATGTGAAGGGCAATGCAAATTCCTGAAGCAGATCAAA 1434
 CVF2 GATAACTTACCTGTCAATTTCAATGTGAAGGGCAATGCAAATTCCTGAAGCAGATCAAA 470
 coC3 GATAACTTACCTGTCAATTTCAATGTGAAGGGCAATGCAAATTCCTGAAGCAGATCAAA 1445

Exon 13

CVF1 TATTTACATACCTCATATTGAATAAAGGGAAGATTTTCAAGGTTGGCAGGCAACCCAGG 1494
 CVF2 TATTTACATACCTCATATTGAATAAAGGGAAGATTTTCAAGGTTGGCAGGCAACCCAGG 530
 coC3 TATTTACATACCTCATATTGAATAAAGGGAAGATTTTCAAGGTTGGCAGGCAACCCAGG 1505

CVF1 AGAGATGGGCAGAATCTGGTGACCATGAATCTGCATATCACTCCAGATCTCATCCCTTCC 1554
 CVF2 AGAGATGGGCAGAATCTGGTGACCATGAATCTGCATATCACTCCAGATCTCATCCCTTCC 590
 coC3 AGAGATGGGCAGAATCTGGTGACCATGAATCTGCATATCACTCCAGATCTCATCCCTTCC 1565

CVF1 TTCCGGTTTGTGGCTTACTACCAAGTGGGAAACAACGAAATTGTGGCTGATTCTGTCTGG 1614
 CVF2 TTCCGGTTTGTGGCTTACTACCAAGTGGGAAACAACGAAATTGTGGCTGATTCTGTCTGG 650
 coC3 TTCCGGTTTGTGGCTTACTACCAAGTGGGAAACAACGAAATTGTGGCTGATTCTGTCTGG 1625

Exon 14

CVF1 GTGGATGTGAAGGATACCTGCATGGGAACCTTGGTTGTGAAAGGAG-----AC 1662
 CVF2 GTGGATGTGAAGGATACCTGCATGGGAACCTTGGTTGTGAAAGGAGCGACTTCCAGAGAC 710
 coC3 GTGGATGTGAAGGATACCTGCATGGGAACCTTGGTTGTGAAAGGAGCGACTTCCAGAGAC 1685

CVF1 AATCTAATACAAAATGCCAGGAGCTGCAATGAAAATCAAATTGGAAGGGGATCCAGGTGCT 1722
 CVF2 AATCTAATACAAAATGCCAGGAGCTGCAATGAAAATCAAATTGGAAGGGGATCCAGGTGCT 770
 coC3 AATCTAATACAAAATGCCAGGAGCTGCAATGAAAATCAAATTGGAAGGGGATCCAGGTGCT 1745

CVF1 CGGGTTGGTCTTGTGGCTGTGGACAAAGCAGTATATGTTCTCAATGATAAATATAAGATT 1782
 CVF2 CGGGTTGGTCTTGTGGCTGTGGACAAAGCAGTATATGTTCTCAATGATAAATATAAGATT 830
 coC3 CGGGTTGGTCTTGTGGCTGTGGACAAAGCAGTATATGTTCTCAATGATAAATATAAGATT 1805

Exon 15

CVF1 AGCCAAGCTAAGATATGGGACACAATAGAAAAGAGTGACTTTGGCTGTACAGCTGGCAGT 1842
 CVF2 AGCCAAGCTAAGATATGGGACACAATAGAAAAGAGTGACTTTGGCTGTACAGCTGGCAGT 890
 coC3 AGCCAAGCTAAGATATGGGACACAATAGAAAAGAGTGACTTTGGCTGTACAGCTGGCAGT 1865

Fortsetzung

| | | |
|------|--|------|
| CVF1 | GGCCAGAATAATCTGGGTGTGTTTGAAGATGCTGGACTGGCTCTGACAACCAGCACTAAT | 1902 |
| CVF2 | GGCCAGAATAATCTGGGTGTGTTTGAAGATGCTGGACTGGCTCTGACAACCAGCACTAAT | 950 |
| coC3 | GGCCAGAATAATCTGGGTGTGTTTGAAGATGCTGGACTGGCTCTGACAACCAGCACTAAT | 1925 |

| | | |
|------|------------------------------|------|
| CVF1 | CTCAACACCAAACAGAGATCAGCTG... | 1927 |
| CVF2 | CTCAACACCAAACAGAGATCAGCTG... | 975 |
| coC3 | CTCAACACCAAACAGAGATCAGCTG... | 1950 |

Abb. 3.19: Darstellung der Exonstrukturen von CVF1, CVF2 und coC3 (Exons 1-15). Die cDNA-Sequenzen sind ab dem Start-Codon ATG angeordnet. Für CVF2 sind die cDNA-Sequenzen erst ab Exon 10 bekannt. Die Nummerierung ist auf die cDNA-Sequenzen bezogen. Die Exongrenzen wurden in Analogie zur humanen C3-Genstruktur vorhergesagt und für CVF1 vollständig sowie für CVF2 und coC3 exemplarisch bestätigt.

In der Folge wurden mittels genomischer PCR diejenigen CVF1-Introns amplifiziert, die nicht durch den Klon 1R67 abgedeckt sind. Die Introns, soweit sie amplifiziert werden konnten, sind beginnend vom 5'-Bereich des CVF1-Gens aufgeführt.

Die erhaltenen Amplifikate der Introns 1-14 sind in Abbildung 3.20 dargestellt.

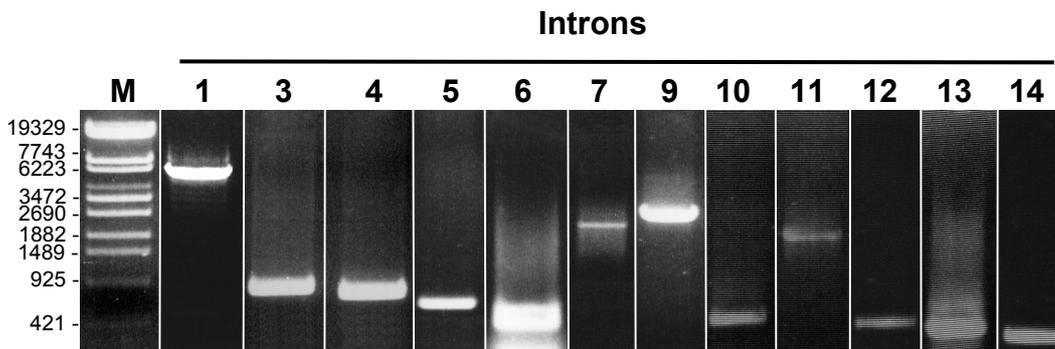


Abb. 3.20: Gelelektrophoretische Analyse genomischer Amplifikate mit Introns 1-14 des CVF1-Gens (1%iges Agarosegel). M: Marker λ -DNA/Eco130I, Intron 1: 6000 bp, Intron 3: 901 bp, Intron 4: 785 bp, Intron 5: 680 bp, Intron 6: 642 bp, Intron 7: 2500 bp, Intron 9: 3000 bp, Intron 10: 560 bp, Intron 11: 1900 bp, Intron 12: 456 bp, Intron 13: 427 bp, Intron 14: 308 bp.

Die für die Amplifikation eingesetzten Primerpaare sind in Tabelle 3.2 aufgelistet. Introns 2 und 8 des CVF1-Gens ließen sich durch genomische PCR nicht amplifizieren, was auf ausgedehnte Intronlängen hinwies. Die Aufklärung der entsprechenden Exon/Intron-Übergänge erfolgte durch Genome Walking (Kapitel 3.1.3).

Tab. 3.2: Übersicht der genomischen CVF1-Amplifikate im 5'-Bereich des CVF1-Gens (Introns 1-14). Mit Hilfe exonspezifischer Primer wurden Introns 1-14 amplifiziert. Größenangaben in Klammern sind Schätzungen.

| Intron | Primerpaar | Größe PCR-Produkt (bp) | Exon-Intron-Übergangssequenzen |
|--------|------------------------|------------------------|--|
| 1 | HB1s-CVF / HB2asII-CVF | (6.000) | TCT gtaagtatca...tcttctgcag CTA |
| 3 | HB3s-CVF / HB4as-CVF | 901 | CAG gtaaccogtt...tctttcttag TAC |
| 4 | HB4s-CVF / HB5as-CVF | 785 | CAG gtgagaatcg...cttatcccag ACT |
| 5 | HB5s-CVF / HB6as-CVF | 680 | CAG gtaatttaac...cttgctgtag TTT |
| 6 | HB6s / HB7as | 642 | ATG gtaagaagga...ttttctgtag TGT |
| 7 | HB7s-CVF / HB8as | (2500) | AAG gtatgcttat...gtcattacag GTA |
| 9 | HB9s / HB10as | (3.000) | CAG gtgagcaggg...tgttttgcag GCA |
| 10 | HB10s / HB11as | 560 | ACG gtatagaaat...gttgtttag GTG |
| 11 | HB11s / HB12as | (1900) | ACT gtaagtcaact...ccctatctag GTT |
| 12 | YL12s / YL13as | 456 | CTC gtaagtatta...tcaccctcag ATA |
| 13 | HB13s / HB14as | 427 | ACG gtaagccagg...gtataaacag TTG |
| 14 | AJS01 / OPA19 | 308 | AAG gtaaggccag...tgtgcactag ATA |

Am Beispiel ausgewählter Introns soll exemplarisch verdeutlicht werden, dass bei geeigneter Primer-Wahl eine Zuordnung der genomischen Sequenz zum entsprechenden Gen (CVF1, CVF2 oder coC3) aufgrund der charakteristischen Sequenzunterschiede möglich ist.

Intron 1

Eine Amplifikation von Intron 1 war zunächst unter Verwendung diverser DNA-Polymerasen (Taq und Taq/Pwo) und variierten PCR-Bedingungen nicht möglich. Erst durch die spezielle Taq/Tgo-DNA-Polymerase-Mischung *Expand 20kbPLUS* wurde mit dem Primerpaar HB1s-CVF/HB2asII-CVF ein singuläres Amplifikat mit einer Länge von ca. 6 kb erzielt (PCR-Programm siehe Anhang 7.6).

Polymerase-Mischungen wie Taq/Pwo und Taq/Tgo verfügen über eine 3'-5'-Exonuklease-Aktivität (Korrekturaktivität), die zu einer höheren Lesegenauigkeit führt und die Amplifikation längerer Sequenzen ermöglicht.

Die Sequenz des erhaltenen Amplifikats ist in Abbildung 3.21 dargestellt.

| | | | | | | | | |
|-----------|-----|-------------------------|------------|------------|-------------|------------|-----|--|
| CVF1 cDNA | 27 | GGTGGCTGCT | CTATTGATTG | GTTTTCCAGG | GTC TTCTCAT | GGGGCTCT | 74 | |
| CVF2 Gen | | GGTGGCTGCT | CTATTGATTG | GTTTTCCAGG | ATCTTCTCAT | GGGACTCT | | |
| coC3 cDNA | 32 | GGTGGCTGCT | CTATTGATTG | GTTTTCCAGG | GTCTTC CAC | GGGGCTCT | 79 | |
| CVF1 Gen | | GGTGGCTGCT | CTATTGATTG | GTTTTCCAGG | GTC TTCTCAT | GGGGCTCTgt | | |
| | | → | | | | | | |
| CVF1 Gen | | aagtatcata | gatatttcag | cagggggaat | gaagcagtgc | tggggaagcg | | |
| CVF1 Gen | | tagtgggctt | | | | | | |
| | |(ca. 5650 bp)..... | | | | | | |
| CVF1 Gen | | | | | | tggagatgcc | | |
| CVF1 Gen | | tcatgaaatg | cctcactttt | cctaagcaaa | ctaactcagg | agatgttctt | | |
| CVF1 cDNA | 75 | | CTACAC | CCTCATCACC | CCTGCTGTTT | TGCGAACAGA | 110 | |
| CVF2 Gen | | | CTACAC | CCTATCACC | CCTGATGTT | TGCGAACAGA | | |
| coC3 cDNA | 80 | | CTATAC | CCTCATCACC | CCTGCTGTTT | TGCGAACAGA | 115 | |
| CVF1 Gen | | ttctttcttct | gcagCTACAC | CCTCATCACC | CCTGCTGTTT | TGCGAACAGA | | |
| CVF1 cDNA | 111 | CACAGAAGAG | CAAATTTTGG | TGGAGGCCA | TGGAGACAGT | ACTCCAAAAC | 160 | |
| CVF2 Gen | | CATCGAAGAG | CAAATTTTGG | TGGAGGCCA | TGGAGACAGT | ACTCCAAAAC | | |
| coC3 cDNA | 116 | CACAGAAGAG | CAAATTTTGG | TGGAGGCCA | TGGAGACAGT | ACTCCAAAAC | 165 | |
| CVF1 Gen | | CACAGAAGAG | CAAATTTTGG | TGGAGGCCA | TGGAGACAGT | ACTCCAAAAC | | |
| CVF1 cDNA | 161 | AGCTTGACAT | CTTTGTTTCA | GATTTTCCAC | GGAAGCAGAA | AACCTTGTTT | 210 | |
| CVF2 Gen | | AGCTTGACAT | CTTTGTTTCA | AATTTTCCAC | AGAAGCAGAA | AACTTGTTT | | |
| coC3 cDNA | 166 | AGCTTGACAT | CTTTGTTTCA | GATTTTCCAC | GGAAGCAGAA | AACCTTGTTT | 215 | |
| CVF1 Gen | | AGCTTGACAT | CTTTGTTTCA | GATTTTCCAC | GGAAGCAGAA | AACCTTGTTT | | |
| CVF1 cDNA | 211 | CAAACCAGAG | TAGATATGAA | TCCAGCAGGA | GGCATGCTTG | TCACTCCAAC | 260 | |
| CVF2 Gen | | CAAACCAGAG | TAGATATGAA | TCCAGCAGGA | GGCATGCTTG | TCACTCCAAC | | |
| coC3 cDNA | 216 | CAAACCAGAG | TAGATATGAA | TCCAGCAGGA | AGCATGCTTG | TCACTCCAAC | 265 | |
| CVF1 Gen | | CAAACCAGAG | TAGATATGAA | TCCAGCAGGA | GGCATGCTTG | TCACTCCAAC | | |
| | | ← | | | | | | |

Abb. 3.21: Sequenz des genomischen PCR-Produkts mit Intron 1 des CVF1-Gens. Als Matrize diente genomische DNA der Kobra *Naja kaouthia*. Zu Vergleichszwecken sind die entsprechenden cDNA-Sequenzen von CVF1 und coC3 sowie die Gensequenz von CVF2 angeordnet. Die cDNA für CVF2 liegt für den entsprechenden Bereich nicht vor. Unterschiede zur CVF1-Sequenz sind durch Boxen markiert. Intron 1 hat eine Größe von ca. 5800 bp. Die flankierenden CVF1-Exons haben Längen von 71 bp (Exon 1, ab Startcodon) und 193 bp (Exon 2).

Die Sequenz zeigte Übereinstimmung mit der CVF1-cDNA und wurde als CVF1-Genabschnitt identifiziert. Die Identifizierung der hier zu Vergleichszwecken dargestellten CVF2-Gensequenz ist in Kapitel 3.2 beschrieben. Die EcoRI-Restriktionsanalyse dieses CVF1-spezifischen Amplifikats ergab Fragmente mit Längen von ca. 4,6 und 1,4 kb.

Intron 2

Die Amplifikation von Intron 2 mittels genomischer PCR führte auch unter Verwendung von speziellen DNA-Polymerase-Mischungen mit Korrekturaktivität nicht zum gesuchten CVF1-Genabschnitt. Stattdessen wurden die entsprechenden Amplifikate des CVF2-Gens (10 kb, siehe Kapitel 3.2.1) und des coC3-Gens (2,5 kb, siehe Kapitel 3.3.1) erhalten. Ein 5 kb-Amplifikat (Abb. 3.67), das möglicherweise den CVF1-Genabschnitt enthält, konnte aufgrund der zu geringen Ausbeute nicht kloniert und sequenziert werden. Die entsprechenden Exon/Intron-Übergänge des CVF1-Gens wurden daher durch die Genome Walking-Methode aufgeklärt (siehe Abschnitt 3.1.4).

Intron 3

Mit dem Primerpaar HB3s-CVF / HB4as-CVF wurde ein 901 bp-Amplifikat erhalten, das als CVF1-Genabschnitt identifiziert wurde (Abb. 3.22).

| | | | | | | | |
|------------------|-----|-------------------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 cDNA | 322 | CAAGTAACTG | GTCCTCAAGT | GAGATTGGAA | AAGGTGGTTC | TCCTTTCTTA | 371 |
| coC3 cDNA | 327 | AAGTAACTG | GTCCTCAAGT | GCATTGGAA | AAGGTGGTTC | TCCTTTCTTA | 376 |
| CVF1 Gen | | CAAGTAACTG | GTCCTCAAGT | GAGATTGGAA | AAGGTGGTTC | TCCTTTCTTA | |
| | | | | → | | | |
| CVF1 cDNA | 372 | CCAGAGTAGC | TTTCTGTTTA | TCCAGACAGA | TAAAGGCATC | TATACACCAG | 421 |
| coC3 cDNA | 377 | CCAGAGTGC | TTTCTGTTC | TCCAGACAGA | TAAAGGCATC | TATACACCAG | 426 |
| CVF1 Gen | | CCAGAGTAGC | TTTCTGTTTA | TCCAGACAGA | TAAAGGCATC | TATACACCAG | |
| CVF1 cDNA | 422 | GGTCTCCAG | | | | | 430 |
| coC3 cDNA | 427 | GCTCTCCAG | | | | | 435 |
| CVF1 Gen | | GGTCTCCAG _g | taaccggttt | tttctgcctc | tatttettca | ctccccctggg | |
| CVF1 Gen | | gtcccactag | ggttatgtga | tgaataatca | aaaccattcc | tgggatttgt | |
| CVF1 Gen | | cacttattcg | | | | | |
| | | | | | (531 bp) | | |
| CVF1 Gen | | | | | | tcagaatatac | |
| CVF1 Gen | | agacaattaa | tacaaacttg | ttttttttac | aatgcaagct | catatttact | |
| CVF1 cDNA | 431 | | | | | TACTCTATCG | 439 |
| coC3 cDNA | 436 | | | | | TCCCTTATCG | 444 |
| CVF1 Gen | | gaatttttacc | acaaattaca | gtagtctct | tctttcttag | TACTCTATCG | |
| CVF1 cDNA | 440 | TGTTTTTCT | ATGGATCACA | ACACAAGCAA | GATGAACAAA | ACTGTGATTG | 490 |
| coC3 cDNA | 445 | TGTCTTTTCT | CTGGATCACA | ACATGCCAC | GATGCACAAA | ACTGTGATTG | 495 |
| CVF1 Gen | | TGTTTTTCT | ATGGATCACA | ACACAAGCAA | GATGAACAAA | ACTGTGATTG | |
| | | | | | | ← | |

Abb. 3.22: Sequenz des genomischen PCR-Produkts mit Intron 3 des CVF1-Gens. Intron 3 hat eine Größe von 732 bp. Die flankierenden CVF1-Exons haben Längen von 163 bp (Exon 3) und 71 bp (Exon 4).

Für CVF2 liegt in dem entsprechenden Bereich keine DNA-Sequenz vor. Aufgrund der Häufigkeit der Unterschiede zwischen CVF1 und coC3 kann jedoch davon ausgegangen werden, dass die CVF2-Sequenz in dem betreffenden Bereich ebenfalls zahlreiche sequenzspezifische Nukleotide enthält. Das hier erhaltene und sequenzierte genomische PCR-Produkt weist in den Bereichen von Exon 3 und 4 jedoch ausschließlich CVF1-spezifische Sequenzen auf, so dass es sich bei dem amplifizierten Intron höchstwahrscheinlich um den gesuchten CVF1-Genabschnitt handelt. Das CVF1-Intron 3 hat eine Länge von 732 bp.

Intron 4

Mit dem Primerpaar HB4s-CVF / HB5as-CVF wurde ein 785 bp-langes Amplifikat erhalten, dessen Randsequenzen nachfolgend gezeigt sind. Die Zuordnung der genomischen Sequenz zum CVF1-Gen wird neben der spezifischen Basenfolge durch unterschiedliche Exongrößen erleichtert.

| | | | | |
|-----------|-----|--|-------------------------|-----|
| CVF1 cDNA | 442 | GTTTTTCTA TGGATCACAA CACAAGCAAG ATGAACAAAA CTGTGATTGT | 491 | |
| coC3 cDNA | 447 | GTCTTTTCTC TGGATCACAA CATGCAACAGG ATGACAAAA CTGTGATTGT | 496 | |
| CVF1 Gen | | GTTTTTCTA TGGATCACAA CACAAGCAAG ATGAACAAAA CTGTGATTGT | | |
| | | → | | |
| CVF1 cDNA | 492 | TGAGTTTCAG | 501 | |
| coC3 cDNA | 497 | CGAGTTTCAG | 506 | |
| CVF1 Gen | | TGAGTTTCAG gtgagaatcg tgtctgcaat cctgtgtgat tggtatagtt | | |
| CVF1 Gen | | ccctgaacaa ggaatgggtg ttaattcctt tttctcatgc aatcatgcca | | |
| CVF1 Gen | | gtagaccaca | | |
| | | (468 bp) | | |
| CVF1 Gen | | atggatttca | | |
| CVF1 Gen | | atcatataga atgcatagga aataacatgc aatgttctat atgttttcaa | | |
| CVF1 cDNA | 502 | | A CTCCAGAAGG CATTCTTGTC | 522 |
| coC3 cDNA | 507 | | A CTCCAGAAGG CATTCTTGTC | 527 |
| CVF1 Gen | | cacacctgta ttttatttcc ttatcccagA CTCCAGAAGG CATTCTTGTC | | |
| CVF1 cDNA | 523 | AGTTCTAATT CAGTTGACCT AAA---CTTC TTCTGGCCTT ACAATTTACC | 569 | |
| coC3 cDNA | 528 | AGTTCTAAAC CAGTCAATCC ATCAGGCTCG ATCAGGCCTT ACAATTTACC | 577 | |
| CVF1 Gen | | AGTTCTAATT CAGTTGACCT AAA---CTTC TTCTGGCCTT ACAATTTACC | | |
| | | ← | | |

Abb. 3.23: Sequenz des genomischen PCR-Produkts mit Intron 4 des CVF1-Gens. In Exon 5 befindet sich stromaufwärts der Primer-Bindungsstelle im coC3-Gen ein Einschub von 3 bp, der in der CVF1-Sequenz entsprechend als Lücke gekennzeichnet ist. Intron 4 hat eine Länge von 657 bp. Die flankierenden CVF1-Exons haben Längen von 71 bp (Exon 4) und 80 bp (Exon 5).

Introns 5, 6 und 7 konnten ebenfalls aufgrund der Übereinstimmung mit der CVF1-cDNA als die gesuchten CVF1-Genabschnitte identifiziert werden.

Intron 9

Mit dem Primerpaar HB9s/HB10as und Verwendung einer Taq/Pwo-Polymerase-Mischung wurde ein singuläres Amplifikat von ca. 3000 bp Länge erhalten, welches als CVF1-Genesequenz identifiziert wurde.

| | | | |
|-----------|-----|--|-----|
| CVF1 cDNA | 914 | CAAATCTCAA TGAGCTTGTT GGCATACTC TGTATGCATC TGTAACAGTC | 963 |
| coC3 cDNA | 925 | AAGATCTCAA TCAGCTTGTT GGTCATACTC TGTATGATC TGTAACAGTC | 974 |
| CVF1 Gen | | CAAATCTCAA TGAGCTTGTT GGCATACTC TGTATGCATC TGTAACAGTC | |
| | | → | |
| CVF1 cDNA | 964 | ATGACAGAAT CAG | 976 |
| CVF2 cDNA | 8 | AT CAG | 12 |
| coC3 cDNA | 975 | ATGACAGAAT CAG | 987 |
| CVF1 Gen | | ATGACAGAAT CAGgtgagca ggggtgggag tagagaaggg gtgtttagtc | |

Fortsetzung

| | | | | | | | |
|------------------|--|-------------------|-------------------|-------------------|--------------------|-------------------|------|
| CVF1 Gen | caattcaggt ctggtgcttg tgccgagcc gatcaagaaa accngtcagg | | | | | | |
| CVF1 Gen | aggtgcagat | (ca. 2800 bp) | | | | | |
| CVF1 Gen | | | | | tcgaggattt | | |
| CVF1 Gen | ttcatttttc tcttgcattt ctgagtttga cttcctattt taagttgcat | | | | | | |
| CVF1 cDNA | 977 | | GCAGTGATAT | GGTAGTGACT | GAGCAAAGCG | 1006 | |
| CVF2 cDNA | 13 | | GAGTGATAT | GGTAGTGACT | GAGCAAAGCG | 42 | |
| coC3 cDNA | 988 | | GCAGTGATAT | GGTAGTGACT | GAGCAAAGCG | 1017 | |
| CVF1 Gen | tccttctttt tgttttgacag | | GCAGTGATAT | GGTAGTGACT | GAGCAAAGCG | | |
| CVF1 cDNA | 1007 | GCATTCATAT | TGTGGCATCT | CCCTATCAGA | TCCACTTCAC | AAAAACCCCC | 1056 |
| CVF2 cDNA | 43 | GCATTCATAT | TGTGACATCT | CCCTATCAGA | TCCTACTTCAC | AAAAACCCCC | 92 |
| coC3 cDNA | 1018 | GCATTCATAT | TGTGACATCT | CCCTATCAGA | TCCTACTTCAC | AAAAACCCCC | 1067 |
| CVF1 Gen | | GCATTCATAT | TGTGGCATCT | CCCTATCAGA | TCCACTTCAC | AAAAACCCCC | |

Abb. 3.24: Sequenz des genomischen PCR-Produkts mit Intron 9 des CVF1-Gens. Zu Vergleichszwecken wurde außer den cDNA-Sequenzen von CVF1 und coC3 auch die cDNA-Teilsequenz von CVF2 angeordnet, die ab der Basenfolge ATCAG (Exon 9) bekannt ist. Die Nummerierung von CVF2 geschieht in Anlehnung an die von Fritzingler *et al.* 1994 publizierte Patentschrift¹. Unterschiede der CVF2-Sequenz zur CVF1-Sequenz wurden durch blau hinterlegte Boxen markiert. Intron 9 hat eine Größe von ca. 3000 bp. Die flankierenden CVF1-Exons haben Längen von 121 bp (Exon 9) und 116 bp (Exon 10).

Durch die Möglichkeit des Vergleichs mit der CVF2-cDNA kann das amplifizierte Intron 9 zweifelsfrei dem CVF1-Gen zugeordnet werden.

Intron 11

Mit den nicht gemäss der ARMS-Methode gewählten Primern HB11s/HB12as wurde ein Amplifikat mit einer Länge von ca. 1900 bp erhalten. Die in Abbildung 3.25 gezeigte Sequenz wurde trotz einer Abweichung in Exon 12 als die gesuchte CVF1-Gensequenz identifiziert.

| | | | | | | | |
|------------------|------|-------------------|-------------------|-------------------|-------------------|-------------------|------------|
| CVF1 cDNA | 1128 | CCATGTGCCA | GTGGTATCAG | AGGCCTTTCA | TTCTATGGGA | ACCACTTTGA | 1177 |
| CVF2 cDNA | 164 | CCATGTGCCA | GTGGTATCAG | AGGCCATTCA | TTCTGAGGGA | ACCACTTTGA | 213 |
| coC3 cDNA | 1139 | CCATGTGCCA | GTGGTATCAG | AGGCCATTCA | TTCTGAGGGA | ACCACTTTGA | 1188 |
| CVF1 Gen | | CCATGTGCCA | GTGGTATCAG | AGGCCTTTCA | TTCTATGGGA | ACCACTTTGA | |
| CVF1 cDNA | 1178 | GTGATGGGAC | TGCTAAGCTC | ATCCTGAACA | TACCATTGAA | TGCTCAAAGC | 1227 |
| CVF2 cDNA | 214 | GTGATGGGAC | TGCTAAGCTC | ATCCTGAACA | CACCACAA | TGCTCAAAGC | 263 |
| coC3 cDNA | 1189 | GTGATGGGAC | TGCTAAGCTC | ATCCTGAACA | CACCACTGAA | CATTCAAAGC | 1238 |
| CVF1 Gen | | GTGATGGGAC | TGCTAAGCTC | ATCCTGAACA | TACCATTGAA | TGCTCAAAGC | |
| CVF1 cDNA | 1228 | CTACCAATCA | CT | | | | 1239 |
| CVF2 cDNA | 264 | CTACC | CATCA | CT | | | 275 |
| coC3 cDNA | 1239 | CTACC | CATCA | CT | | | 1250 |
| CVF1 Gen | | CTACCAATCA | CT | gtaagtca | ctctgacatt | ttgctaaatc | ccattcagtg |
| CVF1 Gen | | tgctgatgag | agggaaatgct | atgacatgga | ggtggattcc | caggagtaag | |
| CVF1 Gen | | ggggtagat | | | | | |
| CVF1 Gen | | | (ca. 1700 bp) | | | | |
| CVF1 Gen | | | | | | agtgccagcc | |
| CVF1 Gen | | ctaaaaggag | ttctagtttc | ttcctaatta | ggttactgct | tcattgtttac | |

¹ Die Nummerierung der CVF2-Teilsequenz beginnt erst mit 8, da die publizierte Sequenz irrtümlich mit der Sequenz GAATTC des Klonierungsvektors beginnt (siehe auch Erratum in Fritzingler *et al.* 1995).

Fortsetzung

| | | | | | | | |
|-----------|------|---------------------|------------|--------------------|---------------------|---------------------|------|
| CVF1 cDNA | 1240 | G | TTAGAACTAA | CCATGGAGAC | CTCCAAGAG | AACGCCAGGC | 1280 |
| CVF2 cDNA | 276 | G | TTAGAACTAA | CCATGGAGAC | CTCC <u>T</u> AAGAG | AACGCCAGGC | 316 |
| coC3 cDNA | 1251 | G | TTAGAACTAA | CCATGGAGAC | CTCCAAGAG | AACGCCAGGC | 1291 |
| CVF1 Gen | | cctatctagG | TTAGAACTAA | CCATGGAGAC | CTC <u>T</u> CAAGAG | AACGCCAGGC | |
| | | | | | | | |
| CVF1 cDNA | 1281 | AACAAAGTCC | ATGACAGCCA | TAGCCTACCA | AACCCAGGGA | GGATCTGGAA | 1330 |
| CVF2 cDNA | 317 | AA <u>T</u> AAAGTCC | ATGACAGCCA | <u>C</u> AGCCTACCA | AACCCAGGGA | GGATCTGGAA | 366 |
| coC3 cDNA | 1292 | AA <u>T</u> AAAGTCC | ATGACAGCCA | <u>C</u> AGCCTACCA | AACCCAGGGA | GG <u>C</u> TCTGAAA | 1341 |
| CVF1 Gen | | AACAAAGTCC | ATGACAGCCA | TAGCCTACCA | AACCCAGGGA | GGATCTGGAA | |

Abb. 3.25: Sequenz des genomischen PCR-Produkts mit Intron 11 des CVF1-Gens. Intron 11 hat eine Länge von ca. 1900 bp. Die flankierenden CVF1-Exons haben Längen von 147 bp (Exon 11) und 210 bp (Exon 12). In der genomischen CVF1-Sequenz wurde an Position 1264 der entsprechenden cDNA ein T anstelle von C identifiziert (umrahmt).

Die genomische Sequenz weicht in Position 1264 von der CVF1-cDNA ab (C-T), wobei es sich um eine Mutation handeln könnte, die entweder während der PCR oder bei der Klonierung in *E. coli* entstanden ist. Diese Veränderung würde für Serin (S) anstelle von Prolin (P) codieren. Zur Absicherung dieser Sequenz wurde wie nachfolgend für Intron 12 bis 14 beschrieben zusätzlich eine *long range*-PCR durchgeführt, wobei die Sequenz des erhaltenen Amplifikats an der entsprechenden Stelle dieselbe Mutation aufwies.

Intron 12 bis Intron 14

Zunächst wurde mit dem Primerpaar YL12s/YL13as mittels genomischer PCR Intron 12 amplifiziert und dessen Größe von 123 bp bestimmt. Sowohl Intron 12 als auch die nachfolgenden Introns 13 und 14 sind mit Längen zwischen 100 bp und 209 bp relativ kurz, weshalb eine genomische *long range* PCR von Exon 11 bis Exon 15 möglich war. Diese wurde mit dem Primerpaar HB11s/OPA19 durchgeführt und führte zu einem Amplifikat mit einer Länge von ca. 2500 bp (Abb. 3.27, Bahn 1).

Darüber hinaus wurde mit dem Primerpaar AJS01/HB-1R67-5'-as ein *long range*-PCR-Produkt von Exon 14 bis Intron 15 erhalten, das eine überlappende Sequenz mit dem Cosmid-Klon 1R67 aufwies.

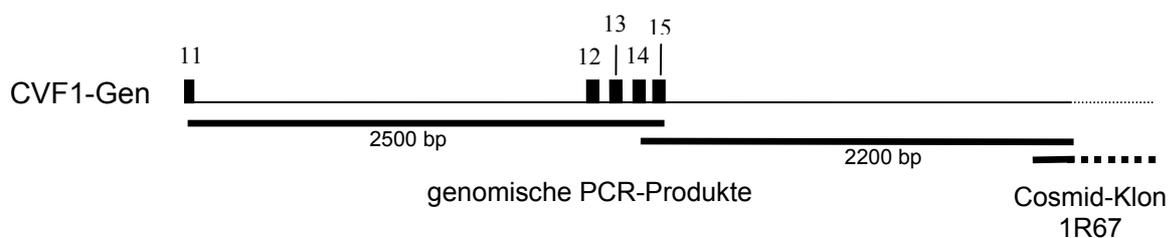


Abb. 3.26: Darstellung der Überlappung genomischer PCR-Produkte mit dem Cosmid-Klon 1R67 im Bereich von CVF1-Exon 11 bis Intron 15. Die nummerierten Kästchen geben die Exons und die dazwischenliegenden Bereiche die Introns wieder. Vom Cosmid-Klon 1R67 ist nur der 5'-Bereich dargestellt, der mit einem genomischen *long range*-PCR-Amplifikat überlappt.

Die beiden *long range*-PCR-Produkte, die sich über mehrere Exons erstrecken, sind nachfolgend dargestellt.

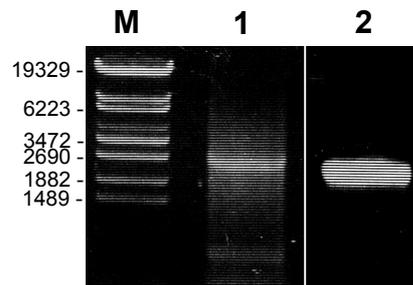


Abb. 3.27: Gelelektrophoretische Analyse der genomischen *longe range*-PCR-Produkte (1%iges Agarosegel). M: DNA-Marker λ /Eco130I, Bahn 1: Introns 11-14-Amplifikat (2500 bp, Primerpaar HB11s/OPA19), Bahn 2: Introns 14-15-Amplifikat (2200 bp, Primerpaar AJS01/HB-1R67-5'-as).

Intron 14 und Intron 15 (5'-Bereich)

Eine vollständige Amplifikation von Intron 15 des CVF1-Gens war aufgrund der Länge von ca. 9.000 bp nicht möglich. Für eine partielle Intron 15-Amplifikation bis zu der Gensequenz, die im Cosmid-Klon 1R67 enthalten ist, wurden außer Exon 15 auch noch Intron 14 und Exon 14 einbezogen und durch eine *long range*-PCR amplifiziert. Da die Gene CVF1, CVF2 und coC3 in Exon 15 identisch sind (Abb. 3.19), konnten durch die Einbeziehung der weiter stromaufwärts gelegenen Gensequenzen Unterscheidungsmerkmale genutzt werden, die das erhaltene *long range*-Amplifikat als CVF1-Gensequenz identifizierten.

Intron 14 wurde zunächst durch Sequenzierung eines genomischen PCR-Amplifikats, das mit dem Primerpaar AJS01/OPA19 erhalten wurde (Abb. 3.20), analysiert und wies eine Länge von nur 100 bp auf. Die *long range*-PCR mit dem Primerpaar AJS01/HB-1R67-5'as führte zu einem Amplifikat von ca. 2,2 kb Länge (Abb. 3.27), dessen Sequenz nachfolgend in Abbildung 3.28 gezeigt ist.

| | | | | | | | | | |
|-------------|-------------|------|-------------------|---------------------------|--------------------------|--------------------|-------------------------|-------------------|--|
| CVF1 | cDNA | 1710 | GGATCCAGGT | GCTCGGGTTG | GTCTTGTGGC | TGTGGACAAA | GCAGTATATG | 1759 | |
| CVF2 | cDNA | 758 | GGATCCAGGT | GCTTGGATTG | GTCTTGTGGC | TGTGGACAAA | GCAGTATATG | 807 | |
| coC3 | cDNA | 1733 | GGATCCAGGT | GCTCGGGTTG | GTCTTGTGGC | TGTGGACAAA | GCAGTATATG | 1782 | |
| CVF1 | Gen | | GGATCCAGGT | GCTCGGGTTG | GTCTTGTGGC | TGTGGACAAA | GCAGTATATG | | |
| | | | → | | | | | | |
| CVF1 | cDNA | 1760 | TTCTCAATGA | TAAATATAAG | ATTAGCCAAG | CTAAG | | 1794 | |
| CVF2 | cDNA | 808 | TTCTCAATGA | TAAATATAAG | ATTAGCCAAG | CTAAG | | 842 | |
| coC3 | cDNA | 1783 | TTCTCAATGA | TAAATATAAG | ATTAGCCAAG | CTAAG | | 1817 | |
| CVF1 | Gen | | TTCTCAATGA | TAAATATAAG | ATTAGCCAAG | CTAAG | <i>gtaag gccagaaatg</i> | | |
| CVF1 | Gen | 101 | <i>tottaagtta</i> | <i>caaacattca</i> | <i>gctttcaaga</i> | <i>gttttccaac</i> | <i>ttacagtact</i> | | |
| CVF1 | cDNA | 1795 | | | | ATATG | GGACACAATA | 1809 | |
| CVF2 | cDNA | 843 | | | | ATATG | GGACACAATA | 857 | |
| coC3 | cDNA | 1818 | | | | ATATG | GGACACAATA | 1832 | |
| CVF1 | Gen | | <i>gaagaggaac</i> | <i>atTTTTTTTtc</i> | <i>tctgatgtgc</i> | <i>actag</i> | ATATG | GGACACAATA | |
| CVF1 | cDNA | 1810 | GAAAAGAGTG | ACTTTGGCTG | TACAGCTGGC | AGTGGCCAGA | ATAATCTGGG | 1859 | |
| CVF2 | cDNA | 858 | GAAAAGAGTG | ACTTTGGCTG | TACAGCTGGC | AGTGGCCAGA | ATAATCTGGG | 907 | |
| coC3 | cDNA | 1833 | GAAAAGAGTG | ACTTTGGCTG | TACAGCTGGC | AGTGGCCAGA | ATAATCTGGG | 1882 | |
| CVF1 | Gen | | GAAAAGAGTG | ACTTTGGCTG | TACAGCTGGC | AGTGGCCAGA | ATAATCTGGG | | |
| CVF1 | cDNA | 1860 | TGTGTTTGAA | GATGCTGGAC | TGGCTCTGAC | AACCAGCACT | AATCTCAACA | 1909 | |
| CVF2 | cDNA | 908 | TGTGTTTGAA | GATGCTGGAC | TGGCTCTGAC | AACCAGCACT | AATCTCAACA | 957 | |
| coC3 | cDNA | 1883 | TGTGTTTGAA | GATGCTGGAC | TGGCTCTGAC | AACCAGCACT | AATCTCAACA | 1932 | |
| CVF1 | Gen | | TGTGTTTGAA | GATGCTGGAC | TGGCTCTGAC | AACCAGCACT | AATCTCAACA | | |
| | | | → | | | | | | |
| CVF1 | cDNA | 1910 | CCAAACAGAG | ATCAG | | | | 1924 | |
| CVF2 | cDNA | 958 | CCAAACAGAG | ATCAG | | | | 972 | |
| coC3 | cDNA | 1933 | CCAAACAGAG | ATCAG | | | | 1947 | |
| CVF1 | Gen | | CCAAACAGAG | ATCAG | <i>gtagt</i> | <i>atTTTTgggaa</i> | <i>tggatgctaa</i> | <i>tgtctgggag</i> | |
| CVF1 | Gen | | <i>atggcagagg</i> | <i>TTTTccaatc</i> | <i>catttagggg</i> | <i>ccatgggaag</i> | <i>agcattgctc</i> | | |
| CVF1 | Gen | | <i>catcatatac</i> | (ca. 1400 bp) | | | | | |
| CVF1 | Gen | | | | | | | <i>gtgagagtga</i> | |
| CVF1 | Gen | | <i>ccatgtcata</i> | <i>ctagaattca</i> | <i>acataacaca</i> | <i>aacacaagca</i> | <i>atagaacata</i> | | |
| CVF1 | Gen | | <i>atgacaccag</i> | <i>agtcoctggat</i> | <u><i>cttaaaagag</i></u> | <i>ctggTTTTaa</i> | <i>taaactcaga</i> | | |
| CVF1 | Gen | | <i>gagaacttgg</i> | <i>gaaagattcc</i> | <i>atggatgaaa</i> | <i>atcctaaag</i> | <i>ggaaaaacaac</i> | | |
| CVF1 | Gen | | <i>ttaaggatct</i> | <i>tggaaaactc</i> | <i>tgaaaagtac</i> | <i>aatcatagaa</i> | <i>gccagacca</i> | | |
| | | | ← | | | | | | |

Abb. 3.28: Genomische CVF1-Sequenz von Exon 14 bis Intron 15. Die kursiv dargestellte Intron 15-Sequenz am 3'-Ende des genomischen PCR-Produkts (stromabwärts der MboI-Schnittstelle *gatc*) ist überlappend mit dem 5'-Bereich des Cosmid-Klons 1R67. Eine EcoRI-Schnittstelle ist unterstrichen dargestellt.

Aufgrund der spezifischen Sequenzmerkmale in Exon 14, der Sequenzidentität in Intron 14 und der mehr als 100 bp-langen Sequenz-Übereinstimmung mit dem 5'-Ende des Cosmid-Klons 1R67 konnte das obige genomische Amplifikat als CVF1-Genabschnitt identifiziert werden.

3'-Bereich des CVF1-Gens

Für eine Aufklärung der Exon/Intron-Strukturen im 3'-Bereich des CVF1-Gens wurden wie bereits beschrieben die Exonstrukturen in Analogie zur humanen C3-Genstruktur vorhergesagt. In Abbildung 3.29 ist der entsprechende Ausschnitt aus den cDNA-Sequenzen (Exons 28-41) gezeigt. Die Introns wurden mittels genomischer PCR unter Verwendung exonspezifischer Primer amplifiziert (Abb. 3.30).

Exon 28

CVF1 ... AATAGTCTAGACAGCAGCATCAAGAAGGCCACAAATTATTTACTCAAAAAGTATGAG 3495
 CVF2 ... AATAATCTAGACAGCAGCATCACTAAGGCCACAATTATTTACTCAAAAAGTATGAG 2543
 coC3 ... AATAATCTAGACAGCAGCATCAATAAGGCCACAATTATTTACTCAAAAAGTATGAG 3518

CVF1 AACTGCAAAGGCCTTACACTACAGCCCTCACAGCCTATGCTTTGGCTGCTGCAGACCAA 3555
 CVF2 AACTGCAAAGGCCTTACACTACAGCCCTCACAGCCTATGCTTTGGCTGCTGCAGCAACA 2603
 coC3 AACTGCAAAGGCCTTACACTACAGCCCTCACAGCCTATGCTTTGGCTGCTGCAGACCAA 3578

Exon 29

CVF1 CTC AATGATGACAGGGTACTCATGGCAGCATCAACAGGAAGGGATCATTTGGGAAGAATAC 3615
 CVF2 CTC AATGATGACAGGGTACTCATGGCAGCATCAACAGGAAGGAATCTTTGGGAAGAACTT 2663
 coC3 CTC AATGATGACAGGGTACTCATGGCAGCATCAACAGGAAGGAATCTTTGGGAAGAATAT 3638

CVF1 AATGCTCACACCCACAACATTGAAGGCACTTCCATATGCCTTGTGGCCCTGCTGAAAATG 3675
 CVF2 AAAGCCACACCCATAAACATTGAAGGCACTTCCATATGCCTTGTGGCCCTGCTGAAAATG 2723
 coC3 AATGCTCGCACCCATAAATTATTGAAGGCACTTCCATATGCCTTGTGGCCCTGCTGAAAATG 3698

CVF1 AAGAAATTTGATCAAACCTGGTCCCATAGTCAGATGGCTGACAGATCAGAAATTTTATGGG 3735
 CVF2 AAGAAATTTGTTGAGCCCGGTCCCTAGTCCAAATGGCTGATAGATCAGCAATATTTATGGG 2783
 coC3 AAGAAATTTGCTGAGGTCGGTCCCTAGTCCAAATGGCTGATAGATCAGAAATATTTATGGG 3758

Exon 30

CVF1 GAAACATATGGACAAACCAAAGCAACAGTTATGGCATTTCAGCTTTGCTGAATATGAG 3795
 CVF2 GAAACATATGGACAAACCAAAGCAACAGTTATGATCTTTCAAGCTTTGCTGAATATGAG 2843
 coC3 GAAACATATGGACAAACCAAAGCAACAGTTATGGTCTTTCAAGCTTTGCTGAATATGAG 3818

CVF1 ATTCAGATGCCTACCCATAAGGACTTAACTTAGATATTACTATTGAACTGCCAGATCGA 3855
 CVF2 ATTCAGATGCCTACCCATAAGGACTTAACTTAGATATTACTATTGAACTGCCAGATCGA 2903
 coC3 ATTCAGATGCCTACCCATCAGGACTTAAATTTTAGATATTCTATTAACTGCCAGAACGA 3878

CVF1 GAAGTACCTATAAGGTACAGAATTAATTATGAAAATGCTCTCCTGGCTCGGACAGTAGAG 3915
 CVF2 GAAGTACCTATAAGGTACAGAATTAATTATGAAAATGCTCTCCTGGCTCAGACAGTAGAG 2963
 coC3 GAAGTACCTGAAGGTACAGCATTAAATGATAGAAATGCTGTCCAGGCCCGGACAGTAGAG 3938

Exon 31

CVF1 ACCAAACTCAACCAAGACATCACTGTGACAGCATCAGGTGATGGAAAAGCAACAATGACC 3975
 CVF2 ACCAAACTCAACCAAGACTTTCACGTGTTCAGCATCAGGTGATGGAAAAGCAACAATGACC 3023
 coC3 ACCAAACTCAACCAAGACTTTCACGTGTTCAGCATCAGGTGATGGAAAAGCAACAATGACC 3998

Exon 32

CVF1 ATTTTGACATTTCTATAACGCACAGTTGCAGGAGAAGGCAAATGTTTGCAATAAATTTTCAT 4035
 CVF2 ATTTTGACGGTCTATAATGCACAATTGAGGGAGGATGCAAATGTTTGCAACAAATTTTCAT 3083
 coC3 ATTTTGACGGTCTATAATGCACAATTGAGGGAGGATGCAAATGTTTGCAACAAATTTTCAT 4058

Exon 33

CVF1 CTTAATGTTTCTGTTGAAAACATCCACTTGA-----ATGCAATGGGAGCCAAGGGA 4086
 CVF2 CTTGATGTTTCTGTTGAAAACCTCCAGTTGAACTTAAAACAGGCAAGGGGAGCCAAGGGA 3143
 coC3 CTTGATGTTTCTGTTGAAAACCTCCAAATTGAACTTAAAACAGGCAAGGGGAGCCAAGGCA 4118

Exon 34

CVF1 GCCCTCATGCTCAAGATCTGCACAAGGTATCTGGGAGAAGTTGATTCTACAATGACAATA 4146
 CVF2 GCCCTCAAGCTCAAAATCTGCACCTAGGTATCTGGGAGAAGTTGATTCTACAATGACAATA 3203
 coC3 GCCCTCAAGCTTAAATCTGCACCTAGGTATCTGGGAGAAGTTGATTCTACAATGACAATA 4178

CVF1 ATTGATATTTCTATGCTGACTGGTTTTCTCCCTGATGCTGAAGACCTTACAAGCCTTTTCT 4206
 CVF2 ATTGATGTTTTCTATGCTGACTGGTTTTCTCCCTGATGCTGAAGACCTTACGAGCCTTTCT 3263
 coC3 ATTGATATTTCTATGCTGACTGGTTTTCTCCCTGATGCTGAAGACCTTAAAGGCTTTTCT 4238

Exon 35

CVF1 AAAGGAGTGGACAGATACATCTCCAGATATGAAGTTGACAATAATATGGCTCAGAAAGTA 4266
 CVF2 AAAGGAGTGGACAGATATCTCCATGTTTGAAATTAACAATAATATGGCTCAGAAAGCA 3323
 coC3 AAAGGAGTGGACAGATACATCTCCAGTTTGAAATTTGACAATAATATGGCTCAGAAAGCA 4298

Fortsetzung

| | | | Exon 36 | |
|------|--|--|----------------|------|
| CVF1 | GCTGTTATCATTTACTTAAACAAG | GTCTCCCCTCTGAAGATGAATGCCTGCAC | TTAAG | 4326 |
| CVF2 | A CTGTTATCATTTACTT A GACAAG | GTCTCCCCTCTGAAGATGAATGCCTGCAC | TTAAG | 3383 |
| coC3 | A CTGTT C TATTACTT A GACAAG | GTCTCCCCTCTGAAGATGAATGCCTGCAC | TTAAG | 4358 |
| | | | | |
| CVF1 | ATTCTCAAGCATTTTGAAGTTGGCTTCATT | CAGCCAGGATCAGTCAAGGTGTACAGCTAC | | 4386 |
| CVF2 | ATTCTCAAGCATTTTGAAGTTGGCTTCATT | CAGCCAGGATCAGTCAAGGTGTACAGCTAC | | 3443 |
| coC3 | ATT C AAGCATTTTGAAGTTGGCTTCATT | CAGCCAGGATCAGTCAAGGTGTACAGCTAC | | 4418 |
| | | | | |
| | | | Exon 37 | |
| CVF1 | TACAATCTAGATGAAAAATGTACCAAGTTCT | ACCATCCAGATAAAGGAACAGGCCTTCTC | | 4446 |
| CVF2 | TACAATCTAGATGAAAAATGTACCAAG A | ATCTACCATCCAGAT C AAG C AACAGGCCTTCTC | | 3503 |
| coC3 | TACAATCTAGATGAA C AATGTACCAAGTTCT | ACCATCCAGATAAAG A AACAGG T CTTCTC | | 4478 |
| | | | | |
| | | | Exon 38 | |
| CVF1 | AATAAGATATGTATTGGTAACGTTTGCCGAT | GTGCAGGAGAAACCTGTTCCCTCGCTCAAC | | 4506 |
| CVF2 | AATAAGATATGT G TTGGTAACGTTTGCCGAT | GTGCAG A AGAAACCTGTTCC T GCTCAAC | | 3563 |
| coC3 | AATAAGATATGT C ATGGTAAC A TTTGCCGAT | GTGCAG A AGAAACCTGTTCC T GCTCAAC | | 4538 |
| | | | | |
| CVF1 | CATCAGGAAAGGATTGATGTTCCATTACAAAT | TGAAAAAGCCTGCGAGACGAATGTGGAT | | 4566 |
| CVF2 | C A GCAG A AG A AT G TT A CT C CG C A A TT G CG A ATT C AG A AGCCT T CG A T C C A AAATGTGGAT | | | 3623 |
| coC3 | C A GCAG A AA A AGATTGAT C TT C A A TTAC G AATT C AAAAAGCCT G CG C AA A ATGTGGAT | | | 4598 |
| | | | | |
| | | | Exon 39 | |
| CVF1 | TATG T CTACAAAACCAAGCTGCTTCGAATAGA | AACAAGATGGTAATGATATCTATGTC | | 4626 |
| CVF2 | TATG T CTA T AAAACCAAGCTGCTTCGAATAGA | AAGAA A AGATGGTAATGATATCTATGTC | | 3683 |
| coC3 | TATG T CTACAAAACCAAGCTGCTTCGAATAGA | AAGAA A AGATGGTAATGATATCTAT T TC | | 4658 |
| | | | | |
| | | | Exon 40 | |
| CVF1 | ATGGATGTTT T AGAAGTTATTAACAAG G | TACTGACGAAAA T CCACGAGCAAAGACCCAC | | 4686 |
| CVF2 | ATGG A CGTTT T AGAAGTT C TTAAACAAG G | C A CTGAC C AAAAT C A A AG T AAAG C TCC G C | | 3743 |
| coC3 | ATGGATGTTT T AGAAGTTATTA A AG G AG G C A CTGAC C GAAAT C C A AGCAAAG C CC G C | | | 4718 |
| | | | | |
| CVF1 | CAGTACATAAGTCAAAGGAAATGCCAGGAGG | CTCTGAATCTGAAGGTGAATGATGATTAT | | 4746 |
| CVF2 | CAGTA T GTAAGTCAAAGGAAATGCCAGGAGG | CTTGAATCTGA T GGTGAAT A ATGATTAT | | 3803 |
| coC3 | CAGTA T GTAAGTCAAAGGAAATGCCAGGAGG | CTTGAATCTGAAG C TG C ATA A ATGATTAT | | 4778 |
| | | | | |
| | | | Exon 41 | |
| CVF1 | CTGATCTGGGGTTCCAGGAGTGACCTGTTGCC | ACGAAAGATAAAATTTCC T TACATCATT | | 4806 |
| CVF2 | CTGATCTGGGGT C CAAG C AGTGACCTGT G GCC A T | GAAAGATAAAATTTCC T A T C T CATT | | 3863 |
| coC3 | CTGATCTGGGGT C T C AG C AGTGACCTGT G GCC A T | GAAAGAT C A T A T C T C T TAC T CATT | | 4838 |
| | | | | |
| CVF1 | ACAAAGAACACATGGATTGAGAGATGGCCAC | ATGAAGACGAATGTCAGGAAGAAGAATTC | | 4866 |
| CVF2 | ACAAAGAACAC C TGGATTGAGAGATGGCCAC | ATGAAGAC A AAATGTCAGGAAGAAGAATTC | | 3923 |
| coC3 | ACAAAGAACAC C TGGATTGAGAGATGGCC A | ATGAAGAT G AAAT G C C AGGA T GAAAGAATTC | | 4898 |
| | | | | |
| CVF1 | CAAAAGTTGTGTGATGACTTTGCTCAGTTT | AGCTACACATTGACTGAGTTTGGCTGCCCT | | 4926 |
| CVF2 | CAAAAGTTGTGTGATGACTTTGCT C T T GTT | AGCTAC G CA T GA---GTTTG-CTGCCCT | | 3978 |
| coC3 | CA G AA T TTGTGTGATGACTTTGCTCAGTT G | T C CA T TAC A CTGACT A TTTTGGCTGCCCT | | 4958 |
| | | | | |
| CVF1 | ACTTAA... | | | 4932 |
| CVF2 | ACTTAA... | | | 3984 |
| coC3 | ACTTAA... | | | 4964 |

Abb. 3.29: Darstellung der Exonstrukturen von CVF1, CVF2 und coC3. Auszug aus den cDNA-Sequenzen (Exons 28-41, bis zum Stop-Codon TAA). Dargestellt sind die tatsächlich bestätigten Exongrenzen. Die ursprünglich zwischen Exon 31 und Exon 32 vermutete Exongrenze ist gestrichelt. Die Nummerierung bezieht sich auf die cDNA-Sequenzen.

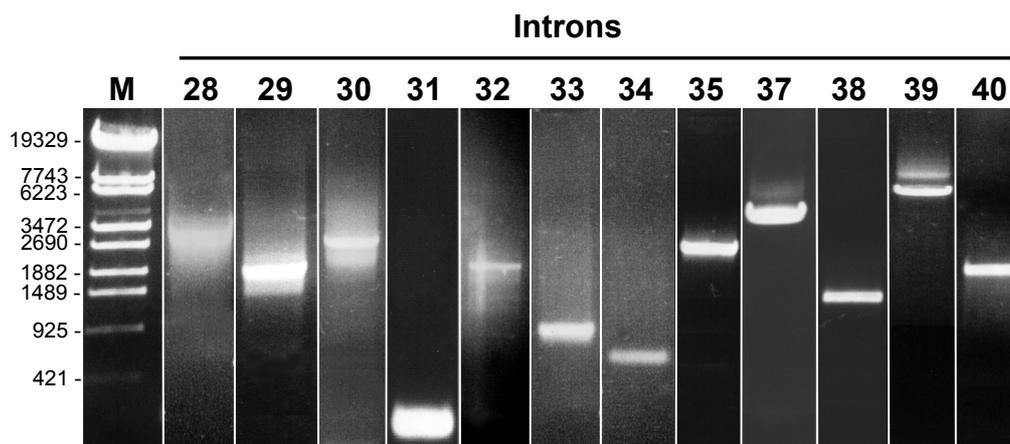


Abb. 3.30: Gelelektrophoretische Analyse genomischer Amplifikate mit Introns 28-40 des CVF1-Gens (1%iges Agarosegel). M: Marker λ -DNA/Eco130I, Intron 28: 2800 bp, Intron 29: 1800 bp, Intron 30: 2600 bp, Intron 31: 123 bp, Intron 32: 2200 bp, Intron 33: 898 bp, Intron 34: 560 bp, Intron 35: 2600 bp, Intron 37: 3600 bp, Intron 38: 1400 bp, Intron 39: 7000 bp, Intron 40: 2500 bp.

Für die in Abbildung 3.30 gezeigten Amplifikate wurden die in Tabelle 3.3 aufgeführten Primerpaare verwendet.

Tab. 3.3: Übersicht der genomischen CVF1-Amplifikate im 3'-Bereich des CVF1-Gens (Introns 28-40). Angaben in Klammern sind Schätzungen. Intron 27 und 36 des CVF1-Gens waren durch genomische PCR nicht zu amplifizieren.

| Intron | Primerpaar | Größe PCR-Produkt (bp) | Exon-Intron-Übergangssequenzen |
|--------|----------------------------|------------------------|--|
| 28 | HB28s / HB29as | (2.800) | CAG gtacacatat...tttatttttag GAA |
| 29 | HB29s-CVF / HB30as-CVF | (1.800) | CAAG tgagtcctt...ctcattctag GCA |
| 30 | HB30s-CVF / HB31as-CVF | (2.600) | GAG gtatagtatt...tgcttttcag ACC |
| 31 | HB31s-CVFgenom./HB32as-CVF | 123 | kein Intron 31 |
| 32 | HB32s-CVF / HB33as-CVF | (2.200) | TGA gtaagtttga...aacttcaaag ATG |
| 33 | HB33s-CVF / HB34as-CVF | 898 | AAG gtaagaatgc...ttgatttttag GTA |
| 34 | HB34s-CVF / HB35as-CVF | 560 | AGG gtaggcaagc...cgttatttag CTT |
| 35 | HB35s-CVF / HB36as-CVF | (2600) | AAG gtaaataactt...tccatcccag GTC |
| 37 | HB37s-CVF / HB38as-CVF | (3600) | GAG gtaagttggg...ccccttgaag AAA |
| 38 | HB38s-CVF / HB39as-CVF | (1400) | ATG gtatttatat...ttcttaaaag TCT |
| 39 | HB39s-CVF / HB40as-CVF | (7000) | AAG gtaagaattt...atctctgcag GTA |
| 40 | HB40s-CVF / HB41as-CVF | (2500) | TAA gtacgtattt...ctctttccag AAT |

Eine Amplifikation von Intron 27 und Intron 36 des CVF1-Gens durch genomische PCR war auch unter Anwendung unterschiedlichster Strategien nicht möglich, was auf ausgedehnte Intronlängen hinwies. Die stattdessen amplifizierten Introns homologer Gene sind in Kapitel 3.4 beschrieben.

Die Exon/Intron-Übergänge von Intron 27 und Intron 36 des CVF1-Gens wurden stattdessen durch den Cosmid-Klon 1R67 bzw. durch Genome Walking aufgeklärt (Kap. 3.1.4).

Bemerkenswert an obigen Resultaten (Tab. 3.3) ist das Fehlen von Intron 31, was nachfolgend ausführlich erläutert wird.

Fehlen von Intron 31

Bei einer mit den Primern HB31s-CVFgenom. und HB32as-CVF durchgeführten genomischen PCR wurde ein Produkt von 123 bp Länge erhalten (Abb. 3.30), dessen Sequenz mit der entsprechenden CVF1-cDNA identisch war. Um eine Kontamination mit cDNA als Template auszuschliessen und sicherzustellen, dass an der entsprechenden Position im CVF1-Gen anders als im humanen C3-Gen kein Intron vorhanden ist, wurde Genome Walking (siehe Kapitel 3.1.3 und Abb. 3.46) durchgeführt. Die Sequenz-Daten bestätigten, dass im CVF1-Gen kein Intron 31 vorhanden ist. Zur weiteren Absicherung dieses Befunds wurden mit den Primerpaaren HB30s-CVF/HB-CVF-32as-II und HB31s-CVFgenom./HB33as-CVF zwei überlappende PCRs wie in Abbildung 3.31B illustriert durchgeführt und die erhaltenen Amplifikate sequenziert.

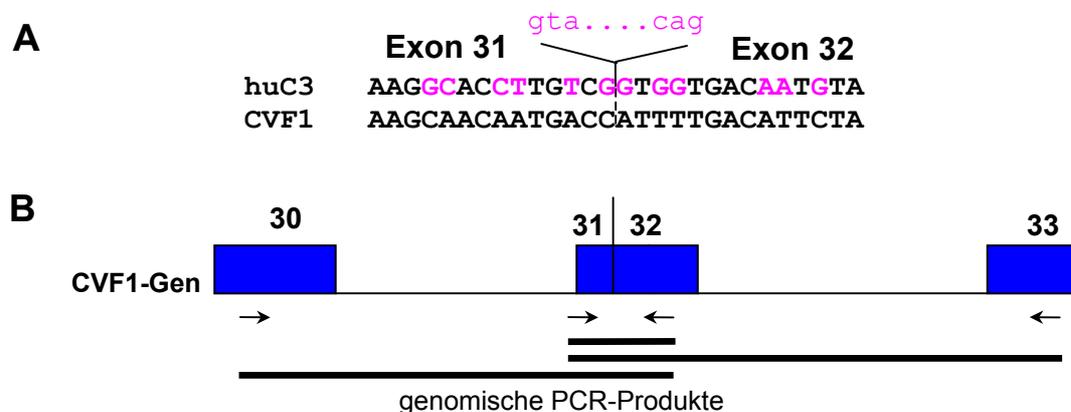


Abb. 3.31: Fehlen von Intron 31 im CVF1-Gen. **A:** Vergleich der Gensequenzen von human C3 und CVF1 im Grenzbereich Exon 31/Exon 32. Intron 31 im humanen C3-Gen ist durch einen Einschub an der Exongrenze dargestellt. Die Sequenzunterschiede von huC3 zu CVF1 sind rot markiert. **B:** Schematische Darstellung zur Überlappung der genomischen PCR-Produkte. Die Exons im CVF1-Gen sind als nummerierte Boxen dargestellt. Die Primer-Bindungsstellen sind als Pfeile markiert.

Alle Experimente bestätigten, dass im Gegensatz zur humanen C3-Genstruktur Exons 31 und 32 im CVF1-Gen kontinuierlich und nicht durch ein Intron unterbrochen sind. Die Exon- und Intron-Nummerierung stromabwärts von Exon 31 wurde aufgrund der ausgeprägten Konservierung analog zur C3-Genstruktur beibehalten.

Intron 32

Zur Amplifikation von Intron 32 wurde zunächst eine genomische PCR mit dem Primerpaar HB32s-CVF/HB33as-CVF durchgeführt, welche zu einem spezifischen Amplifikat von 2,2 kb-Länge führte (Abb. 3.30). Dieses Amplifikat wurde kloniert und aufgrund seiner Übereinstimmung mit der CVF1-Sequenz dem CVF1-Gen zugeordnet.

Eine zusätzliche genomische PCR mit dem Primerpaar HB31s-CVFgenom./HB33as-CVF führte zu einem geringfügig schwereren PCR-Produkt von ca. 2250 bp und konnte den Beweis erbringen, dass Intron 31 nicht im CVF1-Gen enthalten ist. Die Sequenz dieses PCR-Produkts ist nachfolgend dargestellt und stimmt mit der Intron 32-Sequenz des zuerst gewonnenen Amplifikats überein.

| | | Exon 31 | | | | | |
|-----------|------|----------------|------------|------------|------------|---------------|------------|
| CVF1 cDNA | 3916 | ACCAA | ACT | CAACCAAGAC | ATCACTGTGA | CAGCATCAGG | 3953 |
| CVF2 cDNA | 2964 | ACCAA | ACT | CAACCAAGAC | ATCACTGTGA | CAGCATCAGG | 3001 |
| coC3 cDNA | 3939 | ACCAA | ACT | CAACCAAGAC | ATCACTGTGA | CAGCATCAGG | 3976 |
| CVF1 Gen | | gttgcttttc ag | ACCAA | ACT | CAACCAAGAC | ATCACTGTGA | CAGCATCAGG |
| | | Exon 32 | | | | | |
| CVF1 cDNA | 3954 | TGATGGAAAA | GCAACAATGA | CCATTTTGAC | ATTCTATAAC | GCACAGTTGC | 4003 |
| CVF2 cDNA | 3002 | TGATGGAAAA | GCAACAATGA | CCATTTTGAC | GGTCTATAAT | GCACAGTTGA | 3051 |
| coC3 cDNA | 3977 | TGATGGAAAA | GCAACAATGA | CCATTTTGAC | GGTCTATAAT | GCACAGTTGA | 4026 |
| CVF1 Gen | | TGATGGAAAA | GCAACAATGA | CCATTTTGAC | ATTCTATAAC | GCACAGTTGC | |
| CVF1 cDNA | 4004 | AGGAGAAGGC | AAATGTTTGC | AATAAATTC | ATCTTAATGT | TTCTGTTGAA | 4053 |
| CVF2 cDNA | 3052 | GGGAGCATGC | AAATGTTTGC | AACAATTC | ATCTTCATGT | TTCTGTTGAA | 3101 |
| coC3 cDNA | 4027 | GGGAGCATGC | AAATGTTTGC | AACAATTC | ATCTTCATGT | TTCTGTTGAA | 4076 |
| CVF1 Gen | | AGGAGAAGGC | AAATGTTTGC | AATAAATTC | ATCTTAATGT | TTCTGTTGAA | |
| CVF1 cDNA | 4054 | AACATCCACT | TGA | | | | 4066 |
| CVF2 cDNA | 3102 | AACATCCACT | TGA | | | | 3114 |
| coC3 cDNA | 4077 | AACATCCACT | TGA | | | | 4089 |
| CVF1 Gen | | AACATCCACT | TGA | gtaagtt | tgactggggg | cctttggcct | gctagaaata |
| CVF1 Gen | | gaatcta | aatg | tatgg | tgggggag | ttaa | agtgtt |
| CVF1 Gen | | atgtcag | gat | | | tctatt | catc |
| CVF1 Gen | | | | | | | |
| CVF1 Gen | | | | | | (ca. 2000 bp) | |
| CVF1 Gen | | | | | | | ccaaggagat |
| CVF1 Gen | | gttacag | cca | gacaat | caaaa | ctttcag | ct |
| CVF1 cDNA | | | | | | | - |
| CVF2 cDNA | | | | | | | A 3115 |
| coC3 cDNA | | | | | | | A 4090 |
| CVF1 Gen | | gtgcagagtc | cgattttctc | ttttcttttg | taatttttaa | acttcaaag- | |
| CVF1 cDNA | 4067 | -----AT | GCAATGGGAG | CCAAGGGAGC | CCTCATGCTC | AAGATCTGCA | 4108 |
| CVF2 cDNA | 3116 | CTTAAAAGAG | GCAAAGGGAG | CCAAGGGAGC | CCTCAAGCTC | AAATCTGCA | 3165 |
| coC3 cDNA | 4091 | CTTAAAACAG | GCAAAGGGAG | CCAAGGGAGC | CCTCAAGCTC | AAATCTGCA | 4140 |
| CVF1 Gen | | -----AT | GCAATGGGAG | CCAAGGGAGC | CCTCATGCTC | AAGATCTGCA | |

Abb. 3.32: Sequenz des genomischen PCR-Produkts mit Intron 32 des CVF1-Gens. Intron 32 hat eine Größe von ca. 2200 bp. Eine DraI-Schnittstelle (ttaaa) im 5'-Bereich des Introns ist unterstrichen dargestellt (vgl. GenomeWalking Abb. 3.35).

3.1.4 Aufklärung der CVF1-Genstruktur durch Genome Walking

Als weitere Methode zur Aufklärung der CVF1-Genstruktur wurde Genome Walking eingesetzt. Diese Methode ist geeignet, um unbekannte genomische DNA-Sequenzen in Nachbarschaft zu einer bekannten Sequenz aufzufinden.

Durch die Analyse des Cosmidklons 1R67 sowie mittels genomischer PCR konnte der überwiegende Teil der CVF1-Genstruktur aufgeklärt werden. Jedoch war es mit genomischer PCR nicht möglich, die noch nicht identifizierten CVF1-Intron 2, 8, 27 und 36 zu amplifizieren. Für diese und weitere CVF1-Introns sowie für die Promotorregion und den 3'-Bereich wurden Sequenzinformationen durch Genome Walking erhalten. Nachfolgend sind die Ergebnisse beginnend mit dem 5'-Ende des CVF1-Gens aufgeführt. Die ebenfalls erhaltenen Amplifikate homologer Gene werden in den nachfolgenden Kapiteln 3.2.2, 3.3.2 und 3.4 beschrieben.

Tabelle 3.4 gibt eine Übersicht über die durch Genome Walking aufgeklärten Sequenzen des CVF1-Gens.

Tab. 3.4: Übersicht der Genome Walking-Amplifikate des CVF1-Gens. Das jeweils größte sequenzierte PCR-Produkt sowie die dazu eingesetzte Genome Walking (GW)-Bibliothek sind angegeben. Größenangaben sind Schätzungen.

| Exon-Intron-Grenze | Primerpaare | Größe PCR-Produkt (GW-Bibliothek) | Exon-Intron- Übergangssequenz |
|--------------------------|--|--------------------------------------|---|
| Promotor/Exon 1/Intron 1 | CVF-Intr.1as-I/AP1 CVF-Intr.1as-II/AP2 | 1100 bp (PvuII) | ¹⁾ TCT gtaagtatca... |
| Exon 1/Intron 1 | HB1s-II/AP1 HB1s/AP2 | 1050 bp (DraI) | TCT gtaagtatca... |
| Exon 2/Intron 2 | HB2s-CVF/AP1 HB2sII/AP2 | 680 bp (StuI) | GAG gtgggtgtat... |
| Intron 2/Exon 3 | HB3as-CVF/AP1 HB3as-CVF-II/AP2 | 1300 bp (PvuII) | ...tgtcatccag ATT |
| Exon 8/Intron 8 | HB8s-CVF-Intr./AP1 HB8s-CVF-Intr./AP2 | 1800 bp (PvuII) | CCG gtaaaatgta... |
| Intron 8/Exon 9 | HB9as-CVF/AP1 HB9as-II-CVF/AP2 | 420 bp (DraI) | ...ttattttcag ATT |
| Intron 27/Exon 28 | HB28as-CVF/AP1 HB28as-CVF/AP2 | 550 bp (PvuII) | ...tccttcag AGT |
| Exon 31/Exon 32/Intr. 32 | HB31s-CVFgenom./AP1 HB31s-CVFgenom./AP2 | 850 bp (StuI) | ²⁾ TGA gtaagttga... |
| Exon 36/Intron 36 | HB36s-CVF/AP1 HB36s-CVF/AP2 | 1200 bp (DraI) | TAG gtaatggcag... |
| Intron 36/Exon 37 | HB37as-CVF/AP1 HB37as-CVF/AP2 | 900 bp (DraI) | ...cttcattag ATG |
| Terminator | HB-CVF-41s-I/AP1 HB-CVF-41s-II/AP2 | 1500 bp (EcoRV) | ³⁾ |

1) Die Promotorregion des CVF1-Gens ist in Abbildung 3.34 dargestellt. Die Sequenzangaben hier beziehen sich auf Exon 1/Intron 1.

2) Kein Intron 31 enthalten. Die Sequenzangaben beziehen sich auf Exon 32/Intron 32.

3) Die Sequenz des Terminationsbereichs des CVF1-Gens ist in Abbildung 3.52 gezeigt.

Identifizierung der Promotor-Region von CVF1

Die cDNA-Sequenzen von CVF1 und coC3 unterscheiden sich im Exon 1-Bereich (vgl. Abb. 3.19) in nur vier Nukleotiden, die alle für eine zweifelsfreie Zuordnung der Promotorregion zum entsprechenden Gen berücksichtigt werden sollten. Für die Aufklärung der Promotor-Region des CVF1-Gens wurde daher wie weiter unten beschrieben zunächst die Sequenz von Intron 1 aufgeklärt (Abb. 3.36). Die Aufklärung der Intron 1-Sequenz des CVF1-Gens durch Genome Walking und genomische PCR ermöglichte die Generierung CVF1-spezifischer Intron 1-Primer (CVF-Intr.1as-I und CVF-Intr.1as-II), die zur Aufklärung der Promotorregion verwendet wurden. Mit Hilfe dieser Primer sowie den Adapter-Primern (AP1 und AP2) wurden verschachtelte PCRs mit allen vier Genome Walking-Bibliotheken durchgeführt und mit drei Bibliotheken Amplifikate erhalten (Abb. 3.33).

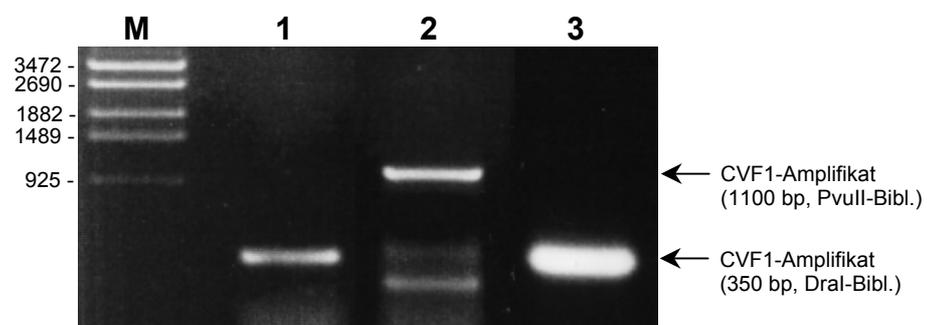


Abb. 3.33: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Intron 1 stromaufwärts). 1%iges Agarosegel, M: DNA-Marker λ /Eco130I, Bahn 1: EcoRV-Bibliothek, Bahn 2: PvuII-Bibliothek, Bahn 3: DraI-Bibliothek.

Das ca. 1100 bp schwere Amplifikat aus der PvuII-Bibliothek (Bahn 2) wurde kloniert und sequenziert. In Abbildung 3.34 ist die Sequenz dieses Amplifikats dargestellt.

3 Ergebnisse

| | | | | | | | |
|---------------|------|--------------------|-----------------------------|---------------------|--------------------|---------------------|------|
| CVF1 Gen | -892 | <u>CAGCTG</u> | GCAGAGTGCT | GCTGGAGGCT | GGGAGGCCG | AAAATGGCC | -847 |
| CVF1 Gen | -846 | GTGGAGGTTG | GGGGCAACAT | GCGCCCCTTT | ATGGCCGTT | TTGGCCAGCA | -797 |
| CVF1 Gen | -796 | GAGGCACTGC | AGGCCAGTCC | TTTGATATTT | CCAGGCCGTC | ACCGCAGGCC | -747 |
| | | CACCC | | | | | |
| CVF1 Gen | -746 | AAGTTTAAGC | ACCC ANCTGG | CTGGCCCCAG | CCTGCGGGCC | TTGAGTTTGN | -697 |
| | | | | PEA-3 | | CACCC (-) | |
| CVF1 Gen | -696 | CAGCCCTGCA | CTAGAAAGAA | AGGAAA TAT | TATTCCTGTG | T GGTGG TGC | -647 |
| CVF1 Gen | -646 | TTTGTTAAAA | GACCCACGT | CCAAGTGTT | GCCAGCTACA | AAAAGGCTAA | -597 |
| CVF1 Gen | -596 | AAGAGTACTG | CACAGTGAAG | CACATTCCTT | TATTTTTTCT | TACTTGCTTG | -547 |
| CVF1 Gen | -546 | CTTACTTGTT | TGTTTATTTT | TCTATTCATT | CACTTGTTTC | TTCCTTAGC | -497 |
| | | | | TATA | | GATA-1 (-) | |
| CVF1 Gen | -496 | ATTGTCTGTT | TCAAAGATAC | TCTACA TATA | AATA ATATAG | AA TTATCT GT | -447 |
| | | | | GR (-) | | | |
| CVF1 Gen | -446 | ACATTAAAAA | ATCTAATT TG | TTCT AACATA | TTAAAATCAC | ATCCAGAAAG | -397 |
| CVF1 Gen | -396 | TATCATCAGA | GTCCCGTATC | GCCTGTGGGA | AACCAACCCC | TCCCAGAAGA | -347 |
| | | | | PEA-3 | | | |
| CVF1 Gen | -346 | ATGAAATATT | TT AGGAAA TATA | GTAAAAGGC | AGAATATTAT | ATTTCCCCTA | -297 |
| | | | | | | Sp1 (-) | |
| CVF1 Gen | -296 | AA AGATAAAT | AGAA TCTTG | GGGAGACAGA | AACTCAGCCC | CAGC TCCCTG | -247 |
| | | | | | | GATA-1 (-) | |
| CVF1 Gen | -246 | CCCC CAGCAG | GTATTTGGTG | CCCATTAAGA | AAGACTGGGT | CATCC TATCT | -197 |
| CVF1 Gen | -196 | ACAAGACAAA | AGACCTTCAA | CCTCAGAGTG | ATGGGATTAG | CAAGACAAAT | -147 |
| CVF1 Gen | -146 | AATCCATTGT | GAGTTTAAA | GGTATTAATT | TTGTTGGTAA | TTCTCAAGCT | -97 |
| CVF1 Gen | -96 | CCTCCTTTAT | ACTGATTAAA | CATTCAGCAA | TTGCAAAGGG | GCTTGCTTTC | -47 |
| | | | | TATA | | TIS (+1) | |
| CVF1 Gen | -46 | AAAGCAGGCT | CTCTGG ATAA | ATACT GGAGT | TGTCAGCCCT | TCTCT GCCC | 4 |
| Exon 1 | | | | | | | |
| CVF1 cDNA | 1 | | C CCATGGAGAG | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | 41 |
| CVF2 Gen | | | A CCATGGAG GG | GA C GGCTCTC | T CTCTGGTGG | CTGCTCTATT | |
| coC3 cDNA | 1 | GGACTA | CCATGGAG GG | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | 46 |
| CVF1 Gen | 5 | TTCTGGACTA | CCATGGAGAG | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | 54 |
| CVF1 cDNA | 42 | GATTGGTTTT | CCAGGGTCTT | CTCATGGGGC | TCT | | 74 |
| CVF2 Gen | | GATTGGTTTT | CCAG G ATCTT | CTCATGG G A | TCT | | |
| coC3 cDNA | 47 | GATTGGTTTT | CCAGGGTCTT | CCCA GGGGC | TCT | | 79 |
| CVF1 Gen | 55 | GATTGGTTTT | CCAGGGTCTT | CTCATGGGGC | TCTgtaagta | tcatagatat | 104 |
| CVF1 Gen | 105 | ttcagcaggg | ggaatgaagc | agtgctgggg | aagcgtagtg | ggctttctct | 154 |
| CVF1 Gen | 155 | tccgtgatgg | tatgtattaa | acggaactat | cacatccaag | gctgtggtcc | 204 |

Abb. 3.34: Darstellung der Promotorregion des CVF1-Genes. Die Promotorregion stromaufwärts von Exon 1 ist in Großbuchstaben dargestellt. Putative Bindungsstellen für Transkriptionsfaktoren, TATA-Boxen sowie der putative Transkriptions-Initiations-Start (TIS) sind fett und kursiv dargestellt. Die Gen-Sequenz weicht in einem Nukleotid von der CVF1-cDNA (Position 1) ab (umrahmt). Die PvuII-Schnittstelle (cagctg) am 5'-Ende sowie die DraI-Schnittstelle (tttaa) sind unterstrichen. Die Sequenz des inneren Primers CVF-Intr. Ias-II ist als Pfeil markiert.

Zur vollständigen Sequenz-Bestimmung wurde die Sequenzierung zusätzlich mit dem internen Primer HB-CVF1-Prom.as durchgeführt.

Das ca. 350 bp schwere Amplifikat der DraI-Bibliothek korreliert in seiner Größe mit der Position einer entsprechenden Schnittstelle (tttaa) in der CVF1-Sequenz. Eine EcoRV-Schnittstelle ist in obiger Sequenz nicht vorhanden, sodass dieses Amplifikat wahrscheinlich homologen Genen (CVF2, coC3) zuzuordnen ist.

Exon 1/Intron 1

Zusätzlich zur Amplifizierung von Intron 1 des CVF1-Gens durch genomische PCR (Kap. 3.1.2) wurde der Exon1/Intron1-Übergang der CVF/C3-Genfamilie durch Genome Walking charakterisiert. Als exonspezifische Primer wurden solche Primer verwendet, die am 3'-Ende nicht zwischen CVF1 und coC3 diskriminieren. Eine verschachtelte PCR mit den Primerpaaren HB1s-II/AP1 und HB1s/AP2 führte erwartungsgemäß zu Amplifikaten der beiden Gene CVF1 und coC3.

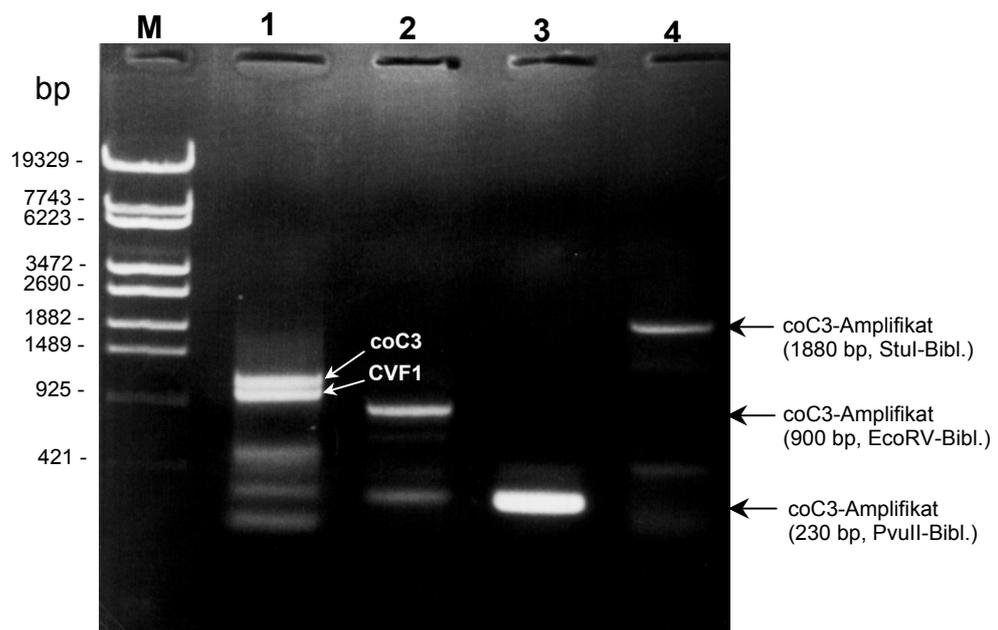


Abb. 3.35: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Exon 1 stromabwärts). M: DNA-Marker λ /Eco130I, Bahn 1: DraI-Bibliothek, Bahn 2: EcoRV-Bibliothek, Bahn 3: PvuII-Bibliothek, Bahn 4: StuI-Bibliothek. Die markierten Banden der DraI-Bibliothek (Bahn 1) und StuI-Bibliothek (Bahn 4) wurden sequenziert. Die markierten Banden der EcoRV-Bibliothek (Bahn 2) und PvuII-Bibliothek (Bahn 3) stimmen in ihrer Größe mit den entsprechenden EcoRV- und PvuII-Schnittstellen der coC3-Gensequenz überein.

Nach einer weiteren Auftrennung wurden die beiden schweren Banden der DraI-Bibliothek (ca. 1050 bzw. 1200 bp) kloniert und sequenziert sowie das schwere Amplifikat der StuI-Bibliothek (ca. 1880 bp) charakterisiert. Das 1050 bp-Amplifikat der DraI-Bibliothek wurde als CVF1-Genabschnitt identifiziert. Bei den anderen markierten Amplifikaten handelte es sich um coC3-Genabschnitte, deren Beschreibung in Kapitel 3.3.2 erfolgt.

Die nachfolgend gezeigte Gensequenz des DraI-Amplifikats (1050 bp) ist bis auf eine Abweichung (Position 146 in Intron 1, siehe Anhang) identisch mit dem genomischen PCR-Amplifikat mit CVF1-Intron 1 (Kapitel 3.1.3). Aufgrund von Unterschieden zur CVF2-Gensequenz sowie zur coC3-cDNA wurde die Sequenz dem CVF1-Gen zugeordnet.

| | | | | | | | |
|-----------|----|------------|------------|------------|------------|------------|----|
| CVF1 cDNA | 4 | ATGGAGAGGA | TGGCTCTCTA | TCTGGTGGCT | GCTCTATTGA | TTGGTTTCC | 53 |
| CVF2 Gen | | ATGGAGCGGA | CGGCTCTCTG | TCTGGTGGCT | GCTCTATTGA | TTGGTTTCC | |
| coC3 cDNA | 9 | ATGGAGCGGA | TGGCTCTCTA | TCTGGTGGCT | GCTCTATTGA | TTGGTTTCC | 58 |
| CVF1 Gen | | | | GGTGGCT | GCTCTATTGA | TTGGTTTCC | |
| CVF1 cDNA | 54 | AGGGTCTTCT | CATGGGGCTC | T | | | 74 |
| CVF2 Gen | | AGGATCTTCT | CATGGGACTC | T | | | |
| coC3 cDNA | 59 | AGGGTCTTCG | CACGGGGCTC | T | | | 79 |
| CVF1 Gen | | AGGGTCTTCT | CATGGGGCTC | Tgtaagtatc | atagatattt | cagcaggggg | |
| CVF1 Gen | | aatgaagcag | tgctggggaa | gcgtagtggg | ctttctcttc | cgtgatggta | |
| CVF1 Gen | | tgtattaaac | ggaactatca | catccaaggc | tgtggtcctt | | |

Abb. 3.36: Darstellung der partiellen Intron 1-Sequenz des CVF1-Gens.

Die hier zu Vergleichszwecken dargestellte CVF2-Gensequenz ist in Kapitel 3.2 näher charakterisiert.

Intron 1/Exon 2

Bei der verschachtelten PCR mit den Primerpaaren HB2as-CVF-II/AP1 und HB2as-CVF/AP2 wurden nur solche Amplifikate erhalten, die nicht dem CVF1-Gen zuzuordnen waren. Eine Charakterisierung erfolgt in Kapiteln 3.2.2 und 3.4.

Exon 2/Intron 2

Da eine Amplifikation von Intron 2 des CVF1-Gens durch genomische PCR nicht möglich war, wurden die interessierenden Exon/Intron-Übergänge durch Genome Walking aufgeklärt. Durch verschachtelte PCRs mit den Primerpaaren HB2s-CVF/AP1 und HB2sII/AP2 wurden die in Abbildung 3.37 gezeigten Amplifikate erhalten.

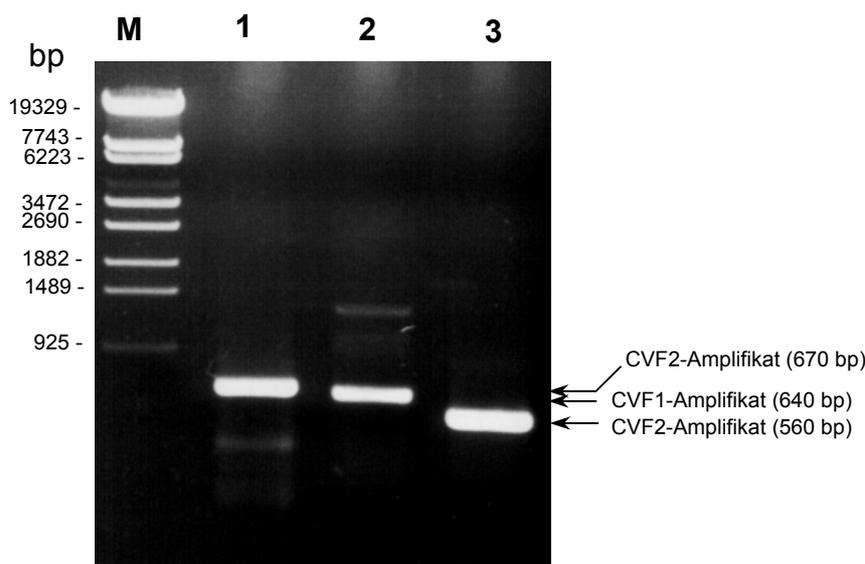


Abb. 3.37: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Exon 2 stromabwärts). M: DNA-Marker λ /Eco130I, Bahn 1: DraI-Bibliothek, CVF2-Amplifikat (670 bp), Bahn 2: StuI-Bibliothek, CVF1-Amplifikat (640 bp), Bahn 3: PvuII-Bibliothek, CVF2-Amplifikat (560 bp).

Alle drei Hauptamplifikate wurden kloniert und sequenziert. Die Sequenz des StuI-Amplifikats (Abb. 3.37, Bahn 2) ist nachfolgend in Abbildung 3.38 gezeigt und konnte dem CVF1-Gen zugeordnet werden.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------|------|-----|-----|------|------|------|------|------|------|-----|------|------|-----|-----|-----|----|-----|------|-------|-----|---|---|---|---|---|---|---|---|---|---|---|---|-----|---|-----|
| CVF1 | cDNA | 137 | CCC | ATGG | GAGA | CAG | TACT | TCCA | AA | CAG | CTTG | AC | ATC | TTT | GT | TC | ATG | ATTT | 186 | | | | | | | | | | | | | | | | |
| CVF2 | Gen | | CC | TAT | GG | GAGA | CAG | TACT | TCCA | AA | CAG | CTTG | AC | ATC | TTT | GT | TC | CA | AATTT | | | | | | | | | | | | | | | | |
| coC3 | cDNA | 142 | CCC | ATGG | GAGA | CAG | TACT | TCCA | AA | TC | CG | CTTG | AC | ATC | TTT | GT | TC | ATG | ATTT | 191 | | | | | | | | | | | | | | | |
| CVF1 | Gen | | CCC | ATGG | GAGA | CAG | TACT | TCCA | AA | CAG | CTTG | AC | ATC | TTT | GT | TC | ATG | ATTT | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CVF1 | cDNA | 187 | CC | AC | GG | AA | GC | AG | AA | AC | CTT | GT | T | C | C | AA | AC | AG | AG | T | A | G | A | T | G | A | T | C | C | A | G | C | 236 | | |
| CVF2 | Gen | | CC | AC | AG | AA | GC | AG | AA | AC | CTT | GT | A | C | C | AA | AC | AG | AG | T | A | G | A | T | G | A | T | C | C | A | G | A | | | |
| coC3 | cDNA | 192 | CC | AC | GG | AA | GC | AG | AA | AC | CTT | GT | T | C | C | AA | AC | AG | AG | T | A | G | A | T | G | A | T | C | C | A | G | C | 241 | | |
| CVF1 | Gen | | CC | AC | GG | AA | GC | AG | AA | AC | CTT | GT | T | C | C | AA | AC | AG | AG | T | A | G | A | T | G | A | T | C | C | A | G | C | | | |
| CVF1 | cDNA | 237 | AG | G | A | G | G | C | A | T | G | C | A | T | T | A | G | A | G | A | G | G | C | A | T | G | C | A | T | A | T | A | G | A | 267 |
| CVF2 | Gen | | G | G | A | G | G | C | A | T | G | C | A | T | T | A | G | A | G | A | G | G | C | A | T | G | C | A | T | A | T | A | G | A | |
| coC3 | cDNA | 242 | AG | G | A | G | G | C | A | T | G | C | A | T | T | A | G | A | G | A | G | G | C | A | T | G | C | A | T | A | T | A | G | A | 272 |
| CVF1 | Gen | | AG | G | A | G | G | C | A | T | G | C | A | T | T | A | G | A | G | A | G | G | C | A | T | G | C | A | T | A | T | A | G | A | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CVF1 | Gen | | tg | ca | a | t | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | |
| CVF1 | Gen | | ct | g | t | c | c | a | g | t | t | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | a | g | |

Abb. 3.38: Darstellung der partiellen Intron 2-Sequenz des CVF1-Gens (5'-Bereich mit Exon 2). Exon 2 hat eine Größe von 193 bp.

Die in Abbildung 3.38 für Vergleichszwecke aufgeführte Exon-Sequenz des CVF2-Gens wurde u.a. durch die Sequenzierung der obigen CVF2-Amplifikate der DraI- und PvuII-Bibliotheken erhalten. Die Intron-Sequenz des CVF2-Amplifikats ist in Kapitel 3.2.2 aufgeführt.

Intron 2/Exon 3

Eine verschachtelte PCR mit den Primerpaaren HB3as-CVF/AP1 und HB3as-CVF-II/AP2 führte zu den in Abbildung 3.39 gezeigten Amplifikaten:

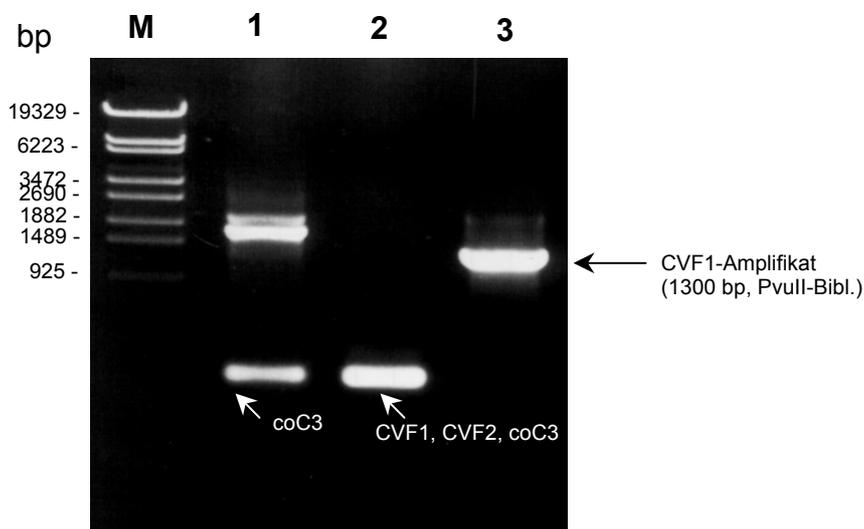


Abb. 3.39: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Intron 2/Exon 3). 1%iges Agarosegel. M: DNA-Marker λ /Eco130I; Bahn 1: DraI-Bibliothek, Bahn 2: StuI-Bibliothek, Bahn 3: PvuII-Bibliothek.

Das Amplifikat der PvuII-Bibliothek (1300 bp) wurde kloniert und sequenziert und konnte dem CVF1-Gen zugeordnet werden.

| | | | | | | | |
|------------------|-------|---------------------|--------------------|--------------------|--------------------|--------------------|-----|
| CVF1 Gen | | tttttgatc | aatgttttag | tgttgttgcc | taggctgag | | |
| CVF1 Gen | | gaatgaggcc | tcaagtggcc | tttagtgatg | attaagtttg | aaacctctgc | |
| CVF1 cDNA | 268 | | | ATTCCAGCA | AAAGAAGTGA | GTACGGACTC | 296 |
| CVF2 Gen | | | | GTTCC TGCA | AAAGAA CTGA | ATA AGGACTC | |
| coC3 cDNA | 273 | | | GTTCC TGCA | AAAGAA CTGA | ATA AGGACTC | 301 |
| CVF1 Gen | | tctactcatc | ttgtcatcca | g ATTCCAGCA | AAAGAAGTGA | GTACGGACTC | |
| CVF1 cDNA | 297 | CAGGCAAAAAT | CAATATGTGG | TTGTGCAAGT | AACTGGTCCT | CAAGTGAGAT | 346 |
| CVF2 Gen | | CA AGCAAAAAT | CAA... | | | | |
| coC3 cDNA | 302 | CA AGCAAAAAT | CA CTATGTGG | TTGTG AAAGT | AACTGGTCCT | CAAGTG GCAT | 351 |
| CVF1 Gen | | CAGGCAAAAAT | CAATATGTGG | TTGTGCAAGT | AACTGGTCCT | C | |

Abb. 3.40: Darstellung der partiellen Intron 2-Sequenz des CVF1-Gens (Intron 2/Exon 3). Eine StuI-Schnittstelle (aggcct) im Intron ist unterstrichen dargestellt.

Eine StuI-Schnittstelle (aggcct) befindet sich in den Gensequenzen von CVF1 und auch von CVF2 und coC3 66 bp stromaufwärts von Exon 3 (siehe Abb. 3.40, Anhang 7.2 und Anhang 7.3). Die Position dieser StuI-Schnittstelle korreliert damit mit der Größe eines 200 bp-Amplifikats mit der StuI-Bibliothek (Abb. 3.39, Bahn 2).

Das ca. 300 bp-Amplifikat der DraI-Bibliothek (Abb. 3.39, Bahn 1) korreliert in seiner Größe mit der Position einer DraI-Schnittstelle im coC3-Gen, die sich 182 bp stromaufwärts von Exon 3 befindet (siehe Anhang 7.3).

Aufgrund der Spezifität der Methode werden allerdings nur Amplifikate für CVF1 und CVF2 erwartet.

Exon 8/Intron 8

Mit dem genspezifischen Primer HB8s-CVF-Intr. und den Adapterprimern AP1 und AP2 wurde eine semi-verschachtelte PCR durchgeführt. Mit den EcoRV- und PvuII-Bibliotheken wurden zunächst in schwacher Ausbeute Amplifikate von 1,5 bzw. 1.8 kb Länge erhalten, die gereinigt und als Template für weitere Amplifikationen mit dem Primerpaar HB8s-CVF-Intr./AP2 eingesetzt wurden.

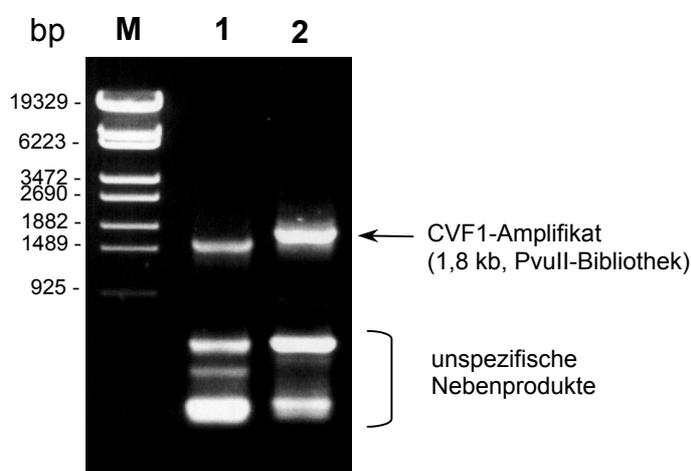


Abb. 3.41: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Exon 8/Intron 8). 1%iges Agarosegel. M: DNA-Marker λ /Eco130I, Bahn 1: EcoRV-Bibliothek, Bahn 2: PvuII-Bibliothek.

Das 1,8 kb-Produkt (Abb. 3.41) wurde sequenziert und konnte aufgrund der Übereinstimmung mit der entsprechenden CVF1-cDNA dem CVF1-Gen zugeordnet werden.

| | | | | | | | |
|------------------|-----|-------------------|-------------------|-------------------|-------------------|-------------------|----------------|
| CVF1 cDNA | 753 | GTACTTGT | ATGGAGAGGA | AGTGAAGGT | GTGGCCTTTG | 790 | |
| coC3 cDNA | 764 | GTACTTAT | ATGGAAAGGA | AGTGAAGGT | GTGGCCTTTG | 801 | |
| CVF1 Gen | | gtgtcattac ag | <u>GTACTTGT</u> | <u>ATGGAGAGGA</u> | <u>AGTGAAGGT</u> | <u>GTGGCCTTTG</u> | |
| | | | → | | | | |
| CVF1 cDNA | 791 | TCCTCTTTGG | AGTGAAAATA | GATGATGCTA | AAAAGAGTAT | TCCAGACTCA | 840 |
| coC3 cDNA | 802 | TCCTCTTTGG | AGTGAAAATA | GATGATGCTA | AAAAGAGTAT | TCCAGACTCA | 851 |
| CVF1 Gen | | TCCTCTTTGG | AGTGAAAATA | GATGATGCTA | AAAAGAGTAT | TCCAGACTCA | |
| CVF1 cDNA | 841 | CTCACGAGAA | TTCCG | | | | 855 |
| coC3 cDNA | 852 | CTCACGAGAA | TTCCG | | | | 866 |
| CVF1 Gen | | CTCACGAGAA | TTCCG | gtaaa | atgtaacag | ggtgttt | gtagcattggagtg |
| CVF1 Gen | | tgccctgtggg | ttttt | gtgtc | agatctacta | cgggttgata | ggttaaggcgg |
| CVF1 Gen | | ggctgacatc | ctagcgatac | aacatgcctc | acagagatct | | |

Abb. 3.42: Darstellung der partiellen Intron 8-Sequenz des CVF1-Gens. Eine EcoRI-Schnittstelle (GAATTC) in Exon 8 ist unterstrichen dargestellt.

Eine EcoRV-Schnittstelle (gatatc) ist an der erwarteten Position nicht vorhanden, so dass das EcoRV-Amplifikat wahrscheinlich den Genabschnitt eines homologen Gens enthält.

Zusätzlich wurde ein ca. 460 bp-Amplifikat aus der DraI-Bibliothek erhalten (ohne Abbildung) und direkt mit dem Primer AP2 sequenziert. Beide Sequenzen waren bis auf ein Nukleotid im Intron ($\underline{n} = a/g$) identisch. Da es bei der Konsensus-Sequenz keine Widersprüche zur CVF1-Sequenz gab, wurde obige Sequenz dem CVF1-Gen zugeordnet.

Intron 8/Exon 9

Mit den Primerpaaren HB9as-CVF/AP1 und HB9as-II-CVF/AP2 wurde nach Durchführung einer verschachtelten PCR mit der DraI-Bibliothek ein Amplifikat von ca. 420 bp erhalten (ohne Abbildung), dessen partielle Sequenz in Abb. 3.43 dargestellt ist.

| | | | | | | | |
|------------------|-------|-------------------|-------------------|-------------------|-------------------|--------------------|-----|
| CVF1 Gen | | ttaaacata | ttttaccct | tgctgctatg | gccacagagc | | |
| CVF1 Gen | | ttgcatatct | tcatocatca | tcttcttat | tctacattgt | ggcagaccac | |
| CVF1 cDNA | 856 | | | | ATTATT | 861 | |
| coC3 cDNA | 867 | | | | ATTATT | 872 | |
| CVF1 Gen | | ttgatttct | tccttatttt | tacttatttt | tcgtttattt | tcag ATTATT | |
| CVF1 cDNA | 862 | GATGGAGATG | GGAAAGCAAC | ACTAAAAAGA | GATACATTCC | GTCCTCGATT | 911 |
| coC3 cDNA | 873 | GATGGAGATG | GGCAAGCAAC | ACTAAAAAGA | GATACACTAC | GTCCTCGATT | 922 |
| CVF1 Gen | | GATGGAGATG | GGAAAGCAAC | ACTAAAAAGA | GATACATTCC | GTCCTCGATT | |
| CVF1 cDNA | 912 | TCCAAATCTC | AATGAGCTTG | TTGGGCATAC | TCTGTATGCA | TCTGTAACAG | 961 |
| coC3 cDNA | 923 | TCAAGATCTC | AATCAGCTTG | TTGGTCATAC | TCTGTATGTA | TCTGTAACAG | 972 |
| CVF1 Gen | | TCCAAATCTC | AATGAGCTTG | TTGGGCATAC | TCTGTATGCA | TC | |

Abb. 3.43: Darstellung der partiellen Intron 8-Sequenz (Intron 8/Exon 9) des CVF1-Gens.

Das Amplifikat ist im Bereich von Exon 9 mit der CVF1-cDNA identisch und wurde daher dem CVF1-Gen zugeordnet. Zusammenfassend wurde für das CVF1-Intron 8 eine Mindestgröße von 2.1 kb ermittelt.

Intron 27/Exon 28

Mit dem exonspezifischen Primer HB28as-CVF und den Adapter-Primern AP1 und AP2 wurde eine semi-verschachtelte PCR durchgeführt.

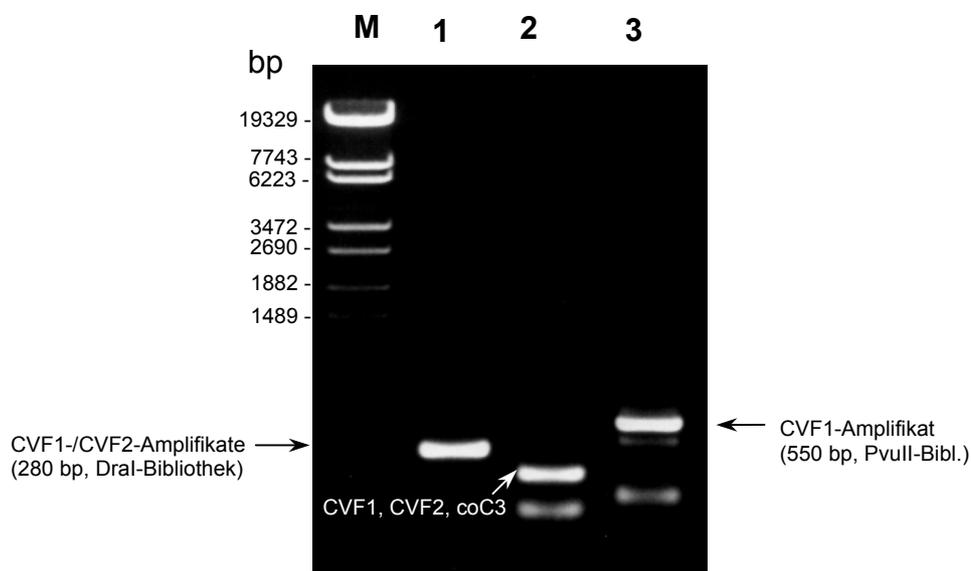


Abb. 3.44: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Intron 27/Exon 28). 1%iges Agarosegel. M: DNA-Marker λ /Eco130I, Bahn 1: DraI-Bibliothek, Bahn 2: StuI-Bibliothek, Bahn 3: PvuII-Bibliothek.

Das größte Amplifikat der PvuII-Bibliothek (ca. 550 bp) wurde kloniert und sequenziert. Die Intron/Exon-Übergangssequenz ist nachfolgend in Abbildung 3.45 dargestellt.

| | | | | | | |
|------------------|-------|-------------------|---------------------------|---------------------------|-------------------|------------------------|
| CVF1 Gen | | tgatagctca | gggtatcata | cctcagggaa | atgctgttcc | |
| CVF1 Gen | | atagagatgc | ctaaacatt | cctaattggat | ctcttctott | ccctctcccc |
| CVF1 cDNA | 3442 | | AGTCTAGACA | GCAGCATCAA | GAAGGCCACA | AATTATTTAC 3481 |
| CVF2 cDNA | 2490 | | ATTCTAGACA | GCAGCATCA G | TAAGGCCACA | GATTATTTAC 2529 |
| coC3 cDNA | 3465 | | ATTCTAGACA | GCAGCATCAA | TAAGGCCACA | GATTATTTAC 3504 |
| CVF1 Gen | | tcctttccag | AGTCTAGACA | GCAGCATCAA | GAAGGCCACA | AATTATTTAC |
| CVF1 cDNA | 3482 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC 3531 |
| CVF2 cDNA | 2530 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC 2579 |
| coC3 cDNA | 3505 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC 3554 |
| CVF1 Gen | | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC |
| CVF1 cDNA | 3532 | TATGCTTTGG | CTGCTGCAGA | CCAACTCAAT | GATGACAGGG | TACTCATGGC 3581 |
| CVF2 cDNA | 2580 | TATGCTTTGG | CTGCTGCAG G | ACCACTCAAT | GATGACAGGG | TACTCATGGC 2629 |
| coC3 cDNA | 3555 | TATGCTTTGG | CTGCTGCAGA | CCCACTCAAT | GATGACAGGG | TACTCATGGC 3604 |
| CVF1 Gen | | TATGCTTTGG | CTGCTGCAGA | CCAACTCAAT | GATGACAGGG | TACTCATGGC |

Abb. 3.45: Darstellung der partiellen Intron 27-Sequenz des CVF1-Gens (Intron 27/Exon 28). Eine StuI-Schnittstelle (agcct) in Exon 28 der drei Gene ist unterstrichen.

Das Amplifikat wurde aufgrund seiner spezifischen Exonsequenz dem CVF1-Gen zugeordnet. Sowohl in den Gensequenzen von CVF1 als auch von CVF2 (Anhang 7.1 und

7.2) befindet sich eine DraI-Schnittstelle (tttaaa) 232 bp stromaufwärts von Exon 28. Die Position dieser Schnittstelle korreliert mit der Größe der Bande der DraI-Bibliothek (Bahn 1), so dass es sich dabei vermutlich um Amplifikate beider Gene handelt. Aufgrund einer StuI-Schnittstelle (aggcct) in Exon 28 aller drei Gene ist die Bildung von jeweils ca. 110 bp-Amplifikaten mit der StuI-Bibliothek möglich (Abb. 3.44, Bahn 2). Für Intron 27 des CVF1-Gens wurde eine Mindestgröße von 1333 bp ermittelt.

Exon 31 stromabwärts

Bei der semi-verschachtelten PCR mit den Primerpaaren HB31s-CVFgenom./AP1 und HB31s-CVFgenom./AP2 zur Charakterisierung der Region stromabwärts von Exon 31 wurden Amplifikate mit der DraI-Bibliothek (250 bp) und mit der StuI-Bibliothek (850 bp) erhalten (Abb. 3.46).

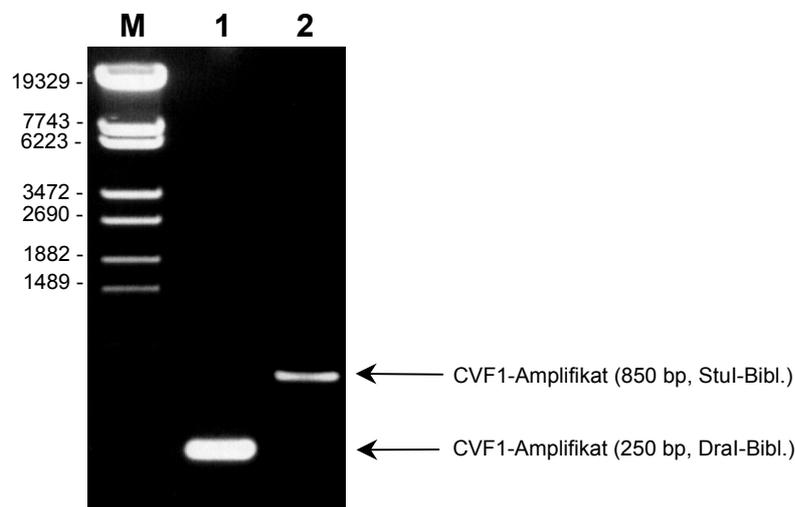


Abb. 3.46: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Exon 31 stromabwärts). 1%iges Agarosegel. M: λ -DNA/*Eco130I*-Marker; Bahn 1: DraI-Bibliothek, Bahn 2: StuI-Bibliothek.

Das Amplifikat der StuI-Bibliothek bei 850 bp wurde kloniert und sequenziert. Die Sequenz, in der auch die DraI-Schnittstelle nach ca. 250 bp identifiziert wurde, konnte zweifelsfrei dem CVF1-Gen zugeordnet werden. Es zeigte sich, dass im CVF1-Gen im Gegensatz zum humanen C3-Gen kein Intron 31 vorhanden ist und bestätigte die Ergebnisse der überlappenden genomischen PCR (Abb. 3.31 und Abb. 3.32).

Exon 36/Intron 36

Im gesamten Exon 36 unterscheiden sich die cDNA-Sequenzen von CVF1 und CVF2 nur in einem singulären Nukleotid von coC3 (Abb. 3.29). Für CVF1 und CVF2 gibt es im Exonbereich keine Unterscheidungsmerkmale. Der Primer HB36s-CVF wurde so gewählt,

dass eine Verlängerung nur im CVF1- bzw. im CVF2-Gen möglich sein sollte. Eine semi-verschachtelte PCR wurde durchgeführt, wobei in beiden Fällen der Primer HB36s-CVF als sense-Primer fungierte. Anstelle von maximal zwei erwarteten PCR-Amplifikaten je Genome Walking-Bibliothek (für CVF1 und CVF2) wurden mit der DraI-Bibliothek drei Amplifikate der Größen von ca. 1200, 1050 und 900 bp erhalten (Abb. 3.47).

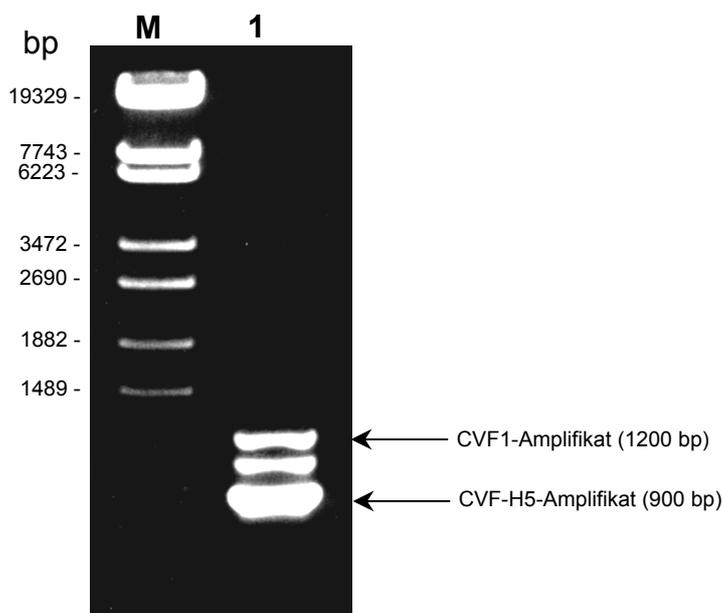


Abb. 3.47: Gelelektrophoretische Analyse der Genome Walking-Amplifikate (Exon 36/Intron 36). 1%iges Agarosegel. M: DNA-Marker λ Eco130I, Bahn 1: DraI-Bibliothek.

Die markierten Banden wurden kloniert und sequenziert. Das 1200 bp-Amplifikat wurde aufgrund der nachfolgend dargestellten Sequenz als CVF1-Genabschnitt identifiziert.

| | | | | | | | |
|------------------|------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| CVF1 cDNA | 4305 | AGATGAATGC | CTGCACTTTA | AGATTCTCAA | GCATTTTGAA | GTTGGCTTCA | 4354 |
| CVF2 cDNA | 3362 | AGATGAATGC | CTGCACTTTA | AGATTCTCAA | GCATTTTGAA | GTTGGCTTCA | 3411 |
| coC3 cDNA | 4337 | AGATGAATGC | CTGCACTTTA | AGATTCTCAA | GCATTTTGAA | GTTGGCTTCA | 4386 |
| CVF1 Gen | | <u>AGATGAATGC</u> | <u>CTGCACTTTA</u> | <u>AGATTCTCAA</u> | <u>GCATTTTGAA</u> | <u>GTTGGCTTCA</u> | |
| CVF1 cDNA | 4355 | TTCAGCCAGG | ATCAGTCAAG | GTGTACAGCT | ACTACAATCT | AG | 4396 |
| CVF2 cDNA | 3412 | TTCAGCCAGG | ATCAGTCAAG | GTGTACAGCT | ACTACAATCT | AG | 3453 |
| coC3 cDNA | 4387 | TTCAGCCAGG | ATCAGTCAAG | GTGTACAGCT | ACTACAATCT | AG | 4428 |
| CVF1 Gen | | TTCAGCCAGG | ATCAGTCAAG | GTGTACAGCT | ACTACAATCT | AGgtaatggc | |
| CVF1 Gen | | agacgtactg | aatacacctc | cctgctccta | atttccccac | aacaaccact | |
| CVF1 Gen | | ctgtgaagtg | agtcaagatt | gtgagcgaat | ggcttgGCCA | | |

Abb. 3.48: Darstellung der partiellen Intron 36-Sequenz des CVF1-Gens.

Die Sequenz des 900 bp-Amplifikats zeigte Übereinstimmung mit der Intron 36-Sequenz eines genomischen Amplifikats von 2,0 kb Länge (vgl. Kapitel 3.4, Abb. 3.72), welches aufgrund von Abweichungen zu den Sequenzen von CVF1, CVF2 und coC3 als CVF-Homolog 5 (CVF-H5) bezeichnet wurde (ausführliche Charakterisierung in Kapitel 3.4).

Möglicherweise enthält die mittlere Bande (1050 bp) der DraI-Bibliothek (Abb. 3.47) den entsprechenden CVF2-Genabschnitt.

Die Intron 36-Sequenzen von CVF1 und CVF-H5 sind im 5'-Bereich auf einer Länge von ca. 340 bp bis auf drei Nukleotide identisch (ohne Abbildung).

Um die Zahl der Amplifikate zu reduzieren und möglichst nur das CVF1-Amplifikat zu erhalten, wurden neue Primer gewählt (HB-CVF1-Intr.35s und HB-CVF1-Intr.35s-II), die mit der Intron 35-Sequenz von CVF1 ca. 200 bp stromaufwärts von Exon 36 identisch waren. Es wurde davon ausgegangen, dass sich die Intronsequenzen der drei homologen Gene in einem Abstand von 200 bp zum Exon soweit unterscheiden, dass die Primer nur mit der gesuchten Sequenz hybridisieren. Überraschenderweise wurden bei der verschachtelten PCR mit den Primern HB-CVF1-Intr.35s/AP1 und HB-CVF1-Intr.35s-II/AP2 erneut drei Amplifikate erhalten, deren Bandenmuster sich von den bereits oben beschriebenen Amplifikaten nur durch eine Größenzunahme um jeweils ca. 200 bp unterschied (ohne Abbildung). Es wurde wiederum die schwerste Bande kloniert und sequenziert. Die Sequenz des schwersten Amplifikats stimmte auch im Intron 35-Bereich (200 bp) mit CVF1 überein, so dass die schwerste Bande der DraI-Bibliothek als CVF1-Genabschnitt identifiziert wurde.

Intron 36/Exon 37

Mit dem genspezifischen Primer HB37as-CVF und den Adapter-Primern AP1 und AP2 wurde eine semi-verschachtelte PCR durchgeführt.

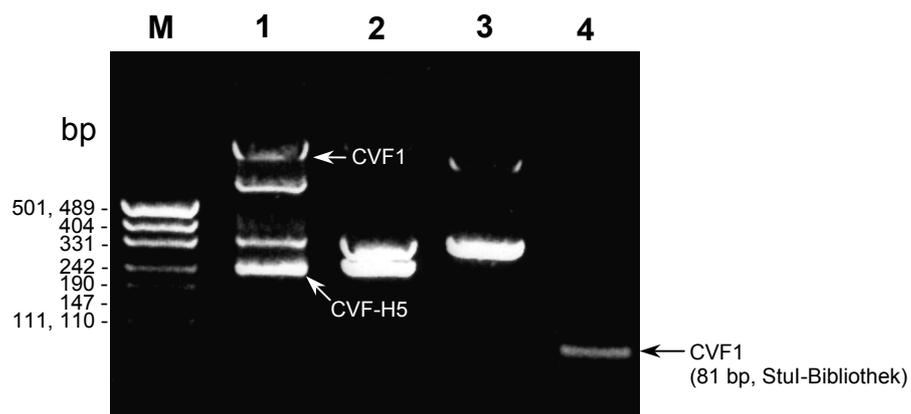


Abb. 3.49: Gelelektrophoretische Analyse der GenomeWalker-Amplifikate (Intron 36/Exon 37). 2%iges Agarosegel. M: DNA-Marker pUC19/MspI, Bahn 1: DraI-Bibliothek, Bahn 2: EcoRV-Bibliothek, Bahn 3: PvuII-Bibliothek, Bahn 4: StuI-Bibliothek. Die markierten Amplifikate der DraI- und StuI-Bibliotheken korrelieren mit ihrer Größe mit entsprechenden DraI- bzw. StuI-Schnittstellen in den jeweiligen Gensequenzen.

Das längste Amplifikat der DraI-Bibliothek (ca. 900 bp) wurde kloniert, sequenziert und als CVF1-Genabschnitt identifiziert (Abb. 3.50). Das kleinste Amplifikat der DraI-Bibliothek wurde aufgrund einer entsprechend vorhandenen DraI-Schnittstelle in der CVF-homologen Sequenz CVF-H5 zugeordnet (Kapitel 3.4 und Anhang 7.4).

Aufgrund einer StuI-Schnittstelle in Exon 37 des CVF1-Gens kann auch die Bande in Abbildung 3.49, Bahn 4, zugeordnet werden.

Das Vorkommen von mindestens vier distinkten Banden in der DraI-Bibliothek ist möglicherweise ein Hinweis auf das Vorhandensein von mehr als drei Mitgliedern in der CVF/C3-Genfamilie.

Nachfolgend ist die Randsequenz des CVF1-Amplifikats der DraI-Bibliothek (900 bp) dargestellt.

```

CVF1 Gen      ..... ttaaatgaat gattgtaagt tgaggactag ctgtaattaa
CVF1 Gen      atctatactc accagagttt cttcatgtga tgtatggata taacttttaa

CVF1 cDNA 4397      ATGAAAAA TGTACCAAGT TCTACCATCC AGATAAAGGA 4434
CVF2 cDNA 3454      ATGAAAAA TGTACCAAGA TCTACCATCC AGATCAAGCA 3491
coC3 cDNA 4429      ATGAACAA TGTACCAAGT TCTACCATCC AGATAAAGAA 4466
CVF1 Gen      ttcttccatt agATGAAAAA TGTACCAAGT TCTACCATCC AGATAAAGGA

CVF1 cDNA 4435      ACAGGCCTTC TCAATAAGAT ATGTATTGGT AACGTTTGCC GATGTGCAGG 4484
CVF2 cDNA 3492      ACAGGCCTTC TCAATAAGAT ATGTATTGGT AACGTTTGCC GATGTGCAGA 3541
coC3 cDNA 4467      ACAGGCCTTC TCAATAAGAT ATGTATTGGT AACGTTTGCC GATGTGCAGA 4516
CVF1 Gen      ACAGGCCTTC TCAATAAGAT ATGTATTGGT AACGTTTGCC GATGTGCAGG

```

Abb. 3.50: Partielle Sequenz des Genome Walking-Amplifikats (900 bp) mit Intron 36 des CVF1-Gens. Eine StuI-Schnittstelle (aggcct) in Exon 37 von CVF1 und CVF2 ist unterstrichen.

Für Intron 36 des CVF1-Gens wurde durch Genome Walking eine Mindestgröße von 1,9 kb ermittelt.

Exon 41 stromabwärts

Bei einer verschachtelten PCR mit den Primerpaaren HB-CVF-41s-I / AP1 und HB-CVF-41s-II / AP2 wurde mit der PvuII-Bibliothek ein ca. 600 bp-Amplifikat und mit der EcoRV-Bibliothek ein ca. 1500 bp-Amplifikat erhalten (Abb. 3.51).

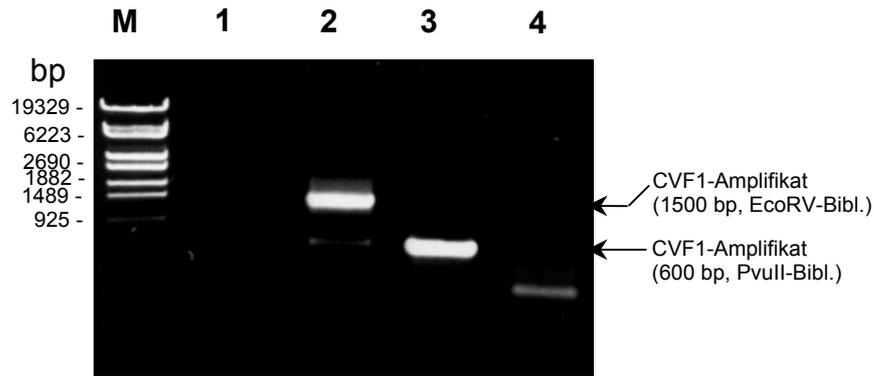


Abb. 3.51: Gelelektrophoretische Analyse der GenomeWalker-Amplifikate (Exon 41 stromabwärts). 1%iges Agarosegel. M: DNA-Marker λ /Eco130I, Bahn 1: DraI-Bibliothek, Bahn 2: EcoRV-Bibliothek, Bahn 3: PvuII-Bibliothek, Bahn 4: StuI-Bibliothek.

Das 1500 bp-Amplifikat der EcoRV-Bibliothek wurde kloniert und sequenziert. Die Sequenz ist nachfolgend in Abbildung 3.52 dargestellt. Eine PvuII-Schnittstelle (cagctg) befindet sich an Position 5395-5400 der CVF1-cDNA, und ihre Position korreliert mit der Größe des Amplifikats der PvuII-Bibliothek (Abb. 3.51, Bahn 3).

| | | | | | | | |
|------------------|------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| CVF1 cDNA | 4811 | AGAACACATG | GATTGAGAGA | TGGCCACATG | AAGACGAATG | TCAGGAAGAA | 4860 |
| CVF2 cDNA | 3868 | AGAACACCTG | GATTGAGAGA | TGGCCACATG | AAGACGAATG | TCAGGAAGAA | 3917 |
| coC3 cDNA | 4843 | AGAACACCTG | GATTGAGAGA | TGGCCAAATG | AAGATGAATG | CCAGGATGAA | 4892 |
| CVF1 Gen | 1 | | GAGAGA | TGGCCACATG | AAGACGAATG | TCAGGAAGAA | 36 |
| CVF1 cDNA | 4861 | GAATTCCAAA | AGTTGTGTGA | TGACTTTGCT | CAGTTTAGCT | ACACATTGAC | 4910 |
| CVF2 cDNA | 3918 | GAATTCCAAA | AGTTGTGTGA | TGACTTTGCT | CAGTTTAGCT | ACCAATGA- | 3966 |
| coC3 cDNA | 4893 | GAATTCCAA | AGTTGTGTGA | TGACTTTGCT | CAGTTGTCCA | ATACACTGAC | 4942 |
| CVF1 Gen | 37 | GAATTCCAAA | AGTTGTGTGA | TGACTTTGCT | CAGTTTAGCT | ACACATTGAC | 86 |
| CVF1 cDNA | 4911 | TGAGTTTGGC | TGCCCTACTT | AAAAGTTCAG | AAGAATCAAT | GATAGGAAGG | 4960 |
| CVF2 cDNA | 3967 | ---GTTTG-C | TGCCCTACTT | AA | | | 3984 |
| coC3 cDNA | 4943 | TATTTTGGC | TGCCCTACTT | AA | | | 4964 |
| CVF1 Gen | 87 | TGAGTTTGGC | TGCCCTACTT | AAAAGTTCAG | AAGAATCAAT | GATAGGAAGG | 136 |
| CVF1 cDNA | 4961 | AAATTCTCAG | AAGACAGATT | TTTGAGCCAA | TGCATATATG | TTACTTTGCC | 5010 |
| CVF1 Gen | 137 | AAATTCTCAG | AAGACAGATT | TTTGAGCCAA | TGCATATATG | TTACTTTGCC | 186 |
| CVF1 cDNA | 5011 | TCTTGATCTT | TTAGTTTTAT | GTCAATTTGC | TCTGTTATTT | TCCCTTAAAT | 5060 |
| CVF1 Gen | 187 | TCTTGATCTT | TTAGTTTTAT | GTCAATTTGC | TCTGTTATTT | TCCCTTAAAT | 236 |
| | | Poly-A-Signal | | | | | |
| CVF1 cDNA | 5061 | TGTTTATACA | TAAAATAAAT | AATCGATTC | TTACTTTGAT | ATGTTCTTGA | 5110 |
| CVF1 Gen | 237 | TGTTTATACA | TAAAATAAAT | AATCGATTC | TTACTTTGAT | ATGTTCTTGA | 286 |
| | | Poly-A-Signal | | | | | |
| CVF1 cDNA | 5111 | TTTTTAATAA | ACAATGGTGA | TTCATGATTA | TTTTTTTCTT | CTTCTGATCC | 5160 |
| CVF1 Gen | 287 | TTTTTAATAA | ACAATGGTGA | TTCATGATTA | TTTTTTTCTT | CTTCTGATCC | 336 |
| CVF1 cDNA | 5161 | ATCCAATATT | TGAAGTGCTC | TGAACAGAGC | ACTTATGGAG | TAATGTTTTA | 5210 |
| CVF1 Gen | 337 | ATCCAATATT | TGAAGTGCTC | TGAACAGAGC | ACTTATGGAG | TAATGTTTTA | 386 |
| CVF1 cDNA | 5211 | GTGATGGATG | AATAAGTTGG | TGAGTCAATA | TTATCAGGCC | CTATATACTC | 5260 |
| CVF1 Gen | 387 | GTGATGGATG | AATAAGTTGG | TGAGTCAATA | TTATCAGGCC | CTATATACTC | 436 |
| CVF1 cDNA | 5261 | TTATGGAAGA | TCGATTTGTA | CCCAAAGAAA | CATAGATTGA | AATGTGTTAC | 5310 |
| CVF1 Gen | 437 | TTATGGAAGA | TCGATTTGTA | CCCAAAGAAA | CATAGATTGA | AATGTGTTAC | 486 |
| CVF1 cDNA | 5311 | TTTGAAAACA | GAGGTTTCAG | TTGTATATGT | TTACACTTGG | ATACAATCTT | 5360 |
| CVF1 Gen | 487 | TTTGAAAACA | GAGGTTTCAG | TTGTATATGT | TTACACTTGG | ATACAATCTT | 536 |

| | | Poly-A-Signal | | | | | | |
|------------------|------|-------------------|-------------------|--------------------|--------------------|--------------------|------|--|
| CVF1 cDNA | 5361 | <u>AACTCTTAAT</u> | <u>AAACACTGAT</u> | CTCAGAACAT | TTAACAGCTG | CTATTTAATA | 5410 | |
| CVF1 Gen | 537 | <u>AACTCTTAAT</u> | <u>AAACACTGAT</u> | CTCAGAACAT | <u>TTAACAGCTG</u> | CTATTTAATA | 586 | |
| CVF1 cDNA | 5411 | ATGACAAAAAT | ATTCTTTGAC | TGCACCCACA | GAAAACATTG | CATTACATTA | 5460 | |
| CVF1 Gen | 587 | ATGACAAAAAT | ATTCTTTGAC | TGCACCCACA | GAAAACATTG | CATTACATTA | 636 | |
| CVF1 cDNA | 5461 | GAATGGGTTT | TATCAGATGA | CTAAGTCTGC | TAGACTTGCC | ATCTGTCAAA | 5510 | |
| CVF1 Gen | 637 | GAATGGGTTT | TATCAGATGA | CTAAGTCTGC | TAGACTTGCC | ATCTGTCAAA | 686 | |
| CVF1 cDNA | 5511 | ATGTGCCTCT | TCCCCAGCTC | CAACTTTAAG | GATAGTAACT | AATAGATGTT | 5560 | |
| CVF1 Gen | 687 | ATGTGCCTCT | TCCCCAGCTC | CAACTTTAAG | GATAGTAACT | AATAGATGTT | 736 | |
| CVF1 cDNA | 5561 | CTCTCATTGG | CTCCTGACAG | AGGTGTGGTA | GCCACTGAGT | TTCCCTGGAT | 5610 | |
| CVF1 Gen | 737 | CTCTCATTGG | CTCCTGACAG | AGGTGTGGTA | GCCACTGAGT | TTCCCTGGAT | 786 | |
| CVF1 cDNA | 5611 | GACACTAGAA | GCTGGCAGCA | CACTGCAGCC | TGGTGGAGGG | G-CCTCTTTT | 5659 | |
| CVF1 Gen | 787 | GACACTAGAA | GCTGGCAGCA | CACTGCAGCC | TGGTGGAGGG | <u>GG</u> CCTCTTTT | 836 | |
| CVF1 cDNA | 5660 | GCTATCCCAT | GAGCTTCTAT | TCATCCTCTT | ATCTGTTGGG | ATGGGGATGG | 5709 | |
| CVF1 Gen | 837 | GCTATCCCAT | GAGCTTCTAT | TCATCCTCTT | ATCTGTTGGG | ATGGGGATGG | 886 | |
| CVF1 cDNA | 5711 | GACGTCTCTG | ATTTTCCAGG | TATACAGGTG | ATCTCATTTA | CTAACATCAC | 5759 | |
| CVF1 Gen | 887 | GACGTCTCTG | ATTTTCCAGG | TATACAGGTG | ATCTCATTTA | CTAACATCAC | 936 | |
| CVF1 cDNA | 5760 | CACTAACTTC | AAGGATTGGT | TGAGGGGTTA | <u>TGCCAAT</u> GTG | ATTGAAGG-T | 5808 | |
| CVF1 Gen | 937 | CACTAACTTC | AAGGATTGGT | TGAGGGGTTA | <u>TGCCAAT</u> GTG | ATTGAAGG <u>GT</u> | 986 | |
| CVF1 cDNA | 5809 | TTCACCCATG | TGAATCTATT | <u>CTCCAAT</u> CCC | <u>AATGCT</u> GTAT | CTATGCTGCT | 5858 | |
| CVF1 Gen | 987 | TTCACCCATG | TGAATCTATT | <u>CTCCAAT</u> CCC | <u>AATGCT</u> GTAT | CTATGCTGCT | 1036 | |
| CVF1 cDNA | 5859 | CATTTCTGCT | TGTAAAAATG | GTATAAAAAG | <u>AATAAA</u> CACT | GCCCAGGCAG | 5908 | |
| CVF1 Gen | 1037 | CATTTCTGCT | TGTAAAAATG | GTATAAAAAG | <u>AATAAA</u> CACT | GCCCAGGCAG | 1086 | |
| CVF1 cDNA | 5909 | TCAGACATCT | TTGGACACTG | | | | 5928 | |
| CVF1 Gen | 1087 | TCAGACATCT | TTGGACACTG | aatgactgg | ttattgaatc | cctatacaga | 1136 | |
| | | | <u>MAZ</u> | | | | | |
| CVF1 Gen | 1137 | gggaaggag | <u>ggaggagg</u> | agtatccatc | agtcctggt | ctgtatacag | 1186 | |
| CVF1 Gen | 1187 | tcccccttgt | atactataaa | taagaacagc | atatcaatct | agggaaagt | 1236 | |
| CVF1 Gen | 1237 | <u>tgaattcgcc</u> | agtgtggca | gcaaagtta | tgccagtata | gtcccaaac | 1286 | |
| CVF1 Gen | 1287 | tcagtaagga | ttactatata | gcattctggg | agctggtgta | gttgtttgca | 1336 | |
| CVF1 Gen | 1337 | gtgctgaaga | aatcttttag | acattacaac | atTTTTattt | atgTTTTtct | 1386 | |
| CVF1 Gen | 1387 | ctgaaccatc | <u>ttgaattctg</u> | atgagagaag | gcagcaacat | tgtcagccac | 1436 | |
| CVF1 Gen | 1437 | aggggatgaa | ctggtggggg | ggggtcaatg | taatgcacgc | tccaccatt | 1486 | |
| CVF1 Gen | 1487 | taatatcaag | <u>atatac</u> | | | | 1501 | |

Abb. 3.52: Darstellung des 3'-Bereichs des CVF1-Gens. Die für Vergleichszwecke aufgeführten cDNA-Sequenzen von CVF2 und coC3 wurden zwecks besserer Übersichtlichkeit nur bis zum Stop-Codon TAA (in allen Sequenzen unterstrichen) dargestellt. Die Exonsequenzen sind bis zum Stop-Codon in fetten Großbuchstaben dargestellt. Der 3'-untranslatierte Bereich bis zum Ende der CVF1-cDNA ist in Großbuchstaben dargestellt. Die darüber hinaus führende genomische Sequenz ist in Kleinbuchstaben dargestellt. Das Poly-Adenylierungssignal AATAAA in CVF1 (Position 5889-5894 der cDNA) ist fett und kursiv dargestellt. Drei zusätzliche AATAAA-Sequenzen weiter stromaufwärts sind unterstrichen. An Position 571 der genomischen Sequenz beginnt die PvuII-Schnittstelle CAGCTG (unterstrichen), und am 3'-Ende ist die EcoRV-Schnittstelle gatac (unterstrichen) markiert. Unterschiede der genomischen Sequenz zur CVF1-cDNA existieren im 3'-untranslatierten Bereich an den Positionen 5161 (A-G), 5648-5651 (4G-5G) und 5806-5807 (2G-3G) und wurden fett und unterstrichen markiert.* Eine EcoRI-Schnittstelle im translatierten Bereich von Exon 41 ist in allen Sequenzen vorhanden und unterstrichen, ebenso sind zwei EcoRI-Schnittstellen im 3'-Bereich markiert. Im 3'-Bereich sind ferner drei CCAAT-Boxen und drei verschachtelte Bindungsstellen für MAZ gekennzeichnet.

* Diese drei Abweichungen der genomischen Sequenz von der publizierten CVF1-cDNA wurden kürzlich durch Re-Sequenzierung der cDNA bestätigt (Fritzinger, persönliche Mitteilung).

Aus der genomischen Sequenz wird ersichtlich, dass der 3'-untranslatierte Bereich des CVF1-Gens nicht durch weitere Introns unterbrochen ist. Somit erstreckt sich das terminale Exon 41 über eine Länge von 1140 bp (bzw. 139 bp ohne 3'-UTR). Die hier erstmalig als Poly-Adenylierungssignal des CVF1-Gens identifizierte Sequenz AATAAA beginnt 40 bp stromaufwärts des Poly-A-Anhangs der CVF1-mRNA. Die dargestellte Gen-Sequenz reicht noch 395 bp über das Ende der cDNA hinaus. Auffällig ist in diesem Sequenzabschnitt eine vierfache GGGA-Wiederholung, die das Bindungsmotiv des Transkriptionsfaktors MAZ (GGGAGGG) dreifach verschachtelt enthält.

Die PvuII-Schnittstelle an Position 571 der genomischen Sequenz korreliert mit der Größe des PCR-Produkts der PvuII-Bibliothek (Abb. 3.51), so dass es sich dabei höchstwahrscheinlich ebenfalls um ein CVF1-Amplifikat handelt.

Unter Berücksichtigung, dass die Gene der CVF/C3-Familie stark konserviert sind und das Stop-Codon sich in allen drei Genen an konservierter Position befindet, erscheinen in der CVF2-cDNA-Sequenz die Lücken (4+1 bp) stromaufwärts des Stop-Codons insofern ungewöhnlich, da sie nicht dem Vielfachen eines Codons entsprechen und deswegen zu einer Leserasterverschiebung und Negation des Stop-Codons führen würden.

Tab. 3.5: Darstellung der Exon-Intron-Struktur des CVF1-Gens. Zum Vergleich wurden die Daten des humanen C3-Gens (Vik *et al.* 1991) ergänzt (schattiert dargestellt). Für Vergleichszwecke der terminalen Exons 1 (ab Startcodon) und 41 (bis Stopcodon) wurden die untranslatierten Bereiche am 5'- und 3'-Ende getrennt dargestellt.

| Exon | Position in cDNA | Länge (bp) CVF1 | Länge (bp) huC3 | Spleiß-Akzeptor | Spleiß-Donor | Intron-Größe (bp) CVF1 | Intron-Größe (bp) huC3 |
|--------|------------------------|-----------------|-----------------|-----------------|---------------|------------------------|------------------------|
| 5'-UTR | 1-3 | 3 ¹ | 60 | | | | |
| 1 | 4-74 | 71 | 74 | | TCT / gta agt | (5800) | (1100) |
| 2 | 75-267 | 193 | 193 | ctg cag / CTA | GAG / gtg ggt | >1000 | (760) |
| 3 | 268-430 | 163 | 166 | atc cag / ATT | CAG / gta acc | 732 | 82 |
| 4 | 431-501 | 71 | 71 | tct tag / TAC | CAG / gtg aga | 657 | (3500) |
| 5 | 502-581 | 80 | 95 | tcc cag / ACT | CAG / gta att | 548 | 102 |
| 6 | 582-661 | 80 | 83 | ctg tag / TTT | ATG / gta aga | 524 | 83 |
| 7 | 662-752 | 91 | 91 | ctg tag / TGT | AAG / gta tgc | (2500) | (500) |
| 8 | 753-855 | 103 | 103 | tta cag / GTA | CCG / gta aaa | >1000 | 91 |
| 9 | 856-976 | 121 | 127 | ttt cag / ATT | CAG / gtg agc | (2900) | 437 |
| 10 | 977-1092 | 116 | 116 | ttg cag / GCA | ACG / gta tag | 393 | 98 |
| 11 | 1093-1239 | 147 | 150 | ttt tag / GTG | ACT / gta agt | (1900) | (780) |
| 12 | 1240-1449 | 210 | 210 | atc tag / GTT | CTC / gta agt | 123 | 141 |
| 13 | 1450-1644 | 195 | 207 | cct cag / ATA | ACG / gta agc | 209 | (700) |
| 14 | 1645-1794 | 150 | 159 | aaa cag / TTG | AAG / gta agg | 100 | (1800) |
| 15 | 1795-1924 | 130 | 130 | cac tag / ATA | CAG / gta ttt | (9000) | 262 |
| 16 | 1925-1999 | 75 | 72 | ctt cag / CTG | AAG / gta tgt | (1400) | 192 |
| 17 | 2000-2200 | 201 | 198 | ttt tag / CGG | GAG / gta agt | (7000) | (4400) |
| 18 | 2201-2312 | 112 | 109 | ttt cag / ATG | AGG / gta ttg | 1438 | 242 |
| 19 | 2313-2398 | 86 | 87 | tcc tag / GAT | AAG / gta aag | 108 | (4200) |
| 20 | 2399-2538 | 140 | 142 | ttg cag / GGA | TAT / gta agt | 887 | 94 |
| 21 | 2539-2751 | 213 | 213 | cac tag / GTG | GTA / gta agt | 506 | (700) |
| 22 | 2752-2815 | 64 | 67 | taa cag / CCT | TTG / gtg agt | 321 | 126 |
| 23 | 2816-2902 | 87 | 87 ² | ttt cag / GTG | AAG / gta gga | 681 | (1750) |
| 24 | 2903-3106 | 204 | 204 | att cag / GTG | CTG / gta agg | 595 | (1000) |
| 25 | 3107-3182 | 76 | 76 | ttt cag / GTT | TTG / gta ggg | (8000) | (250) |
| 26 | 3183-3345 | 163 | 160 | tta cag / GCT | CAG / gta tgg | (3000) | (900) |
| 27 | 3346-3441 | 96 | 99 | ttc tag / GGA | AAT / gta cgt | >1300 | (3500) |
| 28 | 3442-3592 | 151 | 157 | ttc cag / AGT | CAG / gta cac | (2600) | (470) |
| 29 | 3593-3756 | 164 | 164 | ttt tag / GAA | CAA / gtg agt | (1500) | (1000) |
| 30 | 3757-3915 | 159 | 159 | ttc tag / GCA | GAG / gta tag | (2500) | 155 |
| 31 | 3916-(3975) | (60) | 60 | ttt cag / ACC | ACC | kein Intron | 124 |
| 32 | (3976)-4066 | (91) | 91 | ATT | TGA / gta agt | (2200) | 138 |
| 33 | 4067-4112 | 46 | 52 | tca aag / ATG | AAG / gta aga | 773 | (2300) |
| 34 | 4113-4200 | 88 | 88 | ttt tag / GTA | AGG / gta ggc | 441 | 111 |
| 35 | 4201-4290 | 90 | 90 | att tag / CTT | AAG / gta aat | (2500) | (1700) |
| 36 | 4291-4396 | 106 | 106 | tcc cag / GTC | TAG / gta atg | >1900 | (900) |
| 37 | 4397-4486 | 90 | 90 | cat tag / ATG | GAG / gta agt | (3500) | 196 |
| 38 | 4487-4570 | 84 | 84 | ttg aag / AAA | ATG / gta ttt | (1250) | (600) |
| 39 | 4571-4654 | 84 | 84 | taa aag / TCT | AAG / gta aga | (7000) | 85 |
| 40 | 4655-4790 | 136 | 136 | ctg cag / GTA | TAA / gta cgt | (2400) | 128 |
| 41 | 4791-4929 | 139 | 142 | ttc cag / AAT | | | |
| 3'-UTR | 4930-5930 ³ | 1001 | 46 | | | | |

Anmerkungen:

- 1) Die exakte Größe von Exon 1 (inklusive 5'-UTR) kann erst nach experimenteller Bestätigung des Transkriptionsstarts bestimmt werden.
- 2) Die Größe von Exon 23 des humanen C3-Gens ist in der Literatur (Vik *et al.* 1991) irrtümlich mit 88 bp angegeben.
- 3) Die CVF1-cDNA ist im 3'-untranslatierten Bereich um 2 bp länger als publiziert (siehe Abb. 3.52 und Fritzing *et al.* 1995).

Eine anschauliche graphische Zusammenfassung der oben beschriebenen Ergebnisse ist in Abbildung 3.53 gezeigt. Die dargestellte Karte des CVF1-Gens gibt eine Übersicht über die Positionen der Exons im CVF1-Gen und die Größen der Introns.

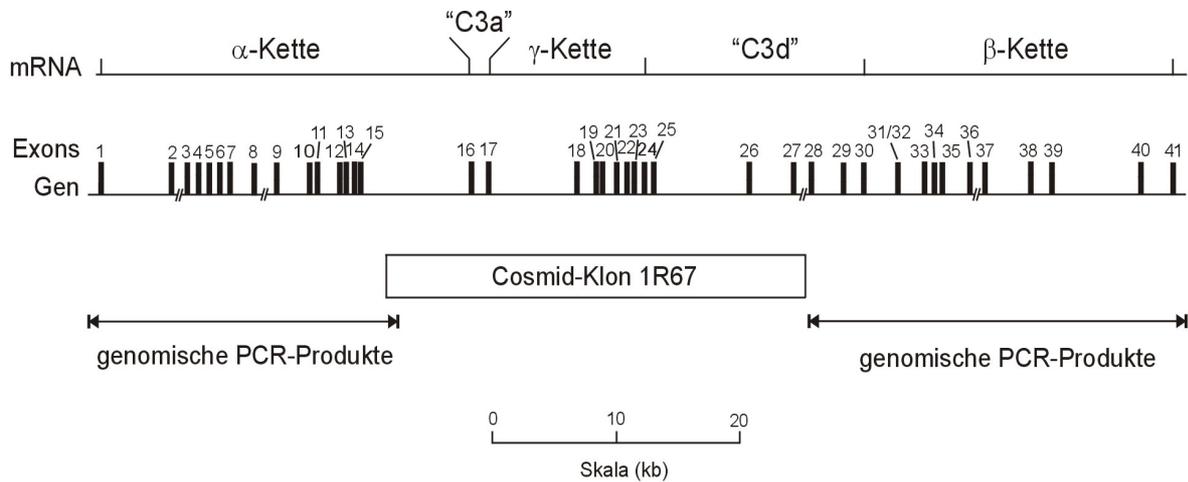


Abb. 3.53: Karte des CVF1-Gens.

In der Genkarte (mittleres Panel) sind die Positionen der Exons und die Größen der Introns dargestellt. Nicht vollständig bestimmte Introngrößen sind durch unterbrochene Striche gekennzeichnet. Darüber ist die korrelierende mRNA bzw. Protein-Kettenstruktur gezeigt. Unten sind die Bereiche markiert, die durch den Cosmidklon 1R67 sowie durch PCR-Produkte (genomische PCR und Genome Walking) abgedeckt wurden.

In der folgenden Übersicht sind die Strukturelemente der CVF1-mRNA und des CVF1-Proteins sowie die entsprechenden Exonbereiche dargestellt.

| Strukturelemente der CVF1-mRNA / des CVF1-Proteins | Exonbereiche |
|---|--------------|
| 5'-UTR und SS | Exon 1 |
| α-Kette | Exons 1-16 |
| „C3a“ | Exons 16-17 |
| γ-Kette | Exons 17-24 |
| „C3d“ | Exons 24-30 |
| β-Kette | Exons 30-41 |
| 3'-UTR | Exon 41 |

Eine detaillierte Analyse der Intronphasen des CVF1-Gens wurde durchgeführt, um durch Vergleich mit homologen Genen Aussagen über die Konservierung von Genstrukturen treffen zu können.

Bei den Introns unterscheidet man zwischen Typ 0 (Intron ist zwischen zwei Codons inseriert), Typ 1 (Intron nach Nukleotid 1 eines Codons) und Typ 2 (Intron nach Nukleotid 2 eines Codons). Tabelle 3.6 gibt eine Übersicht über die Intronphasen des CVF1-Gens.

Tab. 3.6: Intronphasen des CVF1-Gens. Die Aminosäuren sind mittig oberhalb der Codons dargestellt.

| Intron | Spleiß-Sequenzen | Phasen-Typ | Intron | Spleiß-Sequenzen | Phasen-Typ |
|--------|---|------------|--------|---|-------------|
| 1 | ^L TCT/gta...cag/ ^Y CTAC | 2 | 21 | ^V GTA/gta...cag/ ^P CCT | 0 |
| 2 | ^E GAG/gtg...cag/ ^I ATT | 0 | 22 | ^V TTG/gtg...cag/ ^G GTG | 1 |
| 3 | ^P CAG/gta...tag/ ^V TAC | 1 | 23 | ^Q AAG/gta...cag/ ^G GTG | 1 |
| 4 | ^Q CAG/gtg...cag/ ^T ACT | 0 | 24 | ^T CTG/gta...cag/ ^G GTT | 1 |
| 5 | ^S CAG/gta...tag/ ^L TTTG | 2 | 25 | ^W TTG/gta...cag/ ^L GCTA | 2 |
| 6 | ^Y ATG/gta...tag/ ^V TGT | 1 | 26 | ^Q CAG/gta...tag/ ^G GGA | 0 |
| 7 | ^R AAG/gta...cag/ ^Y GTAC | 2 | 27 | ^N AAT/gta...cag/ ^S AGT | 0 |
| 8 | ^P CCG/gta...cag/ ^I ATT | 0 | 28 | ^T CAG/gta...tag/ ^G GAA | 1 |
| 9 | ^S CAG/gtg...cag/ ^G GCA | 1 | 29 | ^Q CAA/gtg...tag/ ^A GCA | 0 |
| 10 | ^T ACG/gta...tag/ ^V GTG | 0 | 30 | ^E GAG/gta...cag/ ^T ACC | 0 |
| 11 | ^T ACT/gta...tag/ ^V GTT | 0 | 31 | ^T ^I ACC/ATT | kein Intron |
| 12 | ^L CTC/gta...cag/ ^I ATA | 0 | 32 | ^L TGA/gta...aag/ ^N ATG | 1 |
| 13 | ^T ACG/gta...cag/ ^L TTG | 0 | 33 | ^R AAG/gta...tag/ ^Y GTAT | 2 |
| 14 | ^K AAG/gta...tag/ ^I ATA | 0 | 34 | ^R AGG/gta...tag/ ^L CTT | 0 |
| 15 | ^S CAG/gta...cag/ ^A CTG | 1 | 35 | ^K AAG/gta...cag/ ^V GTC | 0 |
| 16 | ^K AAG/gta...tag/ ^A CGG | 1 | 36 | ^L TAG/gta...tag/ ^D ATG | 1 |
| 17 | ^R GAG/gta...cag/ ^D ATG | 1 | 37 | ^G GAG/gta...aag/ ^E AAA | 1 |
| 18 | ^G AGG/gta...tag/ ^I GATT | 2 | 38 | ^Y ATG/gta...aag/ ^V TCT | 1 |
| 19 | ^K AAG/gta...cag/ ^G GGA | 1 | 39 | ^Q AAG/gta...cag/ ^G GTA | 1 |
| 20 | ^Y TAT/gta...tag/ ^V GTG | 0 | 40 | ^K TAA/gta...cag/ ^I AATT | 2 |

Die Intron-Typen des CVF1-Gens sind mit denen des huC3-Gens weitgehend identisch und unterscheiden sich lediglich in Intron 19 (Typ 1 in CVF1, Typ 2 in huC3) und dem Fehlen von Intron 31 in CVF1 (Typ 0 in huC3). Die Häufigkeit der drei Intron-Phasen-Typen im CVF1-Gen und huC3-Gen ist nachfolgend dargestellt:

| | CVF1 | huC3 |
|---------------|----------|------------|
| Typ 0-Introns | 16 (41%) | 17 (42,5%) |
| Typ 1-Introns | 16 (41%) | 15 (37,5%) |
| Typ 2-Introns | 7 (18%) | 8 (20%) |

3.2 Struktur des CVF2-Gens

Die Struktur des CVF2-Gens wurde mit dem Ziel analysiert, Gemeinsamkeiten und Unterschiede zur CVF1-Genstruktur aufzuzeigen. Die Sequenzen von CVF1, CVF2 und coC3 sind auf cDNA-Ebene zu mehr als 91% identisch, sodass von einer starken Konservierung der Exon-Strukturen auszugehen war. Um die Vorhersagbarkeit von Exongrenzen in den Genen CVF2- und coC3 durch Experimente zu bestätigen, wurden ausgewählte Introns durch genomische PCR amplifiziert, kloniert und sequenziert. Für die Generierung CVF2-spezifischer Primer wurden die Exongrenzen in Analogie zur CVF1-Struktur vorhergesagt. Für eine Analyse des Promotorbereichs sowie von Exon/Intron-Übergangssequenzen im 5'-Bereich des CVF2-Gens wurde die Genome Walking-Methode eingesetzt.

3.2.1 Aufklärung der CVF2-Genstruktur durch genomische PCR

Zur Aufklärung der CVF2-Genstruktur wurden durch genomische PCR exemplarisch Introns 18 und 32 mit CVF2-spezifischen Primern amplifiziert. Darüber hinaus wurde Intron 2 des CVF2-Gens unter Verwendung CVF1-spezifischer Primer erhalten. In Abbildung 3.54 sind die erhaltenen Amplifikate dargestellt.

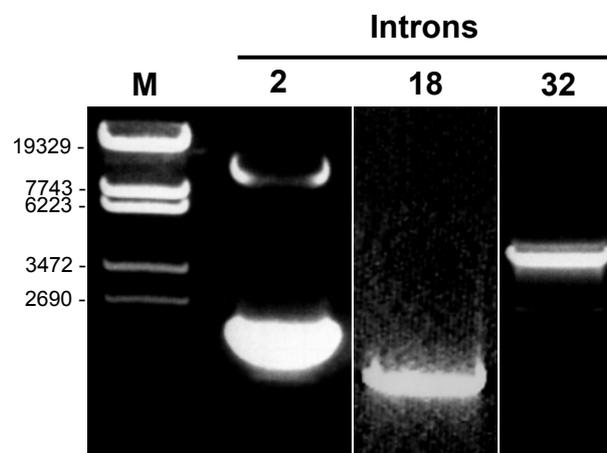


Abb. 3.54: Gelelektrophoretische Analyse der genomischen PCR-Amplifikate des CVF2-Gens. M: DNA-Marker λ /Eco130I; Intron 2: CVF2 (10 kb), coC3 (2,5 kb); Intron 18: 1400 bp; Intron 32: 3900 bp.

Alle Amplifikate wurden kloniert und sequenziert. Tabelle 3.7 gibt eine Übersicht über die verwendeten Primerpaare und die Exon/Intron-Übergangssequenzen der erhaltenen Amplifikate.

Tab. 3.7: Übersicht der genomischen PCR-Amplifikate des CVF2-Gens. Größenangaben in Klammern sind Schätzungen.

| Intron | Primerpaar | Größe PCR-Produkt (bp) | Exon-Intron-Übergangssequenzen |
|--------|--------------------------------|------------------------|---|
| 2 | HB2s-CVF / HB3as-II-CVF | (10.000) | AAG gtgggtgat...tctcatccag GTT |
| 18 | CVF2-Ex18s / CVF2-Ex19as | (1400) | AGG gtattggagt...ttgtccctag GAT |
| 32 | HB-31s-CVF2,coC3 / HB33as-CVF2 | (3900) | TGA gtaagtttg...caattcttag ACT |

Im weiteren erfolgt eine Analyse aller erhaltenen Sequenzen, um eine zweifelsfreie Zuordnung zum CVF2-Gen zu zeigen.

Intron 2

Mit den CVF1-spezifischen Primern HB2s-CVF und HB3as-II-CVF sollte ursprünglich Intron 2 des CVF1-Gens amplifiziert werden. Da unter Verwendung von Taq-Polymerasen keine Amplifikation möglich war, wurde die PCR mit der Taq/Tgo-Polymerase-Mischung *Expand* (Roche, Boehringer) durchgeführt. Auf diese Weise wurden Amplifikate von ca. 2,5 und 10 kb Länge erhalten, wobei das 10 kb-Amplifikat die in Abbildung 3.55 dargestellte Sequenz enthielt.

| | | | | | | | |
|------------------|-----|---------------------------|-------------------------|--------------------------|--------------------------|--------------------------|-----|
| CVF1 cDNA | 137 | CCCATGGAGA | CAGTACTCCA | AAACAGCTTG | ACATCTTTGT | TCATGATTTT | 186 |
| coC3 cDNA | 142 | CCCATGGAGA | CAGTACTCCA | AAATCGCTTG | ACATCTTTGT | TCATGATTTT | 191 |
| CVF2 Gen | | CCCATGGAGA | CAGTACTCCA | AAACAGCTTG | ACATCTTTGT | TCACAATTTT | |
| CVF1 cDNA | 187 | CCACGGAAGC | AGAAAACCTT | GTTCCAAACC | AGAGTAGATA | TGAATCCAGC | 236 |
| coC3 cDNA | 192 | CCACGGAAGC | AGAAAACCTT | GTTCCAAACC | AGAGTAGATA | TGAATCCAGC | 241 |
| CVF2 Gen | | CCACAGAAGC | AGAAATCTT | GTACCAAACC | AGAGTAGATA | TGAATCCAGA | |
| CVF1 cDNA | 237 | AGGAGGCATG | CTTGTCACTC | CAACTATAGA | G | | 267 |
| coC3 cDNA | 242 | AGGAAGCATG | CTTGTCACTC | CAACTATAAA | G | | 272 |
| CVF2 Gen | | GGGAGGCATG | CTTGTCACTC | CAACTATAAA | G gtgggtgta | tggtgatggt | |
| CVF2 Gen | | tgcaatagag | tcttttggtat | cctgcgaaa | gaacaaaaca | aacaaacaaa | |
| CVF2 Gen | | <u>cccaactaa</u> | | | | | |
| CVF2 Gen | | | (ca. 9,5 kb) | | | | |
| CVF2 Gen | | | | | | gctggtgccc | |
| CVF2 Gen | | tgagcgaag | aatgaggcct | caaatggccc | ttagtgatga | ttaagtttga | |
| CVF1 cDNA | 268 | | | | ATTCCAGCAA | AAGAAGTGAG | 287 |
| coC3 cDNA | 273 | | | | GTTCCTGCAA | AAGAACTGAA | 292 |
| CVF2 Gen | | agcctgtgct | ntactcatct | tctcatccag | GTTCCTGCAA | AAGAACTGAA | |
| CVF1 cDNA | 288 | TACGGACTCC | AGGCAAAATC | AATATGTGGT | TGTGCAAGTA | ACTGGTCCTC | 337 |
| coC3 cDNA | 293 | TAAAGGACTCC | AAGCAAAATC | AGTATGTGGT | TGTGAAAGTA | ACTGGTCCTC | 342 |
| CVF2 Gen | | TAAGGACTCC | AAGCAAAATC | AATATGTGGT | TGTGCAAGTA | ACTGGTCCTC | |

Abb. 3.55: Sequenz des genomischen PCR-Produkts mit Intron 2 des CVF2-Gens. Intron 2 hat eine Größe von ca. 10.000 bp und enthält eine vierfach repetitive aaac-Sequenz (unterstrichen). Eine Stuf-Schnittstelle (aggcct) im Intron ist unterstrichen.

Zwar fehlte für Vergleichszwecke der 5'-Bereich der CVF2-cDNA, doch aufgrund zahlreicher Unterschiede zu den korrespondierenden CVF1- und coC3-Sequenzen wurde die vorliegende DNA-Sequenz als CVF2-Genabschnitt identifiziert.

Intron 18

Der mit dem CVF2-spezifischen Primerpaar CVF2-Ex18s/CVF2-Ex19as gezielt amplifizierte CVF2-Genabschnitt (Abb. 3.54) war in seiner Größe von ca. 1400 bp mit dem entsprechenden CVF1-Homolog (Abb. 3.3, 1531 bp) annähernd vergleichbar. Aufgrund der charakteristischen Sequenz im Exonbereich war das Amplifikat eindeutig dem CVF2-Gen zuzuordnen. Unterschiede zur CVF1-Sequenz waren auch im Intronbereich detektierbar (ohne Abbildung).

| | | | | | | | | |
|-----------|------|-------------|------------------|------------|------------|------------|-------|--|
| CVF1 cDNA | 2201 | ATGATAATG | AAGATGGTTT | CATAGCAGAT | AGTGATATCA | TCTCAAGGTC | 2249 | |
| CVF2 cDNA | 1249 | GTGATTTTG | AAGATGATTT | ATTGGAGAA | GGTAACATCA | CCTCAAGGTC | 1297 | |
| coC3 cDNA | 2224 | GTGATTTTG | AAGATCAACT | CTTGGAGAT | GACAACATCA | TCTCAGGTC | 2272 | |
| CVF2 Gen | | gGTGATTTTG | AAGATGATTT | ATTGGAGAA | GGTAACATCA | CCTCAAGGTC | | |
| CVF1 cDNA | 2250 | TGATTTCCCC | AAGAGTTGGT | TGTGGCTAAC | AAAGGACTTG | ACCGAGGAGC | 2299 | |
| CVF2 cDNA | 1298 | TGATTTTCC | CAGAGTTGGT | TGTGGCTAAT | GGAGCAGCTG | TCTGAACATC | 1347 | |
| coC3 cDNA | 2273 | TGATTTTCC | CAGAGTTGGT | TGTGGCTAAC | ACAGGAATTG | ACCGGGGAGC | 2322 | |
| CVF2 Gen | | TGATTTTCC | CAGAGTTGGT | TGTGGCTAAT | GGAGCAGCTG | TCTGAACATC | | |
| CVF1 cDNA | 2300 | CTAACAGTCA | AGG | | | | 2312 | |
| CVF2 cDNA | 1348 | CTAACAGTAA | AGG | | | | 1360 | |
| coC3 cDNA | 2323 | CTAACAAATCA | AGG | | | | 2335 | |
| CVF2 Gen | | CTAACAGTAA | AGGgtattgg | agttcttctg | tttatattag | cccacttttc | | |
| CVF2 Gen | | agttgaaagg | aaaatctaaa | tattgcaaaa | tttatacaga | aatggggaag | | |
| CVF2 Gen | | atggtgtctc | | | | | | |
| CVF2 Gen | | | (ca. 1000 bp) | | | | | |
| CVF2 Gen | | |tatctaattgc | | | | | |
| CVF2 Gen | | tagttttaa | atgcacagta | taatagctct | gtaaattacg | cgacacattt | | |
| CVF1 cDNA | 2313 | | GATTTT | AAGCAAGACA | ATGTCTTTTT | ATCTGAGGGA | 2348 | |
| CVF2 cDNA | 1361 | | GATTTT | AAGCAAGATA | GTACCTTTTT | ATCTGAGGGA | 1396 | |
| coC3 cDNA | 2336 | | GATTTT | AAGCAAGACA | GTACCTTTTT | ATCTGAGGGA | 2371 | |
| CVF2 Gen | | cattttgtcc | ctagGATTTT | AAGCAAGATA | GTACCTTTTT | ATCTGAGGGA | | |
| CVF1 cDNA | 2349 | TTCCATCACA | ACCTGGGTGG | TGCTGGCTGT | AAGCTTTACA | CCCACCAAAG | 2398 | |
| CVF2 cDNA | 1397 | TTCCATCACA | ACCTGGGAGT | TGCTGGCTGT | GGGCCTTTCA | CCCACCAAAG | 1446 | |
| coC3 cDNA | 2372 | TTCCATCACA | ACCTGGGAGT | TGCTGGCTGT | GGGCCTTTCA | CCCACCAAAG | 2421 | |
| CVF2 Gen | | TTCCATCACA | ACCTGGGAGT | TGCTGGCTGT | GGGCCTTTCA | CCCACCAAAG | | |

Abb. 3.56: Sequenz des genomischen PCR-Produkts mit Intron 18 des CVF2-Gens. Intron 18 hat eine Größe von ca. 1200 bp.

Das genomische Amplifikat enthält vollständig Exons 18 und 19 des CVF2-Gens und diente als Grundlage für weitere Experimente zum proteinchemischen Nachweis des CVF2-Proteins (siehe Kapitel 3.2.3).

Intron 32

Da entgegen der Erwartungen aus der human C3-Genstruktur im CVF1-Gen Intron 31 fehlt, sollte geklärt werden, ob es sich dabei um eine charakteristische Besonderheit nur des CVF1-Gens handelt und wie die Gene CVF2 und coC3 an der entsprechenden Stelle strukturiert sind. Dazu wurden Primer eingesetzt, die eine Analyse der entsprechenden Exon 31/Exon 32-Grenze ermöglichten. Eine genomische PCR mit den Primern HB31s-CVF2/coC3 und HB33as-CVF2 führte zu einem Amplifikat von 3.9 kb Länge.

| | | Exon 31 | | | | | | | | |
|-------------|-------------|----------------|--|---------|-------|-------|-------|-------|------|------|
| CVF1 | cDNA | 3916 | ACCAA | AAGACA | TCACT | AGCAT | GATGG | 3964 | | |
| CVF2 | cDNA | 2964 | ACCAA | AAGACA | TCACT | AGCAT | GATGG | 3012 | | |
| coC3 | cDNA | 3939 | ACCAA | AAGACA | TCACT | AGCAT | GATGG | 3987 | | |
| CVF2 | Gen | | gACCAA | AAGACA | TCACT | AGCAT | GATGG | | | |
| | | | ↓ | | | | | | | |
| | | Exon 32 | | | | | | | | |
| CVF1 | cDNA | 3965 | CAACA | ATTGACA | TTCTA | CACAG | GGAGA | 4014 | | |
| CVF2 | cDNA | 3013 | CAACA | ATTGACA | GTCTA | CACA | GGAG | 3062 | | |
| coC3 | cDNA | 3988 | CAACA | ATTGACA | GTCTA | CACA | GGAG | 4037 | | |
| CVF2 | Gen | | CAACA | ATTGACA | GTCTA | CACA | GGAG | | | |
| CVF1 | cDNA | 4015 | AATGT | TTGCA | ATAAA | TCTTA | TCTGT | 4064 | | |
| CVF2 | cDNA | 3063 | AATGT | TTGCA | ACAA | TCTT | TCTGT | 3112 | | |
| coC3 | cDNA | 4038 | AATGT | TTGCA | ACAA | TCTT | TCTGT | 4087 | | |
| CVF2 | Gen | | AATGT | TTGCA | ACAA | TCTT | TCTGT | | | |
| CVF1 | cDNA | 4065 | GA | | | | | | 4066 | |
| CVF2 | cDNA | 3113 | GA | | | | | | 3114 | |
| coC3 | cDNA | 4088 | GA | | | | | | 4089 | |
| CVF2 | Gen | | GA gtaagttt ggcttcctag aaatagtagg ttgtaaggta ttaaaatttt | | | | | | | |
| CVF2 | Gen | | agtgttcacc atgtcagtat ctcaatatga aacaagtttt ctctgctatt | | | | | | | |
| CVF2 | Gen | | tattcataaa | | | | | | | |
| CVF2 | Gen | | (ca. 3500 bp) | | | | | | | |
| CVF2 | Gen | | ccaaggagat | | | | | | | |
| CVF2 | Gen | | gttatggcca gttagacaat caaacctttc acctttaaaa aaactgctaa | | | | | | | |
| CVF1 | cDNA | | --- | | | | | | | |
| CVF2 | cDNA | 3115 | ACT | | | | | | 3117 | |
| coC3 | cDNA | 4090 | ACT | | | | | | 4092 | |
| CVF2 | Gen | | gaatttgttc agagtctgat tttccccttt tctttggcaa ttcttag ACT | | | | | | | |
| CVF1 | cDNA | 4067 | ----- | ATGC | AATGG | AAGGG | TCATG | GATCT | 4110 | |
| CVF2 | cDNA | 3118 | TAAA | AGC | AA | AGG | AGCC | TCA | AGCT | 3167 |
| coC3 | cDNA | 4093 | TAAA | AGC | AA | AGG | AGCC | TCA | AGCT | 4142 |
| CVF2 | Gen | | TAAA | AGC | AA | AGG | AGCC | TCA | AGCT | |

Abb. 3.57: Sequenz des genomischen PCR-Produkts mit Intron 32 des CVF2-Gens. Intron 32 hat eine Größe von ca. 3700 bp. In Exon 33 weicht die genomische CVF2-Sequenz in einem Nukleotid (umrahmt) von der CVF2-cDNA-Sequenz ab. Intron 31 ist im CVF2-Gen nicht vorhanden.

Trotz ausgeprägter Sequenzhomologie zu coC3 war aufgrund der spezifischen Sequenz eine Zuordnung zum CVF2-Gen möglich. Lediglich an einer Position in Exon 33 (umrahmt) stimmt die genomische Sequenz nicht mit der CVF2-cDNA überein.

Ebenso wie im CVF1-Gen sind auch im CVF2-Gen Exons 31 und 32 kontinuierlich und nicht durch Intronsequenzen unterbrochen.

3.2.2 Aufklärung der CVF2-Genstruktur durch Genome Walking

Die Genome Walking-Methode wurde eingesetzt, um die Promotorregion des CVF2-Gens aufzuklären und somit eine Identifizierung regulatorischer Elemente zu ermöglichen. Darüber hinaus wurde die Methode genutzt, um Sequenzinformationen aus dem noch unbekanntem 5'-Bereich der codierenden CVF2-Sequenz zu erhalten. Alle erhaltenen CVF2-Produkte wurden mit Hilfe von Primern aufgeklärt, die nicht hochspezifisch zwischen CVF2 und den homologen Sequenzen differenzierten (Tab. 3.8).

Tab. 3.8: Übersicht der Genome Walker-Amplifikate des CVF2-Gens. Das jeweils größte sequenzierte PCR-Produkt sowie die dazu eingesetzte Genome Walking (GW)-Bibliothek sind angegeben. Größenangaben sind Schätzungen.

| Exon-Intron-Grenze | Primerpaare | Größe PCR-Produkt (GW-Bibliothek) | Exon-Intron-Übergangssequenz |
|--------------------------|---|-----------------------------------|------------------------------|
| Promotor/Exon 1/Intron 1 | CoC3-Intr.1as-I/AP1 CoC3-Intr.1as-II/AP2 | 1480 bp (StuI) | TCT gtaagtatca... |
| Intron 1/Exon 2 | HB2as-CVF-II/AP1 HB2as-CVF/AP2 | 1000 bp (StuI) | ...tcttctgcag CTA |
| Exon 2/Intron 2 | HB2s-CVF/AP1 HB2sII/AP2 | 680 bp (DraI) | AAG gtgggtgtat... |

Promotorregion

Da für das 5'-Ende des CVF2-Gens keine Sequenzinformationen verfügbar waren, war eine gezielte Amplifikation mit Hilfe CVF2-spezifischer Primer nicht möglich. Stattdessen wurde im Zuge der Aufklärung des coC3-Promotors (Kapitel 3.3) zusätzlich zum coC3-Genabschnitt auch der gesuchte CVF2-Genabschnitt mit der Promotorregion amplifiziert (Abb. 3.58). Zunächst wurde die Intron 1-Sequenz des coC3-Gens aufgeklärt, um anschließend mit Hilfe von coC3-spezifischen Intronprimern den coC3-Promotor zu amplifizieren. Eine verschachtelte PCR mit den Primerpaaren CoC3-Intr.1as-I /AP1 und CoC3-Intr.1as-II / AP2 führte zu den entsprechenden Amplifikaten sowohl des coC3- als auch des CVF2-Gens (Abb. 3.58).

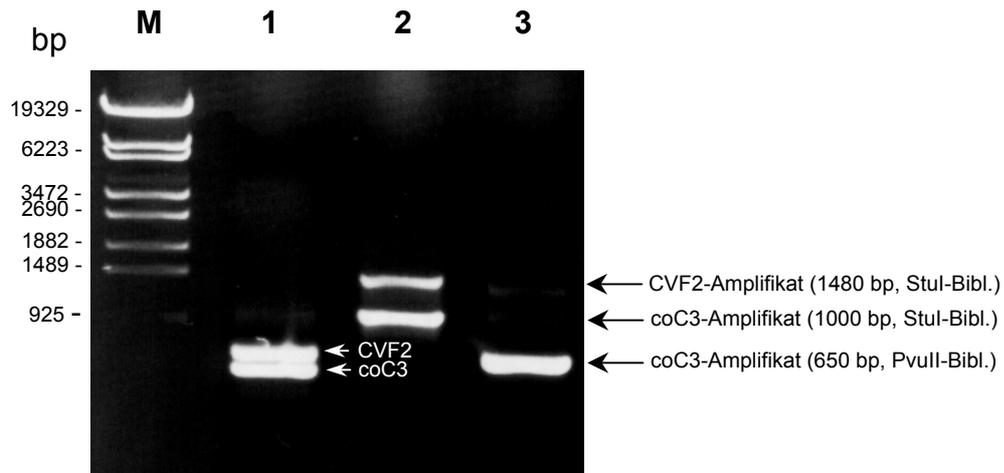


Abb. 3.58: Gelelektrophoretische Analyse der GenomeWalker-Amplifikate (Promotorbereich, Intron 1 stromaufwärts). M: DNA-Marker, Bahn 1: DraI-Bibliothek, Bahn 2: StuI-Bibliothek, Bahn 3: PvuII-Bibliothek. Die markierten Banden der StuI-Bibliothek (Bahn 2) wurden sequenziert. Die weiteren markierten Banden stimmen in ihrer Größe mit den entsprechenden CVF2- bzw. coC3-Gensequenzen überein.

Die schwerste Bande der StuI-Bibliothek (ca. 1480 bp) wurde sequenziert. Aufgrund charakteristischer Unterschiede in Exon 1 zu den cDNA-Sequenzen von CVF1 und coC3 wurde die Sequenz als CVF2-Genabschnitt identifiziert (Abb. 3.59).

| | | | | | | | |
|------------------|-------|---------------------------|--------------------|--------------------|--------------------|----------------------------|-------|
| CVF2 Gen | -1301 | | | <u>AGGCC</u> | <u>TGTTAATCT</u> | <u>TTTTCATGGC</u> | -1277 |
| CVF2 Gen | -1276 | CGCAGCTGAG | GATTTGGTGG | CTTCCGCTGA | AACAGCCTCT | TCAATGGCCT | -1227 |
| CVF2 Gen | -1226 | <u>CTTTAAACGT</u> | GATCTTAGGT | TTCATTAAAA | GTTTGCTTTG | TAAGATCTTA | -1177 |
| CVF2 Gen | -1176 | TCTTTAAGGC | CCAAAACAAA | CCGATCACGC | ATGGCCTCTT | CTACGTTGGT | -1127 |
| | | | CCAAT | | | | |
| CVF2 Gen | -1126 | AAAATCACAA | GCACCC CCAA | TGTCCCTTAA | GCTTGCCACG | AAGTCATTAA | -1077 |
| CVF2 Gen | -1076 | CACTCTCACT | ATCCCCTTGA | TCCCGACGGT | AAAATTCGTG | TCTCTTAATG | -1027 |
| | | | CCAAT | | | | |
| CVF2 Gen | -1026 | AGGCACGATA | CTTGTGGGGT | GTAATGGGCT | CTTAAAAGTC | TTT CCAAT TC | -977 |
| | | | | | SP1 | | |
| CVF2 Gen | -976 | GGGCCATTCC | ATTACCTTCA | ATTCTGCTGC | TCCCGCCAGC | GCCTTCGCCA | -927 |
| | | | | | SP1 | | |
| CVF2 Gen | -926 | CCTTGAAGAC | TGCCGGACCG | CAGTGTGTGA | GAAAG CCGC | CC TCTTCCGT | -877 |
| CVF2 Gen | -876 | CTCTCGGGGA | TGCCGGCGAG | ATCATTTCCT | TCCAGGAACA | TGTGGAAACG | -827 |
| CVF2 Gen | -826 | ATCCGCATAC | GATTCCCAGT | CTTCCGTCGC | CGGATCGAAC | GGTGTGGGG | -777 |
| CVF2 Gen | -776 | GGGTGTTTCA | CGCAGCCATC | TCACTCTGNT | GCTTGTTCGT | CGCGTCGCTC | -727 |
| | | | NF-kB | | | | |
| CVF2 Gen | -726 | TCTCTT CAGG | GAAATCCCAC | CTTCGTCGCC | AGTGTTAAGT | TTGCATGAAG | -677 |
| CVF2 Gen | -676 | AGTCANACGC | GATGAGTTCC | AGCAAAAGGC | TTTATTAGTG | NCTNCAGCTC | -627 |
| CVF2 Gen | -626 | GAATGGNCCC | GTCCGGGTCA | GGCTNCGGCT | GCTTCTTATA | GCCTGCAGCC | -577 |
| | | | CCAAT | | | | |
| CVF2 Gen | -576 | GTAGCTGCCA | GCCAAT GGGA | GTGCAGGGTT | TCTTCCCACA | CTATTTAAAC | -527 |
| | | CCAAT | | GR | | CCAAT | |
| CVF2 Gen | -526 | CAAT ACCATT | TGACATTTTA | CAAGAACAGT | TCAAATAAAC | CAAT AACATT | -477 |
| CVF2 Gen | -476 | TAAACCTACA | TACATAACAA | TTTCTTTTTC | AAAAAACCTG | TGGAAACAAC | -427 |
| | | | GATA | | | | |
| CVF2 Gen | -426 | AGTAATGCCT | TAATACAAAT | ACTG TGATAA | GTTTAGGAAC | AGACTTTGGG | -377 |
| | | | CACCC | | | | |
| CVF2 Gen | -376 | CTGTTGAGAA | GTTGGGCT CA | CCC AAAATT | GGGCTACTTG | AAAACCTCTT | -327 |
| CVF2 Gen | -326 | CCAAATATGT | GCATGAGGAA | GAGTGGCGAA | GTGAGCTTGT | <u>GAATTC</u> CAGA | -277 |
| | | GATA | HNF-5 | | | | |
| CVF2 Gen | -276 | TCACA AGATA | AAGGAA TATT | TGC AGATACA | ATAGCAGAAT | ACCAAAACCA | -227 |
| CVF2 Gen | -226 | AGTTGCATAA | AAACTGCCAT | TTAGCCATTG | CATTGAAACC | AGAGTGAGAA | -177 |
| | | | CCAAT | | | | |
| CVF2 Gen | -176 | GACAAATAAT | CCAAT GTGAT | TTTTTTGTTA | ATTTCTCTG | AAATTCCCTG | -127 |
| | | | | AP-2 | | | |
| CVF2 Gen | -126 | TCTGCTCAAG | CTCCTCCTCA | CCTC CCCCAC | CC TCTTTTAT | ACTGATTAAA | -77 |
| | | | | | | TATA | |
| CVF2 Gen | -76 | CACTCATCAA | TTGCAAAGGG | GCTTGGCTTC | AAAGCAGGCT | CTCTGG ATAA | -27 |
| | | | | | | Exon 1 | |
| CVF1 cDNA | 1 | | | | | C CCATGGAGAG | 11 |
| coC3 cDNA | 1 | | | TIS (+1) | GGACTA | CCATGGAGGG | 16 |
| CVF2 Gen | -26 | ATA CTGGAGT | TGTCAGCCCT | TCTCTT CCCC | CTCTGGACTA | CCATGGAGGG | 24 |
| CVF1 cDNA | 12 | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGGGTCTT | 61 |
| coC3 cDNA | 17 | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGGGTCTT | 66 |
| CVF2 Gen | 25 | GAC GGCTCTC | TCT CTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGG ATCTT | 74 |
| CVF1 cDNA | 62 | CTCATGGGGC | TCT | | | | 74 |
| coC3 cDNA | 67 | CC CA GGGGC | TCT | | | | 79 |
| CVF2 Gen | 75 | CTCATGGG AC | TCT gtaagta | tcatagatat | ttcagcaggg | ggaataaagc | 124 |
| CVF2 Gen | 125 | agtgctgata | aacccttgty | gtcctttgca | gggtaatgaa | tttccttcca | 174 |

Abb. 3.59: Promotorregion des CVF2-Gens. Die Region stromaufwärts von Exon 1 ist in Großbuchstaben dargestellt. Putative Bindungstellen für Transkriptionsfaktoren sowie die TATA-Box und der putative Transkriptions-Initiations-Start (TIS) sind fett und kursiv dargestellt. Die StuI-Schnittstelle (AGGCCT) am 5'-Ende sowie Schnittstellen für DraI (zwei, TTTAAA), PvuII (CAGCTG) und EcoRI (GAATTC) sind unterstrichen.

Die vollständige Sequenz des Amplifikats wurde durch zusätzliche Sequenzierung mit den internen Primern HB-coC3-II-Prom-s und HB-coC3-II-Prom-as erhalten. Der putative Transkriptionsstart von CVF2 wurde mit TIS (+1) gekennzeichnet und die Sequenz von Exon 1 in fetten Großbuchstaben markiert. Im Promotorbereich sind putative Bindungsstellen für Transkriptionsfaktoren, deren regulatorische Funktionen in Kapitel 4 beschrieben sind, markiert.

Intron 1/Exon 2

Bei der verschachtelten PCR mit den Primerpaaren HB2as-CVF-II/AP1 und HB2as-CVF/AP2 zur Charakterisierung des Übergangs Intron 1/Exon 2 wurden nur schwache Amplifikate erhalten (Abb. 3.60).

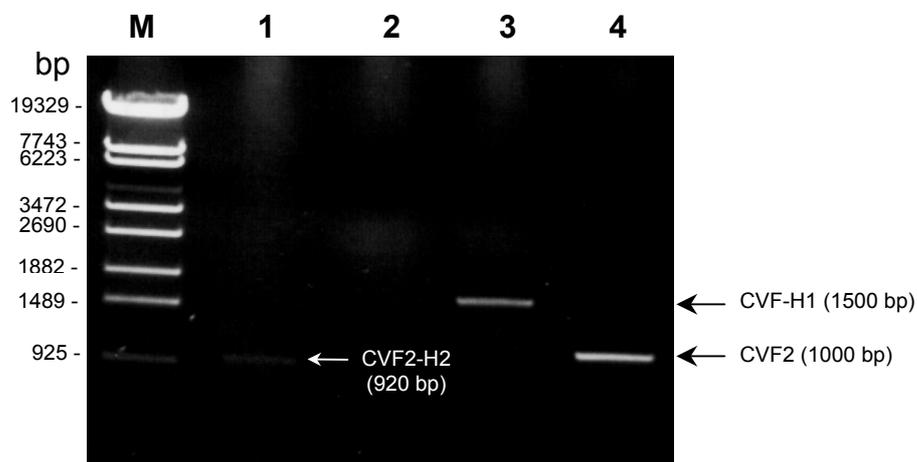


Abb. 3.60: Gelelektrophoretische Analyse der GenomeWalker-Amplifikate (Exon 2 stromaufwärts). M: DNA-Marker λ /Eco130I, Bahn 1: DraI-Bibliothek, Bahn 2: EcoRV-Bibliothek, Bahn 3: PvuII-Bibliothek, Bahn 4: StuI-Bibliothek.

Das StuI-Amplifikat von ca. 1000 bp (Abb. 3.60, Bahn 4) wurde gereinigt und als Template für eine neue Sekundär-Amplifikation mit dem Primerpaar HB2as-CVF/AP2 eingesetzt, um zwecks Klonierung eine hohe Ausbeute zu erhalten. Die Sequenz dieses Amplifikats ist in Abbildung 3.61 dargestellt.

| | | | | | | | |
|------------------|-------|-------------------|--------------------|--------------------|--------------------|-------------------|-------------------|
| CVF2 Gen | | cacatcttta | accatattatt | tcttcagttt | tctaaaagaa | | |
| CVF2 Gen | | gagatcatga | aaaacttcat | caaatgcctc | acttttccta | agcaaactaa | |
| CVF1 cDNA | 75 | | | CTACACCCCTC | ATCACCCCTG | | 94 |
| coC3 cDNA | 80 | | | CTAACCCTC | ATCACCCCTG | | 99 |
| CVF2 Gen | | ctcaggagat | gttcttttct | tcttctgcag | CTACACCCCTC | ATCACCCCTG | |
| CVF1 cDNA | 95 | CTGTTTTGCG | AACAGACACA | GAAGAGCAA | TTTTGGTGG | GGCCCATGGA | 144 |
| coC3 cDNA | 100 | CTGTTTTGCG | AACAGACACA | GAAGAGCAA | TTTTGGTGG | GGCCCATGGA | 149 |
| CVF2 Gen | | ATGTTCTGCG | AACAGACA | TC | GAAGAGCAA | TTTTGGTGG | GGCCATGGA |
| CVF1 cDNA | 145 | GACAGTACTC | CAAAACAGCT | TGACATCTTT | GTTCATGATT | TTCCACGGAA | 194 |
| coC3 cDNA | 150 | GACAGTACTC | CAAAAACAGCT | TGACATCTTT | GTTCATGATT | TTCCACGGAA | 199 |
| CVF2 Gen | | GACAGTACTC | CAAAACAGCT | TGACATCTTT | GTTCA | CAATT | TTCCACAGAA |
| CVF1 cDNA | 195 | GCAGAAAACC | TTGTTCCAAA | CCAGAGTAGA | TATGAATCCA | GCAGGAGGCA | 244 |
| coC3 cDNA | 200 | GCAGAAAACC | TTGTTCCAAA | CCAGAGTAGA | TATGAATCAG | GCAGGAGGCA | 249 |
| CVF2 Gen | | GCAGAAAATC | TTGTACAAA | CCAGAGTAGA | TATGAATCCA | GCAGGAGGCA | |

Abb. 3.61: Darstellung der partiellen Intron 1/Exon 2-Sequenz des CVF2-Gen. Die hier dargestellte und mit einem Pfeil markierte CVF1-Primersequenz weicht an zwei Positionen (umrahmt) von der tatsächlichen CVF2-Sequenz ab: CCAGAGTAGATATGAATCCAGAGGAGGCA.

Aufgrund der Unterschiede zu den cDNA-Sequenzen von CVF1 und coC3 sowie der überlappenden Sequenz-Identität mit weiteren CVF2-Amplifikaten (Abb. 3.55 und 3.62) wurde die Sequenz dem CVF2-Gen zugeordnet.

Das ebenfalls sequenzierte PvuII-Amplifikat (1500 bp) wies auf einen neuen homologen Genabschnitt hin und wurde mit „CVF-Homolog 1“ (CVF-H1) bezeichnet. Eine ausführliche Charakterisierung erfolgt in Kapitel 3.4. Bei der sehr schwachen Bande der DraI-Bibliothek (Bahn 1, ca. 920 bp) handelt es sich um ein Amplifikat eines weiteren neuen, mit „CVF-Homolog 2“ (CVF-H2) bezeichneten Genabschnitts (siehe Kapitel 3.4), da es in seiner Größe mit dem dort beschriebenen DraI-Amplifikat korreliert.

Exon 2/Intron 2

Bei einer verschachtelten PCR mit den adapterligierten DraI- und PvuII-Bibliotheken wurden unter Verwendung der Primerpaare HB2s-CVF/AP1 und HB2sII/AP2 Amplifikate von ca. 670 bp bzw. 560 bp erhalten (Abb. 3.37, Kapitel 3.1.3), welche durch Sequenzierung als CVF2-Genabschnitte identifiziert wurden. Die Konsensus-Sequenz beider Amplifikate ist in Abb. 3.62 dargestellt.

| | | | | | | | |
|-----------|-----|-------------|------------|-------------|------------|------------|-----|
| CVF1 cDNA | 137 | CCCATGGAGA | CAGTACTCCA | AAACAGCTTG | ACATCTTTGT | TCATGATTTT | 186 |
| coC3 cDNA | 142 | CCCATGGAGA | CAGTACTCCA | AAATCGCTTG | ACATCTTTGT | TCATGATTTT | 191 |
| CVF2 Gen | | CCCATGGAGA | CAGTACTCCA | AAACAGCTTG | ACATCTTTGT | TCACAATTTT | |
| | | | | → | | | |
| CVF1 cDNA | 187 | CCACGGAAGC | AGAAAACCTT | GTTCCAAACC | AGAGTAGATA | TGAATCCAGC | 236 |
| coC3 cDNA | 192 | CCACGGAAGC | AGAAAACCTT | GTTCCAAACC | AGAGTAGATA | TGAATCCAGC | 241 |
| CVF2 Gen | | CCACA GAAGC | AGAAAATCTT | GTA CCAAACC | AGAGTAGATA | TGAATCCAGA | |
| CVF1 cDNA | 237 | AGGAGGCATG | CTTGTCACTC | CAACTATAGA | G | | 267 |
| coC3 cDNA | 242 | AGGAAGCATG | CTTGTCACTC | CAACTATAAA | G | | 272 |
| CVF2 Gen | | CGGAGGCATG | CTTGTCACTC | CAACTATAAA | Ggtgggtgta | tggtgatggt | |
| CVF2 Gen | | tgcaatagag | tctttggtat | cctgcgaaa | gaacagaaca | aacaaacaaa | |
| CVF2 Gen | | cccaaactaa | aatcctgtcc | agtttggtt | gagggtaagt | | |

Abb. 3.62: Darstellung der partiellen Intron 2-Sequenz des CVF2-Gens.

Die hier gezeigte Sequenz ist identisch mit dem 10 kb-genomischen Amplifikat des CVF2-Gens (Kapitel 3.2.1).

Zusammenfassende Darstellung der neu aufgeklärten CVF2-Sequenz

Vom zuvor unbekanntem 5'-Bereich des CVF2-Gens konnte durch genomische PCR und Genome Walking der Bereich vom Promotor bis Exon 3 aufgeklärt werden. In Abbildung 3.63 ist die neu identifizierte Konsensus-Sequenz von CVF2 ab dem Startcodon zusammengefaßt und mit den entsprechenden CVF1- und coC3-Sequenzen verglichen.

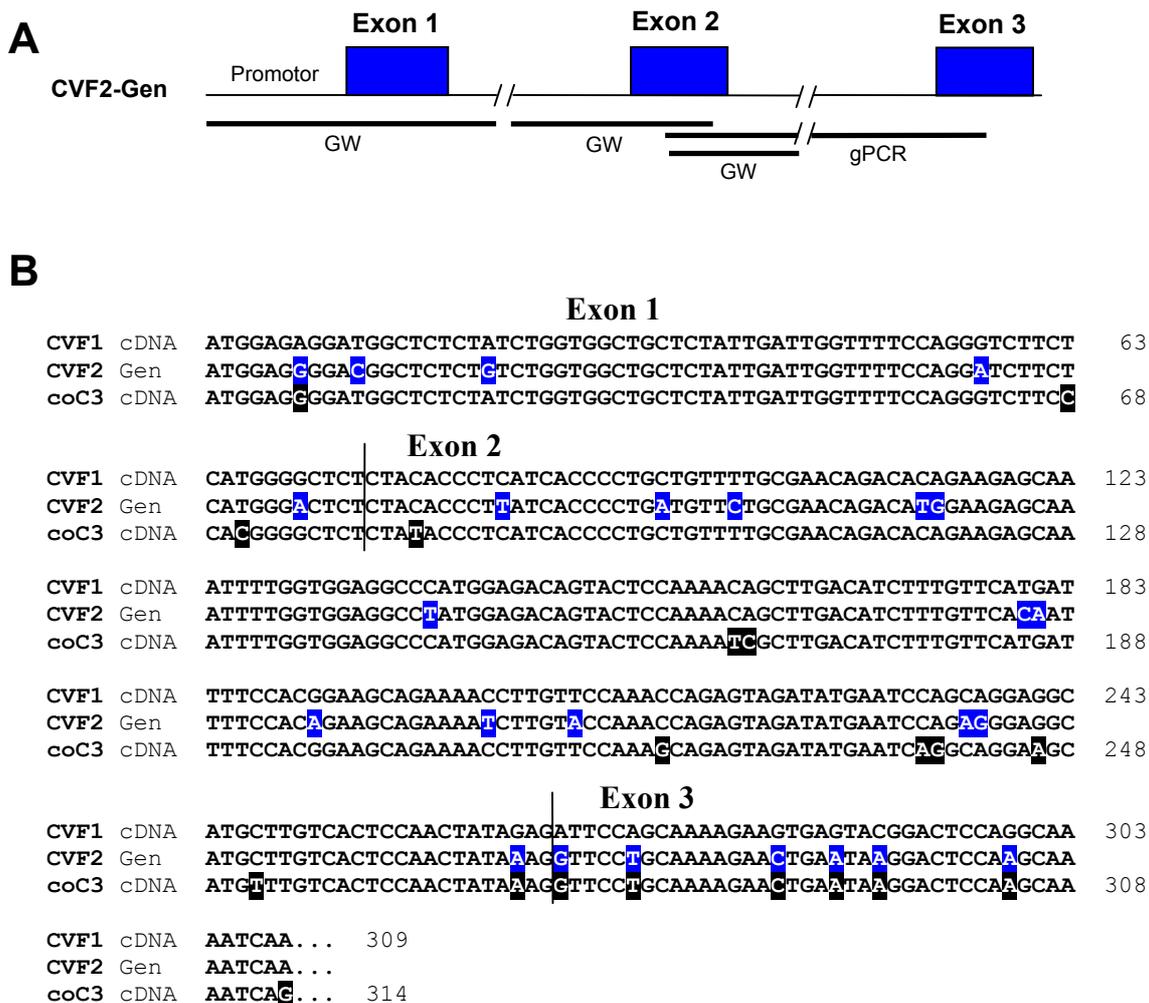


Abb. 3.63: Aufklärung der Exon-Sequenzen des CVF2-Gens (5'-Bereich).

A: Graphische Darstellung zur Position der PCR-Produkte. Die durch Genome Walking erhaltenen PCR-Produkte sind mit "GW" und das genomische PCR-Produkt mit "gPCR" gekennzeichnet.

B: Darstellung der Exonsequenzen des CVF2-Gens. Die CVF2-Sequenzen wurden durch genomische PCR und Genome Walking ermittelt und sind vom Startcodon bis Exon 3 zusammenhängend dargestellt. Für Vergleichszwecke sind die cDNA-Sequenzen von CVF1 und coC3 gezeigt. Unterschiede zur CVF1-Sequenz sind durch Boxen markiert.

3.2.3 Ansatz zum proteinbiochemischen Nachweis von CVF2

Die Transkription von CVF2 und deren Nachweis auf cDNA-Ebene deuten darauf hin, dass CVF2 möglicherweise als Protein exprimiert wird, dessen Identifizierung bisher jedoch noch nicht gelungen war. Als Nachweismethoden des CVF2-Proteins bieten sich zweidimensionale Gelelektrophorese von Gesamt-Giftproteinen mit anschließendem Immunoblot mit anti-CVF-Antikörpern sowie Immunoblot mit CVF2-spezifischen Antikörpern an. Nachfolgend ist die Expression eines CVF2-Peptids beschrieben, welches für die Generierung CVF2-spezifischer monoklonaler Antikörper mittels Phage Display geeignet erscheint.

Die CVF2-Peptidsequenz wurde aufgrund eines Sequenzvergleichs mit CVF1 ausgewählt und zeichnet sich durch eine relativ geringe Sequenzidentität (62%) mit CVF1 aus (Abb. 3.64). Die ausgewählte Region entspricht Exons 18 und 19 und codiert für einen Bereich, der in CVF1 am N-Terminus der γ -Kette liegt.

A

N-Terminus der γ -Kette (CVF1)

```
CVF1 DDNEDGFIADSDIISRSDFPKSWLWLTkdLTEEPNSQGISSKtMSfYLRDSITTWVVLAVSfTPTK
CVF2 SDFEDDLfGEGNITSRSDFPESWLWlMEQLSEHPNSKGISSKIVPFYLRDSITTWELLAVGLSPTK
* ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

B

```

Exon 18
| S D F E D D L F G E G N I T S R S D F P E S W L W
| agtgat|tttgaagatgattt|at|tttg|gagaag|gtaacatcacctcaaggtctgatttt|cctgagag|ttggttg|gg
|
| L M E Q L S E H P N S K G | I S S K I V P F Y L R D
| ctaatggagcagctgtctgaacatcctaacagtaaagg|gatttcaagcaagatag|taccttttatctgagggat
|
| S I T T W E L L A V G L S P T K |
| tccatcacaacctgggaggttgctggctgtgggcctttcaccaccaaag|

```

Abb. 3.64: Sequenz des CVF2-Peptides. A: Vergleich der abgeleiteten Aminosäuresequenzen von CVF1 (AS 711-776) und CVF2 im 5'-Bereich der CVF1- γ -Kette. Identische Aminosäuren sind durch Sternchen markiert. B: Nukleotid- und abgeleitete Aminosäuresequenz des CVF2-Peptides (66 AS). Die Exongrenzen sind markiert.

Unter Verwendung genomischer Kobra-DNA als Template wurde mit dem CVF2-spezifischen Primerpaar CVF2-Ex18s/CVF2-Ex19as zunächst Intron 18 unter vollständiger Einbeziehung der flankierenden Exons amplifiziert (Abb. 3.54) und seine Identität durch Sequenzierung bestätigt (Abb. 3.56).

Anschließend wurde das genomische CVF2-Amplifikat als Template eingesetzt, um daraus Exon 18 und Exon 19 mit den Primerpaaren CVF2-Ex18s/CVF2-Ex18as-Ü und CVF2-Ex19s-Ü/CVF2-Ex19as zu amplifizieren. Die beiden inneren Primer waren mit einem Sequenz-Überhang der jeweils benachbarten Exonsequenz so modifiziert, dass anschließend eine überlappende PCR durchgeführt und mit den Primern CVF2-Ex18s/CVF2-Ex19as sodann das intronfreie Amplifikat erhalten werden konnte (Abb. 3.65). Dieses wurde in den Vektor pCR2.1-TOPO kloniert und sequenziert.

Zum Zwecke der Umklonierung in den Expressionsvektor pET-26b(+) wurde mit den modifizierten Primern CVF2-Ex18s-NcoI und CVF2-Ex19as-XhoI, welche die gewünschten Restriktionsschnittstellen NcoI bzw. XhoI enthalten, die CVF2-Sequenz aus dem Plasmid pCR2.1-TOPO-CVF2-Ex18/19 amplifiziert, mit NcoI und XhoI verdaut und in den NcoI/XhoI-geschnittenen Vektor pET-26b(+) kloniert. Das erhaltene Plasmid enthält stromaufwärts der CVF2-Sequenz die pelB-Signalsequenz für periplasmatische Sekretion sowie am 3'-Ende ein His-Tag für Affinitätsreinigung mittels IMAC (*immobilized metal ion affinity chromatography*). Für Expressionszwecke wurden *E. coli* BL21(DE3)-pLysS mit dem Plasmid pET-26b(+)-CVF2-Ex18/19 transformiert und eine exponentiell wachsende Kultur mit einer finalen Konzentration von 1 mM IPTG induziert. Nach einem denaturierenden Zellaufschluss erfolgte die Proteinreinigung mittels IMAC. In Abbildung 3.66 ist das gereinigte CVF2-Peptid gezeigt.

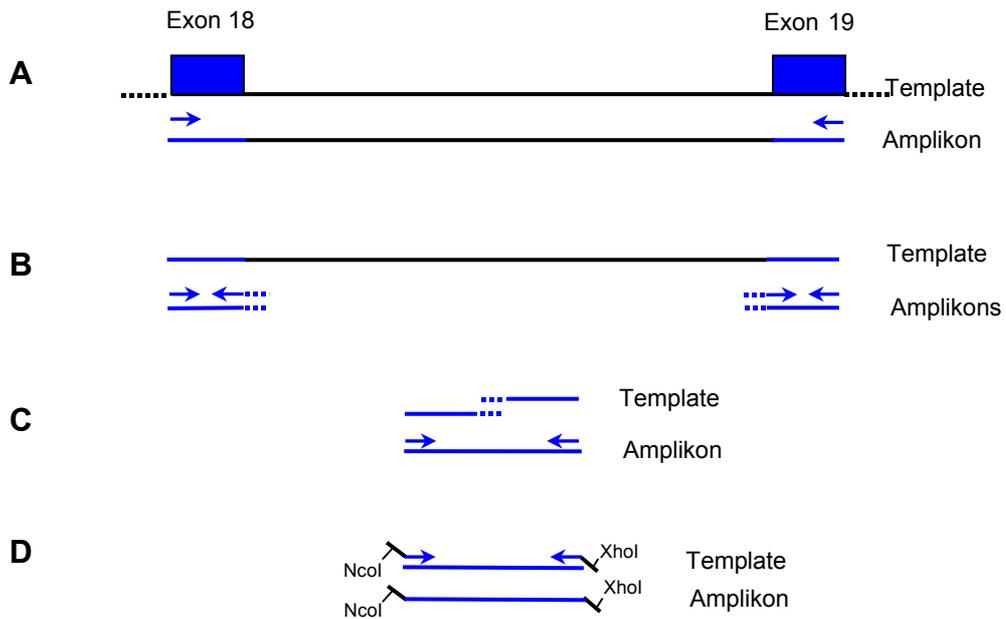


Abb. 3.65: Strategie zur Klonierung des CVF2-Fragments.

- A: Genomische Amplifikation von Exon 18-Exon 19 des CVF2-Gens
- B: Separate Amplifikation von Exon 18 und Exon 19 mit partiell überlappenden Primern
- C: Verknüpfungs-PCR von Exon 18 und Exon 19
- D: Amplifikation mit modifizierten Primern

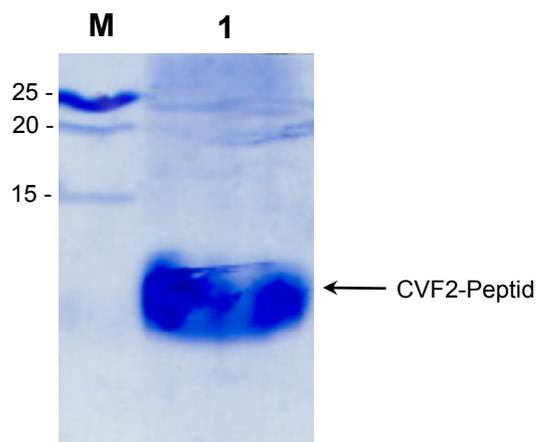


Abb. 3.66: SDS-PAGE-Analyse des bakteriell exprimierten CVF2-Peptids. M: Marker, Bahn 1: Gereinigtes His-Tag-Peptid CVF2-Ex18/19 (8 kDa) nach Nickelchelate-Affinitätschromatographie. Coomassie-gefärbtes SDS-Polyacrylamidgel (17.5%, reduzierend).

3.3 Struktur des Cobra C3-Gens

Die Struktur des coC3-Gens wurde mit dem Ziel analysiert, Gemeinsamkeiten und Unterschiede zu den Genstrukturen von CVF1 und CVF2 aufzuzeigen. Dazu wurde in Analogie zur Aufklärung der CVF2-Genstruktur verfahren.

3.3.1 Aufklärung der Cobra C3-Genstruktur durch genomische PCR

Zur Analyse der Struktur des Cobra C3 (coC3)-Gens wurden mittels genomischer PCR exemplarisch einige ausgewählte coC3-Introns amplifiziert. Die Amplifikate entstanden teilweise als Co-Amplifikate der entsprechenden homologen Genabschnitte oder wurden als ausschließliches Produkt gebildet (Abb. 3.67).

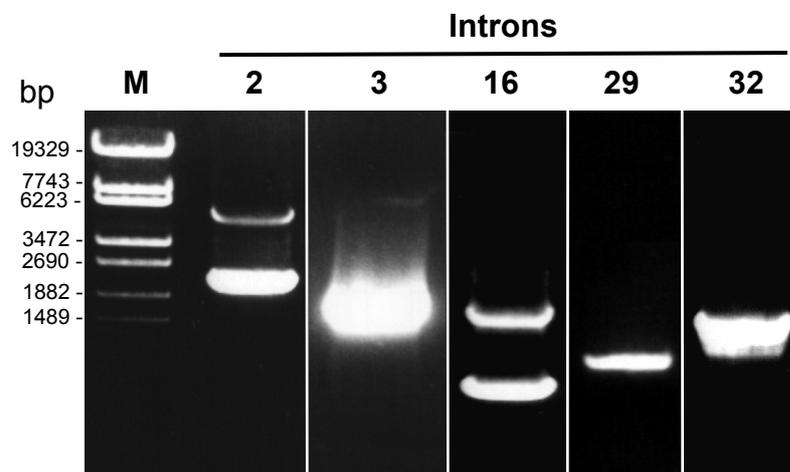


Abb. 3.67: Gelelektrophoretische Analyse der genomischen PCR-Amplifikate des coC3-Gens. M: DNA-Marker λ /Eco130I; Intron 2: oben CVF1? (5 kb), unten coC3 (2,5 kb); Intron 3: 1,8 kb; Intron 16: oben CVF1 (1,6 kb), unten coC3 (903 bp); Intron 29: 1250 bp; Intron 32: 1,7 kb.

Die oben in Abbildung 3.67 dargestellten PCR-Amplifikate sind in Tabelle 3.9 bezüglich der verwendeten Primerpaare und Exon/Intron-Übergangssequenzen näher charakterisiert. Die vollständigen Sequenzen sind in Anhang 7.3 dargestellt.

Tab. 3.9: Übersicht der genomischen PCR-Amplifikate des coC3-Gens. Die Größenangaben in Klammern sind Schätzungen.

| Intron | Primerpaar | Größe PCR-Produkt (bp) | Exon-Intron-Übergangssequenzen |
|--------|-------------------------------|------------------------|---|
| 2* | HB2s-CVF / HB3as-CVF | (2600) | AAG gtgggtgat...tctcatccag GTT |
| 3* | YL3s-C3 / YL4as-C3 | (1800) | CAG gtaaccggt...tttttcttag TGC |
| 16 | OPS72 / OPA24 | 903 | AAG gtatgtggca...catatcttag CGG |
| 29 | YL29s-C3/YL30as-C3 | (1250) | CAA gtgagtcttt...cccattctag GCA |
| 32 | HB31s-CVF2,coC3 / HB33as-coC3 | (1700) | TGA gtaagtttg...taatttttag ACT |

*Eine Unterscheidung von Intron 2 vom entsprechenden CVF2-Genabschnitt war trotz fehlender CVF2-cDNA-Sequenzen mittels in dieser Arbeit aufgeklärten homologen Gensequenzen (siehe Kapitel 3.2, Abb. 3.63B, und Kapitel 3.4, Abb. 3.80) möglich. Intron 3 wurde aufgrund Übereinstimmung der Exonsequenzen mit der entsprechenden coC3-cDNA-Sequenz zugeordnet.

Intron 2

Mit den CVF1-spezifischen Primern HB2s-CVF und HB3as-CVF sollte ursprünglich der entsprechende CVF1-Genabschnitt amplifiziert werden. Durch Verwendung der Taq/Tgo-Polymerase-Mischung *Expand* (Roche), welche eine 3'-5'-Exonuklease-Aktivität (Korrektur-Aktivität) aufweist, wurde ein ca. 2,6 kb langes Amplifikat erhalten (Abb. 3.67), das als coC3-Genabschnitt identifiziert wurde (Abb. 3.68).

| | | | |
|------------------|-----|---|-----|
| CVF1 cDNA | 137 | CCC ATGGAGA CAG TACTCCA AAAC AGCTTG ACAT C'TTTGT TCAT GATTTT | 186 |
| coC3 cDNA | 142 | CCC ATGGAGA CAG TACTCCA AAAT CGCTTG ACAT C'TTTGT TCAT GATTTT | 191 |
| coC3 Gen | | CCC ATGGAGA CAG TACTCCA AAAC CGCTTG ACAT C'TTTGT TCAT GATTTT | |
| CVF1 cDNA | 187 | CCAC GGAAGC AGAAA ACCTT GTT CCAAACC AGAG TAGATA TGAAT CCAGC | 236 |
| coC3 cDNA | 192 | CCAC GGAAGC AGAAA ACCTT GTT CCAAAAC AGAG TAGATA TGAAT CAGGC | 241 |
| coC3 Gen | | CCAC GGAAGC AGAAA ACCTT GTT CCAAAAC AGAG TAGATA TGAAT CAGGC | |
| CVF1 cDNA | 237 | AGG AGGCATG CTT GTCACTC CAACT TATAGA G | 267 |
| coC3 cDNA | 242 | AGG AAGCATG TTT GTCACTC CAACT TATAA G | 272 |
| coC3 Gen | | AGG AAGCATG TTT GTCACTC CAACT TATAA Ggtgggtgta tggatgatggt | |
| coC3 Gen | | tgcaatagag tctactgtat tctgcaaat atctatatct atctatttat | |
| coC3 Gen | | ataaaaatgg (ca. 2300 bp) tagggctgag | |
| coC3 Gen | | gaatgaggcc tcaggtggcc cttagtgatg attaagtttg aagcctctgc | |
| CVF1 cDNA | 268 | ATT CCAGCA AAAGA AGTGA GTAC GGACTC | 296 |
| coC3 cDNA | 273 | GTT CCAGCA AAAGA ACTGA ATA AGGACTC | 301 |
| coC3 Gen | | tctactcatc ttctcatcca g GTT CCAGCA AAAGA ACTGA ATA AGGACTC | |
| CVF1 cDNA | 297 | CAG GCAAAT CAAT TATGTGG TTG TGCAAGT AACT GGTCCT CAAG TGAGAT | 346 |
| coC3 cDNA | 302 | CAAG CAAAAT CAG TATGTGG TTG TGAAAGT AACT GGTCCT CAAG TGCAAT | 351 |
| coC3 Gen | | CAAG CAAAAT CAG TATGTGG TTG TGAAAGT AACT GGTCCT CAAG TGAGAT | |

Abb. 3.68: Sequenz des genomischen PCR-Produkts mit Intron 2 des coC3-Gens. Exon 2 hat eine Größe von 193 bp. Intron 2 hat eine Größe von ca. 2500 bp. Eine StuI-Schnittstelle (aggcct) im Intron ist unterstrichen.

Die Sequenz wies die Besonderheit auf, dass die 3'-Enden beider CVF1-spezifischer Primer im Sinne der coC3-Sequenz verändert waren. Im Falle des Vorwärts-Primers HB2s-CVF wurde von den beiden CVF1-spezifischen Nukleotiden am 3'-Ende nur das äußere Nukleotid im Sinne der coC3-Sequenz verändert (A-C). Das bedeutet, dass durch die Korrektur-Aktivität der Tgo-Polymerase die 3'-Fehlpaarung der hybridisierten Primer an ihrem jeweiligen 3'-Ende entsprechend der coC3-Gensequenz verändert und anschließend verlängert wurden.

Mit Ausnahme der Primersequenzen zeigt die genomische Sequenz im Exonbereich vollständige Übereinstimmung mit der korrespondierenden coC3-cDNA-Sequenz. Intron 2 des coC3-Gens ist kleiner als das entsprechende Intron des CVF2-Gens und vermutlich auch des CVF1-Gens, und wurde aufgrund seiner relativ geringen Größe bevorzugt amplifiziert.

Im Intron 2 befindet sich ca. 90 bp stromabwärts von Exon 2 eine alternierende Region aus ct und gt (Anhang 7.3). Solche Sequenzen aus alternierenden Purinen und Pyrimidinen bilden *in vitro* bevorzugt eine Z-DNA aus (Rich *et al.* 1984).

Introns 3, 16, 29

Für weitere Größenvergleiche zu Introns des CVF1-Gens wurden mit coC3-spezifischen Primern Intron 3 und Intron 29 des coC3-Gens amplifiziert, deren CVF1-Homologe bereits oben charakterisiert sind.

Um weitere Charakteristika des coC3-Gens aufzuklären, wurde mit den coC3-spezifischen Primern YL3s-C3/YL4as-C3 und YL29s-C3/YL30as-C3 entsprechend der ARMS-Methode die gesuchten Genabschnitte amplifiziert (Abb. 3.67) und aufgrund ihrer Übereinstimmung mit der coC3-cDNA im Exonbereich dem coC3-Gen zugeordnet (siehe Anhang 7.3).

Mit dem relativ unspezifischen Primerpaar OPS72/OPA24 (vgl. Abb. 3.4, Kapitel 3.1) wurde zusätzlich zu Intron 16 des CVF1-Gens auch der korrespondierende coC3-Genabschnitt amplifiziert (Abb. 3.18 und 3.67). Aufgrund eindeutiger Sequenzmerkmale war eine Zuordnung zum coC3-Gen möglich (Sequenz siehe Anhang 7.3).

Intron 32

Mit dem nachfolgenden Experiment sollte untersucht werden, ob Exon 31 und Exon 32 des coC3-Gens ebenso wie im Falle von CVF1 und CVF2 kontinuierlich und nicht durch ein Intron unterbrochen sind. Eine genomische PCR mit dem Primerpaar HB31s-CVF2,coC3 und HB33as-coC3 führte zu einem Amplifikat von ca. 1700 bp Länge (Abb. 3.67), das aufgrund seiner Sequenz dem coC3-Gen zugeordnet wurde (Abb. 3.69).

| | | Exon 31 | | | | | | |
|--|------|----------------|------------|---------------|------------|------------|------------|------------|
| CVF1 | cDNA | 3916 | ACCAAAC | TCACTGTGAC | AGCATCAGGT | GATGGAAAAG | 3964 | |
| CVF2 | cDNA | 2964 | ACCAAAC | TCCTATAACG | AGCATCAGGT | GATGGAAAAG | 3012 | |
| coC3 | cDNA | 3939 | ACCAAAC | TCCTATAACG | AGCATCAGGT | GATGGAAAAG | 3987 | |
| coC3 | Gen | | gACCAAAC | TCCTATAACG | AGCATCAGGT | GATGGAAAAG | | |
| <div style="display: flex; justify-content: center; align-items: center; gap: 10px;"> <div style="border-top: 1px solid black; width: 100%;"></div> <div style="font-size: 12px;">↓</div> </div> | | | | | | | | |
| | | Exon 32 | | | | | | |
| CVF1 | cDNA | 3965 | CAACAATGAC | TTCTATAACG | CACAGTTGCA | GGAGAAGGCA | 4014 | |
| CVF2 | cDNA | 3013 | CAACAATGAC | GTCTATAATG | CACAATTGAG | GGAGGATGCA | 3062 | |
| coC3 | cDNA | 3988 | CAACAATGAC | GTCTATAATG | CACAATTGAG | GGAGGATGCA | 4037 | |
| coC3 | Gen | | CAACAATGAC | GTCTATAATG | CACAATTGAG | GGAGGATGCA | | |
| CVF1 | cDNA | 4015 | AATGTTTGCA | TCTTAATGTT | TCTGTTGAAA | ACATCCACTT | 4064 | |
| CVF2 | cDNA | 3063 | AATGTTTGCA | TCTTATGATGTT | TCTGTTGAAA | ACATCCACTT | 3112 | |
| coC3 | cDNA | 4038 | AATGTTTGCA | TCTTATGATGTT | TCTGTTGAAA | ACATCCACTT | 4087 | |
| coC3 | Gen | | AATGTTTGCA | TCTTATGATGTT | TCTGTTGAAA | ACATCCACTT | | |
| CVF1 | cDNA | 4065 | GA | | | | 4066 | |
| CVF2 | cDNA | 3113 | GA | | | | 3114 | |
| coC3 | cDNA | 4088 | GA | | | | 4089 | |
| coC3 | Gen | | GA | gtaagttt | ggcttcctag | aaatagtagg | ttgtaatgaa | gtttctgagg |
| coC3 | Gen | | agagatatta | aaattttagt | gttcaccatg | tcagaatctc | aatatggaac | |
| coC3 | Gen | | aagttatctc | | | | | |
| coC3 | Gen | | | (ca. 1500 bp) | | | | |
| coC3 | Gen | | | | | aggttagcca | | |
| coC3 | Gen | | ccgctgggct | agttgggtcaa | tcaaactttt | catcttaaaa | aggctgctaa | |
| CVF1 | cDNA | | | | | --- | | |
| CVF2 | cDNA | 3115 | | | | ACT | 3117 | |
| coC3 | cDNA | 4090 | | | | ACT | 4092 | |
| coC3 | Gen | | gaat | ttgtgc | agagtctgat | tttccccctt | tcttttgtaa | tttttagACT |
| CVF1 | cDNA | 4067 | -----ATGC | AATGGGAGCC | AAGGGAGCCC | TCATGCTCAA | GATCTGCACA | 4110 |
| CVF2 | cDNA | 3118 | TAAAAGAGGC | AAAGGGAGCC | AAGGGAGCCC | TCAAGCTCAA | AATCTGCACT | 3167 |
| coC3 | cDNA | 4093 | TAAAAGAGGC | AAAGGGAGCC | AAGGGAGCCC | TCAGGCTTAA | AATCTGCACT | 4142 |
| coC3 | Gen | | TAAAAGAGGC | AAAGGGAGCC | AAGGGAGCCC | TCAGGCTTAA | AATCTGCACT | |

Abb. 3.69: Sequenz des genomischen PCR-Produkts mit Intron 32 des coC3-Gens. Intron 32 hat eine Größe von ca. 1700 bp. Exon 31 und Exon 32 sind im coC3-Gen nicht durch ein Intron unterbrochen. Die genomische coC3-Sequenz weicht in Exon 32 von der cDNA-Sequenz (Position 4026, G-A) ab (unterstrichen). Exon 33 des CVF1-Gens ist um 9 Nukleotide kürzer (gestrichelt dargestellt) als die entsprechenden Exons von CVF2 und coC3.

Die in Abbildung 3.69 gezeigte genomische Sequenz stimmt bis auf eine Abweichung in Exon 32 mit der coC3-cDNA überein. Da Polymerasefehler nicht auszuschliessen sind, wird vermutet, dass es sich bei der gezeigten Sequenz um den coC3-Genabschnitt handelt. Die genomische Sequenz zeigt, dass Exon 31 und Exon 32 auch im coC3-Gen nicht durch ein Intron unterbrochen sind.

3.3.2 Aufklärung der coC3-Genstruktur durch Genome Walking

Die 5'-Region des coC3-Gens, welche für die Identifizierung regulatorischer Elemente und ein Vergleich mit den Promotorregionen von CVF1 und CVF2 von Interesse war, wurde durch Genome Walking aufgeklärt.

Tab. 3.10: Übersicht der Genome Walker-Amplifikate des coC3-Gens. Das jeweils größte sequenzierte PCR-Produkt sowie die dazu eingesetzte Genome Walking (GW)-Bibliothek sind angegeben. Größenangaben sind Schätzungen.

| Exon-Intron-Grenze | Primerpaare | Größe PCR-Produkt (GW-Bibliothek) | Exon-Intron- Übergangssequenz |
|--------------------------|---|--------------------------------------|----------------------------------|
| Promotor/Exon 1/Intron 1 | CoC3-Intr.1as-I/AP1 CoC3-Intr.1as-II/AP2 | 1000 bp (StuI) | TCT gtaagtatca... |
| Exon 1/Intron 1 | HB1s-II/AP1 HB1s/AP2 | 1880 bp (StuI) | TCT gtaagtatca... |

Promotorregion

Für die Aufklärung der Promotorregion des coC3-Gens wurde auf Basis der Intron 1-Sequenz (siehe weiter unten) verschachtelte PCRs mit den vier adapter-ligierten Genome Walking-Bibliotheken durchgeführt. Als genspezifische Primer wurden CoC3-Intr.1as-I (außen) und CoC3-Intr.1as-II (innen) eingesetzt. Mit der DraI- und der StuI-Bibliothek wurden Doppelbanden und mit der PvuII-Bibliothek wurde eine singuläre Bande erhalten (siehe Kapitel 3.2.2, Abb. 3.58, CVF2-Promotor). Das ca. 1000 bp lange Amplifikat der StuI-Bibliothek enthielt die nachfolgend gezeigte Promotorregion des coC3-Gens. Die entsprechenden DraI- und PvuII-Schnittstellen sind in der Sequenz markiert.

Für eine vollständige Sequenzierung des Promotorbereichs wurde zusätzlich der interne sense-Primers HB-coC3-Prom-s eingesetzt. Putative Bindungsstellen für Transkriptionsfaktoren sind markiert. Eine mögliche Rolle dieser Transkriptionsfaktoren ist in Kapitel 4 erläutert.

Exon 1/Intron 1

Bei der verschachtelt durchgeführten PCR mit den Primerpaaren HB1s-II/AP1 und HB1s/AP2 wurden sowohl Amplifikate des CVF1-Gens als auch des coC3-Gens erhalten (Abb. 3.35). Das schwerste Amplifikat der DraI-Bibliothek (1200 bp) sowie das Amplifikat der StuI-Bibliothek (1880 bp) wurden sequenziert und als coC3-Genabschnitte identifiziert. Die Konsensus-Sequenz der beiden coC3-Amplifikate ist in Abbildung 3.71 dargestellt.

3 Ergebnisse

| | | | | | | |
|------------------|------|----------------------|--------------------|---------------------|--------------------|-------------------------|
| coC3 Gen | -818 | <u>AG</u> GCCTTTCCTG | TCATTTTGT | GTCTCTGACC | AAT CTTCAAA | -777 |
| coC3 Gen | -776 | AATAGTAAGC | TTGTTTGTCT | GAAAAAATAC | TTTCATCCTA | ATCATTCTGA -727 |
| | | | | | C CCAAT | |
| coC3 Gen | -726 | TTTGGTTTAT | AGTTATGCAT | ATGTTCCCAT | AAAGTGCTCA | CCAAT GCATT -677 |
| | | | | | TBP | |
| coC3 Gen | -676 | CT TATAA AGT | TATACTGGTT | TTAATTGTTC | TATACTTAAG | TTTAAATGCCT -627 |
| coC3 Gen | -626 | GTAAATATTG | TAGATAGGTG | TGCAAGGATA | GGGAGGAGAG | AAATTTGAAA -577 |
| | | | | | IRS | |
| coC3 Gen | -576 | TGGGACCCTG | CAGTAGCTTC | AAGGGAAGTA | CTCGGG GAAA | AGAAA ATTAA -527 |
| coC3 Gen | -526 | GAATACTGAG | AGCTACTATA | TCTGGTTGTA | GACATTAAAG | TAACTATAAC -477 |
| | | | | | CCAAT | |
| coC3 Gen | -476 | ATAACTGTCC | CATATC CCA | AT CCTCAAAG | <u>CAGCTGCCTC</u> | CAAATAAGTG -427 |
| | | | | | PEA3 (-) | |
| coC3 Gen | -426 | GCA TTTCCTG | TTCAAAAAA | CTG AGGAAAC | AACAATGCCT | TAATAGCCTT -377 |
| | | | | | PEA-3 | |
| coC3 Gen | -376 | ATTGTGATCT | GTTTAGGAGC | GGATTTGGGT | TACTTTAAAG | CCCTTTCAAA -327 |
| | | | | | MAZ | |
| coC3 Gen | -326 | TCTGTGTGAG | GGAGGGAAGG | AAA GTGAGTT | TGTGAGTTCC | TGATCTCAAG -277 |
| | | | | | PEA-3 | |
| coC3 Gen | -276 | ATAA AGGCTA | TTTGCAGGTG | TAATAGCAGA | ATATCT AGAA | CAA ATTGCAC -227 |
| | | | | | GATA-1 | |
| coC3 Gen | -226 | AAAAATGCTG | TTTAGCAATC | ATGTTTAACT | AAGAGAGAAA | GAAGACAAAT -177 |
| | | | | | NF-kB | |
| coC3 Gen | -176 | AATCCACTGT | GGTTTTTAAG | GGCCTCATTA | ATTTCCCTCTG | AAATTCCCTG -127 |
| | | | | | HNF-5 | |
| coC3 Gen | -126 | TTTGCT CAAG | CTCCGCCTCA | CC TCCCCAC | CCT CCTTTAC | ACTGATTAAA -77 |
| | | | | | Sp1 | |
| coC3 Gen | -76 | CATTCATCAG | TTGCAAAGGG | GCTTGGCTTC | AAAGCAGGCT | CTCTGG ATAA -27 |
| | | | | | AP-2/PuF | |
| | | | | | TATA | |
| | | | | | Exon 1 | |
| CVF1 cDNA | 1 | | | | C | CCATGGAGAG 11 |
| CVF2 Gen | | | | | GGACTA | CCATGGAGCG |
| coC3 cDNA | 1 | | | TIS (+1) | GGACTA | CCATGGAGGG 16 |
| coC3 Gen | -26 | ATA CTGGAGT | TGTCAGCCCT | TCTCTT GCCC | CTCTGGACTA | CCATGGAGGG 24 |
| CVF1 cDNA | 12 | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGGGTCTT 61 |
| CVF2 Gen | | GACGGCTCTC | TCTCTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGGATCTT |
| coC3 cDNA | 17 | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGGGTCTT 66 |
| coC3 Gen | 25 | GATGGCTCTC | TATCTGGTGG | CTGCTCTATT | GATTGGTTTT | CCAGGGTCTT 74 |
| CVF1 cDNA | 62 | CTCATGGGGC | TCT | | | 74 |
| CVF2 Gen | | CTCATGGGAC | TCT | | | |
| coC3 cDNA | 67 | CCCACGGGGC | TCT | | | 79 |
| coC3 Gen | 75 | CCCACGGGGC | TCT gtaagta | tca g agatat | ttcagcaggg | ggaataaagc 124 |
| coC3 Gen | 125 | agtgctgata | aacccttgctg | gtcctttgca | gggtaatgaa | tttccttcca 174 |

Abb. 3.70: Promotorregion des coC3-Gens. Die Region stromaufwärts von Exon 1 ist in Großbuchstaben dargestellt. Zu Vergleichszwecken wurde die genomische Sequenz von CVF2-Exon 1 mit aufgeführt. Putative Bindungsstellen für Transkriptionsfaktoren sowie die proximale TATA-Box und der putative Transkriptions-Initiations-Start (TIS) sind fett und kursiv dargestellt. Die StuI-Schnittstelle (aggcct) am 5'-Ende sowie zwei DraI-Schnittstelle (tttaaa) und eine PvuII-Schnittstelle (cagctg) sind unterstrichen.

Das Amplifikat der StuI-Bibliothek (1880 bp) wurde direkt sequenziert und ergab auf der Seite des sense-Primers HB1s Übereinstimmung mit der coC3-cDNA.

Das ca. 230 bp schwere Amplifikat der EcoRV-Bibliothek und das ca. 900 bp-Amplifikat der PvuII-Bibliothek (Abb. 3.35) stimmen in ihrer Größe mit den Positionen entsprechender Schnittstellen in der coC3-Gensequenz überein (Anhang 7.3) und wurden demzufolge als coC3-Amplifikate identifiziert.

| | | | | |
|-------------|-------------|----|---|----|
| CVF1 | cDNA | 4 | <u>ATGGAGAGGA TGGCTCTCTA TCTGGTGGCT GCTCTATTGA TTGGTTTTCC</u> | 53 |
| CVF2 | Gen | | <u>ATGGAGGGGA CGGCTCTCTC TCTGGTGGCT GCTCTATTGA TTGGTTTTCC</u> | |
| coC3 | cDNA | 9 | <u>ATGGAGGGGA TGGCTCTCTA TCTGGTGGCT GCTCTATTGA TTGGTTTTCC</u> | 58 |
| coC3 | Gen | | <u>GGTGGCT GCTCTATTGA TTGGTTTTCC</u> | |
| CVF1 | cDNA | 54 | <u>AGGGTCTTCT CATGGGGCTC T</u> | 74 |
| CVF2 | Gen | | <u>AGGATCTTCT CATGGGACTC T</u> | |
| coC3 | cDNA | 59 | <u>AGGGTCTTCC CACGGGGCTC T</u> | 79 |
| coC3 | Gen | | <u>AGGGTCTTCC CACGGGGCTC Tgtaagtatc agagatattt cagcaggggg</u> | |
| coC3 | Gen | | aataaagcag tgctgataaa cccttggtgt cctttgcagg gtaatgaatt | |
| coC3 | Gen | | tccttcagc tgtctctaaa cactttgagc tttctcagaa | |

Abb. 3.71: Darstellung der partiellen Intron 1-Sequenz des coC3-Gens.

3.4 Hinweise auf die Existenz weiterer Gene der CVF/C3-Familie

Im Zuge der Aufklärung der CVF1-Genstruktur durch genomische PCR und Genome Walking wurden einige Sequenzdaten erhalten, die Abweichungen von den cDNA-Sequenzen der drei bekannten Gene CVF1, CVF2 und coC3 von meist mehreren Nukleotiden zeigten und daher als Hinweis auf zusätzliche Mitglieder der CVF/C3-Genfamilie gewertet wurden. Die neuen Gensequenzen wurden als "CVF-Homologe" (CVF-H) bezeichnet und beginnend mit der am weitesten stromaufwärts gelegenen Sequenz durchnummeriert (CVF-H1 bis CVF-H5).

Genomische PCR

In diesem Abschnitt sind die durch genomische PCR erhaltenen Sequenzen dargestellt, die in den flankierenden Exonsequenzen mit keinem der drei korrespondierenden cDNA-Sequenzen von CVF1, CVF2 und coC3 zu 100% identisch sind und infolgedessen auf die Existenz mindestens eines neuen homologen Gens der CVF/C3-Genfamilie schliessen lassen.

Tab. 3.11: Übersicht der genomischen PCR-Produkte putativer neuer homologer Gene. Größenangaben in Klammern sind Schätzungen.

| Intron | Primerpaar | Größe PCR-Produkt (bp) | Exon-Intron-Übergangssequenz |
|--------|---|------------------------|--|
| 12 | YL12s / YL13as | 456 | CTC gtaagtatta...tcaccctcag ATA |
| 27 | HB27s-CVF / HB28as-CVF | (1300) | AAT gtatggtgat...ttacttctag ATT |
| 36 | HB-CVF1-Intr.35s/HB37as-CVF HB36s-CVF/HB37as-CVF | (2200) (2000) | TAG gtaatggcag...ctttcattag ATG TAG gtaatggcag...ctttcattag ATG |

Die erhaltenen PCR-Amplifikate sind in Abbildung 3.72 dargestellt.

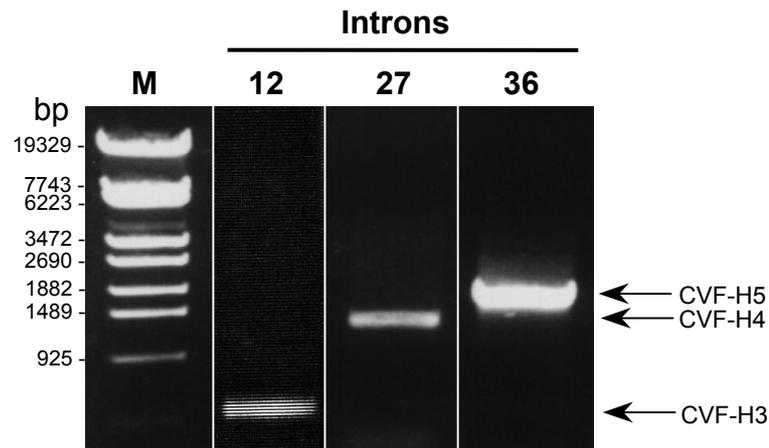


Abb. 3.72: Gelelektrophoretische Analyse genomischer PCR-Amplifikate putativer neuer homologer Gene der CVF/C3-Familie (1%iges Agarosegel). M: Marker λ -DNA/Eco130I, Intron 12 (456 bp), Intron 27 (1300 bp), Intron 36 (2000 bp). Das Intron 36-Amplifikat wurde durch Verwendung der Primer HB36s-CVF/HB37as-CVF erzeugt.

Intron 12

Ursprünglich sollte mit dem CVF1-spezifischen Primerpaar YL12s / YL13as mittels genomischer PCR das gesuchte CVF1-Intron 12 amplifiziert werden. Das erhaltene Amplifikat mit einer Länge von 456 bp wurde kloniert und sequenziert, wobei zusätzlich zu CVF1-Klonen (Kapitel 3.1) ein Klon identifiziert wurde, dessen Sequenz sich in zwei charakteristischen Nukleotiden von der CVF1-Sequenz unterscheidet (Abb. 3.73). Die Gensequenz zeigt stromabwärts der unveränderten Primersequenz in Exon 12 100% Übereinstimmung mit CVF2, wohingegen im Bereich von Exon 13 stromaufwärts der unveränderten Primersequenz 100% Sequenzidentität mit CVF1 bzw. coC3 zu finden ist. Da das Amplifikat unter Verwendung von Proofreading-Polymerase erhalten wurde, ist diese charakteristische Sequenz mit mindestens zwei Abweichungen von den bisher bekannten cDNA-Sequenzen höchstwahrscheinlich nicht auf Polymerase-Fehler zurückzuführen, sondern scheint vielmehr auf die Existenz eines neuen homologen Gens hinzuweisen.

3 Ergebnisse

| | | | |
|------------------|------|---|-------------------------------|
| CVF1 cDNA | 1279 | GCAACAAAGT CCATGACAGC CATAGCCTAC CAAACCCAGG GAGGATCTGG | 1328 |
| CVF2 cDNA | 315 | GCAAT^AAAAGT CCATGACAGC CA^CAGCCTAC CAAACCCAGG GAGGATCTGG | 364 |
| coC3 cDNA | 1290 | GCAAT^AAAAGT CCATGACAGC CA^CAGCCTAC CAAACCCAGG GAGG^CCTG^A | 1339 |
| CVF-H3 | | GCAACAAAGT CCATGACAGC CATAGCCTAC CAAACCCAGG GAGGATCTGG | |
| CVF1 cDNA | 1329 | AAACTATCTT CATGTAGCCA TTACATCTAC AGAGATTAAG CCCGGAGATA | 1378 |
| CVF2 cDNA | 365 | AAACTATCTT CATGTAGCCA TTACATCTAC AGAGATTAAG CCCGGAGATA | 414 |
| coC3 cDNA | 1340 | AAACTATCTT CATGTAGCCA TTACATCTAC AGAGATTAAG CCCGGAGATA | 1389 |
| CVF-H3 | | AAACTATCTT CATGTAGCCA TTACATCTAC AGAGATTAAG CCCGGAGATA | |
| CVF1 cDNA | 1379 | ACTTACCTGT CAATTTCAAT GTGAAGGGCA ATGCAAATTC ACTGAAGCAG | 1428 |
| CVF2 cDNA | 415 | ACTTACCTGT CAATTTCAAT GTGA^CGGGCA ATGCAAATTC ACTGAA^CCAG | 464 |
| coC3 cDNA | 1390 | ACTTACCTGT CAATTTCAAT GTGA^CGGGCA ATGCAAATTC ACTGAA^CCAG | 1439 |
| CVF-H3 | | ACTTACCTGT CAATTTCAAT GTGA^CGGGCA ATGCAAATTC ACTGAA^CCAG | |
| CVF1 cDNA | 1429 | ATCAAATATT TCACATACCT C | 1449 |
| CVF2 cDNA | 465 | ATCAAATATT TCACATACCT C | 485 |
| coC3 cDNA | 1440 | ATCAAATATT TCACATACCT C | 1460 |
| CVF-H3 | | ATCAAATATT TCACATACCT Cgtaagtatt atagatctgg aggaaatcca | |
| CVF-H3 | | ggggctgcat ttcaattcat attgtagcct tgaagccact ggcttaatct | |
| CVF1 cDNA | 1450 | | ATATTG 1455 |
| CVF2 cDNA | 486 | | ATA^CCTG 491 |
| coC3 cDNA | 1461 | | ATATTG 1466 |
| CVF-H3 | | atcatttcta accttcccat taatcagtat ttgttcacc tcag ATATTG | |
| CVF1 cDNA | 1456 | AATAAAGGGA AGATTTTCAA GGTGGCAGG CAACCCAGGA GAGATGGGCA | 1505 |
| CVF2 cDNA | 492 | AATAAAGGGA AGATTTTCAA GGTGGCAGG CAAC^ACAGG^G GAGATGGG^CA | 541 |
| coC3 cDNA | 1467 | AATAAAGGGA AGATTTTCAA GGTGGCAGG CAACCCAGGA GAGATGGGCA | 1516 |
| CVF-H3 | | AATAAAGGGA AGATTTTCAA GGTGGCAGG CAACCCAGGA GAGATGGGCA | |
| CVF1 cDNA | 1506 | GAATCTGGTG ACCATGAATC TGCATATCAC TCCAGATCTC ATCCCTTCCT | 1555 |
| CVF2 cDNA | 542 | GAATCTGGTG ACCATGAATC TA^ACATATCAC TCCAGATCTC AT^TCCCTTCCT | 591 |
| coC3 cDNA | 1517 | GAATCTGGTG ACCATGAATC TGCATATCAC TCCAGATCTC ATCCCTTCCT | 1566 |
| CVF-H3 | | GAATCTGGTG ACCATGAATC TGCATATCAC TCCAGATCTC ATCCCTTCCT | |
| CVF1 cDNA | 1556 | TCCGGTTTGT GGCTTACTAC CAAGTGGGAA ACAACGAAAT TGTGGCTGAT | 1605 |
| CVF2 cDNA | 592 | TCCGGTTTGT GGCTTACTAC CAAGTGGGAA ACA^ATGAAAT TGTGGCTGAT | 641 |
| coC3 cDNA | 1567 | TCCGGTTTGT GGCTTACTAC CAAGTGGGAA A^TTAACGAAAT TGTGGCTGAT | 1616 |
| CVF-H3 | | TCCGGTTTGT GGCTTACTAC CAAGTGGGAA ACAACGAAAT TGTGGCTGAT | |
| CVF1 cDNA | 1606 | TCTGTC | 1611 |
| CVF2 cDNA | 642 | TCTGTC | 647 |
| coC3 cDNA | 1617 | TCTGTC | 1622 |
| CVF-H3 | | <u>TCTGTC</u> | |

Abb. 3.73: Gensequenz mit Intron 12 eines putativ neuen Mitglieds der CVF/C3-Genfamilie. Intron 12 hat eine Größe von 123 bp. Abweichungen von der CVF1-Sequenz sind markiert.

Die oben gezeigte Sequenz stimmt außer den beiden markierten Nucleotiden in allen Merkmalen mit der korrespondierenden CVF1-Sequenz (Abb. 3.20, Anhang 7.1) überein, auch in bezug auf die Intronsequenz. In dem Sequenzausschnitt in Abbildung 3.74 sind die beiden von der CVF1-Sequenz abweichenden Nucleotide in Exon 12 deutlich zu erkennen. Die Besonderheit dieser Sequenzfolge liegt darin, dass sie mit den korrespondierenden cDNA-Sequenzen von CVF2 und coC3 identisch ist. Daher scheiden zufällige Polymerasefehler, die sich während der PCR-Amplifikation ereignet haben könnten, als Ursache dieser Gensequenz aus. Vielmehr scheint es sich um die Sequenz eines neuen homologen Gens der CVF/C3-Genfamilie zu handeln, welche die Charakteristika einer partiellen Hybridsequenz aus CVF1 und CVF2 (bzw. CVF1 und coC3) trägt. Die hier gezeigte partielle Gensequenz eines putativ neuen homologen Gens der CVF/C3-Genfamilie wurde mit "CVF-H3" bezeichnet.

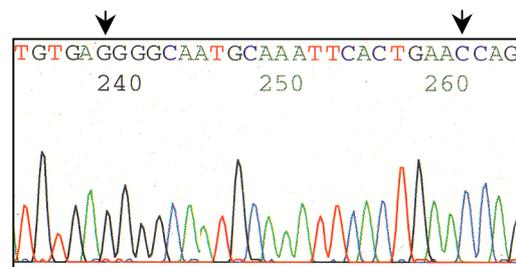


Abb. 3.74: Sequenzierung des genomischen Amplifikats ("CVF-H3") mit Intron 12. Der Ausschnitt zeigt deutlich die beiden charakteristischen Merkmale in Exon 12 (Pfeile) innerhalb der sonst CVF1-ähnlichen Sequenz, die auf ein neues Gen der CVF/C3-Familie hinweisen.

Intron 27

Bei dem Versuch, Intron 27 des CVF1-Gens unter Verwendung der CVF1-spezifischen Primer HB27s-CVF und HB28as-CVF mittels genomischer PCR zu amplifizieren, wurde anstelle des gesuchten Genabschnitts ein Amplifikat von 1,3 kb Länge erhalten (Abb. 3.72). Die Sequenz des durch Verwendung von Proofreading-Polymerasen erhaltenen 1300 bp-Amplifikats ist nachfolgend gezeigt.

| | | | | | | | |
|-----------|------|------------------------------|---------------------|---------------------|-----------------------------|----------------------|------|
| CVF1 cDNA | 3346 | GGAGGAATTC | AAGGTGCTGA | AGAAGAAGTA | TATTTAACAG | CTTTCATTCT | 3395 |
| CVF2 cDNA | 2394 | GGAGG C AATTC | AAGGTGCTGA | ACCAGAAG C A | TCTTTAACAG | CTTTCATTCT | 2443 |
| coC3 cDNA | 3369 | GGAGGAA C TA | AAGGTGCTGA | ACCAGAAG C A | TCTTTAACAG | C A TTTCATTCT | 3418 |
| CVF-H4 | | GGAGGAATTC | AAGGTGCTGA | AGAAGAAGTA | TCTTTAACAG | CTTTCATTCT | |
| | | | | ↓ | | | |
| CVF1 cDNA | 3396 | GTTTGC G TTG | TTGGAATCCA | AAACAATCTG | CAATGACTAT | GTC A AT | 3441 |
| CVF2 cDNA | 2444 | GTTTGC G TTG | TTGGAATCCA | GATCAATCTG | CAATG C ATAT | AT C AAT | 2489 |
| coC3 cDNA | 3419 | G A CTGC A TTA | TTGGAATCCA | GATCA C TCTG | CAA A GA C AA | AT C AAT | 3464 |
| CVF-H4 | | GTTTGC G TTG | TTGGAATCCA | GATCAATCTG | CAATG C ATAT | AT C AATgtat | |
| CVF-H4 | | gttgatggtg | tgatgggaag | tggtacgaga | gggccaaaat | ataatctcat | |
| CVF-H4 | | ccattggggg | | | | | |
| | | | | | (ca. 900 bp) | | |
| CVF-H4 | | | | | | atgctgttcc | |
| CVF-H4 | | atagagatgc | ctaaaacatt | cctaattgat | ctcttctctt | tcctctgcct | |
| CVF1 cDNA | 3442 | | AGTCTAGACA | GCAGCATCAA | GAAGGCCACA | AATTATTTAC | 3481 |
| CVF2 cDNA | 2490 | | AT T CTAGACA | GCAGCAT C A | TAAGGCCACA | GATTATTTAC | 2529 |
| coC3 cDNA | 3465 | | AT T CTAGACA | GCAGCATCAA | TAAGGCCACA | GATTATTTAC | 3504 |
| CVF-H4 | | ttacttctag | AT T CTAGACA | GCAGCAT C A | TAAGGCCACA | GATTATTTAC | |
| CVF1 cDNA | 3482 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC | 3531 |
| CVF2 cDNA | 2530 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC | 2579 |
| coC3 cDNA | 3505 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC | 3554 |
| CVF-H4 | | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC | |
| CVF1 cDNA | 3532 | TATGCTTTGG | CTGCTGCAGA | CCAACTCAAT | GATGACAGGG | TACTCATGGC | 3581 |
| CVF2 cDNA | 2580 | TATGCTTTGG | CTGCTGCAG C | ACC A CTCAAT | GATGACAGGG | TACTCATGGC | 2629 |
| coC3 cDNA | 3555 | TATGCTTTGG | CTGCTGCAGA | CC G ACTCAAT | GATGACAGGG | TACTCATGGC | 3604 |
| CVF-H4 | | TATGCTTTGG | CTGCTGCAG A | ACC A CTCAAT | GATGACAGGG | TACTCATGGC | |

Abb. 3.75: Gensequenz mit Intron 27 eines putativ neuen Mitglieds der CVF/C3-Familie. Intron 27 hat eine Größe von ca. 1000 bp. Trotz weitgehender Übereinstimmung mit der CVF2-Sequenz weicht die gefundene Gensequenz in drei Positionen (umrahmt) von dieser ab. Eine EcoRI-Schnittstelle in Exon 27 und eine StuI-Schnittstelle in Exon 28 der drei Gene sind unterstrichen.

Beide CVF1-spezifischen Primer enthalten eine 3'-Fehlpaarung zu den korrespondierenden Sequenzen von CVF2 und coC3 und sollten daher nur im CVF1-Gen verlängert werden. Da dies mit Taq-Polymerase nicht zu einem Amplifikat führte, wurde eine PCR mit der Proofreading-Polymerase *Expand* durchgeführt. Die Exon-Sequenzen des erhaltenen Amplifikats haben zwar mit der CVF2-cDNA die größte Ähnlichkeit, doch weisen sie an drei Positionen Unterschiede zur bekannten CVF2-cDNA-Sequenz auf: sowohl die 3'-Nukleotide beider Primer als auch eine weitere Position in Exon 28 stromaufwärts des Primers HB28as-CVF (G-A an Position 2599 der CVF2-cDNA, umrahmt) stimmen nicht

mit der CVF2-cDNA-Sequenz überein. Diese drei Abweichungen wurden in zwei unabhängig voneinander amplifizierten und klonierten PCR-Produkten festgestellt. Da zur Amplifikation in beiden Fällen Proofreading-Polymerasen verwendet wurden, ist zu erwarten, dass bei der Amplifikation des entsprechenden CVF2-Genabschnitts die Primer-Enden im Sinne der CVF2-Sequenz verändert worden wären (vgl. Intron 2 des coC3-Gens, Kapitel 3.3.1). Zudem wäre eine Hybridisierung des CVF1-spezifischen Primers HB27s-CVF mit dem coC3- oder CVF2-Gen unter den gewählten PCR-Bedingungen insofern bemerkenswert, da der Primer vier (CVF2) bzw. fünf (coC3) Fehlpaarungen zu diesen Genen aufweist. Es muß aufgrund der Divergenz zwischen den 3'-Enden der Primer und der CVF2-cDNA sowie einem zusätzlichen Unterschied vermutet werden, dass es sich bei dem vorliegenden Amplifikat um den Genabschnitt eines neuen Mitglieds der CVF/C3-Genfamilie handelt.

Intron 36

Mit der Zielsetzung, Intron 36 des CVF1-Gens zu amplifizieren, wurde mit den CVF1-spezifischen Primern HB36s-CVF und HB37as-CVF unter Verwendung einer Proofreading-Polymerase eine PCR durchgeführt, die zu einem Amplifikat von 2,0 kb Länge führte. Die Sequenz dieses Amplifikats zeichnet sich zwar durch eine weitgehende Ähnlichkeit mit den Exonsequenzen des CVF2-Gens aus, aber unterscheidet sich von dieser in der 3'-Fehlpaarung des Rückwärtsprimers HB37as-CVF (Abb. 3.76).

Daraufhin wurde in einer zusätzlichen PCR der CVF1-spezifische Primer HB-CVF1-Intr.35s als Vorwärtsprimer eingesetzt, der in Intron 35 des CVF1-Gens bindet. Es wurde davon ausgegangen, dass sich die Introns der CVF/C3-Genfamilie in einer Entfernung von 200 bp zum nächsten Exon soweit unterscheiden, dass unter den gewählten Bedingungen nur das gewünschte CVF1-Amplifikat entstehen sollte. Das auf diese Weise erhaltene Amplifikat der Größe 2,2 kb erwies sich erneut sehr homolog zur entsprechenden CVF2-cDNA-Sequenz mit dem bereits charakteristischen Unterschied in Exon 37. Die Konsensus-Sequenzen beider PCR-Produkte wurden als CVF-H5 bezeichnet und sind in Abbildung 3.76 zusammenfassend dargestellt.

| | | | | | | | |
|------------------|------|-------------------|--------------------|-------------------|-------------------|-------------------|------------|
| CVF-H5 | | ccttctgttc | ctctcacctt | ttccattgct | tccctttatt | ctgggaaacc | |
| CVF-H5 | | tggaacagatt | tgacagaat | agtccttcaa | ccactgggtc | aaacctgagg | |
| CVF-H5 | | aacttggaa | aagcagttt | cttgaggta | ttcggatgga | gtttcgggtg | |
| CVF-H5 | | ggggggaatt | atctccatgg | gtaaactggc | accacaaat | gatcttctct | |
| CVF1 cDNA | 4291 | | GTCTCCAC | TCTGAAGATG | AATGCCTGCA | CTTTAAGATT | 4329 |
| CVF2 cDNA | 3348 | | GTCTCCAC | TCTGAAGATG | AATGCCTGCA | CTTTAAGATT | 3386 |
| coC3 cDNA | 4323 | | GTCTCCAC | TCTGAAGATG | AATGCCTGCA | CTTTAAGATT | 4361 |
| CVF-H5 | | ctccatcca | g GTCTCCAC | TCTGAAGATG | AATGCCTGCA | CTTTAAGATT | |
| CVF1 cDNA | 4330 | CTCAAGCATT | TTGAAGTTGG | CTTCATTCAG | CCAGGATCAG | TCAAGGTGTA | 4379 |
| CVF2 cDNA | 3387 | CTCAAGCATT | TTGAAGTTGG | CTTCATTCAG | CCAGGATCAG | TCAAGGTGTA | 3436 |
| coC3 cDNA | 4362 | CTCAAGCATT | TTGAAGTTGG | CTTCATTCAG | CCAGGATCAG | TCAAGGTGTA | 4411 |
| CVF-H5 | | CTCAAGCATT | TTGAAGTTGG | CTTCATTCAG | CCAGGATCAG | TCAAGGTGTA | |
| CVF1 cDNA | 4380 | CAGCTACTAC | AATCTAG | | | | 4396 |
| CVF2 cDNA | 3437 | CAGCTACTAC | AATCTAG | | | | 3453 |
| coC3 cDNA | 4412 | CAGCTACTAC | AATCTAG | | | | 4428 |
| CVF-H5 | | CAGCTACTAC | AATCTAG | gta | atggcagacg | tactgaatat | acctccctgc |
| CVF-H5 | | tcctaatttc | cccacaacaa | ccactctgtg | aagtgagtca | agattgtgag | |
| CVF-H5 | | cgaatggctt | | | | | |
| CVF-H5 | | | | | | | |
| CVF-H5 | | | | | | | |
| CVF-H5 | | cctataatca | taataaattt | atcgatgtga | cctatggaca | taacttttaa | |
| CVF1 cDNA | 4397 | | ATGAAAAA | TGTACCAAGT | TCTACCATCC | AGATAAAGGA | 4434 |
| CVF2 cDNA | 3454 | | ATGAAAAA | TGTACCAAGA | TCTACCATCC | AGATCAAGCA | 3491 |
| coC3 cDNA | 4429 | | ATGAACAA | TGTACCAAGT | TCTACCATCC | AGATAAAGGA | 4466 |
| CVF-H5 | | ttctttcatt | ag ATGAAAAA | TGTACCAAGA | TCTACCATCC | AGATCAAGCA | |
| CVF1 cDNA | 4435 | ACAGGCCTTC | TCAATAAGAT | ATGTATTGGT | AACGTTTGCC | GATGTGCAGG | 4484 |
| CVF2 cDNA | 3492 | ACAGGCCTTC | TCAATAAGAT | ATGTCTTGGT | AACGTTTGCC | GATGTGCAGA | 3541 |
| coC3 cDNA | 4467 | ACAGGCCTTC | TCAATAAGAT | ATGTCTTGGT | AACGTTTGCC | GATGTGCAGA | 4516 |
| CVF-H5 | | ACAGGCCTTC | TCAATAAGAT | ATGTATTGGT | AACGTTTGCC | GATGTGCAGG | |

Abb. 3.76: Sequenz des genomischen PCR-Produkts (CVF-H5) mit Intron 36 eines putativ neuen Gens. Für die Amplifikation der hier gezeigten Sequenz waren die Primer HB-CVF1-Intr.35s/HB37as-CVF eingesetzt worden. Intron 36 hat eine Größe von ca. 1850 bp. Im ebenfalls dargestellten 3'-Ende von Intron 35 sind drei charakteristische Unterschiede zur CVF1-Gensequenz markiert. Zusätzlich ist die Sequenz des Primers HB36s-CVF als Pfeil markiert. Abweichungen der Primersequenz zur CVF2-cDNA in Exon 37 sind umrahmt.

Mit Ausnahme der Primersequenzen stimmen die Exonsequenzen zu 100% mit der CVF2-cDNA-Sequenz überein. Da jedoch das Amplifikat unter Verwendung einer Proofreading-Polymerase erhalten wurde, hätte die 3'-Fehlpaarung des CVF1-spezifischen Rückwärts-Primers HB37as-CVF im Sinne der CVF2-Sequenz verändert werden müssen. Da die vorliegende Sequenz CVF-H5 keine derartigen Veränderungen zeigt, wurde dies als Hinweis auf eine neue Gensequenz interpretiert.

Innerhalb von Intron 35 zeigt die Sequenz CVF-H5 die drei markierten Unterschiede zur CVF1-Sequenz. Da das benachbarte Exon 36 in allen drei Genen mit einer einzigen Abweichung im coC3-Gen identische Sequenzen hat, läßt dieses Resultat den Schluß zu, dass Intronsequenzen in unmittelbarer Exon-Nähe umso ähnlicher sind, je größer die Sequenzidentität im benachbarten Exon ist.

Genome Walking

Hinweise auf neue homologe Gene der CVF/C3-Familie wurden ebenfalls bei einigen Amplifikaten der Genome Walking-Bibliotheken erhalten. Bei der Aufklärung der Intron 1/Exon 2-Grenze wurde unter Verwendung der Primerpaare HB2as-CVF-II/AP1 und HB2as-CVF/AP2 mit der PvuII-Bibliothek ein Amplifikat von ca. 1500 bp Länge erhalten (Abb. 3.60, Kapitel 3.2), dessen Sequenz nachfolgend dargestellt ist. Aufgrund der vorhandenen Unterschiede sowohl zu den cDNA-Sequenzen der Gene CVF1 und coC3 als auch zur neu aufgeklärten CVF2-Gensequenz wurde diese homologe Gensequenz mit "CVF-H1" bezeichnet.

| | | | | | | | |
|-----------|-------|------------|------------|------------|-------------|------------|-----|
| CVF-H1 | | gttcttgtaa | ctatztat | cttcactttt | ccagaagaag | | |
| CVF-H1 | | agattatgga | ggtgcctcat | gaaatgctc | actttttcta | agcaaactaa | |
| CVF1 cDNA | 75 | | | CTACACCCTC | ATCACCCCTG | 94 | |
| coC3 cDNA | 80 | | | CTAACCCTC | ATCACCCCTG | 99 | |
| CVF-H1 | | ctcaggagat | gttcttttct | tcttctgcag | CTACACCCTC | ATCACCCCTG | |
| CVF1 cDNA | 95 | CTGTTTTGCG | AACAGACACA | GAAGAGCAAA | TTTGGTGGA | GGCCCATGGA | 144 |
| coC3 cDNA | 100 | CTGTTTTGCG | AACAGACACA | GAAGAGCAAA | TTTGGTGGA | GGCCCATGGA | 149 |
| CVF-H1 | | CTGTTTTGCG | AACAGACACA | GAAGAGCAAA | TCTGGTGGA | GGCCCATGGA | |
| CVF1 cDNA | 145 | GACAGTACTC | CAAAACAGCT | TGACATCTTT | GTTTCATGATT | TTCACGGAA | 194 |
| coC3 cDNA | 150 | GACAGTACTC | CAAAACAGCT | TGACATCTTT | GTTTCATGATT | TTCACGGAA | 199 |
| CVF-H1 | | GACAATACTC | CAAAACAGCT | TGACATCTTT | GTTTCATGATT | TTCACGGAA | |
| CVF1 cDNA | 195 | GCAGAAAACC | TTGTTCCAAA | CCAGAGTAGA | TATGAATCCA | GCAGGAGGCA | 244 |
| coC3 cDNA | 200 | GCAGAAAACC | TTGTTCCAAA | CCAGAGTAGA | TATGAATCAG | GCAGGAGGCA | 249 |
| CVF-H1 | | GCAGAAAACC | TTGTTCCAAA | CCAGAGTAGA | TATGAATCCA | GCAGGAGGCA | |

Abb. 3.77: Sequenz des Genome Walking Amplifikats (CVF-H1) mit Exon 2 eines putativ neuen Gens der CVF/C3-Familie. Zum Vergleich sind die cDNA-Sequenzen von CVF1 und coC3 gegenübergestellt. Der Rückwärtsprimer HB2as-CVF ist unterstrichen.

Genome Walking mit den Primerpaaren HB-CVF1-Intr.2as-I/AP1 (äußere PCR) und HB-CVF1-Intr.2as-II/AP2 (verschachtelte PCR) führte mit der DraI-Bibliothek zu einem ca. 1000 bp schweren Amplifikat, das mit den bis dato bekannten Sequenzen der CVF/C3-Genfamilie zwar Homologien aufweist, aber nicht übereinstimmt, und daher mit CVF-H2 bezeichnet wurde.

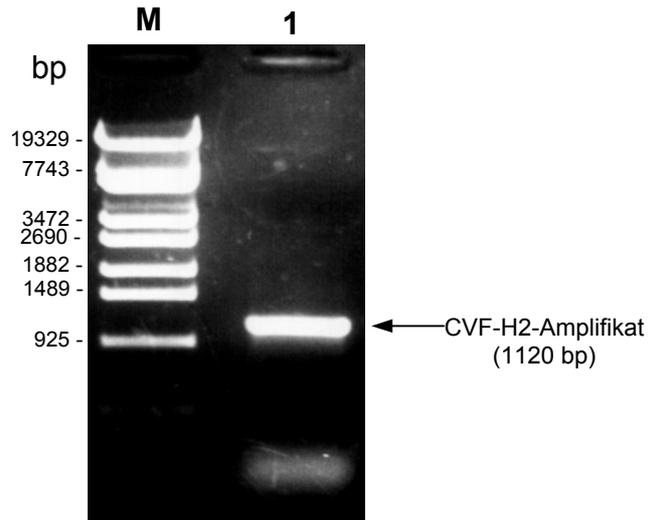


Abb. 3.78: Gelelektrophoretische Analyse des Genome Walking-Amplifikats (Intron 2 stromaufwärts). M: DNA-Marker λ /Eco130I, Bahn 1: Amplifikat der DraI-Bibliothek.

Das hier gezeigte singuläre Amplifikat der DraI-Bibliothek entspricht in seiner Größe dem sehr schwachen DraI-Amplifikat aus Abbildung 3.60 (Kapitel 3.2). Die partielle Sequenz des mit „CVF-H2“ bezeichneten Genabschnitts ist nachfolgend dargestellt.

| | | | | | | | |
|-----------|-----|------------|------------|------------|------------|------------|------------|
| CVF-H2 | | | cttttctaaa | agaagagatt | atggagatgc | ctcatgaaat | |
| CVF-H2 | | gcctcacttt | tcctaagcaa | actaactcag | gagatgttct | tttcttcttc | |
| CVF1 cDNA | 75 | CTACA | CCCTCATCAC | CCCTGCTGTT | TTGCGAACAG | ACACAGAAGA | 119 |
| coC3 cDNA | 80 | CTA | CCCTCATCAC | CCCTGCTGTT | TTGCGAACAG | ACACAGAAGA | 124 |
| CVF-H2 | | tg | cagCTACA | CCCTCATCAC | CCCTGCTGTT | TTGCGAACAG | ACACAGAAGA |
| CVF1 cDNA | 120 | GCAAATTTTG | GTGGAGGCC | ATGGAGACAG | TACTCCAAA | CAGCTTGACA | 169 |
| coC3 cDNA | 125 | GCAAATTTTG | GTGGAGGCC | ATGGAGACAG | TACTCCAAA | TCGCTTGACA | 174 |
| CVF-H2 | | GCAAATTTTG | GTGGAGGCC | ATGGAGACAG | TACTCCAAA | CAGCTTGACA | |
| CVF1 cDNA | 170 | TCTTTGTTCA | TGATTTTCCA | CGGAAGCAGA | AAACCTTGTT | CCAAACCAGA | 219 |
| coC3 cDNA | 175 | TCTTTGTTCA | TGATTTTCCA | CGGAAGCAGA | AAACCTTGTT | CCAAACCAGA | 224 |
| CVF-H2 | | TCTTTGTTC | TGATTTTCCA | CGGAAGCAGA | AAACCTTGTT | CCAAACCAGA | |
| CVF1 cDNA | 220 | GTAGATATGA | ATCCAGCAGG | AGGCATGCTT | GTCACTCCAA | CTATAGAG | 267 |
| coC3 cDNA | 225 | GTAGATATGA | ATCCAGCAGG | AGGCATGCTT | GTCACTCCAA | CTATAGAG | 272 |
| CVF-H2 | | GTAGATATGA | ATCCAGCAGG | GGGCATACTT | GTCACTCCAA | CTATAGAGgt | |
| CVF-H2 | | gggtgtatgg | tgatgggttc | aatagagtct | atggtatccc | tgcaaaacaa | |
| CVF-H2 | | aaaaccaagc | taaaatcctg | tccagtcg | cttgagggga | aaccctatgc | |
| CVF-H2 | | actttagtca | aacatgacta | aattgtcaga | tattgagaga | aactgtgcat | |

Abb. 3.79: Sequenz eines Genome Walking-Amplifikats (CVF-H2) mit Exon 2 eines putativ neuen Mitglieds der CVF/C3-Genfamilie. Zum Vergleich sind die cDNA-Sequenzen von CVF1 und coC3 gegenübergestellt. Die Sequenz des Primers HB-CVF1-Intr.2as-II ist als Pfeil markiert.

Das CVF1-Gen sowie die homologen Sequenzen CVF-H1 und CVF-H2 enthalten in Intron 1 Tandem-Wiederholungen von ac und ta, wobei die verkürzte ta-Wiederholung in CVF-H1 als Unterscheidungsmerkmal fungieren kann.

Die Kombination aus cDNA-Sequenzen, genomischen Sequenzen und Genome Walking-Resultaten für die Region um Exon 2 (Abb. 3.80) ergab deutliche Hinweise auf die Existenz von insgesamt fünf verschiedenen Gensequenzen. Die Zuordnung einer der Sequenzen zum CVF2-Gen muss noch als vorläufig gelten, bis eine Vergleichsmöglichkeit mit der entsprechenden cDNA gegeben ist. Grundsätzlich sind die darüber hinaus gefundenen zwei neuen Sequenzen, die mit CVF-H1 und CVF-H2 bezeichnet wurden, aufgrund der erheblichen Unterschiede als ein Hinweis auf die Existenz von zwei zusätzlichen Genen der CVF/C3-Multigenfamilie aufzufassen.

Der dargestellte Bereich weist für CVF1, CVF2 und coC3 eine ausgeprägte Heterogenität auf. Hingegen zeigen die Sequenzen CVF-H1 und CVF-H2 eine große Ähnlichkeit mit der CVF1-Sequenz.

4 Diskussion

4.1 Aspekte der angewandten Methoden

Im Rahmen dieser Arbeit sollte die Struktur des CVF1-Gens aufgeklärt werden. Auf der Grundlage eines Cosmid-Klons mit partiell enthaltenem CVF1 sollten weiterführend die Intron/Exon-Grenzen des gesamten Gens bestimmt und kartiert sowie regulatorische Bereiche analysiert werden. Um eindeutige Aussagen bezüglich der Identität von Sequenzinformationen treffen zu können, wurden unterschiedlichste molekularbiologische Techniken etabliert und eingesetzt.

Screening einer Cosmid-Bibliothek

Als Grundlage dieser Arbeit diente eine Bibliothek genomischer Kobra-DNA im Cosmidvektor sCos-1, die ca. $4-5 \cdot 10^5$ verschiedene Klone mit Insertgrößen von 30-45 kb umfasste. Durch ein in Vorarbeiten durchgeführtes Screening dieser Bibliothek nach CVF1-Klonen konnte ein singulärer Klon mit der Bezeichnung 1R67 identifiziert werden (Kunze, unveröffentlichte Ergebnisse). Als Sonde wurde hier P-32 markierte cDNA von Exon 24 (204 bp) eingesetzt. Zur Identifizierung weiterer positiver Klone wurde in dieser Arbeit anstelle der radioaktiven Markierung das Digoxigenin-System verwendet, welches über eine vergleichbare Sensitivität verfügt. Um CVF1-Klone zu identifizieren, deren genomischen Bereiche stromaufwärts bzw. stromabwärts der 1R67-Region (Exons 16-27) liegen, wurden zwei DIG-markierte Sonden, die Exon 12 und dem Intron 26/Exon 27-Bereich entsprechen, hergestellt. Zusätzlich konnten als Positiv-Kontrollen beim Screening sowohl 1R67-DNA als auch 1R67-Klone aufgetragen werden. Durch Verwendung geeigneter Positiv-Kontrollen wurde die notwendige Sensitivität der DIG-markierten Sonden und der Screening-Prozedur als solche bestätigt.

Trotz Durchmusterung von insgesamt $2,5 \cdot 10^5$ Klonen konnten keine positiven Klone erhalten werden. Die detaillierte Analyse eines vermeintlich positiven Klones M26 konnte dabei nur Kreuzreaktivitäten aufzeigen, die die Schwierigkeiten beim Screening illustrieren. Dabei ist der Mangel an Spezifität und der Erhalt falsch-positiver Klone ein ausreichender Beleg für inherente Schwierigkeiten beim Screening hoher Klon-Zahlen. Darüber hinaus muss die Diversität der Cosmid-Bibliothek kritisch betrachtet werden. Legt man für die Kobra eine Genomgröße von ca. 2,5 Mrd. Basenpaaren zugrunde, so wäre für die erfolgreiche Identifizierung überlappender Klone der CVF/C3-Genfamilie eine Mindestgröße von ca. $1 \cdot 10^6$ Klonen erforderlich. Inwieweit zudem die angenommene

Diversität der Bibliothek realistisch ist oder eventuell durch langfristige Lagerung o.ä. Einbußen erlitten haben könnte, bleibt offen. Demzufolge muss das Screening der vorhandenen Bibliothek für die Identifizierung positiver Klone über das gesamte CVF1-Gen als außerordentlich schwierig angenommen werden, was durch die erfolglose Durchmusterung unterstrichen wird.

Aus diesem Grunde wurden andere Methoden wie genomische PCR und Genome Walking eingesetzt, um die Genstruktur von CVF1 zu analysieren.

Genomische PCR

Die genomische PCR bietet die Möglichkeit, ohne Klonierung von Bibliotheken direkt aus primären Quellen Informationen über die Struktur des CVF1-Gens zu erhalten. Dabei ist allerdings zu berücksichtigen, dass der Vermeidung von Sequenzhomologien bei der Auswahl der Primer höchste Beachtung beigemessen werden muss. Gerade die Homologie zu C3 kann hier als Kriterium genommen werden, um die Spezifität der PCR zu evaluieren.

Für die Etablierung der genomischen PCR zur Aufklärung der CVF1-Genstruktur wurde daher zunächst genomische DNA aus der Leber der Monokel-Kobra *Naja kaouthia* isoliert, auf Größe und Reinheit untersucht und hinsichtlich ihrer Eignung für genomische PCR getestet. Dazu wurde exemplarisch für die Amplifikation von ausgewählten CVF1-Introns, deren Größen durch Aufklärung des Cosmid-Klons 1R67 bekannt waren, sowohl genomische als auch 1R67-DNA als Matrize eingesetzt. Es zeigte sich, dass die genomische PCR für die Amplifikation von Introns geringer Größe (bis ca. 3 kb) problemlos einsetzbar ist. Da die hierfür verwendeten Primer nicht hochspezifisch gewählt waren und daher nicht zwischen den drei Gensequenzen (CVF1, CVF2, coC3) differenzieren, wäre theoretisch die Co-Amplifikation von jeweils drei Intronsequenzen möglich gewesen. Tatsächlich wurde jedoch meist nur ein Amplifikat der erwarteten Größe erhalten. Für Intron 16 allerdings wurde neben dem CVF1-Intron (1,5 kb) noch ein kleineres coC3-Amplifikat (903 bp) gebildet, was auf unterschiedliche Introngrößen innerhalb der CVF/C3-Genfamilie hindeutet.

Für die gezielte genspezifische Amplifikation von CVF1-Introns aus einem Pool homologer Sequenzen wurde die ARMS-Methode angewandt. Das *amplification refractory mutation system* ist eine PCR-Anwendung, bei der die Primer so gewählt sind, dass sie zwischen Templates differenzieren können, die sich in einem einzigen Nukleotid unterscheiden (Newton *et al.* 1989). ARMS basiert auf dem Prinzip, dass der Taq DNA-

Polymerase eine 3'-5'-Exonuklease-Aktivität fehlt. Aufgrund dessen führt eine Fehlpaarung zwischen dem 3'-Ende eines PCR-Primers und dem Template zu keiner bzw. nur zu einer sehr eingeschränkten Amplifikation. Die Spezifität der Methode kann dadurch erhöht werden, dass zwei ARMS-Primer verwendet werden und zusätzlich interne Fehlpaarungen der Primersequenz zu den homologen Templates vorhanden sind, welche zu einer verminderten Hybridisierungseffizienz führen. Unter diesen Bedingungen sollte nur dann ein positives Signal entstehen, wenn das richtige Gen amplifiziert wird. Die Wahl der Primer war durch die Sequenzen der CVF/C3-Genfamilie vorgegeben. Die ARMS-Methode wurde mit Erfolg zur genspezifischen Amplifikation innerhalb der CVF/C3-Multigenfamilie eingesetzt. Da sich jedoch einige Introns auch unter variierten Bedingungen nicht amplifizieren ließen (Introns 2, 8, 27 und 36), wurde geschlossen, dass es sich dabei um lange Introns mit einer Größe von mehr als 5 kb handelt. Eine Amplifikation von Genabschnitten dieser Größenordnung gelingt oft mit Mischungen aus Taq- und Pwo-DNA-Polymerasen (bzw. Taq/Tgo-DNA-Polymerasen), wobei die Pwo- (bzw. Tgo-) Polymerase aufgrund ihrer 3'-5'-Exonuklease-Aktivität eine Korrekturfunktion besitzt (sogenannte Proofreading-Polymerase). Die Verwendung von Proofreading-Polymerasen in der genomischen PCR erlaubt die Amplifikation von bis zu 35 kb genomischer Sequenzen, sofern die genomische Matrizen-DNA höchst intakt und kontaminationsfrei ist (Barnes 1994). Bei einer in dieser Arbeit eingesetzten Kombination von CVF1-spezifischen Primern mit interner und 3'-Fehlpaarung zu den entsprechenden CVF2/coC3-Sequenzen und Proofreading-Polymerasen wurden teilweise Amplifikate erhalten, bei denen es sich trotz Verwendung von CVF1-spezifischen Primern um Sequenzen der homologen Gene CVF2 und coC3 handelte. Aus den Sequenzen der amplifizierten Produkte war ersichtlich, dass durch die Proofreading-Polymerase jeweils das letzte Nukleotid der 3'-Fehlpaarung der Primer korrigiert wurde. Auch bei einer doppelten Fehlpaarung am 3'-Ende hatte eine Korrektur des jeweils letzten Nukleotids genügt, um eine Verlängerung des Primers zu ermöglichen. Vieles deutete darauf hin, dass durch Verwendung von Proofreading-Polymerasen das jeweils kleinste Amplifikat der CVF/C3-Genfamilie bevorzugt gebildet wurde.

Ein weiterer Punkt, der bei Amplifikationen generell berücksichtigt werden muss, ist die Möglichkeit einer fehlerhaften Insertion von Nukleotiden durch die verwendeten Polymerasen. Die Fehlerrate der Taq-Polymerase liegt bei ungefähr 10^{-5} je eingebautem Nukleotid, während die Fehlerrate für Polymerasen mit Korrekturaktivität etwa um den

Faktor 10 niedriger ist (Cline *et al.* 1996). Abweichungen der genomischen Sequenzen von den cDNA-Sequenzen können daher auch von Polymerase-Fehlern herrühren.

Insgesamt erwies sich die Etablierung der genomischen PCR als vorteilhaft zur Aufklärung der genomischen Struktur der unterschiedlichen Gene, da innerhalb kurzer Zeit spezifische Sequenzinformationen erhalten werden konnten.

Genome Walking

Da die genomische PCR aufgrund der Notwendigkeit zur Auswahl hochspezifischer Primer keine Möglichkeit bietet, unbekannte Genabschnitte zu analysieren, wurde für die Aufklärung derjenigen Exon/Intron-Übergänge, bei denen eine Amplifikation aufgrund der erheblichen Intron-Längen nicht möglich war, die Methode des Genome Walking etabliert. Diese beruht auf einer verschachtelten PCR mit adapter-ligierten Bibliotheken und benötigt jeweils zwei genspezifische Primer in derselben Orientierung. Weiterhin wurde diese Methode verwendet, um die 5'- und 3'-Regionen der CVF/C3-Genfamilie aufzuklären.

Bei der Genome Walking-Methode wurden ebenso wie bei der oben beschriebenen ARMS-Methode genspezifische Primer eingesetzt, die eine 3'-Fehlpaarung zu den entsprechenden homologen Sequenzen aufwiesen. Anders als erwartet wurden bei der Amplifikation mit einer einzigen Genome Walking-Bibliothek jedoch nicht nur singuläre genspezifische Amplifikate, sondern bis zu drei verschieden große Produkte erhalten (z.B. Exon 36/Intron 36). Da alle Amplifikationen mit einer Hot-Start Taq-DNA-Polymerase ohne 3'-5'-Exonuklease-Aktivität durchgeführt wurden und demzufolge keine Veränderung der Primersequenz möglich war, deutete die Bildung mehrerer Amplifikate in einigen Fällen auf die Existenz weiterer Mitglieder der CVF/C3-Genfamilie hin (z.B. Exon 36/Intron 36, Intron 36/Exon 37).

Zur Charakterisierung der CVF1-Genstruktur wurde meist das schwerste Amplifikat der vier Bibliotheken kloniert und sequenziert. Die gesuchte CVF1-Sequenz galt erst dann als identifiziert, wenn die erhaltene Sequenz zweifelsfrei dem CVF1-Gen zuzuordnen war. Einige Amplifikate konnten dadurch zugeordnet werden, dass ihre Größe mit einer entsprechend vorhandenen Schnittstelle in der Sequenz korrelierte.

4.2 Die Struktur des CVF1-Gens

Die Kombination der etablierten Techniken ermöglichte insgesamt eine nahezu lückenlose Charakterisierung der genomischen Struktur von CVF1 und darüber hinaus die partielle Charakterisierung der Genstrukturen von coC3 und CVF2. Die sich ergebenden Daten aus den unterschiedlichen Ansätzen wurden dabei zu der Gesamtinformation vereint.

Anhand der erhaltenen Ergebnisse konnte gezeigt werden, dass das CVF1-Gen eine Gesamtgröße von mehr als 89 kb aufweist und aus 40 Exons und 39 Introns zusammengesetzt ist. Der Promotor und der Terminationsbereich des CVF1-Gens sind in Kapitel 4.5 beschrieben.

Die Größen der CVF-1-Exons variieren zwischen 46-1140 bp. Das erste Exon enthält den 5'-untranslatierten Bereich und codiert für die Signalsequenz sowie die ersten beiden Aminosäuren der α -Kette. Das 3'-terminale Exon (Exon 41) ist mit einer Gesamtgröße von 1140 bp das größte Exon und setzt sich sowohl aus 142 bp der 3'-codierenden Sequenz inklusive Stop-Codon sowie 998 bp der 3'-UTR zusammen. Das klassische Poly-Adenylierungssignal AATAAA (Gil und Proudfoot 1984) kommt im 3'-untranslatierten Bereich des CVF1-Transkripts vierfach vor, wobei sich das am weitesten stromabwärts gelegene Signal 40 bp vor dem Poly(A)-Anhang der reifen mRNA befindet (Position 5889-5894 der CVF1-cDNA). Das Vorkommen multipler Polyadenylierungssignale im 3'-Bereich der Gene CVF1, CVF2 und coC3 wird in Kapitel 4.6 diskutiert.

Die Exon-Sequenzen stimmen mit den publizierten cDNA-Sequenzen mit Ausnahme von fünf Nukleotiden überein. Im 5'-untranslatierten Bereich in unmittelbarer Nähe des Startcodons ATG wurde anstelle von C an Position 1 der cDNA ein A nachgewiesen. Dadurch stimmt die genomische Sequenz von CVF1 ebenso wie die von CVF2 und coC3 mit den Erfordernissen einer Kozak-Sequenz ACCATGG überein (siehe Abb. 4.4), derzufolge in der mRNA von höheren Eukaryoten meistens ein A an der Position -3 und ein A oder G an der Position +4 vorkommt (Kozak 1986, 1987). Im codierenden Bereich wurde an Position 1264 ein T anstelle von C identifiziert. Das veränderte Codon TCA codiert für Serin anstelle von Prolin (CCA). Diese Nukleotidabweichung in Exon 12 wurde auch bei einem *long range*-PCR-Amplifikat festgestellt, so dass es sich bei dieser Abweichung von der cDNA möglicherweise um eine Variation zwischen Mitgliedern derselben Spezies handelt (*single nucleotide polymorphism*, SNP). Im 3'-untranslatierten Bereich wurden 3 Unterschiede festgestellt: Nukleotid 5161 wurde als G anstelle von A identifiziert, Positionen 5648-5651 enthalten fünf G anstelle von vier, und Positionen

5806-5807 enthalten drei G anstelle von zwei. Diese Änderungen im 3'-Bereich wurden alle durch Re-Sequenzierung der CVF1-cDNA bestätigt (Fritzinger, persönliche Mitteilung). Demzufolge ist die 3'-UTR um 2 Nukleotide länger als früher berichtet (Fritzinger *et al.* 1994b), und die Länge der CVF1-cDNA erhöht sich auf 5930 bp.

Die Größen der Introns variieren zwischen 100 und ca. 9000 bp, wobei alle 5'- und 3'-Spleiss-Donor- und Akzeptor-Sequenzen der gt/ag-Regel folgen (Mount 1982). Bemerkenswerterweise konnte das komplette Fehlen von Intron 31 im CVF1-Gen nachgewiesen werden. Dieser Befund wurde sowohl durch Genome Walking als auch durch überlappende genomische PCR bestätigt. Dieses Ergebnis unterstreicht damit nicht nur die grundsätzliche Vereinbarkeit der beiden Methoden, sondern zeigt auf, dass beide Methoden für sich selbst geeignet sind, korrekte Ergebnisse zu liefern. Die Nummerierung stromabwärts von Exon 31 wurde aufgrund der stark konservierten Exongrößen in Anlehnung an die verwandten Komplementgene beibehalten (siehe Tab. 4.1). Für Introns 2, 8, 27 und 36, die durch die Genome Walking-Methode partiell charakterisiert wurden, wurden Mindestgrößen ermittelt. Da eine vollständige Amplifikation dieser Introns durch genomische PCR auch unter Verwendung von Proofreading-Polymerasen nicht möglich war, werden für diese Introns Größen von mehr als 5 kb vermutet, da Introns dieser Größenordnung häufig im CVF1-Gen vorkommen (Introns 1, 15, 17, 25, 39). Diese Größen erscheinen allerdings durchaus nicht als unrealistisch, da Introns mit mehr als 200 kb im humanen Dystrophin-Gen nachgewiesen wurden, die damit zu den bisher größten identifizierten Introns zählen (Coffey *et al.* 1992). In diesem Gen, das mit einer Gesamtlänge von mehr als 2 Megabasen das grösste und komplexeste humane Gen darstellt (Roberts *et al.* 1993, Mandel 1989), sind Introns der Größenordnung >200 kb zudem kein singuläres Ereignis, sondern finden sich über das gesamte Gen verteilt. Daher werden auch für die nicht vollständig bestimmten Introns des CVF1-Gens Längen in der Größenordnung von 5-9 kb vermutet.

In anderen Genen wurden in Einzelfällen innerhalb größerer Introns verschachtelt vollständige Gene identifiziert, was der ursprünglichen Annahme widerspricht, Introns seien nicht-exprimierte Regionen. So ist z.B. ein Cuticula-Gen der Puppe von *Drosophila* innerhalb eines Introns des GART-Gens enthalten (Moriyama und Gojobori 1989). Eine ähnliche Entdeckung wurde auch bei den humanen Komplementgenen C4A und C4B gemacht, welche eine ungewöhnliche Gengrößen-Dichotomie (14,2 kb bzw. 20,6 kb) aufweisen. Der Größenunterschied ist allein auf ein verlängertes Intron 9 zurückzuführen, innerhalb dessen das humane endogene Retrovirus HERV-K(C4) mit

einer Länge von 6,4 kb identifiziert wurde (Dangel *et al.* 1994). In den partiell sequenzierten Introns der CVF/C3-Genfamilie wurde allerdings kein Hinweis auf ein verschachteltes neues Gen gefunden.

4.3 Die Struktur des CVF1-Gens im Vergleich zu verwandten Komplementgenen

Die Charakterisierung der Struktur des CVF1-Gens wurde unter Berücksichtigung der Struktur des huC3-Gens vorgenommen, wobei die Exon/Intron-Struktur durch Vergleich der cDNA-Sequenzen vorhergesagt werden konnte. Die tatsächliche Übereinstimmung der genomischen Struktur belegt damit eine starke Konservierung der Strukturen beider Gene. Als grundsätzliches Merkmal sind die Exongrößen beider Gene sehr ähnlich. 23 Exons zeigen identische Größen, insbesondere im 3'-Bereich ab Exon 29. Bei Exons unterschiedlicher Größe beträgt der Unterschied meist ein Vielfaches eines Codons, wobei sich die Größenunterschiede im Bereich von 3 bp (Bsp. Exons 3 und 11) bis 15 bp (Exon 5) bewegen.

Ausnahmen von der vorhergesagten Genstruktur bilden das Fehlen von Intron 31 in CVF1-Gen sowie eine Phasenverschiebung in Intron 19. Bei den Exons 19 und 20 ist der Unterschied von 3 bp in Exon 20 aufgrund einer Phasenverschiebung auf beide Exons verteilt (Abb. 3.5). Daher weist Exon 20 von huC3 eine Größe von 142 bp auf, wohingegen die aller weiteren in Tabelle 4.1 aufgelisteten Komplementgene 140 bp beträgt. Ebenso unterscheidet sich nur das huC3-Exon 19 von allen vergleichbaren Exons dadurch, dass die Differenz nicht dem Vielfachen eines Codons entspricht. Demzufolge handelt es sich bei der Phasenverschiebung in Intron 19 um eine Besonderheit des huC3-Gens. Der wesentlichste Unterschied zur humanen C3-Genstruktur ist jedoch das Fehlen von Intron 31 im CVF1-Gen und den homologen Kobra-Genen.

Während die Positionen der Exons und der Introns im CVF1- und huC3-Gen weitgehend konserviert sind, zeigen die Intronsequenzen selbst deutliche Unterschiede. Beim Vergleich ausgewählter Intronsequenzen von CVF1 und human C3 (*Homo sapiens* Chromosom 19, Klon CTD-3218G10, Genbank-Nr. AC008760.7) konnte keine Homologie festgestellt werden. Dieser Befund deutet auf die Akkumulation von zahlreichen Mutationen in den Introns hin, so dass außerhalb der funktionell bedeutsamen Spleiss-region keine Ähnlichkeiten mehr zu erkennen sind. Schätzungen zufolge lag der Zeitpunkt der Divergenz der Mammalia-Linie von der Reptilien-Linie vor etwa 300 bis 350

Millionen Jahren. Selbst bei einem Genvergleich zwischen verschiedenen Mammalia-Arten sind oft keine Homologien der Introns festzustellen, wie ein Vergleich der humanen und murinen C4-Introns ergab (Ogata *et al.* 1989).

Das größte identifizierte Intron des CVF1-Gens (9000 bp, Intron 15) ist etwa doppelt so groß wie das größte huC3-Intron (4400 bp, Intron 17). Zudem ist das CVF1-Gen mit einer Mindestgröße von 89 kb mehr als doppelt so groß wie das huC3-Gen (42 kb) und nahezu viermal so groß wie das murine C3-Gen (24 kb). Dieses Größenverhältnis des CVF1-Gens zu den beiden homologen C3-Genen wurde bereits nach Aufklärung des Cosmidklons 1R67 erkennbar, der bei einer Insertlänge von ca. 33 kb den Genabschnitt von Intron 15 bis Intron 27 abdeckt. Der 3'-untranslatierte Bereich ist im CVF1-Gen mit 998 bp deutlich größer als der entsprechende Bereich im huC3-Gen mit 46 bp. Insgesamt ist das CVF1-Gen mit einer Länge von mindestens 89 kb deutlich größer als alle bisher charakterisierten C3-, C4- und C5-Gene.

Ein Vergleich mit dem murinen C3-Gen ist nur eingeschränkt möglich, da dessen partielle Struktur noch vor der Erfindung der Polymerase-Kettenreaktion beschrieben wurde (Wiebauer *et al.* 1982, Lundwall *et al.* 1984). Neben der Gesamt-Gengröße (24 kb) sind die flankierenden Regionen charakterisiert. Die Größen von Exon 1 (136 bp), Exon 2 (190 bp) und Exon 41 (139 bp) liegen im Bereich der Erwartungen.

Der 3'-untranslatierte Bereich ist in allen hier beschriebenen Komplementgenen von großer Varianz und reicht von 39 bp im murinen C3-Gen bis 998 bp im CVF1-Gen.

In den Fällen, bei denen sich die gleichen Exons in ihrer Größe unterscheiden, beträgt der Unterschied immer ein Vielfaches von 3 bp und somit einem Codon. In zwei Ausnahmen sind aufgrund einer Phasenverschiebung ein Codon auf 2 benachbarte Exons verteilt. So beträgt der Unterschied zwischen Exons 19 und 20 in CVF1 und huC3 insgesamt 3 bp, ebenso wie bei Exons 2 und 3 in huC5 und muC5. Die Exongrößen von CVF2 und coC3 sind aufgrund von Homologievergleichen mit CVF1 vorhergesagt und wurden teilweise durch Sequenzierung von genomischen PCR-Produkten bestätigt. Obwohl die cDNAs der dargestellten Gene mit ca. 5000 bp etwa gleich groß sind, variieren die Gen-Größen zwischen 16 und >89 kb. Die Größen von Exon 1 und Exon 41 variieren zum Teil erheblich aufgrund unterschiedlich langer untranslatierter Bereiche. Innerhalb der CVF/C3-Genfamilie existieren Größenvariationen nur bei Exons 5, 6, 14 und 33. Die Größen-Unterschiede betragen 3 bp (Exons 5 und 6), 9 bp (Exon 33) und 12 bp (Exon 14).

Tab. 4.1: Vergleich der Exongrößen von CVF1 und verwandter Komplementgene. Die Zahlen geben die Exongrößen in Basenpaaren wieder (n.a.: nicht analysiert). Bei den grau schattierten Zeilen handelt es sich um Exons, deren Größen in allen Genen identisch sind. Die Größen der umrandeten Exons im humanen C3-Gen unterscheiden sich aufgrund einer Phasenverschiebung von den homologen Exons (siehe Text). Die Exongrößen für CVF2 und coC3 wurden aufgrund der stark konservierten Exon-Spleissstellen vorhergesagt und exemplarisch bestätigt. Die Genstrukturen der weiteren Komplement-Gene wurden der Literatur entnommen: human C3 (Vik *et al.* 1991), human C4A (Yu 1991), murin C4 (Ogata *et al.* 1989), human C5 (Carney *et al.* 1991), murin C5 (Haviland *et al.* 1991b).

| Exon | C3 und Homologe | | | | C4 | | C5 | |
|-----------------|-----------------|-------|------|------|-------|------|------|------|
| | CVF1 | CVF2 | coC3 | huC3 | huC4A | muC4 | huC5 | muC5 |
| 5'-UTR | (3) | n.a. | (8) | 60 | | 56 | 29 | 110- |
| 1 | 71 | 71 | 71 | 74 | 126 | 65 | 65 | 65 |
| 2 | 193 | 193 | 193 | 193 | 199 | 193 | 193 | 193 |
| 3 | 163 | n.a. | 163 | 166 | 202 | 202 | 163 | 163 |
| 4 | 71 | n.a. | 71 | 71 | 71 | 71 | 71 | 71 |
| 5 | 80 | n.a. | 83 | 95 | 89 | 89 | 92 | 92 |
| 6 | 80 | n.a. | 83 | 83 | 83 | 83 | 83 | 83 |
| 7 | 91 | n.a. | 91 | 91 | 97 | 97 | 91 | 91 |
| 8 | 103 | n.a. | 103 | 103 | 106 | 106 | 115 | 115 |
| 9 | 121 | n.a. | 121 | 127 | 133 | 133 | 127 | 127 |
| 10 | 116 | 116 | 116 | 116 | 116 | 116 | 116 | 116 |
| 11 | 147 | 147 | 147 | 150 | 180 | 177 | 186 | 186 |
| 12 | 210 | 210 | 210 | 210 | 183 | 186 | 204 | 204 |
| 13 | 195 | 195 | 195 | 207 | 186 | 186 | 210 | 210 |
| 14 | 150 | 162 | 162 | 159 | 159 | 159 | 150 | 150 |
| 15 | 130 | 130 | 130 | 130 | 127 | 127 | 130 | 133 |
| 16 | 75 | 75 | 75 | 72 | 75 | 75 | 63 | 72 |
| 17 | 201 | 201 | 201 | 198 | 198 | 195 | 198 | 198 |
| 18 | 112 | 112 | 112 | 109 | 112 | 115 | 91 | 91 |
| 19 | 86 | 86 | 86 | 87 | 71 | 71 | 74 | 74 |
| 20 | 140 | 140 | 140 | 142 | 140 | 140 | 140 | 140 |
| 21 | 213 | 213 | 213 | 213 | 210 | 204 | 228 | 228 |
| 22 | 64 | 64 | 64 | 67 | 52 | 52 | 61 | 61 |
| 23 | 87 | 87 | 87 | 87 | 90 | 90 | 90 | 90 |
| 24 | 204 | 204 | 204 | 204 | 210 | 210 | 213 | 213 |
| 25 | 76 | 76 | 76 | 76 | 76 | 76 | 76 | 76 |
| 26 | 163 | 163 | 163 | 160 | 157 | 157 | 160 | 160 |
| 27 | 96 | 96 | 96 | 99 | 117 | 117 | 96 | 96 |
| 28 | 151 | 151 | 151 | 157 | 172 | 172 | 172 | 172 |
| 29 | 164 | 164 | 164 | 164 | 233 | 233 | 206 | 206 |
| 30 | 159 | 159 | 159 | 159 | 168 | 168 | 153 | 153 |
| 31 | (60) | (60) | (60) | 60 | 60 | 60 | 63 | 63 |
| 32 | (91) | (91) | (91) | 91 | 94 | 94 | 82 | 82 |
| 33 | 46 | 55 | 55 | 52 | 187 | 181 | 58 | 58 |
| 34 | 88 | 88 | 88 | 88 | 91 | 91 | 103 | 103 |
| 35 | 90 | 90 | 90 | 90 | 75 | 75 | 75 | 75 |
| 36 | 106 | 106 | 106 | 106 | 103 | 103 | 106 | 106 |
| 37 | 90 | 90 | 90 | 90 | 90 | 90 | 84 | 84 |
| 38 | 84 | 84 | 84 | 84 | 99 | 99 | 90 | 90 |
| 39 | 84 | 84 | 84 | 84 | 84 | 84 | 84 | 84 |
| 40 | 136 | 136 | 136 | 136 | 133 | 133 | 139 | 136 |
| 41 | 142 | 137 | 142 | 142 | 142 | 142 | 130 | 133 |
| 3'-UTR | 998 | >147 | 240 | 46 | 119 | 99 | 401 | 356 |
| cDNA-Größe (bp) | 5930 | >4119 | 5204 | 5099 | 5415 | 5372 | 5444 | 5403 |
| Gen-Größe (kb) | > 89 | n.a. | n.a. | 42 | 17 | 16 | 79 | 78 |

Der überwiegende Teil der Exons weist eine identische Größe auf. Bemerkenswert dabei ist, dass CVF2 und coC3 im vorliegenden Bereich identische Exonstrukturen aufweisen. Dies ist ebenfalls ein Hinweis auf die enge genetische Verwandtschaft von CVF2 und coC3, wie sie im phylogenetischen Stammbaum (Abb. 4.6) dargestellt ist.

Beim Vergleich der Genstrukturen von CVF1 und huC3 zeigen sich bei 23 Exons (56%) identische Größen (Exons 2, 4, 7, 8, 10, 12, 15, 21, 23, 24, 25, 29, 30, (31, 32), 34, 35, 36, 37, 38, 39, 40 und 41 (ohne 3'-UTR)). Auffällig ist dabei die große Anzahl konservierter Exongrößen im 3'-Bereich der Gene.

Bei allen in Tabelle 4.1 dargestellten Komplementgenen finden sich konservierte Größen bei Exons 4, 10, 25 und 39 (schattiert dargestellt). Diese Tatsache unterstreicht die genetische Verwandtschaft dieser Genfamilie und die Hypothese, dass diese Gene aus einem gemeinsamen Vorläufergen entstanden sein könnten, wie auch bereits nach vergleichenden Analysen der cDNAs und Proteinsequenzen vermutet wurde (Fritzinger *et al.* 1994b, Sottrup-Jensen *et al.* 1985, Hughes 1994, Dodds und Law 1998).

Die CVF/C3-, C4- und C5-Gene weisen bei einigen Exons eine jeweils charakteristische Länge auf (Exons 8, 30, 32, 34 und 38), was zusätzlich auf eine verwandtschaftliche Beziehung dieser Gene schließen lässt. Jeweils gleiche Exongrößen der CVF/C3- und C4-Gene (Bsp. Exon 37), der CVF/C3- und C5-Gene (Bsp. Exons 7, 36) sowie der C4- und C5-Gene (Exons 23, 35) lassen ebenfalls Rückschlüsse zu Verwandtschaftsgrad und Funktion zu. Weitere Aufschlüsse über die Beziehungen der Genstruktur und Proteinfunktion können sicherlich aus zukünftigen Erkenntnissen der dreidimensionalen Proteinstruktur gewonnen werden.

Der markanteste Größenunterschied ist bei Exon 33 zwischen dem CVF1- und dem humanen C4A-Gen zu finden. Das CVF1-Exon 33, das mit 46 Basen zugleich das kleinste Exon der dargestellten Komplementgene im allgemeinen und des CVF1-Gens im besonderen ist (siehe auch Abb. 4.1), weist zum entsprechenden C4A-Exon (187 bp) einen Unterschied von 141 bp (entsprechend 47 Aminosäuren) auf. Ein noch kleineres Exon der Komplementgene als das CVF1-Exon 33 befindet sich mit 21 bp in Exon 7 des humanen Faktor I-Gens (Vyse *et al.* 1994). Die kleinsten Exons mit Größen von nur 3 bp sind für das N-CAM-Gen (Prediger *et al.* 1988) und das Troponin T-Gen von *Drosophila* (Benoist *et al.* 1998) beschrieben.

4.4 Vergleich der Genstrukturen von CVF1, CVF2 und coC3

In Analogie zur Vorhersage der Genstruktur von CVF1 konnten die Genstrukturen von CVF2 und coC3 aufgrund der erhaltenen Informationen über CVF1 vorhergesagt werden. Diese Vorhersage konnte durch Amplifikationen exemplarisch ausgewählter Introns mittels genomischer PCR und Genome Walking erfolgreich verifiziert werden. Dabei konnte gezeigt werden, dass für die C3/CVF-Genfamilie der Kobra *Naja kaouthia* die Exon-Spleissstellen in allen drei Genen hochkonserviert sind. Die Introns variieren meist in ihrer Größe. Am Beispiel von Intron 32 wurde sowohl die Größenvarianz als auch die Konservierung der Intronsequenzen demonstriert (Abb. 4.1). In keinem der drei Gene ist entgegen der Erwartung aus der humanen C3-Genstruktur an der ursprünglich postulierten Position Intron 31 vorhanden. Eine Vorhersage der Exongrößen von CVF2 und coC3, die auf den Daten von CVF1 basiert, ist daher für die gesamte codierende Sequenz möglich und in Tabelle 4.1 gezeigt.

Dabei ist auffallend, dass alle Exons von CVF2 und coC3 identische Größen aufweisen, wohingegen sich Exons 5, 6, 14 und 33 des CVF1-Gens von denen der coC3-/CVF2-Gene unterscheiden. Durch diesen Größenunterschied in vier Exons ist die CVF1-cDNA im codierenden Bereich um 27 bp kleiner als CVF2 bzw. coC3. Die identischen Exongrößen von coC3 und CVF2 unterstreichen ebenfalls die enge phylogenetische Verwandtschaft dieser Gene. Für die noch nicht identifizierte 5'-Sequenz von CVF2 kann daher auch von identischen Exongrößen wie bei coC3 ausgegangen werden. Dies wiederum würde bestätigen, dass es sich bei den amplifizierten und dem CVF1-Gen zugeordneten Introns 4, 5 und 6 nicht um CVF2-Introns handeln kann. Bei Exon 41 des CVF2-Gens entspricht die Differenz zu allen anderen Genen (5 bp zu Exon 41 von CVF1 und coC3) nicht dem Vielfachen eines Codons. Daher könnte davon ausgegangen werden, dass die vorläufigen cDNA-Sequenzdaten von CVF2 (Fritzinger *et al.* 1994b) in diesem Bereich eine Abweichung von der tatsächlichen Sequenz von bis zu fünf Nukleotiden aufweisen. Andernfalls würde dies zu einer Leserasterverschiebung führen. Im Falle einer Deletion würde es sich um eine Exongrößenvarianz zwischen CVF2 und coC3 handeln. Das Stop-Codon ist in allen drei Genen identisch (TAA, *ochre*) und an konservierter Position.

Eine Analyse der 3'-untranslatierten Bereiche von CVF1, CVF2 und coC3 erfolgt in Kapitel 4.6.

Die Sequenzdaten im 5'-Bereich des CVF2-Gens wurden für den Promotor sowie für Exons 1 und 2 aufgeklärt und sind geeignet, um auf cDNA-Ebene durch PCR die CVF2-Sequenz zu komplettieren.

4.5 Vergleich von Intronsequenzen

Um die genetische Verwandtschaft der CVF/C3-Genfamilie der Kobra *Naja kaouthia* zu analysieren, wurden neben einer Charakterisierung der Exongrößen und -sequenzen (cDNA-Bereich) und des Promotorbereichs auch die Intronsequenzen auf Homologien hin untersucht. Diese Resultate können Aufschluss über den Verwandtschaftsgrad der drei Gene sowie über ihre Entstehung im Verlauf der Evolution geben. Exemplarisch werden in Abbildung 4.1 die Intron 32-Randsequenzen näher charakterisiert.

```

CVF1  TGAgtaagtt tgactggggg cctttggcct gctagaaata gaatctaatg
CVF2  TGAgtaagtt tg----- -----gctt cctagaaata gtaggttgta
coC3  TGAgtaagtt tg----- -----gctt cctagaaata gtaggttgta

CVF1  a--aatgtat ggtgggggga gtttaaagtg tttctattca tcatgtcagg
CVF2  a-----g taattaaatt ttagtgttca ccatgtcagt
coC3  atgaa gtttc tgaggagaga taattaaatt ttagtgttca ccatgtcaga

CVF1  atcccaatat ggaacaagct ttctctatta cttattcata aaaa--gttt
CVF2  atctcaatat gaacaagct ttctctgcta ttattcata aaaaaagctt
coC3  atctcaatat ggaacaagct atctctgcta ttattcata aaaaa-gtct

CVF1  ctcatttgat agaaaataac attattaatt ccatacttag ttgtaatcca
CVF2  ctaatttgat agaaaataat cttattaact tcatacttag ttgagatcca
coC3  ctaatttgat agaaaataat cttattaact tcatacttag ttgagatcca

CVF1  gttcttgaag tcaacaaa.....
CVF2  gtaacttcag tcaacaaa.....
coC3  gtaacttcag tcaacaaa.....

CVF1  ..... (ca. 1700 bp) .....
CVF2  ..... (ca. 3400 bp) .....
coC3  ..... (ca. 1400 bp) .....

CVF1  .....agccagaca
CVF2  .....agttagaca
coC3  .....agttggtca

CVF1  atcaaacttt tcagctt-ta aaaaactgct a-gaatttgt gcagagtcgg
CVF2  atcaaacttt tcaacttta aaaaactgct aagaatttgt tcagagtcgg
coC3  atcaaacttt tcaactt-aa aaaggctgct aagaatttgt gcagagtcgg

CVF1  attttctc-t tttcttttgt aattttttaa cttcaaag-- -----ATG
CVF2  attttcccct tttcttttgc aattctt--- -----agAC TTAAAAGACG
coC3  attttccc c tttcttttgt aattttt--- -----agAC TTAAAACAG

```

Abb. 4.1: Homologievergleich der genomischen Intron 32-Sequenzen von CVF1, CVF2 und coC3. Dargestellt sind die Intron-Randbereiche, soweit alle drei Sequenzen zueinander homolog sind. Exonsequenzen sind in fetten Großbuchstaben dargestellt, Lücken sind gestrichelt. Die Position der Lücke im 3'-Intron-Bereich von CVF2 und coC3 zeigt, dass nur die Nukleotide ag im 3'-Bereich der Introns konserviert sind (ag/gt-Regel).

Bei einem Vergleich der Introns der CVF/C3-Genfamilie wird erkennbar, dass die Introns nicht nur an den Randbereichen, sondern auch an internen Positionen über ausgedehnte Bereiche Homologien aufweisen können. Lücken variieren von einem Nukleotid bis zu mehreren Hundert Nukleotiden. Daraus resultieren die partiell stark divergierenden Introngrößen innerhalb der CVF/C3-Genfamilie. Eine schematische Vereinfachung der Intronstrukturen der CVF/C3-Genfamilie ist in Abbildung 4.3 gezeigt.

Intron 32

Am Beispiel von Intron 32, das für alle drei Gene in bezug auf Größe und Homologien der Randsequenzen charakterisiert wurde, lässt sich die genetische Verwandtschaft detaillierter untersuchen.

Dabei ist im 5'-Bereich der Intronsequenzen eine deutliche Ähnlichkeit der CVF2- und coC3-Sequenzen festzustellen. Dies manifestiert sich an der 14 bp-Lücke an identischer Position in Exon 32-Nähe, an einer weitgehend identischen Nukleotidfolge, sowie an einer Homologie, die sich über einen größeren Bereich als zu CVF1 erstreckt. Im 3'-Bereich ist diese Ähnlichkeit der CVF2- und coC3-Sequenzen durch eine 9 bp-Lücke an identischer Position in unmittelbarer Exonnähe charakterisiert. Die Homologie der Intronsequenzen und Positionen der Lücken deuten auf eine enge genetische Verwandtschaft zwischen CVF2 und coC3.

Zwar überwiegt weiter stromaufwärts die Ähnlichkeit von CVF1 zu CVF2, doch sind diese Sequenzhomologien auf Intronebene insgesamt konsistent mit dem Dendogramm der CVF/C3-Genfamilie der Kobra (Abb. 4.6), das auf cDNA-Sequenzen beruht und eine engere Verwandtschaft von CVF2 und coC3 anzeigt. Unterstützt wird diese Verwandtschaft durch die Tatsache, dass auch im Promotorbereich die CVF2- und coC3-Sequenzen eine höhere Homologie zueinander aufweisen als zur CVF1-Sequenz. Da jedoch CVF2 im Gegensatz zu coC3 in der Giftdrüse transkribiert wird, könnte eine genauere Analyse der cis-aktiven Elemente in der Promotorregion ergeben, welche Sequenzen essentiell für eine Transkription in der Giftdrüse (CVF1, CVF2) bzw. in der Leber (coC3) sind. Die DNA-Sequenzidentität (ab Exon 9 bis zum Stop-Codon) von CVF2 mit CVF1 und coC3 beträgt in beiden Fällen ca. 92%.

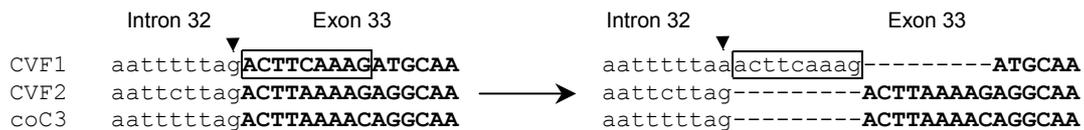


Abb. 4.2: Modell der Verkürzung des CVF1-Exons 33 durch eine singuläre Substitution anstelle einer Deletion. Aufgrund einer singulären g-a-Substitution in der Intron 32-Spleissstelle von CVF1 entsteht aus einer Exonsequenz eine Intronsequenz. Die Position des veränderten Nukleotids ist durch eine Pfeilspitze markiert. Als neue Spleissstelle fungiert die stromabwärts gelegene Nukleotidfolge AG. Der vom Exon zum Intron verwandelte Sequenzbereich ist umrahmt.

Bei einer exakteren Analyse der Sequenzen im Übergangsbereich Intron 32/Exon 33 in Abbildung 4.1 wird ersichtlich, dass sich die Lücken in Intron 32 von CVF2 und coC3 in unmittelbarer Nachbarschaft der Exon 33-Lücke von CVF1 befinden und alle Lücken dieselbe Länge von 9 bp aufweisen. Bei einer Verschiebung der Sequenzen unter Aufhebung der Lücken bleibt jedoch eine große Homologie erhalten. Abbildung 4.2 zeigt ein Modell, wie durch Substitution eines singulären Nukleotids an der Intron-Spleissstelle aus einer Exonsequenz eine Intronsequenz wird. Durch Veränderung der Intron-Konsensus-Sequenz ag, die am 3'-Ende aller Introns konserviert ist, kann das Intron nicht herausgespleisst werden. Infolgedessen wird Intron 32 erst 9 bp stromabwärts herausgespleisst, wo sich in der Gensequenz bei der Nukleotidfolge AG eine kryptische Spleissstelle befindet. Auf diese Weise könnte ein um 9 bp verkürztes Exon 33 des CVF1-Gens entstanden sein, welches mit 46 bp zugleich das kleinste Exon der in Tabelle 4.1 dargestellten Komplementgene ist. Bei allen weiteren Exonunterschieden innerhalb der CVF/C3-Genfamilie (Exons 5, 6, 14, 41) befinden sich die Lücken an interner Position, so dass die Größenunterschiede dort wahrscheinlich durch Deletion bzw. Insertion hervorgerufen wurden. Exon 33 des CVF1-Gens weist somit die Besonderheit auf, dass die 9 bp-Lücke am Exon-Randbereich nicht durch Deletion, sondern höchstwahrscheinlich aufgrund einer singulären g-a-Mutation entstanden ist.

Für die Expression von CVF1 als funktionell aktives Protein ist essentiell, dass keine Leserasterverschiebung zu einem Stop-Codon und vorzeitigen Abbruch der Expression führen. Zahlreiche ähnliche Mutationen, die zu einer Veränderung des natürlichen Spleissmusters führen und in defekten Proteinen resultieren, sind für das humane Globin-Gen beschrieben und verursachen verschiedene Formen der Thalassämie (Busslinger *et al.* 1981, Baird *et al.* 1981, Orkin *et al.* 1981, Weatherall 2001). Mutationen der Exon-Spleissstelle wurden auch in defekten Genen von C2 (Johnson *et al.* 1992), C3 (Botto *et al.* 1990) und C6 (Fernie *et al.* 1995) gefunden.

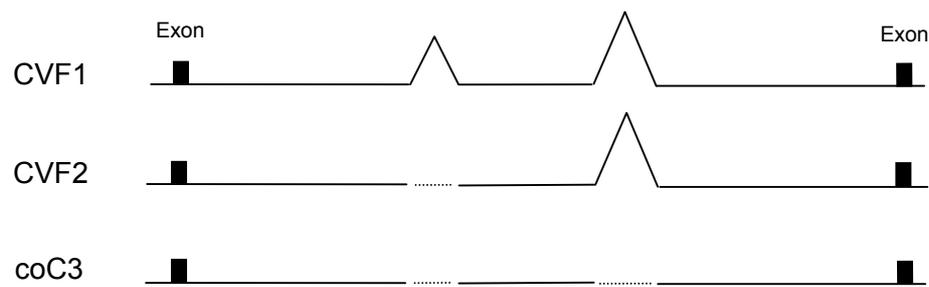


Abb. 4.3: Vergleich der Intronstrukturen von CVF1, CVF2 und coC3. Vereinfachte Darstellung der unterschiedlichen Größen und Strukturen am Beispiel eines hypothetischen Introns. Weitgehend identische Intronsequenzen sind durch durchgehende Linien gekennzeichnet. Größere Sequenz-Einschübe sind durch Dreiecksspitzen und entsprechende Lücken durch gepunktete Linien dargestellt.

Die Position der 9 bp-Lücke im 3'-Bereich von Intron 32 von CVF2 und coC3 zeigt, dass im 3'-Bereich von Introns nur die Nukleotide ag konserviert sind. Stromaufwärts davon befindet sich meist eine ca. 20 bp-lange Purin-reiche Sequenz. Allerdings befindet sich im CVF1-Gen an der Position direkt stromaufwärts von ag in wenigen Fällen auch ein Adenosin (Introns 32, 37, 38; siehe Tab. 4.1).

Im Gegensatz zu den hochkonservierten Genstrukturen der C3/CVF-Genfamilie mit meist unterschiedlichen Introngrößen (Tab. 4.2) verfügen andere Multicopy-Genfamilien der Kobra sogar über identische Intronsequenzen. So wurden für die sechs Cardiotoxin-Gene und drei Neurotoxin-Gene der Kobra *Naja sputatrix* bei identischen Intronsequenzen Variationen nur innerhalb Exons berichtet (Lachumanan *et al.* 1998, Afifiyan *et al.* 1999).

Tab. 4.2: Vergleich der Introngrößen der CVF/C3-Genfamilie. Die Introngrößen sind in Basenpaaren angegeben. Angaben in Klammern sind geschätzt. Nicht analysierte Introns sind durch einen Punkt markiert. Die charakterisierten Introns der CVF-homologen Sequenzen stammen möglicherweise von mehreren neuen Genen.

| Intron | CVF1 | CVF2 | coC3 | CVF-Homologe |
|--------|--------|----------|--------|--------------|
| 2 | > 1000 | (10.000) | (2500) | • |
| 3 | 732 | • | (1700) | • |
| 12 | 123 | • | • | 123 |
| 16 | (1400) | • | 795 | • |
| 18 | 1438 | (1400) | • | • |
| 27 | > 1300 | • | • | (1300) |
| 29 | (1800) | • | 952 | • |
| 32 | (2200) | (3700) | (1700) | • |
| 36 | > 1900 | • | • | (1900) |

Die Bestimmung vollständiger Intronsequenzen durch serielle Genome Walking-Schritte wird bei der CVF/C3-Multicopy-Genfamilie dadurch erschwert, dass die Introns nicht nur am Randbereich, sondern auch an internen Positionen konserviert sein können. Zudem wurden für Intron-Sequenzen signifikante Präsenz repetitiver Sequenzen gezeigt. Intron-Primer könnten daher eventuell in mehreren Genen hybridisieren und verlängert werden, wodurch eine Zuordnung der Sequenzen zu einem bestimmten Gen nur eingeschränkt möglich ist. In dieser Arbeit wurden daher die PCR-Primer so gewählt, dass bei allen entstehenden Amplifikaten Unterscheidungsmöglichkeiten auf Basis der Exonsequenzen gegeben sind.

Dabei ist nach wie vor ungeklärt, inwieweit repetitive Einheiten eine funktionelle Bedeutung aufweisen oder lediglich Ausdruck der Redundanz von Intronsequenzen sind. Zu den auffälligsten Tandem-Wiederholungen der CVF/C3-Genfamilie zählen eine 12fache (cttct)-Wiederholung in Intron 6 des CVF1-Gens sowie eine 11fache (tagaa)-Wiederholung in Intron 16 des coC3-Gens.

4.6 Regulation von CVF1, CVF2 und coC3

Weitere genomische Bereiche, die Aufschluss über die Verwandtschaft von CVF1, CVF2 und coC3 geben könnte, sind die für die Transkription und Translation der jeweiligen Proteine essentiellen regulatorischen Bereiche. Aus diesem Grunde wurden sowohl der Promotor-Bereich als auch der Terminationsbereich einer detaillierten Analyse unterzogen und putative Bindungsstellen regulatorisch relevanter Faktoren identifiziert.

Analyse der Promotorregionen von CVF1, CVF2 und coC3

Die Promotorregion von CVF1, CVF2 und coC3 enthalten mehrere AT-Heptanukleotid-Sequenzen, die als potentielle TATA-Box infrage kommen. Für die Identifizierung der tatsächlichen TATA-Box, die nach ihren Entdeckern auch als Goldberg-Hogness-Box bezeichnet wird, wurden die Promotorregionen auf Sequenz-Homologien untersucht (Abb. 4.4). Da die drei Gene im codierenden Bereich über eine sehr ausgeprägte Konservierung verfügen, sollte auch der Transkriptionsstart konserviert sein.

```

CVF1 -----
CVF2 ATTTCTTTTCAAAAAACCTGTGGAAACAACAGTAATGCCTTAATACAAATACTGTGATA
coC3 ATTTCTCTTTCAAAAAAACTGAGGAAACAACA---ATGCCTTAATAGCCTTATTTGTGATC

CVF1 -----
CVF2 AGTTTAGGAACAGACTTTGGGCTGTTGAGAAGTTGGGCTCACCCAAAATTGGGCTACTT
coC3 TGTTTAGGAGCCGATTT-----GGGTTACTT

CVF1 -----
CVF2 GAAACCTCTTCCAAATATGTGCATGAGGAAGAGTGGCGAAGTGAGCTTGTGAATCCAG
coC3 TAAAGCC-CTTTCAAATCTGTG--TGAGGGAGGGAAGGAAAGTGAGTTTGTGAGTTCCTG

CVF1 -----
CVF2 ATCACAAGATAAAGGAATATTTGCAGATACAATAGCAGAATACCAAACAAGTTGCATA
coC3 ATCTCAAGATAAAGGC-TATTTGCAGGTGTAATAGCAGAATACTTAGAACAAATTGCACA
TATA 3

CVF1 -----AAGACAAAATAATCCATTGT
CVF2 AAAACTGCCATTTAGCCATTGCATTGAAACCAGAGTGA--GAAGACAAAATAATCCAATGT
coC3 AAAA-TGCTGTTTAGCAATCATGTTTAACTAAGAGAGAAAGAAGACAAATAATCCACTGT

TATA 2
CVF1 GAGTTTTAAAGG---TATTAATTTTGTTGGTAATTC-----TCAAGCTCC-----
CVF2 GATTTTT-----TTGTTAATTTCCCTCTGAAAATTCCTGTCTGCCTCAAGCTCCTCTCA
coC3 GTTTTTAAAGGCCTCATTAATTTCCCTCTGAAAATTCCTGTTTGCTCAAGCTCCCCCTCA

CVF1 -----TCCTTTATACTGATTAACATTCAGCAATTGCAAAGGGGCTTGGCTTC
CVF2 CCTCCCCACCCTCCTTTATACTGATTAACACTCATCAATTGCAAAGGGGCTTGGCTTC
coC3 CCTCCCCACCCTCCTTTACACTGATTAACATTCATCAGTTGCAAAGGGGCTTGGCTTC

TATA 1 TIS
CVF1 AAAGCAGGCTCTCTGGATAAAATACTGGAGTTGTCAGCCCTTCTCTTGCCCTTCTGGACTA
CVF2 AAAGCAGGCTCTCTGGATAAAATACTGGAGTTGTCAGCCCTTCTCTTGCCCCTCTGGACTA
coC3 AAAGCAGGCTCTCTGGATAAAATACTGGAGTTGTCAGCCCTTCTCTTGCCCCTCTGGACTA

Kozak Exon 1
CVF1 CCATGGAGGGATGGCTCTCTATCTGGTGGCTGCTCTATTGATTGGTTTTCCAGGGTCTT
CVF2 CCATGGAGGGAGCGGCTCTCTCCTGGTGGCTGCTCTATTGATTGGTTTTCCAGGATCTT
coC3 CCATGGAGGGATGGCTCTCTATCTGGTGGCTGCTCTATTGATTGGTTTTCCAGGGTCTT

CVF1 CTCATGGGGCTCTgta...
CVF2 CTCATGGGACTCTgta...
coC3 CCCAGGGGGCTCTgta...

```

Abb. 4.4: Vergleich der Promotorregionen von CVF1, CVF2 und coC3. Die Sequenzanordnung der Promotorregion ist soweit dargestellt, wie mindestens zwei der Sequenzen homolog zueinander sind. Die Exon 1-Sequenzen sind ab dem Start-Codon fett dargestellt. Putative TATA-Boxen sind umrahmt. Der putative Transkriptions-Initiations-Start (TIS) ist markiert. Die Kozak-Sequenz ACCATGG ist unterstrichen.

Alle drei Gensequenzen sind unmittelbar stromaufwärts von Exon 1 weitgehend identisch. Zur Identifizierung von TATA-Boxen wurde dieser Bereich nach AT-Heptanukleotidsequenzen untersucht, wobei drei putative TATA-Boxen identifiziert und mit TATA 1-3 bezeichnet wurden. Es ist naheliegend, dass die mit TATA 1 bezeichnete Sequenz die tatsächliche Bindungsstelle des *TATA-Box binding protein* (TBP) und nachfolgend der RNA-Polymerase II darstellt, da diese TATA-Sequenz am nächsten zu den Start-Codons liegt und in allen drei Genen konserviert ist, ohne dass bei einem Vergleich der Gene Lücken zwischen der TATA-Sequenz und dem Start-Codon vorhanden sind. Bei TATA 1 handelt es sich um ein nicht-kanonisches Motiv, das als tatsächliche Bindungsstelle von TBP häufig vorkommt. Direkt stromaufwärts von TATA 3 besteht eine Homologie nur

zwischen CVF2 und coC3, was auf eine enge genetische Verwandtschaft dieser beiden Gene hinweist (siehe auch Abb. 4.6).

TATA-Boxen befinden sich gewöhnlich 26-34 bp stromaufwärts des Transkriptions-Initiationsstarts TIS (Breathnach und Chambon 1981). Da der Transkriptionsstart meist mit einem Purin beginnt und sich in allen drei Genen in dem definierten Abstand von der TATA-Box nur ein Purin inmitten von 7-10 Pyrimidinen befindet, wurde der putative Transkriptions-Initiations-Start (TIS) auf ein Nukleotid eingegrenzt. Demzufolge wären die 5'-untranslatierten Regionen der CVF/C3-Genfamilie mit einer Länge von 16 bp verhältnismäßig kurz.

Im CVF1-Promotor befinden sich stromaufwärts der proximalen TATA-Box zusätzlich zwei distale TATA-Motive, von denen eines (Sequenz ATAAATA, Position -292 bis -286) ebenfalls nicht-kanonisch (wie das proximale Motiv) ist und das andere (Sequenz TATAAATA, Position -470 bis -463) sowohl die kanonische Sequenz TATAAA als auch die nicht-kanonische Sequenz ATAAATA ineinander verschachtelt enthält. Diese beiden zusätzlichen TATA-Motive sind möglicherweise ebenfalls Bindungsstellen des *TATA-binding protein* (TBP), so dass eventuell die CVF1-Transkription von mehreren Startpunkten beginnt, ähnlich wie auch für das humane *single copy* C3-Gen zwei Transkriptionsstartpunkte nachgewiesen wurden (Vik *et al.* 1991).

Im coC3-Promotor wurden zusätzlich zur proximalen TATA-Box zwei weitere kanonische TATA-Sequenzen (Position -721 bis -716, invers, und Position -674 bis -669) identifiziert.

Putative Bindungsstellen für Transkriptionsfaktoren

Die Expressionsprofile von CVF1, CVF2 und coC3 sind derzeit noch nicht vollständig geklärt. Möglicherweise könnten jedoch Analysen der regulatorischen Elemente der jeweiligen Gene gewebsspezifische Expression untermauern. Daher wurden die 5'-flankierenden Regionen der Gene CVF1, CVF2 und coC3 anhand von Datenbankanalysen auf Bindungsstellen von Transkriptionsfaktoren hin untersucht.

Dabei konnten zahlreiche Bindungsstellen für Transkriptionsfaktoren wie das Activator Protein (AP) 1, Spezifitäts-Protein (SP) 1, sowie GATA-Motive, CACCC-Bindungsstellen und CCAAT-Boxen identifiziert werden. CCAAT-Boxen binden spezifisch Transkriptionsfaktoren wie CCAAT/enhancer binding protein (C/EBP), CP1, CP2 und

auch NF1 (Chodosh *et al.* 1988). Die GATA-Familie umfaßt bislang sechs Mitglieder (GATA-1 bis GATA-6), die das GATA-Motiv WGATAR erkennen.

Im coC3-Promotor wurden Bindungsmotive für die leber-spezifischen Transkriptionsfaktoren HNF-5 (*hepatocyte nuclear factor 5*, Grange *et al.* 1991) und NF-1L (*liver-enriched nuclear factor 1*, Schweizer-Groyer *et al.* 1992) identifiziert. Das HNF-5-Bindungsmotiv TRTTTGY zeigt sich allerdings auch im CVF2-Promotor, weshalb die Überprüfung der Expression von CVF2 in Leberzellen Hinweise auf die Bedeutung dieses Motives geben könnte, zumal die Expression von Toxinen in anderen Geweben als in Giftdrüsen dokumentiert ist (Ohno *et al.* 1998).

In den Promotorregionen aller drei Gene wurden ferner Glucocorticoid-Rezeptor (GR)-Elemente identifiziert. Solche GR-Elemente, die aus je einer Hälfte der palindromen Sequenz AGAACANNNTGTTCT des *Glucocorticoid response element* (GRE) bestehen, können ebenfalls regulatorische Funktionen haben; ihnen wurde eine Bedeutung für die Expression von Giftdrüsen-Proteinen beigemessen (Ma *et al.* 2001).

Eine Übersicht wichtiger regulatorischer Elemente in den Promotorregionen der Gene CVF1, CVF2 und coC3 ist in Tabelle 4.3 wiedergegeben. Inwieweit die putativen Bindungsstellen für die Regulation der Transkription tatsächlich von Bedeutung sind, muss hier unklar bleiben.

Tab. 4.3: Putative regulatorische Elemente in den Promotorregionen von CVF1, CVF2 und coC3. Putative Bindungsmotive für Transkriptionsfaktoren wurden durch Datenbankanalyse identifiziert und deren Vorkommen im jeweiligen Gen durch (+) markiert.

| Gen | | | Sequenz | Bindungsmotiv für Transkriptionsfaktor |
|------|------|------|--|--|
| CVF1 | CVF2 | coC3 | | |
| | + | + | CCAAT | CCAAT-Box |
| + | + | + | WGATAR | GATA |
| + | + | + | CACCC | CACCC |
| + | + | | CCGCCC/CCCGCC (CVF2) GGGGCAGGG (CVF1) | SP1 (GC-Box) |
| + | | + | AGGAAA | PEA-3 |
| | + | + | TRTTTGC | HNF-5 |
| | | + | TGGCAT | NF-1L |
| | + | + | GGGAAATCCC (CVF2) GAAATTCCC (coC3) | NF-κB |
| | | + | GGGAGGG | MAZ |
| + | + | + | AGAACA (CVF2, coC3) TGTTCT (CVF1) | GR-Element* |

*Halbsequenz des Palindroms AGAACANNNTGTTCT (*glucocorticoid response element*, GRE)

Terminationssignale des CVF1-Gens

Erwägungen zur Bedeutung der regulatorischen Elemente betreffen nicht nur den 5'-Bereich, sondern auch den 3'-untranslatierten Bereich der Gene von CVF1, CVF2 und coC3. Aus diesem Grunde wurden sowohl die Polyadenylierungssignale als auch eventuelle Sequenzmotive, die für korrekte Termination verantwortlich zeichnen könnten, detailliert untersucht.

Polyadenylierungssignale

Der 3'-untranslatierte Bereich von CVF1 auf der Ebene der cDNA konnte vollständig bestimmt werden. Hier war eine Zuordnung des klassischen Polyadenylierungssignals AATAAA möglich, das sich in CVF1 40 bp stromaufwärts der Poly-A-Sequenz in der reifen mRNA befindet, obwohl der 3'-UTR von CVF1 die Sequenz AATAAA insgesamt vierfach enthält. Dies weist darauf hin, dass diese Sequenz zwar ein notwendiges Signal für eine Polyadenylierung, aber nicht zugleich das Signal für die Termination der Transkription darstellt. So ist in allen drei Genen die Sequenz AATAAA bereits 141-145 bp stromabwärts des Stop-Codons an konservierter Position vorhanden, ohne jedoch als Polyadenylierungssignal zu fungieren. Bisher wurden zahlreiche Gene mit multiplen Poly-A-Signalen identifiziert, von denen einige funktionell aktiv sind (Edwards-Gilbert *et al.* 1997).

Bislang konnten erst wenige Sequenzen identifiziert werden, die den Terminations-Stop in Eukaryonten vermitteln. Dies ist vermutlich auf die Ungenauigkeit des Terminationsprozesses zurückzuführen, d.h. die Primär-Transkripte weisen neben dem stark konservierten Polyadenylierungssignal kaum weitere homogene 3'-Sequenzen auf.

Ein Konsensus-mRNA-Degradierungssignal in Form einer AT-reichen Sequenz (Shaw und Kamen 1986) direkt stromabwärts des Polyadenylierungssignals ist nicht vorhanden, was auf eine hohe Stabilität der CVF1-mRNA hinweist. Sequenzmotive wie YGTGTTY (McLauchlan *et al.* 1985), welche als Terminationssignale in anderen Genen identifiziert wurden, sind im Terminationsbereich des CVF1-Gens nicht vorhanden. Jedoch befinden sich drei CCAAT-Boxen, die ebenfalls als wichtige Signale für die Termination der RNA-Polymerase II fungieren (Connelly und Manley 1989), im CVF1-Gen unmittelbar stromaufwärts des Polyadenylierungssignals.

MAZ-Bindungsmotive

Als Terminationssignal für kurz aufeinanderfolgende Gene ist die Bindungsstelle GGGGGAGGGGG des Transkriptionsfaktors MAZ (*myc-associated zinc finger protein*) bekannt, die zuerst im humanen Komplementgen C2 als solche identifiziert wurde (Ashfield *et al.* 1994, Moreira *et al.* 1995). Da dieses Signal in den untersuchten flankierenden Sequenzen der CVF/C3-Genfamilie nicht vorhanden ist, kann ein eng benachbarter Gen-Cluster vermutlich ausgeschlossen werden.

Als Bindungsstelle für MAZ im Promotorbereich ist jedoch auch die verkürzte Sequenz GGGAGGG bekannt. Diese Sequenz ist im Terminationsbereich von CVF1 als dreifach verschachteltes Motiv enthalten und hat möglicherweise einen Einfluss auf die Termination.

Sekundärstrukturen

Als alternativer Ansatz zum besseren Verständnis der Termination im CVF1-Gen wird im folgenden die Ausbildung von charakteristischen RNA-Sekundärstrukturen diskutiert. In Prokaryonten sind die Terminationssignale durch komplementäre Sequenzen festgelegt, die zur Ausbildung von Haarnadelschleifen führen. Ähnliche Sequenzen, die eine RNA-Duplex ermöglichen, wurden im 3'-Bereich von CVF1 identifiziert. Die Duplex-Bereiche der in Abb. 4.5 aufgezeigten Sekundärstrukturen mit Längen von 5 bzw. 6 bp sind zwar unter thermodynamischen Aspekten nicht so stabil wie prokaryontische Terminatoren. So besteht beispielsweise die Haarnadelschleife der *trp*-mRNA aus *E. coli* aus einer 8 bp langen GC-reichen RNA-Duplex (Ryan und Chamberlin 1983). Dennoch könnte die Ausbildung von Sekundärstrukturen in der CVF1-mRNA möglicherweise als additiver Effekt zu einem Transkriptionsstopp durch die RNA-Polymerase II und anschließender Dissoziation der neuen RNA von der DNA-Matrize beitragen und somit erklären, warum drei weiter stromaufwärts gelegene AATAAA-Sequenzen im 3'-untranslatierten Bereich von CVF1 nicht als Polyadenylierungssignale fungieren. Bei Eukaryonten wurde die Ausbildung von Haarnadelstrukturen als spezifisches Terminationssignal auch bei Histon-mRNAs diskutiert, welchen das Poly-Adenylierungssignal fehlt (Krieg und Melton 1984). Denkbar ist auch, dass es sich bei den oben beschriebenen palindromen Sequenzen um Bindungsstellen für Transkriptionsfaktoren auf DNA-Ebene handelt.

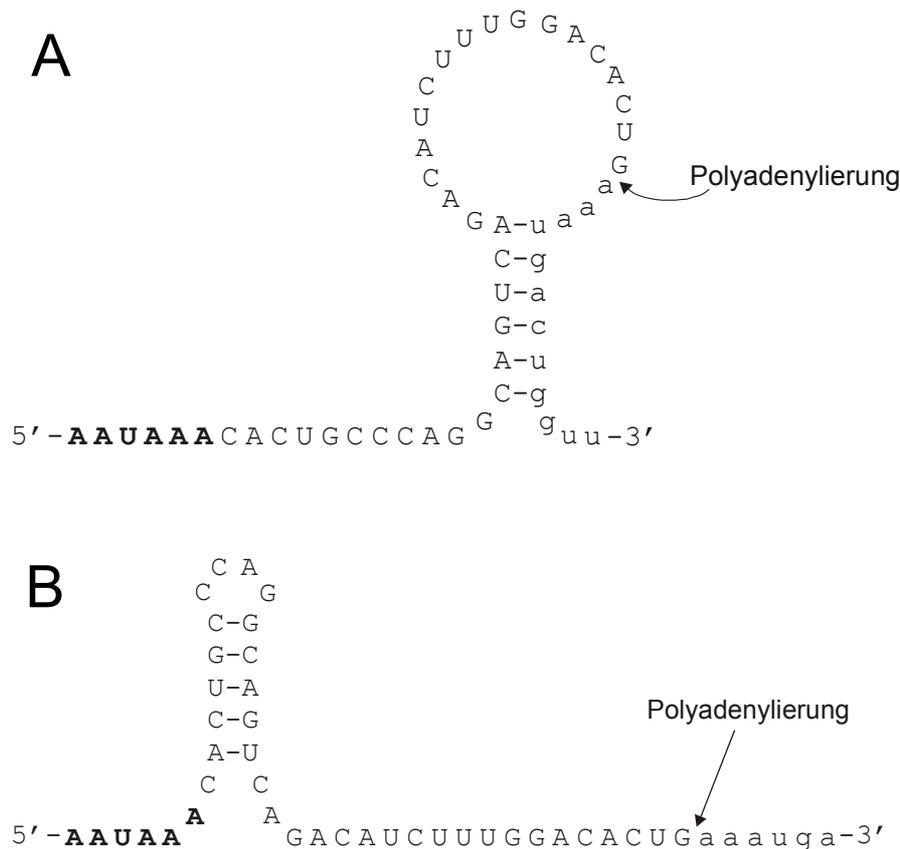


Abb. 4.5: Alternative Sekundärstrukturen als mögliche Terminationssignale der CVF1-mRNA.

Stromabwärts des Polyadenylierungssignals AAUAAA befinden sich Sequenzfolgen, die revers komplementär zueinander sind und durch Formation von RNA-Duplices besondere Sekundärstrukturen ausbilden könnten. Die Sequenz in Grossbuchstaben entspricht der CVF1-cDNA; in Kleinbuchstaben ist die entsprechende Gen-Sequenz stromabwärts davon dargestellt.

A: Pfannenstiel-Struktur. Ausbildung einer Stem-loop-Sekundärstruktur (Pfannenstiel). Der doppelsträngige Pfannenstiel besteht aus 6 Basenpaaren.

B: Haarnadelstruktur. Ausbildung einer Haarnadelstruktur. Die RNA-Duplex besteht aus 5 Basenpaaren.

Eine Analyse der 3'-UTRs von CVF1, CVF2 und coC3 ergab, dass die Sequenzen der drei Gene direkt stromabwärts des Stop-Codons stark homolog sind (ohne Abbildung). Neben dem Sequenzmotiv AATAAAA, das sich in allen drei Genen ca. 140 bp stromabwärts des Stop-Codons an konservierter Position befindet, enthält die coC3-cDNA-Sequenz stromabwärts noch ein weiteres AATAAAA-Motiv, das als funktionell aktives Polyadenylierungssignal für das coC3-Transkripts dient. Mit einer Länge von 240 bp ist die 3'-UTR von coC3 im Vergleich zum CVF1-Transkript (998 bp) relativ kurz, und die 3'-UTR von CVF2 (147 bp) liegt nur partiell vor.

Eine Aufklärung der terminalen Gensequenzen von CVF2 und coC3 könnte zeigen, ob auch in diesen Genen im Terminationsbereich die Ausbildung von Sekundärstrukturen wie im CVF1-Gen möglich wäre.

4.7 Hinweise auf weitere Mitglieder der CVF/C3-Genfamilie

Durch Genome Walking-Analyse sowie durch genomische PCR konnte am Beispiel von Exon 2 gezeigt werden, dass es deutliche Hinweise auf die Existenz von bis zu fünf Mitgliedern der CVF/C3-Genfamilie gibt (Abb. 3.80). Dabei wurde diejenige überlappende Sequenz dem CVF2-Gen zugeordnet, welche die größten Abweichungen von den entsprechenden CVF1- und coC3-Sequenzen aufwies, da die Exon 2-Sequenzen ebenfalls sehr heterogen sind. Bis zur Aufklärung der fehlenden CVF2-cDNA-Sequenz in diesem Bereich muss diese Zuordnung allerdings als vorläufig betrachtet werden.

Die zusätzlichen, spezifischen Sequenzen wurden als CVF-Homologe 1-5 bezeichnet und zeigen sowohl im Exon- als auch im Intronbereich eine weitgehende Übereinstimmung mit der CVF1-Sequenz. Da die PCR-basierten Nachweise allesamt mit der DNA einer einzigen Kobra-Leber als Template durchgeführt wurden, kann ausgeschlossen werden, dass es sich um Sequenzen unterschiedlicher Spezies bzw. Individuen handelt. Es bleibt allerdings die Frage offen, ob es sich bei der Vielzahl der Sequenzen in allen Fällen um vollständig transkribierte (und translatierte) Sequenzen oder möglicherweise um Pseudogen-Sequenzen handelt. Zudem ist das Vorhandensein allelischer Unterschiede denkbar.

Einen Hinweis auf mindestens 5 Mitglieder liefert auch Abbildung 3.35 (Genome Walking Exon 1 stromabwärts). Zwar wurden von den fünf distinkten Banden nur zwei hinsichtlich ihrer Sequenz charakterisiert, doch kann aufgrund der PCR-Bedingungen von einer hohen Spezifität der Banden ausgegangen werden. Möglicherweise werden nur die Gene CVF1, CVF2 und coC3 transkribiert, da nur für diese beim Screening von cDNA-Banken unterschiedliche Klone gefunden wurden (Fritzinger *et al.* 1994b). Dies ist im Einklang mit dem Befund, dass auch im Medakafisch von insgesamt 11 verschiedenen C3-Genen, die in einer BAC-Bibliothek identifiziert wurden, nur 3 Gene transkribiert werden (Nonaka *et al.* 2002). Von den 11 C3-Genen wurden fünf als verkürzte Pseudogene identifiziert.

4.8 Evolution und Phylogenie der CVF/C3-Genfamilie

Aufgrund der vorliegenden Daten, die überwiegend eine enge genetische Verwandtschaft von CVF2 und coC3 unterstreichen, kommt für die Phylogenie der CVF/C3-Genfamilie nur eine Möglichkeit in Betracht (Abb. 4.6). Demnach könnte CVF1 im Verlauf der Evolution durch eine frühe Genduplikation eines coC3-Vorläufergens entstanden sein. Das CVF2-Gen könnte aus einer weiteren, später erfolgten, coC3-Duplikation hervorgegangen sein. Die teilweise stark ausgeprägten Sequenzhomologien nur zwischen CVF1 und CVF2 können dadurch erklärt werden, dass seit den Genduplikationen durch einen höheren

Evolutionsdruck auf C3 als wesentliche Serumkomponente nur dieses Gen selektiv verändert wurde, während CVF1 und CVF2 in den entsprechenden Bereichen weitgehend unverändert blieben.

Aufgrund der Transkription sowohl von CVF1 als auch von CVF2 in der Giftdrüse ist die Vermutung naheliegend, dass CVF1 und CVF2 durch Duplikation eines gemeinsamen Vorläufers entstanden sein könnten. Hinweise dafür müssten auch in der Promotorregion zu finden sein, wo die Transkription reguliert wird, doch zeigt insbesondere der Promotorbereich eher Ähnlichkeiten von CVF2 mit coC3. Kritisch ist in diesem Zusammenhang jedoch zu bewerten, dass das Bindeglied zwischen dieser Promotorregion und der CVF2-cDNA-Sequenz (entsprechend ab Exon 9) noch fehlt und demzufolge unklar ist, ob es sich bei der mit CVF2 bezeichneten Promotorregion tatsächlich um den vermuteten CVF2-Genabschnitt handelt.

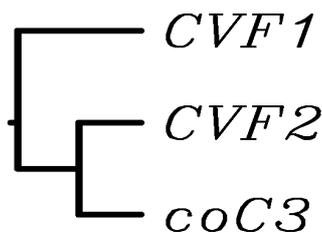


Abb. 4.6: Phylogenetischer Stammbaum der CVF/C3-Genfamilie. Dendogramm erstellt mit ClustalW.

Die in dieser Arbeit zusätzlich aufgeklärten homologen Gensequenzen zeigen die größte Übereinstimmung mit CVF1 und CVF2.

Vergleich der C3-Multicopy-Genfamilien verschiedener Spezies

Seit den ersten Veröffentlichungen über die funktionellen und strukturellen Gemeinsamkeiten von CVF und human C3 (Alper und Balavitch 1976, Vogel *et al.* 1984) war bekannt, dass es sich bei der Giftkomponente CVF um eine Isoform der Serumkomponente C3 handelt, die in der Folge aus Kobraserum aufgereinigt und charakterisiert wurde (Petrella *et al.* 1989). Nachdem auch in der Regenbogenforelle *Salmo gairdneri* das Vorkommen von zunächst zwei C3-Isoformen nachgewiesen werden konnte (Nonaka *et al.* 1985b), wurden multiple C3-Proteine oder -Gene auch in anderen Teleosten beschrieben (Gongara *et al.* 1998; Nakao und Yano 1998; Sunyer *et al.* 1996, 1997a, 1997b). Durch das Vorhandensein mehrerer C3-Isoformen entsteht eine größere strukturelle und funktionelle Diversität als bei höheren Vertebraten. Für die multiplen

Formen von C3 in diversen Teleost-Fischen wurde gezeigt, dass sie mit unterschiedlicher Effizienz an komplement-aktivierende Oberflächen binden. Möglicherweise wird durch diese Diversität das Spektrum der angeborenen Immunabwehr erweitert. Für die Klasse der Amphibien wurde am Beispiel des Krallenfroschs *Xenopus laevis* neben C3 auch das Vorkommen eines verkürzten C3-cDNA-Transkripts nachgewiesen (Lambris *et al.* 1995). Da allerdings noch ungeklärt ist, ob es sich dabei um das Produkt eines weiteren C3-Gens handelt, enthalten bisher nur Fische und die Kobra *Naja kaouthia* erwiesenermaßen multiple C3-Gene (Sunyer *et al.* 1998b).

Sowohl Teleosten als auch Amphibien und Reptilien zählen zu den poikilothermen (wechselwarmen) Vertebraten, die ähnlich wie homeotherme Vertebraten über ein gut entwickeltes Komplementsystem zu verfügen scheinen (Sunyer und Lambris 1998). Im Gegensatz zu Homeothermen, bei denen für die Klasse der Mammalia u.a. bei Maus (Wiebauer *et al.* 1982, Lundall *et al.* 1984) und Mensch (Vik *et al.* 1991) und auch für die Klasse der Vögel beim Huhn (Mavroidis *et al.* 1995) C3 als single copy-Gen nachgewiesen wurde, enthalten viele Spezies von Poikilothermen allerdings mehrere Formen der Komplement-Komponenten C3 und Faktor B (Sunyer *et al.* 1998a).

Einzig die Kobra enthält mit CVF1 und CVF2 zwei weitere Gene der C3-Familie, deren Produkte nicht im Serum zirkulieren, sondern nur in der Giftdrüse gebildet werden (Fritzinger *et al.* 1991, Fritzinger *et al.* 1994b). Der Nachweis für die Proteinexpression und Aktivität von CVF2 muss allerdings noch erbracht werden.

Die Funktion von CVF1 als nicht-toxische Gift-Komponente ist tatsächlich noch nicht vollständig geklärt. Wohl ist bekannt, dass durch die Aktivierung des alternativen Komplement-Wegs durch CVF1 im Beutetier ein Überschuss der Anaphylatoxine C3a und C5a entsteht, und dass die folgende Erhöhung der Permeabilität der Blutgefäße eine schnellere Wirkung der toxischen Komponenten (Cardiotoxine, Neurotoxine, Cytotoxine) verursacht, doch letztlich kann ein mit Vorhandensein von CVF1 und CVF2 im Gift verbundener evolutionärer Vorteil für die Kobra nur vermutet werden.

Bei der in dieser Arbeit gezeigten Besonderheit der Genstrukturen von C3 und seinen Homologen CVF1 und CVF2 in der Kobra, nämlich dem Fehlen von Intron 31 im Vergleich zu den homologen Mammalia-Genen, wurde vermutet, dass es sich um ein gemeinsames Merkmal aller Reptilien bzw. Poikilothermen handeln könnte. Für die im Rahmen der Genom-Forschungsprojekte analysierte Klasse der Fische (Modellorganismen Zebrafisch *Danio rerio*, Kugelfisch *Takifugu rubripes*, Medakafisch *Oryzias latipes*) sind

bis dato keine genomischen C3-Sequenzen im Bereich von Exon 31 und Exon 32 veröffentlicht. Vorläufige Untersuchungen (in Kooperation mit D.C. Fritzingler und M. Nonaka) zur C3-Struktur im Medakafisch weisen darauf hin, dass die Exons 31 und 32 durch ein Intron von ca. 420 bp Länge unterbrochen sind. Möglicherweise ist daher die partiell intronfreie C3-Struktur auf Kobras bzw. auf die Klasse der Reptilien begrenzt.

Beim Vergleich der CVF/C3-Gene mit dem huC3-Gen im Bereich der Exon 31/Exon 32-Grenze fällt auf, dass die ansonsten teilweise stark ausgeprägte Sequenzhomologie im Bereich der Exongrenzen hier fehlt. Obwohl die Varianz der Spleiss-Donor-Sequenzen im CVF1-Gen groß ist (vgl. Tab. 3.5), überwiegt doch die Sequenzfolge *AG*, die an den Exon-Spleissstellen den Intronsequenzen vorausgeht. Im Gegensatz dazu ist in der CVF/C3-Genfamilie an der postulierten Spleissstelle für Intron 31 die Sequenz *ACC* vorhanden, die ansonsten in keinem der in Tabelle 4.1 aufgeführten Komplementgene als Spleiss-Donor-Sequenz vorhanden ist¹. Stattdessen wird die Sequenz *ACC* aber als Spleiss-Akzeptor für Exon 31 des CVF1-Gens gefunden. Es ist theoretisch möglich, dass in den „ursprünglichen“ CVF/C3-Genen die Sequenzen im Bereich Exon 31/Exon 32 so beschaffen sind, dass sie zunächst nicht als Spleissstelle für Intron 31 fungieren können. Erst durch die Veränderung dieser genomischen Sequenz im Verlauf der Evolution könnte sich eine Proto-Spleissstelle gebildet haben, wie sie heute bei den höher entwickelten Mammalia vorgefunden wird.

Im Zusammenhang mit dem Entstehen der CVF/C3-Multigenfamilie ist von Interesse, ob sich diese Gen-Vervielfältigung und -Variation im Kobra-Genom im Rahmen einer Tetraploidisierung ereignet hat, wie sie für frühe Vertebraten diskutiert wird (Sankoff 2001). Jedoch haben nicht alle Organismen mit multiplen C3-Isoformen ein tetraploides Genom. So wurde die bisher höchste C3-Diversität mit fünf Isoformen im diploiden Fisch *Sparus aurata* identifiziert (Sunyer *et al.* 1997a, 1997b). Ferner gibt es keine Hinweise darauf, dass Kobras im speziellen und Reptilien im allgemeinen ein tetraploides Genom enthalten (W. Wüster, persönliche Mitteilung), sondern höchstwahrscheinlich sind sie diploid wie alle höheren Vertebraten. Es ist daher denkbar, dass die CVF/C3-

¹ Für die humane C3-Genstruktur ist zwar in Exon 27 die Sequenz *ACC* als Spleiss-Donor-Sequenz beschrieben (Vik *et al.* 1991), doch findet sich sowohl in der huC3-cDNA-Sequenz (de Bruijn und Fey 1985, Genbank-Nr. NM_000064.1) als auch in einem genomischen Klon (Genbank-Nr. AC008760.7) an derselben Position die Sequenz *AAC*.

Im C3-1B-Gen des Medakafisch würden die Spleiss-Donor/Akzeptor-Sequenzen um Intron 31 ebenfalls nicht *ACC* lauten.

Multigenfamilie der Kobra *Naja kaouthia* durch lokale Duplikation entstanden ist, wie sie bei den humanen C4A und C4B-Genen und deren murinen Homologen, C4 und Slp, beschrieben ist (Yang und Yu 2000, Blanchong *et al.* 2001).

4.9 Ansatz zum Nachweis des CVF2-Proteins

Obwohl der Nachweis von CVF2 auf genomischer Ebene und auf mRNA-Basis erfolgreich geführt werden konnte, ist eine Identifizierung und Charakterisierung des CVF2-Proteins bislang noch nicht gelungen. Unter der Annahme, dass CVF2 tatsächlich exprimiert wird und in der Giftdrüse als aktives Glykoprotein vorliegt, ist es denkbar, dass die CVF1- und CVF2-Proteine in ihrem molekularen Aufbau so ähnlich sind, dass eine Trennung der Proteine nicht möglich war. Bei dem in drei Stufen gereinigten CVF-Protein (Vogel und Müller-Eberhard 1984) konnten Inhomogenitäten zweifelsfrei auf eine CVF1-spezifische γ -Ketten-Heterogenität zurückgeführt werden. Neben diesen konnten keinerlei Hinweise auf die Präsenz sekundärer Giftproteine erhalten werden, wobei unklar bleibt, inwieweit eine Detektion und Sequenzierung geringster Quantitäten von CVF2 apparativ durchführbar gewesen wäre.

Bei einem kürzlich beschriebenen neuen komplement-verbrauchenden Faktor aus dem Gift der Kobra *Naja kaouthia* (Sun *et al.* 2001) handelt es sich höchstwahrscheinlich nicht um CVF2, da die N-terminalen Aminosäuresequenzen des dreikettigen Proteins weder mit CVF1 noch mit der abgeleiteten CVF2-Sequenz übereinstimmen.

Ganz allgemein stellt sich bei der Protein- und Genomanalyse von Schlangen das Problem der korrekten Bestimmung der Spezies. Obwohl die auf dem indischen Subkontinent beheimateten Kobra-Spezies *Naja kaouthia* (Monokelschlange) und *Naja naja* (Brillenschlange) aufgrund ihres phänotypischen Aussehens unterscheidbar sein sollten, ist eine Verwechslung nicht ausgeschlossen.

Eine Erklärung dafür, warum CVF2 als Protein bisher nicht nachgewiesen werden konnte, ist möglicherweise in den Promotorsequenzen von CVF1, CVF2 und coC3 begründet. Wie in Abbildung 4.1 verdeutlicht, sind unmittelbar stromaufwärts von "TATA3" keine Homologien zwischen CVF1 und CVF2 erkennbar. Das widerspricht den Erwartungen, da für beide Gene Transkripte in der Giftdrüse nachweisbar waren (Fritzinger *et al.* 1991). Aufgrund der Sequenzunterschiede im Bereich der Bindungsstellen für Transkriptionsfaktoren könnte es sich allerdings bei CVF2 im Gegensatz zu CVF1 um ein schwach

transkribiertes Gen handeln, dessen Expressionsrate um ein Vielfaches niedriger als die von CVF1 liegt.

Als Möglichkeiten eines Nachweises des CVF2-Proteins bieten sich MALDI-Analyse bzw. Detektion mit CVF2-spezifischen Antikörpern an.

Auf Proteinebene könnte der Nachweis von mehreren CVF-Isoformen eventuell durch 2D-Gelelektrophorese möglich sein. Eine nachfolgende Sequenzanalyse/Massenbestimmung durch MALDI wäre direkt im Anschluss an 2D-Gelelektrophorese möglich. Desweiteren könnten durch MALDI-Analysen von fragmentierten Proteinbanden mögliche Massenunterschiede detektiert werden, wodurch aufgrund der bekannten DNA-Sequenz Rückschlüsse auf die enthaltenen Proteine denkbar wären.

Prinzipiell könnte die Identifizierung und Reinigung von CVF2 unter Verwendung CVF2-spezifischer Antikörper durchgeführt werden. Monoklonale CVF2-spezifische Antikörper böten die Möglichkeit des Nachweises von CVF2 durch Immunoprint sowie der Reinigung mittels Affinitätschromatographie.

Für die Generierung solcher Antikörper wurde eine CVF2-Region ausgewählt und exprimiert, deren abgeleitete Aminosäuren-Sequenz eine ausgeprägte Differenz zur entsprechenden CVF1-Sequenz aufweist. Diese Region wird auf genomischer Ebene durch Exons 18 und 19 codiert, welche im CVF1-Gen den N-Terminus der γ -Kette codieren.

Kritisch bei der Auswahl der CVF2-Peptidsequenz ist die unbekannte Kettenstruktur des putativen CVF2-Proteins. Da im nativen CVF1 die „C3a“- und „C3d“-Regionen des pro-Proteins nicht mehr enthalten sind, wird vermutlich auch das CVF2-Protein in bisher unbekannter Weise prozessiert. Die „C3a“-Region in CVF1 liegt unmittelbar stromaufwärts der γ -Kette. Da die Spezifität der für die Prozessierung verantwortlichen Proteasen nicht bekannt ist und die ausgewählte CVF2-Peptidsequenz besonders im N-Terminus ausgeprägte Differenzen zur CVF1-Sequenz aufweist, ist eine Relevanz dieser Region für die Prozessierung nicht ausgeschlossen.

Einen Hinweis auf das Vorhandensein von CVF2 als Gift-Komponente der Kobraspezies *Naja kaouthia* könnte der Nachweis von Kohlenhydratresten in allen drei Proteinketten liefern. Da in der γ -Kette von CVF1 im Gegensatz zur CVF2-Sequenz keine N-Glykosylierungsstelle vorhanden ist (siehe Abb. 4.7), könnte es sich bei der potentiell

glykosylierten γ -Kette von CVF2 um ein geeignetes Kriterium zum Nachweis von CVF2 und zur Unterscheidung von CVF1 mittels Lektin-Blots handeln.

| | | |
|------|-----------------------------------|------|
| | S D I I S | |
| CVF1 | AGTGATATCATCTCA | 2244 |
| CVF2 | GGT A ACATCA C CTCA | 1292 |
| | G N I T S | |
| | CHO | |

Abb. 4.7: Potentielle N-Glykosylierungsstelle von CVF2 im Bereich der CVF1- γ -Kette (Exon 18). Die homologen DNA- und Protein-Sequenzen von CVF1 und CVF2 sind einander gegenübergestellt. Die putative Glykosylierungsstelle von Asparagin (N) in der CVF2-Sequenz ist mit CHO markiert.

4.10 Ausblick

In nachfolgenden Untersuchungen könnte durch Southern-Blot-Analyse von genomischer Kobra-DNA geprüft werden, ob und wieviele homologe Gene zusätzlich zu CVF1, CVF2 und coC3 im Genom der Kobra *Naja kaouthia* existieren. Promotor-Deletionsstudien und DNase I-Footprinting können dazu beitragen, die für die Regulation wichtigen Elemente zu identifizieren. Möglicherweise kann dadurch die molekulare Ursache für die niedrige Expressionsrate von CVF2 identifiziert werden.

Die in dieser Arbeit identifizierten 5'-Sequenzen des CVF2-Gens könnten genutzt werden, um durch RT-PCR die CVF2-Sequenz zu vervollständigen. Für die Detektion, Reinigung und Charakterisierung von CVF2 wären prinzipiell CVF2-spezifische monoklonale Antikörper geeignet, für deren Selektion im Rahmen dieser Arbeit durch die Expression eines CVF2-Peptids die Grundlagen gelegt wurden. Eine Trennung von mehreren CVF-Isoformen aus Kobragift könnte eventuell durch Affinitätschromatographie mit einem monoklonalen Antikörper gelingen, welcher die Isoformen mit unterschiedlicher Affinität bindet. Zunächst müsste geklärt werden, ob CVF2 tatsächlich translatiert wird und als Protein im Kobra-Gift vorhanden ist.

Das CVF2-Protein könnte rekombinant exprimiert und in seiner Aktivität mit CVF1 verglichen werden. Struktur-/Funktionsstudien mit beiden Proteinen können dazu beitragen, die für die Stabilität der CVF-abhängigen C3-Konvertase wichtigen Regionen zu identifizieren. Nachweis und Charakterisierung eines potentiell aktiven CVF2-Proteins könnten für die Identifizierung von Bindungsstellen von anderen Komplement-Komponenten sowie für die Generierung von rekombinanten Hybridmolekülen zum therapeutischen Einsatz von CVF im Menschen von Interesse sein.

Da die Kobra unter den Vertebraten eine relativ niedrige Entwicklungsstufe einnimmt, kann ihr Komplementsystem als einfaches Modellsystem für höhere Vertebraten wie Menschen betrachtet werden. Detailliertere Analysen bezüglich der Diversität und der Regulation könnten dann Auskunft geben über evolutive Mechanismen zur Entwicklung komplexerer Interaktionen sowie der zugehörigen genetischen Grundlagen. Die Anwendung dieser Mechanismen könnte für eine gezielte Einflußnahme auf das Komplementsystem und damit für eine verbesserte Therapie genutzt werden.

5 Literatur

- Affiyan F, Armugam A, Tan CH, Gopalakrishnakone P, Jeyaseelan K.** Postsynaptic α -neurotoxin gene of the spitting cobra, *Naja naja sputatrix*: structure, organization, and phylogenetic analysis. *Genome Res.* 9, 259-266 (1999).
- Alper CA, Balavitch D.** Cobra venom factor: Evidence for its being altered cobra C3 (the third component of complement). *Science* 191, 1275-1276 (1976).
- Al-Sharif WZ, Sunyer JO, Lambris JD, Smith LC.** Sea urchin coelomocytes specifically express a homologue of the complement component C3. *J. Immunol.* 160, 2983-2997 (1998).
- Andrä J, Halter R, Kock M, Niemann H, Vogel CW, Paul D.** Generation and characterization of transgenic mice expressing cobra venom factor. *Mol. Immunol.* 39, 357-365 (2002)
- Ashfield R, Patel AJ, Bossone SA, Brown H, Campbell RD, Marcu KB, Proudfoot NJ.** MAZ-dependent termination between closely spaced human complement genes. *EMBO J.* 13, 5656-5667 (1994).
- Baird M, Driscoll C, Schreiner H, Sciarratta GV, Sansone G, Niazi G, Ramirez F, Bank A.** A nucleotide change at a splice junction in the human beta-globin gene is associated with beta 0-thalassemia. *Proc. Natl. Acad. Sci. USA* 78, 4218-4221 (1981).
- Barnes WM.** PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91, 2216-2220 (1994).
- Bates PF, Swift RA.** Double *cos* site vectors: Simplified cosmid cloning. *Gene* 26, 315-321 (1983).
- Belt KT, Carroll MC, Porter RR.** The structural basis of the multiple forms of human complement component C4. *Cell* 36, 907-914 (1984).
- Benoist P, Mas JA, Marco R, Cervera M.** Differential muscle-type expression of the *Drosophila* troponin T gene. A 3-base pair microexon is involved in visceral and adult hypodermic muscle specification. *J. Biol. Chem.* 273, 7538-7546 (1998).
- Berget SM.** Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411-2414 (1995).
- Birnboim HC, Doly J.** A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7, 1513-1514 (1979).
- Birren B, Green EC, Klapholz S, Myers RM, Riethman H, Roskams J** (Hrsg.). *Genome Analysis: A laboratory manual*, Vol. 3 Cosmides, 87-201. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1997).

- Blanchong CA, Chung EK, Rupert KL, Yang Y, Yang Z, Zhou B, Moulds JM, Yu CY.** Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Sfp and C4. *Int. Immunopharmacol.* 1, 365-392 (2001).
- Bossone SA, Asselin C, Patel AJ, Marcu KB.** MAZ, a zinc finger protein, binds to *c-MYC* and *C2* gene sequences regulating transcriptional initiation and termination. *Proc. Natl. Acad. Sci. USA.* 89, 7452-7456 (1992).
- Botto M, Fong KY, So AK, Rudge A, Walport MJ.** Molecular basis of hereditary C3 deficiency. *J. Clin. Invest.* 86, 1158-1163 (1990).
- Breathnach R, Chambon P.** Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50, 349-383 (1981).
- Burke DT, Carle GF, Olson MV.** Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236, 806-812 (1987).
- Busslinger M, Moschonas N, Flavell RA.** Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell* 27, 289-298 (1981).
- Campbell RD, Law SKA, Reid KBM, Sim RB.** Structure, organization and regulation of the complement genes. *Ann. Rev. Immunol.* 6, 161-195 (1988).
- Carney DF, Haviland DL, Noack D, Wetsel RA, Vik DP, Tack BF.** Structural aspects of the human C5 gene. *J. Biol. Chem.* 266, 18786-18791 (1991).
- Chang LS, Lin J, Wu PF.** cDNA sequence analysis of cardiotoxin variants from Taiwan cobra. *Biochem. Mol. Biol. Int.* 42, 85-92 (1997).
- Cheng S, Fockler C, Barnes WM, Higuchi R.** Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* 91, 5695-5699 (1994).
- Cheung AK, Parker CJ, Wilcox L.** Effects of two types of cobra venom factor on porcine complement activation and pulmonary artery pressure. *Clin. Exp. Immunol.* 78, 299-306 (1989).
- Chodosh LA, Baldwin AS, Carthew RW, Sharp PA.** Human CCAAT-binding proteins have heterologous subunits. *Cell* 53, 11-24 (1988).
- Chung CT, Niemela SL, Miller RH.** One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* 86, 2172-2175 (1989).
- Church GM, Gilbert W.** Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81, 1991-1995 (1984).
- Clark JM.** Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* 16, 9677-86 (1988).

- Cline J, Braman JC, Hogrefe HH.** PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* 24, 3546-3551 (1996).
- Cochrane CG, Müller-Eberhard HJ, Aikin BS.** Depletion of plasma complement in vivo by a protein of cobra venom: its effect on various immunologic reactions. *J. Immunol.* 105, 55-69 (1970).
- Coffey AJ, Roberts RG, Green ED, Cole CG, Butler R, Anand R, Giannelli F, Bentley DR.** Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach. *Genomics* 12, 474-484 (1992).
- Collins J, Hohn B.** Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc. Natl. Acad. Sci. USA* 75, 4242-4246 (1978).
- Connelly S, Manley JL.** RNA polymerase II transcription termination is mediated specifically by protein binding to a CCAAT box sequence. *Mol. Cell. Biol.* 9, 5245-5259 (1989).
- Daha MR, Fearon DT, Austen KF.** C3 requirements for formation of alternative pathway C5 convertase. *J. Immunol.* 117, 630-634 (1976).
- Dahl MR, Thiel S, Matsushita M, Fujita T, Willis AC, Christensen T, Vorup-Jensen T, Jensenius JC.** MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway. *Immunity* 15, 127-135. (2001).
- Dangel AW, Mendoza AR, Baker BJ, Daniel CM, Carroll MC, Wu LC, Yu CY.** The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 40, 425-436 (1994).
- D'Aquila RT, Bechtel LJ, Videler JA, Eron JJ, Gorczyca P, Kaplan JC.** Maximizing sensitivity and specificity of PCR by pre-amplification heating. *Nucleic Acids Res.* 19, 3749 (1991).
- de Bruijn MH, Fey GH.** Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl. Acad. Sci. USA* 82, 708-112 (1985).
- Dibb NJ, Newman AJ.** Evidence that introns arose at proto-splice sites. *EMBO J.* 8, 2015-2021 (1989).
- DiLella AG, Woo SL.** Cloning large segments of genomic DNA using cosmid vectors. *Methods Enzymol.* 152, 199-212 (1987).
- Dodds AW, Law SK.** The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and alpha 2-macroglobulin. *Immunol. Rev.* 166, 15-26 (1998).
- Dodds AW, Smith SL, Levine RP, Willis AC.** Isolation and initial characterisation of complement components C3 and C4 of the nurse shark and the channel catfish. *Dev. Comp. Immunol.* 22, 207-216 (1998).

- Don RH, Cox PT, Wainright BJ, Baker K, Mattick JS.** 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 19, 4008 (1991).
- Dower WJ, Miller JF, Ragsdale CW.** High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* 16, 6127-6145 (1988).
- Edwalds-Gilbert G, Veraldi KL, Milcarek C.** Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* 25, 2547-2561 (1997).
- Eggertsen G, Lind P, Sjoquist J.** Molecular characterization of the complement activating protein in the venom of the Indian cobra (*Naja n. siamensis*). *Mol. Immunol.* 18, 125-133 (1981).
- Ehrlich P, Morgenroth J.** Zur Theorie der Lysinwirkung. *Berl. Klin. Wochschr.* 36, 6-9 (1899).
- Evans GA, Lewis K, Rothenberg BE.** High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* 79, 9-20 (1989).
- Fernie BA, Orren A, Wurzner R, Jones AM, Potter PC, Lachmann PJ, Hobart MJ.** Complement component C6 and C7 haplotypes associated with deficiencies of C6. *Ann. Hum. Genet.* 59, 183-195 (1995).
- Flexner S, Noguchi H.** Snake venom in relation to haemolysis, bacteriolysis and toxicity. *J. Exp. Med.* 6, 277-301 (1903).
- Fong KY, Botto M, Walport MJ, So AK.** Genomic organization of human complement component C3. *Genomics* 7, 579-586 (1990).
- Frank MM, Fries LF.** Role of complement in inflammation and phagocytosis. *Immunol. Today* 12, 322-327 (1991).
- Fritzinger DC, Bredehorst R, Vogel CW.** Sequence analysis of cobra venom factor mRNA reveals the existence of two different mRNA species. *Complement Inflamm.* 8, 152 (Abstract) (1991).
- Fritzinger DC, Bredehorst R, Vogel CW.** Complete sequence of two different cobra venom factor cDNAs. *FASEB J.* 6, A2998 (Abstract) (1992a).
- Fritzinger DC, Bredehorst R, Vogel CW.** DNA encoding cobra C3, CVF1, and CVF2. *International Patent*, Publication Number WO 94/23024 (1994a).
- Fritzinger DC, Connelly M, Petrella EC, Bredehorst R, Vogel CW.** Primary structure of cobra complement component C3. *J. Immunol.* 149, 3554-3562 (1992b).
- Fritzinger DC, Bredehorst R, Vogel CW.** Molecular cloning and derived primary structure of cobra venom factor. *Proc. Natl. Acad. Sci. USA* 91, 12775-12779 (1994b), und Erratum *Proc. Natl. Acad. Sci. USA* 92, 7605 (1995).

- Fujita T.** Evolution of the lectin-complement pathway and its role in innate immunity. *Nat. Rev. Immunol.* 2, 346-353 (2002).
- Furth ME, Wickner SH.** Lambda DNA replication. In: *Lambda II* (Ed. RW Hendrix *et al.*), pp. 145-173. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1983).
- Gil A, Proudfoot NJ.** A sequence downstream of AAUAAA is required for rabbit beta-globulin mRNA 3'-end formation. *Nature* 312, 473-474 (1984).
- Gil A, Proudfoot NJ.** Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globulin mRNA 3'-end formation. *Cell* 49, 399-406 (1987).
- Gongora R, Figueroa F, Klein J.** Independent duplications of Bf and C3 complement genes in the zebrafish. *Scand. J. Immunol.* 48, 651-658 (1998).
- Gowda DC, Glushka J, Halbeek Hv, Thotakura RN, Bredehorst R, Vogel CW.** N-linked oligosaccharides of cobra venom factor contain novel alpha(1-3)galactosylated Le(x) structures. *Glycobiology* 11, 195-208 (2001).
- Gowda DC, Petrella EC, Raj TT, Bredehorst R, Vogel CW.** Immunoreactivity and function of oligosaccharides in cobra venom factor. *J. Immunol.* 152, 2977-2986 (1994).
- Grange T, Roux J, Rigaud G, Pictet R.** Cell-type specific activity of two glucocorticoid responsive units of rat tyrosine aminotransferase gene is associated with multiple binding sites for C/EBP and a novel liver-specific nuclear factor. *Nucleic Acids Res.* 19, 131-139 (1991).
- Green ED, Olson MV.** Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* 87, 1213-1217 (1990).
- Gros-Bellard M, Oudet P, Chambon P.** Isolation of high-molecular-weight DNA from mammalian cells. *Eur. J. Biochem.* 36, 32-38 (1973).
- Grunstein M, Hogness DS.** Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* 72, 3961-3965 (1975).
- Hanahan D, Jessee J, Bloom FR.** Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol.* 204, 63-113 (1991).
- Haviland DL, Haviland JC, Fleischer DT, Hunt A, Wetsel RA.** Complete cDNA sequence of human complement pro-C5. Evidence of truncated transcripts derived from a single copy gene. *J. Immunol.* 146, 362-368 (1991a).
- Haviland DL, Haviland JC, Fleischer DJ, Wetsel RA.** Structure of the murine fifth complement component (C5) gene. *J. Biol. Chem.* 266, 11818-11825 (1991b).
- Hentschel CC, Birnstiel ML.** The organization and expression of histone gene families. *Cell* 25, 301-313 (1981).

- Hoheisel JD, Maier E, Mott R, McCarthy L, Grigoriev AV, Schalkwyk LC, Nizetic D, Francis F, Lehrach H.** High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* 73, 109-120 (1993).
- Hourcade D, Holers VM, Atkinson JP.** The regulators of complement activation (RCA) gene cluster. *Adv. Immunol.* 45, 381-416 (1989).
- Hu G.** DNA polymerase-catalyzed addition of non-templated extra nucleotides to the 3' end of a DNA fragment. *DNA Cell Biol.* 12, 763-770 (1993).
- Hughes AL.** Phylogeny of the C3/C4/C5 complement-component gene family indicates that C5 diverged first. *Mol. Biol. Evol.* 11, 417-425 (1994).
- Imagawa S, Izumi T, Miura Y.** Positive and negative regulation of the erythropoietin gene. *J. Biol. Chem.* 269, 9038-9044 (1994).
- Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ.** A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* 6, 84-89 (1994).
- Iseli C, Stevenson BJ, de Souza SJ, Samaia HB, Camargo AA, Buetow KH, Strausberg RL, Simpson AJ, Bucher P, Jongeneel CV.** Long-range heterogeneity at the 3' ends of human mRNAs. *Genome Res.* 12, 1068-1074 (2002).
- Isenman DE, Kells DI, Cooper NR, Muller-Eberhard HJ, Pangburn MK.** Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b-like functional properties. *Biochemistry* 20, 4458-4467 (1981).
- Isenman DE, Young JR.** The molecular basis for the difference in immune hemolysis activity of the Chido and Rodgers isotypes of human complement component C4. *J. Immunol.* 132, 3019-3027 (1984).
- Ji X, Azumi K, Sasaki M, Nonaka M.** Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian, *Halocynthia roretzi*. *Proc. Natl. Acad. Sci. USA* 94, 6340-6345 (1997).
- Johnson CA, Densen P, Hurford RK Jr, Colten HR, Wetsel RA.** Type I human complement C2 deficiency. A 28-base pair gene deletion causes skipping of exon 6 during RNA splicing. *J. Biol. Chem.* 267, 9347-9353 (1992) und Erratum in: *J. Biol. Chem.* 268, 2268 (1993).
- Juhl H, Sievers M, Baltzer K, Helmig F, Wolf H, Brenner W, Kalthoff H.** A monoclonal antibody-cobra venom factor conjugate increases the tumor-specific uptake of a ^{99m}Tc-labeled anti-carcinoembryonic antigen antibody by a two-step approach. *Cancer Res.* 55, 5749-5755 (1995).
- Kato Y, Salter-Cid L, Flajnik MF, Namikawa C, Sasaki M, Nonaka M.** Duplication of the MHC-linked *Xenopus* complement factor B gene. *Immunogenetics* 42, 196-203 (1995).

- Kellogg DE, Rybalkin I, Chen S, Mukhamedova N, Vlasik T, Siebert P, Chenchik A.** TaqStart Antibody: Hotstart PCR facilitated by a neutralizing monoclonal antibody directed against *Taq* DNA polymerase. *BioTechniques* 16, 1134-1137 (1994).
- Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI.** Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids. Res.* 20, 1083-1085 (1992).
- Kinoshita T.** Biology of complement: the overture. *Immunol. Today* 12, 291-295 (1991).
- Kock MA.** Expression and characterization of recombinant cobra venom factor. Dissertation, Universität Hamburg, Fachbereich Chemie, Wissenschaft und Technik Verlag (1996).
- Kohara Y, Akiyama K, Isono K.** The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* 50, 495-508 (1987).
- Kozak M.** An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15, 8125-8132 (1987).
- Kozak M.** Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283-292 (1986).
- Krieg PA, Melton DA.** Formation of the 3' end of histone mRNA by post-transcriptional processing. *Nature* 308, 203-206 (1984).
- Kuroda N, Naruse K, Shima A, Nonaka M, Sasaki M, Nonaka M.** Molecular cloning and linkage analysis of complement C3 and C4 genes of the Japanese medaka fish. *Immunogenetics* 51, 117-128 (2000).
- Lachumanan R, Armugam A, Tan CH, Jeyaseelan K.** Structure and organization of the cardiotoxin genes in *Naja naja sputatrix*. *FEBS Lett.* 433, 119-124 (1998).
- Laemmli UK.** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685 (1970).
- Lambris JD, Pappas J, Mavroidis M, Wang Y, Manzone H, Schwager J, Pasquier LD, Silibovsky R.** The third component of *Xenopus* complement: cDNA cloning, structural and functional analysis, and evidence for an alternate C3 transcript. *Eur. J. Immunol.* 25, 572-578 (1995) und Erratum *Eur. J. Immunol.* 25, 1485 (1995).
- Larhammar D, Risinger C.** Why so few pseudogenes in tetraploid species? *Trends Genet.* 10, 418-9 (1994).
- Leonard LA, Strandberg JD, Winkelstein JA.** Complement-like activity in the sea star, *Asterias forbesi*. *Dev. Comp. Immunol.* 14, 19-30 (1990).
- Leventhal JR, Dalmasso AP, Cromwell JW, Platt JL, Manivel CJ, Bolman RM 3rd, Matas AJ.** Prolongation of cardiac xenograft survival by depletion of complement. *Transplantation* 55, 857-866 (1993).

- Li Y.** Screening einer Cosmid-Genbibliothek der indischen Kobra *Naja naja kaouthia* nach Klonen des Cobra Venom Factor (CVF) sowie Identifizierung eines CVF-Klons. *Diplomarbeit*, Fachbereich Chemie, Universität Hamburg (2000).
- Lin SR, Chang LS, Chang KL.** Separation and structure-function studies of Taiwan cobra cardiotoxins. *J. Protein Chem.* 21, 81-86 (2002).
- Lundwall Å, Wetsel RA, Domdey H, Tack BF, Fey GH.** Structure of murine complement component C3. I. Nucleotide sequence of cloned complementary and genomic DNA coding for the β chain. *J. Biol. Chem.* 259, 13851-13856 (1984).
- Ma D, Armugam A, Jeyaseelan K.** Expression of cardiotoxin-2 gene. Cloning, characterization and deletion analysis of the promoter. *Eur. J. Biochem.* 268, 1844-1850 (2001).
- Mandel JL.** Dystrophin. The gene and its product. *Nature* 339, 584-586 (1989).
- Mandel M, Higa A.** Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* 53, 159-162 (1970).
- Marchuck D, Drum M, Saulino A, Collins FS.** Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* 19, 1154 (1991).
- Marino R, Kimura Y, De Santis R, Lambris JD, Pinto MR.** Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian *Ciona intestinalis*. *Immunogenetics* 53, 1055-1064 (2002).
- Matthijs G, Devriendt K, Cassiman JJ, van den Berghe H, Marynen P.** Structure of the human alpha-2 macroglobulin gene and its promoter. *Biochem. Biophys. Res. Commun.* 184, 596-603 (1992).
- Mavrodis M, Sunyer JO, Lambris JD.** Isolation, primary structure, and evolution of the third component of chicken complement and evidence for a new member of the alpha 2-macroglobulin family. *J. Immunol.* 154, 2164-2174 (1995).
- McLauchlan J, Gaffney D, Whitton JL, Clements JB.** The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* 13, 1347-1368 (1985).
- Medicus RG, Götze O, Müller-Eberhard HJ.** Alternative pathway of complement: recruitment of precursor properdin by the labile C3/C5 convertase and the potentiation of the pathway. *J. Exp. Med.* 144, 1076-1093 (1976).
- Moreira A, Wollerton M, Monks J, Proudfoot NJ.** Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *EMBO J.* 14, 3809-3819 (1995).
- Moriyama EN, Gojobori T.** Evolution of nested genes with special reference to cuticle proteins in *Drosophila melanogaster*. *J. Mol. Evol.* 28, 391-397 (1989).

- Mount SM.** A catalogue of splice junction sequences. *Nucleic Acids Res.* 10, 459-472 (1982).
- Müller-Eberhard HJ.** Molecular organization and function of the complement system. *Annu. Rev. Biochem.* 57, 321-347 (1988).
- Mullis KB, Faloona FA.** Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155, 335-350 (1987).
- Nakao M, Yano T.** Structural and functional identification of complement components of the bony fish, carp (*Cyprinus carpio*). *Immunol. Rev.* 166, 27-38 (1998).
- Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF.** Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 17, 2503-2516 (1989).
- Nonaka M, Kuroda N, Naruse K, Shima A.** Molecular genetics of the complement C3 convertases in lower vertebrates. *Immunol. Rev.* 166, 59-65 (1998).
- Nonaka M, Nakayama K, Yeul YD, Takahashi M.** Complete nucleotide and derived amino acid sequences of the fourth component of mouse complement (C4): Evolutionary aspects. *J. Biol. Chem.* 260, 10936-10943 (1985 a).
- Nonaka M, Nonaka M, Irie M, Tanabe K, Kaidoh T, Natsuume-Sakai S, Takahashi M.** Identification and characterization of a variant of the third component of complement (C3) in rainbow trout (*Salmo gairdneri*) serum. *J. Biol. Chem.* 260, 809-815 (1985 b).
- Nonaka M, Umetsu D, Matsuo MY.** High degree C3 amplification in a bony fish, medaka. *Int. Immunopharm.* 2, 1291-1292 (2002) (Abstract).
- Ogata RT, Rosa PA, Zepf NE.** Sequence of the gene for murine complement component C4. *J. Biol. Chem.* 264, 16565-16572 (1989).
- Ohno M, Menez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC, Tamiya T, Menez A.** Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? *Prog. Nucleic Acid Res. Mol. Biol.* 59, 307-364 Review (1998).
- Orkin SH, Goff SC, Hechtman RL.** Mutation in an intervening sequence splice junction in man. *Proc. Natl. Acad. Sci. USA* 78, 5041-5045 (1981).
- Pangburn MK, Müller-Eberhard HJ.** The alternative pathway of complement. *Springer Semin. Immunopathol.* 7, 163-192 (1984).
- Patthy L.** Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* 214, 1-7 (1987).
- Petrella EC, Bredehorst R, Vogel CW.** Purification of cobra C3: initial characterization and comparison to cobra venom factor. *Complement Inflamm.* 6, 386 (Abstract) (1989).

- Petrella EC, Wilkie SD, Smith CA, Morgan AC Jr, Vogel CW.** Antibody conjugates with cobra venom factor. Synthesis and biochemical characterization. *J. Immunol. Methods* 104, 159-172 (1987).
- Pierce JC, Sauer B, Sternberg N.** A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy. *Proc. Natl. Acad. Sci. USA* 89, 2056-2060 (1992).
- Prediger EA, Hoffman S, Edelman GM, Cunningham BA.** Four exons encode a 93-base-pair insert in three neural cell adhesion molecule mRNAs specific for chicken heart and skeletal muscle. *Proc. Natl. Acad. Sci. USA* 85, 9616-9620 (1988).
- Reilly JD, Silva RF.** A double-*cos*-site vector containing a multiple cloning site flanked by T7 and T3 RNA polymerase promoters. *Gene* 124, 135-136 (1993).
- Rich A, Nordheim A, Wang AJH.** The chemistry and biology of left-handed Z-DNA. *Ann. Rev. Biochem.* 53, 791-846 (1984).
- Riles L, Dutchik JE, Baktha A, McCauley BK, Thayer EC, Leckie MP, Braden VV, Depke JE, Olson MV.** Physical maps of the six smallest chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6 kilobase pairs. *Genetics* 134, 81-150 (1993).
- Risinger C, Larhammar D.** Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. *Proc. Natl. Acad. Sci. USA* 90, 10598-10602 (1993).
- Roberts RG, Coffey AJ, Bobrow M, Bentley DR.** Exon structure of the human dystrophin gene. *Genomics* 16, 536-538 (1993).
- Rossi V, Cseh S, Bally I, Thielens NM, Jensenius JC, Arlaud GJ.** Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and -2. *J. Biol. Chem.* 276, 40880-40887 (2001).
- Ryan AF, Catanzaro A, Wasserman SI, Harris JP, Vogel CW.** The effect of complement depletion on immunologically mediated middle ear effusion and inflammation. *Clin. Immunol. Immunopathol.* 40, 410-421 (1986).
- Ryan T, Chamberlin MJ.** Transcription analyses with heteroduplex trp attenuator templates indicate that the transcript stem and loop structure serves as the termination signal. *J. Biol. Chem.* 258, 4690-4693 (1983).
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA.** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491 (1988).
- Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA.** Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324, 163-166 (1986).
- Saitou N, Nei M.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425 (1987).

- Sambrook J, Fritsch EF, Maniatis T.** *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
- Sanger F, Nicklen S, Coulson AR.** DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463-5467 (1977).
- Sankoff D.** Gene and genome duplication. *Curr. Opin. Genet. Dev.* 11, 681-684 (2001).
- Sarkar G, Kapelner S, Sommer SS.** Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* 18, 7465 (1990).
- Schneider PM, Rittner C.** Complement genetics. In: *Complement, a practical approach* (Ed. AW Dodds, RB Sim) Oxford University Press, Oxford, New York, Tokyo (1997).
- Schumaker VN, Zavodszky P, Poon PH.** Activation of the first component of complement. *Annu. Rev. Immunol.* 5:21-42 (1987).
- Schwaeble W, Dahl MR, Thiel S, Stover C, Jensenius JC.** The mannan-binding lectin-associated serine proteases (MASPs) and MASP19: four components of the lectin pathway activation complex encoded by two genes. *Immunobiology* 205, 455-466 (2002).
- Schweizer-Groyer G, Groyer A, Cadepond F, Grange T, Baulieu EE, Pictet R.** Two liver-enriched trans-acting factors support the tissue-specific basal transcription from the rat tyrosine aminotransferase promoter. *J. Steroid Biochem. Mol. Biol.* 41, 747-752 (1992).
- Sepich DS, Noonan DJ, Ogata RT.** Complete cDNA sequence of the fourth component of murine complement. *Proc. Natl. Acad. Sci. USA* 82, 5895-5899 (1985).
- Sharkey DJ, Scalice ER, Christy KG Jr, Atwood SM, Daiss JL.** Antibodies as thermolabile switches: high temperature triggering for the polymerase chain reaction. *Biotechnology* 12, 506-509 (1994).
- Shaw G, Kamen R.** A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46, 659-667 (1986).
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M.** Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* 89, 8794-8797 (1992).
- Shuman S.** Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* 269, 32678-32684 (1994).
- Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA.** An improved method for walking in uncloned genomic DNA. *Nucleic Acids Res.* 23, 1087-1088 (1995).

- Sim RB, Kolble K, McAleer MA, Dominguez O, Dee VM.** Genetics and deficiencies of the soluble regulatory proteins of the complement system. *Int. Rev. Immunol.* 10, 65-86 (1993).
- Sim RB, Reid KB.** C1: molecular interactions with activating systems. *Immunol. Today* 12, 307-311 (1991).
- Sottrup-Jensen L, Stepanik TM, Kristensen T, Lonblad PB, Jones CM, Wierzbicki DM, Magnusson S, Domdey H, Wetsel RA, Lundwall A, Tack BF, Fey GH.** Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. *Proc. Natl. Acad. Sci. USA* 82, 9-13 (1985).
- Southern EM.** Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503-517 (1975).
- Sun QY, Lu QM, Wang WY, Xiong YL.** A highly active anticomplement factor from the venom of *Naja kaouthia*. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* (Shanghai) 33, 483-488 (2001).
- Sunyer JO, Lambris JD.** Evolution and diversity of the complement system of poikilothermic vertebrates. *Immunol. Rev.* 166, 39-57 (1998).
- Sunyer JO, Tort L, Lambris JD.** Structural C3 diversity in fish: characterization of five forms of C3 in the diploid fish *Sparus aurata*. *J. Immunol.* 158, 2813-2821 (1997a).
- Sunyer JO, Tort L, Lambris JD.** Diversity of the third form of complement, C3, in fish: functional characterization of five forms of C3 in the diploid fish *Sparus aurata*. *Biochem. J.* 326, 877-881 (1997b).
- Sunyer JO, Zarkadis IK, Sahu A, Lambris JD.** Multiple forms of complement C3 in trout that differ in binding to complement activators. *Proc. Natl. Acad. Sci. USA* 93, 8546-8551 (1996).
- Sunyer JO, Zarkadis I, Sarrias MR, Hansen JD, Lambris JD.** Cloning, structure, and function of two rainbow trout Bf molecules. *J. Immunol.* 161, 4106-4114 (1998a).
- Sunyer JO, Zarkadis IK, Lambris JD.** Complement diversity: a mechanism for generating immune diversity? *Immunology Today* 19, 519-523 (1998b).
- Tack BF, Harrison RA, Janatova J, Thomas ML, Prah JW.** Evidence for presence of an internal thiolester bond in third component of human complement. *Proc. Natl. Acad. Sci. USA* 77, 5764-5768 (1980).
- Tassabehji M, Strachan T, Anderson M, Campbell RD, Collier S, Lako M.** Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. *Nucleic Acids Res.* 22, 5211-5217 (1994).

- Thiel S, Petersen SV, Vorup-Jensen T, Matsushita M, Fujita T, Stover CM.** Interaction of C1q and mannan-binding lectin (MBL) with C1r, C1s, MBL-associated serine proteases 1 and 2, and the MBL-associated protein MAb19. *J. Immunol.* 165, 878-887 (2000).
- Thompson C.** Protein proves to be a key link in innate immunity. *Science* 269, 301-302 (1995).
- Thompson JD, Higgins DG, Gibson TJ.** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680 (1994).
- Tilghman SM, Tiemeier DC, Seidman JG, Peterlin BM, Sullivan M, Maizel JV, Leder P.** Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. *Proc. Natl. Acad. Sci. USA* 75, 725-729 (1978).
- Vik DP, Amiguet P, Moffat GJ, Fey M, Amiguet-Barras F, Wetsel RA, Tack BF.** Structural features of the human C3 gene: intron/exon organization, transcriptional start site, and promoter region sequence. *Biochemistry* 30, 1080-1085 (1991).
- Vogel CW.** (Ed.) Immunoconjugates. Antibody conjugates in radioimaging and therapy of cancer. Oxford University Press, New York (1987).
- Vogel CW, Bredehorst R, Fritzinger DC, Grunwald T, Ziegelmüller P, Kock MA.** Structure and function of cobra venom factor, the complement-activating protein in cobra venom. In: Natural toxins II (Ed.: Singh BR und Tu AT), pp. 97-113, Plenum Press, New York (1996).
- Vogel CW, Müller-Eberhard HJ.** Induction of immune cytolysis: tumor-cell killing by complement is initiated by covalent complex of monoclonal antibody and stable C3/C5 convertase. *Proc. Natl. Acad. Sci. USA* 78, 7707-7711 (1981).
- Vogel CW, Müller-Eberhard HJ.** The cobra venom factor-dependent C3 convertase of human complement. A kinetic and thermodynamic analysis of a protease acting on its natural high molecular weight substrate. *J. Biol. Chem.* 257, 8292-8299 (1982).
- Vogel CW, Müller-Eberhard HJ.** Cobra venom factor: improved method for purification and biochemical characterization. *J. Immunol. Meth.* 73, 203-220 (1984).
- Vogel CW, Smith CA, Müller-Eberhard HJ.** Cobra venom factor: structural homology with the third component of human complement. *J. Immunol.* 133, 3235-3241 (1984).
- Vogelstein B, Gillespie D.** Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* 76, 615-619 (1979).
- Vogt W, Schmidt G.** Abtrennung des anaphylatoxinbildenden Prinzips aus Cobragift von anderen Giftkomponenten. *Experientia* 20, 207-208 (1964).
- Volonakis JE.** Transcriptional regulation of complement genes. *Annu. Rev. Immunol.* 13, 277-305 (1995).

- Vorup-Jensen T, Jensenius JC, Thiel S.** MASP-2, the C3 convertase generating protease of the MBLectin complement activating pathway. *Immunobiology* 199, 348-357 (1998).
- Vyse TJ, Bates GP, Walport MJ, Morley BJ.** The organization of the human complement factor I gene (IF): a member of the serine protease gene family. *Genomics* 24, 90-98 (1994).
- Waterston R, Sulston J.** The genome of *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci USA* 92, 10836-10840 (1995).
- Weatherall DJ.** Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.* 2, 245-255 (2001).
- Wetsel RA, Fleischer DT, Haviland DL.** Deficiency of the murine fifth complement component (C5): A 2- base pair gene deletion in a 5'-exon. *J. Biol. Chem.* 265, 2435-2440 (1990).
- Wetsel RA, Lemons RS, Le Beau MM, Barnum SR, Noack D, Tack BF.** Molecular analysis of human complement component C5: localization of the structural gene to chromosome 9. *Biochemistry* 27, 1474-1482 (1988).
- Wiebauer K, Domdey H, Diggelmann H, Fey G.** Isolation and analysis of genomic DNA clones encoding the third component of mouse complement. *Proc. Natl. Acad. Sci. USA* 79, 7077-7081 (1982).
- Wüster W.** Taxonomic changes and toxinology: systematic revisions of the Asiatic cobras (*Naja naja* species complex). *Toxicon* 34, 399-406 (1996).
- Yang Z, Yu CY.** Organizations and gene duplications of the human and mouse MHC complement gene clusters. *Exp. Clin. Immunogenet.* 17, 1-17 (2000).
- Yonaha M, Proudfoot NJ.** Transcriptional termination and coupled polyadenylation in vitro. *EMBO J.* 19, 3770-3777 (2000).
- Yu CY.** The complete exon-intron structure of a human complement component C4A gene. DNA sequences, polymorphism, and linkage to the 21-hydroxylase gene. *J. Immunol.* 146, 1057-1066 (1991).
- Yu CY, Belt KT, Giles CM, Campbell RD, Porter RR.** Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity. *EMBO J.* 5, 2873-2881 (1986).

Zusammenfassung

Cobra Venom Factor (CVF), das komplement-aktivierende Protein aus dem Gift der Monokel-Kobra *Naja kaouthia*, ist strukturell und funktionell homolog zur Komplement-Komponente C3. Ähnlich wie C3b, die aktive Form von C3, bildet CVF mit Faktor B eine C3-Konvertase. In vorangegangenen Arbeiten wurde beim Screening von cDNA-Bibliotheken aus Giftdrüsen-mRNA die Sequenz eines weiteren CVF-Moleküls, genannt CVF2, identifiziert. Die cDNA-Sequenzen von CVF1, CVF2 und C3 sind ca. 92% identisch.

In der vorliegenden Arbeit wurde die Struktur des CVF1-Gens aufgeklärt und mit der des humanen C3-Gens verglichen. Für die Aufklärung der Genstruktur wurden eine genomische Cosmid-Bibliothek, genomische PCR und Genome Walking eingesetzt.

Die sich aus den erhaltenen Daten ergebende Struktur des vollständigen CVF1-Gens umfaßt 40 Exons und 39 Introns. Die Exongrößen variieren von 46 bp bis 1140 bp, und Introngrößen variieren von 100 bp bis ca. 9000 bp. Es konnte gezeigt werden, dass das CVF1-Gen mit einer Mindestgröße von 89 kb mehr als doppelt so groß wie das humane C3-Gen (42 kb) ist. Dennoch sind die Strukturen beider Gene in hohem Ausmaß konserviert. Mehr als die Hälfte der einander entsprechenden Exons zeigen identische Größen. Unterschiede in den Exongrößen variieren meist von 3 bp bis 15 bp und repräsentieren stets das Vielfache eines Codons. Dabei sind die Intronphasen nahezu identisch. Sequenzhomologien bei den Introns wurden nur am äußersten Randbereich festgestellt. Ein bedeutsamer Unterschied jedoch besteht darin, dass die konservierten Exons 31 und 32 im CVF1-Gen kontinuierlich und nicht durch ein Intron unterbrochen sind, was im Gegensatz zu den Strukturen der humanen C3, C4 und C5-Gene steht.

In den darüber hinaus untersuchten Strukturen der Gene von CVF2 und coC3 sind die Exons 31 und 32 ebenfalls kontinuierlich und nicht durch ein Intron unterbrochen. Es zeigte sich, dass die Exongrößen der CVF/C3-Genfamilie stark konserviert sind. Als Ursache für einen Größenunterschied von 9 bp in Exon 33 des CVF1-Gens im Vergleich zu den entsprechenden Exons in den CVF2- und coC3-Genen konnte eine singuläre Mutation in der Intron-Spleißstelle identifiziert werden. Obwohl die Introngrößen der drei Gene teilweise stark variieren, sind ausgeprägte Sequenzhomologien auch an internen Positionen vorhanden.

Zusätzlich wurden die Promotorregionen der drei homologen Gene näher analysiert, um mögliche Ursachen für die Expression in unterschiedlichen Geweben zu finden. Putative TATA-Boxen wurden in allen drei Genen an konservierten Positionen gefunden. Es konnte

gezeigt werden, dass sich die stromaufwärts gelegenen Regionen allerdings in den möglichen Bindungsstellen für Transkriptionsfaktoren unterscheiden.

Einige Ergebnisse wiesen darauf hin, dass das Kobra-Genom zusätzlich zu den bereits bekannten Genen CVF1, CVF2 und coC3 möglicherweise noch zwei weitere homologe Gene bzw. Pseudogene enthält. Diese Ergebnisse sind im Einklang mit multiplen C3-Genen in Teleost-Fischen und multiplen Toxin-Genen in verschiedenen *Naja*-Spezies.

Summary

Cobra Venom Factor (CVF), the complement-activating protein in the venom of the monocle cobra (*Naja kaouthia*), is a structural and functional analog of complement component C3. Like C3b, the active form of C3, CVF can form a C3 convertase with factor B. The screening of cDNA libraries prepared from cobra venom gland mRNA revealed the presence of a second cobra venom factor, named CVF2. The cDNA sequences of CVF (now named CVF1), CVF2, and cobra C3 are approximately 92% identical.

As an approach to understand the molecular basis for the structural and functional homology and difference between CVF1 and C3, the CVF1 gene structure was elucidated using a genomic cosmid library, genomic PCR, and genome walking. The CVF1 gene consists of 40 exons and 39 introns. Exon sizes vary from 46 bp to 1,140 bp, and introns vary from 100 bp to 9,000 bp. Whereas the CVF1 gene is with greater than 89 kb more than twice as large as the human C3 gene (42 kb), the structures of both genes exhibit a very high degree of conservation. More than half of the corresponding exons show identical sizes. Minor variations in exon sizes of 3 bp to 15 bp represent multiples of a codon. Intron phases are almost identical. In contrast, intron sequences and sizes show no homology between the CVF1 and human C3 genes. The most important difference is that the conserved exons 31 and 32 are contiguous in CVF1 and not separated by an intron, which is in contrast to the human C3, C4, and C5 gene structures.

As is the case for the CVF1 gene, exons 31 and 32 in the CVF2 and cobra C3 genes are also not separated by an intron. Intron sizes vary between the three genes. Exon 33 of the CVF1 gene is 9 bp smaller than the corresponding exons of the CVF2 and cobra C3 genes. The exon size difference is caused by a single nucleotide substitution in the intron splice site rather than by deletion.

Moreover, the promoter regions of the three homologous genes were analyzed in more detail to find the molecular basis for expression in different tissues. Putative TATA boxes were found in all three genes at conserved positions. However, the upstream regions of the three genes differ in putative binding sites for transcription factors. In addition to the genes for CVF1, CVF2, and cobra C3, some results indicate that the cobra genome may contain two more homologous genes or pseudogenes. These findings are consistent with multiple C3 genes in teleost fish and multiple toxin genes in various *Naja* species.

Anhang 7.1: Sequenzen des CVF1-Gens

Alle nachfolgenden Sequenzen sind in 5'-3'-Richtung dargestellt. Exonsequenzen sind in fetten Großbuchstaben und Intronsequenzen in Kleinbuchstaben dargestellt. Die Primer-Sequenzen sind mit Pfeilen unterlegt. Die für Genome Walking relevanten Schnittstellen (DraI, EcoRV, PvuII, StuI), EcoRI-Schnittstellen sowie charakteristische repetitive Sequenzen sind markiert. Die Nummerierung der Sequenzen ist fortlaufend.

Die 5'- und 3'- flankierenden Sequenzen des CVF1-Gens sind bereits vollständig im Ergebnisteil aufgeführt.

Intron 1

Dargestellt ist die Konsensus-Sequenz aus dem genomischen PCR-Amplifikat (ca. 6 kb) und dem Genome Walking-Amplifikat der DraI-Bibliothek (ca. 1050 bp). Die DraI-Schnittstelle im 5'-Bereich sowie häufige Wiederholungen der Dinukleotide ac und ta im 3'-Bereich sind unterstrichen. Die EcoRI-Analyse ergab Fragmente von 4,6 kb und 1,4 kb. Vorwärtsprimer: HB1s-CVF, Rückwärtsprimer: HB2as-II-CVF.

| | | | | | | |
|------|---------------------------|---------------------|-------------------|-------------------|--------------------|------|
| CVF1 | GGTGGCTGCT | CTATTGATTG | GTTTTCCAGG | GTCTTCTCAT | GGGGCTCT gt | 50 |
| CVF1 | aagtatcata | gatatttcag | cagggggaat | gaagcagtgc | tggggaagcg | 100 |
| CVF1 | tagtgggctt | tctcttccgt | gatggtatgt | attaaacgga | actatcacat | 150 |
| CVF1 | ccaaggctgt | ggtcctttgc | agggtaatgg | atctccttgc | aggyagctcc | 200 |
| CVF1 | aaacactttg | gagctttttc | agaaagtcca | atataatcca | cgacaggagt | 250 |
| CVF1 | aaaagtatta | taaccatcaa | tagggggaac | tgacaggccc | ttaagttcta | 300 |
| CVF1 | agatgaggaa | ttgtccttga | agtgaagatt | ctctggacaa | atatctcttg | 350 |
| CVF1 | ttctcagtta | tggtctgggg | gaaggggtgc | atctaaatta | ttaccattgc | 400 |
| CVF1 | tacgcatact | atctctgatg | catcaacaac | ttccagaaaa | caataagtt | 450 |
| CVF1 | gaatatctag | aaaacagcca | tcccacattc | tttgtttctg | atgtcttcaa | 500 |
| CVF1 | gtaaaccttg | gccatcaggg | taaggcttca | gtgacaaaaga | gactagtcta | 550 |
| CVF1 | cctgcttggt | ttgagaaaat | atctgtttca | ttgtgcacac | ctatagcaga | 600 |
| CVF1 | tatgactttg | cagcaggcat | gcagtatgtg | taagtaaatg | acgtatgagt | 650 |
| CVF1 | gatagcagca | acgaggctga | ctatatttct | ctgggtcccag | tatggagaat | 700 |
| CVF1 | cagctactga | aaaagcaaaa | gacaggagaa | ggaggaggca | gagaaggtgg | 750 |
| CVF1 | tttgtcagta | cacatataac | tcattgcagg | aaagaaaagg | aagagagaga | 800 |
| CVF1 | cgaacaaatc | tcaaaaactat | gctagatggg | gattttgtta | gggacttggc | 850 |
| CVF1 | atggaatcca | gattgattta | ttgcttttct | tttttttcca | aaaagtttt | 900 |
| CVF1 | tttattttta | gacaaacata | cacaaacaaa | aaatacaatg | tgctggtagg | 950 |
| CVF1 | ttacaaattt | ttttgtgcat | cttctccata | atcctcaatt | atgattattt | 1000 |
| CVF1 | atataatattc | tcatattgtt | atgtgaaata | ttccatattt | ttaaa..... | 1045 |
| | (ca. 4500 bp) | | | | | |
| CVF1 |ccccacaat | gtttgtttcc | agagagtaag | ccattttttgt | | 40 |
| CVF1 | atgaagtttg | gttgaaattg | cttgaggctt | tacagagtta | tgctggaaca | 90 |
| CVF1 | <u>cacacacaca</u> | <u>cacacacgga</u> | <u>gccatttata</u> | <u>tatatatata</u> | <u>tatatataga</u> | 140 |
| CVF1 | aaacagcacc | ccaaatccct | tttaagtgat | ataaagataa | agatttacta | 190 |
| CVF1 | atcttccaac | tgggttttga | agctttctaa | gtcacaaaag | agtagttctt | 240 |
| CVF1 | gtaactatct | atctcttcac | ttttctaaaa | gaagagatta | tggagatgcc | 290 |
| CVF1 | tcatgaaatg | cctcactttt | cctaagcaaa | ctaactcagg | agatgttctt | 340 |
| CVF1 | ttcttctctt | gcag CTACAC | CCTCATCACC | CCTGCTGTTT | TGCGAACAGA | 390 |
| CVF1 | CACAGAAGAG | CAAATTTTGG | TGGAGGCCCA | TGGAGACAGT | ACTCCAAAAC | 440 |
| CVF1 | AGCTTGACAT | CTTTGTTTTCAT | GATTTTCCAC | GGAAGCAGAA | AACCTTGTTT | 490 |
| CVF1 | CAAACCAGAG | TAGATATGAA | TCCAGCAGGA | GGCATGCTTG | TCACTCCAAC | 540 |

An Position 146 der Intronsequenz (y = c oder t) wurde im genomischen PCR-Amplifikat ein t und im Genome Walking-Amplifikat ein c gefunden.

Intron 2*Exon 2 stromabwärts*

Die StuI-Schnittstelle am 3'-Ende ist unterstrichen dargestellt. Äußerer Vorwärtsprimer: HB2s-CVF, innerer Vorwärtsprimer: HB2s-II.

```

CVF1  CCCATGGAGA CAGTACTCCA AAACAGCTTG ACATCTTTGT TCATGATTTT  50
CVF1  CCACGGAAGC AGAAAACCTT GTTCCAAACC AGAGTAGATA TGAATCCAGC 100
CVF1  AGGAGGCATG CTTGTCACTC CAACTATAGA Ggtgggtgta tggatgatggt 150
CVF1  tgcaatagag tctatgggat ccctgcaaaa caaaaaacca agctaaaaatc 200
CVF1  ctgtccagtc cggcttgagg ggaaatccta tgcactttag tcaaacatga 250
CVF1  ctaaattgtc agatattgag agaaactgtg catataattc tagggtggaa 300
CVF1  aggaaaatag agttatgggt tccatataac tctatacgtt tttagttact 350
CVF1  gttcgtatct ctacaattgc ttgaaagact ctgggcagca tttcacagca 400
CVF1  acaaaataaa tatttagatc aaaatcggta ttaaaaacaa atcaaccata 450
CVF1  gaaagaaatg aaaaacaaat ctgacagata caaagtatga taaaattatg 500
CVF1  atgattaata aaacaaaatt gctgagaacc agattgcttg agatactgaa 550
CVF1  tcagtctaaa tttagtgggt agtaaaagta aaaaagttag taaagaagtt 600
CVF1  agtaaaagta ggaaggcct 619

```

Exon 3 stromaufwärts

Dargestellt ist die partielle Sequenz eines Amplifikats der PvuII-Bibliothek (1,3 kb). Eine StuI-Schnittstelle (aggcct) ist unterstrichen dargestellt. Äußerer Rückwärtsprimer: HB3as-CVF, innerer Rückwärtsprimer: HB3as-II-CVF.

```

CVF1  .....cttgggggtgt gggttccctg actctccctc  30
CVF1  ccacttgttt tgagtttaga atttggtttt gaattttcctt ggcattgtaa  80
CVF1  ttaggagggg aaagtggcag ctatagtaag gccagggaga agaaggtgag 130
CVF1  gtttataggt gtactgcagt ggacaaaggg caccaaagt gatttgtaat 180
CVF1  agaagagtga taatgtgaga tctcaggaat atattctgca ttagtttcaa 230
CVF1  tgactgggct gctttaaactg gacactatat ggtctttcca ttgccttctg 280
CVF1  ccagaaaaga tcaaaaactg tcttaataga agcaagcaac acatttttaa 330
CVF1  ggtaagccta tgtttggcat tctgtaatcc tctttgtttt catttttgac 380
CVF1  tcaggagaaa tttttggatc aatgttttag tgttgttggc tagggctgag 430
CVF1  gaatgaggcc tcaagtggcc tttagtgatg attaaagtgg aaacctctgc 480
CVF1  tctactcacc ttgtcatoca gATTCCAGCA AAAGAAGTGA GTACGGACTC 530
CVF1  CAGGCAAAAT CAATATGTGG TTGTGCAAGT AACTGGTCCT CAAGTGAGAT 571

```

Intron 3

Vorwärtsprimer: HB3s-CVF, Rückwärtsprimer: HB4as-CVF.

| | | | | | | |
|------|-------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | CAAGTAACTG | GTCCTCAAGT | GAGATTGGAA | AAGGTGGTTC | TCCTTTCTTA | 50 |
| CVF1 | CCAGAGTAGC | TTTCTGTTTA | TCCAGACAGA | TAAAGGCATC | TATACACCAG | 100 |
| CVF1 | GGTCTCCAGg | taaccgctt | tttctgcctc | tatttcttca | ctcccctggg | 150 |
| CVF1 | gtcccactag | ggttatgtga | tgaataatca | aaaccattcc | tgggatttgt | 200 |
| CVF1 | cacttattcg | agaaacttcc | agagagtggg | ggaatgtatg | tcttgtgagt | 250 |
| CVF1 | gttggcattg | gtgatattga | gaaattttag | gaagttcaga | aaagatttta | 300 |
| CVF1 | gaaacacatt | gatagatggt | tcagtggaat | gtaatcaca | attaggaagg | 350 |
| CVF1 | gaacggagaa | gtccaagaag | atgtcacctc | gactaattgc | aagttataaa | 400 |
| CVF1 | agcgggaaa | gccttcaggg | ttatgtgagt | gatctttttt | tctaaaaaaa | 450 |
| CVF1 | aacaaccaat | tattattatg | cttctttaag | agaagaagaa | aatttgtcat | 500 |
| CVF1 | cttgaatgtc | ttttcaagaa | aggccttttt | aaaaaagatt | ttctgaagaa | 550 |
| CVF1 | ttaggtagtg | cttttctctc | aggtacagaa | gatccaggct | gaagtaatct | 600 |
| CVF1 | agaaacattt | atgctgtaat | aaaccatttg | gcctttaaag | tgcttgagga | 650 |
| CVF1 | tgctgctggt | ttggccaaat | ggaccaatga | gctgctttct | agaaactgtc | 700 |
| CVF1 | ccatatctgc | tttttaggat | tttggaaatc | ggcaaacagc | ttcagaatat | 750 |
| CVF1 | cagacaatta | atacaaaact | gtttttttta | caatgcaagc | tcatatttac | 800 |
| CVF1 | tgaattttac | cacaaattac | agttagtctc | ttctttctta | gTACTCTATC | 850 |
| CVF1 | GTGTTTTTTC | TATGGATCAC | AACACAAGCA | AGATGAACAA | AACTGTGATT | 900 |
| CVF1 | G | | | | | 901 |

Intron 4

Der Größenunterschied zu Exon 5 des coC3-Gens ist durch eine Lücke markiert. Vorwärtsprimer: HB4s-CVF, Rückwärtsprimer: HB5as-CVF.

| | | | | | | |
|------|-------------------|--------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | GTTTTTCTA | TGGATCACAA | CACAAGCAAG | ATGAACAAA | CTGTGATTGT | 50 |
| CVF1 | TGAGTTTCAG | gtgagaatcg | tgtctgcaat | cctgtgtgat | tggatagtt | 100 |
| CVF1 | ccctgaacaa | ggaatggttg | ttaattcctt | tttctcatgc | aatcatgcca | 150 |
| CVF1 | ggtagaccga | accagccac | tctatgaagc | tttggttggc | caacatttta | 200 |
| CVF1 | tatgtgcagt | ttttcttaaa | tgttttattg | taagttctgt | gctaaatatt | 250 |
| CVF1 | aggctgtatg | aaagatttaa | aaagtttcac | tatgaaattg | gaaaataatg | 300 |
| CVF1 | ccaagttga | atgtatgat | ttgtgtgtg | gcatgtatat | ggtggagata | 350 |
| CVF1 | ggagctaatt | gtttaattcc | tgttctgtca | gcagtaagcc | ttttttcata | 400 |
| CVF1 | tgaccatgt | ttatcttgcc | aaaaagggtc | taggggtggc | aatgagcaaa | 450 |
| CVF1 | tgttctattt | tcaatgcaa | ttagtttaa | ttatctgttt | tgtattatca | 500 |
| CVF1 | atcgacattt | gcatntagat | gaaaattcaa | ataatagtct | tacatcttat | 550 |
| CVF1 | atcccagtg | cccattctgct | gggttgcaat | tgctctctat | attatgtatg | 600 |
| CVF1 | tcatcacatg | taaaatctaa | tatttcagat | ggatttcaat | catatagaat | 650 |
| CVF1 | gcataggaaa | taacatgcaa | tgttctatat | gttttcaaca | cacctgtatt | 700 |
| CVF1 | ttatttctct | atcccag ACT | CCAGAAGGCA | TTCTTGTCAG | TTCTAATTCA | 750 |
| CVF1 | GTTGACCTAA | A---CTTCTT | CTGGCCTTAC | AATTTACC | | 785 |

Intron 5

Unterschiede zu den Exongrößen der coC3-Sequenz in Exon 5 und Exon 6 sind durch eine Lücke markiert. Auffällig ist ein 12facher G-Stretch (unterstrichen dargestellt). Vorwärtsprimer: HB5s-CVF, Rückwärtsprimer: HB6as-CVF.

| | | | | | | |
|------|-------------------|--------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | TCTTGTCAGT | TCTAATTCAG | TTGACCTAAA | ---CTTCTTC | TGGCCTTACA | 47 |
| CVF1 | ATTACCAGA | CCTTGTCAGg | taatttaaca | tggatattgt | cataagacaa | 97 |
| CVF1 | cctgttttga | aaccaactg | aaacgggggg | ggggggcatt | tttctattaa | 147 |
| CVF1 | agattggtta | aagaaggcca | tatagcattt | ttccttttag | cttacttttc | 197 |
| CVF1 | taatcaatta | gtaaagaaag | aaaaagtcca | ggtttgatat | tgtgtatttg | 247 |
| CVF1 | tccatcacat | ttctatttcc | tcttgttttt | tggatagact | ggagcagcac | 297 |
| CVF1 | agggtgggct | ataatgttat | ttgtatgaca | attttgtgaa | gttgaatgga | 347 |
| CVF1 | taaaataata | gatacaacct | catcaccacc | atttaagctc | aatgagtaag | 397 |
| CVF1 | aggaaattta | atacagaagt | aagagtcaca | tttcagcttc | cacaaatgtg | 447 |
| CVF1 | aagatttttg | aatgaggcat | tttcttctc | tcttataact | attctccaca | 497 |
| CVF1 | agacaagtaa | acctatcaat | atcaggaaga | agccttggat | ctcttgttca | 547 |
| CVF1 | tatattttatt | gtcataaatc | ggttataaag | ccaaagtaat | gaaaaatatt | 597 |
| CVF1 | ttcattttct | tgctgtag TT | TGGGGACTTG | GAGGATTGTG | GCCAAATATG | 647 |
| CVF1 | AACATTCCCC | AGA---GAAT | TATACTGCAT | ATTTTG | | 680 |

Intron 6

Auffällig sind die 12fach repetitive Sequenz aus cttct, eine fünffache gtt-Wiederholung, ein dreifacher agg-Repeat, sowie eine at-reiche Region. Ein Größenunterschied zu Exon 6 der coC3-Sequenz ist durch eine Lücke markiert. Vorwärtsprimer: HB6s, Rückwärtsprimer: HB7as.

```

CVF1 GACTTGGAGG ATTGTGGCCA AATATGAACA TCCCCAGA- --GAATTATA 50
CVF1 CTGCATATTT TGATGTCAGG AAATATGgta agaaggatta tttattgtct 100
CVF1 tgtcttctct tctcttctct tctcttctct tctcttctct tctcttctct 150
CVF1 tctcttctct tctttccact ccctgatcct gggtttctcc ccattccatt 200
CVF1 ttatgtttaa aaaacctagt tctaacattt gtagaagtgg aaatggttgc 250
CVF1 tatgatcatt tctgaaaagt taggagagcc aatttggctct agtagttaag 300
CVF1 gtagcagtct agaactctgg agactgagtt ctaatctcac tttagccatg 350
CVF1 aaagtcagct ggggtggcctt ggccagtcgt tctctttcag cccaattcac 400
CVF1 ctcacagggt gttgttgggt gttgtgggga aaataggagg aggaaagagt 450
CVF1 gtcagatatg tccaccactt tgagtgttt ataaaagtaa taaagtagg 500
CVF1 atataaataa ataataaata aaatatattg aattaaactc agcagataca 550
CVF1 tcttctgtca gtgctagagt tctacccttt caccogttcc ttttctgta 600
CVF1 gTTTGCCAA GCTTTGAAGT CCGTCTGCAA CCATCAGAGA AG 642

```

Intron 7

Die unterstrichenen Regionen sind zueinander revers komplementär und ermöglichen die Ausbildung von Sekundärstrukturen im Fall von Einzelstrang-DNA oder RNA. Vorwärtsprimer: HB7s-CVF, Rückwärtsprimer: HB8as.

```

CVF1 TCCGCTGCA ACCATCAGAG AAGTTTTTTT ACATTGACGG CAATGAAAAT 50
CVF1 TTCCACGTGT CTATCACTGC AAGgtatgct tattacaac tatgagtagc 100
CVF1 taagaaaatg tattttggca ctggtaata tctgcatgaa agggattttt 150
CVF1 ctttgggtca taatattcag aaaaataaga tttggagact agtcctgatt 200
CVF1 tgcaaaaattt acttgaaatt tatttatgaa atacattatg tcagagttcc 250
CVF1 aagtaacata cccatagcaa acaaaactct gaggcaggca atttcttaa 300
CVF1 agaatagggg tgggattcta ccagttcgga ccggttctat agaaccagtt 350
CVF1 gttatgattt gcgcagttcg gagaaccaga aaattgtgcc ccttgctggt 400
CVF1 cctgcccctc cccgtccaacc cctcccctcc ccccagtcga gcccttatca 450
CVF1 cgctgcatg acattgtgcg catgctcagc attgctgact gcattttccc 500
CVF1 ctgcctagtc actccttaag..... 520
.....(ca. 1500 bp).....
.....
CVF1 .....ccaggcatag 10
CVF1 cagtgataag aggaagaata atcaaatcca ttttccaggg aaatattaac 60
CVF1 acccccccac aaaaatgcc ccaatccctt tatttaggaa gttacaataa 110
CVF1 tgttcctttc acaaaactct tagttatttt aataattagt ttaattgta 160
CVF1 ttaaggctttt aattgtcaca tattttctac tccatacttt ttatttttgc 210
CVF1 aagccacctt gagtgttga atggacagaa aggtgaaatt tagaaatcag 260
CVF1 aaaataaaat gggcagaagg atgaaggac tgaacttagc atcctcctcc 310
CVF1 ccactctcaa attctgtttc ctctgctca cctttagcat cagaagtgac 360
CVF1 aaaggcttgg agaaagatgc ctgaatgagt agagttgaga aaataaaaag 410
CVF1 ctgtgagaaa tagcttaaca ccctactga ttctcttgt gtcaaagaag 460
CVF1 atcggcagga aaagcaaagt tgtataattg ggcctttact ttcattgtgt 510
CVF1 cattacagGT ACTTGATGG AGAGGAAGTG GAAGGTGTGG CTTTGTCTT 560
CVF1 CTTTGGAGTG AAAATAGATG ATGCTAAAAA GAGTATCCA GACTCACTCA 610

```

Intron 8

Exon 8 stromabwärts

Dargestellt ist die Konsensus-Sequenz eines DraI-Amplifikats (ca. 500 bp) und eines PvuII-Amplifikats (ca. 1,8 kb). Die EcoRI-Schnittstelle (GAATTC) im Exonbereich, eine DraI-Schnittstelle (tttaa) und die PvuII-Schnittstelle (cagctg) am 3'-Ende sind unterstrichen dargestellt. An der umrahmten Position in Exon 8 wurde im DraI-Amplifikat ein T wie in CVF1 und im PvuII-Amplifikat ein C gefunden. Vorwärtsprimer: HB8s-CVF-Intr.

| | | | | | | |
|------|--------------------------|--------------------|--------------------|-------------------|--|-----|
| CVF1 | <u>gtgtcattac</u> | ag GTACTTGT | ATGGAGAGGA | AGTGAAGGT | GTGGCCTTTG | 50 |
| CVF1 | TCCTCTTTGG | AGTGAAAATA | GATGATGCTA | AAAAGAGTAT | TCCAGAC T CA | 100 |
| CVF1 | CTCACGAGAA | TTCCG gtaaa | atgtaacagg | ggtgtttgtg | cattggagtg | 150 |
| CVF1 | tgcctgtggg | tttttgtgtc | agatctacta | ccggttgata | ggtaaggcgg | 200 |
| CVF1 | ggctgacatc | ctagcgatac | aacatgcctc | acagagatct | ctgtcagact | 250 |
| CVF1 | gtcaaaatct | agctgtcaga | gaggtcctaa | atggacccan | atcaccttgg | 300 |
| CVF1 | agggatgatg | agtgtagcga | gcacactctg | gtgactgcag | gtctgcaatc | 350 |
| CVF1 | tccctgatcg | gcctatctaa | cagcaaggaa | gacagccatt | gtgtgccgtt | 400 |
| CVF1 | ccaaagctgg | agcagtcact | gcaaagaaag | tggaagtggg | agacagtgtg | 450 |
| CVF1 | aacggactgg | gcgagataat | <u>tttaa</u> atctg | aatatttaag | aacaa..... | 495 |
| | | | | | | |
| | (ca. 860 bp) | | | | | |
| | | | | | | |
| CVF1 |ccccncaat | taaaataaac | aaaccccaag | agtcagcaca | | 39 |
| CVF1 | gcccagattg | taaaaacagc | caggttttta | gctcgtggtg | gaaggtgcag | 89 |
| CVF1 | agatcgggag | ccagccagag | ctccagcggg | agaagattcc | ataaggtggg | 139 |
| CVF1 | ggccgccaca | gagtaggcc | tccccgggt | ccctgccaaag | caacatgag | 189 |
| CVF1 | tcacagacgg | caccctgaag | agaccctgtc | tgtgcgccgg | taccaagcgg | 239 |
| CVF1 | ttggtacttg | aagctgggag | gagtcgttct | cataagtatc | ctggacccat | 289 |
| CVF1 | gctatgtagg | actttgaatg | ttaataccaa | aaccttgaat | ctcaccaga | 339 |
| CVF1 | gggcaaccag | aagccaatgc | aggccgtgca | ggaggggtgt | gatatgggag | 389 |
| CVF1 | tatcaggaag | ctcccagcac | tacctgcgcg | gctgcattct | gcaccagctg | 439 |

Exon 9 stromaufwärts

Die DraI-Schnittstelle am 5'-Ende ist unterstrichen dargestellt. Äußerer Rückwärtsprimer: HB9as-CVF, innerer Rückwärtsprimer: HB9as-II-CVF.

| | | | | | | |
|------|-------------------|-------------------|-------------------|-------------------|--------------------|-----|
| CVF1 | | | <u>ttta</u> | <u>aaagatctac</u> | taggtagaac | 24 |
| CVF1 | aaaccacaaa | accaattcct | caagaaggct | acttttactg | gctataataa | 74 |
| CVF1 | cagaatthtg | caagcctagt | tgtgctacat | ttttttctcc | ttctaataat | 124 |
| CVF1 | tcagtgaata | agggagcacg | tgctgagtt | tcttctgaat | gttatctgga | 174 |
| CVF1 | tgttctggac | ttaaaacata | ttttaccctt | tgctgctatg | gccacagagc | 224 |
| CVF1 | ttgcataatc | tcattccatca | tcttccttat | tctacattgt | ggcagaccac | 274 |
| CVF1 | ttgatttcct | tccttatttt | tacttatttt | tcgtttattt | tcag ATTATT | 324 |
| CVF1 | GATGGAGATG | GGAAAGCAAC | ACTAAAAAGA | GATACATTCC | GTTCTCGATT | 374 |
| CVF1 | TCCAAATCTC | AATGAGCTTG | TTGGGCATAC | TCTGTATGCA | TCTGTAACAG | 416 |

Intron 9

Eine vierfache tta-Wiederholung ist unterstrichen. Vorwärtsprimer: HB9s, Rückwärtsprimer: HB10as.

| | | | | | | |
|------|-------------------|--------------------|-------------------|--------------------|-------------------|-----|
| CVF1 | CAAATCTCAA | TGAGCTTGTT | GGGCATACTC | TGTATGCATC | TGTAACAGTC | 50 |
| CVF1 | ATGACAGAAT | CAG gtgagca | ggggtgggag | tagagaaggg | gtgttgtagc | 100 |
| CVF1 | caattcaggt | ctggtgcttg | tgccgaggcc | gatcaagaaa | accngtcagg | 150 |
| CVF1 | aggtgcagat | ttgtttcctt | ttattgaagc | tctccaagaa | aaagggtagt | 200 |
| CVF1 | tgacccaaat | gccaaagAAC | cttgaacaat | agaatcaaga | aagctttata | 250 |
| CVF1 | cattttcatt | aaagcagaaa | ataagataaa | tatataatat | acaaggatat | 300 |
| CVF1 | atagaaata | tcatttaata | aacatttagt | aatgatctaa | caattattct | 350 |
| CVF1 | attggctctc | agctgtccct | attcctaagc | ctnggctaata | taatgcttca | 400 |
| CVF1 | ggagttatct | aatatagact | ctatttgctc | cagagcatct | ctattcctaa | 450 |
| CVF1 | cttctaccgt | gtttccccga | aaatangacc | taccctgaaa | gtaagacctt | 500 |
| CVF1 | gcac..... | | | | | 504 |
| | | | | | | |
| | | (ca. 2100 bp) | | | | |
| | | | | | | |
| CVF1 |tgta | tgaaaatgtg | attatgttca | | | 24 |
| CVF1 | gtgagaaaa | gaagagagag | aaagataaaa | tcaggcttcc | tttgcttggt | 74 |
| CVF1 | attttaacaa | cttccttggt | ctgaaaaaaa | aatgaaaatc | cgttttataa | 124 |
| CVF1 | tgtattgttg | ctgttttggt | <u>tttatttatt</u> | <u>tatttatttag</u> | atttttatcc | 174 |
| CVF1 | tacccttttc | atttaggggt | ttctagaaaa | agagaggaag | aggaggagag | 224 |
| CVF1 | aaaagagaga | cagggagaaat | ttgtgaactg | ttaatgtact | gggtttatta | 274 |
| CVF1 | ggtttgaaac | caatagtgat | acctctaggt | tcaaagtctt | tgcaggattt | 324 |
| CVF1 | ttcatttttc | tcttgcaatt | ctcagtttga | cttcctattt | taagttgcat | 374 |
| CVF1 | tccttctttt | tgttttgcag | GCAGTGATAT | GGTAGTGACT | GAGCAAAGCG | 424 |
| CVF1 | GCATTCATAT | TGTGGCATCT | CCCTATCAGA | TCCACTTCAC | AAAAACCCCC | 474 |

Intron 10

Vorwärtsprimer: HB10s, Rückwärtsprimer: HB11as.

| | | | | | | |
|------|-------------------|--------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | GCATCTCCCT | ATCAGATCCA | CTTCACAAAA | ACCCCAAAT | ATTTCAAGCC | 50 |
| CVF1 | AGGAATGCCA | TATGAAGTGA | CG gtagaaa | attgaagtct | cctattttgt | 100 |
| CVF1 | caaagagcca | ccttagcaat | gtctttctcc | ccttactggg | cttctgagga | 150 |
| CVF1 | gctgtctaga | cttgcttggt | caagatgcaa | gtgggagctt | cagcttatct | 200 |
| CVF1 | aggagttag | gctgccactg | tagattgatt | gattgattgg | tattttatac | 250 |
| CVF1 | actggatggt | ttgctgtaat | gtatctgcac | agaagttagt | tnatttagtg | 300 |
| CVF1 | acaaaaagcc | ttaataaatt | aatctgtaag | ccacctggaa | tactatgtg | 350 |
| CVF1 | atthgggcag | ccatataaat | cttgtaaatg | attaataaat | aagtgaataa | 400 |
| CVF1 | actgatthtag | atcagagga | agctgccatg | tttcatgctg | aatcacctct | 450 |
| CVF1 | tctgggttgt | tttag GTGTA | TGTTACCAAC | CCTGATGGCT | CACCAGCTGC | 500 |
| CVF1 | CCATGTGCCA | GTGGTATCAG | AGGCCTTTCA | TTCTATGGGA | ACCACTTTGA | 550 |
| CVF1 | GTGATGGGAC | | | | | 560 |

Intron 11

Das genomische Amplifikat stimmte an der umrahmten Position in Exon 12 nicht mit der CVF1-cDNA (Position 1264) überein (C-T). Eine EcoRI-Schnittstelle in Intron 11 ist unterstrichen. Vorwärtsprimer: HB11s, Rückwärtsprimer: HB12as.

| | | | | | | |
|------|--------------------------|-------------------|-------------------|--------------------------|-------------------|-----|
| CVF1 | <u>CCATGTGCCA</u> | <u>GTGGTATCAG</u> | <u>AGGCCTTTCA</u> | <u>TTCTATGGGA</u> | <u>ACCACTTTGA</u> | 50 |
| CVF1 | <u>GTGATGGGAC</u> | <u>TGCTAAGCTC</u> | <u>ATCCTGAACA</u> | <u>TACCATTGAA</u> | <u>TGCTCAAAGC</u> | 100 |
| CVF1 | <u>CTACCAATCA</u> | CTgtaagtca | ctctgacatt | ttgctaaatc | ccattcagtg | 150 |
| CVF1 | tgctgatgag | agggaatgct | atgacatgga | ggtggattcc | caggagtaag | 200 |
| CVF1 | gggggtagat | tccaaaggag | taagaggcat | ctcaatcagg | atgctttgtg | 250 |
| CVF1 | agagagaaat | agagtgagga | cagccctcc | ctaatagttc | tcagagtata | 300 |
| CVF1 | ttttagatac | agatagtaga | ctcctcagtc | tggcaagatt | ttgcctatag | 350 |
| CVF1 | gtttgaaagc | aataaagaac | tcaacttggg | aaaagtgtta | ttcttggctt | 400 |
| CVF1 | tgggcctgac | atctatgttg | gaacataggg | ccatcacagt | ctgggtgtgca | 450 |
| CVF1 | aattgtgcac | cacctaattg | cttgacctaa | agagaggatt | gtagttgcc | 500 |
| CVF1 | agttacaaag | aaagagtagg | agatacagtg | ccagtaacat | caacttacc | 550 |
| CVF1 | gctttgccca | tacacagact | ctgggagcag | agacaggcct | agcaggtcta | 600 |
| CVF1 | tgcattgtgg | <u>aattccgcac</u> | a..... | | | 621 |
| | | | | | | |
| | (ca. 1050 bp)..... | | | | | |
| | | | | | | |
| CVF1 | | aacaggatgc | tagcaggtcc | attctagcca | | 30 |
| CVF1 | tcaatgtatt | ctttcagcag | tcttcgtgtg | aacatcacta | agtgccagcc | 80 |
| CVF1 | ctaaaaggag | ttctagtttc | ttcctaatta | ggttactgct | tcatgtttac | 130 |
| CVF1 | cctatctagG | <u>TTAGAATAA</u> | <u>CCATGGAGAC</u> | <u>CTC</u> <u>CAAGAG</u> | <u>AACGCCAGGC</u> | 180 |
| CVF1 | <u>AACAAAGTCC</u> | <u>ATGACAGCCA</u> | <u>TAGCCTACCA</u> | <u>AACCCAGGGA</u> | <u>GGATCTGGAA</u> | 230 |

Intron 12

Vorwärtsprimer: YL12s, Rückwärtsprimer: YL13as.

| | | | | | | |
|------|-------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | <u>GCAACAAAGT</u> | <u>CCATGACAGC</u> | <u>CATAGCCTAC</u> | <u>CAAACCCAGG</u> | <u>GAGGATCTGG</u> | 50 |
| CVF1 | <u>AAACTATCTT</u> | <u>CATGTAGCCA</u> | <u>TTACATCTAC</u> | <u>AGAGATTAAG</u> | <u>CCCGGAGATA</u> | 100 |
| CVF1 | <u>ACTTACCTGT</u> | <u>CAATTTCAAT</u> | <u>GTGAAGGGCA</u> | <u>ATGCAAATTC</u> | <u>ACTGAAGCAG</u> | 150 |
| CVF1 | <u>ATCAAATATT</u> | <u>TCACATACCT</u> | Cgtaagtatt | atagatctgg | aggaaatcca | 200 |
| CVF1 | ggggctgcat | ttcaattcat | attgtagcct | tgaagccact | ggcttaatct | 250 |
| CVF1 | atcatttcta | accttcccat | taatcagtat | ttgttcaccc | tcagATATTG | 300 |
| CVF1 | <u>AATAAAGGGA</u> | <u>AGATTTTCAA</u> | <u>GGTTGGCAGG</u> | <u>CAACCCAGGA</u> | <u>GAGATGGGCA</u> | 350 |
| CVF1 | <u>GAATCTGGTG</u> | <u>ACCATGAATC</u> | <u>TGCATATCAC</u> | <u>TCCAGATCTC</u> | <u>ATCCCTTCCT</u> | 400 |
| CVF1 | <u>TCCGGTTTGT</u> | <u>GGCTTACTAC</u> | <u>CAAGTGGGAA</u> | <u>ACAACGAAAT</u> | <u>TGTGGCTGAT</u> | 450 |
| CVF1 | <u>TCTGTC</u> | | | | | 456 |

Intron 13

Die Lücke in Exon 14 markiert einen charakteristischen Größenunterschied zu den entsprechenden Sequenzen von CVF2 und coC3. Vorwärtsprimer: HB13s, Rückwärtsprimer: HB14as.

| | | | | | | |
|------|-------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | <u>CACTCCAGAT</u> | <u>CTCATCCCTT</u> | <u>CCTTCCGGTT</u> | <u>TGTGGCTTAC</u> | <u>TACCAAGTGG</u> | 50 |
| CVF1 | <u>GAAACAACGA</u> | <u>AATTGTGGCT</u> | <u>GATTCTGTCT</u> | <u>GGGTGGATGT</u> | <u>GAAGGATACC</u> | 100 |
| CVF1 | <u>TGCATGGGAA</u> | CGgtaagcca | ggcttcccag | tagctttctt | agcagtcatg | 150 |
| CVF1 | ggatctccct | tattatagct | gcataaatca | gatgtgcata | atthaaagca | 200 |
| CVF1 | ctttcatggc | atthtcccca | aaacatttat | ttctgttgag | ctagagaaac | 250 |
| CVF1 | tatttccct | taaattggca | atthttctctg | tctgttagag | atthaaagctg | 300 |
| CVF1 | ttctttgctt | tgtataaaca | gTTGGTTGTG | AAAG----- | -----GAGA | 338 |
| CVF1 | <u>CAATCTAATA</u> | <u>CAAATGCCAG</u> | <u>GAGCTGCAAT</u> | <u>GAAAATCAA</u> | <u>TTGGAAGGGG</u> | 388 |
| CVF1 | <u>ATCCAGGTGC</u> | <u>TCGGGTTGGT</u> | <u>CTTGTGGCTG</u> | <u>TGGACAAAG</u> | | 427 |

Intron 14

Vorwärtsprimer: AJS01, Rückwärtsprimer: OPA19.

```

CVF1 GGATCCAGGT GCTCGGGTTG GTC'TTGTGGC TGTGGACAAA GCAGTATATG 50
CVF1 TTCTCAATGA TAAATATAAG ATTAGCCAAG CTAAGgtaag gccagaaatg 100
CVF1 tcttaagtta caaacattca gctttcaaga gttttccaac ttacagtact 150
CVF1 gaagaggaac attttttttc tctgatgtgc actagATATG GGACACAATA 200
CVF1 GAAAAGAGTG ACTTTGGCTG TACAGCTGGC AGTGGCCAGA ATAATCTGGG 250
CVF1 TGTGTTTGAA GATGCTGGAC TGGCTCTGAC AACCAGCACT AATCTCAACA 300
CVF1 CCAAACAG 308

```

Intron 15*Exon 15 stromabwärts*

Gezeigt ist die Sequenz eines genomischen PCR-Produkts stromabwärts von Exon 15. Die kursiv dargestellte Intron 15-Sequenz am 3'-Ende (stromabwärts der ersten MboI-Schnittstelle gatc) ist überlappend mit dem 5'-Bereich des Cosmid-Klons 1R67. Auffällig ist eine vierfache caag-Wiederholung (unterstrichen). Eine EcoRI-Schnittstelle ist unterstrichen dargestellt. Vorwärtsprimer: HB15s, Rückwärtsprimer: HB-1R67-5'as.

```

CVF1 TGGCTGTACA GCTGGCAGTG GCCAGAATAA TCTGGGTGTG TTTGAAGATG 50
CVF1 CTGGACTGGC TCTGACAACC AGCATAATC TCAACACCAA ACAGAGATCA 100
CVF1 Ggtagtattt tgggaatgga tgctaagtgc tgggagatgg cagaggtttt 150
CVF1 ccaatccatt taggggcat gggaagagca ttgctccatc atatactatt 200
CVF1 ggaatcatat gggaatcaat tcacaaggcc cagagcaaat agtcaagccc 250
CVF1 cactgcagag tcacttgggt gagatgggca gtttagaaat tacttaatta 300
CVF1 aacaagcaag caagcaagca aaacgtattg gtatcaatgt gccctatagg 350
CVF1 aactacctcc agaa..... 364
.....
..... (ca. 1000 bp) .....
.....
CVF1 .....c ctttttgcca 11
CVF1 accaactaac aaacatatgc aggaaacaca caatagtaat ggggcacttc 61
CVF1 aattatcctg atatcaattg gaaaactaat tctgcaccaa atgagaaatc 111
CVF1 aaataagttc ctaatcaacc tagctgacaa ctttgttgtc caaaaggtag 161
CVF1 agggtgaaac tagagggaca gccatactgg acataattct tacaacacaga 211
CVF1 gaagaagtga tagagggagt tgaaactgca ggaaccttgg gtgagagtga 261
CVF1 ccatgtcata ctagaattca acataacaca aacacaagca atagaacata 311
CVF1 atgacaccag agtcctggat cttaaaagag ctggtttta taaactcaga 361
CVF1 gagaacttg gaaagattcc atggatgaaa atcctaaagg ggaaaacaac 411
CVF1 ttaaggatct tggaaaactc tgaaaagtac aatcatagaa gccagacca 461

```

Exon 16 stromaufwärts

Dargestellt ist der 5'-Sequenzbereich (7,5 kb) des Cosmid-Klons 1R67. Die MboI-Schnittstelle (gatac) am 5'-Ende des CVF1-Inserts und drei weitere MboI-Schnittstellen sowie eine EcoRI-Schnittstelle (gaattc) sind unterstrichen. Die mit dem genomischen Amplifikat (siehe Exon 15 stromabwärts) überlappende Sequenz ist kursiv dargestellt. 5'-PCR-Produkt: Vorwärtsprimer: sCos1-T3s, Rückwärtsprimer: HB-Con2eco-as; 3'-PCR-Produkt: Vorwärtsprimer: Con2eco, Rückwärtsprimer: OPA72.

| | Vektor sCos-1 | | | | | |
|------|---------------------------|--------------------------|--------------------------|--------------------------|--------------------------|------|
| CVF1 | <u>CGCAATTAAC</u> | <u>CCTCACTAAA</u> | <u>GGgatcttaa</u> | <i>aagagctggt</i> | <i>ttaataaac</i> | 50 |
| CVF1 | <i>tcagagagaa</i> | <i>cttgggaaaag</i> | <i>attccatgga</i> | <i>tgaaaatcct</i> | <i>aaaggggaaa</i> | 100 |
| CVF1 | <i>acaacttaag</i> | <i>gatcttggaa</i> | <i>aactctgaaa</i> | <i>agtacaatca</i> | <i>tagaagccca</i> | 150 |
| CVF1 | <i>gaccagcaca</i> | <i>ataccactga</i> | <i>gaaagaaaa</i> | <i>caagaaatcc</i> | <i>aagaagaac</i> | 200 |
| CVF1 | <i>cagcatggcc</i> | <i>acataaagat</i> | <i>ctctctggca</i> | <i>aattgaaagg</i> | <i>aaaaaaggac</i> | 250 |
| CVF1 | <i>aagtagaaaa</i> | <i>aatggaaaaga</i> | <i>gggcatacaa</i> | <i>ttaaggcaga</i> | <i>atatcagcag</i> | 300 |
| CVF1 | <i>atagccagaa</i> | <i>tctgcaaaga</i> | <i>tgaagtcag</i> | <i>aaagcaaaaa</i> | <i>ctcagaacga</i> | 350 |
| CVF1 | <i>acaaagactt</i> | <i>gcaacaaaag</i> | <i>tcaaagataa</i> | <i>tttttataaa</i> | <i>aaatcttcca</i> | 400 |
| CVF1 | <i>acataataat</i> | <i>aacaagaaaa</i> | <i>aaatcaagga</i> | <i>aacagttggt</i> | <i>ccactaaaga</i> | 450 |
| CVF1 | <i>gagaagatgg</i> | <i>caaggaagta</i> | <i>acaggcaata</i> | <i>gagagaaaag</i> | <i>agaactgctt</i> | 500 |
| CVF1 | <i>aactcaatct</i> | <i>ttgcatcagt</i> | <i>cttcatgcaa</i> | <i>aaaaaaaaacc</i> | <i>tatagctcaa</i> | 550 |
| CVF1 | <i>cccacaaaa</i> | <i>acataactgt</i> | <i>aaaagacaga</i> | <i>ccagaaataa</i> | <i>aagttaaaat</i> | 600 |
| CVF1 | <i>aagcaaaaaca</i> | <i>aaaagtagta</i> | <i>agagaacgcc</i> | <i>tatctgatct</i> | <i>tgatgaacac</i> | 650 |
| CVF1 | <i>aaatcaccag</i> | <i>gcctgatgg</i> | <i>attacatccc</i> | <i>agagttttta</i> | <i>ag.....</i> | 692 |
| | (ca. 3300 bp) | | | | | |
| CVF1 | <i>.....t</i> | <i>ttagcaaata</i> | <i>ctgtaccct</i> | <i>gccaaaggaa</i> | <i>tgagcagat</i> | 41 |
| CVF1 | <i>gatgcacaac</i> | <i>aggtacagg</i> | <i>tatgcctct</i> | <i>gctgcagcg</i> | <i>atggttaaga</i> | 91 |
| CVF1 | <i>gtccccttgt</i> | <i>agtcaccaca</i> | <i>gattctcact</i> | <i>gaccatcag</i> | <i>ccttgatgg</i> | 141 |
| CVF1 | <i>tgtgactatg</i> | <i>ggtgtttccc</i> | <i>atctggagt</i> | <i>gtcagtaggt</i> | <i>tccaaaacc</i> | 191 |
| CVF1 | <i>cttgggcat</i> | <i>gagcttatct</i> | <i>aattgtgcat</i> | <i>caattttttc</i> | <i>ctttaaagca</i> | 241 |
| CVF1 | <i>aatggaacc</i> | <i>tcttgggctt</i> | <i>taaacgaatt</i> | <i>gggggacact</i> | <i>tggggtctaa</i> | 291 |
| CVF1 | <i>gctgaaagca</i> | <i>ataggggtcc</i> | <i>ccatatactt</i> | <i>gccagtcct</i> | <i>tcgctgaaga</i> | 341 |
| CVF1 | <i>cctctggaaa</i> | <i>ctcagacttc</i> | <i>aaggactctt</i> | <i>cccagtcaga</i> | <i>actgctcacg</i> | 391 |
| CVF1 | <i>tgatggatgc</i> | <i>cagtcacctc</i> | <i>gatgccaagt</i> | <i>ggcttgaacc</i> | <i>actcaagtcc</i> | 441 |
| CVF1 | <i>cagcaaaact</i> | <i>gggcgatcct</i> | <i>tgtccaagat</i> | <i>ggtcagtggc</i> | <i>aaacgtttgg</i> | 491 |
| CVF1 | <i>tggtggcctt</i> | <i>aaactgcacg</i> | <i>ttgaattcc</i> | <i>ctcgtcccaa</i> | <i>caactggaatc</i> | 541 |
| CVF1 | <i>ttgttccctt</i> | <i>ggtagtctt</i> | <i>gagggtcagt</i> | <i>gcttgcctct</i> | <i>ggaaatgctt</i> | 591 |
| CVF1 | <i>tttctgaatt</i> | <i>tttgaaaaga</i> | <i>gtctgtgat</i> | <i>agtgtccat</i> | <i>gacaccatag</i> | 641 |
| CVF1 | <i>ataaataaga</i> | <i>gcctgagtc</i> | <i>acttccatgt</i> | <i>ggcacggtct</i> | <i>cccttcaata</i> | 691 |
| CVF1 | <i>aacaccttca</i> | <i>catgtacttt</i> | <i>tctgctattt</i> | <i>aaatgtgaag</i> | <i>aatggccaat</i> | 741 |
| CVF1 | <i>ggtgattgtt</i> | <i>tggtgagttt</i> | <i>cactcttgcc</i> | <i>atctgctttt</i> | <i>aaatttcgcc</i> | 791 |
| CVF1 | <i>ccatttttcc</i> | <i>aattaaattg</i> | <i>ttgggcttaa</i> | <i>aatcttctcg</i> | <i>aggaaatttc</i> | 841 |
| CVF1 | <i>ttcctaaatg</i> | <i>gtggagttct</i> | <i>ggatgatttg</i> | <i>cagactctgg</i> | <i>ctatatgacc</i> | 891 |
| CVF1 | <i>ccgttttctca</i> | <i>caacgtaggc</i> | <i>aaatggagtc</i> | <i>aaagaacttg</i> | <i>cattcagtcg</i> | 941 |
| CVF1 | <i>gggggtgcc</i> | <i>ccccccacag</i> | <i>cctctgcaat</i> | <i>agaatccttg</i> | <i>ctgtttttc</i> | 991 |
| CVF1 | <i>tcagagtctt</i> | <i>ttcttttctt</i> | <i>ctccattctg</i> | <i>agcttaaaaa</i> | <i>tcccctcttc</i> | 1041 |
| CVF1 | <i>ttcctcactg</i> | <i>ctatccgaat</i> | <i>cagagtctat</i> | <i>ttgtttgca.....</i> | | 1080 |
| | (ca. 2300 bp) | | | | | |
| CVF1 | <i>.....ccaa</i> | <i>gattgttggg</i> | <i>ggcaataagc</i> | <i>tgactttgta</i> | | 34 |
| CVF1 | <i>aaaatgccgc</i> | <i>ttgggggggg</i> | <i>gggggcaaaa</i> | <i>aagccctagt</i> | <i>aataagcagg</i> | 84 |
| CVF1 | <i>catatacaat</i> | <i>ggaatgaagg</i> | <i>ctattgctga</i> | <i>gcgcatgggg</i> | <i>ttgtgctccc</i> | 134 |
| CVF1 | <i>aaaaaaatg</i> | <i>taagccgcc</i> | <i>tgagtctcca</i> | <i>gagaagggcg</i> | <i>gcatataaat</i> | 184 |
| CVF1 | <i>gtgaataaaa</i> | <i>gaaaacaacc</i> | <i>atgaaaaagc</i> | <i>tgagatggt</i> | <i>ctttgacttc</i> | 234 |
| CVF1 | <i>ttcaactgtg</i> | <i>agcccttcaa</i> | <i>ctcagacca</i> | <i>atgaagagtt</i> | <i>gagctatgct</i> | 284 |
| CVF1 | <i>ttttacccta</i> | <i>ctgcacctc</i> | <i>ccagctctaa</i> | <i>agtgaacttt</i> | <i>ctatgagctg</i> | 334 |
| CVF1 | <i>ctttggatct</i> | <i>gtagatata</i> | <i>gatccaaaata</i> | <i>actcataatt</i> | <i>aaggcttcat</i> | 384 |
| CVF1 | <i>tcccagtcta</i> | <i>ttttaaatgg</i> | <i>gggagtgctc</i> | <i>catagaaatc</i> | <i>agcagaactc</i> | 434 |
| CVF1 | <i>ccttactgta</i> | <i>aatccatttt</i> | <i>aaaattaatc</i> | <i>ttcacaggaa</i> | <i>ctagtgtctga</i> | 484 |
| CVF1 | <i>atgttacctt</i> | <i>cttactgtaa</i> | <i>agaataattg</i> | <i>ttcccacctc</i> | <i>ttctcccctg</i> | 534 |
| CVF1 | <i>cctccttcag</i> | <u>CTGCAAAGTG</u> | <u>TCCTCAGCCT</u> | <u>GCAAATCGGA</u> | <u>GGCGTCGCAG</u> | 584 |

Intron 16

Eine EcoRI-Schnittstelle im 5'-Bereich von Intron 16 ist unterstrichen dargestellt. Vorwärtsprimer: OPS72, Rückwärtsprimer: OPA24.

| | | | | | | |
|------|---------------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | <u>CTGCAAAATCG</u> | GAGGCGTCGC | AGTTCTGTTT | TGCTGCTTGA | CAGCAACGCA | 50 |
| CVF1 | AGCAAAG gta | tgtggcaaat | aatgaagttt | ctgtttactc | tgagataaaa | 100 |
| CVF1 | gaataaaaact | acagtccact | tgcattgatt | gaatcaatca | ggaaatggag | 150 |
| CVF1 | gaaggaatgg | aggaagatta | tccccctttc | ccttctttct | caagcaagag | 200 |
| CVF1 | ggtagaaaat | taataatctc | tattaaatag | gatcattctt | aacctatcaa | 250 |
| CVF1 | taactttgag | <u>tgcttgaatt</u> | cagagtgcaa | tggggggagt | acttaagggg | 300 |
| CVF1 | cccaatagtt | taatgtagaa | tatacagtaa | ctggttccag | ctaaagagat | 350 |
| CVF1 | tacatttgat | acacaaaatt | ttgaagcatt | gcttttgggt | gttctcggat | 400 |
| CVF1 | cgcagcactg | atcaagaatt | tctcttgatg | aggaggaagt | aagcagtcag | 450 |
| CVF1 | tggccagtga | actacaagca | acccccaaac | caggaacatc | tgtgggaaatg | 500 |
| CVF1 | cctgcatatc | aacatggcag | ttccaactga | acaatcaaag | ctctgcccc. | 549 |
| | | | | | | |
| | (ca. 400 bp) | | | | | |
| | | | | | | |
| CVF1 |aggcg | tatactcag | aactgtcatt | tggactcctc | cgtaagcaag | 45 |
| CVF1 | gttgtgggat | ggcttgggtg | aaagaagagg | gaaaactgaa | gaaaaggcc | 95 |
| CVF1 | tttatgtgat | ttaaagccca | tggtttttcc | aggagctggt | gatggcttac | 145 |
| CVF1 | tcaagcctcc | tgtagcaga | tggcacctag | aatgctccat | ttaaggggta | 195 |
| CVF1 | gagatacact | gcaacatttg | attgattttc | aaatacagaa | gataagtttg | 245 |
| CVF1 | aatgttctgg | caaaactgca | gatgcagtaa | taagccataa | agaaatctta | 295 |
| CVF1 | tgaaaatgac | agtccttagt | aactacagta | gtgaaaagtc | aaaggatggt | 345 |
| CVF1 | cagctgcccc | ttccggctgt | ctgggacagc | ataaacaact | aactcatttg | 395 |
| CVF1 | attggtttat | gtagactc | gttgtctctt | gtttcccttt | atgcacatga | 445 |
| CVF1 | ctatgagaaa | taggattctc | taatgtttca | cgcattttgc | atatttttagC | 495 |
| CVF1 | GGCAGAATTT | CAGGATCAAG | ACCTGCCTAA | ATGCTGTGAA | GATGTCATGC | 545 |

Intron 17

Beim EcoRI-Verdau entstanden Fragmente von ca. 3,5 kb, 3,2 kb und 250 bp. Eine EcoRI-Schnittstelle im 3'-Bereich von Intron 17 ist unterstrichen dargestellt. Vorwärtsprimer: OPS24, Rückwärtsprimer: OPA22.

| | | | | | | |
|------|---------------------------|--------------------|--------------------|-------------------|-------------------|-----|
| CVF1 | CGTAAATGCT | GTGAAGATGT | CATGCATGAG | AACCCCATGG | GGTACACTTG | 50 |
| CVF1 | TGAAAAGCGT | GCAAAAATACA | TCCAGGAGGG | AGATGCTTGT | AAGGCTGCCT | 100 |
| CVF1 | TCCTTGAATG | CTGTCGCTAC | ATCAAGGGGG | TCCGAGATGA | AAACCAACGG | 150 |
| CVF1 | GAGAGCGAGT | TGTTTCTGGC | AAGAG gtaag | tagcaggttc | cgggtttatt | 200 |
| CVF1 | gaaagaaagg | aacattcaat | tcacaatacc | tggaaaacat | ggcacataac | 250 |
| CVF1 | agcatacca | aggcttgaa | acccatttgg | agtatattgt | tattattatt | 300 |
| CVF1 | aggtatttat | atgtcattgt | tgetaatttg | tagctgagct | ttagtgcaaa | 350 |
| CVF1 | atacagctct | gtaaatcag | gggaatatta | gaagtgggt | tgagtgagac | 400 |
| CVF1 | tctacttgty | ggaatcacag | cacatccagc | ctcccagtca | gtccctgggc | 450 |
| CVF1 | tctctcagtc | cccatcaagg | ttagctgcat | tttttaaaac | attggtgaca | 500 |
| CVF1 | tttgggttca | gctaaagcac | cagcagcttt | tggggaggtc | tcagcagttg | 550 |
| CVF1 | aagaaagctg | tcagtgaaaa | atgatttggg | gagcttgttg | agtgaagtg | 600 |
| CVF1 | atggaatac | agatctttct | ttggtggtag | tctgcttcca | ttt..... | 643 |
| | | | | | | |
| | (ca. 5700 bp) | | | | | |
| | | | | | | |
| CVF1 | | ccctc | | | | 5 |
| CVF1 | gatgtgatgg | tccggtgtct | ggaggcggty | cagatctgga | tggggaggaa | 55 |
| CVF1 | cagacttgg | caatcactca | actgggagtt | aaagacagcc | tgcagcagca | 105 |
| CVF1 | acaggytcag | agtttgtagt | tgtccagcca | gcacacaagt | aaataattca | 155 |
| CVF1 | gcaggttcat | ttatagaaga | aacatcttta | tataccataa | tataaaacat | 205 |
| CVF1 | gatgcatgta | tacaataccg | atgcagcttt | tgcaatgtgg | tagctagtac | 255 |
| CVF1 | aggacaagaa | gtagaagtac | tctcaagcaa | agagacaaac | gaactctttt | 305 |
| CVF1 | atacacagaa | cagttaagcc | aagccatacc | ccggctccaa | gctctctcag | 355 |
| CVF1 | tcacaaaaca | gccccattg | gctgacttgt | tgccatgtct | atctcagtc | 405 |
| CVF1 | <u>tgaattctta</u> | ttcactgaca | aacaggctac | tcaggacctc | tgtaggatgt | 455 |
| CVF1 | gtccattcat | ggtcaccac | tccattcaat | ttatgaaata | tttaattaac | 505 |
| CVF1 | aaagtacagt | taaagaata | atcaattata | ccaatgtcct | ttctattttt | 555 |
| CVF1 | cag ATGATAA | TGAAGGTGGT | TTCATAGCAG | ATAGTGATAT | CATCTCAAGG | 605 |
| CVF1 | TCTGATTTCC | CCAAGAGTTG | GTTGGGCTAA | CAAAGGACTT | GACCGAGGAG | 655 |

Intron 18

Eine NdeI-Schnittstelle (catatg) ist unterstrichen dargestellt. Vorwärtsprimer: OPS22, Rückwärtsprimer: OPA21. Die Sequenz des internen Primers HB-Intr.18as (Sequenzier-Primer) ist als Pfeil markiert.

| | | | | | | |
|------|-------------------|-------------------|-------------------|--------------------|-------------------|------|
| CVF1 | AAGGACTTGA | CCGAGGAGCC | TAACAGTCAA | GG gtattgga | gttctgctgt | 50 |
| CVF1 | ttctattagc | ccacttttca | gntgaaatgc | aaattcaatg | actgcaaaaat | 100 |
| CVF1 | ttatacagga | atggggcagt | gtctcttcca | agaggaatca | atcacaaaaat | 150 |
| CVF1 | gttctaagtg | aaacttgggg | tttatttttg | acctccccta | aaagagaaaa | 200 |
| CVF1 | gcttttgctc | ttttctatgg | aagcaacttg | ttctttatta | tgggtcaatga | 250 |
| CVF1 | ccaaattcca | ttttaaaaag | gtagacatta | aaacttacaa | cataaaaaata | 300 |
| CVF1 | tgcaacacta | attaaataat | caatatgatt | gattacactg | agttttacaa | 350 |
| CVF1 | acagtaaaac | aaatgtttac | aacattgtcc | ctaaaataaa | tctgacttac | 400 |
| CVF1 | ccagaaattt | agctaataaa | aaggaattta | agacatggc | atgcttttct | 450 |
| CVF1 | aaaaccagca | gtgtctcaat | acatggactg | tattttcaac | tgctcttttg | 500 |
| CVF1 | ccactaggac | aaagtctttc | aaaatacagg | atctcccaa | ctttcctaca | 550 |
| CVF1 | ggaatagctg | gtggaaatgt | gtaaaaatgt | gctgtttagg | ctgttctaaa | 600 |
| CVF1 | tttaaagatt | aattttaata | tcttgctggt | ttaaaaaaag | gaaaactggc | 650 |
| CVF1 | tgctttttta | aaaaaacatc | cttaaggagg | gttgacaaat | gtaattatct | 700 |
| CVF1 | ttctggatct | ataattgatt | atthaaggct | tgatacacc | tgcttaaac | 750 |
| CVF1 | cacaccatcc | tttaaaagcc | atgaattttg | ataaattaat | cttagccaag | 800 |
| CVF1 | ttgatccaga | gtcatcagcc | aggataaaaa | aaaaccctag | gtccattgga | 850 |
| CVF1 | aaaaaaatag | tttcagccat | acgttggcat | <u>atgctgtttt</u> | ccccttctca | 900 |
| CVF1 | catcgcttca | cacttgaagg | actctattag | cagcaataat | gcatctttct | 950 |
| CVF1 | cttgggttct | actatatgga | gaacagtaga | aggatcagaa | aaaatggtag | 1000 |
| CVF1 | ttgtggagaa | agctgctttt | <u>cagctcttgc</u> | cacttggact | agaccggttt | 1050 |
| CVF1 | gttatgcatc | atctctcacc | acagggcaac | tttggagct | atcagtagaa | 1100 |
| CVF1 | tcttgaagga | aatgcagata | aaggaaatgt | gactgtgagt | cacctgaag | 1150 |
| CVF1 | ttgatttcaa | agttatgaat | aggtagttgg | tgattgttta | tccagcagca | 1200 |
| CVF1 | atcaagttgc | gttgttatat | ctatttcccc | ttcaaaataa | atgagtattt | 1250 |
| CVF1 | tattattaaa | tctactgtc | acaacatag | aatttaatga | aggaagtaac | 1300 |
| CVF1 | tgatgacaac | tttgatttgg | aactcatttg | tgtttcccnc | tgattctcag | 1350 |
| CVF1 | atthtttttt | gatctgtatg | attggtagct | ctaggaattg | tctatctggt | 1400 |
| CVF1 | ataagcttaa | tagcttatac | tgtaatagct | ctgcacagga | ataatgtaac | 1450 |
| CVF1 | acccttaatt | ttgttcctag | GATTTCAAGC | AAGACAATGT | CTTTTTATCT | 1500 |
| CVF1 | GAGGGATTCC | ATCACAACT | GGGTGGTGCT | G | | 1531 |

Intron 19

Vorwärtsprimer: OPS21, Rückwärtsprimer: OPA16.

| | | | | | | |
|------|-------------------|---------------------------|-------------------|-------------------|--------------------|-----|
| CVF1 | CAACCTGGGT | GGTGC TGGCT | GTAAGCTTTA | CACCCACCAA | AG gtaaagaa | 50 |
| CVF1 | ttatgaacaa | tggaactgaa | gttgacaggt | gtgaaaagca | ccttanaagt | 100 |
| CVF1 | aatgtanaat | taatgttttg | ggtgggcttt | gccattggct | ttccttgag | 150 |
| CVF1 | GGATCTGTGT | GGCTGAACCT | TATGAAATAA | GAGTCATGAA | AGTCTTCTTC | 200 |
| CVF1 | ATTGATCTTC | AAATGCCATA | TTCAGTAGTG | AAGAATGAGC | AGGTGGAGAT | 250 |
| CVF1 | TCGAGCTAT | CTGCACAACT | ACGT | | | 274 |

Intron 20

Vorwärtsprimer: OPS16, Rückwärtsprimer: OPA17.

| | | | | | | |
|------|-------------------|--------------------|--------------------|--------------------|--------------------|-----|
| CVF1 | GCTATTCTGC | ACA ACTACGT | TAACGAGGAT | ATTTAT gtaa | gtgttgggac | 50 |
| CVF1 | tgtgataaga | gcatgggagg | gtagagatgg | gagaaatatt | tagactgaat | 100 |
| CVF1 | agaaaagata | tgggtttaga | tagagacaaa | gatccaagag | aggagtatgg | 150 |
| CVF1 | gaggcaagca | aaaaaatgaa | tgagaattgg | agtattaaaa | cagacgagat | 200 |
| CVF1 | tacggataga | aaggaaatgc | cctttaatat | aatggtggcc | tgactgaaag | 250 |
| CVF1 | ctagaataga | ggtagaaata | agaatacaca | ttttatataa | tagaataaaa | 300 |
| CVF1 | gcaatagcta | gagggaaaga | ttagaggttg | aaagatggta | ttattgtgta | 350 |
| CVF1 | tgcttaaata | atatggcact | agaacgggtc | attgaaacat | tgaacaacta | 400 |
| CVF1 | aaataatgaa | ggattatatac | tcaatacaaa | catgtgtatt | aacactagga | 450 |
| CVF1 | ttacataagt | ttgactaatg | caacaataac | tattaccatt | aacattctta | 500 |
| CVF1 | ttgctctctg | tatcaaaacc | aacaatgcca | cactgttctg | tattcatggt | 550 |
| CVF1 | gatgtatatt | taaaaaataa | aataaaaaata | tataccaaaa | aaaagaaata | 600 |
| CVF1 | ctggtttaga | tcataatctc | tactccactg | ggcgccctta | ttgcataaaa | 650 |
| CVF1 | actctgatgc | attgaattgt | tcacacagtc | ctattttcta | ccagtctatg | 700 |
| CVF1 | aagtagcaaa | ggggaggata | aattgcaaga | tatgatgtat | gtgtatttgt | 750 |
| CVF1 | gctctgtgtg | taacccttt | ctgtctttcc | ttccatgtgt | cacaatgact | 800 |
| CVF1 | tgctggtgca | ggaacagatt | tggtggggaa | tgtgtttgtg | tcatctatgg | 850 |
| CVF1 | gaaagctgga | ttttaaataa | ttcttatggc | caaagctgcc | ttgtggctta | 900 |
| CVF1 | tgctcttttt | tcacctccac | tag GTGCGAG | TGGA ACTGTT | ATACA ACCCA | 950 |
| CVF1 | GCCTTCTGCA | GTGCTTCC | | | | 968 |

Intron 21

Vorwärtsprimer: OPS17, Rückwärtsprimer: OPA25.

| | | | | | | |
|------|-------------------|-------------------|-------------------|-------------------|--------------------|-----|
| CVF1 | CCCAGCCTTC | TGCAGTGCTT | CCACAAAAGG | ACAAAGATAC | CGACAGCAGT | 50 |
| CVF1 | TCCAATTA A | AGCCCTGTCC | TCCAGAGCAG | TACCGTTTGT | GATAGTCCCA | 100 |
| CVF1 | TTAGAGCAAG | GATTGCATGA | TGTTGAGATT | AAAGCAAGTG | TCCAGGAAGC | 150 |
| CVF1 | GTTGTGGTCA | GACGGTGTGA | GGAAGAACT | GAAAGTTGTA | gtaagtaaca | 200 |
| CVF1 | agtggggtta | cattttggct | cttggagaaa | gagagtcaat | aaccagagat | 250 |
| CVF1 | cgctgtagat | aaagactacc | ataattttac | taatgaatgc | atcttccoct | 300 |
| CVF1 | attgcaacaa | aaagacattt | tgagctcaaa | cattctgctt | gtaagtgcac | 350 |
| CVF1 | aaaataacaa | gtctaccact | gtggctaaaa | tggttctgat | tgtttttttt | 400 |
| CVF1 | taaactcttg | gaatcttcoct | atagcaaaat | ttggttgctt | tgaaaacggc | 450 |
| CVF1 | agaacggagg | cagagctccc | tgcttgggga | gttatttcat | gcaaagttaa | 500 |
| CVF1 | ttcatccttt | tcctactgng | catccaagca | ccaattctat | tactgaatat | 550 |
| CVF1 | tttactacat | ttctaagaga | tgctacaggg | gagtctttgt | attctgtagt | 600 |
| CVF1 | tgttcttttg | aaagtatttg | ttgctgtcat | gcataatctt | tttatttgca | 650 |
| CVF1 | atgaggctct | cagaagagat | tattctgggtg | tctttttcta | taacag CCTG | 700 |
| CVF1 | AAGGGTACA | GAAATCCATT | GT | | | 722 |

Intron 22

Vorwärtsprimer: OPS25, Rückwärtsprimer: OPA14.

| | | | | | | |
|------|-------------------|-------------------|-------------------|--------------------|-------------------|-----|
| CVF1 | GAAGGGGTAC | AGAAATCCAT | TGTGACTATT | GTAAACTGG | ACCCAAGGGC | 50 |
| CVF1 | AAAAGGAGTT | Ggtgagtaac | tctaattctt | actcattaat | tcccatcata | 100 |
| CVF1 | cgctgcgaga | tggcaaagtc | ctttcaccca | catgttttat | tgtttccagg | 150 |
| CVF1 | gaaaggagag | gatcttttta | ccatgccagt | ttcaataaaa | agttatttta | 200 |
| CVF1 | ctttgcaaaa | gttatgtgct | tgacagacaat | tatattcctt | cctattaagt | 250 |
| CVF1 | gcctgttaag | gtaattgtgt | taaatccat | ttacaattga | ttattagaaa | 300 |
| CVF1 | caatgcaaaa | gaggaatctc | atgtggaagt | aacatatttt | tcttttcttg | 350 |
| CVF1 | gaattttctt | cattttcttt | ttcatttttc | ag GTGGAACA | CAGCTAGAAG | 400 |
| CVF1 | TGATCAAAGC | CCGCAAATTA | GATGACAGAG | TGCCTGACAC | AGAA | 444 |

Intron 23

Vorwärtsprimer: HB23s, Rückwärtsprimer: HB24as.

| | | | | | | |
|------|-------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | GATGACAGAG | TGCCTGACAC | AGAAATTGAA | ACCAAGATTA | TCATCCAAGg | 50 |
| CVF1 | taggaactag | ttatttgaaa | tccacnctac | tacacctggt | gaacagagaa | 100 |
| CVF1 | aataaaatag | tcaccagccc | aatttccctg | tccttccatt | catgaattga | 150 |
| CVF1 | tgcagtgtag | accagagcct | tccttattgg | attcacaggt | acaatcgatc | 200 |
| CVF1 | tgtcccactt | ctaacagttt | ttataacaaa | taaggttatt | taactttatt | 250 |
| CVF1 | acttttgccc | atctgggtgag | atcagaccaa | gtgggtgtgc | tccaagtccct | 300 |
| CVF1 | gatatttaaa | cagtgtaac | ctgggggcc | tatgaatnct | tacttctgng | 350 |
| CVF1 | ctcagggcct | gcccctntgg | aatgatnttn | tctcctanat | ccatgcattn | 400 |
| CVF1 | aagaaatcat | ttaaacattt | ttntntctgt | gagcccctgn | acacagngtg | 450 |
| CVF1 | gttttctact | ggtattgggt | taatcttctg | atcagctagg | tctttttcta | 500 |
| CVF1 | tatatattgt | ttcttctttt | taactgtaag | ccagccagaa | ttgctaagga | 550 |
| CVF1 | atcatggaat | cttgaaactg | aagtaaataa | attagttaat | aaaccttcta | 600 |
| CVF1 | attagtctctg | gaattatcct | ggaagggtag | tggagtggga | agcaagatta | 650 |
| CVF1 | gactaatcag | atactgcagc | ttaaagcaag | atcacacact | accagcttc | 700 |
| CVF1 | aaccatgcct | ggcaattttt | tcacattcag | GTGACCCTGT | GGCTCAGATT | 750 |
| CVF1 | ATTGAA | | | | | 756 |

Intron 24

Vorwärtsprimer: SD24s, Rückwärtsprimer: HB25as.

| | | | | | | |
|------|--------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | GTGACCCTGT | GGCTCAGATT | ATTGAAA | CAATTGATGG | AAGTAAACTC | 50 |
| CVF1 | AACCATCTCA | TTATCACTCC | TTCTGGCTGT | GGGAGCAAA | ATATGATCCG | 100 |
| CVF1 | CATGGCCGCA | CCAGTTATTG | CCACCTACTA | CCTGGACACC | ACAGAGCAGT | 150 |
| CVF1 | GGGAGACTCT | CGGCATAAAT | CGCAGGACTG | AAGCTGTCAA | TCAGATCGTG | 200 |
| CVF1 | ACTG gtaagg | caaattgttt | tottcagggg | atggaaatga | gagatccttt | 250 |
| CVF1 | catccctaaa | atgaagaaat | gtaagacca | cagaagtcaa | gaagccaagt | 300 |
| CVF1 | gttctaaatg | tatggcaggg | ttcttttttg | tagttgcttt | ctttcacaag | 350 |
| CVF1 | gtaagtttcc | aatagctttg | ttaaggatat | cgcttcttta | acctggaagg | 400 |
| CVF1 | gtggaggtgg | tcatagtaac | tgctgagaaa | tatttaaata | tgcactcntt | 450 |
| CVF1 | atcacaatat | tctattaaaa | gacattttaa | tctcttttgt | ggtaagtttg | 500 |
| CVF1 | gaagtattgt | tctgatataca | aatgtctgtt | tctcagttgt | agaggaggt | 550 |
| CVF1 | aatatatgtt | ctattcatca | gctccttttg | tggttggtgt | taaggttgct | 600 |
| CVF1 | tgctttcata | atctcaacag | cagagtggga | tagttttact | ctgagaccag | 650 |
| CVF1 | aacacaaatc | aaaacaattg | tgcaaaggaa | aagcagctcc | aatgaggcac | 700 |
| CVF1 | gcagcagaga | aagaagagtg | caaaactggt | ctggatgcag | ttgtgatttt | 750 |
| CVF1 | gtacgactgg | agctaaattt | cacagctctc | tctatcttgt | cccttttcag | 800 |
| CVF1 | GTTATGCCCA | GCAGATGGTG | TACAAGA | | | 827 |

Intron 25

Beim EcoRI-Verdau entstanden Fragmente von ca. 3.3 kb, 2.6 kb, 800 bp, 750 bp und 400 bp.
Vorwärtsprimer: OPS15, Rückwärtsprimer: OPA26.

| | | | | | | |
|------|----------------------------------|-------------------|--------------------|-------------------|-------------------|-----|
| CVF1 | <u>ATGCCCAGCA</u> | <u>GATGGTGTAC</u> | <u>AAGAAAGCAG</u> | <u>ATCATTCCTA</u> | <u>TGCAGCATTT</u> | 50 |
| CVF1 | <u>ACA AACCGTG</u> | <u>CATCTAGTTC</u> | <u>TTG</u> gtaggga | caaaaatata | gccattactt | 100 |
| CVF1 | tcagatcaaa | agctttcctc | taacttcctc | tataaacggt | catatatttt | 150 |
| CVF1 | cattttctaa | cttctgctgg | tgcatctgat | tctttgtctt | tacattttgt | 200 |
| CVF1 | ttcttgcttt | cttcatttcc | tcatttctac | ctaccatttc | cttctttctc | 250 |
| CVF1 | gtagcttctt | ctagtgtgct | tcagtccttg | acaaaatact | cttttcta | 300 |
| CVF1 | aatggcttac | ccatataata | tgtcaccctc | ttgtctcctc | atgtctagca | 350 |
| CVF1 | tttagtatca | atgagttgta | ttgttttttc | cccttaaaat | aagaatcaga | 400 |
| CVF1 | atgtcctgct | tcaagatagc | gaaagcactt | gaaacatgga | ttccctgaga | 450 |
| CVF1 | ttgaattggt | gttgctgnta | ttattactag | attttttggg | atcaactcaa | 500 |
| CVF1 | gccagtgaac | aaacttaata | ttcctgccat | ttatttccc..... | | 539 |
| |(ca. 7000 bp)..... | | | | | |
| CVF1 |cagaga | | | | gagggaggga | 16 |
| CVF1 | aggaaagaga | gtggggaaag | agacagagag | agagaaagag | agaggctgaa | 66 |
| CVF1 | ctctgcctta | tttatcatta | ttttaggtag | taaaatctca | gtattcaggt | 116 |
| CVF1 | taaattgco | gtttggcact | ttgcaataat | taaatgggtt | ttgggttgca | 166 |
| CVF1 | gtttgagcac | ttggcctcta | aaaggtttcc | catcattgcc | ctagactcat | 216 |
| CVF1 | gcacaggttc | tactcttgct | gtgccataat | gtagtggtgt | ggtgactctg | 266 |
| CVF1 | gacaagccac | tcttcccctc | tccctcctag | ctagcagcaa | cctagccata | 316 |
| CVF1 | tttgtcttac | aaaatcaagc | aagcaactac | ttgtattcta | ctacttgat | 366 |
| CVF1 | taagatcttc | attgaaatgg | ttggttaatg | cattacattc | tttacagGCT | 416 |
| CVF1 | <u>AACAGCATAT</u> | <u>GTCGTAAAAG</u> | <u>TCTTTGCCAT</u> | <u>GGCTGCCAAA</u> | <u>ATGGTAGCAG</u> | 466 |

Intron 26

Die EcoRI-Schnittstellen im 5'-Bereich von Intron 26 und in Exon 27 sind unterstrichen dargestellt. Eine siebenfache Wiederholung der Sequenz ccct im 5'-Bereich sowie eine dreifache Wiederholung der Sequenz gtttt im 3'-Bereich sind ebenfalls unterstrichen. Vorwärtsprimer: OPS20, Rückwärtsprimer: HB27as.

| | | | | | | |
|------|----------------------------------|--------------------|--------------------|-------------------|-------------------|-----|
| CVF1 | <u>ATGCCCTGT</u> | <u>ACTTCTGGA</u> | <u>ACAATGCAG</u> g | tatggtttcc | <u>ttccaccctc</u> | 50 |
| CVF1 | <u>cctccctccc</u> | <u>tccctcccctc</u> | <u>cctctctctg</u> | ggattttttt | taagttctta | 100 |
| CVF1 | actgcaaaaca | atccaagccc | aggtattaaa | gaaaaaattg | cactgtacaa | 150 |
| CVF1 | agttctagct | tcttctctaa | taatattggc | cagtcaagag | ctgtacagga | 200 |
| CVF1 | <u>attcagacag</u> | <u>gacatgaggg</u> | <u>tgaattatta</u> | <u>tatctataga</u> | <u>tagagttgtc</u> | 250 |
| CVF1 | caggatattc | catcattaac | tataagtgga | gaggtgttgc | ccccgtaga | 300 |
| CVF1 | tagggcgcg | aacttgagc | tctcctgga | tgcgcgngt | agtttgagg | 350 |
| CVF1 | agcaggtggg | agctgtggc | aggggggat | ttgccaggt | tcgctgggt | 400 |
| CVF1 | caccagttgc | accctaacct | ggatcgagat | gccctacgta | cgg..... | 443 |
| |(ca. 2000 bp)..... | | | | | |
| CVF1 | ..ggaagt | ctcagaatat | gacactcctt | gaaattgtct | agaactaggc | 47 |
| CVF1 | catgtcccgc | ttaaagcaaa | taattaaac | attaaataaa | agaactcatt | 97 |
| CVF1 | tctcattctt | ttatagacta | tctagcagaa | aattcaattc | ccagagtgtg | 147 |
| CVF1 | ctgctccttg | ttcttatttt | ctctcctcat | gtaatgttct | taggataggt | 197 |
| CVF1 | tatagctccc | atataaccaa | agttaaatgg | ttcctcaaac | tccaaccctt | 247 |
| CVF1 | caacactgca | tcttcaagat | ggacaaacat | ggatggccac | tagatggcag | 297 |
| CVF1 | aacaagacat | ataaaaaacc | ttaactcttt | actccatata | cttgctatgt | 347 |
| CVF1 | attagatagc | atltattcat | tagcattggt | <u>tttgtttttg</u> | <u>ttttttcatt</u> | 397 |
| CVF1 | atattttatc | tctctcccct | gtccccccc | cctgcccccc | atccccatt | 447 |
| CVF1 | ttcttctagG | <u>GAGGAATTCA</u> | <u>AGGTGCTGAA</u> | <u>GAAGAAGTAT</u> | <u>ATTTAACAGC</u> | 497 |

Intron 27*Exon 27 stromabwärts*

Dargestellt ist der 3'-Bereich des Cosmid-Klons 1R67 von der EcoRI-Schnittstelle in Exon 27 bis zur MboI-Klonierungsstelle des Vektors. Vier weitere stromaufwärts gelegene MboI-Schnittstellen sind ebenfalls unterstrichen.

| | | | | | | |
|------|-------------------------|--------------------------|-------------------------|--------------------------------|-------------------------|------|
| CVF1 | GAATTC AAGG | TGCTGA AGAA | GAAGTAT AT | TAACAG C TTT | CATTCT GGTT | 50 |
| CVF1 | GCGTTG TTGG | AATCCA AAAC | AATCTG CAAT | GACTAT GTCA | <u>AT</u> gtacgtga | 100 |
| CVF1 | tg tt gtgatg | ggaaatggta | tgagggggcc | aaaataagat | tacatctctc | 150 |
| CVF1 | cacatagtat | ttttttttaa | tcatagtatt | ccat ttt ccc | accttcttga | 200 |
| CVF1 | ccccagg ttc | agtattaaaa | gctttaaagta | tttcaaaaata | gcctaaacag | 250 |
| CVF1 | cattgctgta | g ttt tatactt | tatataaacc | agcttcagct | ggtgcagaat | 300 |
| CVF1 | gcagccgcgc | gggtt gt gat | gggggtatcc | cgatactccc | atgttacagc | 350 |
| CVF1 | cctcctgcgc | ggcttgcatt | ggttgccggt | tg cct cccgg | gtcgccttca | 400 |
| CVF1 | agg t attagt | tacaacttat | aaagccctta | atggcttggg | tccagggtac | 450 |
| CVF1 | ttgcaggaac | gcctcctgcc | agcatcactt | <u>tccgat</u> cgcc | ctgttcgctc | 500 |
| CVF1 | tcataggagg | ggcctccttg | gggtgcccgc | agtcagatat | tgtaggcttg | 550 |
| CVF1 | cgg t ccccag | ggctagggcc | ttctctgtgg | gagcacctgc | cctttggaat | 600 |
| CVF1 | cagctccccg | <u>tcgat</u> ctccg | tttggcccct | gacctccgct | ccttccgccc | 650 |
| CVF1 | ggacctgaag | acttgg tt gt | tccaacgggc | cggcctt gaa | tgacctgggg | 700 |
| CVF1 | gggg tt gggt | ttat ttt aat | tttatttggc | ttg ttt actg | tttaattttg | 750 |
| CVF1 | ttttattgta | aat ttt tctc | tg ttt taatg | ttcgccgccc | tgagtcctta | 800 |
| CVF1 | gggaataggg | cggcatacaa | gtcgaataaa | taaataataa | aataaattca | 850 |
| CVF1 | tttcaaatca | aatgaaacca | aaataaatag | tcaattcaat | aaggatacta | 900 |
| CVF1 | ttaggaaatt | acacatatac | tttcaagcca | gcaacatatt | <u>tagat</u> cagtt | 950 |
| CVF1 | aat gat cttt | aaatatcaaa | gagcttttca | ctgctgcagt | ataat ttt tc | 1000 |
| CVF1 | atctgcctca | aaggcagatt | tagaaag atc | | | 1030 |

Exon 28 stromaufwärts

Dargestellt ist die Sequenz eines PvuII-Amplifikats (550 bp). Zwei StuI-Schnittstellen (aggcct) im Intron und in Exon 28 sowie eine DraI-Schnittstelle (tttaa) im Intron sind unterstrichen. Die erwartete PvuII-Schnittstelle (cagctg) am 5'-Ende konnte nicht identifiziert werden. Rückwärtsprimer: HB28as-CVF.

| | | | | | | |
|------|--------------------|-------------------|--------------------|-------------------|-------------------|-----|
| CVF1 | | ...caaag | ggcaggtgct | cccacagaga | aggccctagc | 35 |
| CVF1 | cctggggacc | gccagcgaca | atgtctgact | gacggcacc | caaggaggcc | 85 |
| CVF1 | cctcctatga | gagcgaacag | ggcgcctgga | aagtgatgct | ggcaggaggc | 135 |
| CVF1 | gttccccgaa | gtaccctgga | cgcgaatacc | tacaacattt | <u>aaaaatccac</u> | 185 |
| CVF1 | cacataacaa | accataaac | cactcccacc | ccgaaatca | aatcaaccac | 235 |
| CVF1 | ttgatctgct | ccatctccat | ttgtggctcc | caatgacaaa | acctggtttt | 285 |
| CVF1 | <u>taaggc</u> cttt | tgatagctca | gggtatcata | cctcagggaa | atgctgttcc | 335 |
| CVF1 | atagagatgc | ctaaaacatt | cctaattgat | ctcttctctt | ccctctcccc | 385 |
| CVF1 | tccttccag | AGTCTAGACA | GCAGCATCAA | GAAGGCCACA | AATTATTTAC | 435 |
| CVF1 | TCAAAAAGTA | TGAGAAACTG | CAAAGGCCCTT | ACACTACAGC | CCTCACAGCC | 485 |
| CVF1 | TATGCTTTGG | CTGCTGCAGA | CCAACTCAAT | GATGACAGGG | TACTCATGGC | 535 |

Intron 28

Vorwärtsprimer: HB28s, Rückwärtsprimer: HB29as.

| | | |
|------|--|-----|
| CVF1 | CAGACCAACT CAATGATGAC AGGGTACTCA TGGCAGCATC AACAG gtaca | 50 |
| CVF1 | catattgctt tgttttaaat agaagaacc aggtaagggtt gggttttgaa | 100 |
| CVF1 | aacagatgat tccactaaga agcagaatga aggattagct aggaatacaa | 150 |
| CVF1 | atcacgatgg gcagatTTTT agatattgca gccagttaca aatccatctt | 200 |
| CVF1 | gttggtgtgc catctctccc acattttgct agtttatcaa gaagtagatt | 250 |
| CVF1 | gttatctggt ttgttaaagt ccttgctgaa gtccaagtat attatTTTgc | 300 |
| CVF1 | tggctacta atttaagcat tttctggtc tactaattta gtcactttgt | 350 |
| CVF1 | taaagaatga aataggattg gatggcgatt atttgtttt aacanaccg | 400 |
| CVF1 | tgctggctac tacttaaac tatttgtttc tagatggtaa cagatctggt | 450 |
| CVF1 | TTTTgattt tttttcagt atcttcccag gtattgatgt taggctgatt | 500 |
| CVF1 | gctctgtatt t..... | 511 |
| | (ca. 1700 bp) | |
| CVF1 |tcag | 4 |
| CVF1 | tctgatgtt aatgaacca tccacaggtg tgaagatgt atggaacgg | 54 |
| CVF1 | tcctntgag aaagagataa ggagtcaccc caatgcgctt cttccatgag | 104 |
| CVF1 | gtttccgaa tgaaagaaa cagatggccc tcccctcttg tacattcctc | 154 |
| CVF1 | aagttaaatc aaagcacagg gcagaatagg tgacaggtaa aagttataaa | 204 |
| CVF1 | ctagggaaat gacaggatac ataggaaaca ggatgagagc aataaatgac | 254 |
| CVF1 | taaaagaatc cggggttccc ccaccacatg aatggcagta tcatagtagg | 304 |
| CVF1 | gatctgttaa gttgattctt catctacctg gtctacctgg aagtgtgat | 354 |
| CVF1 | cataactggg tatcatcagc aagtcacatg cattgagtaa cctcaccaca | 404 |
| CVF1 | caactcacac atagcatctg ctttcttgaa gaatgggaga cttaatcaat | 454 |
| CVF1 | agttaatTTT cccaatctt tcccttcaat caaaagtac catcaaggta | 504 |
| CVF1 | agtgaccaat attttatct gtttggctaa attctgtttc tgtgattggt | 554 |
| CVF1 | ttatTTtag AAGGGATCAT TGGGAAGAAT ACAATGCTCA CACCCACAAC | 604 |

Intron 29

Eine zehnfache taaa-Sequenzwiederholung ist unterstrichen. Vorwärtsprimer: HB29s-CVF, Rückwärtsprimer: HB30as-CVF.

| | | |
|------|--|-----|
| CVF1 | gGAAGGGATC ATTGGGAAGA ATACAATGCT CACACCCACA ACATTGAAGG | 50 |
| CVF1 | CACTTCCTAT GCCTTGTGG CCTTGCTGAA AATGAAGAAA TTTGATCAAA | 100 |
| CVF1 | CTGGTCCCAT AGTCAGATGG CTGACAGATC AGAATTTTA TGGGGAACA | 150 |
| CVF1 | TATGGACAAA CCCAA gtgag tctttgcttt attgtcagag tcttatcttc | 200 |
| CVF1 | tgagctctgt tctactttag atgcttaaaa caaacagta gtctgaattg | 250 |
| CVF1 | ttgcatgttt tgcctaaga agagaaggca aagtctttta atatttgaaa | 300 |
| CVF1 | tcctttcccg gggaagagggt tgtggattta ttctccacag gagaatagga | 350 |
| CVF1 | taagataaga aacagtaaat ggaagcttta cagagaggaa cttccanaaa | 400 |
| CVF1 | aagaatanga aggacttttc agattatgag tgctattaan caatgggcat | 450 |
| CVF1 | aacggtccat caggaggct..... | 469 |
| | (ca. 850 bp) | |
| CVF1 |ccacc tgagtcctta gggagaagg cggcatacaa | 36 |
| CVF1 | gtgtaaataa ataaataaat <u>aaataaataa ataaataaat</u> aaatataagt | 86 |
| CVF1 | caactctcct ctgtttgttt acttgttcag tttgtagccc atcttctcat | 136 |
| CVF1 | ttatgatcaa atgatctaac aaacaattgt aaaaatcaga aagatagtct | 186 |
| CVF1 | gggtatttct tgcaacaagc cccatcacca tgttcttccc tatgtgttaa | 236 |
| CVF1 | acatttatgt ctctggtaaa atatatcact ggattttcag ttcttttggt | 286 |
| CVF1 | agatactcca agacaaaca tcttattttt tccttctctc attctag GCA | 336 |
| CVF1 | ACAGTTATGG CATTCAAGC TCTTGCTGAA TATGAGATTC AGATGCCTAC | 386 |
| CVF1 | CCATAAGGAC TTAACCTAG ATATTACTAT TGAAGTCCA GATCGAGAAG | 436 |
| CVF1 | TACCTATAAG GTACAGAATT AATTATGAAA ATGCTCTCCT GGCTCGGACA | 486 |

Intron 30

Vorwärtsprimer: HB30s-CVF, Rückwärtsprimer: HB31as-CVF.

| | | | | | | |
|------|--------------------|--------------------|--------------------|--------------------|--------------------|-----|
| CVF1 | GATGCCTACC | CATAAGGACT | TAAACTTAGA | TATTACTATT | GAACTGCCAG | 50 |
| CVF1 | ATCGAGAAGT | ACCTATAAGG | TACAGAATTA | ATTATGAAAA | TGCTCTCCTG | 100 |
| CVF1 | GCTCGGACAG | TAGAG gtata | gtatttcagc | aagcatat | attactcctt | 150 |
| CVF1 | tcattcaaaa | taattat | ggttcaaat | caggttcaa | tgcactctt | 200 |
| CVF1 | ttcaagaact | aggaaaaact | atacagaatt | tagtgaaaca | ggaaaaaaaa | 250 |
| CVF1 | ctaaaccagt | tctctctatt | ggaattgtgt | agcagattgc | gggagactgg | 300 |
| CVF1 | cttaagacta | ctagcttctt | tatcaacata | caatgtagtt | ttaaattgact | 350 |
| CVF1 | tccctctcct | cacaaaaagc | tgcaaaaaggc | tgtatttaca | acaatattgc | 400 |
| CVF1 | aaaattgtgt | ctgtacttcc | aaggactatt | ccatctgtga | taagaaagag | 450 |
| CVF1 | at | ctctggggga | cttctggagc | acttcatgac | ttaggaaatg | 500 |
| CVF1 | aagataatat | cattaaaaca | attgygttgc | tacttctgat | gttgcataag | 550 |
| CVF1 | gaat | tcaaagtaga | tgcatgatac | caatagagnt | atatgtttaa | 600 |
| CVF1 | atatggatg..... | | | | | 609 |
| | | | | | | |
| | | (ca. 1500 bp) | | | | |
| | | | | | | |
| CVF1 | | | | ccattttt | tataataaaa | 18 |
| CVF1 | aagtcttgat | aatggatatg | gtagcacctg | gagacagcag | aagagaaggg | 68 |
| CVF1 | aaaggactgn | agaaaaccac | atcatataaa | cacctgcaaa | cagaaataga | 118 |
| CVF1 | aaataganca | actgtggcaa | agatagtatt | aaaagtaata | agacacttcg | 168 |
| CVF1 | gatacaatcc | ctaaacatct | ggaatccttt | attgctacga | atgtatatca | 218 |
| CVF1 | atacaaaaat | cgctaagatt | ggagcacaat | at | gttttgatt | 268 |
| CVF1 | ca | ctgtgtct | tcagaaagta | gaaggaaagg | agttttcttt | 318 |
| CVF1 | ttt | gttgata | cctttcta | atactgtcat | ttcttcattc | 368 |
| CVF1 | gtgag | tcttt | ttatatttgg | aaacctattc | caccactagc | 418 |
| CVF1 | ttaca | acaaa | gtcaataaaa | tactagcact | aataacaagt | 468 |
| CVF1 | tagcca | agga | actgcagctg | aaagtc | atgc | 518 |
| CVF1 | ACCAA ACTCA | ACCA AGACAT | CACT GTGACA | GCAT CAGGTG | ATG AAAAAGC | 568 |

Intron 32

Eine DraI-Schnittstelle (tttaaa) im 5'-Bereich des Introns ist unterstrichen dargestellt. Eine putative StuI-Schnittstelle (aggcct) befindet sich ca. 700 bp stromabwärts von Exon 32. Die 9 bp-Lücke in Exon 33 markiert den Größenunterschied zu den entsprechenden CVF2- und coC3-Exons. Vorwärtsprimer: HB31s-CVFgenom., Rückwärtsprimer: HB33as-CVF.

| | | Exon 31 | | | | |
|------|---|---------------------------|--|--|--|-----|
| CVF1 | <u>gttgcttttc</u> ag ACCAA ACT CAACCAAGAC ATCACTGTGA CAGCATCAGG | | | | | 50 |
| | | Exon 32 | | | | |
| CVF1 | TGATGGAAAA GCAACAATGA COATTTTGAC ATTCTATAAC GCACAGTTGC | | | | | 100 |
| CVF1 | AGGAGAAGGC AAATGTTTGC AATAAATTTT ATCTTAATGT TTCTGTTGAA | | | | | 150 |
| CVF1 | AACATCCACT TGA gtaagtt tgactggggg ctttggctt gctagaata | | | | | 200 |
| CVF1 | gaatctaataa aaatgatgg tggggggagt ttaaagtgt tctatcatc | | | | | 250 |
| CVF1 | atgtcaggat cccaatatgg aacaagcttt ctctattact tattcataaa | | | | | 300 |
| CVF1 | aagtttctca tttgatagaa aataacatta ttaattccat acttagttgt | | | | | 350 |
| CVF1 | aatccagttc ttgaagtcaa caaataatt tatttccttt tgcctggac | | | | | 400 |
| CVF1 | aaacagggcat tgagagcact gaaggacagg ttttgggtatt gaaatgta | | | | | 450 |
| CVF1 | gattggacag gacttgaagt tgtaggcaac agacatgta ctgtcattta | | | | | 500 |
| CVF1 | tatttcctttg gagacaaaa gaaaatccta ttagctgatg gctagcaaag | | | | | 550 |
| CVF1 | aagaggatat tagatccata ttgccaatat aattcctaac taatccaaa | | | | | 600 |
| CVF1 | ctaacagcaa ttgcagtcca cgagaggctt gcaaaacaa ttaggatttg | | | | | 650 |
| CVF1 | gac..... | | | | | 653 |
| | | (ca. 1100 bp) | | | | |
| CVF1 |ctgcaaga tgagcttaat gctttactac | | | | | 28 |
| CVF1 | ctgttgtggt attcttgcaa tatttcaata caatagggtta ggaaaaacag | | | | | 78 |
| CVF1 | tgatacaccc acacacgcac acatagatac acacacaaaa catgtaacag | | | | | 128 |
| CVF1 | caaattgtat aaaaaaac cctaaatttg gttcttgcaa aacgaagaat | | | | | 178 |
| CVF1 | aaattgtatt ttacactcag ctgggaaatt tctagcactt acactttaga | | | | | 228 |
| CVF1 | atcaaaacaa ccaattctcc agttattgaa actgggtttg ttattctttc | | | | | 278 |
| CVF1 | tttaaacatt ccttattcac atgatttctt tcaagaagga accaatcata | | | | | 328 |
| CVF1 | aaaatagaga gtaaaaatta ttagattaat ctttgggtt ccaaggagat | | | | | 378 |
| CVF1 | gttacagcca gacaatcaaa ctttcagct ttaaaaaact gctagaattt | | | | | 428 |
| CVF1 | gtgcagagtc cgattttctc ttttcttttg taatttttaa acttcaaag- | | | | | 477 |
| CVF1 | ----- AT GCAATGGGAG CCAAGGGAGC CCTCATGCTC AAGATCTGCA | | | | | 520 |

Intron 33

Vorwärtsprimer: HB33s-CVF, Rückwärtsprimer: HB34as-CVF.

| | | | | | | |
|------|---|--|--|--|--|-----|
| CVF1 | GGGAGCCAAG GGAGCCCTCA TGCTCAAGAT CTGCACAAGg taagaatgct | | | | | 50 |
| CVF1 | gaacactaaa gaaaactgga ttgtaacagg atatttagaa ttataacttt | | | | | 100 |
| CVF1 | gagaagcaat tatgttactt ttcaagaaaa taattatact ttcccaact | | | | | 150 |
| CVF1 | ttacactggt tctctgcttg tatattgatc tcttttctgt gcatcatcct | | | | | 200 |
| CVF1 | aatttghtaat attttghtgt ttgtggatga acaaaagagc agaatttaag | | | | | 250 |
| CVF1 | tactgcaga gtgcagatct tttagattga tttagggatg cagaaattgc | | | | | 300 |
| CVF1 | aatatgctac agctcgatgg ccagtacatg aatgggaaac tccaggagac | | | | | 350 |
| CVF1 | tctanaactg ttggctagac tcagaagttg aggaaaaataa cccaggagaa | | | | | 400 |
| CVF1 | ggtaatggca attatattcc atgttgatgg gaaaatttca caaacttggt | | | | | 450 |
| CVF1 | catcagacat tgaactggac tcaagagaac tatagagtga gaccaagtct | | | | | 500 |
| CVF1 | tcaacttgggt tttgccattt aaagaaaaaa aatacattcc attgtatgcc | | | | | 550 |
| CVF1 | ttaatcaatg ctttttactc agttttcac caaattcata agctttgcaa | | | | | 600 |
| CVF1 | aaacttatat gaaaattggt atgggagtta ggagacata cattgcggtg | | | | | 650 |
| CVF1 | gctcatgtgc cagcgtataa agagctgtca attgctagat ttctccctt | | | | | 700 |
| CVF1 | caagatgaat cattttccaa aacacagcac tgcctcata gcaatgaagg | | | | | 750 |
| CVF1 | ggaaactggt tgtgaattag gaaatccatc tttggtcatt tttcctccat | | | | | 800 |
| CVF1 | tcttgatttt ag GTATCTGG GAGAAGTTGA TTCTACAATG ACAATAATTG | | | | | 850 |
| CVF1 | ATATTTCTAT GCTGACTGGT TTTCTCCCTG ATGCTGAAGA CCTTACAA | | | | | 898 |

Intron 34

Vorwärtsprimer: HB34s-CVF, Rückwärtsprimer: HB35as-CVF.

| | | | | | | |
|------|---------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-----|
| CVF1 | <u>TTGATATTTCT</u> | <u>TATGCTGACT</u> | <u>GGTTTTCTCC</u> | <u>CTGATGCTGA</u> | <u>AGACCTTACA</u> | 50 |
| CVF1 | AGG gtaggca | agctttgcat | gggccttgaa | atggctgttt | ctgtagttca | 100 |
| CVF1 | gcaaagatgc | caaattagaa | aatgtgtaat | agttaatgcc | tatatgatta | 150 |
| CVF1 | aaactgtgga | gagaatgaat | gaatattcat | tattagtaaa | tgaatcaata | 200 |
| CVF1 | agatgctgaa | ttttatgaca | cataaaaagg | cattatataat | taataaatca | 250 |
| CVF1 | gaaaagtgca | taacttctaa | tgattctgga | atctctctga | caaattaata | 300 |
| CVF1 | atatgaaaac | tcttttttgt | tgtctccaag | tctctcaggt | gtacatcctg | 350 |
| CVF1 | cgaaccttga | aaatattttc | agatttttgc | ataaatgtaa | cacagtgcca | 400 |
| CVF1 | gtataaatgg | cttcagttat | ctgaaattcc | atgtacttga | atcatctgcg | 450 |
| CVF1 | actcctgaag | tcaatgtcat | atcttctttc | ctttcgttat | ttag CTTTCT | 500 |
| CVF1 | AAAGGAGTGG | ACAGATACAT | CTCCAGATAT | GAAGTTGACA | ATAATATGGC | 550 |
| CVF1 | <u>TCAGAAAGTA</u> | | | | | 560 |

Intron 35

Eine DraI-Schnittstelle im 3'-Bereich von Intron 35 ist unterstrichen. Vorwärtsprimer: HB35s-CVF, Rückwärtsprimer: HB36as-CVF.

| | | | | | | |
|------|--------------------------|--------------------------|--------------------------|--------------------------|---------------------------|-----|
| CVF1 | <u>TGGACAGATA</u> | <u>CATCTCCAGA</u> | <u>TATGAAGTTG</u> | <u>ACAATAATAT</u> | <u>GGCTCAGAAA</u> | 50 |
| CVF1 | GTAGCTGTTA | TCATTTACTT | AAACAAG gta | aatacttatg | gaaccattt | 100 |
| CVF1 | ctttgagtaa | tcacatggtt | ggttatttgg | tttgctagcc | aaactccgga | 150 |
| CVF1 | gtgtattata | ctgaaaaaga | tagagtgtat | atntagcaaa | ataaacctac | 200 |
| CVF1 | ataatccatc | attaactgca | acagatcatg | cttcccacat | atattagtta | 250 |
| CVF1 | actgattggc | aaaaagttta | tctatttctt | cgaatgggt | gggtggctca | 300 |
| CVF1 | tagtatacac | ctatggcaat | ctcactcgtt | tttcctttta | tcttgacca | 350 |
| CVF1 | tatacaagtt | tttatctttg | atcatgggct | tttctgtaga | tatataatga | 400 |
| CVF1 | tctcttttaa | tatagtgcaa | ctccacttcc | tctttttgta | gatctgtttc | 450 |
| CVF1 | ttttgaacaa | tttgtatcca | ttcatcaatt | cattccagtc | atgagtttca | 500 |
| CVF1 | ccccaccaa | tttttggaat | ggcaactata | tcatacctac | cctcatgcac | 550 |
| CVF1 | tattgcttct | agttcatcct | gcttggttct | ttcaagaatt | aa..... | 592 |
| | | | | | (ca. 1400 bp) | |
| CVF1 | | | |aatcctg | cactctccct | 17 |
| CVF1 | gtgtgttcaa | cagctgccgc | atggaaatcc | attgtgccgc | gtgaagcagc | 67 |
| CVF1 | tgaacttcaa | aaccaggaaa | taaaggccag | aatagagcca | gggcaggcag | 117 |
| CVF1 | gcaggcccag | caactgatca | ctggagtgaa | cgggtttgct | tacaaatgat | 167 |
| CVF1 | ccttgaggct | accggttcag | caaaagcggg | ctgaatcagt | agaatatac | 217 |
| CVF1 | ccctggatac | atgacagaat | agtgtgtaaa | ttctctaact | tttgaccagt | 267 |
| CVF1 | atcttctgag | acaccattta | aaccttctgt | tcctctcacc | ttttccattg | 317 |
| CVF1 | cttcccttta | ttctgggaaa | cctggacaga | tttgacagaat | ttagtccttc | 367 |
| CVF1 | aaccactggg | tcaaacctgt | ggaacttggg | ataagcagtt | tccttgaggt | 417 |
| CVF1 | tattcggatg | gagttttggg | tgggggggaa | ttatctccat | gggtaactg | 467 |
| CVF1 | gcaccacaaa | gtgatcttct | ctctccatcc | cag GTCTCC | ACTCTGAAGA | 517 |
| CVF1 | TGAATGCCTG | CACTTTAAGA | TTCTCAAGCA | TTTTGAAGTT | GGCTTCATTC | 567 |

Intron 36*Exon 36 stromabwärts*

Im Intronbereich ist eine Wiederholung der Sequenz aagatatataaa auffällig (unterstrichen). Zwei PvuII-Schnittstellen (cagctg) sind unterstrichen. Die erwartete DraI-Schnittstelle (tttaa) am 3'-Ende konnte nur partiell identifiziert werden. Vorwärtsprimer: HB36s-CVF.

| | | | | | | |
|------|-------------------|---------------------|-------------------|-------------------|--------------------|------|
| CVF1 | AGATGAATGC | CTGCACTTTA | AGATTCTCAA | GCATTTTGAA | GTTGGCTTCA | 50 |
| CVF1 | TTCAGCCAGG | ATCAGTCAAG | GTGTACAGCT | ACTACAATCT | AG gtaatggc | 100 |
| CVF1 | agacgtactg | aatac <u>ac</u> ctc | cctgctccta | atccccac | aacaaccact | 150 |
| CVF1 | ctgtgaagtg | agtcaagatt | gtgagcgaat | ggcttggcca | aggcatocca | 200 |
| CVF1 | gttgctttc | atggctaagg | cgggacttta | ttcatggctc | ccccctttct | 250 |
| CVF1 | agtctggtgc | tttaaccatt | aggccacact | ggctctttcc | aaggaactca | 300 |
| CVF1 | gtggtactgg | agcttgagaa | aattagccaa | aaagcccaac | aacaatgcat | 350 |
| CVF1 | ggttgataaa | acattcttga | ttgattttcc | aaaaactct | ttataacaat | 400 |
| CVF1 | tttataaaag | ctgaagtga | gaatgaatca | tttgaatcat | tatcaatact | 450 |
| CVF1 | <u>gaagatata</u> | <u>taaagttata</u> | <u>ttaaagatat</u> | <u>attaaacaca</u> | aagctctgcc | 500 |
| CVF1 | caactactcg | acagacttct | ggtctgagtt | tcctttgctt | gccttctgag | 550 |
| CVF1 | cagagttggt | attcagcagg | ttctgaccag | ttctggagaa | ctggtagcgg | 600 |
| CVF1 | aaattttgag | tagttcggag | aactggcaaa | tgccacctct | gactggccct | 650 |
| CVF1 | gcccccttct | attatctgcc | ttctgactcc | <u>cagctgatca</u> | <u>gctgggtctt</u> | 700 |
| CVF1 | ctttggttgc | cctgcacagg | agaatggagc | tggaaagcag | gtagtgggg | 750 |
| CVF1 | tggggacgga | atgggatttt | gcagtaactt | tcccctgcca | ttcccaccaa | 800 |
| CVF1 | gccacgcccc | cagaaccagt | agtaaaaaaa | tgtttgaatc | ccaccgctgc | 850 |
| CVF1 | ttctgagggg | cagaggacaa | agccagtgac | accaaggtcc | tgtttttctc | 900 |
| CVF1 | tgctctggtg | ggtaacgtga | tgctcatctt | agggggggag | actttggaag | 950 |
| CVF1 | ctgagcttgt | catggtctgag | gaggagtttt | gccattctaa | taagtatttt | 1000 |
| CVF1 | gaaagtcat | taccattttt | atctctgaat | agcattttta | gtgcagctct | 1050 |
| CVF1 | ttaaccagga | attgcgcaaa | atcatgaaa | tacttcagag | aaaattacct | 1100 |
| CVF1 | gtaaaacctg | tatatagtag | gcacaattgc | aaaaaaaaag | cattctttgg | 1150 |
| CVF1 | aaaacatact | ctcaaatgca | tacacattcc | taaaatgta | aaactataca | 1200 |
| CVF1 | aaataattgc | tt | | | | 1212 |

Exon 37 stromaufwärts

Die DraI-Schnittstelle (tttaa) am 5'-Ende sowie eine EcoRI-Schnittstelle (gaattc) sind unterstrichen. Rückwärtsprimer: HB37as-CVF.

| | | | | | | |
|------|-------------------|--------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | | | | <u>tt</u> | <u>taaac</u> taca | 12 |
| CVF1 | tacataacag | taagcattgt | actgtcacac | cacaatccct | ttcttatggt | 62 |
| CVF1 | aatggagca | tcacatgcat | gccccaaaaca | tacaacacat | tgccattttc | 112 |
| CVF1 | atcttttttg | acaaaaggag | <u>cagaattctg</u> | caggtatcca | tgtttcaatt | 162 |
| CVF1 | tcctaactcc | attagataat | ggcatatcct | ctcctacatt | tcttgatgca | 212 |
| CVF1 | ttgtagcact | ctctgcttcc | tgggtaaaatg | cttaacaaaa | ctcattccgt | 262 |
| CVF1 | caaccaacc | tcttgataag | ttccattgt | actctcaggt | tagacatgcc | 312 |
| CVF1 | ccaaatttga | aatctctttg | agaataatct | ttgtacttct | aaatacaggt | 362 |
| CVF1 | agtattcctc | aacttatgaa | cacaattgaa | cccaacattt | atggtgctaa | 412 |
| CVF1 | gtgagacatt | tgtaagtga | gttttgcccc | attttacaac | ctttcttgcc | 462 |
| CVF1 | acagtcgta | agtgaatcac | tgagttggt | aagtgagtaa | tacagttttt | 512 |
| CVF1 | aagcgaatct | ggctccgct | ttgattttgc | ttgtgaaaag | gtctcaaaaag | 562 |
| CVF1 | gtggttacat | aactcacgga | taccgcaacc | gtcataaata | tgaactagtt | 612 |
| CVF1 | gtcaagcatc | caaatgtaaa | tcacatgacc | atggggattt | tgcaatggtc | 662 |
| CVF1 | ataagtgtga | aaaatggcca | taagtcaact | tttcagtgct | attacaactt | 712 |
| CVF1 | tgaacagtca | ttaaatgaat | gattgtaagt | tgaggactag | ctgtaattaa | 762 |
| CVF1 | atctatactc | accagagttt | cttcatgtga | tgatggata | taacttttaa | 812 |
| CVF1 | ttctccatt | ag ATGAAAAA | TGTACCAAGT | TCTACCATCC | AGATAAAGGA | 862 |
| CVF1 | ACAGGCCTTC | TCAATAAGAT | ATGTATTGGT | AACGTTTGCC | GATGTGCAGG | 912 |

Intron 37

Vorwärtsprimer: HB37s-CVF, Rückwärtsprimer: HB38as-CVF.

| | | | | | | |
|------|--------------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | TACCAAGTTC | TACCATCCAG | ATAAAGGAAC | AGGCCTTCTC | AATAAGATAT | 50 |
| CVF1 | GTATTGGTAA | CGTTTGCCGA | TGTGCAGGAG | gtaagttggg | agaaataagg | 100 |
| CVF1 | taaacgttga | ctctgagcat | cagaatggaa | gcggtgggga | agagttagta | 150 |
| CVF1 | tgggatctat | ggcaaata | aatttaattc | aaattgaagc | aatgataatt | 200 |
| CVF1 | tttctcaatc | tgaatctgtt | tggaaacttt | ggactctgta | ctctgtaaaa | 250 |
| CVF1 | taatatgagg | agagagttgc | tggatcttaa | tagttgcagt | acttcataaa | 300 |
| CVF1 | atgtttacta | aggagtttac | atgtgattga | agatgcattg | tttctacaca | 350 |
| CVF1 | tttctatggt | tatttatgag | ctgccccact | ttttaaatcc | aacgtagtac | 400 |
| CVF1 | caaagcaatt | aatatcaagg | aaataaagta | aaacaaaaat | gtatccatac | 450 |
| CVF1 | aatatgaata | tttaacaaat | accattctcc | tgcgggacat | gcccctctct | 500 |
| CVF1 | gatgacctg | gccggtttcc | ttccatttgc | ctgttactca | ttcccctgag | 550 |
| CVF1 | agtcagtcag | cagttgggca | agc..... | | | 573 |
| | | | | | | |
| | (ca. 2500 bp)..... | | | | | |
| | | | | | | |
| CVF1 |ctatgtagg | | | | | 9 |
| CVF1 | attaggttaa | caactattaa | tgccttatct | aaggggcagc | cacattctcc | 59 |
| CVF1 | aaagcaccta | aaggcagaac | atgaagcaaa | ggatgagagg | caacctagaa | 109 |
| CVF1 | ctaaggagaa | atctactctg | aagagaacaa | ttaaccagtg | gaacaacttg | 159 |
| CVF1 | cttccagaag | ttgtgagtgc | tcaaacaatg | gaagttttta | aaaagagatt | 209 |
| CVF1 | ggacaacat | tcgtctgaaa | tggtacagag | tttcctgcct | gagcagggtt | 259 |
| CVF1 | tggaactagat | ctccaaggtt | ccttacaacc | ctatctgttg | ttctgctggt | 309 |
| CVF1 | tgaataaaca | cattatcaaa | taactatgct | tatagatgag | ccctagatat | 359 |
| CVF1 | acggtgccac | atntaaatta | gtgtttcttg | tttctgatta | tcaaacacag | 409 |
| CVF1 | atgtcttcta | caaatgagat | gttatctgct | ctgggctggc | tactttttgc | 459 |
| CVF1 | cctaacaatg | tcttttaaat | caaaacatgt | gtcctcaaca | atcaacccaa | 509 |
| CVF1 | cattttgtct | tctcttttct | ccccttgaag | AAACCTGTTC | CTCGCTCAAC | 559 |
| CVF1 | CATCAGGAAA | GGATTGATGT | TCCATTACAA | ATTGAAAAAG | CCTGCGAGAC | 609 |

Intron 38

Vorwärtsprimer: HB38s-CVF, Rückwärtsprimer: HB39as-CVF.

| | | | | | | |
|------|--|-------------------|-------------------|-------------------|-------------------|-----|
| CVF1 | CTCAACCATC | AGGAAAGGAT | TGATGTTCCA | TTACAAATG | AAAAAGCCTG | 50 |
| CVF1 | CGAGACGAAT | GTGGATTATG | gtatttata | tgtagaatct | tattaataag | 100 |
| CVF1 | gtagtagtgt | ctttcattgt | taacagtggt | ctaactgtca | agaagtcatg | 150 |
| CVF1 | atntgaagca | accattttca | gtttattcat | aatgaagcaa | ccattttcaa | 200 |
| CVF1 | tttatttgta | atattcatgg | atnttttacc | cttggaatgc | tttgctgctg | 250 |
| CVF1 | ctacacacct | tcacccctga | gtggtatttt | tctagttaat | gtggtgctac | 300 |
| CVF1 | ttgttggtgt | gttattgacc | ctccagtgta | gcataacaat | tcggggctcc | 350 |
| CVF1 | tagaaataga | tctgtatcaa | tgagtaagca | gagccaaaac | accgggattt | 400 |
| CVF1 | ggagaccaat | tocacgctcc | acacatttac | tgntttcact | ttcaatcctt | 450 |
| CVF1 | caaaaagcat | catgaagcca | aactgcagta | atctgcatgg | ccccgtatat | 500 |
| CVF1 | caaaattgag | agatgaaatg | ctaccttgct | gtgaagtggg | gtgagatgg. | 549 |
| | | | | | | |
| | (ca. 670 bp)..... | | | | | |
| | | | | | | |
| CVF1 |cctttgg gacaacctg acctggatga ttcggaatct | | | | | 37 |
| CVF1 | tcgtagagtg | tgccctcactg | cttatgccat | tcatgggtga | aaataatcag | 87 |
| CVF1 | ctgtcttaat | gaggattggt | tctgtctttg | ttcttaaaag | TCTACAAAAC | 137 |
| CVF1 | CAAGCTGCTT | CGAATAGAAG | AACAAGATGG | TAATGATATC | TATGTCATGG | 187 |

Intron 39

Beim EcoRI-Verdau entstanden Fragmente von ca. 6000, 200 und 100 bp. Eine EcoRI-Schnittstelle im 3'-Bereich von Intron 39 ist unterstrichen dargestellt. Vorwärtsprimer: HB39s-CVF, Rückwärtsprimer: HB40as-CVF.

| | | | | | | |
|------|---------------------------|--------------------------|---|--------------------------|--------------------------|-----|
| CVF1 | <u>ACCAAGCTGC</u> | <u>TTCGAATAGA</u> | <u>AGAA</u> <u>CAAGAT</u> | <u>GGTAATGATA</u> | <u>TCTATGTCAT</u> | 50 |
| CVF1 | <u>GGATGTTTTA</u> | <u>GAAGTTATTA</u> | <u>AACAAG</u> gtaa | gaatttcccc | ctaaaatata | 100 |
| CVF1 | tcacaggtgt | caaacttatg | ttgccatggt | ggcatcatgt | gatgtattgg | 150 |
| CVF1 | gattttcccc | cttcgctaaa | ctgggtgggg | gcagggccag | cacgtgatgc | 200 |
| CVF1 | atccggcctg | tgggccacga | gttaacaccc | ctgatataatc | acataattaaa | 250 |
| CVF1 | tagtggaaac | atthtcaacta | aacctgaaag | tttattttgaa | ccccatttga | 300 |
| CVF1 | aagttcacat | tcatgtatth | agcttgatth | aatatattca | ttttaattgc | 350 |
| CVF1 | ccccatatct | aaagcgactc | tgggcaatgt | acgatacaat | acaacaacaaa | 400 |
| CVF1 | ctgaaaatgc | aatatagtga | aattattaat | ggaagaggac | aatgaaataa | 450 |
| CVF1 | agtacaggcg | cacaggctga | atccagctga | taacatccac | actcccaaat | 500 |
| CVF1 | gtacctgacg | ggaactcctc | taaccacaaag | tcttgaag..... | | 538 |
| | | | | | | |
| | (ca. 5400 bp) | | | | | |
| | | | | | | |
| CVF1 | ...ccaagta | tagatcctct | acctaggatc | gaactcacag | cctcntgatt | 47 |
| CVF1 | gtgaggcaag | agcttcacca | ntagggcacc | accacatttt | tattgttcta | 97 |
| CVF1 | ccaattttct | gtatcctgat | aaatctgcc | cagtgtgtca | gaagatcctc | 147 |
| CVF1 | cagatgagtg | ggggaaggat | ctggtttagg | ggagttgccg | atagcctccc | 197 |
| CVF1 | ttccctctgc | aaaaatgcc | cccttttctg | tttgttagtt | gaactttacc | 247 |
| CVF1 | atccacagct | ctagtgtctga | gatagggaac | tggtttttca | gggtcagatg | 297 |
| CVF1 | ctgctgcagt | atctgtgtgg | cgtataatgg | cctctctgag | gaaaaaggta | 347 |
| CVF1 | gaactgagaa | gatggtttat | ttgcccagg | gtgataaata | tgggcaata | 397 |
| CVF1 | gcagttgaag | aaagaattca | cccactaaat | acagctgctt | cctctttgcc | 447 |
| CVF1 | agttacttgc | cattagccat | gtaatacata | cagtatgtga | attaaaccaa | 497 |
| CVF1 | tgaaggcta | taccctaata | gaaatattta | tccagattct | aatttctgca | 547 |
| CVF1 | g <u>GTACTGACG</u> | <u>AAAATCCAG</u> | <u>AGCAAAGACC</u> | <u>CACCAGTACA</u> | <u>TAAGTCAAAG</u> | 597 |

Intron 40

Zwei EcoRI-Schnittstellen in Intron 40 sowie eine EcoRI-Schnittstelle in Exon 41 sind unterstrichen dargestellt. Vorwärtsprimer: HB40s-CVF, Rückwärtsprimer: HB41as-CVF.

| | | | | | | |
|------|---------------------------|--------------------------|---------------------------|--------------------------|--------------------------|-----|
| CVF1 | <u>AGGCTCTGAA</u> | <u>TCTGAAGGTG</u> | <u>AATGATGATT</u> | <u>ATCTGATCTG</u> | <u>GGGTCCAGG</u> | 50 |
| CVF1 | <u>AGTGACCTGT</u> | <u>TGCCACGAA</u> | <u>AGATAA</u> gtac | gtatttcaact | atthtttttt | 100 |
| CVF1 | aattttctaa | aattttatag | actttgcttt | tctgctatca | aatccacaga | 150 |
| CVF1 | acatctaaca | atataattac | agtgtcatag | aatgttgtat | tacaactttg | 200 |
| CVF1 | aaagcaacca | taatctaaat | ctgttaatat | caaaatagtc | attcttctctc | 250 |
| CVF1 | agttttatca | gatgccatgg | aagacttcta | taggttagat | gtgacattgt | 300 |
| CVF1 | ggggaagact | gtgaaataat | tttaactaca | atthtttagat | acctagataa | 350 |
| CVF1 | caccatctgc | ccaactatat | ctactggga | atatgtaaaa | cttttaccca | 400 |
| CVF1 | gatgcttcaa | accacacggg | ccgtatgctc | tatgtcctgc | tgaagagctg | 450 |
| CVF1 | tgcttgctg | caggatctt | ttctatcaca | acctcacatt | tcccagtgcc | 500 |
| CVF1 | tccagtgcag | gcctgcaagg | atctatattt | tattctattt | ttattttatt | 550 |
| CVF1 | ngaatgtcaa | acatgtacaa | gataacaagt | ataggtatga | a..... | 591 |
| | | | | | | |
| | (ca. 1300 bp) | | | | | |
| | | | | | | |
| CVF1 |tctt | aaagtctccc | tgaatttagg | | | 24 |
| CVF1 | attgtcttag | ttatcacaag | cctgatgcat | ctthtattct | tcatgacctt | 74 |
| CVF1 | tgagaagttc | gaactgcagg | ccactggagg | ggtttttcat | tccatttggg | 124 |
| CVF1 | ttccacaaat | acacctacta | ccatctgttg | atagtatata | tctgtagcat | 174 |
| CVF1 | tcagatccaa | ttttctgctc | gacatctcaa | ctgaaatgct | taaatcttag | 224 |
| CVF1 | aactggggtg | agaaatctat | gagacaaaac | tgatttatca | ttgaattcat | 274 |
| CVF1 | tcgttcattg | ttgtagcaat | catgttttat | ggcagcagcc | atttgaattt | 324 |
| CVF1 | tctagaattc | agttgtttta | ggattcagat | tttaaggaaa | aagaaaagtt | 374 |
| CVF1 | tcttctctta | ccctaaaatc | aaaatagaaa | ttaaaatgct | atattcatag | 424 |
| CVF1 | gatgaaatga | ttaaaaacta | tattggattc | agtccaagca | cagcacatta | 474 |
| CVF1 | ctccaacagg | agacagttga | actaatctcc | ttgcagggct | aagttttaag | 524 |
| CVF1 | ctagcctgct | tcctctttcc | ag <u>AATTTCCT</u> | <u>ACATCATTAC</u> | <u>AAAGAACACA</u> | 574 |
| CVF1 | <u>TGGATTGAGA</u> | <u>GATGGCCACA</u> | <u>TGAAGACGAA</u> | <u>TGTCAGGAAG</u> | <u>AAGAATTCCA</u> | 624 |

Anhang 7.2: Sequenzen des CVF2-Gens

Die Promotorregion des CVF2-Gens ist im Ergebnisteil (Kapitel 3.2) aufgeführt.

Nachfolgend sind in den Exonbereichen der CVF2-Sequenzen Unterschiede zu den entsprechenden CVF1-Sequenzen durch Markierungen hervorgehoben (wie in Abb. 3.2, 3.19 und 3.29).

Intron 1

Exon 2 stromaufwärts

Die StuI-Schnittstelle am 5'-Ende ist unterstrichen. Die dargestellte Primersequenz (CVF1-Sequenz, unterstrichen) weicht in zwei Nukleotiden (eingerahmt) von der CVF2-Gensequenz ab (ca-ag). Rückwärtsprimer: HB2as-CVF.

| | | | | | | | |
|------|--------------------------|------------------|-------------------|-------------------|-------------------|-------------------|-----|
| CVF2 | <u>aggcctcagg</u> | ctcaatccct | gccgcgccgg | tggaaggatt | ccagatgctt | 50 | |
| CVF2 | ccaccggcgg | agactggagc | agcggtgac | ccgtccgctt | gggctgccgt | 100 | |
| CVF2 | ccaggaatca | actgaactgt | gcaggtccaa | tcaagccccg | aagagccccg | 150 | |
| CVF2 | cctacttggtg | tgactacgta | aggtaggcag | ccccggattg | gtcgactgga | 200 | |
| CVF2 | ctaagaagg | aggggtgttc | tatatgccca | tatctggaca | atgagaggag | 250 | |
| CVF2 | tcggctggag | actcggetca | ttggctcagc | cactggcggg | aaaaacaaag | 300 | |
| CVF2 | tcccgtcagt | ggctgaagcg | gcccagacagc | ccggctataa | aagccggcgc | 350 | |
| CVF2 | ccgccgcgtg | ttcttttagtc | tatthttgaa | ataaagagct | g..... | 391 | |
| | (ca. 250 bp) | | | | | | |
| CVF2 | | | |gcat | tggtcattat | 14 | |
| CVF2 | aacgtgttga | ttacattaata | tatcacagag | gttactgtac | cagtaatggt | 64 | |
| CVF2 | cagtctcatt | ttataactaa | aaaaaccca | gcaccccaa | tatcttttaa | 114 | |
| CVF2 | gtgatataaa | gatgcaatag | atttactcat | ttttcagcta | tatttgaaaa | 164 | |
| CVF2 | ctttgtgggt | cacatcttta | accatttatt | tcttcagttt | tctaaaagaa | 214 | |
| CVF2 | gagatcatga | aaaacttcat | caaatgcctc | acttttccta | agcaaaactaa | 264 | |
| CVF2 | ctcaggagat | gttcttttct | tcttctgcag | CTACACCCTT | ATCACCCTG | 314 | |
| CVF2 | ATGTTCTGCG | AACAGACA | GAAGAGCAA | TTTTGGTGGA | GGCCATGGA | 364 | |
| CVF2 | GACAGTACTC | CAAACAGCT | TGACATCTTT | GTTCA | CAATT | TTCCACAGAA | 414 |
| CVF2 | GCAGAAA | TTGTACCAA | CCAGAGTAGA | TATGAATCCA | GCAGGAGGCA | 464 | |

Intron 2

Für den 5'-Bereich ist die Konsensus-Sequenz aus einem genomischen PCR-Produkt (ca. 10 kb) und einem entsprechenden Amplifikat der DraI-Bibliothek (ca. 680 bp, Exon 2 stromabwärts) dargestellt. Schnittstellen für PvuII (cagctg) und DraI (ttaaa) im 5'-Bereich der Intronsequenz sowie für StuI (aggcct) im 3'-Bereich sind unterstrichen. Auffällig ist eine dreifach repetitive aac-Sequenz (unterstrichen). Die dargestellte Sequenz des sense-Primers (unterstrichen) weicht in einem Nukleotid (eingerahmt) von der putativen CVF2-Gensequenz ab (C-T). Vorwärtsprimer: HB2s-CVF, Rückwärtsprimer: HB3as-II-CVF.

| | | |
|------|---|-----|
| CVF2 | <u>CC</u> C <u>ATGGAGA</u> CAGTACTCCA AAACAGCTTG ACATCTTTGT TCAC <u>AATTTT</u> | 50 |
| CVF2 | CCAC <u>GAAGC</u> AGAAA <u>TCTT</u> GT <u>ACCAAACC</u> AGAGTAGATA TGAATCCAGA | 100 |
| CVF2 | GGGAGGCATG CTTGTCACCTC CAACTATA AA Ggtgggtgta tggatgatggt | 150 |
| CVF2 | tgcaatagag tcttttggtat ccctgcgaaa gaacaraaca <u>aacaaacaaa</u> | 200 |
| CVF2 | <u>cccaaaactaa</u> aatcctgtcc agtttgctt gagggtaagt ccyatgcacc | 250 |
| CVF2 | ttaatcaaac atgattagat tgtcagatat tgagagaaat tgagcatata | 300 |
| CVF2 | attctaagga ggaaagcaaa atgaaatfff gggatatttg tgtatttatt | 350 |
| CVF2 | ttacagntct tattgccaca attgctgaaa gattctctgc agcatttcac | 400 |
| CVF2 | agcaaataag taaaatataa taaatattta gatcaaaaac agtatcaaaa | 450 |
| CVF2 | ccaaaaccac aggtctgact ctgctaacta cttgaaccca atgaaatgta | 500 |
| CVF2 | <u>aacagctgct</u> ggcactgga gggacaata gactgagaaa agttctgcaa | 550 |
| CVF2 | attggaccat gatcatacta acacaaatgc tacatttcta tatgcagtga | 600 |
| CVF2 | tttccatgag taggcatatt aaatgtccta acctgacagt gatgaaat <u>tt</u> | 650 |
| CVF2 | <u>taaa</u> | 654 |
| |(ca. 9000 bp)..... | |
| CVF2 |agagagaaa aggccagaga gancaagggtg | 29 |
| CVF2 | gggtttgtgg gtgtactgca gtggacagat gtggcttgta acataatgta | 79 |
| CVF2 | gcaagatgtg gcaaggcgag atgtcaggga aatgttctgt attagtttct | 129 |
| CVF2 | ttgactgggn tgetttaatt ggatgttaca tggcttttcc attgccttct | 179 |
| CVF2 | gccaggaaaa atcagaaatt tccttagtag aagcaagaag caagtttttc | 229 |
| CVF2 | agattaaata catctatggt tggcattctg tgatcctctt tgttttcatt | 279 |
| CVF2 | tttgactcag caggcaagat ttttgacca atgttttagt gctgttgccc | 329 |
| CVF2 | tgagcgaagg aatg <u>aggcct</u> caaatggccc ttagtgatga ttaagtttga | 379 |
| CVF2 | agcctgtgct ntactcatct tctcatccag GTTC TGCAA AAGAA CTGAA | 429 |
| CVF2 | TA AGGACTCC A AGCAAAATC AATATGTGGT TGTGCAAGTA ACTGGTCTCTC | 479 |

An Position 55 der Intronsequenz (r = a oder g) wurde im 10 kb-Amplifikat ein „a“ und im entsprechenden Genome Walking-Amplifikat der DraI-Bibliothek ein „g“ gefunden. An Position 112 der Intronsequenz (y = c oder t) wurde im 10 kb-Amplifikat ein „t“ und im Genome Walking-Amplifikat ein „c“ gefunden.

Intron 18

Vorwärtsprimer: CVF2-Ex18s, Rückwärtsprimer: CVF2-Ex19as.

```

CVF2  gCTGATTTTG AAGATGATT ATTTCAGAA GGTAACATCA CCTCAAGGTC 50
CVF2  TGATTTTCC CAGAGTTGGT TGTGGCTAAT GGAGCAGCTG TCTGAACATC 100
CVF2  CTAACAGTAA AGGgtattgg agttcttctg tttatattag cccacttttc 150
CVF2  agttgaaag aaaatctaaa tattgcaaaa tttatacaga aatggggaag 200
CVF2  atgggtgtctc ttccaagagg aaaggygctg aatgaaacct gggatttatt 250
CVF2  tttaataaat ctcttctctt tagaagagaa a..... 281
.....
..... (ca. 550 bp) .....
.....
CVF2  .ccaggaaaa ggtagaagtg gacaaagctg cttttcagct cttgccactt 49
CVF2  ggactagacc ggtttgttg tctcctacca aagggcaact ttggaagcta 99
CVF2  ctagtagaat cttgaaggaa atgcagataa aggaaatgtg acagtgagt 149
CVF2  accttgacgt tgatttcaaa gttatgaata ggtagtttat tatttatcca 199
CVF2  gcagcaatca aattgcattg tcatatctgt tccccatatt actgttgatt 249
CVF2  actttcaaaa tgagtgtttt attttgaac cgctctgttg caaacttaga 299
CVF2  atttaatggg agaaatagct gatgacaact tccatttga acttgttgt 349
CVF2  atttccaatc aacgctgtat ttgtagttgt aggagttgtc tatctaagc 399
CVF2  tagtttttaa atgcacagta taatagctct gtaaattacg cgacacattt 449
CVF2  cattttgtcc ctagGATTTC AAGCAAGATA GTACCTTTT ATCTGAGGGA 499
CVF2  TTCCATCACA ACCTGGGACT TGCTGGCTGT GGGCCTTTCA CCCACCAAAG 549

```

Intron 32

Eine Abweichung zur CVF2-cDNA-Sequenz besteht an Position 497 der genomischen CVF2-Sequenz (A-G) auf Exon 33-Seite (umrahmt dargestellt). Eine EcoRI-Schnittstelle im 3'-Bereich von Intron 32 ist unterstrichen dargestellt. Intron 31 ist im CVF2-Gen nicht vorhanden. Vorwärtsprimer: HB-31s-CVF2,coC3; Rückwärtsprimer: HB33as-CVF2.

```

                Exon 31
CVF2  gACCAAATC AACGAAGACT TCACTGTGTC AGCATCAGGT GATGGAAAAG 50
                |
                Exon 32
CVF2  CAACAATGAC CATTTTGACG GTCTATAATG CACAATTGAG GGAGCATGCA 100
CVF2  AATGTTTGCA ACAAATTC TCCTGATGTT TCTGTTGAAA ACGTCAGTT 150
CVF2  GAgtaagttt ggcttcctag aaatagtagg ttgtaaggta ttaaaatttt 200
CVF2  agtgttcacc atgtcagtat ctcaatatga aacaagtttt ctctgctatt 250
CVF2  tattcataaa aaaagtctct aatttgatag aaaataatct tattaacttc 300
CVF2  atacttagtt gagatccagt acttgtagtc aacaaaccag ttgggatttg 350
CVF2  gacagagggg caaatagcca caggccagaa ggcctctgtt gtccactaag 400
CVF2  tcttcaactgt tctatcattt aactagttgc aacaactctt aagatttgag 450
CVF2  gcgaaagatg ccaaagagaa aatgactta atagcattct taattttaga 500
CVF2  aactctttag aaacttttga ctgg..... 524
.....
..... (ca. 2850 bp) .....
.....
CVF2  .....aa aaactggttc ttgcaaaatt atgatgaaag gcattttaca 42
CVF2  ccagcttga aagttcctag ctcatacatt ttagaacagg agtgtcaaat 92
CVF2  tggaggcca tgggcccagat gcatcacgtg caggtcatgc ccaccccagc 142
CVF2  tcagcacaag ggaaaaacat tacaaaacag cacatgacag caacatgatg 192
CVF2  cagcgagttt gacatgctgt ctttagaagc tacacaacca attcttcagt 242
CVF2  aattggaat gggctctggt tttatttatt attctttggt taacattcc 292
CVF2  ctattcatat gatatcttct aggaattctt tcaaaaacca atccatttta 342
CVF2  gaaatagagg gcaaaaagtga ttagattaat tattggtttt ccaaggagat 392
CVF2  gttatggcca gttagacaat caaaccttct acctttaaaa aaactgctaa 442
CVF2  gaatttggtc agagtctgat tttccccttt tctttggcaa ttcttagACT 492
CVF2  TAAAGGAGCG AAAGGGAGCC AAGGGAGCCC TCAAGCTCAA AATCTGCAC 542

```

Anhang 7.3: Sequenzen des coC3-Gens

Die Promotorregion des coC3-Gens ist im Ergebnisteil (Kapitel 3.3) dargestellt.

Nachfolgend sind in den Exonbereichen der coC3-Sequenzen Unterschiede zu den entsprechenden CVF1-Sequenzen durch Markierungen hervorgehoben (wie in Abb. 3.2, 3.19 und 3.29).

Intron 1

Exon 1 stromabwärts

Dargestellt ist die Sequenz eines Amplifikats der StuI-Bibliothek ab der Bindungsstelle des Primers HB1s. Eine PvuII-Schnittstelle (cagctg), eine EcoRV-Schnittstelle (gatatc) sowie zwei DraI-Schnittstellen (ttaaa) sind unterstrichen. Vorwärtsprimer: HB1s.

| | | | | | | |
|------|-------------------------|--|------------------------|-------------------|-------------------|------|
| coC3 | | | <u>GGTGGCT</u> | <u>GCTCTATTGA</u> | <u>TTGGTTTTCC</u> | 27 |
| coC3 | <u>AGGGTCTTCC</u> | <u>CA[■]GGGGCTC</u> | Tgtaagtatc | agagatattt | cagcagggg | 77 |
| coC3 | aataaacgag | tgctgataaa | cccttggtg | cctttgcagg | gtaatgaatt | 127 |
| coC3 | tccttccagg | tgtctctaaa | cactttgagc | tttctcagaa | agtccaatat | 177 |
| coC3 | aacctatgac | aagagtaaaa | gtggctcagg | ctgttaaatg | cctgctctta | 227 |
| coC3 | agatccagct | gtctacaatt | actgtaggct | cgaatcccac | caggctcaag | 277 |
| coC3 | gttgactcag | ccttccatcc | tcctggggta | ggtaaattga | ggaccaaga | 327 |
| coC3 | ttgttggggg | caattagctg | accttgtaaa | aatgccgctt | agtaataagc | 377 |
| coC3 | aggcatatac | aatggaatga | aggctattgc | tgagcacact | gggttgatc | 427 |
| coC3 | ccccaaaaat | tgtaagccgc | cctgagtctt | cggagaaggg | cggcatataa | 477 |
| coC3 | atgtaattaa | aaaaaataat | ataatcaata | aggggaactg | acaggccctg | 527 |
| coC3 | aagtctctaa | ctgaggaatt | gtccttgagg | tgaagattct | ctggacaaat | 577 |
| coC3 | ataaagacct | cttgttctca | gttacagcgg | tggggtgcgt | cagtggtagg | 627 |
| coC3 | gaatctgtat | taccattgct | aagcagagtc | aaaaacttcc | agaaaacaat | 677 |
| coC3 | aaagttgaat | atthagaaaa | cagtcatccc | acattctttg | ttcctgatgt | 727 |
| coC3 | cttcatgatg | cctttgttca | tgatgccatc | atggttaaggc | ttcagtgaca | 777 |
| coC3 | tagagactgg | actacctgct | tgttttgaga | aaatatttat | ttcactgtgt | 827 |
| coC3 | acctatactg | gaaaaccaga | tatgactttg | cagcaggcat | acaatatgtg | 877 |
| coC3 | taagtaaatg | acgtatgatg | <u>atatcagcaa</u> | tgaggatgac | tatatttctc | 927 |
| coC3 | tgggtgccagt | atggagaatc | agctactgaa | aaagcaaaag | acaggagaag | 977 |
| coC3 | gaggaggcag | agaaggtgat | ttgtcagtat | acataataa | actcttattg | 1027 |
| coC3 | caggaaagaa | agataagagg | tcagttctgg | tcccatttgg | gatttggcaa | 1077 |
| coC3 | acaaaactctc | tgtagatgta | aatccaaagc | agggccatcc | aaaagagacg | 1127 |
| coC3 | aacaaatctc | aaaactacgc | taggtgggga | ttttgttagg | aacttggcat | 1177 |
| coC3 | ggaatccaaa | ttgatttatt | <u>gcttttaaa</u> | | | 1206 |
| | | | | | | |
| | (ca. 80 bp) | | | | | |
| | | | | | | |
| coC3 | atccatggtc | aggctgccta | agacattggt | <u>ttaaaggagc</u> | agctcttaaa | 50 |
| coC3 | gttcataaaa | gaatgataac | tcctaattgt | tcatgctcaa | ttactctttg | 100 |
| coC3 | aattgatgtg | aagaatataa | tgtaatata | atatggttac | aactgtcagg | 150 |
| coC3 | ccattataac | acaagcaaca | gctgtgcaat | attgggtatg | tgatcattgc | 200 |
| coC3 | agcagaagaa | agcagtgttt | gtaaaggcat | atgtggatta | gcacaattaa | 250 |
| coC3 | ttattgacaa | ttatcaccat | gaaaccttat | ctgaacttca | gttcttggct | 300 |
| coC3 | aagaggaagt | aaatTTTTgt | taattcaagg | aaacgtatga | agggtttcat | 350 |
| coC3 | ggtgatcatt | cttaggcatt | catactgtat | tcctctgcac | tctgtggaat | 400 |
| coC3 | gctgaccagt | ggtgttattt | gaagtgtctt | gtactttttt | ttagtacact | 450 |
| coC3 | gggcctggga | gatcctgctg | gccaggccct | ttatggataa | acttgcctcat | 500 |
| coC3 | atctgctgta | agttggatgc | tggatgggca | aggttttctc | atggcacaag | 550 |
| coC3 | agaaata..... | (ca. 40 bp bis StuI-Schnittstelle aggccct) | | | | 575 |

Intron 2

Eine EcoRI-Schnittstelle (gaattc), eine DraI-Schnittstelle (ttaaa) sowie eine StuI-Schnittstelle (aggcct) im 3'-Bereich von Intron 2 sind unterstrichen dargestellt. Auffällig ist neben einer vierfachen ct-Wiederholung vor allem die mehrfache gt-Wiederholung (unterstrichen). y = c oder t. Vorwärtsprimer: HB2s-CVF, Rückwärtsprimer: HB3as-CVF. Die 3'-Enden beider Primer sind im Sinne der CVF2-Sequenz modifiziert.

| | | | | | | |
|------|---------------------------|--------------------|---------------------|-------------------|--------------------|-----|
| coC3 | <u>CCC</u> ATGGAGA | CAGTACTCCA | AAAC <u>CG</u> CTTG | ACATCTTTGT | TCATGATTTT | 50 |
| coC3 | CCACGGGAAGC | AGAAAACCTT | GTTCCAAAC | AGAGTAGATA | TGAATCAGGC | 100 |
| coC3 | AGGAAGCATG | <u>TTTG</u> TCACTC | CAACTATAAA | Ggtgggtgta | tggtgatggt | 150 |
| coC3 | tgcaatagag | tctactgtat | tctgtcaaat | atctatatct | atctattat | 200 |
| coC3 | ataaaaatgg | <u>ctctctctgt</u> | <u>gtgtacgtgt</u> | <u>gtgtgtgtgt</u> | <u>gtttatgatc</u> | 250 |
| coC3 | cagcataact | ctggaacgcc | ttgagcaatt | tcaaccaa | at | 300 |
| coC3 | gatgacttac | tctctggaag | caaataattgt | gggggtaaga | caccctaac | 350 |
| coC3 | agccctcggg | ggatatgttc | tgттаagata | cagcctacag | tgcaatggag | 400 |
| coC3 | ttgctatggt | aacggcttca | cagtactcca | cgagggggcg | tttghtaagg | 450 |
| coC3 | ggcgtgyctg | caccactacg | agatcgacaa | aatgtacaca | cagataacat | 500 |
| coC3 | acaaaatgac | atctcttgat | gctt..... | | | 524 |
| | (ca. 1600 bp) | | | | | |
| coC3 |gtgtgtt | | | | | 7 |
| coC3 | aatctgtgtg | tatctccatg | tatttgata | atccaataat | ttcccaatga | 57 |
| coC3 | aaggaaaaga | gagaatagct | ctcatagccc | agatggtaga | attcattaat | 107 |
| coC3 | gtgctgggat | tcatgtttct | ccaggttgag | gcggtcgacc | aggaaagt | 157 |
| coC3 | attgggggtga | gagttccctc | attctccctc | ccacttgtct | tgagcttgga | 207 |
| coC3 | attaggggct | gaattttctt | ggtgtgtaaa | ttagggagga | aaaggttggtg | 257 |
| coC3 | ggtacacggc | agtgggcaa | ggacaccaag | tgtgacttgt | aacagcagaa | 307 |
| coC3 | tggaaggtg | agagatctca | gggatatggt | cttcattagt | ttcattgact | 357 |
| coC3 | gggctgcttt | agctggacac | tataaggcct | ctccttagcc | ttctgccagg | 407 |
| coC3 | aaagatcaga | aattgattta | gcagaagcaa | gcaatatggt | ttaagattt | 457 |
| coC3 | aatagttcta | tgtgtggcat | tctgtgatcc | tctttgtttt | catttttgac | 507 |
| coC3 | tcagaagaaa | attttgatc | aatgttttag | tgttgttgcc | tagggctgag | 557 |
| coC3 | gaatgaggcc | tcagggtggc | cttagtgatg | attaagttt | aagcctctgc | 607 |
| coC3 | tctactcatc | ttctcatcca | g <u>CTTCC</u> TGCA | AAAGAACTGA | <u>ATAAGG</u> ACTC | 657 |
| coC3 | <u>CAAG</u> CAAAAT | <u>CAG</u> TATGTGG | <u>TTGTG</u> AAAGT | AACTGGTCCT | CAAGTGAAT | 707 |

Intron 3

Vorwärtsprimer: YL3s-C3, Rückwärtsprimer: YL4as-C3.

| | | | | | | |
|------|--------------------------|-------------------|--------------------|-------------------|-------------------|-----|
| coC3 | TGAAAGTAAC | TGGTCCTCAA | GTGCCATTGG | AAAAGGTGGT | TCTCCTTTCT | 50 |
| coC3 | TACCAGAGTG | GCTTTCTGTT | CATCCAGACA | GATAAAGGCA | TCTATACACC | 100 |
| coC3 | AGGCTCTCCA | Ggtaaccogt | tttttctgcc | tctattttctt | cactcccctg | 150 |
| coC3 | gggtccact | atggttatgt | gatgaataat | actataaaac | aaaaccattc | 200 |
| coC3 | ctgggatttg | ttgtttattc | aatatatttc | cagggaatgg | agaatgtat | 250 |
| coC3 | gtcttgtgag | tggtggcatt | ggtgatattg | agaaatcaag | cttagaaagt | 300 |
| coC3 | tcaggaaaga | tttaagaaac | acaaagatag | atggttcagt | gcagtgggtg | 350 |
| coC3 | tattctacca | gttcggacca | gttcacctga | accggtagga | aaaaagcct | 400 |
| coC3 | cccagatcag | ctgagaggag | aaagaagaga | aaaacgcagg | atcagccctc | 450 |
| coC3 | ctaccacac | ctgcactttt | atctccttta | tcctctcacc | tgattgggtc | 500 |
| coC3 | tgca..... | | | | | 504 |
| | (ca. 650 bp) | | | | | |
| coC3 | ..tttacagt | gagttcggca | gacagctgct | gtcatggttg | aggcttggct | 48 |
| coC3 | cactttgtaa | accccgctta | cagccagaca | agaagaatgg | ggggaggag | 98 |
| coC3 | ggaatagatt | ggaatgcagg | ggaggcataa | cggggtcaaa | atggggggga | 148 |
| coC3 | aaggactgca | aagagcacta | agttggagcc | tgagaaccaa | tgtcatctgt | 198 |
| coC3 | tgagaaagtc | tgctaggtgg | agaatccgtc | acccaacaa | actgaactga | 248 |
| coC3 | actgaccgta | ttgattggat | ggaacaaagg | ggaacattta | actatgctgt | 298 |
| coC3 | ttgctcatga | aaacgttagt | gttggtcttg | ctacttctgt | atccatggtg | 348 |
| coC3 | gctctgtaa | ttatactctt | ggtaataaaa | atgtcccagg | agtgtttatt | 398 |
| coC3 | tgctgaatga | accttgtgga | cggttgacag | ctacaaaatg | tcatacaatt | 448 |
| coC3 | tatgacatta | ctttttgctg | atgatgatgc | aaacacatca | atcaatcaat | 498 |
| coC3 | ttttttatta | tggtcactgg | ccagcatgat | gcaagcacat | atcttctgaa | 548 |
| coC3 | tactatcaca | aattacagtt | cgtctctttt | tttcttagTG | CGTTATCGTG | 598 |
| coC3 | TCTTTTCTGT | GGATCACAAC | ATGCA CAGGA | TGCACAAAAC | TGTGATTGTC | 648 |

Intron 16

Eine EcoRI-Schnittstelle (unterstrichen) befindet sich 209 bp stromabwärts von Exon 16. Auffällig ist eine 11fache Wiederholung der Sequenz tagaa (unterstrichen). Die Sequenz des CVF1-spezifischen Rückwärtsprimers weicht an einer Position (umrahmt) von der entsprechenden coC3-cDNA-Sequenz ab(G-T). Vorwärtsprimer: OPS72, Rückwärtsprimer: OPA24.

| | | | | | | |
|------|--------------------|-------------------|-------------------|-------------------|-------------------|-----|
| coC3 | CTGCAAATCG | GAGGCGTCGC | AGTTCTGTTT | TGCTGCTTGA | CAGCAAAGCA | 50 |
| coC3 | AGCAAAG gta | tgtggcaaat | aatgaagttt | ctgtttactc | tgaggtaaaa | 100 |
| coC3 | gaataaaact | acagtccact | tgcattgatt | gaatcaatca | ggaaatggag | 150 |
| coC3 | gaaggaatgg | gggaagatta | tcccctttc | ccttctttct | caagcaagag | 200 |
| coC3 | ggtagaaaat | taataatctc | tattaaatag | gatcattctt | aacctatcaa | 250 |
| coC3 | taactttgag | tgcttgaatt | cagagtgcaa | aggggggagt | acttaagggg | 300 |
| coC3 | cccaatagtt | taatgtagaa | tatacagtaa | ctggttccag | ctaaagagat | 350 |
| coC3 | tacatttgat | acacaagctt | tttgagtgtt | ctcagatcgc | agcactgatc | 400 |
| coC3 | agaatttgct | cttcatgagg | aggaagtaag | cagtcagtgg | ccagtgaact | 450 |
| coC3 | acaagcagct | cccaaaccag | gaacatctgt | gggaatgcct | gcatatcaac | 500 |
| coC3 | atggcagttc | caactgaaca | atcaaagctc | tgcccctttt | atatgtatgg | 550 |
| coC3 | ggcttcaate | atccccttcc | taaagacttc | catgcccaca | agcctttttt | 600 |
| coC3 | taaaaactta | tgaatcacca | ggacagaata | gaatgaatag | <u>aatagaatag</u> | 650 |
| coC3 | <u>aatagaatag</u> | <u>aatagaatag</u> | <u>aatagaatag</u> | <u>aatagaatag</u> | <u>aatagaatag</u> | 700 |
| coC3 | aatgaatcac | caggacagta | taacagcaga | ataccagaa | attatttcat | 750 |
| coC3 | ttgattgttt | tatgtagata | ctcattgtct | cttgtttccc | gttatgcaca | 800 |
| coC3 | tgactatgag | aaataggatt | ctctaagttt | tcacgcattt | tgcatatttt | 850 |
| coC3 | ag CGGCACAG | TTTCAGGATC | AAGCCCTGCG | TAAATGCTGT | GAAGATGTC | 900 |
| coC3 | TGC | | | | | 903 |

Intron 29

Auffällig ist eine fünffache acc-Wiederholung (unterstrichen). Vorwärtsprimer: YL29s-C3, Rückwärtsprimer: YL30as-C3.

| | | | | | | | |
|------|-------------------|-------------------------|-------------------|-------------------|-------------------|--------------------|------|
| coC3 | ag | GAAGG AAT | CGTTGGAAG | AATATAATGC | TCGCACCCAT | AAATATTGAAG | 50 |
| coC3 | GCACTTCCTA | TGCCTTGTTG | GCCCTGCTGA | AAATGAAGAA | ATTTGCTCAG | | 100 |
| coC3 | GTCGGTCCTC | TAGTCAGATG | GCTGATAGAT | CAGAAATATT | ATGGGGCAAC | | 150 |
| coC3 | ATATGGACAA | ACCCAA gtga | gtctttgttt | tattgtcaga | ttataacctc | | 200 |
| coC3 | ctgaattgta | ttctatttta | ggctggttta | aacaaaataa | taatctgaat | | 250 |
| coC3 | tgttgcattg | ttcacttaaa | gaaggggagg | caaagttgag | acactatagc | | 300 |
| coC3 | agtcttttaa | tatttgaagt | gctgtcacag | ggaagaggtt | gtggatttat | | 350 |
| coC3 | tctctatagt | gcctaaggat | aagataagaa | acagtgaatg | aaaactttac | | 400 |
| coC3 | agagaaagac | <u>caccaccacc</u> | <u>acccaaaaaa</u> | <u>taaaaaagaa</u> | taaggagtgt | | 450 |
| coC3 | tttttttact | gtacttttga | tactactcag | aataagggat | tgaataactg | | 500 |
| coC3 | ctgtaggtct | tgactaattc | tgatagactg | aaataaactt | cagtcactctg | | 550 |
| coC3 | caagcttttt | tttggcacia | aatatccata | ggagaaaaatg | ttttgtgctg | | 600 |
| coC3 | agaaaagatc | agggtgtata | cttaccatgt | ggatagacaa | ttgtttatct | | 650 |
| coC3 | ctacttgatt | agcagacaag | agcatggatc | ttttgatgtg | caggacgaag | | 700 |
| coC3 | aatagttttt | gaaaggaatt | tttgcctaaa | gattatctag | ggtttatgtc | | 750 |
| coC3 | cttcatccat | gctgaaatta | agggaaacca | gagaaatctc | atccatagac | | 800 |
| coC3 | ttcctattgt | cctagatcag | ttagatatac | agataaaaag | tagtttaaag | | 850 |
| coC3 | atgtatgtat | gtgtatgtaa | gtcaactctt | ctctgtttgt | ttactcatca | | 900 |
| coC3 | atztatagcc | catcttctca | tttatgatca | aatgatctaa | cagacaattg | | 950 |
| coC3 | taaaaattag | aaggatagtt | tgggtatttc | ttgcaacaag | ccacatcacc | | 1000 |
| coC3 | atgttcttcc | ctatgtgtta | aacattttatt | tctctggtta | aatatatcac | | 1050 |
| coC3 | tggattttca | gttcttttgt | gagatactcc | aagacaaaaca | atcttatttt | | 1100 |
| coC3 | ttccttctcc | cattctagGC | AACAGTTATG | GTCCTTCAAG | CTCTTGCTGA | | 1150 |
| coC3 | ATATGAGATT | CAGATGCCTA | CCCATCAGGA | CTTAAATTTA | GATATTCTA | | 1200 |
| coC3 | TTAAACTGCC | AGAACGAGAA | GTACCTGAAA | GGTACAGCAT | TAATGATAGA | | 1250 |
| coC3 | AATGCTCTCC | AGGCCCGGAC | AG | | | | 1272 |

Intron 32

Die genomische coC3-Sequenz weicht in Exon 32 von der entsprechenden cDNA-Sequenz (Position 4026, A-G) ab (umrahmt). Vorwärtsprimer: HB31s-CVF2, coC3; Rückwärtsprimer: HB33as-coC3.

| | | | | | | | |
|--------------------------|-----------------------------------|---------------------------|-------------------|-------------------|--------------------|-------------------|-----|
| | | Exon 31 | | | | | |
| coC3 | g | ACCAA ACTC | AACGAAGACT | TCACTGTGTC | AGCATCAGGT | GATGGAAAAG | 50 |
| | | Exon 32 | | | | | |
| coC3 | CAACAATGAC | CATTTTGACG | GTCTATAATG | CACAATTGGG | GGAGGATGCA | | 100 |
| coC3 | AATGTTTGCA | ACAAATTC CA | TCTTCATGTT | TCTGTTGAAA | ACCTCCAAAT | | 150 |
| coC3 | GA gtaagttt | ggcttcctag | aaatagtagg | ttgtaatgaa | gtttctgagg | | 200 |
| coC3 | agagatatta | aaattttagt | gttcaccatg | tcagaatctc | aatatggaac | | 250 |
| coC3 | aagttatctc | tgctatttat | tcataaaaaa | gtctctaatt | tgatagaaaa | | 300 |
| coC3 | taatcttatt | aacttcatac | ttagttgaga | tccagtactt | gcagtcaaca | | 350 |
| coC3 | aaccagttgg | gatttggaca | gaggggcaaa | tgtccacag | ccagaaggcc | | 400 |
| coC3 | tcttgttgtc | cactaagtct | tcaactgtct | cctatcactt | aactagttgc | | 450 |
| coC3 | acaactctt | aagatttgag | gtgaaagatg | ccaaagagaa | aatgactta | | 500 |
| coC3 | atagcattat | taattttaga | aactccagaa | agagagatta | atttcagaga | | 550 |
| coC3 | tttttgactg | g..... | | | | | 561 |
| (ca. 900 bp) | | | | | | | |
| coC3 | |aaag | gcctcaaacc | agccaaaatc | agctgtgaag | | 34 |
| coC3 | tgcagtgag | gtaagggggg | gcagggggat | tgcgcacaca | tgcacagggg | | 84 |
| coC3 | gcgtgggggc | gcatgcattg | ccttctgggt | gctggcatgt | gagcgcatga | | 134 |
| coC3 | gtgccagtg | gcacacaatt | tttgtcacag | gagcccaaaa | aggttagcca | | 184 |
| coC3 | cgctggggct | agttgggtcaa | tcaaactttt | catcttaaaa | aggctgctaa | | 234 |
| coC3 | gaatttgtgc | agagtctgat | tttccccctt | tcttttgtaa | tttttga ACT | | 284 |
| coC3 | TAAAA CA CGC | AAAGGGAG GC | AAGGCAGCCC | TCAGGCTTAA | AATCTGCACT | | 334 |

Anhang 7.4: Sequenzen weiterer putativer Gene der CVF/C3-Familie

Homologe Sequenzen, die auf die Existenz mindestens eines neuen Gens der CVF/C3-Genfamilie hindeuten, wurden beginnend vom 5'-Ende mit CVF-Homolog 1 (CVF-H1), CVF-H2 usw. durchnummeriert.

Unterschiede zu den entsprechenden CVF1-Sequenzen (Exonbereiche) sind durch Markierungen hervorgehoben.

Intron 1

Exon 2 stromaufwärts

Die Sequenz des entsprechenden Amplifikats der PvuII-Bibliothek (ca. 1,5 kb) ist mit CVF-H1 bezeichnet. Die PvuII-Schnittstelle (cagctg) am 5'-Ende sowie eine EcoRI-Schnittstelle (gaattc) sind unterstrichen. Charakteristische Tandem-Wiederholungen der Dinukleotide ac und ta sind unterstrichen dargestellt. Die Sequenz von CVF-H1 weicht an zwei Positionen von Exon 2 der CVF1-Sequenz ab (markiert). Rückwärtsprimer: HB2as-CVF.

| | | | | | | | |
|--------|--------------------|-------------------|--------------------|-------------------|--------------------|-------------------|-----|
| CVF-H1 | <u>cagctg</u> gaat | taagtgctag | gcgtttaaca | gcctgcacca | ccaaggccct | 50 | |
| CVF-H1 | ttaataat | agtagtagag | ctatctgtct | gtctggagac | tgagatagat | 100 | |
| CVF-H1 | tcatatataa | agacatgaat | tagcaagatc | tgtctgcaa | agacgtaatt | 150 | |
| CVF-H1 | tattagccct | tggtttctga | ggttgatact | gcatctgttc | tgcaacagag | 200 | |
| CVF-H1 | caatactgtg | ttgaatgaat | <u>tctgg</u> tatgg | ggattagggg | gtgtgtgggc | 250 | |
| CVF-H1 | agcttctgaa | tggtgaggag | tctcaggcga | ctacgtgggc | tacagagatg | 300 | |
| CVF-H1 | ttactaagag | gagttgctca | acatttccac | ctcattttta | gcaaagggtg | 350 | |
| CVF-H1 | tgggttaaag | atgggaatcc | ctaactccta | aaaaacccaa | cttgaoctca | 400 | |
| CVF-H1 | ttttggcaaa | atttagccat | tttgccgata | agtttcagca | gcgtagtttc | 450 | |
| CVF-H1 | aggtgccagg | gtctgtagat | gggttactgg | aggactgtat | gcaggctagg | 500 | |
| CVF-H1 | aattgcctaa | ccttggatga | aggcac..... | | | 526 | |
| | | | | | | | |
| | | | (ca. 400 bp) | | | | |
| | | | | | | | |
| CVF-H1 | | | actgcgc | tgtaggctg | tatcttaaca | 26 | |
| CVF-H1 | gaacacacac | cctgaggggt | gtaggggtg | tcttaccccc | cacaatgttt | 76 | |
| CVF-H1 | gttccagag | agtaagccat | ttttgatga | agtttggtg | aaattgcttg | 126 | |
| CVF-H1 | aggctttaca | gagttatgct | ggaacacaca | cacacacaca | cacggagcca | 176 | |
| CVF-H1 | <u>ttatatata</u> | <u>tatagaaaac</u> | agcaccocaa | atccctttta | agtgatataa | 226 | |
| CVF-H1 | agataaagat | ttactaattt | tccaactagg | cttggagct | ttctaagtca | 276 | |
| CVF-H1 | caaaagagta | gttcttgtaa | ctatttattt | cttcactttt | ccagaagaag | 326 | |
| CVF-H1 | agattatgga | ggtgcctcat | gaaatgcctc | actttttcta | agcaaaactaa | 376 | |
| CVF-H1 | ctcaggagat | gttcttttct | tcttctgcag | CTACACCCTC | ATCACCCCTG | 426 | |
| CVF-H1 | CTGTTTTCG | AACAGACACA | GAAGAGCAAA | T | TTGGTGGG | GGCCCATGGA | 476 |
| CVF-H1 | GACA | TACTC | CAAAAACAGCT | TGACATCTTT | GTTTCATGATT | TTCCACGGAA | 526 |
| CVF-H1 | GCAGAAAACC | TTGTTCCAAA | CCAGAGTAGA | TATGAATCCA | GCAGGAGGCA | 576 | |

Intron 2 stromaufwärts

Die Sequenz des entsprechenden Amplifikats der DraI-Bibliothek (ca. 1120 bp) ist mit CVF-H2 bezeichnet. Die DraI-Schnittstelle am 5'-Ende ist unterstrichen. Häufige Wiederholungen der Dinukleotide ac und ta sind unterstrichen dargestellt. Die Sequenz von CVF-H2 weicht in Exon 2 in drei Positionen von der CVF1-Sequenz ab (markiert). Rückwärtsprimer: HB-Intr.2as-CVF1-II.

| | | | | | | |
|--------|--------------------|-------------------|-------------------|-------------------|--------------------|------|
| CVF-H2 | | | | ttttaaacc | gtagccgat | 20 |
| CVF-H2 | aaccagcgt | tgctgggta | tttctttatc | ctaactcctgt | attagacagg | 70 |
| CVF-H2 | aatggatg | tcaagtgca | acgtaagtgt | ggaatgggct | gtcctccctc | 120 |
| CVF-H2 | tttccattaa | catctccaca | attcctgaaa | agacgatggc | aagagcaact | 170 |
| CVF-H2 | tctccctaa | ccctgatttc | ggtgacaatt | agttaaaaat | gttttccctg | 220 |
| CVF-H2 | taccccaga | agcatcaaga | aatgtcatct | cgtatgtgat | ctggggaagt | 270 |
| CVF-H2 | acatcatgtc | gatctcatag | tggtgcaggc | aggccccctt | acaaacaccc | 320 |
| CVF-H2 | ccttgtggag | tattgtgaag | ccattaccat | ggcaactcca | ctgcgctgta | 370 |
| CVF-H2 | ggctgtatct | taacagaaca | cacaccctga | gggggtgttag | gggtgtctta | 420 |
| CVF-H2 | ccccccaaa | tgtttgtttc | cagagagtaa | gccatttttg | tatgaagttt | 470 |
| CVF-H2 | ggttgaatt | gcttgaggct | ttacagagtt | atgctggaac | <u>acacacacac</u> | 520 |
| CVF-H2 | <u>acacacacgg</u> | agccatttat | atataatat | atataatat | aaancagcac | 570 |
| CVF-H2 | cccaaatccc | ttttaagtga | tataaagata | aagatttact | aattttccaa | 620 |
| CVF-H2 | ctgggttttg | aagctttcta | agtcacaaaa | gagtagctct | tgtaactatt | 670 |
| CVF-H2 | tatttcttca | cttttctaaa | agaagagatt | atggagatgc | ctcatgaaat | 720 |
| CVF-H2 | gcctcacttt | tcctaagcaa | actaactcag | gagatgttct | tttcttcttc | 770 |
| CVF-H2 | tgcag CTACA | CCCTCATCAC | CCCTGCTGTT | TTGCGAACAG | ACACAGAAGA | 820 |
| CVF-H2 | GCAAAATTTG | GTGGAGGCC | ATGGAGACAG | TACTCCAAA | CAGCTTGACA | 870 |
| CVF-H2 | TCTTTGTTG | TGATTTTCCA | CGGAAGCAGA | AAACCTTGTT | CCAAACCAGA | 920 |
| CVF-H2 | GTAGATATGA | ATCCAGCAGG | GGGCATACTT | GTCACTCCAA | CTATAGAG gt | 970 |
| CVF-H2 | gggtgtatgg | tgatggttgc | aatagagtct | atggtatccc | tgcaaaaaca | 1020 |
| CVF-H2 | aaaaccaagc | taaaatcctg | tccagtcagg | cttgagggga | aaccctatgc | 1070 |
| CVF-H2 | actttagtca | aacatgacta | aattgtcaga | tattgagaga | aactgtgcat | 1120 |

Intron 12

Vorwärtsprimer: YL12s, Rückwärtsprimer: YL13as. Unterschiede zur CVF1-Sequenz sind markiert.

| | | | | | | |
|--------|-------------------|-------------------|-------------------|-------------------|---------------------|-----|
| CVF-H3 | <u>GCAACAAAGT</u> | <u>CCATGACAGC</u> | <u>CATAGCCTAC</u> | CAAACCCAGG | GAGGATCTGG | 50 |
| CVF-H3 | AAACTATCTT | CATGTAGCCA | TTACATCTAC | AGAGATTAAG | CCCGGAGATA | 100 |
| CVF-H3 | ACTTACCTGT | CAATTTCAAT | GTGAGGGGCA | ATGCAAATTC | ACTGAA C CAG | 150 |
| CVF-H3 | ATCAAATATT | TCACATACCT | Cgtaagtatt | atagatctgg | aggaaatcca | 200 |
| CVF-H3 | ggggctgcat | ttcaattcat | atgttagcct | tgaagccact | ggcttaatct | 250 |
| CVF-H3 | atcatttcta | accttcccat | taatcagtat | ttgttcaccc | tcag ATATTG | 300 |
| CVF-H3 | AATAAGGGA | AGATTTTCAA | GGTTGGCAGG | CAACCCAGGA | GAGATGGGCA | 350 |
| CVF-H3 | GAATCTGGTG | ACCATGAATC | TGCATATCAC | TCCAGATCTC | ATCCCTTCCT | 400 |
| CVF-H3 | TCCGGTTTGT | GGCTTACTAC | CAAGTGGGAA | ACAACGAAAT | TGTGGCTGAT | 450 |
| CVF-H3 | <u>TCTGTC</u> | | | | | 456 |

Intron 27

Eine EcoRI-Schnittstelle in Exon 27 (Primersequenz), eine StuI-Schnittstelle in Exon 28 sowie zwei DraI-Schnittstellen (tttaa) und eine StuI-Schnittstelle (aggcct) im Intron sind unterstrichen. Vorwärtsprimer: HB27s-CVF, Rückwärtsprimer: HB28as-CVF. Unterschiede zur CVF1-cDNA-Sequenz sind markiert.

| | | | | | | |
|--------|--------------------------|-------------------|-------------------|-------------------|--------------------|-----|
| CVF-H4 | <u>GGAGGAATTC</u> | <u>AAGGTGCTGA</u> | <u>AGAAGAAGTA</u> | TCTTTAACAG | CTTTCATTCT | 50 |
| CVF-H4 | GGTTGCGTTG | TTGGAATCCA | GATCAATCTG | CAATGCATAT | ATCAAT gtat | 100 |
| CVF-H4 | gttgatggtg | tgatgggaag | tggtacgaga | gggccaaaat | ataatctcat | 150 |
| CVF-H4 | ccattggggg | tttttcacag | tatttaattt | tcacaacttc | ttggccccag | 200 |
| CVF-H4 | tattaaaagc | cttcaggctc | tccaaatagc | ctaatacagc | tggtataaat | 250 |
| CVF-H4 | ttatacttta | ttaatttcaa | atgaaactaa | gataaataat | caattcaatt | 300 |
| CVF-H4 | aggatactat | caggaaatta | cacatatact | ttcaagccag | caaggatatt | 350 |
| CVF-H4 | agatcagtta | ataccgtggt | tccccgaaaa | taagacactg | tcttatattt | 400 |
| CVF-H4 | ttttttggct | ccccaaaaca | cactaggtct | tattttcggg | gtaggctcta | 450 |
| CVF-H4 | tatatattatg | gtgccaaaata | gtgggtgtga | gtgtgta..... | | 487 |
| | (ca. 300 bp) | | | | | |
| CVF-H4 |cct | atttaagaaa | ttcctaataga | atztatgttc | | 33 |
| CVF-H4 | aacacctctt | cctgtataaa | aactttaaaa | taagacatat | gatggctaaa | 83 |
| CVF-H4 | tttgaattat | tagcatgtaa | aatatacaag | caggaaatgg | cagcatattt | 133 |
| CVF-H4 | ttatttattt | attatttatt | gaacaaattt | atatggctgt | ccaactcaca | 183 |
| CVF-H4 | agtagttgac | cctggccagt | ttacaagatt | taaaaatcta | ctacataaaa | 233 |
| CVF-H4 | aaatccataa | cttatctgct | cccgacccat | ggcagcagga | gcacattcaa | 283 |
| CVF-H4 | atcaaccatt | tgtggttctc | agggccaatg | acaaaacat | gtttttaagg | 333 |
| CVF-H4 | <u>ccttttgaaa</u> | aagagtgtgg | accaatcata | cctcagggaa | atgctgttcc | 383 |
| CVF-H4 | atagagatgc | ctaaaacatt | cctaattggat | ctctctctct | tcctctgcct | 433 |
| CVF-H4 | ttacttctag | ATTCTAGACA | GCAGCATCAG | TAAGGCCACA | GATTATTTAC | 483 |
| CVF-H4 | TCAAAAAGTA | TGAGAACTG | CAAAGGCCTT | ACACTACAGC | CCTCACAGCC | 533 |
| CVF-H4 | TATGCTTTGG | CTGCTGCAGA | CAACTCAAT | GATGACAGGG | TACTCATGGC | 583 |

Intron 36

Konsensus-Sequenz aus genomischen PCR-Produkt (2000 bp) und Genome Walking-Amplifikat der DraI-Bibliothek (900 bp). Charakteristische Unterschiede zur CVF1-Sequenz sind im 3'-Ende von Intron 35 und in Exon 37 markiert.

Beim EcoRI-Verdau entstehen Fragmente von ca. 1.5 kb, 400 bp und 220 bp. EcoRI-Schnittstellen (gaattc) und DraI-Schnittstellen (tttaaa) in Intron 36 sowie eine StuI-Schnittstelle (AGGCCT) in Exon 37 sind unterstrichen. Die Sequenz des antisense-Primers (CVF1-Sequenz) weicht in zwei Nukleotiden von der CVF2-cDNA ab (umrahmt). Vorwärtsprimer: HB-CVF1-Intr.35s, HB36s-CVF; Rückwärtsprimer: HB37as-CVF.

```

CVF-H5  ccttctgttc ctctcacctt ttccattgct tccctttatt ctgggaaacc 50
CVF-H5  tggacagatt tgcagaatth agtccttcaa ccactgggtc aaacctgagg 100
CVF-H5  aacttggaat aagcagtttc cttgaggtta ttcggatgga gtttcgggtg 150
CVF-H5  ggggggaatt atctccatgg gtaaactggc acccacaat gatcctctct 200
CVF-H5  ctccatccca gTCTCCAC TCTGAAGATG AATGCCTGCA CTTTAAGATT 250
CVF-H5  CTCAAGCATT TTGAAGTTGG CTTTCATTCAG CCAGGATCAG TCAAGGTGTA 300
CVF-H5  CAGCTACTAC AATCTAGta atggcagacg tactgaatat acctccctgc 350
CVF-H5  tcctaatttc cccacaacaa ccactctgtg aagtgaatca agattgtgag 400
CVF-H5  cgaatggcct ggccaaggct atccagttgg ctttcatggc taaggcggga 450
CVF-H5  ctttattcat ggtctcccc tttctagtct ggtgctttaa ccattaggcc 500
CVF-H5  atactggctc tttccaagga actcagtggt actggagctt gagaaaatta 550
CVF-H5  gccaaaaagc ccaacaacaa tgcattggtg ataaaacatt cttgattgat 600
CVF-H5  tttccaaaaa actcctttata acaattttat aaaagctgaa gtaaagaatg 650
CVF-H5  aatcatttgg gaatattgaa ggtatattaa aggtagtac acagagctct 700
CVF-H5  acccaatgct cgggcagcct tctggtctga gtttcctttg cttgcctttt 750
CVF-H5  gaggggcaga gaacaaacac caagttcctg ttttctctg ctctggtggg 800
CVF-H5  cagcatgatg ctcatctttg gggggaaaag tggatgaga ttttggaaag 850
CVF-H5  tgagctgtgc atggctgagg aagagttttg ccattctaataa taatattttg 900
CVF-H5  aaagttcatt atcattttta tttctgtata gcattttcag tgcagctctt 950
CVF-H5  aaaccaggaa atgtgcaaaa tcatgaaaat acttcagaga aaattacctg 1000
CVF-H5  taaaacctgt atatagtagg cacaattgca aaaaaaggc tttcctttga 1050
CVF-H5  aaacatactc tcaaatgcat acacatttct gaaatgtaa aactatacaa 1100
CVF-H5  aataattggt taaa..... 1114
.....
..... (ca. 500 bp) .....
.....
CVF-H5  .....cag caatgattta tgaccggaat 23
CVF-H5  tcagggtcctca attgtggtta taagttgaga actacctata gtctatttta 73
CVF-H5  tgctatggaa aagtggcaca tacaaggaag catcaaagg gctaaaatga 123
CVF-H5  gaaatgtaaa cattttattg tcagaccaga agccccacac cttttgtatg 173
CVF-H5  ttaattgtag catcgcatgc atgccaaaaa cttgcaaggc aataccattt 223
CVF-H5  aataaaagga gcagaattct acagatgtcc atgttccaac ttcctagctc 273
CVF-H5  cactaaataa tgggatatac tctcttacct ttcttattgc attgtagggc 323
CVF-H5  tctgtgcttc tcagttaaat gtttaaaaaa agtcaatcag ccaatcaaca 373
CVF-H5  ctcatcataa gtttccattg cactgtcagt tcacacgtgt cccaaattta 423
CVF-H5  aaatctcctt gagaaaaaag ggttcttctt tgtctttcta agtaatacaa 473
CVF-H5  cctataatca taataaattt atcgatgtga cctatggaca taacttttaa 523
CVF-H5  ttctttcatt agATGAAAAA TGTACCAAGA TCTACCATCC AGATGAAGCA 573
CVF-H5  ACAGGCCTTC TCAATAAGAT ATGTAATTGGT AACGTTTGCC GATGTGCAGG 623

```

Anhang 7.5: Primersequenzen

Primer für CVF1-Genamplifikate

| Name | Sequenz 5'-3' | Bindungsstelle CVF1-cDNA | Bindungsstelle im CVF1-Gen |
|---------------------|---|-----------------------------|-------------------------------|
| HB-CVF1-Prom-as | TGT CTT GCT AAT CCC ATC ACT CTG | --- | Promotor |
| HB1as | AGA GCC CCA TGA GAA GAC CCT GGA A | 50-74 | Exon 1 |
| HB1as-II | CAT GAG AAG ACC CTG GAA AAC CAA TC | 42-67 | Exon 1 |
| HB1s-II | ATG GAG AGG ATG GCT CTC TAT CTG G | 4-28 | Exon 1 |
| HB1s-CVF | GGT GGC TGC TCT ATT GAT TGG TTT TCC AGG | 27-56 | Exon 1 |
| CVF-Intr.1as-I | ccc tgc aaa gga cca cag cct tgg atg tg | --- | Intron 1 |
| CVF-Intr.1as-II | gga cca cag cct tgg atg tga tag ttc cg | --- | Intron 1 |
| HB2as-CVF | TGC CTC CTG CTG GAT TCA TAT CTA CTC TGG | 215-244 | Exon 2 |
| HB2as-II-CVF | GTT GGA GTG ACA AGC ATG CCT CCT GCT G | 233-260 | Exon 2 |
| HB2s-CVF | CCC ATG GAG ACA GTA CTC CAA AAC A | 137-161 | Exon 2 |
| HB2s-II | CAG TAC TCC AAA ACA GCT TGA CAT C | 147-171 | Exon 2 |
| HB-Intr.2as-CVF1-I | ccc tag aat tat atg cac agt ttc tct c | --- | Intron 2 |
| HB-Intr.2as-CVF1-II | atg cac agt ttc tct caa tat ctg aca | --- | Intron 2 |
| HB3as-II-CVF | GAG GAC CAG TTA CTT GCA CAA CCA CAT AT | 309-337 | Exon 3 |
| HB3as-CVF | ATC TCA CTT GAG GAC CAG TTA CTT G | 322-346 | Exon 3 |
| HB3s-CVF | CAA GTA ACT GGT CCT CAA GTG AG | 322-344 | Exon 3 |
| HB4as-CVF | CAA TCA CAG TTT TGT TCA TCT TGC T | 466-490 | Exon 4 |
| HB4s-CVF | GTT TTT TCT ATG GAT CAC AAC ACA AG | 442-467 | Exon 4 |
| HB5as-CVF | GGT AAA TTG TAA GGC CAG AAG AAG | 546-569 | Exon 5 |
| HB5s-CVF | TCT TGT CAG TTC TAA TTC AGT TGA C | 516-540 | Exon 5 |
| HB6as-CVF | CAA AAT ATG CAG TAT AAT TCT CTG GG | 621-646 | Exon 6 |
| HB6s | GAC TTG GAG GAT TGT GGC CAA ATA TG | 588-613 | Exon 6 |
| HB7as | CTT CTC TGA TGG TTG CAG ACG GAC | 679-702 | Exon 7 |
| HB7s-CVF | TCC GTC TGC AAC CAT CAG AG | 680-699 | Exon 7 |
| HB-CVF1-Intr.7s | acc ttg agt gtt gga atg gac aga aag gtg | --- | Intron 7 |
| HB8as | TGA GTG AGT CTG GAA TAC TC | 825-844 | Exon 8 |
| HB8s-CVF-Intr. | gtg tea tta cag GTA CTT GTA TGG AGA GG | 753-769 | Intron 7/Exon 8 |
| HB8s-CVF | GAA GTG GAA GGT GTG GCC TTT GTC C | 769-793 | Exon 8 |
| HB8s-II-CVF | AGG AAG TGG AAG GTG TGG CCT TTG TCC | 767-793 | Exon 8 |
| HB9as-CVF | CAG ATG CAT ACA GAG TAT GC | 936-955 | Exon 9 |
| HB9as-CVF | CTG TTA CAG ATG CAT ACA GAG TAT GC | 936-961 | Exon 9 |
| HB9as-II-CVF | GAT GCA TAC AGA GTA TGC CCA ACA AGC TC | 925-953 | Exon 9 |
| HB9s | CAA ATC TCA ATG AGC TTG TTG GGC | 914-937 | Exon 9 |
| HB-CVF1-Intr.9as | caa ggt tcc ttg gca ttt tgg tca cat acc | --- | Intron 9 |
| HB10as | GGG GGT TTT TGT GAA GTG GAT CTG | 1033-1056 | Exon 10 |
| HB10s | GCA TCT CCC TAT CAG ATC CAC TTC | 1021-1044 | Exon 10 |
| HB11as | GTC CCA TCA CTC AAA GTG GTT CC | 1165-1187 | Exon 11 |
| HB11s | CCA TGT GCC AGT GGT ATC AGA GG | 1128-1150 | Exon 11 |
| HB12as | TTC CAG ATC CTC CCT GGG TTT GG | 1308-1330 | Exon 12 |
| YL12s | GCA ACA AAG TCC ATG ACA GCC AT | 1279-1301 | Exon 12 |
| YL13as | GAC AGA ATC AGC CAC AAT TTC GTT G | 1587-1611 | Exon 13 |
| HB13s | CAC TCC AGA TCT CAT CCC TTC C | 1533-1554 | Exon 13 |
| HB14as | CTT TGT CCA CAG CCA CAA GAC C | 1729-1750 | Exon 14 |
| AJS01 | GGA TCC AGG TGC TCG GGT TGG | 1710-1730 | Exon 14 |
| OPA19 | CTG TTT GGT GTT GAG ATT AGT G | 1896-1917 | Exon 15 |
| HB15s | TGG CTG TAC AGC TGG CAG TGG CCA G | 1824-1848 | Exon 15 |
| HB-1R67-5'as | tgg tct ggg ctt cta tga ttg tac | --- | Intron 15 |
| Con2eco | tca gtg gca aac gtt tgg tgt gg | --- | Intron 15 |
| HB-Con2eco-as | gat tcc agt tgt tgg agc gag g | --- | Intron 15 |
| OPA72 | CTG CGA CGC CTC CGA TTT GCA GGC | 1941-1964 | Exon 16 |
| OPS72 | CTG CAA ATC GGA GGC GTC GCA GT | 1943-1965 | Exon 16 |
| OPA24 | GCA TGA CAT CTT CAC AGC ATT | 2030-2050 | Exon 17 |
| OPS24 | CGT AAA TGC TGT GAA GAT GTC ATG C | 2026-2050 | Exon 17 |
| OPA22 | CTC CTC GGT CAA GTC CTT TG | 2279-2298 | Exon 18 |
| OPS22 | AAG GAC TTG ACC GAG GAG | 2281-2298 | Exon 18 |
| HB-Intr.18as | tag tcc aag tgg caa gag ctg | --- | Intron 18 |
| OPA21 | CAG CAC CAC CCA GGT TG | 2357-2373 | Exon 19 |

Fortsetzung

| | | | |
|---------------------|---|-----------|------------------|
| OPS21 | CAA CCT GGG TGG TGC TG | 2357-2373 | Exon 19 |
| OPA16 | ACG TAG TTG TGC AGA ATA GC | 2503-2522 | Exon 20 |
| OPS16 | GCT ATT CTG CAC AAC TAC GTT | 2503-2523 | Exon 20 |
| OPA17 | GGA AGC ACT GCA GAA GGC TGG G | 2562-2583 | Exon 21 |
| OPS17 | CCC AGC CTT CTG CAG TGC TTC C | 2562-2583 | Exon 21 |
| OPA25 | ACA ATG GAT TTC TGT ACC CCT TC | 2755-2777 | Exon 22 |
| OPS25 | GAA GGG GTA CAG AAA TCC ATT GT | 2755-2777 | Exon 22 |
| OPA14 | TTC TGT GTC AGG CAC TCT GTC | 2857-2877 | Exon 23 |
| HB23s | GAT GAC AGA GTG CCT GAC ACA G | 2854-2875 | Exon 23 |
| HB24as | TTC AAT AAT CTG AGC CAC AGG G | 2907-2928 | Exon 24 |
| HB24s | CAG GAC TGA AGC TGT CAA TCA G | 3075-3096 | Exon 24 |
| SD24s | GTG ACC CTG TGG CTC AGA TTA TTG | 2903-2926 | Exon 24 |
| HB25as | TCT TGT ACA CCA TCT GCT GGG C | 3112-3133 | Exon 25 |
| OPS15 | ATG CCC AGC AGA TGG TGT ACA A | 3110-3131 | Exon 25 |
| OPA26 | CTG CTA CCA TTT TGG CAG CCA TGG CA | 3210-3235 | Exon 26 |
| OPS20 | ATG CCC CTG TAC TTT CT | 3317-3333 | Exon 26 |
| YLIntr.26s | ctc tcc tca tgt aat gtt ctt agg | --- | Intron 26 |
| YL27as | cgt acA TTG ACA TAG TCA TTG CAG | 3423-3441 | Intr. 27/Exon 27 |
| HB27as | GCT GTT AAA TAT ACT TCT TCT TCA GCA CC | 3358-3386 | Exon 27 |
| HB27s-CVF | GGA GGA ATT CAA GGT GCT GAA GAA GAA GT | 3346-3374 | Exon 27 |
| HB-1R67-F5-s | GGA GGA ATT CAA GGT GCT GAA G | 3346-3367 | Exon 27 |
| HB Intr.27s | ttt tca tct gcc tca aag gca gat | --- | Intron 27 |
| CVF-Intr.27as | caa gta ccc tgg acc caa gc | --- | Intron 27 |
| HB-1R67-F5-as | tac tga acc tgg ggt caa gaa g | --- | Intron 27 |
| HB28as-CVF | GCC ATG AGT ACC CTG TCA TCA TTG AGT T | 3554-3581 | Exon 28 |
| HB28as | AGG CCT TTG CAG TTT CTC ATA C | 3489-3510 | Exon 28 |
| HB28s | CAG ACC AAC TCA ATG ATG ACA GGG | 3548-3571 | Exon 28 |
| HB29as | GTT GTG GGT GTG AGC ATT GTA TTC | 3610-3633 | Exon 29 |
| HB29s-CVF | gGA AGG GAT CAT TGG GAA GAA TAC | 3593-3615 | Intr. 28/Exon 29 |
| HB30as-CVF | TGT CCG AGC CAG GAG AGC ATT TTC | 3886-3909 | Exon 30 |
| HB30s-CVF | GAT GCC TAC CCA TAA GGA CTT AAA C | 3801-3825 | Exon 30 |
| HB31as-CVF | GCT TTT CCA TCA CCT GAT GCT GT | 3943-3965 | Exon 31 |
| HB31s-CVFgenom | ggt gct ttt cag ACC AAA CTC AAC C | 3916-3928 | Intr. 30/Exon 31 |
| HB32as-CVF | ATT GCA AAC ATT TGC CTT CTC CTG | 4003-4026 | Exon 32 |
| HB-CVF-32as-II | GAA ATT TAT TGC AAA CAT TTG CCT TCT CCT | 4004-4033 | Exon 32 |
| HB32s-CVF | TAT AAC GCA CAG TTG CAG GAG AAG | 3988-4011 | Exon 32 |
| HB33as-CVF | TGC AGA TCT TGA GCA TGA GGG CTC | 4085-4108 | Exon 33 |
| HB33s-CVF | GGG AGC CAA GGG AGC CCT CAT | 4074-4094 | Exon 33 |
| HB34as-CVF | TTG TAA GGT CTT CAG CAT CAG GGA G | 4174-4198 | Exon 34 |
| HB34s-CVF | TTG ATA TTT CTA TGC TGA CTG GTT TTC | 4148-4174 | Exon 34 |
| HB35as-CVF | TAC TTT CTG AGC CAT ATT ATT GTC AAC | 4240-4266 | Exon 35 |
| HB35s-CVF | TGG ACA GAT ACA TCT CCA GAT ATG AAG | 4214-4240 | Exon 35 |
| HB-CVF1-Intr.35s | cct tet gtt cct ctc acc ttt tcc att gc | --- | Intron 35 |
| HB-CVF1-Intr.35s-II | tag agc cag ggc agg cag gca g | --- | Intron 35 |
| HB36as-CVF | GAA TGA AGC CAA CTT CAA AAT GCT TGA | 4331-4357 | Exon 36 |
| HB36s-CVF | AGA TGA ATG CCT GCA CTT TAA GAT TCT | 4305-4331 | Exon 36 |
| HB37as-CVF | CCT GCA CAT CGG CAA ACG TTA CCA AT | 4459-4484 | Exon 37 |
| HB37s-CVF | TAC CAA GTT CTA CCA TCC AGA TAA AGG | 4407-4433 | Exon 37 |
| HB38as-CVF | GTC TCG CAG GCT TTT TCA ATT TGT AAT G | 4529-4556 | Exon 38 |
| HB38s-CVF | CTC AAC CAT CAG GAA AGG ATT GAT G | 4501-4525 | Exon 38 |
| HB39as-CVF | CCA TGA CAT AGA TAT CAT TAC CAT CTT G | 4603-4630 | Exon 39 |
| HB39s-CVF | ACC AAG CTG CTT CGA ATA GAA GAA C | 4579-4603 | Exon 39 |
| HB40as-CVF | CTT TGA CTT ATG TAC TGG TGG GTC | 4680-4703 | Exon 40 |
| HB40s-CVF | AGG CTC TGA ATC TGA AGG TGA ATG | 4715-4738 | Exon 40 |
| HB41as-CVF | TGG AAT TCT TCT TCC TGA CAT TCG | 4845-4868 | Exon 41 |
| HB-CVF-41s-I | AGA ACA CAT GGA TTG AGA GAT GGC CAC | 4811-4837 | Exon 41 |
| HB-CVF-41s-II | GAG AGA TGG CCA CAT GAA GAC GAA TGT | 4825-4851 | Exon 41 |

Primer für CVF2-Genamplifikate

| Name | Sequenz 5'-3' | Bindungsstelle CVF2-cDNA* | Bindungsstelle CVF2-Gen |
|--------------------|--|---------------------------|-------------------------|
| HB-coC3-II-Prom-s | AGT CTT TCC AAT TCG GGC CAT TCC | --- | Promotor |
| HB-coC3-II-Prom-as | TCA CTC TGG TTT CAA TGC AAT GG | --- | Promotor |
| CVF2-Ex18s | gGT GAT TTT GAA GAT GAT TTA TTT GGA GAA GG | 1249-1279 | Intr.17/Exon 18 |
| CVF2-Ex19as | CTT TGG TGG GTG AAA GGC CCA CAG CCA G | 1419-1446 | Exon 19 |
| HB-31s-CVF2/coC3 | gAC CAA ACT CAA CGA AGA CTT CAC TGT GT | 2964-2991 | Intr. 30/Exon 31 |
| HB33as-CVF2 | AGT GCA GAT TTT GAG CTT GAG GGC TC | 3142-3167 | Exon 33 |
| HB-CVF2-41s | GTG TGA TGA CTT TGC TCT GTT TAG CTA CG | 3932-3960 | Exon 41 |

*bezogen auf die vorläufige CVF2-cDNA-Nummerierung (Fritzinger *et al.* 1994a)

Primer für Cobra C3-Genamplifikate

| Name | Sequenz 5'-3' | Bindungsstelle in coC3-cDNA | Bindungsstelle Cobra C3-Gen |
|------------------|---|-----------------------------|-----------------------------|
| HB-coC3-Prom-s | AAG GGA AGT ACT CGG GGA AAA GA | --- | Promotor |
| CoC3-Intr.1as-I | aga cac ctg gaa gga aat tca tta ccc tgc | --- | Intron 1 |
| CoC3-Intr.1as-II | tgg aag gaa att cat tac cct gca aag gac | --- | Intron 1 |
| YL3s-C3 | TGA AAG TAA CTG GTC CTC AAG TGG C | 325-349 | Exon 3 |
| YL4as-C3 | GAC AAT CAC AGT TTT GTC CAT CCT GTG | 471-497 | Exon 4 |
| YL29s-C3 | agG AAG GAA TCG TTG GGA AGA ATA T | 3616-3638 | Intr. 28/Exon 29 |
| YL30as-C3 | CTG TCC GGG CCT GGA CAG CAT TTC | 3910-3933 | Exon 30 |
| HB-31s-CVF2/coC3 | gAC CAA ACT CAA CGA AGA CTT CAC TGT GT | 3939-3966 | Intr. 30/Exon 31 |
| HB33as-coC3 | AGT GCA GAT TTT AAG CCT GAG GGC TG | 4117-4142 | Exon 33 |
| HB-coC3-41s | GAA TTT GTG TGA TGA CTT TGC TCA GTT GTC | 4901-4930 | Exon 41 |

Modifizierte Primer für CVF2-Peptidexpression

| Name | Sequenz 5'-3' |
|------------------|---|
| CVF2-Ex18as-Ü | GTA CTA TCT TGC TTG AAA TCC <i>CTT TAC TGT TAG GAT GTT C</i> |
| CVF2-Ex19s-Ü | <i>GAA CAT CCT AAC AGT AAA GGG</i> ATT TCA AGC AAG ATA GTA CC |
| CVF2-Ex18s-NcoI | ACT <u>GCC ATG GCA</u> AGT GAT TTT GAA GAT GAT TTA TTT G |
| CVF2-Ex19as-XhoI | ACT <u>GCT CGA GTT</u> TGG TGG GTG AAA GGC CCA CA |

Bei den modifizierten Primern mit Überhängen (Ü) sind die Sequenzen der jeweils benachbarten Exons kursiv dargestellt.

Bei den modifizierten Primern mit Restriktionsschnittstellen (NcoI bzw. XhoI) sind diese unterstrichen.

Sonstige Primer

| Name | Sequenz 5'-3' | Bindungsstelle |
|---------------|---------------------------------|-----------------------------|
| sCos1-T7as | CGC ATA ATA CGA CTC ACT ATA GG | Vektor sCos1 |
| sCos1-T3s | CGC AAT TAA CCC TCA CTA AAG G | Vektor sCos1 |
| sCos1-5's | CAG GGT TAT TGT CTC ATG AGC GGA | Vektor sCos1 |
| sCos1-3'as | GAA ACG ACA GGT GCT GAA AGC GAG | Vektor sCos1 |
| M13 universal | GTT GTA AAA CGA CGG CCA GTG | Vektoren pUC18, pCR2.1-TOPO |
| M13 reverse | CAG GAA ACA GCT ATG AC | Vektoren pUC18, pCR2.1-TOPO |
| AP1 | GTA ATA CGA CTC ACT ATA GGG C | Genome Walking-Adaptor |
| AP2 | ACT ATA GGG CAC GCG TGG T | Genome Walking-Adaptor |

Anhang 7.6: PCR-Programme

CVF1-Amplifikate

Intron 1:

Primärdenaturierung: 2 min bei 92°C,
Denaturierung 10 s bei 92°C, Annealing 30 s bei 61°C, Elongation 10 min bei 68°C (10 Zyklen),
Denaturierung 10 s bei 92°C, Annealing 30 s bei 60°C, Elongation 12 min bei 68°C (20 Zyklen),
finale Elongation 7 min bei 68°C.
Besonderheiten: Verwendung von dünnwandigen Tubes
Polymerase: *Expand 20 kbPLUS* (Roche)

Intron 3 und Intron 4:

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 10 s bei 94°C, Annealing: 20 s bei 57.5°C, Elongation: 4 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid-AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 5:

Primärdenaturierung: 1 min 15 s bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 56°C, Elongation: 6 min bei 68°C (28 Zyklen),
finale Elongation: 20 min bei 68°C.
Polymerase: *Proof Sprinter Taq/Pwo* (Hybaid-AGS)

Intron 6:

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 10 s bei 94°C, Annealing: 20 s bei 60°C, Elongation: 4 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 7:

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 10 s bei 94°C, Annealing: 20 s bei 55°C, Elongation: 4 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid-AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 9:

Primärdenaturierung: 1 min 15 s bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 56°C, Elongation: 6 min bei 68°C (28 Zyklen),
finale Elongation: 20 min bei 68°C.
Polymerase: *Proof Sprinter Taq/Pwo* (Hybaid-AGS)

Intron 10 und Intron 11:

Primärdenaturierung: 30 s bei 94°C,
Denaturierung: 30 s bei 94°C, Annealing: 40 s bei 57°C, Elongation: 2 min bei 72°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 12:

Primärdenaturierung: 2 min bei 94°C,
Denaturierung: 30 s bei 94°C, Annealing: 40 s bei 57°C, Elongation: 1 min bei 72°C (28 Zyklen),
finale Elongation: 5 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 13:

Primärdenaturierung: 2 min bei 94°C,
Denaturierung: 40 s bei 94°C, Annealing: 40 s bei 57°C, Elongation: 6 min bei 72°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 14:

Primerpaar: AJS01/OPA19
Primärdenaturierung: 2 min bei 94°C,
Denaturierung: 40 s bei 94°C, Annealing: 40 s bei 56°C, Elongation: 2 min bei 72°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 14 und Intron 15 (partiell):

Primerpaar: AJS01/HB-1R67-5'as
Primärdenaturierung: 20 s bei 94°C,
Denaturierung: 10 s bei 94°C, Annealing: 30 s bei 60°C, Elongation: 3 min bei 72°C (5 Zyklen),
Denaturierung: 10 s bei 94°C, Annealing: 30 s bei 58°C, Elongation: 3 min bei 72°C (25 Zyklen),
finale Elongation: 5 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 11 bis Intron 15 (long range-PCR):

Primerpaar: HB11s/OPA19
Primärdenaturierung: 40 s bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 40 s bei 60°C, Elongation: 3 min bei 72°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Introns 16-26:

Allgemeines PCR-Programm für Amplifikationen mit Cosmid-1R67-DNA als Template.

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 53-58°C, Elongation: 1-10 min bei 72°C (25-30
Zyklen), finale Elongation: 10 min bei 72°C.
alternativ: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 28:

Primärdenaturierung: 30 s bei 94°C,
Denaturierung: 10 s bei 94°C, Annealing: 30 s bei 60°C, Elongation: 4 min bei 68°C (5 Zyklen),
Denaturierung: 10 s bei 94°C, Annealing: 30 s bei 58°C, Elongation: 4 min bei 68°C (25 Zyklen),
finale Elongation: 10 min bei 72°C.

Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 29 und Intron 30:

Primärdenaturierung: 15 s bei 94°C,
Denaturierung: 10 s bei 94°C, Annealing: 20 s bei 59°C, Elongation: 4 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von Enhancer (Hybaid AGS)
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 32:

Primerpaar HB32s-CVF/HB33as-CVF
Primärdenaturierung: 15 s bei 94°C,
Denaturierung: 15 s bei 94°C, Annealing: 30 s bei 56°C, Elongation: 4 min bei 68°C (28 Zyklen),
finale Elongation: 10 min bei 68°C.
Polymerase: *Proof Sprinter Taq/Pwo* (Hybaid-AGS)

Primerpaar HB31s-CVFgenom./HB33as-CVF
Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 63.5°C, Elongation: 3 min bei 68°C (5 Zyklen),
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 61°C, Elongation: 3 min 30 s bei 68°C (30 Zyklen), finale Elongation: 7 min bei 68°C.
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Intron 33 und Intron 34:

Primärdenaturierung: 4 min bei 95°C,
Denaturierung: 20s bei 95°C, Annealing: 30 s bei 57°C, Elongation: 3 min bei 68°C (30 Zyklen),
finale Elongation: 3 min bei 68°C.
Polymerase: *Fast Start Taq-DNA-Polymerase* (Roche)

Intron 35:

Primärdenaturierung: 2 min bei 94°C,
Denaturierung: 15 s bei 94°C, Annealing: 30 s bei 57°C, Elongation: 3 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 68°C.
Polymerase: *Expand Taq-DNA-Polymerase* (Roche)

Intron 37:

Primärdenaturierung: 2 min bei 94°C,
Denaturierung: 15 s bei 94°C, Annealing: 30 s bei 58°C, Elongation: 5 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 68°C.
Polymerase: *Expand Taq-DNA-Polymerase* (Roche)

Intron 38:

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 30 s bei 94°C, Annealing: 30 s bei 59°C, Elongation: 7 min bei 68°C (28 Zyklen),
finale Elongation: 7 min bei 68°C.
Polymerase: *Sawady Mid Range PCR System Taq/Pwo* (Peqlab)

Intron 39:

Primärdenaturierung: 1 min 15 s bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 56°C, Elongation: 6 min bei 68°C (28 Zyklen),
finale Elongation: 20 min bei 68°C.
Polymerase: *Proof Sprinter Taq/Pwo* (Hybaid-AGS)

Intron 40:

Primärdenaturierung: 4 min bei 95°C,
Denaturierung: 25 s bei 95°C, Annealing: 30 s bei 56.5°C, Elongation: 6 min bei 72°C (32 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von *GC-rich Solution* (Roche)
Polymerase: *Fast Start Taq-DNA-Polymerase* (Roche)

CVF2-Amplifikate

Intron 2:

Primärdenaturierung: 2 min bei 92°C,
Denaturierung 10 s bei 92°C, Annealing 30 s bei 59°C, Elongation 10 min bei 68°C (10 Zyklen),
Denaturierung 10 s bei 92°C, Annealing 30 s bei 58.5°C, Elongation 12 min bei 68°C (10 Zyklen),
Denaturierung 10 s bei 92°C, Annealing 30 s bei 58°C, Elongation 14 min bei 68°C (10 Zyklen),
finale Elongation: 10 min bei 68°C.
Besonderheiten: Verwendung von dünnwandigen Reaktionsgefäßen
Polymerase: *Expand 20 kbPLUS* (Roche)

Intron 18:

Primärdenaturierung: 4 min bei 95°C,
Denaturierung: 20 s bei 95°C, Annealing: 30 s bei 60°C, Elongation: 3:30 s bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 72°C.
Polymerase: *Fast Start Taq-DNA-Polymerase* (Roche)

Intron 32:

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 63.5°C, Elongation: 3 min bei 68°C (5 Zyklen),
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 61°C, Elongation: 3 min 30 s bei 68°C (30 Zyklen), finale Elongation: 7 min bei 68°C.
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

coC3-Amplifikate

Intron 2:

Primärdenaturierung: 2 min bei 92°C,
Denaturierung 10 s bei 92°C, Annealing 1 min bei 63°C, Elongation 12 min bei 68°C (10 Zyklen),
Denaturierung 10 s bei 92°C, Annealing 1 min bei 63°C, Elongation 13 min bei 68°C (10 Zyklen),
Denaturierung 10 s bei 92°C, Annealing 1 min bei 63°C, Elongation 15 min bei 68°C (10 Zyklen),
finale Elongation: 10 min bei 68°C.
Besonderheiten: Verwendung von dünnwandigen Reaktionsgefäßen
Polymerase: *Expand 20 kbPLUS* (Roche)

siehe auch CVF2-Amplifikate, identische Bedingungen

Intron 3:

Primärdenaturierung: 4 min bei 95°C,
Denaturierung: 25 s bei 95°C, Annealing: 30 s bei 56.5°C, Elongation: 6 min bei 72°C (32 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von *GC-rich Solution* (Roche)
Polymerase: *Fast Start* Taq-DNA-Polymerase (Roche)

Intron 16:

Primärdenaturierung: 40 s bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 56.5°C, Elongation: 4 min 30 s bei 68°C (32
Zyklen), finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von *GC-rich Solution* und *FastStart-Puffer* (Roche)
Polymerase: Taq-DNA-Polymerase (Roche)

Intron 29:

Primärdenaturierung: 1 min 30 s bei 94°C,
Denaturierung: 15 s bei 94°C, Annealing: 30 s bei 57°C, Elongation: 7 min bei 68°C (10 Zyklen),
Denaturierung: 15 s bei 94°C, Annealing: 30 s bei 56°C, Elongation: 10 min bei 68°C (20 Zyklen),
finale Elongation: 7 min bei 68°C.
Polymerase: *Expand Long Template PCR System* (Roche)

Intron 32:

Primärdenaturierung: 1 min bei 94°C,
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 63.5°C, Elongation: 3 min bei 68°C (5 Zyklen),
Denaturierung: 20 s bei 94°C, Annealing: 30 s bei 61°C, Elongation: 3 min 30 s bei 68°C (30
Zyklen), finale Elongation: 7 min bei 68°C.
Polymerase: *AGS Gold Polymerase* (Hybaid-AGS)

Amplifikate neuer homologer Gene**Intron 12 (CVF-H3):**

siehe CVF1-Amplifikate, identische Bedingungen

Intron 27 (CVF-H4):

Primärdenaturierung: 4 min bei 95°C,
Denaturierung: 20 s bei 95°C, Annealing: 30 s bei 56.5°C, Elongation: 5 min bei 71°C (32 Zyklen),
finale Elongation: 10 min bei 72°C.
Besonderheiten: Verwendung von *GC-rich Solution* (Roche)
Polymerase: *Fast Start* Taq-DNA-Polymerase (Roche)

Intron 36 (CVF-H5):

Primärdenaturierung: 2 min bei 94°C,
Denaturierung: 15 s bei 94°C, Annealing: 30 s bei 57°C, Elongation: 3 min bei 68°C (30 Zyklen),
finale Elongation: 10 min bei 68°C.
Polymerase: *Expand* Taq-DNA-Polymerase (Roche)

PCR-Programme für Genome Walking:

1. Runde

Primärdenaturierung: 4 min bei 95°C

Denaturierung: 20 s bei 95°C, Annealing + Elongation: 3 min bei 72°C (7 Zyklen),

Denaturierung: 20 s bei 95°C, Annealing + Elongation: 3 min bei 67°C (32 Zyklen),

finale Elongation: 7 min bei 67°C

2. Runde

Primärdenaturierung: 4 min bei 95°C

Denaturierung: 20 s bei 95°C, Annealing + Elongation: 3 min bei 72°C (5 Zyklen),

Denaturierung: 20 s bei 95°C, Annealing + Elongation: 3 min bei 67°C (20 Zyklen),

finale Elongation: 7 min bei 67°C

Polymerase: *Fast Start* Taq-DNA-Polymerase (Roche)

Veröffentlichungen

Teilergebnisse dieser Arbeit wurden als Poster-Präsentationen auf Kongressen veröffentlicht.

1. Initial characterization of the cobra venom factor gene.

Bammert H, Kunze B, Fritzing DC, Bredehorst R, Andrä J, Vogel CW.
XVIII International Complement Workshop, Snowbird/Utah 2000.
Immunopharmacology 49, 12 (Abstract) (2000).

2. Structure of the cobra venom factor (CVF) gene and comparison with the human C3 gene.

Bammert H, Kunze B, Li Y, Fritzing DC, Bredehorst R, Vogel CW.
XIX International Complement Workshop, Palermo 2002.
Int. Immunopharmacology 2, 1275 (Abstract) (2002).

3. Cobra venom factor is a member of a multi-copy gene family in cobra.

Bammert H, Fritzing DC, Bredehorst R, Vogel CW.
XIX International Complement Workshop, Palermo 2002.
Int. Immunopharmacology 2, 1288 (Abstract) (2002).

Danksagung

An dieser Stelle möchte ich all jenen danken, ohne die die Fertigstellung dieser Arbeit so nicht denkbar gewesen wäre.

Herrn Prof. Dr. Dr. C.-W. Vogel danke ich für die Überlassung des interessanten Themas, seine Fürsprache und Unterstützung. Mein Dank gilt ebenso Herrn Prof. Dr. R. Bredehorst für sein Interesse an meiner Arbeit, seine stete Diskussionsbereitschaft sowie die Möglichkeit, diese Arbeit in seinem Arbeitskreis zu vollenden.

Allen Mitgliedern des Arbeitskreises danke ich für die freundschaftliche Zusammenarbeit und die nette Atmosphäre. Bärbel Kunze danke ich besonders für die Einarbeitung in die Thematik, viele Anregungen und Hilfestellungen und ihr stetes Interesse am Fortgang dieser Arbeit.

Danken möchte ich Dr. Jörg Andrä, Dr. Bijan Bambai, Benjamin Bockisch, Dr. Kerstin David, Susanne Deckers, Dr. Thomas Grunwald, Dr. Sven Heiligtag, Khadiga Jenjan, Dr. Hyun-Jin Kim, Johanna Kölln, Dr. Eric Nebling, Alexander Paul, Lars Redecke, Mladen Simonovic, Dr. Daniel Wehrhahn und Dr. Patrick Ziegel Müller und vielen anderen für ihre Unterstützung, ihr Interesse, für Vorschläge und kritische Kommentare und auch für ihre Freundschaft. Ganz besonders möchte ich daher Dr. Edzard Spillner danken.

Mein Dank gilt auch Connie West und Leonie Bubenheim.

Darüber hinaus möchte ich Dr. Ute Haker, Olaf Nagel, Ankica Manojlovic und Ute Lipprandt für die Unterstützung bei der Durchsicht der Arbeit danken.

Ingke Braren, Sonja During und Stefan Schiebeler danke ich für ihre experimentelle Unterstützung im Rahmen ihrer Schwerpunktarbeiten. Yan Li gilt mein Dank für ihren experimentellen Beitrag im Rahmen ihrer Diplomarbeit.

Mein Dank gilt außerdem Dr. R. Werner für Vorschläge und methodische Tipps zum Screening mit dem Digoxigenin-System sowie Dr. D.C. Fritzing für wertvolle Diskussionen.

Lebenslauf

Persönliche Angaben

Name: Holger Bammert
geboren am: 07. März 1968
Geburtsort: Eltville am Rhein
Staatsangehörigkeit: deutsch

Schulbildung

1974 - 1978 Grundschule in Rauenthal
1978 - 1987 Gymnasium in Eltville und Geisenheim
26. Mai 1987 Erlangung der Hochschulreife

Zivildienst

10/1987 - 05/1989 Schwerbehindertenbetreuung beim Caritasverband Wiesbaden

Studium

10/1989 - 08/1992 Chemie an der Universität Hamburg
11. August 1992 Vordiplom
09/1992 - 06/1993 Chinesisch an der Universität Nanjing (VR China)
10/1993 - 11/1997 Fortsetzung Chemie an der Universität Hamburg
Thema der Diplomarbeit: „Untersuchungen zur Expression des rekombinanten Kobrafaktors im Baculovirus-System“
17. November 1997 Diplom

Promotion

01/1998 - 01/2003 Promotion am Fachbereich Chemie der Universität Hamburg,
Abteilung für Biochemie und Molekularbiologie.
31. Januar 2003 Disputation

Stipendien

09/1992 - 06/1993 Stipendium der Alfried Krupp von Bohlen und Halbach-Stiftung
07/1993 - 06/1994 Stipendium der Studienstiftung des deutschen Volkes

Berufstätigkeit

01/1998 - 12/2001 Wissenschaftlicher Angestellter am Fachbereich Chemie der
Universität Hamburg