

**Analyse der Struktur, Funktion und Evolution der
Astacin-Protein-Familie**

**Frank Möhrlen
Heidelberg, 2002**

INAUGURAL-DISSERTATION

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-mathematischen Gesamtfakultät

der

Ruprecht-Karls-Universität

Heidelberg

vorgelegt von **Frank Möhrle**

aus Heidelberg

Tag der mündlichen Prüfung: 28.06.2002

Thema

**Analyse der Struktur, Funktion und Evolution
der Astacin-Protein-Familie**

Gutachter:

Prof. Dr. Robert Zwillig

Prof. Dr. Richard Herrmann

Die vorliegende Arbeit wurde am Lehrstuhl Physiologie des Instituts für Zoologie der Universität Heidelberg unter der wissenschaftlichen Leitung von Prof. Dr. R. Zwillig angefertigt. Für die Überlassung des interessanten Themas und die Betreuung meiner Arbeit gilt Herrn Prof. Dr. Robert Zwillig mein bester Dank.

Prof. Dr. Richard Herrmann danke ich für die Übernahme des Koreferates.

Ich möchte auch allen herzlich danken, die darüber hinaus diese Arbeit unterstützt haben:

Prof. Dr. Werner Müller für die weitere Bereitstellung meines Arbeitsplatzes und die Förderung nach der Emeritierung von Prof. Dr. R. Zwillig.

Der Landesgraduiertenförderung Baden-Württemberg für die einjährige finanzielle Unterstützung.

Priv. Doz. Dr. Günter Vogt für die wertvolle Unterstützung und seine Diskussionsbereitschaft bei den Flusskrebs-Arbeiten.

Vom deutschen Krebsforschungszentrum Dr. Martina Schnölzer für die gute Zusammenarbeit bei den massenspektrometrischen Analysen, Dr. Hans-Richard Rackwitz für die Peptidsynthesen und Dr. Hans Heid für die Durchführung der Edman-Sequenzierungen.

Prof. Dr. Günter Plickert (Universität Köln) für die schöne Zusammenarbeit im Rahmen der *Hydractinia* Astacine.

Dr. Harald Hutter vom MPI für medizinische Forschung in Heidelberg für die Einführung in die Mikroinjektionstechnik und GFP-Analyse bei *C. elegans*.

Prof. Dr. Vitalinao Pallini und Dr. Luca Bini (Universität Siena) für die Zusammenarbeit bei den 2D-Elektrophoresen.

Dr. Uri Frank für die vielen Gespräche und Diskussionen, sowie die gemeinsame Arbeit im „Labor 403“.

Prof. Dr. Monika Hassel, Prof. Dr. Thomas Leitz und Dr. Gebhard Geier die mir in ihrer Heidelberger Zeit in vielen Diskussionen wichtige Anregungen gaben.

Allen ehemaligen Mitgliedern des Labors „Zwillig“ Anja, Stefanie, Tobias, Patrick und die Ehemaligen in der Physiologie Lutz, Heike, Jürgen, Steffi und Tobias. Für die gute Arbeitsatmosphäre im „Labor 403“ Regina, Jasenka, Mark, Ibrahim und Gabi.

Rüdi für die Hilfe bei der Korrektur des Manuskripts.

Und nicht zuletzt meiner Frau Claudia und meinen Eltern, dass ihr immer für mich da wart.

INHALTSVERZEICHNIS

I. ZUSAMMENFASSUNG	1
II. EINLEITUNG	3
1. Strukturelle und funktionelle Vielfalt proteolytischer Enzyme	3
2. Klassen proteolytischer Enzyme	4
3. Die Astacin-Protein-Familie	5
4. Die Metzinkine	8
5. Ausgangspunkt und Problemstellung	9
III. MATERIAL UND METHODEN	11
1. Allgemeine Materialien	11
1.1. Geräte	11
1.2. Chemikalien und Artikel	11
1.3. Puffer und Lösungen	12
1.4. Bakterienstämme	12
1.5. Cosmide	12
1.6. Plasmide	12
1.7. Oligonukleotide	12
1.8. Enzyme und Produkte für molekularbiologische Methoden	15
1.9. Peptide	16
2. Tierhaltung	16
2.1. <i>Hydractinia echinata</i>	16
2.2. <i>Caenorhabditis elegans</i>	17
2.3. <i>Astacus astacus</i>	18
3. Allgemeine molekularbiologische Methoden	18
3.1. Arbeiten mit <i>Escherichia coli</i>	18
3.2. RNA-Isolierung	20
3.3. DNA-Isolierung	21
3.4. Reverse Transkription und Polymerasekettenreaktion	21
3.5. Rapid Amplification of cDNA Ends (RACE)	23
3.6. Gelelektrophorese und Klonierung von DNA	25
3.7. DNA-Sequenzierung	27
4. Allgemeine biochemische Methoden	27
4.1. Herstellung von Antikörpern	27
4.2. Protein-Isolierung	28
4.3. SDS- Polyacrylamid-Gelelektrophorese (SDS-PAGE)	29
4.4. Immunoblot	30

5. Zweidimensionale (2D-) Gelelektrophorese	31
5.1. 2D-Gelelektrophorese nach der IPG-SDS-PAGE Methode	31
5.2. Westernblot von 2D-Gelen	34
5.3. Immunoblot von 2D-Gelen	34
5.4. Mikrosequenzierung	35
5.5. Massenspektrometrische Analyse	35
6. Chromatographische Methoden	36
6.1. Gewinnung des Krestrypsins durch Affinitätschromatographie	36
6.2. Gewinnung des Astacins durch Ionenaustausch-Chromatographie und Gelfiltration	37
6.3. Reinigung von Astacin und Astacin-homologen Proteinen durch Affinitätschromatographie	37
7. Aktivitätsmessung proteolytischer Enzyme	37
7.1. Aktivitätsmessung von Trypsin mit BAPNA	38
7.2. Aktivitätsmessung von Astacin mit STANA	38
7.3. Aktivitätsmessungen mit dansylierten Substraten	38
7.4. Aktivitätsmessungen mit FITC-Casein	38
8. GFP-Fusionsproteine und Expressionsnachweis in <i>C. elegans</i>	39
8.1. Expressionsvektoren	39
8.2. Herstellung der Injektionslösung	39
8.3. Herstellung der Injektionskanülen	39
8.4. Injektion	39
8.5. Identifizierung von Transformanten	40
8.6. Dokumentation	40
9. Expressionsnachweis durch <i>in situ</i> Hybridisierung in <i>Hydractinia echinata</i>	40
9.1. Vorbehandlung der Tiere	40
9.2. Herstellung der <i>in situ</i> Sonden	41
9.3. <i>In situ</i> Hybridisierung	41
10. Präparation und Immunhistochemie von <i>Astacus astacus</i>	41
10.1. Magensaftentnahme und Präparation des Hepatopankreas	41
10.2. Herstellung von Paraffin-Schnitten	42
10.3. Übersichtsfärbungen	42
10.4. Immunhistochemie	43
11. Proteolyse und massenspektrometrische Analyse von Modellpeptiden	44
11.1. Spaltungsversuche mit synthetischen Modellpeptiden	44
11.2. MALDI-TOF Massenspektrometrie	44
12. Computergestützte Analysen	44
12.1. Programme für molekulargenetische Analysen	44
12.2. Datenbankanalyse Astacin-homologer Gene	45
12.3. Phylogenetische Analyse	46
12.4. <i>C. elegans</i> und das Internet	46

IV. ERGEBNISSE	48
1. Astacin-homologe Proteine in <i>C. elegans</i>	48
1.1. Datenbankanalyse	48
1.2. Transkriptom-Analyse	51
1.2.1. RT-PCR und EST-Projekte	51
1.2.2. Abgleich der cDNA mit den genomischen Sequenzdaten	55
1.3. Funktions-Analyse	57
1.3.1. GFP-Fusionsproteine	57
1.3.2. Auswertung der GFP, RNAi und microarray Projekte	62
1.4. Proteom-Analyse	65
2. Astacin-homologe Proteine in <i>Hydractinia echinata</i>	74
2.1. RT-PCR/RACE	74
2.2. Klonierung und Sequenzdaten von <i>Hydractinia echinata</i> Astacin 1 und 2	75
2.3. <i>In situ</i> Hybridisierung	79
3. Datenbankanalyse und phylogenetische Analyse	82
3.1. Identifizierung Astacin-homologer Gene (BLAST, FASTA)	82
3.2. Analyse der Domänenstruktur	87
3.3. Molekularphylogenetische Analyse	91
4. Aktivierung von Pro-Astacin	99
4.1. Immunhistochemie	99
4.2. Immunoblotting	103
4.3. Analyse der Aktivierungsstelle (Prozessierung)	104
4.4. Massenspektrometrische Analyse	107
V. DISKUSSION	109
1. Astacin-homologe Proteine in <i>C. elegans</i>	109
2. Astacin-homologe Proteine in <i>Hydractinia echinata</i>	117
3. Molekularphylogenetische und strukturelle Untersuchungen der Astacin-Protein-Familie	119
4. Aktivierung von Pro-Astacin	125
VI. LITERATUR	129

VII. ANHANG	139
1. Abkürzungen	139
2. Aminosäuren und Nukleotide	140
3. Abbildungsverzeichnis	141
4. Tabellenverzeichnis	142
5. Eigene Publikationen und Kongressberichte	143
6. Sequenzdaten	144
6.1. Sequenzdaten der Astacin-homologen Gene von <i>C. elegans</i>	144
6.2. Korrigierte Proteinsequenzen von <i>C. elegans</i> (FASTA Format)	177
6.3. Sequenzvergleich der Protease-Domänen der Astacin-Protein-Familie	181

I. ZUSAMMENFASSUNG

Astacin (E.C. 3.4.24.21) aus dem Kaumagen des Flusskrebse *Astacus astacus* L. ist der Prototyp einer Familie von Zink-Endopeptidasen (Familie M12A, nach der Nomenklatur von Rawlings und Barrett), deren Mitglieder bei Vertebraten, Invertebraten und Bakterien gefunden werden. Die Mehrzahl der bisher beschriebenen Astacine enthalten dabei neben einem katalytischen Modul, der eigentlichen Astacin-Domäne, noch weitere Domänen die der Verankerung in der Zellmembran oder regulatorischen Aufgaben, etwa im Rahmen der Morphogenese und Differenzierung, dienen. In den letzten Jahren haben insbesondere die Genom- und Transkriptom-Analysen vieler Organismen die Anzahl annotierter Sequenzen, die Astacin-ähnliche Domänen enthalten, stark anwachsen lassen.

(1) Bei dem Nematoden *Caenorhabditis elegans* konnte durch Datenbanksuchen mit dem BLAST-Algorithmus die überraschend große Zahl von 40 Astacin-homologen Genen identifiziert werden. Die Transkriptom-Analyse durch RT-PCR Experimente sowie die Auswertung der EST-Sequenzen und der *microarray* Ergebnisse von Kim et al. (2001) ergab, dass mit Ausnahme eines Pseudogens alle anderen Astacin-homologen Proteine auch tatsächlich exprimiert werden. Die gefundenen cDNA-Sequenzen wichen dabei zum Teil erheblich von den Spleißvorhersagen der *wormpep* oder *Genie Gene* Datenbanken ab und wurden für eine richtige Übersetzung der Transkripte in Protein-Sequenzen anhand der experimentellen Daten korrigiert. Um erste Aussagen über die Funktion dieser Proteine machen zu können erfolgte an ausgewählten Beispielen die Lokalisierung *in vivo* durch GFP-Reporter-Proteine. Anhand des Expressionmusters konnte für zwei Proteine auf eine Verdauungsfunktion und auf eine entwicklungsrelevante Rolle geschlossen werden. Die Auswertung der RNAi Untersuchungen von Maeda et al. (2001) und der *microarray* Analyse von Jiang et al. (2001) legen dabei für einige weitere Astacin-homologe Proteine in *C. elegans* eine entwicklungssteuernde Funktion nahe. Bei einer Proteom-Analyse durch 2D-Elektrophorese konnten zwar keine Astacin-homologen Proteine identifiziert, jedoch einige neue Proteine der *C. elegans* Proteinkarte (*reference map*) zugeordnet werden.

(2) Aus den Basisgruppen des Tierreichs (z.B. Cnidaria) lagen bisher kaum Daten über Astacin-homologe Proteine vor. Hier konnten nun zwei neue Astacin-homologe Gene bei dem Modellorganismus *Hydractinia echinata* (Cnidaria, Hydrozoa) charakterisiert werden, die als *Hydractinia echinata* Astacin 1 und 2 (HEA-1 und HEA-2) bezeichnet wurden. *In situ* Hybridisierungen deuten eine Rolle bei der Kopfbildung während der Metamorphose und bei der Differenzierung von Tentakelzellen an.

(3) Durch BLAST- und FASTA-Suchen in den allen verfügbaren Datenbanken konnten im Rahmen dieser Arbeit insgesamt 106 verschiedene Astacin-homologe Proteine und 77 weitere unvollständige Genom-, EST- oder N-terminale Sequenzfragmente gefunden werden. Die phylogenetische Analyse und die Analyse der Domänenstruktur ließen ein allgemeines Evolutions-Prinzip erkennen, das an der Astacin-Protein-Familie insgesamt beobachtet werden kann, welches aber bei den Astacinen von *C. elegans* besonders ausgeprägt und mit neuen Elementen in Erscheinung tritt. Die katalytische Kette des Astacins kann nach Anhängung von regulatorischen Einheiten in verschiedenen

Kompositionen als ein Modul in unterschiedlicher Umgebung eingesetzt werden. Somit entstehen strukturell neue Proteine, die verschiedene neue Funktionen übernehmen können. Weiterhin erlaubte die Stammbaum-Analyse (*Neighbor-Joining*) eine Einteilung der Astacine in mehrere phylogenetische Gruppen (A: Cnidaria-Astacine, B: Meprine, C: BMP-1/Tolloid-ähnliche Astacine) oder Subgruppen (D3: *hatching*-Enzyme, D2: *Drosophila* Astacine, D1 und D4: *C. elegans* Astacine) die auf jeweils ein Vorläufer-Protein zurückgehen müssen, das sich nach Genduplikationen divergent weiterentwickelte.

(4) Als Modellprotein für die Untersuchung eines möglicherweise gemeinsamen Aktivierungsmechanismus aller Pro-Astacine wurde das Flusskrebs-Astacin verwendet. Die immunhistochemischen, proteinbiochemischen und die massenspektrometrischen Untersuchungen von Substrat-Peptiden zeigten, dass Pro-Astacin in *Astacus astacus* autokatalytisch im proximalen Bereich des Hepatopankreas vor Übertritt in den Magen aktiviert wird. Darüber hinaus konnte auch gezeigt werden, dass Astacin in der Lage ist, eine Arginin-Alanin Bindung zu hydrolysieren, die sich bei vielen Astacinen an der Aktivierungsstelle findet. Dadurch kann die Mehrzahl der Pro-Astacine ebenfalls autokatalytisch aktiviert werden.

II. EINLEITUNG

1. Strukturelle und funktionelle Vielfalt proteolytischer Enzyme

Eiweißspaltende Enzyme spielen eine Schlüsselrolle bei den verschiedensten physiologischen Vorgängen. Die Pionierarbeiten des Heidelberger Physiologen Wilhelm Kühne in der zweiten Hälfte des 19. Jahrhunderts lieferten erste Aufschlüsse über die Funktion von proteolytischen Enzymen bei der Proteinverdauung der Säugetiere (für eine Übersicht siehe Neurath und Zwilling 1986). Der intestinale Abbau von Nahrungsproteinen ergibt dabei allerdings weniger als die Hälfte der für die Proteinbiosynthese benötigten Aminosäuren (Barrett 1977). Ein wesentlicher Anteil wird durch intrazelluläre Degradation körpereigener Proteine bereitgestellt. Dafür verantwortlich sind grundlegende physiologische Abläufe, wie die lysosomale Proteolyse oder die ubiquitin-vermittelte Proteolyse im Cytoplasma (Bond und Butler 1987; Orłowski 1990; Rock und Goldberg 1999).

Aufgrund ihrer Fähigkeit, das Zielsubstrat spezifisch erkennen zu können, sind Peptidasen auch für die Aktivierung von Proteinvorstufen zuständig. Das Prinzip der spezifischen („limitierten“) Proteolyse wurde am Beispiel der Verdauungsenzyme der Säuger erschlossen, die als Pro-Enzyme (Zymogene) in Magen und Pankreas synthetisiert werden (Palade 1975; Stroud et al. 1977; Huber und Bode 1978; Neurath 1984).

Während der Biosynthese von Proteinen kommt es zur co- und posttranslationalen Prozessierung von Proteinvorstufen wie zum Beispiel bei der Abspaltung der Signalsequenz durch den membranständigen *signal peptidase complex* (SPC) oder der in den Mitochondrien vorliegenden *mitochondrial processing peptidase* (MPP) (Shelness et al. 1993; Ito 1999). Die Blutgerinnung wird über eine Kaskade von proteolytischen Aktivierungen durch Serinproteasen vermittelt, die eine schnelle und wirksame Antwort auf Verletzungen gewährleisten. Das proteolytische Kaskadensystem der Fibrinolyse ist zuständig für den Abbau von Blutgerinnsel, während die Komplementreaktion bei der Beseitigung von inflammatorischen Partikeln und pathogenen Organismen zum Einsatz kommt (Neurath 1986; Patthy 1990). Weitere proteolytische Kaskadensysteme finden sich bei dem Renin-Angiotensinsystems (Lumbers 1999), dem Kallikrein-Kinin-System (Margolius 1984; Margolius 1998) oder der Akrosomenreaktion bei der Befruchtung (Neurath 1986). Peptidhormone wie Insulin (Steiner et al. 1974; Steiner et al. 1996) und Glukagon (Steiner et al. 1996) oder die Neuropeptide neuroendokriner Zellen (Muller und Lindberg 1999) werden räumlich und zeitlich proteolytisch aktiviert und in ihrer physiologischen Wirkung reguliert. Im Rahmen der Abwehr von Infektionen oder Entzündungen werden verschiedene proteolytische Systeme gleichzeitig angeschaltet. Neben den Komplementenzymen werden die Proteasen aus Mastzellen, neutrophilen Granulozyten und cytotoxischen T-Lymphozyten aktiviert (Neurath 1989; Bond und Butler

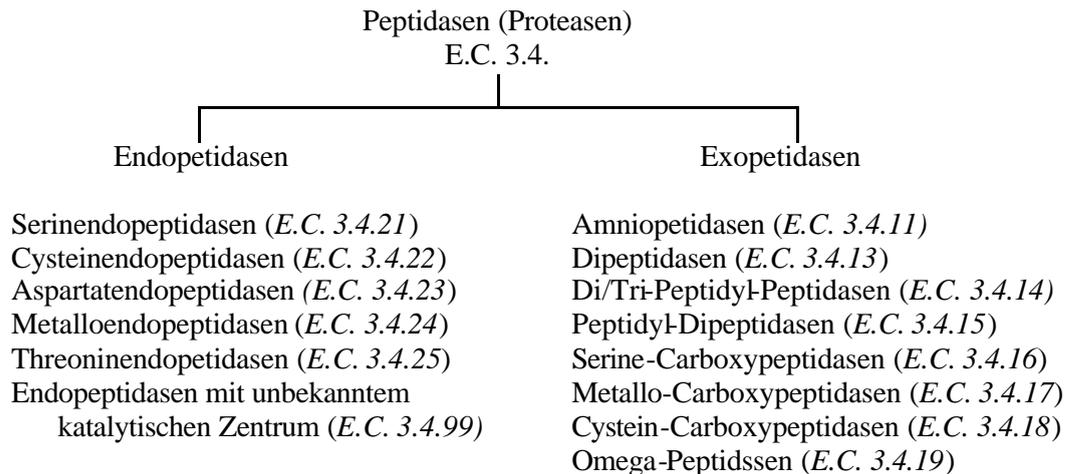
1987). Weiterhin erfolgt in verschiedenen B- und T-Zelltypen der intrazelluläre Abbau von Antigenstrukturen über das 26S- und das Immuno-Proteasom (York et al. 1999; Cascio et al. 2001) und die Aktivierung des *p105* Vorläufers des Transkriptionsfaktors NFκB, welcher bei der Differenzierung der Immunzellen eine zentrale Bedeutung besitzt (Coux und Goldberg 1998).

Im Verlauf der Entwicklung und Differenzierung von Organen oder Organismen spielen stadienspezifisch exprimierte Proteasen eine große Rolle. Dazu zählen beispielsweise Proteasen, die den Transkriptionsfaktor *Dorsal* aus der NFκB-Familie proteolytisch aktivieren und somit indirekt eine Rolle bei der Regulation der dorso-ventralen Körpersymmetrie in *Drosophila melanogaster* spielen (Mohier 1993; Morisato und Anderson 1995). Am Schlüpfvorgang von Embryonen beteiligt sind die sogenannten *hatching enzymes* (Inohaya et al. 1997; Geier und Zwillig 1998; Lei et al. 1999; Fan und Katagiri 2001). Intensiv untersucht werden gegenwärtig jene Proteasen (MMP, *matrix metallo proteases*) die am Umbau der extrazellulären Matrix beteiligt sind. Ihr Wirkungsspektrum ist dabei nicht auf die Strukturproteine der extrazellulären Matrix beschränkt, sondern umfasst auch die Regulation und Aktivierung anderer Proteasen, von Protease-Inhibitoren, Wachstumsfaktoren, Zelloberflächenrezeptoren oder Zell-Adhäsionsmolekülen. Dementsprechend vielfältig sind die physiologischen Aufgaben, die Matrix-Metalloproteasen übernommen haben (Khasigov et al. 2001; Sternlicht und Werb 2001).

Fehlfunktionen der proteolytischen Steuerungssysteme haben unter Umständen schwere pathologische Störungen zur Folge. Beispiele sind Hämophilien (Geddes et al. 1989), Lungenemphysem (Thompson 1982), die Auflösung des interstitiellen Bindegewebes durch tumorassoziierte Enzyme (Curran and Murray 2000), die Alzheimer'sche und die Huntington'sche Krankheit (Nixon et al. 2000; Wellington und Hayden 2000; Checler et al. 2000).

2. Klassen proteolytischer Enzyme

Zur Klassifizierung proteolytischer Enzyme (Proteasen, Peptidasen) schlug Hartley 1960 eine Einteilung nach dem katalytischen Mechanismus vor, wobei das Katalyse-Zentrum für die Namensgebung der Gruppe verwendet werden sollte (Hartley 1960). Auf dieser Grundlage wurden von der internationalen *Enzyme Commission (E.C.)* ab 1972 die Peptidasen in Klassen eingeteilt, wobei zusätzlich zwischen Endopeptidasen, die innerhalb einer Polypeptidkette schneiden, und Exopeptidasen, die eine oder mehrere endständige Aminosäurereste abspalten, unterschieden wird. Seit 1992 wird diese aktuelle Liste (*Enzyme List*), die darüber hinaus auch alle weiteren bekannten Enzyme systematisch katalogisiert, von dem *Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)* herausgegeben (<http://www.chem.qmw.ac.uk/iubmb/enzyme>). Das NC-IUBMB sieht heute folgende systematische Einteilung der proteolytischen Enzyme vor:



Ein Problem der Einteilung proteolytischer Enzyme allein aufgrund des katalytischen Mechanismus ist, dass molekulare Strukturen oder Evolutionszusammenhänge nicht berücksichtigt werden. Rawlings und Barrett schlugen deshalb 1993 die Einführung eines neuen Systems vor. Eine Protein-Familie muss danach eine statistisch signifikante Übereinstimmung in der Aminosäuresequenz aufweisen (Rawlings und Barrett 1993). Dies basiert auf dem Evolutionsgedanken, dass Proteine aus der selben Familie sich aus einem gemeinsamen Vorläufer entwickelt haben und nach der Definition von (Reeck et al. 1987) homologe Proteine sind. Jede Protein-Familie wird in diesem System mit einem Buchstaben nach dem Katalyse-Mechanismus (S, T, C, A, M oder U für Serin, Threonin, Cystein, Aspartat, Metallo- oder *unknown*) und einer arabischen Nummer benannt. Darüber hinaus werden homologe Familien zu *clans* zusammengefasst, bei denen die einzelnen Mitglieder soweit divergiert sind, dass eine Verwandtschaft auf Primärstrukturebene nicht mehr erkennbar ist. Daher wird hier eine Analyse der dreidimensionalen Struktur mit einbezogen (Barrett et al. 1998).

3. Die Astacin-Protein-Familie

Astacin (*E.C. 3.4.24.21*), eine kollagenolytische Zink-Endopeptidase aus dem Kaumagen des Flusskrebses *Astacus astacus* L., wurde 1967 von Pfeleiderer und Mitarbeiter erstmals beschrieben (Pfeleiderer et al. 1967; Stöcker und Zwilling 1995). Das aus 200 Aminosäuren bestehende, aktive Verdauungsenzym besitzt ein Molekulargewicht von 22614 Dalton (Titani et al. 1987). Es wird in den F-Zellen des Hepatopankreas als Pro-Enzym synthetisiert und in aktiver Form extrazellulär im Kaumagen gespeichert (Vogt et al. 1989; Geier et al. 1997). Die Aminosäuresequenz und die dreidimensionale Struktur des Flusskrebsenzym zeigt, dass es sich bei diesem Protein um den Prototyp einer neuen Familie von Proteasen handelt (Titani et al. 1987; Stöcker et al. 1993).

Im Unterschied zu den bisher bekannten Metalloproteasen, deren zentrales Zinkion im Grundzustand tetraedrisch koordiniert vorliegt, besitzt das Zinkion des Astacins fünf

Zinkliganden in trigonal bipyramidaler Anordnung. Dabei sind die Histidinreste His-92, His-102 und das katalytische Wassermolekül (Sol300) koplanar zum Zinkion angeordnet. Ein weiterer Histidinrest (His-96) befindet sich oberhalb und ein Tyrosinrest (Tyr-149) unterhalb dieser Ebene. Die dreidimensionale Gesamt-Struktur des Astacinmoleküls ist von nierenförmiger Gestalt. Es ist aus einer aminoterminalen und carboxyterminalen Domäne aufgebaut, die durch den ausgedehnten Spalt des aktiven Zentrums voneinander getrennt sind. Das katalytisch essentielle Zinkion ist am Boden dieses Spalts lokalisiert. Die aminoterminal Domäne besteht im wesentlichen aus vier parallelen und einem antiparallelen β -Strang, und zwei langen α -Helices. Über die Cysteinreste Cys42 und Cys198 ist der antiparallele Strang mit der zentralen α -Helix B verknüpft. Die carboxyterminale Domäne weist bis auf eine α -Helix kaum reguläre Strukturen auf. Beide Domänen sind mittels einer Disulfidbrücke zwischen Cys42 und Cys198 kovalent miteinander verbunden. Eine zusätzliche Verbindung der Domänen wird durch den Aminoterminus des aktiven Proteins (Ala-1) hergestellt, der über eine Salzbrücke mit Glu-103 verbunden ist. Betrachtet man die Aminosäuresequenz, enthält das bei allen Astacinen stark konservierte, zinkbindende Motiv **HExxHxxGFxHExxRxDRD** die drei Histidin-Zink-Liganden, die katalytische Glutaminsäure Glu-93 und das strukturell wichtige Gly-99, welches die zentrale α -Helix beendet und dem Protein die Rückfaltung des dritten Histidins in die Nähe des Zink-Ions ermöglicht. Der Zinkligand Tyr-149 ist Teil eines weiteren charakteristischen Strukturelements, des Methionin-haltigen 1,4- β -Turns mit dem Motiv **SxMHY** (Bode et al. 1992; Gomis-Rüth et al. 1993; Stöcker et al. 1993).

Die von Geier et al. analysierte genomische Organisation des Astacins weist auf eine Pro-Form des Enzyms hin, die auf Proteinebene im Flusskrebis nicht isoliert werden konnte (Geier et al. 1997). Demzufolge hat das Prä-Proenzym eine aminoterminal Ver längerung von 49 Aminosäureresten, sodass in diesem Zustand die Salzbrücke zwischen Ala-1 und Glu-103 offensichtlich nicht ausgebildet werden kann. Möglicherweise ist hier ein proteolytischer Aktivierungsmechanismus ausgebildet, wie er in vergleichbarer Form bei den trypsinähnlichen Serinproteasen beschrieben wurde (Huber und Bode 1978). Diese werden aktiviert, indem das Propeptid vom inaktiven Zymogen abgespalten wird und der neu entstandene Aminoterminus mit Asp-194, dem Nachbar des katalytisch aktiven Ser-195, interagieren und so das aktive Zentrum des Enzyms stabilisieren kann (Stöcker et al. 1993).

Die Astacin-Protein-Familie (Peptidase-Familie M12A, nach der Nomenklatur von Rawlings und Barrett) kann strukturell eindeutig durch das konservierte zinkbindende Motiv, den konservierten Methionin-haltigen Turn und die räumliche Anordnung der Zink-Liganden von anderen Metalloproteasen unterschieden werden (Dumermuth et al. 1991; Stöcker et al. 1993; Bond and Beynon 1995).

Die Astacin-Protein-Familie umfasst nach dem aktuellen Stand der Literatur inzwischen mehr als 40 Mitglieder. Dazu gehören die membrangebundenen Meprine A und B aus der Niere und dem Dünndarm von Maus, Ratte und Mensch. Ein morphogenetisch aktives

Protein ist BMP-1 (*bone morphogenetic protein 1*), von dem gezeigt werden konnte, dass es an der Knochen- und Skelettbildung beim Menschen beteiligt ist. BMP-1 konnte als Procollagen C-Proteinase I identifiziert werden, welche durch die Bildung reifer Kollagenmonomere für die Assemblierung der Kollagenfasern und damit der extrazellulären Matrix eine entscheidende Rolle spielt. HCE und LCE beim Knochenfisch *Oryzias latipes*, CAM-1 bei der Wachtel und AEA bei *Astacus astacus* spielen eine Rolle beim Schlüpfvorgang der Embryonen. An der Musterbildung sind Tolloid und Tolkin bei *Drosophila melanogaster*, UVS.2 bei *Xenopus laevis*, BP10 und SpAN bei *Paracentrotus lividus* und HMP1, HMP2 sowie FARM-1 bei *Hydra vulgaris* beteiligt. Ein prokaryotischer Vertreter aus der Astacin Familie ist ebenfalls bekannt, das Flavastacin aus *Chryseobacterium meningosepticum*.

Die folgende Auflistung gibt einen Überblick und Literaturhinweise zu den bisher beschriebenen Mitglieder der Astacin-Protein-Familie:

Protein	Vorkommen	Literatur (Übersichtsartikel)
Astacin	<i>Astacus astacus</i>	Barrett et al. 1998, Kap. 405
	<i>Chionoecetes opilio</i>	Klimova et al. 1991
AEA	<i>Astacus astacus</i>	Geier and Zwilling 1998
Alveolin	<i>Oryzias latipes</i>	Shibata et al. 2000
BMP-1	Mensch, Maus, Ratte, <i>S. purpuratus</i> , <i>Xenopus laevis</i> , <i>Gallus gallus</i>	Barrett et al. 1998, Kap.411 Reynolds et al. 2000 (<i>Gallus</i>)
BP-10	<i>Paracentrotus lividus</i>	Lhomond et al. 1996
CAM-1	<i>Coturnix coturnix</i>	Elaroussi and DeLuca 1994
Colloid	<i>Gallus gallus</i>	Liaubet et al. 2000
HCE-1	<i>Oryzias latipes</i>	Barrett et al. 1998, Kap. 410
HCE-2	<i>Oryzias latipes</i>	Barrett et al. 1998, Kap. 410
FARM-1	<i>Hydra vulgaris</i>	Kumpfmüller et al. 1999
Flavastacin	<i>Chryseobacterium meningosepticum</i>	Tarentino et al. 1995
HCH-1	<i>Caenorhabditis elegans</i>	Hishida et al. 1996
HMP-1	<i>Hydra vulgaris</i>	Yan et al. 1995
HMP-2	<i>Hydra vulgaris</i>	Yan et al. 2000a
LCE	<i>Oryzias latipes</i>	Barrett et al. 1998, Kap. 409
Meprin α	Mensch, Maus, Ratte,	Barrett et al. 1998, Kap.406, 408
Meprin β	Mensch, Maus, Ratte,	Barrett et al. 1998, Kap.407, 408
mTLL-1, colloid	Mensch, Maus, <i>Gallus gallus</i>	Scott et al. 1999
mTLL-2	Mensch, Maus	Scott et al. 1999
Nephrosin	<i>Cyprinus carpio</i>	Hung et al. 1997
PMP-1	<i>Podocoryne carnea</i>	Pan et al. 1998
SPAN	<i>Strongylocentrotus purpuratus</i>	Wardle et al. 1999
TBL-1	<i>Aplysia californica</i>	Liu et al. 1997
Tolkin	<i>Drosophila melanogaster</i>	Barrett et al. 1998, Kap. 413
Tolloid	<i>Drosophila melanogaster</i>	Barrett et al. 1998, Kap. 413
UVS-2	<i>Xenopus laevis</i>	Fan and Katagiri 2001
Xolloid	<i>Xenopus laevis</i>	Piccolo et al. 1997
Zebrafisch-Tolloid	<i>Danio rerio</i>	Blader et al. 1997

In dieser Liste sind ausschließlich in wissenschaftlichen Publikationen veröffentlichte Daten berücksichtigt. Hierin nicht eingeschlossen sind Gene oder Proteine, die allein durch Genom- oder EST-Sequenzierungsprojekte identifiziert wurden und als nicht annotierte, reine Sequenzdaten in verschiedenen Datenbanken gespeichert sind. Barrett et al. listen auf Basis dieser Sequenzierungsprojekte aber schon eine Reihe weiterer Astacin-homologer Proteine auf, wobei beispielsweise bei dem Nematoden *Caenorhabditis elegans*, neben dem HCH-1 Protein (Hishida et al. 1996), nun noch die überraschende Anzahl von 14 weiteren Astacin-homologen Genen postuliert wird (Barrett et al. 1998).

Choriolysin L und H sind wie das Flusskrebsastacin kompakte kleine Proteasen. Die anderen Mitglieder der Astacin-Familie besitzen dagegen carboxyterminal der Protease-Domäne eine Vielzahl von regulatorischen Einheiten. Viele der Astacin-Familien-Mitglieder enthalten EGF- (*epidermal growth factor*) ähnliche Domänen, sowie die sogenannten CUB- (*complement subcomponents Cqr/C1s, embryonic sea urchin protein Uegf, BMP-1*) Domänen. Eine threoninreiche Domäne ist bei SPAN und BP10, eine serin-/threoninreiche Domäne bei BMP-1 des Seeigels zu finden. Die Meprine enthalten zusätzlich eine MAM- (*meprin subunits, A-5 protein, and receptor protein tyrosin phosphatase C m*) Domäne, eine transmembrane und eine cytoplasmatische Domäne (Übersicht: Bond and Beynon 1995).

Alle Mitglieder der Astacin-Familie weisen aminoterminal der Protease-Domäne weitere Sequenzabschnitte auf. Das Signal- oder Prä-Peptid dient als intrazelluläres Lokalisationssignal und wird bereits während der Proteinsynthese bei der Passage durch das Endoplasmatische Reticulum abgespalten. Der Pro-Teil ist vermutlich für die Inaktivierung der Protease verantwortlich und muss zu deren Aktivierung abgespalten werden. Obwohl alle Astacine als Pro-Enzym synthetisiert werden war bisher in keinem Fall der Mechanismus oder der an der Aktivierung beteiligten Proteasen bekannt.

4. Die Metzinkine

Phylogenetische Untersuchungen zeigen, dass die Familie der Astacine sich mit verwandten Proteinfamilien zur Super-Familie der Metzinkine zusammenfassen lässt. Diese enthält als gemeinsame Struktur-Elemente das verkürzte zinkbindende Motiv HExxHxxGxxH und den methioninhaltigen Abschnitt mit der Sequenz MxY. Zur Super-Familie der Metzinkine fassen Stöcker et al. neben den Astacinen (M12A) die Familien der Adamalysine (M12B,C; Schlangengiftproteinasen), der Serralysine (M10B; bakterielle extrazelluläre Proteinase) und der Matrixine/Matrix-Metalloproteasen (M10A; Säugetierkollagenasen, MMP's) zusammen (Stöcker et al. 1995). Schlangenhauf et al. konnten 1998 den Metzinkinen als eine weitere Familie die Leishmanolysin-ähnlichen Enzyme (M8) zuordnen (Schlangenhauf et al. 1998). Obwohl diese Familien nur geringe Sequenzähnlichkeiten zeigen, bestehen doch starke topologische Übereinstimmungen zwischen den Sekundärstrukturen der jeweiligen katalytischen Module der Proteine.

Gemeinsame Strukturelemente sind die vier parallelen und der antiparallele β -Strang der aminoterminalen Domäne, sowie die drei langen α -Helices, die bei allen Metzinkinen in identischer Reihenfolge auftreten und die in der gleichen Weise räumlich angeordnet sind. Die Metzinkine haben sich demnach durch divergente Evolution aus einem gemeinsamen Vorläufer-Enzym entwickelt. Die große Zahl der Punktmutationen in den Proteinsequenzen und das Vorkommen der Metzinkine sowohl bei Pro- wie auch bei Eukaryonten spricht dafür, dass ein gemeinsames Vorläufer-Enzym bereits in einem Prokaryonten vorhanden war, der entwicklungsgeschichtlich noch vor der Prokaryonten/Eukaryonten-Divergenz einzuordnen ist (Stöcker et al. 1995; Rawlings und Barrett 1995).

Rawlings und Barrett fassen die Metzinkine wiederum mit anderen Familien zu einem Klan von Metalloproteasen (*Clan* MA) zusammen. Diese enthalten als gemeinsames Element die kurze zinkbindende Konsensussequenz HExxH und einen Glutamaterest 18-72 Aminosäurereste carboxyterminal zu diesem Motiv. Zu diesem *clan* gehören die Familien der *immune inhibitor A* Proteine (M6), Snapalysine (M7), Säugetierkollagenasen (M10), Gametolysine (M11), Astacine (M12), Deuterolysine (M35), Cytophagalysine (M43), *pregnancy-associated plasma protein A* (M46) und prtB (M57) (Barrett et al. 1998). Allerdings liegen bisher kaum dreidimensionale Strukturanalysen für diese Proteine vor. Aufgrund der geringen Sequenzhomologie muss bei dieser Klassifizierung deshalb die Frage offen bleiben, ob sich hier noch eine ursprüngliche Verwandtschaft widerspiegelt, oder ob es sich in manchen Fällen um konvergente Bildungen handelt.

5. Ausgangspunkt und Problemstellung

Durch die fast ubiquitäre Verbreitung der Astacine von den Bakterien bis zum Menschen und die unterschiedlichen physiologischen Funktionen dieser Enzyme, wird die Frage aufgeworfen, welche molekularen Veränderungen und Evolutionsabläufe dem zugrunde liegen. Es stellt sich etwa das Problem, wie ein Verdauungsenzym des Flusskrebses im Verlauf der Evolution zu einem menschlichen Knochenwachstumsfaktor werden kann. Anhand solcher orthologen Sequenzen kann versucht werden den phylogenetischen Stammbaum der Aufspaltung der Spezies und der damit einhergehenden Proteinevolution nachzuvollziehen. Auffällig ist in diesem Zusammenhang auch, dass innerhalb dieser Protein-Familie eine Vielzahl regulatorischer Domänen als carboxyterminale Erweiterungen der Protease-Domäne auftreten kann. So findet sich die CUB-Domäne als Regulations-Modul sowohl im Flusskrebs (AEA), als auch beim BMP-1 des Menschen oder sogar als Bestandteil von Serinproteasen des Komplementsystems. Weiterhin stellt sich die Frage weshalb bei einigen wenigen Organismen im Tierreich (*C. elegans*) sich eine große Anzahl von paralogenen Genen entwickelt hat, während andere offenbar nur wenige Genkopien aufweisen. Die innerhalb von *C. elegans* auftretenden Gene müssen durch Duplikationen entstanden sein und haben danach wahrscheinlich unterschiedliche physiologische Funktionen übernommen. Durch den Vergleich der Struktur und Funktion

von solchen orthologen und paralogen Proteinen können darüber hinaus auch neue Einblicke in die molekulare Evolution überhaupt gewonnen werden. Neben allen Veränderungen, die diese Proteine im Verlauf der Evolution erfahren haben, sind möglicherweise einige allgemeine Mechanismen hoch konserviert. Ein Ansatzpunkt bietet hier die Analyse der Propeptide und der Zymogenaktivierung, welche bei allen Astacin-homologen Proteinen gefunden werden.

Um anhand dieser Fragestellungen einen weiteren, vertieften Einblick in die molekulare Evolution der Astacine zu bekommen, wurden verschiedene Ansätze gewählt:

1. Untersuchung der überraschend großen Anzahl Astacin-homologer Gene bei *C. elegans*, die gleichzeitig einen Beitrag zur Erforschung des Genoms und Proteoms dieses Organismus liefern soll.
2. Identifizierung und Charakterisierung neuer Astacin-homologer Proteine aus Organismen an der Basis der Metazoen (*Hydractinia echinata*, Hydrozoa). Diese sollen vergleichend in die phylogenetischen Studien einbezogen werden.
3. Datenbankanalyse und phylogenetische Analyse aller Astacin-homologen Proteine. Dies schließt auch die Analyse der im Rahmen der Genomsequenzierungsprojekte angefallenen, nicht annotierten Sequenzdaten ein.
4. Untersuchung des Aktivierungsmechanismus der Zymogene am Beispiel des Flusskrebsses *Astacus astacus*.

III. MATERIAL UND METHODEN

1. Allgemeine Materialien

1.1. Geräte

Gerätetyp	Lieferant/Hersteller
Branson Sonifier® L15	Branson
Chromatographie (HPLC)	
HPLC-Pumpe 2248	LKB, Bromma
HPLC-Controller 2152	LKB, Bromma
Spectroflow 757	Krates
Fraktionssammler Foxy	ISCO
Fluoreszenz Mikroskop	Zeiss
Geldokumentation BioDocAnalyze®	Biometra, Göttingen
Lumineszenz Spectrometer LS 50	Perkin-Elmer, Überlingen
PCR-Maschinen	
Mini Cycler™	MJ Research, Watertown
TGradient™	Biometra, Göttingen
Semi-Dry Blotkammer CE	Sigma Aldrich, Deisendorf
Speed Vac Concentrator	Savant
Stereomikroskop	Leica
Ultrazentrifuge L70	Beckmann
UV/VIS Spectrometer Lambda 2S	Perkin-Elmer, Überlingen
Zentrifugen	
Centrikon H401	Kontron
Labofuge 400R	Heraeus

1.2. Chemikalien und Artikel

Hier nicht verzeichnete laborübliche Chemikalien wurden in p.A. Qualität verwendet und von den Firmen Merck KGaA (Darmstadt), Carl Roth GmbH (Darmstadt), Serva Feinbiochemika GmbH (Heidelberg), Sigma Aldrich Chemikalien GmbH (Deisenhofen) oder Boehringer Mannheim GmbH (Mannheim) bezogen.

Bezeichnung	Lieferant
Agarose (Elektrophoresegrad, 100%)	Roth, Darmstadt
Antibiotika (Ampicillin, Carbenicillin, Tertracyclin, Kanamycin)	Serva, Heidelberg
Dextransulfat	Sigma, Deisendorf
Diethylpyrocarbonat (DEPC)	Sigma, Deisendorf
Ficoll Type 400	Sigma, Deisendorf
Inject Keyhole Limpet Hemocyanin (KLH)	Pierce, Rockford, Illinois
Saccharose, Rnase/Dnase frei	Roth, Darmstadt
Standard-I-Nährbouillon	Merck, Darmstadt

1.3. Puffer und Lösungen

Für das Ansetzen von Puffern und Lösungen wurde Wasser aus einer Milli-Q®-Anlage verwendet. Die Speisung dieser Anlage geschieht mit voll entsalztem Wasser, der Ohm'sche Widerstand des Wassers bei der Entnahme beträgt 18,2 MΩ.

Alle Lösungen, die für die Präparation von RNA verwendet wurden, enthielten DEPC-H₂O statt Milli-Q-Wasser. Zur DEPC-Behandlung wurde 1 l Milli-Q-Wasser mit 1 ml DEPC versetzt, 24 h bei Raumtemperatur inkubiert und 30 min autoklaviert.

1.4. Bakterienstämme

Escherichia coli XL-1-Blue Stamm zur Vermehrung von bluescript und pBC Vektoren (Stratagene, Heidelberg).

Escherichia coli OP 50 Der Uracil-auxotrophe Stamm dient als Nahrungsquelle für *Caenorhabditis elegans* Kulturen und wurde von Prof. E. Schierenberg, Universität Köln, zur Verfügung gestellt.

1.5. Cosmide

Die *C. elegans* Cosmide C05D11, C07D10, K04E7, F42A10 und T04G9, die im Rahmen *C. elegans* sequencing project zur Verfügung stehen, wurden von Dr. A. Coulson, Sanger Center, Cambridge, bezogen.

1.6. Plasmide

Klonierungsvektoren wurden von folgenden Quellen bezogen:

pBluescript II SK+ (Genbank/EMBL accession number X53328), Stratagene, Heidelberg
puC18 (Genbank/EMBL accession number L09136), Amersham Pharmacia, Freiburg
pGEM®-T Vector, Promega, Mannheim.

Die Vektoren pPD95.85, pPD114.35 zur Herstellung von *C. elegans* Fusionsproteinen stammen aus dem „Fire Lab Vector Kit“ und wurden von Dr. A. Fire, Department of Embryology, Carnegie Institution of Washington, Baltimore bezogen.

Das Plasmid pBx, das als Injektionsmarker für die Etablierung transgener *C. elegans* Linien benutzt wurde, stammt von Prof. R. Schnabel, Institut für Genetik, Technische Universität Braunschweig.

1.7. Oligonukleotide

Alle Oligonukleotide wurden im Synthesemaßstab 0,02 µmol von Interaktiva, Ulm bezogen und wurden vom Hersteller RP-HPLC gereinigt. Aus den lyophilisierten Proben wurden nach Angaben des Herstellers Stammlösungen mit einer Konzentration von 100 pmol/µl in DEPC-H₂O hergestellt und bei -20°C gelagert. Die Konzentration der Gebrauchslösungen betrug, wenn nicht anders vermerkt, 10 pmol/µl.

1.7.1. Oligonucleotide für RT-PCR (*C. elegans*)

Bezeichnung	Sequenz
C05D11.6/C48/F1	TCT GTA CCA CAA ACC GAA AC
C05D11.6/C48/R3	GCA TAA TGG AAG CGT AGT CA
C07D10.4/C13/F1	CGG TGG AAG TGG TAA ATG
C07D10.4/C13/R3	ATA ATT AAC CCG TCC TGA GC
C24F3/F1	ATT GGC AGC TGA TTG TAT TG
C24F3/R3	TGT TTG TCC TCG CAC TTT AC
C34D4/F1	TCC TAG AAC CGC TGT AGA CA
C34D4/R3	GTA GTT CGG TGG TGA TGC T
D1022/F1	TGC TAA TGT CGG GTC AAG
D1022/R3	ACC TTT ACA ATC GCT TCC TC
F09E8/F1	TTC GTA TGT TGG CAG AGC
F09E8/R3	TCG TCG TAC TTG TCA CCT TC
F20G2/F1	GGC AAC GTT TGA AGA TTT
F20G2/R3	GC CGT ACA TAT CAC ATT TTC
F39D8/F1	CTT TCA CTT ATG CCA TCA CC
F39D8/R3	CCA TCG CCA AGA CTT ACT
F42A10/C00/F1	GCT GCC TTA AAC TAC GGT
F42A10/C00/R3	GCT TTG ATA TGC CAG TAA CA
F46C5/F1	TGG GAT CGA GCA TAC TCA
F46C5/R3	GAC AAT TAT TGG GGT TTG GT
F54B8/F1	GTC AAG ACC ACC CAA ACT
F54B8/R1	TGG AGC CAT TTA TCT CTG AG
F56A4/F1	GCC GGA AAT TTT TAG AGG
F56A4/R3	CTT TCG GAG CAA GTT TAC TG
F58A6/F1	GAA ATG ACG ATG GGT GTT
F58A6/R3	GTT ACG GGA ATT AGT GAA GC
F58B4/F1	GGG ACA TGC TCT TGG ATT
F58B4/R3	ATC ACG AAG TGT TGG AAG TC
K03B8.1/F1	TCA CAA ACT TGC GTT ACT TTC
K03B8.1/R3	TCT TTC GGA GCA TTT ATT GA
K03B8.2/F1	GCG ATC CGA AAA GAT GTG
K03B8.2/R3	CAT TGT ACG TCG TTT GTT CA
K03B8.3/F1	ATC GCC TTA ACT TTA TCT GC
K03B8.3/R3	ACT TCT TGG GAA ACA TTG TC
K03B8.5/F1	TCC CCA CCG ATT TAC ATA
K03B8.5/R3	ACT TCC GAT GGT ATC ACA AC
K04E7.3/C39/F1	CAC TGT AAT CCC GTC CAG
K04E7.3/C39/R3	CCG AAA CAA TTC TCC AAT AA
K06A4.1/F1	CGA ACC TGA AAA AGA CGA T
K06A4.1/R3	GAC TTG GCA CGT TGA TAG AT
T02B11/F1	CCG GAT TGA ATG AGT ATT TG
T02B11/R3	GAT CTT TGG GCA CTA TGG
T04G9.2/C41/F1	GTC TGA AGA CGG AAA GTG GT
T04G9.2/C41/R3	ATG AAA TTG GGC AGT TGG
T11F9.3/F1	CTT GTT CCG TTT GAT TAT GG
T11F9.3/R3	TCG CAT GTA TGT TGA CAG AA
T11F9.5/F1	CTT TAC GCA TGG ACA ATG A
T11F9.5/R3	CTT CCC CTT CAC AAG AAC AT

T11F9.6/F1	GAA GAA ATG TCG CCG AAA AA
T11F9.6/R3	ATT GCA TGA CAC TCC CAA CA
T23F4/F1	TTC CGA TGC CTT ACT CTT T
T23F4/R3	ACT TTT GAT CTC TGG CGT AG
T23H4/F1	TTT TGA CTG TTT CGG TGG TC
T23H4/R3	GAA GCA TCT TTG GGG TGA C
Y95B8A.1/F1	CAC CGG AGC CAT TCA CAT T
Y95B8A.1/R3	GGA CTG CTC GTG CCA AAA G
4R79.1/F1	TTT GAC AGT TAC CGA CGA AC
4R79.1/R3	GAC CGA ATT TTTATC CAC AAC
Gpd3/F1	GGA GAC TAC CTT CTT GTC GC
Gpd3/R1	AGT GGT CAT AAG TCC CTC AA
Oligo dT primer	TTT TTT TTT TTT TTT T

1.7.2. Oligonukleotide für GFP-Fusionsproteine (*C.elegans*)

C05D11:GFP/Kpn1/F1	GGG GTA CCC TGA CAC GCT GAC CCA TAC
C05D11:GFP/Sac1/R1	CGA GCT CTT GAG TGA AGA TGC CAA GA
C07D19:GFP/BamH1/F1	GGG GTA CCC GAT CCT CGC ATT CTA
C07D19:GFP/BamH1/R1	CGG GAT CCT TCC GCC AAA GTC ATT TAG
K04E7:GFP/Kpn1/F1	GGG GTA CCC CGG ACC ACA GTA AAG AAT
K04E7:GFP/Sac1/R1	CGA GCT CTG ACA AGA AAG GCA CAA AG
T04G9:GFP/Kpn1/R1	GGG GTA CCC GCT GGG TAG TGG AGT TG
T04G9:GFP/Pst1/F1	AAC TGC AGC TTT TCG GAA GAC TTT TGC

1.7.3. Oligonukleotide für RT-PCR und RACE (*Hydractinia echinata*)

HeAst(mp1)/F1	CAA ACG CAA TCA ACG AAG
HeAst(mp1)/R1	AAA TCA GCT TGG CCA AGT G
HeAst-mp1/F2	CAA GGC TGC AAA CGC AAT CAA CGA
HeAst-mp1/R2	GCT TTG CCA AGT GGA GCT GAC GAT G
HEA1/PstI/F6	AAC TGC AGC AAG GCT GCA AAC GCA ATC AAC GA
HEA1/NotI/R6	TAG CGG CCG CGC TTT GCC AAG TGG AGC TGA CGA TG
HEA2/PstI/F7	AAC TGC AGA GGG CAT ATG GAA AAC TGG AAA TGC TT
HEA2/NotI/R7	TAG CGG CCG CGT GAA GAG CAG CCG CCT CCT TTA TAG
HEA1/F8	AGG CTA ACA CCA AAT GCA TCA CAT
HEA1/R8	GGT TCA AAC CTA AGG CGT GTC CTA
HEA2/F9	GAA GAT GGC CCG GTG CAA AAA C
HEA2/F10	ATC CAG ACG AGA AGG AGG GC
HEA2/R10	GAT CGA TAA CCA ACT GGT GAA
SMART II	AAG CAG TGG TAA CAA CGC AGA GTA CGC GGG
3'-RACE-CDS	AAG CAG TGG TAA CAA CGC AGA GTA C(T) ₃₀ N ₋₁ N (N = A,C,G oder T; N ₋₁ = A,G oder C)
5'-RACE-CDS	(T) ₂₅ N ₋₁ N (N = A,C,G oder T; N ₋₁ = A,G oder C)
UPM long	CTA TAT CGA CTC ACT ATA GGG CAA GCA TGT GTA ACA ACG CAG AGT
UPM short	AAG CAG TGG TAA CAA CGC AGA GT

1.7.4. Sequenzierungs- und Transkriptions-Oligonukleotide

M13/22-mer sequencing primer	GCC AGG GTT TTC CCA GTC ACG A
M13revers/24-mer sequencing primer	GAG CGG ATA ACA ATT TCA CAC AGG
GFP/F1 sequencing primer	GGG CAC AAA TTT TCT GTC AGT G
GFP/R1 sequencing primer	TCT GCT AGT TGA ACG CTT CCA T
SP6/18-mer promoter primer	ATT TAG GTG ACA CTA TAG
T7/20-mer promoter primer	TAA TAC GAC TCA CTA TAG GG
T3/24-mer promoter primer	GCG CGA AAT TAA CCC TCA CTA AAG

1.8. Enzyme und Produkte für molekularbiologische Methoden

Alle Enzyme wurden mit den vom Hersteller mitgelieferten Puffer-Stammlösungen verwendet.

Enzym	Aktivität	Hersteller/Lieferant
Calf Intestine Alkaline Phosphatase (CIAP)	1 U/μl	MBI Fermentas, St. Leon-Rot
DNase I	10 U/μl	Boehringer, Mannheim
Elongase® Enzym Mix		Life Technology, Karlsruhe
Klenow Fragment, exo ⁻	5 U/μl	MBI Fermentas, St. Leon-Rot
Pfu DNA Polymerase	3 U/μl	Promega, Mannheim
PowerScript™ Reverse Transcriptase	200 U/μl	Clontech, Heidelberg
RevertAid H minus M-MuLV Reverse-Transkriptase	200 U/μl	MBI Fermentas, St. Leon-Rot
SP6 RNA Polymerase	20 U/μl	MBI Fermentas, St. Leon-Rot
T3 RNA Polymerase	20 U/μl	MBI Fermentas, St. Leon-Rot
T4 DNA Ligase	5 U/μl	MBI Fermentas, St. Leon-Rot
T4 Polynucleotid Kinase	10 U/μl	MBI Fermentas, St. Leon-Rot
T7 RNA Polymerase	20 U/μl	MBI Fermentas, St. Leon-Rot
Taq DNA Polymerase LC, recombinant	1 U/μl	MBI Fermentas, St. Leon-Rot
Taq DNA Polymerase, recombinant	5 U/μl	Life Technology, Karlsruhe
Restriktionsendonukleasen	10 U/μl	MBI Fermentas, St. Leon-Rot
Bam HI; Bsh TI (Age I); Eco 32 I (Eco RV); Eco RI; Hind III; Kpn I; Nhe I; Not I; Pst I; SacI; Sal I; Sma I		

Reaktionskomplettausstattungen	Hersteller/Lieferant
Oligotex mRNA Mini Kit	QIAGEN, Hilden
Oligotex direct mRNA Micro Kit	QIAGEN, Hilden
QIAamp Blood And Tissue Kit	QIAGEN, Hilden
QIAGEN Plasmid Midi Kit	QIAGEN, Hilden
NucleoSpin® Plasmid Mini	Macherey & Nagel, Düren
NucleoSpin® Extract	Macherey & Nagel, Düren
SureClone™ Ligation Kit	Amersham Pharmacia, Freiburg
DNA T7 Sequencing Kit	Amersham Pharmacia, Freiburg

Sonstige Produkte	Konzentration	Hersteller/Lieferant
(α - ³⁵ S)dATP	10 mCi/ml	Amersham Pharmacia, Freiburg
ATP	100 mM	MBI Fermentas, St. Leon-Rot
dNTP	10 mM	MBI Fermentas, St. Leon-Rot
dNTP für <i>long range PCR</i>	10 mM	Life Technology, Karlsruhe
NTP	10 mM	MBI Fermentas, St. Leon-Rot
DNA-Marker		MBI Fermentas, St. Leon-Rot
GeneRuler 100bp Ladder plus		
GeneRuler 1kb Ladder		
MassRuler low range		
MassRuler high range		
RNA-Marker low range		MBI Fermentas, St. Leon-Rot
Ribonuclease Inhibitor	50 U/ μ l	MBI Fermentas, St. Leon-Rot

1.9. Peptide

Peptide wurden auf einem modifizierten *Applied Biosystems 430A Peptide Synthesizer* durch das *in situ* Neutralisations-Protokoll für die Festphasen Boc-Chemie Peptidsynthese nach Schnölzer et al. (1992) durchgeführt. Die Entfernung der Schutzgruppen und die Abspaltung des neu synthetisierten Peptids von der Säule wurde mit Fluorsäure durchgeführt. Die Peptide wurden anschließend RP-HPLC gereinigt und auf einem *matrix-assisted laser desorption/ionisation time-of-flight* (MALDI-TOF) Massenspektrometer (MS) überprüft. Die Synthese erfolgte durch Dr. H.-R. Rackwitz, Deutsches Krebsforschungszentrum, Heidelberg.

Bezeichnung	Aminosäuresequenz
CePepZinc_C07D10.6	HEMMHVVGIFYHEHERC
CePepZinc_K04E7.3	HEVGHALGFWHEQARC
ProAst-N	SPIIPEAARALYYNDGMFEGDIKLR
ProAst-C	CDGMFEGDIKLRAGRQPARVG
AstG	RAGRQPARVGAAILGDEY
AstR	RAGRQPARVRAAILGDEY
AstR*	R*AGR*QPAR*VGAAILGDEY R* = D(-)-Arginin

2. Tierhaltung

2.1. *Hydractinia echinata*

Hydractinia echinata Kolonien als Auswuchs auf Schneckenschalen der Gattung *Littorina* wurden im Wattenmeer der Insel Sylt, Deutschland, gesammelt. Die Kultivierung im Labor erfolgt wie in Frank et al. (2001) beschrieben. Zur Herstellung des künstlichen Seewassers werden 4 kg Tropic Marine™ (Dr. Biener GmbH Aquarientechnik, Deutschland) in 95l destilliertem Wasser gelöst. Die erhaltene Lösung sollte eine Dichte von 1.023 g/cm³ und eine Osmolarität von 960 mOsm aufweisen. Der pH-Wert wird mit Salzsäure auf pH 8,2-8,3 eingestellt.

2.2. *Caenorhabditis elegans*

Als Labortier wurde der Wildtyp „N2 Variante Bristol“ von *Caenorhabditis elegans* verwendet, den Sydney Brenner 1974 isolierte. Diese Variante wird seit 1988 in unserem Labor kultiviert und stammte ursprünglich von Prof. E. Schierenberg, Zoologisches Institut, Universität Köln. Die Kulturtechniken wurden modifiziert nach Wood (1988) und Lewis et al. (1995).

2.2.1. Kultivierung von *C. elegans* auf Agarplatten

Als Nahrungsquelle für die Nematoden wurde der Bakterienstamm *Escherichia coli* OP50 verwendet. Dieser Stamm ist Uracil auxotroph, so dass im Kulturmedium, das kein Uracil enthält, ein zu starkes Wachstum der Bakterien verhindert wird. Als Anzuchtmedium wurde Standard-I-Medium (Merck, Darmstadt) in einer Konzentration von 25 g/l eingesetzt.

Für die Herstellung von Agarplatten wurde das Kulturmedium (0,05 M NaCl; 0,1 M KH_2PO_4 , pH 6; 0,25% Pepton; 1,7% Agar) autoklaviert. Anschließend wurde bei etwa 70°C 1mM MgSO_4 , 1mM CaCl_2 , 0,005% Cholesterol zugegeben und das Medium in 14 cm Petrischalen gegossen. Nach dem Aushärten wurde auf der Agarfläche ca. 1 ml *E. coli* Über-Nacht-Kultur zu einem Bakterienrasen ausgestrichen.

Diese Platten wurden mit Nematoden angeimpft, die sich bei 15°C und einer Generationszeit von ungefähr drei Tagen normal entwickeln können. Alle drei bis sieben Tage wurden die Platten unter dem Binokular kontrolliert und nach Verschwinden des Bakterienrasens (Futtermangel) wurden 10 bis 100 Nematoden mit einem Agarstück der alten Platte auf eine neue Platte transferiert. Die alten Platten wurden verworfen.

2.2.2. Kultivierung von *C. elegans* in Flüssigkultur

Aus der Kultivierung von *C. elegans* im Flüssigmedium können größere Mengen für RNA-, DNA- oder Protein-Präparationen erhalten werden.

Wachstum

1 Liter Nährmedium (1000 ml S-Basal (0,1 M NaCl, 0,05 M KH_2PO_4 , pH 6), 10 ml Schwermetalllösung (5,0 mM Na_2EDTA , 2,5 mM $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 1,0 mM $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 1,0 mM $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0,1 mM $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$), 10 ml 1M Kaliumcitrat, pH 6,0, 3 ml 1M CaCl_2 , 3 ml 1M MgSO_4 , 1 ml 5mg/ml Cholesterol und Bakterienpellet von 2 Litern *E. coli* OP50 Über-Nacht-Kultur, siehe 2.2.1) wurde auf 4 Fermbach-Kolben verteilt und mit *C. elegans* angeimpft. Zur Deckung des hohen Sauerstoffbedarfs einer dicht wachsenden Flüssigkultur wurden die Kolben bei 100 rpm auf einem Bakterenschüttler ständig bewegt. Um eine genügend große Ausbeute an Nematoden zu erhalten, wurden die Kulturen 8 bis 14 Tage bei 20°C inkubiert.

Reinigung einer Bakterien - Nematoden Suspension

Die aus der Flüssigkultur gewonnene Nematodensuspension enthält neben *C. elegans* noch *E. coli* und Fragmente von toten Würmern. Bevor mit einer Aufarbeitung begonnen werden kann, müssen diese Komponenten abgetrennt werden. Die Nematodensuspension aus 1 Liter Flüssigkultur wurde in einen hohen Standzylinder überführt und über Nacht bei 4°C sedimentiert. Das Pellet wurde einmal mit M9-Puffer (44 mM KH_2PO_4 , 21 mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 86 mM NaCl, pH 7,0) gewaschen und durch Zentrifugation bei 4500 g, 4°C, 3 min erneut pelletiert. Die Suspension wurde in M9-Puffer aufgenommen, auf 15 ml Zentrifugenröhrchen verteilt, das gleiche Volumen 70%ige (w/v) Saccharose zugegeben und 10 min bei 3000 g zentrifugiert. Bakterien und Wurmfragmente befinden sich im

Pellet, lebende Nematoden im Überstand. Der Überstand wurde vorsichtig abgenommen und noch zweimal mit M9-Puffer gewaschen. Das so präparierte Pellet enthält eine Nematoden-Gesamtfraction, eine Mischung aus adulten Tieren, allen Larvenstadien und Eiern. Das Pellet wurde bis zur RNA-, DNA- oder Protein-Isolierung in flüssigem Stickstoff eingefroren und bei -80°C gelagert.

2.2.3. Transgene *C. elegans* Linien

Für die Herstellung transgener *C. elegans* Linien wurde die pha-1 (*pharynx development abnormal*) III.5.93 e2123 Mutante bei einer Temperatur von 15°C kultiviert. Diese Linie entwickelt sich bei 15°C normal ist aber bei 25°C durch eine ausbleibende Pharynx-Entwicklung zu 100% lethal. Nach der Injektion des *rescue*-Plasmids pBx wurden die transgenen Würmer dann auf Agarplatten bei 25°C gehalten (siehe auch 8.).

2.3. *Astacus astacus*

Astacus astacus (L.) 1758 wurde von der Krebszuchtstation Max Keller (Teichwirtschaft Bernbachtal, Augsburg) bezogen und über mehrere Monate in 300 l fassenden Aquarien mit konstantem Wasser- und Belüftungskreislauf bei 18°C im Zoologischen Institut gehalten. Die Tierhaltung erfolgte darüber hinaus wie in Vogt et al. (1989) beschrieben. Eine Woche vor Präparation und Magensaftentnahme wurden einige Tiere ohne Nahrung in ein gesäubertes Aquarium gesetzt.

3. Allgemeine molekularbiologische Methoden

Alle nicht näher beschriebenen Methoden wurden nach den Protokollen des Laborhandbuchs *Molecular Cloning* durchgeführt (Sambrook und Russell 2001).

3.1. Arbeiten mit *Escherichia coli*

3.1.1. Kultivierung in Flüssigkultur

E. coli wurden bei 37°C bei ausreichender Belüftung in LB-Medium (1% Tryptone, 0,05% Hefeextrakt, 0,05% Kochsalz, pH 7,5) kultiviert. Eine Übernachtskultur enthielt ca. $1-2 \times 10^9$ Zellen/ml. Die Bestimmung des Bakterientiters erfolgte durch photometrische Messung der Zellsuspension ($1 \text{ OD}_{600\text{nm}} = 2,5 \times 10^8$ Bakterien/ml). Transformierte *E. coli* Stämme wurden unter Antibiotika-Selektionsdruck kultiviert.

3.1.2. Kultivierung auf Agarplatten und Blau-Weiß-Selektion

Die Kultivierung von *E. coli* auf LB-Agarplatten (1% Tryptone, 0,05% Hefeextrakt, 0,05% Kochsalz, pH 7,5, 1,5% Agar) erfolgte bei 37°C. Zur Selektion transformierter Bakterien wurden entsprechend der erworbenen Resistenz dem Medium kurz vor der Herstellung der Platten eines der folgenden Antibiotika zugegeben:

	Stammlösung	Endkonzentration
• Ampicilin	50mg/ml in water	100 µg/ml
• Carbenicillin	50mg/ml in water	100 µg/ml
• Kanamycin	10mg/ml in water	100 µg/ml

Rekombinante wurden von nicht-rekombinanten Plasmiden durch die Blau-Weiß Selektion unterschieden. Dazu wurden als Substrate 1 h vor dem Ausplattieren der Bakterien 40 µl einer 2%igen X-Gal Lösung in DMSO in und 7 µl einer 20%igen IPTG Lösung in Wasser ausgestrichen.

3.1.3. Lagerung

Übernachtskulturen wurden mit Glycerin (Endkonzentration 20%) versetzt, in flüssigem Stickstoff schockgefroren und bei -80°C aufbewahrt. Sie können mehrere Jahre gelagert werden.

3.1.4. Transformation

Zur Transformation extrachromosomaler DNA in *E. coli* wurde eine modifizierte Methode von Hanahan et al. oder für große Plasmide und Cosmide die Elektroporation nach Dower et al. angewandt (Hanahan 1983; Dower et al. 1988).

3.1.4.1. Herstellung kompetenter Zellen nach Hanahan

200 ml LB-Medium wurde mit 2 ml einer Übernachtskultur angeimpft und bis zum Erreichen des Bakterientiters von $\text{OD}_{600\text{nm}} = 0,4$ bei 37°C inkubiert. Die Kultur wurde für 15 min auf Eis gestellt und anschließend 10 min bei 4°C , 6000 g zentrifugiert. Die sedimentierten Zellen wurden in 80 ml TFB1 (10 mM MES; 100 mM RbCl; 10 mM $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$, 50 mM $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, pH 5,8) resuspendiert, weitere 15 min auf Eis gestellt und erneut zentrifugiert. Das Pellet wurde in 8 ml TFB2 (10 mM MOPS, 75 mM $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$, 10 mM RbCl, 15% Glycerin, pH 6,5) resuspendiert. Die kompetenten Zellen wurden in 100 µl Portionen in flüssigem Stickstoff schockgefroren und bei -80°C gelagert.

3.1.4.2. Transformation nach Hanahan

Gefrorene kompetente Zellen (100 µl) wurden bei Raumtemperatur aufgetaut und anschließend 5 min auf Eis gestellt. Nach Zugabe von 0,01 bis 0,05 pmol Plasmid-DNA (in maximal 5 µl Volumen) kühlte man 30 min auf Eis. Nach dreiminütiger Inkubation bei 42°C und erneutem Kühlen für 2 min auf Eis gab man 1 ml TY-Medium (2% Tryptone, 0,05% Hefeextrakt, 0,0058% Kochsalz, 20 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, pH 7,0) zu und inkubierte unter Schütteln für 60 min bei 37°C . Von diesem Ansatz wurden 10 µl und 100 µl und der Rest auf LB-Agarose-Platten verteilt und unter Antibiotika- und Blau-Weiß-Selektion über Nacht bei 37°C kultiviert.

3.1.4.3. Herstellung elektroporationskompetenter Zellen

500 ml LB-Medium wurde mit 2 ml einer Übernachtskultur angeimpft und bis zum Erreichen des Bakterientiters von $\text{OD}_{600\text{nm}} = 0,4$ bei 37°C inkubiert. Die Kultur wurde für 15 min auf Eis gestellt und anschließend 10 min bei 4°C , 4000 g zentrifugiert. Die sedimentierten Zellen wurden zweimal mit 500 ml eiskaltem Milli-Q-Wasser gewaschen und wie angegeben zentrifugiert. Das Sediment wurde dann in 2 ml eiskaltem GYT-Medium (10% Glycerin, 0,125% Hefeextrakt, 0,25% Tryptone) suspendiert. Von einer 1:100 Verdünnung dieser Suspension wurde die $\text{OD}_{600\text{nm}}$ bestimmt. Die Zellen wurden daraufhin bis zu einer Konzentration von $2-3 \times 10^{10}$ Zellen/ml ($1 \text{ OD}_{600\text{nm}} = 2,5 \times 10^8$ Zellen/ml) mit GYT verdünnt. Die Konduktivität der Suspension wurde in einer Elektroporations-Küvette bestimmt und musste unter 5 mEq liegen. Andernfalls müssen die Zellen weiter verdünnt werden. Die kompetenten Zellen wurden in 40 µl Portionen in flüssigem Stickstoff schockgefroren und bei -80°C gelagert.

3.1.4.4. Elektroporation

Gefrorene elektrokompetente Zellen (40 μ l) wurden auf Eis aufgetaut. Nach Zugabe von 1 μ l Ligationsansatz wurde die Suspension in eine auf 4°C vorgekühlte Elektroporationsküvette (Bridge, Spaltweite 2,5 mm) überführt. Die Elektroporation wurde mit dem *EQUIBIO Easyject Plus Pulser* bei 2,5 kV, 25 μ F, 201 Ω für 5 ms durchgeführt. Sofort danach wurde 1 ml TY-Medium in die Küvette gegeben, gemischt und die Suspension in einem 2 ml Eppendorfgefäß unter Schütteln für 60 min bei 37°C inkubiert. Von diesem Ansatz wurden 10 μ l und 100 μ l und der Rest auf LB-Agarose-Platten verteilt und unter Antibiotika- und Blau-Weiß-Selektion über Nacht bei 37°C kultiviert.

3.1.5. Reinigung von Plasmid/Cosmid-DNA

Für Plasmid/Cosmid-Präparationen aus *E. coli* im kleinen Maßstab (*mini prep*) wurde der *NucleoSpin[®] Plasmid Purification Kit* und für Präparationen bis 100 μ g DNA (*midi prep*) wurde der *Qiagen[®] Plasmid Midi Kit* nach Angaben der Hersteller Mascherey & Nagel und Qiagen verwendet.

3.2. RNA-Isolierung

3.2.1. Total-RNA Isolierung

Für die RNA-Isolierung wurde die Guanidin-Thiocyanat-Phenol-Chloroform Methode nach Chomczynski und Sacchi (1987) durchgeführt.

Für eine RNA-Aufarbeitung wurden 10 μ l Pellet aus der *C. elegans* Flüssigkultur (siehe 2.2.2) oder 1000-3000 Primärpolypen von *Hydractinia echinata* (siehe 9.) mit 1 ml Solution D (4 M Guanidin Thiocyanat, 25 mM Natriumcitrat, 0,5% Sarcosyl, 0,1 M 2-Mercaptoethanol) versetzt. Zur Homogenisierung der Tiere wurde die Suspension in einen Glashomogenisator gegeben und 5 min auf Eis behandelt. Das Zell-Lysat wurde in ein Zentrifugenröhrchen überführt, mit dem „VortexerTM“ suspendiert und mit 0,1 ml 2 M Natriumacetat, pH 4, 1 ml Phenol (wassergesättigt), 0,2 ml Chloroform-Isoamylalkohol (49:1) versetzt. Diese Suspension wurde gut gemischt, für 15 min auf Eis gestellt und bei 4°C, 10000g zentrifugiert. Die obere wässrige Phase wurde vorsichtig in ein frisches 2 ml Eppendorfgefäß überführt, während DNA und Proteine in der Interphase bzw. organischen Phase verblieben. Nach Zugabe von 1 ml Isopropanol wurde die RNA 1 h bei -20°C präzipitiert und bei 10000g, 20 min in einer Eppendorf Zentrifuge sedimentiert. Das erhaltene Pellet wurde in 0,3 ml Lösung D aufgenommen, die Lösung in ein neues Eppendorf-Gefäß überführt, das gleiche Volumen Isopropanol zugegeben und die RNA wie oben beschrieben präzipitiert. Nach der zweiten Fällung wurde das RNA-Pellet in 0,5 ml 75% Ethanol resuspendiert, sedimentiert, 15 min an der Luft getrocknet und bis zur weiteren Verwendung bei -80°C gelagert.

3.2.2. mRNA Isolierung

Die Isolierung der poly-A⁺ mRNA aus einer Total RNA Probe wurde mit dem „OligotexTM mRNA Mini Kit“ nach Angaben des Herstellers durchgeführt.

Das aus der Total-RNA Extraktion erhaltene Pellet wurde in 200 μ l DEPC-H₂O gelöst, die RNA-Menge photometrisch bestimmt (vgl. 3.2.3) und mindestens 50 μ g für die poly A⁺ Selektion eingesetzt. Die Elution wurde zur Erhöhung der Ausbeute nochmals wiederholt. Als Ergebnis lag die poly A⁺ mRNA in 80 μ l Elutionspuffer vor. Zur Konzentrierung kann die mRNA mit dem gleichen Volumen Ethanol versetzt, 1h bei -20°C präzipitiert und bei

10000g 20 min sedimentiert werden. Das erhaltene Pellet wird luftgetrocknet, in 10 µl DEPC-H₂O aufgenommen und bei -80°C gelagert.

3.2.3. Photometrische Quantifizierung und Überprüfung der RNA

Die einfachste Methode zur Mengenbestimmung von Nukleinsäuren ist die photometrische Quantifizierung. Dazu wurde ein Aliquot der RNA in einer Quarzküvette verdünnt und die Extinktion bei 260 nm im Photometer gemessen. Mit der nachstehenden Formel kann die RNA-Konzentration direkt ermittelt werden:

$$[\text{RNA}] \mu\text{g/ml} = E_{260} \times \text{Verdünnungsfaktor} \times 40$$

Um die Reinheit der RNA zu bestimmen, wurde ein Spektrum im Wellenlängenbereich von 220 nm bis 320 nm aufgenommen. Saubere RNA hat das Absorptionsmaximum bei 260 nm. Der Quotient E_{260} / E_{280} beträgt $2.0 \pm 0,1$.

3.2.4. RNA-Agarosegel-Elektrophorese

Für ein 1 % iges Gel wurde 1 g Agarose in 62,1 ml DEPC-H₂O-a aufgekocht, auf 60°C abgekühlt, 20 ml 5 x Laufpuffer (0,1 M MOPS (3-[N-Morpholino]propansulfonsäure) pH 7,0, 40 mM NaAc, 5 mM EDTA pH 8,0) und 17,9 ml 37 % Formamid deionisiert zugegeben und in die Form gegossen. Zur Probenvorbereitung wurden 4,5 µl RNA (2- 15 µg) mit 2,0 µl 5 x Laufpuffer, 3,5 µl Formaldehyd und 10 µl Formamid (deionisiert) gemischt, 15 min bei 65°C denaturiert und kurz auf Eis gestellt. Dann wurden 2 µl Ladepuffer (50% Glycerin, 1 mM EDTA pH 8,0, 0,1 % Bromphenolblau, 0,1% Xylencyanol) zugegeben und auf das Gel aufgetragen. Die Elektrophorese erfolgte in 1 x Laufpuffer bei 5 V/cm und wurde gestoppt, sobald der Farbmaler 2/3 des Gels durchlaufen hatte. Abschließend wurde 30 min in 0,5 µg/ml Ethidiumbromid in 0,1 M Ammoniumacetat gefärbt und die RNA auf einem UV-Transluminator überprüft.

3.3. DNA-Isolierung

Zur Isolierung von genomischer DNA aus *C. elegans* wurde das Protokoll von E. Moss, Dartmouth College, Hannover, benutzt. Das Protokoll ist in einer *C. elegans* spezifischen Methodensammlung unter <http://cobweb.dartmouth.edu/~ambros/worms/index.html> abrufbar. Für eine Präparation wurden 100 µl Pellet aus einer *C. elegans* Flüssigkultur (siehe 2.2.2) eingesetzt. Die Phenol/Chloroform Extraktion wurde zur Vermeidung von RNA-Kontaminationen zweimal wiederholt.

Abschließend wurde die Konzentration der in 200 µl Puffer vorliegenden DNA im Photometer wie in 3.2.3 für RNA beschrieben bestimmt. Die Konzentration muss allerdings für DNA mit der folgenden Formel berechnet werden:

$$[\text{DNA}] \mu\text{g/ml} = E_{260} \times \text{Verdünnungsfaktor} \times 50.$$

Die genomische DNA wurde bis zur Verwendung bei 4°C gelagert.

3.4. Reverse Transkription und Polymerasekettenreaktion

3.4.1. Reverse Transkription (RT)

Als Standardprotokoll für die Reverse Transkription wurde die RevertAid H minus M-MuLV Reverse-Transkriptase nach den Angaben des Herstellers MBI Fermentas benutzt. Für einen Reaktionsansatz wurden verwendet:

Komponente	eingesetzt	Endkonzentration
5 x RT-Buffer	4 µl	50 mM Tris pH 8,3, 50 mM KCl, 4 mM MgCl ₂
dNTP-Mix	2 µl	10 mM
<i>antisense</i> bzw. oligo-dT <i>primer</i>	2 µl	20 pmol bzw. 0,5 µg
RNase Inhibitor	1 µl	50 U
Enzym	1 µl	200 U
DEPC-H ₂ O	ad 15 µl	

Es wurden 100 ng poly(A)⁺mRNA oder 3 µg RNA in 5 µl DEPC-H₂O gelöst, 5 min auf 65°C erhitzt und sofort auf Eis gestellt. Die RNA-Probe wurde zum Reaktionsansatz gegeben, 60 min bei 42°C inkubiert und bei Reaktionsende 10 min auf 70°C erhitzt. Die neu synthetisierte cDNA wurde mit DEPC-H₂O auf 100 µl aufgefüllt und bis zum Start der PCR bei -20°C gelagert.

3.4.2. Polymerasekettenreaktion (PCR)

3.4.2.1. Standardprotokoll RT-PCR

Für die Amplifizierung von cDNA wurde die Taq-DNA Polymerase nach dem modifizierten Protokoll des Herstellers MBI Fermentas benutzt. Zur Vermeidung von Synthesefehlern wurden 100 U Taq-Polymerase 1 U Pfu DNA Polymerase zugegeben. Für einen 50 µl Ansatz wurden eingesetzt:

Komponente	eingesetzt	Endkonzentration
MgCl ₂ 25mM	6 µl	3 mM
dNTP-Mix	2 µl	0,2 mM
<i>antisense</i> Primer	2 µl	2 µM
<i>sense</i> Primer	2 µl	2 µM
10 x Puffer	5 µl	10 mM Tris-HCl pH 8,8, 50 mM KCl, 0,08% Nonident P40
Taq-Polymerase	2 µl	2 U
RT-Ansatz	10 µl	
H ₂ O	ad 50 µl	

Die Reaktion wurde unter folgenden Reaktionsbedingungen durchgeführt:

Denaturierung:	94°C für 2 min
Zugabe der Taq-Polymerase (<i>hot start</i>)	
Zyklen: 35 x Denaturierung:	94 °C für 45 sec
<i>Annealing</i>	52–64°C für 1 min
Synthese	72 °C für 2 min
Syntheseende:	72 °C für 10 min

Die amplifizierten DNA Moleküle wurden bei -20°C gelagert.

3.4.2.2. Standardprotokoll PCR

Für die Amplifizierung von DNA Molekülen bis 4000 bp wurde das Standard RT-PCR Protokoll (3.4.2.1) mit folgenden Änderungen verwendet:

- A. Es wurden 10 pg-1 µg DNA *template* anstelle des RT-Ansatz eingesetzt.
- B. PCR-Bedingungen für Oligonukleotide mit einer Schmelztemperatur unter 66°C:
- | | |
|--|-------------------------------------|
| Denaturierung: | 94°C für 2 min |
| Zugabe der Taq-Polymerase (<i>hot start</i>) | |
| Zyklen: 35 x Denaturierung: | 94 °C für 45 sec |
| <i>Annealing</i> | 52–64°C für 1 min |
| Synthese | 72 °C für 1 min pro kb Produktlänge |
| Syntheseende: | 72 °C für 10 min |
- C. PCR-Bedingungen für Oligonukleotide mit einer Schmelztemperatur über 66°C:
- | | |
|--|------------------------------------|
| Denaturierung: | 94°C für 2 min |
| Zugabe der Taq-Polymerase (<i>hot start</i>) | |
| Zyklen: 35 x Denaturierung: | 94 °C für 45 sec |
| <i>Annealing</i> und | |
| Synthese | 68°C für 1 min pro kb Produktlänge |
| Syntheseende: | 72 °C für 10 min |

3.4.2.3. Amplifizierung von großen DNA Kopien (*long range PCR*)

Die Amplifizierung von Molekülen über 4000 bp aus genomischer DNA erfolgte mit dem Elongase® Enzym Mix und nach Angaben des Herstellers Life Technologies, Karlsruhe.

Für einen 50 µl Ansatz wurden eingesetzt:

Komponente	eingesetzt	Endkonzentration
dNTP-Mix <i>long range</i>	2 µl	0,2 mM
antisense Primer	2 µl	2 µM
sense Primer	2 µl	2 µM
Puffer B	5 µl	60 mM Tris-SO ₄ pH 9,1, 18 mM (NH ₄)SO ₄ , 2 mM MgSO ₄
Elongase® Enzym	1 µl	
DNA <i>template</i>	100 ng	
H ₂ O	ad 50 µl	

Die Reaktionen erfolgten wie unter 3.4.2.2 beschrieben.

3.5. Rapid Amplification of cDNA Ends (RACE)

Die Darstellung der 5` und 3` Enden von cDNA Molekülen wurde in Anlehnung an den *SMART[®]RACE cDNA Amplification Kit* des Herstellers Clontech, Heidelberg durchgeführt.

3.5.1. Herstellung der cDNA

Für eine Reaktion wurden 100 ng poly(A)⁺ selektierte mRNA in 10 µl DEPC-H₂O gelöst. Für einen 5` oder 3` RACE wurden getrennt eingesetzt:

Komponente	eingesetzt	Endkonzentration
RNA-Lösung	10 µl	
SMART II primer (20pmol/µl)	1 µl	2 µM
5` oder 3` CDS primer (20pmol/µl)	1 µl	2 µM

Die Lösung wurde 2 min auf 70°C erhitzt, sofort auf Eis gestellt und kurz abzentrifugiert. Dann gab man folgende Komponenten zu:

5 x RT-Buffer	4 µl	
dNTP-Mix	2 µl	20 mM
PowerScript™ RT	1 µl	200 U
DTT	0,5 µl	
DEPC-H ₂ O	ad 20 µl	

Der Reaktionsansatz wurde 90 min bei 42°C inkubiert, dann mit DEPC-H₂O auf 100 µl aufgefüllt, 7 min auf 72°C erhitzt und anschließend bei -20°C eingefroren.

3.5.2. 5` und 3` PCR

Für die Amplifizierung der 5` und 3` cDNA wurde die Taq-DNA Polymerase nach dem modifizierten Protokoll des Herstellers MBI Fermentas benutzt. Zur Vermeidung von Synthesefehlern wurden 100 U Taq-Polymerase 1 U Pfu DNA-Polymerase zugegeben. Für einen 50 µl Ansatz wurden eingesetzt:

Komponente	eingesetzt	Endkonzentration
MgSO ₄ 50mM	5 µl	2,5 mM
dNTP-Mix	2 µl	0,2 mM
antisense Primer (5` RACE) oder sense Primer (3` RACE)	2 µl	2 µM
UPM Primer Mix (0,4 µM UPM short, 2 µM UPM long)	2,5 µl	
10 x Puffer mit (NH ₄) ₂ SO ₄	5 µl	75 mM Tris-HCl pH 8,8, 20 mM (NH ₄) ₂ SO ₄ , 0,01% Tween 20
RT-Ansatz 5` oder 3` cDNA	2 µl	
Taq-Polymerase	2 µl	2 U
H ₂ O	ad 50 µl	

PCR-Bedingungen (*touchdown PCR*):

Denaturierung:	94°C für 2 min
Zugabe der Taq-Polymerase (<i>hot start</i>)	
Zyklen: 4 x Denaturierung:	94 °C für 40 sec
<i>Annealing</i> und	
Synthese	72 °C für 2 min 40 sec
Zyklen: 4 x Denaturierung:	94 °C für 40 sec
<i>Annealing</i>	69°C für 40 sec
Synthese	72 °C für 2 min
Zyklen: 30 x Denaturierung:	94 °C für 40 sec
<i>Annealing</i>	66°C für 40 sec
Synthese	72 °C für 2 min
Syntheseende:	72 °C für 10 min

Als Kontrolle wurde von jedem 5' und 3' cDNA-Ansatz eine Reaktion mit zwei genspezifischen (*sense* und *antisense*) Oligonukleotiden unter den gleichen PCR-Bedingungen durchgeführt. An Stelle des UPM Primer Mix wurde 2 µl des komplementären Oligonukleotids verwendet.

3.6. Gelelektrophorese und Klonierung von DNA

3.6.1. DNA-Gelelektrophorese

Entsprechend der Größe der aufzutrennenden DNA wurden Gele von 0,8% bis 2% Agarose in 1 x TAE (50 x TAE: 2 M Tris-Acetat, pH 8,0, 50 mM EDTA) verwendet. Sollten PCR-Proben analysiert werden, wurde zur Probenvorbereitung die DNA 5 min bei 65°C denaturiert und kurz auf Eis gestellt. Dann wurde 5 x Ladepuffer (50% Glycerin, 0,2% SDS, 0,05% Bromphenolblau, 0,05% Xylencyanol in 5 x TAE) zugegeben (Endkonzentration 1 x) und auf das Gel aufgetragen. Von dem entsprechenden DNA-Größenmarker wurden jeweils 500 ng einer gleich behandelten Probe aufgetragen. Die Elektrophorese erfolgte in 1 x TAE bei 5 V/cm und wurde gestoppt, sobald der Farbmarker 2/3 des Gels durchlaufen hatte. Abschließend wurde 15 min in 0,5 µg / ml Ethidiumbromid Lösung gefärbt und die DNA auf einem UV-Transluminator überprüft.

3.6.2. Quantifizierung von DNA

Zur Konzentrationsbestimmung von Plasmiden, Cosmiden oder PCR-Fragmenten wurde ein Aliquot der Probe zusammen mit einem DNA-Standard elektrophoretisch aufgetrennt und die Bandenintensität mit Hilfe der BioDocAnalyse® Software bestimmt. Durch Vergleich mit den bekannten Konzentrationen des Standards wurde so die DNA Konzentration der Probe berechnet.

3.6.3. Isolierung von DNA-Fragmenten aus Agarosegelen

DNA-Fragmente wurden aus Agarosegelen mit Hilfe des *NucleoSpin[®] Extract Kits* nach Angaben des Herstellers Macherey & Nagel, Düren, extrahiert.

3.6.4. Reinigung von DNA-Fragmenten

Zur Reinigung von wässrigen DNA-Lösungen wurde die Probe im Verhältnis 1:1 mit Phenol-Chloroform (1:1) vermischt, 1 min bei 13000g zentrifugiert und die wässrige Phase in ein neues Eppendorfgefäß gegeben. Dazu gab man die zweieinhalbfache Menge 100% Ethanol, präzipitierte die DNA für 1 Stunde bei -20°C, zentrifugierte bei 13000g für 5 min und verwarf den Überstand. Das Pellet wurde noch mal mit 70% Ethanol gewaschen und anschließend für 5-10 min an der Luft getrocknet.

3.6.5. Klonierungen

Die Experimente wurden nach den allgemein gebräuchlichen Standardvorschriften für molekulare Klonierungen durchgeführt (Sambrook und Russell 2001).

3.6.5.1. Restriktion von DNA

Die Reaktionsbedingungen für die sequenzspezifische DNA-Hydrolyse mit Hilfe von Restriktionsendonukleasen wurden entsprechend den Angaben des Herstellers MBI Fermentas gewählt, von dem alle verwendeten Enzyme bezogen wurden. In der Regel

wurde die Enzymmenge für eine Reaktionszeit von einer Stunde berechnet (1U Enzym hydrolysiert 1 µg DNA pro Stunde). Für analytische Proben wurden 0,5-1 µg und für präparative Zwecke 1–10 µg DNA in 20 µl Gesamtvolumen eingesetzt. Für den gleichzeitigen Einsatz von zwei Enzymen in einer Reaktion (Doppelverdau, *double digestion*) wurde nach Möglichkeit das *Universal Y+/Tango[®] Buffer* System des Herstellers verwendet. Für nicht kompatible Enzyme wurde nach der ersten Reaktion das Enzym bei 80°C deaktiviert und der Ansatz Phenol-Chloroform gereinigt.

3.6.5.2. Auffüllen von überstehenden DNA-Enden (*blunting*)

Mit Hilfe der T4-DNA-Polymerase wurden überhängende einzelsträngige DNA-Enden, wie sie beispielsweise bei PCR-Reaktionen entstehen, aufgefüllt. Ein 20 µl Ansatz setzte sich aus 1 µg DNA, 4 µl Reaktionspuffer (Endkonzentration 67 mM Tris-HCl pH 8,8, 6,6 mM MgCl₂, 1 mM DTT), 0,1 mM dNTP und 1 U T4-DNA-Polymerase zusammen. Nach einer Inkubationszeit von 20 min bei 37°C wurde die Reaktion durch Erhitzen auf 70°C für 10 min gestoppt. Anschließend wurde der Reaktionsansatz Phenol-Chloroform gereinigt.

3.6.5.3. Kinasierung von DNA-Fragmenten

Nicht-phosphorierte DNA (aus Protokoll 3.6.5.2) muss vor der Ligation in dephosphorierte Vektoren mit dem Enzym T4-Polynukleotidkinase kinasiert werden. Für einen 20 µl Ansatz wurden 20 pmol DNA, 2 µl Reaktionspuffer (Endkonzentration 50mM Tris pH 7,6, 10 mM MgCl₂, 5 mM DTT, 0,1 mM Spemidine, 0,1 mM EDTA), 40 pmol ATP und 10 U T4-Polynukleotidkinase 30 min bei 37°C inkubiert. Die Inaktivierung erfolgte durch Zugabe von 1 µl 0,5 M EDTA und 10 min bei 70°C.

3.6.5.4. Dephosphorylierung von DNA-Fragmenten

Zur Vermeidung einer Religation von einfach hydrolysierten Vektoren, werden die 5'Enden der DNA dephosphoriert. 10 pmol DNA wurden mit 2,5 µl Reaktionspuffer (Endkonzentration 10mM Tris pH 7,5, 10 mM MgCl₂), 1 U CIAP vermischt und mit Wasser auf 25 µl aufgefüllt. Die Reaktion wurde 30 min bei 37°C durchgeführt und 15 min bei 85°C gestoppt. Bei Einsatz einiger Restriktionsendonukleasen war es auch möglich die Reaktion direkt nach der DNA-Hydrolyse (3.6.5.1) durchzuführen, indem lediglich 1 U CIAP zugegeben wurde.

3.6.5.5. Ligation von DNA-Fragmenten in Vektor DNA

Für Ligasereaktionen wurden ausschließlich Ge/eluierte oder Phenol-Chloroformgereinigte DNA eingesetzt. Die Ligation erfolgte mit T4-Ligase nach Angaben des Herstellers MBI Fermentas. Für einen Reaktionsansatz wurden verwendet:

Komponente	eingesetzt	Endkonzentration
Vektor-DNA in TE-Puffer	variabel	50-400 ng / 20 µl
DNA in TE-Puffer		
mit glatten Enden (<i>blunt ends</i>)	variabel	Molarität Vektor:Insert = 3:1
mit überhängenden Enden (<i>sticky ends</i>)	variabel	Molarität Vektor:Insert = 1:1
10 x Ligase-Buffer	2 µl	40 mM Tris pH 8,3, 10 mM DTT, 10 mM MgCl ₂ , 0,5mM ATP, pH 7,8
T4-Ligase		
Für <i>blunt ends</i>	4 µl	4 U
Für <i>sticky ends</i>	2 µl	2 U
50% PEG 4000 (nur für <i>blunt ends</i>)	2 µl	
H ₂ O	ad 20 µl	

Zur Berechnung der Molaritäten wurde folgende Formel benutzt: DNA-Enden in pmol = $[2 \times \text{DNA-Menge (in ng)} \times 10^3] / [\text{DNA-Länge (in bp)} \times 649 \text{ Dalton}]$.
Die Reaktion erfolgte für 1 h bei 22°C und wurde für 10 min bei 65°C gestoppt.

3.6.5.6. Klonierung von PCR-Fragmenten

Für die schnelle Klonierung von PCR-Fragmenten, die keine Restriktionsschnittstellen enthielten, wurde der *SureClone* \hat{a} *Ligation Kit* nach Angaben des Herstellers Amersham Pharmacia, Freiburg verwendet.

3.7. DNA-Sequenzierung

Die Sequenzierung von Plasmid-DNA wurde zu Beginn der Arbeit manuell nach der Sanger-Methode durchgeführt (Sambrook and Russell 2001). Die Sequenzreaktion wurde mit dem *T7-Sequencing Kit* nach Angaben des Herstellers Amersham Pharmacia, Freiburg durchgeführt. Für die Primer Anlagerung wurde dabei die *Standard Annealing* Methode verwendet. Als radioaktives Nukleotid wurde $[\alpha\text{-}^{32}\text{S}]\text{dATP}$ eingesetzt. Die Gelelektrophorese und Herstellung der Autoradiogramme wurde nach den Protokollen aus Sambrook und Russell (2001) durchgeführt. Als Standard wurde ein 6%iges Polyacrylamidgel mit 42% Harnstoff in der Sequenzgelkammer Model S2 (Life Technologies, Karlsruhe) bei 1800 V und 40 W drei bis acht Stunden gefahren. Das Autoradiogramm wurde auf einem hochauflösendem Film (HyperfilmTMMP, Amersham Pharmacia) aufgenommen und mit einer automatischen Entwicklermaschine (*HyperProcessor*, Amersham Pharmacia) entwickelt. Die Auswertung erfolgte wie in Sambrook und Russell (2001) beschrieben.

Mit dem Aufkommen kommerzieller Anbieter wurde die DNA-Sequenzierung durch die Firma SeqLab, Göttingen durchgeführt. Der Standard Sequenzierungs Service umfasste hier eine *Cycle Sequencing* Reaktion mit Standardprimern, Gelelektrophorese, sowie die Grundeditierung der Sequenzdaten und Elektrophogramme.

4. Allgemeine biochemische Methoden

4.1. Herstellung von Antikörpern

Für die immunhistochemischen und *immunoblotting* Methoden wurden verschiedene polyklonale Antikörper hergestellt. Das Anti-Astacin Antiserum wurde mit aus dem Flusskrebs isolierten aktiven Enzym Astacin wie in Vogt et al. (1989) beschrieben gewonnen. Die Antiseren Anti-ProAst-N, Anti-ProAst-C, Anti-CePepZinc_C07D10.6 und Anti-CePepZinc_K04E7.3 wurden durch Immunsierung von Kaninchen mit den entsprechenden synthetischen Peptiden (ProAst-N, ProAst-C, CePepZinc_C07D10.6 und CePepZinc_K04E7.3) gewonnen. Für eine bessere Immunantwort der Versuchstiere müssen die Peptide jedoch zunächst an ein Trägermolekül gekoppelt werden. Die Kopplung erfolgte carboxy- oder aminoterminal über einen aktivierten Cysteinrest an *Keyhole Limpet Hemocyanin* (KLH) der Firma Pierce. KLH gilt als schwach antigen und als gutes Adjuvans. Zu 200 μl wässriger KLH Lösung (10 mg/ml) wurden 300 μl DMSO und 2 mg MBS in 700 μl Wasser gegeben und 1 h bei Raumtemperatur unter leichtem Schütteln inkubiert. Gleichzeitig wurde eine Entsalzungssäule (D-Salt Dextran) mit 15 ml

Konjugationspuffer (0,083 M Natriumphosphat pH 7,2, 0,9 M NaCl, 0,1 M EDTA) equilibriert. Die Reaktionsmischung wurde auf die Säule gegeben und mit Konjugationspuffer in 0,5 ml Portionen eluiert. Die Fraktionen wurden spektrophotometrisch ($E_{280\text{nm}}$) vermessen, wobei aktiviertes KLH im ersten Absorptionsmaximum eluiert. Danach wurden 2 mg eines Peptids in 500 μl Konjugationspuffer gelöst, zu der KLH Lösung gegeben und 2 h bei Raumtemperatur unter leichtem Schütteln inkubiert. Der Reaktion wurde anschließend gegen PBS (10 mM Natriumphosphat pH 7,2, 140 mM NaCl) dialysiert und konnte direkt zur Immunisierung von Kaninchen eingesetzt werden. Die Immunisierung wurde von der Firma Pineda Antibody Service, Hamburg, durchgeführt. *Booster injections* mit 500 μg Protein wurden 14, 28 und 56 Tage nach der ersten Immunisierung durchgeführt. Nach den Immunisierungen wurde die Antikörperreaktion des entnommenen Serums jeweils in einem Immunoblot verfolgt. Das beste Ergebnis erhielt man 5 Tage nach der letzten Injektion. Das danach gewonnene Antiserum wurde in Portionen bei -80°C gelagert.

4.2. Protein-Isolierung

4.2.1. Gewinnung von Proteinextrakten aus *C. elegans*

Die gereinigten Nematoden aus einer Flüssigkultur wurden einmal mit Milli-Q-Wasser gewaschen und in ein 15 ml Zentrifugenröhrchen überführt. Zu 0,5 ml verdichteter Wurm suspension wurden 4 ml 50 mM Natriumphosphat, pH 7,5 gegeben. Diese Lösung wurde auf Eis direkt mit Ultraschall behandelt. Dazu wurde ein *Sonifier* (Cell Disrupter B 15, Branson) auf der Intensitätsstufe 3 bei 50% für dreimal 5 Minuten benutzt. Dazwischen wurde für jeweils 5 Minuten auf Eis stehen gelassen, um eine zu starke Erwärmung des Homogenats zu vermeiden. Zur Abtrennung der unlöslichen Bestandteile zentrifugierte man für 30 min bei 20000g. Der klare Überstand (Gesamtproteinextrakt) wurde in Eppendorfgefäße überführt und bis zur Verwendung bei -80°C gelagert.

4.2.2. Proteinextraktion aus Magensaft und Hepatopankreas von *Astacus astacus*

Magensaft wurde aus *Astacus astacus* wie in 10.1 beschrieben mit einer Pipette entnommen. Dieses Extrakt kann dann direkt für eine chromatographische Aufreinigung verwendet werden. Für die Immunoblot Experimente wurde der frisch entnommene Magensaft in SDS-Probenpuffer (50 mM Tris-HCl, pH 6.8; 100 mM Dithiothreitol, 2% SDS, 0.1% Bromphenol Blau, 10% Glycerol), der zusätzlich 50 mM EDTA und 50 mM EGTA als Metalloprotease-Inhibitoren enthielt, überführt und für 5 min auf 100°C erhitzt. Danach zentrifugierte man eventuell vorhandene, unlösliche Bestandteile für 5 min bei 13000 g ab. Die Lagerung erfolgte bei -20°C .

Zur Herstellung von Gesamtproteinextrakten aus dem Hepatopankreas für die Immunoblot-Experimente wurde das Gewebe in SDS-Probenpuffer mit 50 mM EDTA und 50 mM EGTA überführt. Die Proben wurden im Wasserbad auf 100°C erhitzt und mit einem Glashomogenisator 10 min behandelt. Danach zentrifugierte man für 10 min bei 13000g die unlöslichen Zellbestandteile ab und lagerte die Proben bis zum Start der Elektrophorese bei -20°C .

4.2.3. Bestimmung der Proteinkonzentration nach Bradford

Die Proteinbestimmung nach Bradford beruht auf der Beobachtung, dass der Farbstoff Coomassie Brilliant Blue G-250 in Abhängigkeit von der Proteinkonzentration in saurer Lösung einer Verschiebung des Absorptionsmaximum von 465 nm nach 595 nm unterliegt (Bradford 1976). Zu 100 µl Probe wurde 1 ml Färbelösung (0.01% CBB G-250, 4.7% Ethanol, 8.5% Phosphorsäure) gegeben. Nach 5 min Inkubationszeit wurde die OD bei 595 nm vermessen. Die Konzentration der gemessenen Probe musste zwischen 10 µg/ml und 200 µg/ml liegen (linearer Bereich), andernfalls mußte vor der Messung verdünnt werden. Als Standard wurde eine definierte Reihe von 10 µg/ml bis 200 µg/ml BSA in 50 mM Phosphatpuffer, pH 7,5 verwendet, wobei die Proteinkonzentration der Eichlösungen wie in 4.2.4 beschrieben bestimmt wurde ($\epsilon_{280\text{nm},\text{BSA}} = 66,7 \cdot 10^3 \text{ l}/(\text{mol}^1 \cdot \text{cm}^1)$). Die Ermittlung der Proteinkonzentration der unbekanntenen Probe erfolgte dann über lineare Regression mit einer erstellten Standardgeraden.

4.2.4. Bestimmung der Protein/Peptidkonzentration im Spektrophotometer bei 280 nm

Eine Möglichkeit der Konzentrationsbestimmung von Peptiden oder Proteinen, die in reiner Lösung vorliegen, ist die Messung der OD bei 280 nm. Die unbekanntene Proteinkonzentration kann dann nach dem Lambert-Beerschen Gesetz

$$E_{280\text{nm}} = \epsilon_{280\text{nm}} [\text{cm}^2 \mu\text{mol}^{-1}] \cdot c [\text{mol/l}] \cdot d [\text{cm}]$$

Voraussetzung für diese Meßmethode ist allerdings, dass der molare Extinktionskoeffizient (ϵ) bekannt ist und das untersuchte Peptid/Protein als Reinsubstanz vorliegt.

4.3. SDS- Polyacrylamid-Gelelektrophorese (SDS-PAGE)

Die eindimensionale SDS-Polyacrylamid-Gelelektrophorese wurde nach der Methode von Laemmli (1970) durchgeführt. Eine Übersicht der Methoden findet sich auch in Sambrook und Russell (2001).

4.3.1. Zusammensetzung der Gellösungen

Zusammensetzung des Trenngels

Acrylamidkonzentration	10 %	12,5 %	15 %	18 %
H ₂ O [ml]	5,4	3,8	2,1	0,1
Acrylamid (29/1) [ml]	6,7	8,3	10,0	12,0
1 M Tris pH 8,8 [ml]	7,5	7,5	7,5	7,5
10% SDS [ml]	0,2	0,2	0,2	0,2
10% APS [ml]	0,2	0,2	0,2	0,2
TEMED [ml]	0,02	0,02	0,02	0,02

Zusammensetzung des Sammelgels

Acrylamidkonzentration	3%
H ₂ O [ml]	7,5
Acrylamid (29/1) [ml]	1
1 M Tris pH 6,8 [ml]	1,25
10% SDS [ml]	0,1
10% APS [ml]	0,1
TEMED [ml]	0,01

4.3.2. Probenvorbereitung

Es wurden jeweils 10 µl Probe mit Proteinkonzentrationen von 10 mg/ml für komplexe Proteingemische und 0,5 mg/ml für gereinigte Proteine in SDS-Probenpuffer (50 mM Tris-HCl, pH 6,8; 100 mM Dithiothreitol, 2% SDS, 0,1% Bromphenol Blau, 10% Glycerol) 3 min auf 100°C erhitzt, kurz abzentrifugiert und auf ein Gel aufgetragen. Als Größenmarker wurde der Dalton Mark VII-L Proteinstandard (Albumin, Rind, 66 000; Albumin, Huhn, 45 000; Glycerinaldehyd-3-P-Dehydrogenase, 36 000; Carboanhydrase, Rind, 29 000; Trypsinogen, Rinderpankreas, 24 000; Trypsin Inhibitor, Soyabohne, 20 000; α-Lactalbumin, Kuhmilch, 14 200) nach Angaben des Herstellers Sigma eingesetzt.

4.3.3. Elektrophorese

Der Elektrophoreselauf wurde bei maximal 4 mA / cm durchgeführt und wurde gestoppt, sobald die Bromphenolblau-Laufmittelfront den unteren Rand des Trenngels erreicht hatte.

4.3.4. Coomassie Färbung

Nach Durchführung der Elektrophorese wurde das Polyacrylamidgel für 1 Stunde bei Raumtemperatur in Coomassie Färbelösung (0,06% Coomassie Brillant Blue G250 in 50% Methanol, 10% Eisessig) unter leichtem Schütteln inkubiert. Danach wurde das Gel durch mehrmaliges Wechseln der Entfärbelösung (7,5% Methanol, 5% Eisessig) bis zum Verschwinden des Hintergrunds differenziert.

4.4. Immunoblot

Der Immunoblot wurde modifiziert nach den Anleitungen in Sambrook und Russell (2001) durchgeführt.

4.4.1. Transfer auf PVDF-Membranen (*Western Blot*)

Für den Proteintransfer auf die PVDF-Membran (Porablot, Porenweite 0,2 µm, Machery-Nagel) wurde die *semi-dry* Methode nach Angaben des Herstellers durchgeführt. Die Membran wurde zunächst für 15 Minuten in 100% Methanol aktiviert. Membran und Gel wurden anschließend für 5 Minuten in Transferpuffer (25 mM Tris, 192 mM Glycin pH 8,3 und 20 % Methanol) äquilibriert. Dann wurde auf die untere Stahlelektrodenplatte einer *Semi Dry Blotter Unit* (Sigma) 6 Lagen mit Transferpuffer getränktes Filterpapier (Whatmann 3MM) gelegt. Darauf wurden die Membran, das Gel und weitere 6 Lagen getränktes Filterpapier aufgelegt und entstehende Luftblasen durch leichtes Rollen einer Pasteurpipette vorsichtig herausgedrückt. Der Aufbau wurde dann mit der zweiten platinbeschichteten Kathodenplatte verschlossen. Der Transfer fand bei Raumtemperatur und 1,5 mA / cm² für 90 Minuten statt. Die Membran kann bei -20°C gelagert werden.

4.4.2. Färbung der Membran mit Coomassie-Färbelösung

Die Membran wurde in 40% Methanol, 10% Essigsäure, 0,1% Coomassie Brilliant Blue G250 für 5 min unter leichtem Schütteln gefärbt. Anschließend wurde in 40% Methanol, 10% Essigsäure bis zum Verschwinden des Hintergrunds entfärbt.

4.4.3. Immunologischer Nachweis

Die Immunoblot-Analyse dient dem Nachweis von elektrophoretisch aufgetrennten Proteinen nach dem Transfer auf Membranen (*Western Blot*) mit spezifischen Antikörpern.

Die Membran wurde nach dem Transfer über Nacht bei 4°C und leichtem Schütteln in 1% *Blocking Powder*[®] (Schleicher und Schüll) in PBS (0,05 M Natriumphosphate, pH 7,4, 0,15 M NaCl) abgesättigt. Die Inkubation mit dem ersten Antikörper erfolgte bei Raumtemperatur für 2 Stunden in 0,5% *Blocking Powder*[®] in PBS. Die Antiseren wurden in einer Verdünnung von 1:200 eingesetzt. Überschüssiges Antiserum wurde durch dreimaliges Waschen für 10 min mit PBST (0,05 M Natriumphosphate, pH 7,4, 0,15 M NaCl, 0,1% Triton X100) entfernt. Die Membran wurde anschließend für 2 h mit dem zweiten Antikörper (1:15000, alkalische Phosphatase gekoppelter *goat anti rabbit IgG*, Sigma) in 0,5% *Blocking Powder*[®] in PBS inkubiert. Danach wurde dreimal 10 min in PBST und einmal 5 min in Wasser gewaschen. Anschließend wurde die Membran in Färbelösung (0,1 M Tris/HCl pH 9,5, 0,1 M NaCl, 5mM MgCl₂) überführt und pro 10 ml Lösung 7 µl NBT (50 mg/ml Nitrobluetetrazolium in 70% DMF) sowie 3,5 µl BCIP (20 mg/ml 5-Bromo-4-Chloro-3-indolylphosphat-Na) zugegeben. Die löslichen Farbsubstrate werden von der an den zweiten Antikörper gekoppelten Phosphatase zu einem unlöslichen dunkelvioletten Farbstoff umgesetzt und machen so die Proteinbanden indirekt sichtbar. Gefärbt wurde für höchstens 30 Minuten im Dunklen unter leichtem Schütteln. Die Reaktion wurde für zweimal 10 min in *aqua dest.* gestoppt. Die Dokumentation erfolgte mit Hilfe des BioDocAnalyse[®]-Gelscanners (Biometra).

5. Zweidimensionale (2D-) Gelelektrophorese

Mit Hilfe der zweidimensionalen Gelelektrophorese sollte versucht werden, durch spezifische Antikörper, Astacin-homologe Proteine in Protein-Extrakten von *C. elegans* zu identifizieren. Die Arbeiten wurden, wenn nicht anders vermerkt, im Labor von Prof. V. Pallini, Department für Molekulare Biologie, Universität Siena, Italien, durchgeführt.

5.1. 2D-Gelelektrophorese nach der IPG-SDS-PAGE Methode

Für die 2D-Gelelektrophorese wurde für die erste Dimension die IPG-SDS-PAGE Methode nach Görg et al. (1988), mit Änderungen durch Bjellqvist et al. (1993b) und für die zweite Dimension die Methode nach Laemmli (1970), modifiziert durch Hochstrasser et al. (1988a,b) verwendet. Als System wurde das MultiphorII[™] und Hoefer DALT[™] System des Lieferanten Amersham Pharmacia, Freiburg, benutzt.

5.1.1. Vorbereitung der Probe

Lyophilisierte *C. elegans* Gesamt-Proteinextrakte (siehe 4.2.1) wurden in 8 M Harnstoff, 4% (w/v) CHAPS, 40 mM Tris, 65 mM Dithioerythritol und 0,001% Bromphenolblau resuspendiert. Die Proteinkonzentration wurde nach der Methode von (Bradford 1976) bestimmt. Für einen Gellauf wurden 50 µg (analytisches Gel) oder bis 1000 µg (präparatives Gel) Proteinextrakt eingesetzt.

5.1.2. Isoelektrische Fokussierung

Immobiline DryStrip[™] Gele für die erste Dimension der Gelelektrophorese wurden vorgefertigt in Form von dehydrierten Streifen von Amersham Pharmacia, Freiburg bezogen und bei -20° C aufbewahrt. Es wurden Gel-Streifen mit einer Länge von 180 mm und einem pH Gradienten von 3 bis 10 verwendet.

5.1.2.1. Rehydrierung der Gel Streifen

Die Gel-Streifen müssen vor Gebrauch 10 Minuten bei Raumtemperatur aufgetaut und rehydriert werden. Die Rehydrierungslösung (8M Harnstoff, 2 % (w/v)CHAPS, 10 mM DTE, 0,8 % (w/v) Anfolin pH 3-10 und 0,001 % (w/v) Bromphenolblau) wurde mit der entsprechenden Probenmenge des Proteinextrakts vermischt. Die Gelstreifen verblieben in dieser Lösung 6-18 Stunden.

5.1.2.2. Laufbedingungen für die isoelektrische Fokussierung

Die isoelektrische Fokussierung wurde bei 10°C und unter Mineralöl durchgeführt.

Die Spannung wurde linear wie folgt erhöht:

300 V	30 min
600 V	30 min
1100 V	30 min
1700 V	30 min
2500 V	30 min
3500 V	30 min

Die letzte Phase wurde bei 5000 Volt durchgeführt. Für einen analytischen Lauf betrug die Laufzeit 20 Stunden (insgesamt 100 kWh); für einen präparativen Lauf wurde bis zu 4 Tagen fokussiert.

5.1.2.3. Äquilibration der Streifen

Bevor die Proteine in der zweiten Dimension getrennt werden können, müssen die Gelstreifen mit SDS, DTE und Jodacetamid beladen werden. Nach Beendigung der ersten Dimension, erfolgte die Äquilibration in 2 % (w/v) SDS, 6M Harnstoff, 30 % (v/v) Glycerol, 0,5 M Tris HCl pH 6,8 und 2 % (w/v) DTE für 12 Minuten und anschließend in 2 % (w/v) SDS, 6M Harnstoff, 30 % (v/v) Glycerol, 0,5 M Tris HCL pH 6,8 und 25 % (w/v) Jodacetamid für 4 Minuten.

5.1.3. Zweite Dimension: SDS-PAGE

Für die zweite Dimension wurden 9-16%ige lineare Polyacrylamid Gradientengele (18 cm x 20 cm x 1,5 mm) verwendet.

5.1.3.1. Herstellung der Gele

Lösung 1: 9 % Acrylamid (Endvolumen für 6 Gele 180 ml): 54 ml Acrylamid 30%-PDA 0,8%, 45,1 ml Tris-HCL 1,5 M pH 8,8, 79,27 ml Milli-Q-H₂O, 0,881 ml Na-Thiosulfat 5%, 0,0678 ml TEMED und 0,678 ml APS 10 %

Lösung 2: 16 % Acrylamid (Endvolumen für 6 Gele 180 ml): 99ml Acrylamid 30 %-PDA 0,8 % (w/v), 45,1 ml Tris-HCL 1,5 M pH 8,8, 34,27 ml H₂O Bidest, 0,881 ml Na-Thiosulfat 5 % (w/v), 0,0678 ml TEMED und 0,678 APS 10 % (w/v).

Zur Erzeugung des linearen Polyacrylamid Gradienten wurde ein Gradienten-Mischbehälter benutzt. Nachdem die vermischten Gellösungen zwischen den Glasplatten eingefüllt waren, wurde mit wassergesättigtem Isobutanol (1ml/Gel) überschichtet, wodurch eine Polymerisierung des Gels unter Luftabschluss ermöglicht wird. Nach Beendigung der Polymerisierung wurde *aqua dest.* aufgetragen (1ml), um das Austrocknen des Gels zu verhindern.

5.1.3.2. Beladen der zweiten Dimension mit den Gelstreifen

Die Gelstreifen wurden auf die Oberfläche der Gele der 2. Dimension gelegt und mit heißer Agarose Lösung (0,5% Agarose in Elektrodenpuffer) auf der Oberfläche der Gele fixiert.

5.1.3.3. Elektrophoreselauf

Die Gele wurden 5 Stunden bei 10 °C und einer konstanten Stromstärke von 40 mA/Gel gefahren. Als Elektrodenpuffer wurde 2,88 % Glycin, 0,6 % Tris, 0,1 % SDS verwendet.

5.1.4. Silberfärbung

Die analytischen Gele wurden routinemäßig nach der Methode von Oakley et al. (1980) und Hochstrasser et al. (1988b) mit Silbernitrat gefärbt.

Nach Beendigung des Elektrophoreselaufs wurden die Gele 5 Minuten in H₂O gewaschen. Alle folgenden Schritte mit Ausnahme der Entwicklung, wurden in Glasschalen (500 ml Volumen) unter Bewegung ausgeführt:

- Fixierung in Ethanol 40 % (v/v) und Essigsäure 10 % (v/v)
- Fixierung für mindestens 3 h bis 3 Tage in Ethanol 5 % (v/v)/ Essigsäure 5 % (v/v)
- 10 min in Milli-Q-H₂O bei 4°C waschen
- 30 min in Natriumacetat 0,5 M, Glutaraldehyd 2 %
- 3 x 10 min in Milli-Q-H₂O bei 4°C waschen
- 2 x 30 min in 0,05 % NDS Lösung
- 4 x 10 min in Milli-Q-H₂O bei 4°C waschen
- 30 min in Silbernitrat Lösung (9 g Silbernitrat, 15 ml Ammoniak 25 %, 2,25 ml NaOH 10 N auf 1125 ml mit Bidest aufgefüllt, reicht für 12 Gele)
- 4 x 4 min in Milli-Q-H₂O waschen
- Entwicklung für 12 min in 0,05 g/l Zitronensäure und 1ml/l Formaldehyd 37%
- Sobald die Proteine dunkel gefärbt waren wurde die Reaktion in 5 % (v/v) Essigsäure gestoppt und mehrmals mit Milli-Q-Wasser gewaschen.

5.1.5. Kolloidale Coomassie-Färbung

Die empfindliche kolloidale Coomassie-Färbung wurde nach der Methode von Neuhoff et al. (1988) mit dem *Blue Stain Kit* nach Angaben des Herstellers Novex durchgeführt. Die Fixierung erfolgte in 40% Methanol, 10% Essigsäure über Nacht. Die Färbung erfolgte in 500 ml Färbelösung (*Stainer A* und *B* in 20% Methanol). Nach 12 h konnte mit *aqua dest.* entfärbt werden, bis der Hintergrund klar erschien.

5.1.6. Auswertung der 2D-Gele

Gefärbte Gele wurden auf einem *Computing Densitometer 300S Scanner* des Herstellers Molecular Dynamics digitalisiert (4000 x 5000 *pixels*; 12 bits/*pixel*). Die Datenanalyse wurde mit Hilfe des Programms MELANIE II (*Medical Electrophoresis Analysis Interactive Expert System*; (Appel et al. 1997)) auf einer *Sun PARCstation* durchgeführt. Damit konnten die Proteinspots durch einen Suchalgorithmus detektiert und ein Gelbild mit allen quantitativ identifizierten Spots erstellt werden (*matching*). Für die Kalibrierung des pI und M_r wurden humane Serumproteine als interner Standard benutzt (Bjellqvist et al. 1993a). Durch Vergleich verschiedener Gelläufe ergibt sich so ein reproduzierbares Proteinmuster, das in einer Referenzkarte festgehalten wird (*reference map*). Dies erlaubt später auch die Zuordnung einzelner Proteinspots die beispielsweise durch eine Immunoreaktion detektiert wurden. Die automatische Spoterkennung stellt allerdings ein

großes Problem dar, da sowohl unterschiedliche Laufverhalten als auch große Intensitätsunterschiede einzelner Proteinspots die Interpretation erschweren können. In solchen Fällen wurde in den Untersuchungen zum Teil auch auf manuelle Zurordnung zurückgegriffen.

5.2. Westernblot von 2D-Gelen

5.2.1. Western-Transfer

Für den immunologischen Nachweis oder eine Microsequenzierung von einzelnen Proteinen wurden die Proteine direkt nach dem Elektrophoreselauf auf eine PVDF-Membran übertragen. Für 2D-Gele wurde die *semi-dry* Methode nach Matsudaira (1987) verwendet. Die Membran hatte eine Größe von 18 x 16 und eine Porenweite von 0,2 µm. Membran und Gel wurden für zunächst 15 Minuten in Transferpuffer (25 mM Tris, 192 mM Glycin und 20 % Methanol) äquilibriert; der Transfer fand bei 4° C und 100 V für 1 Stunde statt. Die Membran kann bei -20°C gelagert werden.

5.2.2. Ponceau S Färbung

Als Kontrollfärbung des Western-Transfers wurde die reversible Färbung mit Ponceau S verwendet. Hierzu wurde die Membran kurz in 0,2% Ponceau S in 3 % Essigsäure inkubiert und der überschüssige Farbstoff mit Wasser abgewaschen.

5.2.3. Coomassie Färbung der Membranen

Für die Mikrosequenzierung wurde die Membran in 40% Methanol, 10% Essigsäure, 0,1% Coomassie Brilliant Blue G250 für 5 min unter leichtem Schütteln gefärbt. Anschließend wurde in 40% Methanol, 10% Essigsäure bis zum Verschwinden des Hintergrunds entfärbt.

5.3. Immunoblot von 2D-Gelen

Die Membran wurde nach dem Transfer dreimal 10 Minuten in 3% Milchpulver in PBST (0,05 M Natriumphosphate, pH 7,4, 0,15 M NaCl, 0,1 % Triton) blockiert. Die Inkubation mit dem ersten Antikörper erfolgte über Nacht in 3% Milchpulver in PBST. Die Antiseren wurden in einer Verdünnung von 1:500 eingesetzt. Überschüssiges Antiserum wurde durch dreimaliges Waschen für 10 min mit 3% Milchpulver in PBST entfernt. Die Membran wurde anschließend für 2 h mit dem zweiten Antikörper (Meerrettich-Peroxidase gekoppelter *goat anti-rabbit IgG*, Sigma) in 3% Milchpulver in PBST inkubiert. Danach wurde dreimal 10 min in 3% Milchpulver in PBST, einmal 30 min in PBS (0,05 M Natriumphosphate, pH 7,4, 0,15 M NaCl) mit 0,5 % Triton und dreimal 30 min in 0,05 M Tris pH 6,8 gewaschen. Für die Chemolumineszenz-Reaktion wurde die ECL (*Enhanced Chemoluminescence*) Reaktionkomplettausstattung nach Angaben des Herstellers Amersham Pharmacia benutzt. Die Expositionszeit der Autoradiographiefilme (Hyperfilm-ECL, Amersham Pharmacia) betrug zwischen 1 und 5 min.

Diese identifizierten Proteinsignale der Filme wurden mit dem erstellten Referenzgel verglichen (siehe 5.1.6) und die Identifizierung der Proteine erfolgte anschließend über Mikrosequenzierung und massenspektrometrische Analyse der verschiedenen normalisierten Membranen oder Gelen.

5.4. Mikrosequenzierung

Für eine Identifizierung von Proteinspots mit Hilfe der aminoterminalen Sequenzierung wurde 1 mg Gesamtprotein benutzt und die Proteine nach dem Gellauf auf PVDF-Membran transferiert. Einzelne Proteinflecke wurden aus der Coomassie gefärbten PVDF-Membran ausgeschnitten und auf einem *Procise 494 Protein Sequence Apparatus* der Firma Applied Biosystems, Weiterstadt, analysiert. Die aminoterminalen Sequenzierung wurde von Dr. Hans Heid, Zentrale Proteinanalytik, Deutsches Krebsforschungszentrum, Heidelberg, durchgeführt.

Die Sequenzidentifizierung wurde mit Hilfe der *C. elegans* BLAST Programme des Sanger Centers (<http://www.sanger.ac.uk>) durchgeführt. Dabei wurde der Suchalgorithmus BLASTP zur Suche in der übersetzten Peptiddatenbank WORMPEP und TBLASTN für die genomische Datenbank verwendet.

5.5. Massenspektrometrische Analyse

Eine Silberfärbung der Proteingele ist für die Analyse der Proteinspots durch MALDI-TOF (*matrix-assisted laser desorption/ionisation time-of-flight*) Massenspektrometrie nicht geeignet. Deshalb wurden die Gele hierfür mit der kolloidalen Coomassie-Färbelösung entwickelt. Es wurde 1 mg Gesamtprotein für einen Gellauf eingesetzt. Die massenspektrometrische Analyse wurde im Labor von Dr. M. Schnölzer am Deutschen Krebsforschungszentrum, Zentrale Proteinanalytik, Heidelberg durchgeführt. Die Durchführung erfolgte wie in Jensen et al. (1997) und Möhrlein et al. (2001) beschrieben.

5.5.1. Probenvorbereitung

Die Proteinflecken wurden aus dem Gel ausgeschnitten und je zweimal 15 min mit 100 µl Wasser und 50 % Acetonitril gewaschen. Anschließend wurde für zweimal 15 min in 100% Acetonitril dehydriert und an der Luft getrocknet. Pro Gelstück wurden 20 µl 40 mM Ammoniumbicarbonat und 0,8 µl Schweine-Trypsin (0,5 µg/µl in 1 mM HCl, *sequencing grade*, Promega) zugegeben und 15 min bei Raumtemperatur inkubiert. Hierbei quoll das Gelstück wieder auf die ursprüngliche Größe. Anschließend wurde kontrolliert, ob das Gelstück vollständig mit Puffer bedeckt war und über Nacht bei 37°C inkubiert.

5.5.2. Herstellung der Matrix

Für die Herstellung der Matrix wurde von α -Cyano-4-Hydroxy-Zimtsäure und Nitrozellulose eine gesättigte Lösung in Aceton hergestellt. Außerdem wurde eine Lösung von 10 mg/ml Nitrocellulose in Aceton/Isopropanol (1:1) hergestellt. Als Gebrauchslösung wurden 40 µl der gesättigten Lösung mit 10 µl der Nitrozellulose Lösung gemischt und 0,3 µl dieses Gemischs auf die einzelnen Positionen des Probenträgers (Metalltarget) aufgetragen und antrocknen lassen. Danach wurden jeweils 0,8 µl 10% Ameisensäure und 0,5 µl der zu analysierenden Probe aufgetragen. Nach kurzem Trocknen wurde überschüssiges Salz durch Waschen mit einem Tropfen 10%iger Ameisensäure und einem Tropfen Wasser entfernt. Der getrocknete Probenträger wurde in das Massenspektrometer eingebracht und vermessen.

5.5.3. Aufnahme der Massenspektren

Die MALDI-TOF Massenspektren wurden in einem Reflex II Flugzeitmassenspektrometer der Firma Bruker Daltonik (Bremen) aufgenommen, der mit einem SCOUT *multiprobe*

inlet und einem 337 nm Stickstofflaser ausgestattet ist. Die Beschleunigungsspannung betrug 20 kV, die Reflektorspannung 21,5 kV und die Spannung an der ersten Extraktionsplatte 15,4 kV. Die Massenspektren wurden im Positiv-Ionen-Modus mit verzögerter Extraktion aufgenommen und durch 10-100 einzelne Laserschüsse erhalten. Die Kalibrierung des Massenspektrometers erfolgte als lineare Zweipunktmessung mit zwei externen Standardproteinen.

5.5.4. Auswertung der Spektren

Die Identifizierung der Proteine erfolgte mit Hilfe der Programme ProFound (<http://129.85.19.192/prowl/cgi/Profound.exe>), ProteinProspector (<http://prospector.ucsf.edu>) und MatrixScience (<http://www.matrixscience.com>) anhand der tryptischen Peptidmassen. Die Ergebnisse wurden mit der nicht redundanten, genomischen *wormbase* Datenbank und der *wormpep* Datenbank (beide <http://www.wormbase.org>) für *C. elegans* Proteine verglichen.

6. Chromatographische Methoden

Astacin und Trypsin wurden aus dem Magensaft von *Astacus astacus* nach den Methoden von Zwilling und Neurath (1981) präpariert. Astacin wurde darüber hinaus nach Reyda et al. (1999) affinitätschromatographisch aufgereinigt.

6.1. Gewinnung des Krepstrypsins durch Affinitätschromatographie

Mit Bromcyan-aktivierter Sepharose 4 B (Amersham-Pharmacia) können alle Arten von Proteinen und andere Liganden über primäre Aminogruppen kovalent an den Träger gekoppelt werden. 1 Gramm Sepharose-Gel wurden in 50 ml 1 mM HCl 30 min vorgequollen, auf einem G3-Glasfilter 15 min mit 200 ml 1 mM HCl gewaschen und anschließend mit dem gleichen Volumen Kopplungs-Puffer (0,1 M NaHCO₃, 0,5 M NaCl, pH 8.3) nachgewaschen. 15 mg SBTI (Sojabohnen-Trypsin-Inhibitor, Sigma) wurden in 5 ml Kupplungspuffer gelöst und mit dem Gel unter leichtem Schütteln 4 h bei Raumtemperatur inkubiert. Anschließend wurde ungebundenes Material mit 50 ml Kupplungspuffer entfernt. Zur Beseitigung etwa verbliebener aktiver Gruppen wurde das Sepharose-Gel 2 h mit 30 ml einer 1 M Äthanolaminlösung bei pH 8.0 behandelt und in abwechselnd je 3 Zyklen 30 ml Acetat-Puffer (1 M NaCl, 0,1 M Natriumacetat, pH 4.0) und 30 ml Borat-Puffer (1 M NaCl, 0,1 M Natriumborat pH 8.0) gewaschen. Das Säulenmaterial wurde in Tris-Puffer (50 mM Tris-HCl pH 8,0, 10 mM CaCl₂) gelöst und in eine 10 ml Säule (Amersham-Pharmacia) gefüllt. 10 ml Magensaft wurde mit Tris-Puffer 1: 50 verdünnt und auf die Säule aufgetragen. Danach wurde die Säule bei einer Durchflussgeschwindigkeit von 1 ml/min mit 100 ml Tris-Puffer bis zum Erreichen der Grundlinie im Photometer bei 280 nm equilibriert. Die Elution erfolgte mit Natriumacetat pH 3,0 in 1 ml Fraktionen. Zur Neutralisation wurden pro Fraktionsgefäß 100 µl 1 M Tris-HCl vorgelegt. Die Trypsinfraktionen wurden durch Aktivitätsmessung (BTNA-Substrat) bestimmt und auf einem SDS-PAGE überprüft. Die vereinigten Fraktionen wurden über Nacht gegen 10 mM Tris-HCl dialysiert, anschließend lyophilisiert und bei -20°C gelagert. Als Ausbeute waren etwa 10 mg Trypsin zu erwarten.

6.2. Gewinnung des Astacins durch Ionenaustausch-Chromatographie und Gelfiltration

Eine DEAE-Sephacel[®] Säule (Amersham-Pharmacia) wurde mit Trispuffer (50 mM Tris-HCl pH 8,0) equilibriert. 10 ml Magensaft wurden aufgetragen und bei einer Durchflussgeschwindigkeit von 1 ml/min 60 min bis zum Erreichen der Grundlinie bei 280 nm mit Trispuffer nachgespült. Die Elution erfolgte mit einem linearen Salzgradienten (0,05 bis 0,5 M NaCl in 50 mM Tris-HCl pH 8,0) in 1 ml Fraktionen. Die Astacin Fraktionen wurde durch Aktivitätsmessung (STANA-Substrat) bestimmt, die vereinten Fraktionen mit 70% Ammoniumsulfat gefällt, in einem kleinen Volumen Tris-Puffer aufgenommen und über Nacht gegen 50 mM Tris-HCl dialysiert. Die Lösung wurde auf eine Sephacel S 3000 Gelfiltrationssäule aufgetragen und mit Tris-Puffer wieder eluiert. Durch einen für Astacin charakteristischen Retentionseffekt wird dieses stark verzögert bei einem scheinbaren Molekulargewicht von < 10000 Da eluiert. Für weitere Experimente wurden die Fraktionen wieder durch Aktivitätsmessungen bestimmt, über Nacht dialysiert, lyophilisiert und bei -20°C gelagert. Aus 10 ml Magensaft konnten 5-6 mg lyophilisiertes reines Astacin erhalten werden.

6.3. Reinigung von Astacin und Astacin-homologen Proteinen durch Affinitätschromatographie

Der Astacin spezifische Inhibitor Pro-Leu-Gly-Hydroxamat (Bachem) wurde an aktivierte CH-Sepharose 4B nach Angaben des Herstellers Amersham-Pharmacia und wie in 6.1. beschrieben gekoppelt. 5 mg lyophilisiertes Flusskrebs-Astacin wurden in 50 ml Tris-Puffer gelöst (50 mM Tris-HCl pH 8,0) und auf die Säule aufgetragen. Danach wurde bei einer Durchflussgeschwindigkeit von 1 ml/min mit mindestens 50 ml 0,5 M NaCl, 50 mM Tris-HCl pH 8,0 bis zum Erreichen der Grundlinie im Photometer bei 280 nm equilibriert. Die Elution erfolgte mit Tris pH 11,0 in 1 ml Fraktionen. Zur Neutralisation wurden pro Fraktionsgefäß 100 µl 1 M Tris-HCl vorgelegt. Die Astacin-Fraktionen wurden durch Aktivitätsmessungen (STANA-Substrat) bestimmt und auf einem SDS-PAGE überprüft. Die vereinigten Astacin-Fraktionen wurden über Nacht gegen 10 mM Tris-HCl dialysiert, anschließend lyophilisiert und bei -20°C gelagert.

Darüber hinaus wurde versucht, aus 10 ml Gesamtproteinextrakt von *C. elegans* Astacin-homologe Proteine in einem Schritt zu isolieren. Der Rohextrakt wurde mit Tris-Puffer 1:50 verdünnt und auf die Säule aufgetragen. Die Elution erfolgte wie oben beschrieben. Die Fraktionen wurden durch Aktivitätsmessungen (dansyliertes Oligopeptidsubstrat) und SDS-PAGE analysiert.

7. Aktivitätsmessung proteolytischer Enzyme

Für die Bestimmung der spezifischen Aktivität proteolytischer Enzyme wurden verschiedene Messmethoden eingesetzt.

Nitroanelidsubstrate setzen nach einer spezifischen proteolytischen Spaltung gelbes Nitroanelin frei und die Reaktion kann bei 405 nm photometrisch verfolgt werden. Nach dem Lambert-Beerschen Gesetz kann aus der erhaltenen Extinktionsänderung der Substratumsatz pro Minute errechnet werden ($\Delta E_{405} / \text{min} = \epsilon \cdot c \cdot d$). Die Formel wird nach c (Substratkonzentration in mol/l) aufgelöst. Der Extinktionskoeffizient ϵ für Nitroanelin beträgt $\epsilon_{405\text{nm}} = 10200 \text{ M}^{-1} \text{ cm}^{-1}$. Daraus lässt sich die Aktivität eines Enzyms in

Units [U] berechnen: $1 \text{ U} = 1 \mu\text{mol Substratumsatz (Konzentrationsänderung) / min}$. Die Spezifische Aktivität [U / mg] bezieht dann die Aktivität auf ein mg Enzym.

Dansylierte Oligopeptide wie Dansyl-Pro-Lys-Phe-Ala-Pro-Trp-Val (Bachem) werden spezifisch zwischen den zentralen Aminosäureresten Phe/Ala gespalten. Dies führt zu einer 10 fach erhöhten Tryptophan-Fluoreszenz bei 340 nm bei einer Anregungswellenlänge von 280 nm. Im intakten Substrat ist die messbare Tryptophan-Fluoreszenz durch die Dansyl-Gruppe unterdrückt (*quenching*). Die Berechnung der Substratkonzentrationen und spezifischen Aktivität erfolgt analog zu den Nitroanilidsubstraten ($\epsilon_{340\text{nm}}$ der Dansyl-Gruppe = $4300 \text{ M}^{-1}\text{cm}^{-1}$)

7.1. Aktivitätsmessung von Trypsin mit BAPNA

Benzoyl-L-Arginin-p-Nitroanilid (BAPNA) ist ein Nitroanilid-Substrat, welches spezifisch von Trypsin gespalten wird. Für einen Reaktionsansatz wurden in eine 1 cm-Quarzküvette 750 μl BAPNA-Substratlösung (1 mM BAPNA, 1,75% DMF, 50 mM Tris-HCl pH 8,0, 10 mM CaCl_2) mit 0,1-20 μM Enzymlösung vermischt und in einem Spektralphotometer (Lambda S2, Perkin Elmer) bei 405 nm und 25°C die Extinktionszunahme pro Zeit 10 min lang aufgezeichnet.

7.2. Aktivitätsmessung von Astacin mit STANA

Succinyl-Tri-Analin-4-Nitroanilid (STANA) ist ein spezifisches Nitroanilid-Substrat für Astacin (Stöcker und Zwilling 1995). Für einen Reaktionsansatz wurden in eine 1 cm-Quarzküvette 750 μl STANA-Substratlösung (1 mM STANA, 50 mM Tris-HCl pH 8,0) mit 0,1-20 μM Enzymlösung vermischt und in einem computergestützten Spektralphotometer (Lambda S2, Perkin Elmer) bei 405 nm und 25°C die Extinktionszunahme pro Zeit 10 min aufgezeichnet.

7.3. Aktivitätsmessungen mit dansylierten Substraten

Die spezifische Messung der Astacin Aktivität mit Dns-PKFAPWV (dansyl-Pro-Lys-Phe-Ala-Pro-Trp-Val) Substrat erfolgte wie in Stöcker und Zwilling (1995) und Wagner et al. (2001) beschrieben. Die Reaktion wurde bei 25°C in einer 300 μl Quarzküvette in einem Fluoreszenz-Photometer (LS 50, Perkin Elmer) durchgeführt. 280 μl Substratlösung (0,1 mM Dns-Substrat in 50 mM Tris-HCl pH 8,0) wurden mit 20 μl Enzymlösung (0,1-4 nM) vermischt und einer Anregungswellenlänge von 280 nm (Spaltbreite 2,5 mm) bei 340 nm (Spaltbreite 5,0 mm) für 10 min gemessen. Als Enzymlösung wurden sowohl aufgereinigtes Astacin als auch verdünnte Protein-Rohextrakte von *C. elegans* eingesetzt.

7.4. Aktivitätsmessungen mit FITC-Casein

Bei Anregung mit Licht der Wellenlänge von 490 nm emittiert Fluoreszeinisothiocyanat (FITC) Licht mit einer Wellenlänge von 515 nm. Dieser Farbstoff ist an Casein gekoppelt, das als Substrat für die Hydrolyse durch Proteasen dient. Nach der Spaltung können die nicht-gespaltenen Proteine mit TCA ausgefällt werden; Spaltprodukte verbleiben in Lösung. Die Fluoreszenzintensität des Überstands kann im Fluorimeter gemessen werden und ist ein direktes Maß für die proteolytische Aktivität der Probenlösung.

In einem Eppendorfgefäß wurden 100 μl Na-Acetat Puffer (0,25 M, pH 8) mit 1 μl Probenlösung, 20 μl Dithiothreitol-Lösung (8 mM), und 30 μl FITC-Casein Lösung (0,5 mg/ml) gemischt. Dieser Ansatz wurde für 2 Stunden bei 25°C im Wasserbad inkubiert.

Die Proteolyse wurde durch Hinzufügen des gleichen Volumens 10%iger Trichloressigsäure gestoppt und die Proteine ausgefällt. Nach zehnmütigem Abzentrifugieren bei 14000 rpm wurden 0,2 ml des Überstands zu 1,8 ml 0,2 M Natriumhydroxidlösung gegeben. Die Messung erfolgte in einer 3 ml Quarzküvette mit dem Luminiszenz-Fluoreszenzspektrometer LS 50 des Herstellers Perkin Elmer. Die Anregungswellenlänge betrug 490 nm, bei einer Spaltbreite von 3,0. Die Fluoreszenzmessung erfolgte bei 515 nm, bei einer Spaltbreite von 5,0.

8. GFP-Fusionsproteine und Expressionsnachweis in *C. elegans*

Die Herstellung und Expressionsanalyse von Fusionsproteinen in *C. elegans* wurde nach den Standard-Methoden in Hope (1999) durchgeführt. Sie werden im Folgenden zusammenfassend erläutert.

8.1. Expressionsvektoren

Als Expressionsvektoren wurden die Plasmide pPD95.85 und pPD114.35 aus dem *Fire Lab Vector Kit* (Dr. A. Fire, Carnegie Institution of Washington, Baltimore) verwendet. Diese pUC19 Derivate enthalten, neben einer Klonierungsstelle (*multiple cloning site*) und einem Ampicillinresistenzgen, GFP (*green fluorescent protein*) als Reporter gen. Über eine *long range* PCR (siehe 3.4.2.3) wurden die Astacin-homologen Gene C05D11.6, C07D10.4, K04E7.3 und T04G9.2 von genomischer DNA mit genspezifischen Oligonukleotiden amplifiziert. Dabei wurde der genomische DNA-Abschnitt von ungefähr 5000 bp aufwärts des Startkodons (Promoterregion) bis kurz vor Ende des letzten Exons vervielfältigt. Die Oligonukleotide enthielten jeweils verschiedene Restriktionsschnittstellen, so dass nach einem Restriktionsverdau und der Gelelektion diese direkt in die Expressionsvektoren kloniert werden konnten.

8.2. Herstellung der Injektionslösung

Für die Herstellung der Injektionslösung wurden mit Hilfe des NucloSpin® Plasmid Mini Kits aufgereinigtes Plasmid mit Wasser auf 50 ng/µl verdünnt. Der Transformationsmarker (pBX) wurde ebenfalls auf 50 ng/µl verdünnt und 1:1 mit dem Expressionsvektor vermischt.

8.3. Herstellung der Injektionskanülen

Zur Herstellung der Injektionskanülen benutzte man Borsilikat-Glaskapillaren von 1mm Durchmesser, deren feine Spitze auf einem Mikropipetten-Auszieher (*needle puller*) hergestellt wurde. Die Beladung der Kanülen mit 2 µl Injektionslösung geschah durch Kapillarkräfte. Bevor mit einer Injektion begonnen werden konnte, wurde die Spitze an kleinen Salzkristallen auf einem Objektträger unter dem Mikroskop aufgebrochen.

8.4. Injektion

Die Würmer wurden für die Injektion auf getrockneten, Agarose-beschichteten Objektträgern immobilisiert. Zu deren Herstellung wurde 5% Agarose in Wasser auf 80°C

erhitzt, ein Tropfen auf einen Objektträger gegeben, sofort mit einem weiteren Objektträger abgedeckt, der obere Objektträger wieder entfernt und die Agarschicht eintrocknen lassen. Mit einer feinen Präpariernadel aus Platindraht wurden 5-10 Tiere auf den Objektträger gesetzt und zur Vermeidung des Austrocknens mit Mineralöl überschichtet. Die Injektion von ca. 100 nl Injektionslösung erfolgte mit Hilfe eines Mikromanipulators und einem inversen Lichtmikroskop (Zeiss) direkt in die syncytiale Gonade der Tiere. Die behandelten Würmer wurden mit der Präpariernadel so schnell als möglich einzeln zurück auf Agarplatten gesetzt.

8.5. Identifizierung von Transformanten

Die transgene F1 Generation der behandelten Tiere wurden durch Selektion des mutanten Phenotyps pha-1 bei 25°C identifiziert (vergleiche 2.2.3). Bei dieser Temperatur konnten von der pha-1 Linie nur Tiere überleben, die den Coinjektionsmarker pBx (*rescue-plasmid*) als extrachromosomales DNA-arrays stabil integriert hatten. Von jeder Agaroseplatte wurde jeweils nur ein überlebendes Tier der F1 Generation weiter als eigenständige Linie kultiviert. Die fluoreszenzmikroskopische Analyse wurde ab der F2 Generation begonnen.

8.6. Dokumentation

Zur Mikroskopie wurden die Tiere von einer 3 bis 6 Tage alten Agaroseplatte mit 1ml Levamisole (50 mM) abespült, in einem Eppendorfgefäß sedimentiert und 10 µl dieses Pellets auf einen Agarose-beschichteten Objektträger vorsichtig aufgebracht. Die Dokumentation erfolgte auf einem Fluoreszenzmikroskop der Firma Zeiss mit angeschlossener Digitalkamera. Für die Fluoreszenzaufnahmen wurde das vom Hersteller mitgelieferte GFP- oder FITC-Filterset benutzt.

9. Expressionsnachweis durch *in situ* Hybridisierung in *Hydractinia echinata*

Die *in situ* Hybridisierung von *Hydractinia echinata* wurde nach der Methode von Gajewski et al. (1996) und Plickert et al. (1997) durchgeführt. Die automatisierte *in situ* Hybridisierung in dem Apparat der Firma Abimed wurde im Labor von Prof. G. Plickert, Zoologisches Institut, Universität Köln, durchgeführt.

9.1. Vorbehandlung der Tiere

Für die *in situ* Experimente sollten Larven, verschiedene Metamorphosestadien und adulte Polypen untersucht werden. Larven wurden aus dem Seewasser genommen und direkt fixiert. Adulte Polypen wurden vorsichtig von einer Kolonie abgeschnitten und sofort, sowie nach 4 h Regenerationszeit, fixiert. Darüberhinaus wurden adulte Polypen im Bereich der Körperachse verletzt und wiederum sofort und nach 4 h fixiert. Zur Untersuchung der Metamorphosestadien wurden ein bis drei Wochen alte Larven für 3 h in 112 mM CsCl in Seewasser inkubiert und anschließend dreimal mit Seewasser ausgewaschen. Die sich entwickelnden Tiere wurden sofort, sowie 4 h, 12 h, 24 h und 48 h nach Metamorphoseauslösung fixiert.

9.2. Herstellung der *in situ* Sonden

Die *in situ* Sonden wurden durch PCR-gestützte *in vitro* Transkription hergestellt. Dazu wurden die in bluescript oder pGMT Vektoren klonierten DNA-Fragmente über PCR mit T3/T7 oder SP6/T7 Oligonukleotiden amplifiziert. Nach der Kontrolle und Isolierung des PCR-Ansatzes aus einem Agarosegel wurden 100 ng des Produkts zur *in vitro* Transkription eingesetzt. Ein 10 µl Ansatz enthielt 0,5 µl Digoxigenin-RNA *Labeling* Mix, 1 µl Transkriptionspuffer, 10 U T3, T7 oder SP6 Polymerase, 5 U RNase Inhibitor (alles Boehringer Mannheim), 100 ng PCR Template und DEPC-behandeltes Wasser. Die Transkription wurde über Nacht bei 37°C durchgeführt. 1 µl dieses Ansatzes wurde auf einem Agarosegel überprüft. Der restliche Ansatz wurde mit 1 µl 200 mM EDTA pH 8,0 abgestoppt. Die Präzipitation erfolgte mit 1,25 µl LiCl und 75 µl 100% Ethanol. Das Pellet wurde in 12,5 µl DEPC-H₂O und 5 U RNase Inhibitor bei 37°C für 30 min gelöst.

9.3. *In situ* Hybridisierung

10 bis 50 Tiere wurden 1 h bei Raumtemperatur in 1 ml 4% Paraformaldehyd in 0,1 M HEPES pH 7,5, 0,42 M NaCl, 2 mM MgSO₄ fixiert. Anschließend wurde dreimal 10 min mit 1 ml PBST (10 mM Na₂HPO₄, 150 mM NaCl, pH 7,5, 0,1% Tween 20) gewaschen und für 10 min auf 90°C erhitzt. Die weitere Prozedur erfolgte dann wie in Plickert et al. (1997) beschrieben in einem automatisierten *in situ* Apparat (Abimed). Zusammengefasst wurde zunächst Proteinase K verdaut (20 µg proteinase K in PBST), gewaschen (2 mg/ml Glycin in PBST und in PBST), mit Paraformaldehyd nachfixiert, erneut gewaschen (PBST), für 2 h prähybridisiert (0,5 mg/ml Heringsperma DNA, 50% Formamid, 5x SSC [0,75 M NaCl, 75 mM Natriumcitrat pH 7,0], 0,1 mg/ml Heparin, 0,1% Tween 20) über Nacht mit 4-20 ng Digoxigenin markierter RNA bei 50°C hybridisiert und überschüssige Sonde durch mehrmaliges Waschen (50% Formamid, 2x SSC, 0,1% Tween 20; 2x SSC, 0,1% Tween 20; PBST) entfernt. Die Antikörperreaktion erfolgte für 2 h mit Anti-Digoxigenin Alkalische Phosphatase gekoppeltem IgG (Boehringer) und wurde anschließend in AP-Puffer (0,1 M Tris-HCl pH 9,5, 0,1 M NaCl, 50 mM MgCl₂, 0,1% Tween 20) gestoppt. Für die Substratreaktion wurden die Tiere aus dem Automaten in 24 Lochplatten überführt, mit den Farbsubstraten NBT/BCIP in AP-Puffer versetzt und die Reaktion unter dem Mikroskop kontrolliert. Nach 15 min bis 4 h wurde die Reaktion in 10 mM EDTA in PBST gestoppt und die Tiere in 90% Glycerol in 10 mM EDTA in PBST auf Objektträger überführt. Die Präparate können so bei -20°C gelagert werden. Die Dokumentation erfolgte an einem Phasenkontrastmikroskop der Firma Zeiss mit angeschlossener Digitalkamera (Nikon, DN100).

10. Präparation und Immunhistochemie von *Astacus astacus*

10.1. Magensaftentnahme und Präparation des Hepatopankreas

Zur Gewinnung von Magensaft für die Immunoblot-Experimente und zur Isolierung von Astacin und Trypsin wurde Tieren, die eine Woche gehungert hatten, der Magensaft entnommen. Dazu wurde eine Plastikpipette vorsichtig zwischen den Mandibeln in den Ösophagus eingeführt (Pfleiderer et al., 1967). Der Magensaft floss in die Pipette, wurde in Eppendorfgefäßen gesammelt und bis zur Verwendung bei -80°C eingefroren.

Die Entnahme von Magensaft führt gleichzeitig zu einer Stimulierung der Enzym-Neusynthese im Hepatopankreas (Vogt et al. 1989). Deshalb wurden die so behandelten Tieren nach 3 oder 7 Stunden für die Präparation des Hepatopankreasgewebes durch Abtrennung des Carapaxes getötet. Die bräunlichen Hepatopankreashälften wurden mit einem Skalpell herauspräpariert. Als Kontrolle wurde zusätzlich der Hepatopankreas von einigen unbehandelten Tieren präpariert.

Die Fixierung für immunhistologische Experimente erfolgte über Nacht bei 4°C mit Bouin-Lösung (60 ml Pikrinsäure, 20 ml 35 %Formol) oder Susa-Lösung (4,5 g HgCl₂-Sublimat, 0,5 g NaCl in 70 ml Aqua dest werden mit 10 ml 2% Trichloressigsäure, 20 ml 35 % Formol und 4 ml Eisessig vermischt).

10.2. Herstellung von Paraffin-Schnitten

10.2.1. Einbettung

Die fixierten Gewebestücke wurden mehrmals in Wasser gespült und anschließend je eine Stunde in 70, 80, 90 und 96%igem Ethanol entwässert. Danach wurden sie zweimal 30 Minuten in Isopropanol und über Nacht in Methylbenzoat inkubiert. Die Methylbenzoatlösung wurde am nächsten Tag mehrmals gewechselt und erneut über Nacht in Methylbenzoat belassen. Das Gewebe wurde in Paraffin (Paraplast, Sherwood Medical) 3 Tage bei 60°C inkubiert wobei das Paraffin nach 2 Tagen einmal gewechselt wurde. Erst dann erfolgte die endgültige Aushärtung bei Raumtemperatur in vorgeformten Blöcken.

10.2.2. Herstellung der Schnitte

Von den fertigen Paraffinblöckchen wurden 6 µm dicke Schnitte hergestellt (Schneidegerät: Jung, Heidelberg). Die Schnitte wurden auf einem 37°C-Wasserbad gestreckt und auf beschichtete Glasobjektträger (HistoBond Superior, Marienfeld) aufgenommen.

10.2.3. Entparaffinierung

Für die anschließenden Färbungen und immunhistologischen Experimente müssen die Schnitte wieder entparaffiniert und in wässriges Milieu überführt werden. Dazu wurden diese für dreimal 10 Minuten in Xylol inkubiert und über eine absteigende Alkoholreihe von zweimal 5 min Isopropanol, je viermal 3 min 96%, 90%, 80%, 70% Ethanol in Wasser überführt und abschließend mit *aqua dest.* gespült.

10.3. Übersichtsfärbungen

Um einen Überblick über den Gewebsaufbau des Hepatopankreas zu erhalten wurden die histologischen Schnitte gefärbt.

Verwendete Färbelösungen:

- Eisenhämatoxylin nach Weigert:

Lösung A: 1 g Hämatoxylin in 100 ml 96% Ethanol

Lösung B: 1,16 g FeCl₃ wurden in 98 ml Aqua dest. gelöst und mit 1 ml 25% HCl versetzt; die Lösungen wurden vor Gebrauch 1:1 gemischt.

- Säurefuchsin-Ponceau-Azophloxin:

Lösung A: 0,02 g Ponceau de Xylidine, 0,01 g Säurefuchsin und 0,06 ml Eisessig in 30 ml *aqua dest.*

Lösung B: 0,05 g Azophloxin und 0,002 ml Eisessig in 10 ml *aqua dest.*

Gebrauchslösung: 16 ml Lösung A und 4 ml Lösung B in 176 ml 0,2% Essigsäure

- Phosphormolybdänsäure-Orange G:
8 g Phosphormolybdänsäure und 4 g Orange G in 200 ml *aqua dest.*
- Lichtgrün: 0,4 g Lichtgrün und 0,4ml Eisessig in 200 ml *aqua dest.*

Die erste Färbung erfolgt mit Eisenhämatoxylin nach Weigert für 1 min. Hierbei werden, durch die Anlagerung des basischen Farbstoffes Hämatoxylin an die sauren Phosphatgruppen der DNA, die Zellkerne angefärbt. Durch anschließendes Spülen in Leitungswasser (10 min) kommt es infolge einer pH-Verschiebung zur Blaufärbung der Zellkerne. Danach wurde 5 Minuten mit Säurefuchsin-Ponceau-Azophloxin gefärbt, wenige Sekunden in 1% Essigsäure gespült und dann in Phosphormolybdänsäure-Orange G für weitere 5 Minuten gefärbt. Nach einer kurzen Spülung in 1% Essigsäure wurde 5 Minuten mit Lichtgrün gefärbt und wiederum für 5 Minuten in 1% Essigsäure gespült. Die letztgenannten Farbstoffe sind Dispersionsfarbstoffe, die durch ihre unterschiedliche Diffusionsgeschwindigkeit die verschiedenen Gewebeteile unterschiedlich anfärben. Feindispersive Komponenten wie Ponceau, Azophloxin und Orange G diffundieren im Gewebe schneller als grobdispersive wie Säurefuchsin und Lichtgrün.

Über 100% Isopropanol und dreimal 10 Minuten Xylol wurden die Schnitte in DePeX (Serva) eingebettet und mit einem Leitz Aristoplan Mikroskop ausgewertet (vergleiche 10.4.3).

10.4. Immunhistochemie

Die Methode der Immunhistochemie ermöglicht es, direkt in Gewebeschnitten mit Hilfe von spezifischen Antikörpern ein gesuchtes Protein zu lokalisieren.

Als Gegenprobe für die Anti-Astacin und Anti-Proastacin Antikörper wurde ein Anti-Hämocyanin Serum verwendet. Dieses Serum wurde von Prof. W. Stöcker, Institut für Zoophysologie, Universität Münster, zur Verfügung gestellt. Als Zweitantikörper wurde immer FITC-konjugiertes Anti-Kaninchen IgG aus der Ziege (Sigma) benutzt.

10.4.1. Anti-Hämocyanin Nachweis

Die Paraffinschnitte wurden in Xylol und Alkohol entparaffiniert (siehe 10.2.3) und für zwei Stunden bei Zimmertemperatur in einer Feuchtkammer mit Anti-Hämocyanin, 1:50 verdünnt in PBS-Puffer (0,15 M Natriumphosphat Puffer, pH 7,2) inkubiert. Nach drei 15-minütigen Spülungen in PBS wurden die Schnitte für eine Stunde mit 1:40 verdünntem FITC-konjugiertem Anti-Kaninchen-IgG aus der Ziege (Sigma) inkubiert und wiederum dreimal 15 Minuten in PBS gewaschen. Die Gegenfärbung erfolgte für 5 Minuten mit 0,1% Evans Blue (Sigma) in PBS. Vor dem Mikroskopieren wurden die Schnitte mit PBS und DABCO (25 mg/ml) eingedeckt, um ein Ausbleichen der Fluoreszenz zu verhindern.

10.4.2. Anti-Astacin und Anti-Pro-Astacin Nachweis

Die entparaffinierten Schnitte wurden für 10 Minuten mit 0,3% Triton behandelt, um ein besseres Eindringen der Antikörper zu ermöglichen, dann für zwei Stunden bei Zimmertemperatur in einer Feuchtkammer mit Anti-Astacin bzw. Anti-Pro-Astacin, 1:10 verdünnt in PBS, inkubiert, dreimal 15 Minuten in PBS gewaschen und anschließend für eine Stunde mit 1:200 verdünntem FITC-konjugiertem Anti-Kaninchen-IgG inkubiert. Nach dreimal 15 Minuten Waschen in PBS erfolgte die Gegenfärbung für 10 Minuten mit 0,1 % Evans Blue in PBS. Eingedeckt wurde mit PBS und DABCO (25 mg/ml).

10.4.3. Mikroskopische Auswertung

Die Auswertung erfolgte mit einem Leitz Aristoplan (Leitz, Wetzlar) Fluoreszenzmikroskop. Bilder wurden mit einem Photoautomat (MPS 45, Wild) auf verschiedenen Kodak Diafilmen dokumentiert.

11. Proteolyse und massenspektrometrische Analyse von Modellpeptiden

11.1. Spaltungsversuche mit synthetischen Modellpeptiden

Die Spaltungsversuche verschiedener Modellpeptide wurden mit affinitätsgereinigtem Astacin und Trypsin (siehe 6) durchgeführt. Die lyophilisierten Enzyme und die Peptide wurden in 50 mM Tris-HCl pH 8,0 aufgenommen. Für einen Probenansatz von 20 µl in 50 mM Tris-HCl pH 8,0 wurden 0,5 µg des jeweiligen Enzyms und 5 µg der Peptide AstG, AstR oder AstR* für 2 h bei 30°C inkubiert. Für die Ansätze mit Trypsin wurde dem Puffer zusätzlich 10 mM CaCl₂ zugegeben. Die Quantifizierung der Enzymmenge basierte auf der Messung bei 280 nm, die Substratkonzentration wurde durch Wägen bestimmt. Die entstandenen Spaltprodukte wurden bis zur massenspektrometrischen Analyse bei -80°C gelagert.

11.2. MALDI-TOF Massenspektrometrie

Die Herstellung der Matrix und die MALDI-TOF Messungen erfolgten wie in 5.5 beschrieben. Die aus den Spaltungsansätzen erhaltenen Proben wurden 1:10 mit 10% Ameisensäure verdünnt und davon 0,5 µl direkt auf das vorbereitete Metalltarget aufgetragen. Die Kalibrierung des Massenspektrometers geschah als lineare Zweipunktmessung mit Angiotensin I (m/z 1269,65) und der oxidierten B-Kette des Rinderinsulins (m/z 3494,65). Die protonierten Molekülmassen der Peptide und enzymatischen Fragmente wurden mit Hilfe des Programms MacProMass Version 1.03 berechnet.

12. Computergestützte Analysen

12.1. Programme für molekulargenetische Analysen

- Alle genspezifischen Oligonukleotide wurden mit dem Programm OLIGO 5.0, National Bioscience, Plymouth erstellt. Die Schmelztemperatur wurde entsprechend der jeweiligen Anwendung zwischen 58°C und 74°C gewählt.
- Die Erstellung von Plasmidkarten und die Vorhersage von Restriktionsschnittstellen wurde mit dem Programm CLONE 3.11, Scientific & Education Software durchgeführt.
- Für das Editieren von Sequenzdaten wurde das STADEN PROGRAM PACKAGE 2001, Medical Research Council, Cambridge (<http://www.mrc-lmb.cam.ac.uk/pubseq>) verwendet.
- Die Berechnung der theoretischen Molekulargewichte und des isoelektrischen Punktes eines Proteins erfolgte mit dem Programm *Compute pI/Mw tool* (http://www.expasy.ch/tools/pi_tool.html).

12.2. Datenbankanalyse Astacin-homologer Gene

Für die Ähnlichkeitssuche paraloger und orthologer Astacin Gene wurde die BLAST (*Basic Local Alignment Search Tool*) und FASTA (*Fast Alignment*) Methode verwendet. Als Vergleichssequenzen wurde die Astacin Sequenz (*Astacus*, Position +50 bis +251, *Swissprot Acc.* P07584), die Protease-Domäne von BMP-1 (Mensch, Position +121 bis +321, *Swissprot Acc.* P13497), sowie das zinkbindende Motif und der Met-Turn (*motif 2/3* CVYHGTIIHELMHAIGFYH/ EHTRMDRDNYVTINYQNV und *motif 4* DYQYY SIMHYGKYSFS; SPRINT Datenbank <http://www.bioinf.man.ac.uk/dbbrowser/sprint>) benutzt.

Die erste Identifizierung erfolgte mit Hilfe des NCBI-BLAST (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) und des EBI-BLAST Servers (<http://www.ebi.ac.uk/blast2>). Als Programm wurde TBLASTN verwendet, welches die Aminosäuresequenz mit Nukleinsäuredatenbanken vergleicht, die in allen Leserastern translatiert sind. Als Parameter wurden die Standard-Einstellungen mit einer Blosom62 Matrix und einem *expected threshold* Wert (E) = 10 verwendet. Die Suche erfolgte in der NCBI Datenbank *nr* (alle nicht-redundanten *GenBank CDS translations*, *PDB*, *SwissProt*, *PIR*, *PRF* Datenbanken) und in der *EMBL database (all divisions)*.

Die Identifizierung wurde im nächsten Schritt durch eine FASTA3 Suche auf Proteinebene (<http://www.ebi.ac.uk/fasta33/>) und auf Nukleinsäureebene (<http://www.ebi.ac.uk/fasta33/genomes.html>) in der EMBL-Datenbank vervollständigt. Als Parameter wurden die Standard-Einstellungen verwendet.

Die Sequenzdaten verschiedener Organismen, insbesondere solche mit laufenden Genom-Sequenzierungsprojekten sind nicht vollständig in den bisher genannten Datenbanken eingeschlossen. Deshalb mussten diese Daten mit Hilfe verschiedener BLAST-Suchen vervollständigt werden. Die Programm-Parameter wurden immer auf die Standard-Werte mit einer Blosom62 Matrix und einem *expected threshold* Wert (E) = 10 eingestellt.

Folgende BLAST Server wurden verwendet:

- TBLASTN in EMBL *updates*: <http://www.ebi.ac.uk/service/tmp/762905.9086-125474.html>
- TBLASTN in *microbial genomes* NCBI (133 Bakterien/14 Archaeabakterien): http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/genom_table.cgi
- TBLASTN in *C. elegans genomic* und BLASTP in *wormpep*: http://www.sanger.ac.uk/Projects/C_elegans/blast_server.shtml
- TBLASTN in *All Drosophila sequences* (release2.9) und BLASTP in *All Predicted Proteins*: <http://edgp.ebi.ac.uk/www-blast.html>
- TBLASTN in *human „golden path“* und BLASTP in *human genscan/ensembl peptides*: http://www.ensembl.org/Homo_sapiens/blastview
- TBLASTN in *mouse cDNA/latestgp* und BLASTP in *mouse peptides*: http://mouse.ensembl.org/Mus_musculus/blastview
- TBLASTN in der *Sequencing Projects Database* des Sanger-Centers <http://www.sanger.ac.uk/DataSearch>

Alle Suchergebnisse wurden auf Mehrfacheinträge untersucht. Zur Überprüfung erfolgte ein Mehrfach-Sequenzvergleich (*multiple alignment*) aller durch BLAST oder FASTA identifizierten Proteine. Dazu wurden zunächst die vollständigen Protein-Sequenzen mit Hilfe des *Sequence Retrieval System* (SRS, <http://genius.embnet.dk/fz-heidelberg.de/menu/srs>) kopiert. Die *Drosophila* bzw. *C. elegans* Sequenzen wurde direkt aus der *FlyBase* Datenbank (<http://flybase.bio.indiana.edu>) bzw. *wormbase* Datenbank (<http://www.wormbase.org>) erhalten. Gene, die ausschließlich in genomischen

Datenbanken vorhanden waren, wurden mit Hilfe des Programms *GenScan* (<http://genome.dkfz-heidelberg.de/cgi-bin/GENSCAN/genscan.cgi>) oder *Genfinder* (<http://www.softberry.com/nucleo.html>) in Proteinsequenzen übersetzt. Der Sequenzvergleich wurde mit dem Programm ClustalX (<ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX>) durchgeführt.

Die Daten wurden abschließend mit der MEROPS Datenbank (<http://www.merops.co.uk>) verglichen, welche inzwischen einen umfassenden Überblick über alle bekannten proteolytischen Enzyme gibt.

Für die Ermittlung der Domänenstruktur eines Proteins wurden die Programme SMART (*Simple Modular Architecture Research Tool*, <http://smart.embl-heidelberg.de>), PFAM (*Protein families database of alignments and HMMs*, <http://www.sanger.ac.uk/Software/Pfam>) und ProfileScan (<http://hits.isb-sib.ch/cgi-bin/PFSCAN>) benutzt.

12.3. Phylogenetische Analyse

Der erste Schritt einer molekular-phylogenetischen Untersuchung ist das Erstellen eines Mehrfach-Sequenzvergleichs (*multiple alignment*). Dieser wurde für alle Astacin-homologen Proteine mit der gemeinsamen, 200 Aminosäurereste langen Protease-Domäne durchgeführt. Dieser Abschnitt entspricht beim Astacin des Flusskrebse dem Teilstück von Position +50 bis +251 (*Swissprot Acc. P07584*). Der Sequenzvergleich wurde mit dem Programm ClustalX (<ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX>) und folgenden Parametern berechnet: (1) *pairwise alignment parameter: gap opening = 10, gap extension = 0,2, protein weight matrix gonnet 250*; (2) *multiple alignment parameter: gap opening = 10, gap extension = 0,2, delay divergent sequences 30%, protein weight matrix gonnet series*; (3) *protein gap parameters: gap separation distance = 5*. Der erhaltene Sequenzvergleich wurde in das Programm GeneDoc (www.psc.edu/biomed/genedoc) geladen und manuell anhand der Strukturdaten des Astacins korrigiert.

Die Stammbaumerstellung erfolgte auf Basis dieses Sequenzvergleichs mit dem Programmpaket PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>). Dazu wurde zunächst eine Distanzmatrix mit dem Programm PROTDIST (PAM dayhoff matrix) berechnet. Der Stammbaum wurde nach der *Neighbor-joining* Methode (Programm NEIGHBOR) erstellt. Die Darstellung erfolgte in dem Programm *Treeview* (<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>). Zur statistischen Überprüfung der Ergebnisse wurde eine *bootstrap* Analyse (Programme NEIGHBOR, CONSENSE) mit 100 Wiederholungen durchgeführt.

12.4. *C. elegans* und das Internet

Jede Information zu *C. elegans* wird seit Beginn des Genom Projekts im Jahre 1991 in der Datenbank *wormbase* (alter Name: ACeDB, *A C. elegans Database*) gespeichert und ist unter der Internet-Adresse <http://www.wormbase.org> frei zugänglich. Diese Seite erlaubt den Zugang über verschiedene Funktionen:

- BLAST Suche gegen die genomische Sequenzdaten oder die *genefinder* Vorhersage der Proteine (siehe 12.2.).
- Textsuchen über die gesamte Datenbank.
- Textsuche nach Genen, Klonen, EST oder genomischen Sequenzen.
- Suche über die physikalische Genkarte oder Klonkarte.

Über diese reinen Sequenzdaten hinaus enthält die Datenbank für jedes vorhergesagte oder bereits identifizierte Gen eine Reihe weiterer Informationen. Dies beinhaltet zum einen alle

bibliographischen Daten der bereits untersuchten Gene. Zum anderen sind alle Informationen der funktionellen Genomanalyse (EST, *ORFeome*, *gene expression map*, *reverse genetics*, *gene expression patterns*) für das jeweilige Gen gespeichert.

Das *C. elegans Sequencing Consortium* hat mit Fertigstellung der genomischen Sequenzierung im Jahr 1998 begonnen, systematisch alle durch *genefinder* vorhergesagten Gene von *C. elegans* näher zu charakterisieren. Diese Großprojekte werden in Kooperation mit verschiedenen Laboratorien durchgeführt. Alle Daten werden an die *wormbase* Datenbank übermittelt und deren Herkunft mit einem Buchstabenkürzel gekennzeichnet. Ein Überblick der an den jeweiligen Projekten beteiligten Forscher ist unter <http://biosci.umn.edu/CGC/CGChomepage.htm> abrufbar. Im Folgenden sind die wichtigsten Projekte näher beschrieben:

- **Alternative Spleißmuster (*Intronerator*)**

Kent und Zahler haben im Jahr 2000 eine Datenbank mit alternativen Spleißmustern (*Intronerator*, <http://www.cse.ucsc.edu/~kent/intronerator/index.html>) aller Gene erstellt, die z.T. erheblich von den *genefinder* Vorhersagen abweichen. Der Vorteil dieser Datenbank liegt in der Einbeziehung aller bis zu diesem Zeitpunkt verfügbaren cDNA- oder EST-Daten in die Computervorhersage.

- **Transkriptom-Projekt (*ORFeome*)**

Das *ORFeome* Projekt (<http://worfdb.dfci.harvard.edu>) versucht dagegen systematisch über RT-PCR Experimente und Sequenzierung alle durch *genefinder* vorhergesagten Gene zu bestätigen oder alternative Spleißvarianten aufzuzeigen.

- **EST-Projekt**

Die Analyse aller exprimierten Gene auf cDNA Ebene wurde in erster Linie von Kohara und Mitarbeiter (<http://www.ddbj.nig.ac.jp>) durchgeführt. EST-Klone sind inzwischen für ca. 40% der vorhergesagten Gene gefunden worden.

- **Expressionskarte (*Topomap*)**

Kim und Mitarbeiter erstellten 2001 eine Expressionskarte (*Sk-map*) mit 17,661 Genen zur Analyse coregulierter Gene. Insgesamt wurden 537 *microarray*-Experimente durchgeführt, die auf der Internetseite http://cmgm.stanford.edu/~kimlab/topomap/c._elegans_topomap.htm zusammengefasst sind.

- **microarray Experimente**

Jiang und Mitarbeiter erstellten zeitgleich eine 17 871 Gene umfassende *microarray* Analyse, welche die relativen Expressionsraten während der Entwicklung vom Ei zum adulten Organismus aufzeigen sollten (<http://www.proteome.com/databases>).

- **RNA interference (RNAi)**

Für eine funktionelle Genomanalyse durch RNA-vermittelte Gen Inaktivierung (RNAi) haben verschiedene Laboratorien bis heute Experimente mit 7200 Genen durchgeführt (http://www.wormbase.org/db/searches/rnai_search).

- **Expressionsmuster**

Für ungefähr 1000 Gene wurde das Expressionsmuster durch GFP/lacZ Reporter Konstrukte oder durch *in situ* Hybridisierung (nur für Embryonen) ermittelt (http://www.wormbase.org/db/searches/expr_search).

- **Mutanten**

Alle verfügbaren und charakterisierten Mutanten von *C. elegans* können unter <http://www.wormbase.org/db/searches/strains> gefunden werden.

Alle oben genannten Datenbanken wurden nach Einträgen auf die durch BLAST/FASTA identifizierten Astacin-homologen Gene durchsucht.

IV. ERGEBNISSE

Ziel dieser Arbeit war es, einen umfassenden Überblick über die Struktur, Funktion und Evolution der Astacin-Protein-Familie zu erhalten. Dazu wurden am Beispiel der Modellorganismen *C. elegans* (1.) und *Hydractinia echinata* (2.) neue Astacin-homologe Proteine identifiziert und näher charakterisiert. Außerdem wurde eine vollständige Datenbank-Analyse und phylogenetische Analyse aller Astacin-homologen Gene durchgeführt (3.). Darüber hinaus sollte der Aktivierungsmechanismus der Pro-Enzyme am Modell des Flusskrebse *Astacus astacus* untersucht werden (4.).

1. Astacin-homologe Proteine in *C. elegans*

Vor Beginn dieser Arbeit im Jahre 1996 ergaben Voruntersuchungen (Dr. G. Geier, Zoologisches Institut, Universität Heidelberg), dass im Genom des Nematoden *Caenorhabditis elegans* mindestens 12 unterschiedliche Astacin-homologe Gene gefunden werden. Diese Zahl erschien zunächst überraschend groß, waren in anderen Spezies doch nur maximal 3 Gene (Maus, Mensch) beschrieben worden. Die Untersuchung der *C. elegans* Astacine sollte daher neue Aufschlüsse für die Funktion und Evolution der Astacin-Protein-Familie liefern.

1.1. Datenbankanalyse

Ein Projekt zur Sequenzierung des 97 Megabasen umfassenden Genoms von *Caenorhabditis elegans* wurde 1991 unter Leitung des *C. elegans* Sequencing Consortium begonnen und konnte im Dezember 1998 nahezu vollständig publiziert werden (The *C. elegans* Sequencing Consortium 1998). Diese frei zugänglichen Daten bildeten die Basis für alle weiteren Analysen Astacin-homologer Gene bei *C. elegans*.

Bei Beginn dieser Arbeit zeigte eine erste BLAST Suche nach Astacin-homologen Proteinen in der zu dieser Zeit aktuellen *wormpep11* Datenbank (Juli 1996, enthielt 7299 vorhergesagte Proteine) die Zahl von 12 Astacin-homologen Proteinen. In der kurze Zeit später veröffentlichten *wormpep12* (November 1997, 12178 Proteine) konnten bereits über 20 homologe Proteine gefunden werden. Die hier präsentierten Daten beziehen sich nun, wenn nicht anders vermerkt, auf die aktuellen *wormbase 73* und *wormpep 73 (released 7th February 2002, 20,379 predicted protein sequences)* Datenbanken.

Um zunächst überhaupt einen Überblick über das Vorkommen von proteolytischen Enzymen im Genom von *C. elegans* zu bekommen, wurde am Ende dieser Arbeit auch eine Suche in der aktuellen Protease-Datenbank MEROPS (*Release 5.7: 17-Dec-2001*) durchgeführt. Darin sind für *C. elegans* nun 353 Protease-Gene (E.C. 3.4.) insgesamt verzeichnet. 149 Gene davon gehören zur funktionellen Gruppe der Metalloproteasen (E.C. 3.4.24). Tabelle 1 zeigt die Zuordnung dieser Proteine zu den Clans und Familien nach der Einteilung von Barrett et al. (1998). Die Metzinkine (Clan MA(M)) stellen mit 57 Vertretern dabei mehr als ein Drittel aller Metalloproteasen.

Clan	Familie	Anzahl	Clan	Familie	Anzahl
MA(E)	M1 aminopeptidase	15	ME	M16X	1
	M2 peptidyl-dipeptidase	1	MF	M17 leucyl aminopeptidase	2
	M3A oligopeptidase	2	MG	M24A methionyl aminopeptidase I	6
	M13 neprilysin	22		M24B aminopeptidase P	5
	M41 <i>E. coli</i> endopeptidase	3	MH	M18 aminopeptidase I	1
	M48A Ste24 endopeptidase	1		M20A glutamate carboxypeptidase	1
MA(M)	M8 leishmanolysin	1		M20B Gly-X carboxypeptidase	4
	M10A MMP	6		M28B aminopeptidase Y	2
	M12A Astacin	35		M28C IAP aminopeptidase	1
	M12B ADAM	8	M28X	3	
	M12C ADAM	2	MJ	M38 beta-aspartyl dipeptidase	1
MC	M14A carboxypeptidase A	9	MK	M22 O-sialoglycoprotein endopeptidase	2
	M14B carboxypeptidase E	3	MM	M50 S2P protease	1
ME	M16A pitrilysin	5	MX	M49 dipeptidyl-peptidase III	1
	M16B mitochondrial processing peptidase	3			

Tabelle 1: 152 Gene für Metalloproteasen in *C. elegans* zusammengestellt nach der MEROPS Protease-Datenbank (*Release 5.7: 17-Dec-2001*). Die Gen-Anzahl für die einzelnen Protease-Familien ist jeweils in der rechten Spalte angegeben.

Um einen vollständigen Überblick aller im Genom von *C. elegans* vorhandenen Astacin-homologen Gene zu erhalten, mussten unabhängig der in der aktuellen MERPOS Datenbank (*Release 5.7: 17-Dec-2001*) postulierten Astacin-homologen, BLAST-Suchen mit der Astacin-Domäne, dem Zinkbindungsmotiv oder dem Bereich des Met-turns in den *C. elegans* Datenbanken *wormbase* (Genom-Sequenz) und *wormpep* (vorhergesagte Proteine) durchgeführt werden (vergleiche III.12.2). Tabelle 2 zeigt das Ergebnis dieser Analyse. Angegeben ist jeweils der Protein-Name der *wormpep 73* Datenbank, die in dieser Arbeit verwendete Bezeichnung und die Identifizierungsnummer (*accession number*) der an die SWISSPROT, EMBL oder GENBANK übermittelten Sequenzen. Es konnten genau 40 Gene identifiziert werden, wobei im Vergleich zur MEROPS Datenbank fünf weitere Proteine gefunden werden. Drei Proteine (F54B8, D1022, F38E9) sind zudem nicht in der *wormpep* Proteindatenbank verzeichnet, konnten aber nach der genomischen Lokalisierung durch die TBLASTN-Suche mit Hilfe des Programms *genescan* postuliert werden. Für F54B8.m ergab die *genescan* Vorhersage allerdings kein vollständiges Protein, sondern nur ein insgesamt 88 Aminosäurereste umfassendes Fragment (vergleiche VII 6.2). Im Genom findet sich dieses Fragment in nur einem einzigen Leseraster, wobei innerhalb der DNA-Sequenz zudem zwei Basentriplets für Stopkodons kodieren und somit zumindest eine Protein-Translation unmöglich machen.

Bis auf eine Ausnahme (HCH-1, Hishida et al. 1996) waren die in der *wormbase* gefundenen Astacin-homologen Gene nicht näher charakterisiert. Es lagen bisher nur die durch das Programm *genefinder* postulierten Proteine und die partielle Transkriptom-Analyse durch die EST/ORFeome-Projekte (siehe III.12.4) vor. Darüber hinaus sind die

Astacin-homologen Gene aber experimentell weder auf RNA- noch auf Proteinebene bestätigt worden. Im nächsten Schritt sollte deshalb zunächst die Expression dieser Gene auf RNA Ebene untersucht werden.

	Gen Name	MERNUM (Release 5.8)	Swiss-Prot	TrEMBL	EMBL / GenBank
1	4R79.1	MER05846		Q9U3S9	AL031254
2	C05D11.6	MER02346	P55112		U00048
3	C07D10.4	MER02282	P55113		U13072
4	C17G1.6	MER03171		Q93243	Z78415
5	C24F3.3	MER05120		Q9XTD6	AL022716, Z81055
6	C26C6.3	MER02349		Q18206	Z72503
7	C34D4.9	MER03819		Q18439	U58755
8	C37H5.9	MER03493		P91137	U88315
9	D1022.m	MER05843			U23517
10	F09E8.6	MER03972		Q19269	Z73896
11	F20G2.4	MER03172		Q93542	Z79753
12	F38E9.2	MER03097		Q20176	U46668
13	F39D8.4	MER02352		Q20191	Z69791
14	F42A10.8	MER02171	P98061		U10414
15	F45G2.1	MER05123		O62243	Z93382
16	F46C5.3	MER02353		Q20459	Z54281
17	F54B8.m	MER05847			Z93383
18	F56A4.1	MER12259		Q9N5R0	AC006645, AC006722
19	F57C12.1	MER02287		Q20942	U41554
20	F58A6.4	MER02354		Q20958	U53339
21	F58B4.1	MER03814		Q20975	Z74038
22	HCH-1	MER03163			D85744, Z69792
23	K03B8.1	MER03164		Q21180	Z74039
24	K03B8.2	MER03115		Q21178	Z74039
25	K03B8.3	MER03116		Q21179	Z74039
26	K03B8.5	MER03165		Q21181	Z74039
27	K04E7.3	MER02288	P55114		U39666
28	K06A4.1	MER03096		Q21252	Z70755
29	K09C8.3	MER03166		Q21388	Z68006
30	K11G12.1	MER03167		Q21432	U23525
31	R151.5	MER02172	P98060		U00036
32	T02B11.7	MER03831		O16977	AF022979
33	T04G9.2	MER02291	P55115		U41274
34	T11F9.3	MER03174		Q22396	Z74042
35	T11F9.5	MER03169		Q22401	Z74042
36	T11F9.6	MER03168		Q22398	Z74042
37	T23F4.4	MER03833		O17264	AF025466
38	T23H4.3	MER03173		P91828	Z83240
39	T24A11.3	MER01112		Q22710	Z49072
40	Y95B8A.1	MER12278		Q9N2V2	AC024877

Tabelle 2: 40 Astacin-homologe Proteine in *C. elegans*.

1.2. Transkriptom-Analyse

1.2.1. RT-PCR und EST Projekte

Die über BLAST Suche der genomischen DNA identifizierten Astacin-homologen Gene wurden mit den in der *wormbase* Datenbank gespeicherten experimentellen cDNA Daten abgeglichen. Dabei zeigte sich, dass vor Beginn des hier durchgeführten Expressionsnachweises von 11 der insgesamt 40 Gene bereits EST- oder ORFeome-Klone (vergleiche Tabelle 4) bekannt waren. Die Expression dieser Gene ist auf mRNA Ebene damit eindeutig nachgewiesen. Die verbleibenden 29 Gene wurden durch RT-PCR und anschließende Sequenzierung der DNA-Fragmente untersucht. Dazu wurden für jedes dieser Gene spezifische Primer-Paare synthetisiert und die Genfragmente über PCR amplifiziert. Die Orientierung und Lage der verwendeten Primer ist im Anhang (VII.6.1.1. bis VII.6.1.40.) wiedergegeben. Abbildung 1 zeigt die Analyse der amplifizierten Produkte auf Ethidiumbromid gefärbten Agarosegelen. Für jedes Gen wurde eine PCR Reaktion mit revers transkribierter RNA (cDNA, c) und eine Kontrollreaktion mit genomischer DNA (g) durchgeführt. Da in allen Fällen ein oder mehrere Introns innerhalb des amplifizierten DNA-Bereichs lagen, ergaben die Kontrollreaktionen jeweils größere DNA-Fragmente. Tabelle 3 (rechter Teil) gibt einen Überblick der zu erwartenden Größe der genomischen PCR-Produkte (Spalte Genom) und der *genefinder* Vorhersagen der RNA-Transkripte (Spalte ORF cDNA) im Vergleich zu den experimentell gefundenen DNA-Fragmentgrößen (Spalten PCR Genom und PCR cDNA). Danach konnten für alle genomischen PCR-Ansätze Fragmente der entsprechenden Größe erhalten werden. Auf cDNA Ebene ergaben die Gene F54B8.m, F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 auch unter modifizierten PCR Bedingungen kein Produkt. Alle anderen untersuchten Gene ergaben in den RT-PCR Reaktionen aber ein spezifisches DNA-Fragment (Abbildung 1). Allerdings zeigten die Gene C05D11.6, C07D10.4, C34D4.9, F42A10.8 und K04E7.3 eine gegenüber der *genefinder* Vorhersage abweichende DNA-Größe. Um alle erhaltenen DNA-Fragmente eindeutig zu identifizieren und von der theoretischen Vorhersage abweichende Spleißmuster aufzuzeigen, wurden die PCR-Produkte aus einem Agarosegel isoliert, mit glatten Enden (*blunt end*) in den Vektor pUC 18 kloniert und sequenziert. Die Bezeichnung der Klone sind in Tabelle 23 (rechte Spalte) und die cDNA Sequenzen im Anhang (VII.6.1.1 bis VII.6.1.40) wiedergegeben. Dabei ist ersichtlich, dass die Gene C05D11.6, C07D10.4, C34D4.9, F42A10.8 und K04E7.3 tatsächlich eine von der *genefinder* Vorhersage abweichendes Spleißmuster aufweisen. Die Expression der verbleibenden Gene, welche nach den PCR-Ergebnisse kein anderes Spleißmuster erwarten lassen, konnten durch die Sequenzierung alle bestätigt werden. Für die Gene F54B8.m, F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 konnten anhand dieser Daten noch keine eindeutigen Aussagen über deren Expression getroffen werden. Das im Jahre 2001 begonnene ORFeome Projekt (vgl. Tabelle 3) und die Expressionskarte von Kim et al. (Tabelle 3 und Abbildung 7) können darüber aber weitere Aussagen machen.

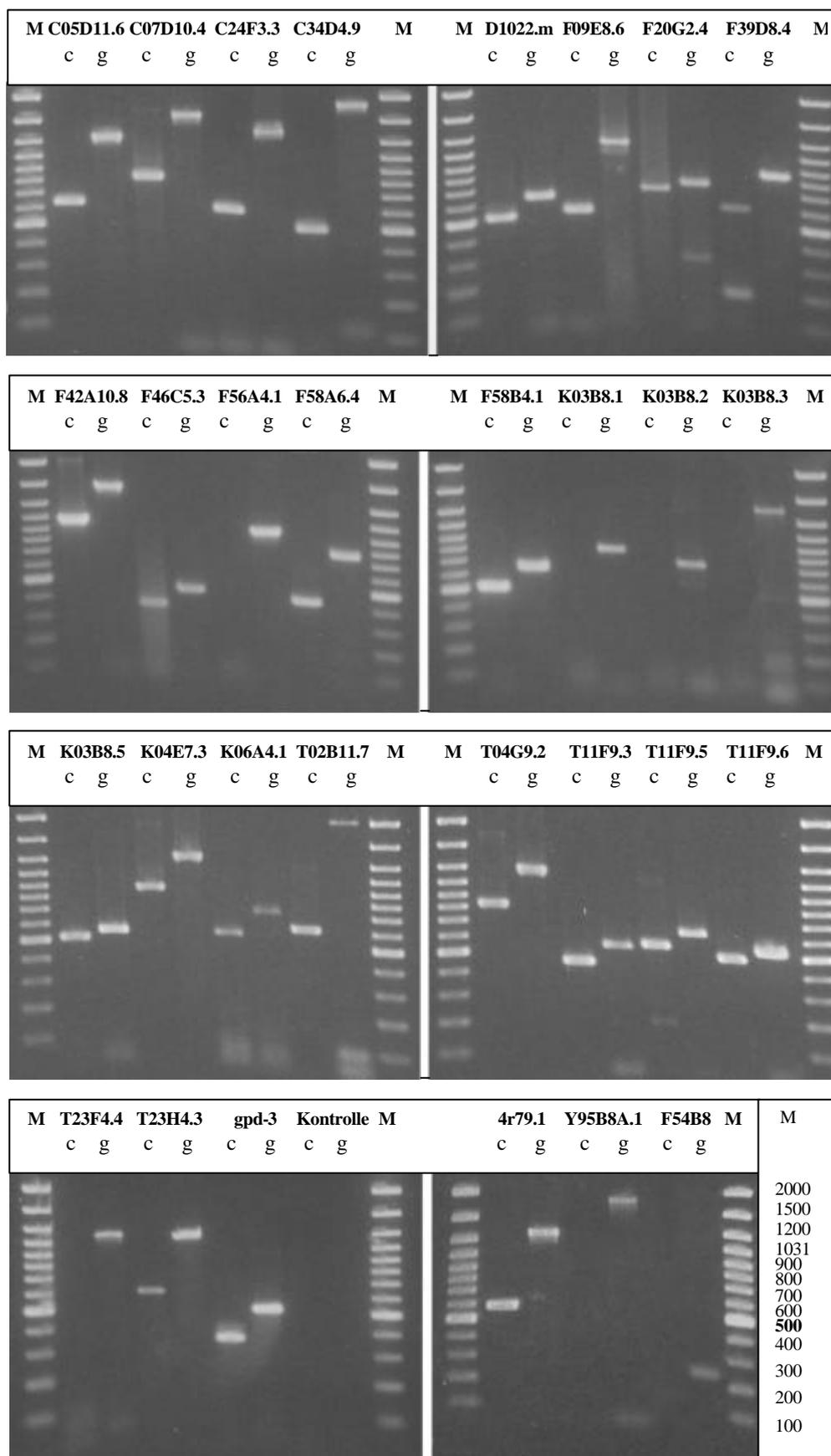


Abbildung 1: 1,5 %ige Ethidiumbromid gefärbte Agarosegele mit DNA-Fragmenten aus den RT-PCR Experimenten. c: RT-PCR; g: Kontroll PCR mit genomischer DNA; M DNA-Größenmarker

Gen	Expressionsdatenbank ORFeome				EST	SK- map <i>mount</i>	Ergebnisse RT-PCR / Klonierung vgl. auch Abbildung 1				
	Experiment	ORF cDNA (bp)	PCR (bp)	ORF eome (bp)			ORF cDNA (bp)	Genom (bp)	PCR cDNA (bp)	PCR Genom (bp)	Sequenz/ Klon siehe
4R79.1	208 / D4	768	800			10	567	1256	600	1250	VII 6.1.1
C05D11.6	044 / F2	1005	1200	1		14	742	1268	650	1250	VII 6.1.2
C07D10.4	108 / B5	1257	1300			16	905	1585	850	1600	VII 6.1.3
C17G1.6	222 / G1	2211	2200		2	38					
C24F3.3	138 / F2	1197	1200			03	603	1315	600	1300	VII 6.1.5
C26C6.3	088 / B6	1956	2000		3	29					
C34D4.9	138 / D5	1176	-			03	688	1765	500	1800	VII 6.1.7
C37H5.9	196 / G6	1554	1650		4	08					
D1022.m							536	619	530	650	VII 6.1.9
F09E8.6	140 / E9	1554	1600			14	560	1119	560	1150	VII 6.1.10
F20G2.4	179 / D3	1101	1150			00	658	753	680	750	VII 6.1.11
F38E9.2	222 / H10	2304	2500		5	03					
F39D8.4	215 / E11	1947	-			09	557	854	600	850	VII 6.1.13
F42A10.8	045 / B3	1254	1300	6		17	773	1427	950	1400	VII 6.1.14
F45G2.1	043 / G3	777	750	7							
F46C5.3	105 / H1	1038	1300		8	13	362	448	350	450	VII 6.1.16
F54B8.m							258	258	-	260	
F56A4.1	163 / C11	669	-				798	937	-	950	VII 6.1.18
F57C12.1	222 / G11	2313	2300		9	14					
F58A6.4	108 / E2	1320	1400			14	476	768	480	800	VII 6.1.20
F58B4.1	188 / F5	1866	2000			06	511	652	500	650	VII 6.1.21
HCH-1	154 / E6	1818	1800		10	12					
K03B8.1	184 / B10	1356	-			00	683	810	-	800	
K03B8.2	184 / A5	1347	-			00	535	704	-	700	
K03B8.3	182 / D6	1233	-			01	642	1350	-	1350	
K03B8.5	185 / A9	1440	1600			26	509	552	500	550	VII 6.1.26
K04E7.3	215 / B5	1656	-			13	929	1313	850	1250	VII 6.1.27
K06A4.1	186 / C6	1536	-			21	558	712	550	750	VII 6.1.28
K09C8.3	221 / F3	1683	-		11	06					
K11G12.1	221 / H1	1764	-		12						
R151.5	058 / F8	1833	-		13	01					
T02B11.7	189 / D9	2103	-		14	06	636	1970	640	2000	VII 6.1.32
T04G9.2	222 / A5	1866	1400		15	14	836	1204	830	1200	VII 6.1.33
T11F9.3	171 / G11	972	1000			22	458	557	450	550	VII 6.1.34
T11F9.5	166 / C11	798	850			28	576	664	560	650	VII 6.1.35
T11F9.6	162 / B4	603	650			03	513	556	500	550	VII 6.1.36
T23F4.4	109 / B6	1395	1600			29	519	1251		1250	VII 6.1.37
T23H4.3	081 / B9	1083	1100			00	664	1481	650	1500	VII 6.1.38
T24A11.3	057 / C6	1179	-		16	16					
Y95B8A.1	089 / D5	2241	-			01	773	1850	-	1900	

Tabelle 3: Transkriptom-Analyse der Astacin-homologen Gene. Dargestellt sind vergleichend die Ergebnisse des ORFeome Projekts (<http://worfdb.dfci.harvard.edu>), des EST Projekts (<http://www.ddbj.nig.ac.jp>), der Expressionskarte (*SK-map*) von Kim et al. (2001) und der RT-PCR Analyse dieser Arbeit. Angegeben sind jeweils die theoretischen Fragmentlängen (**ORF-cDNA** und **Genom**) sowie die experimentell gefundenen Fragmentlängen (**PCR**, **PCR-cDNA** und **PCR-Genom**) der PCR-Produkte in Basenpaaren (bp). Zur Erläuterung der ORFeome- und EST-Klone (Nr. **1-16**, Fettdruck) siehe Tabelle 4. Für die *SK-map* ist die Lokalisation in der jeweiligen Koexpressionsgruppe (*mount*) angegeben.

Nr.	Gen	EST/ORFeome Klone
1	C05D11.6	BII75418 (neu)
2	C17G1.6	yk130e2.5 yk180a11.5 yk180a11.3 yk355c4.3 yk355c4.5 yk581h6.3 yk581h6.5
3	C26C6.3	yk11f8.3 yk11f8.5
4	C37H5.9	yk852b05.3 yk892d11.3 yk827h04.3 yk892d11.5 yk827h04.5 yk852b05.5 yk120a2.3 yk120a2.5 yk164b6.3 yk164b6.5 yk298f4.3 yk298f4.5 yk300b8.3 yk300b8.5 yk581h5.3 yk581h5.5 yk722h5.5 yk1010b03.3 yk890g03.3 yk1011b03.3 yk722h5.3
5	F38E9.2	yk414e5.3 yk414e5.5 yk839g10.5
6	F42A10.8	BII74472, BII75435 (neu)
7	F45G2.1	BII74418, BII75363, BII75385
8	F46C5.3	yk701a10.3 yk701a10.5
9	F57C12.1	yk148g6.3 yk148g6.5 yk378d10.3 yk378d10.5 yk384c9.3 yk384c9.5
10	HCH-1	CEESG73F yk82b7.3 yk82b7.5 yk167a12.3 yk167a12.5 yk628d7.3 yk628d7.5 yk435f3.3 yk435f3.5 yk376d4.3 yk376d4.5 yk486e2.5 yk508c7.3 yk645g3.3 yk657c10.3 yk663d6.3 yk508c7.5 yk645g3.5 yk657c10.5 yk663d6.5 Kompletter cDNA Klon: D85744
11	K09C8.3	yk439a6.5
12	K11G12.1	yk721e4.5, yk753c12.3, yk721e4.3, yk129a7.3, yk129a7.5
13	R151.5	yk709e2.5, yk820f06.5, yk702c10.5, yk730d8.5, yk462f12.5, yk168f6.5, yk456f9.5, yk728d4.5, yk146b10.5, yk189d3.5, yk101h3.5, yk820f06.3, yk168f6.3, yk728d4.3, yk462f12.3, yk189d3.3, yk730d8.3, yk709e2.3, yk146b10.3, yk456f9.3, yk702c10.3, yk101h3.3
14	T02B11.7	yk726g11.3, yk726g11.5 (neu)
15	T04G9.2	yk118h9.3 yk118h9.5 yk275h5.5 yk366f6.3 yk366f6.5 (neu)
16	T24A11.3	yk729e4.5 cm01a12 yk296d6.5 yk73a3.5 yk296d6.3 yk73a3.3

Tabelle 4: Bezeichnung der EST und ORFeome Klone nach der *wormbase* Datenbank (vergleiche Tabelle 3).

Das ORFeome Projekt hat dabei zum Ziel alle 20 000 vorhergesagten *C. elegans* Transkripte durch spezifische Primer zu amplifizieren und durch vollständige Sequenzierung zu bestätigen. Die PCR Ergebnisse zeigen aber, dass viele der durchgeführten Reaktionen unter Standard-PCR Bedingungen keine Produkte ergeben. Im Falle der *C. elegans* Gene gilt dies beispielsweise für C34D4.9, K04E7.3, K06A4.1, K09C8.3, K11G12.1, R151.5 und T24A11.3. Darüber hinaus sind bisher nur wenige der tatsächlich amplifizierten PCR-Produkte durch Sequenzierung bestätigt worden (Astacine: C05D11.6, F42A10.8, F45G2.1). Die Aussagen die hier getroffen werden können sind also nach derzeitigem Stand des Projektes größtenteils nicht eindeutig.

Kim und Mitarbeitern ist es gelungen 1991 auf der Expressionskarte (*SK-map*) über 90% der vorhergesagten *C. elegans* Gene zu lokalisieren (Kim et al. 2001). Die Astacin-homologen Gene D1022.m, F54B8.m, F56A4.1 und K11G12.1 sind allerdings nicht repräsentiert. Die bisher über PCR nicht nachgewiesenen Gene F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 können dagegen auf der Expressionskarte gefunden werden (Abbildung 7). Die hier verwendete *microarray* und Hybridisierungstechnik ist aber hauptsächlich auf große Analysedaten von koexprimierten Gengruppen ausgelegt und soll nicht explizit dem Nachweis einzelner Gene dienen. Deshalb wird auch in Kauf genommen, dass im Einzelfall ein nicht signifikantes Hybridisierungssignal (Fluoreszenz) positiv gewertet wird und damit fälschlicherweise die Expression eines Genes als gegeben

angenommen wird. Dennoch gibt die Lokalisierung der Gene F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 auf der *SK-map* einen deutlichen Hinweis auf die tatsächliche Expression dieser Gene.

Zusammenfassend kann festgestellt werden, dass mit Ausnahme F54B8.m alle Astacin-homologen Gene auf Transkriptionsebene nachgewiesen werden konnten, wenn auch für die Gene F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 eine endgültige Bestätigung aussteht. Das experimentell gefundene Spleißmuster weicht in manchen Fällen deutlich von der theoretischen Vorhersage von *genefinder* ab. Im nächsten Abschnitt sollen die bereits bekannten cDNA-Sequenzen deshalb mit den genomischen Sequenzdaten abgeglichen werden.

1.2.2. Abgleich der cDNA Sequenzen mit den genomischen Sequenzdaten

Für eine korrekte Übersetzung der Transkripte in Proteinsequenzen war es nötig, die experimentell noch nicht überprüften RNA-Vorhersagen mit den bereits bekannten cDNA-Sequenzen zu vergleichen. Darüber hinaus sollte damit auch die Zuverlässigkeit verschiedener Programme in der Strukturvorhersage analysiert werden. Dazu wurden Sequenzvergleiche der *genefinder* Vorhersagen der *wormpep 73* Datenbank und die alternativen *intronerator* Vorhersagen von Kent et al. (Kent und Zahler 2000, <http://www.soe.ucsc.edu/~kent/intronerator>, *released 07-Dec-2001*) mit den experimentell ermittelten cDNA Sequenzen durchgeführt. Die Sequenzvergleiche sind im Anhang VII.6.1 dargestellt. Tabelle 5 zeigt die jeweiligen Gen-Bezeichnungen der *wormpep 73* Datenbank, der Genie Gene (*Intronerator*) Datenbank und der sequenzierten EST/OST/cDNA-Klone. Vollständige cDNA Sequenzen liegen nur für die Gene F45G2.1 HCH-1 und R151.5 vor. Für die EST Klone fehlen in der Regel die 5' Bereiche der mRNA, da in den cDNA Banken oft verkürzte Transkripte vorhanden sind. Für die mit RT-PCR und genspezifischen Primern ermittelten cDNA Sequenzen fehlt, methodisch bedingt, immer die 5' und die 3' Region der mRNA. Die tatsächlichen Transkriptions-Start- und End-Punkte wären nur durch aufwendige 5' oder 3' RACE Experimente zugänglich und wurden im Rahmen dieser Arbeit nicht durchgeführt. Darüber hinaus fehlen bei einigen sehr langen Transkripten noch Sequenzdaten aus den mittleren Bereichen. Für die vollständige Übersetzung der mRNA in Proteinsequenzen wurde deshalb die Vorhersage (*wormpep* oder Genie Gene) verwendet, die den bisherigen experimentellen cDNA Daten entspricht. Gegebenenfalls wurden zusätzliche alternative Spleißmuster berücksichtigt. Die Übersetzung in Proteinsequenz geschah anschließend mit dem *Staden Programm Package* und ist im FASTA Format im Anhang VII.6.2 dargestellt. In Tabelle 25 (rechte Spalte) ist die Kurzbezeichnung dieser korrigierten Proteinsequenzen angegeben, wie sie für die folgenden phylogenetischen und strukturellen Untersuchungen (Ergebnisse IV. 3) verwendet wurde.

<i>genefinder</i> (<i>wormpep</i>)	<i>Intronator</i> (<i>Genie Gene</i>)	genomische Position (Chromosom)	EST/ ORF- eome	Klon	Beste Übereinstimmung	ID Korrigierte Protein Sequenz
4R79.1	g-IV-3005	IV: 30.16		VII 6.1.1	Genie Gene	C_4R79.1i
C05D11.6	g-III-1042	III:-1.33	1	VII 6.1.2	Genie Gene	C_C05D11.6i
C07D10.4	g-II-1703	II:0.41		VII 6.1.3	Genie Gene	C_C07D10.4i
C17G1.6	g-X-2881	X:1.48	2		Genie Gene	C_C17G1.6i
C24F3.3	g-IV-1838	IV:4.54		VII 6.1.5	wormpep	C_C24F3.3
C26C6.3	g-I-1289	I:2.05	3		wormpep	C_C26C6.3
C34D4.9	g-IV-1119	IV:3.29		VII 6.1.7	Genie Gene	C_C34D4.9i
C37H5.9	g-V-856	V:-6.52	4		Genie Gene	C_C37H5.9i
D1022.m		II:0.45		VII 6.1.9	ORF Mernum	C_D1022.m
F09E8.6	g-IV-2471	IV:8.02		VII 6.1.10	Genie Gene	C_F09E8.6i
F20G2.4	g-V-2804	V:5.42		VII 6.1.11	Genie Gene	C_F20G2.4i
F38E9.2	g-X-2544	X:23.83	5		wormpep 74	C_F38E9.2n
F39D8.4	g-X-2412	X:21.46		VII 6.1.13	Genie Gene	C_F39D8.4i
F42A10.8	g-III-995	III:-1.38	6	VII 6.1.14	cDNA RT-PCR	C_F42A10.8m
F45G2.1	g-III-2312	III:22.1	7		OST (vollständig)	C_F45G2.1o
F46C5.3	g-II-2024	II:0.92	8	VII 6.1.16	wormpep	C_F46C5.3n
F54B8.m		V:9.77			ORF Mernum	C_F54B8.m
F56A4.1	g-V-409	V:-13.99		VII 6.1.18	Genie Gene	C_F56A4.1i
F57C12.1	g-X-41	X:-19.47	9		Genie Gene	C_F57C12.1i
F58A6.4	g-II-1161	II:-1.98		VII 6.1.20	Genie Gene	C_F58A6.4i
F58B4.1	g-V-2200	V:2.87		VII 6.1.21	wormpep	C_F58B4.1
HCH-1	g-X-2296	X:19.9925	10		cDNA Hishida et al.	C_HCH1
K03B8.1	g-V-2307	V:3.16			wormpep	C_K03B8.1
K03B8.2	g-V-2310	V:3.16			wormpep	C_K03B8.2
K03B8.3	g-V-2309	V:3.16			wormpep	C_K03B8.3
K03B8.5	g-V-2311	V:3.16		VII 6.1.26	wormpep	C_K03B8.5
K04E7.3	g-X-945	X:-2.93		VII 6.1.27	wormpep	C_K04E7.3
K06A4.1	g_V-1836	V:1.98		VII 6.1.28	Genie Gene	C_K06A4.1i
K09C8.3	g-X-1670 g- X-1671	X:2.51	11		wormpep	C_K09C8.3
K11G12.1	g-X-993	X:-2.66	12		wormpep	C_K11G12.1
R151.5	g-III-1186	III:-0.76	13		EST (vollständig)	C_R151.5
T02B11.7	g-V-166	V:-19.07	14	VII 6.1.32	wormpep 74	C_T02B11.7n
T04G9.2	g-X-2732	X:-19.12	15	VII 6.1.33	Genie Gene	C_T04G9.2i
T11F9.3	g-V-2325	V:3.2		VII 6.1.34	Genie Gene	C_T11F9.3i
T11F9.5	g-V-2326	V:3.2		VII 6.1.35	Genie Gene	C_T11F9.5i
T11F9.6	g-V-2327	V:3.2		VII 6.1.36	Genie Gene	C_T11F9.6i
T23F4.4	g-II-266	II:-13.27		VII 6.1.37	wormpep	C_T23F4.4
T23H4.3	g-I-1809	I:4.03		VII 6.1.38	wormpep	C_T23H4.3
T24A11.3	g-III-483	III:-4.54	16		Genie Gene	C_T24A11.3i
Y95B8A.1	g-I-117	I:-20.88			wormpep	C_Y95B8A.1

Tabelle 5: Astacin-homologen Gene in *C. elegans*: Alternative Vorhersage der Spleißmuster (*wormpep*, *Genie Gene*), genomische Lokalisation (Chromosom), EST/ORFeome-Klone (Nr. 1-16, siehe Tabelle 24), die im Rahmen dieser Arbeit sequenzierten cDNA-Klone (vergleiche Anhang VII.6.1.) und Kurzbezeichnung der anhand dieser Daten übersetzten Proteinsequenzen (vergleiche Anhang VII.6.2.).

1.3. Funktions-Analyse

Die Expression der Astacin-homologen Gene wurde bereits bei der Transkriptomanalyse nachgewiesen. Weitergehende Untersuchungen sollten nun an ausgewählten Beispielen die Lokalisierung *in vivo* durch GFP-Reporter Proteine zeigen, um erste Aussagen über die Funktion dieser Gene machen zu können. Darüber hinaus sollten die bereits vorliegenden Daten der RNAi, *microarray* oder GFP-Reporter Großprojekte (siehe III.12.4) mit in die Funktions-Analyse einbezogen werden.

1.3.1. GFP-Fusionsproteine

Zur Herstellung der GFP-Reporter Konstrukte wurden die Gene C05D11.6, C07D10.4, K04E7.3 und T04G9.2 mit spezifischen Primern amplifiziert, mit Restriktionsendonukleasen verdaut und gerichtet in die Vektoren bluescript SK+/pBD114_35-GFP oder pBD 95_35 kloniert (Abbildung 2 bis 5).

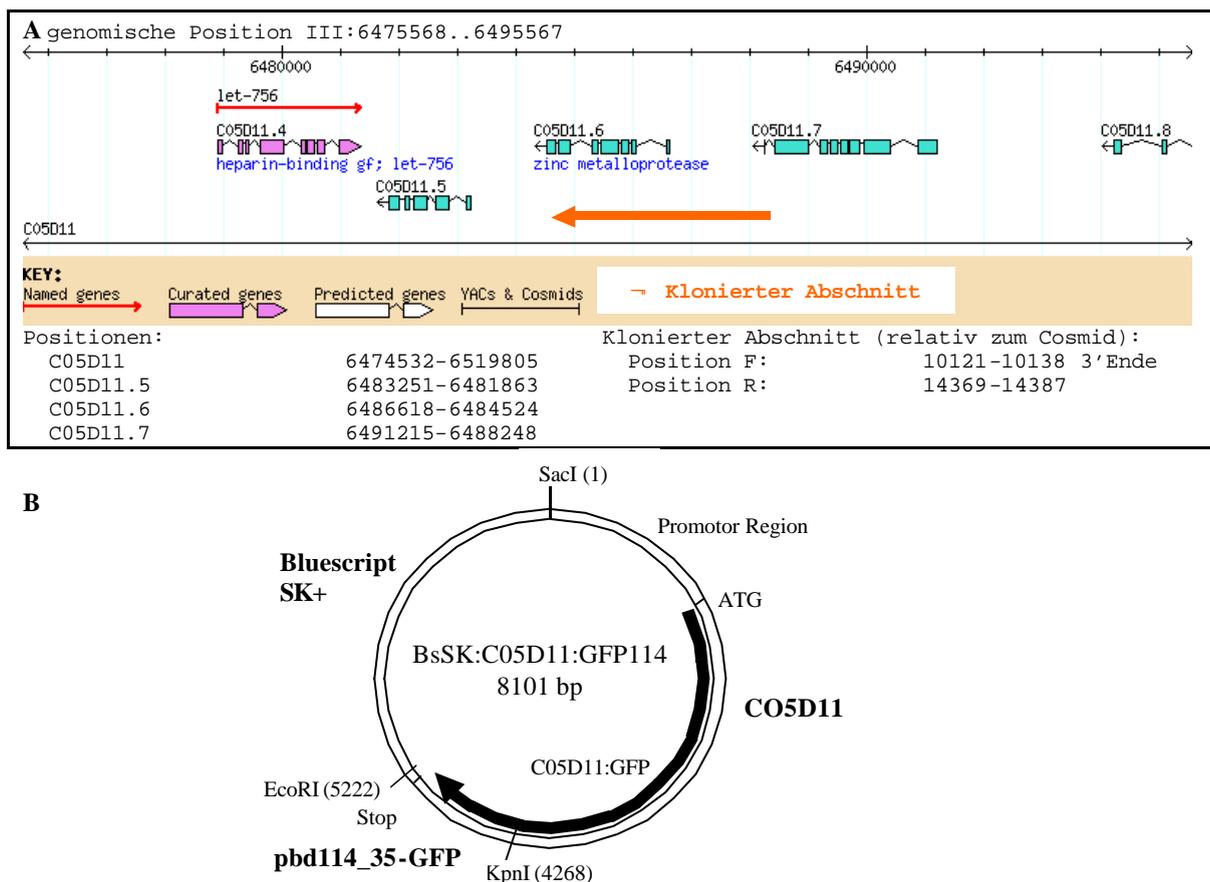


Abbildung 2: A: Genomische Position des klonierten Abschnitts.
 B: Bluescript Vektor mit kloniertem GFP und C05D11.

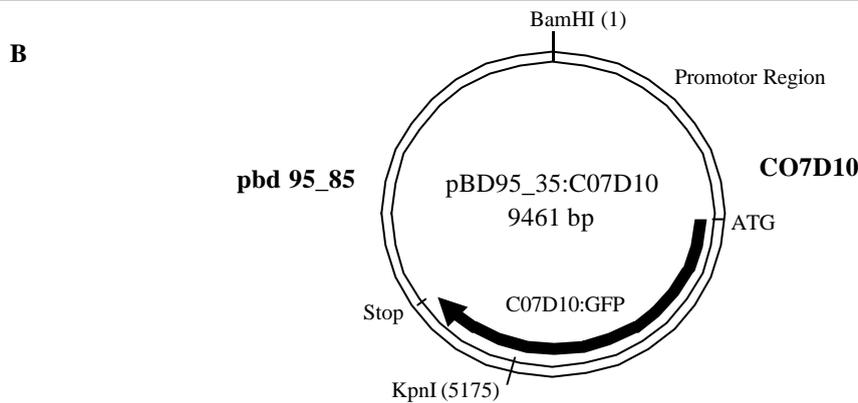
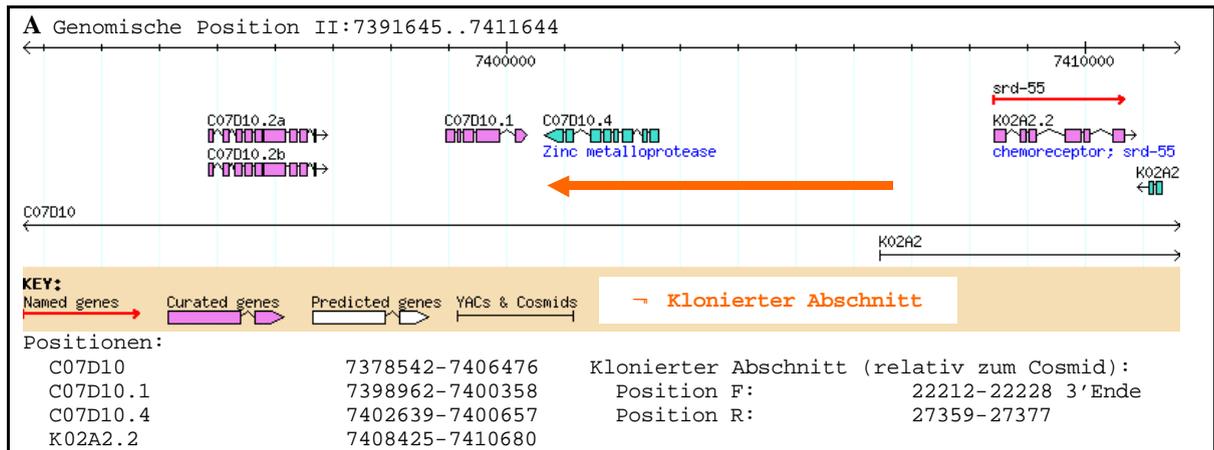


Abbildung 3: A: Position des klonierten Abschnitts. B: Vektor mit kloniertem C07D10.

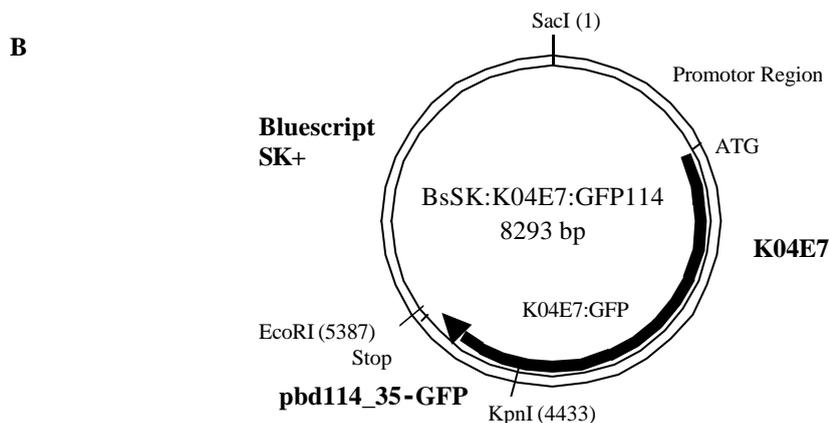
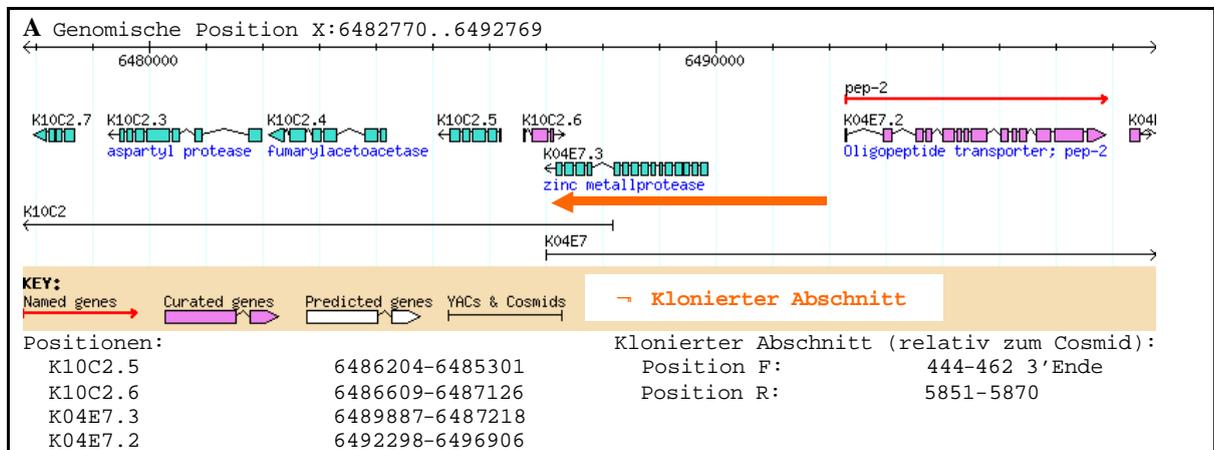


Abbildung 4: A: Position des klonierten Abschnitts. B: Vektor mit kloniertem K04E7.

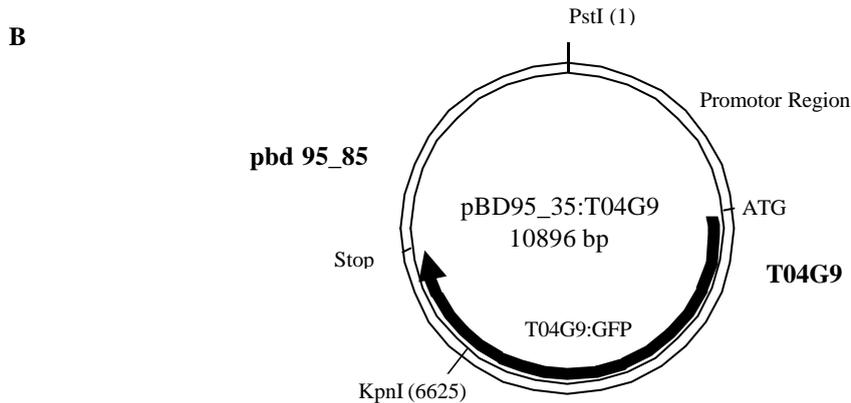
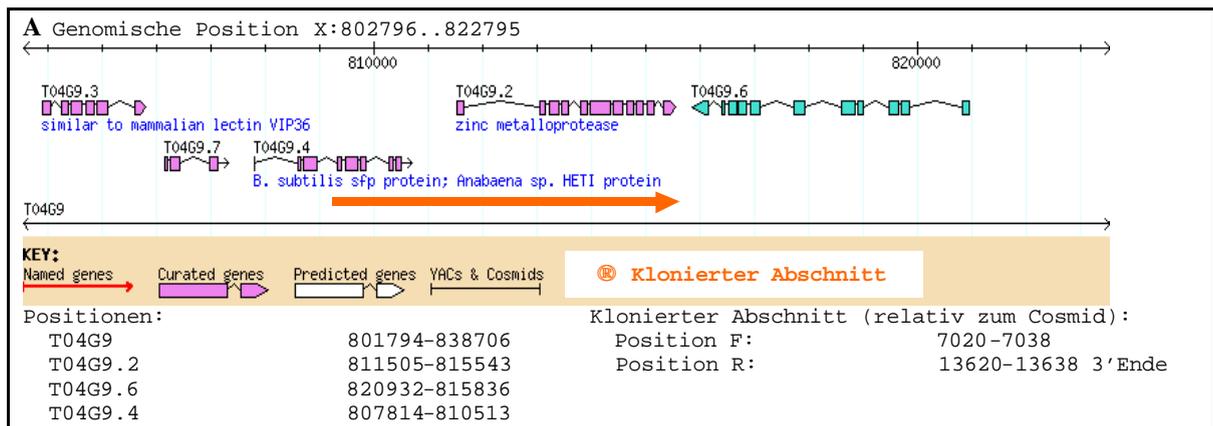


Abbildung 5: A: Genomische Position des klonierten Abschnitts.

B: Vektor mit kloniertem T04G9.

Im Falle der pBD 95_85 Vektoren (Abbildung 3 und 5) wurde der proteinkodierende Teil des jeweiligen Gens unmittelbar vor dem Stop-Codon direkt im Leseraster vor die GFP-Sequenz kloniert. Durch die beschränkte Verfügbarkeit geeigneter Restriktionsschnittstellen mussten für die Gene C05D11.6 und K04E7.3 (Abbildung 2 und 4) die GFP-Sequenz des Vektors pBD 114_35 zunächst in den bluescript Vektor umklont werden, bevor die Gensequenzen in gleicher Weise wie für pBD 95_85 eingefügt werden konnten. Durch die Lokalisation des strangaufwärts gelegenen Primers (*upper* oder *forward primer*) am Startpunkt des nächsten Gens wurde die gesamte Promoterregion in die Vektoren kloniert. Damit soll sichergestellt werden, dass die *in vivo* Expression allein unter der Kontrolle des genspezifischen Promotors erfolgt. Die schwarzen Pfeile in den Abbildungen 2 bis 5 verdeutlichen den Bereich des zu amplifizierenden Primärtranskripts (prä-mRNA). Das Start-Codon des jeweiligen Gens (ATG) und das Stop-Codon des GFP (Stop) wurden gesondert markiert. In der Darstellung ist nicht berücksichtigt, dass sowohl der proteinkodierende Teil der Gene, als auch das GFP noch Intronsequenzen enthalten die *in vivo* im Spleißapparat von *C. elegans* entfernt werden und somit letztendlich eine korrekte Translation erlauben.

Nach den Klonierungen wurden alle Plasmide durch PCR und Restriktionsanalyse sowie durch Sequenzierung überprüft.

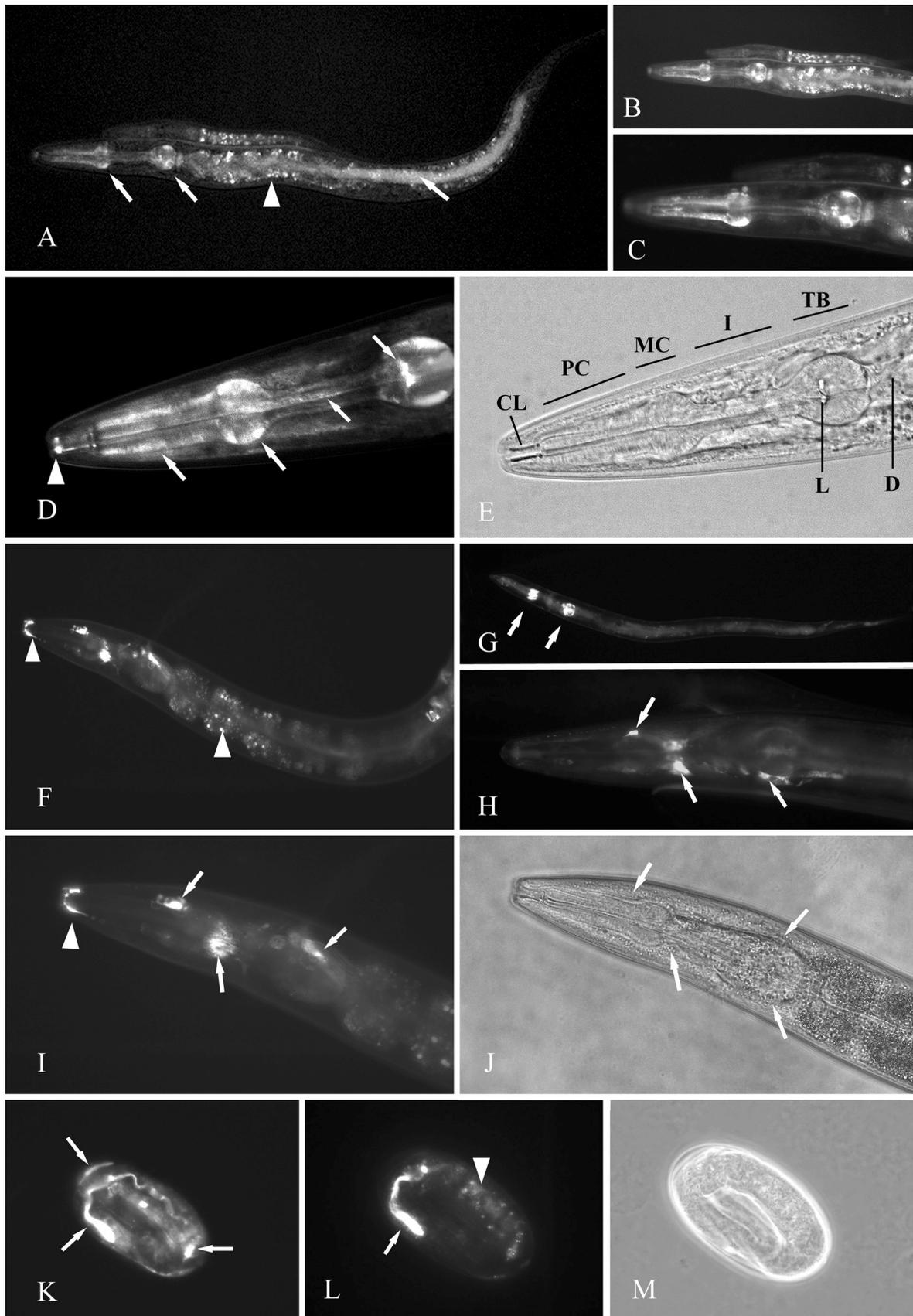


Abbildung 6: Expression der GFP-Fusionsproteine BsSK:C05D11:GFP (A-E) und pBD95_35:C07D10 (F-M) in *C. elegans*. Legende siehe nächste Seite.

Abbildung 6, vorangegangene Seite: Expression der GFP-Fusionsproteine BsSK: C05D11:GFP114 (A – E) und pBD95_35:C07D10 (F - M) in *C. elegans*. (A) GFP-Fluoreszenzsignale in adulten Hermaphroditen und L2 Larven, Vergrößerung x18; (B) und (C) Ausschnittsvergrößerung adulter Hermaphrodit und L2 Larve, x24 und x28 (D) Pharynxregion eines Hermaphroditen, die intrazelluläre GFP-Fluoreszenz zeigt, x40 (E) Lichtmikroskopische Aufnahme des in Bild (D) gezeigten Hermaphroditen. x 40; CL Cuticula Leisten, PC Procorpus, MC, Metacorpus, I Isthmus, TB terminale Verdickung (*terminal bulb*), L Lumen, D Darm. (F), (G), (H), (I) Adulte Hermaphroditen mit GFP-Fluoreszenzen außerhalb des Pharynx (Pfeile). Vergrößerung x30, x15, x30, x40 (J) Lichtmikroskopische Aufnahme des in Bild (I) gezeigten Hermaphroditen, x40 (K), (L) In Eiern kurz vor dem Schlüpfvorgang wird das Fusionsprotein C07D10:GFP in der Pharynxregion und an der Eihülle gefunden (Pfeile), x65 (M) Lichtmikroskopische Aufnahme, um den Entwicklungszustand des Eies zu demonstrieren, x65. Die Dreiecke markieren die Autofluoreszenzen der Cuticulaleisten oder des Darms.

Die Transformation der Plasmide erfolgte zusammen mit einem Koinjektionsmarker in die pha-1 Linie von *C. elegans*. Die transgene F1 Generation wurde durch Selektion des mutanten Phenotyps bei 25°C identifiziert. Bei dieser Temperatur konnten von der pha-1 Linie nur Tiere überleben, welche das Plasmid und den Coinjektionsmarker pBx (*rescue-plasmid*) als extrachromosomale DNA (*arrays*) integriert hatten. Von jeder Agaroseplatte wurde dann ein überlebendes Tier der F1 Generation weiter als eigenständige Linie kultiviert. Die fluoreszenzmikroskopische Analyse wurde ab der F2 Generation begonnen. Mit den Plasmiden BsSK:K04E7:GFP114 und pBD95_35:T04G9 konnten in keiner der transgenen Linien ein GFP-Fluoreszenzsignal beobachtet werden. Dagegen zeigten die Konstrukte BsSK:C05D11:GFP114 und pBD95_35:C07D10 die in Abbildung 6 dargestellten GFP-Fluoreszenzen.

In adulten Hermaphroditen ist das Fusionsprotein C05D11:GFP in der Pharynxregion und weiter posterior auch im gesamten Darmlumen (Abbildung 6 A, Pfeile) lokalisiert. Das Dreieck markiert die im Fluoreszenzmikroskop immer sichtbare gelb-grüne Autofluoreszenz von granulären Darmstrukturen und stellt keine auf GFP basierende Fluoreszenz dar. Frühe Entwicklungsstadien oder Larven zeigen keine Expression von C05D11:GFP. Die L2 Larve in Bild A und B lässt ausschließlich die unspezifische Darmfluoreszenz erkennen. Abbildung 6 B, C und D zeigen in höherer Vergrößerung die Lokalisierung von C05D11:GFP in der Pharynxregion eines adulten Hermaphroditen. Das Organ (Phasenkontrastaufnahme, Bild E) beginnt mit zwei starken Cuticulaleisten (CL) die unmittelbar hinter der Mundregion liegen. Daran schließt sich der Procorpus (PC), Metacorpus (MC), Isthmus (I) und die terminale Verdickung (*terminal bulb*, TB) an. Das gesamte System wird von nur 20 Muskelzellen (Zelllinie pm 5-8), 20 Neuronen (I 4-6, M1,2,5), 9 Epithelzellen und 9 spezialisierten Epithelzellen (z.B. Drüsenzellen g1 und 2) gebildet (Wood 1988, Seite 81 bis 122). Posterior geht der Pharynx direkt in den langen Mitteldarm (D) über, der mit dem Anus endet. Aufgrund seiner zellulären Verteilung (Abbildung 6 D, Pfeile) scheint das Fusionsprotein C05D11:GFP in den das Pharynxlumen umschließenden Muskelzellen der terminalen Verdickung, des Metacorpus und etwas schwächer entlang des Procorpus und Isthmus lokalisiert zu sein. Auch das Lumen (L) der terminalen Verdickung zeigt ein starkes Signal. Das Dreieck in Bild D markiert die Autofluoreszenzen der Cuticulaleisten, die entsprechend den granulären Darmstrukturen (siehe oben) keine auf GFP basierende Fluoreszenz darstellen.

Das Fusionsprotein C07D10:GFP wird sowohl in Eiern (Abbildung 6 K, L) als auch in adulten Hermaphroditen (F-I) gefunden. In Larven konnte keine GFP-Fluoreszenz beobachtet werden. Abbildung 6 G zeigt in adulten Hermaphroditen zwei sehr starke Signale ausschließlich im Bereich des Pharynx. Bei höherer Vergrößerung (Bild F, H und I) wird sichtbar, dass die C07D10:GFP Expression lateral auf den Procorpus und die terminale Verdickung beschränkt ist (Pfeile). Das Fusionsprotein ist aber nicht direkt in den Pharynxzellen lokalisiert (Bild I, J). Strukturell könnte das Protein mit dem in dieser Region liegenden Exkretionssystem (Drüsenzellen, H-Zelle, Exkretionsgang) assoziiert sein. Die Dreiecke markieren wiederum die Autofluoreszenzen der Cutikulaleisten oder des Darms. In späten Entwicklungsstadien kurz vor dem Schlüpfvorgang aus der Eihülle wird eine sehr starke Expression in Pharynxstrukturen bis in den vorderen Bereich des Darmes gefunden (Abbildung 6 K und L). Auch an der Eihülle finden sich an verschiedenen Stellen Fluoreszenzsignale. Ein möglicher Sekretionsweg von den C07D10:GFP synthetisierenden Zellen nach außen ist allerdings nicht klar erkennbar.

1.3.2. Auswertung der GFP, RNAi und *microarray* Projekte

Die Expressionskarte (*SK-map*) von Kim et al. (2001) kann der Identifizierung koregulierter Gene während verschiedener Wachstumsbedingungen, Entwicklungsstadien oder Stoffwechselwegen dienen. Die Auswertung ergab, dass 36 der insgesamt 40 Astacin-homologen Gene in sehr unterschiedlichen Koexpressions-Gruppen repräsentiert sind (Abbildung 7). Die Gene D1022.m, F54B8.m, F56A4.1 und K11G12.1 sind nicht vorhanden. In Tabelle 6 ist die Zuordnung der Gene zu den Koexpressionsgruppen (*mounts*) und gegebenenfalls zu funktionellen Gruppen aufgeführt. Dabei fällt auf, dass in keinem Fall eine Koexpression in einer spezifischen Gruppe von Genen festgestellt werden kann. In einigen Fällen kann lediglich die recht ungenaue Zuordnung zur Gruppe der nicht-proteasomalen Proteasen gefunden werden. Aussagen über die Rolle dieser Proteasen im Wechselspiel mit anderen, koexprimierten Proteinen können aber nicht getroffen werden.

Das *microarray* Projekt von Jiang et al. (2001) hatte zum Ziel für alle 20 000 postulierten *C. elegans* Gene die relativen Expressionsraten in Eiern, Larvenstadien (L1-L4), adulten Hermaphroditen und Männchen im Vergleich zu einer Gesamtpopulation zu ermitteln. 17800 Gene konnten tatsächlich auf den *microarrays* fixiert werden. Damit sollten Gene, die eine Entwicklungs- oder Sex-spezifische Rolle haben, identifiziert werden. 20 Astacin-homologe Proteine weisen dabei eine veränderte Expressionsrate in mindestens einem Entwicklungsstadium auf (Tabelle 7). Endgültige Aussagen können aber auch hier nicht getroffen werden, da die ein ganzes Genom umfassende Datenauswertung noch hohe Fehlerquoten aufweist und in vielen Fällen von anderen Experimenten abweicht. Beispielsweise postulieren Jiang und Mitarbeiter für das *hch-1* Gen eine geringere Expression in L3 Larven und adulten Hermaphroditen im Vergleich zu einer Gesamtkultur. Hishida et al. konnten aber eindeutig zeigen, dass *hch-1* ausschließlich in Eiern kurz vor dem Schlüpfvorgang und in L1 Larven exprimiert wird. Gleiches gilt für das in dieser Arbeit untersuchte Gen C07D10.4 (Tabelle 7).

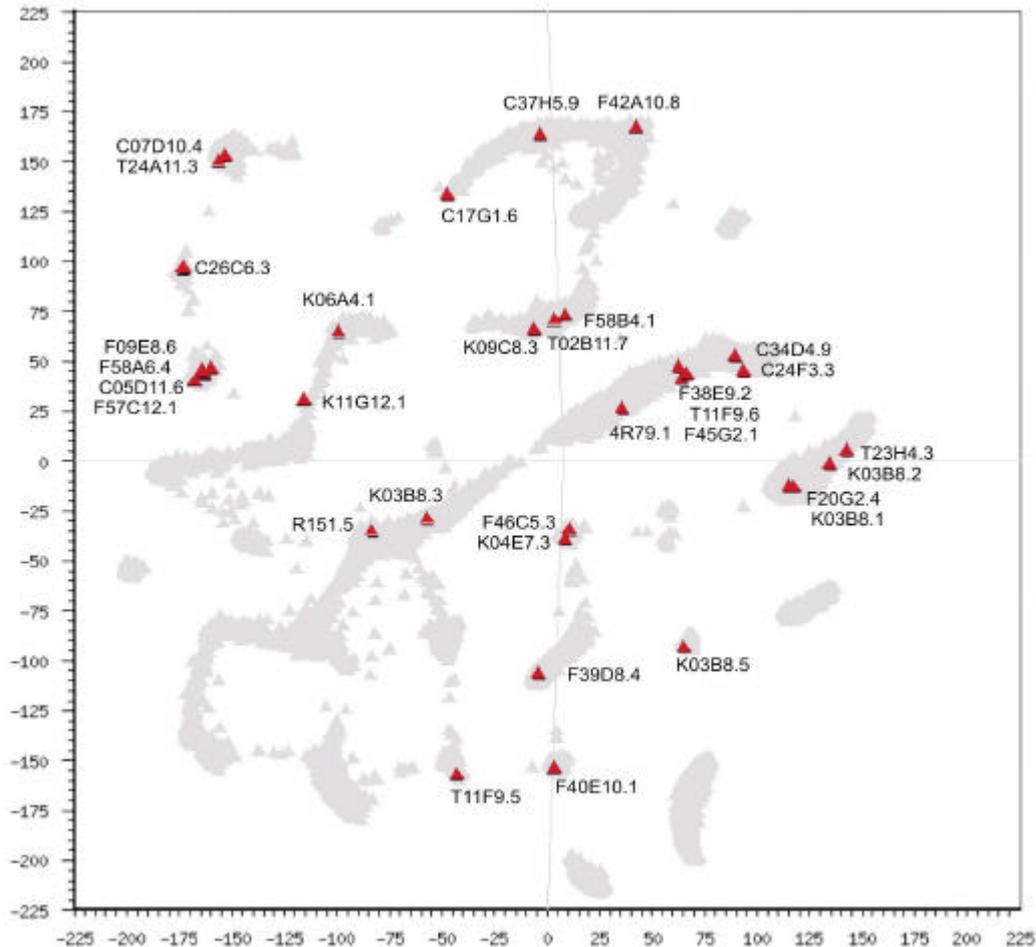


Abbildung 7: Lokalisation Astacin-homologer Gene auf der Expressionskarte (*SK-map*) von Kim et al. (2001).

Gen	SK-map koexprimiert in Gruppe, Gen-Anzahl	Gen	SK-map koexprimiert in Gruppe, Gen-Anzahl
4R79.1	mount10, R=0.19, 635	HCH-1	mount12, R=0.29, 462; Proteases, 116
C05D11.6	mount14, R=0.38, 353	K03B8.1	mount00, R=0.11, 2703
C07D10.4	mount16, R=0.40, 230	K03B8.2	mount00, R=0.11, 2703
C17G1.6	mount38, R=0.44, 8; Proteases, 116	K03B8.3	mount01, R=0.15, 1818
C24F3.3	mount03, R=0.13, 1363; Proteases, 116	K03B8.5	mount26, R=0.43, 95
C26C6.3	mount29, R=0.53, 40; Proteases, 116	K04E7.3	mount13, R=0.10, 396
C34D4.9	mount03, R=0.13, 1363	K06A4.1	mount21, R=0.30, 154
C37H5.9	mount08, R=0.21, 803	K09C8.3	mount06, R=0.21, 909
F09E8.6	mount14, R=0.38, 353; Proteases, 116	R151.5	mount01, R=0.15, 1818
F20G2.4	mount00, R=0.11, 2703; Proteases, 116	T02B11.7	mount06, R=0.21, 909
F38E9.2	mount03, R=0.13, 1363	T04G9.2	mount14, R=0.38, 353
F39D8.4	mount09, R=0.16, 786; Proteases, 116	T11F9.3	mount22, R=0.58, 151
F42A10.8	mount17, R=0.37, 210	T11F9.5	mount28, R=0.28, 61
F45G2.1	Proteases, 116	T11F9.6	mount03, R=0.13, 1363
F46C5.3	mount13, R=0.10, 396; Proteases, 116	T23F4.4	mount29, R=0.53, 40
F57C12.1	mount14, R=0.38, 353; Proteases, 116	T23H4.3	mount00, R=0.11, 2703
F58A6.4	mount14, R=0.38, 353	T24A11.3	mount16, R=0.40, 230
F58B4.1	mount06, R=0.21, 909; Proteases, 116	Y95B8A.1	mount01, R=0.15, 1818

Tabelle 6: Mit Astacin-homologen koexprimierte Gene in der Expressionskarte (*SK-map*) von Kim et al. (2001). R = Pearson Korrelation

	RNAi Projekt Maeda et al. (2001)	<i>microarray</i> Projekt Jiang et al. (2001)	GFP Analyse siehe 1.3.1
Gen	Experiment / Phenotyp	Expression im Vergleich zu Gesamtkultur (↑ höher, ↓ geringer)	Lokalisation
4R79.1		↓ L2-L4	
C05D11.6	TH:304A8 / normal	↓ Eier, L2-L4	Pharynx/Darm in adulten Hermaphroditen
C07D10.4		↑ Eier, L1-L4	Eier und Pharynxregion adulter Hermaphroditen
C17G1.6	SA:yk355c4 / 21% Defekt ¹	↑ L2-L4	
C24F3.3			
C26C6.3	JA:C26C6.3 / normal		
C34D4.9			
C37H5.9	Maeda et al. / 6% lethal	↑ L2-L4	
F09E8.6		↑ Eier, L2-L4	
F20G2.4		↓ L4	
F38E9.2			
F39D8.4			
F42A10.8			
F45G2.1			
F46C5.3		↑ Eier, L1-L4	
F57C12.1	SA:yk384c9 / normal	↑ L2-L4	
F58A6.4		↑ Eier L2 und L4	
F58B4.1		↓ L4	
HCH-1 ²		↑ <i>hatching</i> und Larven ²	<i>hatching</i> /Neuroblasten (Hishida et al. 1996)
K03B8.1		↑ ♂	
K03B8.2		↓ L4, ↑ ♂	
K03B8.3		↓ L3; ↑ Eier, L4 und Adult	
K03B8.5			
K04E7.3	SA:yk296d6 / normal		nicht detektiert
K06A4.1		↓ L2-L4	
K09C8.3			
K11G12.1	SA:yk300b8 / 100% lethal	↑ L1-L4 und Adult	
R151.5			
T02B11.7			
T04G9.2	SA:yk366f6 / normal		nicht detektiert
T11F9.3		↓ Eier, L1-L4 und Adult	
T11F9.5			
T11F9.6		↑ Adult	
T23F4.4			
T23H4.3	JA:T23H4.3 / normal	↑ Eier L2 und L4	
T24A11.3		↑ Eier L2 und L4	
Y95B8A	JA:Y95B8A_85 / normal		

Tabelle 7: Auswertung der RNAi, *microarray* und GFP Projekte. In Spalte 2 ist die Identifikationsnummer des jeweiligen RNAi-Experiments und der beobachtete Phenotyp angegeben (Fußnote ¹: Die Larven zeigen eine anormale morphologische Entwicklung in der Nähe der Vulva). Die Pfeile in Spalte 3 geben symbolisch eine höhere (↑) oder geringere (↓) Expressionsrate verschiedener Entwicklungsstadien (Eier, L1- bis L4-Larven, adulte Hermaphroditen) im Vergleich zu einer Gesamtkultur an.

Das Genom-weite RNAi Projekt von Maeda et al. (2001) soll durch das spezifische Ausschalten einzelner Gene durch doppelsträngige RNA (RNAi) weitere funktionelle Aussagen ermöglichen. Sieben der zehn bereits untersuchten Astacin-homologen Gene zeigen aber keinen auffälligen Phänotyp (Tabelle 7). Dagegen scheint das Gen C17G1.6 an der Vulva-Entwicklung in Larven beteiligt zu sein (Experiment SA:yk355c4). C37H5.9

und K11G12.1 müssen zumindest eine entscheidende Rolle in der frühen Embryonalentwicklung spielen, da 6% bzw. 100% der Embryonen nach der RNAi Behandlung absterben.

1.4. Proteom-Analyse

Zusätzlich zur Transkriptomanalyse sollte für die Astacin-homologen Gene eine Analyse auf Proteinebene durchgeführt werden. In den publizierten zweidimensionalen (2D) Proteinkarten war bisher noch keines dieser Proteine beschrieben worden (Bini et al. 1997; Kaji et al. 2000; Schrimpf et al. 2001). Dies war insofern überraschend, da insgesamt bereits über 400 Proteine in den verschiedenen 2D-Proteingelen sequenziert wurden. Da nicht zu erwarten war durch einfache Sequenzierung weiterer Spots auch nur einige Astacin-homologe Proteine im Proteom von *C. elegans* zu identifizieren, sollten deshalb spezifischere Methoden angewandt werden.

Als erster Ansatz wurde die Detektion verschiedener Astacin-homologer Gene durch Antikörper auf einem zweidimensionalen Immunoblot gewählt. Die Antiseren wurden durch Immunisierung von Kaninchen mit den synthetischen Peptiden CePepZinc_C07D10.6 und CePepZinc_K04E7.3 gewonnen. Die Sequenz dieser Peptide umfasst dabei die konservierte Zinkbindungsregion der Protease-Domäne und ist im aktiven Zentrum des Moleküls lokalisiert.

Für die Auswertung der Immunoblot Experimente sollte es hilfreich sein die ungefähre Lokalisation der gesuchten Proteine auf einem 2D-Gel zu kennen. Zu diesem Zweck wurden die theoretischen isoelektrischen Punkte und Molekulargewichte der postulierten Protein-Sequenzen berechnet (Tabelle 8) und ein simuliertes 2D-Experiment mit den 40 Astacin-homologen Proteinen durchgeführt (Abbildung 8). Die Mehrzahl der Proteine findet sich bei dieser theoretischen Platzierung in einem Feld zwischen 45-75 kDA bei einem Isoelektrischen Punkt von 6,5 bis 9.

Die Immunodetektion mit dem CePepZinc_C07D10.6 Antiserum (Abbildung 9 A) ergab ungefähr 30 deutliche ECL (*Enhanced Chemiluminescence*) Signale auf dem Röntgenfilm, während mit dem CePepZinc_K04E7.3 Antiserum (Abbildung 9 B) ungefähr 25 Signale detektiert werden konnten. Mit Hilfe des MELANIE II (*Medical Electrophoresis Analysis Interactive Expert System*) Programms wurde nun versucht, alle erhaltenen Signale dem 2D-Referenzgel von Bini et al. (1997) zuzuordnen (Abbildung 10). Dies gelang für 26 Proteine der Anti-CePepZinc_C07D10.6 Reaktion (Abbildung 9A und 10A, Pfeile) und für 22 Proteine der Anti-CePepZinc_K04E7.3 Reaktion (Abbildung 9B und 10B, Pfeile). Die Signale 3, 15 und 20 des Anti-CePepZinc_C07D10.6 Antikörpers wurden dabei auch von dem Anti-CePepZinc_K04E7.3 Antikörper erkannt (Signale 11,17 und 19). Bereits auf dieser Stufe musste festgestellt werden, dass beide Antiseren nicht allein die Sequenz der zur Immunisierung verwendeten Peptide CePepZinc_C07D10.6 und CePepZinc_K04E7.3 erkennen können. Vielmehr schienen die Antikörper auch an andere, vielleicht ähnliche Epitope auf anderen Proteinen zu binden. Die vorangegangene Transkriptom- und Datenbankanalyse hatte gezeigt, dass das hochkonservierte Zinkbindungsmotiv (HExxHxxGFxHExxRxDRD) aber in 38 weiteren Astacin-homologen Proteinen und mit einigen Aminosäureaustauschen in 17 anderen Metalloproteasen vorkommt.

Protein	IP	MW	Protein	IP	MW
C_4R79.1i	6,94	39277,63	C_F58B4.1	8,04	70066,33
C_C05D11.6i	6,89	39830,50	C_HCH1	8,02	66259,16
C_C07D10.4i	8,22	45547,02	C_K03B8.1	8,63	51507,96
C_C17G1.6i	8,73	85179,71	C_K03B8.2	7,81	50992,62
C_C24F3.3	8,19	45740,35	C_K03B8.3	8,46	46270,84
C_C26C6.3	5,91	73372,50	C_K03B8.5	7,19	54848,77
C_C34D4.9i	5,85	46095,16	C_K04E7.3	8,65	61672,96
C_C37H5.9i	8,62	60977,22	C_K06A4.1i	9,73	34400,20
C_D1022.m	9,13	44140,70	C_K09C8.3	9,31	63041,13
C_F09E8.6i	6,58	56911,27	C_K11G12.1	8,55	66096,01
C_F20G2.4i	8,24	45161,84	C_R151.5	8,28	69488,88
C_F38E9.2n	6,20	105287,84	C_T02B11.7i	8,44	74176,86
C_F39D8.4i	6,79	50760,09	C_T04G9.2i	8,11	62830,61
C_F42A10.8i	6,91	55787,05	C_T11F9.3i	5,51	43377,66
C_F45G2.1o	8,87	30756,03	C_T11F9.5i	6,67	37474,51
C_F46C5.3n	5,90	45456,20	C_T11F9.6i	5,12	68146,29
C_F54B8.m	6,02	10222,07	C_T23F4.4	9,13	53291,28
C_F56A4.1	9,62	33765,86	C_T23H4.3	7,92	40772,43
C_F57C12.1i	6,31	81023,43	C_T24A11.3i	8,10	47170,45
C_F58A6.4i	7,80	60394,59	C_Y95B8A.1	9,33	84191,38

Tabelle 8: Theoretischer isoelektrischer Punkt (IP) und Molekulargewicht (MW) aller Astacin-homologer Proteine in *C. elegans*.

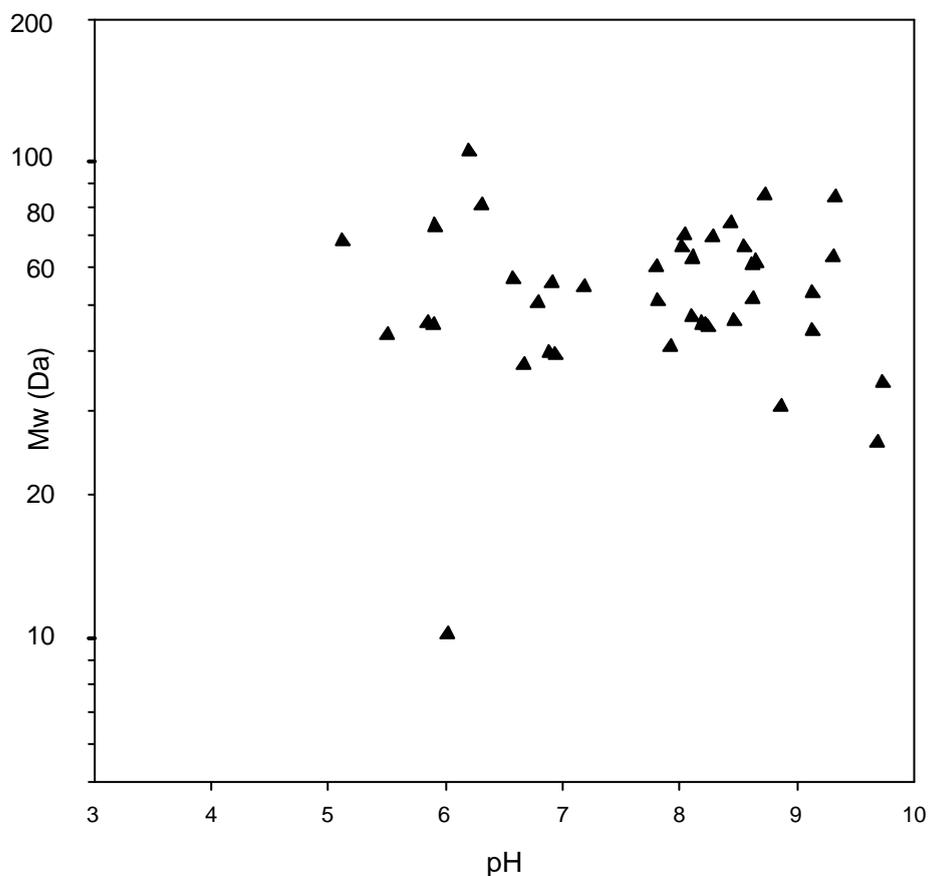


Abbildung 8: Simuliertes 2D-Experiment mit den postulierten Astacin-homologen Proteinen von *C. elegans*. Auf der x-Achse wurde der isoelektrische Punkt (pH) und auf der y-Achse das relative Molekulargewicht (Mw) der theoretischen Proteine aufgetragen.

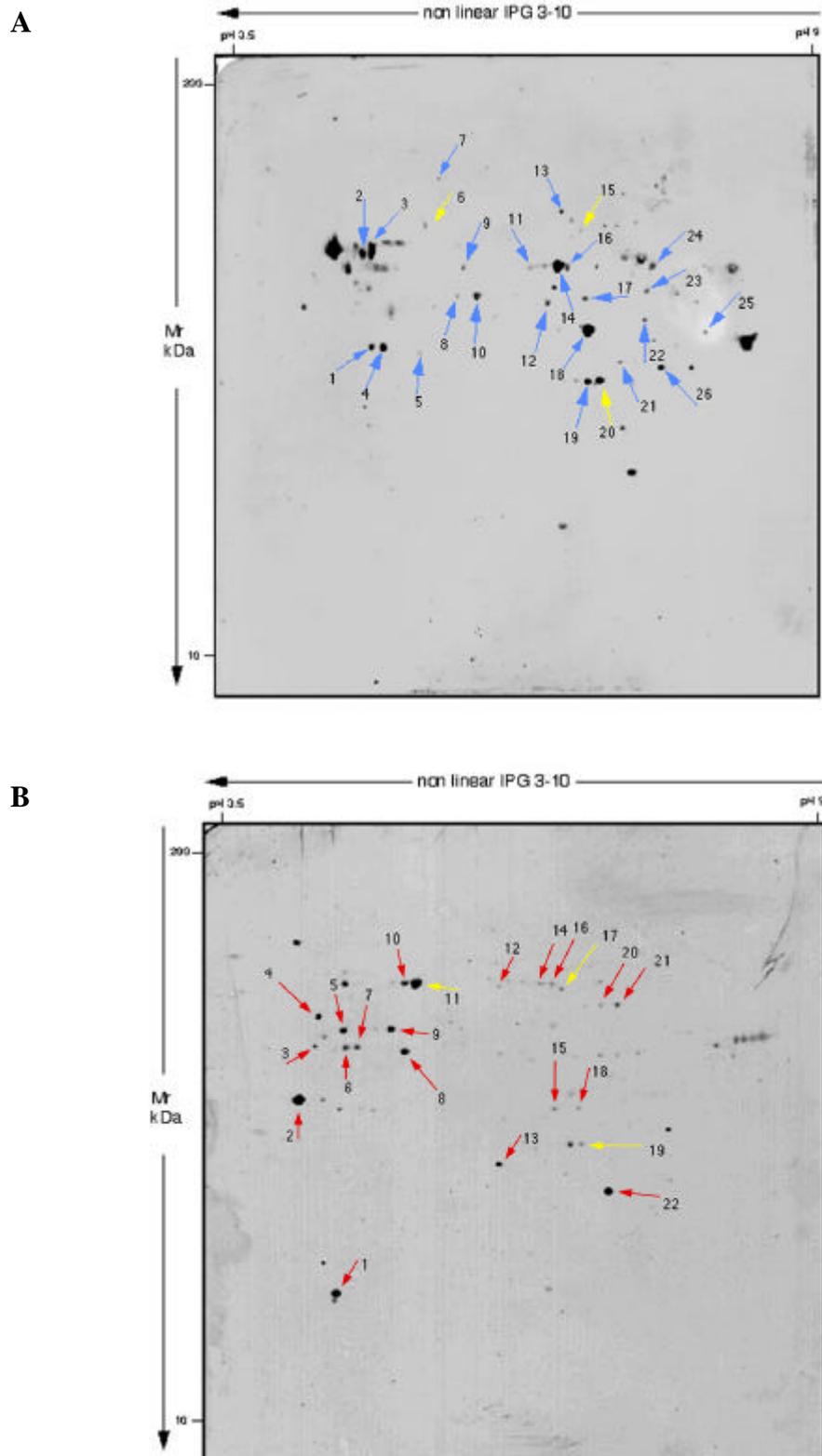


Abbildung 9: 2D-Immunoblot mit dem Anti-CePepZinc_C07D10.6 (**A**) und CePepZinc_K04E7.3 (**B**) Antiserum. Die Signale wurden über eine Chemolumineszenz-Reaktion (ECL, *Enhanced Chemioluminescence*) auf Röntgenfilmen sichtbar gemacht. Die Pfeile und Nummern markieren die Proteine, welche dem Referenzgel (Abbildung 10) zugeordnet werden konnten. Die unterschiedliche Farbgebung der Pfeile markiert Proteine, die ausschließlich von dem Anti-CePepZinc_C07D10.6 (blau) oder CePepZinc_K04E7.3 (rot) Antikörper, sowie von beiden (gelb) detektiert wurden.

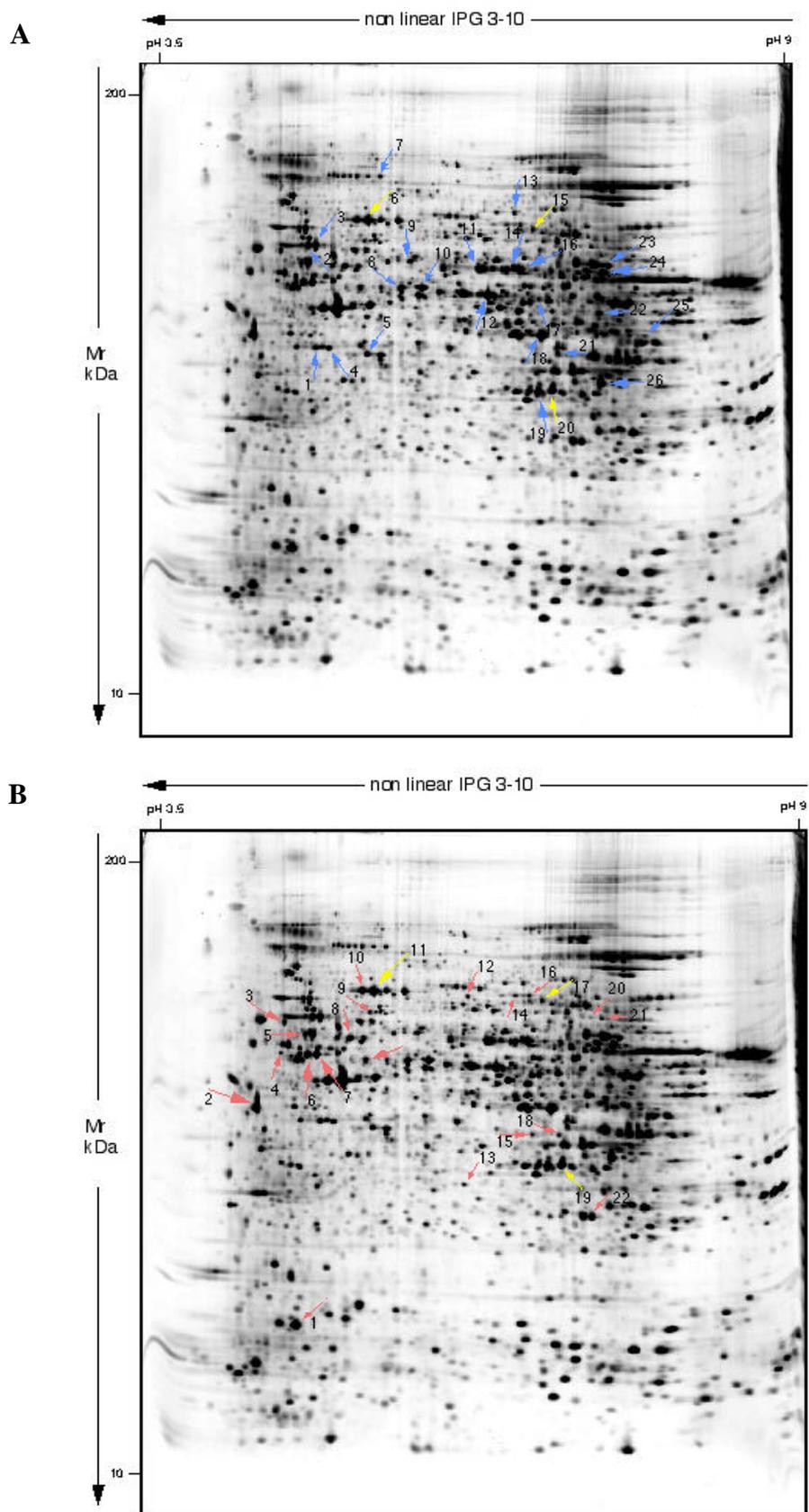


Abbildung 10: Silber-gefärbtes 2D-Referenzgel von *C. elegans* Gesamtproteinextrakt (Bini et al., 1997; abrufbar unter <http://www.bio-mol.unisi.it/2d/2d.html>). Bild (A) zeigt die Lokalisation der durch immunoblotting mit dem Anti-CePepZinc_C07D10.6 Antiserum identifizierten Proteine (Nummerierung und Pfeile); Bild (B) desgl., Antiserum: Anti-CePepZinc_K04E7.3. Farbgebung der Pfeile siehe Abbildung 9.

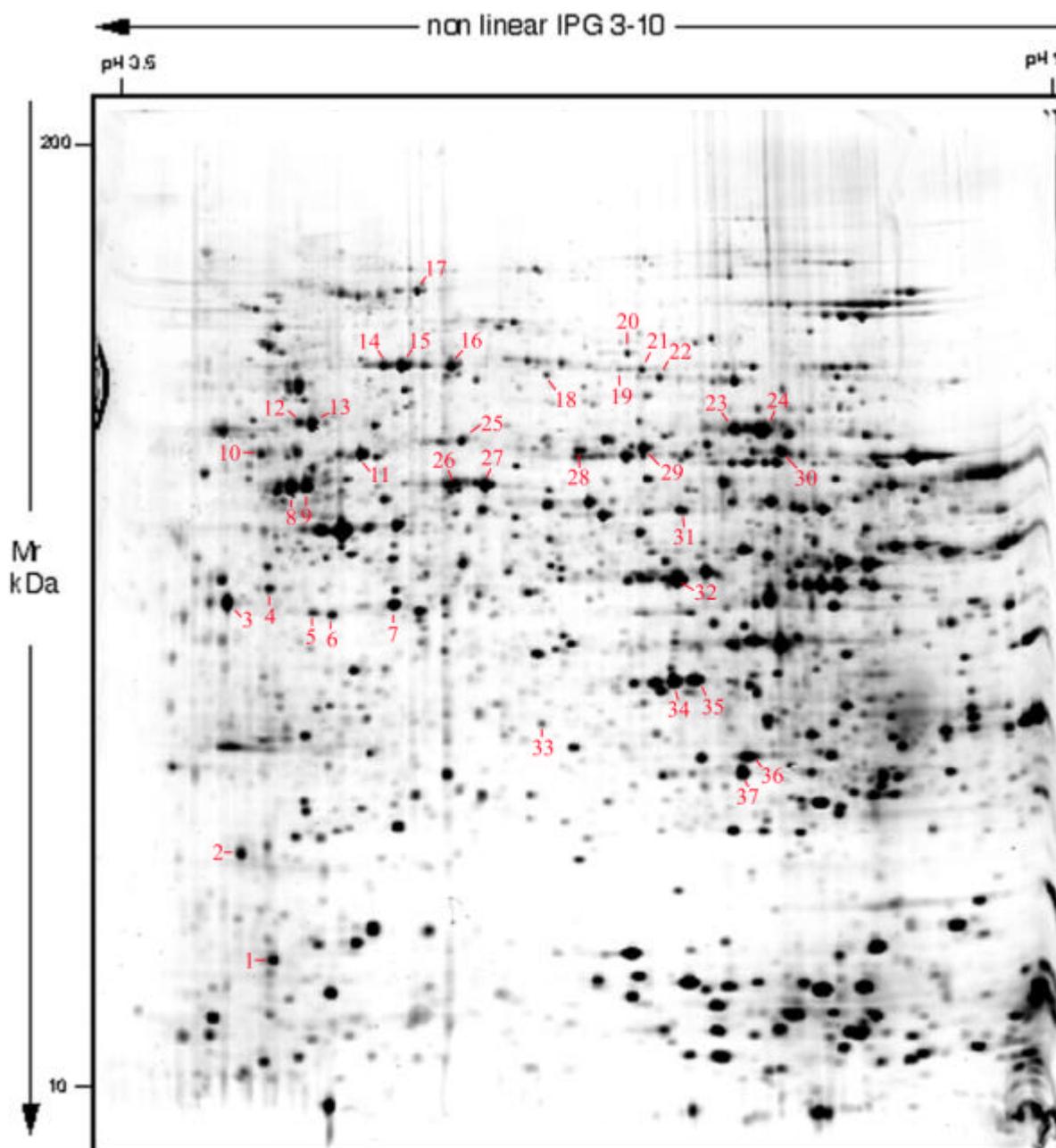


Abbildung 11: Coomassie-gefärbtes 2D-Gel von *C. elegans* Gesamtproteinextrakt. Die 37 markierten Proteine sind über Massenspektrometrie oder Edman-Sequenzierung analysiert worden (siehe Tabelle 9).

Spot Nr.	Detektion	Blot	Protein	Funktion
1	Edman (blockiert)	K1		
2	MS		F25H2.11 ²	Translationally controlled tumor protein
3	MS	K2	T07C4.9	nex-2, Annexin
4	MS		C47E12.4 ^{1,2}	inorganic pyrophosphatase
5	Edman (blockiert) / MS	C1	C18A11.7 ^{1,2}	dim-1, immunoglobulin family
6	Edman (PDGKAPXFXQ) / MS	C4	C18A11.7 ^{1,2}	dim-1, immunoglobulin family
7	Edman (blockiert) / MS		C07D8.6 ²	aldehyde reductase
8	MS	K6	C34E10.6 ²	ATP synthase beta chain
9	Edman (AAAATKVS AK)	K7	C34E10.6 ²	ATP synthase beta chain
10	MS	K3	K01G5.7 ²	tubulin beta-chain
11	Edman (blockiert)	K8		
12	-	C2		
13	-	C3		
14	MS	K10	F26D10.3 ²	hsp-1, heat shock 70 protein A
15	MS	C6 / K11	F26D10.3 ²	hsp-1, heat shock 70 protein A
16	MS		C37H5.8 ^{1,2}	hsp-6, heat shock 70 protein
17	-	C7		
18	MS	K12	F22B5.9 ²	frs-2, phenylalanyl-tRNA synthetase
19	-	K14		
20	MS	C13	F32B6.2	carbamoyl-phosphate carboxylase
21	-	K16		
22	MS	C15 / K17	C03G5.1 ²	succinate dehydrogenase
23	-			
24	Edman (AATLDAHSQ)	C16	ZK829.4 ^{1,2}	Glu/Leu/Phe/Val dehydrogenase
25		C9		
26		C8		
27	Edman (PITKI HARQIY)	C10	T21B10.2 ^{1,2,3}	enolase
28	-	C11		
29	-	C14		
30	MS	C24	LLC1.3 ²	dihydrolipoamide dehydrogenase
31	MS (schlechtes Spektrum)	C17		
32	Edman (blockiert) / MS	C18	F46H5.3 ^{1,2}	arginine kinase
33	MS	K13	F39H11.5	pbs-7, Proteasome A- and B-type
34	Edman (blockiert) / MS	C19		
35	Edman (blockiert) / MS	C20 / K19	W09H1.6 ^{1,2}	lec-1, galactoside-binding lectins
36	Edman (XXKFFVGGNX)		F42F12.2	zig-2, IG-like C2-type domains
37	Edman (TRKFFVGNW) / MS		Y17G7B.7 ²	tpi-1, Triosephosphate isomerase

Tabelle 9: Übersicht der in dieser Arbeit nachgewiesenen Proteine im *C. elegans* Gesamtproteinextrakt. K1-19 bezeichnet die Protein-Nummer im CePepZinc_C07D10.6 und C1-20 die Nummerierung im CePepZinc_K04E7.3 Immunoblot (Abbildungen 9 und 10). Die Fußnoten markieren die Proteine, deren Identifizierung bereits auf verschiedenen 2D-Karten durch Kaji et al. (2000) (¹), Schimpf et al. (2001) (²) oder Bini et al. (1997) (³) erfolgte.

Auch der Vergleich mit dem simulierten 2D-Experiment (Abbildung 8), wonach die Proteine im Bereich von 45-75 kDA bei einem Isoelektrischen Punkt von 6,5 bis 9 liegen sollten, sprach für diese Tatsache.

Deshalb wurden eine Coomassie-gefärbte PVDF-Membran zur Edman Sequenzierung und ein Coomassie-gefärbtes Gel zur massenspektrometrischen Identifizierung der detektierten Proteinsignale angefertigt. Abbildung 11 zeigt das für die Sequenzierung verwendete 2D-Gel eines *C. elegans* Gesamtproteinextrakts. Wiederum wurde mit Hilfe der MELANIE II (*Medical Electrophoresis Analysis Interactive Expert System*) Software die detektierten Signale dem 2D-Sequenzgel zugeordnet (Abbildung 11, Nr. 1-37). Die hier verwendete Nummerierung ist in Tabelle 29 zusammen mit der Bezeichnung der Immunoblot Experimente aufgeführt. Die Edman-Sequenzierungen und die massenspektrometrischen Analysen ergaben nun in keinem Fall die Detektion eines Astacin-homologen Proteins (Tabelle 29). Im Vergleich zu den bereits publizierten 2D-Proteingelen von Bini et al. (1997), Kaji et al. (2000) und Schrimpf et al. (2001) konnten aber einige neue Proteine identifiziert werden.

Als zweiter Ansatz sollte deshalb versucht werden, die Astacin-homologen Proteine durch Affinitätschromatographie mit einem spezifischen Inhibitor aufzureinigen und die isolierten Proteine anschließend durch massenspektrometrische Analyse nachzuweisen. Für die Affinitätschromatographie stand der bereits gut charakterisierte Inhibitor Pro-Leu-Gly-Hydroxamat zur Verfügung (Moore and Spilburg 1986; Reyda et al. 1999; Köhler et al. 2000).

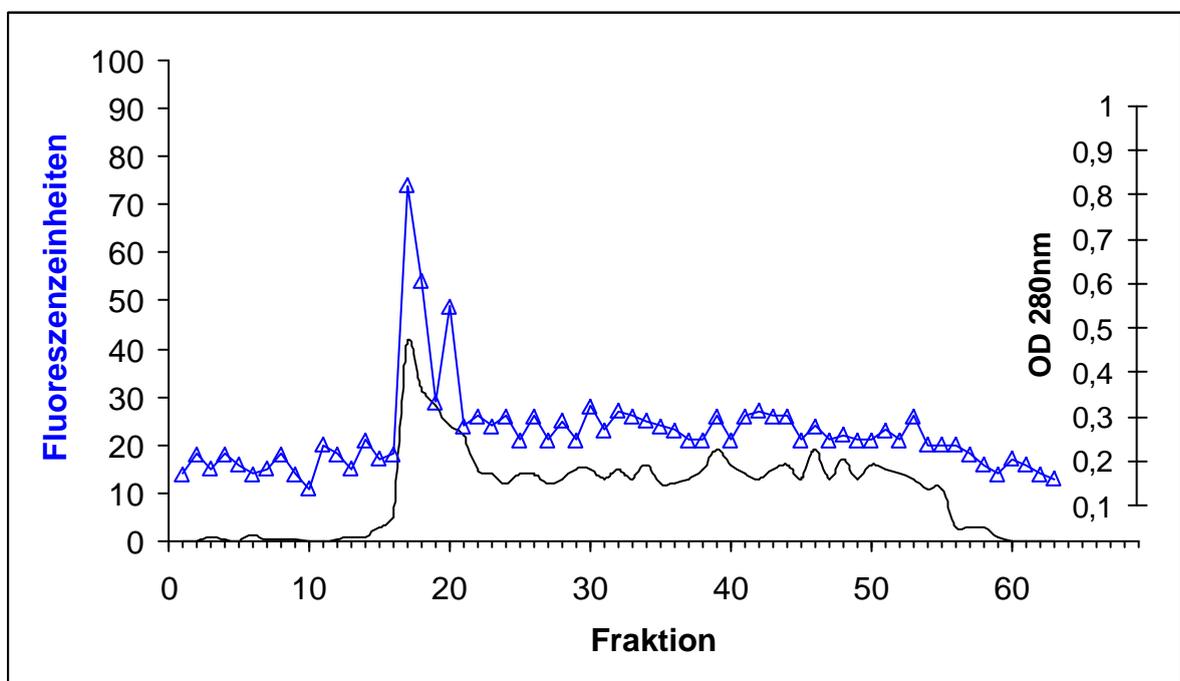


Abbildung 12: Aktivitätsmessung der durch Affinitätschromatographie mit dem Liganden Pro-Leu-Gly-NHOH aufgereinigten Enzymfraktionen gegen FITC-Casein (Symbol Δ) und Proteinkonzentration (OD_{280nm} durchgehende Linie). Die Elution mit Tris-HCl, pH 11 wurde bei Fraktion 10 begonnen.

Ein durch Ammoniumsulfat-Präzipitation vorgereinigter Protein-Gesamtextrakt wurde bei pH 8,0 auf die Säule gegeben, nachgewaschen und mit einer pH Änderung (Tris-Puffer, pH 11) wieder eluiert. Als spezifischer Enzymtest wurde das dansylierte Substrat DnsPKFAPWV (Stöcker et al. 1990; Wagner et al. 2002) verwendet. Abbildung 12 zeigt den Verlauf der Elution (Extinktion bei 280 nm) und die mit dem Substrat DnsPKFAPWV gemessene proteolytische Aktivität (relative Fluoreszenzeinheiten). Im ersten Elutionsmaximum (Fraktionen 17-21) ist eine deutliche Zunahme einer Astacin-ähnlichen proteolytischen Aktivität zu beobachten. Diese Fraktionen wurden vereinigt und biochemisch weiter charakterisiert (Tabelle 10). Die Aktivitätsmessung erfolgte wiederum mit dem Substrat DnsPKFAPWV. Die Reaktion konnte durch den Einsatz des allgemeinen Metalloprotease-Inhibitors 1, 10- Phenantrolin fast vollständig gehemmt werden.

	relative Fluoreszenzeinheiten
Durchlauf	21,95
Fraktion 17-21	84,21
Phenantrolin Hemmung	20,31

Tabelle 10: Fluoreszenzintensitäten der vereinigten Fraktionen (17-21) gemessen mit dem DnsPKFAPWV Substrat und biochemische Charakterisierung mit den Inhibitoren 1, 10 Phenantrolin sowie Pro-Leu-Gly-Hydroxamat.

Die vereinigten Fraktionen wurden anschließend direkt auf einem zweidimensionalen Gel aufgetrennt und die Proteine mit Coomassie Farbstoff sichtbar gemacht (Abbildung 13). Dabei fällt auf, dass von der Affinitätssäule noch eine große Zahl verschiedener Proteine eluiert wurden. Dies könnte wiederum mit dem Vorhandensein der 40 Astacin-homologen Proteine erklärt werden. Durch die massenspektrometrische Analyse von 6 deutlichen Spots konnte aber kein Astacin-homologes Protein identifiziert werden. Vielmehr wurden zwei proteasomale Metalloproteasen, sowie eine Hydratase, Dehydrogenase oder Isomerase (Tabelle 11) gefunden. Für die mit dem Substrat DnsPKFAPWV gemessene proteolytische Aktivität könnten also möglicherweise die proteasomalen Proteasen oder andere hier nicht identifizierte Proteasen verantwortlich sein. Auf eine weitere Untersuchung der Proteine wurde allerdings verzichtet, da die affinitätschromatographische Aufreinigung alleine nun nicht mehr geeignet erschien, Astacin-homologe Proteine mit einem einfachen Verfahren anzureichern.

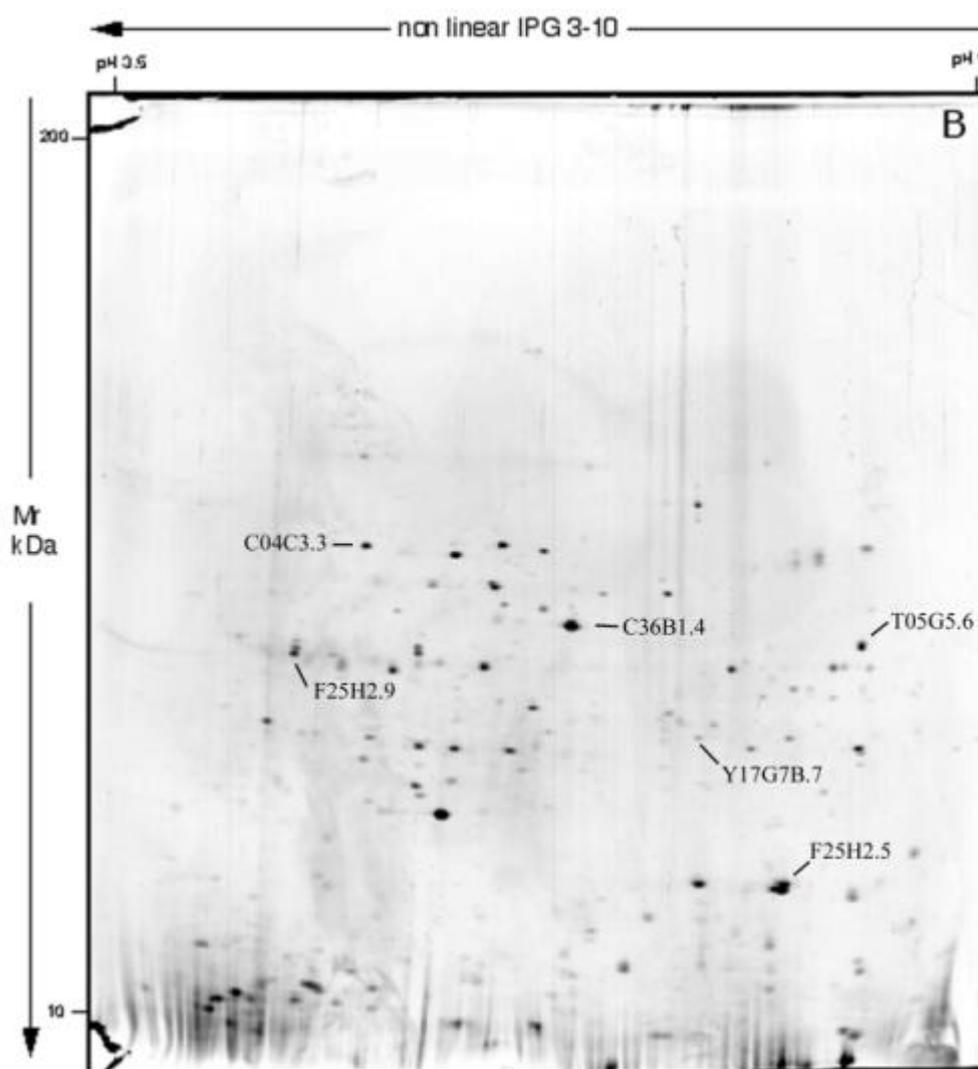


Abbildung 13: Coomassie-gefärbtes 2D-Gel der mit Pro-Leu-Gly-Hydroxamat affinitätsgereinigten Proteinen. Die über Massenspektrometrie identifizierten Proteine sind mit ihren *wormpep 73* Nummern im Gel verzeichnet.

Spot Nr.	Gen	Protein
1	C04C3.3	pyruvate dehydrogenase
2	F25H2.9	pas-5, proteasome zeta chain (Multispecific protease of the proteasome)
3	C36B1.4	pas-4, proteasome A-type submit (Multispecific protease of the proteasome)
4	T05G5.6	Enoyl-CoA hydratase
5	Y17G7B.7	tpi-1, Triosephosphate isomerase
6	F25H2.5	nucleoside diphosphate kinase

Tabelle 11: Übersicht der mit Pro-Leu-Gly-Hydroxamat affinitätsgereinigten Proteine.

2. Astacin-homologe Proteine in *Hydractinia echinata*

Aus den Basisgruppen des Tierreichs (Porifera, Ctenophora, Cnidaria) lagen bisher nur wenige Daten über Astacin-homologe Proteine vor. Da mit Flavastacin (Tarentino et al. 1995) aber auch schon ein bakterieller Vertreter gefunden wurde, war es für die Analyse der molekularen Evolution der Protein-Familie wichtig, hier weitere vergleichende Informationen zu erhalten. In dieser Arbeit sollten deshalb neue Astacin-homologe Gene bei dem Modellorganismus *Hydractinia echinata* (Cnidaria, Hydrozoa) näher charakterisiert werden.

2.1. RT-PCR/RACE

Ausgangspunkt dieser Untersuchungen waren zwei kurze Genfragmente, die von Prof G. Plickert (Zoologisches Institut, Universität Köln) in einem *gene expression screening* als Astacin-homologe Sequenzen identifiziert worden sind. Diese Gene werden im folgenden als *Hydractinia echinata* Astacin 1 und 2 (HEA-1, HEA-2) bezeichnet. Anhand dieser Sequenzinformationen wurden in einem ersten Schritt genspezifische Oligonukleotide hergestellt. Damit wurde über eine Reverse Transkription (RT) und Polymerase-Ketten-Reaktion (PCR) die entsprechende DNA vervielfältigt (Abbildung 14, rechter Teil). Nach einer Gelelektion wurden die Produkte direkt mit glatten Enden (*blunt end*) in den pUC 18 Vektor oder mit A/T Überhängen in den pGEM-T Vektor kloniert.

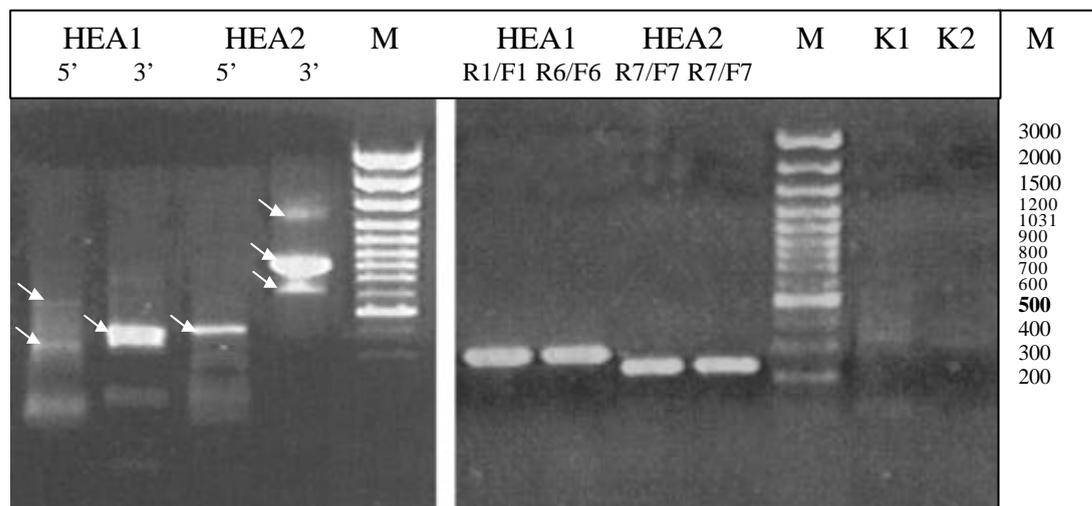


Abbildung 14: 1,5 %iges, Ethidiumbromid-gefärbtes Agarosegel mit DNA-Fragmenten aus den RACE/PCR Experimenten. HEA1 5': 5' RACE mit UAP und R6 *primern*; HEA1 3': 3' RACE mit UAP und F6 *primern*; HEA2 5': 5' RACE mit UAP und R7 *primern*; HEA2 3': 3' RACE mit UAP und F7 *primern*; HEA1 und HEA2 R1/F1 R6/F6 R7/F7: PCR Fragmente mit den angegebenen *primer* Kombinationen; M Marker DNA; K1 und K2: Kontroll-PCR mit R6/F6 bzw. R7/F7 ohne *template*; Pfeile: klonierte DNA-Fragmente der RACE Experimente.

Klon3 Bezeichnung HEA1-R6-5'RACE; Vektorsequenz kursiv
RACE Fragment SmaI kloniert in pUC 18

GTAAAACGACGGCCAGTGC^{M13}CAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGAT^{SmaI}CCCCCTAATACGACTC^{UPM long}
 ACTATAGGGCAAGCAGTGGTAACAACGCAGAGTACGGGGGACTGAAAAATAAGATGGTTGATGCAAGAAA
 TTAATTTAAATATGCAATCCGTTGAGTGTCTTCTGCTTGTGTGCGTTCTTTACCTTTTATACGTGTTATGTA
 TTTCTAAAATTTCTAAAATCAAATTTTTACTACCCCTTATTACTTTGGTGTGTGCTTCAAACAACAT
 CTTCTGAGAGCAAGGTATAAATGTAGGTAACATGTTTTTTATTTAATTGTCTGTTTTGAAAACCTCGAC
 GTTCAACATGACTAACTTTACTTTCTTGGTTTTATGCGCGGTTGTGGGTGTGCGTATGAGCCAGAGGTTG
 TGGCCGAATGGTAACGTTCCATACGTGAGCACTATCAAGGCTGTAAACGCAATCAACGAAGCTTTGCAAG
 CAATCGAGGCTAACACCAAATGCATTACATTCACACAAAAGACAGCCGCTGACAAGAATTCATTGAATT
 CAAACATGGAGGAGGATGTTACTCCTACGTTGGTATGATGAGGCGTACAAACCCGATTTATCTTTACCCA
 CCTTGTTACCGAAGCAAGGATATTGTATGCATGAAGTAGGACACGCCCTTAGGTTTGAACCAACATCCTG
 GTCAAAAGCCATCGTCAGCTCCACTTGGCAAAGCGCGGCCGCTAGGGTACCGAGCTCGAATTCGTAATCA
TGGTCATAGCTGTTTCTCTG^{M13 revers}

Abbildung 17: Nukleinsäuresequenz des Klons 3 (HEA1-R6-5'RACE)

Klon4 Bezeichnung HEA1-F6-3'RACE; Vektorsequenz kursiv
RACE Fragment SmaI kloniert in pUC 18

CAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCCAACTGCAGCAAGGCTGCAAACGCAA^{M13 revers}
^{SmaI}HEA1/PstI/F6
 TCAACGAAGCTTTTGAAGCAATCGAGGCTAACACCAAATGCATCACATTCACACAAAAGACAGCCGCTGA
 CAAGAAGCTTCAATGAAATCAAACATGGAGGAGGATGTTACTCCTACGTTGGTATGATGAGGCGTACAAAC
 CCGATTTATCTTTACACACCTTGTACCAGCAAGGATATGTCATGCATGAAGTAGGACACGCCTTAG
 GTTTTGAACCAACATCCTGGTCAAAAAGCCATCGTCAGCTCCACTTGGCAAAGCTGATTTTGAATAATTA
 CAAAATGTATACATGTTAAGAAAAAAGAAAAATCTTTTCCCTAAAAAAGAAAAAAGAAAAAAGAAAAA
 AAGTACTCTGCGTTGTTACCACTGCTTGCCCTATAGTGAGTCGTATTAGGGGATCCTCTAGAGTCGACC
^{SmaI}
^{UPM long}
TGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTAC^{M13}

Abbildung 18: Nukleinsäuresequenz des Klons 4 (HEA1-F6-3'RACE)

Klon5 Bezeichnung: HEA2-GP-52; kursiv: Vektor Sequenz
PCR Fragment mit Adaptoren kloniert in pGEM-T

ATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTCGACC
^{SP6 promotor}
TGCAGGCGGCCGCACTAGTGATTACCGACGTCGACTATCCATGAACAATCCAGACGAGAAGGAGGGCATA
^{Adaptor}
 TGAAAACTGGAAATGCTTTTGCCAGCATAAAAAGGTGGAAGATGGCCCGGTGCAAAAATTCATATGTGCG
 TTGAGGGGTCAATCAGTAGCAGAGGGATTTACGCTATTAATGCAGCCATTGCTGATTATCACAAATTCAC
 CTGCTTGAGATTTACAGGAGAACTAATGAGAGAGCTCATATCAGCTTCTATAAAGGAGGCGGCTGCTCT
 TCACCAGTTGGTTATCGATCTGTTTCATGGATAGTCGACGTCGGTAATCCCGCGCCATGGCGGCCGGGAG
^{Adaptor}
CATGCGACGTCGGGCCAATTCGCCCTATAGTGAGTCGATTA
^{17 promotor}

Abbildung 19: Nukleinsäuresequenz des Klons 5 (HEA2-GP-52)

Klon6 Bezeichnung HEA2-R7-5'RACE-7.5; Vektorsequenz kursiv
RACE Fragment SmaI kloniert in pUC 18

CAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCCTAATACGACTCACTATAGGGCAAG
^{M13 revers}
^{SmaI}
^{UAP long}
 CAGTGGTAACAACGCAGAGTACGCGGGACATTAAGTGTTCAAAACAGACACCACAATGAATTCGATTACT
 GTCGCATTACTGTTGCTGGCTTCTTGCTTATCTGTACATGTGTTTGGGTAAAAGAAATGGAAAACCCCTG
 GTGTTTTTTGAGGGAGATATGGTTTTGGATCCAGACGAGAAGGAGGGCATATGGAAAACGGAAATGCTTT
 TGCCAGCATAAAAAGGTGGAAGATGGCCCGGTGCAAAAATTCATATGTCGTTGAGGGTCAATCAGTAGC
 AGAGGGATTTACGCTATTAATGCAGCCATTGCTGATTATCACAAATTCACCTGCTTGAGATTTACAGGA
 GAACTAATGAGAGAGCTCATATCAGCTTCTATAAAGGAGGCGGCTGCTTTCACGCGGCCGCTAGGGGAT
^{SmaI}
^{HEA2/NotI/R7}
CCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTAC
^{M13}

Abbildung 20: Nukleinsäuresequenz des Klons 6 (HEA2-R7-5'RACE-7.5)

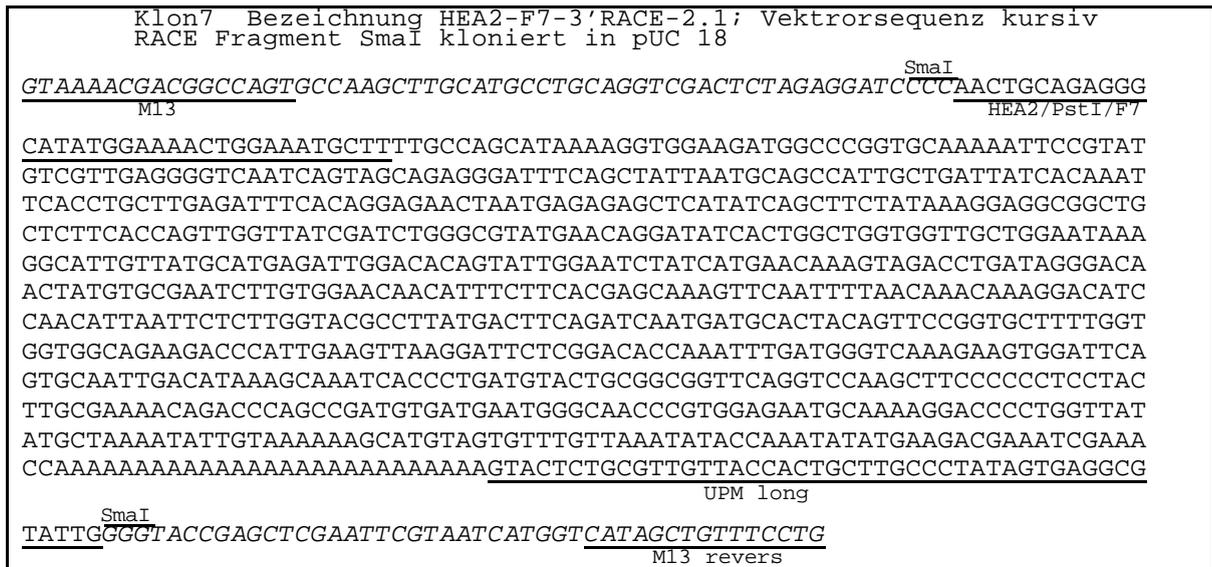


Abbildung 21: Nukleinsäuresequenz des Klons 7 (HEA2-F7-3'RACE-2.1)

Abbildung 22 und 23 zeigen als Übersicht die vollständige DNA- und abgeleitete Proteinsequenz von HEA-1 und HEA-2. Die cDNA Sequenz beginnt mit dem ersten Nukleotid nach dem 5'-RACE und endet mit dem letzten Nukleotid vor dem 3'-RACE adaptor primer.

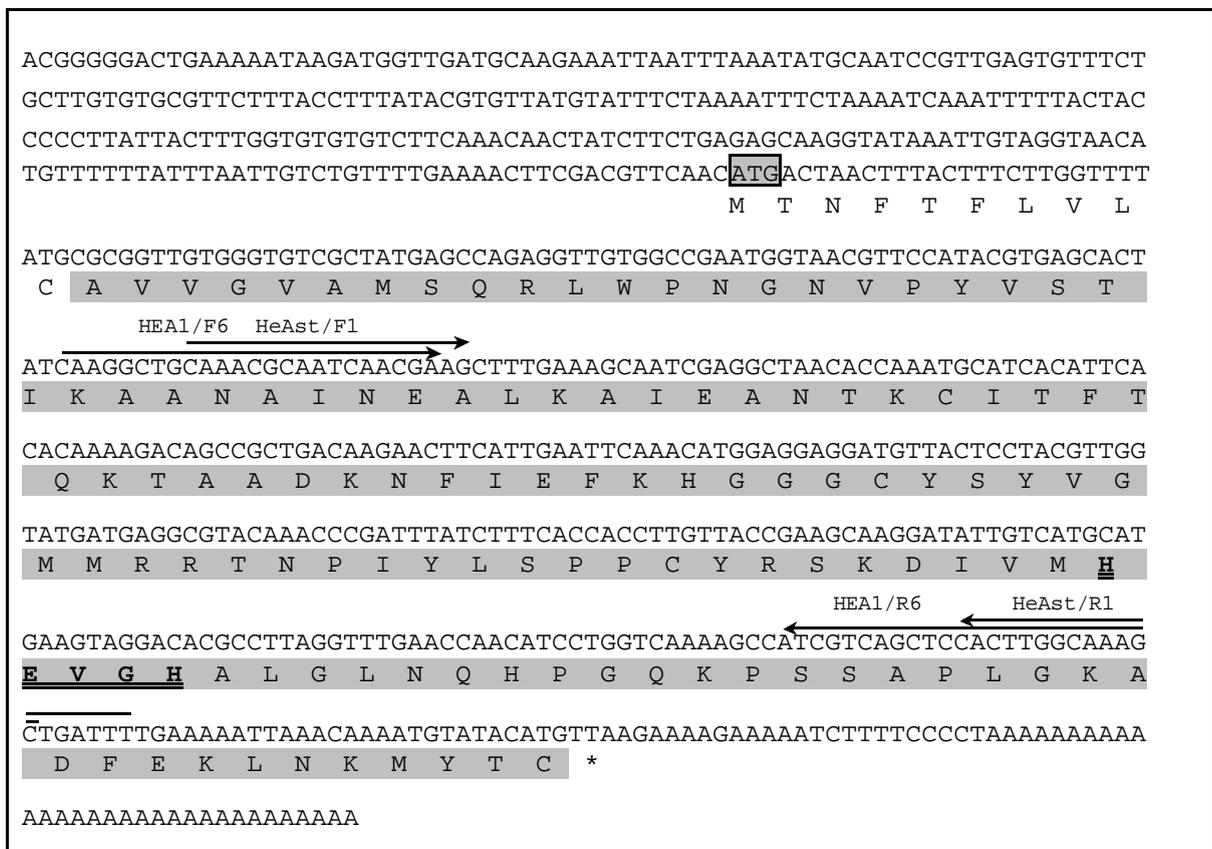


Abbildung 22: DNA- und abgeleitete Proteinsequenz von HEA 1. Hervorgehoben wurde das Startkodon (graue Box), die Astacin Domäne (grau) und das Zink-Bindungs-motif (doppelt unterstrichen). Die Position und Orientierung der Oligonukleotide für die PCR/RACE Experimente sind mit Pfeilen markiert.

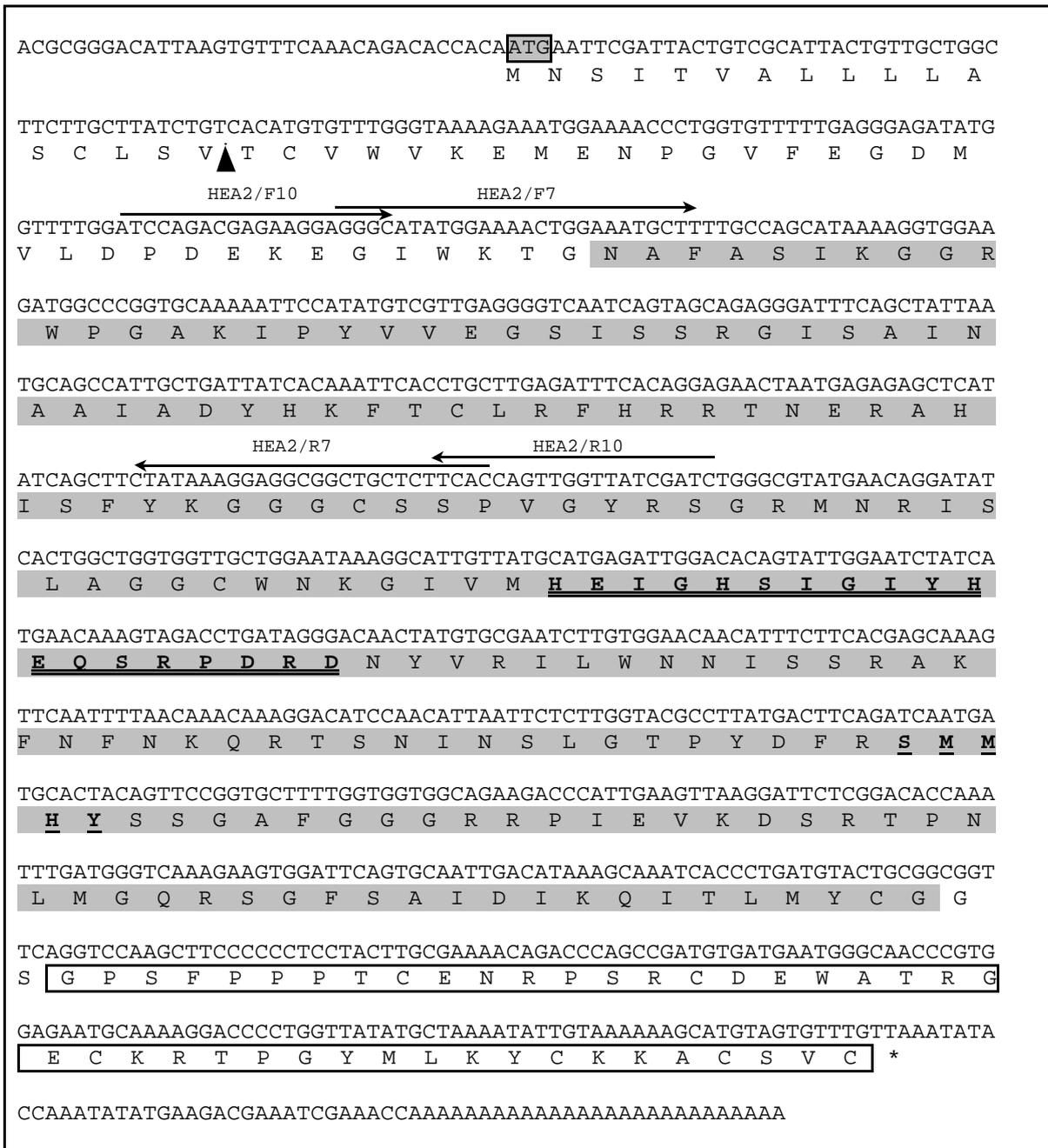


Abbildung 23: DNA- und abgeleitete Proteinsequenz von HEA 2. Hervorgehoben wurde das Startkodon (graue Box), die Astacin Domäne (grau), das Zink-Bindungsmotif (doppelt unterstrichen), der *Met-Turn* (einfach unterstrichen) und die Toxin-Domäne (weiße Box). Das Dreieck markiert die durch das Programm SignalSeq vorhergesagte Spaltstelle des Signalpeptides im Prä-Pro-Teil des Proteins. Die Position und Orientierung der Oligonukleotide für die PCR/RACE Experimente sind mit Pfeilen markiert.

In der HEA-1 cDNA-Sequenz (Abbildung 22) fällt zunächst auf, dass ein sehr langer 5' nicht-translatierter Bereich vorliegt. Auf Proteinebene konnte kein Signalpeptid identifiziert werden. Außerdem ist die normalerweise 200 AS lange Astacin-Domäne (Abbildung 22, grau hervorgehoben und Abbildung 24) unmittelbar nach dem Zink-Bindungsmotiv um ca. 50 Aminosäurereste verkürzt. Carboxyterminal finden sich keine weiteren Domänen. Das Protein endet mit einem Cysteinrest, der im Flusskrebs Astacin am carboxyterminalen Ende bei Position +198 liegt und Bestandteil der zweiten

Disulfidbrücke ist. Im HEA-2 Protein (Abbildung 23) findet sich demgegenüber nach dem 17 Aminosäuren langen Propetid zunächst ein kurzer Pro-Teil. Die Astacin Domäne ist mit 200 Aminosäureresten vollständig vorhanden. Carboxyterminal schließt sich eine sogenannte Toxin-Domäne (ShKT) an. Abbildung 24 zeigt einen schematischen Vergleich der Domänenstruktur von HEA-1 und HEA-2 mit den anderen bei Cnidaria beschriebenen Astacinen (*Hydra* HMP-1, HMP-2, FARM-1 und *Podocoryne* PMP-1). Weitere strukturelle und phylogenetische Analysen sind in Abschnitt IV.3. beschrieben.

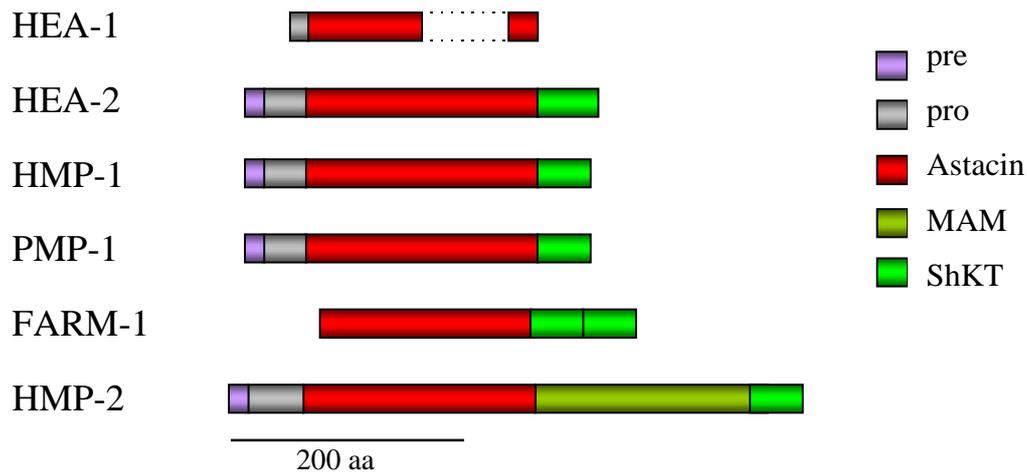


Abbildung 24: Domänenstruktur Astacin-homologer Proteine bei Cnidaria. Der Farbcode rechts bezeichnet die unterschiedlichen Domänen.

2.3. *In situ* Hybridisierung

Um erste Aussagen über die Funktion von HEA-1 und HEA-2 machen zu können, wurde das Expressionsmuster durch *in situ* Hybridisierung bestimmt. Als Sonde wurde Digoxigenin markierte *antisense*-RNA von Klon2 (HEA-1) und Klon5 (HEA-2) durch *in vitro* RNA-Synthese hergestellt. Dabei wurde für Kontrollreaktionen jeweils auch eine *sense*-RNA Sonde synthetisiert. Die Farbreaktion wurde mit Anti-Digoxigenin Alkalische Phosphatase gekoppelten Antikörpern und NBT/BCIP als Substrat durchgeführt.

Die *in situ* Hybridisierung zeigte keine Expression von HEA-2 mRNA in Gonozoiden oder Keimzellen (nicht dargestellt). HEA-2 mRNA konnte ebenfalls nicht während der frühen Embryogenese oder in Planula Larven (Abbildung 25 a, links) dedektiert werden. 12 Stunden nach Metamorphose Auslösung wurden in Primärpolypen die ersten Transkripte durch *in situ* Hybridisierung nachweisbar (25 a, rechts). In 24 –30 Stunden alten Primärpolypen wurde das HEA-2 Signal ausschließlich in endodermalen Zellen gefunden (Abbildung 25 b). Dabei handelte es sich vermutlich um differenzierende Drüsen-Zellen. Die Expressionsstärke, repräsentiert durch die Intensität der Farbsubstrat-Reaktion, nimmt innerhalb der Körpersäule vom Hypostome zur Fußregion deutlich ab. Abbildung 25 c zeigt eine Aufsicht auf das Hypostome und die Mundregion eines Primärpolypen. In adulten Polypen wurde eine schwache Expression unterhalb der Tentakelbasis gefunden (Abbildung 25 d, Pfeil).

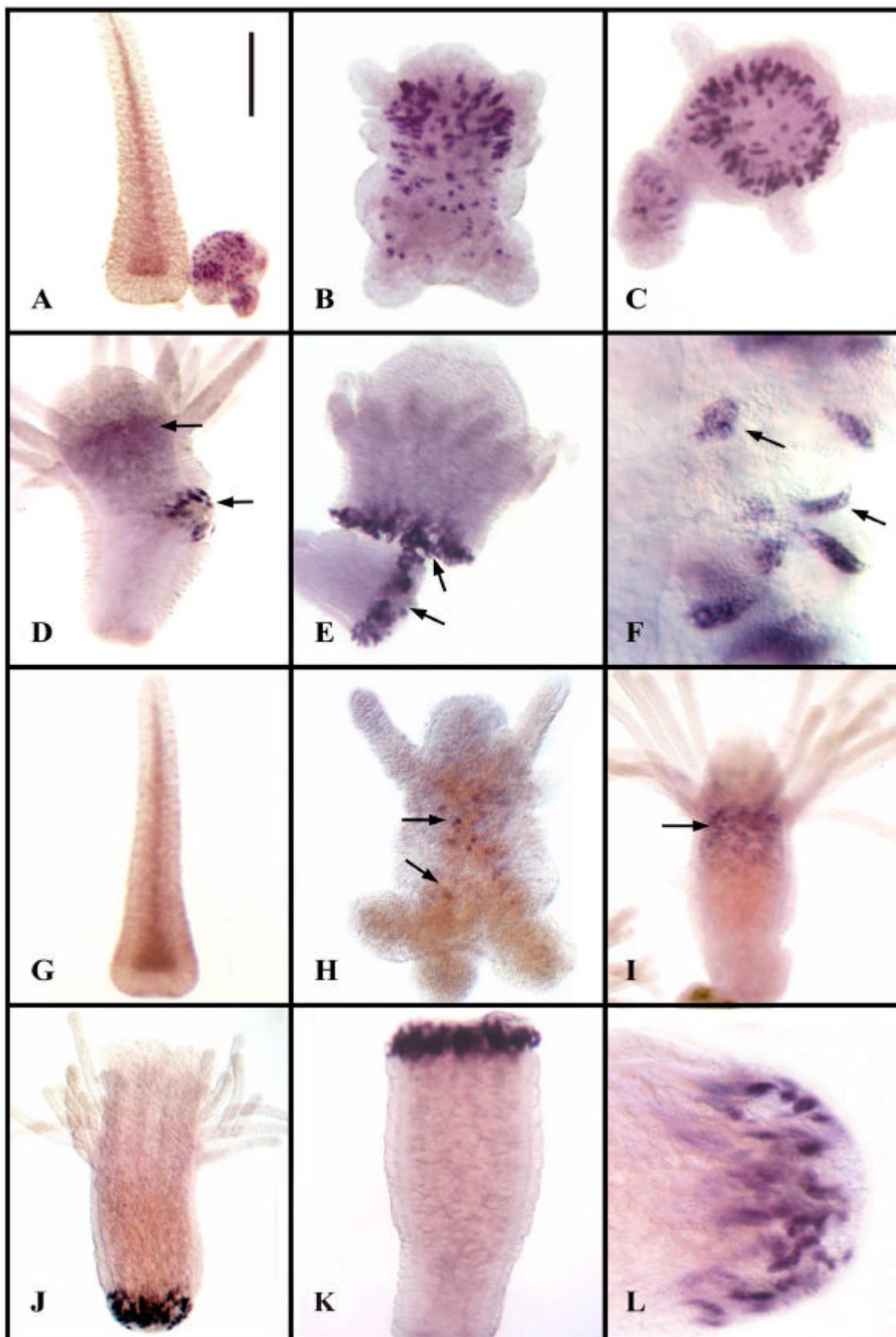


Abbildung 25: *in situ* Hybridisierung von HEA-2 (A-F) und HEA-1 (G-L). Legende siehe Seite 81.

Abbildung 25, vorangegangene Seite: *in situ* Hybridisierung von HEA-2 (A-F) und HEA-1 (G-L). (A) Larve und 12 Stunden alter Primärpolyp. (B) Primärpolyp 24 Stunden nach Metamorphoseauslösung (C) Aufsicht auf einen Primärpolypen (24 h) (D) Gastrozoid, lateral verletzt (E) Gastrozoid, median geschnitten (F) Ausschnittsvergrößerung einer Schnittwunde (G) Larve (H) 24 Stunden alter Primärpolyp (I) adulter Polyp (J) Gastrozoid, an der Fußregion geschnitten (K) Gastrozoid, median geschnitten (L) Ausschnittsvergrößerung einer Wundfläche. Die Länge des Balkens gibt die Bildvergrößerung an: A,G 100µm; B,C,H 150µm; D,J 350 µm; E,K 400 µm; F 40 µm; I 550 µm; L 80 µm

Weiterführende Experimente wurden mit adulten Polypen durchgeführt, die von einer Kolonie isoliert wurden und einen Tag regenerieren konnten. Diese Gastrozoide wurden dann lateral oder quer geschnitten und 30 Minuten später mit der *in situ* Hybridisierung begonnen. Zellen der Wundflächen reaktivierten dabei transient die HEA-2 Expression (Abbildung 25 d und e, Pfeile). Abbildung 25 e zeigt in höherer Vergrößerung, dass die HEA-2 Expression der Wundflächen auf einzelne, endodermale Drüsenzellen (Pfeile) beschränkt ist. In der Nachbarschaft liegende Zellen des offensichtlich gleichen Typs wurden nicht markiert.

Die *in situ* Hybridisierung mit den HEA-1 Sonden ergab ein ähnliches Expressionsmuster wie für HEA-2. HEA-1 ist weder in Gonozoiden, während der Embryonalentwicklung, noch in Planula Larven exprimiert (Abbildung 25 g). mRNA Transkripte wurden erst 12 Stunden nach Metamorphose-Auslösung in Primärpolypen gefunden (25 h). Allerdings ist die Expression hier im Gegensatz zu HEA-2 auf wenige endodermale Zellen (Pfeile) der Körperachse beschränkt. In adulten Polypen wurde eine schwache Expression unterhalb der Tentakelbasis gefunden (Abbildung 25 i, Pfeil). In verletzten oder geschnittenen Gastrozoiden konnte wiederum an den Schnittflächen eine starke, transiente Expression nachgewiesen werden (Abbildung 25 j und k). Das Gen wird also wie HEA-2 im Verlauf einer Wund-Antwort oder Geweberegeneration aktiviert. Dabei wurden die Transkripte für HEA-1 in endodermalen Drüsenzellen gefunden (Abbildung 25 l).

In den Kontrollreaktionen mit *sense*-Sonden konnte in keinem Fall ein Signal durch *in situ* Hybridisierung nachgewiesen werden.

Durch die hier gezeigten Expressionsmuster konnte eine Beteiligung sowohl von HEA-2 als auch von HEA-1 an der Geweberegeneration und an der Normalentwicklung während der Metamorphose gezeigt werden. In adulten Polypen deutet die Expression unterhalb der Tentakelbasis eine Beteiligung beider Gene an der ständigen Bildung oder Neubildung von Hypostom- oder Tentakel-Strukturen an.

3. Datenbankanalyse und phylogenetische Analyse

Mit der Aufklärung der Aminosäuresequenz von Astacin durch Titani et al. zeigte sich, dass diese Protease in keine bisher bekannte Enzymfamilie eingeordnet werden konnte und den Prototyp einer neuen Familie von Zink-Endopeptidasen repräsentieren muss (Zwilling et al., 1981, Krauhs et al., 1982, Titani et al., 1987, Stöcker et al., 1993). Neue Mitglieder der „Astacin-Protein-Familie“ wurden tatsächlich kurze Zeit später von verschiedenen Arbeitsgruppen entdeckt: BMP-1 (*bone morphogenetic protein 1*) im Menschen (Wozney et al. 1988), UVS.2 beim Krallenfroschs *Xenopus laevis* (Sato und Sargent 1990) und die Meprine von Maus und Mensch (Dummermuth et al. 1991). Bond und Beynon (1995) sowie Zwilling und Stöcker (1997) gaben dann letztmals einen Überblick der Struktur und Funktion der bis dahin etwa 20 Mitglieder der Astacin-Protein-Familie. Durch die in den letzten Jahren stark angewachsenen Datenbank-Informationen sollte deshalb ein aktueller Überblick der gesamten Astacin-Protein-Familie gegeben werden. Phylogenetische Analysen aller in den Datenbanken gefundenen Astacine und die in dieser Arbeit neu charakterisierten Astacin-homologen Proteine von *C. elegans* und *Hydractinia echinata* sollten darüber hinaus erstmals Aussagen über die Evolution der Astacin Protein Familie insgesamt ermöglichen.

3.1. Identifizierung Astacin-homologer Gene (BLAST, FASTA)

Der erste Schritt einer weitergehenden strukturellen und phylogenetischen Analyse war die Identifizierung aller Astacin-homologen Sequenzen in den Datenbanken. Dazu wurden als Suchprogramme die BLAST und FASTA Methoden verwendet. Als Vergleichssequenzen wurde die Astacin Sequenz, die Protease-Domäne von BMP-1 (Mensch), sowie das zinkbindende Motif und der Met-Turn (*motif 2/3* und *motif 4* der SPRINT Datenbank) benutzt, um auch unvollständig sequenzierte Genfragmente zu erhalten. Alle Suchergebnisse wurden auf Mehrfacheinträge untersucht. Zur Überprüfung erfolgte ein Sequenzvergleich (*multiple alignment*) aller durch BLAST oder FASTA identifizierten Proteine. Insgesamt konnten 106 verschiedene Astacin-homologe Proteine identifiziert werden (Tabelle 12). Davon werden 6 Proteine beim Menschen, 5 bei der Maus, 3 bei der Ratte, 4 bei Vögeln (3 Huhn, 1 Wachtel), 5 bei Amphibien (*Xenopus*), 15 bei Fischen (5 Medaka, 7 Aal, 1 Zebrafisch, 1 Fugufisch, 1 Karpfen), 4 bei Seeigeln (1 *Paracentrotus lividus*, 2 *Strongylocentrotus purpuratus*, 1 *Strongyloides stercoralis*), 14 bei Insekten (13 *Drosophila melanogaster*, 1 *Glossina morsitans*), 2 bei Crustacea (*Astacus astacus*), 40 bei *Caenorhabditis elegans* (siehe auch IV.1.), 6 bei Cnidaria (3 *Hydra*, 1 *Podocoryne*, 2 *Hydractinia*: siehe IV.2.), 1 bei *Aplysia californicum*, 1 bei *Ostertagia ostertagi*, 1 bei *Ancylostoma caninum* und ein bakterielles Protein (*Flavobacterium meningosepticum*) gefunden. Unterschiedliche Spleißvarianten eines Genes werden beim BMP-1 des Menschen (6 Formen), der Maus (3 Formen) oder *Xenopus laevis* (1 Form) gefunden.

Organismus	Gen Name	Bezeichnung	MERNUM	Swiss-Prot / TrEMBL	EMBL / GenBank	Kommentar / korrigierte Sequenz (siehe IV.1.2.2)
<i>Astacus astacus</i>	AEA	AA_AEA		O44072	AJ003190	
	Astacin	AA_AST	MER01104	P07584	AJ242595 X95684	Protein Sequenz Genom, PIR: A58830
<i>Aplysia californica</i>	TBL-1,	AC_TBL1	MER03792	P91972	U57369	PIR: T31069
<i>Ancylostoma caninum</i>	MTP-1	ACA_MTP1	MER15241		AY036056	
<i>Anguilla japonica</i>	EHE10	AJ_EHE10		Q90Y86	AB071426	
	EHE12	AJ_EHE12		Q90Y85	AB071427	
	EHE13	AJ_EHE13		Q90Y84	AB071428	
	EHE14	AJ_EHE14		Q90Y83	AB071429	
	EHE4	AJ_EHE4		Q90Y89	AB071423	
	EHE6	AJ_EHE6		Q90Y88	AB071424	
	EHE7	AJ_EHE7		Q90Y87	AB071425	
<i>Brachydanio rerio</i>	tolloid	BR_TLD	MER05126	O57460	AF027596	
<i>Caenorhabditis elegans</i>	4R79.1	C_4R79.1i	MER05846	Q9U3S9	AL031254	4R79.1i
	C05D11.6	C_C05D11.6i	MER02346	P55112	U00048	C05D11.6i
	C07D10.4	C_C07D10.4i	MER02282	P55113	U13072	C07D10.4i
	C17G1.6	C_C17G1.6i	MER03171	Q93243	Z78415	C17G1.6i
	C24F3.3	C_C24F3.3	MER05120	Q9XTD6	AL022716	= Z81055 / C24F3.3
	C26C6.3	C_C26C6.3	MER02349	Q18206	Z72503	C26C6.3
	C34D4.9	C_C34D4.9i	MER03819	Q18439	U58755	C34D4.9i
	C37H5.9	C_C37H5.9i	MER03493	P91137	U88315	C37H5.9i
	D1022.m	C_D1022.m	MER05843		U23517	D1022.m
	F09E8.6	C_F09E8.6i	MER03972	Q19269	Z73896	F09E8.6i
	F20G2.4	C_F20G2.4i	MER03172	Q93542	Z79753	F20G2.4i
	F38E9.2	C_F38E9.2n	MER03097	Q20176	U46668	F38E9.2n
	F39D8.4	C_F39D8.4i	MER02352	Q20191	Z69791	F39D8.4i
	F42A10.8	C_F42A10.8m	MER02171	P98061	U10414	F42A10.8m
	F45G2.1	C_F45G2.1o	MER05123	O62243	Z93382	F45G2.1o
	F46C5.3	C_F46C5.3n	MER02353	Q20459	Z54281	F46C5.3n
	F54B8.m	C_F54B8.m	MER05847		Z93383	F54B8.m
	F56A4.1	C_F56A4.1i	MER12259	Q9N5R0	AC006645	= AC006722 / F56A4.1i
	F57C12.1	C_F57C12.1i	MER02287	Q20942	U41554	F57C12.1i
	F58A6.4	C_F58A6.4i	MER02354	Q20958	U53339	F58A6.4i
	F58B4.1	C_F58B4.1	MER03814	Q20975	Z74038	F58B4.1
	HCH1	C_HCH1	MER03163		D85744	= Z69792 / HCH1
	K03B8.1	C_K03B8.1	MER03164	Q21180	Z74039	K03B8.1
	K03B8.2	C_K03B8.2	MER03115	Q21178	Z74039	K03B8.2
	K03B8.3	C_K03B8.3	MER03116	Q21179	Z74039	K03B8.3
	K03B8.5	C_K03B8.5	MER03165	Q21181	Z74039	K03B8.5
	K04E7.3	C_K04E7.3	MER02288	P55114	U39666	K04E7.3
	K06A4.1	C_K06A4.1i	MER03096	Q21252	Z70755	K06A4.1i
	K09C8.3	C_K09C8.3	MER03166	Q21388	Z68006	K09C8.3
	K11G12.1	C_K11G12.1	MER03167	Q21432	U23525	K11G12.1
	R151.5	C_R151.5	MER02172	P98060	U00036	R151.5
	T02B11.7	C_T02B11.7n	MER03831	O16977	AF022979	T02B11.7n
	T04G9.2	C_T04G9.2i	MER02291	P55115	U41274	T04G9.2i
	T11F9.3	C_T11F9.3i	MER03174	Q22396	Z74042	T11F9.3i
	T11F9.5	C_T11F9.5i	MER03169	Q22401	Z74042	T11F9.5i
	T11F9.6	C_T11F9.6i	MER03168	Q22398	Z74042	T11F9.6i
T23F4.4	C_T23F4.4	MER03833	O17264	AF025466	T23F4.4	
T23H4.3	C_T23H4.3	MER03173	P91828	Z83240	T23H4.3	
T24A11.3	C_T24A11.3i	MER01112	Q22710	Z49072	T24A11.3i	
Y95B8A.1	C_Y95B8A.1	MER12278	Q9N2V2	AC024877	Y95B8A.1	

Tabelle 12: Fortsetzung nächste Seite

Organismus	Gen Name	Bezeichnung	MERNUM	Swiss-Prot / TrEMBL	EMBL / GenBank	Kommentar
<i>Cyprinus carpio</i>	Nephrosin	CC_Nephr	MER03149	O42326	U62621	
<i>Coturnix coturnix</i>	CAM-1	CJ_AST	MER01114	P42662	U12642	N-terminus fehlt
<i>Drosophila melanogaster</i>	CG10280	DM_CG10280	MER11298	Q9VVK5	AE003601	
	CG11864	DM_CG11864	MER11333	Q9VJN9	AE003648	
	CG11865	DM_CG11865	MER11336		AE003648	
	CG15253	DM_CG15253	MER11335		AE003648	
	CG15254	DM_CG15254	MER11334	Q9V3M3	AE003413	= AE003648
	CG15255	DM_CG15255	MER11332		AC093096	= AE003648
	CG5715	DM_CG5715	MER11416	Q9VC98	AC008204 AE003747	
	CG6696	DM_CG6696	MER03918	Q9VWR6	AC012164 AE003509	= AC012373
	CG6763	DM_CG6763	MER11411	Q9VCN5	AC010670 AE003742	
	CG6974	DM_CG6974	MER11378	Q9VFD6	AC009457 AE003707	N-terminus fehlt
	CG7631	DM_CG7631	MER11337		AE003648	
	Tolloid, TLD	DM_TLD	MER01121	P25723	AE003749 M76976 U04239	PIR: A39288
	Tolkin, TOK	DM_TOK	MER02301	Q9VC47 Q23995 Q24132	AE003749 AY051585 U12634 U34777	
	<i>Flavobacterium meningosepticum</i>	Flavastacin	FM_Flavast	MER01612	Q47899	L37784
<i>Gallus gallus</i>	BMP-1	GG_BMP1	MER05127	O57658	U75331	
	Colloid, TLL-1	GG_COLLOID	MER12218	Q9DER7	AJ012462	
	TOLLOID	GG_TOLLOID		O57659		nur EGF/CUB-Domäne
<i>Glossina morsitans</i>	Astacin	GM_M15242	MER15242		AF329481	
<i>Hydractinia echinata</i>	HEA -1	HE_HEA1				Deletion von 50 Aminosäureresten
	HEA -2	HE_HEA2				
<i>Homo sapiens</i>	BMP-1, PCP, Bone morph. protein 1, Pro Collagen C Proteinase	HS_BMP-1	MER01113	P13497 Q13872	BC002593 BC009305 L35278	weitere Accession Nummern: Q13292; Q99421; Q99422; Q99423; Q14874; Q9UL38, I08629 (Patent) Splice Form, PIR: A37278
		HS_BMP1spT		Q14874	L35279 L40486	Splice Form BMP1-2 5' UTR
		HS_BMP1sp1		P13497	M22488 U50330	Splice Form BMP1-1
		HS_BMP1sp4		P13497	Y08723	Splice Form BMP1-4
		HS_BMP1sp5		P13497	Y08724	Splice Form BMP1-5
		HS_BMP1sp6		P13497	Y08725	Splice Form BMP1-6
		HS_BMP1sp7		P13497		Splice Form BMP1-7
	Hatch 1	HS_HCH1				Acc. in HUMAN PROT XM_065448
	Mepriin a	HS_MEPA	MER01111	Q16819	AL161618 M82962	PIR: S60193
	Mepriin b	HS_MEPB	MER05213	Q16820	X81333	PIR: S49383
	TLL1, Tolloid like 1	HS_TLL	MER05124	Q9NQS4 O43897	AF282732 U91963	
	TLL2, Tolloid like 2	HS_TLL2	MER05866	Q9UQ00 Q9Y6L7	AB023149 AF059516	weitere Accession Nummern: AK026106, AL136181, AL391136, BC013871

Tabelle 12: Fortsetzung nächste Seite

Organismus	Gen Name	Bezeichnung	MERNUM	Swiss-Prot / TrEMBL	EMBL / GenBank	Kommentar
<i>Hydra vulgaris</i>	FARM-1,	HV_FARM1	MER11835	<i>Q9U4X9</i>	AF125506	N-terminus fehlt
	HMP-1	HV_HMP1	MER01593	<i>Q25174</i>	U22380	
	HMP-2	HV_HMP2	MER05667	<i>Q9XZG0</i>	AF140020	
<i>Mus musculus</i>	BMP-1, PCP	MM_BMP1	MER02164	P98063 <i>Q61356</i>	AK004995 L24755 L35280 L35281	Tolloid Splice Form, PIR: I49540 Splice Form Splice Form
	Meprin a	MM_MEPA	MER01108	P28825	M74897 U62765	Fragment
	Meprin b	MM_MEPB	MER05214	Q61847	L15193	
	TLL1	MM_TLL	MER03170	<i>Q62381</i>	U34042	
	TLL2	MM_TLL2	MER05867	<i>Q9WVM6</i>	AF073526	kein Signalpeptid
<i>Oryzias latipes</i>	Alveolin	OL_Alveo	MER12290	<i>Q9IBE7</i>	AB030957	
	Choriolysin H	OL_CHOR_H	MER01105	P31580 <i>O13116</i>	D83950	Precursor von HCE-1 / 2, in Merops = HCE-1
	HCE-1	OL_HCE1			M96170	
	HCE-2	OL_HCE2	MER01106	P31581	M96171	
	LCE	OL_LCE	MER01107	P31579	D83949	= M96169
<i>Ostertagia ostertagi</i>		OO_AJ419180		<i>Q95PN5</i>	OOS419180	Fragment
<i>Podocoryne carnea</i>	PMP-1	PC_PMP1	MER05122	<i>O62558</i>	AJ005052	
<i>Paracentrotus lividus</i>	BP10	PL_BP10	MER01122	P42674	S99978	
					X56224	
				<i>Q26051</i>	X65721	
<i>Rattus norvegicus</i>	Meprin a	RN_MEPA	MER01109	Q64230	S43408	PIR: S24134
	Meprin b	RN_MEPB	MER05215	P28826	M88601	
	BMP-1		MER05387	<i>Q9Z135</i>	AB012139	Fragment
<i>Strongylocentrotus purpuratus</i>	BMP-1	SP_BMPH	MER02169	P98069	L23838	
	SPAN	SP_SPAN	MER01123	<i>P98068</i>	M84144	
<i>Strongyloides stercoralis</i>	Metalloprot.	StS_MP		<i>Q968Y6</i>	AF118022	Fragment
<i>Takifugu rubripes</i>	HCE-1	TR_HCE1		<i>AAL40376</i>	AC087333	
<i>Xenopus laevis</i>	BMP-1	XL_BMP1	MER01120	P98070	L12249	PIR: JC2218
	BMP-1b	XL_BMP1b		<i>O57381</i>	Y09660	
	UVS-2	XL_UVS2	MER01115	P42664	D89632 M27162	Fragment
	Xolloid related	XL_XOL_R		<i>Q91925</i>	D83476	BMP-1 Splice Form
	XOLLOID	XL_XOLLOID	MER05125	<i>O57382</i>	Y09661	

Tabelle 12: Übersicht aller 107 annotierter Astacin-homologen Proteine.

Tabelle 12 beinhaltet keine EST-Klone, die identisch mit bereits bekannten, vollständig sequenzierten Klonen sind und auch nicht kurze, unvollständige Genom (GSC)-, EST- oder N-terminale Sequenzfragmente. Tabelle 13 gibt eine Übersicht dieser 77 Protein-Fragmente. Hierunter finden sich beispielsweise ein weiteres bakterielles Astacin-homologes Sequenzfragment von *Legionella pneumophila* sowie N-terminale Aminosäuresequenzen von *Chionoectes opilio* (Crustacea) und eines Verdauungsenzymes des Maikäfers (*Melolontha melolontha*). *Xenopus laevis* weist neben den bereits bekannten BMP-1, UVS-2 und tolloid-ähnlichen Proteinen noch 4 weitere, bisher nicht charakterisierte Sequenzen auf. Bei dem Nematoden *Caenorhabditis briggsae* konnten analog zu *C. elegans* eine überraschend große Kopienzahl von insgesamt 34 Genfragmente identifiziert werden.

Organismus	Quelldaten	Gen Bezeichnung	EMBL / Genbank	Kommentar
<i>Ancylostoma ceylanicum</i>	EST	AC_BI704652	BI704652	identisch mit BM130670
		AC_BM130670	BM130670	identisch mit BM130670
		AC_BI704826	BI704826	
<i>Caenorhabditis briggsae</i>	GSC	34 Gene, nicht annotiert, nur in Wormbase: c013600731; c010500107; c002600794; c001200679; c008200278; c002500858; c012300800; c004900169; c010301013; c003401041; c013200824; c002300559; c001500442; c006900571; c001700704; c002500668; c006500442; c013100148; c006700548; c009300834; c010200914; c006900534; c011800840; c003800791; c008900679; c011300623; c009100728; c000100457; c007900654; c007500103; c003500775; c010400911; c011600772; c003800198		
<i>Ciona intestinalis</i>	EST	CI_AV877936	AV877936	
<i>Chionoecetes opilio</i>	Edman	CO_Ast.F		20 Aminosäuren N-Terminus, SWISSPROT: P34156, Merops: MER01893
<i>Danio rerio</i>	EST	DR_BI705073	BI705073.5	Identisch mit BM025358, BM155001
		DR_BI708379	BI708379	Identisch mit BI708513, BI709736, BI866941, BM082481, BM082751
		DR_BI884917	BI884917	
<i>Drosophila simulans</i>	GSC	DS_AF264913	AF264913	TrEMBL: Q9GTN4, MEROPS: MER12216
		DS_AF264915	AF264915	TrEMBL: Q9GTN2, MEROPS: MER12238
<i>Drosophila yakuba</i>	GSC	DY_AF264914	AF264914	TrEMBL: Q9GTN3, MEROPS: MER12226
<i>Heterodera glycines</i>	EST	HG_BI748309	BI748309	
<i>Legionella pneumophila</i>	GSC	LP_MP-ORF-2		gnl CUCGC_446 19, genomisches Sequenzierungsfragment von NCBI Bacteria
<i>Meloidogyne arenaria</i>	EST	MA_BI747415	BI747415	
		MA_BI747765	BI747765	
<i>Melolontha melolontha</i>	Edman	MeM_Ast.F		N-Terminus (18 Aminosäurereste); Moehren und Wagner, unpublished
<i>Mus musculus</i>	EST	MM_BB847373	BB847373	
<i>Nippostrongylus brasiliensis</i>	EST		BM279324	Neu 22.12.01
<i>Oryzias latipes</i>	EST	OL_BJ000733	BJ000733	Identisch mit: BJ012377, BJ013120, BJ015289, BJ026641, BJ027405, BJ027922
		OL_BJ006532	BJ006532	Identisch mit BJ010078, BJ010496, BJ011497, BJ020406, BJ023882, BJ025747
		OL_BJ011407	BJ011407	Identisch mit BJ025661
		OL_BJ013894	BJ013894	Identisch mit BJ022807, BJ024300, BJ028053
<i>Parastrongyloides trichosuri</i>	EST	PT_BI742807	BI742807	
		PT_BI743922	BI743922	
<i>Salmo salar</i>	EST	SS_BG933928	BG933928	
		SS_BG933935	BG933935	
		SS_BG933936	BG933936	
		SS_BG934059	BG934059	
		SS_BG934150	BG934150	
		SS_BG934168	BG934168	
		SS_BG934270	BG934270	
		SS_BG934609	BG934609	
		SS_BG934688	BG934688	
SS_G934654	G934654			
<i>Teladorsagia circumcincta</i>	EST	TC_BM052081	BM052081	
		TC_BM052081	BM052081	
<i>Xenopus laevis</i>	EST	XL_BJ037887	BJ037887	
		XL_BJ068096	BJ068096	
		XL_BM179683	BM179683	
<i>Xenopus tropicalis</i>	EST	XT_AL641908	AL641908	
		XT_AL644278	AL644278	Identisch mit AL652763
		XT_AL650582	AL650582	
		XT_AL662162	AL662162	

Tabelle 13: Übersicht aller 77 GSCs, ESTs und Edman-sequenzierter Fragmente.

3.2. Analyse der Domänenstruktur

Über die Struktur der meisten Astacin-homologen Proteine ist wenig bekannt. Alle Proteine sollten deshalb zunächst anhand ihrer Domänenstruktur analysiert werden. Dazu wurden die Programme SMART (*Simple Modular Architecture Research Tool*), PFAM (*Protein families database of alignments and HMMs*) und ProfileScan benutzt. Die Aminosäuresequenzen sind mit Ausnahme des Astacins von *Astacus astacus* (vollständige Edman-Sequenzierung, Titani et al. 1987) ausnahmslos von cDNA oder genomischen Sequenzdaten abgeleitet. Dabei wurden, soweit verfügbar die Proteinsequenzen der SwissProt, TrEMBL, EMBL oder GenBank Datenbanken verwendet, deren Identifikationsnummern (*Accession number*) in Tabelle 12 angegeben sind. Lagen in den Datenbank nur DNA-Sequenzen vor, wurden die offenen Leseraster mit dem *StadenProgramPackage* in Aminosäure-Sequenzen übersetzt. Für *C. elegans* wurden die nach den cDNA Daten korrigierten genomischen Sequenzen (vergleiche IV.1.2.2) verwendet.

Die Abbildungen 26, 27 und 28 zeigen schematisch die Domänenstruktur der 106 identifizierten Astacin-homologen Proteine. Fast alle Proteine weisen danach eine Pre-Pro-Struktur aminoterminal der Proteasedomäne auf. Pre-Pro-Sequenzen werden bei CAM-1 der Wachtel, FARM-1 von *Hydra*, F54B8.m von *C. elegans* und CG6974 von *Drosophila* nicht gefunden. Zudem zeigen diese Proteine eine aminoterminal verkürzte Proteasedomäne. Signalpeptide konnten zudem nicht für BMP-1 von *Strongylocentrotus*, die *C. elegans* Proteine K09C8.3, K03B8.3, K03B8.1, K03B8.5, F20G2.4, D1022.m, Y95B8A.1, K04E7.3, F38E9.2 und die *Drosophila* Proteine CG15255, CG10280 postuliert werden. Einige Astacine bestehen, wie das Verdauungsenzym Astacin von *Astacus Astacus*, nur aus der Protease-Domäne. Das gilt für 6 *C. elegans* Proteine (Abbildung 27), 11 *Drosophila* Proteine (Abbildung 28), 1 *Hydractinia* Astacin (HEA-1), die *hatching* Enzyme von *Anguilla* (EHE-4, -6, -7, -10, -13 und -14), *Cyprinus* (Nephrosin), *Oryzias* (Alveolin, HCE-1, HCE-2, LCE), *Takifugu* (HCE-1) und die Spleißvariante 4 des humanen BMP-1 (alle Abbildung 26). Alle anderen Proteine weisen carboxyterminal verschiedene Domänen in einer oder mehreren Kopien auf (Abbildung 26 – 28). Häufig findet man EGF- (*epidermal growth factor*) oder CUB-Domänen (*complement components C1r/C1s*, *embryonic sea urchin protein Uegf*) die eine wichtige Rolle bei der Substratbindung durch die Protease spielen könnten. Die BMP-1/Tolloid ähnlichen Proteine (Abbildung 26) besitzen dabei die Anordnung von zwei CUB-, einer EGF-, einer CUB-, einer EGF- und zwei CUB-Domänem. Verschiedenen BMP-1 Spleißvarianten des Menschen, BMP-H des Seeigels und BMP1 des Huhns fehlen dabei die drei letzten Abschnitte. SpAN und BP10 besitzen in anderer Anordnung eine EGF und zwei CUB-Domänen, die durch Ser/Thr reiche Regionen voneinander getrennt sind. Die *hatching* Enzyme AEA von *Astacus*, Astacin von *Cortunix* und UVS.2 von *Xenopus* enthalten dagegen nur CUB-Domänen. Bei *C. elegans* (Abbildung 27) findet man ein Protein (F38E9.2) welches hinsichtlich der Domänenstruktur mit den BMP-1/Tolloid ähnlichen Proteinen übereinstimmt. 23 weitere *C. elegans* Proteine weisen mindestens eine EGF- oder CUB-Domäne carboxyterminal der

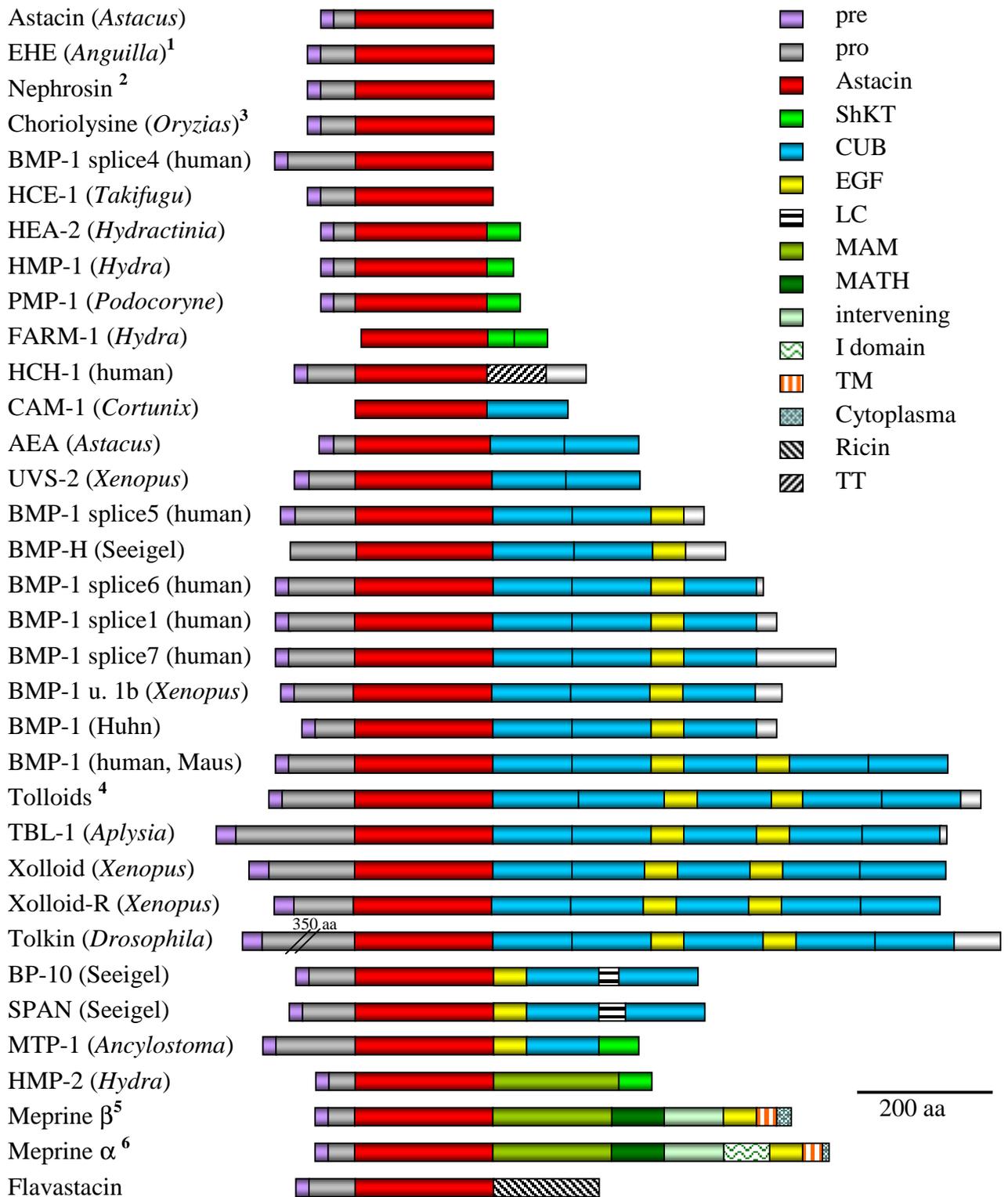


Abbildung 26: Vergleich der Domänenstruktur Astacin-homologer Proteine (*C. elegans* und *Drosophila melanogaster* Proteine wurden nur exemplarisch berücksichtigt). Proteine mit ähnlicher Struktur wurden zusammengefasst: 1 EHE-10, EHE-12, EHE-13, EHE-14, EHE-4, EHE-6, EHE-7; 2 Nephrosin (*Cyprinus*), M15242 (*Glossina*), HCE-1 (*Takifugu*); 3 Alveolin, Choriolysin H, HCE-1, HCE-2, LCE; 4 Tolloid (*Drosophila*), TLL (human, Maus), TLL-2 (human, Maus), Tolloid (*Brachydanio*), Colloid (Huhn); 5 Meprin β , (human, Maus, Ratte); 6 Meprin α , (human, Maus, Ratte). Der Farbcode oben rechts bezeichnet die unterschiedlichen Domänen.

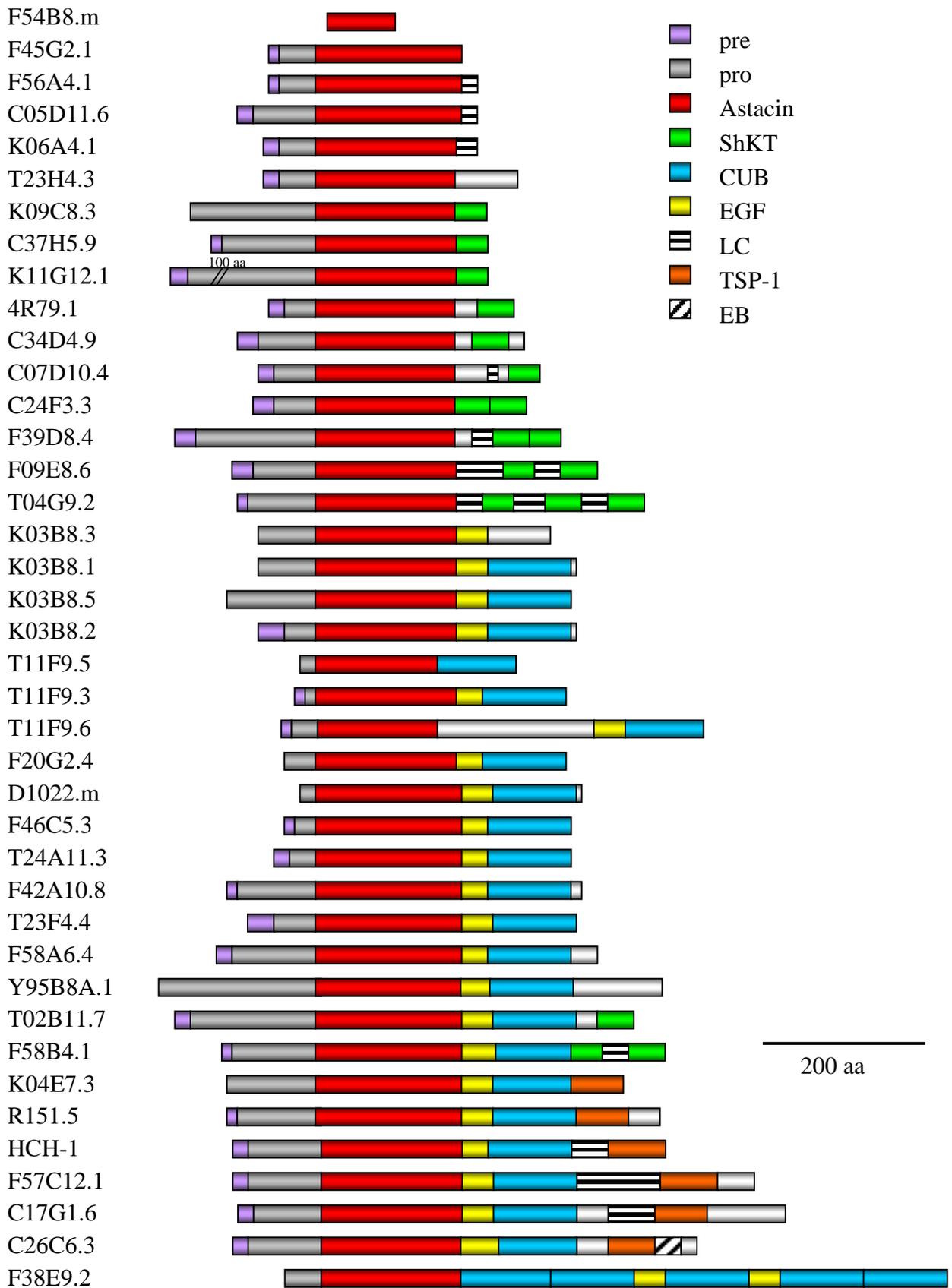


Abbildung 27: Domänenstruktur Astacin-homologer Proteine bei *C. elegans*. Die Proteine wurden entsprechend den Gen-Namen in Tabelle 12 bezeichnet. Der Farbcodex oben rechts bezeichnet die unterschiedlichen Domänen.

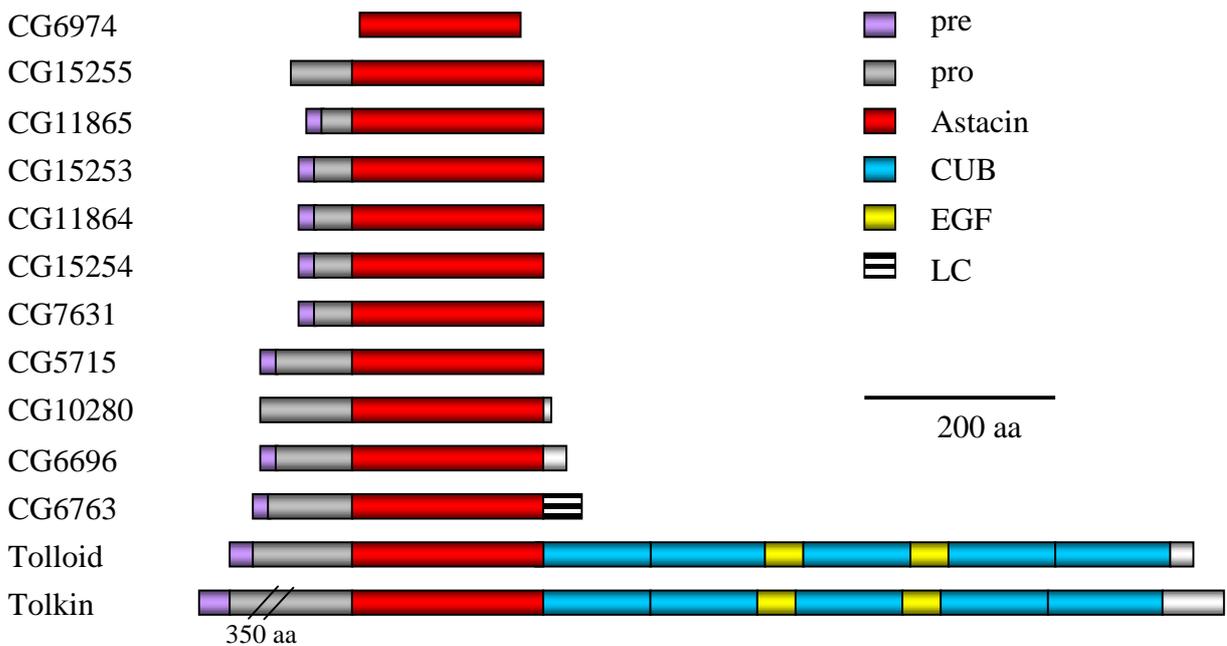


Abbildung 28: Domänenstruktur Astacin-homologer Proteine bei *Drosophila melanogaster*. Der Farbcode oben rechts bezeichnet die unterschiedlichen Domänen.

Protease-Domäne auf HCH-1, K04E7.3, R151.5, F57C12.1, C17G1.6 und C26C6.3 von *C. elegans* enthalten interessanterweise zusätzlich eine TSP-1-ähnliche Domäne (*Thrombospondin type 1*). T02B11.7 und F58B4.1 (Abbildung 27) sowie MTP1 des Nematoden *Ancylostoma* (Abbildung 26) besitzen neben einem EGF- und CUB-Motiv eine sogenannte ShKT-Domäne, ein ursprünglich bei Seeanemonen zum ersten Mal beschriebenes Toxin (*ShK toxin domain*). Dieses Motif findet sich ohne EGF/CUB-Domänen auch bei HEA-2, HMP-1, PMP-1, FARM-1 (Abbildung 26) und bei weiteren 10 *C. elegans* Astacinen (Abbildung 27). LC-Regionen (*low compositional complexity*) sind wie bei SpAN und BP10 (Abbildung 26) meist Ser/Thr-reiche Sequenzabschnitte, die als O-Glykolisierungsstelle der Proteine dienen können und die bei vielen Astacinen gefunden werden. MAM- (*A5 receptor protein, tyrosine phosphatase μ and others*), MATH- (*meprin and TRAF homology*) oder Transmembran-Domänen (TM) werden ausschließlich bei den Meprin und bei *Hydra* HMP-2 gefunden (Abbildung 26). Die Meprine bestehen darüber hinaus aus EGF- und noch nicht näher charakterisierten Zwischenbereichen (I-Domäne, *intervening*). Ungewöhnlich erscheint das Vorkommen einer TT-Domäne (*open reading frame 2* des humanen TT-Virus) bei HCH-1 des Menschen, einer Ricin-Domäne (Lektin in *Ricinus communis*) bei Flavastacin oder eines sogenannten EB Moduls (bestehend aus 8 konservierten Cysteinen) (Abbildung 26).

3.3. Molekularphylogenetische Analyse

Die Analyse der Domänenstruktur gibt bereits erste Hinweise auf die Evolution der Astacin-Protein-Familie. Die unterschiedlichen Kompositionen verschiedener Domänen zu neuen Proteinen innerhalb eines Organismus oder bei der Aufspaltung der Spezies im Verlauf der Evolution müssen sich auch in der Primärstruktur der Protease-Domäne widerspiegeln. Um einen tieferen Einblick in die Entwicklung der orthologen und paralogen Proteinsequenzen zu bekommen wurde darüber hinaus eine molekularphylogenetische Analyse durchgeführt. Ausgangspunkt war die Erstellung eines Mehrfach-Sequenzvergleichs (*multiple alignment*) der Protease-Domäne. Hierbei können aber nur Proteine berücksichtigt werden, die keine Deletionen in diesem Bereich aufweisen. Die Proteine F54B8.m (*C. elegans*), CG6974 (*Drosophila*), CAM-1 (Wachtel), FARM-1 (*Hydra*) und HEA-1 (*Hydractinia*) konnten hier deshalb nicht einbezogen werden. Für alle übrigen Proteine wurde mit Hilfe des Programms CLUSTALX ein computergestütztes *alignment* hergestellt und anhand der bekannten Sekundärstrukturen und konservierten Regionen des Astacins manuell korrigiert. Abbildung 29 zeigt den Sequenzvergleich der 200 Aminosäurereste umfassenden Astacin-Domäne beginnend mit dem Aminoterminus der aktiven Enzyme bis zu dem Cysteinrest der zweiten Disulfidbrücke. Die Zinkliganden sind mit einem Dreieck und die zwei Disulfidbrücken mit dem Symbol © hervorgehoben. Oberhalb der Aminosäure-Nummerierung ist die Konsensussequenz der Astacin-Protein-Familie angegeben.

Da die hohe Kopienzahl der Astacin-homologen Gene bei *C. elegans* und *Drosophila melanogaster* die Interpretation der folgenden phylogenetischen Analyse erschwert, wurde ein weiterer Sequenzvergleich durchgeführt, der exemplarisch nur jeweils drei Proteine dieser Organismen enthielt (Anhang VII.6.3).

Anhand beider Sequenzvergleiche wurden nun die phylogenetischen Verwandtschaftsverhältnisse mit Hilfe der *Neighbor-joining* Methode berechnet und ein Stammbaum erstellt (Abbildung 30 und 31). Als Außengruppe (*outgroup*) wurde das entwicklungsgeschichtlich vermutlich älteste, bakterielle Flavastacin verwendet. Zur statistischen Überprüfung der Ergebnisse (*bootstrap analysis*) wurde ein Konsensusstammbaum mit 100 Datensätzen ermittelt. Die statistische Unterstützung der einzelnen Äste ist jeweils an den Knotenpunkten angegeben.

Die Astacine können in verschiedene Stammbaum-Gruppen (Abbildung 30 A-D) unterteilt werden. Die Cnidaria Proteine HMP-2, HEA-2 und PMP-1 (Gruppe A) liegen dabei basal als nächste Verwandte des Flavastacins. Als Schwestergruppe erscheinen die Meprine (Gruppe B), die aber überraschenderweise nur bei Säugetieren gefunden werden. HMP-2 weist mit dem Vorhandensein einer MAM-Domäne als einziges Astacin-Familien-Mitglied eine weitere Übereinstimmung mit den Meprinen auf.

Abbildung 29, nächste Seite: Sequenzvergleich der Astacin-Protein-Familie. © Cysteinreste von Disulfidbrücken, ▼ Zink-Liganden; oberhalb der Aminosäure-Nummerierung ist die Konsensussequenz der Astacin-Protein-Familie (Kleinbuchstaben 80%, Großbuchstaben 100% konservierte Aminosäurereste) angegeben; Aminosäure-Positionen, die in mehr als 80 % der Proteine konserviert sind, wurden zudem schwarz hinterlegt. Zur Erläuterung der hier verwendeten Protein-Bezeichnung siehe Tabelle 12.

	w	20	40	a	c	60	80		
AA_AST	:	--AAILG--DEYIWSGGV	IPYTFAG-VS--GADQSA	LLSGMQELEEKTCIRFVPR	-----	TTESD	VEIFTPSGS--		
HV_HMP2	:	--AALSN--SSILWLPNNKV	VPWSTIKQLENTA	EATFCGMAAFREWEERSCL	PKRR	-----	TDEKDYIEFFQGS		
HE_HEA2	:	NAFAS--IKGGRWPGAK	IPYVVEGSIS--SRGIS	ANAAADYHKFTCLRPHRR	-----	TNERAF	ISFYKGG--		
PC_PMP1	:	NAFAS--IKGGRWPGAK	IAVQLIERSIG--SGGRT	AIANAIAQYHKHTCLRPHKR	-----	TNERSF	ISFYKGG--		
HS_MEPA	:	NLGRD--PNTRWTFP	IPYILADNLG--LNAKGA	ILYAFEMFRLKSCVD	PKPY	-----	EGESSYIIFQGF--		
MM_MEPA	:	NAMRD--PSSRWKLP	IPYILADNLE--LNAKGA	ILHAFEMFRLKSCVD	PKPY	-----	EGESSYIIFQKLS--		
HS_MEPB	:	NSIIG--EKYRWPH	IPYVLEDSE--MNAKGI	LNAFERYRLKTCID	PKPW	-----	AGETNYISVFKGS--		
MM_MEPB	:	NSIIG--DHKRWPH	IPYVLEDSE--MNAKGI	LNAFERYRLKTCID	PKPW	-----	AGETNYISVFKGS--		
AC_TBL1	:	--ATPAD--RSKIWDHGV	IPYVLESNYS--GENKDL	FKLAMRHWENLTCL	VEPKDKG	-----	PEDTNYILFTQTDC--		
C_F38E9.2n	:	--AATAK--KERIWP	EPVIASNFS--GEHQHL	FLRAMRHWENLT	CVSEVPRQ	-----	PHHKYITFTVDDK--		
SP_BMPH	:	--AVTAR--PERRTDAV	IPYVIDGNFT--GSQRA	MFKCAMRHWENY	TCIFVERN	PAN	-----	EIHDNYIVFTYQAC--	
BR_TLD	:	--AATSR--AEKIWPGV	IPYVIGGNFT--GSQRA	MFKCAMRHWEK	TCVIFIEK	-----	TDEESYIVFTYRCP--		
HS_TLL	:	--AATSR--TERIWP	PGGVIPYVIGGNFT--GSQRA	MFKCAMRHWEK	TCVIFLER	-----	SDEESYIVFTYRCP--		
HS_TLL2	:	--ATTSR--TERIWP	GGVIPYVIGGNFT--GSQRA	MFKCAMRHWEK	TCVIFLER	-----	TDEESYIVFVYRCC--		
XL_XOLLOID	:	--AATSR--TERIWP	GGIIPYVLAGNFT--GTQRA	IFKCAMRHKKHT	TCVIFVER	-----	TDEESYIVFTYRCP--		
XL_BMP1	:	--AATSR--PERVW	PDGVIIPYVISGNFS--GSQRA	IFRCAMRHWEK	TCVIFLER	-----	TDEESYIVFTYRCP--		
HS_BMP1	:	--AATSR--PERVW	PDGVIIPYVIGGNFT--GSQRA	IFRCAMRHWEK	TCVIFLER	-----	TDEESYIVFTYRCP--		
MM_BMP1	:	--AATSR--PERVW	PDGVIIPYVIGGNFT--GSQRA	IFRCAMRHWEK	TCVIFLER	-----	TDEESYIVFTYRCP--		
DM_TLD	:	--AVTVR--KERIWD	YGVIPYVIDTIFS--GAHKAL	FKLAMRHWENLT	TCIKFVERDP	-----	NLHANYIYFTVKNC--		
DM_TOK	:	--AVTAK--KERIWD	YGVIPYVIDGNFS--GLHKAL	FKLAMRHWENLT	TCIKFVERDP	-----	EIHPNYIYFTVYRSC--		
PL_BP10	:	--ATIY--ESQRW	PKYVIPYVISPSSS--GQSS--I	IRNADHWEQNTCLR	EPRTSSHSRQLGHNA	YLSFFRGS	-----		
SP_SPAN	:	--ATIY--ESQRW	SYKIIPYVIESSSS--GQSS--I	IRNADHWEQNTCLR	EPLTSSHSRLGHTS	YLSFFRGN	-----		
C_R151.5	:	----FIGSNLR	WDASRP	IIWAFDGSHT--QREQR	IELALEHWHNITCL	NFORNDQ	-----	ANSNGNYIVFTD	
C_HCH1	:	SMTSF--LSE	RSF	VPYVITDSSG--VNTNA	VLAVAKWEQTCAR	TRLNSYSS	-----	SSRQNALRFISGN	
C_F58B4.1	:	----YYPST	WSS	VYVYDRTAT--PKIV	KAFECVAFWQVNT	TCINMQS	-----	STAINIRVFKQ	
ACA_MTP1	:	NY-----PRTL	SKG	IVPFRHRNAT--PEVRS	VFVKGAKLWMD	TCIDPFES	-----	NSAPDIRVFKEN	
OO_MP	:	QAFKDHKY--PKT	WISQ	VNYVYFHDMA--KQMK	SFVVKGAKWWEK	TCINFTEN	-----	RSAEDRIMVFPK	
C_F46C5.3n	:	QVQRD--LTY	WPNNT	QNVVYVGN--VT--	STIKSVRLATEEL	QAWTCIR	QNVNEK	-----	YSDGDSVRIVDL
C_F20G2.4i	:	----EFERL	GSKLGGT	INVYVADNNT--SVKE	KVKSATAYIANH	TCIKFNED	-----	PTHWYIKFITSELS	
C_K03B8.5	:	QFERL--GTK	SYGV	VNYVYADKNN--EIKEM	VESAAYIANH	TCIRFNED	-----	QNAVQRVQIRMQNW	
C_K03B8.2	:	QITKI--WK	WPAK	VFYVYENFT--SLKR	IMSYAMAHISSN	TCVKFOES	-----	NSATNIRFTNTG	
C_K03B8.1	:	QVVTKL--FSP	QNSAK	VFYVGNADSF--KLQL	IMFKAMFNFISS	TCVIFTEEN	-----	CTISTRKIFVTS	
C_K03B8.3	:	----S	QLPSSAK	LLVCFDSSFS--TEKRE	IMLFSMNFISS	TCVIFTEEN	-----	CSIPNRYQLVNGK	
C_T02B11.7n	:	QIDN--LAQ	WPGKV	VYVYFDSGLT--TTVQ	QVIRDAITFLESN	TCLKFEIN	-----	STATN--RIFSGV	
C_T11F9.3i	:	QAMRD--YAK	ENNKM	SLVIFYNLPL--EMQ	AMFRDANYLENH	TCLKFEYN	-----	ENAETAVIRKGN	
C_D1022.m	:	NAMY--PKT	WLPN	VPVFLHGSLS--AKSRS	SIIVAAAFWEKH	TCVAFKRR	-----	TSEKVIYLLMSGQEE	
C_F58A6.4i	:	QAYLDNN--YPAT	WKNG	VAFMFHSEL--PIAKT	ILKAVHFWR	TCIEFHR	-----	TFQKRYLLFVGNDD	
C_T24A11.3i	:	----EVIAG	QIYDINSYE	IPVQIWGDY--NFQ	SILRRGIRMWEDS	TCLRFKEN	-----	QCRDRIYVLEKGD	
C_K04E7.3	:	----MNL	NGTWSRN	IPYRFLD--TD--	GNWOSITNGLRHY	ERNTCIR	SLN	-----	GGGSDYIVFSKGE
C_C17G1.6i	:	QAHPD--PRN	WPNLT	ISVEFYG--GE--	ETWRQLRSARH	VEQVNF	CFKFKEN	-----	GGDRDCLRYRGN
C_Y95B8A.1	:	----VIT	GSVYRKS	IPVPRFKGGDA--KWK	ILREGILWEKE	TCVWSEN	-----	GGPDYIVIFFRGS	
C_C26C6.3	:	----SFV	SDKTATWKTMP	IKYRFHESID--FYT	LSQIIAARFVWEDS	TCIFENVSD	-----	SPDGDYIEFFSGQ	
C_F57C12.1i	:	----IGRN	PLYKWDTRGP	ISIDYAESIP--FQTR	QKRSAMLLWQQH	TCLRFEEG	-----	GNVNRLEFFDGG	
C_F42A10.8m	:	QALVDT--TNE	FSVSPV	IFVQFDTKLS--ATNIA	NRKAIQFVNDN	SCLSPKED	-----	NNAKN--RLFLSSAG	
C_T23F4.4	:	--AVIRQ--KHK	KWKLMP	YSFDRNFREF--LREK	NLEAMQFWEK	TCVIFPHEN	-----	RYVYPVSIIFEGN	
OL_Alveo	:	NA-----VNM	IAQGV	IPYITIMBLA--HRQ	--SBEKEAFSMIT	SATCIR	QER	-----	TREANVLDIKDGD
DM_CG11865	:	VVKSW--SEY	WKGR	IPYVSYAGFS--SLDIA	ISESAMAEISSK	TCVKFRTEY	-----	KYEP--CVVIQKES	
C_K09C8.3	:	--ASIFFEQ	NLIQKWPSTSP	IPYTFDSSLD--NLD	ONDVRGAISEIEQ	TCIRFKYFAS	-----	PPKGNFINYQKVN	
C_C37H5.9i	:	--SGVFFQ	ESAVQKWDIWK	IQYILDDSE--ESDK	KDIRDAHEISINT	TCIFRYNA	-----	TPKGYHLNYMKVDS	
C_K11G12.1	:	--SALYFEG	NLIKWDPS	IRVYILDDSE--DLK	NDVRGAIYIEIK	TCIRFKELSS	-----	PTGSGEIVYKVD	
DM_CG6696	:	NGLLN--ERL	WPEAA	VPYILDPQDFN--ANQ	TMVILKAFKEYH	DRTCIR	RPYE	-----	QGDKHWLLIKGNYS
DM_CG15255	:	NGLIN--TEK	WPGNV	VVYRISDDFD--TAHK	AKAQGTIDTLELH	TCLRFREAT	-----	DEDKAYLTVTAKSG	
DM_CG7631	:	QQLS--ETR	WPNAT	VPYRISSEFD--APHE	VYKMGQFIEYS	SCIRFVPAD	-----	EDEENYIVFLPST	
DM_CG11864	:	NGIVN--QIY	WPNRT	VPYVLEDDAFA--DSHY	RELRASIEEN	SCVIFPKPATE	-----	MDFPMAIVITSKGL	
GM_M15242	:	NGLR--DEV	WPNNT	VYVYKFTVFD--EDH	NYILRGKIIIEE	ISCIREEADA	-----	TT--PNVYVITG	
DM_CG15253	:	NLWRN--ETY	WPNRI	IVYVFNYSID--EHR	NILAKIYKIESI	SCLIFKEAT	-----	TDQKYVYVNTSEAG	
DM_CG15254	:	NGLRN--ET	WPNRI	VYVYINRDID--TEH	RNHLRIRLIEQS	SCLIFKEAT	-----	TDQKYVYVNTSEAG	
DM_CG5715	:	NGILA--LSS	WPGGV	VPYIKGFTF--SQE	LGNHNAFKEYHT	TCVRFKPRT	-----	T--ERDYISISGGKS	
DM_CG6763	:	NGLPT--QSS	WPNV	VPYVIRG--NFN--	ARDMATAIGEYH	RRTCIR	VKRS	-----	S-KEDYISIRGDNS
C_F09E8.6i	:	ALNLVTY--PDK	WPEGO	VPYVLEEGMT--NDQ	RTAIAQAFDEYK	TKTCV	VPKT	-----	DDDFDYIVVKNVA
XL_UVS2	:	SAINDA--RFT	WPKSADGI	VPVYVNLSSYSN--ADQ	LALFKKATQEF	FEALTCV	VPW	-----	TTEVNFNLNMSG
AA_AEA	:	NAGG--SM	APKASEGF	IPYVVTQYD--KENI	IKLEKAFDSYRR	TCVRFVPR	-----	RTETDFLID--SRS	
OL_LCE	:	NAMKCFGAP	DSCRPKSSNGI	--VKVYV	VDNYSDEK	ETIRNAMKEFAE	TCIFVPR	-----	NNERAYLSLEPRF
OL_HCE1	:	NAMKCW--SSS	CFWKKASNGL	--VVIPYV	ISSEYS--GGE	VATEGAMRAFNG	TCIRFVRR	-----	TNEYDFISVVSKT
AJ_EHE4	:	NAMKCW--NNQ	CLWRKSSDGL	--VEVPYV	IVSNEFS--YHKK	RLEKAMNKT	FNTFTCIR	-----	SSQRDFISVSRD
AJ_EHE7	:	NAMKCR--NNR	CLWKKSSNGL	--VEVPYV	IVSRQFS--YYQ	KRVKAMKT	FNTFTCIR	-----	SRQRDYISIKSRG
CC_Neph	:	NADPCT--ARR	CKWRKSRNGL	--VYVPYV	IKISQYS--PDE	IKVKKQGLNS	FTGISCI	-----	NGQRDFLNIQSDS
TR_MP	:	NADACT--SRG	CMWAKSSDGL	--VYVPYV	IVASHYS--SRES	RTERGILSFHDV	TCIRFTPR	-----	RNQRDFLISVSDS
DM_CG10280	:	NPMRH--PKR	WPNNG	TIPEEISPRYA--NQE	RAIICAVKTFNSL	TCVIFVPYD	-----	GEVDDYLLIEPPL	
HS_HCH1	:	--LLSA--TSN	KWPMGGSGV	VEVPELSSKYD--EPS	RQVLEAEFER	STCIR	VTY	-----	QQRDFISIPMY
C_4R79.1i	:	NALKN--KQL	WEGG	IPYVEMDTAFS--PNE	IKLEKAFDSYRR	TCIRFEKR	-----	EGQDYVNIIVKGY	
C_C24F3.3	:	--GVS	IKGSMNRRNSNNI	--VPYV	ISPOYS--PAQ	QIIVSSLRVFER	VSCFKEVERTT	-----	QNDYLPVPLD
C_C45G2.1o	:	--GGV	VAAEKDWPNGR	--VPYV	ILSAAYT--SAQ	RADARAFDTYAK	TCIRFVPKS	-----	PADKDYIVIQKLD
C_C07D10.4i	:	NGVSR--AAK	LWPNAR	IPYVILSPHYS--PHER	ALAKAVQYHEK	TCIRFVPRQ	-----	TGEPDYVIFGKVD	
C_C34D4.9i	:	NGVIT--GTR	KWPNGR	IPYVILSNQYN--DRE	RAVARSFOAYHDK	TCVRFVVRT	-----	AVDNDYVYIGKID	
HV_HMP1	:	KGVT--LDR	WPGAQ	IPYVFDNFD--NRR	RQVWQATISY	NQHTCV	VPAPR	-----	TNQDFVHMVNEQ
C_T23H4.3	:	NALLSN--SPL	RWKMQLDGNYL	IPYVILSGNYD--TVER	DTIKTAMEKIANN	TCIRLIPR	-----	TNQDFVYAEINNKGGQ	
C_F56A4.1i	:	GIALHPLQ	WASYLWPNAE	--VPYV	ILATHYT--STEK	SILLSAMEAFKNV	TCVRFPRRA	-----	ATDKHYLQINKYFN
C_K06A4.1i	:	GIAIHPQ	WESHLPWNAE	--VPYV	ILASHYT--ATER	SILLSAMEAFRDV	TCVRFPRRR	-----	STDKHYLQINKHYQ
C_C05D11.6i	:	NAIKQ--I	YRRWPNNE	--IPYV	ILSSQYG--SYARS	VIANAMNEYHTK	TCVRFVARDP	-----	SKHDYVLIHHPDE
C_F39D8.4i	:	NAVRQT--Y	LKWEQAR	--IPYV	ILSSQYS--SYS	RSKIAEAEIEYR	TCIDFSPKS	-----	AGDLDYVHIYVDP
C_T04G9.2i	:	NRLQ--LW	PEGR	--IPYV	ILSSQYS--SYS	RSLAASQYASH	TCIRWVPKE	-----	AADVNYVHIYVDP
FM_Flavast	:	STIVS--SFI	KWPNATV	--YVTL	PSQSLSTQAYNT	LTKNKA	FDMISSKTSVKE	-----	TNQTETVITFTYST

Abbildung 29: Sequenzvergleich der Astacin-Protein-Familie. Legende siehe vorangegangene Seite.

	gc	s	g	100	120	c	HE	H	G	H	r	drd	140	160																																	
AA_AST	: -GCWS	-YVGRIS	-GA	-----	QQVSLQAN	-----	GCVYHG	---	T	I	H	E	I	M	H	A	I	G	F	Y	H	E	H	T	R	M	D	R	D	N	Y	V	T	I	N	Y	Q	N	V	D	P	S	M	T	S		
HV_HMP2	: -GCWS	-YLGVRG	-GL	-----	QNLSLDD	-----	GCWNGK	---	T	I	V	H	E	I	G	H	A	M	G	F	G	H	E	Q	N	R	P	D	R	D	C	Y	T	I	R	W	E	N	I	P	E	S	K	K	H		
HE_HEA2	: -GCSS	-PVGYRS	-GRM	-----	NRISLAG	-----	CCWNGK	---	I	V	M	H	E	I	G	H	S	I	G	I	Y	H	E	Q	S	R	P	D	R	D	Y	V	R	L	M	N	N	I	S	S	R	A	K	F			
PC_PMP1	: -GCSS	-PVGYRS	-GRQ	-----	NRISLAG	-----	CCWRTG	---	I	V	M	H	E	I	G	H	S	I	G	I	Y	H	E	Q	S	R	P	D	R	D	S	Y	V	E	I	V	W	G	N	I	P	S	S	K	K	F	
HS_MEPA	: -GCWS	-EVDQDQHV	-----	QNLSIGQ	-----	CCAYKA	---	T	I	H	E	I	H	A	L	G	F	Y	H	E	Q	S	R	T	D	R	D	D	Y	V	N	I	W	W	D	Q	I	L	S	G	Y	Q	H				
MM_MEPA	: -GCWS	-NIGDQVQV	-----	QNLSIGE	-----	CCDFKA	---	T	I	H	E	I	H	A	L	G	F	F	H	E	Q	S	R	T	D	R	D	D	Y	V	N	I	W	W	D	Q	I	L	T	D	Y	R	H				
HS_MEPB	: -GCWS	-SVGNRRVVK	-----	QELSIGA	-----	NCDRIA	---	T	V	Q	H	E	F	T	H	A	L	G	F	W	H	E	Q	S	R	S	D	R	D	D	Y	V	R	I	M	W	D	R	I	L	S	G	R	E	H		
MM_MEPB	: -GCWS	-SVGNHACK	-----	QELSIGT	-----	NCDRIA	---	T	V	Q	H	E	F	T	H	A	L	G	F	W	H	E	Q	S	R	A	D	R	D	D	Y	V	I	I	V	W	D	R	I	O	P	G	K	E	H		
AC_TBL1	: -GCSS	-FVKGHGSCA	-----	QVLSL GK	-----	CCYYFG	---	T	V	V	H	E	I	G	H	V	G	F	W	H	E	H	N	R	P	D	R	D	K	Y	V	O	I	R	K	N	I	M	P	G	K	E	S				
C_F38E9.2n	: -GCSS	-YVGRRGKGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	G	F	W	H	E	H	T	R	P	D	R	D	M	Y	V	D	I	F	Y	K	S	I	O	T	G	Q	D	Y			
SP_BMPH	: -GCSS	-FVGRKGDCA	-----	QAVSVGK	-----	NCDKFG	---	V	V	V	H	E	I	G	H	V	G	F	W	H	E	H	T	R	P	D	R	N	E	F	V	G	I	V	H	Q	N	I	V	P	G	Q	E	Y			
BR_TLD	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	I	G	F	W	H	E	H	T	R	P	D	R	D	H	V	S	I	R	D	N	I	P	G	Q	E	Y					
HS_TLL	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	I	G	F	W	H	E	H	T	R	P	D	R	D	N	H	V	S	I	R	E	N	I	P	G	Q	E	Y				
HS_TLL2	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	A	H	E	I	G	H	V	G	F	W	H	E	H	T	R	P	D	R	D	O	H	V	S	I	R	E	N	I	P	G	Q	E	Y					
XL_XOLLOID	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	G	F	W	H	E	H	T	R	P	D	R	D	H	V	S	I	R	E	N	I	P	G	Q	E	Y						
XL_BMP1	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	I	G	F	W	H	E	H	T	R	P	D	R	D	H	V	S	I	R	E	N	I	P	G	Q	E	Y					
HS_BMP1	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	G	F	W	H	E	H	T	R	P	D	R	D	R	H	V	S	I	R	E	N	I	P	G	Q	E	Y					
MM_BMP1	: -GCSS	-YVGRRGNGP	-----	QVLSL GK	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	I	G	F	W	H	E	H	T	R	P	D	R	D	H	V	S	I	R	E	N	I	P	G	Q	E	Y					
DM_TLD	: -GCSS	-FLGKNGNCR	-----	QVLSL GR	-----	NCEKFG	---	I	I	I	H	E	I	G	H	T	I	G	F	H	E	H	A	R	G	D	R	D	K	H	I	V	N	K	G	N	I	M	R	G	E	Y					
DM_TOK	: -GCSS	-FVGRKNGCP	-----	QVLSL GR	-----	NCDKFG	---	I	V	V	H	E	I	G	H	V	G	F	W	H	E	H	T	R	P	D	R	E	K	H	V	I	E	H	N	I	M	K	G	Q	D	Y					
PL_BP10	: -GCWS	-YVGRAFNGE	-----	QVLSL GN	-----	CCAYFG	---	T	I	V	H	E	I	G	H	A	I	G	F	H	E	H	E	Q	S	R	P	D	R	D	H	V	I	N	L	Y	Q	N	I	P	G	R	O	H			
SP_SPAN	: -GCWS	-HVGRSFTNQ	-----	QVLSL GP	-----	CCGYFG	---	T	I	V	H	E	I	G	H	A	I	G	F	H	E	H	E	Q	S	R	P	D	R	D	E	Y	I	N	V	H	F	E	N	V	Q	S	G	R	E	H	
C_R151.5	: -GCAS	-NVGRHPLCEE	-----	QVLSL AP	-----	ECIRIC	---	V	I	A	H	E	V	A	H	A	L	G	F	W	H	E	Q	S	R	P	D	R	D	C	Y	V	T	R	W	E	N	I	D	K	S	G					
C_HCH1	: -GCYS	-NIGKVSFRFP	-----	QVLSL GW	-----	CCSTIG	---	T	V	C	H	E	I	G	H	A	L	G	F	Y	H	E	Q	A	R	Y	D	R	D	V	S	I	L	T	Q	N	I	Q	D	M	Y	L	S				
C_F58B4.1	: -GCYS	-YVGRIS	-GV	-----	QVLSL GT	-----	CCEEFG	---	T	A	A	H	E	I	G	H	A	L	G	F	F	H	T	Q	S	R	Y	D	R	D	N	Y	I	S	I	N	Y	A	N	I	D	P	S	Y	V	E	
ACA_MTP1	: -GCWS	-YVGRIG	-GE	-----	QVLSL GE	-----	CCQSVG	---	T	A	A	H	E	I	G	H	A	I	G	F	Y	H	T	H	A	R	D	R	D	N	F	I	T	F	N	A	Q	N	V	K	P	D	W	L	S		
OO_MP	: -GCWS	-NIGKIG	-GE	-----	QVLSL GG	-----	CCHSVS	---	I	A	A	H	E	I	G	H	A	I	G	F	F	H	T	M	S	R	H	D	R	D	F	I	T	W	M	N	H	V	D	W	L	S					
C_F46C5.3n	: -GCSS	-PIGRQQIT	-----	QVLSL TK	-----	NCWGMG	---	T	A	I	H	E	I	M	H	A	I	G	I	B	T	Q	S	R	S	D	R	N	R	Y	L	D	I	L	A	Q	N	I	D	N	R	D	L	P			
C_F20G2.4i	: -HCRS	-TI GAPG	-TRS	-----	GSAGELSMETG	-----	WCANIC	---	S	I	V	H	E	F	S	H	S	L	G	R	Y	H	E	H	T	R	P	D	R	D	N	S	L	K	V	T	S	T	D	Y	E	A	R	P	--		
C_K03B8.5	: -LCQS	-YV GAPG	-MSMS	-----	KPIGELSMIVQ	-----	SCDTIC	---	S	I	V	H	E	F	S	H	S	L	G	R	F	H	E	H	T	R	P	D	R	D	H	V	S	I	T	W	M	N	H	V	P	H	E	A	R	P	--
C_K03B8.2	: -GCAS	-YI GMNG	-GE	-----	QVLSL FGD	-----	CCLIIFG	---	T	A	V	H	E	I	M	H	S	L	G	L	F	H	T	H	S	R	F	D	R	D	N	F	L	S	Y	K	D	V	P	E	N	M	V	G			
C_K03B8.1	: -FCAS	-YV GMIN	-SV	-----	QVLSL FPD	-----	WCMRFG	---	S	A	V	H	E	I	M	H	A	L	G	V	I	H	T	H	A	R	D	R	D	N	F	L	N	V	N	L	N	L	N	K	D	E	D	D	S		
C_K03B8.3	: -GCSS	-YI GMNN	-IV	-----	QVLSL FND	-----	CCIDFG	---	T	A	V	H	E	I	M	H	A	L	G	V	I	H	T	H	S	R	L	D	R	D	N	F	L	N	L	N	T	N	S	K	E	M	H				
C_T02B11.7n	: -GCYS	-DT GMLG	-GE	-----	QVLSL GY	-----	CCCVTG	---	T	A	A	H	E	I	A	H	T	L	G	L	F	H	T	Q	M	R	S	D	R	D	Y	V	T	I	D	L	T	D	V	P	E	S	S	Q			
C_T11F9.3i	: -GCYS	-LY GMHA	-GEV	-----	QVLSL TDY	-----	NCASFG	---	T	A	V	H	E	I	M	H	A	L	G	T	A	H	Q	A	R	S	D	R	D	D	Y	L	I	V	D	S	T	N	S	N	D	G	I	E	N		
C_D1022.m	: -GCWS	-YVGRDEAOGA	-----	QVLSL GT	-----	CCEMFG	---	I	T	S	H	E	I	A	H	A	L	G	L	F	H	E	H	Q	S	R	Y	D	R	D	V	O	I	K	S	R	I	A	Q	T	N	F	Y				
C_F58A6.4i	: -GCWS	-YVGRDASQOK	-----	QVLSL GN	-----	CCHEFG	---	V	T	S	H	E	I	A	H	A	L	G	L	F	H	E	H	Q	S	R	F	D	R	D	E	S	V	F	N	P	R	V	E	R	D	L	L	F			
C_T24A11.3i	: -SCFT	-EYIGRNG	-GH	-----	QVLSL IG	-----	CCAEYF	---	V	A	A	H	E	I	G	H	A	L	G	F	W	H	T	H	O	R	P	D	R	D	R	H	S	I	N	N	K	N	V	M	E	E	A	T	A		
C_K04E7.3	: -GCAS	-YVGRIG	-GP	-----	QVLSL GD	-----	CCETIC	---	T	I	T	H	E	V	G	H	A	L	G	F	W	H	E	Q	A	R	P	D	R	D	S	Y	V	R	L	N	I	N	S	K	E	M	L	E	G	L	
C_C17G1.6i	: -GCWS	-NVGRVG	-GR	-----	QVLSL GY	-----	CCDSIG	---	I	V	S	H	E	I	T	H	A	L	G	L	W	H	E	Q	S	R	D	R	D	N	F	I	S	I	V	A	D	K	I	R	G	T	E	G			
C_Y95B8A.1	: -GCYS	-SVGRIG	-GS	-----	QVLSL GY	-----	CCEDKG	---	I	V	A	H	E	V	G	H	S	L	G	F	W	H	E	Q	S	R	P	D	R	D	Y	I	H	R	K	D	M	I	K	G	T	D	G				
C_C26C6.3	: -GCYS	-MIGRNG	-GR	-----	QVLSL GESCVK	-----	ASFFVMG	---	V	I	E	H	E	I	G	H	A	L	G	L	W	H	E	Q	S	R	P	D	R	D	C	Y	T	I	R	D	F	L	P	S	Y	I	S				
C_F57C12.1i	: -GCSS	-FVGRVG	-GT	-----	QVLSL IST	-----	PCDDVVG	---	I	I	S	H	E	I	G	H	A	L	G	L	F	H	E	Q	A	R	P	D	Q	E	R	H	I	A	N	Y	N	N	I	P	L	S	R	W	N		
C_F42A10.8m	: -GCWS	-YV GKQVDMPY	-----	QVLSL VGP	-----	NCDTFG	---	T	A	T	H	E	I	M	H	A	I	G	F	W	H	E	Q	S	R	A	D	R	D	N	Y	V	I	D	F	S	N	I	P	S	Q	A	Y				
C_T23F4.4	: -GCWS	-FVGRQPSLRE	-----	QVLSL ER	-----	CCDTHF	---	V	A	A	H	E	I	A	H	T	L	G	F	Y	H	E	H	A	R	G	D	R	D	O	F	I	S	D	S	N	I	P	N	L	T	F					
OL_Alveo	: -GCAS	-YV GCSG	-GA	-----	QVLSL YFGS	-----	TCS_VG	---	N	I	C	H	E	I	M	H	A	V	G	M	Y	H	E	H	T	R	Y	D	R	D	C	F	I	S	V	L	W	D	N	I	K	P	G	K	Q		
DM_CG11865	: -GCWS	-YVGYLGRAD	-----	QVLSL NGS	-----	CCMSNR	---	T	I	Q	H	E	I	H	A	L	G	F	F	H	T	H	S	D	P	Q	R	D	K	Y	V	R	I	Q	D	N	I	R	S	G	H	E	K				
C_K09C8.3	: FCGLS	-YVGRVEPA	-----	NPVYLSF	-----	CCGNGRG	---	I	A	V	H	E	I	M	H	A</																															

	f	yd	S	mhY	t	g	d	n	y	c
AA_AST	: NFDI-DTYSR--YVGED	YQYYSIMHYGKYS	FSIQW----	GVLETTIVPLQN--	GIDLTD	PDYK	KAHMLQ	TANQ	INNYTNE	CSLR
HV_HMP2	: NERL-YNSLVDLSLNSP	YDYSIMQYSKTAF	GGIND----	SVLIDPKLP--	GIF--	OLGOR	VGFTEHD	OYQAMOLYR--	COGK	
HE_HEA2	: NENK-QRTSNINSLGTP	YDFRSMHYSSGAF	GGG-----	RRFTEVKDSR--	TPNIM	SGHSA	IDTKQ	TIMY--	CGGS	
PC_PMP1	: NENK-KSSSSINSLGTP	YDFRSMHYSTTA	IGGG-----	RITIRTKNPS--	NQRIL	LGOR	SGHSE	IDTKQ	INLMY--	CNGV
HS_MEPA	: NEDT-YDDSLITDLNTP	YDYESLMHYQPF	FNKNA-----	SVPTITAKIPE--	FNSI	IGOR	LDFSA	ADLER	INRMYN--	CTTT
MM_MEPA	: NENT-YDNTITDLNTP	YDYESLMHYGPF	FNKNE-----	SIFITTKIPE--	FNTI	IGOL	PDFSA	IDIR	INRMYN--	CTAT
HS_MEPB	: NENT-YSDDISDSLNV	YDYISVMHYSKTAF	FQNGT-----	EPTIVTRISD--	FEDVI	IGORM	DFSD	DLK	LNQLYN--	CSSS
MM_MEPB	: NENI-YNDSVSDSLNVP	YDYISVMHYSKTAF	FQNGT-----	ESTIVTRISE--	FEDVI	IGORM	DFSD	DLK	LNQLYN--	CTSS
AC_TBL1	: ENNI-LDEDKVDLSGEP	YDYSIMHYSRDK	FSKHS-----	YLDTRPFRQRMIAL	PRIC	ONIR	INDG	VRO	INRKYK--	CPTC
C_F38E9.2n	: NEFK-SKPEEVDLSGEP	YDFSSIMHYARDT	FSRGA-----	FYDTILPKPNSG--	FRLE	IGOR	VQLSEG	IRQT	KKLYK--	CAEC
SP_BMPH	: NERV-LDAAEVDSLGET	YDFASIMHYARNT	FSRGI-----	WLDTILPRKDPE	SGIRPE	IGOR	KHLS	EGD	IIQAN	ILYK--
BR_TLD	: NERK-MEPGDVNSLGET	YDFDLSIMHYARNT	FSRGM-----	FLDTILPSRD--	ENGR	VPAL	SGOR	TRLS	KGDI	IAQARKLYR--
HS_TLL	: NERK-MEPGEVNSLGER	YDFDLSIMHYARNT	FSRGM-----	FLDTILPSRD--	DNGI	RPA	IGOR	TRLS	KGDI	IAQARKLYR--
HS_TLL2	: NERK-MEAGEVSSLGET	YDFDLSIMHYARNT	FSRGM-----	FLDTILPRQD--	DNGV	RPI	IGOR	TRLS	QGD	IAQARKLYR--
XL_XOLLOID	: NERK-MEPGEVSSLGET	YDFDLSIMHYARNT	FSRGM-----	FLDTILPRRI--	DTSV	RPI	IGOR	TRLS	QGD	IAQARKLYK--
XL_BMP1	: NERK-MEPPEVESLGET	YDFDLSIMHYARNT	FSRGI-----	FLDTILPKYD--	VNGV	RPE	IGOR	TRLS	SSG	VAQARKLYK--
HS_BMP1	: NERK-MEPQEVESLGET	YDFDLSIMHYARNT	FSRGI-----	FLDTIVPKYE--	VNGV	KPE	IGOR	TRLS	SGD	IAQARKLYK--
MM_BMP1	: NERK-MEPQEVESLGET	YDFDLSIMHYARNT	FSRGI-----	FLDTIVPKYE--	VNGV	KPE	IGOR	TRLS	SGD	IAQARKLYK--
DM_TLD	: NEDV-LSPPEVDLPLLP	YDLNSIMHYAKNS	FSKSP-----	YLDITPIGI--	PPG	THLE	IGOR	KRL	LSRQD	IVQAN
DM_TOK	: NENM-LSPDEVDSLGMA	YDYDSIMHYARNT	FSKGT-----	YLDITLPIEM--	KGRK	RPE	IGOR	LR	LSQD	IAQAN
PL_BP10	: NERK-YTWGRVTSRNV	YDYSIMHYGYG	FSNG-----	RFTITTRDPR--	LNSR	IGOR	TAL	SPAD	IELAN	LIYE--
SP_SPAN	: NERK-YTWGSVTSNV	YDYSIMHYGYG	FSNG-----	RFTITTRDPR--	LNSR	IGOR	TAL	SAAD	IELAN	LIYE--
C_R151.5	: QERK-EDPDDVDNAGVP	YDYGSIMHYRSKA	FSKFD-----	DLYITSTYVD--	YQKI	IGOR	DOL	SFND	IRL	INRKYK--
C_HCH1	: QERK-QASSMVDYGVG	YDYGSIMHYDQA	AFSST-----	GGNTIATRDPN--	FQAI	IGOR	VAF	SFAD	VKR	INRKYK--
C_F58B4.1	: QERK-ETSNTNFYGM	YDYGSIMYAGTAS	ASN-----	DKATMIARDE--	YQD	IMG	SDF	VGF	YD	ISM
ACA_MTP1	: QERK-QTPATNENYGIT	YDYGSIMHYGANS	AFSON-----	GRFTMVPHPDK--	YVE	ILG	SP	IS	FYB	ELM
OO_MP	: QERK-ETTRNDNYGMT	YDYGSIMHYGDT	SASYN-----	NKFTMVPFVD--	YQQL	IG	SP	IS	FYB	ELM
C_F46C5.3n	: NEFK-LLSPRLWANLVP	YDYGSIMHYADS	FSNK-----	DDEQIMLPKDRS--	FIE	TMG	SMI--	PNFY	EDQ	INQYQY--
C_F20G2.4i	: -----PWGTTMYGP	FEHGSIMHYH--	SSNY-----	GVGKMEPYDME--	YKN	IMG	SRR--	VTF	YDM	KINQYQY--
C_K03B8.5	: -----PSGTTMYGP	FEHGSIMHYH--	SSNY-----	TYGPTMDPLMD--	YKQ	IMG	NRN--	VTF	YDM	KINQYQY--
C_K03B8.2	: NLE--KETEQTYNVAVP	FEYGSIMLYRYN--	-----	TFGEGILVSKNED--	YQK	IMG	LRR--	VSF	YD	IVN
C_K03B8.1	: NFE--IVSPPFSINVVP	FEYGSIMLYRYN--	-----	VSGINSLPKQME--	YRIL	GNRR--	VTF	YD	MLT	INTAYN--
C_K03B8.3	: NYA--IFN--QSTNVVP	FEYGSIMHYAD--	-----	ISTMFPKKE--	YSA	ILG	IGR--	VTF	YD	MLT
C_T02B11.7n	: NERK-LTBEATSTNLVD	YDYGSIMHYGRAN	PT-----	YGGVDSIVPKDPV--	MVY	TMG	RI--	VTF	FLD	KMLN
C_T11F9.3i	: -----TENLVP	FDYGSIMLYARDP	H-----	SKRILPIDPE--	YNF	TMG	SLR--	VAF	YD	MLT
C_D1022.m	: DEAV-VGKKNMETYGQK	YDYSIMHYRTE	FS-----	LDGGNSIAKDVN--	MQNT	IMG	RFG	SE	ID	VAK
C_F58A6.4i	: NERK-ISPROMSTYGLP	YDYSIMHYRTE	FS-----	IPSIITLAAIDTN--	LQT	IMG	LE	GF	SF	V
C_T24A11.3i	: SFMA-FGIRQVSPRRVP	YDYGSIMHYHTVA	HAV-----	KVSDFTVPKELK--	YVT	IMG--	TEK	MA	F	LD
C_K04E7.3	: QERK-RSWSEVNEYSLP	YDYGSIMHYGPK	FSK-----	SSTMTNVEPDDPA--	FINT	IGNR	VE	ES	FLD	TK
C_C17G1.6i	: NERK-RTAANSNDLQGP	YDLGSIMHYGAKS	FAY-----	DWSSDITKTRDWR--	YQNT	IGOR	DGL	SF	KD	AKM
C_Y95B8A.1	: NERK-RSWEIEDMGVVP	YDYGSIMHYGSNA	FATK-----	DWDQITVETKDSR--	YQNT	IGOR	QK	SF	ID	VQ
C_C26C6.3	: DEKL--QRDEIDTLGIP	YDLGSIMHYGSNA	FATK-----	DQKSTIVTRDSL--	YQQI	IGOR	KLS	F	VD	AT
C_F57C12.1i	: NERK-VGENHAETYNLP	YDYSIMHYGYPG	FAS-----	DPYTEITRTERLV--	QQST	IGOR	AGE	SF	YD	QAN
C_F42A10.8i	: NERK-MAVDQAQLNLP	YDYGSIMHYYPY	FAV-----	DSSKYITLAKENG--	FQNS	MGOR	EAP	SF	IG	VN
C_T23F4.4	: NERK-ESEKQLDHEAA	YDYGSIMHYSDQ	FAV-----	NTNRPVLYARDQK--	FQAQ	MGNR	RAF	TQ	DV	SR
OL_Alveo	: NEDV-KMG--DTLNLVP	YDYGSIMHYGTSY	FASK-----	DGSEITQPNEG--	VN	IGOR	KRL	SQD	IQK	IN
DM_CG11865	: NERK-LRANGVTNYGFG	YDYDSIMHYGPF	FSK-----	NGQSTIVPLK--	SHAK	IGQAT	MSP	KD	Q	TR
C_K09C8.3	: AERV-ADSKLYTYGVK	YDYDSIMHYNYAT	TGAV-----	NIAKITMPLVNQQA--	NIGL	IGOR	AK	MS	NAD	VEI
C_C37H5.9i	: TBAI-SDAKLYTSYGTK	YDYDSIMHYNYAL	LGAK-----	DPNKITMPLVNPQE--	NTPK	IGOR	AK	L	TR	DL
C_K11G12.1	: AERV-VDSKLYTSYGK	YDYDSIMHYNYTA	QAQ-----	NIATMNPKTNSAV--	NLKV	IGOR	QK	M	TR	DL
DM_CG6696	: NERK-YARTHITNFVGE	YDYGSIMHYSSRA	FSK-----	NGKATIEPLD--	PYAS	IGOR	RGL	S	D	V
DM_CG15255	: NERK-YADTVVDFEYV	YDYDSIMHYRPGA	FAS-----	NGEDITVPLD--	SSAV	IGOR	VGL	S	S	D
DM_CG7631	: NERK-YEEDEVGDFDQP	YDYGSIMHYSSRA	FSK-----	NGEATIVALNPE--	GQE	IGOR	L	M	S	D
DM_CG11864	: NERVNNDNSTAWHDFEG	YDYESVMHYVPRA	FSR-----	NGQITVPLRE--	GAEN	IGOR	F	M	S	E
GM_M15242	: NERK-YNETQVNDFDQE	YDYGSIMHYGPKA	FSK-----	NGEDITVPLRE--	AAGN	IGOR	L	M	S	E
DM_CG15253	: NERK-YTETVNDFGEK	YDYGSIMHYGPF	FSK-----	NGERTIVALLEG--	KEDV	IGOR	L	M	S	E
DM_CG15254	: NERK-YDNETVEDYGE	YDYSSVLHYTAY	FAV-----	NGEMITVPLQEG--	AEE	IGOR	L	M	S	E
DM_CG5715	: NERK-SSRTQYGFVGE	YDYGSIMHYSP	FTS-----	NGQPTIKALRATS--	DASO	IGOR	K	G	S	D
DM_CG6763	: NERK--AARTEA-FGV	YDYGSIMHYSKNA	FSK-----	NGQPTILAMQAN--	GADK	IGOR	N	G	S	D
C_F09E8.6i	: NERK-YPRKIIDSGLMP	YDYESVMHYHKL	FAV-----	RNGKPTIIPKD--	NEAD	IGOR	Y	L	S	E
XL_UVS2	: NERK-YNT--NNLGI	YDYASVMHYSR	HYSI-----	NGDITIEPKD--	PNVE	IGOR	D	L	S	E
AA_AEA	: NERK-EMN--MQLGLP	YDYTSIMHYARY	VYSI-----	EGDESIDDPAN--	GNVP	IGOR	D	L	S	E
OL_LCE	: NERK-NDT--DNLGTP	YDYGSIMHYGR	TAFK-----	DRKETITPPIPN--	PKAA	IGOR	T	M	S	D
OL_HCE1	: NERK-HDT--NNLNT	YDYSSIMHYGR	DAFSI-----	AYGRSITPPIPN--	PNVE	IGOR	N	G	S	D
AJ_EHE4	: NERK-QDT--NNLNT	YDYTSIMHYGR	TAFS-----	NGMDITTPVPIPN--	PNQ	IGOR	N	G	S	D
AJ_EHE7	: NERK-QNT--NNLNT	YDYTSIMHYG	KAAFS-----	NGKDTITPPIPN--	PKQ	IGOR	R	S	M	S
CC_Nephr	: NERK---RKTNNL	YDYSVMHYSR	FAFSK-----	NEPTMIPPIPN--	RNVV	IGOR	E	A	R	K
TR_MP	: NERK-MNT--LNLNT	YDYSVMHYHRY	AFS-----	NGKPTMAPIPN--	ANIQ	IGOR	E	A	R	K
DM_CG10280	: NERK-VS-KKKGKYAFD	YDYSVMHYGE	FYSK-----	RKGEKPTMPLQ--	PGV	IGOR	K	T	S	K
HS_HCH1	: NERK-QSS--NMLTP	YDYSSVMHYGR	LAFS-----	RGLTITPPLWA--	PSVE	IGOR	N	L	S	D
C_4R79.1i	: QERK-ISAVLQDLQGEN	YDYKSIMHYD	STAFS-----	RNGRITIEVENG--	FTQV	IGOR	T	M	S	D
C_C24F3.3	: NERK-LKTSH-VEYPTI	YDFKSIMHYD	GAYFGRVDTARRVRLA	TMTPLKPG--	VIL	EDNM	K	F	T	A
C_F45G2.1o	: QERK-LSNLGLSYYGEP	YDYSSIMHYE	ANEGS-----	RNGKNTIEAKNSH--	FTA	IMG	K	S	D	S
C_C07D10.4i	: QERK-VLDSKTSYYGEP	YDYKSIMHYD	SLAFS-----	KNGFTMLPKV--	KSAT	IGOR	N	A	R	D
C_C34D4.9i	: QERK-VDLAESSYYGQL	YDYYSIMHYD	SLAFS-----	KNGFTMVAQOSE--	MTAV	IGOR	G	A	I	D
HV_HMP1	: NERK-YRT--EYQG	YDHSIMHYR	NDEFA-----	ANGKNTIAQAN--	NPGL	IGOR	K	L	G	N
C_T23H4.3	: QERK-LSSRDATTYSVP	YDYSVMHYD	ENAFS-----	KPKGISMMTKDSK--	FQKV	IGOR	T	M	S	D
C_F56A4.1i	: NERK-I-YRRA-KTLYMGA	YDANSIMHYNF	QNLP-----	WQRR-----	DHF	ST	S	D	I	N
C_K06A4.1i	: NERK-I-YORA-KSYMGG	YDANSIMHYNF	GSVP-----	WQRR-----	DYH	SP	S	D	I	N
C_C05D11.6i	: QERK-YNLNVI	SHLDEP	YDYASIMHYG	PYAFS-----	GSGK	IGOR	V	P	K	S
C_F39D8.4i	: QERK-YSLNMIDHLG	YDYGSIMHYA	PYAFS-----	KNGKPTIEPIE--	KNVE	IGOR	A	G	S	E
C_T04G9.2i	: QERK-YGHGTIQSLG	YDYGSIMHYG	TAFS-----	RNGQPTMIPK--	NGAT	IGOR	N	G	S	D
FM_Flavast	: NERL-YND--YAGHGE	YDFGSVMHYK	STDAIDP-----	NLPVMTKLDG--	STFG	IGOR	D	L	S	A

Abbildung 29: Fortsetzung

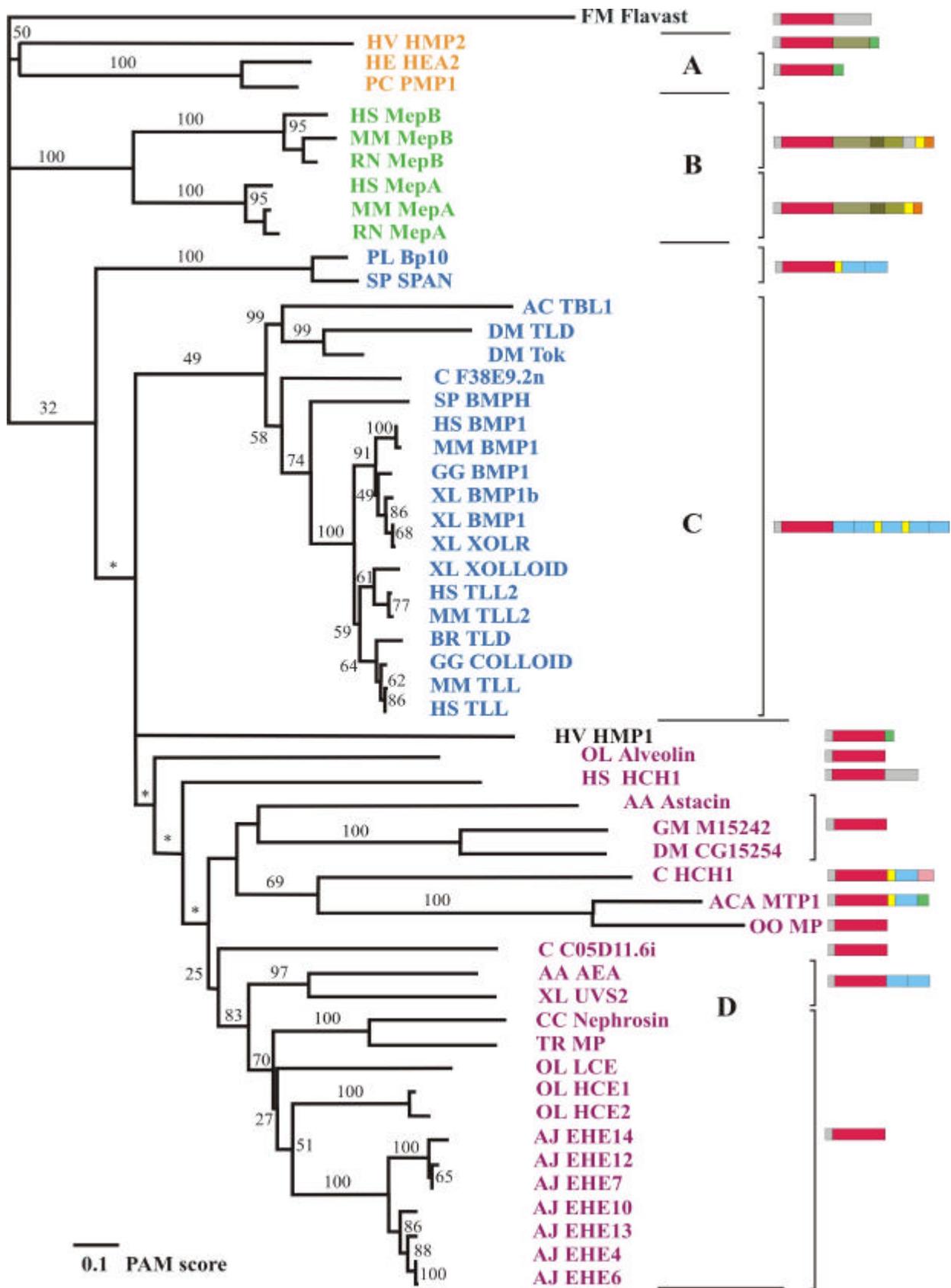


Abbildung 30: Stammbaum der Astacin-Protein-Familie, erstellt nach der *Neighbor-joining* Methode. Der besseren Übersichtlichkeit wegen wurden bei *C. elegans* und *Drosophila* nur einige Proteine exemplarisch berücksichtigt. *Bootstrap*-Werte (Maß für die statistische Unterstützung, siehe III.12.3) sind an den Verzweigungspunkten angegeben, wobei Werte <20 durch einen Stern (*) dargestellt sind. A bis D bezeichnen die im Text näher beschriebenen Stammbaum-Gruppen. Zur Erläuterung der Domänenstruktur am rechten Bildrand siehe Abbildungen 26 und 27. Die Protein-Bezeichnung ist in Tabelle 12 erläutert. PAM = *Accepted Point Mutations*.

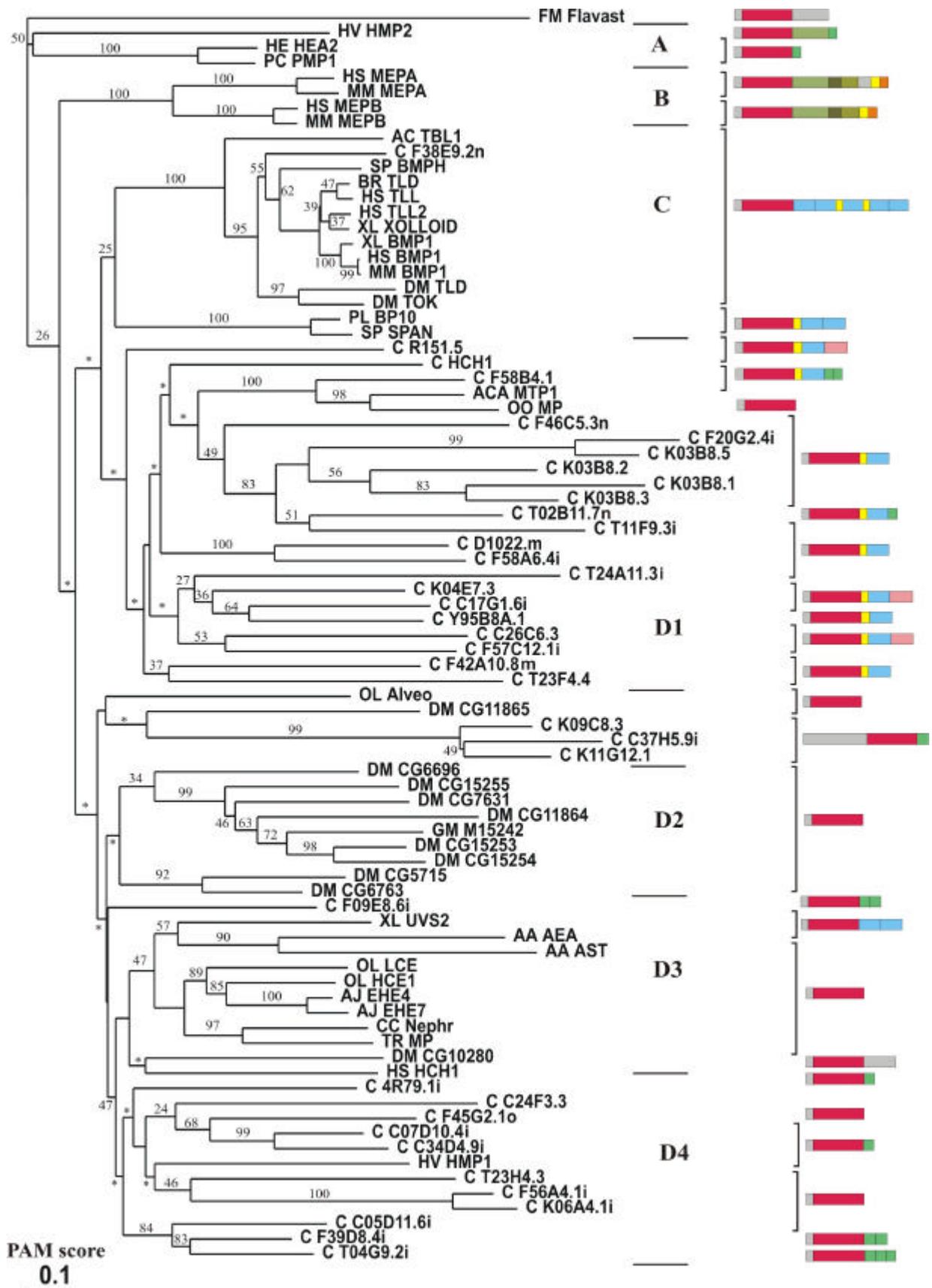


Abbildung 31: Stammbaum der Astacin-Protein-Familie erstellt nach der *Neighbor-joining* Methode. Alle *C. elegans* und *Drosophila* Proteine wurden in die Analyse einbezogen. Die *Bootstrap*-Werte (Maß für die statistische Unterstützung, siehe III.12.3) sind an den Verzweigungspunkten angegeben, wobei Werte <20 durch einen Stern (*) dargestellt sind. A bis D bezeichnen die im Text näher beschriebenen Stammbaum-Gruppen. Zur Erläuterung der Domänenstruktur am rechten Bildrand siehe Abbildungen 26 und 27. Die Protein-Bezeichnung ist in Tabelle 12 erläutert. PAM = *Accepted Point Mutations*.

Die BMP-1/Tolloid-ähnlichen Proteine bilden eine Stammbaum-Gruppe (Abbildung 30, Gruppe C) mit gleicher Domänenstruktur und werden bei Nematoden (*C. elegans* F38E9.2), Mollusken (TBL-1), Insekten (Tolloid, Tolkin), Seeigeln (BMPH) und Vertebraten (BMP-1, TLL, TLL-2, Xolloid, Colloid) gefunden. Diesen stehen die Proteine BP10 und SPAN des Seeigels nahe, die strukturell als verkürzte BMP-1/Tolloid-Verwandte angesehen werden können.

Vor der nächsten großen Stammbaum-Gruppe der Astacine (Gruppe D), die funktionell überwiegend als *hatching*-Enzyme charakterisiert werden können, erscheint HMP-1. Auch wenn dieser Ast des Stammbaums nach der statistischen Überprüfung nicht eindeutig zugeordnet werden kann, scheint dieses Protein offensichtlich stark von den übrigen Cnidaria Astacinen (Gruppe A) abzuweichen.

Allgemein scheint innerhalb der Gruppe D, aufgrund der zum Teil schlechten statistischen Unterstützung der Verzweigungen, eine phylogenetische Interpretation schwierig. Dennoch zeigen die *hatching* Enzyme der Fische (LCE, HCE, EHE, Nephrosin) mit den *hatching* Enzymen von *Astacus* (AEA) und *Xenopus* (UVS-2) hohe Übereinstimmungen. Diese scheinen ebenfalls mit dem Verdauungsenzym Astacin des Flusskrebsses, zwei Insektenproteinen (M15242, CG15254) oder einem *hatching*-Enzym von *C. elegans* (HCH-1) strukturell eng verwandt zu sein. Bemerkenswert ist noch, dass das neu identifizierte Astacin-homologe Protein des Menschen HCH-1 ebenfalls in dieser Gruppe zu finden ist. Die *hatching* Enzyme der Fische, Astacin, Alveolin und die zwei Insektenproteinen (M15242, CG15254) zeigen dabei keine carboxyterminalen Anhänge, wohingegen die *hatching*-Enzyme AEA (*Astacus*), UVS-2 (Seeigel), und HCH-1 (*C. elegans*) sowie die Proteine MTP-1 (*Ancylostoma*) oder HCH-1 (Mensch) unterschiedliche Domänenanordnungen besitzen.

Bezieht man in die phylogenetische Analyse nun alle Astacin-homologen Proteine von *C. elegans* und *Drosophila melanogaster* ein, ergibt sich prinzipiell das gleiche Bild (Abbildung 31). Die Gruppen der Cnidaria-Astacine (Gruppe A), der Meprine (Gruppe B), der BMP-1/Tolloid homologen Proteine mit BP-10 und SPAN (Gruppe C) unterscheiden sich klar von der großen Gruppe (D1 bis D4) der restlichen Astacine. Aufgrund der zum Teil geringen Sequenzidentität, verbunden mit einer schlechten statistischen Unterstützung der Verzweigungen, ist eine vollständige phylogenetische Interpretation innerhalb der Subgruppen D1 bis D4 wiederum schwer nachvollziehbar. Dennoch zeichnen sich einige wichtige Tendenzen ab. Unter D1 ist eine Subgruppe von *C. elegans* Proteinen zusammengefasst, die auch hinsichtlich ihrer Domänenstruktur große Übereinstimmungen (EGF, CUB, ShKT) zeigt und zu der beispielsweise auch das HCH-1 Protein von *C. elegans* gehört. In Subgruppe D2 finden sich die Mehrzahl der kurzen, kompakten *Drosophila*-Proteine. In Subgruppe D3, sind wie für Abbildung 30, Gruppe D, beschrieben, die *hatching* Enzyme der Fische, die von *Astacus* (AEA) und *Xenopus* (UVS2), sowie das Astacin des Flusskrebsses und das Nephrosin repräsentiert. Wiederum scheint das menschliche HCH-1 Protein dieser Gruppe nahezustehen. HMP-1 bildet in diesem Stammbaum zusammen mit einigen weiteren *C. elegans* Proteinen unterschiedlicher Domänenstruktur eine eigene, klar abgegrenzte Subgruppe (D4). Das Alveolin, das *Drosophila* Protein CG11865 und die *C. elegans* Proteine K09C8.3, C37H5.9 und K11G12.1 können keiner der genannten Subgruppen eindeutig zugeordnet werden.

Die verschiedenen Stammbaum-Gruppen (Gruppen A bis D) der Astacine sind nach diesen Ergebnissen wahrscheinlich durch mehrere Genduplikationen entstanden und haben sich als orthologe Proteine getrennt weiterentwickelt. Dabei kann eindeutig zwischen den Cnidaria-Astacinen (Gruppe A), den Meprinen (Gruppe B), den BMP-1/Tolloid-ähnlichen Proteinen (Gruppe C) und einer heterogenen Restgruppe (Gruppe D), welche die *hatching* Enzyme, Astacin und die Mehrzahl der *C. elegans* und *Drosophila* Proteine einschliesst, unterschieden werden. Darüber hinaus kann festgestellt werden, dass es sich bei den *C. elegans* Proteinen der Subgruppen D1 und D4 (Abbildung 31), der *Drosophila* Proteine der Subgruppe D2 (Abbildung 31), sowie der *hatching*-Enzyme von *Anguilla* (EHE4,6,7,10,12,13,14; Abbildung 30 D) offenbar um paraloge Proteine handelt, die durch zahlreiche Genduplikationen entstanden sind. Die Domänenanordnung scheint innerhalb der Gruppen A bis C annähernd konserviert zu sein, ist aber in der Gruppe D und vor allem bei den paralogenen Genen von *C. elegans* sehr variabel. Bemerkenswert ist letztlich noch die Tatsache, dass von *C. elegans* und *Drosophila* in der Gruppe D eine große Anzahl von Genkopien, bei den BMP-1/Tolloid-ähnlichen Proteinen aber nur jeweils ein (F38E9.2) bzw. zwei (Tolloid und Tolkin) Vertreter gefunden werden. Umgekehrt besitzt der Mensch offensichtlich mehrere Genkopien für BMP-1/Tolloid-ähnliche Proteine (TLL, TLL2, BMP-1), aber nur ein Gen, das für eine kurze, kompakte Protease kodiert (humanes HCH-1).

4. Aktivierung von Pro-Astacin

Der Aktivierungsmechanismus der Pro-Enzyme der Astacin-Protein-Familie wurde am Modell des Astacins aus *Astacus Astacus* untersucht. Dazu sollte der Synthese- und Aktivierungsort von Pro-Astacin im Flusskrebis mit Hilfe spezifischer Antikörper lokalisiert werden. Darüber hinaus sollten die an der Aktivierung beteiligten proteolytischen Enzyme durch Spaltung und massenspektrometrische Analyse verschiedener Substrat-Modellpeptide charakterisiert werden.

4.1. Immunhistochemie

Um einen Überblick über den Gewebeaufbau des Verdauungstraktes von *Astacus astacus* zu erhalten, wurden vor Beginn der immunhistochemischen Experimente zunächst verschiedene histologische Präparate angefertigt. Die morphologische Beschreibung bezieht sich dabei auf Arbeiten von (Vogt et al. 1989; Vogt 2002).

Das Verdauungssystem von *Astacus astacus* besteht aus Darm und Anhangsdrüsen (Abbildung 32). Der Darm setzt sich aus Vorderdarm, Mitteldarm und Enddarm zusammen. Der Vorderdarm beginnt mit dem Mund und einem kurzen Ösophagus. Posterior schließen sich zwei Kammern an, deren spezielle Vorrichtungen die Nahrung zerkleinern und filtern (Cardia=Kaugagen und Pylorus=Filtermagen). Die Mitteldarmdrüse (Hepatopankreas) mündet kurz vor dem Mitteldarm in das Atrium, einen Übergangsbereich zwischen Mitteldarm und Pylorus. Resorption und Produktion der Verdauungssäfte finden ausschließlich im Hepatopankreas statt. In der Mitteldarmdrüse laufen dementsprechend zwei Flüssigkeitsströme: der Verdauungssaft Richtung Cardia und der Nahrungssaft zu den resorbierenden Zellen. Mitteldarm und Enddarm sind sehr kurz und enden mit dem Anus.

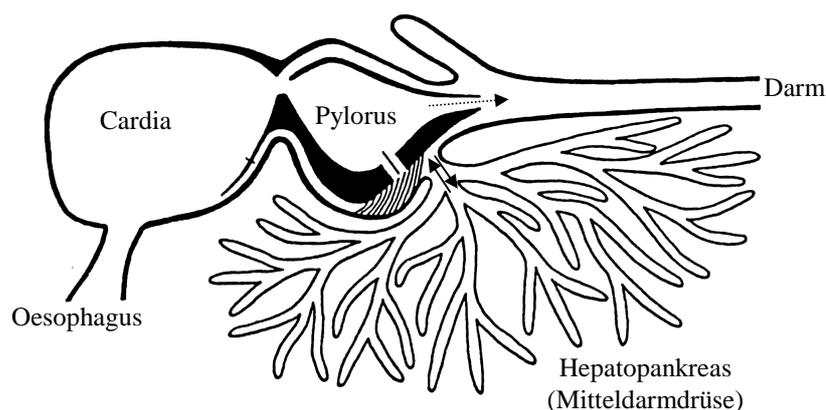


Abbildung 32: Schematische Darstellung des Verdauungstraktes und des Weges der Verdauungsflüssigkeiten von *Astacus astacus*. Durchgezogene Pfeile: Weg der im Magen verdauten Nahrungsflüssigkeit (Richtung Hepatopankreas) und der sezernierten Verdauungsenzyme (Richtung Pylorus/Cardia). Gestrichelter Pfeil: Weg der festen, unverdaulichen Bestandteile (modifiziert nach Vogt et al. 1989).

In den immunhistochemischen Versuchen zeigte der Anti-Astacin Antikörper, nach Stimulierung der Enzymneusynthese durch Entnahme von Magensaft (vergleiche III.10.1), eine Astacin-gekoppelte Fluoreszenz in allen F-Zellen des Tubulus-Epithels. Andere Zelltypen wurden nicht markiert. Abbildung 34 d zeigt einen Schnitt durch das Epithel mit F-, R-, und B-Zellen. Innerhalb der F-Zellen ist das Fluoreszenzsignal auf Golgi-Vesikel beschränkt (Abbildung 34 f, Pfeil). Extrazellulär wurde Astacin in allen Teilen der Tubuli-Lumina sichtbar (Abbildung 34 c).

In den immunhistochemischen Versuchen mit dem Anti-Pro-Astacin Antikörper fand man nach Stimulierung der Enzymneusynthese die gleichen Fluoreszenzsignale wie mit dem Anti-Astacin Antikörper. Nur die F-Zellen zeigten eine Pro-Astacin-gekoppelte Fluoreszenz (Abbildung 34 i). Innerhalb der F-Zellen wurde die Immunfluoreszenz wiederum in den Golgi-Vesikeln sichtbar (Abbildung 34 g, h, Pfeile). Abbildung 34 g zeigt darüber hinaus ein Signal am Mikrovillisaum einer F-Zelle, dem Ort der Enzym-Sekretion in das Tubulus-Lumen. Extrazellulär wurde Pro-Astacin interessanterweise auch noch in allen Teilen des Tubulus-Lumens und des Sammelkanals gefunden (Abbildung 34 i-k). Die ungleichmäßige Färbung der großen, extrazellulären Kompartimente ist auf eine Auswaschung der Enzyme während der experimentellen Prozedur (vergleiche III.10.4) zurückzuführen.

In Kontroll-Versuchen wurde als primärer Antikörper Anti-Hämocyanin anstelle von Anti-Astacin verwendet. Haemocyanin ist ausschließlich in der Haemolymphflüssigkeit lokalisiert. In Abbildung 34 b ist deutlich erkennbar, dass der Haemolymphraum, nicht aber das Tubulus-Lumen oder die Epithelzellen markiert wurden.

Zusammenfassend kann festgestellt werden, dass Pro-Astacin auf dem gesamten Synthese und Sekretionsweg von den F-Zellen bis in die Sammelkanäle nachgewiesen werden konnte. Eine Aktivierung des Pro-Enzyms direkt nach der Sekretion aus den F-Zellen kann damit ausgeschlossen werden. Der Übergang des Hepatopankreas in den Magen-Darm-Trakt (Atrium) sowie der gesamte Kaumagen (Cardia) und Filtermagen (Pylorus) entzogen sich wegen ihrer Größe einer direkten immunhistochemischen Untersuchung. Für den Nachweis des Pro-Enzyms oder des bereits aktiven Enzyms im Magen musste deshalb ein anderer Ansatz gewählt werden. Dazu wurde dem Flusskrebs artifiziell der Magensaft entnommen und auf einem Immunoblot mit dem Astacin-Antikörper und den Pro-Astacin-Antikörpern untersucht.

Abbildung 34, nächste Seite: Immunhistochemische Lokalisation von Astacin und Pro Astacin im Hepatopankreas von *Astacus astacus*. (a) Schnitt durch Hepatopankreas Tubuli (HT) und Sammelkanal (CD); Goldner Färbung, x 34. (b) Querschnitt durch einen Tubulus, Anti-Haemocyanin Antiserum (Kontrolle). Haemocyanin Fluoreszenz (gelb-grün) in der Haemolymph (HS). HE: Hepatopankreas Epithel; TL: Tubulus Lumen; x 87. (c) Tubulus Lumen mit Astacin (gelb-grün) erhalten mit Anti-Astacin Antiserum; x 80. (d) Schnitt durch das Hepatopankreas Epithel mit intrazellulärer Astacin Immunofluoreszenz in F-Zellen (Pfeile). B-Zellen (B) und R-Zellen (R). N: Zellkern; x 370. (e) Schnitt durch R-Zellen and F-Zellen des Hepatopankreas Epithel. Goldner Färbung; x 490. (f-h) Vergrößerung von F-Zellen nach Inkubation mit Anti-Astacin (f), Anti-ProAst-N (g) und Anti-ProAst-C (h) Antiserum.; x 480 (f), x 550 (g), x 820 (h). (i) Schnitt durch einen Tubulus. ProAst-N Immunofluoreszenz in F-Zellen; x 180. (j-k) Pro-Astacin Fluoreszenz im Lumen der Tubuli und im Sammelkanal erhalten mit Anti-ProAst-N (j) und Anti-ProAst-C (k) Antiserum; x 80.

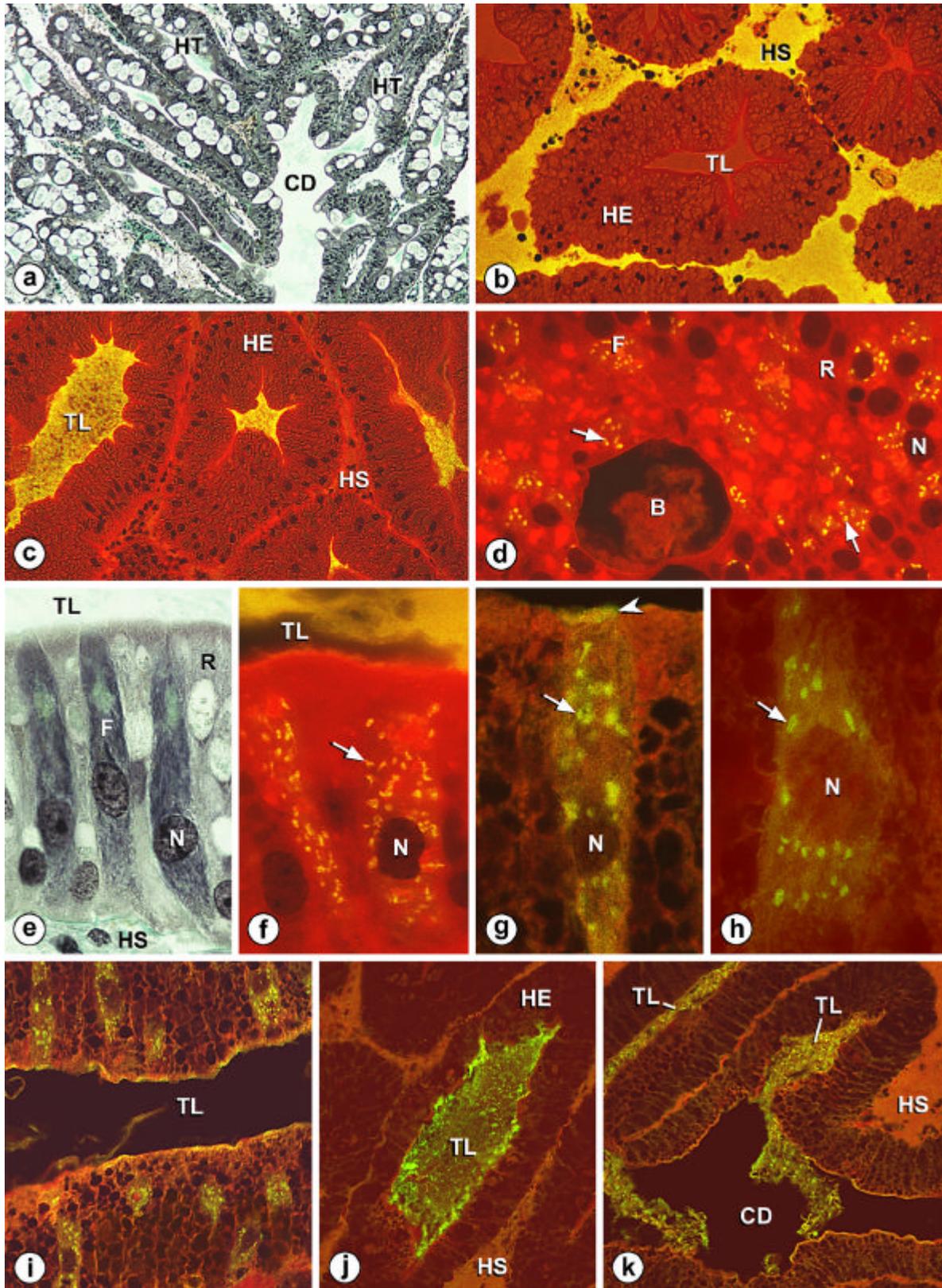


Abbildung 34: Immunhistochemische Lokalisation von Astacin und Pro-Astacin im Hepatopankreas von *Astacus astacus*. Legende siehe vorherige Seite.

4.2. Immunoblotting

Die *Immunoblotting*-Experimente sollten einerseits die Lokalisation von Astacin und Pro-Astacin im Magen des Flusskrebsees ermöglichen (vergleiche IV 4.1). Zum Anderen sollte damit auch die Spezifität der verwendeten Antikörper nachgewiesen werden.

Zur Gewinnung der Protein-Gesamtextrakte aus Mageninhalt und Mitteldarmdrüse wurden mehreren Tieren, die für eine Woche gehungert hatten, der Magensaft entnommen („0h“). Von einem Tier wurde gleichzeitig der Hepatopankreas für eine Proteinextraktion herauspräpariert. Die vollständige Entnahme des Magensaftes führte dabei gleichzeitig zu einer Stimulierung der Enzym-Neusynthese im Hepatopankreas. Deshalb wurden den behandelten Tieren nach 3 oder 7 Stunden erneut Magensaft entnommen und das Hepatopankreasgewebe präpariert.

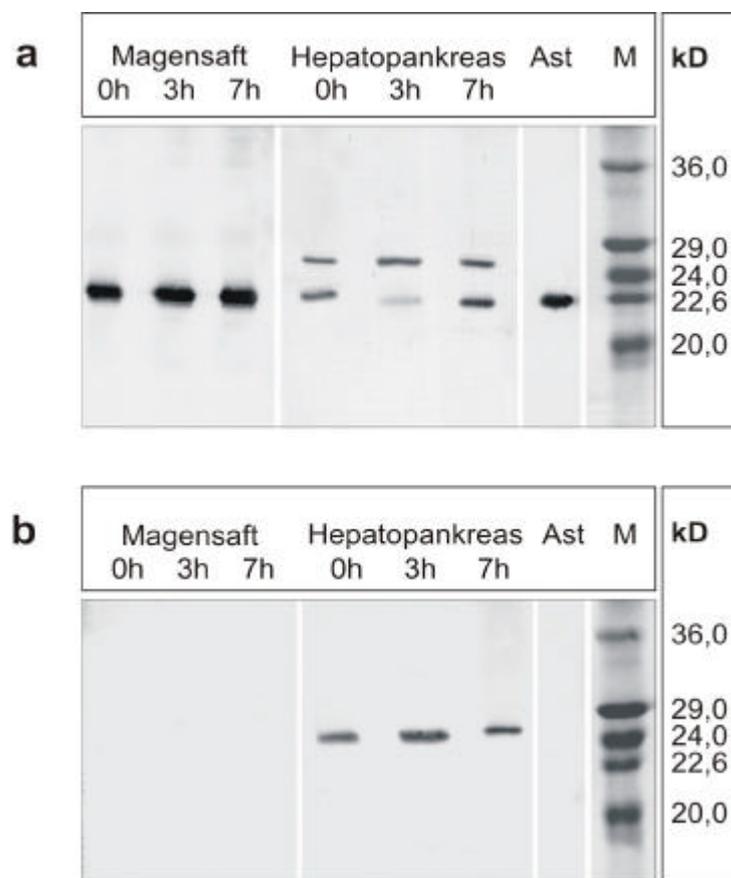


Abbildung 35: Nachweis von Astacin und Pro-Astacin im Magensaft und Hepatopankreas Homogenaten durch *Immunoblotting*. **(a)** Antiserum: Anti-Astacin; Antigene: Magensaft oder Hepatopankreas Homogenate 0 h, 3 h, 7 h nach Stimulierung der Enzymneusynthese durch Entleerung des Mageninhalts. Ast, aufgereinigtes, aktives Astacin; M, Marker Proteine. **(b)** Antiserum: Anti-Pro-Astacin (Anti-ProAstN); Antigene: wie in (a).

Abbildung 35 zeigt, dass der aus der Cardia entnommene Magensaft mit dem Anti-Astacin Antikörper eine deutliche Bande bei 22,6 kDa ergibt (a). Dies weist das Vorkommen von aktivem Verdauungsenzym im Magen nach. Dabei spielt es keine Rolle, ob die Tiere vor der Entnahme eine Woche gehungert hatten (0h) oder der Magen, durch die Entnahme von

Magensaft stimuliert, nach 3 oder 7 Stunden (3h, 7h) mit neusynthetisiertem Enzym gefüllt war. Im Gegensatz dazu wurde im Magensaft mit dem Anti-ProAstN Antiserum überhaupt kein Pro-Astacin gefunden (b). Gleichzeitig ist erkennbar, dass der Anti-ProAstN Antikörper nicht mit dem aktiven Enzym des Magensaftes kreuzreagiert.

In den Hepatopankreas Homogenaten erkennt der Anti-Astacin Antikörper beide Formen: Astacin (22,6 kDa) und Pro-Astacin (28 kDa) (a). Betrachtet man hier die Kinetik der Enzym-Neusynthese, so fällt auf, dass auch im Hungerzustand (0h) Pro-Astacin vorhanden ist, was auf eine stetige Neusynthese des Enzyms schließen lässt. Drei Stunden nach Stimulierung hatte die Menge an Pro-Astacin im Vergleich zu Astacin deutlich zugenommen (3h). Dies könnte eine erhöhte Neusynthese-Rate des Zymogens widerspiegeln. Nach sieben Stunden wird das ursprüngliche Mengenverhältnis wieder erreicht (7h). Teil (b) zeigt wiederum, dass der Anti-Pro-Astacin Antikörper (Anti-ProAstN) spezifisch mit dem Pro-Enzym des Hepatopankreas Homogenats, nicht aber mit dem ebenfalls vorhandenen aktiven Astacin reagiert.

In einer Kontrollreaktion wurde aufgereinigtes, aktives Astacin nur durch den Astacin-Antikörper (a, Ast) und nicht durch den Pro-Astacin-Antikörper (b, Ast) erkannt.

Zusammenfassend zeigen die *immunoblotting* Experimente, dass im Magen des Flusskrebses ausschließlich aktives Verdauungsenzym vorhanden ist. Während im Sammelkanal des Hepatopankreas noch Pro-Astacin nachgewiesen werden konnte (vgl. IV 4.1) ist schließlich in der Cardia kein Pro-Enzym mehr vorhanden. Die Abspaltung des Pro-Peptides muss also in den unteren Teilen der Sammelkanäle oder beim Übertritt in den Pylorus (im Artrium) erfolgen.

4.3. Analyse der Aktivierungsstelle (Prozessierung)

Im Verdauungstrakt von *Astacus astacus* werden nur zwei Endopeptidasen gefunden, welche für die Aktivierung von Pro-Astacin in Frage kommen: Flusskrebs Trypsin und Astacin (Pfleiderer et al. 1967; Vogt 2002). Die unterschiedliche Spaltspezifität dieser beiden Proteasen (Abbildung 36) und die Aminosäuresequenz der Aktivierungsstelle sollten dabei einen Hinweis auf die korrekte Prozessierung von Pro-Astacin im Flusskrebs geben.

Astacin	P₄	P₃	P₂	P₁	P₁'	P₂'	P₃'	P₄'
	apolar		basisch		A	klein	apolar	
					T			
					S			
					G			
Trypsin	P₄	P₃	P₂	P₁	P₁'	P₂'	P₃'	P₄'
	X	X	X	R	X	X	X	X
				K				

Abbildung 36: Spaltungsspezifität von Trypsin und Astacin. Die Substrat-Spezifität von Astacin wurde durch Degradation der α und β Kette des Tubulins, sowie durch synthetische Substrate bestimmt (Krauh et al. 1982; Stöcker et al. 1990; Stöcker et al. 1991); die Substrat-Spezifität von Flusskrebs-Trypsin wurde anhand der β Kette des Insulins und synthetischer Substrate ermittelt (Zwilling und Neurath 1981).

4.4. Massenspektrometrische Analyse

Mit Hilfe der Massenspektrometrie sollte der Effekt von Trypsin und Astacin auf die synthetische Peptide, welche die potentielle Aktivierungsstelle des Astacins überlappen, untersucht werden (siehe IV 4.3). Dazu wurden jeweils 0,5 µg Astacin oder Trypsin mit 5 µg des entsprechenden Substrats (AstG, AstR, AstR*) 2 Stunden bei 30°C inkubiert und die entstandenen Spaltprodukte durch MALD-TOF-Massenspektrometrie analysiert. Auch beim Arbeiten mit affinitätsgereinigtem Astacin oder Trypsin stellte sich in den ersten Versuchen heraus, dass schwache Kontaminationen mit anderen proteolytischen Enzymen in den Massenspektren zu falschen Interpretationen führen können. Deshalb wurde in den Trypsin-Ansätzen eine mögliche Verunreinigung mit dem Astacin-spezifischen Inhibitor Pro-Leu-Gly-Hydroxamat oder mit dem Metalloprotease Inhibitor 1,10 Phenanthrolin unterdrückt. In den Astacin Ansätzen wurde etwaige tryptische Verunreinigung durch den Einsatz von Sojabohnen Trypsin Inhibitor (SBTI) oder Phenyl-Methyl-Sulfonyl Fluorid (PMSF) inhibiert. Zusätzlich wurden als Kontrolle einer Astacin-spezifischen Spaltung in dem Peptid AstG alle L(-)-Argininreste durch die isostereometrischen D(-)-Argininreste ausgetauscht und damit jede tryptische Spaltung verhindert.

In Kontrollansätzen wurden alle Peptide zunächst ohne Enzym-Zugabe inkubiert, um eine spontane Hydrolyse der Peptide unter den Reaktionsbedingungen auszuschließen. In den Massenspektren konnte ausschließlich die monoisotopische einfach protonierte Masse der intakten Peptide (1899,0 Da für AstG und AstR*; 1998,1 Da für AstR) beobachtet werden. Abbildung 39 zeigt die Massenspektren der Spaltprodukte von AstG, AstR und AstR* bei Inkubation mit Astacin (a-c) oder Trypsin (d-f). Astacin hydrolysierte im Peptid AstG (Abbildung 39 a) ausschließlich die Gly-1/Ala+1 Bindung, welche die tatsächliche Aktivierungsstelle im Flusskrebs-Astacin repräsentiert. Astacin spaltete auch dann noch exklusiv an dieser Position, wenn Gly-1 durch einen Argininrest substituiert wurde (AstR, b). Ein basischer Rest kommt natürlicherweise im Verdauungsenzym Astacin in dieser Position nicht vor, wohl aber im embryonal exprimierten AEA (*Astacus Egg Astacin*) und vielen anderen Vertretern der Astacin-Protein-Familie. Trypsin hydrolysierte die Peptide erwartungsgemäß an allen vorhandenen basischen Aminosäureresten. AstG wurde in drei Peptide gespalten, wobei der tatsächliche N-Terminus des reifen Astacin-Moleküls nicht entstanden ist (d). Nur bei Einführung eines Argininrestes in Position -1, welcher die bei vielen Astacinen vorhandene Sequenz widerspiegelt, wurde die -1/+1 Bindung auch von Trypsin gespalten (AstR, e). In dem modifizierten Peptid AstR* produziert Astacin immer noch den korrekten Aminoterminus von Astacin (c), während Trypsin das Substrat überhaupt nicht mehr umsetzen kann (f).

Zusammenfassend kann aus diesen Ergebnissen geschlossen werden, dass das Astacin des Flusskrebses in der Lage ist, seine eigene Aktivierung zu katalysieren. Eine Aktivierung durch Trypsin würde zusätzliche Abspaltungen durch eine Aminopeptidase erfordern. Darüber hinaus konnte gezeigt werden, dass Astacin auch in der Lage ist eine Arg-Ala Bindung zu hydrolisieren. Damit könnte für die Mehrzahl der Astacine neben einer Aktivierung durch Trypsin (oder andere Proteasen mit derselben Spaltspezifität) auch eine autokatalytische Aktivierung in Frage kommen.

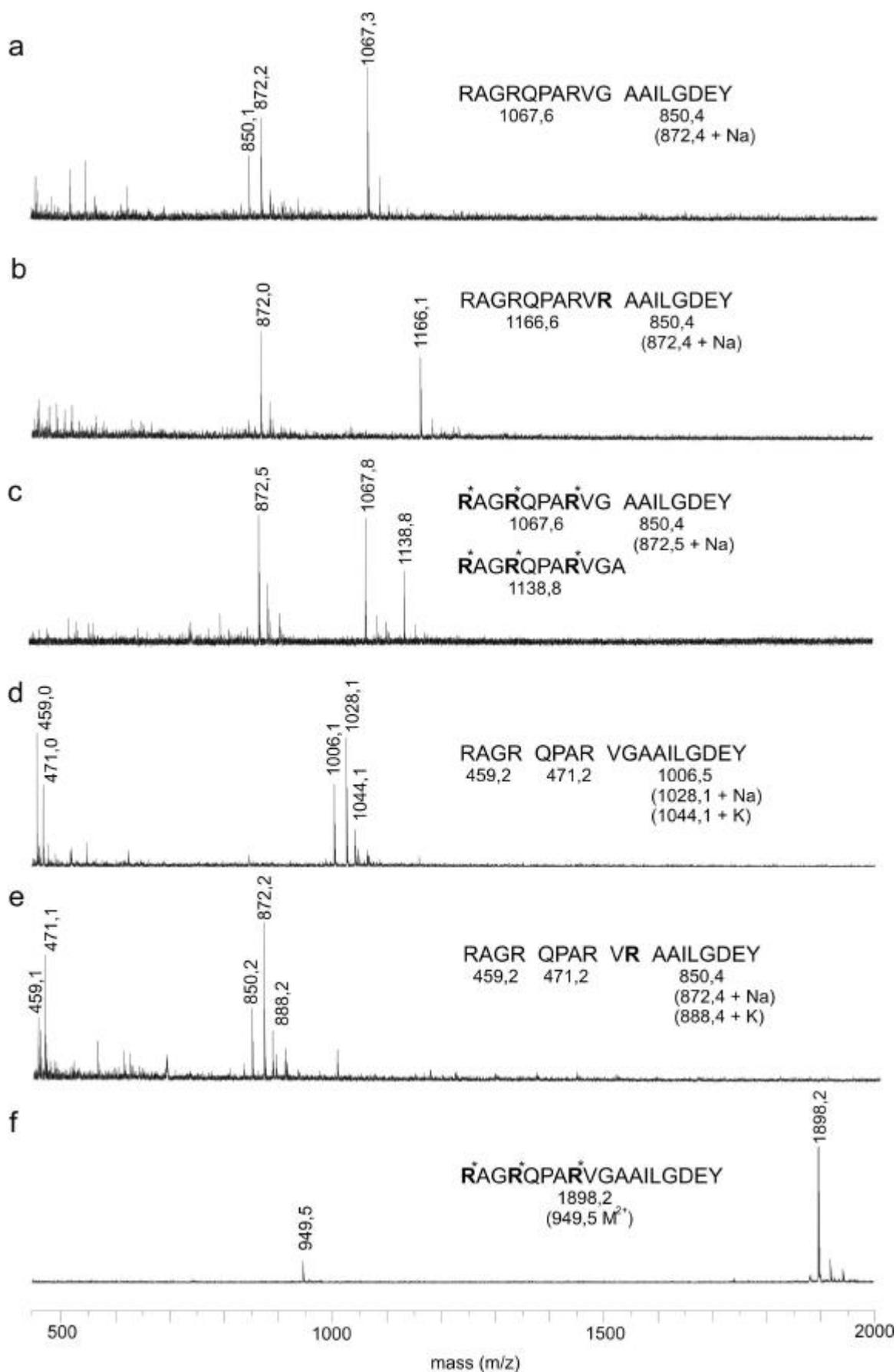


Abbildung 39: Identifizierung der Peptid-Spaltprodukte durch MALDI-TOF Massenspektrometrie. (a) Enzym: Astacin, Substrat: AstG; (b) Astacin, AstR, (c) Astacin, AstR*; (d) Trypsin, AstG; (e) Trypsin, AstR; (f) Trypsin, AstR*

V. DISKUSSION

Astacin (*E.C. 3.4.24.21*) ist eine kollagenolytische Zink-Endopeptidase aus dem Kaumagen des Flusskrebse *Astacus astacus* L. und wurde 1967 von Pfeleiderer und Mitarbeiter erstmals beschrieben (Pfeleiderer et al. 1967; Stöcker und Zwilling 1995). Mit der Aufklärung der Primärstruktur durch Titani et al. zeigte sich, dass diese Invertebraten-Protease in keine bisher bekannte Enzymfamilie eingeordnet werden konnte und den Archetyp einer neuen Familie von Proteasen repräsentiert (Zwilling et al., 1981, Krauhs et al., 1982, Titani et al., 1987).

Neue Mitglieder der „Astacin-Familie“ wurden kurze Zeit später von verschiedenen Arbeitsgruppen entdeckt: BMP-1 (*bone morphogenetic protein 1*), von dem gezeigt werden konnte, dass es an der Knochen- und Skelettbildung beim Menschen beteiligt ist (Wozney et al., 1988); UVS.2, welches bei der Embryonalentwicklung des Krallenfroschs *Xenopus laevis* stadienspezifisch exprimiert wird (Sato und Sargent, 1990) und die Meprine aus Niere und Dünndarm von Mensch und Maus (Dummermuth et al., 1991). Die Aufklärung der dreidimensionalen Struktur des Flusskrebsenzym bestätigte dann nochmals die Bedeutung des Flusskrebs-Astacin als einen Prototyp einer neuen Familie von Zink-Endopeptidasen (Stöcker et al. 1993).

In den letzten Jahren hat die intensive Charakterisierung verschiedener morphogenetisch aktiver Astacine von Vertebraten, Seeigeln und *Drosophila* (für einen Überblick siehe Einleitung) und insbesondere die Genom- oder Transkriptom-Analyse vieler Organismen die Anzahl annotierter Sequenzen, die Astacin-ähnliche Domänen enthalten, stark anwachsen lassen. Mitglieder der Astacin-Protein-Familie werden nunmehr fast ubiquitär von den Bakterien bis zum Menschen gefunden (MEROPS, The Protease Database; Release 5.5). Dennoch liegen seit 1995 (Bond und Beynon 1995) keine aktuelleren Untersuchungen über die molekularen Veränderungen und Evolutionsabläufe der Astacin-Protein-Familie überhaupt vor. Auch sind weder die aus den Genom-Sequenzierungsprojekten erhaltenen Sequenzdaten bisher experimentell überprüft worden, noch lagen für phylogenetische Vergleiche aus den Basisgruppen des Tierreichs (z.B. Cnidaria) genügend Informationen vor. Hinsichtlich einer gemeinsamen, funktionellen Charakterisierung der Astacine war darüberhinaus die Frage nach einem möglicherweise konservierten Aktivierungsmechanismus der als Pro-Enzyme synthetisierten Proteine nicht näher untersucht worden. In dieser Arbeit sollte nun neben einer funktionellen Charakterisierung verschiedener Astacine ein umfassender Überblick über die Struktur und Evolution der Astacin-Protein-Familie gegeben werden.

1. Astacin-homologe Proteine in *C. elegans*

Mit *C. elegans* wurde im Jahre 1991 erstmals ein Projekt zur vollständigen Sequenzierung des Genoms eines vielzelligen Organismus begonnen, das 1998 seinen Abschluss fand (The *C. elegans* Sequencing Consortium 1998). Als erste Untersuchung sollte im Rahmen dieser Arbeit eine Analyse der im Laufe des Sequenzierungsprojektes identifizierten Astacin-homologen Genen durchgeführt werden, womit gleichzeitig ein Beitrag zur Erforschung des Transkriptoms und Proteoms dieses Organismus geliefert werden sollte.

Die Datenbanksuchen nach Astacin-homologen Proteinen in den *C. elegans* spezifischen Datenbanken *wormbase* und *wormpep* wurden mit dem BLAST2 Algorithmus (Altschul et al. 1990; Altschul und Gish 1996) durchgeführt. Zu Beginn dieser Arbeit ergab eine erste BLASTP-Suche in der zu dieser Zeit aktuellen *wormpep11* Protein-Datenbank (Juli 1996, enthielt 7299 vorhergesagte Proteine) bereits die überraschende Anzahl von 12 Astacin-homologen Proteinen; in der kurze Zeit später veröffentlichten *wormpep12* Datenbank (November 1997, 12178 Proteine) konnten bereits 20 Proteine und mit dem Vorliegen der gesamten Genom-Sequenz (Dezember 1998, 18891 Proteine) schließlich 37 Astacin-homologe Proteine gefunden werden. Damit wurden aber nur Proteine erfasst, die durch das Programm *genefinder* überhaupt erkannt werden und die mit dem richtigen Spleißmuster in der *wormpep* Datenbank gespeichert sind. Um tatsächlich alle im Genom vorhandenen Astacin-Gene zu identifizieren wurden deshalb auch TBLASTN-Suchen direkt in der in alle Leserastern übersetzten genomischen *wormbase* Datenbank durchgeführt und am Ende dieser Arbeit anhand der neuen *wormbase 73* Datenbank (*released 7th February 2002, 20,379 predicted protein sequences*) nochmals überprüft. Als Vergleichssequenzen wurden neben der gesamten Protease-Domäne des Flusskrebsenzym und der Protease-Domäne von BMP-1 auch das zinkbindende Motif und der Met-Turn der Astacine allein verwendet. Damit ist sichergestellt, dass auch durch *genefinder* nicht erkannte Proteine, verkürzte Genfragmente oder Proteine mit zunächst falsch vorhergesagten Spleißmustern erkannt werden. In der Tat konnten dadurch zusätzlich drei Sequenzen (F54B8, D1022 und F38E9) gefunden werden, die in der *wormpep* Datenbank nicht vorhanden waren. Die genomischen Sequenzabschnitte wurden daraufhin mit dem Programm *genescan* in mRNA und Protein übersetzt. Durch die ständige Überarbeitung der Computervorhersagen anhand der zunehmenden cDNA Sequenzdaten ist das postulierte Protein F38E9.2 allerdings inzwischen auch in die *wormpep 76* (29 März 2002) aufgenommen worden. Durch einen abschließenden Sequenzvergleich (*alignment*) konnte die Identifizierung von insgesamt nun 40 Astacin-homologen Proteinen auf Primärstrukturebene eindeutig anhand der 200 Aminosäurereste umfassenden Protease-Domäne mit dem stark konservierten, zinkbindenden Motiv HExxHxxGFxHExxRxDRD und dem Methionin-haltigen 1,4- β -Turns mit dem Motiv SxMHY erfolgen (Bode et al. 1992; Gomis-Rüth et al. 1993; Stöcker et al. 1993).

Vergleicht man die Astacin-homologen Proteine mit anderen Proteasen von *C. elegans*, so stellt man fest, dass in keinem Fall eine ähnliche Häufung homologer Gene beobachtet wird (vergleiche Tabelle 1). Im Genom von *C. elegans* sind 353 Protease-Gene (E.C. 3.4.) insgesamt zu finden, wovon 152 Gene zur funktionellen Gruppe der Metalloproteasen (E.C. 3.4.24) gehören. Die Metzinkine (Clan MA(M), nach der Nomenklatur von Barrett et al. 1998) stellen mit 57 Vertretern dabei mehr als ein Drittel aller vorhandenen Metalloproteasen. Homologe *C. elegans* Proteine werden dabei in allen Familien der Metzinkine mit Ausnahme der Serralysine (M10B) gefunden. Bereits an dieser Stelle muss man sich die Frage stellen, wozu ein relativ einfacher Organismus, der aus nur 959 Zellen besteht, eine so große Zahl verschiedener Proteasen benötigt.

Bis auf eine Ausnahme (HCH-1, Hishida et al. 1996) waren die im Genom gefundenen Astacin-homologen Gene bisher nicht näher charakterisiert. Es lagen nur die durch die Programme *genefinder* oder *genescan* postulierten Proteine und die partielle Transkriptom-Analyse durch die EST/ORFeome oder *microarray* Großprojekte vor. Ein Abgleich mit der EST-Datenbank von Kohara et al. (<http://www.ddbj.nig.ac.jp>) und dem im Jahre 2001 begonnenen ORFeome Projekt (<http://worfdb.dfci.harvard.edu>) ergab, dass heute von 16 der insgesamt 40 Gene bereits ein oder mehrere EST/ORFeome-Klone vorliegen, womit zunächst die Expression dieser Gene auf mRNA-Ebene eindeutig nachgewiesen ist. Neue cDNA-Sequenzen sind hier in Zukunft kaum noch zu erwarten, da das EST-Projekt nun weitgehend abgeschlossen ist und die bisherigen Ergebnisse des ORFeome-Projekts zeigten, dass viele der durchgeführten Reaktionen unter Standard-PCR Bedingungen überhaupt keine Produkte ergaben, auch wenn die Expression der Gene auf anderem Wege (z.B. klonierte cDNA oder EST) eindeutig nachgewiesen werden konnte. Das hochgesteckte Ziel des ORFeome-Projekts, in den nächsten Jahren routinemäßig das gesamte Transkriptom (rund 20 000 vorhergesagten Transkripte) durch spezifische Primer zu amplifizieren und durch Sequenzierung zu bestätigen, scheint damit nicht weiter erfolgversprechend zu sein. Im Rahmen dieser Arbeit wurden insgesamt 28 Astacin-homologe Gene durch RT-PCR mit genspezifischen Primern und anschließende Sequenzierung der DNA-Fragmente untersucht, wovon inzwischen die vier Gene C05D11.6, F42A10.8, T02B11.7 und T04G9.2 zusätzlich als EST/ORFeome Klone (siehe oben) bestätigt worden sind. Für die Gene F54B8.m, F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 wurde auch unter modifizierten PCR-Bedingungen kein Produkt erhalten, wobei Kontrollreaktionen mit genomischer DNA die erwarteten PCR-Fragmente zeigten. Daraus kann geschlossen werden, dass diese Gene entweder tatsächlich nicht exprimiert werden oder einer der verwendeten Primer möglicherweise durch eine fehlerhafte Voraussage der cDNA Sequenz durch *genefinder* in einem Intron-Bereich des Genes liegt. Es wäre aber auch vorstellbar, dass diese Gene nur unter spezifischen Umweltbedingungen (Hitzestress, Infektionen, Dauerlarven, etc.) exprimiert werden und die mRNA damit in der verwendeten *C. elegans* Gesamtkultur nicht repräsentiert ist. Die Expression aller anderen untersuchten Gene konnte eindeutig nachgewiesen werden, wobei die Sequenzierung der Gene C05D11.6, C07D10.4, C34D4.9, F42A10.8, K04E7.3 zeigte, dass diese eine von der *genefinder*-Vorhersage abweichendes Spleißmuster aufweisen. Weitere Hinweise über die Expression der Gene F54B8.m, F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1, die kein Produkt in den RT-PCR Experimenten ergaben, konnte aber die Expressionskarte von Kim et al. geben, auf der über 90% der vorhergesagten *C. elegans* Gene vertreten sind (Kim et al. 2001). Die Lokalisierung dieser Transkripte auf der Expressionskarte (*SK-map*) ergab einen deutlichen Hinweis auf die tatsächliche Expression dieser Gene. Allerdings ist die hier verwendete *microarray* und Hybridisierungstechnik mit Oligonukleotiden hauptsächlich auf große Analysedaten von koexprimierten Gengruppen ausgelegt und soll nicht explizit dem Nachweis eines einzelnen Genes dienen. Deshalb wird auch in Kauf genommen, dass im Einzelfall in der Computeranalyse ein nicht signifikantes Hybridisierungssignal (Fluoreszenz) positiv gewertet wird und damit fälschlicherweise die Expression eines Genes als gegeben angenommen wird. Weiterhin

können mit dieser Technik direkt keine neuen Sequenzinformationen gewonnen werden, sodass auch keine alternativen Spleißmuster erkannt werden.

Zusammenfassend kann dennoch festgestellt werden, dass mit Ausnahme von F54B8.m die Expression aller Astacin-homologen Gene auf Transkriptionsebene nachgewiesen werden konnten, wenn auch für die Gene F56A4.1, K03B8.1, K03B8.2, K03B8.3 und Y95B8A.1 eine endgültige Bestätigung durch Sequenzierung offen bleibt. Im Fall von F54B8.m scheint es sich dagegen um ein Pseudogen zu handeln. Dafür spricht auch die Tatsache, dass sich nur ein sehr kurzes Genfragment erkennen lässt (vergleiche VII 6.1.17), das sich in einem einzigen Leseraster befindet und zudem zwei Basentriplets für Stopkodons enthält (vergleiche VII 6.2). Weiterhin werden in diesem Chromosomenbereich keine weiteren offenen Leseraster gefunden, deren übersetzte Proteinsequenzen Homologien zu der normalerweise 200 Aminosäuren umfassenden Astacin-Domäne aufweisen.

Für die folgenden phylogenetischen und strukturellen Untersuchungen war es nötig, die RNA-Transkripte korrekt in Proteinsequenzen zu übersetzen. Dabei stellte sich das Problem, dass den mit RT-PCR und genspezifischen Primern ermittelten cDNA Sequenzen methodisch bedingt immer die terminale 5' und der 3' Region der mRNA und den EST-Klone oft der 5' Bereich der Transkripte fehlt. Die tatsächlichen Transkriptions-Start und Endpunkte wären nur durch aufwendige 5' oder 3' RACE Experimente zugänglich, die im Rahmen dieser Arbeit nicht durchgeführt wurden. Darüber hinaus wurden bei langen Transkripten noch keine vollständige Sequenzierung aus den mittleren Bereichen durchgeführt. Dies bedeutet, dass bisher nur für die Gene F45G2.1, HCH-1 und R151.5 vollständige cDNA Sequenzen vom 5'-nichttranslatierten Bereich bis zum Polyadenylierungssignal vorliegen. Für die richtige Vorhersage der restlichen Transkripte wurden deshalb Sequenzvergleiche der *genefinder* Vorhersagen der *wormpep 73* Datenbank und die alternativen *intronerator* Vorhersagen von Kent und Zahler (2000) mit den bekannten cDNA Fragmenten durchgeführt (Anhang VII. 6.1). Es zeigte sich, dass keine Präferenz in der Zuverlässigkeit der *genefinder* oder *intronerator* Vorhersagen erkannt werden konnte, da beide Programme etwa gleich häufig mit den experimentellen Daten übereinstimmten oder davon abwichen (Tabelle 5). Es wurden die Sequenzen ausgewählt, welche in Übereinstimmung mit den experimentellen cDNA Daten waren, wobei gegebenenfalls zusätzliche alternative Spleißmuster berücksichtigt wurden. Damit konnten viele Spleißstellen experimentell abgesichert werden und es ist davon auszugehen, dass diese Sequenzvorhersagen nun den tatsächlichen Spleißmustern weitgehend entsprechen und eine korrekte Übersetzung der mRNA Sequenzen in Proteinstrukturen erlauben. Die Transkriptions-Start und Endpunkte beruhen allerdings weiterhin allein auf den Computermodellen. Die Übersetzung der so ermittelten mRNA-Sequenzen in Aminosäure-Sequenzen geschah abschließend mit dem *Staden Programm Package* (die Proteinsequenzen im FASTA-Format finden sich im Anhang VII.6.2). Dabei konnte in einem Alignment der Astacin-Domäne mit den korrigierten Proteinsequenzen nochmals bestätigt werden, dass die Spleißmuster richtig ermittelt wurden, da die Protease-Domäne (Abbildung 29) keine Insertionen oder Deletionen mehr zeigte, wie sie durch falsche Spleißvorhersagen entstehen würden.

Die Analyse des Expressionsmuster eines Gens kann erste Aussagen auch über dessen Funktion machen. An ausgewählten Beispielen sollte daher die Lokalisierung Astacin-homologer Proteine *in vivo* durch GFP- (*green fluorescent protein*) Reporter-Proteine erfolgen. Diese Technik ist für *C. elegans* gut etabliert (Chalfie et al. 1994) und mit dem *Fire Lab Vector Kit* (Dr. A. Fire, Department of Embryology, Carnegie Institution of Washington, Baltimore) stehen eine Reihe von Vektoren zur Verfügung, die eine einfache Klonierung des zu untersuchenden Gens ermöglichen. Dabei wird der gesamte Promotorbereich und der proteinkodierende Teil des jeweiligen Gens direkt im Leseraster vor die GFP-Sequenz kloniert. Nach erfolgreicher Transformation dieser Vektoren in *C. elegans* kann das unter der Kontrolle des genspezifischen Promotors exprimierte Fusionsprotein dann direkt im lebenden Tier beobachtet werden (für eine Übersicht siehe Hope 1999). Diese Untersuchungen wurden nun exemplarisch an vier Genen mit unterschiedlicher Domänenkomposition (vergleiche Abbildung 27) durchgeführt, die dementsprechend auch verschiedene physiologische Funktionen erwarten lassen.

Für die Gene K04E7.3 und T04G9.2 konnten allerdings in keiner der transgenen Linien ein GFP-Fusionsprotein nachgewiesen werden. Beide Vektoren wurden daraufhin nochmals durch Sequenzierung überprüft, um sicherzustellen, dass durch die über PCR amplifizierte DNA keine Mutationen eingeführt wurden. Ein eventuell auftretendes Stop-Codon würde nämlich zu einem Abbruch der Translation führen und es entstünde kein funktionelles Fusionsprotein. Das Ausbleiben eines Fluoreszenzsignals kann dann nur damit erklärt werden, dass die Tiere nach der Transformation entweder die Vektoren nicht stabil in den Zellkern integriert haben oder dass das Fusionsprotein nur kurzweilig während spezifischer Entwicklungszeiten exprimiert wird und ein entsprechend schwaches Signal im Fluoreszenzmikroskop nicht detektiert werden konnte. Für eine transiente Expression dieser Gene und einer damit einhergehenden entwicklungssteuernden Funktion spricht auch das Vorkommen der regulatorischen carboxyterminalen Erweiterungen: Im Falle von K04E7.3 sind eine EGF, eine CUB und eine TSP-1 ähnliche Domäne, bei T04G9.2 sind drei ShKT ähnliche Domänen zu finden. Die möglichen Funktionen solcher carboxyterminalen Domänen bei Astacinen wird bei der Diskussion der molekularphylogenetischen Untersuchungen noch ausführlich behandelt werden. Leider lagen für beide Gene noch keine Expressionsdaten der *microarray* Analyse von Jiang et al. (2001) vor, die eine transiente Expression hätten bestätigen können. Bezieht man aber die funktionellen Analysen der RNA-Interferenz (RNAi) Experimente von Maeda et al. mit ein (vergleiche Tabelle 7), so scheinen diese Gene allerdings nicht essentiell für eine normale Entwicklung zu sein, da bei deren Ausfall kein auffälliger Phenotyp beobachtet werden konnte.

Das Fusionsprotein C05D11.6:GFP ist ausschließlich in adulten Hermaphroditen in der Pharynxregion und weiter posterior im Darmlumen lokalisiert. Offenbar wird das Protein in einigen das Pharynxlumen umschließenden Muskelzellen synthetisiert (Zelllinien pm 5-8), exocytotisch in das Lumen entlassen und schließlich im Darmkanal akkumuliert. Funktionell steht das gesamte Pharynxsystem im Dienste der Nahrungsaufnahme, der Konzentrierung und Vorverdauung der Nahrungsbestandteile, bevor diese in den Darm gepumpt werden (einen Überblick gibt Wood 1988). Da in *C. elegans* spezialisierte Organe für die Sekretion von Verdauungsenzymen fehlen, könnte sich bei diesem Protein also

tatsächlich um ein von den Pharynxmuskelzellen sezerniertes Enzym im Dienste der Nahrungsverdauung handeln. Dafür spricht auch die Tatsache, dass das C05D11.6 Protein wie das Verdauungsenzym des Flusskrebse keine carboxyterminalen regulatorischen Domänen aufweist. Die RNAi Studien von Maeda et al. (2001) zeigen keinen auffälligen Phänotyp, wobei diese Technik auf die Analyse von Entwicklungsstadien beschränkt ist und im adulten Tier ohnehin keinen Effekt mehr zeigen kann. Auch ist nach Untersuchungen von Geier et al. (1999) der Hauptteil proteolytischer Aktivität in *C. elegans* auf Asparatproteasen zurückzuführen, welche damit auch größtenteils die Nahrungsverdauung übernehmen sollten. Dies schließt aber keineswegs die Beteiligung anderer Verdauungsenzyme aus. So werden beispielsweise auch im Verdauungstrakt des Flusskrebse sowohl Serinproteasen (Trypsin) als auch Astacin gemeinsam gefunden (Pfleiderer et al., 1967, Zwilling und Neurath 1981). Ebenso sind bei Säugetieren mit Pepsin und Trypsin Vertreter verschiedener Protease-Familien gleichzeitig an der Nahrungsverdauung beteiligt. Die GFP-Reporter Experimente dieser Arbeit und die *microarray* Analyse von Jiang et al (2001) zeigen in Übereinstimmung aber keine Expression des C05D11.6 Proteins in Eiern oder den Larvenstadien. Geht man von der Annahme einer Verdauungsfunktion für dieses Protein aus, müssen die Larven dann offenbar eine andere Zusammensetzung ihrer Verdauungsenzyme aufweisen. Möglicherweise könnten, neben den Asparatproteasen, eins oder mehrere der Astacin-homologen Proteine F45G2.1, F56A4.1, K06A4.1 und T23H4.3, die wie das C05D11.6 Protein keine carboxyterminalen Domänen besitzen, diese Aufgabe übernehmen.

Das Fusionsprotein C07D10.4:GFP wird sowohl in Eiern als auch in der Pharynxregion adulter Hermaphroditen aber nicht in Larvenstadien gefunden. Im Gegensatz zu C05D11.6 ist die Expression von C07D10.4 extrazellulär auf den Bereich des Procorpus und der terminalen Verdickung des Pharynx beschränkt. Damit könnte das Protein mit dem in dieser Region liegenden Exkretionssystem (Drüsenzellen, H-Zelle, Exkretionsgang) assoziiert sein. Andere Drüsenzellen des Pharynxsystems, deren Funktion nicht vollständig geklärt ist, haben direkten Kontakt zum Lumen und können Golgi-Vesikel exocytotisch entlassen (Wood 1988). Allerdings kann hier nur spekuliert werden, ob diese Protease damit etwa eine Rolle bei der mikrobiellen Abwehr spielen könnte. Das Vorkommen einer ShK-Toxin Domäne in diesem Protein würde diese Hypothese jedenfalls unterstützen. Überraschenderweise findet sich aber auch eine Expression dieses Gens in Entwicklungsstadien kurz vor dem Schlüpfvorgang der Embryonen aus der Eihülle. Eine starke Expression ist dabei im Bereich des Pharynx (Exkretionssystem) und extrazellulär direkt an der Eihülle zu finden. Dieses Muster gleicht dem von Hishida et al. (1996) beschriebenen Astacin-homologen HCH-1 Protein, von dem gezeigt werden konnte, dass es eine Rolle beim Schlüpfvorgang (*hatching*) der Embryonen spielt.

C. elegans bot durch einige im Jahre 2001 veröffentlichte Großprojekte zur Charakterisierung des gesamten Transkriptom (Kim et al. 2001; Jiang et al. 2001; Maeda et al. 2001) darüber hinaus die Möglichkeit, funktionelle Aussagen auch über die weiteren, hier nicht näher charakterisierten Astacin-homologen Gene machen zu können. Die Analyse der Expressionskarte (*SK-map*) von Kim et al. (2001) hinsichtlich der Astacin-

homologen Gene ergab aber keine signifikante Koexpression mit anderen Genen bei verschiedenen Wachstumsbedingungen, Entwicklungsstadien oder Stoffwechselfvorgängen. Das *microarray* Projekt von Jiang et al. (2001) zeigte dagegen, dass 20 Astacin-homologe Proteine eine veränderte Expressionsrate in mindestens einem Entwicklungsstadium aufweisen. Die nahezu das gesamte Transkriptom umfassende Datenauswertung (17800 Gene) weist aber die gleichen Fehlerquellen wie die *SK-map* von Kim et al. (2001) auf. Jiang und Mitarbeiter geben anhand der statistischen Auswertung der durchgeführten Hybridisierungsexperimente immerhin eine Fehlerquote von ungefähr 5% an. So postulieren Jiang et al. beispielsweise für das *hch-1* Gen eine geringere Expression in L3-Larven und adulten Hermaphroditen im Vergleich zu einer Gesamtkultur; Hishida et al. (1996) konnten aber eindeutig zeigen, dass HCH-1 ausschließlich in Eiern kurz vor dem Schlüpfvorgang und in L1 Larven exprimiert wird. Gleiches gilt für die in vorliegender Arbeit untersuchten Gene C05D11.6 und C07D10.4. Für C05D11.6 wird von Jiang et al. eine geringere Expression nur für Eier und L2- bis L4-Larven, nicht aber für L1-Larven angegeben; C07D10.4 sollte nach Jiang et al. vermehrt auch in Larven exprimiert werden, welches ebenfalls im Widerspruch zu unseren GFP-Experimenten steht. Die Auswertung der *microarray*-Daten kann also lediglich einen Hinweis darauf geben, dass einige Astacin-homologe Gene tatsächlich stadienspezifisch exprimiert werden. Vielversprechender wird auch in Zukunft sicher das RNAi Projekt von Maeda et al. (2001) sein, welches durch das spezifische Ausschalten einzelner Gene durch doppelsträngige RNA (Fire et al. 1998) weitere funktionelle Aussagen ermöglicht. Sieben der zehn bereits untersuchten Astacin-homologen Gene zeigen aber keinen auffälligen Phenotyp (unter anderem K04E7.3 und T04G9.2, siehe oben). Dagegen scheint das Gen C17G1.6 an der Vulva-Entwicklung in Larven beteiligt zu sein. C37H5.9 und K11G12.1 müssen zumindest eine entscheidende Rolle in der frühen Embryonalentwicklung spielen, da 6% bzw. 100% der Embryonen nach der RNAi-Behandlung absterben.

Zusätzlich zur Transkriptomanalyse sollte für die Astacin-homologen Gene eine Analyse auf Proteinebene durchgeführt werden. Als Schlüsseltechnologie der Proteomanalyse gilt dabei die zweidimensionale (2D-) Gelelektrophorese. Die bisher erstellten 2D-Proteinkarten von *C. elegans* mit über 400 identifizierten Proteinen (Bini et al. 1997; Kaji et al. 2000; Schrimpf et al. 2001) führten bisher zu keinem Nachweis eines Astacin-homologen Proteins. Für die Identifikation dieser Proteine auf der 2D-Proteinkarte sollten deshalb spezifischere Methoden angewandt werden.

Als erster Ansatz wurde die Detektion verschiedener Astacin-homologer Proteine durch Antikörper auf einem zweidimensionalen Immunoblot gewählt. Die Antiseren wurden durch Immunisierung von Kaninchen mit zwei synthetischen Peptiden (CePepZinc_C07D10.6 und CePepZinc_K04E7.3) gewonnen, deren Sequenz die konservierte Zinkbindungsregion der Protease-Domäne umfasst, welche im aktiven Zentrum des Moleküls lokalisiert ist. Die Immunodetektion mit dem CePepZinc_C07D10.6 Antiserum und dem CePepZinc_K04E7.3 Antiserum ergab jeweils ungefähr 30 deutliche Signale (Abbildung 9), wovon drei von beiden Antikörpern erkannt wurden. Für diese Ergebnisse gab es zunächst keine schlüssige Erklärung. Durch die Transkriptomanalyse war aber die Expression von 37 weiteren Astacin-homologen

Proteinen nachgewiesen worden, die ebenfalls das hochkonservierte Zinkbindungsmotiv (HE_{xx}H_{xx}GF_xHE_{xx}R_xDRD) enthalten und deshalb im Falle einer ausreichenden Expression und Translation, mit den verwendeten Antikörpern kreuzreagieren können. Die abschließenden Sequenzierungen der reagierenden „Spots“ durch Edman-Abbau oder die Identifizierung mit Hilfe der MALDI-TOF Massenspektrometrie ergaben aber in keinem Fall die Detektion eines Astacin-homologen Proteins. Vielmehr wurden eine Reihe sehr unterschiedlicher Proteine wie Dehydrogenasen, ATP-Synthasen, Kinasen, Isomerasen oder *heat shock* Proteine erkannt (Tabelle 9). Die Frage, welche Epitope die verwendeten Antiseren nun tatsächlich erkennen, muss dementsprechend offen bleiben. Alle Sequenz- und massenspektrometrischen Daten werden in Kürze in Zusammenarbeit mit Dr. Luca Bini, (Department of Molecular Biology, University of Siena) in der *C. elegans* Referenzkarte der SIENA-2DPAGE Datenbank (Bini et al. 1997, <http://www.bio-mol.unisi.it/2d/2d.html>) bereitgestellt, sodass unabhängig von der Identifizierung Astacin-homologer Proteine ein Beitrag zur Aufklärung der Proteomorganisation von *C. elegans* geleistet werden konnte. Die Mehrzahl der hier charakterisierten Proteine sind zwar bereits in den 2D-Karten von Kaji et al. (2000) und Schrimpf et al. 2001 repräsentiert (Tabelle 9). Da diese unterschiedlichen 2D-Karten aufgrund der gewählten Versuchsbedingungen aber untereinander kaum vergleichbar sind, ist es dennoch nützlich und sinnvoll, diese in die ursprüngliche Referenzkarte von Bini et al. (1997) aufzunehmen. Zukünftige Arbeiten sollen diese Karte weiter vervollständigen.

Als weiterer Ansatz sollte versucht werden, die Astacin-homologen Proteine durch Affinitätschromatographie mit dem Astacin-spezifischen Inhibitor Pro-Leu-Gly-Hydroxamat aufzureinigen (Moore und Spilburg 1986) und anschließend durch massenspektrometrische Analyse des gereinigten Proteinextrakts zu identifizieren. Als spezifischer Enzymtest wurde das dansylierte Substrat DnsPKFAPWV (Spaltung der Arg/Ala-Bindung mit einem $K_{cat}/K_m = 1 \times 10^5$ für Astacin) verwendet (Stöcker et al. 1990; Stöcker und Zwilling 1995; Wagner et al. 2002). Im ersten Elutionsmaximum der Affinitätschromatographie war eine deutliche Zunahme einer Astacin-ähnlichen proteolytischen Aktivität zu beobachten. Die Aufreinigung Astacin-ähnlicher Proteasen schien also zunächst gelungen zu sein. Die Analyse dieser Fraktion auf einem 2D-Gel und der anschließenden massenspektrometrischen Identifizierung der stärksten Proteinspots zeigte dann aber lediglich zwei proteosomale Metalloproteasen, sowie eine Hydratase, Dehydrogenase und Isomerase (Tabelle 11). Die anhand der Spaltung des Heptapeptidsubstrats DnsPKFAPWV gefundene proteolytische Aktivität kann demnach nur durch die proteosomalen Metalloproteasen oder hier nicht identifizierte Proteasen erklärt werden. Eine weitere Untersuchung erschien allerdings nicht mehr sinnvoll, da es sich zeigte, dass die affinitätschromatographische Aufreinigung nicht geeignet war, Astacin-homologe Proteine mit einem einfachen Verfahren anzureichern. Mit einem vergleichbaren Ansatz ist es Geier et al. (1999) dagegen über eine Pepstatin-Affinitätschromatographie in einem Schritt gelungen fünf der insgesamt 12 Aspartatproteasen von *C. elegans* aufzureinigen. Für die Astacine scheint das Problem in der Wahl eines geeigneten Inhibitors zu liegen. Pro-Leu-Gly-Hydroxamat wurde ausgewählt, weil er der einzige verfügbare, reversible und weitgehend spezifische Inhibitor mit einer hohen Bindungsaffinität für Astacine und einige Metzinkine (Adamlysinase und

MMP) ist (Moore and Spilburg 1986; Wolz et al. 1990; Stöcker und Zwilling 1995). Allerdings haben vorangegangene Untersuchungen auch gezeigt, dass die K_i Werte innerhalb der Astacin-Familie große Unterschiede aufweisen. Die α Untereinheit des Meprins [EC 3.4.24.18] wird beispielsweise mit einem K_i von 0.45 μM inhibiert, während die homologe β -Untereinheit einen 100-fach höheren Wert zeigt (Köhler et al. 2000; persönliche Mitteilung Prof. W. Stöcker, Universität Münster). Der entsprechende Wert für Astacin [EC 3.4.24.21] liegt bei 16 μM . Dies bedeutet aber auch, dass Pro-Leu-Gly-Hydroxamat schon für eine Aufreinigung oder Inhibierung von Meprin β ungeeignet ist. Möglicherweise liegen die K_i Werte der Astacin-homologen Proteine von *C. elegans* in eben diesem Bereich ($K_i > 30$), sodass eine Aufreinigung mit Pro-Leu-Gly-Hydroxamat nicht möglich ist.

2. Astacin-homologe Proteine in *Hydractinia echinata*

Aus den Basisgruppen des Tierreichs (Porifera, Ctenophora, Cnidaria) lagen mit Ausnahme von *Hydra vulgaris* (Kumpfmüller et al. 1999; Yan et al. 2000a; Yan et al. 2000b; Yan et al. 1995) nur wenige Daten über Astacin-homologe Proteine vor. Da Astacine aber schon auf der Ebene der Eubakterien gefunden werden (Tarentino et al. 1995), ist es für die Analyse der molekularen Evolution der Protein-Familie unerlässlich, hier weitere vergleichende Informationen zu erhalten. Besonders interessant ist in diesem Zusammenhang etwa die Frage, ob schon früh in der Evolution Genduplikationen stattgefunden haben, die sich in der Folge getrennt zu strukturell oder funktionell unterschiedlichen Astacin-homologen Proteinen entwickelt haben.

Mit vorliegender Arbeit konnten nun zwei neue Astacin-homologe Gene bei dem Modellorganismus *Hydractinia echinata* (Cnidaria, Hydrozoa) charakterisiert werden. Diese Proteine wurden als *Hydractinia echinata* Astacin 1 und 2 (HEA-1 und HEA-2) bezeichnet.

Die vollständige Sequenzierung der HEA-1 cDNA-Sequenz und der daraus abgeleiteten Aminosäuresequenz zeigte zunächst, dass die normalerweise 200 Aminosäurereste lange Astacin-Domäne unmittelbar nach dem Zink-Bindungsmotiv um ca. 50 Aminosäurereste verkürzt ist (Abbildung 22 und 24). Das Protein endet mit einem Cysteinrest, der im Flusskrebs Astacin am carboxyterminalen Ende bei Position +198 liegt und Bestandteil der zweiten Disulfidbrücke ist. Der deletierte Bereich umfasst also den Abschnitt von dem dritten Histidin-Zink-Ligand (His-102 im Flusskrebs-Astacin) bis zu dem Met-Turn, welcher den vierten Zinkliganden des Moleküls enthält (Tyr-149 im Astacin). Das Fehlen dieses strukturell wichtigen Abschnittes erlaubt damit keine korrekte dreidimensionale Struktur des Proteins mehr (Bode et al. 1992) und selbst die Komplexbildung eines Zink-Ions im aktiven Zentrum erscheint unmöglich. Dadurch muss dieses Protein auch seine proteolytische Aktivität verloren haben. Da die Expression von HEA-1 auf mRNA-Ebene aber eindeutig nachgewiesen werden konnte und es damit höchstwahrscheinlich auch zu einer Translation kommt, könnte dieses Protein andere physiologische Aufgaben übernommen haben. Das klassische Beispiel hierfür ist das Haptoglobin des Menschen, welches zur Familie der Serinproteasen gehört (Maeda et al. 1986). Ein weiteres gut

untersuchtes Beispiel, in dem Enzyme ihre katalytischen Fähigkeiten verloren haben, trotzdem aber weitere physiologisch wichtige Aufgaben übernehmen, sind die Pseudo-Hämocyanine der Insekten (Burmester 1999; Burmester 2001; Burmester et al. 1998).

Das HEA-2 Protein (Abbildung 23) zeigt demgegenüber eine vollständig vorhandene Astacin-Domäne und außerhalb dieser eine zusätzliche carboxyterminale Toxin-Domäne (ShKT).

Die *in situ* Hybridisierung ergibt für die mRNA von HEA-1 und HEA-2 zunächst ein sehr ähnliches Expressionsmuster (Abbildung 25). Beide mRNA Transkripte sind weder in Gonozoiden, während der Embryonalentwicklung, noch in Planula Larven exprimiert. Erst 12 Stunden nach Metamorphose-Auslösung werden in Primärpolypen die ersten Transkripte gefunden, welche bei 24 –30 Stunden alten Primärpolypen ausschließlich in endodermalen, sich differenzierenden Drüsen-Zellen gefunden werden. Die Expressionsstärke nimmt innerhalb der Körpersäule vom Hypostom zur Fußregion deutlich ab. Die Expression von HEA-1 ist hier aber im Gegensatz zu HEA-2 auf wenige endodermale Zellen der Körperachse beschränkt. In adulten Polypen wird für beide Gene eine schwache Expression unterhalb der Tentakelbasis gefunden. Weiterhin scheinen sowohl HEA-2 als auch HEA-1 im Verlauf einer Wund-Antwort oder Geweberegeneration aktiviert zu werden. Während der Metamorphose scheinen beide Gene aufgrund der gradientenartigen Verteilung in die Kopfbildung involviert zu sein. In adulten Polypen deutet die Expression unterhalb der Tentakelbasis eine Beteiligung an der ständigen Neubildung von Hypostom- oder Tentakel-Zellen an. So könnten hier diese Proteasen, durch Degradation der extrazellulären Matrix, im Rahmen der Wegfindung von Zellen aus der Tentakelbasis in die apikalen Bereiche beteiligt sein. Vorläufige, hier nicht gezeigte Experimente mit doppelsträngiger RNA (RNAi) unterstützen diese Hypothesen, da behandelte Larven nach ca. 12 Stunden die Metamorphose nicht weiter fortsetzen und adulte Polypen nach einiger Zeit ihre Tentakeln verlieren. Vergleicht man diese Daten mit den bereits näher charakterisierten Astacin-homologen Proteinen von *Hydra vulgaris* oder *Podocoryne carnea* fallen ähnliche Expressionsmuster und entwicklungssteuernde Mechanismen auf. Die *Hydra* Metalloprotease 1 (HMP-1) wurde in der extrazellulären Matrix an der Basis der Tentakeln gefunden und scheint eine Rolle bei der Kopf-Regeneration und Differenzierung/Transdifferenzierung von Batteriezellen der Tentakeln zu spielen (Yan et al. 1995; Yan et al. 2000). HMP-2 zeigt auch in Regenerationsexperimenten das gleiche Expressionsmuster wie HEA-2 in *Hydractinia*, wobei allerdings noch keine weitergehende funktionelle Charakterisierung erfolgte (Yan et al. 2000). FARM-1 (*Hydra foot activator responsive metalloprotease 1*) scheint dagegen eine andere Aufgabe bei der Fußdifferenzierung übernommen zu haben (Kumpfmüller et al. 1999). Die Metalloproteinase 1 von *Podocoryne carnea* (PMP-1), die von allen Cnidaria-Astacinen die höchste Sequenz-Homologie zu HEA-2 besitzt (vergleiche Abschnitt V 3. dieser Diskussion), zeigt in Polypen dieses Tieres dagegen eine völlig andere Funktion. Man findet das Protein ausschließlich im Manubrium, wobei die Expression durch Nahrungsaufnahme eindeutig induziert werden kann. In dem Medusenstadium von *Podocoryne*, welches bei *Hydra* und *Hydractinia* nicht ausgebildet wird, findet man dagegen eine transiente Expression bei der Etablierung der

Körpergrundgestalt. Damit scheint dieses Protein eine Doppelfunktion als Verdauungsenzym im Polypen und eines Morphogens bei der Medusenentwicklung übernommen zu haben (Pan et al. 1998). Sowohl bei *Hydractinia* als auch bei *Hydra* gibt es keine Hinweise auf eine Funktion der Astacine bei der Nahrungsverdauung. Die Frage, ob diese Verdauungsfunktion ein ursprüngliches Merkmal darstellt, das auch bei Crustacea (Flusskrebbs-Astacin) noch gefunden wird oder ob sich diese Funktion mindestens zweimal im Tierreich unabhängig voneinander bei einigen Spezies entwickelt hat, muss aber zunächst noch offen bleiben.

3. Molekularphylogenetische und strukturelle Untersuchungen der Astacin-Protein-Familie

Ausgangspunkt jeder molekular-phylogenetischen Untersuchung ist die Identifizierung aller in den Datenbanken verfügbaren Sequenzen, die für eine entsprechende Protein-Familie kodieren. Einen sehr guten, wenn auch nicht vollständigen Überblick für proteolytische Enzyme bietet inzwischen die MEROPS-Datenbank (Barrett et al. 2001). Alle Datenbanksuchen nach Astacin-homologen Proteinen mussten dennoch mit dem BLAST2 Algorithmus (Altschul et al. 1990; Altschul und Gish 1996) und dem FASTA3 Algorithmus (Pearson 2000) ergänzt werden. Als Vergleichssequenzen wurden, wie schon für *C. elegans* beschrieben, neben der gesamten Protease-Domäne des Flusskrebsenzym und der Protease-Domäne von BMP-1 auch das zinkbindende Motif und der Met-Turn, jeweils allein, verwendet. Insgesamt werden heute 106 verschiedene Astacin-homologe Proteine in den Datenbanken gefunden (Tabelle 12). Darüber hinaus befinden sich weitere 77 kurze, unvollständige Genom-, EST- oder N-terminale Sequenzfragmente in den Datenbanken (Tabelle 13). Durch die Vielzahl der noch laufenden Genom- oder EST-Projekte wird aber die Anzahl Astacin-homologer Sequenzen in Zukunft rasch weiter zunehmen. Für eine strukturelle und molekularphylogenetische Analyse konnten natürlich nur vollständige Protein-Sequenzen benutzt werden. Die 77 Sequenzfragmente wurden deshalb nicht in die Untersuchungen einbezogen.

Die Analyse der Domänenstruktur (Abbildung 26-28) zeigte zunächst für nahezu alle Proteine eine Pre-Pro-Struktur aminoterminal von der Proteasedomäne. Die bei FARM-1 (*Hydra*), CAM-1 (*Coturnix*) und CG6974 (*Drosophila*) fehlende Pre-Pro-Struktur ist vermutlich auf unvollständige cDNA-Klone (FARM-1, CAM-1) oder eine falsche Spleißvorhersage der genomischen DNA (CG6974) zurückzuführen. Das Pre- oder Signalpeptid deutet dabei auf die Tatsache hin, dass es sich um sekretierte Proteine handelt; die Pro-Sequenz spielt eine Rolle bei der Aktivierung dieser Proteasen und wird im nächsten Abschnitt (V.4.) ausführlich diskutiert werden.

Carboxyterminal der Protease-Domäne treten bei den Astacinen nun in unterschiedlichen Kompositionen eine Vielzahl weiterer Proteindomänen auf. So finden sich CUB-Domänen (*complement subcomponents Cqr/C1s*, *embryonic sea urchin protein Uegf*, **BMP-1**, SMART accession number: SM0042) bei allen BMP-1 / Tolloid-ähnlichen Proteasen (BMP-1, Tolloid, Tolkin, TLL-1, TLL-2, BP10, SpAN, u.a.; vergleiche auch Abbildung 30), beim *hatching*-Enzym AEA (*Astacus embryonic astacin*) aus *Astacus astacus*, beim UVS-2 aus

Xenopus, beim MTP-1 in dem Nematoden *Anclystoma canium*, sowie bei einer Reihe von *C. elegans* Astacinen. CUB-Domänen können Calcium binden, spielen vermutlich eine entscheidende Rolle bei der Substratbindung der Protease und können durch CUB-CUB oder CUB-Substrat Interaktionen diese beeinflussen. Sieron et al. konnten solche Wechselwirkungen am Beispiel des menschlichen BMP-1 zeigen (Sieron et al. 2000). Bekannt wurde die CUB-Domäne als Bestandteil der Serinproteasen C1r/C1s, die in der klassischen Immunantwort des Komplementsystems aktiv sind (Bork und Beckmann 1993). EGF-ähnliche (*epidermal growth factor*, PFAM accession number: PF00008) Domänen treten mit Ausnahme der Meprine und eines *C. elegans* Proteins (K03B8.3) bei den Astacinen immer in Kombination mit CUB-Domänen auf, wobei nicht alle Astacine, die CUB-Domänen besitzen auch EGF-Domänen aufweisen (AEA, UVS-1, CAM-1). EGF-ähnliche Domänen werden darüber hinaus häufig bei anderen sezernierten Proteasen gefunden, wie z.B. den Koagulationsfaktoren IX und X der Blutgerinnungskaskade, oder den Plasminogenaktivatoren in Geweben von Säugern. Sie beinhalten Threonin-reiche Abschnitte, die Anknüpfungspunkte der O-Glykosylierung bei der posttranslationalen Modifikation während der Proteinreifung darstellen. EGF-ähnliche Domänen können aber auch andere regulatorische Aufgaben übernehmen (Bond und Beynon 1995). Die Meprine sowie HMP-2 von *Hydra* besitzen als einzige Vertreter der Astacin-Familie eine MAM-Domäne (*meprin subunits, A-5 protein, and receptor protein tyrosin phosphatase C m*, SMART accession number: SM0137), welche häufig bei membranständigen Enzymen vorkommt und als Adhäsionsdomäne dient (Marchand et al. 1996). Unterstützt wird diese Annahme durch vergleichbare Eigenschaften intestinaler Enteropeptidasen (E.C.3.4.21.9) bei Vertebraten, die nach Bindung des Vorläufers Trypsinogen, an denen die MAM und CUB Domänen beteiligt sind, diesen zu reifem Trypsin proteolytisch spalten und dadurch aktivieren (Barrett et al. 1998). SpAN und BP10 aus Seeigeln, das *Drosophila* Protein CG6763, sowie einige *C. elegans* Proteine enthalten eine Ser/Thr-reiche Region, die auch als LC- (*low complexity*) Domäne bezeichnet wird. Diese Bereiche können bei der posttranslationalen Modifikation der Proteine Stellen der O-Glykosylierung sein. Die sogenannte ShKT-Domäne (SMART accession number: SM0254) wird ausschließlich bei den Cnidaria-Astacinen (HMP-2, FARM-1, PMP-1, HEA-2, HMP-1) und bei Nematoden (*Ancylostoma* MTP1 und 12 *C. elegans* Astacine) gefunden. Das ShK-Toxin Modul wurde erstmals als eine 50 Aminosäurereste große Domäne in Metridin, einem Toxin der Seeanemonen beschrieben. Dieses Toxin kann spannungsgesteuerte Kaliumkanäle blockieren und so als Nervengift wirken (Schweitz et al. 1995). Bei Nematoden ist dieselbe Domäne auch als SXC (*six cysteine*) benannt worden, die sich im Genom von *C. elegans* in 184 Kopien findet (Gems et al. 1995; Blaxter 1998). Diese Module finden sich entweder einzeln, in Clustern von einer bis zu vier Wiederholungen oder als carboxyterminale Domänen in Thyrosinasen, Myeloperoxidasen, Phosphatidylethanolamin-Bindeproteinen, Ionenkanälen und Astacinen wieder. Es wird angenommen, dass die SXC-Domäne hier in Protein-Protein Interaktionen, möglicherweise spezifisch mit der Nematoden Cuticula in der extrazellulären Matrix, beteiligt ist (Blaxter 1998). TSP-1-ähnliche (Thrombospondin type 1, SMART accession number: SM0209) Domänen werden innerhalb der Astacine exklusiv bei 6 *C. elegans* Proteinen (vergleiche auch Abbildung 31, Stammbaum-Gruppe D1) gefunden. Das Thrombospondin 1 der Säugetiere ist dabei ein Glycoprotein, das Zell-

Zell oder Zell-Matrix Interaktionen vermittelt (Lawler und Hynes 1986; Lawler und Hynes 1987). Von den carboxyterminal gelegenen Typ 1 Domänen innerhalb des Thrombospondin-Proteins konnte gezeigt werden, dass diese den TGF-beta Signal-Transduktionsweg aktivieren können und dadurch auch an der Proliferation, Migration und Differenzierung von Zellen direkt beteiligt sind (Crawford et al. 1998; Lawler et al. 1998). Bei dem bakteriellen Flavastacin wird im carboxyterminalen Bereich weiterhin eine Ricin-Domäne (SMART accession number: SM0458) gefunden. Ricin ist ein toxisches pflanzliches Lektin, welches erstmals bei *Ricinus communis* beschrieben wurde. Einige weitere bei den Astacinen gefundene Domänen sind bisher ausschließlich aufgrund von offensichtlichen Sequenzhomologien beschrieben, aber in keinem Fall näher untersucht worden. Dazu gehören die EB-Domäne (PFAM accession number: PF01683), welche durch 8 konservierte Cysteinreste charakterisiert ist und ausschließlich in verschiedenen *C. elegans* Proteinen gefunden wird (Astacine: C26C6.3), die MATH-Domäne (SMART accession number: SM0061) der Meprine und die TT-Domäne (PFAM accession number: PF02957), welche Homologien zu einem humanpathogenen TT-Virus aufweist. Für die Meprine hat man aus praktischen Gesichtspunkten weitere Domänen benannt (Intervening, I-Domäne, Transmembran-Domäne), ohne dass hier bisher Homologien zu anderen Proteinen bekannt wären (Kumar und Bond 2001).

Analysiert man die in den Abbildungen 26 bis 28 dargestellten Domänenstrukturen aller Mitglieder der Astacin-Protein-Familie auch hinsichtlich der physiologischen Funktionen, die diese Proteine übernommen haben (für eine Literaturübersicht siehe Auflistung in der Einleitung), können bereits einige Gemeinsamkeiten erkannt werden. Um aber gleichzeitig weitere Aussagen über die Proteinevolution der Astacin-Familie machen zu können, wurde die molekularphylogenetische Analyse (*Neighbor-Joining*-Stammbaum, Abbildungen 30 und 31) mit in die Diskussion einbezogen.

Zunächst konnte festgestellt werden, dass die Astacine in mehrere phylogenetische Gruppen unterteilt werden können. Die Stammbaum-Gruppen der Cnidaria-Astacine (Gruppe A), der Meprine (Gruppe B), der BMP-1/Tolloid-homologen Proteine (Gruppe C) unterscheiden sich im Stammbaum klar von der großen Gruppe D der restlichen Astacine (Abbildung 30). Unter Einbeziehung aller *C. elegans* und *Drosophila* Proteine, konnte die Gruppe D nochmals in die vier Subgruppen D1 bis D4 eingeteilt werden (Abbildung 31). Aufgrund der geringen Sequenzidentität, verbunden mit einer schlechten statistischen Unterstützung der Verzweigungen (*Bootstrap*-Werte <20), ist eine vollständige phylogenetische Interpretation innerhalb der Gruppe D zum Teil jedoch schwer durchführbar.

Innerhalb der Stammbaum-Gruppe A finden sich nun die Cnidaria-Astacine HMP-2, HEA-2 und PMP-1, entsprechend ihrer basalen phylogenetischen Stellung im Tierreich, als nächste Verwandte des bakteriellen Flavastacins. HMP-1 besitzt zwar die gleiche Domänenstruktur wie die anderen Cnidaria-Astacine, zeigt aber innerhalb der Protease-Domäne eine größere Homologie zu den Astacinen der Gruppe D. Als nächste Verwandte erscheinen hier die *C. elegans* Proteine der Subgruppe D4. Überraschend ist auch, dass HMP-2 aufgrund der Sequenzhomologie der Protease-Domäne der Gruppe A zugeordnet wird, aber als einziges Astacin dieser Gruppe darüber hinaus auch eine Meprin-spezifische

MAM-Domäne aufweist. Strukturell und funktionell ist allen Cnidaria-Astacinen (HEA-2, HMP-1, HMP-2, PMP-1, FARM-1), unabhängig von ihrer Stellung im Stammbaum, das Auftreten einer ShKT-Domäne und die Beteiligung an entwicklungssteuernden Prozessen (vergleiche V.2.) gemeinsam.

In der phylogenetischen Analyse erscheinen die Meprine als klar abgegrenzte Gruppe (B), die zwischen den Cnidaria-Astacinen der Gruppe A und den BMP-1/Tolloid-ähnlichen Proteinen der Gruppe C stehen. Interessanterweise werden die Meprine aber nur bei Säugetieren gefunden. Die Meprine besitzen innerhalb der Astacine eine einzigartige Domänenstruktur mit einer MAM-, MATH-, *intervening*-, Transmembran- und cytoplasmatischen Domäne. Sie sind die einzigen membranständigen Enzyme und nur in dimerisiertem Zustand aktiv. Die Domänenstruktur der Meprine ist funktionell also durch diesen Zusammenhang begründet.

Alle Proteine der Stammbaum-Gruppe C lassen sich von der Multi-Domänenstruktur des BMP-1 (Mensch) bzw. Tolloids (*Drosophila*) ableiten, die carboxyterminal mehrere EGF- und CUB-Domänen besitzen. Dazu gehören Tolkin, TLL (Mensch, Maus), TLL-2 (Mensch, Maus), Tolloid (Zebrafisch), Colloid (Huhn) und die um zwei CUB-Domänen verkürzten Proteine BMP-H (Seeigel) und BMP-1 (*Xenopus*, Huhn), sowie BP10 (Seeigel) und SPAN (Seeigel). Diese Proteasen sind alle als Morphogene beschrieben worden und spielen spezifische Rollen in der Embryonalentwicklung des jeweiligen Tieres. So konnte beispielsweise für Tolloid eine Beteiligung an der Etablierung der dorso-ventralen Körperachse gezeigt werden, in dem der Transkriptionsfaktor *short gastrulation* (SOG), einem Antagonisten des TGFβ-ähnlichen *decapentaplegic* (DPP), proteolytisch gespalten wird. Eine ähnliche Aktivität konnte für BMP-1 gezeigt werden, wobei in Vertebraten (*Xenopus*, Zebrafisch, Maus) das SOG-homologe Chordin gespalten wird und das TGFβ-ähnliche BMP-4 freisetzt (für einen Überblick siehe auch Weinmaster 1998). Von besonderem Interesse ist hier noch, dass in *C. elegans* ein Protein (F38E9.2) identifiziert werden konnte, das sowohl hinsichtlich der Domänenstruktur als auch der Sequenzhomologie innerhalb der Protease-Domäne (74% Sequenz-Identität zu BMP-1 des Menschen), eindeutig den BMP-1/Tolloid-homologen Proteinen zugeordnet werden kann.

Weisen die Gruppen A, B und C sowohl hinsichtlich der Funktion, Domänenstruktur und der Sequenzhomologie der Proteasedomäne hohe Übereinstimmungen auf, so erweist sich die Gruppe D als recht heterogen. Dennoch fallen zunächst die *hatching*-Enzyme der Fische (Abbildung 30: *Anguilla* EHE-4, -6, -7, -10, -13 und -14, *Oryzias* HCE-1, HCE-2 und LCE, *Cyprinus* Nephrosin, *Takifugu* HCE-1) auf, die nur aus der Protease-Domäne bestehen und hier sehr hohe Sequenzhomologien aufweisen. Die sechs *Anguilla*-Astacine müssen dabei aufgrund der hohen Sequenzidentität (>90%) innerhalb dieser Spezies durch Duplikation eines ursprünglichen Gens entstanden sein. Gleiches gilt für die drei *Oryzias*-Enzyme.

Bezieht man nun alle *C. elegans* und *Drosophila* Proteine in die Analyse ein, so lassen sich die Subgruppen D1 bis D4 bilden (Abbildung 31). Subgruppe D3 enthält danach die eben genannten *hatching*-Enzyme der Fische und zusätzlich das Verdauungsenzym Astacin von *Astacus* sowie die *hatching*-Enzyme AEA von *Astacus* und UVS-2 von *Xenopus*. Außer AEA und UVS-2 bestehen diese Proteine alle nur aus der Protease-Domäne. Für die Ausübung ihrer physiologischen Funktion werden offenbar auch keine regulatorischen

Elemente benötigt, da sie an der einfachen Degradation der Eihülle oder von Nahrungsbestandteilen beteiligt sind. Die zusätzlichen CUB-Domänen von AEA und UVS-2 weisen möglicherweise auf eine unterstützende Funktion bei der Substratbindung hin. Besonders interessant ist hier noch, dass innerhalb der Subgruppe D3 auch das neu identifizierte humane HCH-1 Protein gefunden wird, welches carboxyterminal der Protease-Domäne nur eine kurze TT-Domäne unbekannter Funktion enthält. Zukünftige Untersuchungen sollen Aufschlüsse über die Funktion dieses Astacin-homologen Proteins beim Menschen geben.

Die Subgruppe D2 wird von 8 kleinen, kompakten *Drosophila*-Astacinen gebildet (CGx), die vergleichbar den *Anguilla*- oder *Oryzias*-Astacinen, direkt als paraloge *Drosophila*-Proteine erkannt werden können. Ein weiteres Insekten Astacin (*Glossina*, M15242) wird ebenfalls in dieser Subgruppe gefunden. Die Proteine CG10280 und CG11865 sind wahrscheinlich ebenfalls dieser Subgruppe zuzuordnen und erscheinen im Stammbaum, nur aufgrund der schlechten Unterstützung der Knotenpunkte (*Bootstrap* Werte <20) in einer anderen Subgruppe. Da die genannten Insekten Astacine keine carboxyterminalen Domänen enthalten, könnten diese Gene in *Drosophila* oder *Glossina* möglicherweise wie die Proteine der D3-Subgruppe eine Rolle bei *hatching*-Prozessen oder der Nahrungsverdauung spielen. Für letzteres spricht auch die Tatsache, dass vor kurzem ein Astacin-ähnliches Verdauungsenzym bei dem Maikäfer *Melolontha melolontha* (Wagner et al. 2002) beschrieben wurde. Somit scheint sich innerhalb der Arthropoden nicht nur bei den dekapoden Crustaceen, sondern auch den Insekten eine Verdauungsprotease vom Astacin-Typ zu finden.

In den Subgruppen D1 und D4 sind nun fast ausschließlich *C. elegans* Astacine repräsentiert. Hier treten nun weitere neue Domänenkompositionen auf, wobei keine Übereinstimmung zwischen der phylogenetischen Verwandtschaft dieser Proteine und der Domänenanordnung mehr erkannt werden kann. Zusammenfassend finden sich hier 5 Proteine ohne carboxyterminale Domänen, 10 Proteine mit einer bis drei ShKT Domänen, sowie 15 Proteine mit einer EGF- und einer CUB-Domäne, 2 Proteine mit einer EGF-, einer CUB- und einer ShKT-Domäne, sowie 6 Proteine mit einer EGF-, einer CUB- und einer TSP-1-Domäne. Hinsichtlich der Funktion dieser unterschiedlichen Proteine konnten in dieser Arbeit exemplarisch die Proteine C05D11.6 (keine carboxyterminalen Domänen) und C07D10.4 (mit ShKT-Domänen) bearbeitet werden (vergleiche V.1). Die Expressionsanalysen durch GFP-Reporterproteine legten dabei eine Verdauungsfunktion für C05D11.6 und eine Beteiligung von C07D10.4 am Schlüpfvorgang der Embryonen nahe. Die RNAi Experimente von Maeda et al. (2001) konnten darüber hinaus für die Proteine C37H5.9 und K11G12.1 (beide eine ShKT-Domäne) eine entwicklungssteuernde Funktion zeigen. Für das HCH-1 Protein (carboxyterminal eine EGF-, eine CUB- und eine TSP-1-Domäne) konnten Hishida et al. (1996) ein Doppelfunktion beim Schlüpfen der Embryonen und eine gestörte Wegfindung bestimmter Neuroblasten nachweisen. Das C17G1.6 Protein hat die gleiche Domänenanordnung wie HCH-1, zeigt nach den RNAi Experimenten von Maeda et al. (2001) aber eine völlig andere physiologische Funktion. Danach ist es an der Vulva-Entwicklung in frühen Embryonalstadien beteiligt. Dennoch deuten zusammenfassend alle bisherigen funktionellen Daten zusammen mit der komplexen Domänenanordnung für die Mehrzahl der Astacin-homologen Proteine in *C.*

elegans verschiedene spezifische Rollen bei der Zelldifferenzierung oder Entwicklung an. Die Funktionen aller Astacin-homologen Proteine von *C. elegans* müssen für ein besseres Verständnis der Evolution dieser Proteine aber sicherlich noch weiter analysiert werden.

Besonders interessant ist hier auch das Evolutions-Prinzip, welches an der Astacin-Protein-Familie insgesamt beobachtet wird, welches aber bei den Astacinen von *C. elegans* besonders ausgeprägt und mit neuen Elementen in Erscheinung tritt. An die katalytische Kette, die eigentliche Astacin-Domäne, werden die carboxyterminalen Domänen in unterschiedlicher Komposition angehängt und erlauben es, die Astacin-Domäne wie ein Modul in wechselnder Umgebung und zu unterschiedlichen Zwecken einzusetzen. Auffällig ist bei *C. elegans* darüber hinaus auch, dass die Astacin-homologen Proteine ohne erkennbare Präferenz auf alle Chromosomen verteilt sind, und dass sie untereinander und im Vergleich zu anderen Astacinen nur geringe Sequenzhomologien aufweisen. Hier zeichnet sich für *C. elegans* das von Baxter (1998) beschriebene Phänomen ab, dass viele Gene in diesem Organismus eine viel höhere Evolutionsgeschwindigkeit als dieselben Gene in anderen Spezies aufweisen (Baxter 1998). Überraschenderweise bildet aber das BMP-1/Tolloid-ähnliche Protein F38E9.2 aus *C. elegans* (Gruppe C) eine Ausnahme. In diesem Astacin-homologen Protein findet sich eine hohe Sequenzidentität und eine exakt gleiche Domänenstruktur wie bei den BMP-1-ähnlichen Astacinen der Vertebraten oder den Tolloiden von *Drosophila* (siehe oben). Für diese Genlinie scheint also eine hohe Konservierung vorzuliegen, die in der spezifischen Funktion dieses Proteins begründet sein könnte. Die restlichen 39 Astacin-homologen Proteine von *C. elegans* lassen sich aufgrund der beobachteten hohen Evolutionsgeschwindigkeit wahrscheinlich auf nur zwei (Subgruppen D1 und D4) Vorläufer-Gene zurückführen, wobei deren Entwicklungslinien aber wegen der geringen Sequenzhomologien im Stammbaum nicht mehr klar erkennbar sind.

Auf die Frage, ob die hohe Divergenz der Astacin-homologen Proteine von *C. elegans* nun ein exklusives Merkmal dieses Organismus darstellt, konnte das Genomprojekt von *Caenorhabditis briggsae* Auskunft geben. Bei diesem Nematoden konnten analog zu *C. elegans* die große Kopienzahl von insgesamt 34 Genfragmenten identifiziert werden (Tabelle 13). Vorläufige Untersuchungen anhand der nur teilweise vorliegenden Sequenzdaten zeigen, dass für jedes *C. briggsae* Astacin ein orthologes Protein in *C. elegans* gefunden werden kann. Inwiefern solche Genduplikationen nun aber auch bei anderen weniger verwandten Spezies innerhalb des Tierstamms der Nematoden vorkommen, werden aber erst die laufenden Genom- oder EST-Projekte zeigen. In Kürze sollen die ersten Daten des Genomprojekts des Sanger Instituts für *Brugia malayi* (<http://www.sanger.ac.uk/Projects/Nematodes>) und das Nematoden *EST Sequencing Project* des Sanger Instituts in Kollaboration mit Mark Blaxter, University of Edinburgh, für *Ascaris suum*, *Haemonchus contortus*, *Necator americanus*, *Teladorsagia circumcincta*, *Trichuris muris* (<http://www.nematodes.org>) vorliegen. Erst dann kann die interessante Frage beantwortet werden, ob die große Anzahl von Genduplikationen Astacin-homologer Proteine eine Sonderentwicklung innerhalb der Ordnung Rhabdita (*C. elegans* und *C. briggsae*) darstellt oder ob dieses Phänomen auch bei anderen Ordnungen innerhalb der Nematoden gefunden wird.

Eine vollständige Analyse aller paralogen und orthologen Astacine in einem Organismus kann derzeit nur bei *C. elegans*, *Drosophila* und dem Menschen erfolgen, da nur hier die Genom-Sequenzierungs-Projekte abgeschlossen sind. Im Vergleich zu der bei *C. elegans* bisher einzigartigen Komplexität, finden sich bei *Drosophila* nur 11 kleine kompakte Proteasen (CG10280, CG11865, Subgruppe D2) und die beiden BMP-1-ähnlichen Proteine Tolloid und Tolkin. Beim Menschen findet man dagegen nur 1 kompaktes Astacin (HCH-1) sowie 3 Gene für BMP-1 ähnliche Proteine (BMP-1, TLL, TLL-2), die aber in unterschiedlichen Spleißvarianten (BMP-1 splice 1 bis 7) vorkommen. Möglicherweise ist dies ein Hinweis auf die Frage, wieso der Mensch viel weniger Gene besitzt als ursprünglich angenommen. Derzeit geht man bei einer Genomgröße von 3000 Mio. Basenpaaren von 24.000 (*genescan prediction*) bis maximal 80 000 Gene aus (Hubbard et al. 2002). Zum Vergleich werden bei *C. elegans* mit 97 Mio. Basenpaaren 20.000 Gene (*wormbase*) und bei *Drosophila* mit 180 Mio. Basenpaaren 14.000 Gene (Adams et al. 2000) angenommen. Interessant ist in diesem Zusammenhang ebenfalls die Tatsache, dass im Hefegenom (12 000 Kilo-Basenpaare, 7000 Gene; Goffeau et al. 1996) kein Astacin-homologes Gen gefunden wird, wobei aber in zwei Bakterienspezies (Flavastacin in *Flavobacterium*; genomisches Sequenzfragment MP-ORF-2 in *Legionella pneumophila*) Astacin-homologe Proteine identifiziert werden konnten.

Zusammenfassend kann festgestellt werden, dass die verschiedenen Gruppen und Subgruppen der Astacin-homologen Proteine jeweils auf ein Vorläufer-Protein zurückzuführen sein müssen, die ihrerseits durch Genduplikationen eines ursprünglichen Vorläufer-Gens entstanden sind. Dieser Vorgang muss sich in der Astacin-Protein-Familie im Laufe der Evolution auf mehreren Ebenen wiederholt haben. In Bakterien wird lediglich ein Astacin-homologes Protein gefunden (*Legionella pneumophila*, *Flacobacterium meningosepticum*), während innerhalb des basalen Tierstamms der Cnidaria bereits bis zu drei Kopien (*Hydra*) vorkommen. Die BMP-1/Tolloid-ähnlichen Proteine und Meprine bilden dann klar abgegrenzte Gruppen, die im Verlauf der Evolution auch ähnliche Funktionen beibehalten haben. Die große Gruppe aller anderen Astacine ist entwicklungs-geschichtlich nicht mehr eindeutig differenzierbar, doch grenzen sich verschiedene Subgruppen, wie die *hatching*-Enzyme der Fische (Subgruppe D3) und die *Drosophila*-Proteine der Subgruppe D2 voneinander ab.

4. Aktivierung von Pro-Astacin

Trotz aller Veränderungen, welche die Astacine im Verlauf der Evolution erfahren haben, sind möglicherweise einige allgemeine Mechanismen hoch konserviert. Ein Ansatzpunkt sollte hier die Analyse der Propeptide (Zymogene) sein, welche bei allen Astacin-homologen Proteinen gefunden werden.

Als Modellprotein wurde für diese Untersuchungen das Astacin des Flusskrebs *Astacus astacus* verwendet, welches das innerhalb der gesamten Protein-Familie am besten charakterisierte Enzym darstellt. So steht seit langem ein einfaches biochemisches Reinigungsverfahren der nativen Protease zur Verfügung (Pfleiderer et al. 1967). Auf biochemischer Ebene konnte daraufhin die Spaltspezifität (Sonneborn et al. 1969; Krauhs

et al., 1982) und die vollständige Aminosäuresequenz des nativen Proteins durch Edman-Abbau (Titani et al. 1987) ermittelt werden. Kinetische Untersuchungen mit dansylierten Oligopeptid-Substrate (Ng and Auld 1989; Stöcker et al. 1990), Nitroanilin-Substraten (Stöcker et al. 1991) oder verschiedene Inhibitoren (Grams et al. 1996; Yiallourous et al. 1998) ergänzten die biochemischen Charakterisierungen. Schließlich konnte vom Flusskrebs-Astacin als einzigem Mitglied dieser Protein-Familie auch die Raumstruktur mit 1,8Å^o Auflösung ermittelt werden (Bode et al. 1992; Gomis-Rüth et al. 1993). Die Isolierung der cDNA und die bakterielle Expression des nativen Enzyms (Reyda et al. 1999) sowie verschiedener mutagenisierter Varianten erlaubten detaillierte Aussagen über den Katalyse-Mechanismus des Astacins (Yiallourous et al. 2000). Darüber hinaus gelang die Aufklärung der genomischen Struktur durch (Geier et al. 1997). Der Synthesort des Enzyms im Flusskrebs-Hepatopankreas konnte ebenfalls lokalisiert werden (Vogt et al., 1989). Weiterhin sind das Verdauungssystem und die an dem Verdauungs-Prozess beteiligten proteolytischen Enzyme des Flusskrebses *Astacus astacus* sehr gut charakterisiert worden (einen Überblick geben Zwilling und Neurath 1981; Vogt 1996 ; Vogt 2002). Trotz dieser weitreichenden Kenntnisse ist weder für das Pro-Enzym (Zymogen) des Flusskrebs-Astacins noch für irgendeinen anderen Vertreter der Astacin-Protein-Familie bisher der Aktivierungs-Mechanismus, der Ort der Aktivierung oder der *in vivo* an der Aktivierung beteiligten Proteasen untersucht worden.

Immunhistochemische Experimente mit Antikörpern die gegen das reife Astacin gerichtet sind und damit auch das Pro-Enzym erkennen können, bestätigten zunächst die Untersuchungen von Vogt et al. (1989), wonach dieses Enzym ausschließlich in den F-Zellen des Hepatopankreas synthetisiert wird. Die punktförmige intrazelluläre Fluoreszenz konnte bereits von Vogt et al. (1989) durch elektronenmikroskopische Vergleiche den Golgi-Vesikeln zugeordnet werden. Extrazellulär konnte Astacin in allen Bereichen des hepatopankreatischen Tubulus-Systems gefunden werden. Überraschenderweise wurde nun mit zwei verschiedenen Antikörpern (Pro-Ast-N, Pro-Ast-C), die spezifisch nur den N-terminalen oder C-terminalen Teil des Pro-Peptids erkennen, Pro-Astacin vom Synthesort in den F-Zellen bis in die Lumina des Tubulus-Systems und selbst noch in dem zum Magen führenden Sammelkanal gefunden. Die Immunoblot Experimente zeigten dann einerseits die Spezifität der verwendeten Antikörper, aber auch das ausschließliche Vorkommen des aktiven Enzyms im Magen. Danach kann Pro-Astacin erst in den proximalen Teilen des Sammelkanals oder direkt vor Übertritt in den Magen aktiviert werden. In den Immunoblot Experimenten konnte ebenfalls festgestellt werden, dass drei Stunden nach der Stimulierung der Enzymneusynthese durch Entnahme von Magensaft die Menge an Pro-Astacin im Vergleich zum reifen Enzym deutlich zugenommen hatte. Dies stimmt mit den von Vogt et al. (1989) durchgeführten Stimulierungs-Experimenten überein. Eine mögliche Kreuzreaktion der Pro-Astain Antikörper mit dem freien Propeptid konnte dadurch ausgeschlossen werden, dass in den Immunoblot-Experimenten keine Detektion des ca. 5 kDa großen Peptids beobachtet werden konnte. Um zu überprüfen, ob die geringe Größe der Propeptide eine Immunodetektion möglicherweise verhindern könnte, wurden in Voruntersuchungen auch die synthetischen Pro-Peptide (Ast-C, Ast-N) auf PVDF-Membranen übertragen, wobei diese eindeutig eine immunopositive Reaktion

zeigten (Thesa, 1998). Diese Versuche sollten also eine Interferenz von freiem, vom Pro-Enzym bereits abgespaltenen Pro-Peptid mit der tatsächlichen Pro-Astacin Lokalisation im Hepatopankreas ausschließen. Darüberhinaus würde ein freies Pro-Peptid natürlich auch bedeuten, dass bereits genügend aktive Proteasen vorliegen, die das Peptid rasch weiter degradieren würden.

Ein Sequenzvergleich der gesamten Astacin-Protein-Familie im Bereich der Aktivierungsstelle zeigt, dass in Position P1 häufig basische Aminosäurereste lokalisiert sind und eine Häufung oder *Cluster*-Bildung von Arginin- oder Lysin-Resten aminoterminal der Aktivierungsstelle (Abbildung 37 und 38) zu beobachten. Das Vorkommen der basischen Reste in Position P1 legt in diesen Fällen eine Beteiligung von Trypsin-ähnlichen Proteasen am Aktivierungsmechanismus nahe, da deren Spaltungsspezifität ausschließlich auf basische Reste (Arg, Lys) begrenzt ist (BRENDA *enzyme database*). Bei BP10, SPAN den Tolloid/BMP-1-ähnlichen und einigen *C. elegans* Proteinen findet sich in den Positionen P4 bis P1 auch eine Spaltstelle für Furin (E.C. 3.4.21.75). Diese Serinprotease spaltet nach dem Motif Arg-Xaa-Yaa-Arg, wobei Xaa eine beliebige und Yaa ein Arg oder Lys Rest repräsentiert. Furin-ähnliche Serinproteasen sind tatsächlich bei Vertebraten, *Drosophila* und *C. elegans* (F11A6.1a) beschrieben worden (BRENDA *enzyme database*). Auch konnte beispielsweise für die Endopeptidase-24.18 der Ratte ein solcher Aktivierungsmechanismus gezeigt werden (Milhiet et al. 1995). Auf der anderen Seite fehlen in mehreren anderen Astacinen basische Aminosäuren in Position P1. Dies gilt für die bei Cnidariern gefundenen Astacine (HEA-2, PMP-1, HMP-1), für das bei Fischen gefundene Nephrosin und für einige *C. elegans* Astacine. Aber auch im Astacin des Flusskrebses ist ein basischer Aminosäurerest erst in Position P3 vorhanden. Hier müssen also sicher auch andere proteolytische Enzyme in den Aktivierungsprozess involviert sein.

Im Verdauungstrakt von *Astacus astacus* werden allerdings nur zwei Endopeptidasen gefunden, welche überhaupt für die Aktivierung von Pro-Astacin in Frage kommen: Flusskrebs Trypsin und Astacin (Pfleiderer et al. 1967, Vogt 2002). Die Beteiligung dieser Proteasen an der Aktivierung konnte anhand des Spaltmusters synthetischer Modellpeptide, welche die Aktivierungsstelle des Astacins umfassen (Abbildung 37) überprüft werden. Astacin oder Trypsin wurden dazu mit den Peptiden AstG (repräsentiert dabei die Astacin Sequenz), AstR (Glycin -1 wurde durch das häufig bei anderen Astacinen in dieser Position vorkommende Arginin ersetzt) und AstR* (enthält D(-) Arginin in Position -3, -7 und -10, um als Kontrolle eine Kontamination der Astacin-Präparation mit Trypsin auszuschließen) inkubiert und die Produkte massenspektrometrisch analysiert (Abbildung 39). Astacin hydrolysierte im Peptid AstG ausschließlich die Gly-1/Ala+1 Bindung, welche die tatsächliche Aktivierungsstelle im Protein repräsentiert und spaltete auch dann noch immer exklusiv an dieser Position, wenn Gly-1 durch einen Argininrest substituiert wurde. Trypsin hydrolysierte die Peptide erwartungsgemäß an allen vorhandenen basischen Aminosäureresten, wobei der tatsächliche N-Terminus der reifen Astacin-Moleküls nicht entstehen kann. Nur bei Einführung eines Argininrestes in Position -1, welcher die Sequenz bei vielen anderen

Astacinen widerspiegelt, wurde die -1/+1 Bindung auch von Trypsin gespalten. Die Kontrollreaktionen mit dem AstR*-Peptid konnten jede tryptische Verunreinigung in den Astacin-Präparationen ausschließen. Diese Ergebnisse zeigen, dass das Astacin des Flusskrebses in der Lage ist, seine eigene Aktivierung zu katalysieren. Eine Aktivierung durch Trypsin würde die zusätzliche Einwirkung durch eine Aminopeptidase erfordern. Für *Astacus astacus* kann somit zusammenfassend festgestellt werden, dass Pro-Astacin vermutlich autokatalytisch im proximalen Bereich des Sammelkanals vor Übertritt in den Magen (Cardia) aktiviert wird. Ob eine erste initiale Aktivierung allerdings einer noch unbekanntem membranständigen Protease oder einer intrinsischen Aktivität von Pro-Astacin zugeschrieben werden kann, müssen weitere Studien zeigen. Darüber hinaus konnte aber auch gezeigt werden, dass Astacin in der Lage ist, eine Arg-Ala Bindung an der Aktivierungsstelle zu hydrolisieren. Damit könnte für die Mehrzahl der Astacine neben einer Aktivierung durch tryptische Enzyme ebenfalls eine autokatalytische Aktivierung in Frage kommen.

Vergleicht man die Erkennungssequenzen der Aktivierungsstelle mit der im vorigen Abschnitt diskutierten Unterteilung der Astacin-Familie in die Stammbaum-Gruppen, fallen einige Übereinstimmungen auf. Die Vertreter der BMP-1/Tolloid Gruppe zeigen mit Ausnahme des *C. elegans* Proteins F38E9.2 alle eine Furin-Erkennungssequenz. Zusammen mit den ähnlichen physiologischen Funktionen, die diese Proteine übernommen haben, könnte sich hier ein besonderer Aktivierungsmechanismus entwickelt haben. Das als ursprünglich anzusehende bakterielle Flavastacin, die *hatching*-Enzyme, die kurzen, kompakten *Drosophila* Proteine und die Meprine weisen dagegen einen einzelnen basischen Rest in Position P1 auf. Dieser scheint bei den Cnidaria-Astacinen, Nephrosin und Astacin verloren gegangen zu sein. Dennoch könnte sich aufgrund der hier dargestellten Ergebnisse für alle diese Subgruppen ein gemeinsamer autokatalytischer Aktivierungsmechanismus erhalten haben. In den 40 *C. elegans* Astacinen scheinen dagegen alle Möglichkeiten unterschiedlicher Aktivierungen realisiert zu sein, soweit dies die Pro-Peptid Struktur angeht. Es finden sich Furin-Erkennungssequenzen, basische Aminosäurereste in P1 oder auch keine basischen Aminosäurereste in dieser Position. Diese lassen sich aber weder mit der Domänenstruktur noch mit den im Stammbaum (Abbildung 31) abgeleiteten Verwandtschaftsverhältnissen homologisieren. Möglicherweise unterliegt der hier jeweils etablierte Aktivierungsmechanismus, begünstigt durch die große Variabilität dieser Gene in *C. elegans*, allein der Selektion durch die Expression dieser Proteine in unterschiedlich spezialisierten Zelltypen. Je nach der vorgefundenen Ausstattung mit anderen Proteasen kann sich dann eine autokatalytische, tryptische oder Furin-ähnliche Aktivierung etabliert haben.

VI. LITERATUR

Adams MD et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195.

Altschul SF, Gish W. 1996. Local alignment statistics. *Methods Enzymol* 266:460-480.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-410.

Appel RD, Palagi PM, Walther D, Vargas JR, Sanchez JC, Ravier F, Pasquali C, Hochstrasser DF. 1997. Melanie II--a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface. *Electrophoresis* 18:2724-2734.

Barrett AJ. 1977. Preface. In: Barrett AJ, editor. *Proteinases in mammalian cells and tissues*. Amsterdam - New York - Oxford: North-Holland Publishing Company.

Barrett AJ, Rawlings ND, O'Brien EA. 2001. The MEROPS database as a protease information system. *J Struct Biol* 134:95-102.

Barrett AJ, Rawlings ND, Woessner JF. 1998. *Handbook of Proteolytic Enzymes*. London: Academic Press.

Bini L, Heid H, Liberatori S, Geier G, Pallini V, Zwillig R. 1997. Two-dimensional gel electrophoresis of *Caenorhabditis elegans* homogenates and identification of protein spots by microsequencing. *Electrophoresis* 18:557-562.

Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D. 1993a. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 14:1023-1031.

Bjellqvist B, Pasquali C, Ravier F, Sanchez JC, Hochstrasser D. 1993b. A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis* 14:1357-1365.

Blader P, Rastegar S, Fischer N, Strahle U. 1997. Cleavage of the BMP-4 antagonist chordin by zebrafish tolloid. *Science* 278:1937-1940.

Blaxter M. 1998. *Caenorhabditis elegans* is a nematode. *Science* 282:2041-2046.

Bode W, Gomis-Rüth F, Huber R, Zwillig R, Stöcker W. 1992. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* 358:164-167.

Bond JS, Beynon RJ. 1995. The astacin family of metalloendopeptidases. *Protein Sci* 4:1247-1261.

Bond JS, Butler PE. 1987. Intracellular proteases. *Annu Rev Biochem* 56:333-64:-64.

Bork P, Beckmann G. 1993. The CUB domain. A widespread module in developmentally regulated proteins. *J Mol Biol* 231:539-545.

- Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-54.
- Burmester T. 1999. Identification, molecular cloning, and phylogenetic analysis of a non-respiratory pseudo-hemocyanin of *Homarus americanus*. *J Biol Chem* 274:13217-13222.
- Burmester T. 2001. Molecular evolution of the arthropod hemocyanin superfamily. *Mol Biol Evol* 18:184-195.
- Burmester T, Massey HC, Zakharkin SO, Benes H. 1998. The evolution of hexamerins and the phylogeny of insects. *J Mol Evol* 47:93-108.
- Cascio P, Hilton C, Kisselev AF, Rock KL, Goldberg AL. 2001. 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *EMBO J* 20:2357-2366.
- Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263:802-805.
- Checler F, da Costa CA, Ancolio K, Chevallier N, Lopez-Perez E, Marambaud P. 2000. Role of the proteasome in Alzheimer's disease. *Biochim Biophys Acta* 1502:133-138.
- Chomczynski P, Sacchi N. 1987. Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Anal Biochem* 162:156-159.
- Coux O, Goldberg AL. 1998. Enzymes catalyzing ubiquitination and proteolytic processing of the p105 precursor of nuclear factor kappaB1. *J Biol Chem* 273:8820-8828.
- Crawford SE, Stellmach V, Murphy-Ullrich JE, Ribeiro SM, Lawler J, Hynes RO, Boivin GP, Bouck N. 1998. Thrombospondin-1 is a major activator of TGF-beta1 in vivo. *Cell* 93:1159-1170.
- Curran S, Murray GI. 2000. Matrix metalloproteinases: molecular aspects of their roles in tumour invasion and metastasis. *Eur J Cancer* 36:1621-1630.
- Dower WJ, Miller JF, Ragsdale LW. 1988. High-efficiency transformation of *E. coli* by high voltage electroporation. *Nucl Acid Res* 16:6127-6145.
- Dumermuth E, Sterchi EE, Jiang WP, Wolz RL, Bond JS, Flannery AV, Beynon RJ. 1991. The astacin family of metalloendopeptidases. *J Biol Chem* 266:21381-21385.
- Elaroussi MA, DeLuca HF. 1994. A new member to the astacin family of metalloendopeptidases: a novel 1,25-dihydroxyvitamin D-3-stimulated mRNA from chorioallantoic membrane of quail. *Biochim Biophys Acta* 1217:1-8.
- Fan TJ, Katagiri C. 2001. Properties of the hatching enzyme from *Xenopus laevis*. *Eur J Biochem* 268:4892-4898.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.

- Frank U, Leitz T, Müller WA. 2001. The hydroid *Hydractinia*: a versatile, informative cnidarian representative. *Bioessays* 23:963-971.
- Gajewski M, Leitz T, Schlossherr J, Plickert G. 1996. LWamides from Cnidaria constitute a novel family of neuropeptides with morphogenetic activity. *Roux's Archives of Developmental Biology* 205:232-242.
- Geddes VA, Le Bonniec BF, Louie GV, Brayer GD, Thompson AR, MacGillivray RT. 1989. A moderate form of hemophilia B is caused by a novel mutation in the protease domain of factor IXVancouver. *J Biol Chem* 264:4689-4697.
- Geier G, Jacob E, Stocker W, Zwilling R. 1997. Genomic organization of the zinc-endopeptidase astacin. *Arch Biochem Biophys* 337:300-307.
- Geier G, Zwilling R. 1998. Cloning and characterization of a cDNA coding for *Astacus* embryonic astacin, a member of the astacin family of metalloproteases from the crayfish *Astacus astacus*. *Eur J Biochem* 253:796-803.
- Gems D, Ferguson CJ, Robertson BD, Nieves R, Page AP, Blaxter ML, Maizels RM. 1995. An abundant, trans-spliced mRNA from *Toxocara canis* infective larvae encodes a 26-kDa protein with homology to phosphatidylethanolamine-binding proteins. *J Biol Chem* 270:18517-18522.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 genes. *Science* 274:546,563-546,567.
- Gomis-Rüth F, Stöcker W, Huber R, Zwilling R, Bode W. 1993. Refined 1.8 Å X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. *J Mol Biol* 229:945-968.
- Görg A, Postel W, Gunther S. 1988. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 9:531-546.
- Grams F, Dive V, Yiotakis A, Yiallouris I, Vassiliou S, Zwilling R, Bode W, Stöcker W. 1996. Structure of astacin with a transition-state analogue inhibitor. *Nat Struct Biol* 3:671-675.
- Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166:577-580.
- Hartley BS. 1960. Proteolytic enzymes. *Ann Rev Biochem* 29:45-72.
- Hishida R, Ishihara T, Kondo K, Katsura I. 1996. *hch-1*, a gene required for normal hatching and normal migration of a neuroblast in *C. elegans*, encodes a protein related to TOLLOID and BMP-1. *EMBO J* 15:4111-4122.
- Hochstrasser DF, Harrington MG, Hochstrasser AC, Miller MJ, Merrill CR. 1988a. Methods for increasing the resolution of two-dimensional protein electrophoresis. *Anal Biochem* 173:424-435.

Hochstrasser DF, Patchornik A, Merrill CR. 1988b. Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Anal Biochem* 173:412-423.

Hope IA. 1999. *C. elegans*. New York: Oxford University Press.

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M. 2002. The Ensembl genome database project. *Nucleic Acids Res* 30:38-41.

Huber R, Bode W. 1978. Structural Basis of the Activation and Action of Trypsin. *Acc Chem Res* 11:114-122.

Hung CH, Huang HR, Huang CJ, Huang FL, Chang GD. 1997. Purification and cloning of carp nephrosin, a secreted zinc endopeptidase of the astacin family. *J Biol Chem* 272:13772-13778.

Inohaya K, Yasumasu S, Araki K, Naruse K, Yamazaki K, Yasumasu I, Iuchi I, Yamagami K. 1997. Species-dependent migration of fish hatching gland cells that express astacin-like proteases in common [corrected]. *Dev Growth Differ* 39:191-197.

Ito A. 1999. Mitochondrial processing peptidase: multiple-site recognition of precursor proteins. *Biochem Biophys Res Commun* 265:611-616.

Jensen ON, Podtelejnikov AV, Mann M. 1997. Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. *Anal Chem* 69:4741-4750.

Jiang M, Ryu J, Kiraly M, Duke K, Reinke V, Kim SK. 2001. Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 98:218-223.

Kaji H, Tsuji T, Mawuenyega KG, Wakamiya A, Taoka M, Isobe T. 2000. Profiling of *Caenorhabditis elegans* proteins using two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21:1755-1765.

Kent WJ, Zahler AM. 2000. The intronerator: exploring introns and alternative splicing in *Caenorhabditis elegans*. *Nucleic Acids Res* 28:91-93.

Khasigov PZ, Podobed OV, Ktzoeva SA, Gatagonova TM, Grachev SV, Shishkin SS, Berezov TT. 2001. Matrix metalloproteinases of normal human tissues. *Biochemistry (Mosc)* 66:130-140.

Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS. 2001. A gene expression map for *Caenorhabditis elegans*. *Science* 293:2087-2092.

Klimova OA, Vedishcheva IV, Strongin AI. 1991. Vydelenie i kharakteristika kollagenoliticheskikh fermentov iz gepatopankreasa kraba-striguna *Chionoecetes opilio*.

[Isolation and characteristics of collagenolytic enzymes from the hepatopancreas of the crab *Chionoecetes opilio*]. Dokl Akad Nauk SSSR 317:482-484.

Köhler D, Kruse M, Stöcker W, Sterchi EE. 2000. Heterologously overexpressed, affinity-purified human meprin alpha is functionally active and cleaves components of the basement membrane in vitro. FEBS Lett 465:2-7.

Krauhs E, Dorsam H, Little M, Zwilling R, Ponstingl H. 1982. A protease from *Astacus fluviatilis* as an aid in protein sequencing. Anal Biochem 119:153-157.

Kumar JM, Bond JS. 2001. Developmental expression of meprin metalloprotease subunits in ICR and C3H/He mouse kidney and intestine in the embryo, postnatally and after weaning. Biochim Biophys Acta 1518:106-114.

Kumpfmüller G, Rybakine V, Takahashi T, Fujisawa T, Bosch TC. 1999. Identification of an astacin matrix metalloprotease as target gene for Hydra foot activator peptides. Dev Genes Evol 209:601-607.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

Lawler J, Hynes RO. 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. J Cell Biol 103:1635-1648.

Lawler J, Hynes RO. 1987. Structural organization of the thrombospondin molecule. Semin Thromb Hemost 13:245-254.

Lawler J, Sunday M, Thibert V, Duquette M, George EL, Rayburn H, Hynes RO. 1998. Thrombospondin-1 is required for normal murine pulmonary homeostasis and its absence causes pneumonia. J Clin Invest 101:982-992.

Lei H, Furth EE, Kalluri R, Wakenell P, Kallen CB, Jeffrey JJ, Leboy PS, Strauss JF. 1999. Induction of matrix metalloproteinases and collagenolysis in chick embryonic membranes before hatching. Biol Reprod 60:183-189.

Lewis JA, Fleming JT. 1995. Basic Culture Methods. In: Epstein HF, Shakes DC, editors. *Caenorhabditis elegans: Modern Biological Analysis of an Organism*. San Diego: Academic Press. p 4-29.

Lhomond G, Ghiglione C, Lepage T, Gache C. 1996. Structure of the gene encoding the sea urchin blastula protease 10 (BP10), a member of the astacin family of Zn²⁺-metalloproteases. Eur J Biochem 238:744-751.

Liaubet L, Bertrand N, Medevielle F, Pituello F. 2000. Identification by differential display of a chicken tolloid-related metalloprotease specifically expressed in the caudal notochord. Mech Dev 96:101-105.

Liu QR, Hattar S, Endo S, MacPhee K, Zhang H, Cleary LJ, Byrne JH, Eskin A. 1997. A developmental gene (Tolloid/BMP-1) is regulated in *Aplysia* neurons by treatments that induce long-term sensitization. J Neurosci 17:755-764.

Lumbers ER. 1999. Angiotensin and aldosterone. Regul Pept 80:91-100.

- Maeda I, Kohara Y, Yamamoto M, Sugimoto A. 2001. Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr Biol* 11:171-176.
- Maeda O, Smithies O. 1986. The evolution of multigene families: human haptoglobin genes. *Annu Rev Genet* 20:81-108.
- Marchand P, Volkman M, Bond JS. 1996. Cysteine mutations in the MAM domain result in monomeric meprin and alter stability and activity of the proteinase. *J Biol Chem* 271:24236-24241.
- Margolius HS. 1984. The kallikrein-kinin system and the kidney. *Annu Rev Physiol* 26:46309-26.
- Margolius HS. 1998. Kallikreins, kinins and cardiovascular diseases: a short review. *Biol Res* 31:135-141.
- Matsudaira P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* 262:10035-10038.
- Milhiet PE, Chevallier S, Corbeil D, Seidah NG, Crine P, Boileau G. 1995. Proteolytic processing of the alpha-subunit of rat endopeptidase-24.18 by furin. *Biochem J* 309:683-688.
- Mohier E. 1993. [Determination of the dorso-ventral polarity of the *Drosophila* embryo]. *Ann Genet* 36:5-15.
- Moore WM, Spilburg CA. 1986. Purification of human collagenases with a hydroxamic acid affinity column. *Biochemistry* 25:5189-5195.
- Morisato D, Anderson KV. 1995. Signaling pathways that establish the dorsal-ventral pattern of the *Drosophila* embryo. *Annu Rev Genet* 29:371-99:-99.
- Möhrlen F, Baus S, Gruber A, Rackwitz HR, Schnölzer M, Vogt G, Zwillig R. 2001. Activation of pro-astacin. Immunological and model peptide studies on the processing of immature astacin, a zinc-endopeptidase from the crayfish *Astacus astacus*. *Eur J Biochem* 268:2540-2546.
- Muller L, Lindberg I. 1999. The cell biology of the prohormone convertases PC1 and PC2. *Prog Nucleic Acid Res Mol Biol* 63:69-108:-108.
- Neuhoff V, Arold N, Taube D, Ehrhardt W. 1988. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9:255-262.
- Neurath H. 1984. Evolution of proteolytic enzymes. *Science* 224:350-357.
- Neurath H. 1986. The versatility of proteolytic enzymes. *J Cell Biochem* 32:35-49.
- Neurath H. 1989. Proteolytic processing and physiological regulation. *Trends Biochem Sci* 14:268-271.

Neurath H, Zwilling R. 1986. Willy Kühne und die Anfänge der Enzymologie. Semper Apertus, Sechshundert Jahre Ruprecht-Karls-Universität Heidelberg 1386-1986, Festschrift in sechs Bänden, Band II. Berlin - Heidelberg - New York: Springer Verlag.p 361-374.

Ng M, Auld DS. 1989. A fluorescent oligopeptide energy transfer assay with broad applications for neutral proteases. *Anal Biochem* 183:50-56.

Nixon RA, Cataldo AM, Mathews PM. 2000. The endosomal-lysosomal system of neurons in Alzheimer's disease pathogenesis: a review. *Neurochem Res* 25:1161-1172.

Oakley BR, Kirsch DR, Morris NR. 1980. A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal Biochem* 105:361-363.

Orlowski M. 1990. The multicatalytic proteinase complex, a major extralysosomal proteolytic system. *Biochemistry* 29:10289-10297.

Palade G. 1975. Intracellular aspects of the process of protein synthesis. *Science* 189:347-358.

Pan T, Groger H, Schmid V, Spring J. 1998. A toxin homology domain in an astacin-like metalloproteinase of the jellyfish *Podocoryne carnea* with a dual role in digestion and development. *Dev Genes Evol* 208:259-266.

Pathy L. 1990. Evolution of blood coagulation and fibrinolysis. *Blood Coagul Fibrinolysis* 1:153-166.

Pearson WR. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 132:185-219.

Pfleiderer G, Zwilling R, Sonneborn HH. 1967. [On the evolution of endopeptidases, 3. A protease of molecular weight 11,000 and a trypsin-like fraction from *Astacus fluviatilis* Fabr] Eine Protease vom Molekulargewicht 11000 und eine trypsinähnliche Fraktion aus *Astacus fluviatilis* Fabr. *Hoppe Seylers Z Physiol Chem* 348:1319-1331.

Piccolo S, Agius E, Lu B, Goodman S, Dale L, De Robertis EM. 1997. Cleavage of Chordin by Xolloid metalloprotease suggests a role for proteolytic processing in the regulation of Spemann organizer activity. *Cell* 91:407-416.

Plickert G, Gajewski M, Gehrke G, Gausepohl H, Schlossherr J, Ibrahim H. 1997. Automated in situ detection (AISD) of biomolecules. *Development Genes and Evolution* 207:362-367.

Rawlings ND, Barrett AJ. 1993. Evolutionary families of peptidases. *Biochem J* 290:-18.

Rawlings ND, Barrett AJ. 1995. Evolutionary families of metallopeptidases. *Methods Enzymol* 248:183-228.

Reeck GR, de Haen C, Teller DC, Doolittle RF, Fitch WM, Dickerson RE, Chambon P, McLachlan AD, Margoliash E, Jukes TH, et al. 1987. "Homology" in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell* 50:667.

- Reyda S, Jacob E, Zwillig R, Stöcker W. 1999. cDNA cloning, bacterial expression, in vitro renaturation and affinity purification of the zinc endopeptidase astacin. *Biochem J* 344 Pt 3:851-857.
- Reynolds SD, Zhang D, Puzas JE, O'Keefe RJ, Rosier RN, Reynolds PR. 2000. Cloning of the chick BMP1/Tolloid cDNA and expression in skeletal tissues. *Gene* 248:233-243.
- Rock KL, Goldberg AL. 1999. Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu Rev Immunol* 17:739-79:79.
- Sambrook J, Russell DW. 2001. *Molecular Cloning, A laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.
- Sato SM, Sargent TD. 1990. Molecular approach to dorsoanterior development in *Xenopus laevis*. *Dev Biol* 137:135-141.
- Schlangenhauf E, Etges R, Metcalf P. 1998. The crystal structure of the *Leishmania major* surface proteinase leishmanolysin (gp63). *Structure* 6:1035-1046.
- Schnölzer M, Alewood P, Jones A, Alewood D, Kent SB. 1992. In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* 40:180-193.
- Schrimpf SP, Langen H, Gomes AV, Wahlestedt C. 2001. A two-dimensional protein map of *Caenorhabditis elegans*. *Electrophoresis* 22:1224-1232.
- Schweitz H, Bruhn T, Guillemare E, Moinier D, Lancelin JM, Beress L, Lazdunski M. 1995. Kalicludines and kaliseptine. Two different classes of sea anemone toxins for voltage sensitive K⁺ channels. *J Biol Chem* 270:25121-25126.
- Scott IC, Blitz IL, Pappano WN, Imamura Y, Clark TG, Steiglitz BM, Thomas CL, Maas SA, Takahara K, Cho KW, Greenspan DS. 1999. Mammalian BMP-1/Tolloid-related metalloproteinases, including novel family member mammalian Tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. *Dev Biol* 213:283-300.
- Shelness GS, Lin L, Nicchitta CV. 1993. Membrane topology and biogenesis of eukaryotic signal peptidase. *J Biol Chem* 268:5201-5208.
- Shibata Y, Iwamatsu T, Oba Y, Kobayashi D, Tanaka M, Nagahama Y, Suzuki N, Yoshikuni M. 2000. Identification and cDNA cloning of alveolin, an extracellular metalloproteinase, which induces chorion hardening of medaka (*Oryzias latipes*) eggs upon fertilization. *J Biol Chem* 275:8349-8354.
- Sieron AL, Tretiakova A, Jameson BA, Segall ML, Lund KS, Khan MT, Li S, Stöcker W. 2000. Structure and function of procollagen C-proteinase (mTolloid) domains determined by protease digestion, circular dichroism, binding to procollagen type I, and computer modeling. *Biochemistry* 39:3231-3239.
- Sonneborn HH, Zwillig R, Pfliegerer G. 1969. Zur Evolution der Endopeptidasen, X. Die Spaltungsspezifität der niedermolekularen Protease aus *Astacus leptodactylus* Esch. *Hoppe Seylers Z Physiol Chem* 350:1097-1102.

Steiner DF, Kemmler W, Tager HS, Rubenstein AH. 1974. Molecular events taking place during intracellular transport of exportable proteins. The conversion of peptide hormone precursors. *Adv Cytopharmacol* 2195-205:-205.

Steiner DF, Rouille Y, Gong Q, Martin S, Carroll R, Chan SJ. 1996. The role of prohormone convertases in insulin biosynthesis: evidence for inherited defects in their action in man and experimental animals. *Diabetes Metab* 22:94-104.

Sternlicht MD, Werb Z. 2001. How matrix metalloproteinases regulate cell behavior. *Annu Rev Cell Dev Biol* 17:463-516.

Stöcker W, Gomis-Rüth F, Bode W, Zwilling R. 1993. Implications of the three-dimensional structure of astacin for the structure and function of the astacin family of zinc-endopeptidases. *Eur J Biochem* 214:215-231.

Stöcker W, Grams F, Baumann U, Reinemer P, Gomis-Rüth F, McKay DB, Bode W. 1995. The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci* 4:823-840.

Stöcker W, Ng M, Auld DS. 1990. Fluorescent oligopeptide substrates for kinetic characterization of the specificity of *Astacus* protease. *Biochemistry* 29:10418-10425.

Stöcker W, Sauer B, Zwilling R. 1991. Kinetics of nitroanilide cleavage by astacin. *Biol Chem Hoppe Seyler* 372:385-392.

Stöcker W, Zwilling R. 1995. Astacin. *Methods Enzymol* 248:305-325.

Stroud RM, Kossiakoff AA, Chambers JL. 1977. Mechanisms of zymogen activation. *Annu Rev Biophys Bioeng* 6177-93:-93.

Tarentino AL, Quinones G, Grimwood BG, Hauer CR, Plummer-TH J. 1995. Molecular cloning and sequence analysis of flavastacin: an O-glycosylated prokaryotic zinc metalloendopeptidase. *Arch Biochem Biophys* 319:281-285.

Tesar U. 1998. Untersuchung von Proastacin im Hepatopankreas des Flußkrebsses *Astacus astacus*. Staatsexamensarbeit, Universität Heidelberg

The *C.elegans* Sequencing Consortium. 1998. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* 282:2012-2018.

Thompson RJ. 1982. α_1 -Antitrypsin Deficiency - The European Sickle Cell Anemia? *Trends Biochem Sci* 7:349-350.

Titani K, Torff HJ, Hormel S, Kumar S, Walsh KA, Rodl J, Neurath H, Zwilling R. 1987. Amino acid sequence of a unique protease from the crayfish *Astacus fluviatilis*. *Biochemistry* 26:222-226.

Vogt G, Stöcker W, Storch V, Zwilling R. 1989. Biosynthesis of *Astacus* protease, a digestive enzyme from crayfish. *Histochemistry* 91:373-381.

Vogt G. 1996. Morphology and physiology of digestive epithelia in Decapod crustaceans. *Pflügers Arch* 431:R239-R240.

- Vogt G. 2002. Functional Anatomy. In: Holdich DM, editor. Biology of the Freshwater Crayfish. Ames: Blackwell Science.p 53-151.
- Wagner W, Möhrle F, Schnetter W. 2002. Characterization of the proteolytic enzymes in the midgut of the European Cockchafer, *Melolontha melolontha* (Coleoptera: Scarabaeidae). Journal of insect molecular biology and biochemistry.
- Wardle FC, Angerer LM, Angerer RC, Dale L. 1999. Regulation of BMP signaling by the BMP1/TLD-related metalloprotease, SpAN. Dev Biol 206:63-72.
- Weinmaster G. 1998. Reprolysins and astacins...alive, alive-o. Science 279:336-337.
- Wellington CL, Hayden MR. 2000. Caspases and neurodegeneration: on the cutting edge of new therapeutic approaches. Clin Genet 57:1-10.
- Wolz RL, Zeggaf C, Stocker W, Zwilling R. 1990. Thiol containing compounds and amino acid hydroxamates as reversible synthetic inhibitors of *Astacus* protease. Arch Biochem Biophys 281:275-281.
- Wood WB. 1988. The Nematode *Caenorhabditis elegans*. New York: Cold Spring Harbor.
- Wozney JM, Rosen V, Celeste AJ, Mitsock LM, Whitters MJ, Kriz RW, Hewick RM, Wang EA. 1988. Novel regulators of bone formation: molecular clones and activities. Science 242:1528-1534.
- Yan L, Fei K, Zhang J, Dexter S, Sarras-MP J. 2000a. Identification and characterization of hydra metalloproteinase 2 (HMP2): a meprin-like astacin metalloproteinase that functions in foot morphogenesis. Development 127:129-141.
- Yan L, Leontovich A, Fei K, Sarras-MP J. 2000b. Hydra metalloproteinase 1: a secreted astacin metalloproteinase whose apical axis expression is differentially regulated during head regeneration. Dev Biol 219:115-128.
- Yan L, Pollock GH, Nagase H, Sarras-MP J. 1995. A 25.7 x 10(3) M(r) hydra metalloproteinase (HMP1), a member of the astacin family, localizes to the extracellular matrix of *Hydra vulgaris* in a head-specific manner and has a developmental function. Development 121:1591-1602.
- Yiallourous I, Grosse-Berkhoff E, Stöcker W. 2000. The roles of Glu93 and Tyr149 in astacin-like zinc peptidases. FEBS Lett 484:224-228.
- Yiallourous I, Vassiliou S, Yiotakis A, Zwilling R, Stöcker W, Dive V. 1998. Phosphinic peptides, the first potent inhibitors of astacin, behave as extremely slow-binding inhibitors. Biochem J 331:375-379.
- York IA, Goldberg AL, Mo XY, Rock KL. 1999. Proteolysis and class I major histocompatibility complex antigen presentation. Immunol Rev 17249-66:-66.
- Zwilling, R. and Neurath, H. Invertebrate Proteases. 80, 633-664. 1981. Academic Press, Inc. Methods in Enzymology. Ref Type: Report
- Zwilling R, Stöcker W. 1997. The Astacins: structure and function of a new protein family. Verlag Dr. Kovac, Hamburg.

VII. ANHANG

1. Abkürzungen

2D	zweidimensional
APS	Amoniumpersulfat
as	<i>antisense</i>
ATP	Adenosintriphosphat
bp	<i>basepairs</i>
BSA	<i>bovine serum albumine</i>
Ci	Curie
Da	Dalton
DEPC	Diethylpyrocarbonat
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
Dns	Dansyl
ds	<i>double strand</i>
EDTA	Ethylendiamintetraacetat
EST	<i>expressed sequence tag</i>
GFP	<i>green fluorescent protein</i>
GSC	<i>geneome sequencing center</i>
HPLC	<i>high performance liquid chromatography</i>
IEP	Isoelektrischer Punkt
IPG	<i>Immobiline polyacrylamide gel</i>
IPTG	Isopropylthio- β -D-Galaktose
kb	<i>kilobases</i>
MALDI-TOF MS	<i>matrix-assisted laser desorption/ionisation time-of-flight</i>
	Massenspektrometrie
M_r	Molekulargewicht
OD	optische Dichte
ORF	<i>open reading frame</i>
PAGE	Polyacrylamid-Gelelektrophorese
PCR	<i>polymerase chain reaction</i>
RACE	<i>rapid amplification of cDNA ends</i>
RNA	Ribonukleinsäure
RNAi	<i>RNA interference</i>
RP	<i>reversed phase</i>
rpm	<i>rounds per minute</i>
RT	Reverse Transkription
s	<i>sense</i>
SDS	Sodiumdodecylsulfat
TEMED	N',N',N',N'-Tetramethylethylendiamin
Tris	Tris-(hydroxymethyl)-aminomethan
UTP	Uraciltriphosphat
UV	ultraviolett
X	Beliebige Aminosäure

2. Aminosäuren und Nukleotide

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

A	Adenosin
C	Cytosin
G	Guanosin
T	Thymidin
U	Uracil
dATP	Desoxyadenosintriphosphat
dCTP	Desoxycytosintriphosphat
dGTP	Desoxyguanosintriphosphat
dTTP	Desoxythimidintriphosphat

3. Abbildungsverzeichnis

Abbildung 1:	Ethidiumbromid gefärbte Agarosegele mit DNA-Fragmenten aus den RT-PCR Experimenten (<i>C. elegans</i>).	52
Abbildung 2:	Vektor mit kloniertem GFP und C05D11.	57
Abbildung 3:	Vektor mit kloniertem GFP und C07D10.	58
Abbildung 4:	Vektor mit kloniertem GFP und K04E7.	58
Abbildung 5:	Vektor mit kloniertem GFP und T04G9.	59
Abbildung 6:	Expression der GFP-Fusionsproteine C05D11:GFP und GFP:C07D10 in <i>C. elegans</i> .	60
Abbildung 7:	Lokalisation Astacin-homologer Gene auf der Expressionskarte (<i>SK-map</i>) von Kim et al. (2001).	63
Abbildung 8:	Simuliertes 2D-Experiment mit den Astacin-homologen Proteinen von <i>C. elegans</i> .	66
Abbildung 9:	2D-Immunoblot mit dem Anti-C07D10.6 und Anti-K04E7.3 Antiserum.	67
Abbildung 10:	Silber-gefärbtes 2D-Referenzgel von <i>C. elegans</i> Gesamtproteinextrakt.	68
Abbildung 11:	Coomassie-gefärbtes 2D-Gel von <i>C. elegans</i> Gesamtproteinextrakt.	69
Abbildung 12:	Aktivitätsmessung der durch Affinitätschromatographie mit dem Liganden Pro-Leu-Gly-NHOH aufgereinigten Enzymfraktionen.	71
Abbildung 13:	Coomassie-gefärbtes 2D-Gel der mit Pro-Leu-Gly-Hydroxamat affinitätsgereinigten Proteinen.	73
Abbildung 14:	Ethidiumbromid gefärbtes Agarosegel mit DNA-Fragmenten aus den RACE/PCR Experimenten (<i>Hydractinia echinata</i>).	74
Abbildung 15:	Nukleinsäuresequenz des Klons 1 (HEA1-VI-Gen3).	75
Abbildung 16:	Nukleinsäuresequenz des Klons 2 (HEA1-R1/F1-bs).	75
Abbildung 17:	Nukleinsäuresequenz des Klons 3 (HEA1-R6-5'RACE).	76
Abbildung 18:	Nukleinsäuresequenz des Klons 4 (HEA1-F6-3'RACE).	76
Abbildung 19:	Nukleinsäuresequenz des Klons 5 (HEA2-GP-52).	76
Abbildung 20:	Nukleinsäuresequenz des Klons 6 (HEA2-R7-5'RACE-7.5).	76
Abbildung 21:	Nukleinsäuresequenz des Klons 7 (HEA2-F7-3'RACE-2.1).	77
Abbildung 22:	DNA- und abgeleitete Proteinsequenz von HEA-1.	77
Abbildung 23:	DNA- und abgeleitete Proteinsequenz von HEA-2.	78
Abbildung 24:	Domänenstruktur Astacin-homologer Proteine bei Cnidaria.	79
Abbildung 25:	<i>in situ</i> Hybridisierung von HEA-2 und HEA-1.	80
Abbildung 26:	Vergleich der Domänenstruktur Astacin-homologer Proteine (ohne <i>C. elegans</i> und <i>Drosophila</i> Astacine).	88
Abbildung 27:	Domänenstruktur Astacin-homologer Proteine bei <i>C. elegans</i> .	89
Abbildung 28:	Domänenstruktur Astacin-homologer Proteine bei <i>Drosophila</i> .	90
Abbildung 29:	Sequenzvergleich der Astacin-Protein-Familie.	92
Abbildung 30:	Stammbaum der Astacin-Protein-Familie, erstellt nach der <i>Neighbor-joining</i> Methode (ohne <i>C. elegans</i> und <i>Drosophila</i> Astacine).	95

Abbildung 31: Stammbaum der Astacin-Protein-Familie erstellt nach der <i>Neighbor-joining</i> Methode (mit allen <i>C. elegans</i> und <i>Drosophila</i> Astacinen).	96
Abbildung 32: Schematische Darstellung des Verdauungstraktes von <i>Astacus astacus</i> .	99
Abbildung 33: Aminosäure Sequenz des Pre-Pro-Peptides und N-terminale Teil-Sequenz von Astacin.	100
Abbildung 34: Immunhistochemische Lokalisation von Astacin und Pro-Astacin im Hepatopankreas von <i>Astacus astacus</i> .	102
Abbildung 35: Nachweis von Astacin und Pro-Astacin im Magensaft und Hepatopankreas Homogenaten durch <i>Immunoblotting</i> .	103
Abbildung 36: Spaltungsspezifität von Trypsin und Astacin.	104
Abbildung 37: Aktivierungsstelle der Astacin-Protein-Familie (ohne <i>C. elegans</i> und <i>Drosophila</i> Astacine).	105
Abbildung 38: Potentielle Aktivierungsstellen bei den Astacinen von <i>C. elegans</i> und <i>Drosophila melanogaster</i> .	106
Abbildung 39: Identifizierung der Peptid Spaltprodukte durch MALDI-TOF Massenspektrometrie.	108

4. Tabellenverzeichnis

Tabelle 1: 152 Gene für Metalloproteasen in <i>C. elegans</i> .	49
Tabelle 2: 40 Astacin-homologe Proteine in <i>C. elegans</i> .	50
Tabelle 3: Transkriptom-Analyse der Astacin-homologen Gene.	53
Tabelle 4: Bezeichnung der EST und ORFeome Klone.	54
Tabelle 5: Astacin-homologe Gene in <i>C. elegans</i> .	56
Tabelle 6: Koexpressiongruppen Astacin-homologer Gene.	63
Tabelle 7: Auswertung der RNAi, <i>microarray</i> und GFP Projekte.	64
Tabelle 8: Theoretischer isoelektrischer Punkt (IP) und Molekulargewicht (MW) aller Astacin-homologer Proteine in <i>C. elegans</i> .	66
Tabelle 9: Übersicht der nachgewiesenen Proteine im <i>C. elegans</i> Gesamtproteinextrakt.	70
Tabelle 10: Fluoreszenzintensitäten der vereinigten Fraktionen.	72
Tabelle 11: Übersicht der mit PLG-NHOH affinitätsgereinigten Proteine.	73
Tabelle 12: Astacin-homologe Proteine.	83
Tabelle 13: GSC, EST und Edman-sequenzierte Fragmente.	86

5. Eigene Publikationen und Kongressberichte

Möhrlen F, Baus S, Gruber A, Rackwitz H-R, Schnölzer M, Vogt G, Zwilling R. 2001. Activation of pro-astacin: Immunological and model peptide studies on the processing of immature astacin, a zinc-endopeptidase from the crayfish *Astacus astacus*. Eur. J. Biochem. 268: 2540-2546

Wagner W, Möhrlen F, Schnetter W. 2002. Characterization of the proteolytic enzymes in the midgut of the European Cockchafer, *Melolontha melolontha* (Coleoptera: Scarabaeidae). Insect Biochem Mol Biol. 32

Möhrlen F, Zwilling R. 2002. Astacin-like proteins in the Nematode *Caenorhabditis elegans*. in Vorbereitung

Bond JS, Möhrlen F, Stöcker W. 2002. Miscellaneous astacin homologs. In: Barrett AJ, Rawlings ND, Woessner JF. Handbook of Proteolytic Enzymes. Academic Press, London and San Diego. Neue Auflage Ende 2002

Möhrlen F, Zwilling R. 2001. Activation of pro-astacins. International meeting: Metzincins at the Cell Surface and the Extracellular Matrix, Ascona

Möhrlen F, Rebscher N, Gajewski M, Maniura M, Plickert G, Zwilling R. 2001. Astacin-like metalloproteinases are involved in *Hydractinia* development. International meeting: Hydra and the Evolution of Metazoan Development, Tutzing

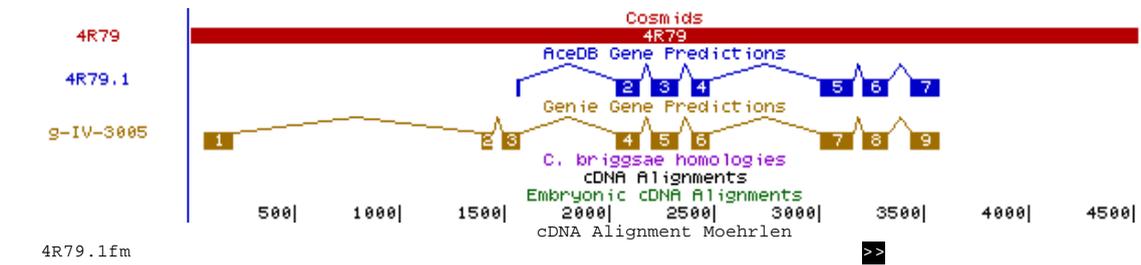
Möhrlen F, Frank U. 2002 Functional studies (RNAi) on *Hydractinia echinata* astacin (HEA) 1 and 2. International meeting: 1st Hydractinia-meeting, Schmitten

6. Sequenzdaten

6.1. Sequenzdaten der Astacin-homologen Gene von *C. elegans*

Die Graphiken zeigen die von *wormpep* oder *Genie Gene* (vergleiche IV.1.2.2.) vorhergesagte genomische Organisation der Astacin-homologen Gene und die Abschnitte, von denen experimentell ermittelte cDNA-Sequenzen vorliegen (modifiziert nach Kent und Zahler, <http://www.soe.ucsc.edu/~kent/intronerator>). Die *alignments* zeigen in Ausschnitten die Sequenzen der *wormpep* oder *Genie Gene* Vorhersagen im Vergleich zu allen bekannten cDNA-Sequenzen. Die in vorliegender Arbeit für die Transkriptomanalyse verwendeten Oligonukleotide (vergleiche IV.1.2.) sind jeweils unterstrichen.

6.1.1. Sequenzdaten 4R79.1



```

4R79.1fm
4R79.1 : 1930bp // acctgcaatataagcgaaatTTTgaaccccatgTTTTccagCACCGAATGAAATAAAAAATTTGGAAAA : 501
g-IV-3005 : 1930bp // acctgcaatataagcgaaatTTTgaaccccatgTTTTccagCACCGAATGAAATAAAAAATTTGGAAAA : 2000

4R79.1 : AGCGTTTGACAGTTACCGACGAACAACATGCATTAGATTGAAAAACGAGAAGGTCAAACAGACTACTTAAACATTGTAA : 581
g-IV-3005 : AGCGTTTGACAGTTACCGACGAACAACATGCATTAGATTGAAAAACGAGAAGGTCAAACAGACTACTTAAACATTGTAA : 2080

4R79.1 : AAGGATATGGgtaagaatatataatcttatacaacccctactgaatTTtaattttccagGTGCTACTCTCAAGTTGGACGA : 661
g-IV-3005 : AAGGATATGGgtaagaatatataatcttatacaacccctactgaatTTtaattttccagGTGCTACTCTCAAGTTGGACGA : 2160

4R79.1 : ACTGGAGGAAAAACAGGAAATTTCTTTGGGACGTGGCTGCTTTTTTCATGAAATAATTGTACACGAACTGATGCATTCGGT : 741
g-IV-3005 : ACTGGAGGAAAAACAGGAAATTTCTTTGGGACGTGGCTGCTTTTTTCATGAAATAATTGTACACGAACTGATGCATTCGGT : 2240

4R79.1 : CGGATTTTGGCAGCAACTCGAGAGCTGgtcagttcgtgtaaaaaaacttTgtttacttaagatcttaaaactgaataat : 821
g-IV-3005 : CGGATTTTGGCAGCAACTCGAGAGCTGgtcagttcgtgtaaaaaaacttTgtttacttaagatcttaaaactgaataat : 2320

4R79.1 : acttacttttttagATCGCGATGATCACATTAAGATCAACTGGGATAATATTCTGCCTGGAATGAAGTCTCAATTCGATAA : 901
g-IV-3005 : acttacttttttagATCGCGATGATCACATTAAGATCAACTGGGATAATATTCTGCCTGGAATGAAGTCTCAATTCGATAA : 2400

4R79.1 : AATTCGGCAGTGTACAAgtgagcattataaatcacaaaaaatcatcggaactcgtgtaactcaaaaatTTTTgacggc : 981
g-IV-3005 : AATTCGGCAGTGTACAAgtgagcattataaatcacaaaaaatcatcggaactcgtgtaactcaaaaatTTTTgacggc : 2480

4R79.1 : //Intron 420 bp // caatcgaagaagctgcgTTTTgtgggaaagttaaatagtaatttctagGATCTTCAAGGAGA : 1461
g-IV-3005 : //Intron 420 bp // caatcgaagaagctgcgTTTTgtgggaaagttaaatagtaatttctagGATCTTCAAGGAGA : 2960

4R79.1 : AAACACGATTACAATCAATAATGCATATGACAGCACTGCGTTTTCAAGAAACGGACGAAACACAATAGAAACTGTAG : 1541
g-IV-3005 : AAACACGATTACAATCAATAATGCATATGACAGCACTGCGTTTTCAAGAAACGGACGAAACACAATAGAAACTGTAG : 3040

4R79.1 : AAAATGGATTACACAGGTTATTGGGACCGCTATGGACTTGTACCTCTGGATATTGTGAAAgtaataatTTTggatata : 1621
g-IV-3005 : AAAATGGATTACACAGGTTATTGGGACCGCTATGGACTTGTACCTCTGGATATTGTGAAAgtaataatTTTggatata : 3120

4R79.1 : ttttcaaaataaattttaacataaaaagtTTtagATCAACAACTGATTCGTGTAAAACCTAAGAAGAAAGAGAAAGTAAA : 1701
g-IV-3005 : ttttcaaaataaattttaacataaaaagtTTtagATCAACAACTGATTCGTGTAAAACCTAAGAAGAAAGAGAAAGTAAA : 3200
4R79.1fm : -----CAAACCTGTATTTCGTGTAAAACCTAAGAAGAAAGAGAAAGTAAA :

4R79.1 : GCCTGCAACTACCGAGGAACCTCATCAACTGATTCGGCAA GTTGTGGATAAAAAATTCGGTTCGATTCTGGAGAAAgttagt : 1781
g-IV-3005 : GCCTGCAACTACCGAGGAACCTCATCAACTGATTCGGCAA GTTGTGGATAAAAAATTCGGTTCGATTCTGGAGAAAgttagt : 3280
4R79.1fm : GCCTGCAACTACCGAGGAACCTCATCAACTGATTCGGCAA GTTGTGGATAAAAAATTCGGTTCGATTCTGGAGAAAgttagt :

4R79.1 : gTTTTgtttcatttctgtgatcgcattgaaaaatgTTtacagaaaagTTTTcaactgtaacattaaaaaaaactcccaa : 1861
g-IV-3005 : gTTTTgtttcatttctgtgatcgcattgaaaaatgTTtacagaaaagTTTTcaactgtaacattaaaaaaaactcccaa : 3360

4R79.1 : caaactttatTTTcagAATGTGTGATCATTTCGGGATTTGCCGCACCTTTGCACAATACTGTACCCGTGCTTCATTTTT : 1941
g-IV-3005 : caaactttatTTTcagAATGTGTGATCATTTCGGGATTTGCCGCACCTTTGCACAATACTGTACCCGTGCTTCATTTTT : 3440

4R79.1 : CTTTGTATGAAATCGTACTGTCCATTTACGTGTAACATTGCTCTGGTGTGCGAAAACCTAAAAAATCGGGATGA - : 2017
g-IV-3005 : CTTTGTATGAAATCGTACTGTCCATTTACGTGTAACATTGCTCTGGTGTGCGAAAACCTAAAAAATCGGGATGA - : 3516
    
```

6.1.2. Sequenzdaten C05D11.6

Cosmids C05D11

AceDB Gene Predictions (1-7)

Genie Gene Predictions (1-9)

C. briggsae homologies

cDNA Alignments

Embryonic cDNA Alignments

Scale: 500 | 1000 | 1500 | 2000 | 2500 | 3000

OST

cDNA Alignment Moehrlen

BI175418

C05D11.6fm

C05D11.6 : 490bp // tccgcatttttctaaaattatcttctattttcacaactaaattgtaccaattttattgtttttagGTAAT : 560

g-III-1042 : 490bp // tccgcatttttctaaaattatcttctattttcacaactaaattgtaccaattttattgtttttagGTAAT : 560

C05D11.6 : AGATACATCTGTACCACAAACCGAAACGACACTGAATGATGCTGATTCCATAGTGATTTACACCAGAGATACGATTGTC : 640

g-III-1042 : AGATACATCTGTACCACAAACCGAAACGACACTGAATGATGCTGATTCCATAGTGATTTACACCAGAGATACGATTGTC : 640

C05D11.6fm :TCTGTACCACAAACCGAAACGACACTGAATGATGCTGATTCCATAGTGATTTACACCAGAGATACGATTGTC : 73

C05D11.6 : AGACTCTTGtaagaactttttaaactgtttaataataaatttccaactttccagGAATAAAAGTAAAAGACGATCCAAC : 720

g-III-1042 : AGACTCTTGtaagaactttttaaactgtttaataataaatttccaactttccagGAATAAAAGTAAAAGACGATCCAAC : 720

C05D11.6fm : AGACTCTTG-----GAATAAAAGTAAAAGACGATCCAAC : 109

C05D11.6 : TTGGAAACTACAGTGAGGGAGATATTTTGTGGAAAGCCCTAAGAAATTTGTGGAGGAGAACAATAAGTTGGGACGAA : 800

g-III-1042 : TTGGAAACTACAGTGAGGGAGATATTTTGTGGAAAGCCCTAAGAAATTTGTGGAGGAGAACAATAAGTTGGGACGAA : 800

C05D11.6fm : TTGGAAACTACAGTGAGGGAGATATTTTGTGGAAAGCCCTAAGAAATTTGTGGAGGAGAACAATAAGTTGGGACGAA : 189

C05D11.6 : GCGATTAACAAATATACAGgtgtggaatgttttagttccatatttaaagtataaactttcaagttgaaaatttcaagTTT : 880

g-III-1042 : GCGATTAACAAATATACAGgtgtggaatgttttagttccatatttaaagtataaactttcaagttgaaaatttcaagTTT : 880

C05D11.6fm : GCGATTAACAAATATACAG----- : 209

C05D11.6 : TGGAGCTACATTGTGGCAAATCCTCTAGAGGTTGACAAAGTAAATTTGACATTACCGATTTTCATGTTCACTACAGTAG : 960

g-III-1042 : tggagctacattgtggcaaatcctctagaggttgacaaagtaaatgtacattaccgatttttcatgttccactacagtag : 960

C05D11.6fm : ----- : -

C05D11.6 : TATAATGTCAACTCAATTCAGACGTTGGCCAAATAACGAGATTCATACACCCTGTCTTCACAATACGGTTCCTATGCAA : 1040

g-III-1042 : tataatgtcaactcaattcagACGTTGGCCAAATAACGAGATTCATACACCCTGTCTTCACAATACGGTTCCTATGCAA : 1040

C05D11.6fm : -----ACGTTGGCCAAATAACGAGATTCATACACCCTGTCTTCACAATACGGTTCCTATGCAA : 268

C05D11.6 : GAAGTGTGATTGCCAACGCAATGAACGAATATCATACAAAACATGTGTCAAATTCGTTGCCAGAGATCCATCGAAGCAT : 1120

g-III-1042 : GAAGTGTGATTGCCAACGCAATGAACGAATATCATACAAAACATGTGTCAAATTCGTTGCCAGAGATCCATCGAAGCAT : 1120

BI175418 : -----CAT : 3

C05D11.6fm : GAAGTGTGATTGCCAACGCAATGAACGAATATCATACAAAACATGTGTCAAATTCGTTGCCAGAGATCCATCGAAGCAT : 348

C05D11.6 : CACGATTATCTATGGATTTCATCCAGACGAAGGATGTTACAGTTTGGTTGGAAAgtagtagtaactgagagaaatgtgac : 1200

g-III-1042 : CACGATTATCTATGGATTTCATCCAGACGAAGGATGTTACAGTTTGGTTGGAAAgtagtagtaactgagagaaatgtgac : 1200

BI175418 : CACGATTATCTATGGATTTCATCCAGACGAAGGATGTTACAGTTTGGTTGGAAA----- : 56

C05D11.6fm : CACGATTATCTATGGATTTCATCCAGACGAAGGATGTTACAGTTTGGTTGGAAA----- : 401

C05D11.6 : aaactagaaaaatttagAACCGGAGGAAAGCAGCCAGTGTCTTTGGATTCTGGATGTATTCAAGTTGGAACAATTGTTCA : 1280

g-III-1042 : aaactagaaaaatttagAACCGGAGGAAAGCAGCCAGTGTCTTTGGATTCTGGATGTATTCAAGTTGGAACAATTGTTCA : 1280

BI175418 : -----AACCGGAGGAAAGCAGCCAGTGTCTTTGGATTCTGGATGTATTCAAGTTGGAACAATTGTTCA : 119

C05D11.6fm : -----AACCGGAGGAAAGCAGCCAGTGTCTTTGGATTCTGGATGTATTCAAGTTGGAACAATTGTTCA : 464

C05D11.6 : CGAGCTGATGCATGCTGTTGGGTTCTTCCATGAGCAGAGCAGGtaaaagagaatatagccatattctcacatggaattggta : 1360

g-III-1042 : CGAGCTGATGCATGCTGTTGGGTTCTTCCATGAGCAGAGCAGGtaaaagagaatatagccatattctcacatggaattggta : 1360

BI175418 : CGAGCTGATGCATGCTGTTGGGTTCTTCCATGAGCAGAGCAGG----- : 161

C05D11.6fm : CGAGCTGATGCATGCTGTTGGGTTCTTCCATGAGCAGAGCAGG----- : 506

C05D11.6 : aaaatgttgcccttgagaaatgggagagaagttcattaaaatgaacaggaaccctattaccaacttctggcagtagta : 1440

g-III-1042 : aaaatgttgcccttgagaaatgggagagaagttcattaaaatgaacaggaaccctattaccaacttctggcagtagta : 1440

C05D11.6 : aatctgggacatttgtttcagaagtgcgctaccaccaaacctttttatgcctacttgctgccccttggcagtagat : 1520

g-III-1042 : aatctgggacatttgtttcagaagtgcgctaccaccaaacctttttatgcctacttgctgccccttggcagtagat : 1520

C05D11.6 : ttgttttttctgaattgattaattcaagaatttaactgggaaattaacacaatataaattttctgtcaaaactttgaaa : 1600

g-III-1042 : ttgttttttctgaattgattaattcaagaatttaactgggaaattaacacaatataaattttctgtcaaaactttgaaa : 1600

C05D11.6 : actttgtcttgaaaaaaaaaactctctgccgaacttcggaaacttgacagaagtgcggtgtctgccttgaatatttttt : 1680

g-III-1042 : actttgtcttgaaaaaaaaaactctctgccgaacttcggaaacttgacagaagtgcggtgtctgccttgaatatttttt : 1680

C05D11.6 : ctgatttcaactcgacttccagACAAGACCGTGACAGTTATATCGACGTAGTCTGGCAGAAATGTGATGAACGGCGCTGATG : 1760

g-III-1042 : ctgatttcaactcgacttccagACAAGACCGTGACAGTTATATCGACGTAGTCTGGCAGAAATGTGATGAACGGCGCTGATG : 1760

BI175418 : -----ACAAGACCGTGACAGTTATATCGACGTAGTCTGGCAGAAATGTGATGAACGGCGCTGATG : 220

C05D11.6fm : -----ACAAGACCGTGACAGTTATATCGACGTAGTCTGGCAGAAATGTGATGAACGGCGCTGATG : 565

C05D11.6 : ATCAATTTGAGAAATACAACCTGAATGTGATTTCCCATTTGGATGAACCATA TGACTACGCTTCCATTATGCATTACGGA : 1840

g-III-1042 : ATCAATTTGAGAAATACAACCTGAATGTGATTTCCCATTTGGATGAACCATATGACTACGCTTCCATTATGCATTACGGA : 1840

BI175418 : ATCAATTTGAGAAATACAACCTGAATGTGATTTCCCATTTGGATGAACCATATGACTACGCTTCCATTATGCATTACGGA : 300

C05D11.6fm : ATCAATTTGAGAAATACAACCTGAATGTGATTTCCCATTTGGATGAACCATATGACTACGCTTCCATTATGC..... : 637

C05D11.6 : CCATACGCATTTCAGTGGATCCGGAAAGAAGACATTGGTGCCAAAGAAAGgtaagacaaactctgcagattacaagaagata : 1920

g-III-1042 : CCATACGCATTTCAGTGGATCCGGAAAGAAGACATTGGTGCCAAAGAAAGgtaagacaaactctgcagattacaagaagata : 1920

BI175418 : CCATACGCATTTCAGTGGATCCGGAAAGAAGACATTGGTGCCAAAGAAAG----- : 348

```

C05D11.6 : taatttaaatatttttagTCTGGGTCCGAACGTATGGGTGACAGCTGTCAAGTTTTCTGACATTGACGTTAGAAAAGATCAATA : 2000
g-III-1042 : taatttaaatatttttagTCTGGGTCCGAACGTATGGGTGACAGCTGTCAAGTTTTCTGACATTGACGTTAGAAAAGATCAATA : 2000
BI175418 : -----TCTGGGTCCGAACGTATGGGTGACAGCTGTCAAGTTTTCTGACATTGACGTTAGAAAAGATCAATA : 412

C05D11.6 : AGCTATACAATTGTCCAGGAGTCTCCGGTAACAATAATAAACAACAACAATCAAATCAACTCAAATTCAAATTTGTC AAT : 2080
g-III-1042 : AGCTATACAATTGTCCAGGAGTCTCCGGTAACAATAATAAACAACAACAATCAAATCAACTCAAATTCAAATTTGTC AAT : 2080
BI175418 : AGCTATACAATTGTCCAGGAGTCTCCGGTAACAATAATAAACAACAACAATCAAATCAACTCAAAT----- : 480

C05D11.6 : CATCCTCAGGTTTGA----- : 2095
g-III-1042 : CATCCTCAGGtttgagcacacctaccatcatcaatctcttttcatcttctatcactaatgcccgtagcgttaaagc : 2160

g-III-1042 : gcttttgcgtgtatccttagtttttgcgtgagcgtggtgattcaccattgtcttcgcactgcacttattacagtac : 2240
g-III-1042 : cccggtcattcatatgacggacgttactgtattttgcgaactgtactccgatgttatttggctccgctactctctgaaa : 2320
g-III-1042 : accaataagtatcgtagtgcacacacacacacacacacacacacacacacacacacacacacacacacacacacac : 2400
g-III-1042 : CCTCTATCAACATCTAACCAAGTAACTAACCAATCGTCGGATGATTTCGAATTTTGGTTAA : 2460

```

6.1.3. Sequenzdaten C07D10.4

```

C07D10.4fm
C07D10.4 : ATGCTTTTGCCATGGATCATCACTATTGTAACCTGTAATACCTGCTACATTAGGTCATAGGAATCGAGTACAAGATGATGA : 80
g-II-1703 : ATGCTTTTGCCATGGATCATCACTATTGTAACCTGTAATACCTGCTACATTAGGTCATAGGAATCGAGTACAAGATGATGA : 80

C07D10.4 : GATGCTTGTATATCAGATAGCACAGATTCATTAATTTGGAAGATTTCGAATTTGCTGATAAGTTAACAGAGgtaata : 160
g-II-1703 : GATGCTTGTATATCAGATAGCACAGATTCATTAATTTGGAAGATTTCGAATTTGCTGATAAGTTAACAGAGgtaata : 160

C07D10.4 : ttgtagtttagatggaattcaataattgtaattccagAGGAACTGTTTGGAAAACACATTCCGGTGGAAAGTGGTAAA : 240
g-II-1703 : ttgtagtttagatggaattcaataattgtaattccagAGGAACTGTTTGGAAAACACATTCCGGTGGAAAGTGGTAAA : 240
C07D10.4fm : .....CGGTGGAAAGTGGTAAA : 16

C07D10.4 : TGATTTCAAGTCAGATATTCGGTTACCAAGACGGCACAAGtaagctatttagcactattaaaaactgaacacacacacac : 320
g-II-1703 : TGATTTCAAGTCAGATATTCGGTTACCAAGACGGCACAAGtaagctatttagcactattaaaaactgaacacacacacac : 320
C07D10.4fm : TGATTTCAAGTCAGATATTCGGTTACCAAGACGGCACAAG----- : 56

C07D10.4 : gatttcttttgaatttcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac : 400
g-II-1703 : gatttcttttgaatttcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac : 400
C07D10.4fm : ----- : -

C07D10.4 : attatctacacatgtaaaattatgtgcccacacacacacacacacacacacacacacacacacacacacacacacacac : 480
g-II-1703 : attatctacacatgtaaaattatgtgcccacacacacacacacacacacacacacacacacacacacacacacacacac : 480
C07D10.4fm : ----- : -

C07D10.4 : ATGAAAATTATAGGCAAAATTATTAATTTTATTCTACACCATTTCAGAGAAATGGAGTTTCCCAGCAGCGAAACTATG : 560
g-II-1703 : atgaaaattatagggcaaaattattaattttattctacaccatttcagAGAAATGGAGTTTCCCAGCAGCGAAACTATG : 560
C07D10.4fm : -----AGAAATGGAGTTTCCCAGCAGCGAAACTATG : 88

C07D10.4 : GCCGAACGCTAGAATTCATATGCAATATCTCCACACTATTCTCCTCATGAAAGAGCACTTCTTGCAGAAAGCTGTGAAAC : 640
g-II-1703 : GCCGAACGCTAGAATTCATATGCAATATCTCCACACTATTCTCCTCATGAAAGAGCACTTCTTGCAGAAAGCTGTGAAAC : 640
C07D10.4fm : GCCGAACGCTAGAATTCATATGCAATATCTCCACACTATTCTCCTCATGAAAGAGCACTTCTTGCAGAAAGCTGTGAAAC : 168

C07D10.4 : AGgtttttaattttatataatttcattgctggcaaatgccttgtaaacacacacacacacacacacacacacacacacac : 720
g-II-1703 : AGgtttttaattttatataatttcattgctggcaaatgccttgtaaacacacacacacacacacacacacacacacacac : 720
C07D10.4fm : AG-----TACCACGAAAAAAGCTGTGATC : 191

C07D10.4 : AGATTTGTTCCACGACAGACCCGGAGAGCCAGATTATCTTTTATTGGAAAAGTTGATGGgtatgtaactaaatgttttcta : 800
g-II-1703 : AGATTTGTTCCACGACAGACCCGGAGAGCCAGATTATCTTTTATTGGAAAAGTTGATGGgtatgtaactaaatgttttcta : 800
C07D10.4fm : AGATTTGTTCCACGACAGACCCGGAGAGCCAGATTATCTTTTATTGGAAAAGTTGATGG----- : 250

C07D10.4 : ttgcatcagcgatacaatttttcagATGTTTCTCTGAAAGTTGGTAGAACATCGGGAGTTCAGGTGTTATCTCTCGATAAT : 880
g-II-1703 : ttgcatcagcgatacaatttttcagATGTTTCTCTGAAAGTTGGTAGAACATCGGGAGTTCAGGTGTTATCTCTCGATAAT : 880
C07D10.4fm : -----ATGTTTCTCTGAAAGTTGGTAGAACATCGGGAGTTCAGGTGTTATCTCTCGATAAT : 305

C07D10.4 : GGATGTATGGAATATGCAACAATTATTCACGAAATGATGATGTTGTCGGATTTTATCATGAACATGgtaagtcacat : 960
g-II-1703 : GGATGTATGGAATATGCAACAATTATTCACGAAATGATGATGTTGTCGGATTTTATCATGAACATGgtaagtcacat : 960
C07D10.4fm : GGATGTATGGAATATGCAACAATTATTCACGAAATGATGATGTTGTCGGATTTTATCATGAACATG----- : 372

C07D10.4 : aaattttaaatcaatattgcaagttattttttctagAACGATGGGATCGCGACAACCTCATTGATATTATCTGGCAAA : 1040
g-II-1703 : aaattttaaatcaatattgcaagttattttttctagAACGATGGGATCGCGACAACCTCATTGATATTATCTGGCAAA : 1040
C07D10.4fm : -----AACGATGGGATCGCGACAACCTCATTGATATTATCTGGCAAA : 414

C07D10.4 : ATATTGATCGTGGAGCTCTTGATCAATTTGGAAAAGTTGATTTATCGAAGACATCGTATTACGGTCAACCATATGATTAT : 1120
g-II-1703 : ATATTGATCGTGGAGCTCTTGATCAATTTGGAAAAGTTGATTTATCGAAGACATCGTATTACGGTCAACCATATGATTAT : 1120
C07D10.4fm : ATATTGATCGTGGAGCTCTTGATCAATTTGGAAAAGTTGATTTATCGAAGACATCGTATTACGGTCAACCATATGATTAT : 494

C07D10.4 : AAATCGATTCTCATTATGATAGTTTGGCATTCTCAAAAAACGGATTCCCAACAATGTTGCCGAAAGTCAAATgtgagtt : 1200
g-II-1703 : AAATCGATTCTCATTATGATAGTTTGGCATTCTCAAAAAACGGATTCCCAACAATGTTGCCGAAAGTCAAATgtgagtt : 1200
C07D10.4fm : AAATCGATTCTCATTATGATAGTTTGGCATTCTCAAAAAACGGATTCCCAACAATGTTGCCGAAAGTCAAAT----- : 567

```

C07D10.4 : tttattattcgaaaaattatataaaaactgtgattaaataccgtaaatcctagttccgatttagtgcgcacacgtaa : 1280
g-II-1703 : tttattattcgaaaaattatataaaaactgtgattaaataccgtaaatcctagttccgatttagtgcgcacacgtaa : 1280

C07D10.4 : cacatTTTTgttttaaaaaaaactatTTctttgtgataaTTgttggttaaaataaagaatcatcaagagTTTTcaaagt : 1360
g-II-1703 : cacatTTTTgttttaaaaaaaactatTTctttgtgataaTTgttggttaaaataaagaatcatcaagagTTTTcaaagt : 1360

C07D10.4 : ggacaatgaggtgcaagactaatatagctaatatagactagatatagactaatataagactagtacggttttattca : 1440
g-II-1703 : ggacaatgaggtgcaagactaatatagctaatatagactagatatagactaatataagactagtacggttttattca : 1440

C07D10.4 : atcaaagTTTTctgaattttaaactttcaaaaatagttaaattacagCTGCAACCATCGGAAACGCTCGTGACTTTTCTGA : 1520
g-II-1703 : atcaaagTTTTctgaattttaaactttcaaaaatagttaaattacagCTGCAACCATCGGAAACGCTCGTGACTTTTCTGA : 1520
C07D10.4fm : -----CTGCAACCATCGGAAACGCTCGTGACTTTTCTGA : 601

C07D10.4 : TGTGGATATTTCAAAAATCAACCGAATGTACAATTGCCAGTTGAAAAGAGTGTGACAGCTCCATTTGCACGAGCACGGg : 1600
g-II-1703 : TGTGGATATTTCAAAAATCAACCGAATGTACAATTGCCAGTTGAAAAGAGTGTGACAGCTCCATTTGCACGAGCACGGg : 1600
C07D10.4fm : TGTGGATATTTCAAAAATCAACCGAATGTACAATTGCCAGTTGAAAAGAGTGTGACAGCTCCATTTGCACGAGCACGG- : 680

C07D10.4 : tttgtatTTTTtcaggaatTTtattcaggaacttcattttcattttccagCACGTTCCAATATATTCACCACAATACCAT : 1680
g-II-1703 : tttgtatTTTTtcaggaatTTtattcaggaacttcattttcattttccagCACGTTCCAATATATTCACCACAATACCAT : 1680
C07D10.4fm : -----CACGTTCCAATATATTCACCACAATACCAT : 710

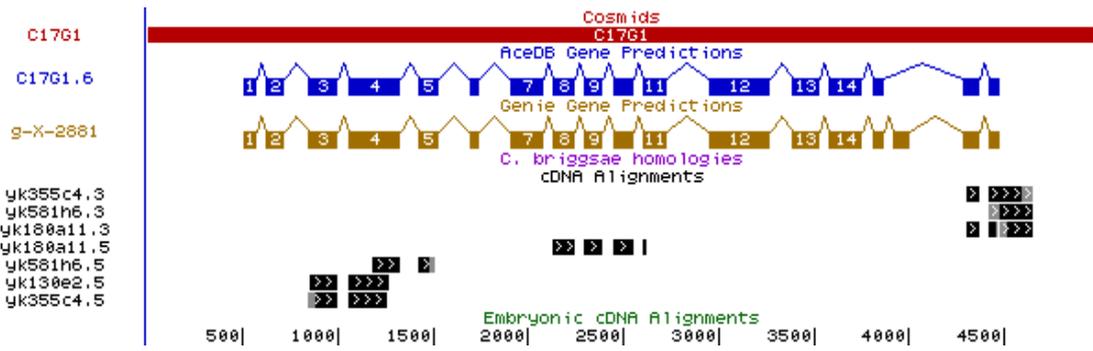
C07D10.4 : AAATATGAAGATCGTCCAAAGATTCCACTTCGATCATTGTATATGCAACAAGGTCCAATTAATCCCAATGGCACAAT : 1760
g-II-1703 : AAATATGAAGATCGTCCAAAGATTCCACTTCGATCATTGTATATGCAACAAGGTCCAATTAATCCCAATGGCACAAT : 1760
C07D10.4fm : AAATATGAAGATCGTCCAAAGATTCCACTTCGATCATTGTATATGCAACAAGGTCCAATTAATCCCAATGGCACAAT : 790

C07D10.4 : ACCATCACAAGTCTTGTGTTTCAAGTAGCTCAGGACGGGTTAATTATAATAGCAATAAACCATCATCTCGTAAGTTAT : 1840
g-II-1703 : ACCATCACAAGTCTTGTGTTTCAAGTAGCTCAGGACGGGTTAATTATAATAGCAATAAACCATCATCTCGTAAGTTAT : 1840
C07D10.4fm : ACCATCACAAGTCTTGTGTTTCAAGTAGCTCAGGACGGGTTAATTAT : 839

C07D10.4 : ATGCATTTTATGGGGAAGAAGAAATTAATTTAGAATGCGAGGATCGGATAACTGTATGCTGGTGGACAGCAGATCGTTGT : 1920
g-II-1703 : ATGCATTTTATGGGGAAGAAGAAATTAATTTAGAATGCGAGGATCGGATAACTGTATGCTGGTGGACAGCAGATCGTTGT : 1920

C07D10.4 : AGGTCTCCTGCAATTTATCAAGTCATGTCTTCATTGTGTCCAAAACATGCAAATTTTGTAG : 1983
g-II-1703 : AGGTCTCCTGCAATTTATCAAGTCATGTCTTCATTGTGTCCAAAACATGCAAATTTTGTAG : 1983

6.1.4. Sequenzdaten C17G1.6



C17G1.6 : 250bp // tgaaaattcccttattgttttccatttttgtgatcaactactgaatttatactgtattatgttattgg : 320
g-X-2881 : 250bp // tgaaaattcccttattgttttccatttttgtgatcaactactgaatttatactgtattatgttattgg : 320

C17G1.6 : tcaacagattagttttccagGATCCTGCTGCTATTCAAGCTATTGCTGAAAATATGGATAAATCAGTGAAGAATGACAAA : 400
g-X-2881 : tcaacagattagttttccagGATCCTGCTGCTATTCAAGCTATTGCTGAAAATATGGATAAATCAGTGAAGAATGACAAA : 400
YK130E2.5 : -----GCTATTCAAGCTATTGCTGAAAATATGGATAAATCAGTGAAGAATGACAAA : 51
YK355C4 : -----ATGATCCTGCTGCTATTCAAGCTATTGCTGAAAATATGGATAAATCAGTGAAGAATGACAAA : 62

C17G1.6 : ACTGAAGCAACTGTGAACAGAAAATTTGGAACGAAGTTTTTGGAAAATGCATAAATTTAACACTTCCACAGGCTGAGTC : 480
g-X-2881 : ACTGAAGCAACTGTGAACAGAAAATTTGGAACGAAGTTTTTGGAAAATGCATAAATTTAACACTTCCACAGGCTGAGTC : 480
YK130E2.5 : ACTGAAGCAACTGTGAACAGAAAATTTGGAACGAAGTTTTTGGAAAATGCATAAATTTAACACTTCCACAGGCTGAGTC : 131
YK355C4 : ACTGAAGCAACTGTGAACAGAAAATTTGGAACGAAGTTTTTGGAAAATGCATAAATTTAACACTTCCACAGGCTGAGTC : 142

C17G1.6 : ATTGTTATCAGAGTCTAgtacgttttttattccagattcataaataagttcaaaataaaataaaatcattgtcagATTACCA : 560
g-X-2881 : ATTGTTATCAGAGTCTAgtacgttttttattccagattcataaataagttcaaaataaaataaaatcattgtcagATTACCA : 560
YK130E2.5 : ATTGTTATCAGAGTCTA-----ATTACCA : 156
YK355C4 : ATTGTTATCAGAGTCTA-----ATTACCA : 167

C17G1.6 : AGATCAAGGCGACAAGCTCATCCAGATCCACGTAATTTCTGGCCAAATCTCACAATTTTCATACGAGTTCTACGGAGGAGA : 640
g-X-2881 : AGATCAAGGCGACAAGCTCATCCAGATCCACGTAATTTCTGGCCAAATCTCACAATTTTCATACGAGTTCTACGGAGGAGA : 640
YK130E2.5 : AGATCAAGGCGACAAGCTCATCCAGATCCACGTAATTTCTGGCCAAATCTCACAATTTTCATACGAGTTCTACGGAGGAGA : 236
YK355C4 : AGATCAAGGCGACAAGCTCATCCAGATCCACGTAATTTCTGGCCAAATCTCACAATTTTCATACGAGTTCTACGGAGGAGA : 247

C17G1.6 : AGAGACTTGGAGACAAGTATCAGAAGTGCTATTAGACACGTTGAGCAGAACGTTTGTCTTTAAATTCAGGAAAATGGAG : 720
g-X-2881 : AGAGACTTGGAGACAAGTATCAGAAGTGCTATTAGACACGTTGAGCAGAACGTTTGTCTTTAAATTCAGGAAAATGGAG : 720
YK130E2.5 : AGAGACTTNGAGACAAGTATCAGAAGTGCTATTAGACACGTTGAGCAGAACGTTTGTCTTTAAATTCAGGAAAATGGAG : 316
YK355C4 : AGAGACTTGGAGACAAGTATCAGAAGTGCTATTAGACACGTTGAGCAGAACGTTTGTCTTTAAATTCAGGAAAATGGAG : 327
YK581H6 : -----TCACACGTTGAGCAGAACGTTTGTCTTTAAATTCAGGAAAATGGAG : 46
YK816E12 : -----AAAAAACCAGGAAAATGGAG : 21

C17G1.6 : GAGATAGAGATGGATTGAGATATTATAGAGGAAACGGGTGTTGGTCAAACGTTGGTAGGGTTGGAGGAAGACAATGGTT : 800
g-X-2881 : GAGATAGAGATGGATTGAGATATTATAGAGGAAACGGGTGTTGGTCAAACGTTGGTAGGGTTGGAGGAAGACAATGGTT : 800
YK130E2.5 : GAGATAGAGATGGATTGAGATATTATA GAGGAAACGGNTGTTGG----- : 360
YK355C4 : GAGATAGAGATGGATTGAGATATTATA GAGGAAA----- : 361
YK581H6 : GAGATAGAGATGGATTGAGATATTATA GAGGAAACGGGTGTTGGTCAAACGTTGGTAGGGTTGGAGGAAGACAATGGTT : 126
YK816E12 : GAGATAGAGATGGATTGAGATATTATA GAGGAAACGGGTGTTGGTCAAACGTTGGTAGGGTTGGAGGAAGACAATGGTT : 101

C17G1.6 : TCTATTGGATACGGATGCGATAGTgtaagtaaatgacagttgtgtgtaaagtagttgtggcctaagtttttggatcaat : 880
g-X-2881 : TCTATTGGATACGGATGCGATAGTgtaagtaaatgacagttgtgtgtaaagtagttgtggcctaagtttttggatcaat : 880
YK581H6 : TCTATTGGATACGGATGCGATAGT----- : 150
YK816E12 : TCTATTGGATACGGATGCGATAGT----- : 125

C17G1.6 : atgataatgtgacatttttattgaaccaaatttcagCTTGGAAATAGTGTCTCATGAGACTCTGCACGCCCTTGGGACTTTG : 960
g-X-2881 : atgataatgtgacatttttattgaaccaaatttcagCTTGGAAATAGTGTCTCATGAGACTCTGCACGCCCTTGGGACTTTG : 960
YK581H6 : -----CTTGGAAATAGTGTCTCATGAGACTCTGCACGCCCTTGGGACTTTG : 194
YK816E12 : -----CTTGGAAATAGTGTCTCATGAGACTCTGCACGCCCTTGGGACTTTG : 169

C17G1.6 : GCACGAACAGTCGAGAGATGATAGAGACAATTTTCATTAGTATTGTTGCTGATAAAAATTACAAGAgtaagtttgacaatgc : 1040
g-X-2881 : GCACGAACAGTCGAGAGATGATAGAGACAATTTTCATTAGTATTGTTGCTGATAAAAATTACAAGAgtaagtttgacaatgc : 1040
YK581H6 : GCACGAACAGTCGAGAGATGATAGAGACAATTTTCATTAGTATTGTTGCTGATAAAAATTACAAG----- : 238
YK816E12 : GCACGAACAGTCGAGAGATGATAGAGACAATTTTCATTAGTATTGTTGCTGATAAAAATTACAAGA----- : 233

C17G1.6 : tagaaagattttaacaacaattttttctcaagccttgggttttcaatcctacaaaaaaaaaaaaacttttagtgatacttta : 1120
g-X-2881 : tagaaagattttaacaacaattttttctcaagccttgggttttcaatcctacaaaaaaaaaaaaacttttagtgatacttta : 1120

C17G1.6 : aacttttttgggatcaagatttttaatacaacagtaaacatttgcatttcaataaaattttcttacagGGAACAGAAGGA : 1200
g-X-2881 : aacttttttgggatcaagatttttaatacaacagtaaacatttgcatttcaataaaattttcttacagGGAACAGAAGGA : 1200
YK816E12 : -----GGAA CAGAAGGA : 245

C17G1.6 : AACTTTGCCAAACGCACAGCTGCCAACAGCGACAACCTTGtaagattttaaacactggtagacagacaaatataaatct : 1280
g-X-2881 : AACTTTGCCAAACGCACAGCTGCCAACAGCGACAACCTTGtaagattttaaacactggtagacagacaaatataaatct : 1280
YK816E12 : AACTTTGCCAAACGCACAGCTGCCAACAGCGACAACCTTG----- : 285

C17G1.6 : gaaaagataatcggttttgatggtttatggcaaacacacggctacacacactatttttcaattttgggtgtgtgt : 1360
g-X-2881 : gaaaagataatcggttttgatggtttatggcaaacacacggctacacacactatttttcaattttgggtgtgtgtgt : 1360

C17G1.6 : gttctttattgttaaaaaatttttataatagacactttttcagGACAACCGTATGATTTGGGATCGGTTATGCATT : 1440
g-X-2881 : gttctttattgttaaaaaatttttataatagacactttttcagGACAACCGTATGATTTGGGATCGGTTATGCATT : 1440
YK816E12 : -----GACAACCGTATGATTTGGGATCGGTTATGCATT : 318

C17G1.6 : ATGGAGCAAATCATTTCATATGATTGGTCTAGTGATACCATAAAAACTAGAGATTGGAGATACCAAAATACCATTGGA : 1520
g-X-2881 : ATGGAGCAAATCATTTCATATGATTGGTCTAGTGATACCATAAAAACTAGAGATTGGAGATACCAAAATACCATTGGA : 1520
YK816E12 : ATGGAGCAAATCATTTCATATGATTGGTCTAGTGATACCATAAAAACTAGAGATTGGAGATACCAAAATACCATTGGA : 398

C17G1.6 : CAAAGAGATGGGCTTTCATTTAAAGATGCCAAAATGATTAATACCAGATATTGCAGCAGtaagtgatttaatttcaaatg : 1600
g-X-2881 : CAAAGAGATGGGCTTTCATTTAAAGATGCCAAAATGATTAATACCAGATATTGCAGCAGtaagtgatttaatttcaaatg : 1600
YK816E12 : CAAAGAGATGGGCTTTCATTTAAAGATGCCAAAATGATTAATACCAGATATTGCAGCA----- : 456

C17G1.6 : aaactagatcaaaaaattttcagACGTTTGTGAGAGAACTTCCATGTTTAAATGAGGGTTACACGGATCCAAACAAC : 1680
g-X-2881 : aaactagatcaaaaaattttcagACGTTTGTGAGAGAACTTCCATGTTTAAATGAGGGTTACACGGATCCAAACAAC : 1680
YK816E12 : -----ACGTTTGTGAGAGAACTTCCATGTTTAAATGAGGGTTACACGGATCCAAACAAC : 513
YK180A11 : -----ACGTTTGTGAGAGAACTTCCATGTTTAAATGAGGGTTACACGGATCCAAACAAC : 57

C17G1.6 : GTGGACGTTGTCGTTTCCATCTGGTTATGGAGGAACATATTGTGAAACTGTAGAGTATAgtagtagttcaaaaagt : 1760
g-X-2881 : GTGGACGTTGTCGTTTCCATCTGGTTATGGAGGAACATATTGTGAAACTGTAGAGTATAgtagtagttcaaaaagt : 1760
YK816E12 : GTGGACGTTGTCGTTTCCATCTGGTTATGGAGGAACATATTGTGAAACTGTAGAGTATA----- : 573
YK180A11 : GTGGACGTTGTCGTTTCCATCTGGTTATGGAGGAACATATTGTGAAACTGTAGAGTATA----- : 117

C17G1.6 : atcgataactaatgtttttaaataatgatcagCATCTTGGGAGGATCACTTACTGCTTCCAGTTCCTACAAAAGATAGAAAG : 1840
g-X-2881 : atcgataactaatgtttttaaataatgatcagCATCTTGGGAGGATCACTTACTGCTTCCAGTTCCTACAAAAGATAGAAAG : 1840
YK816E12 : -----ATCA GCATCTTGGGAGGATCACT TAC----- : 600
YK180A11 : -----CATCTTGGGAGGATCACT TACTGCTTCCAGTTCCTACAAAAGATAGAAAG : 169

C17G1.6 : TGGTATTGTGCAACCAGATGCGAATTGTGTGGCGAATTAGAgtaagcgcacgcaagacatgtagtcaatagaccgtg : 1920
g-X-2881 : TGGTATTGTGCAACCAGATGCGAATTGTGTGGCGAATTAGAgtaagcgcacgcaagacatgtagtcaatagaccgtg : 1920
YK180A11 : TGGTATTGTGCAACCAGATGCGAATTGTGTGGCGAATTAGA----- : 212

C17G1.6 : aaattaactattgttttgcagAATCCTGGAGGAAATGTGGAAGTGTGTTGATCAGGTGAACCTTCCAATGCGCAGAT : 2000
g-X-2881 : aaattaactattgttttgcagAATCCTGGAGGAAATGTGGAAGTGTGTTGATCAGGTGAACCTTCCAATGCGCAGAT : 2000
YK180A11 : -----AATC CTGGAGGAAATGTGGAAGTGTGTTGATCAGGTGAACCTTCCAATGCGCAGAT : 269

C17G1.6 : CCATGCCAGAGTTACGTTGAAGTAAAAATATCTCAGCCAGAAGACATCGACAGGcaagttcactacttaaaatgcattcca : 2080
g-X-2881 : CCATGCCAGAGTTACGTTGAAGTAAAAATATCTCAGCCAGAAGACATCGACAGGcaagttcactacttaaaatgcattcca : 2080
YK180A11 : CCATGCCAGAGTTACGTTGAAGTAAAAATATCTCAGCCAGAAGACATCGACAG----- : 321

C17G1.6 : actcactatatttttagGCGCACGACTATGTTGTAGTTTACCGTCTGTCATTGCTCCGAAGGAGATGATGTAATTATCAT : 2160
g-X-2881 : actcactatatttttagGCGCACGACTATGTTGTAGTTTACCGTCTGTCATTGCTCCGAAGGAGATGATGTAATTATCAT : 2160
YK180A11 : -----GCGCACGACTATGTTGTAGTTTACCGTCTGTCATTGCTCCGAAGGAGATGATGTAATTATCAT : 348

C17G1.6 : // 970bp // ctggtatattctgaaataaattcagGGGAAGCAATTTGGATAGAGAAAGTTGTAACGCTCACGCCTGCG : 3120
g-X-2881 : // 970bp // ctggtatattctgaaataaattcagGGGAAGCAATTTGGATAGAGAAAGTTGTAACGCTCACGCCTGCG : 3120

C17G1.6 : CTAAGCCTAAGAAAGGAATGATCTGTAATGGAGCTCTCCTACTTCCGTGCGATTGCTTGCCAAGCTCAATTTGGATCC : 3200
g-X-2881 : CTAAGCCTAAGAAAGGAATGATCTGTAATGGAGCTCTCCTACTTCCGTGCGATTGCTTGCCAAGCTCAATTTGGATCC : 3200
YK816E12 : ---- GCCTAAGAAAGGAATGATCTGTAATGGAGCTCTCCTACTTCCGTGCGATTGCTTGCCAAGCTCAATTTGGATCC : 676

C17G1.6 : AACAACTACTTGAACCCCAAATGAAACAATCCGGATTGCGAGATgtaagttgtgtttttgtgtttttcaagggggga : 3280
g-X-2881 : AACAACTACTTGAACCCCAAATGAAACAATCCGGATTGCGAGATgtaagttgtgtttttgtgtttttcaagggggga : 3280
YK816E12 : AACAACTACTTGAACCCCAAATGAAACAATCCGGATTGCGAGAT----- : -

C17G1.6 : atatgttgaactataaacgtttttacagCCAGCAGCTTCCGCTACCACGTATCTCGCAACGTAACAGTTTTCGTTAA : 3360
g-X-2881 : atatgttgaactataaacgtttttacagCCAGCAGCTTCCGCTACCACGTATCTCGCAACGTAACAGTTTTCGTTAA : 3360
YK816E12 : -----CCAGCAGCTTCCGCTACCACGTATCTCGCAACGTAACAGTTTTCGTTAA : -

C17G1.6 : CGAgtagttcaaaaaatataattgtgaacgacggagattttttctagattggaagttccaccacaacagaagattc : 3440
g-X-2881 : CGAgtagttcaaaaaatataattgtgaacgacggagattttttctagattggaagttccaccacaacagaagattc : 3440
YK816E12 : CGA-----ATTGGAAGTTTCATCCACCAACAGAAAGATTCC : -

```

C17G1.6 : ctctcatcaagcacaagaagggtgaaaagcaaaactgccaacagattttgtgaaaagtaaggaatagtaaattgaatgat : 3520
g-X-2881 : CTCTCATCAAGCACAAGAAGGGTGAAAAGCACAACACTGCCAACAGATTTTGTGAAAAGtaaggaatagtaaattgaatgat : 3520
YK816E12 : CTCTCATCAAGCACAAGAAGGGTGAAAAGCACAACACTGCCAACAGATTTTGTGAAA----- : -

C17G1.6 : ttgaataataaataatttttacatttttggttaaaaaataatgtaataagtgaaatagtaactttaacatttgggtgaagtcac : 3600
g-X-2881 : ttgaataataaataatttttacatttttggttaaaaaataatgtaataagtgaaatagtaactttaacatttgggtgaagtcac : 3600

C17G1.6 : ttgagaaatcataatctctcagaaagttaaaaaattatcaaaaacgaaactgtgaaatcctaattatctatcttttttaaac : 3680
g-X-2881 : ttgagaaatcataatctctcagaaagttaaaaaattatcaaaaacgaaactgtgaaatcctaattatctatcttttttaaac : 3680

C17G1.6 : atcagtgTTTTTTTacaattgagaaaagtgagaaaaatttcttcaaacataaattactactccttctatgtatgtaataa : 3760
g-X-2881 : atcagtgTTTTTTTacaattgagaaaagtgagaaaaatttcttcaaacataaattactactccttctatgtatgtaataa : 3760

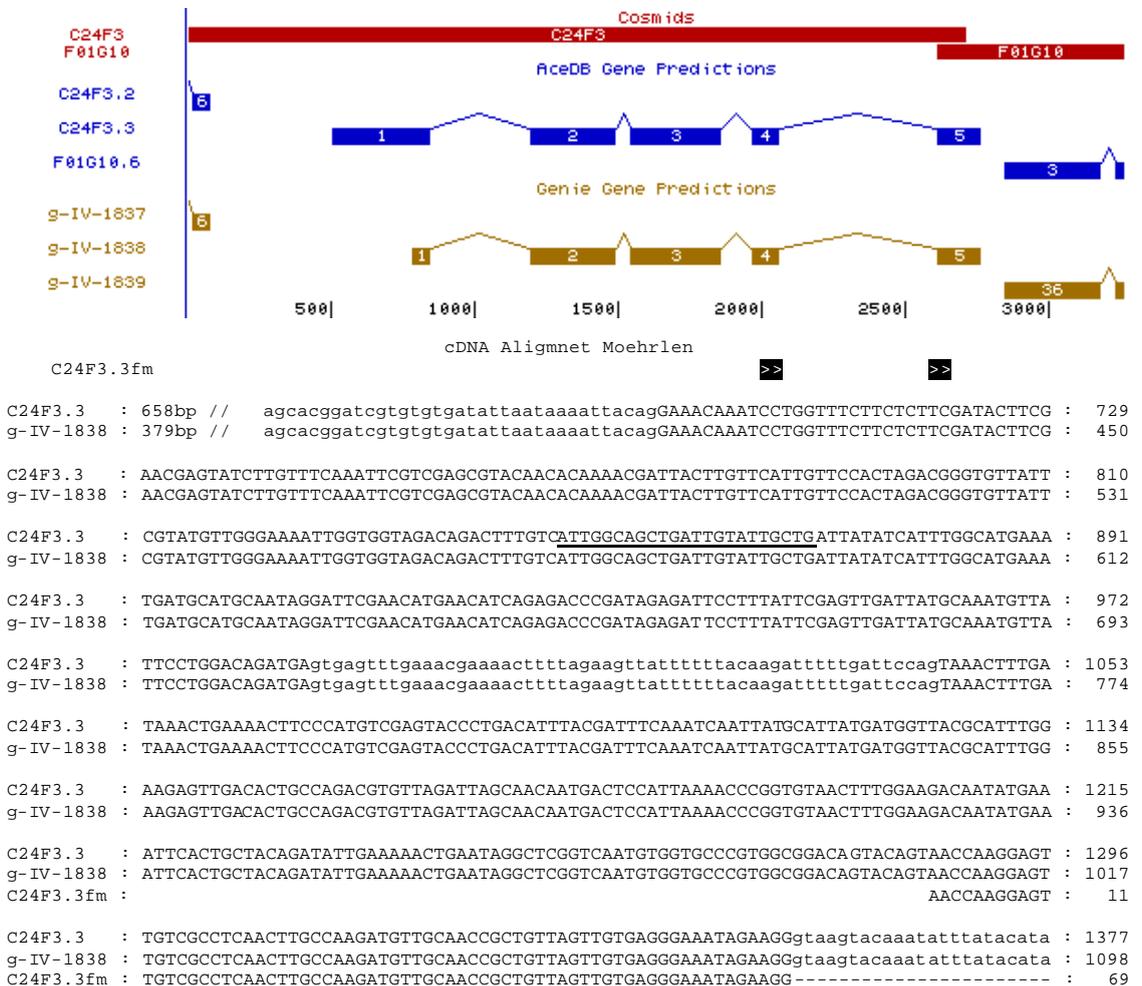
C17G1.6 : accttttttacctttttcagACGATTCATTTACCAGTGCCCAACAGCATTACTCACCATCCAAATGGAATATAAACCCAGAT : 3840
g-X-2881 : accttttttacctttttcagACGATTCATTTACCAGTGCCCAACAGCATTACTCACCATCCAAATGGAATATAAACCCAGAT : 3840
YK355C4 : -----GCATTACTCACCATCCAAATGGAATATAAACCCAGAT : -
YK816E12 : -----ACGATTCA TTTACCAGTGCCCAACAGCATTACTCACCATCCAAATGGAATATAAACCCAGAT : -
YK180A11 : -----CAACAGCATTACTCACCATCCAAATGGAATATAAACCCAGAT : -

C17G1.6 : ACTCAAGGAACCAATGATGCTTATTTTCgtaagtttaaaatgtaacgttcacatccatattaatattttcagCAACAGT : 3920
g-X-2881 : ACTCAAGGAACCAATGATGCTTATTTTCgtaagtttaaaatgtaacgttcacatccatattaatattttcagCAACAGT : 3920
YK355C4 : ACTCAAGGAACCAATGATGCTTATTTTC-----CAACAGT : -
YK581H6 : -----CAACAGT- : -
YK816E12 : ACTCAAGGAACCAATGATGCTTATTTTC-----CAACAGT : -
YK180A11 : ACTCAAGGAACCAATGATGCTTATTTTC-----CAACAGT : -

C17G1.6 : ACCCAGAATGCTGCTCTGGATATACGCCTAGAAGAGGAGTTTGCTACAAAACTGAaaccgatattttttgtaatttttt : 4003
g-X-2881 : ACCCAGAATGCTGCTCTGGATATACGCCTAGAAGAGGAGTTTGCTACAAAACTGAaaccgatattttttgtaatttttt : 4003
YK355C4 : ACCCAGAATGCTGCTCTGGATATACGCCTAGAAGAGGAGTTTGCTACAAAACTGAAACCGATATTTTGTGTAATTTTTT : -
YK581H6 : ACCCAGAATGCTGCTCTGGATATACGCCTAGAAGAGGAGTTTGCTACAAAACTGAAACCGATATTTTGTGTAATTTTTT : -
YK816E12 : ACCCAGAATGCTGCTCTGGATATACGCCTAGAAGAGGAGTTTGCTACAAAACTGAAACCGATATTTTGTGTAATTTTTT : -
YK180A11 : ACCCAGAATGCTGCTCTGGATATACGCCTAGAAGAGGAGTTTGCTACAAAACTGAAACCGATATTTTGTGTAATTTTTT : -

C17G1.6 : gtttttccttccccacacggttgatattgctcatgatctgcttctcaaaactcgctttcttttttatttttggttgataga : 4083
g-X-2881 : gtttttccttccccacacggttgatattgctcatgatctgcttctcaaaactcgctttcttttttatttttggttgataga : 4083
YK355C4 : GTTTTTCCTTCCCCACACGTTTGATATTGCTCATGATCTGCTTCTCAAACCTCGCTTCTTTTTTATTTTTGTTGATAGA : -
YK581H6 : GTTTTTCCTTCCCCACACGTTTGATATTGCTCATGATCTGCTTCTCAAACCTCGCTTCTTTTTTATTTTTGTTGATAGA : -
YK816E12 : GTTTTTCCTTCCCCACACGTTTGATATTGCTCATGATCTGCTTCTCAAACCTCGCTTCTTTTTTATTTTTGTTGATAGA : -
YK180A11 : GTTTTTCCTTCCCCACACGTTTGATATTGCTCATGATCTGCTTCTCAAACCTCGCTTCTTTTTTATTTTTGTTGATAGA : -
    
```

6.1.5. Sequenzdaten C24F3.3



```

C24F3.3 : taacttagctactaacaatggtgttttactactcaaatgttaattgaaagtcaccccttccatctccaaaatctcttggtt : 1458
g-IV-1838 : taacttagctactaacaatggtgttttactactcaaatgttaattgaaagtcaccccttccatctccaaaatctcttggtt : 1179
C24F3.3fm : ----- : -

C24F3.3 : acagCGTGGAAATGTGCAAAAATCCGTTCTACAAACAAATGATGATCAAGTCTTGCCAGAAAAACATGCAGACTATGTTTCGTA : 1539
g-IV-1838 : acagCGTGGAAATGTGCAAAAATCCGTTCTA CAAACAAATGATGATCAAGTCTTGCCAGAAAAACATGCAGACTATGTTTCGTA : 1260
C24F3.3fm : ---CGTGGAAATGTGCAAAAATCCGTTCTACAAACAAATGATGATCAAGTCTTGCCAGAAAAACATGCAGACTATGTTTCGTA : 146

C24F3.3 : CACAAGGATGATTGATGAAGATgtgagtttatatttcaagagtgaatattattataggggaaatgtactgaaaaacaaaat : 1620
g-IV-1838 : CACAAGGATGATTGATGAAGATgtgagtttatatttcaagagtgaatattattataggggaaatgtactgaaaaacaaaat : 1341
C24F3.3fm : CACAAGGATGATTGATGAAGAT----- : 168

C24F3.3 : ataccaaaacatcctaataagaaaaacttttaaaataaaaagattttctcaaactttaaggcaaagtttggttaaagtgcc : 1701
g-IV-1838 : ataccaaaacatcctaataagaaaaacttttaaaataaaaagattttctcaaactttaaggcaaagtttggttaaagtgcc : 1422

C24F3.3 : aattttccaaatagttcactattggtaaatgagcactgctgtaggttcgaaacctatgatattttgacattttagcaa : 1782
g-IV-1838 : aattttccaaatagttcactattggtaaatgagcactgctgtaggttcgaaacctatgatattttgacattttagcaa : 1503

C24F3.3 : ttttaaggagcgggtgtaaaacaattccctactgacaattaaacagtgaaaaataaaaattactttgaagattattcaaaact : 1863
g-IV-1838 : ttttaaggagcgggtgtaaaacaattccctactgacaattaaacagtgaaaaataaaaattactttgaagattattcaaaact : 1584

C24F3.3 : ccagttggttcacaaaattatgttttacatttctagaatttccatatttttggcgtgcacaaatcttgaagttgttcattt : 1944
g-IV-1838 : ccagttggttcacaaaattatgttttacatttctagaatttccatatttttggcgtgcacaaatcttgaagttgttcattt : 1665

C24F3.3 : tcagctttgtctctttaaaattaaaaccattattccaatcgactgtttctggtttattttgggttattttggcaataactt : 2025
g-IV-1838 : tcagctttgtctctttaaaattaaaaccattattccaatcgactgtttctggtttattttgggttattttggcaataactt : 1746

C24F3.3 : aattaattctgaagcctatcttctagtgaacatttacttaaaaatatcaaataagtctaatttttgcacatttcaatttt : 2106
g-IV-1838 : aattaattctgaagcctatcttctagtgaacatttacttaaaaatatcaaataagtctaatttttgcacatttcaatttt : 1827

C24F3.3 : cagGACGATCTAACTCCAAATACTACAGTAAATCTGTAAGTGCAGGACAAACAACCACGATGTGATATTTACTCACAC : 2187
g-IV-1838 : cagGACGATCTAACTCCAAATACTACAGTAAATCTGTAAGTGCAGGACAAACAACCACGATGTGATATTTACTCACAC : 1908
C24F3.3fm : ---GACGATCTAACTCCAAATACTACAGTAAATCTGTAAGTGCAGGACAAACA----- : 209

C24F3.3 : AATGGTTTCTGCACCTTCCATTTTACGATGATGTACGATATCAGTTATGTGCTAAAACGTGTAATTTATGTTAA : 2262
g-IV-1838 : AATGGTTTCTGCACCTTCCATTTTACGATGATGTACGATATCAGTTATGTGCTAAAACGTGTAATTTATGTTAA : 1983

```

6.1.6. Sequenzdaten C26C6.3

```

C26C6.3 : ATGAGATTATGTCATTCAATTATTTTATTTAATTCTTTAATTCTATTCTATTGCTCGAAAGCTGATGATCCTGCACTTT : 82
g-I-1289 : ATGAGATTATGTCATTCAATTATTTTATTTAATTCTTTAATTCTATTCTATTGCTCGAAAGCTGATGATCCTGCACTTT : 82
YK11F8 : ATGAGATTATGTCATTCAATTATTTTATTTAATTCTTTAATTCTATTCTATTGCTCGAAAGCTGATGATCCTGCACTTG : 82

C26C6.3 : TGGTTGCTAGTGAATTTAAGgtttgccaagatttttcagaaacatttccaaaatttcagttaaaaatttggcaatcggagga : 164
g-I-1289 : TGGTTGCTAGTGAATTTAAGgtttgccaagatttttcagaaacatttccaaaatttcagttaaaaatttggcaatcggagga : 164
YK11F8 : TGGTTGCTAGTGAATTTAAG----- : 102

C26C6.3 : ataaaaattaatttagGAGCATTTCATATGTGGAAGAAAAACAATTGGAAACTGTTGAAGAATTGCTTATTTAAAATGAAAAAAT : 656
g-I-1289 : ataaaaattaatttagGAGCATTTCATATGTGGAAGAAAAACAATTGGAAACTGTTGAAGAATTGCTTATTTAAAATGAAAAAAT : 656
YK11F8 : -----GAGCATTTCATATGT GGAAGAAAAACAATTGGAAACTGTTGAAGAATTGCTTATTTAAAATGAAAAAAT : 169

C26C6.3 : TGGCGCATTCAAGAAGTTTAAAGGGAAGAGAATTTGGTCATGATgtatgatttgcatttata //Intron 420bp // : 738
g-I-1289 : TGGCGCATTCAAGAAGTTTAAAGGGAAGAGAATTTGGTCATGATgtatgatttgcatttata //Intron 420bp // : 738
YK11F8 : TGGCGCATTCAAGAAGTTTAAAGGGAAGAGAATTTGGTCATGAT----- : 213

C26C6.3 : gtttagGCAGTTGAAGACTCAAAGAAAGAAGTGGCAATTAGTACTCAACAAGGAACAATTAATAAAAAAGTTTCACCGTTTT : 820
g-I-1289 : gtttagGCAGTTGAAGACTCAAAGAAAGAAGTGGCAATTAGTACTCAACAAGGAACAATTAATAAAAAAGTTTCACCGTTTT : 820
YK11F8 : -----GCAGTTGAAGACTCAAAGAAAGAAGTGGCAATTAGTACTCAACAAGGAACAATTAATAAAAAAGTTTCACCGTTTTG : 289

C26C6.3 : TGTTTGAAGgtatttattaagtttttgtgagatcaataatgctttcagGTGATATTTTCTATCTCGACGTCAGCAGTT : 902
g-I-1289 : TGTTTGAAGgtatttattaagtttttgtgagatcaataatgctttcagGTGATATTTTCTATCTCGACGTCAGCAGTT : 902
YK11F8 : TGTTTGAAG----- : 299

C26C6.3 : GATATTTCTGAAAGCTTTATCAAAAGATAAAAACCAACGATGAGAAGGtaattttttaaattatacattttaaagattgttc : 984
g-I-1289 : GATATTTCTGAAAGCTTTATCAAAAGATAAAAACCAACGATGAGAAGGtaattttttaaattatacattttaaagattgttc : 984

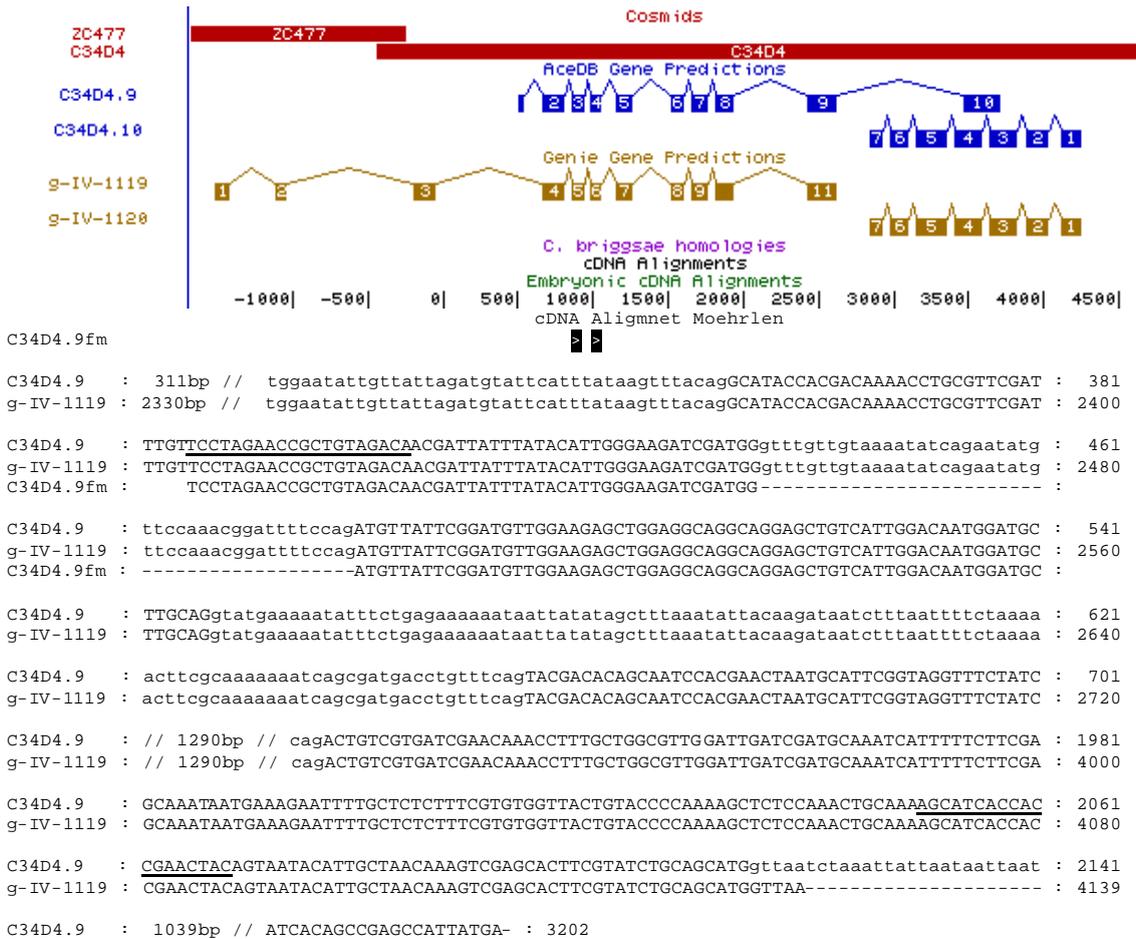
C26C6.3 : 2942bp // tgtattcggaaacattgtatcgacaatttctgctgcaacattttccaacagattttcgttactttgaattct : 3936
g-I-1289 : 2942bp // tgtattcggaaacattgtatcgacaatttctgctgcaacattttccaacagattttcgttactttgaattct : 3936

C26C6.3 : gttgtttttattttcaaaactgcagAAACGACATTGCAATATCGTCATGTGATGCTCCACAATGTTGTCTCCGTTTATTAA : 4018
g-I-1289 : gttgtttttattttcaaaactgcagAAACGACATTGCAATATCGTCATGTGATGCTCCACAATGTTGTCTCCGTTTATTAA : 4018
YK11F8.3 : -----AATGTTGTCNCCGTTTATTAA : 23

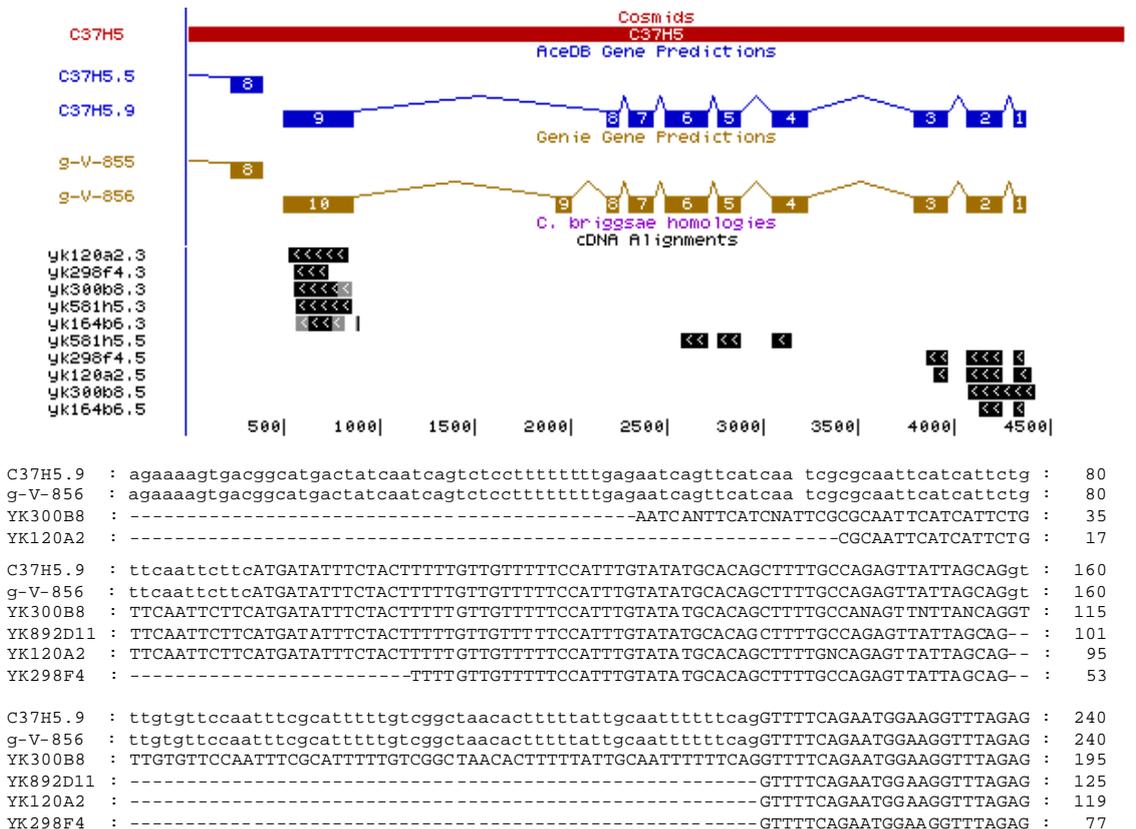
C26C6.3 : TGTGTATGGAGTATGCCAATCTGATCAAGAAAATCAACATGATGAATTATGGCTTCTATTAA --- : 4082
g-I-1289 : TGTGTATGGAGTATGCCAATCTGATCAAGAAAATCAACATGATGAATTATGGCTTCTATTAA --- : 4082
YK11F8.3 : TGTGTATGGAGTATGCCAATCTGATCAAGAAAATCAACATGATGAATTA TGGCTTCTATTAAATCA : 90

```

6.1.7. Sequenzdaten C34D4.9



6.1.8. Sequenzdaten C37H5.9



C37H5.9	: GGGGACCCGATGGTTTTAATAGAGGACCTGGAGGGTTTCATCGTGGACCTGACGGCTTCGGGGGAGACCTCGTGGCGGT	: 320
g-V-856	: GGGGACCCGATGGTTTTAATAGAGGACCTGGAGGGTTTCATCGTGGACCTGACGGCTTCGGGGGAGACCTCGTGGCGGT	: 320
YK300B8	: GGGGACCCGATGGTTTTAATAGAGGACCTGGAGGGTTTCATCGTGGACCTGACGGCTTCGGGGGATANNCTCTTGGCGGT	: 275
YK892D11	: GGGGACCCGATGGTTTTAATAGAGGACCTGGAGGGTTTCATCGTGGACCTGACGGCTTCGGGGGAGACCTCGTGGCGGT	: 205
YK120A2	: GGGGACCCGATGGTTTTAATAGAGGACCTGGAGGGTTTCATCGTGGACCTGACGGCTTCGGGGGAGACCTCGTGGNGGT	: 199
YK298F4	: GGGGACCCGATGGTTTTAATAGAGGACCTGGAGGGTTTCATCGTGGACCTGACGGCTTCGGGGGAGACCTCGTGGCGGT	: 157
C37H5.9	: GTAGACCTCGGCCACTTGATTGGGAATATGTCTGCAAAATGTTGGTCAAGAAATGGGTCTAAACGATGCTGATGTTATTGG	: 400
g-V-856	: GTAGACCTCGGCCACTTGATTGGGAATATGTCTGCAAAATGTTGGTCAAGAAATGGGTCTAAACGATGCTGATGTTATTGG	: 400
YK300B8	: GTAGACCTCGGCCACTTGATTGGGAATATGTCTGCAAAATGTTGGTCAAGAAATGGGTCTAAACGATGCTGATGTT-----	: 350
YK892D11	: GTAGACCTCGGCCACTTGATTGGGAATATGTCTGCAAAATGTTGGTCAAGAAATGGGTCTAAACGATGCTGATGTTATTGG	: 285
YK120A2	: GTAGACCTCGGCCACTTGATTGGGAATATGTCTGCAAAATGTTGGTCAAGAAATGGGTCTAAACGATGCTGATGTTATTGG	: 279
YK298F4	: GTAGACCTCGGCCACTTGATTGGGAATATGTCTGCAAAATGTTGGTCAAGAAATGGGTCTAAACGATGCTGATGTTATTGG	: 237
C37H5.9	: AGATTTAAGAgtaggttattcacagcaaaattatcagaacagtttacagagaaattcaacgtttctcacagaaaaaacta	: 480
g-V-856	: AGATTTAAGAgtaggttattcacagcaaaattatcagaacagtttacagagaaattcaacgtttctcacagaaaaaacta	: 480
YK892D11	: AGATTTAAGA-----	: 295
YK120A2	: AGATTTAAGA-----	: 289
YK298F4	: AGATTTAAGA-----	: 247
C37H5.9	: aacaattttattgaaaccttttatagGGAATCTCTCGAGGACCCCGACCATCATCCATGGAGTGGGGACGGCGAGCCCGGC	: 560
g-V-856	: aacaattttattgaaaccttttatagGGAATCTCTCGAGGACCCCGACCATCATCCATGGAGTGGGGACGGCGAGCCCGGC	: 560
YK892D11	: -----GGAATCTCTCGAGGACCCCGACCATCATCCATGGAGTGGGGACGGCGAGCCCGGC	: 350
YK120A2	: -----GGAATCTCTCGAGGACCCCGACCATCATCCATGGAGTGGGGACGGCGAGCCCGGC	: 344
YK298F4	: -----GGAATCTCTCGAGGACCCCGACCATCATCCATGGAGTGGGGACGGCGAGCCCGGC	: 302
C37H5.9	: ATTTCTGCAGGAGATATCCGGGTCATCCAAAATGTCAGCGAGGACAACCTCCACAGTTTACTGATGTTCCGACTATTATC	: 640
g-V-856	: ATTTCTGCAGGAGATATCCGGGTCATCCAAAATGTCAGCGAGGACAACCTCCACAGTTTACTGATGTTCCGACTATTATC	: 640
YK892D11	: ATTTCTGCAGGAGATATCCGGGTCATCCAAAATGTCAGCGGGGACAACCTCCACAGTTTACTGATGTTCCGACTATTATC	: 430
YK120A2	: ATTTCTGGAGGAGATA-----	: 360
YK298F4	: ATTTCTGCAGGAGATATCCGGGTCATCCAAAATGTCAGCANGACNACTCCACAGTT-----	: 360
C37H5.9	: AACACTATTATCTACAACCGGGGAGATCTTCTCCACGAGTTCCTACGtaagtttcaaatgtgaagagtattagaaaaac	: 720
g-V-856	: AACACTATTATCTACAACCGGGGAGATCTTCTCCACGAGTTCCTACGtaagtttcaaatgtgaagagtattagaaaaac	: 720
YK827H04	: AACACTATTATCTACAACCGGGGAGATCTTCTCCACGAGTTCCTAC-----	: 460
C37H5.9	: tgtctgtttgaaaatttttataacgtttaaaaaatctcctttttgaaatcgatgtgtaagttaggggttttttagtcaac	: 800
g-V-856	: tgtctgtttgaaaatttttataacgtttaaaaaatctcctttttgaaatcgatgtgtaagttaggggttttttagtcaac	: 800
YK892D11	: -----CCTAAACATCCA TGACCCGTTGGCTGGACTAAACAGCGAACTG	: 520
C37H5.9	: aaattcggaaaaatggacaatcgccggtttgcccgaattttgaaattccggcaact//Intron 335bp//	: 880
g-V-856	: aaattcggaaaaatggacaatcgccggtttgcccgaattttgaaattccggcaact//Intron 335bp//	: 880
YK892D11	: GTTGGATTTCATCAAGAGTCTTCAGTCACAATTTGGACAAC-----	: 560
C37H5.9	: cctacctactttactttctgactgatcaatattacagCCTAAACATCCATGACCCGTTGGCTGGACTAAACAGCGAACTG	: 1280
g-V-856	: cctacctactttactttctgactgatcaatattacagCCTAAACATCCATGACCCGTTGGCTGGACTAAACAGCGAACTG	: 1280
C37H5.9	: GTTGGATTTCATCAAGAGTCTTCAGTCACAATTTGGACAACCTCAGTTCCTCAACAGAGAAACGAGATTTCACGATTATCGAG	: 1360
g-V-856	: GTTGGATTTCATCAAGAGTCTTCAGTCACAATTTGGACAACCTCAGTTCCTCAACAGAGAAACGAGATTTCACGATTATCGAG	: 1360
YK581H5	: -----TCAGTTCCTCAACAGAGAAACGAGATTTCACGATTATCGAG	: 41
C37H5.9	: ATCATTCAAGTCCGATCAGCAGTCTCCTCAGAACACACAGGCCAAAGCAGGAGCTATTGACAAAGgtaattacctagtcta	: 1440
g-V-856	: ATCATTCAAGTCCGATCAGCAGTCTCCTCAGAACACACAGGCCAAAGCAGGAGCTATTGACAAAGgtaattacctagtcta	: 1440
YK581H5	: ATCATTCAAGTCCGATCAGCAGTCTCCTCAGAACACACAGGCCAAAGCAGGAGCTATTGACAAAG-----	: 105
C37H5.9	: gggcttccaccctacggtttgctgctgctgctgtaagcgtactgaagcgttaaattgctaaaaagaggcgagagggcaggcg	: 1520
g-V-856	: gggcttccaccctacggtttgctgctgctgctgtaagcgtactgaagcgttaaattgctaaaaagaggcgagagggcaggcg	: 1520
C37H5.9	: tgaggtaggcagggttttaagccgtctgcatggaaccgtaacctggcaagctttaaattcttctcagATGCTCGCCTTCGAT	: 1600
g-V-856	: tgaggtaggcagggttttaagccgtctgcatggaaccgtaacctggcaagctttaaattcttctcagATGCTCGCCTTCGAT	: 1600
YK581H5	: -----ATGCTCGCCTTCGAT	: 120
C37H5.9	: CAAGCGGTTGGCGGTAAGCTGCTCCGGCTCATGACAAGGTGAACCTCCGATTCGACAGAACCTCAGCAAGTGAACAGGC	: 1680
g-V-856	: CAAGCGGTTGGCGGTAAGCTGCTCCGGCTCATGACAAGGTGAACCTCCGATTCGACAGAACCTCAGCAAGTGAACAGGC	: 1680
YK581H5	: CAAGCGGTTGGCGGTAAGCTGCTCCGGCTCATGACAAGGTGAACCTCCGATTCGACAGAACCTCAGCAAGTGAACAGGC	: 200
C37H5.9	: ATTGCTCAAAAAGAGCAAATTTGTGCGATATTgtaagtataccttctggcagtttgaagaaaattcaagtttgcttcagAT	: 1760
g-V-856	: ATTGCTCAAAAAGAGCAAATTTGTGCGATATTgtaagtataccttctggcagtttgaagaaaattcaagtttgcttcagAT	: 1760
YK581H5	: ATTGCTCAAAAAGAGCAAATTTGTGCGATATT-----AT	: 233
C37H5.9	: CGTTCACGCTGACAATGGAGTCTTCGACAGAGATGTGTTGCTCACCAGCATCAGGCTAACTTCTGCTCAATGAGCTCG	: 1840
g-V-856	: CGTTCACGCTGACAATGGAGTCTTCGACAGAGATGTGTTGCTCACCAGCATCAGGCTAACTTCTGCTCAATGAGCTCG	: 1840
YK581H5	: CGTTCACGCTGACAATGGAGTCTTCGACAGAGATGTGTTGCTCACCAGCATCAGGCTAACTTCTGCTCAATGAGCTCG	: 313
C37H5.9	: GCGAGGCCGGGCGTGGAGCAGATGTTGGAGCAGGAGGAGTGGTGGTGGACGAGTTCGAAGATCCGGAGTTTTCTTCCAA	: 1920
g-V-856	: GCGAGGCCGGGCGTGGAGCAGATGTTGGAGCAGGAGGAGTGGTGGTGGACGAGTTCGAAGATCCGGAGTTTTCTTCCAA	: 1920
YK581H5	: GCGAGGCCGGGCGTGGAGCAGATGTTGGAGCAGGAGGAGTGGTGGTGGACGAGTTC-----	: 370
YK1010B03:	CGAGTTCGAAGATCCGGAGTTTCTTCCAA	-
C37H5.9	: GAAAGTGCGGTTTCAGAAATGGGATATTTGGAAGCCTATTTCAGTATACTTTGGATGATTATTAGgtataactctttgtca	: 2000
g-V-856	: GAAAGTGCGGTTTCAGAAATGGGATATTTGGAAGCCTATTTCAGTATACTTTGGATGATTATTAGgtataactctttgtca	: 2000
YK1010B03:	GAAAGTGCGGTTTCAGAAATGGGATATTTGGAAGCCTATTTCAGTATACTTTGGATGATTATTAG-----	-
C37H5.9	: aatttgaagaaaaaccgtcaaaatcttgacaagtttccagAAGAATCCGACAAAAAAGACATCCGGGATGCTCTACACGA	: 2080
g-V-856	: aatttgaagaaaaaccgtcaaaatcttgacaagtttccagAAGAATCCGACAAAAAAGACATCCGGGATGCTCTACACGA	: 2080
YK1010B03:	-----AAGAATCCGACAAAAAAGACATCCGGGATGCTCTACACGA	-
C37H5.9	: AATTTCATCAACACCTGCATCCTATTTCAGGTACAATGCAACGCCAAAAGGATATCACTTGAATTATATGAAAGTTGATA	: 2160
g-V-856	: AATTTCATCAACACCTGCATCCTATTTCAGGTACAATGCAACGCCAAAAGGATATCACTTGAATTATATGAAAGTTGATA	: 2160
YK1010B03:	AATTTCATCAACACCTGCATCCTATTTCAGGTACAATGCAACGCCAAAAGGATATCACTTGAATTATATGAAAGTTGATA	-
C37H5.9	: GCACGACCTTgtaagttgactagagaatctaagtgtagcgtcaactcttctcagCTGCGGCTTATCCTACGTGGGACGG	: 2240
g-V-856	: GCACGACCTTgtaagttgactagagaatctaagtgtagcgtcaactcttctcagCTGCGGCTTATCCTACGTGGGACGG	: 2240
YK164B6	: -----TCCTACGTGGGACGG	-
YK1010B03:	GCACGACCTT-----CTGCGGCTTATCCTACGTGGGACGG	-

```

C37H5.9 : ACCGATCCGGCCAAACCCGATTTACTTGAGCTTCCAGTGTGGGGATgtacggtttcttcttcatTTTTGGGGGAAAAC : 2320
g-V-856 : ACCGATCCGGCCAAACCCGATTTACTTGAGCTTCCAGTGTGGGGATgtacggtttcttcttcatTTTTGGGGGAAAAC : 2320
YK827H04 : ACCGATCCGGCCAAACCCGATTTACTTGAGCTTCCAGTGTGGGGAT----- : -
YK1010B03 : ACCGATCCGGCCAAACCCGATTTACTTGAGCTTCCAGTGTGGGGAT----- : -

C37H5.9 : gtggtgtcAAactgtccattttggtgtgatctacAAAAatgCGGagaatcttcttTgaaatcttTgcgttttttagtag : 2400
g-V-856 : gtggtgtcAAactgtccattttggtgtgatctacAAAAatgCGGagaatcttcttTgaaatcttTgcgttttttagtag : 2400

C37H5.9 : atcacgcccgaatgggacactttgacaccacgtggaagataatgcaaatTCgatcgaaaactttcagaatcgaggagt : 2480
g-V-856 : atcacgcccgaatgggacactttgacaccacgtggaagataatgcaaatTCgatcgaaaactttcagaatcgaggagt : 2480
YK827H04 : -----AATCGAGGAGTT : -

C37H5.9 : gcaatgcacgaaacaatgcacgcgttggcgctcagtcaccaacacttgagactcgacagggataagtatatacaaggtttt : 2560
g-V-856 : GCAATGCACGAAACAATGCACGCGTTGGGCGTCAGTCACCAACACTTGAGACTCGACAGGGATAAGTATATCAAGGtttt : 2560
YK827H04 : GCAATGCACGAAACAATGCACGCGTTGGGCGTCAGTCACCAACACTTGA GACTCGACAGGGATAAGTATATCAAG---- : -

C37H5.9 : // 1050 bp // aattccAAAAatTgcctctTgttcgCgacgagaccagccgatttttAAAAatTcAAAAatTtcc : 3600
g-V-856 : // 1050 bp // aattccAAAAatTgcctctTgttcgCgacgagaccagccgatttttAAAAatTcAAAAatTtcc : 3600

C37H5.9 : agATCGACTGGTCAAAACATCGACCCCCAACACTATGACACATTTGCGATATCCGACGCGAAACTGTACACATCTACGGT : 3680
g-V-856 : agATCGACTGGTCAAAACATCGACCCCCAACACTATGACACATTTGCGATATCCGACGCGAAACTGTACACATCTACGGT : 3680
YK892D11 : --ATCGACTGGTCAAAACATCGACCCCCAACACTATGACACATTTGCGATATCCGACGCGAAACTGTACACATCTACGGT : -
YK120A2 : -----TATGACACATTTGCGATATCCGACGCGAAACTGTACACATCTACGGT : -
YK581H5 : -----TCAAACATCGACCCCCAACACTATGACACATTTGCGATATCCGACGCGAAACTGTACACATCTACGGT : -

C37H5.9 : ACCAAGTACGCCTACGACTCTATAATGCACATAACGCATATCTCGGCGCCAAGGACCCCAACAAGCCGACGATGATCCC : 3760
g-V-856 : ACCAAGTACGCCTACGACTCTATAATGCACATAACGCATATCTCGGCGCCAAGGACCCCAACAAGCCGACGATGATCCC : 3760
YK892D11 : ACCAAGTACGCCTACGACTCTATAATGCACATAACGCATATCTCGGCGCCAAGGACCCCAACAAGCCGACGATGATCCC : -
YK120A2 : ACCAAGTACGCCTACGACTCTATAATGCACATAACGCATATCTCGGCGCCAAGGACCCCAACAAGCCGACGATGATCCC : -
YK298F4 : -----AACAGCCGACGATGATCCC : -
YK581H5 : ACCAAGTACGCCTACGACTCTATAATGCACATAACGCATATCTCGGCGCCAAGGACCCCAACAAGCCGACGATGATCCC : -

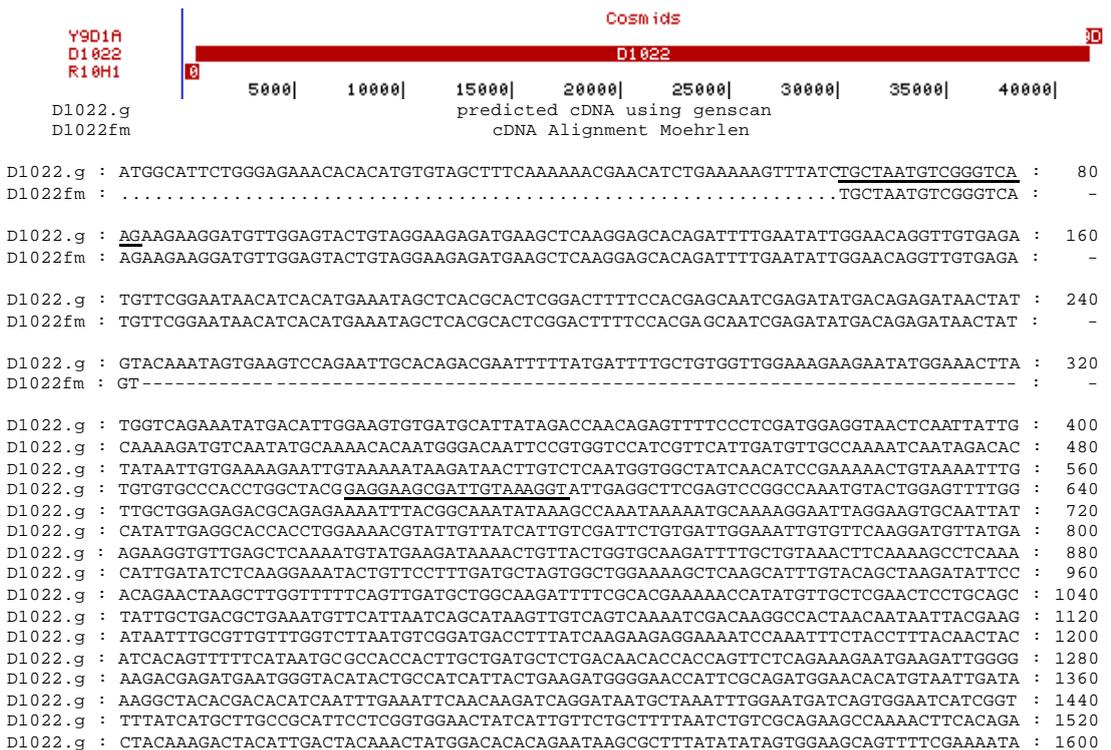
C37H5.9 : GTTGGTGAATCCACAGGAGAACACCCCGAAACTGGGTCAACGAGCCAAACTGACCCGCGCGACATCCGGTTGCTCAAAA : 3840
g-V-856 : GTTGGTGAATCCACAGGAGAACACCCCGAAACTGGGTCAACGAGCCAAACTGACCCGCGCGACATCCGGTTGCTCAAAA : 3840
YK892D11 : GTTGGTGAATCCACAGGAGAACACCCCGAAACTGGGTCAACGAGCCAAACTGACCCGCGCGACATCCGGTTGCTCAAAA : -
YK120A2 : GTTGGTGAATCCACAGGAGAACACCCCGAAACTGGGTCAACGAGCCAAACTGACCCGCGCGACATCCGGTTGCTCAAAA : -
YK164B6 : -----AA : -
YK298F4 : GTTGGTGAATCCACAGGAGAACACCCCGAAACTGGGTCAACGAGCCAAACTGACCCGCGCGACATCCGGTTGCTCAAAA : -
YK581H5 : GTTGGTGAATCCACAGGAGAACACCCCGAAACTGGGTCAACGAGCCAAACTGACCCGCGCGACATCCGGTTGCTCAAAA : -

C37H5.9 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTGCAAGATGAAG : 3920
g-V-856 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTGCAAGATGAAG : 3920
YK892D11 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTGCAAGATGAAG : -
YK120A2 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTGCAAGATGAAG : -
YK164B6 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTGCAAGATGAAG : -
YK298F4 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTGCAAGATGAAG : -
YK581H5 : AAATGTACTGCCGTCGGGATGTGACGATCAGAATGTGCACCTGCGGCACGTGGGCTCTGCATGGCTATTG----- : -

C37H5.9 : GAACAAATGAAATGGATGAATGAGAATTTGAAAGTAGTTGTGATAAATGTTGA----- : 3974
g-V-856 : GAACAAATGAAATGGATGAATGAGAATTTGAAAGTAGTTGTGATAAATGTTGA----- : 3974
YK892D11 : gaacaaa----- : -
YK120a2 : gaacanatgaaatgngtgaatgag----- : -

```

6.1.9. Sequenzdaten D1022.m

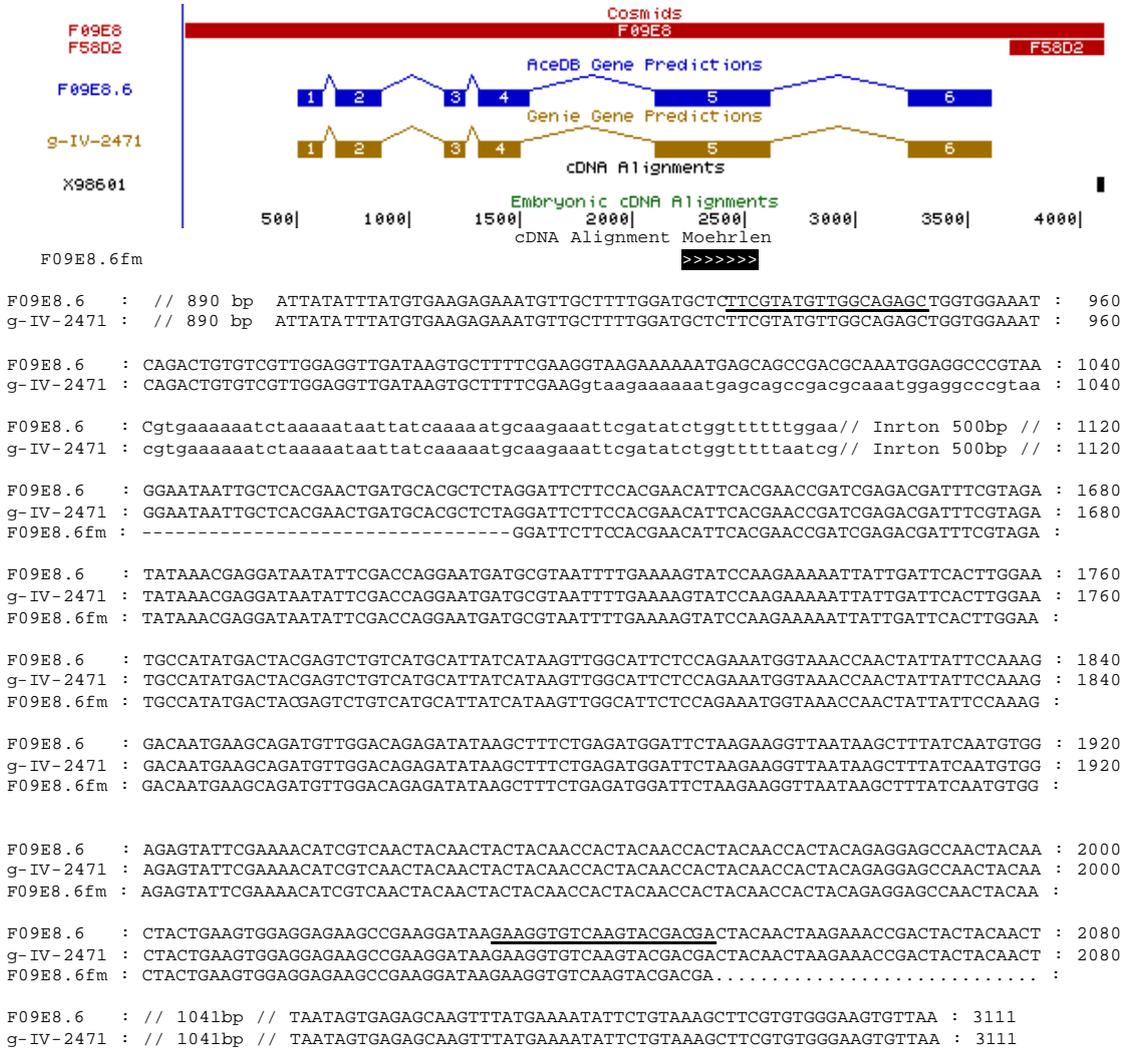


```

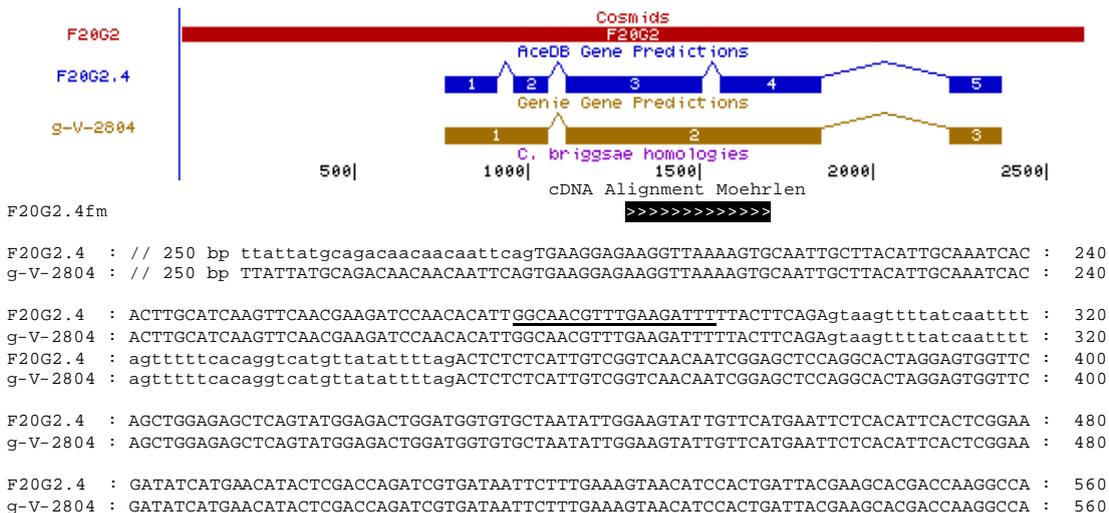
D1022.g : CCGGACAACCTCCTGTGGTCACAGTTAAAGATCTACCAAATCGGCCAACAAATCAATATAACAGGAGGACCAACAATGCCT : 1680
D1022.g : TATCGTTATAAATCCATCAGGAATTTTTAGACCACAGCAAAATAAATTTAATTCAGATAAGCGTACACTTTGGAAGAGC : 1760
D1022.g : AGATGAAGGAGAGAAAGGATCAGAGCATACTGTAGACAGAGTTCGATTTCCGGCGGAAATTCAGCTACTAGCATATAATA : 1840
D1022.g : GTGCACCTTATCCAAATTTTCAGTGTTCGAATGACATACCAAGAGGGCTTCTTGCAGTTTCTGTAATGTTGATTTAATT : 1920
D1022.g : GTTCAGCAGTCTACTGTATCTTCAAGTTGCCATCGAATGTTTACGTGGAATGGGGTACCAAGCGAATCCTGGTAGAAA : 2000
D1022.g : TAAGCGGAATGATTCGGTTTCTCGAAGATATGTTCCAAATCGGAAGTTTTCGAAATCGATAGCATTTCGTCTGATGATG : 2080
D1022.g : TTTCCAAAGCCGGTTCATTTTAG----- : 2103

```

6.1.10. Sequenzdaten F09E8.6



6.1.11. Sequenzdaten F20G2.4



```

F20G2.4 : TGGGGAATGACAACAATGTATGGACCATTGAACTGGAAGTATCATGATGTATCATTCAAGCAACTATGGTGTGGAAA : 640
g-V-2804 : TGGGGAATGACAACAATGTATGGACCATTGAACTGGAAGTATCATGATGTATCATTCAAGCAACTATGGTGTGGAAA : 640

F20G2.4 : AATGGAACCATACGATATGGAATACAAAATACAATGGGGAGTAGAAGAGTGACCTTTTATGATATGTATAAAATCAACC : 720
g-V-2804 : AATGGAACCATACGATATGGAATACAAAATACAATGGGGAGTAGAAGAGTGACCTTTTATGATATGTATAAAATCAACC : 720

F20G2.4 : AATACTACGGATGTGGGTGCTCCACTCagttggaatgcaaaaatggaggatatacttcaccatctgattgttctagATGT : 800
g-V-2804 : AATACTACGGATGTGGGTGCTCCACTCagttggaatgcaaaaatggaggatatacttcaccatctgattgttctagATGT : 800

F20G2.4 : AATTGCCCGAAAGGATTCTTTGGAAAACTTTGTAACTGAAAGACGTCAACAAGATTTCATATGAACTAAAGGCTACCTACGG : 880
g-V-2804 : AATTGCCCGAAAGGATTCTTTGGAAAACTTTGTAACTGAAAGACGTCAACAAGATTTCATATGAACTAAAGGCTACCTACGG : 880

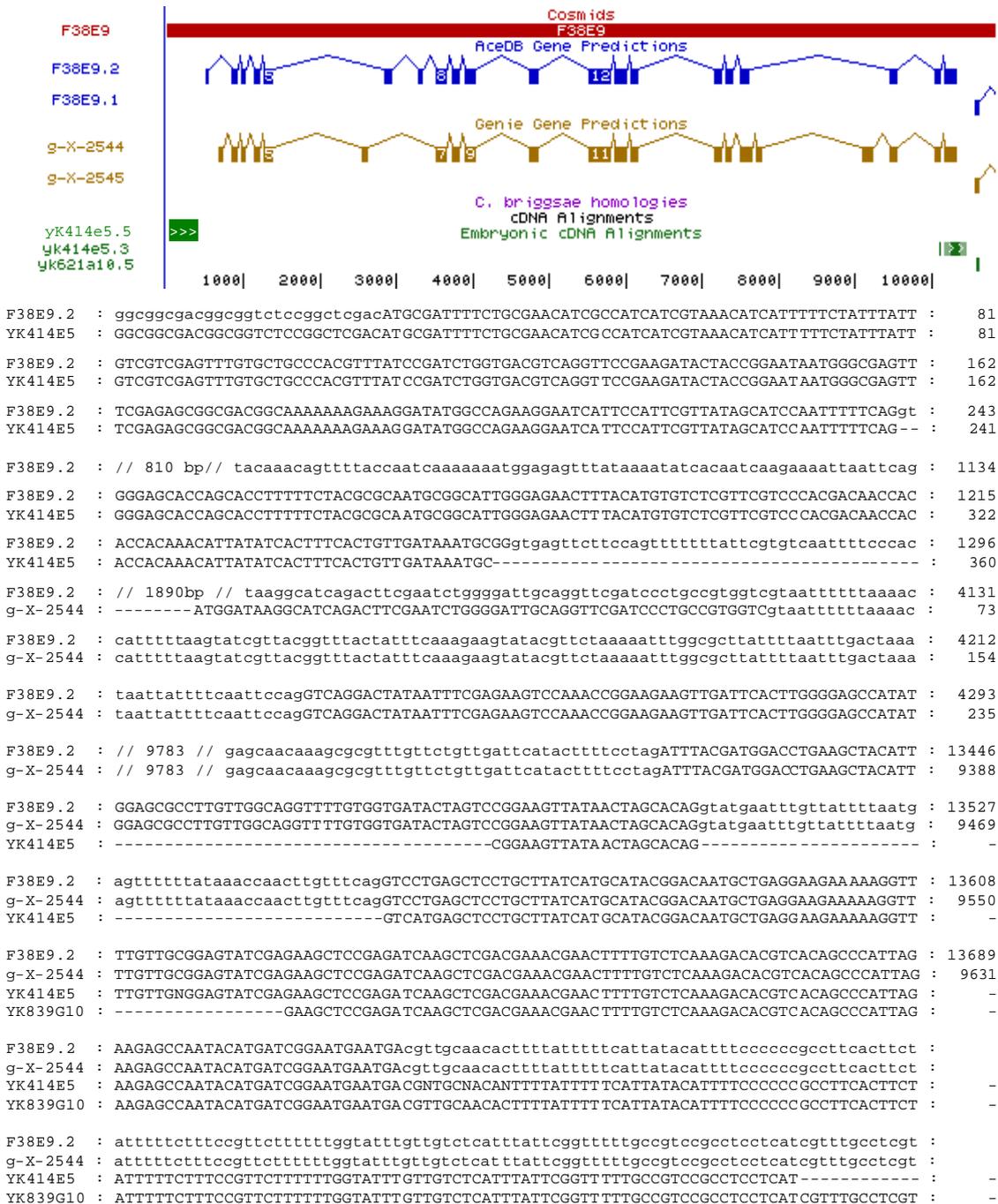
F20G2.4 : TCGATGGCAGACACAGACTATCTCATTAACTACAAGCCAGAACCAGTCTCTGATGGTTTTTATTCAACTTTTGTGTACA : 960
g-V-2804 : TCGATGGCAGACACAGACTATCTCATTAACTACAAGCCAGAACCAGTCTCTGATGGTTTTTATTCAACTTTTGTGTACA : 960
F20G2.4fm: --GATGGCAGACACAGACTATCTCATTAACTACAAGCCAGAACCAGTCTCTGATGGTTTTTATTCAACTTTTGTGTACA :

F20G2.4 : TTACTGGTGAAGCTAACTCAACAATTGAGATTACAATGGAAGGCCCTTGAAAATGTGATATGTACGGCTGGATGACTTGG : 1040
g-V-2804 : TTACTGGTGAAGCTAACTCAACAATTGAGATTACAATGGAAGGCCCTTGAAAATGTGATATGTACGGCTGGATGACTTGG : 1040
F20G2.4fm: TTACTGGTGAAGCTAACTCAACAATTGAGATTACAATGGAAGGCCCTTGAAAATGTGATATGTACGGC..... :

F20G2.4 : // 560bp // TGTTCCGGATAA : 1611
g-V-2804 : // 560bp // TGTTCCGGATAA : 1611

```

6.1.12. Sequenzdaten F38E9.2



F42A10.8	:	ATGTTCTTCCCGTTGTATTCTTCATCCCTTTCGTGCTGGGTGCCCGACTCAAAGGCCCTGGAAAAGATTCTGGTGGA	:	80
g-III-995	:	ATGTTCTTCCCGTTGTATTCTTCATCCCTTTCGTGCTGGGTGCCCGACTCAAAGGCCCTGGAAAAGATTCTGGTGGA	:	80
BI174472	:	-----TTCATCCCTTTCGTGCTGGGTGCCCGACTCAAAGGCCCTGGAAAAGATTCTGGTGGA	:	59
F42A10.8	:	CAATAATCCTGATTTCAGTTACAAATCGAGAGAAGATTTCGTGGAATTATCGATAAGGCGTTTGGAGAACGTGTTCCGAGGG	:	160
g-III-995	:	CAATAATCCTGATTTCAGTTACAAATCGAGAGAAGATTTCGTGGAATTATCGATAAGGCGTTTGGAGAACGTGTTCCGAGGG	:	160
BI174472	:	CAATAATCCTGATTTCAGTTACAAATCGAGAGAAGATTTCGTGGAATTATCGATAAGGCGTTTGGAGAACGTGTTCCGAGGG	:	139
F42A10.8	:	TTCAGGAAGACAAGGTGTTGCTCCTCTGTTACTTTCgtaggttttagtttaacttacaattaaagttaggtatcaaaa	:	240
g-III-995	:	TTCAGGAAGACAAGGTGTTGCTCCTCTGTTACTTTCgtaggttttagtttaacttacaattaaagttaggtatcaaaa	:	240
BI174472	:	TTCAGGAAGACAAGGTGTTGCTCCTCC-----	:	167
F42A10.8	:	atactccaaccttcaatgaaatTTTTatccaatagtaaattgtaaaggttatattccagGCTGCCTTAAACTACGGTC	:	320
g-III-995	:	atactccaaccttcaatgaaatTTTTatccaatagtaaattgtaaaggttatattccagGCTGCCTTAAACTACGGTC	:	320
F42A10FM	:	-----GCTGCCTTAAACTACGGTC	:	19
F42A10.8	:	CAAAGAACAATCAGACAAAGAAATTTGAAGAACTGAATCAAGATATCAACGGCAAgtttcttcgatttttttaggttaaaa	:	400
g-III-995	:	CAAAGAACAATCAGACAAAGAAATTTGAAGAACTGAATCAAGATATCAACGGCAAgtttcttcgatttttttaggttaaaa	:	400
F42A10FM	:	CAAAGAACAATCAGACAAAGAAATTTGAAGAACTGAATCAAGATATCAACG-----	:	70
F42A10.8	:	gaaactgtgtttttagaatacacatttgaatctgacatcatgctgaaatgagaaacaagccaacaacattgcaactgcaa	:	480
g-III-995	:	GAAACTgtgtttttagaatacacatttgaatctgacatcatgctgaaatgagaaacaagccaacaacattgcaactgcaa	:	480
F42A10FM	:	-----AATACACATTTGAATCTGACATCATGCTGAA TGAGAAAACAAGCCAACAACATA TTGCAACTGCAA	:	133
F42A10.8	:	ttgaaaatggaattatcgatcaaagaggcaagcagattgtagatactaccaatTTTTggagtgctcaggttccgatcttt	:	560
g-III-995	:	ttgaaaatggaattatcgatcaaagaggcaagcagattgtagatactaccaatTTTTggagtgctcaggttccgatcttt	:	560
F42A10FM	:	TTGAAAATGGAATTTATCGATCAAAGAGGCAAGCGATTGTAGATACTA CCAATTTTGAGTGTGCTGTTCCGATCTTT	:	213
F42A10.8	:	tatcaatttgatactaaattatgtaagttctagtaattTTTTgtttaaactgggaaaaaatattccagcgcaccaaac	:	640
g-III-995	:	tatcaatttgatactaaattatgtaagttctagtaattTTTTgtttaaactgggaaaaaatattccagcgcaccaaac	:	640
F42A10FM	:	TATCAATTTGATACTAAATTA-----CCGCCACCAAC	:	245
F42A10.8	:	atagccaatgtcagaaaagcaattcaatTTTTggaatgataacagctgtctgagcttcaaggtttccaacaatctattggt	:	720
g-III-995	:	atagccaatgtcagaaaagcaattcaatTTTTggaatgataacagctgtctgagcttcaaggtttccaacaatctattggt	:	720
F42A10FM	:	ATAGCCAATGTGCAAAAAGCAATTCAT TTTGGAATGATAACAGCTGT CTGAGCTTCAAG-----	:	305
F42A10.8	:	tcctttttcaaattataatTTTTcagGAAGACAATAATGCTAAAAACCGTCTCTTTTT ATCATCTGCAGGAGGATGTTGG	:	800
g-III-995	:	tcctttttcaaattataatTTTTcagGAAGACAATAATGCTAAAAACCGTCTCTTTTT ATCATCTGCAGGAGGATGTTGG	:	800
F42A10FM	:	-----GAAGACAATAATGCTAAAAACCGTCTCTTTTT ATCATCTGCAGGAGGATGTTGG	:	359
F42A10.8	:	TCATATGTCGGAAGCAAGTTGACATGCCATACCAATGGTTTCAGTTGGGCCAAACTgtgatacggtaattttattca	:	880
g-III-995	:	TCATATGTCGGAAGCAAGTTGACATGCCATACCAATGGTTTCAGTTGGGCCAAACTGTGATACGgtaattttattca	:	880
F42A10FM	:	TCATATGTCGGAAGCAAGTTGACATGC CATACCAAAATGGTTTCAGTT GGGCCAAACTGTGATACG-----	:	425
F42A10.8	:	tatttagacatttttacagaaaacattcaagtgTTTTatgtcttctctatgaagagctctaaaaacaattttttcagTT	:	960
g-III-995	:	tatttagacatttttacagaaaacattcaagtgTTTTatgtcttctctatgaagagctctaaaaacaattttttcagTT	:	960
F42A10FM	:	-----TT	:	427
F42A10.8	:	TGGTACTGCAACCCACGAGCTCATGCATGCCATTGGCTTCTGGCACCACAATCACGCGCCGATCGGGATAACTACGTCT	:	1040
g-III-995	:	TGGTACTGCAACCCACGAGCTCATGCATGCCATTGGCTTCTGGCACCACAATCACGCGCCGATCGGGATAACTACGTCT	:	1040
F42A10FM	:	TGGTACTGCAACCCACGAGCTCATGCATGCCATTGGCTTCTGGCACCACAATCACGCGCCGATCGGGATAACTACGTCT	:	507
F42A10.8	:	ATGTGGACTTTAGCAATATCATTTCCAAGTCAGGCATACAATTTTCAAAAAATGGCAGTGGATCAGGCACAGCTGTTGAAT	:	1120
g-III-995	:	ATGTGGACTTTAGCAATATCATTTCCAAGTCAGGCATACAATTTTCAAAAAATGGCAGTGGATCAGGCACAGCTGTTGAAT	:	1120
F42A10FM	:	ATGTGGACTTTAGCAATATCATTTCCAAGTCAGGCATACAATTTTCAAAA AAATGGCAGTGGATCAGGCACAGCTGTTGAAT	:	587
F42A10.8	:	TTGCCATATGATTATGGAAGTGTATGCAGTATTACCCATgtaagagcaaaaattaatttgaattggttaaatgaaaaatt	:	1200
g-III-995	:	TTGCCATATGATTATGGAAGTGTATGCAGTATTACCCATgtaagagcaaaaattaatttgaattggttaaatgaaaaatt	:	1200
F42A10FM	:	TTGCCATATGATTATGGAAGTGTATGCAGTATTACCCAT-----	:	627
F42A10.8	:	acagATGCATTTGCTGTGACTCATCGAAGTATAACAATTTGGCAAAGGAGAATGGATTCCAGAATAGTATGGACAACG	:	1280
g-III-995	:	acagATGCATTTGCTGTGACTCATCGAAGTATAACAATTTGGCAAAGGAGAATGGATTCCAGAATAGTATGGACAACG	:	1280
F42A10FM	:	---ATGCATTTGCTGTGACTCATCGAAGTATAACAATTTGGCAAAGGAGAATGGATTCCAGAATAGTATGGACAACG	:	703
F42A10.8	:	AGAGGCACCAGCATTTTCAGATATTATCGGAGTGAACAAGTTGTACAATTGCACAAgtaagcacaattgggtgatgtgtc	:	1360
g-III-995	:	AGAGGCACCAGCATTTTCAGATATTATCGGAGTGAACAAGTTGTACAATTGCACAAgtaagcacaattgggtgatgtgtc	:	1360
F42A10FM	:	AGAGGCACCAGCATTTTCAGATATTATCGGAGTGAACAAGTTGTACAATTGCACAA-----	:	759
F42A10.8	:	ttgtaataatactctttgatgggactgttagggttcaaaattggataaacttactgtagacggactgtagactgtaaaat	:	1440
g-III-995	:	ttgtaataatactctttgatgggactgttagggttcaaaattggataaacttactgtagacggactgtagactgtaaaat	:	1440
F42A10FM	:	-----	:	-
F42A10.8	:	gtagctattaccaagcacaatttttatactattcatttagccattaagcaacttttcattttttccagGCCAATGCAAGAT	:	1520
g-III-995	:	gtagctattaccaagcacaatttttatactattcatttagccattaagcaacttttcattttttccagGCCAATGCAAGAT	:	1520
F42A10FM	:	-----GCCAATGCAAGAT	:	772
F42A10.8	:	ACAAATGAAGTGTTCAAATTTGGAATCACAGACTCCCGTAACTGTAATCAATGCAAAATGTCGCGCTATTTTCACAGGAG	:	1600
g-III-995	:	ACAAATGAAGTGTTCAAATTTGGAATCACAGACTCCCGTAACTGTAATCAATGCAAAATGTCGCGCTATTTTCACAGGAG	:	1600
BI174472	:	-----	:	-
BI175435	:	-----TGTTCAAATTTGGAATCACAGACTCCCGTAACTGTAATCAATGCAAAATGTCGCGCTATTTTCACAGGAG	:	70
F42A10FM	:	ACAAATGAAGTGTTCAAATTTGGAATCACAGACTCCCGTAACTGTAATCAATGCAAAATGTCGCGCTATTTTCACAGGAG	:	852
F42A10.8	:	CTAGCTGCGATAGCTTACCATCTGGTACTGCTCCAAATTTGTAATGGAGCCGTGCTTCAGgtagattgaaacaactttttg	:	1680
g-III-995	:	CTAGCTGCGATAGCTTACCATCTGGTACTGCTCCAAATTTGTAATGGAGCCGTGCTTCAGgtagattgaaacaactttttg	:	1680
BI175435	:	CTAGCTGCGATAGCTTACCATCTGGTACTGCTCCAAATTTGTAATGGAGCCGTGCTTC-----	:	127
F42A10FM	:	CTAGCTGCGATAGCTTACCATCTGGTACTGCTCCAAATTTGTAATGGAGCCGTGCTTCAG-----	:	911
F42A10.8	:	agaactaacagaaacataatgctttcagGCTACATCCTCTTGGGAAACCTTCGACGCAAAGCCGGAGACCCATCCAGCTT	:	1760
g-III-995	:	agaactaacagaaacataatgctttcagGCTACATCCTCTTGGGAAACCTTCGACGCAAAGCCGGAGACCCATCCAGCTT	:	1760
BI175435	:	-----GGCTACATCCTCTTGGGAAACCTTCGACGCAAAGCCGGAGACCCATCCAGCTT	:	181
F42A10FM	:	-----GCTACATCCTCTTGGGAAACCTTCGACGCAAAGCCGGAGACCCATCCAGCTT	:	964

```

F42A10.8 : CTCTTCTTCCACAGACAATTGACCAATGTTACTGGCATATCAAAGCTCCAGAAGGACAACAAATTGAATTCAAAATGA : 1840
g-III-995 : CTCTTCTTCCACAGACAATTGACCAATGTTACTGGCATATCAAAGCTCCAGAAGGACAACAAATTGAATTCAAAATGA : 1840
BI175435 : CTCTTCTTCCACAGACAATTGACCAATGTTACTGGCATATCAAAGCTCCAGAAGGACAACAAATTGAATTCAAAATGA : 261
F42A10FM : CTCTTCTTCCACAGACAATTGACCAATGTTACTGGCATATCAAAGC----- : 1012

F42A10.8 : CAAAACTCCATTGGCAGCAATATGTATGCAAGAATGTCATGGCAATCGATTGAAGTTAATCTTGGAAAGTTTGACTTA : 1920
g-III-995 : CAAAACTCCATTGGCAGCAATATGTATGCAAGAATGTCATGGCAATCGATTGAAGTTAATCTTGGAAAGTTTGACTTA : 1920
BI175435 : CAAAACTCCATTGGCAGCAATATGTATGCAAGAATGTCATGGCAATCGATTGAAGTTAATCTTGGAAAGTTTGACTTA : 341

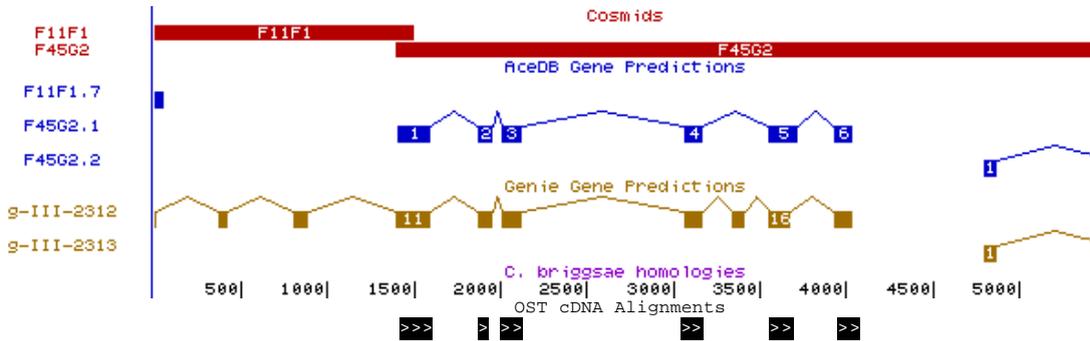
F42A10.8 : TTTGGAATGATgtgagtttaggtatactattgaattttcagttcgaaaattatttccagAACCTGCTGTGACACAATC : 2000
g-III-995 : TTTGGAATGATgtgagtttaggtatactattgaattttcagttcgaaaattatttccagAACCTGCTGTGACACAATC : 2000
BI175435 : TTTGGAATGAT-----AACCTGCTGTGACACAATC : 371

F42A10.8 : CTAAACCAAGTGTTCACCTCTGAGCTTAATATGATTGCTCTTCGTGGAATAATTGCTACAATCAGTTGACGTTTAGCAT : 2080
g-III-995 : CTAAACCAAGTGTTCACCTCTGAGCTTAATATGATTGCTCTTCGTGGAATAATTGCTACAATCAGTTGACGTTTAGCAT : 2080
BI175435 : CTAAACCAAGTGTTCACCTCTGAGCTTAATATGATTGCTCTTCGTGGAATAATTGCTACAATCAGTTGACGTTTAGCAT : 451

F42A10.8 : TCAATATCGAGCAGTGCCAAGTTCGAAGCCTGCTTCAACCAATGCTTGAACCAAGTGA : 2141
g-III-995 : TCAATATCGAGCAGTGCCAAGTTCGAAGCCTGCTTCAACCAATGCTTGAACCAAGTGA : 2141
BI175435 : TCAATATCGAGCAGTGCCAAGTTCGAAGCCTGCTT----- : 487

```

6.1.15. Sequenzdaten F45G2.1



```

F45G2.1 : -----ATGCTCCCGATCGTTGTG : 18
g-III-2312 : // aggaacataaaaaattatttggtaaaatcttcttatcaaaaacttgcagATCCATAATGCTCCCGATCGTTGTG : 80

F45G2.1 : TCCATCCTTTGGCAACTCCCTTGGCACTGTGCCAAGCTCCCTTCTGGATGCCTAATATGCAGCAATTCATTTTAAAC : 98
g-III-2312 : TCCATCCTTTGGCAACTCCCTTGGCACTGTGCCAAGCTCCCTTCTGGATGCCTAATATGCAGCAATTCATTTTAAAC : 160
BI174418 : -CCATCCTTTGGCAACTCCCTTGGCACTGTGCCAAGCTCCCTTCTGGATGCCTAATATGCAGCAATTCATTTTAAAC : 79

F45G2.1 : CGAAACAGATTTTCGTAATGCTCTTCTCTTGCACAACAACTAATACTCGACGGGTCAGAAGTCTTTATCATGATATGAGAA : 178
g-III-2312 : CGAAACAGATTTTCGTAATGCTCTTCTCTTGCACAACAACTAATACTCGACGGGTCAGAAGTCTTTATCATGATATGAGAA : 240
BI174418 : CGAAACAGATTTTCGTAATGCTCTTCTCTTGCACAACAACTAATACTTACGGGTCAGAAGTCTTTATCATGATATGAGAA : 159

F45G2.1 : TTCCATTCAGgtatgatttgggtcaaaagtaatatctatagaacacacacctggtttcaaaaaattataaaataaact : 258
g-III-2312 : TTCCATTCAGgtatgatttgggtcaaaagtaatatctatagaacacacacctggtttcaaaaaattataaaataaact : 320
BI174418 : TTCCATTCAG----- : 170

F45G2.1 : tgaataaacctcctaagttcagtttttcaaatgaaagaatggcttgaataattgaaaaaaatttctccaaataatatt : 338
g-III-2312 : tgaataaacctcctaagttcagtttttcaaatgaaagaatggcttgaataattgaaaaaaatttctccaaataatatt : 400

F45G2.1 : gacttttaaaatcaagattatactggaatccatcatttaaaaaaaactctaataatgagaaagtgggctgaaaaacg : 418
g-III-2312 : gacttttaaaatcaagattatactggaatccatcatttaaaaaaaactctaataatgagaaagtgggctgaaaaacg : 480

F45G2.1 : ttttgataaaactttttctgggtcaatttcttccacttttcagAGATTCAAACGCGCGGGTGGTGTGCGAGTGGCTGTGTA : 498
g-III-2312 : ttttgataaaactttttctgggtcaatttcttccacttttcagAGATTCAAACGCGCGGGTGGTGTGCGAGTGGCTGTGTA : 560
BI174418 : -----AGATTCAAACGCGCGGGTGGTGTGCGAGTGGCTGTGTA : 208
BI175363 : -----AGTGGCTGTGTA : 12

F45G2.1 : AAAGGACAAATGGCCAAATGGAAGAGTTCCTTATATCTATCTGCAGCTTACAgtaattttgctatacttttgattattt : 578
g-III-2312 : AAAGGACAAATGGCCAAATGGAAGAGTTCCTTATATCTATCTGCAGCTTACAgtaattttgctatacttttgattattt : 640
BI174418 : AAAGGACAAATGGCCAAATGGAAGAGTTCCTTATATCTATCTGCAGCTTAC----- : 261
BI175363 : AAAGGACAAATGGCCAAATGGAAGAGTTCCTTATATCTATCTGCAGCTTAC----- : 65

F45G2.1 : aaattgaaaatttggtttcagCAAGTGCACAAAGAGCCGTGCTAGCCCGAGCGTTTGATACATACGCCAAGAGAAGTGT : 658
g-III-2312 : aaattgaaaatttggtttcagCAAGTGCACAAAGAGCCGTGCTAGCCCGAGCGTTTGATACATACGCCAAGAGAAGTGT : 720
BI174418 : -----CAAGTGCACAAAGAGCCGTGCTAGCCCGAGCGTTTGATACATACGCCAAGAGAAGTGT : 320
BI175363 : -----CAAGTGCACAAAGAGCCGTGCTAGCCCGAGCGTTTGATACATACGCCAAGAGAAGTGT : 83

F45G2.1 : ATTCGCTTTGTTCCAAAGTCACCAGCTGATAAGGATTATATTTGTCATTGAGAAGTTGGATGGgtatgttttcaaaaa : 738
g-III-2312 : ATTCGCTTTGTTCCAAAGTCACCAGCTGATAAGGATTATATTTGTCATTGAGAAGTTGGATGGgtatgttttcaaaaa : 800
BI174418 : ATTCGCTTTGTTCCAAAGTCACCAGCTGATAAGGATTATATTTGTCATTGAGAAGTTGGATGG----- : 382
BI175363 : ATTCGCTTTGTTCCAAAGTCACCAGCTGATAAGGATTATATTTGTCATTGAGAAGTTGGATGG----- : 145
BI175385 : -----TCCAAAGTCACCAGCTGATAAGGATTATATTTGTCATTGAGAAGTTGGATGG----- : 51

F45G2.1 : // 810bp // ttaaaattaaaaaagggtgactttaaccatataattataagattcttcgccaactttctagaattttgaa : 1618
g-III-2312 : // 810bp // ttaaaattaaaaaagggtgactttaaccatataattataagattcttcgccaactttctagaattttgaa : 1680

F45G2.1 : ttttcgaaaagctaaaatttttctaaaacttaaaattacagATGTTACGCAGACTTCTCACGTGTGGAGGCCGTCACAAA : 1698
g-III-2312 : ttttcgaaaagctaaaatttttctaaaacttaaaattacagATGTTACGCAGACTTCTCACGTGTGGAGGCCGTCACAAA : 1760
BI174418 : -----ATGTTACGCAGACTTCTCACGTGTGGAGGCCGTCACAAA-- : 420
BI175363 : -----ATGTTACGCAGACTTCTCACGTGTGGAGGCCGTCACAAA : 184
BI175385 : -----ATGTTACGCAGACTTCTCACGTGTGGAGGCCGTCACAAA : 91

```

```

F45G2.1 : GTGTCTCTGGCTGATGAGTGTATTGATTACGCAACTATCATCCATGAGTTGATGCATGTTATTGggttagttgagatcata : 1778
g-III-2312 : GTGTCTCTGGCTGATGAGTGTATTGATTACGCAACTATCATCCATGAGTTGATGCATGTTATTGggttagttgagatcata : 1840
BI175363 : GTGTCTCTGGCTGATGAGTGTATTGATTACGCAACTATCATCCATGAGTTGATGCATGTTATTG----- : 248
BI175385 : GTGTCTCTGGCTGATGAGTGTATTGATTACGCAACTATCATCCATGAGTTGATGCATGTTATTG----- : 155

F45G2.1 : gaatttggtttatttggtttaatttatttggctacttacgagaatctgctcatgagtagacaacacagagaaaatacaaaaa : 1858
g-III-2312 : gaatttggtttatttggtttaatttatttggctacttacgagaatctgctcatgagtagacaacacagagaaaatacaaaaa : 1920

F45G2.1 : aagtttgatgccaatgggtatgggtgacttctgattagagatcccatattttgatgattaaaaaatatatttaaacagga : 1938
g-III-2312 : aagtttgatgccaatgggtatgggtgacttctgattagagatcccatattttgatgattaaaaaatatatttaaacagGA : 2000

F45G2.1 : ttcattccacgaacatcaacgagaagatcggtgattcatatgtctccatcctgtatcaaaaatggttattcaaggtagcatga : 2018
g-III-2312 : TTCATCCACGAACATCAACGAGAAGATCGTGATTCATATGTCTCCATCCTGTATCAAAATGTTATTCAAGtaccgatga : 2080
BI175363 : TTCATCCACGAACATCAACGAGAAGATCGTGATTCATATGTCTCCATCCTGTATCAAAATGTTATTCAAG----- : 320
BI175385 : TTCATCCACGAACATCAACGAGAAGATCGTGATTCATATGTCTCCATCCTGTATCAAAATGTTATTCAAG----- : 227

F45G2.1 : //Intron 95 bp// aaattgaattatgatttttacagGAGCTAACACTGACTTCGACAAGCTGTCAAACCTGGGACT : 2178
g-III-2312 : //Intron 95 bp// aaattgaattatgatttttacagGAGCTAACACTGACTTCGACAAGCTGTCAAACCTGGGACT : 2240
BI175363 : -----GAGCTAACACTGACTTCGACAAGCTGTCAAACCTGGGACT : 360
BI175385 : -----GAGCTAACACTGACTTCGACAAGCTGTCAAACCTGGGACT : 267

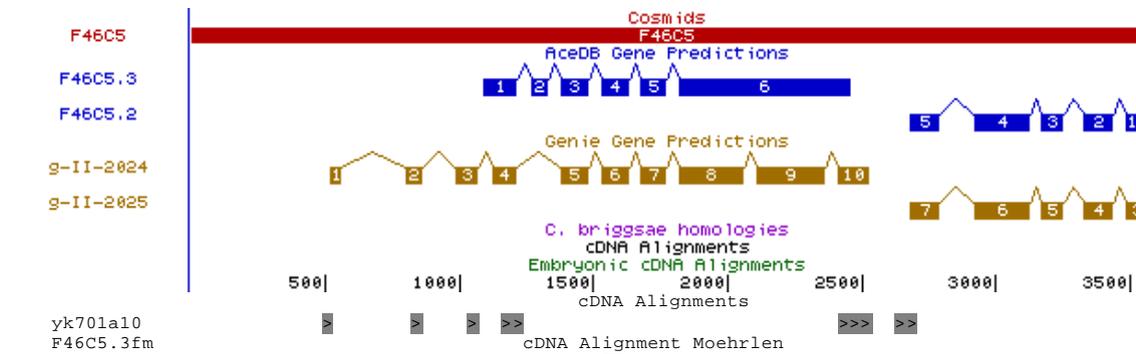
F45G2.1 : ATCCTATTATGGAGAACATTACGATTATAGTTCTATCATGCATTACGAAGCGAATGAAGGATCTCGAAATGGAAAGAACA : 2258
g-III-2312 : ATCCTATTATGGAGAACATTACGATTATAGTTCTATCATGCATTACGAAGCGAATGAAGGATCTCGAAATGGAAAGAACA : 2320
BI175363 : ATCCTATTATGGAGAACATTACGATTATAGTTCTATCATGCATTACGAAGCGAATGAAGGATCTCGAAATGGAAAGAACA : 440
BI175385 : ATCCTATTATGGAGAACATTACGATTATAGTTCTATCATGCATTACGAAGCGAATGAAGGATCTCGAAATGGAAAGAACA : 347

F45G2.1 : CAATTGAGGCAAAGGTGAGATATTTAATTGTGCTGATTATGCCTAATAgtaagtatttctcatagttttatacattga : 2338
g-III-2312 : CAATTGAGGCAAAGgtgagatatttgaattgtgctgattatgctcaatagtaagtatttctcatagttttatacattga : 2400
BI175363 : CAATTGAGGCAAAG----- : 454
BI175385 : CAATTGAGGCAAAG----- : 361

F45G2.1 : // 170 bp // aatttattttcagAACTCCCACTTCACAGCAATAATGGGAAAAGCCAGCGACTTCTCGACTTCAGAT : 2578
g-III-2312 : // 170 bp // aatttattttcagAACTCCCACTTCACAGCAATAATGGGAAAAGCCAGCGACTTCTCGACTTCAGAT : 2640
BI175363 : -----AACTCCCACTTCACAGCAATAATGGGAAAAGCCAGCGACTTCTCGACTTCAGAT : 508
BI175385 : -----AACTCCCACTTCACAGCAATAATGGGAAAAGCCAGCGACTTCTCGACTTCAGAT : 415

F45G2.1 : CTTGACGCTGCAATAGAGCCTATAAATGTCCATTTTTCGCTTTTAA----- : 2626
g-III-2312 : CTTGACGCTGCAATAGAGCCTATAAATGTCCATTTTTCGCTTTTAA----- : 2688
BI175363 : cttgacgctgtcaatagag----- : 527
BI175385 : cttgacgctgtcaatagag----- : 434
    
```

6.1.16. Sequenzdaten F46C5.3



```

g-II-2024 : ATGCAGATCTATTTAGGTATCACAAATTTGTCTCGTGGCTTTTTGggttagtttttaaaggagaatgcaaaaaatatt : 80
YK701A10 : ATGCAGATCTATTTAGGTATCACAAATTTGTCTCGTGGCTTTTTG----- : 45

g-II-2024 : tttgacaagtccaagttaggaattcgaaaaatggttaattacaataatcttcttggttgaggtataaaattttttcaga : 160
g-II-2024 : acagttgcttttacttactggtcatttttgatggaacctataatagattttcatataaataatctcaatctccatttga : 240

g-II-2024 : aatattttgaaaaaatgaattaattcgccatttttaccatttcagACGGTCATCGACTGTGCCATTCCGTATTACCGC : 320
YK701A10 : -----ACGGTCATCGACTGTGCCATTCCGTATTACCGC : 78

g-II-2024 : ACACATTCAAACCTTGGATCACTGGGCAGGgttcgctcattttgtaaaaaaatcaatctcaatctcaaaataacgcgtttg : 400
YK701A10 : ACACATTCAAACCTTGGATCACTGGGCAGG----- : 108

g-II-2024 : ttatgtcaattctaaaaatacaaatgttttaaaactcggacactcttcaatgactcaccaaaataaaattttcagAGAAAA : 480
YK701A10 : -----AGAAAA : 114

g-II-2024 : GTTAGACAAGTCCAGAGAGATCTCACATATAGATGGCCAAACAATACGGTTCCATATTATGTGGGAAATGTGACAAgta : 560
YK701A10 : GTTAGACAAGTCCAGAGAGATCTCACATATAGATGGCCAAACAATACGGTTCCATATTATGTGGGAAATGTGACAA---- : 190

F46C5.3 : -----ATGTTGCTGGTGTACGTTGTGGTTTTAGGTACAATCAAGAAATCTGTGAGATTAGCTATTGA : 62
g-II-2024 : gttaacgaaaagtcaagaaatgttgcgtgtacgttgggttttagGTACAATCAAGAAATCTGTGAGATTAGCTATTGA : 640
YK701A10 : -----GTACAATCAAGAAATCTGTGAGATTAGCTATTGA : 224

F46C5.3 : GGAGCTTCAGGCATGGACTTGTATTTCGGTTTCAAACGTTAATGAGAAATATTCAGgtaaatcttctggcagagtgaatat : 142
g-II-2024 : GGAGCTTCAGGCATGGACTTGTATTTCGGTTTCAAACGTTAATGAGAAATATTCAGgtaaatcttctggcagagtgaatat : 720
YK701A10 : GGAGCTTCAGGCATGGACTTGTATTTCGGTTTCAAACGTTAATGAGAA----- : 271

F46C5.3 : // 250bp //gtaataggatattttgaaatagtaaaataagacaatttaagATGGGAACTGCAATTCACGAATTAATG : 462
g-II-2024 : // 250bp //gtaataggatattttgaaatagtaaaataagacaatttaagATGGGAACTGCAATTCACGAATTAATG : 1040

F46C5.3 : CATGCAATTGGGATCGAGCATACTCAATCCAGAAGTGACAGGAATCGATATTTAGATATTCTAGCTCAAAACATTGATgt : 542
g-II-2024 : CATGCAATTGGGATCGAGCATACTCAATCCAGAAGTGACAGGAATCGATATTTAGATATTCTAGCTCAAAACATTGATgt : 1120
    
```

```

F46C5.3 : aagattttttaaaatatggtttttgaattttcagAATCGGGATCTCCGAATTCGAAGTACTATCGCCTCGT : 622
g-II-2024 : aagattttttaaaatatggtttttgaattttcagAATCGGGATCTCCGAATTCGAAGTACTATCGCCTCGT : 1200

F46C5.3 : CTCTGGGCTAACCTAGTACCGTATGATTATGGAAGTGTATGCATTATTCAGCTGATTCgtaagagaataaggagttttt : 702
g-II-2024 : CTCTGGGCTAACCTAGTACCGTATGATTATGGAAGTGTATGCATTATTCAGCTGATTCgtaagagaataaggagttttt : 1280

F46C5.3 : caatcaaatatcaagacatttttagTTTTTCGAACAAAGATGACGAACAAACAATGCTTCCAAAAGATCGATCATTATATCG : 782
g-II-2024 : caatcaaatatcaagacatttttagTTTTTCGAACAAAGATGACGAACAAACAATGCTTCCAAAAGATCGATCATTATATCG : 1360
F46C5.3fm : -----TTCAAAAGATCGATCATTATATCG : -----

F46C5.3 : AAACAATGGGAAGTATGATTCCAAACTTTTACGACTTCGATCAAATTAATCAGTATTATCAGTGTATGACTCATGCAGA : 862
g-II-2024 : AAACAATGGGAAGTATGATTCCAAACTTTTACGACTTCGATCAAATTAATCAGTATTATCAGTGTATGACTCATGCAGA : 1440
F46C5.3fm : AAACAATGGGAAGTATGATTCCAAACTTTTACGACTTCGATCAAATTAATCAGTATTATCAGTGTATGACTCATGCAGA : -----

F46C5.3 : AATGCCGGCAATTAGCTAATTGTGCAAAATGGGGGAATACCAAACCCCAATAATTGTC AAGTTTGTAACTGTCCAATGGG : 942
g-II-2024 : AATGCCGGCAATTAGCTAATTGTGCAAAATGGGGGAATACCAAACCCCAATAATTGTC AAGTTTGTAACTGTCCAATGGG : 1520
F46C5.3fm : AATGCCGGCAATTAGCTAATTGTGCAAAATGGGGGAATACCAAACCCCAATAATTGTC ----- : -----

F46C5.3 : TTATGGTGGAGATCTCTGTGATCAGAGGGTAATTTGTACGTATACAAATTTAAACCAATATGAACCTGTTTCAGCCAG : 1022
g-II-2024 : TTATGGTGGAGATCTCTGTGATCAGAGGGtaattttgtacgtatacaaaattctaaaaccaaatgaacctgtttcagCCAG : 1600

F46C5.3 : AAGGATGTGGCAGTACACTCGTTGCAACAGATCGCTGGCAAAAACAGAAGCTTTCTGTGAGATTTCAGTAGAAATGATGAT : 1102
g-II-2024 : AAGGATGTGGCAGTACACTCGTTGCAACAGATCGCTGGCAAAAACAGAAGCTTTCTGTGAGATTTCAGTAGAAATGATGAT : 1680

F46C5.3 : CAGTATTTACGTTTTGTAAATTCATGGATTGTGCGGACCTCCGATCGAACATTGCAAGTTATTTATGAAATTACATCCGA : 1182
g-II-2024 : CAGTATTTACGTTTTGTAAATTCATGGATTGTGCGGACCTCCGATCGAACATTGCAAGTTATTTATGAAATTACATCCGA : 1760
YK701A10 : -----TAATTCATGGA TTGTGCGGACCTCCGATCGAACATTGCAAGTTATTTATGAAATTACATCCGA : 334

F46C5.3 : TTCAATCAGACGACAAAATTTGTTTCATTTGGTTGTTATGAAGGAGGAATGAAGTGAACATCTTCAGGACCCAAGGATAA : 1262
g-II-2024 : TTCAATCAGACGACAAAATTTGTTTCATTTGGTTGTTATGAAGGAGGAATGAAGTGAACATCTTCAGGACCCAAGGATAA : 1840
YK701A10 : TTCAATCAGACGACAAAATTTGTTTCATTTGGTTGTTATGAAGGAGGAATGAAGTGAACATCTTCAGGACCCAAGGATAA : 414

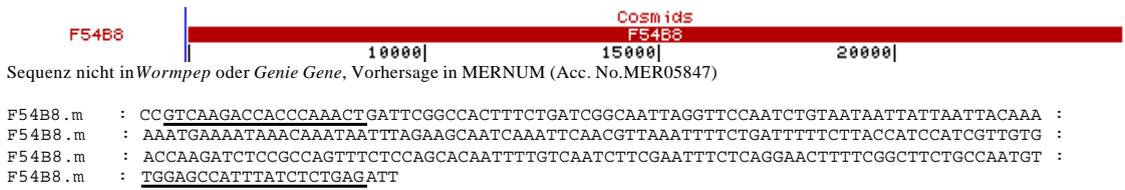
F46C5.3 : CAAATGATAGTACTTTTTGAAGTTATGCAAACTATCAATTGAGTTATTCTTTCAGAGATTGTTGTCTGAATACTCCATT : 1342
g-II-2024 : CAAATGATAGTactcttttgaagttatgcaaaactatcaattgagttattctttcagAGATTGTTGCTGAATACTCCATT : 1920
YK701A10 : CAAATGATAG-----AGATTGTTGTCTGAATACTCCATT : 448

F46C5.3 : AAATCTAACAACTACTGTAA----- : 1363
g-II-2024 : AAATCTAACAACTACTGTAAATCCACTCCCAGTTATTCTATATACGTCGGGAGCAACTGTCACTTATGACTTTTCGTACA : 2000
YK701A10 : AAATCTAACAACTACTGTAAATCCACTCCAGTTATTCTATATACGTCGGGAGCAACTGTCACTTATGACTTTTCGTACA : 528

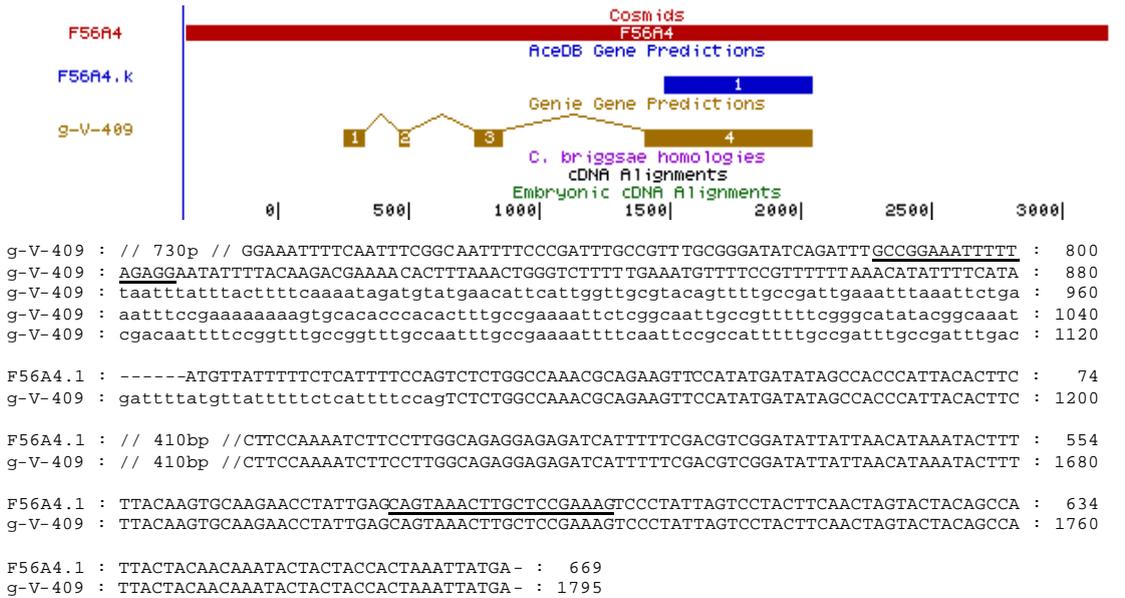
g-II-2024 : GATACGTGTGAgctcgaactttgttaacagttttgtt----- : 2011
YK701A10 : GATACGTGTGAGTCAGAACTTTGTTAAATCAGTTTGT----- : 566

```

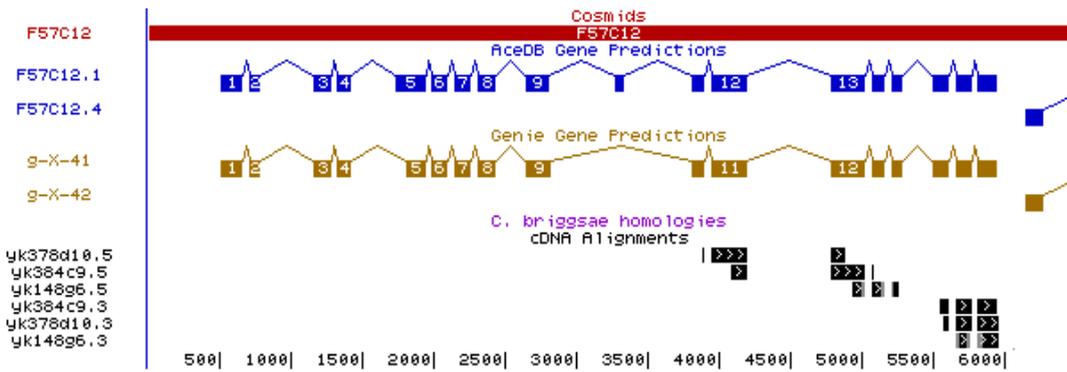
6.1.17. Sequenzdaten F54B8.m



6.1.18. Sequenzdaten F56A4.1



6.1.19. Sequenzdaten F57C12.1



```

F57C12.1 : // 3370bp //TTATCGCAGCGCCAATCGgtaaaatatcagttttaaatttactactatcattatatttcagAAGGAA : 3440
g-X-41 : // 3370bp //TTATCGCAGCGCCAATCGgtaaaatatcagttttaaatttactactatcattatatttcagAAGGAA : 3440
YK378D10 : -----AGGCCAATCG-----AAGGAA :
F57C12.1 : GAGTGTTCATGGAATTTGAAGGAGATTTTGATTTTCTCTGCGAAGACACTTGCACAAAGGCATACGTAGAAGTTAAATAT : 3520
g-X-41 : GAGTGTTCATGGAATTTGAAGGAGATTTTGATTTTCTCTGCGAAGACACTTGCACAAAGGCATACGTAGAAGTTAAATAT : 3520
YK378D10 : GAGTGTTCATGGAATTTGAAGGAGATTTTGATTTTCTCTGCGAAGACACTTGCACAAAGGCATACGTAGAAGTTAAATAT :
F57C12.1 : CATAGTGACAAACGCTTACTGGAGCCAGATACTGTTGCTCGCTGCTTCCAAAAATCGATTATCTCTTTCAAAAATGA : 3600
g-X-41 : CATAGTGACAAACGCTTACTGGAGCCAGATACTGTTGCTCGCTGCTTCCAAAAATCGATTATCTCTTTCAAAAATGA : 3600
yK378d10 : catagtgcacaaacgcttactggagccagatactgttgcctgctgttccaaaaatcgattatctctttcaaaaatga :
YK384C9 : .....TCGATTATCTCTTTCAAAAATGA :
F57C12.1 : GATGATTATTATAATGAGAGGGTATAGATCAAGTGGAGCTGGGTTTAAAGCCAAGTCTGGTCTAATTTGGGAGAGCCAG : 3680
g-X-41 : GATGATTATTATAATGAGAGGGTATAGATCAAGTGGAGCTGGGTTTAAAGCCAAGTCTGGTCTAATTTGGGAGAGCCAG : 3680
YK378D10 : GATGATTATTATAATGAGAGGGTATAGATCAAGTGGAGCTGGGTTTAAAGCCAAGTCTGGTCTAATTTGGGAGAGCCAG :
YK384C9 : GATGATTATTATAATGAGAGGGTATAGATCAAGTGGAGCTGGGTTTAAAGCCAAGTCTGGTCTAATTTGGGAGAGCCAG :
F57C12.1 : AAGGAgtaagtgtgtttttccttagatagtttagaaaaaattgttgcattagttgtgaattaggtgtgtgtttttgaaatg : 3760
g-X-41 : AAGGAgtaagtgtgtttttccttagatagtttagaaaaaattgttgcattagttgtgaattaggtgtgtgtttttgaaatg : 3760
YK378D10 : AAGGA----- :
YK384C9 : AAGGA----- :
F57C12.1 : // 570 bp //caggaatagtagtaataattaaatggtgtagGTTAGCACTCCATTGCCGCCAACAAACAGCTCCGCTCC : 4320
g-X-41 : // 570 bp //caggaatagtagtaataattaaatggtgtagGTTAGCACTCCATTGCCGCCAACAAACAGCTCCGCTCC : 4320
YK378D10 : -----GTTAGCACTCCATTGCCGCCAACAAACAGCTCCGCTCC :
YK384C9 : -----GTTAGCACTCCATTGCCGCCAACAAACAGCTCCGCTCC :
F57C12.1 : CAGAAATTTTCGGAACAACACAAAAGCCAGAACCACCACTGTACAATCTACAACACATATACAACGGCAATTCCGCGG : 4400
g-X-41 : CAGAAATTTTCGGAACAACACAAAAGCCAGAACCACCACTGTACAATCTACAACACATATACAACGGCAATTCCGCGG : 4400
YK378D10 : CAGAAATTTTCGGAACAACACAAAAGCCAGAACCACCACTGTACAATCTACAACACATATACAACGGCAATTCCGCGG :
YK384C9 : CAGAAATTTTCGGAACAACACAAAAGCCAGAACCACCACTGTACAATCTACAACACATATACAACGGCAATTCCGCGG :
F57C12.1 : AGAACCGCAAAAAACAGTTCTTCAAGAAAACCCATCACAATCCCATTGACACCTCTAACTTCCTCGTCAACTACAAC : 4480
g-X-41 : AGAACCGCAAAAAACAGTTCTTCAAGAAAACCCATCACAATCCCATTGACACCTCTAACTTCCTCGTCAACTACAAC : 4480
YK384C9 : AGAACCGCAAAAAACAGTTCTTCAAGAAAACCCATCACAATCCCATTGACACCTCTAACTTCCTCGTCAACTACAAC :
YK148G6 : -----AGAAA-CCATCACACTCCCATTGACACCTCTAACTTCCTCGTCAACTACAAC :
F57C12.1 : GGAGTCTACCACTGTATCTTCTACGACACAATCAACTgtaaggggaaaaacgtgattcggcttacaatcaacatttttag : 4560
g-X-41 : GGAGTCTACCACTGTATCTTCTACGACACAATCAACTgtaaggggaaaaacgtgattcggcttacaatcaacatttttag : 4560
YK384C9 : GGAGTCTACCACTGTATCTTCTACGACACAATCAACT----- :
YK148G6 : GGAGTCTACCACTGTATCTGCTACGACACAATCAANT----- :
F57C12.1 : ACCTGGTTGCCAACTGAACCATCATTCGCTACCGGCGAAACGGAAATCACAACCTGCAAGCCCCACAATCAGTTATTCCC : 4640
g-X-41 : ACCTGGTTGCCAACTGAACCATCATTCGCTACCGGCGAAACGGAAATCACAACCTGCAAGCCCCACAATCAGTTATTCCC : 4640
YK384C9 : ACCTGGTTGCCAACTGA----- :
YK148G6 : ACCTGGTTGCCANCTGAACCATCATTCGCTACCGGNGA .ACGGAAATCA CANCTGNAAGCCCCACAGTCACGTTATTCCC :
F57C12.1 : AAGTTTATCAACTATgtagcttccaagttttttgttgagaaattgttttttccagACTTCTCCTATCAATTCGTTA : 4720
g-X-41 : AAGTTTATCAACTATgtagcttccaagttttttgttgagaaattgttttttccagACTTCTCCTATCAATTCGTTA : 4720
YK148G6 : ANGTTTATCAACTNT-----CTCCTCCTATCAACTTCGTTA :
F57C12.1 : GCTGGAGTTCTACCATCTACACAAGCTCCAGATATTATCAATAGTGTATTAGGtaccggcgctttattcagacaaagttt : 4800
g-X-41 : GCTGGAGTTCTACCATCTACACAAGCTCCAGATATTATCAATAGTGTATTAGGtaccggcgctttattcagacaaagttt : 4800
YK148G6 : GCTGGAGTTCTACCATCTACACAAGCTCCAGA----- :
F57C12.1 : //Intron 175bp// agcaaatgttttcagAATGCGGCTGTGGAGCTTGGTCTGAATGGCAAGGTGAATGCTCCCAGC : 5040
g-X-41 : //Intron 175bp// agcaaatgttttcagAATGCGGCTGTGGAGCTTGGTCTGAATGGCAAGGTGAATGCTCCCAGC : 5040
F57C12.1 : AGTGTGGCGGATGCGGTCATCGACTGAGGAAGCGAGAGTGCAAAAAAGAAGCTTGCAGgtttggaaatataatgtcacat : 5120
g-X-41 : AGTGTGGCGGATGCGGTCATCGACTGAGGAAGCGAGAGTGCAAAAAAGAAGCTTGCAGgtttggaaatataatgtcacat : 5120
YK378D10 : -----CGACTGAGGAAGCGAGAGTGCAAAAAAGAAGCTTGCAG----- :
YK384C9 : ----TGGCGGATGCGGTCATCGACTGAGGAAGCGAGAGTGCAAAAAAGAAGCTTGCAG----- :
YK148G6 : -----ATCGACTG-GGAAGCGAGAGT----- :
F57C12.1 : ttatgtctatcaataaacattttgtagAAAAGAAGAAAAACGCCCATGCAACTTCTCAGCATGCCCCGACGGCACTAACTT : 5200
g-X-41 : ttatgtctatcaataaacattttgtagAAAAGAAGAAAAACGCCCATGCAACTTCTCAGCATGCCCCGACGGCACTAACTT : 5200
YK378D10 : -----AAAAGAAGAAAAACGCCCATGCAACTTCTCAGCATGCCCCGACGGCACTAACTT :
YK384C9 : -----AAAAGAAGAAAAACGCCCATGCAACTTCTCAGCATGCCCCGACGGCACTAACTT :
YK148G6 : -----ANAGAAGAA--ACGCCCATNCAACTTCTCAGCATGCCCCGACGGCACTAACTT :
    
```

```

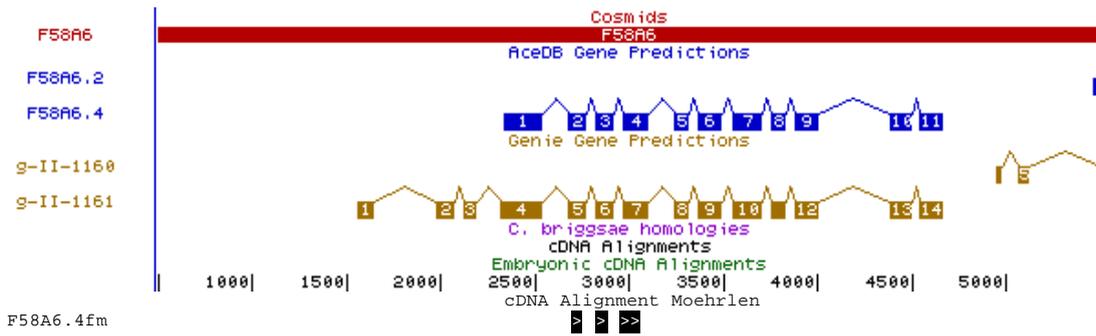
F57C12.1 : CCTAATCAACAATGCTGAATTCCATATTTCTTTGGAGAGGGTGTGTGTTGGGTTGTTTCAGgtattatattttcaataaac : 5280
g-X-41   : CCTAATCAACAATGCTGAATTCCATATTTCTTTGGAGAGGGTGTGTGTTGGGTTGTTTCAGgtattatattttcaataaac : 5280
YK378D10 : CCTAATCAACAATGCTGAATTCCATATTTCTTTGGAGAGGGTGTGTGTTGGGTTGTTTCAG----- :
YK384C9  : CCTAATCAACAATGCTGAATTCCATATTTCTTTGGAGAGGGTGTGTGTTGGGTTGTTTCAG----- :
YK148G6  : CCTAATCAACNATGCTGNATTCATATTTCTTTGGAGAGGGTGTGTGTTGGG----- :

F57C12.1 : aaatttcaaaatagcttattcttagATCCGGAGATCAATGCTCCGCCCTGAAACTGAATCGAACCATTCTTCAAATA : 5360
g-X-41   : aaatttcaaaatagcttattcttagATCCGGAGATCAATGCTCCGCCCTGAAACTGAATCGAACCATTCTTCAAATA : 5360
YK378D10 : -----ATCCGGAGATCAATGCTCCGCCCTGAAACTGAATCGAACCATTCTTCAAATA :
YK384C9  : -----ATCCGGAGATCAATGCTCCGCCCTGAAACTGAATCGAACCATTCTTCAAATA :
YK148G6  : -----AGATCCGGAGATCAATGCTCCGCCCTGAAACTGAATCGAACCATTCTTCAAATA :

F57C12.1 : TCAACAGCCTGCTGAACATTCAAGATGCTAAGAACAATGATACGTTAATAGCAAAAAGAATGATGCGAGGAGAACTAA : 5440
g-X-41   : TCAACAGCCTGCTGAACATTCAAGATGCTAAGAACAATGATACGTTAATAGCAAAAAGAATGATGCGAGGAGAACTAA : 5440
YK378D10 : TCAACAGCCTGCTGAACATTCAAGATGCTAAGAACAATGATACGTTAATAGCAAAAAGAATGATGCGAGGAGAACTAA :
YK384C9  : TCAACAGCCTGCTGAACATTCAAGATGCTAAGAACAATGATACGTTAATAGCAAAAAGAATGATGCGAGGAGAACTAA :
YK148G6  : TCAACAGCCTGCTGATACATTCAAGATNCTAAGAACAATGATACGNTAATATCACCAGAATGATGCGAGGAGAACTAA :

```

6.1.20. Sequenzdaten F58A6.4



```

F58A6.4 : // 428bp //AAAGAGgttatccttttaaaaaatattagaactataattggaattttattcagTATCTCCTTTTATCG : 498
g-II-1161 : //1210bp //AAAGAGgttatccttttaaaaaatattagaactataattggaattttattcagTATCTCCTTTTATCG : 1280

F58A6.4 : GAAATGACGATGGGTGTTGGAGTACAGTGGGACGAGATGCATCACAAGGAAAACAAGTTGTTAGTATTGGAAACGGATGT : 578
g-II-1161 : GAAATGACGATGGGTGTTGGAGTACAGTGGGACGAGATGCATCACAAGGAAAACAAGTTGTTAGTATTGGAAACGGATGT : 1360
F58A6.4fm : GAAATGACGATGGGTGTTGGAGTACAGTGGGACGAGATGCATCACAAGGAAAACAAGTTGTTAGTATTGGAAACGGATGT :

F58A6.4 : GAGCATgttggtttttgttaaatcgtaattgtgctctaaaaaagacgtatttcagTTCGGAGTACATCTCATGAGCTGG : 658
g-II-1161 : GAGCATgttggtttttgt taatcgtaattgtgctctaaaaaagacgtatttcagTTCGGAGTACATCTCATGAGCTGG : 1440
F58A6.4fm : GAGCAT-----TTCGGAGTACATCTCATGAGCTGG :

F58A6.4 : CACACGCATTGGGAATATCCACGAACAATCAAGATTTGATCGGGATGAATCTGTAGTGTTCATCCAAGAGTCGTTGAA : 738
g-II-1161 : CACACGCATTGGGAATATCCACGAACAATCAAGATTTGATCGGGATGAATCTGTAGTGTTCATCCAAGAGTCGTTGAA : 1520
F58A6.4fm : CACACGCATTGGGAATATCCACGAACAATCAAGATTTGATCGGGATGAATCTGTAGTGTTCATCCAAGAGTCGTTGAA :

F58A6.4 : AGAGATCTTTTATTCAATTTTGCAAAAgtaggtatacttttaaatggaacaagatatctaattggaagcaagaaaaat : 818
g-II-1161 : AGAGATCTTTTATTCAATTTTGCAAAAgtaggtatacttttaaatggaacaagatatctaattggaagcaagaaaaat : 1600
F58A6.4fm : AGAGATCTTTTATTCAATTTTGCAAAA----- :

F58A6.4 : tgtttaaacatttaggttgacagctgaaacccatcgtaaaaccattgatgataaaaaatccactttttaaatcagag : 898
g-II-1161 : tgtttaaacatttaggttgacagctgaaacccatcgtaaaaccattgatgataaaaaatccactttttaaatcagag : 1680
F58A6.4fm : ----- :

F58A6.4 : cttcagATATCCCCTCGTCAAATGTCGACTTACGGCTTGCCATATGATATTGGGAGCGTAATGCATTATACTCCAAGTGA : 978
g-II-1161 : cttcagATATCCCCTCGTCAAATGTCGACTTACGGCTTGCCATATGATATTGGGAGCGTAATGCATTATACTCCAAGTGA : 1760
F58A6.4fm : -----ATATCCCCTCGTCAAATGTCGACTTACGGCTTGCCATATGATATTGGGAGCGTAA----- :

F58A6.4 : gtgagtacaaaattgtttttattgaattttcaaatcacacttgagATTCTCAAACATCCCTTCCATACCAACGTTAGCT : 1058
g-II-1161 : gtgagtacaaaattgtttttattgaattttcaaatcacacttgagATTCTCAAACATCCCTTCCATACCAACGTTAGCT : 1840

F58A6.4 : GCAATCGACACAAATCTTCAGCAAACAATGGGTCAATTGGAAGGACCTTCTTTTGTGGATGTGCATATTATGAATCAACA : 1138
g-II-1161 : GCAATCGACACAAATCTTCAGCAAACAATGGGTCAATTGGAAGGACCTTCTTTTGTGGATGTGCATATTATGAATCAACA : 1920

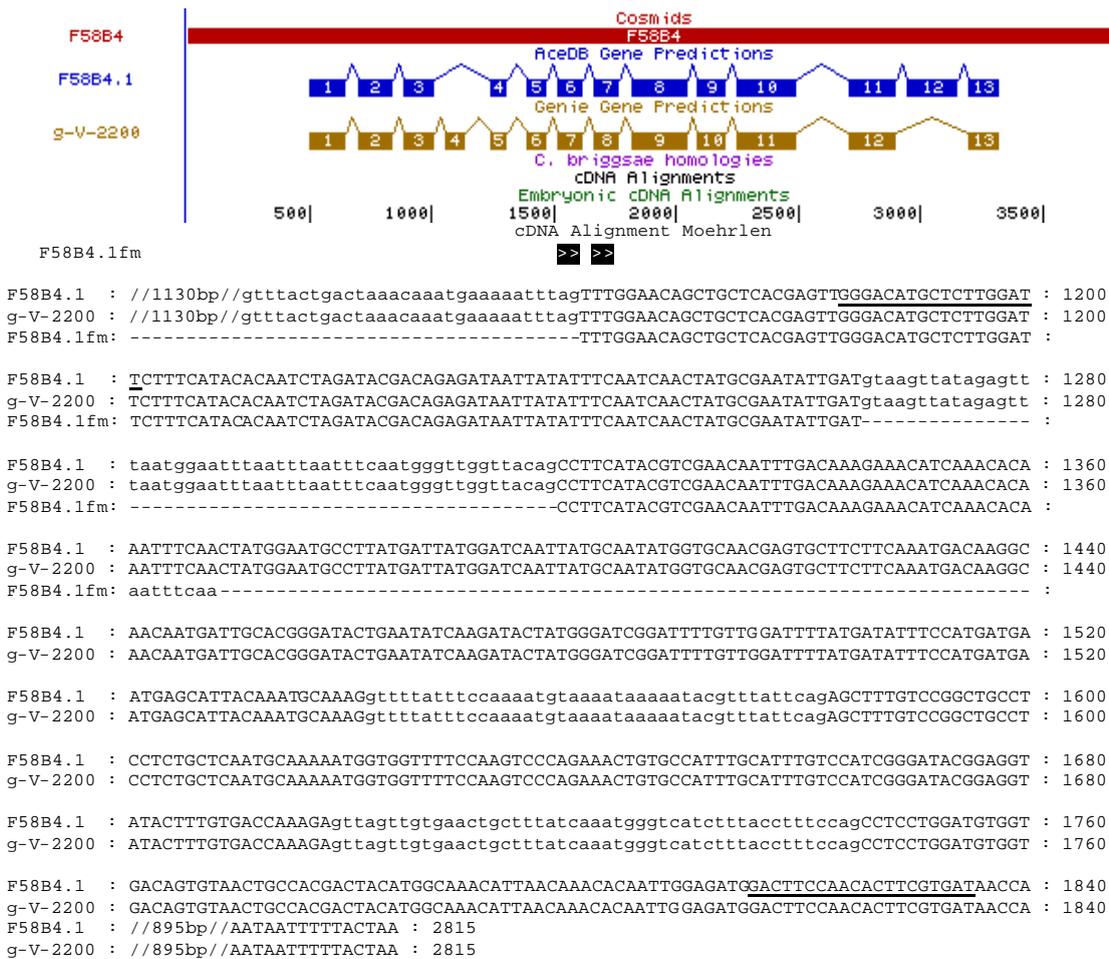
F58A6.4 : TTATCAGTGCCAAAGgtttgggtttttgttagattttgatttaaatgtaaatgtaattgtttttaatttcagAAAAGTGT : 1218
g-II-1161 : TTATCAGTGCCAAAGgtttgggtttttgttagattttgatttaaatgtaaatgtaattgtttttaatttcagAAAAGTGT : 2000

F58A6.4 : CCAACTCAAGCTCCTTGCCAAAATGGTGGCTTCACTAATCCCGTAAC TGCAAAGTTTGCAAATGTCCAACCTGGCTTTGG : 1298
g-II-1161 : CCAACTCAAGCTCCTTGCCAAAATGGTGGCTTCACTAATCCCGTAAC TGCAAAGTTTGCAAATGTCCAACCTGGCTTTGG : 2080

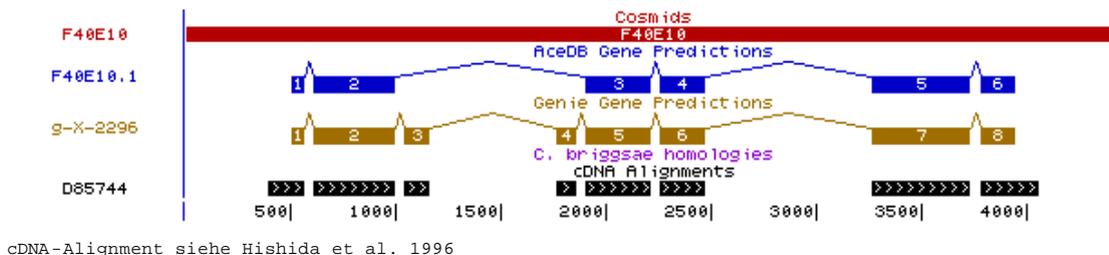
F58A6.4 : // 960bp //GATGGAGTTTTTGTGTCGACACTTCCCTACTTCAAAGGATAGGATTTAGAAGACAGTTATAG : 2332
g-II-1161 : // 960bp //GATGGAGTTTTTGTGTCGACACTTCCCTACTTCAAAGGATAGGATTTAGAAGACAGTTATAG : 3114

```

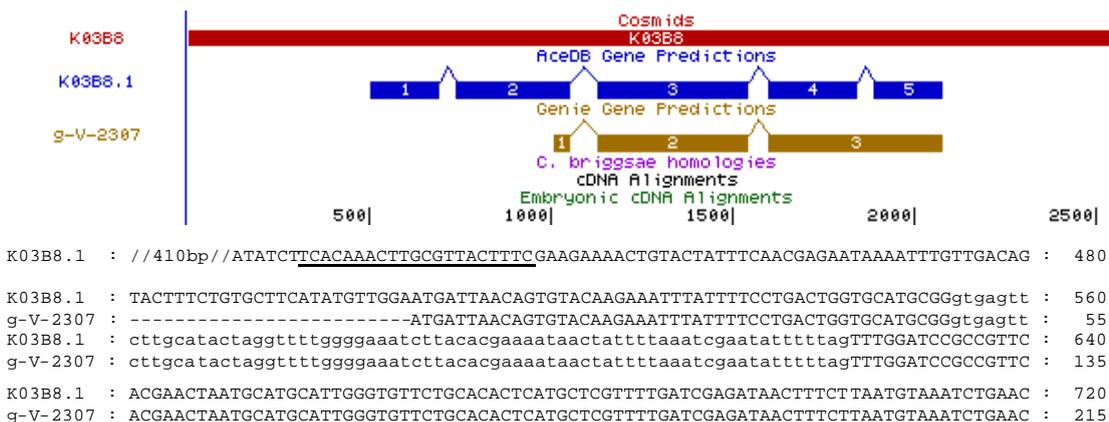
6.1.21. Sequenzdaten F58B4.1



6.1.22. Sequenzdaten F40E10.1 (HCH-1)



6.1.23. Sequenzdaten K03B8.1



```

K03B8.1 : AAAGATGATGAAGACGACAGCAATTTTGAGATGTTTCTCCTCCCTTTTCTATAAACGTTGTACCATATGAGTATGGAAG : 800
g-V-2307 : AAAGATGATGAAGACGACAGCAATTTTGAGATGTTTCTCCTCCCTTTTCTATAAACGTTGTACCATATGAGTATGGAAG : 295

K03B8.1 : CACTTTGCACTATACAGCCGATGTATCGGGAACAAATTCACCTTCTCCGAAACAAATGGAATATTATAGAACACTAGGAA : 880
g-V-2307 : CACTTTGCACTATACAGCCGATGTATCGGGAACAAATTCACCTTCTCCGAAACAAATGGAATATTATAGAACACTAGGAA : 375

K03B8.1 : ATCGAAGAGTTACATTTTATGATATGTTAACTATAAACACTGCATACAACCTGCAAAATGTCCAAGTGAATATTGTGTGCA : 960
g-V-2307 : ATCGAAGAGTTACATTTTATGATATGTTAACTATAAACACTGCATACAACCTGCAAAATGTCCAAGTGAATATTGTGTGCA : 455

K03B8.1 : AATGGAGGATATACAAATCCGTCCAATTTGTTGGAATGATTTGTCCATTGGGATATGGAGGAGTGTATGTGATCGAGT : 1040
g-V-2307 : AATGGAGGATATACAAATCCGTCCAATTTGTTGGAATGATTTGTCCATTGGGATATGGAGGAGTGTATGTGATCGAGT : 535

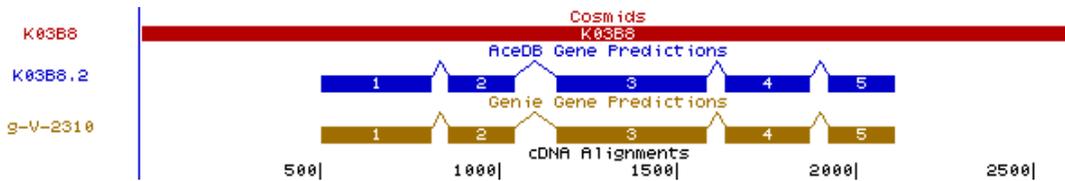
K03B8.1 : AgtttagtttattatctgaattagtttctatcggatTTTTTAAACTGtatttttcagGTTGCTTGTAGTGTCCAGTTAAG : 1120
g-V-2307 : AgtttagtttattatctgaattagtttctatcggatTTTTTAAACTGtatttttcagGTTGCTTGTAGTGTCCAGTTAAG : 615

K03B8.1 : TGCAGATTTCGTATTGGAAGGATCCTGGATTTCTGTTGGAAGTAGTGTTTTGAGAGATACTACAGACCCCGTTAAAGCGT : 1200
g-V-2307 : TGCAGATTTCGTATTGGAAGGATCCTGGATTTCTGTTGGAAGTAGTGTTTTGAGAGATACTACAGACCCCGTTAAAGCGT : 695

K03B8.1 : TTATCTCAATAAATGCTCCGAAAGATAAAATTTATGAAGTTAAAATTTGTGAAAATAGAAAATTTCTCGTGTGACTCTGGT : 1280
g-V-2307 : TTATCTCAATAAATGCTCCGAAAGATAAAATTTATGAAGTTAAAATTTGTGAAAATAGAAAATTTCTCGTGTGACTCTGGT : 775

K03B8.1 : // 240bp // TTGAGTAGCTACAATAAAACCACAAATGGATATGATAACTATGAATATTATGCCTAA : 1577
g-V-2307 : // 240bp // TTGAGTAGCTACAATAAAACCACAAATGGATATGATAACTATGAATATTATGCCTAA : 1072
    
```

6.1.24. Sequenzdaten K03B8.2



```

K03B8.2 : ATGCAAAAATATCAACTGTATAAAAACAATGGAAGATGTCAAAGAAAATCACTTCGACATGTTTCTTCGACCATCTACCTT : 80
g-V-2310 : ATGCAAAAATATCAACTGTATAAAAACAATGGAAGATGTCAAAGAAAATCACTTCGACATGTTTCTTCGACCATCTACCTT : 80

K03B8.2 : ACTTCTAACTTTATTCTTGGCATTGGTTGCTGGATCAGCGATCCGAAAAGATGTGACGAGTTTGATTCAAATAAAGGAA : 160
g-V-2310 : ACTTCTAACTTTATTCTTGGCATTGGTTGCTGGATCAGCGATCCGAAAAGATGTGACGAGTTTGATTCAAATAAAGGAA : 160

K03B8.2 : AAGATGGTATTGTAGATGGAGACATTATGCTTACTGAAGCTCAGCTTCGAATTCCTCAATGGAACCGCAAAACGTTCAAAA : 240
g-V-2310 : AAGATGGTATTGTAGATGGAGACATTATGCTTACTGAAGCTCAGCTTCGAATTCCTCAATGGAACCGCAAAACGTTCAAAA : 240

K03B8.2 : AGGCAGATCACAAAGATTGGAAAAAATGGCCAGATGCAAAAGTATTTTACTACGAAAATGAATTCAgtaagttaa : 320
g-V-2310 : AGGCAGATCACAAAGATTGGAAAAAATGGCCAGATGCAAAAGTATTTTACTACGAAAATGAATTCAgtaagttaa : 320

K03B8.2 : aaaaatttatattcaagttggtatgtaaaacttaattccagCAAGTCTTAAACGAGAATTGATGTCATATGCGATGGCCCA : 400
g-V-2310 : aaaaatttatattcaagttggtatgtaaaacttaattccagCAAGTCTTAAACGAGAATTGATGTCATATGCGATGGCCCA : 400

K03B8.2 : TATCAGTTCGAACACTTGTGTCAAATTCAGAAAAGTAATTCAGCAACGAACAGAATTCGTTTTACAAAATACAGGAGGAT : 480
g-V-2310 : TATCAGTTCGAACACTTGTGTCAAATTCAGAAAAGTAATTCAGCAACGAACAGAATTCGTTTTACAAAATACAGGAGGAT : 480

K03B8.2 : GTGCTTCCTATATTGGAATGAATGGAGGAGAACAACCTTGTGGTTCGGAGATGGATGCTTATCgtgagtttgagacga : 560
g-V-2310 : GTGCTTCCTATATTGGAATGAATGGAGGAGAACAACCTTGTGGTTCGGAGATGGATGCTTATCgtgagtttgagacga : 560

K03B8.2 : ctttgaaaaattagctgattttgagattagtttagagttcaacctgaaaaatatgtttagctgatgttacaacgaaccccc : 640
g-V-2310 : ctttgaaaaattagctgattttgagattagtttagagttcaacctgaaaaatatgtttagctgatgttacaacgaaccccc : 640

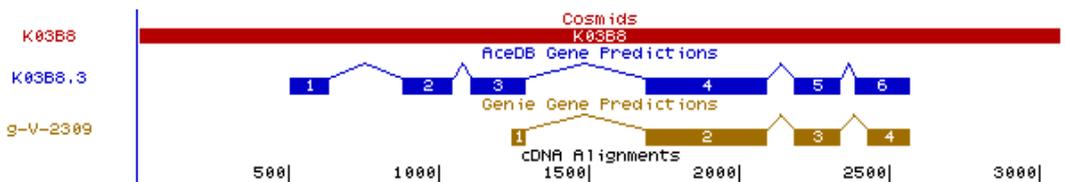
K03B8.2 : gtataaaataagaattaaatttagTTGGTACTGCTGTTACAGAAATCATGCACCTCTCTGGACTGTTCCACACGCACTC : 720
g-V-2310 : gtataaaataagaattaaatttagTTGGTACTGCTGTTACAGAAATCATGCACCTCTCTGGACTGTTCCACACGCACTC : 720

K03B8.2 : TCGTTTTGATCGTGATAACTTCTGTCCGTCAGCTACAAGAGTGTTCCTGAAAATATGGTTGGAACTTGGAAAAAGAAA : 800
g-V-2310 : TCGTTTTGATCGTGATAACTTCTGTCCGTCAGCTACAAGAGTGTTCCTGAAAATATGGTTGGAACTTGGAAAAAGAAA : 800

K03B8.2 : CTGAACAAACGACGCTACAATGCTGTACCTTTGGAATATGGAAGCACTATGTTGTACAGATACAACACCTTCGGAGAGGGA : 880
g-V-2310 : CTGAACAAACGACGCTACAATGCTGTACCTTTGGAATATGGAAGCACTATGTTGTACAGATACAACACCTTCGGAGAGGGA : 880

K03B8.2 : //720 bp// GTATTACGTCTAA : 1613
g-V-2310 : //720 bp// GTATTACGTCTAA : 1613
    
```

6.1.25. Sequenzdaten K03B8.3



```

K03B8.3 : ATGTGTGCATCGTGCTCCACAATGCCAATGTGCGAGTTTCTTAATGGAGACATCTCAAGACATCGCCCTTAACCTTTAT : 80
K03B8.3 : CTGCATAGGTAATAAATGAACATATTTTATAACGTAACATCATTAGATAAACAgtagcattattgatagattttaatacaatt : 160
K03B8.3 : tccttgaagatgaaccatattgttctctcttaacctcatatcagaatttaaattgaagttttggttttcagtaactag : 240
K03B8.3 : catggtcgtcaaaacgtttacaattaacgtatagaacaacacaattcaacttttggtgtaaaatatgtctcaaaagcaataa : 320
K03B8.3 : tatgttcgcaagaaattgattaatacatcttatgtagacactttggcagtttcagAATACTCTTTTACAATCCGAGCGC : 400
K03B8.3 : GCCTGTTCCGCACTCCAGTCACGCCCTATCCTCGCCAGGCGCATTTACCTCTTTTCGCGCCAGTCGCCACCCCTT : 480
K03B8.3 : CGCGCCCGAGGAGGCCAATACCAGAGTAGTCAAAAATTTATATTGCTTTGATTCTTCATTCAgtaagttctattat : 560
K03B8.3 : attttgaagtagtgaatacaacttataaacatgtacttttgagTACTGAAAAGCGAGAATTGATGCTTTTCTCCATGAA : 640
    
```

```

K03B8.3 : CTTTCATATCTTCCAAAACCTTGCCTGACTTTTGAAGAAAACCTGTTCAATTCTTACAGGGTCCAACTTGTAAACGGTAAAG : 720
K03B8.3 : GCTGTTTCATACATATTGGGATGAATAACATAGTACAACATCTAACATTTAACGATACGTGCATAGACGtgagttttct : 800
g-V-2309 : -----ATGAATAACATAGTACAACATCTAACATTTAACGATACGTGCATAGACGtgagttttct : 60

K03B8.3 : //320bp//tccaaacacggcacaccgaaagaatgatagacagatctcaatcacataaaactttcagTTTGGAACTGCTGT : 1200
g-V-2309 : //320bp//tccaaacacggcacaccgaaagaatgatagacagatctcaatcacataaaactttcagTTTGGAACTGCTGT : 460

K03B8.3 : TCATGAATTAATGCATGCATTAGGTGTTATTACACCCACTCTAGACTTGATCGAGATAACTTTTTAAACATTAATCTGA : 1280
g-V-2309 : TCATGAATTAATGCATGCATTAGGTGTTATTACACCCACTCTAGACTTGATCGAGATAACTTTTTAAACATTAATCTGA : 540

K03B8.3 : CAAATGTGTCCAAGGAATGATGCATAATTATGCAATTTTCAACCAATCTACAAACGTTGTACCTTATGAATACGGGAAGC : 1360
g-V-2309 : CAAATGTGTCCAAGGAATGATGCATAATTATGCAATTTTCAACCAATCTACAAACGTTGTACCTTATGAATACGGGAAGC : 620

K03B8.3 : ACTATGCATTACTATGCGAATATATCGACAATGTTTCCCAAGAAGTCAGAATACTCTGCCACACTTGGAAATAGGAAGAGT : 1440
g-V-2309 : ACTATGCATTACTATGCGAATATATCGACAATGTTTCCCAAGAAGTCAGAATACTCTGCCACACTTGGAAATAGGAAGAGT : 700

K03B8.3 : //560bp//ATGTGACTGAACGCTTGAGTGGGTACAATAAAACAACCAATGGATATGATAACTATGAGTATTATGATTAA : 2071
g-V-2309 : //560bp//ATGTGACTGAACGCTTGAGTGGGTACAATAAAACAACCAATGGATATGATAACTATGAGTATTATGATTAA : 1331

```

6.1.26. Sequenzdaten K03B8.5

```

K03B8.5fm
K03B8.5 : ATGTATCTAAAAATATTCAGACTACTATTGGAAATTTAAAGTTC AATTTTGAAAATATGCGTCTATTTTTCTTCGAAAAA : 80
K03B8.5 : CTACGCATATTATCCATTAGTTTTCGATTTCTGTTGATTACTTTGACGgtttattcat taattgctgttccaaatata : 160
K03B8.5 : tcatctctgaaatctccagaacagttttaaactcttataccatgcaactgtttgaagatttttaccactttttgaaac : 240
K03B8.5 : taatgataacgtctccatttaaatgttctctctgcttatgacacatgaaaaaatatgtagTTTTATGAAAAAT TCC : 320
K03B8.5 : CCACCGATTACATATTTCCCTCGATTTTTTCGTATAAGAAATAGGTTGAAAAACGGTATAAATGAACACATTGCTGGTGA : 400

K03B8.5 : AATATCATTACGATCTCGAGATGGTTCGGTTGATTCATCTTATTGGAGCTATTATTCTACTCTTTTCCTACGCCTACTG : 480
g-V-2311 : -----ATGGTTCGGTTGATTCATCTTATTGGAGCTATTATTCTACTCTTTTCCTACGCCTACTG : 59

K03B8.5 : CGGACTTTCTCGGCTCAATGAGCATGATATCGAAGAAAGTTACTCTCATAAACGAGTTAAAAGACAATTCGAACGCTTGG : 560
g-V-2311 : CGGACTTTCTCGGCTCAATGAGCATGATATCGAAGAAAGTTACTCTCATAAACGAGTTAAAAGACAATTCGAACGCTTGG : 139

K03B8.5 : GAACAAAATGGTCATATGGAGTAGTTAACTATTATTATGCGGACAAGAACAATGAAAATAAAGAGATGGTTGAAAGTGCG : 640
g-V-2311 : GAACAAAATGGTCATATGGAGTAGTTAACTATTATTATGCGGACAAGAACAATGAAAATAAAGAGATGGTTGAAAGTGCG : 219
K03B8.5fm: GAACAAAATGGTCATATGGAGTAGTTAACTATTATTATGCGGACAAGAACAATGAAAATAAAGAGATGGTTGAAAGTGCG : 640

K03B8.5 : ATTGCTTACATTGCAAATCACACTTGCATCAGGTTCAATGAGGATCAAATGCTGTACAGCGTGTGCAGATTCCGATGCA : 720
g-V-2311 : ATTGCTTACATTGCAAATCACACTTGCATCAGGTTCAATGAGGATCAAATGCTGTACAGCGTGTGCAGATTCCGATGCA : 299
K03B8.5fm: ATTGCTTACATTGCAAATCACACTTGCATCAGGTTCAATGAGGATCAAATGCTGTACAGCGTGTGCAGATTCCGATGCA : 720

K03B8.5 : ACAGtttgttttcatttaacgacttccctaaacacccgtgtattgcagAAACTGGCTTTGTCAGTCGACAGTCGGAGCACCA : 800
g-V-2311 : ACAGtttgttttcatttaacgacttccctaaacacccgtgtattgcagAAACTGGCTTTGTCAGTCGACAGTCGGAGCACCA : 379
K03B8.5fm: -----AAACTGGCTTTGTCAGTCGACAGTCGGAGCACCA : 379

K03B8.5 : GGTATGAGTATGAGTAAGCCGATTGGAGAGCTTCTATGTTGGTTCAA A TTTGGTATACCATCGGAAGTATCGTTCATGA : 880
g-V-2311 : GGTATGAGTATGAGTAAGCCGATTGGAGAGCTTCTATGTTGGTTCAA A TTTGGTATACCATCGGAAGTATCGTTCATGA : 459
K03B8.5fm: GGTATGAGTATGAGTAAGCCGATTGGAGAGCTTCTATGTTGGTTCAA A TTTGGTATACCATCGGAAGT ----- : 459

K03B8.5 : GTACTCAAGGACGACTTTACCATCAGCTGTTACATTCAAAATATCGCTTCACGAATGACAAAGTTGTGCTTGGTTAA : 1916
g-V-2311 : GTACTCAAGGACGACTTTACCATCAGCTGTTACATTCAAAATATCGCTTCACGAATGACAAAGTTGTGCTTGGTTAA : 1495

```

6.1.27. Sequenzdaten K04E7.3

```

K04E7.3fm
K04E7.3 : -----ATGGGTAACCAGCCTAAAACCTACA AATTTTCATCAAAGGTTTAAAGGGT : 48
g-X-945 : // 490bp//CTCCATGGCATGAGAAGCCTCCATGGgttaaccagcctaaaaactcaaaat tttcacaagggtttaaggGT : 560

K04E7.3 : CTCCCGCCACCCTCCTCCGCTGAGCCGGAACCAACAAGATCAGCCGCAAGTTATGTTTAGTCAGGATATAGACAA : 128
g-X-945 : CTCCCGCCACCCTCCTCCGCTGAGCCGGAACCAACAAGATCAGCCGCAAGTTATGTTTAGTCAGGATATAGACAA : 640
K04E7.3 : GgttagtttaatagaaaaaatatgcaatacgttagaattcagGTTGTCAATAGTGTGAACCAAGTACAGCTGCGTTCC : 208
g-X-945 : GgttagtttaatagaaaaaatatgcaatacgttagaattcagGTTGTCAATAGTGTGAACCAAGTACAGCTGCGTTCC : 720

K04E7.3 : AGCGCCCCGGAGAAAGTTATGATAAAGTTATCCAGATTATGAGTCTTATTTTAAATCGAAAATCAGGATCTCAATATgtg : 288
g-X-945 : AGCGCCCCGGAGAAAGTTATGATAAAGTTATCCAGATTATGAGTCTTATTTTAAATCGAAAATCAGGATCTCAATATgtg : 800

```

```

K04E7.3 : agtcattttttgaaatattttataaaaatttcctatgcagGATATAAA CACTGTAATCCCGTCCAGTGGGATTACAACA : 368
g-X-945 : agtcattttttgaaatattttataaaaatttcctatgcagGATATAAA CACTGTAATCCCGTCCAGTGGGATTACAACA : 880
K04E7.3fm : -----CACTGTAATCCCGTCCAGTGGGATTACAACA : 32

K04E7.3 : ATGAAATGGCTGCAAAATAGCAAAATGCTGCAGTTATGTTTGAAAGTGATATGgttagtttagaaggtttaaaaaatgct : 448
g-X-945 : ATGAAATGGCTGCAAAATAGCAAAATGCTGCAGTTATGTTTGAAAGTGATATGgttagtttagaaggtttaaaaaatgct : 960
K04E7.3fm : ATGAAATGGCTGCAAAATAGCAAAATGCTGCAGTTATGTTTGAAAGTGATATG----- : 85
K04E7.3 : gaataatttttttcagGCTTTAACGGTTTCACAAATGAATAAAGTAGCTCAAAACGGCTTTCGTGTGAAACGAAAAATG : 528
g-X-945 : gaataatttttttcagGCTTTAACGGTTTCACAAATGAATAAAGTAGCTCAAAACGGCTTTCGTGTGAAACGAAAAATG : 1040
K04E7.3fm : -----GCTTTAACGGTTTCACAAATGAATAAAGTAGCTCAAAACGGCTTTCGTGTGAAACGAAAAATG : 148

K04E7.3 : AATTTGAACGGAAACAACGTGGTCTCGAACATTCCTTATCGTTTTTTAGATACTGATGgtatgtacaaataatgtttato : 608
g-X-945 : AATTTGAACGGAAACAACGTGGTCTCGAACATTCCTTATCGTTTTTTAGATACTGATGgtatgtacaaataatgtttato : 1120
K04E7.3fm : AATTTGAACGGAAACAACGTGGTCTCGAACATTCCTTATCGTTTTTTAGATACTGATG----- : 206

K04E7.3 : tgaattatcaacatgtttcagGTAACGGCAATCTCAAATAACAAACGGTTTAAGACACTATGAAAGAAACACTTGCAT : 688
g-X-945 : tgaattatcaacatgtttcagGTAACGGCAATCTCAAATAACAAACGGTTTAAGACACTATGAAAGAAACACTTGCAT : 1200
K04E7.3fm : -----GTAACGGCAATCTCAAATAACAAACGGTTTAAGACACTATGAAAGAAACACTTGCAT : 264

K04E7.3 : ACGTTTTTCTCTCAATGGAGGTGGCTCGGATTACCTGGTTTTTTAGCAAAGGAGAAGGgtccgttttgttttagaaaaag : 768
g-X-945 : ACGTTTTTCTCTCAATGGAGGTGGCTCGGATTACCTGGTTTTTTAGCAAAGGAGAAGGgtccgttttgttttagaaaaag : 1280
K04E7.3fm : ACGTTTTTCTCTCAATGGAGGTGGCTCGGATTACCTGGTTTTTTAGCAAAGGAGAAGG----- : 321

K04E7.3 : aataacaacagcaaatattgattgatttttagATGCTATTCTAGTGTGGCAGATTGGGAGGCCCTCAGGAAATTTCTATT : 848
g-X-945 : aataacaacagcaaatattgattgatttttagATGCTATTCTAGTGTGGCAGATTGGGAGGCCCTCAGGAAATTTCTATT : 1360
K04E7.3fm : -----ATGCTATTCTAGTGTGGCAGATTGGGAGGCCCTCAGGAAATTTCTATT : 370

K04E7.3 : GGTGATGGATGTGAAACGgttagaaaattttttcatatatttactggaattacttttttaatagtggaatgggtgtaat : 928
g-X-945 : GGTGATGGATGTGAAACGgttagaaaattttttcatatatttactggaattacttttttaatagtggaatgggtgtaat : 1440
K04E7.3fm : GGTGATGGATGTGAAACG----- : 388

K04E7.3 : tttagCTCGGAATTATTACTCATGAAGTGGGTATGCTTTAGGATTTTGGCATGAACAAGCGAGACCTGAACGAGATTCA : 1008
g-X-945 : tttagCTCGGAATTATTACTCATGAAGTGGGTATGCTTTAGGATTTTGGCATGAACAAGCGAGACCTGAACGAGATTCA : 1520
K04E7.3fm : -----CTCGGAATTATTACTCATGAAGTGGGTATGCTTTAGGATTTTGGCATGAACAAGCGAGACCTGAACGAGATTCA : 463

K04E7.3 : TACGTTAGgtgagttttgaaaattgaaaaataaaacatgccagtttatcagAATCAATCGCCAAAATGCTATAAATGGTC : 1088
g-X-945 : TACGTTAGgtgagttttgaaaattgaaaaataaaacatgccagtttatcagAATCAATCGCCAAAATGCTATAAATGGTC : 1600
K04E7.3fm : TACGTTAG-----AATCAATCGCCAAAATGCTATAAATGGTC : 500

K04E7.3 : TTGAAGGCCAATTCGACAAGAGATCGTGGTCGGAAGTTAACGAATACAGCTTACCTTATGATTATGGATCGGTTATGCAT : 1168
g-X-945 : TTGAAGGCCAATTCGACAAGAGATCGTGGTCGGAAGTTAACGAATACAGCTTACCTTATGATTATGGATCGGTTATGCAT : 1680
K04E7.3fm : TTGAAGGCCAATTCGACAAGAGATCGTGGTCGGAAGTTAACGAATACAGCTTACCTTATGATTATGGATCGGTTATGCAT : 580

K04E7.3 : TACGGACCAAAAggtgtgttcttacatgttgttccgatgaatttaaccattttcagTCTTTTCCAAATCTTCAACAAT : 1248
g-X-945 : TACGGACCAAAAggtgtgttcttacatgttgttccgatgaatttaaccattttcagTCTTTTCCAAATCTTCAACAAT : 1760
K04E7.3fm : TACGGACCAAA-----TCTTTTCCAAATCTTCAACAAT : 615

K04E7.3 : GAACACCGTTGAACAGTAGATCCTGCATTCATTAACACTATTGGAAACAGAGTGAACCTTCATTCCTTGATCTTAAAGT : 1328
g-X-945 : GAACACCGTTGAACAGTAGATCCTGCATTCATTAACACTATTGGAAACAGAGTGAACCTTCATTCCTTGATCTTAAAGT : 1840
K04E7.3fm : GAACACCGTTGAACAGTAGATCCTGCATTCATTAACACTATTGGAAACAGAGTGAACCTTCATTCCTTGATCTTAAAGT : 695

K04E7.3 : TGCTCAATACGGCTTTCTGTCTAgtcagtttaacggaataattttccacgtttatcttcaacttcagACATTTGCACCAAT : 1408
g-X-945 : TGCTCAATACGGCTTTCTGTCTAgtcagtttaacggaataattttccacgtttatcttcaacttcagACATTTGCACCAAT : 1920
K04E7.3fm : TGCTCAATACGGCTTTCTGTCTA-----ACATTTGCACCAAT : 733

K04E7.3 : CGGATAAACTGCCAACACGGCGGTTACGCGGATCCCAACAATTGTGGGCAATGTACGTGTCCGACAGGACTGGAAGGAAC : 1488
g-X-945 : CGGATAAACTGCCAACACGGCGGTTACGCGGATCCCAACAATTGTGGGCAATGTACGTGTCCGACAGGACTGGAAGGAAC : 2000
K04E7.3fm : CGGATAAACTGCCAACACGGCGGTTACGCGGATCCCAACAATTGTGGGCAATGTACGTGTCCGACAGGACTGGAAGGAAC : 813

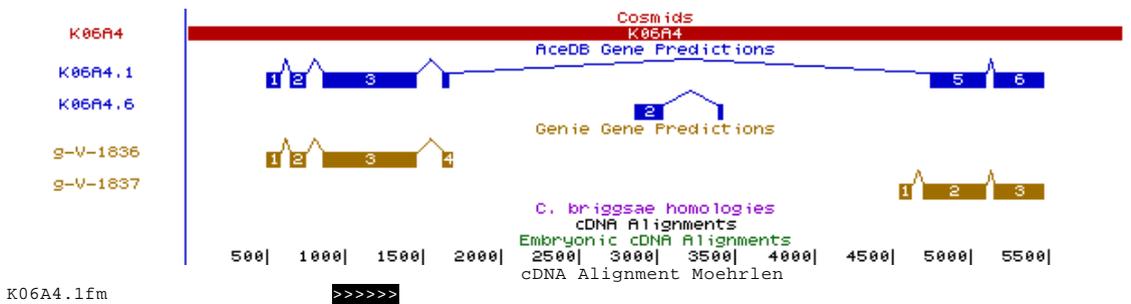
K04E7.3 : ATATTGTGAACGACTGCAAACTTCTAgtgagtcattttccatagaggattttcatcggtgaagaaattcagATTGCGGTG : 1568
g-X-945 : ATATTGTGAACGACTGCAAACTTCTAgtgagtcattttccatagaggattttcatcggtgaagaaattcagATTGCGGTG : 2080
K04E7.3fm : ATATTGTGAACGACTGCAAACTTCTA-----ATTGCGGTG : 848

K04E7.3 : TGGAACTTCCACGGGCAGATTACTCATGGAGAAATATTTCTATTCCGGGATCATCTGACTGTTATTGGAGAATTGTTTCG : 1648
g-X-945 : TGGAACTTCCACGGGCAGATTACTCATGGAGAAATATTTCTATTCCGGGATCATCTGACTGTTATTGGAGAATTGTTTCG : 2160
K04E7.3fm : TGGAACTTCCACGGGCAGATTACTCATGGAGAAATATTTCTATTCCGGGATCATCTGACT----- : 908

K04E7.3 : //970bp// AAGGAACCACCAGGGGAAAACGCAGTGTCTTCAAACCTCAGAGTAAAGAACTAG----- : 2670
g-X-945 : //970bp// AAGGAACCACCAGGGGAAAACGCAGTGTCTTCAAACCTCAGAGTAAAGAACTAG----- : 3200
g-X-945 : //1850bp// ATTAAGGAAGATGTACCTGTTATGCATTCAAACCTTGAATAA : 5099

```

6.1.28. Sequenzdaten K06A4.1



```

K06A4.1fm : >>>>>

K06A4.1 : ATGTACCGATTATTATTCTTCTCTTACTAGCATTGACAGCTTCAAAGTTTC CGAACCCTGAAAAAGACGATGAAAT : 80
g-V-1836 : ATGTACCGATTATTATTCTTCTCTTACTAGCATTGACAGCTTCAAAGTTTC CGAACCCTGAAAAAGACGATGAAAT : 80

K06A4.1 : CGCCGTGAAGgttagtggttcataaataagatttctccaattcaataaaatatttccagATTCCAACAAAACGG : 160
g-V-1836 : CGCCGTGAAGgttagtggttcataaataagatttctccaattcaataaaatatttccagATTCCAACAAAACGG : 160

K06A4.1 : TCAGTCTCCGAACCACCAAGGATGACGATATCGCCGTCAAGATCCCAATGAGGAAGAAACGGGAATTGCTATTACCC : 240
g-V-1836 : TCAGTCTCCGAACCACCAAGGATGACGATATCGCCGTCAAGATCCCAATGAGGAAGAAACGGGAATTGCTATTACCC : 240

```

```

K06A4.1 : ATGGCAATGGGAATCACAgtagtgatcagctctcatattgttcttcttcttactaaactttcctaaaaatcaacattcac : 320
g-V-1836 : ATGGCAATGGGAATCACAgtagtgatcagctctcatattgttcttcttcttactaaactttcctaaaaatcaacattcac : 320

K06A4.1 : tctacttctcgtattcctaaataacaaactgttctctttcagTCTCTGGCCAAACGCTGAAGTACCTTATGATATTGCTTC : 400
g-V-1836 : tctacttctcgtattcctaaataacaaactgttctctttcagTCTCTGGCCAAACGCTGAAGTACCTTATGATATTGCTTC : 400

K06A4.1 : TCACACTACTGCAACTGAAAGAGGAATAATTCTATCTGCAATGGAAGCATTGAGAGATGTCACCTTGTGTTCCGGTCCGTC : 480
g-V-1836 : TCACACTACTGCAACTGAAAGAGGAATAATTCTATCTGCAATGGAAGCATTGAGAGATGTCACCTTGTGTTCCGGTCCGTC : 480

K06A4.1 : CACGCTGTTCTACTGATAAAGCATTATCTCCAATAAACAAGCATTATCAGCTCGAACGATGCTTCTCCTATATTGGTTCGT : 560
g-V-1836 : CACGCTGTTCTACTGATAAAGCATTATCTCCAATAAACAAGCATTATCAGCTCGAACGATGCTTCTCCTATATTGGTTCGT : 560
K06A4.1fm: -----ATGCTTCTCCTATATTGGTTCGT :

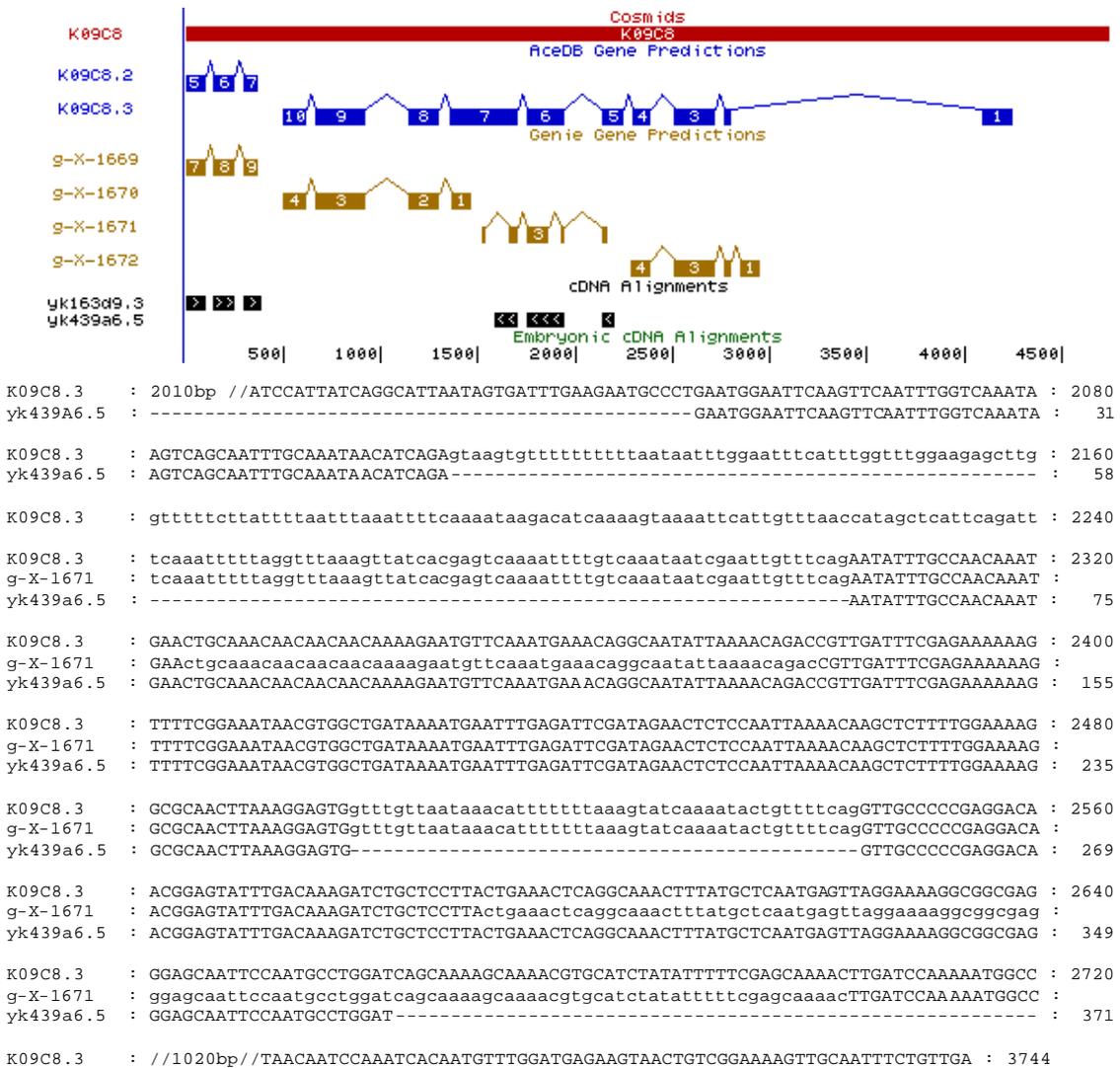
K06A4.1 : CAAAGCAGTCGCTGGCTCTTCGGAACCTCGTGATGGAAAAGTTGAAACCCGCATGAAGCTTGATCCTTCCTGCTTCTCTA : 640
g-V-1836 : CAAAGCAGTCGCTGGCTCTTCGGAACCTCGTGATGGAAAAGTTGAAACCCGCATGAAGCTTGATCCTTCCTGCTTCTCTA : 640
K06A4.1fm: CAAAGCAGTCGCTGGCTCTTCGGAACCTCGTGATGGAAAAGTTGAAACCCGCATGAAGCTTGATCCTTCCTGCTTCTCTA :

K06A4.1 : CAATGGACGGGAACTGTAATGCATGAGTTGATGCACATTCTCGGATTCTACCACGAGCATCAGAGAGACGATCGTGACC : 720
g-V-1836 : CAATGGACGGGAACTGTAATGCATGAGTTGATGCACATTCTCGGATTCTACCACGAGCATCAGAGAGACGATCGTGACC : 720
K06A4.1fm: CAATGGACGGGAACTGTAATGCATGAGTTGATGCACATTCTCGGATTCTACCACGAGCATCAGAGAGACGATCGTGACC :

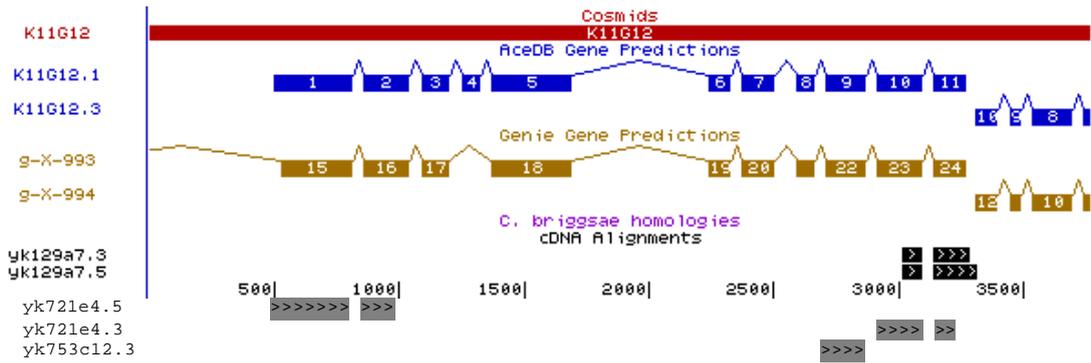
K06A4.1 : GCCGTATTGGAGGATCAGCATCTCATTACAACCTCAAGATCTATCAACGTGCCAAGTCGTACTACATGGGCGGTTACGAT : 800
g-V-1836 : GCCGTATTGGAGGATCAGCATCTCATTACAACCTCAAGATCTATCAACGTGCCAAGTCGTACTACATGGGCGGTTACGAT : 800
K06A4.1fm: GCCGTATTGGAGGATCAGCATCTCATTACAACCTCAAGATCTATCAACGTGCCAAGTC----- : 8

g-V-1836 : // 410bp//ATTCGAATTGTTGCAAAGAAGCAAATCGAGTCGAACAGTTTGTTCGTGAGAAGAGCTTCTCTAAa----- : 1199
K06A4.1 : // 4090 bp // : 4970
    
```

6.1.29. Sequenzdaten K09C8.3



6.1.30. Sequenzdaten K11G12.1



```

g-X-993      : ...// ab Intron 15.....aatttttaataagaaaactaataatctatggctctgtaataagctactga : 49
yk721e4     : TGCCTACTGGGTTTAAATTACCAAGTTG----- : 31

K11G12.1   : -----ATGCAGCAAATAAA : 14
g-X-993    : caatttttgcacctaggttctgattttattgtggtttttcagaaagtagtattgcatttaacactaatgcagcaataaa : 129
K11G12.1   : TTTTTCAGAAATGACACCATCTTTGGTGTTCCTCATTGTGGTGATAGTCGTTGTGGAGGGGCAGGGGTGGCGACCATGGG : 94
g-X-993    : ttttttcagAATGACACCATCTTTGGTGTTCCTCATTGTGGTGATAGTCGTTGTGGAGGGGCAGGGGTGGCGACCATGGG : 209
yk721e4    : -----AGAATGNACCATCTTTGGTGTTCCTCATTGTGGTGATAGTCGTTGTGGAGGGGCAGGGGTGGCGACCATGGG : 104

K11G12.1   : ATAGGTTCAATCATCTCGGAAATTCGGAAACTGGGGTGGAAATAATGGGGCACCCGTCAAAGGAATCAAGAACCACAT : 174
g-X-993    : ATAGGTTCAATCATCTCGGAAATTCGGAAACTGGGGTGGAAATAATGGGGCACCCGTCAAAGGAATCAAGAACCACAT : 289
yk721e4    : ATAGGTTCAATCATCTCGGAAATTCGGAAACTGGGGTGGAAATAATGGGGCACCCGTCAAAGGAATCAAGAACCACAT : 184

K11G12.1   : GATATTCACCTCCAGTTCCTCCCGCCAGGATTCGAGGAAATAATGATCGATTTGGAGGTAACATAATCAAAGTTGTGCGA : 254
g-X-993    : GATATTCACCTCCAGTTCCTCCCGCCAGGATTCGAGGAAATAATGATCGATTTGGAGGTAACATAATCAAAGTTGTGCGA : 369
yk721e4    : GATATTCACCTCCAGTTCCTCCCGCCAGGATTCGAGGAAATAATGATCGATTTGGAGGTAACATAATCAAAGTTGTGCGA : 264

K11G12.1   : AATCATCGACTTGGGTAAGCAAAAACCGCGGAGATATCCTATCGGATTTAAGgtacttgtagggttaaaaaataac : 334
g-X-993    : AATCATCGACTTGGGTAAGCAAAAACCGCGGAGATATCCTATCGGATTTAAGgtacttgtagggttaaaaaataac : 449
yk721e4    : AATCATCGACTTGGGTAAGCAAAAACCGCGGAGATATCCTATCGGATTTAAG----- : 319

K11G12.1   : attaataatagtttttcagGACGTGCACAAAAACACAGGCACCTTGGTCGAAAGAAATGGAAAGGCAAAGTAAACAAT : 414
g-X-993    : attaataatagtttttcagGACGTGCACAAAAACACAGGCACCTTGGTCGAAAGAAATGGAAAGGCAAAGTAAACAAT : 529
yk721e4    : -----GACGTGCACAAAAACACAGGCAC----- : 343

K11G12.1   : //1450bp//attcagAGGATCTTGACAAAAATGACGTTCCGGGCAGCAATTTACGAGATTGAAAAGAACACCTGCATAAG : 1934
g-X-993    : //1450bp//attcagAGGATCTTGACAAAAATGACGTTCCGGGCAGCAATTTACGAGATTGAAAAGAACACCTGCATAAG : 2049
yk753c12.3 : -----TTCGGGCAGCAATTTACGAGATTGAAAAGAACACCTGCATAAG : 43

K11G12.1   : ATTCAAAGAGCTCAGTTCCTCCCACTGGATCTCATATTTGTTTACTACAAAGTTGACAGTCCGACATTgtgagttcata : 2014
g-X-993    : ATTCAAAGAGCTCAGTTCCTCCCACTGGATCTCATATTTGTTTACTACAAAGTTGACAGTCCGACATTgtgagttcata : 2129
yk753c12.3 : ATTCAAAGAGCTCAGTTCCTCCCACTGGATCTCATATTTGTTTACTACAAAGTTGACAGTCCGACATT----- : 112

K11G12.1   : ttacacagaggtgttcttatcgagaggtcaaaactaaaataagtagaacggttgtttaaacttataacgtctgattcagC : 2094
g-X-993    : ttacacagaggtgttcttatcgagaggtcaaaactaaaataagtagaacggttgtttaaacttataacgtctgattcagC : 2209

K11G12.1   : TGTGGTCTTTTCGTATGTTGGTCGAGCTGACCTGCCAACCCAGTTTATCTTAGTTTTGGATGTGACAAATgtaagaatct : 2174
g-X-993    : TGTGGTCTTTTCGTATGTTGGTCGAGCTGACCTGCCAACCCAGTTTATCTTAGTTTTGGATGTGACAAATgtaagaatct : 2289
yk753c12.3 : TGTGG----- : 118

K11G12.1   : ttcattgctattttcccaattgaggaatttcagAACAAAGGAGTTGCAATTCATGAAACAATGCATGCGCTCGGAGTT : 2254
g-X-993    : ttcattgctattttcccaattgaggaatttcagAACAAAGGAGTTGCAATTCATGAAACAATGCATGCGCTCGGAGTT : 2369

K11G12.1   : GCCCATCAGCACCTTCGCAATGATCGTGATCAATTTATTACTATTAATTGGAGTAACATTGATCCACAACAATACGATGC : 2334
g-X-993    : GCCCATCAGCACCTTCGCAATGATCGTGATCAATTTATTACTATTAATTGGAGTAACATTGATCCACAACAATACGATGC : 2449

K11G12.1   : ATTTGTTGTTGGTGGATTCAAACCTGTACACTTCgtaagtgtttagtaatccataggttcaaaacaaaacttctagATAC : 2414
g-X-993    : ATTTGTTGTTGGTGGATTCAAACCTGTACACTTCgtaagtgtttagtaatccataggttcaaaacaaaacttctagATAC : 2529
yk721e4    : -----AACTGTACACTTC-----ATAC : 360

K11G12.1   : GGAGTGAATACGCATACGACAGTATCATGCATTACAACGGTTACTGTGCTCAAAAATTGCGATTCTACTATGAA : 2494
g-X-993    : GGAGTGAATACGCATACGACAGTATCATGCATTACAACGGTTACTGTGCTCAAAAATTGCGATTCTACTATGAA : 2609
yk721e4    : GGAGTGAATACGCATACGACAGTATCATGCATTACAACGGTTACTGTGCTCAAAAATTGCGATTCTACTATGAA : 440

K11G12.1   : CCCAAAAACAACCTCTGCCGTGAATCTTAAAGTCTTGGACAAAGGCAAAAGATGGGGACAACCGATATTGAACCTTCTGA : 2574
g-X-993    : CCCAAAAACAACCTCTGCCGTGAATCTTAAAGTCTTGGACAAAGGCAAAAGATGGGGACAACCGATATTGAACCTTCTGA : 2689
yk721e4    : CCCAAAAACAACCTCTGCCGTGAATCTTAAAGTCTTGGACAAAGGCAAAAGATGGGGACAACCGATATTGAACCTTCTGA : 520
yk129a7    : -----GTGAATCTTAAAGTCTTGGACAAAGGCAAAAGATGGGGACAACCGATATTGAACCTTCTGA : 61

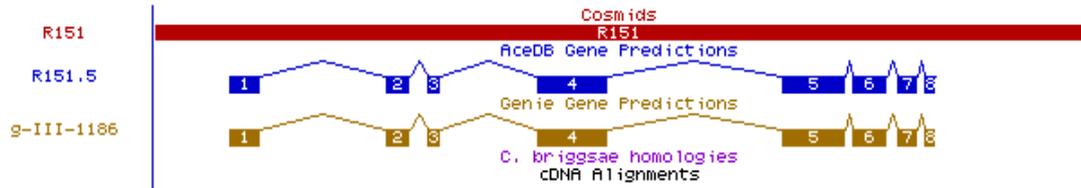
K11G12.1   : AAAAAATGTATTGTCAACCTGgtgagtggaatgtgtgttaataactgaaaatcaaaatttcagGATGCGATGACAAAA : 2654
g-X-993    : AAAAAATGTATTGTCAACCTGgtgagtggaatgtgtgttaataactgaaaatcaaaatttcagGATGCGATGACAAAA : 2769
yk721e4    : AAAAAATGTATTGTCAACCTG-----GATGCGATGACAAAA : 556
yk129a7    : AAAAAATGTATTGTCAACCTG-----GATGCGATGACAAAA : 97

K11G12.1   : ATGTTTATTGCGGAGCATGGGCTCTCAAAGACTTGTGCAAAAAATCCGGGACATGATCAATACATGGCTGCCAATTGTAAG : 2734
g-X-993    : ATGTTTATTGCGGAGCATGGGCTCTCAAAGACTTGTGCAAAAAATCCGGGACATGATCAATACATGGCTGCCAATTGTAAG : 2849
yk721e4    : ATGTTTATTGCGGAGCATGGGCTCTCAAAGACTTGTGCAAAAAATCCGGGACATGATCAATACATGGCTGCCAATTGTAAG : 636
yk129a7    : ATGTTTATTGCGGAGCATGGGCTCTCAAAGACTTGTGCAAAAAATCCGGGACATGATCAATACATGGCTGCCAATTGTAAG : 160

K11G12.1   : AAGAGTTGTGGATTGTGTGCTATTGGAAAAATA----- : 2767
g-X-993    : AAGAGTTGTGGATTGTGTGCTATTGGAAAAATA----- : 2885
yk721e4    : AAGAGTT----- : 643

```

6.1.31. Sequenzdaten R151.5



full lenght cDNA, besethend aus:

yk709e2.5, yk820f06.5, yk702c10.5, yk730d8.5, yk462f12.5, yk168f6.5, yk456f9.5,
 yk728d4.5, yk146b10.5, yk189d3.5, yk101h3.5, yk820f06.3, yk168f6.3, yk728d4.3,
 yk462f12.3, yk189d3.3, yk730d8.3, yk709e2.3, yk146b10.3, yk456f9.3, yk702c10.3, yk101h3.3

```

R151.5/g-III-1186 TGAGGAGGACCCTCCCTCTTCAGCTCCAGGTGTAGAAAACGGAGAATGATGTCAGAGGAAGATCAGAAAACAGTTGACT
YK820F06.5 TGAGGAGGACCCTCCCTCTTCAGCTCCAGGTGTAGAAAACGGAGAATGATGTCAGAGGAAGATCAGAAAACAGTTGACT

R151.5/g-III-1186 ATTATATGGATAAACTGAATAAATTGGCTGATGAGAAGCACCCAGAGgtaaaatTTTTct // Intron 820 bp //
YK820F06.5 ATTATATGGATAAACTGAATAAATTGGCTGATGAGAAGCACCCAGAG-----

R151.5/g-III-1186 ccctcgtcatataaatatTTTTcagGAAATCGAACGACACAAAGAATCCAGAGTTGGTAGCATGGGATCGTAAGAGAGATAG
YK820F06.5 -----GAAATCGAACGACACAAAGAATCCAGAGTTGGTAGCATGGGATCGTAAGAGAGATAG

R151.5/g-III-1186 TGTGCTGAATCCAGAAGAGCAGGGAAAAATCTTTCAAGGAGATATTGTTCTCTATCCAGAGCAAGCAAAGGCTCTTTATG
YK820F06.5 TGTGCTGAATCCAGAAGAGCAGGGAAAAATCTTTCAAGGAGATATTGTTCTCTATCCAGAGCAAGCAAAGGCTCTTTATG

R151.5/g-III-1186 AGCAAGCACTAACTGAAGgttgggtataatattagcgcaccacatttcatttttaaaatgtaccacaaaatttgaaattta
YK820F06.5 AGCAAGCACTAACTGAAG-----

R151.5/g-III-1186 aatTTTTtaagaggcctTTTTttggttgcaaaaaatattaaagaatgaatttcagGAAAAACAGAGTAAAGAGAAAAAT
YK820F06.5 -----GAAAAACAGAGTAAAGAGAAAAAT

R151.5/g-III-1186 CATTGGGTCTAACTTGAGAAGATGGGACGCCAGTCGACCGATAATTTATGCTTTTGATGGAAGTCACAgtgagttttcaa
YK820F06.5 CATTGGGTCTAACTTGAGAAGATGGGACGCCAGTCGACCGATAAATTTATGCTTTTGATGGAAGTCACA-----

R151.5/g-III-1186 // Intron 560bp // agatgtcatcaagtgtggattaagtctaaacgcacctattgtagtataacatgacggaat
YK820F06.5 -----

R151.5/g-III-1186 gCCCAACGCGAACAAAGAATCATCGAACTTGCACTTGAACACTGGCACAATATAACTTGTCTCAACTTTCAAAGAAACGA
YK820F06.5 -CCCAACGCGAACAAAGAATCATCGAACTTGCACTTGAACACTGGCACAATATAACTTGTCTC-----
YK728D4.5 -CCCAACGCGAACAAAGAATCATCGAACTTGCACTTGAACACTGGCACAATATAACTTGTCTC-----

R151.5/g-III-1186 TCAAGCTAACTCTGGAATTCGAATTTGATTTCACTGATGTTGATGGATGTCATCAAATGTTGGGAGACATCCACTCGGAG
YK728D4.5 TCAAGCTAACTCTGGAATTCGAATTTGATTTCACTGATGTTGATGGATGTCATCAAATGTTGGGAGACATCCACTCGG--
YK146B10.5 -----TGATGTTGNTGGATGTCATCAAATGTTGGGAGACATCCACTCGGAG

R151.5/g-III-1186 AAGAACAACCTTGTCTCACTTGACACCTGAATGTATTTCGATTAGGAGTTATTGCTCATGAAGTTGCACATGCTCTGGCTTC
YK146B10.5 AAGAACAACCTTGTCTCACTTGACACCTGAATGTATTTCGATTAGGAGTTATTGCTCATGAAGTTGCACATGCTCTGGNTTC

R151.5/g-III-1186 TGGCATGAACAATCGAGACCCGATAGAGATCAATATGTAACAGTTAGATGGGAAAAATTTGATAAAGATTCAAAGGGGCA
YK146B10.5 TGGCATGAACAATCGAGACCCGATAGAGATCAATATGTAACAGTTAGATGGGAAAAATTTGATAAAGATTCAAAGGGGCA

R151.5/g-III-1186 ATTCTTGAAAGAAGATCCTGATGATGTTGATAATGCTGGAGTACCATATGATTATGGAAGTATCATGCAATACCGGTCCA
YK146B10.5 ATTCTTGAAAGAAGATCCTGATGATGTTGATAATGCTGGAGTACCATATGATTATGGAAGTATCATGCAATACCGGTCCA

R151.5/g-III-1186 AGGCATTTCTCAAAGTTTGTATGATTTGTATACAACTAGTACCTATGTCACGGATTATCAAAAACTATTGgtaagtctatt
YK146B10.5 AGGNATTTCTCAAAGTTTGTATGATTTGTATACAACTAGTACCTATGTCACGGATTATCAAAAACTATTGGTCA-----
YK189D3.5 AGGCAT-CTCAAAGTTTGTATGATTTGTATACAACTAGTACCTATGTCACGGATTATCAAAAACTATTG-----

R151.5/g-III-1186 // Intron 1140bp // ataggaacttgctaaaaataatcttccagGTCAACGTGATCAGTTATCTTTTAAACGAC
YK189D3.5 -----GTCAACGTGATCAGTTATCTTTTAAACGAC
YK101H3.5 -----TAAACGAC

R151.5/g-III-1186 ATCCGACTAATGAATAAAATCTACTGCTCCGAGTCTGCCATCAAACTTCCATGTCAACGTGGAGGATACACGGATCC
YK189D3.5 ATCCGACTAATGAATAAAATCTACTGCTCCGAGTCTGCCATCAAACTTCCATGTCAACGTGGAGGATACACGGATCC
YK101H3.5 ATCCGACTAATGAATAAAATCTACTGCTCCGAGTCTGCCATCAAACTTCCATGTCAACGTGGAGGATACACGGATCC

R151.5/g-III-1186 ACGTCGTTGTGATCGTTTGTGATGTCCTGATGGATTACCCGGCAACTACTGCGAACAAAGTGTGCTGGATACGGAGCCA
YK189D3.5 ACGTCGTTGTGATCGTTTGTGATGTCCTGATGGATTACCCGGCAACTACTGCGAACAAAGTGTGCTGGATACGGAGCCA
YK101H3.5 ACGTCGTTGTGATCGTTTGTGATGTCCTGATGGATTACCCGGCAACTACTGCGAACAAAGTGTGCTGGATACGGAGCCA

R151.5/g-III-1186 CTTGTGGTGGAAAAATCAGTTTAACTAGATCTACAAC TAGAATATCATCACCTGGATATCCAAGAGAATTCAAAGAAGGA
YK189D3.5 CTTGTGGTGGAAAAATCAGTTTAACTAGATCTACAAC TAGAATATCATCACCTGGATATCCAAGAGAATTCAAAGAAGG-
YK101H3.5 CTTGTGGTGGAAAAATCAGTTTAACTAGATCTACAAC TAGAATATCATCACCTGGATATCCAAGAGAATTCAAAGAAGG-

R151.5/g-III-1186 CAAGAATGCTCATGGCTTCTCGTTGCTCCACCAGGTCATATTGTAGAATTTTCAGTTTATTGGAGAATTTGAGATGTACTG
YK101H3.5 CAAGAATGCTCATGGNTTCTCGTNGCTCCACCAGGTCATATTGTAGAATTTTCAGTTTATTGGAGAATTTGAGATGTACTG

R151.5/g-III-1186 CAAAATTAGACATTTCTTGTGCATGGACTACGTAGAAGTTAGAAACTCTACTGATTTTGCAAATACTGGAATGAGgtata
YK101H3.5 CAAAATTAGACATTTCTTGTGCATGGACTACGT- -----GTTTTTGCAAATACTGGAATGAG----
YK820F06.3 -----GTTTTTGCAAATACTGGAATGAG----

R151.5/g-III-1186 ttttgtaattgcatagtaattacctaacattatgataatttccagATATTGTTGCTATGGAACCTCTCCAAC TAGAATCC
YK820F06.3 -----ATATTGTTGCTATGGAACCTCTCCAAC TAGAATCC

R151.5/g-III-1186 GTTCTGCAACTACTGATATGGTTGACTTTTCCGAAGCTTCTACCGTGGTGGTAAAGGCTTTGAAGCAAGAGCCCGTGCA
YK820F06.3 GTTCTGCAACTACTGATATGGTTGACTTTTCCGAAGCTTCTACCGTGGTGGTAAAGGCTTTGAAGCAAGAGCCCGTGCA

R151.5/g-III-1186 GTACCAGAGGCGGGTAACTGGAATTCCTGGAGCCCTGGACGGCATGTTCCGCAACTTGTGGAGCATGCGGAAGTGCAT
YK820F06.3 GTACCAGAGGCGGGTAACTGGAATTCCTGGAGCCCTGGACGGCATGTTCCGCAACTTGTGGAGCATGCGGAAGTGCAT

R151.5/g-III-1186 GAGAACC CGAACATGTCCACCTGGAATGCATGCTCgtacgtttcccaaaactaatcaattggactttatataacatag
YK820F06.3 GAGAACC CGAACATGTCCACCTGGAATGCATGCTC-----

R151.5/g-III-1186 ctaatgaaaatataaaaataactgtaaatTTTTcagTGGGGAGCCAGTAGAAAACACAAAATTTGCAATACTCAAGCTTGTAC
YK820F06.3 -----TGGGGAGCCAGTAGAAAACACAAAATTTGCAATACTCAAGCTTGTAC
    
```

```

R151.5/g-III-1186 GGGTATGTGCGCTCAGAAACGGGAGGAAGAGGGTCAATGTGGAGGTTTCCTTTCTTTGTTACGAGCGCTCCGTTGTAGAC
YK820F06.3 GGGTATGTGCGCTCAGAAACGGGAGGAAGAGGGTCAATGTGGAGGTTTCCTTTCTTTGTTACGAGCGCTCCGTTGTAGAC

R151.5/g-III-1186 AgtaagtctctcaatcaacccaaactttgaattttaaatttcattttttcagAGAGAAGACGGTGATGGCTCCATGCGAG
YK820F06.3 A-----AGAGAAGACGGTGATGGCTCCATGCGAG

R151.5/g-III-1186 AACGCGTGCTGTCCAGGTTTCACTCTTACGCGTGGCAGATGTGTCCGGTGATcaccacaaatttgatccattattttcaattc
YK820F06.3 AACGCGTGCTGTCCAGGTTTCACTCTTACGCGTGGCAGATGTGTCCGGTGATCACCAAATTTGATCCATTATTTCAATTC

R151.5/g-III-1186 catatttttatcctgtaacctacagaactcttacgaagctacgggccaattcctaatacaacccaaattggtataatcctta
YK820F06.3 CATATTTTATCTGTAACTACAGAACTCTTACGAAGCTACGGCCAATTCCTAATCAAACC AAATGTTATAATCCTTA

R151.5/g-III-1186 ttttcaactgtttcattttttcttctctaaataaacacacctctttttatggtgttttcagagtttggtgtgatgattgaa
YK820F06.3 TTTTCACTGTTTCATTTTCTTCTTCTAAATTAATAACACCTCTTTTATGTTGTTTCAGTGTGTGTA-----

```

6.1.32. Sequenzdaten T02B11.7



```

T02B11.7 : //400bp// GACTGCCGACTTCGGGTGCAATCCGATGCCCGTCTTCAGCTTTAATCGGTACTACAAAACACGCTACACA : 3705
g-V-166 : //400bp// GACTGCCGACTTCGGGTGCAATCCGATGCCCGTCTTCAGCTTTAATCGGTACTACAAAACACGCTACACA : 3840
yk726g11.5 : -----CTTTAATCGGTACTACAAAACACGCTACACA : 366

T02B11.7 : TGGTCATATCGATACGgtaagcgctaaattaggcctgcctgcctattttctccgaaaaaccgtaggggtaggcacaaggc : 3785
g-V-166 : TGGTCATATCGATACGgtaagcgctaaattaggcctgcctgcctattttctccgaaaaaccgtaggggtaggcacaaggc : 3920
yk726g11.5 : TGGTCATATCGATACG----- : 382

T02B11.7 : aggccgcagtgctcactctatttgctaattccataaatttcagTGGACTCTAACGTCACCGCATGTGCTGATACATCCGACA : 3865
g-V-166 : aggccgcagtgctcactctatttgctaattccataaatttcagTGGACTCTAACGTCACCGCATGTGCTGATACATCCGACA : 4000
yk726g11.5 : -----TGGACTCTAACGTCACCGCATGTGCTGATACATCCGACA : 421

T02B11.7 : AAGCCACGTGCTCTTTCTTTGAAAAGTGCAAAGAGCAGGGTGTGCTCAATTTATGACTGCTCAGCTGAAGGTGATGTGT : 3945
g-V-166 : AAGCCACGTGCTCTTTCTTTGAAAAGTGCAAAGAGCAGGGTGTGCTCAATTTATGACTGCTCAGCTGAAGGTGATGTGT : 4080
yk726g11.5 : AAGCCACGTGCTCTTTCTTTGAAAAGTGCAAAGAGCAGGGTGTGCTCAATTTATGACTGCTCAGCTGAAGGTGATGTGT : 501

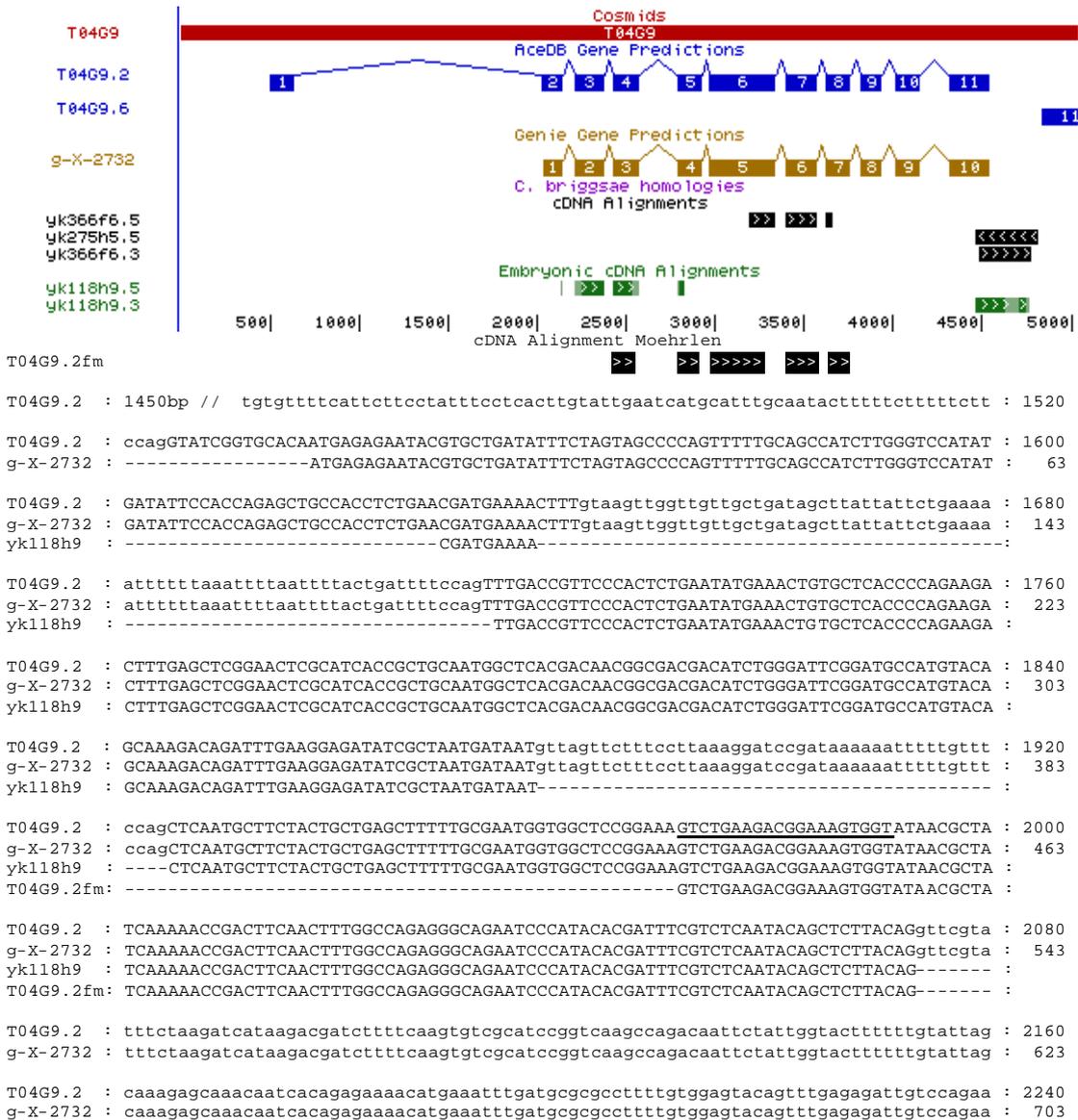
T02B11.7 : GCGGCGACTATGGATTGTGTGGGAAGGTGGCGTCAGATGATGGGACATgtaaggacaggtttcccaagagccagtgagt : 4025
g-V-166 : GCGGCGACTATGGATTGTGTGGGAAGGTGGCGTCAGATGATGGGACATGTAAGGACAGGTTTCCCAAGAGCCAgtgagt : 4160
yk726g11.5 : GCGGCGACTATGGATTGTGTGGGAAGGTGGCGTCAGATGATGGGACATGTAAGGACAGGTTTCCCAAGAGCCA----- : 575

T02B11.7 : tttggggaggtcatggaagaggagtgacgtcacaagtgctcctagatttttaaaatcttcgaaatttcaccttcagATGCT : 4105
g-V-166 : tttggggaggtcatggaagaggagtgacgtcacaagtgctcctagatttttaaaatcttcgaaatttcaccttcagATGCT : 4240
yk726g11.5 : -----ATGCT : 580

T02B11.7 : CTACCTACTCCACGAACGGAATGTGCACCAACAACCGCCACTTGCCGCCGAATTCTCATGCGCTGAAACTTGCGGTTTC : 4185
g-V-166 : CTACCTACTCCACGAACGGAATGTGCACCAACAACCGCCACTTGCCGCCGAATTCTCATGCGCTGAAACTTGCGGTTTC : 4320
yk726g11.5 : CTACCTACTCCACGAACGGAATGTGCACCAACAACCGCCACTTGCCGCCGA----- : 633

T02B11.7 : TGTACTAATCCGGTTTGAAATAAAATATTTTGTGTTCTGTTACTGTTCTTTTGACCAAGTTGAACTTTGATCATTTTACAT : 4265
g-V-166 : TGTACTAATCCGGTTTGA----- : 4338
    
```

6.1.33. Sequenzdaten T04G9.2



```

T04G9.2 : acctgaaaaatcaacaacaagaatcaacaatttattttttgagTCGTTCTCTGATAGCTGCGTCTATGCAAGAGTATGC : 2320
g-X-2732 : acctgaaaaatcaacaacaagaatcaacaatttattttttgagTCGTTCTCTGATAGCTGCGTCTATGCAAGAGTATGC : 783
yk118h9 : -----CGNTCTCTGATAGCTGCGTCTATGCAAGAGTATGC :
T04G9.2fm: -----TCGTTCTCTGATAGCTGCGTCTATGCAAGAGTATGC :

T04G9.2 : ATCACATACTTGCATCAGATGGGTTCCAAAAGAGGCAGCTGATGTCAACTACGTCCACATCTACCCAGACAGAGGATGTT : 2400
g-X-2732 : ATCACATACTTGCATCAGATGGGTTCCAAAAGAGGCAGCTGATGTCAACTACGTCCACATCTACCCAGACAGAGGATGTT : 863
yk118h9 : ATCACATACTTG----- :
T04G9.2fm: ATCACATACTTGCATCAGATGGGTTCCAAAAGAGGCAGCTGATGTCAACTACGTCCACATCTACCCAGACAGAGGATGTT :

T04G9.2 : ACTCGATGGTGGGAAAGATGGGtatgaaactagaaatattttcttagcgactaatcacaactgttaagGAGGAAAGCAA : 2480
g-X-2732 : ACTCGATGGTGGGAAAGATGGGtatgaaactagaaatattttcttagcgactaatcacaactgttaagGAGGAAAGCAA : 943
T04G9.2fm: ACTCGATGGTGGGAAAGATGG-----GAGGAAAGCAA :

T04G9.2 : TCGCTATCACTTGGTTCTGGTTGTATTCAAAAAGGAATCATTCTTCCAGAGTTGATGCACGCCGTTGGTTTTTCCACGA : 2560
g-X-2732 : TCGCTATCACTTGGTTCTGGTTGTATTCAAAAAGGAATCATTCTTCCAGAGTTGATGCACGCCGTTGGTTTTTCCACGA : 1023
T04G9.2fm: TCGCTATCACTTGGTTCTGGTTGTATTCAAAAAGGAATCATTCTTCCAGAGTTGATGCACGCCGTTGGTTTTTCCACGA :

T04G9.2 : ACAATCCCAGACTGATCGTGACGATCATATCACAAATTATGTGGAACAATATTCAAGCAGGAATGCAGGGACAGTTCGAGA : 2640
g-X-2732 : ACAATCCCAGACTGATCGTGACGATCATATCACAAATTATGTGGAACAATATTCAAGCAGGAATGCAGGGACAGTTCGAGA : 1103
T04G9.2fm: ACAATCCCAGACTGATCGTGACGATCATATCACAAATTATGTGGAACAATATTCAAGCAGGAATGCAGGGACAGTTCGAGA :

T04G9.2 : AGTACGGCCACGGGACTATTCAGTCACTGGGTACTGGTTATGATTACGGAAGTATTATGCATTATGGAACAAAGGCTTTC : 2720
g-X-2732 : AGTACGGCCACGGGACTATTCAGTCACTGGGTACTGGTTATGATTACGGAAGTATTATGCATTATGGAACAAAGGCTTTC : 1183
yk366f6.5 : -----TATTATGCATTATGGAACAAAGGCTTTC :
T04G9.2fm: AGTACGGCCACGGGACTATTCAGTCACTGGGTACTGGTTATGATTACGGAAGTATTATGCATTATGGAACAAAGGCTTTC :

T04G9.2 : TCGAGAAATGGACAACCAACTATGATTCCTAAGAAGAATGGAGCCACTATTGGACAGAGAAATGGGTTCTCAAAAGTTGA : 2800
g-X-2732 : TCGAGAAATGGACAACCAACTATGATTCCTAAGAAGAATGGAGCCACTATTGGACAGAGAAATGGGTTCTCAAAAGTTGA : 1263
yk366f6.5 : TCGAGAAATGGACAACCAACTATGATTCCTAAGAAGAATGGAGCCACTATTGGACAGAGAAATGGGTTCTCAAAAGTTGA :
T04G9.2fm: TCGAGAAATGGACAACCAACTATGATTCCTAAGAAGAATGGAGCCACTATTGGACAGAGAAATGGGTTCTCAAAAGTTGA :

T04G9.2 : CAAATTTAAGATCAACACTCTTTATGGATGTCAGTTGgtatggtttacatcaattctcattaacttattcaagaagttaa : 2880
g-X-2732 : CAAATTTAAGATCAACACTCTTTATGGATGTCAGTTGgtatggtttacatcaattctcattaacttattcaagaagttaa : 1343
yk366f6.5 : CAAATTTAAGATCAACACTCTTTATGGATGTCAGTTG----- :
T04G9.2fm: CAAATTTAAGATCAACACTCTTTATGGATGTCAGTTG----- :

T04G9.2 : atgtaaacatthttcagAAGGAGAAAAACCAACCACCTCCGCTCCAACATCGGGACCAATCGTTATCACTGTGAAGCCAGTT : 2960
g-X-2732 : atgtaaacatthttcagAAGGAGAAAAACCAACCACCTCCGCTCCAACATCGGGACCAATCGTTATCACTGTGAAGCCAGTT : 1423
yk366f6.5 : -----AAGGAGAAAAACCAACCACCTCCGCTCCAACATCGGGACCAATCGTTATCACTGTGAAGCCAGTT :
T04G9.2fm: -----AAGGAGAAAAACCAACCACCTCCGCTCCAACATCGGGACCAATCGTTATCACTGTGAAGCCAGTT :

T04G9.2 : GTTATAACAACCGGAAAGCCGCGCAGTCACTTCAAACAGTTTCCCGAGCCGTCCTTGAAGCCAAGCGAGTGCAGAAATCT : 3040
g-X-2732 : GTTATAACAACCGGAAAGCCGCGCAGTCACTTCAAACAGTTTCCCGAGCCGTCCTTGAAGCCAAGCGAGTGCAGAAATCT : 1503
yk366f6.5 : GTTATAACAACCGGAAAGCCGCGCAGTCACTTCAAACAGTTTCCCGAGCCGTCCTTGAAGCCAAGCGAGTGCAGAAATCT :
T04G9.2fm: GTTATAACAACCGGAAAGCCGCGCAGTCACTTCAAACAGTTTCCCGAGCCGTCCTTGAAGCCAAGCGAGTGCAGAAATCT :

T04G9.2 : TAGAGGAGATTGTGATGACCTTGCTAAACAAGgttcgctgaaataaaactttgaaaaaacctgaaaaatggtttttttt : 3120
g-X-2732 : TAGAGGAGATTGTGATGACCTTGCTAAACAAGgttcgctgaaataaaactttgaaaaaacctgaaaaatggtttttttt : 1583
yk366f6.5 : TAGAGGAGATTGTGATGACCTTGCTAAACAAG----- :
T04G9.2fm: TAGAGGAGATTGTGATGACCTTGCTAAACAAG----- :

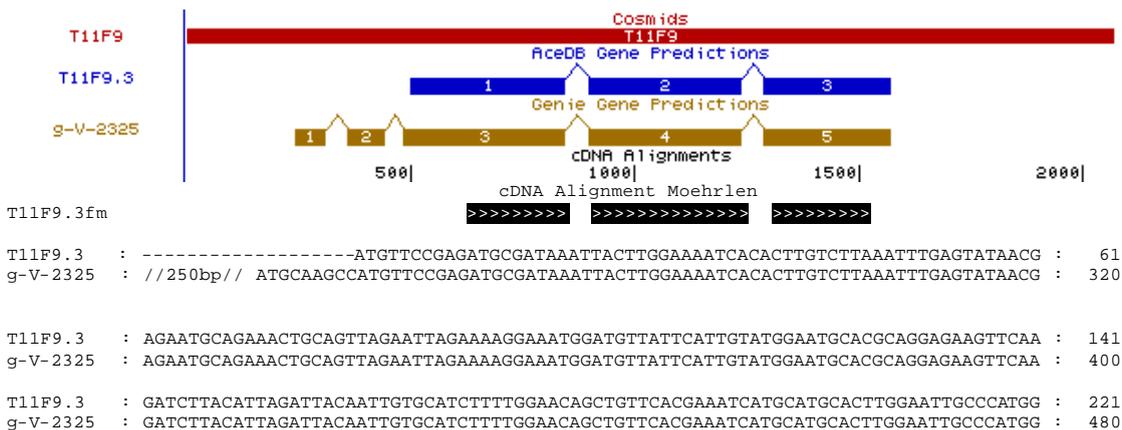
T04G9.2 : tcagGATGGTGCATCAGAAATCCAGGTTGGATGCGTGCCAACTGCCCAATTTTCATGTGGAATGTGCATTCCAACCAAGGA : 3200
g-X-2732 : tcagGATGGTGCATCAGAAATCCAGGTTGGATGCGTGCCAACTGCCCAATTTTCATGTGGAATGTGCATTCCAACCAAGGA : 1663
yk366f6.5 : -----GATGGTGCATCAGAAATCCA----- :
T04G9.2fm: -----GATGGTGCATCAGAAATCCAGGTTGGATGCGTGCCAACTGCCCAATTTTCAT----- :

T04G9.2 : //730bp// CTCCGATCGCAAACATTTTGTCTCACACTGGAAGAGCGCTGGATTCTGTGAAGGAATCTTCATGAACTAC : 4000
g-X-2732 : //730bp// CTCCGATCGCAAACATTTTGTCTCACACTGGAAGAGCGCTGGATTCTGTGAAGGAATCTTCATGAACTAC : 2463
yk118h9.3 : -----GAAGAGCGCTGGATTCTGTGTAGGAATCTTCNTNGCTAC :
yk366f6 : -----GTGAAGGAATCTTCATGAACTAC :

T04G9.2 : ATGAAAAAGAACTGCCAGCATCATGTGGCCTTTGCTAA----- : 4039
g-X-2732 : ATGAAAAAGAACTGCCAGCATCATGTGGCCTTTGCTAA----- : 2502
yk118h9.3 : ATGAAAAAGAACTGCCAGCATCATGTGGCCTTTGCTAAACGGTACCATTCCACAATTTGAGTGCCTTTTTCTTCTATT :
yk366f6 : ATGAAAAAGAACTGCCAGCATCATGTGGCCTTTGCTAAACGGTACCATTCCACAATTTGAGTGCCTTTTTCTTCTATT :

```

6.1.34. Sequenzdaten T11F9.3



```

T11F9.3 : TCAAGCTAGAAGTGATAGGGACGATTATCTTATAGTTGATTCAACTAATTCAAATGATGGTATTGAGAACACTGAAAATC : 301
g-V-2325 : TCAAGCTAGAAGTGATAGGGACGATTATCTTATAGTTGATTCAACTAATTCAAATGATGGTATTGAGAACACTGAAAATC : 560
T11F9.3fm : -----C : -

T11F9.3 : TTGTTCCGTTTGATTATGGAAGCGTCATGCTTTATGCACGAGAgtgagtaaaatcttcttcttcttattaca : 381
g-V-2325 : TTGTTCCGTTTGATTATGGAAGCGTCATGCTTTATGCACGAGAgtgagtaaaatcttcttcttcttattaca : 640
T11F9.3fm : TTGTTCCGTTTGATTATGGAAGCGTCATGCTTTATGCACGAGA----- : 19

T11F9.3 : actaggtggttcagTCCTCACAGTGATAAACGAATCCGATAGATCCAGAATATAACTTCACAATGGGAAGCCTTCGAG : 461
g-V-2325 : actaggtggttcagTCCTCACAGTGATAAACGAATCCGATAGATCCAGAATATAACTTCACAATGGGAAGCCTTCGAG : 720
T11F9.3fm : -----TCCTCACAGTGATAAACGAATCCGATAGATCCAGAATATAACTTCACAATGGGAAGCCTTCGAG : 84

T11F9.3 : TTGCGTTTTATGATATGGTTTTATTGAACAAATCTATGGCTGCAACTGTGATAATCATCCAAGAAAATTGGATTGTAAA : 541
g-V-2325 : TTGCGTTTTATGATATGGTTTTATTGAACAAATCTATGGCTGCAACTGTGATAATCATCCAAGAAAATTGGATTGTAAA : 800
T11F9.3fm : TTGCGTTTTATGATATGGTTTTATTGAACAAATCTATGGCTGCAACTGTGATAATCATCCAAGAAAATTGGATTGTAAA : 164

T11F9.3 : AATGGCGGCTATCAGAATCCGGCCAACGTGGAAGAATGCTTTGTACTGACGGATTCAATGGTCAATTATGCGATCAGCA : 621
g-V-2325 : AATGGCGGCTATCAGAATCCGGCCAACGTGGAAGAATGCTTTGTACTGACGGATTCAATGGTCAATTATGCGATCAGCA : 880
T11F9.3fm : AATGGCGGCTATCAGAATCCGGCCAACGTGGAAGAATGCTTTGTACTGACGGATTCAATGGTCAATTATGCGATCAGCA : 244

T11F9.3 : TGAAGGAGTTTACGTTTGGGAAGCTAAAAAAGAATGGGATGCATCAGGAGTGCCTAATAATTACAGAAAAGGAATTGAAA : 701
g-V-2325 : TGAAGGAGTTTACGTTTGGGAAGCTAAAAAAGAATGGGATGCATCAGGAGTGCCTAATAATTACAGAAAAGGAATTGAAA : 960
T11F9.3fm : TGAAGGAGTTTACGTTTGGGAAGCTAAAAAAGAATGGGATGCATCAGGAGTGCCTAATAATTACAGAAAAGGAATTGAAA : 324

T11F9.3 : CAAATACTATGCCGGAGTATACTTATTTTGCATTGACTgtaagaactatcagagaactaaccagagaattaaaaaaatt : 781
g-V-2325 : CAAATACTATGCCGGAGTATACTTATTTTGCATTGACTgtaagaactatcagagaactaaccagagaattaaaaaaatt : 1040
T11F9.3fm : CAAATACTATGCCGGAGTATACTTATTTTGCATTGACT----- : 362

T11F9.3 : ttcagGCTCCAGAAGGATCCACAATTGAGGTTCTGTATCACCAACTATCCGGATTTTCTGTCAACATACATGCGGATTAT : 861
g-V-2325 : ttcagGCTCCAGAAGGATCCACAATTGAGGTTCTGTATCACCAACTATCCGGATTTTCTGTCAACATACATGCGGATTAT : 1120
T11F9.3fm : -----GCTCCAGAAGGATCCACAATTGAGGTTCTGTATCACCAACTATCCGGATTTTCTGTCAACATACATGCGGAT : 422

T11F9.3 : AATGGTGTGAATTGAAATACAAAACACTGATCGACGAATGTGAGCCATTAGTATGCTGTGATAATGATAATTTATGGAA : 941
g-V-2325 : AATGGTGTGAATTGAAATACAAAACACTGATCGACGAATGTGAGCCATTAGTATGCTGTGATAATGATAATTTATGGAA : 1200

T11F9.3 : TAAGACACGATCATCAACCAATAATCCATTCTATTGCAAATATGGAAATAACCGAACACCTCACTTTGAATTGCAAT : 1021
g-V-2325 : TAAGACACGATCATCAACCAATAATCCATTCTATTGCAAATATGGAAATAACCGAACACCTCACTTTGAATTGCAAT : 1280
    
```

6.1.35. Sequenzdaten T11F9.5

The diagram illustrates the genomic organization of the T11F9.5 region. It features a red bar representing the Cosmids T11F9. Below it, blue bars indicate AceDB Gene Predictions (1, 2, 3), and yellow bars indicate Genie Gene Predictions. Brown bars represent cDNA Alignments (1, 2, 3, 4) from Moehrle. A scale bar at the bottom indicates distances of 500, 1000, and 1500 bp. The sequence data below the diagram shows the DNA sequence for T11F9.5, g-V-2326, and T11F9.5fm, with corresponding coordinates.

```

T11F9.5 : ATGAACTATTTTACTTTCTTTATGCATATTGCTGTATTGAAATTTTATTTTCGATTTTCGAATGGAATAAGAT : 80

T11F9.5 : TGTAATGAGAGTCGGTGGTAGTCTGAAACAAAAGATTGGAGCGATCAAAGCGTCAAGCTTACGCATGGACAATGAAC : 160
g-V-2326 : ---ATGAGAGTCGGTGGTAGTCTGAAACAAAAGATTGGAGCGATCAAAGCGTCAAGCTTACGCATGGACAATGAAC : 76

T11F9.5 : CTCGATGGCCTCGTGGAACTATTAATTTTTCGATGAACAACGTTTTGgtaaaacttcaaattctggcatttttagt : 240
g-V-2326 : CTCGATGGCCTCGTGGAACTATTAATTTTTCGATGAACAACGTTTTGgtaaaacttcaaattctggcatttttagt : 156

T11F9.5 : tattcattttcagATGAAAATTCGAGAGCAACCGTGCTGAGAGCAATGGAAAAATATCTAACCACACATGTATCAAGTT : 320
g-V-2326 : tattcattttcagATGAAAATTCGAGAGCAACCGTGCTGAGAGCAATGGAAAAATATCTAACCACACATGTATCAAGTT : 236

T11F9.5 : TTCTCCAAAAGATGCACGAATAAAACTT CGAATTGTGAGTGATAAAGGATGCCAAGCGGCCATTGGACGAGTTGGGGGAG : 400
g-V-2326 : TTCTCCAAAAGATGCACGAATAAAACTT CGAATTGTGAGTGATAAAGGATGCCAAGCGGCCATTGGACGAGTTGGGGGAG : 316

T11F9.5 : ATCAGCAATATTGAGTTTCTACATCTTGTATAGTGTGGAAGTGCAAGTGAGCTGATTATGTAATTGGCTTTCTT : 480
g-V-2326 : ATCAGCAATATTGAGTTTCTACATCTTGTATAGTGTGGAAGTGCAAGTGAGCTGATTATGTAATTGGCTTTCTT : 396

T11F9.5 : CATTCTCATCAAAGGGCTGATCGTGACGAATACCTTAAGTTGAACTTGCAACCATGGAGATTGAATGATTGGTTTCAAAC : 560
g-V-2326 : CATTCTCATCAAAGGGCTGATCGTGACGAATACCTTAAGTTGAACTTGCAACCATGGAGATTGAATGATTGGTTTCAAAC : 476
T11F9.5fm : -----AATACCTTAAGTTGAACTTG CAACCATGGAGATTGAATGATTGGTTTCAAAC : 52

T11F9.5 : AATGCAGTATAAAAAGTATTTGGATCAATGGTGGATTGTACCATATGATTATGGTAGTATCATGCAATACCATGATTCTG : 640
g-V-2326 : AATGCAGTATAAAAAGTATTTGGATCAATGGTGGATTGTACCATATGATTATGGTAGTATCATGCAATACCATGATTCTG : 556
T11F9.5fm : AATGCAGTATAAAAAGTATTTGGATCAATGGTGGATTGTACATATGATTATGGTAGTATCATGCAATACCATGATTCTG : 132

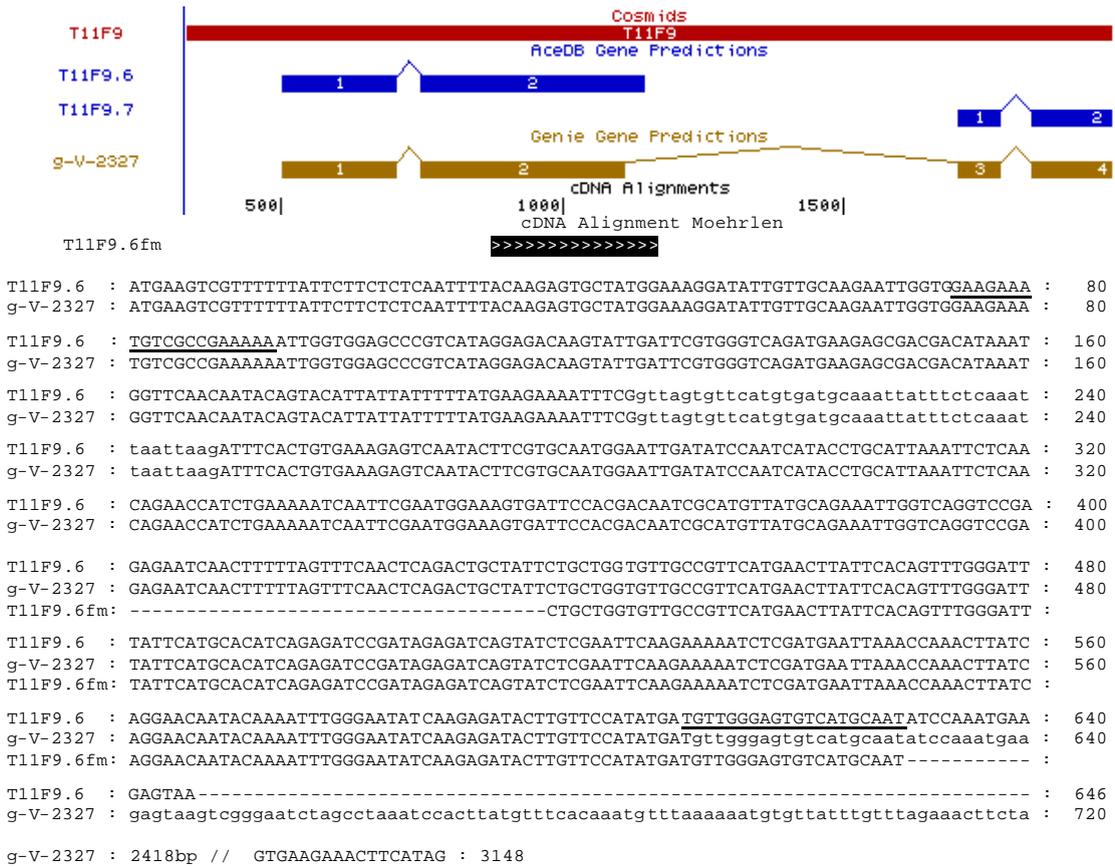
T11F9.5 : AgtaagtagtgttttggtagagaaaataaagtttaagaanaatccagTAATGAATATGGGCCTAAAAATTCAAAATACTT : 720
g-V-2326 : AgtaagtagtgttttggtagagaaaataaagtttaagaanaatccagTAATGAATATGGGCCTAAAAATTCAAAATACTT : 636
T11F9.5fm : A-----TAATGAATATGGGCCTAAAAATTCAAAATACTT : 166

T11F9.5 : CCGAACAAATGGGTTTCGCAAATACCTTCGTATTTTCGATTATCTGATGATTAACGAATACTACCAATGTTCTTGTGAAGGGG : 800
g-V-2326 : ccgaacaatgggtttcgcaaataaccttcgtatttcgattatctgatgattaacgaataactaccaatgttcttctgtgaagggg : 716
T11F9.5fm : CCGAACAAATGGGTTTCGCAAATACCTTCGTATTTTCGATTATCTGATGATTAACGAATACTACCAATGTTCTTGTGAAGGGG : 242

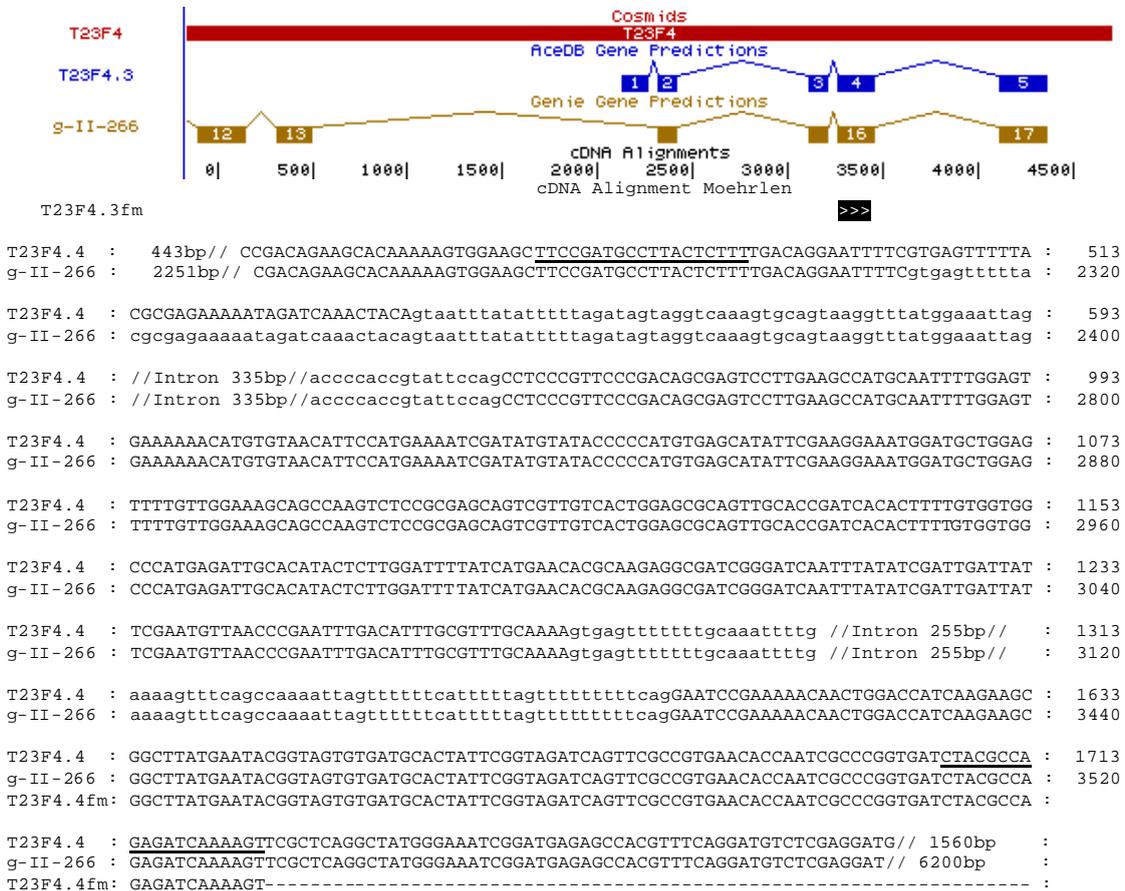
T11F9.5 : AAGAACAATAAATCGCAAGAATCGAGGGTATCCAAATCCAGGAAATGTTCTGAAATGCAATGTCTCTCGGTTTGGG : 880
g-V-2326 : aagaacaataaactgcaagAATCGAGGGTATCCAAATCCAGGAAATGTTCTGAAATGCAATGTCTCTCGGTTTGGG : 796
T11F9.5fm : AAG----- : -

T11F9.5 : GGCTGA----- : 886
g-V-2326 : GGCTGACTGTAGCGACGTTGCAAACACAGCTGTCATTGTAGAATCA //400 bp :
    
```

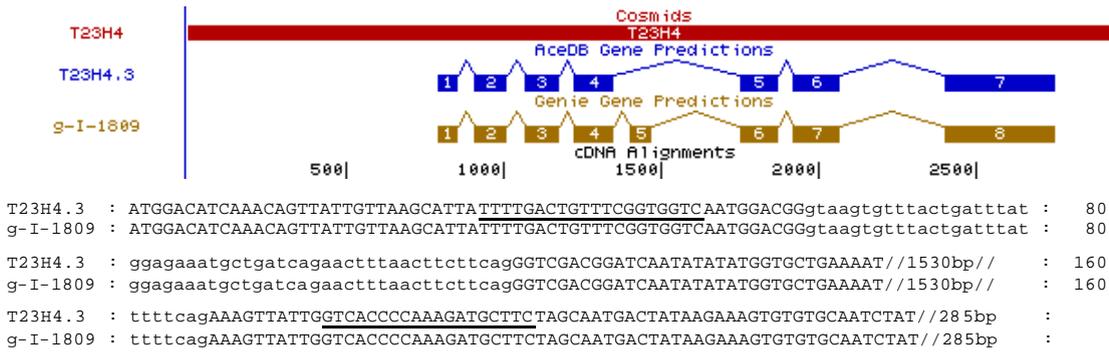
6.1.36. Sequenzdaten T11F9.6



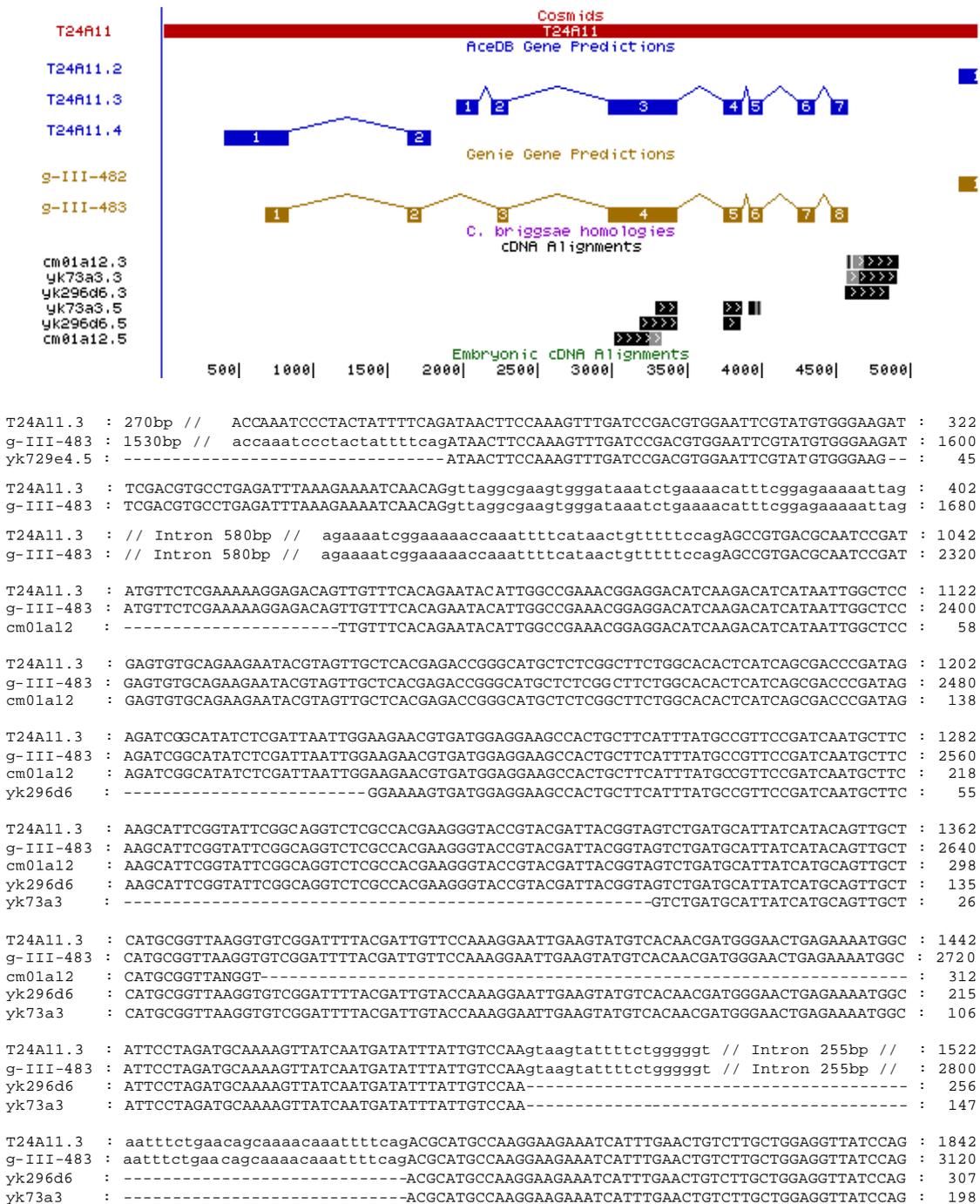
6.1.37. Sequenzdaten T23F4.3



6.1.38. Sequenzdaten T23H4.3



6.1.39. Sequenzdaten T24A11.3



```

T24A11.3 : ATCCAAACAACGTGCAACGTTTGTAGGTGTCTGAAGGACTTGGAGGACCCGACTGTGGTCGACTTCAGCCATCACgtgag : 1922
g-III-483 : ATCCAAACAACGTGCAACGTTTGTAGGTGTCTGAAGGACTTGGAGGACCCGACTGTGGTCGACTTCAGCCATCACgtgag : 3200
yk296d6 : ATCCAAACAACGTGCAACGTTTGTANGTGTCTGAANGACTTGGAGGACCCGACTGTGGTCGACTTCAGC----- : 376
yk73a3 : ATCCAAACAACGTGCAACGTTTGTAGGTGTCTGAAGGACTTGGAGGACCCGACTGTGGTCGACTTCAGCCATCAC----- : 273

T24A11.3 : tttttgatgcgtttcttgatttaaaaatgtatattttagctTGTGGTGGAGAAATTCATGCATCTGATCAATGGCAGAC : 2002
g-III-483 : tttttgatgcgtttcttgatttaaaaatgtatattttagctTGTGGTGGAGAAATTCATGCATCTGATCAATGGCAGAC : 3280
yk296d6 : ----- : -
yk73a3 : -----CTTGTGGTGGAGAAATTCATGCATCTGATCAATGGCAGAC : 313

T24A11.3 : ACTCTCCAGTCCGTCTGGACGTGATGTTCACTGTTATTGGAGGATTTCTgtaagttttattttaaaaattcaaaaatgc : 2082
g-III-483 : ACTCTCCAGTCCGTCTGGACGTGATGTTCACTGTTATTGGAGGATTTCTgtaagttttattttaaaaattcaaaaatgc : 3360
yk296d6 : ----- : -
yk73a3 : ACTCTCCAGTCCGTCTGGACGTGATGTTCACTGTTATTGGNGGGAGNC----- : 361

T24A11.3 : ggtcatattagcgtttacagcttacttttagacagcgcaaaagggcaacgtagcattttatgatgcaatgagctgaca : 2162
g-III-483 : ggtcatattagcgtttacagcttacttttagacagcgcaaaagggcaacgtagcattttatgatgcaatgagctgaca : 3440

T24A11.3 : ggagcactccgacctgtttgaacaatggtgcatcgaattggatttttagatggttcggctcaccatttaaatgcatcaagc : 2242
g-III-483 : ggagcactccgacctgtttgaacaatggtgcatcgaattggatttttagatggttcggctcaccatttaaatgcatcaagc : 3520

T24A11.3 : gctctgcaccatccctgaaattaacaaaatttctaattccagGTCCCCGAAGGCTCCCGAGTTCGTTTCCGTTTGTGAGA : 2322
g-III-483 : gctctgcaccatccctgaaattaacaaaatttctaattccagGTCCCCGAAGGCTCCCGAGTTCGTTTCCGTTTGTGAGA : 3600

T24A11.3 : TGGAGAATTCCCATGCTCCTACGGATGTCAATCGTACGTCGAAATCAAGCACAAGCTTGATGTAAGGTTGACTGGTTTCA : 2402
g-III-483 : TGGAGAATTCCCATGCTCCTACGGATGTCAATCGTACGTCGAAATCAAGCACAAGCTTGATGTAAGGTTGACTGGTTTCA : 3680

T24A11.3 : GAAGgtaagtaagattgctttttgtctgctgcctatcttagtgcctgtgtttctccaacggcctgaacaatggtgcat : 2482
g-III-483 : GAAGgtaagtaagattgctttttgtctgctgcctatcttagtgcctgtgtttctccaacggcctgaacaatggtgcat : 3760

T24A11.3 : caaaactcacgagccaactataaaaactttcagTTGTTGCTACCGTCCAAAAGAGGACACAGTCTCGGAAAGCAATCAAAT : 2562
g-III-483 : caaaactcacgagccaactataaaaactttcagTTGTTGCTACCGTCCAAAAGAGGACACAGTCTCGGAAAGCAATCAAAT : 3840

T24A11.3 : TTTTGTGATCTACCATCCAAATGGCCGTACCGCGGATTTCTCCCTCCGATTCGCGCGGAAGCCTAA----- : 2629
g-III-483 : TTTTGTGATCTACCATCCAAATGGCCGTACCGCGGATTTCTCCCTCCGATTCGCGCGGAAGCCTAA----- : 3907
yk296d6 : -----TTNTCCTTCGGATTCGCGCGGAAGCNCNTAAATTCATTTTNTAG : 419
yk73a3 : -----AAGCCTAAATTCATTTTCTAG : 382

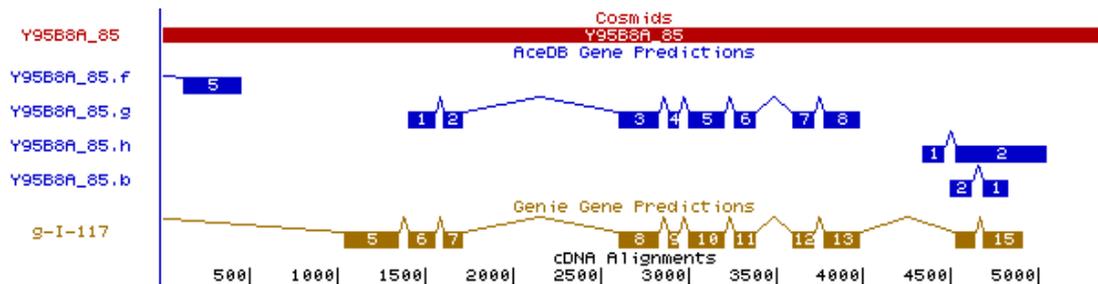
yk296d6 : TTNTTNTACTCNTTCCCATTTNTCTNGAATCCAGGCTTTTGTATTGATCTCTTTGTTCTGTTTTCCAGTTTTTTCAA : 499
yk73a3 : TTCTTCTACTCTTCCCATTTCTCTCGAATCCAGGCTTTTGTATTGATCTCTTTGTTCTGTTTTCCAGTTTTTTCAA : 462

yk296d6 : AATACAATTCCTTTATTACGCCCGGTGACCTACAATAAATGCTCAGTTTCAGGTTTCTCTCCCAAACACCTCGA : 579
yk73a3 : AANACAATTCCTTTATTACGCCCGGTGACCTACAATAAATGCTCAGTTTCAGGTTTCTCTCCCAAACACCTCGA : 542

yk296d6 : ATTTCTGTATATTTCTGCCTCTAATTTTAAATAATTGTCGGTTTTCCCTTATTTTGTCAATTTCTGTCAACTAATTTTG : 659
yk73a3 : ATTTCTGTATATTTCTGCCTCTAATTTTAAATAATTGTCGGTTTTCCCTTATTTTGTCAATTTCTGTCAACTAATTTTG : 622

yk296d6 : TTTCACTTTCATTTTAA----- : 676
yk73a3 : TTTCACTTTCATTTTATACGGGATTNCTATGCACATCTTGTTTTTTCCACATTCAGCGCCCAATG-- : 691
    
```

6.1.40. Sequenzdaten Y95B8A.1 (Y95B8A_85.g)



```

g-I-117 : 5130bp // TGTACGAGGTGAACACCACACAGCTCCTCACTTTGACAACCTTCATCTGGCAACAAAATGCACCAGCTG : 5200
g-I-117 : TCACACCGGAGCCATTCACATTTGCTCCATTCTCTTCCCAACTCTTGCCACCGTCGCACCACACAGCTCCAGGACCCGGC : 5280
g-I-117 : GGCCCGACTCTTGAGCCATTTTACCAGACGACAGCTAGTCCGAAGCTGTTAGCTCATAACACTGCTAGGgtatcaaaaa : 5360
Y95B8A85.g : -----ATGATCCGAGAAATCGCTTCATTTTCCGACGGAGGACG : 38
g-I-117 : tttgaatactgaatacctcctataactttttcaaatccagATGATCCGAGAAATCGCTTCATTTTCCGACGGAGGACG : 5440
Y95B8A85.g : //1530bp//ctactatactatttattttaaattcaaattttctagAAAGGAATCGTTGCTCACGAAGTCGGTCACTCCC : 1638
g-I-117 : //1530bp//ctactatactatttattttaaattcaaattttctagAAAGGAATCGTTGCTCACGAAGTCGGTCACTCCC : 7040
Y95B8A85.g : TAGGCTTTTGGCAGGACGAGTCCCGTCCGGATCGGGATGACTACATCCATTTAAGAAAAGACTGGATTA//890bp : 
g-I-117 : TAGGCTTTTGGCAGGACGAGTCCCGTCCGGATCGGGATGACTACATCCATTTAAGAAAAGACTGGATTA//1760bp : 
    
```

6.2. Korrigierte Proteinsequenzen von *C. elegans* (FASTA Format)

```

>C_4R79.1i Intronerator prediction g-IV-3005
MLDHVLLTYCLVSTVVRSPADVFRSFAGYIPEHRVTHHEWQNSGKFGQDIDGVDPNLLKLPPEGPVLFNALKNKQLTWEQGGVPIPYEMD
TAFSPNEIKILEKAFDSYRRTTCIRFEKREGQTDYLNIVKGYGCYSQVGRGTGGKQEISLGRGCFHEIIVHELMHVSQVGFWEHSRADRDDH
IKINWDNLLPGMKSQFDKI SAVLQDLQGENYDYKSIMHYDSTAFSRNGRNTIETVENGFQVIGTAMDLSPLDIVKINKLYSCKTKKKEKV
KPATTEEPHQLIPQVVDKNSVDSGEKCVDFADCPHFAQYCTRASFFVFMKSYCPFTCKHCPCGDRKLLKKS
>C_C05D11.6i Intronerator prediction g-III-1042
MIYMSSFYCISEKMSHSPKNSTFFNLFQYSMTSEPSLVVIDTSVPQETETLNDADFHSDDLHQRYDLQTLGKVKDDPTIGNYSEGDILLE
SPKKFVEENKLGKRNAIKQIYRRWPNNEI PYTLSSQYGSYARSVIANAMNEYHTKTCVKFVARDP SKHHDYLIWHDPDEGCYSLVVGKGGKQ
PVSLDSGCIQVGTIVHELMHAVGFFHEQSRQDRDSYIDVVQNVMMGADQFEKYNLNVISHLDEPYDYASIMHYGYPAFSGSGKKTLPVK
KSGSERMGQRVKFSDIDVRKINKLYNCPGVSNNNNNNNNQINSIVNHQPSRYAGLPLSTSNQVTNNRRMIRIFG
>C_C07D10.4i Intronerator prediction g-II-1703
MLLPWIITIVTVIIPATLGHNRVQDDEMLVLSDDTSLNLEDFEFADKLTREELFGKHIPVEVVNDFKSDIRLPRRHKRNGVSRAAKLWPN
ARIPYAI SPHYSHERALLAKAVKQYHEKTCIRFVPRQTGEFDYLF IGVVDGCFSEVGRTSQVLSLDNGCMEYATI IHMMHVGFYHE
HERWRDNFDI I I WQNI DRGALDQFGKVDLSKTSYQGYPDYKSLHYDSLAFSKNGFPTMLPKVKSATIGNARDFSDVDISKINRMYNCP
VEKSVTAPFARARHVPIYSPQYHYKEDRPKIPLSRFDMMQGPINPPMAQIPSOQLVSSSSGRVNVNSNKPSSRKLAFYFGEELILECED
RITVCWWTADRCRSPAIYQVMSSCPKTKCF
>C_C17G1.6i Intronerator prediction g-X-2881
MKSQAQLKVLALIGLVSIVSTAYIANDVSDYAEVKELLAAFYRKHAKKYGHYDPAAIQAI AENMDKSVKNDKTEATVNRKLNWVNFEN
DIILTLQAEESLSESNRSPRRQAHPDPRNFWPNTI SYEFYGGEEETWRQLIRSAIRHVEQNVCFKFKENGDRDLRYRNGCWSNVG
RVGGRQLVSIIGYGCDSLGI VSHETLHALGLWHEQSRDRDNFISIVADKIRGTEGNFAKRTAANSNLDGQPYDLGSMHYGAKSFAFDWS
SDTIKTRDWRVQNTIGQRDGLSFKDAKMINTRYCSNVCQSRSLPCLNEGYTDPNNGRCRCRCPSGYGGTYCETVEYTS CGGSLTASSYKKE
SGIVQPDANCVWRIRNPGGNVEVMFDQVNFQCADPCQSYVEVKYLSQKTSSTARLCCSLP SVIRSEGGDVI I I LRGTPTAVGWRGFTLKY
RAIGGTPITPATVTRPYATTTTRPYWTRTASGWIHKNPPLYKPDGQIYTSDEQSAETKYSSEELYDPSTFLSPSSSAPALLLPDASPQ
RPSAQEHDLSQLSONALTRPTPTTTPADPTASWSAWGEWSACSQPCGGCGTKTRVACYGNGQVCPGNSLDRESCNAHACAKPKKGMICNG
RLLLPCDLLAKLNFSGNNYLNPKLQSGFARSSTLPLPRI SQRKPVFVNELEVHPPTERFLSSSTRVRVKRQTANRFCEKRFIYQCP TALLT
IQMEYKPD TQGTNDAYFQQYPECCSGYTPRRGVCYKN
>C_C24F3.3 wormpep prediction
MLYIPQFSYFCLGYLLFCCKISNAVKQSWEINQELITEANKEHTVFGDMRELQLRRPSETAFSVLTPAQLIRYENSKSDLSIRGVSIKG
SSMNRNNSNIVPVIYISPOYSPAQKQILVSSLRVYERFVSCFKFVERTTNDYLFIVPLDGCYSYVGIIGGRQTLADADYI IWHEMMH
AIGFEHSHQRDPDFIRVDYANVIPGQMINFDKLTSHVYEPDIYDFKSIMHYDYAFGRVDTARRVRLATMTPLKPGVTLLEDNMKFTAT
DIEKLNRLGQCGARGGQYSNQGVVASTCQDVATAVSCQEGNRRGMCKNPFYKQMMIKSCQKTCRLCSYTRMIDEDDLTPNTTVKSVKCED
KHPRCDIYSHNGFCTLPFYDDVRYQLCAKTCNLC
>C_C26C6.3 wormpep prediction
MRLCHSIILFNLSISISICKADDPALLVASEFKEHFVNEEKQLETVEELLIKMKLAHSRSFKGREFGHDAVEDSKKEVAISTQOQTINK
KVSPLFEGDIFLSRRQAVDILKALSKDKTKRLRSFVSDKTATWKTMPKIRFHE SIDFYTISQI IAAIRFWEDSTCITFENVSDSPDGD
YIEFFSGQGCYSMIGRNGRQGISIGESC VKASFFVMGVI EHEIGHALGLWHEQSRPDALGYVTIERDFILPSYISDFLQRDEIDTLGIP
YDLGSMHYGSTAFSVDQKSKTVVTRDSLYQQTIGQREKLSFYDVATINTAYCKDECKSEKTCENGGMYPKSKSCELCPPDGLGGKCEK
NEDSKNAECGGI I KLTEEWEKIESPNYPDPGYEADQKCSWLLKAEKGRVEIEFIEDFSFLCTSTCVDFVELKISDDL RNTGFRFACCYDKP
EISFVQPTDAI I I IFRSQ LSTDIGFKIQA KSNELI I KTFKNSQKTKMQLILFPKLSLNNSADAEPRTTIAPTIITTTPLAPDANVWAD
WGEWSMCSRTCGGCGRSRRVSCRSKKCEGRRQFEGTCNLKACPVDKHKAKLLSNRLCNGKVC TKNDIAI SSCDAPQCCPPF INV DGVQC
SDQENQHDELWLSI
>C_C34D4.9i Intronerator prediction g-IV-1119
MRRNDLLNKKITIFLSSLSLVIIIP IYAAEKDLLPSTSETFLTDEDFLRPLNDDETFLTEKDFKNGEKLGEDHVPAGSILWKQVYKKG
DIRGKAAWKLDPKNSES LRNGVITGTRKWPNGRIPYVISNQYNDREAVLARSFQAYHDKTCVRFVPRTAVDNDYLYIGKIDGCYSDVGR
AGGRQELSLDNGC LQYDTA IHELMHVSQVGFYHEHERWDRDEHITILWHNIDREAYDQFGKVDLAESSYGGQLDYYSIMHYDSLAFSKNGFE
TMVAKQSEMTAVIGAAIDFSPIDILKMNLMYQCSVDKLPPTVVGTTDKTIVPVPAPSVTVVEDDCDRDTNLCWRWIDRCKSFFFEQIMKEF
CSLSCGYCTPKALQTAKASPPNYSNLTLTKSSTSYLQG
>C_C37H5.9i Intronerator prediction g-V-856
MIFLLFVVVFFVYAQQLPELLAGFNQGRFRGGPDGFNRGPGGFHRGPDGFGGDPGGVDLGH LIGNIAANVQEMGLNDADVIGDLRGISR
GPRPSSMEWGRRARHFCRRYPGHPKCQRGQLPQFTDVPTIINTI IYNAGDLLPRVPTLNIHDPLAGLNLSELVGF IKSLSQSQFQLSSQQRN
EIHDFSCRFPKCDQSPQNTQAKQELLTKMLAFDQAVGGKAAPAHDKVNLRFDRTTQVQVQALLKRANLSH I IVPADNGVFDRLVLLTEHQAN
FLLNELGEAGRGADVGAGGGGGGRVPRSGVFFQESAVQKWDIWPKI QYTLDDSL EESDKKIRDALHEISINTCILFRYNATPKGYHLNLM
KVDSTTFCGLSYVGRTPANPIYLSFCQGDNRGVAMHETMHALGVSHQLRLDRDKYIKIDWSNIDPQHYDTFAISDAKLYTSYGTKYAYD
SIMHYNAYLGAKDPNKPTMIPLVNPQENTPKLGQRAKLTGRDIRLLKMKYCRPGCDDQNVHCGT WALHG YCKMKEQMKWMNENCKASCDKC
>C_D1022.m not in wormpep/Intronerator MERNUM MER05843
LSFEHLSNIVREQLDHSRTRKRTAFRNAMYPKTIWLPNVFELHGSLSAKSRSSLVAAMAFWEKHTCVAFKRTSEKVVYLLMSGQEEGCWST
VGRDEAQAQILNIGTGCEMFGIT SHEIAHALGLFHEQSR YDRDNYVQIVKSR I AQTNFYDFAVVGKKNMETY GQKYDIGSMHYRPTFEFS
LDGNSI IAKDVNMQNTMGQFRGFSFIDVAKINRHYNCKEKNKNTITCLNGGYQHPKNCKIYYTCRCPGYSGSDCKGIEASSPAKCTGVL
VAGETQRKFTANIKPNKNAKGI RKNYHIEAPPGRK I V I VDSVIGNCVQCYEEGVELKMYEDKTVTGARFCCKLQKQPQLIISQGNVTPL
MLVAGKAQAFVQLRYSTGR LFFNFHRKLP
>C_F09E8.6i Intronerator prediction g-IV-2471
MRLLYSLFHCSAFLVGF TSVGLPI PNEHAASIKAKFDYAEHYLLPEDFHAETA P VVKPTDAEIESMQNSL LFE GDI MGVP EIEKSDI
LKRLRDDPLLDEDEIFRKP FHSALNLV TY PDKLWPEGQVPYMLEEGMTNDQRTAIAQAFDEYKTKTCVRFVVKPTDDDFDYIYVKNRVAFC
SSYVGRAGGNQTVSLEVDKCF SKGI I AHELMHALGFFHEHSRTDRDFVD INEDNIRPGMMRNFEKYPRKI IDSLGMPDYDESVMHYHKL A
FSRNGKPTIIPKDN EADVGQRYKLS E MDSKVNKLYQCGEYSKTSSTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTTTTPK PVERSRNKKCEDLNAHCGMWEQLGHCQHSVKYMAHYCRKACNLCEVEVTTTTTTTTTPKPVPRNKEKENKSASSTTRGTSTATSTPT
KTTTTTTSAPKEKCEDKNLFCSYWAKIGECNSESEKFMKIFCKASCGK

```

>C_F20G2.4i Intronerator prediction g-V-2804
MTRVVHIIIGAAFLSSYAYCGLSRFNEHDIIEGGDSYKRVKREFEERLGSKWLGTTINYYYADNNNSVKEKVKSAIAYIANHTCIKFNEDPTH
WQRLKIFTSELSHCRSTIGAPGTRSGSAGELSMETGWCANIGSIVHEFHSLSGRYHEHTRPDRDNLKVTSTDYEARPRPWGMMTMYGPF
HGSIMMYHSNYSYGVGKMEPYDMEYKNTMGSRRTYFYDMYKINQYQCGCSQLECKNGGYTSPDCSRCNCPKGFGLKCNERRQQDSYEL
KATYGRWQTQITISFNKYKPEPVSDGFYSTFVYITGEANSTIEITMEGLENVICTAGCTWNGVEIKSREDSRITSPVMCKDEPLYKVKFKSL
HNPITIEELYSKETAPSTATFKYRFMNDKIVFG

>C_F38E9.2n new wormep prediction release WS73: 08 Feb 2002
MRFSANIAIIVNIIFLIVVEFVLPFIRSGDVRFRYYRNNGRVSRATAKKEIWPPEGIIPFVIASNFSGEHQHLFLRAMRHWENFTCV
SFVPRQPHHKHYITFTVDKCGCCSYVGRRGEGPQAISIGKNCDFGIIVHELGHVGVFWHEHTRPDRDNDMYVDIFYKSIQTGQDYNFKEKSKP
EEVDLSLGEPYDFSSIMHYARDTFSRGAIFYDITLIPKPNISGFRLEIGQRVQLSEGDIRQTKKLYKCAECGGTLMKESGNLAIQIHAGVCTWHII
SPQGHITIFLNIITGGLKLIIMIIFLTDEKLEDLTEIFKPNILKKNQTYLHWKTFLIYYTNSFFHFEILDRIICGGDSLFRTIASSGNRMLIQV
RSSTPAASLFPFATYAIICGGPIYANEGVIHSPKYPESEYPPNSDCQWTHIVDENSQVAIEFVYFHLQHEKCIYDRLILTEGSKNSKDKGK
EMSETFCGLIEKKTIVSKTNQISLRFSDNSVQKTGFELRFTKELNECATDKNICHHYCVNTVGGFKCACRVGYSLSNGFSCDSTCGGYL
KASDGLSISSEPNFEMYPNSKTCIWEIEAPDGYHIFLNFTKFNVEGKTECAYDYVYKIGDSEKLCGEYHEALLFTTPRNRVRIEFSSDSSVE
RDGFFANFIADFECQNDNAGEHTCQNRGLSYVCTCNPYILAEDKHCKEGSCFFEVNAPAGDINSNPYNDYKPKQNCSSWHFVTTPGH
RLMLTSSFPQVEEHAQCKYDAVSVYDGGDGSAGLWGSFSAWHLPHSFRPPTSSISLFRTRRSLDEAFRLITLLDCSWIVRAKSP
GRGVRIQFSTFNIESEEGCQDYDIEIYDGPPEATLERLVGRFCGDTSPVITSTGPELLELIMHTDNAEEKGFVAEYREAPRSSSTKRTFVS
KTRHSPLEPIHDRNE

>C_F39D8.4i Intronerator prediction g-X-2412
MPSPTSSASVFSHLLFFVFCIFSQIAQSYAQFNRFDPISPEQETVLTDYDYNMIGHFRENQWFVNDSSAMYNPLRFEGDIANSGLNSRS
INTFFGDSPLFGIFGVQRNAVRYTYLKWQARIPYTISSQYSSYSRSKIAEAEIEEYRKKTCIDFSPKSAGLDYIHIIPDDGCYSLVGRIG
GKQVSLGDEGCIQGIIEHELMHVAVGFFHEQSRADRDREYKINWNSNVEAGLQDQFDKYSLNMIIDLHGTKYDYSVSMHYAPTAFSKNGKPTI
EPIEKNEIGQRAGFSENDIYKINMLYNCPTFTATTAPENTKRVKSITKVTSATGSPLSKKEIGSSIGDKGDRGDNVLSLSSLSIHS
LGKCEDRRKCEFLARAGHCESRFSIRFMTENCANSCGKICAEERKKEVCEADARTWCERWANSGMNQTVPFKDYMRQKCAKSCNFC

>C_F42A10.8fm wormep prediction corrected after cDNA sequence this work
MFFPVVFFIPFVLGAPTQALEKILLVDNNPDSVTNREKIRGIIDKAFENRVRVQGRQGVAPPVTFEALNYPKNNQTKKFEELNQDINEY
TFESDIMLNEKQAKHIAATAIENGNYSKRQAIVDTTNFWSVSVPIFYQFDTKLSATNIANVRKAIQFVNDNSCLSFKEDNNAKNNRFLSSA
GGCWSYVGVQVDMYQMVSVGPNCDFTGTATHELMHAIGFWHEQSRADRDNDYVYVDFSNIIIPSAYNFQKMAVDQAQLLNLPHYDYSVMQY
YPYAFVDSKYITLAKENGFQNSMGQREAPAFSDIIGVKNLYNCTSQCKIQMKCSNCGITDSRNCNQCCKPRYFTGASCDLSLPGTAPNC
NGAVLQATSSWETFDAKAGDPSFSSSSTDNSTNCYWHIKAPEGQOIEFKMTKTPLAACIMQECBPWQSIEVNLGKFDLFGMITCCDITLNVQ
FTSELNMIALRGIIRYNQLTFSIQYRAVPSKSPASTNAACLNQ

>C_F45G2.1o full lenght OST clone OSTF043G3
MLPIVVSILLATPLALCQAPFWMPNMQQFSFLTETDFRNALLLPTTNRVRSLSYHDMRIPFQRFKRGGGVAVAAEKDKWPNRVPYILSA
AYTSAQRAVLARAFDITYAKRTICIRFVPKSPADKDYIIVIKLQDGYADFSRVGGRQVSLADECIDYATIIEHELMHVIIGFIEHQREDRDSY
VSILYQNVIQGANTDFDKLSNLGLSYYGEHYDYSSIMHYEANEGRNGKNTIEAKNSHFTAIMGKASDFSTSDLRVNRAYKCSIFAF

>C_F46C5.3n new wormep prediction release WS73: 08 Feb 2002
MQIYLGITICLVAFLTVIDCAIPYRTHSNFSGSLGRRKRVQVQRDLTYRWPNNTVPYVGNVTSTIKKSURLAIEELQAWTCIRFQNVNEK
YSDGDSVRIVDLGSCSSPIGRQIQIGTQDVSLTKNCWGMGTAIHELMHAIGIEHTQSRSDRNRYLDILAQNIENRDLNPFELLSPLWANLV
PYDYSVMHYASDFSNNKDEQTMPLPKDRSFIEETMGSMIPNFYDFDQINQYQCYDSCRNAGQLANCAANGIIPNPNQCVCNCPMGYGGDL
CDQRPEGGSTLVATDRWQKQKLSVRFNRDDQYFTFCNSWIVGSDRDLQVIYEITSDSIRRQICSFQCYEGGIEVKHLQDPRITNDRDC
CLNTPLNLTTTVNPLPVILYTSATVTVYDFSYRYV

>C_F54B8.m not in wormep/Intronerator, genescan, MER05847
AHEIHALGSMHTQSRYDRDEYVIVKTNVSVLHFEELVI**RFSVPGESFPSTSHSIQHFHDFASVMLYHYSDLLKPINPIWKHTIG

>C_F56A4.1i Intronerator prediction g-V-409
MIVLAMTVPKVHLMRIPHDLQESEAYRTAPKVLPEKDDIEAILTKLPTFEPKSKYGHINIPLRKKRGIALHPLQWASYLWPNAEVYDIAT
HYTSTEKSIILSAMESKNTCVFRPRAATDKHYLQINKYFNVERCFYSYIGRQSSRTLFGTPEGNVETRMRLDPACLRGNRGRGIVMHELM
HILGFYHEHQDRDRRIVGSAVHYNFKIYRRAKTYMGAYDANSIMHYNFQNLPLWQRRDHFSTSDIININTFYKCKNLSKLAPKVPIS
PTSTSTAITTTNTTTTKL

>C_F57C12.1i Intronerator prediction g-X-41
MPSPPSNRHI IASCFCLLIFSSAARVPKASKKHARVQLLNDEAERHNTLIQSDSVTVFDDIQRNPNTGVHDELAVNNADEYFQGDV
DLSEQVQKIIIEDQFTQKREKRKIGRNPLYKWDTRGPIISFYDAESIIPFQTRQKIRSAMLLWQHTCLRFEEGGPNVDRLEFFDGGGCSF
VGRVGGTQGISISTPGCDVVGIIISHEIGHALGIFHEQARPQERHAIINYNNIPLSRWNNFQAVGENHAETYNLPYDTGSMHYGPYGFAS
DPYTPITIRTLERVQSTIGQRAGPSFLDYQAINMAYGCTESCADLPLCLRNQYTHPNNCSMCACPEGLSGRYCEQVYPSNAQCGGVIIPATKE
VKYITSPPNYPKFPIDTECNWI IAAPIEGRVFMFEGDFDLCEDTCDKAVVEVKYHSDKRLTGARYCCSLLPKNRFISFKNEMIIMRGY
RSSGAFKAKLAFWSLNGEPEVSTPLPPTTAPLPEISETTQKPEPTTVQSTTTTYYTTAIPRRTAKKQFFTRKPIITPLPLTSSSTTTESTTV
SSTTQSTTWLPTPEPSFATGETEITTASPTITLFPSTLILPINSLAGVLPSTQAPDIINSVLECGCGAWSEWQGECSQQCGGCGHRLRKR
ECKKEACRKEEKRPCNFSACPDGTNFLINNAEFHILWRGCVGLFRSGDQCSALETESNPFKIIINSLLNQDAKNNDTLIAKRMGRGEH

>C_F58A6.4i Intronerator prediction g-II-1161
MISKNTSFCGFLIIVLATCMSAQFVSNESIKLHDLKPSATHRLFDTLQYSVEEQYSDSHLSFDVSTIYNYSEKPISIGKLNKKYRDLILFE
GDMAISYKQLSMIVNGSTEYRKAIKSRRRGNKINGESTDRTRKQAYLDNNYPATIWKNVAFMFHESLTPIAKTAIILKAVHFYRETCTIEF
HPRTFQKEYLLFIGNDDGWSTVGRDASQKQVVISGNGCEHFVTSHELALHALGIFHEQSRFRDRDESVPFNPRVVERDLLFNFAKISPRQ
MSTYGLPYDIGSMVHYTPTEFSNIPSIPTLAAIDTNLQQTMGQLEGPSFVDVHIMNQHYQCQEKCPQAPCQNGGFTNSRNCKVCKCPTGF
GGAYQLIASFSFPFCGGYLNAAETTRRFDITIRQSTTTRSKTCVYHIKAPGKRIIDILKIDSKIEGCWQDGLLELKMKDFRVPYGRF
CCPESSRRKVISSEGNMVPFVMSKEHDFSVSFEYSFVSTAGFDDEKNSDVIDNLDGVFVSDTSLLRIGFRRQL

>C_F58B4.1 wormep prediction
MILQLLFYSLFTHLAVSQIDVQNALNQNKLNIIDTSSSAISDAELEKTFPRTNLSRMRNALKSLRQNSAKLQAMPARNYQNAQTQENGA
TEQKPLREKPRDRVKMEGDTLHQVNKAAGLNDIILYQGDMLVLTEDRYPSSTTWGSSVYYYDRDTAPKIVKAFEQAVAFWQNVTCINIMQS
STAINRIRVFKGQGCYSYVGRISGVQDLSLGTGCEEFGTAAHELGHALGFHTQSRYDRDNYISINYANIDPSYVEQFDKETSNNTNFNYGM
PYDYSIMQYGATSASSNDKATMIARDTEYQDTMGSDFVGFYD ISMMNEHYKCKELCPAASSAQCKNGGFPSPRNCAICICPSGYGGILCD
QRPPGCGDSVTATTTWQTLTNTIGDGLPTLRDNTMCNATYVVKAPDNQAVEIRISGLTTVTIDGCI FGGVEIKTHKDKQLTGYRYCSSADQN
TVHRSTKAVLPIIILFNRYASTKAVLEYRAVTPSVDSATYTTTFAP I VNSQDLHPNCDYKFFGMCRSKRIRSNCKFTCHDCNNNNNASPF
SNFNNNNYNSFNNWYTNKKNYYPYNSNNNKPWICAAAYAWNGFCVNPYFSMQRHYCYACTCGLCWMNNNNNFY

>C_HCH1 full lenght cDNA clone Hishida et al.

MVSYWVPLIVLCLLPICHAKSIFYADFVNGKGFQADALKFMDKMTILNKLQADILGIPQPDEFSAALDFEDKIESKPEIPYLFEGDMVLT
DEQMDLI IKNVDRQYWARSSSTNEFLYAIRGKRSMTSFSLSERWSFVPPYIDTSSGVNTNAVLAVAKWEQETCARFTRLNSYSSSRQNA
LRFISGNGCYSNIGKVSFPQDVSIGWGCTSLGTVCHEIGHALGFYHEQARYDRDDYVSIITQNIQDMYLSQFTKQASASSMVDYGVGYDYG
SVMHYDQAASFSTGGNTIATRDPNFQATIGQRVAPSFADVKRINFAYCNSTCSNYLDCQNGGYINPNDNCNCKCPPFGGQLCDVAGTNSN
GCGAGDITATSSYQTI SASGALTCNYVIKAPVGAQVYFQMTAATFSRYSPTTNYLEINYGDRDFSRVGARFCASYPTI SLSETNTLVVIYK
GVNGARFSLNRYRDPVTFSTSAPTTTSTTTTAPITVPTVSPTTTTTRQTTTTARTSTTTTTTQAPPTTTTSTSQCASWSACSAQCGCGGT
QSRRCGTYVETVYCNTPCTGGYCCRPFFVTSFSGTYGCRRPADTPAAPQRYVEQRKG

>C_K03B8.1 wormpep prediction

MGNLKFVSLNCKIKYRSENLEKSKMFYQYFHLTSLIVLIYGSIDIGKDANEHMSSVPPKIIIEGTRQKRQVVTKLFSPQWSNAKVFCYNAD
SFSKLQLKLLKLNKSDTVKRKLMKFAMNFISSQTCVTFEENCTISTRIKFVDSTFCASYVGMINSVQEIYFPDWCMRFGSAVHELMHALG
VLHTHARFDRDNFLNVLNKNKDEDDSNFEIVSPPFSINVVPYEGSTLHYTADVSGTNSLLPKQMEYRTRLGNRRVTFYDMLTINTAYNCK
CPSELLCANGGYTNPSNCLCICPLGYGGVLCDRVAVACSQVLSADSYWKGWSIVSGSSVLRDTPDKAFISINAPKDKIIIEVKIVKIEINF
SDSGCENNNGVEIKYMGDPRITNPIICENQVDP SNKGYKAKLNPLLINIYTFGLGNKVTFHRYRVNERLSSYNKTTNGYDNYEYYA

>C_K03B8.2 wormpep prediction

MQKYQLYKTMEDVKNHFDMFLRPLSTLLTLFLALVAGSAIRKDVDFEFDNSNKGKDGIVDGDIMLTEAQLRILNGTAKRSKRQITKWKWP
DAKVFYNYEFTSLKRELSYAMAHISSNTCVKQFESNSATNRIRFTNTGGCASYIGMNGGEQTLWFGDGLIFGTAVHEIMHSLGLFHT
HSRFRDNFLSVSYKDVPEMVGLEKETEQTTYNAVPEYFGSTMLRYNTFEGEGLVSKNEDYQKTMGLRRVSYFDLVNINRVYSCGCAK
SLTCENGGYTNPSNCACTVCPTGFAGTLCNEAPSNTIKLTAESYWKGYWVNFYGYSTSIQTTNYLAWITAPADKTEVKIMDLGSGFTCS
YGCNNGVEIKYMGDPRITNPLRCAQDTEYLNQVLISSKQNPPIVMQQRYSKSLTIHYRYVDTPLSSNKKSTNGYDNYQYV

>C_K03B8.3 wormpep prediction

MCASCLHNANAASFLMETFSRHRNLNFCIGKMNIFYNVTSLDNKYSFHNPSAPVPHSSHALLSSRQAHTSFRAFVAPPSPRRSRQLPSSAK
LLYCFDSSFSTEKRELMFMSNFISSQTCVTFEENCSPINRVQLVNGKGCSSYIGMNNIVQHLTFNDTCIDFGTAVHELMHALGVIHSHR
LDRDNFLNINLTVSKEMMHNYAIFNQSTNVVPEYEGSTMHYANI STMPKPKSEY SATLGI GRVTFYDMVLLNTAYNCKCPNELLGNGG
YTNPSKSCSEICPLGYGGVLCDRPSLIGKDTGLPLEVKDKIIIEVKVVGIDNFFSYPTCLINGLEIKYMGDPRITNPIYCSADLINGKSFRS
KLNPLLNHITLFGKNKVTFHRYVTERLSGYNKTTNGYDNYEYYD

>C_K03B8.5 wormpep prediction

MYLKIIFRLLLEIKSSILKICVYFSSKNYGLISISFAFLLIYFDVLYEKFPTDLHIPLDFSYKNKVENGINEHIAAGNIIQHLEMVRLIHLI
GAIILLFSYAYCGLSRLNEHDI EESYSHKRVKRQFERLGTKWSYGVVNYADKNEIKEMVESAIYIANHTCIRFNEDQNAVQRVQIRM
QQNWLQSTVGPAGMSMSKPIGELSMVLVQCDTIGSIVHEFSHSLGRFHEHTRPDRDNFMKVTITVHEARPRPSGMTTMYGPFHEGSMVMY
HADTYGPGTMDPLMDYKQTMGNRRVTFYDMYKINQYGCWCCKLECKNGGYTSPSDCSRNCNPKGFFGNLCDERQDQSYELMAVNNLWQ
SITIPFAYKPEPGSDGFYSTFVYITGKANSTIEITLEGLQDVICTAGCTVNGVEIKFKEDSKITSPVCCCTDKPPYKNVFKSSHNPITIEL
YSRTTLPSAVTFKYRFTNDKVVLG

>C_K04E7.3 wormpep prediction

MGNQPKTTNFHQRFKGLPPPPPEPEPEPQQDQDQVMSQDIDKVVNSVNQNTAAFQRPGESYDKVIQIMSSYFNKSGSQYDINTVIPSSG
IYNNEMAANSKIAAVMFESDMALTVSQMNVAQNGFRVVKRMNLNGTTWSRNIPIYRFLDTDGNWQSQITNGLRHYERNTCIRFSLNNGGSD
YLVFSKGEQCYSSVGRLLGQPQEIISIGDGCETLGIITHEVGHALGFWHEQARPERDSYVRINRQNAINGLEGQFDRKSWSEVNEYSLPYDY
SVMHYGPKSFSSKSNMTEPEVDPDAFINTIGNRVESPLDGLKLLKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNKLNK
CGVELPRADYSWRNISYSGSSDCYWRIVSANGGNVRFELTYVMYRCSPEEVEFVEMKAESHEATGYRQCCKAVLGERISKNSVLIISKA
TQNSQFVLRVREDGTAPTQRPPVRAAPRSYSLWGWTRCSENGSCGTQYRERCTSTTNCLRSKQTRVCNTQPCAQGTTRGKRSVLQ
TQSKN

>C_K06A4.1i Intronerator prediction g-V-1836

MYRFIIFFSLLALTASKVSEPEKDEIAVKIPTKRSVSEPPKDDDIKAVKIPMRKKRGIAIHPQWESHLPNAEVPYDIASHYTATERGII
LSAMAFRDVTCVFRPRRSTDKHYLQINKHYQLERCFYSYIGRQSSRWLFTTRDGKVVETRMKLDPSCLLYNGRGTVMHELMHLLGFVHEHQ
RDDRDRRIGGSASHYFKIYQRAKSYMMGGYDANSIMHYNFGSVPWQKRDYFSPSDIRNINTLYKCNRRVSKFPSTIPSTSTTTTKAPQF
ELFEKKQIESNSLFRRRRS

>C_K09C8.3 wormpep prediction

MYTITDLKTINRQKQAYLQEKLICSRHCWLTHTTSKHINIQTWTLIPHIRCRGFNNGPHNPSRPGSRPNSPLGDI FGNINGMVKGITDQI
GKIAQGLDVNNDLGMAGHPPPPQSEWVEHARRFCRRFPGHPCRGKQLPQFNDIGSMLNGILLVDSGKWLKVPVFINIRDP LSGINSDLKNA
LNGIQVQFGQISQFANNIRNICQQMNCKQQQKNVQMKAQLKQTVDFEKKVFGNNVADKMNLRFDRTLQKQALLEKAQLKGVVAPEDN
GVFDKDLLTETQANFNLNKGEGGAI PMPGSAKAKRASIFFEQNLIQKWPSTSP IPTYFSSLDNLDQNDVRGAISEIEQKTCIRFKY
FASPCKGNHINYQKVNPSFCGLSYIGRVEPANPVYLSFCQNGRGIIVHETMHALGVNHQHLRMDRDKHIVDWSNINPQQYDAFVVADS
KLYTTYGVKYAYDSIMHYNAYTGAVNIAPKPTMIPLVNQANIGLLGQRAKMSNADVEILNKMYCKSAGCDDKNVYCGAWALQDLNPNHN
VWMSNCRKSCNFC

>C_K11G12.1 wormpep prediction

MQQINFRMTPSLVFLIVIVVVEGQGWPRWDRFNHPGNFGNWGNNWGTQRNQEPHDI PPPVPPPGFRGNDRFNGNI IKVVEIIDLK
SKNRGDI LDFDKVHKKHRLGRKEWKGVKQFCHRFPGHPNCRGKVPDQKELEEMIGQFSGKGI GRFLKRVPKIYIEDPLARVDPKLG
FLENAGRGFGHVSSEHVNLKRDICKRRKREQPE SAKKTRELFTQKLADFETKIAGKDKTDSVQLRFDRTLQIKEALLEKGNLTADIVPVD
NGVYDLDTMLTEEQANILLNENLAKAGVGDDEIPLPADDTLDEDDSTNSASGAAPGSSRLKKSALYFEGNLIKKWDPSPIRYVLDSSLE
DLKNDVRAAIYEIEKNTCIRFKELSSPPTGSHIVYKVDSPFCGLSYVGRADPANPVYLSFGCDNNGKVAIHEMHALGVHQLRNDR
DQFITINWSNIDPQQYDAFVVVDSKLYTSYGVKYAYDSIMHYNGYTAQAQNIAPTMNPKTNSAVNLKVLGQRQKMGTTDIELLKKMYCQPG
CDDKNVYCGAWALKDLCKNPGHDQYMAANCKKSCGLCAIGK

>C_R151.5 wormpep prediction confirmed by EST data (full lenght)

MHKIFIIFGLLSLCAAHSLRDLNSNKDEEDPPSSAPGVRKRRMSEEDQKTVDYMDKLNK LADEKHPEEIERHKNPELVADRKRDSVLNP
EEQKGFQGDIVLYPEQAKALYEQALTEGKTRVRRKFIGSNLRRWDASRPIIYAFDGSHTQREQRILELALHWHNITCLNFRNDQANS
NRIVFTDVGCCASNVGRHPLGEEQLVSLAPEICIRLGVIAEVAHALGFWHEQSRPDRDQYVTVRWENIDKDSKGFQFKEDPDDVDNAGVPY
DYGSIMHYRSKAFSKFDDLYTISTYVTDYQKTIQQRDQLSFNDIRLMNKIYCSAVCP SKLPCQRRGYTDRRCDCRCPCDGFQYCEQVM
PGYGATCGGKISLTRSTTRISSPGYPREFKEGQECSWLLVAPPGHIVEFQFIFGEFEMYCKIRHSLCMDYVEVRNSTDFANTGMRYCCYGT
PTRIRSATDMVVLFRSFYRGGKGFPEARARAVPEAGNWSWSPWTAC SATCGACSRMRTRTCCPGNACSGEPVETQICNTQACTGMCAQK
REEEGQCGGFLSLLRGRVRCRQEKTVMAPCENACCPGFTLQRGRCVRYLIRCKKHFTLIIIFVLKN

>C_T02B11.7n new wormep prediction release WS73: 08 Feb 2002

MRRFFICYIGFLSIFLDFILADKDNNEEERDRKFDWKFENENKPEHEHETVIVPKLPDGSYFWKWTWNSRINSTTAATPTSTVTTSTSAPT
 TSPRVYKLGSEARKSLRKALRGVPEPEKRRKQLKKMGKMMKIPKIKTKESNKLHKSyrkVKITENPPALDMFEVNERAGLNEYLFQGDINL
 NNNQIAKISSEQSSKSRKRQIDNLAQFWPGKVVYYFDSGLTTTVQQIVRDAITFLESNTCLKFELNSTATNRFSGVGCYSDTGMGG
 EQTLSLGCGEVTGTAHEIAHTLGLFHTQMRSDRDDYVITDLTDVPESSQNFIKLTEATSTNLVDYEGSFMHYSGRANPTYSTPLKGL
 SRFVSSGGVDSIVPKDPVMVYTMGGRIIVTFLDLKMLNTHYSKSCPTILSCNGGFTNPANCSVCICPYGFGALCTERTDYCGGSTLTATD
 TWQQETTYTFGNASNSATARPSAVYCNHWIQAPVGKQIQFRIDSTYNTQCVYGCTFNGVEPKLKSMDTITQARYCCDEFNAEIMTADFGVNP
 MPVFSFNRYYKTTYTWSYRYVDSNVTACADTSDKATCLSLKSAKEQGCSTYDTAQLKVMCAATMDLCKGKVASDDGTCKDRFPKSCSTYST
 NGMCTQQPPLAAEFSCAETCGFCTNPV

>C_T04G9.2i Intronerator prediction g-X-2732

MREYVLIPLVAVPFAAILGPDIPPELPLNDENFFDRSHSEYETVLTPEDEFELGTRITAAMAHDNGDDIWDSDAMYSKDRFEGDIANDNL
 NASTAELFANGSGKSEDGKWNAINKRLQLWPEGRIPYTISSQYSSYSRSLIAASMQEYASHTCIRWVPKEAADVNVVHIYPDRGCYSMV
 GKMGGKQSLSLGSGCQKGIILHELMHAVGFFHEQSRTRDRDDHITIMWNNIQAGMQQFEKYGHTIQSLGTGYDYGSIMHYGTFKAFSRNG
 QPTMIPKKGATIQQRNFSKVDKFKINTLYGCPVEGEKPTTSAPTSGPIVITVKPVVITTKGPPVIQTVSPAVPLKPESECRNLRGDCDAD
 AKQGWICIRNPGWMRANCPISCGMCIPTKETQKPYVQTTTQAATTTARPQKPVTPQIQPLPPVPLPPTTPEDCEDLRVDCLVLSQRYCKI
 SQNFMKSYCAKSCGCFKPPPEIIPDNRPTVVTTTRPLVTLPPAVIRSRSPAPPVSTTTKAAPTTSTTSAAPYSPTPLPSECSDRKHFCSHW
 KSAGFCEGIFMNYMKKNCASPAGCLC

>C_T11F9.3i Intronerator prediction g-V-2325

MKIVTNVLLVALIGVPSVLSDRHITRDKRQAMRDYAKWENNKMSLFYFNLPLEMQAMFRDAINYLENHTCLKFEYENAEAVRIRKGNCG
 YSLYGMHAGEVQDLTLDYNCASFQTAVHEIMHALGIAHGQARSDDYLDIVDSTNSNDGIENTENLVFPDYGSVMLYARDPHSDKRIPIIDP
 EYNFTMGSLRVAFYDMVLLNKFYGCNCDNHPRKLDCKNGGYQNPANCEECLCTDGFNGQLCDQHEGVVLEAKKEWDASGVRNRYRKGIEY
 NTPMEYTYFALTADPEGSTIEVRIITKLSGFCQHTCDYNGVELKYKTDRIIVSPLVCCDNDNLWNKTRSSSTNNPFI IAKYGNRTPHFEFEY
 RYIPGNATAAPEENN

>C_T11F9.5i Intronerator prediction g-V-2326

MRVGGSPETKRLERSKRQALRMDNEPRWRPRTINYYFFDEQRFDENSRAVLRAMEKISNHTCIKFSPKDARIKLRIVSDKGCQAAIGRVGG
 DQYLSFPPTSCYSYGSASELHVIGFLHSHQRADREYLLKLNLPWRLNDFWQTMQYKYLQDQWVIVPYDYGSIMQYHDSIEIGIQIQEIV
 LNAIVLSVLGADCSDVAKPAVIVESSEEWKNITLNLDAGYMSLQNGTKLHQIDFVFRYLSISAPANKTIEVDIREITGVECNYGCIYGGIE
 VKTHEDRRMTSPRLCCKNDNEIYKSRNPTIVMAFNSEGLDKYNIYFRFTD

>C_T11F9.6i Intronerator prediction g-V-2327

MKSFILLSILQECYKDIVARIGGRNVAEKIGGARHRRQVLRIGSDEERRHKWFNNTVHYFYFEEFNDFTVKESILRAMELISNHTCIKFS
 STEPSEKSI RMESDSTTIACYAEIGQVRENQLFSFNSDCYAGVAVHELHLSLGFIAHQRSDRDQYLEFKKNLDELNQTQYQYKIWEYQ
 EILVYDMEDESSAKYPYRESMPRAERVSMKALQDEWRHLFITRAEGENNIHNPKIPEDTNDFFDDMPIECNESEESITGDREVFVSHHN
 EDEDQEDSSQICEESEPELFEFMDLSLDISGLYLPQTLTGRIDLNDEFVASIEVEDAEKSKELRRRRERPLALRQFHALSRAMANL
 RTYHYDVTTEKIERCPMRNSCKGHLKWEQMRKYRAQCKKNFKDPSLATVEMIDEYYPVRYRTMANTMGSAIVAFYDYLINKYVECS
 ANNLSCKNHGYPNPNSCNCPCPYGFGGADCSQRAEPGATFQATETWQNVITISLDAGYRYLENNQKLPQVDFIYQFLWIMAPANKTTQIRV
 EKFEVGGKCLPGCIRGGVEIKTNEPRLTSPRLCCEETS

>C_T23F4.4 wormep prediction

MFNGKIHTSRKIAARKVSRGLSVNQILPFFPLLITSLHAI PRGRRAVRNRNEGDINSLVGVQYLYQGDIAVVKSRARRAVIRQKHKKW
 KLPMYPSPDRNRFREFLREKNRSNYTSRSRQVLEAMQFWEKTCVTFHENRYVYPHVSI FEGNGCWSFVGKQPSLREQSLERSCTDHTF
 VVAHEIAHTLGFYHEHARGDRDQFISIDYSNVNPNTFAFAKESKQLDQEAAYEYGSVMHYSVDQFVAVNTRNPVIYARDQKFAQAMGNR
 MRATFQDVSRRMNVLYNCHERCANTLNRCCQGGYPAPSDCSQCVCVDFGFGNFCETIEAHSVQKQKNSDCGGVWLAETSQTFYGAVRTRVH
 SNSVLPTEPHCFWHIRASQKXSIEIQIKNIISPCSMSCSFNALELKLNSFTMTGPRFCCDEHIYNRYSQPKVFQSEGFLAVI GAYARYDYL
 DFNIEYRAV

>C_T23H4.3 wormep prediction

MDIKQLLLSII LTVSVVNGRGRINIYGAENKSDIVQLRGPAEQLVYSSPIRERRP IFRNALLSNSPLRWSKMQDLGNYLIPYVISGNY
 DTVERDTIKTAMEKIANNTCIRLIPRTNQPDYAE INNKQGQCYASIGRFPKQNVVLESNDQSCIQEDTVIHELPHVIGLWHEHMRADR
 DAFINLVLYKNI EPAQYQFEKLSRSDATTYSVPYDYNVSMHYDENAFKPKGISMMTKDSKFQKVI GHPKDASSNDYKVKCAIYHCSKCMH
 QDFQQIVEQEHIELNPNIIITNAPVQQGDSCTDRLGICPMLKSREMLNCKVMATFCCSSCSAPTSTTTTSTGTPSDGSLWQRKISIFQ

>C_T24A11.3i Intronerator prediction g-III-483

MTSSLVLILAPLALVAIGEAAFGNSKIFEIPGLEVMASDKYPHFTTIIETVSRTKVHRHREVIAGQIYDWNYSYEPFQI WGGDYNFQSLI
 RRGIRMWEDSTCLRFKENQSRDAIRYVLEKGDSCFTEYIGRNGGHQDIIIGSECAEEYVVAHETGHALGFWHTRQRPDRDRHISINWKNV
 MEEATASFMPFRSMLQAFGIRQVSPRRVPYDYGSLMHYHTVAHAVKVSDFITVPKELKYVTMTGTEKMAFLDAKVINDIYCPNACQGRNHL
 NCLAGGYPDPNNCNCRCPEGLGGPDCGRLQSPSPCGGEI HASDQWQTLSSPSPGRDVHCYWRISVPEGRVRFRLSDGEFPCS YGCQSYVEI
 KHKLDVRLTGFRSCCYRPKEDTVSESNQIFVYIHPNGRTARFSLRFRRA

>C_Y95B8A.1 wormep prediction

MFQKLSFGTILRVTRNTFFVSSFLKFRFIPNIQIEMYMIQGCYIPAKSDKILETIFSDAPPVYRQPRTKAEKIERFRNIARTFSPFVYEV
 NTPPAPHDFDNFIWQONAPAVTPEPFTFAPFSFPTLATVAPPAPGPGGPTLEPFLPTTASPCLLAHNTARMIREIASFSDGGRSRDQDFGAV
 QTLMQAFFEAVSSGNNGGAGAAAGAGTALGDAPMLQAHRDGTELGANRALTNKLFESDMVLTVKQMKAI VLAQAQEARNPGRKRRKVI TGS
 VYRWKSVIPFRFKGGDAKWKLLIREGLGLWEKETCVRWSENGPGKDYVIFRFGSGCYSSVGRGTGGSQLSISIGYGCEDKGI VAHEVGHSLGF
 WHEQSRPDRDDYIHLRKDWIKGTDGNFEKRSWEEIEDMGVYDVGSMHYGSNAFTKDWQITIEETKDSRYQGTIGQRQKLSFIDVKQVN
 RLYCNSVCPVALPCMGGYDPNNCAVCKCPDGLGGKCGRAAKGTDHDKCGGELTATAEWQEMVYKGRKTCNWKVSPSGGRVRLVLT
 RYQCAPACKAYIEIKHNTDFQQTGFRVCCFNKTYDVISDQSEALILSNANIVDYEVSYKQLWIQGTTPPPFLSIERTFFQTTTEKLFHHRN
 PRRPGCPEKRIHSAEWRTRAAPSKSSSKRSQRSETRIGHWRVLLVLRNMVSRQYLYRIMESSFPFGTIFAIVRGIYIYICKFCTIPKI
 QFFHHNFQNLIFHNYSG

6.3. Sequenzvergleich der Protease-Domäne der Astacin-Protein-Familie

C. elegans und *Drosophila* Proteine sind nur exemplarisch berücksichtigt.

		*	20	*	40	*	60	*	80						
AA_AEA	:	NAGG	-----	SMAWPKASEGFVVIPYVVTQYD	--	KENIDIAIDAMAEFDEITCVRVFPVPR	-----	RTETDFLLIDSR	:	65					
AA_AST	:	-AAILG	---	DEYLWGG	-----	VIPYTFAGVS	--	GADQSAILSQMLEEKT	CIRFVPR	-----	TTESDYVEIFTSG	:	61		
AC_TBL1	:	-ATPAD	---	RSKLDHG	-----	VIPYVIESNYS	--	GENKDLFLKAMRHWNLTCLVFKDKG	-----	PEDTNYLFTQTD	:	63			
ACA_MTP1	:	NY	-----	PRTLWSKG	-----	VYFHFHRNAT	--	PEVRSVFKGAKLWMKDTCIDFES	-----	NSAPDRIRVFKEN	:	58			
AJ_EHE10	:	NAMKCW	--	NNQCLWKTSSDGRVEVPYTVSNEFF	--	YQKKRIENAMKTFNTETCIRFVPR	-----	SSQRDFISIESRD	:	69					
AJ_EHE12	:	NAMKCR	--	NNRCLWKKSSNGLVEVPYTVSRQFS	--	YQRRIRVKAMKTFNTQTCIRFVPR	-----	SRQRDIYSIKSRG	:	69					
AJ_EHE13	:	NAINCW	--	NNQCLWRKSSDGLVEVPYTVSSEFS	--	YHKRIENAMKTFNTETCIRFVPR	-----	SSQRDFISIESRD	:	69					
AJ_EHE14	:	NAMKCR	--	NNRCLWKKSSNGLVEVPYTVSRQFS	--	YQRRIRVKAMKTFNTQTCIRFVPR	-----	SRQRDIYSIQSRG	:	69					
AJ_EHE4	:	NAMKCW	--	NNQCLWRKSSDGLVEVPYTVSNEFS	--	YHKRIENAMKTFNTETCIRFVPR	-----	SSQRDFISIESRD	:	69					
AJ_EHE6	:	NAMKCW	--	NNQCLWRKSSDGLVEVPYTVSNEFS	--	YHKRIENAMKTLNTECIRFVPR	-----	SSQRDFISIESRD	:	69					
AJ_EHE7	:	NAMKCR	--	NNRCLWKKSSNGLVEVPYTVSRQFS	--	YQRRIRVKAMKTFNTQTCIRFVPR	-----	SRQRDIYSIKSRG	:	69					
BR_TLD	:	-AATSR	---	AEKIWPGG	-----	VIPYVIGNFT	--	GSQRAMLKQAMRHWEKQTCVTFIEK	-----	TDEESYIVFTYRP	:	62			
C_HCH1	:	SMTSF	---	LSEKWSFP	-----	VPIYIDTSSG	--	VMTN	AVLAGVAKWEQETCARFTRLN	-SYSS-	SSRQNALRFISGN	:	65		
CC_Neph	:	NADPCT	---	ARRCKWRKSRNGLVVPYKISNQYS	--	PDEIKVIKQGLNSFTGIS	CIRFVPH	-----	NGQRDFLNIQSDS	:	69				
DM_TLD	:	-AVTVR	---	KERTWDYG	-----	VIPYEDITIFS	--	GAHKALFKQAMRHWNFTCIKFVERDP	-----	NLHANYIYFTVK	:	64			
DM_TOK	:	-AVTAK	---	KERIDWYG	-----	VIPYEDIGNFS	--	GIHKALFKLAMRHWNSTCIKFVERDP	-----	EIHNYIYFTVRS	:	64			
FM_Flavast	:	STIVS	---	SFIKTPWPNAT	-VYITLPSQGLSTQAYNTFLTNINKAFDMISSKTSVKFVQR	-----	TNQEYITFTYST	:	69						
GG_BMP1	:	-AATSR	---	PERVWPDG	-----	VIPYVIGNFT	--	GSQRVFRQAMRHWEKHTCVTFIER	-----	SDEESYIVFTYRP	:	62			
GG_COLLOID	:	-AATSR	---	TERIWPGG	-----	VIPYVIGNFT	--	GTQRAMFKQAMRHWEKHTCVTFIER	-----	SDEESYIVFTYRP	:	62			
GM_M15242	:	NGLR	---	DEVWRWPN	-----	TVYKFTVFD	--	EDHNYILRGMKIEEIS	CIRFEEADA	-----	TT	-PNYVNI	ITGVF	:	66
HE_HEA2	:	NAFAS	---	IKGGRWPAGK	-----	IPIYVVEGSIS	--	SRGISAINAAIADYHKFTCLRFHRR	-----	TPRAPHYSFYKGG	:	63			
HS_BMP1	:	-AATSR	---	PERVWPDG	-----	VIPYVIGNFT	--	GSQRVFRQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
HS_HCH1	:	-LLSA	---	TSNKWPMGGSGVVEVFLSSKYD	--	EPSRQVILEALAEFERSTCIRFVTY	-----	QDQRDFISIPMY	:	66					
HS_MEP	:	NGLRD	---	PNTRWTFP	-----	IPYILADNLG	--	LNAKGAILYAFEMFRLKSCVDFKPY	-----	EGESSYIIFQQFD	:	61			
HS_MEPB	:	NSIIG	---	EKYRWPHT	-----	IPYVLEDSLE	--	MNAKGVLNNAFERYRLKTCIDFKPW	-----	AGETNYISVFKGS	:	61			
HS_TLL	:	-AATSR	---	TERIWPGG	-----	VIPYVIGNFT	--	GSQRAMFKQAMRHWEKHTCVTFIER	-----	SDEESYIVFTYRP	:	62			
HS_TLL2	:	-ATTSR	---	TERIWPGG	-----	VIPYVIGNFT	--	GSQRIFKQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
HV_HMP1	:	-KGV	---	LDRIWPGA	-----	QIPYVDFNFD	--	NRRRQLVNGAITSYNQHTCVRFPAPR	-----	TNQDFVHMVNEQ	:	61			
HV_HMP2	:	-AALS	---	SSILWPNK	-----	VVPWSITKQLENTAETAFGLMAAFREWEERSCLTFKRR	-----	TDEKDIIEFFQGS	:	66					
MM_BMP1	:	-AATSR	---	PERVWPDG	-----	VIPYVIGNFT	--	GSQRVFRQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
MM_MEP	:	NAMRD	---	PSSRWKLP	-----	IPYILADNLG	--	LNAKGAILHAFEMFRLKSCVDFKPY	-----	EGESSYIIFQKLS	:	61			
MM_MEPB	:	NSIIG	---	DHWRWPHT	-----	IPYVLEDSLE	--	MNAKGVLNNAFERYRLKTCIDFKPW	-----	SGEANYISVFKGS	:	61			
MM_TLL	:	-AATSR	---	TERIWPGG	-----	VIPYVIGNFT	--	GSQRAMFKQAMRHWEKHTCVTFIER	-----	SDEESYIVFTYRP	:	62			
MM_TLL2	:	-ATTSR	---	TERIWPGG	-----	VIPYVIGNFT	--	GTQRIFKQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
OL_Alveo	:	NA	-----	VNMLWAQG	-----	VIPYIISMELA	--	HRQ	-SEIKFAFSMITSATCIRFQER	-----	TYEANYLIDKDG	:	58		
OL_HCE1	:	NAMKCW	--	SSSCFWKKASNGLVVPYVISSEYS	--	GGEVATIEGAMRAFNGKTCIRFVRR	-----	TNEYDFISVSKT	:	69					
OL_HCE2	:	NAMKCW	--	YNSCFWKKASNGFVVPYVISSEYS	--	RGEVATIEGAMRAFNGRTCIRFVRR	-----	TNEYDFISVSKN	:	69					
OL_LCE	:	NAMKCFGADPSCRWPKSSNGLVVKVYVSDNYE	--	SDEKETIRNAMKEFAEKT	CIRHFVPR	-----	NNERAYLSLEPRF	:	71						
OO_MP	:	QAFKDKHY	-PKTLWSQG	-----	VNYVFDHMAS	--	KQMKSVFVKGAKWWEKDT	CINFTE	-----	RSADRIDMVPQK	:	64			
PC_PMP1	:	NAFAS	---	IKGGRWPAGK	-----	IAYQIERSIG	--	SGRTAIAANAIAQYHKHTCLRFHRR	-----	TNERSHISFYKGG	:	63			
PL_BP10	:	-ATIY	---	ESQRWPKYK	-----	IPYVISPSSS	--	QSS	-LIRNAMDHWEQNTCLRFEPRTSSHSRQLGHNA	YLSFFRGS	:	67			
RN_MEP	:	NALRD	---	PSSRWKPP	-----	IPYILADNLG	--	LNAKGAILNFAFEMFRLKSCVDFKPY	-----	EGESSYIIFQQFS	:	61			
RN_MEPB	:	NSIIG	---	DNYRWPHT	-----	IPYVLEDSLE	--	MNAKGVLNNAFERYRLKTCIDFKPW	-----	SGEANYISVFKGS	:	61			
SP_BMP	:	-AVTAR	---	PERRWTDA	-----	VIPYEDIGNFT	--	GSQRAMFKQAMRHWEKHTCVTFIER	-----	SEHNHIVFTYQA	:	66			
SP_SPAN	:	-ATIY	---	ESQRWPKYK	-----	IPYVISPSSS	--	QSS	-LIRSAMDHWEQNTCLRFEPRTSSHSRGLGHTSYISFFRGN	:	67				
TR_MP	:	NADACT	---	SRGCMWAKSSDGKVVVPYVIAISHYS	--	SRERSIERGLLSFHDVSCIRFVPR	-----	RNQRDFLIVSDS	:	69					
XL_BMP1	:	-AGTSR	---	PERVWPDG	-----	VIPYVIGNFT	--	GSQRIFKQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
XL_BMP1b	:	-AATSR	---	PERVWPDG	-----	VIPYVIGNFT	--	GSQRIFKQAMRHWEKHTCVTFIER	-----	TDEESYIIFTYRP	:	62			
XL_UVS2	:	SAINDA	---	RFLWPKSADGIVPVPYVLSYSYN	--	ADQLALFKKAIQEF	EALTCVRFVW	-----	TTEVNFLNIMSNG	:	67				
XL_XOL_R	:	-ASTSR	---	PERVWPDG	-----	VIPYVIGNFT	--	GSQRIFKQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
XL_XOLLOID	:	-AATSR	---	TERIWPGG	-----	IIPYAIAGNFT	--	GTQRIFKQAMRHWEKHTCVTFIER	-----	TDEESYIVFTYRP	:	62			
C_C05D11.6	:	NAIKQ	---	IYRRWPN	-----	EIPYTLSSQYG	--	YSRVSIVANAMNEYHTKTCVDFVARDP	-----	SKHHDYLIWHPDE	:	64			
C_C38E9.2n	:	-ATAK	---	KERIWPEG	-----	IIPFYIASNFS	--	GEHQHLFLRAMRHWNFTCVSFPVPRQ	-----	PHHKYITFTYDK	:	63			
DM_CG15254	:	NGLRN	---	ETFRWPNR	-----	IVYYYINRDID	--	TEHRNHILRGIRIEQSSCLVFK	EAT	-----	TDQEYVNVVTS	:	63		
		*	100	*	120	*	140	*	160						
AA_AEA	:	-GCKSIIGKLG	-EL	-QKISLEKM	-----	GCNMTG	--	IIQHELHALGIFYHEHSRSDRDTYVKIMWENISPDNVRMFDKEMN	:	136					
AA_AST	:	SGCWSYVGRIS	-GA	-QVVSQAN	-----	GCYVHG	-	TIIEHLMHAIGFYHEHTRMDRNYVTINYQNVDP	SMTSNF	DIDTY	:	133			
AC_TBL1	:	CGCCSFVKGHSGA	-QV	ISLKG	-----	GCYVHG	-	TVVHELGHVVGFWHEHNRPRDRKYVQIRKNI	MPGKESEFN	ILDE	:	135			
ACA_MTP1	:	-GCWSYVGRIG	-GE	-QDLSLGE	-----	GCQSVG	-	TAHEIGHAIGFYHHTHARHDRDNFI	TFNAQNVKPD	WLDQFTLQTP	:	128			
AJ_EHE10	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDNKYVRI	NWENAVPHTI	YNFQEQDT	:	140			
AJ_EHE12	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDNKYVRI	NWENAVPRTI	YNFQEQDT	:	140			
AJ_EHE13	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDEYVRI	NWENAVPHTI	YNFQEQDT	:	140			
AJ_EHE14	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDEYVRI	NWENAVPRTI	YNFQEQDT	:	140			
AJ_EHE4	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDEYVRI	NWENAVPHTI	YNFQEQDT	:	140			
AJ_EHE6	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDEYVRI	NWENAVPHTI	YNFQEQDT	:	140			
AJ_EHE7	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDNKYVRI	NWENAVPRTI	YNFQEQDT	:	140			
BR_TLD	:	CGCCSYVGRRNGP	-QA	ISIGK	-----	NCDKFG	-	IVVHELGHVIGFWHEHTRPDRDDHVTI	IRDNI	QPGQYENFLKMEP	:	134			
C_HCH1	:	-GCYSNIGKVSFRP	-QD	VSIGW	-----	CTSLG	-	TVCHEIGHALGIFYHEHTRSDRDEYVRI	QVILQNI	QDMSQFTKQSA	:	136			
CC_Neph	:	-GCYSYLGRTG	-GK	-QVVSQAN	-----	GCYVHG	-	IIQHELSHALGIFYHEHTRSDRDNKYVRI	NWENAVPRTI	YNFQEQDT	:	140			
DM_TLD	:	CGCCSYVGRRNGP	-QA	ISIGR	-----	NCDKFG	-	IVVHELGHVIGFWHEHTRPDRDDHVTI	IRDNI	QPGQYENFLKMEP	:	136			
DM_TOK	:	CGCCSYVGRRNGP	-QA	ISIGR	-----	NCDKFG	-	IVVHELGHVIGFWHEHTRPDRDDHVTI	IRDNI	QPGQYENFLKMEP	:	136			
FM_Flavast	:	-GNSSPLGWKVN	-RVNGIKIYN	-----	TTYPA	-	IIAHEIMHSMHGEQCRPDRDQYI	IVDTN	RAQDGT	RHNFNLNDY	:	139			
GG_BMP1	:	CGCCSYVGRRNGP	-QA	ISIGK	-----	NCDKFG	-	IVVHELGHVIGFWHEHTRPDRDDHVTI	IRDNI	QPGQYENFLKMEP	:	134			
GG_COLLOID	:	CGCCSYVGRRNGP	-QA	ISIGK	-----	NCDKFG	-	IVVHELGHVIGFWHEHTRPDRDDHVTI	IRDNI	QPGQYENFLKMEP	:	134			
GM_M15242	:	GGCYSYVGRWLN	-GA	-QAYNLEMYALDTCFRLG	-----	TVVHELGHVIGFWHEHTRSDRDNKYVRI	-----	QVILQNI	QDMSQFTKQSA	:	140				
HE_HEA2	:	-GCSSPVGYRS	-GRMNR	ISLAGG	-----	CWNGK	-	IVMHEIGHSIGIYHEHTRSDRDNKYVRI	ILWNNIS	SRAKFNPNKQRT	:	134			
HS_BMP1	:	CGCCSYVGRRNGP	-QA	ISIGK	-----	NCDKFG	-	IVVHELGHVIGFWHEHTRPDRDDHVTI	IRDNI	QPGQYENFLKMEP	:	134			
HS_HCH1	:	-GCFSSVGRSG	-GM	-QVVSQAN	-----	TCLQKGRGIVLHELMHVLGFWHEHTRADRDYIRVNVN	-----	ELPGE	INFKSQS	:	138				

