

Eine phylogenetische Analyse von Serpin-Genen aus eukaryotischen Genomprojekten

Von der Technischen Fakultät
der Universität Bielefeld
zur Erlangung des Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)

genehmigte
Dissertation

von
Dipl.-Biochem. Olaf Krüger
geb. in Bielefeld

2003

1. Gutachter: Prof. Dr. Hermann Ragg

AG Zelluläre Genetik

Technische Fakultät, Universität Bielefeld

2. Gutachter: Prof. Dr. Andreas Dress

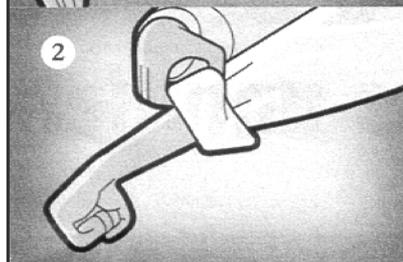
Max-Planck-Institut für Mathematik in den Naturwissenschaften,

Leipzig

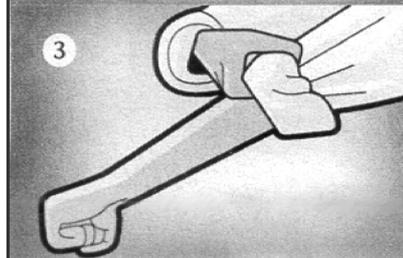
Tag der Disputation: 12. Dezember 2003



Jammern



hilft



nicht.

Inhaltsverzeichnis

1. ZUSAMMENFASSUNG	1
2. EINLEITUNG.....	2
2.1 Serpine: Struktur und Funktion	2
2.1.1 Serpine.....	2
2.1.2 Struktur der Serpine.....	3
2.1.3 Reaktionsmechanismus der Serinproteasen.....	5
2.1.4 Die Inhibierungsreaktion der Serpine.....	7
2.1.5 Physiologische Funktionen der Serpine	9
2.2 Die Evolution der Serpin-Superfamilie - Stand der Forschung.....	11
2.3 Molekulare Evolution	13
2.3.1 Methoden der molekularen Phylogenetik.....	13
2.4 Struktur eukaryotischer Gene	15
2.4.1 Alternatives Spleißen	17
2.5 Zielsetzung der Arbeit	19
3. MATERIAL.....	20
3.1 Verwendete Geräte	20
3.2 Chemikalien und Verbrauchsmaterialien.....	21
3.3 Verwendete Kits.....	22
3.4 Verwendete Enzyme	22
3.4.1 KlenTherm DNA-Polymerase (Genecraft).....	22
3.4.2 <i>Taq</i> DNA-Polymerase (Qiagen).....	23
3.4.3 <i>Pfu</i> DNA-Polymerase (Promega)	23
3.4.4 <i>Taq/Pfu</i> DNA-Polymerasen Mix	23
3.4.5 Restriktionsenzyme	24
3.5 Verwendete Puffer und Lösungen	24
3.6 Primer	25
3.7 Bakterienstämme und Medien.....	27

4. METHODEN.....	29
4.1 Polymerase-Kettenreaktion.....	29
4.1.1 Standard PCR-Ansatz.....	29
4.1.2 Kolonie-PCR.....	30
4.2 Reverse Transkription.....	31
4.2.1 Reverse Transkription mittels „Omniscript RT“-Kit.....	31
4.2.2 Reverse Transkription mittels „GeneRacer TM “-Kit.....	31
4.3 Analyse und Präparation von DNA.....	33
4.3.1 Analyse der PCR-Produkte mittels Agarose-Gelelektrophorese.....	33
4.3.2 Aufreinigung der PCR-Produkte mittels „QIAquick Gel Extraction“-Kit.....	33
4.3.3 Aufreinigung der PCR-Produkte mittels „QIAquick PCR Purification“-Kit.....	34
4.3.4 Plasmidisolierung mittels „QIAprep [®] Miniprep“-Kit.....	34
4.3.5 Anlegen von Glycerinkulturen.....	35
4.3.6 Bestimmung der DNA-Konzentration.....	35
4.4 Klonierung und Klonanalyse.....	36
4.4.1 TA-Klonierung mittels „TOPO TA Cloning [®] Kit for Sequencing“.....	36
4.4.2 Plasmid-Schnell-Charakterisierung.....	38
4.4.3 Restriktionsspaltung.....	38
4.4.4 DNA-Sequenzierung.....	39
4.5 Auswertung von DNA-Sequenzen.....	40
4.6 Identifikation und Analyse von Serpin-Genen.....	41
4.6.1 Serpin-Suche mit AGREP.....	42
4.7 Phylogenetische Analyse.....	44
4.7.1 Alignments.....	44
4.7.2 Maximum-Likelihood-Analyse.....	45
4.7.3 Neighbor-Joining-Analyse.....	46
4.7.4 Maximum-Parsimony-Analyse.....	47
4.7.5 <i>Bootstrap</i> -Testverfahren.....	48
5. ERGEBNISSE.....	49
5.1 Analyse der Serpin-Gene von <i>Caenorhabditis elegans</i>.....	50
5.1.1 Modellorganismus <i>Caenorhabditis elegans</i>	50
5.1.2 Die Serpin-Gene von <i>Caenorhabditis elegans</i>	51
5.1.3 Chromosomale Lokalisation der Serpin-Gene von <i>C. elegans</i>	53
5.1.4 Genstrukturen der Serpin-Gene von <i>Caenorhabditis elegans</i>	55

5.1.5	Intronpositionen der Serpin-Gene von <i>Caenorhabditis elegans</i>	56
5.1.6	Phylogenetische Analyse der Serpine von <i>Caenorhabditis elegans</i>	57
5.2	Analyse der Serpin-Gene von <i>Drosophila melanogaster</i>	60
5.2.1	Modellorganismus <i>Drosophila melanogaster</i>	60
5.2.2	Die Serpin-Gene von <i>Drosophila melanogaster</i>	61
5.2.3	Chromosomale Lokalisation der Serpin-Gene von <i>Drosophila</i>	66
5.2.4	Genstrukturen der Serpin-Gene von <i>Drosophila melanogaster</i>	67
5.2.5	Intronpositionen der Serpin-Gene von <i>Drosophila melanogaster</i>	70
5.2.6	Phylogenetische Analyse der Serpin-Gene von <i>Drosophila melanogaster</i>	72
5.3	Analyse der Serpin-Gene von <i>Arabidopsis thaliana</i>	75
5.3.1	Modellorganismus <i>Arabidopsis thaliana</i>	75
5.3.2	Die Serpin-Gene von <i>Arabidopsis thaliana</i>	76
5.3.3	Chromosomale Lokalisation der Serpin-Gene von <i>Arabidopsis</i>	82
5.3.4	Genstrukturen der Serpin-Gene von <i>Arabidopsis thaliana</i>	84
5.3.5	Intronpositionen der Serpin-Gene von <i>Arabidopsis thaliana</i>	85
5.3.6	Phylogenetische Analyse der Serpine von <i>Arabidopsis thaliana</i>	86
5.4	Analyse der Serpin-Gene von <i>Ciona intestinalis</i>	88
5.4.1	Modellorganismus <i>Ciona intestinalis</i>	88
5.4.2	Die Serpin-Gene von <i>Ciona intestinalis</i>	89
5.4.3	Genstrukturen der Serpin-Gene von <i>Ciona intestinalis</i>	92
5.4.4	Intronpositionen der Serpin-Gene von <i>Ciona intestinalis</i>	93
5.4.5	Phylogenetische Analyse der Serpine von <i>Ciona intestinalis</i>	94
5.5	Interspezies-Stammbäume	97
5.6	Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten.....	103
5.6.1	Experimentelle Untersuchungen zur Expression von <i>Dm-Spn4</i>	104
5.6.2	Die Serpine <i>Ce-Srp-7</i> und <i>Ce-Srp-9</i> von <i>Caenorhabditis elegans</i>	113
5.6.3	Das Serpin-Gen <i>Ci-Spn-10</i> von <i>Ciona intestinalis</i>	115
5.6.4	Weitere Insekten-Serpine mit multiplen RCL-kodierenden Exon-Varianten.....	116
5.6.5	Alignment der Serpine mit multiplen RCL-kodierenden Exon-Varianten	118
6.	DISKUSSION	121
6.1.1	Vergleich der phylogenetischen Kriterien.....	122
6.1.2	Indels als phylogenetische Marker	123
6.2	Die Evolution der Serpin-Gene von <i>Caenorhabditis elegans</i>	124
6.2.1	Die potentiellen Pseudogene <i>Ce-Srp-5</i> und <i>Ce-Srp-10</i>	124
6.2.2	Das Serpin-Gen <i>Ce-Srp-9</i>	125
6.2.3	Die phylogenetische Klassifikation der Serpin-Gene von <i>C. elegans</i>	126

6.3	Evolution der Serpin-Gene von <i>Drosophila melanogaster</i>	127
6.3.1	Variationen der Intronposition 352a.....	127
6.3.2	Die phylogenetische Klassifikation der Serpin-Gene von <i>Drosophila</i>	128
6.4	Evolution der Serpin-Gene von <i>Arabidopsis thaliana</i>	130
6.4.1	Die Serpin-Pseudogene von <i>Arabidopsis thaliana</i>	130
6.4.2	Die phylogenetische Klassifikation der Serpin-Gene von <i>Arabidopsis</i>	131
6.5	Evolution der Serpin-Gene von <i>Ciona intestinalis</i>.....	132
6.5.1	Hotspot-Diversifikation bei Serpin-Genen von <i>Ciona intestinalis</i>	132
6.5.2	Die phylogenetische Klassifikation der Serpin-Gene von <i>Ciona</i>	133
6.6	Interspezies-Stammbäume der Invertebraten-Serpine	134
6.7	Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten.....	136
7.	LITERATURVERZEICHNIS	139
8.	ANHANG	150
8.1	Veröffentlichungen	150
8.2	Abkürzungsverzeichnis	150
8.3	Abbildungsverzeichnis.....	152
8.4	Tabellenverzeichnis.....	153
8.5	Sequenzen und Alignments	155
8.5.1	Aminosäuresequenzen der Serpine im FASTA-Format	155
8.5.2	DNA-Sequenz des Serpin-Gens <i>Dm-Spn4</i> von <i>Drosophila melanogaster</i>	161
8.5.3	Aminosäure-Alignment der Serpine von <i>Caenorhabditis elegans</i>	162
8.5.4	Aminosäure-Alignment der Serpine von <i>Drosophila melanogaster</i> (1/4)	163
8.5.5	Aminosäure-Alignment der Serpine von <i>Arabidopsis thaliana</i>	167
8.5.6	Aminosäure-Alignment der Serpine von <i>Ciona intestinalis</i>	168
DANKSAGUNG	169

1. Zusammenfassung

Die Superfamilie der Serpine (Serinprotease-Inhibitoren) umfasst neben inhibitorischen Serpinen auch nicht-inhibitorische Proteine. Eine geeignete Serinprotease spaltet die auf der Serpin-Oberfläche exponierte, reaktive Schleife (*reactive centre loop*, RCL) als Pseudosubstrat und wird dadurch inhibiert. Als Antagonisten der Serinproteasen sind die Serpine an der Regulation einer Vielzahl von physiologischen Prozessen beteiligt.

In der vorliegenden Arbeit wurden phylogenetische Klassifikationen für die Serpine von *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana* und *Ciona intestinalis* erstellt, die auf mehreren unabhängigen Kriterien basieren. Neben den Aminosäuresequenzen wurden auch die chromosomalen Lokalisationen und die Genstrukturen zur phylogenetischen Klassifikation der Serpine eingesetzt. Dazu wurden die in den Datenbanken zugänglichen Informationen zu Genstrukturen, EST- und cDNA-Sequenzen ausgewertet. Es zeigte sich, dass die zur Zeit annotierten Serpin-Genmodelle teilweise fehlerhafte bzw. unvollständige Exon-Intron-Strukturen aufweisen. Die Genmodelle wurden daher mit einer AGREP-Suchstrategie identifiziert und ihre Genstrukturen überprüft und gegebenenfalls vervollständigt. Die ergänzten und korrigierten Daten wurden mit den Baumrekonstruktions-Verfahren Maximum-Likelihood, Neighbor-Joining und Maximum-Parsimony klassifiziert und die so ermittelten Stammbäume mit den Exon-Intron-Strukturen und den chromosomalen Lokalisationen der Gene verglichen. Die vergleichende Analyse der Parameter „chromosomale Lokalisationen“, „Exon-Intron-Strukturen“ und „Aminosäuresequenzen“ führte bei den Serpin-Genen von *Caenorhabditis elegans*, *Drosophila melanogaster* und *Ciona intestinalis* zu einer weitgehend kongruenten, phylogenetischen Klassifikation der jeweiligen Serpine und ihrer Gene.

Die Serpin-Gen-Analysen der vier Genome gewährten interessante Einblicke in die molekulare Evolution der Serpine und ihrer Gene. Es konnten molekulare Evolutionsprozesse, wie die Entstehung von Serpin-Pseudogenen, Hotspot-Diversifikation und alternatives Spleißen beobachtet werden. Im Rahmen dieser Arbeit konnten Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten in *Caenorhabditis elegans*, *Drosophila melanogaster* und *Ciona intestinalis* identifiziert werden. Durch Spleißen wird eine Variante der mehrfach vorliegenden, spezifitäts-bestimmenden RCL-Exons an das gemeinsame Serpingrundgerüst angefügt. Diese Form des alternativen Spleißens stellt somit eine ökonomische Strategie zur Steigerung der funktionellen Vielfalt der Serpine dar. In der vorliegenden Arbeit wurde die Expression des Serpin-Gens *Dm-Spn4* experimentell untersucht. Es gelang, insgesamt zehn unterschiedliche *Dm-Spn4*-Transkripte zu identifizieren, von denen acht unterschiedliche Protein-Isoformen abgeleitet werden konnten. Das alternative Spleißen und die Verwendung alternativer Transkriptionsstartpunkte bieten die Voraussetzungen zur Generierung von Protein-Isoformen mit unterschiedlicher zellulärer Lokalisation und Protease-Spezifität.

2. Einleitung

2.1 Serpine: Struktur und Funktion

2.1.1 Serpine

Travis und Carrell führten 1985 den Begriff Serpin (Abk. für Serinprotease Inhibitor) als Bezeichnung für die Mitglieder einer neu identifizierten Protein-Superfamilie ein¹. Der Begriff Serpin wurde von der Funktion dieser Proteine abgeleitet, die fast alle die Fähigkeit zur Inhibierung von Serinproteasen zeigen.

Gegenwärtig sind über 500 orthologe und paraloge Mitglieder der Serpin-Superfamilie bekannt. Serpine konnten bisher in einzelnen Prokaryoten², einzelligen Eukaryoten³, Insekten, Pflanzen, Nematoden, höheren Tieren und in Viren identifiziert werden. Die Veröffentlichung der Daten aus den Genomprojekten ermöglicht jetzt erstmals, die Organisation der Serpin-Gene in unterschiedlichen Organismen zu untersuchen (Tabelle 1).

Organismus / Genomprojekt	Jahr der Publikation	Anzahl der Serpin-Gene
<i>Saccharomyces cerevisiae</i> , Hefe	1996	0
<i>Caenorhabditis elegans</i> , Nematode	1998	9
<i>Drosophila melanogaster</i> , Insekt	2000	29
<i>Arabidopsis thaliana</i> , Pflanze	2000	9
<i>Ciona intestinalis</i> , Urochordata	2002	10
<i>Homo sapiens sapiens</i> , Mensch ⁴	2001	33
<i>Giardia lamblia</i> , einzelliger Eukaryot	-	1

Tabelle 1: Genomprojekte - Anzahl der Serpin-Gene pro Genom

Für *Ciona intestinalis* und das menschliche Genom wurden vorläufige Draft-Versionen veröffentlicht. Das Genomprojekt von *Giardia lamblia* ist zur Zeit noch nicht abgeschlossen.

¹ Carrell, C., Travis, J., 1985.

² Irving, J., A., et al. 2002.

³ Genomprojekt von *Giardia lamblia* / Krüger, O., unveröffentlichte Daten.

⁴ van Gent, D., Sharp, P., Morgan, K., Kalsheker, N., 2003.

Serpine sind einkettige Proteine, die eine Kernstruktur von ca. 360 Aminosäuren (As) aufweisen. An diese gemeinsame Kernstruktur schließen sich unterschiedliche amino- und carboxyterminale Verlängerungen an. Obwohl die Mitglieder der Serpin-Superfamilie nur ca. 30 - 60 % Aminosäuresequenz-Identität aufweisen, so ist doch die Serpin-Tertiärstruktur hoch konserviert. Die Serpin-Superfamilie umfasst neben Serinprotease-Inhibitoren auch Cysteinprotease-Inhibitoren und nicht-inhibierende Proteine (Abschnitt 2.1.5).

2.1.2 Struktur der Serpine

Die Kernstruktur der Serpine besteht aus drei β -Faltblattstrukturen (*sheet*, sA, sB, sC) und meist neun α -Helices (*helix*, hA - hI). Die größte der drei β -Faltblattstrukturen, sA, besteht im nativen (ungespaltenen) Serpin aus fünf Strängen (s1A, s2A, s3A, s5A, s6A, Abbildung 1). Die β -Faltblätter sB und sC bestehen aus sechs bzw. vier kürzeren Strängen.

Die reaktive Schleife (*reactive centre loop*, RCL) enthält das reaktive Zentrum des Serpins, welches von einer geeigneten Protease als Substrat erkannt wird. Der exponierte, flexible RCL ist ca. 17 As lang und verbindet die β -Faltblattstränge s5A und s1C. Nach der Nomenklatur von Schechter und Berger¹, wird die Spaltstelle im RCL des Serpins mit P1-P1' bezeichnet. Ausgehend von dieser Spaltstelle werden die Aminosäurereste in Richtung Aminoterminus mit P2, P3, P4 usw. bezeichnet, während die Aminosäurereste in Richtung Carboxyterminus mit P2', P3', P4' usw. bezeichnet werden.

Serinproteasen spalten die P1-P1'-Bindung eines inhibitorischen Serpins nach dem Mechanismus der katalytischen Triade (Abschnitt 2.1.3). Dabei wird der carboxyterminale Teil der reaktiven Schleife freigesetzt. Der aminoternale Teil der nun verkürzten reaktiven Schleife (Abbildung 1, grün markiert, Positionen P18-P1) lagert sich als β -Faltblattstrang s4A antiparallel zwischen die beiden β -Faltblattstränge s3A und s5A ein. Hierbei handelt es sich um einen klassischen Übergang des Moleküls von einer gespannten Form (*stressed*, thermodynamisch ungünstig) in eine entspannte Form (*relaxed*, thermodynamisch begünstigt), die erst durch Abspalten des carboxyterminalen Teils des RCL ermöglicht wird.

¹ Schechter, I., Berger, A., 1967.

Die hoch konservierte Serpin-Kernstruktur zeigt, wie wichtig die einzelnen Strukturelemente wie Faltblattstrukturen und Helices für den Prozess der Serinprotease-Inhibierung sind. Die *Hinge-Region* (*hinge*: Gelenk, Scharnier) ist zwischen P15 - P9 der reaktiven Schleife lokalisiert und ist für das Umklappen der reaktiven Schleife nach Spaltung durch eine Serinprotease essentiell. Die *Breach-Region* (*breach*: Bresche) an der Spitze des β -Faltblatts s3A (Abbildung 1, rosa markiert) bildet den Ort der initialen Insertion der reaktiven Schleife (s4A, antiparallel) in die β -Faltblattstruktur sA.

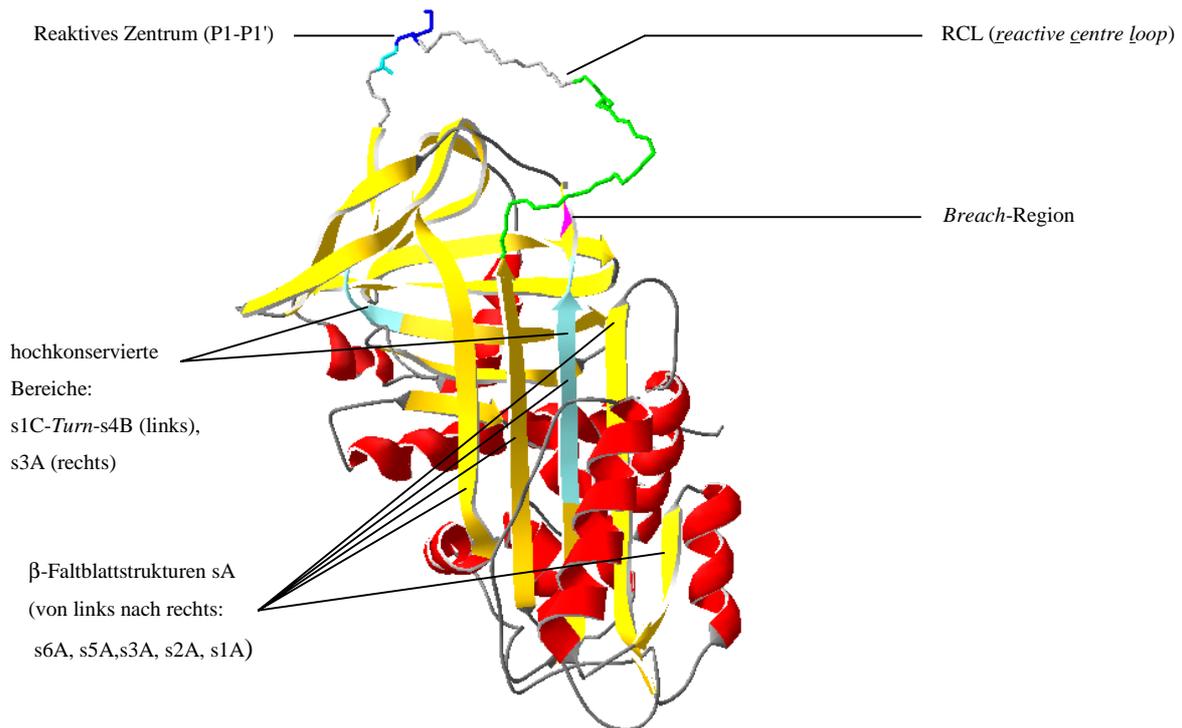


Abbildung 1: Proteinstruktur von α_1 -Antitrypsin (gespannt, nativ)

Darstellung der PDB-Struktur 1HP7¹ (α_1 -Antitrypsin, ungespalten, 2.10 Å) durch das Programm Swiss-PdbViewer 3.7² (<http://www.expasy.org/spdbv/>). Die Sekundärstrukturen sind folgendermaßen markiert: β -Faltblätter, gelb; α -Helices, rot; Loops, grau. Der RCL ist als *Backbone*-Struktur dargestellt, die Positionen P18-P8 sind grün markiert (s4A), die Seitenketten des reaktive Zentrums (P1-P1') sind blau-hellblau markiert. Zusätzlich wurden zwei hoch konservierte Bereiche des Serpins hellblau hervorgehoben. Hierbei handelt es sich um den β -Faltblattstrang s3A, der einerseits nach Spaltung direkt mit dem β -Faltblattstrang s4A (Teil des RCL) wechselwirkt und zweitens die *Breach*-Struktur ausbildet, die für das Exponieren der RCL-Struktur verantwortlich ist. Der zweite Bereich, eine konservierte Struktur (Serpinsignatur, s1C-Turn-s4B), arretiert den RCL auf der Rückseite des Proteins.

¹ Kim, S., *et al.* 2001.

² Guex, N., Peitsch, M., C., 1997.

2.1.3 Reaktionsmechanismus der Serinproteasen

Serinproteasen katalysieren die Spaltung von Peptidbindungen. Bekannte Vertreter der Serinprotease-Familie sind Trypsin und Thrombin. Das Reaktionszentrum aller Serinproteasen wird durch die drei Aminosäuren Histidin, Serin und Aspartat (selten auch Asparagin) gebildet. Nach der auf Chymotrypsin basierenden Nomenklatur¹ besteht das als „katalytische Triade“ bezeichnete Reaktionszentrum der Serinproteasen aus Serin¹⁹⁵, Histidin⁵⁷ und Aspartat¹⁰².

Durch die Anordnung der beteiligten Seitenketten von Histidin⁵⁷ und Aspartat¹⁰² wird die normalerweise protonierte Hydroxylgruppe der Serin¹⁹⁵-Seitenkette in ein starkes Nukleophil umgewandelt. Der Mechanismus der katalytischen Triade^{2,3} ist in Abbildung 2 dargestellt.

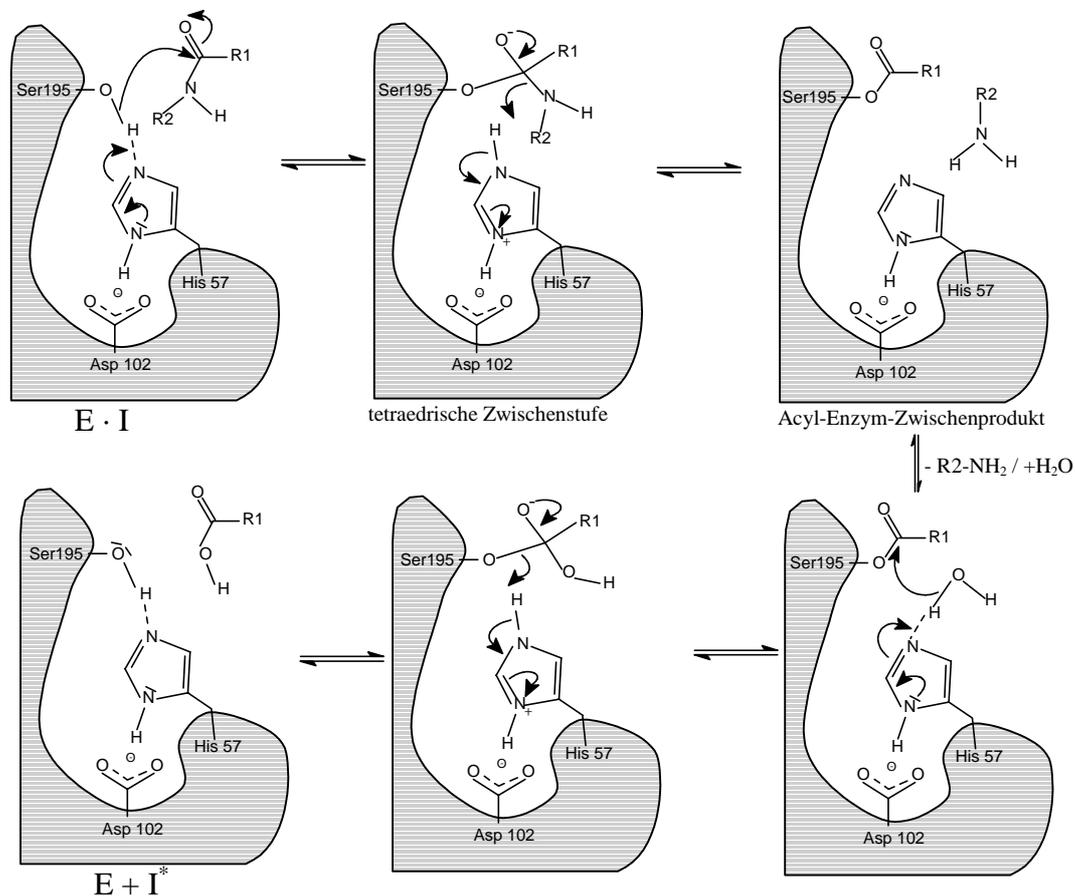


Abbildung 2: Reaktionsmechanismus der katalytischen Triade

Nach Lehninger, A., L., *et al.* 1994, verändert.

¹ Genest, M., Ptak, M., 1982.

² Plotnick, M., I., Mayne, L., Schechter, N., M., Rubin, H., 1996.

³ Frey, P., A., Whitt, S., A., Tobin, J., B., 1994.

Die Hydrolyse der zu spaltenden Peptidbindung kann in einen Acylierungs- und Deacylierungsschritt unterteilt werden. Der Acylierungsschritt beginnt mit dem nukleophilen Angriff des Sauerstoffs der Hydroxylgruppe von Serin¹⁹⁵ auf das Carbonylkohlenstoffatom der Peptidbindung des Substrats (E · I). Dabei bildet sich ein tetraedrischer Übergangszustand mit einem negativ geladenen Carbonyl-Sauerstoffatom. Die negative Ladung des Oxyanion wird innerhalb der so genannten Oxyanionen-Enzymtasche stabilisiert.

Die Auflösung des tetraedrischen Übergangszustandes erfolgt, indem das Proton von Histidin⁵⁷ auf das Stickstoffatom der Peptidbindung übertragen und somit die Peptidbindung gespalten wird. In diesem Zustand ist Serin¹⁹⁵ mit dem Substrat über eine Esterbindung verbunden, das Zwischenprodukt wird als Acyl-Enzym-Intermediat bezeichnet. Anschließend diffundiert die abgespaltene Aminkomponente des Substrates ab und wird durch Wasser ersetzt. Der zur Reaktivierung des Enzyms führende Deacylierungsschritt ist die Umkehrreaktion des Acylierungsreaktion. Zuerst wird ein Proton des Wassers auf das Histidin⁵⁷ übertragen. Das entstehende Hydroxidion greift das Carbonylsauerstoffatom der an Serin¹⁹⁵ gebundenen Acylgruppe unter Bildung eines tetraedrischen Zwischenproduktes an. Wie bei der Hinreaktion wird auch hier die negative Ladung des Oxyanion innerhalb der Oxyanionen-Enzymtasche stabilisiert. Abschließend überträgt Histidin⁵⁷ das übernommene Proton auf das Sauerstoffatom von Serin¹⁹⁵, wodurch der Carboxyteil des Spaltproduktes freigesetzt wird (E + I^{*}).

Die Familie der Thiolproteasen ähnelt den Serinproteasen im Strukturaufbau und Funktionsmechanismus. Die Thiolproteasen verfügen über eine Variante der katalytische Triade aus Cystein, Histidin, Asparagin. Hier wird die Thiol-Gruppe des Cysteins durch die Interaktion mit Histidin und Asparagin in ein starkes Nukleophil umgewandelt. Die für den Mechanismus der katalytischen Triade wichtige Oxyanionen-Enzymtasche ist ebenfalls bei Thiolproteasen vorhanden. Der ähnlichen Mechanismen von Serin- und Thiolproteasen erklärt das Vorhandensein von so genannten „Cross-Class“-Serpinen, deren physiologische Ziel-Proteasen zur Familie der Thiolproteasen gehören.

Die Substratspezifität der verschiedenen Serinproteasen wird zum Teil durch die Struktur der Substrat-Bindetasche realisiert, die beispielsweise gezielt die Aufnahme von großen, kleinen, positiven, negativen, hydrophilen oder hydrophoben Aminosäureresten in die Bindetasche ermöglicht bzw. verhindert.

2.1.4 Die Inhibierungsreaktion der Serpine

Die auf der Serpin-Oberfläche exponierte reaktive Schleife wird von einer geeigneten Ziel-Serinprotease als Pseudosubstrat erkannt und gebunden. Die Serinprotease spaltet die reaktive Schleife und wird durch das Serpin inhibiert. Das Resultat ist ein proteolytisch-inaktives Serpinacyl-Protease-Intermediat¹. Das Serpin ist über eine Esterbindung an die Serin¹⁹⁵-Seitenkette der Serinprotease gebunden. Da auch das Serpin durch die Inhibierungsreaktion die biologische Aktivität verliert, werden Serpine der Gruppe der „Suizid“-Inhibitoren zugeordnet. Das „*branched-pathway*“-Modell in Abbildung 3 stellt die Inhibierungsreaktion der Serpine formal dar.

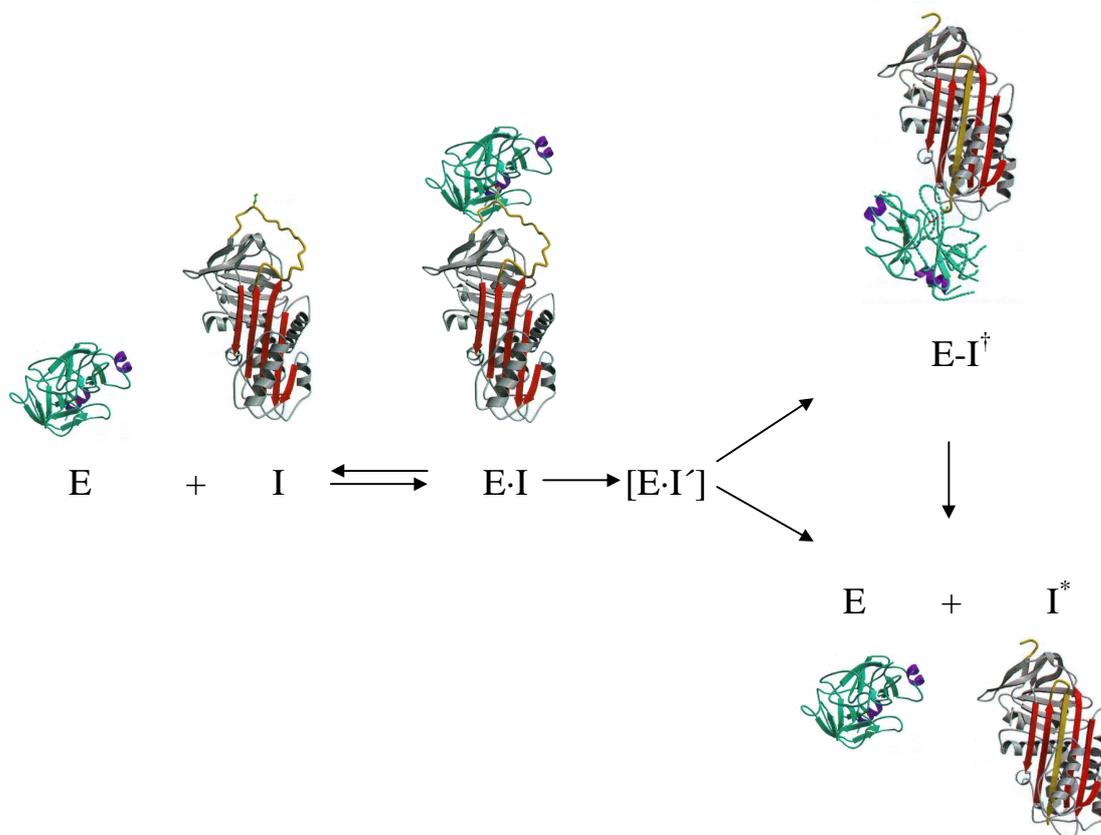


Abbildung 3: „*branched-pathway*“-Modell

Gezeigt ist das „*branched-pathway*“-Modell am Beispiel der Inhibierungsreaktion von Trypsin (Serinprotease; zyan, violett) durch α_1 -Antitrypsin (Serpine; gelb: reaktive Schleife, rot: β -Faltblatt sA, grau). Abbildung verändert nach Gettins *et al.* 1996 und Huntington *et al.* 2000b.

Nach diesem Modell bindet die Serinprotease (E) im ersten Schritt reversibel an die reaktive Schleife des inhibitorischen Serpins (I) und bildet einen so genannten „Michaelis-Komplex“

¹ Lawrence, D., A., Ginsburg, D., Day, D., E., Berkenpas, M., B., Verhamme, I., M., Kvassman, J., O., Shore, J., D., 1995.

(E·I). An dieser nicht-kovalenten Bindung sind unter anderem flankierende Regionen um die zu spaltene Serpin-Peptidbindung (P1-P1') beteiligt, die mit der Substratbindetasche der Protease interagieren. Der „Michaelis-Komplex“ kann entweder unter Regenerierung des freien Enzyms und des aktiven Inhibitors dissoziieren oder die Spaltung der reaktiven Schleife führt zur Bildung eines kovalenten Acyl-Enzym-Intermediates (EI'). Bis zu diesem Schritt entspricht der *branched-pathway*-Reaktionsablauf der Acylierungsreaktion des „katalytischen Triade-Mechanismus“.

Für das Acyl-Enzym-Intermediat gibt es zwei mögliche Wege. Bei der Protease-Inaktivierung inseriert die gespaltene reaktive Schleife des Serpins partiell oder vollständig in das β -Faltblatt sA (Abschnitt 2.1.2) unter Bildung eines irreversiblen Enzym-Inhibitor-Komplexes (EI[†]). Dabei wird die Protease gegen das Serpingerüst gepresst. Im Rahmen dieser Konformationsänderungen im Serpin- und Proteaseanteil des kovalenten Komplexes kommt es zur Deformation der „katalytischen Triade“-Anordnung. Die damit einhergehende biologische Inaktivierung der Protease verhindert die Deacylierungsreaktion. *In vitro* konnte gezeigt werden, dass der Enzym-Inhibitor-Komplex (EI[†]) mit einer Halbwertszeit von Stunden bis Wochen langsam in seine Bestandteile zerfällt (intakte Protease E und gespaltenes inaktives Serpin I^{*}). *In vivo* wird der Enzym-Inhibitor-Komplex (EI[†]) innerhalb von Minuten abgebaut. Als zweite Möglichkeit kann das Acyl-Enzym-Intermediat (EI') auch einen nicht-inhibitorischen Reaktionsweg einschlagen, der über eine Deacylierungsreaktion zur Spaltung des Serpins (I^{*}) und zur Regeneration der Protease (E) führt. Ausschlaggebend für die Wahl des Reaktionsweges ist das Verhältnis der Geschwindigkeiten von der Insertion der reaktiven Schleife im Vergleich zur Deacylierungsreaktion¹.

Die Insertionsfähigkeit einer reaktiven Schleife (RCL) ist dabei von ihrer Größe und von der Aminosäuresequenz abhängig. Damit der aminoternale Teil des RCL, der β -Faltblattstrang s4A, ins β -Faltblatt A inserieren kann (β -Faltblatt-Extensions-Mechanismus), sollten die Positionen P15-P7 mit kleinen Aminosäureresten besetzt sein. Bevorzugt finden sich Alanine in den Positionen P12-P9. Eine weiterer wichtiger Indikator stellt die Position P15 dar, die bei über 95 % der Serpine mit einem Glycin besetzt ist. Position P15 bildet das Zentrum der *Hinge*-Region, die für die Inhibierung der Protease essentiell ist. Ein großer RCL kann bereits ungespalten eine latente Zustandsform einnehmen (partielle Insertion des RCL ins β -Faltblatt sA), wodurch die Fähigkeit zur Inhibition von Serinproteasen gesenkt wird.

¹ Lawrence, D., A., Olson, S., T., Muhammad, S., Day, D., E., Kvassman, J., O., Ginsburg, D., Shore, J., D., 2000.

2.1.5 Physiologische Funktionen der Serpine

Aufgrund der Fähigkeit der Serinproteasen, Peptidbindungen irreversibel zu spalten, ist eine strenge Regulation ihrer Aktivität notwendig. Serinproteasen werden in der Regel als Zymogen (inaktive Vorstufe) produziert und selbst durch Spaltung aktiviert (teilweise autokatalytisch). Die Aktivierung ist somit irreversibel. Das „Abschalten“ bzw. die Inhibierung der aktivierten Serinproteasen wird durch unterschiedliche Klassen von Peptidase-Inhibitoren realisiert. Die Serpine bilden eine hochmolekulare Klasse dieser Serinprotease-Inhibitoren.

Als Antagonisten der Serinproteasen (bzw. Cysteinproteasen) sind die Serpine an der Regulation einer Vielzahl von physiologischen Prozessen beteiligt. Abbildung 4 zeigt einige ausgewählte physiologische Funktionen von inhibitorischen und nicht-inhibitorischen human-Serpinen.

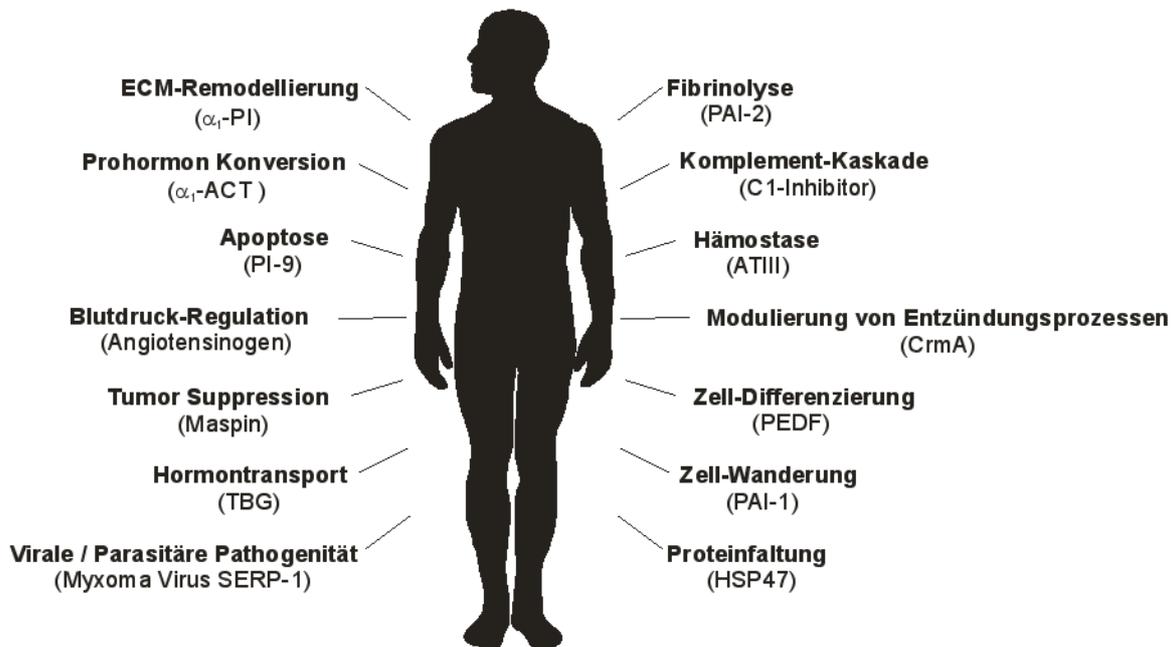


Abbildung 4: Physiologische Funktionen ausgewählter human-Serpine

Darstellung nach Potempa *et al.* 1994, verändert. Abkürzungen: PAI-1/2, plasminogen activator inhibitor 1/2; α_1 -PI, α_1 -protease inhibitor; α_1 -ACT, α_1 -antichymotrypsin; PI-9, protease inhibitor 9; TBG, thyroxine-binding globulin; ATIII, antithrombin III; PEDF, pigment epithelium-derived factor; CrmA, cytokine response modifier protein A; HSP47, heat shock protein 47; ECM, extrazelluläre Matrix.

Inhibitorische human-Serpine sind unter anderem an der Regulation von Blutgerinnung, Fibrinolyse, Prohormon-Konversion und Apoptose beteiligt. Nicht-inhibitorische Serpine können als Hormonvorläufer, Hormontransporter oder intrazelluläre Chaperone dienen.

Einhergehend mit diesem breiten Funktions-Spektrum konnte eine Vielzahl von Krankheitsbildern mit einer „*loss of function*“- bzw. „*gain of function*“-Mutation innerhalb der Serpin-Gene identifiziert werden^{1,2}. Einige seltene Krankheitsbilder entstehen durch Fehlfaltungen und Akkumulation bzw. Polymerisation von Serpinen^{3,4}. So kann die Akkumulation von α_1 -Antitrypsin im endoplasmatischen Retikulum von Leberzellen zu Leberzirrhose und als weitere Folge zum Lungen-Emphysem⁵ führen.

Von den Insekten-Serpinen sind die Serpine von *Drosophila melanogaster* am ausführlichsten untersucht. Das Serpin Nec (*necrotic*) dient als Regulator der Toll-vermittelten Immunantwort, indem es innerhalb einer Serinprotease-Kaskade die Serinprotease Psh (Persephone) inhibiert⁶. Über die physiologischen Funktionen der Nematoden- und Pflanzen-Serpine ist bisher wenig bekannt. Erste Untersuchungen⁷ am Pflanzen-Serpin CmPS-1 des Riesen Kürbis (*Cucurbita maxima*) weisen auf eine Verteidigungsfunktion hin, die den Wirt vor Proteasen von Insekten/Parasiten schützt. Dagegen nutzt der parasitäre Nematode *Brugia malayi* das Serpin Bm-SPN-2, um sich vor Proteasen (*human cathepsin G*, *human neutrophilic elastase*) seines Wirts zu schützen⁸. *Brugia malayi* befällt das Lymphsystem des Menschen und ist der Erreger der malayischen Filariose (Brugiasis).

¹ Carrell, R., W., Stein, P., E., 1996.

² Bruce, D., Perry, D., J., Borg, J., Y., Carrell, R., W., Wardell, M., R., 1994.

³ Carrell, R., W., Lomas, D., A., 2002.

⁴ Lomas, D., A., Carrell, R., W., 2002.

⁵ Lomas D., A., Mahadeva R., 2002

⁶ Robertson, A., S., Belorgey, D., Lilley, K., S., Lomas, D., A., Gubb, D., Dafforn, T., R., 2003.

⁷ Yoo, B., C., *et al.* 2000.

⁸ Zang, X., Yazdanbakhsh, M., Jiang, H., Kanost, M., R., Maizels, R., M., 1999.

2.2 Die Evolution der Serpin-Superfamilie - Stand der Forschung

Hunt und Dayhoff stellten im Jahre 1980 erstmals die These auf, dass sich die Serpine über einen Zeitraum von ca. 500 Millionen Jahren durch Duplikations- und Diversifikationsprozesse entwickelt haben¹. Die erste ausführliche Analyse der Evolution der Serpine wurde von Marshall durchgeführt². Als Grundlage dienten Marshall die Aminosäuresequenzen von ca. 40 Serpinen (orthologe sowie paraloge), mit deren Hilfe die phylogenetischen Verwandtschaftsverhältnisse der Serpine nach der Maximum-Parsimony-Methode (MP, Abschnitt 4.7.4) untersucht wurden. Marshall konnte zeigen, dass sich der Organismen-Stammbaum überraschenderweise nicht im konstruierten MP-Serpin-Stammbaum widerspiegelt. Eine mögliche Erklärung für dieses Ergebnis ist, dass die Erhaltung der Serpin-Inhibierungsfunktion offensichtlich eine hochkonservierte Tertiärstruktur erfordert und es allenfalls zu physikochemisch-neutralen Substitutionen kommt (funktionelle Beschränkung der Evolution).

Nachdem die Anzahl der bekannten Serpine stark angestiegen war, folgten weitere Analysen der Serpin-Phylogenie, die sich jeweils auf unterschiedliche Methoden und unterschiedliche phylogenetische Kriterien stützten.

Basierend auf Aminosäuresequenzen von 219 Serpinen und unter Berücksichtigung von Röntgenkristallstrukturdaten stellten Irving *et al.* ein Serpin-Klassifikationssystem auf, das die Mitglieder der Serpin-Superfamilie (Serpine von Vertebraten und Invertebraten) in 16 Familien einordnet (in Tabelle 2 sind die Serpine der Vertebraten mit a - h bezeichnet)³.

Dagegen hoben Ragg *et al.* in ihren Untersuchungen neben den Aminosäuresequenzen selbst auch die Bedeutung und den Informationsgehalt von seltenen Indels (Insertion/Deletion) und der Exon-Intron-Strukturen der Serpin-Gene hervor⁴. Das Ergebnis war ein auf Exon-Intron-Strukturen basierendes Klassifikationssystem, das die Vertebraten-Serpine in sechs Genfamilien einteilt. Ausgehend von dieser Kern-Einteilung konnten so genannte diagnostische Aminosäuresequenz-Positionen (*diagnostic sites*) von 111 Serpinen bestimmt werden.

¹ Hunt, L., T., Dayhoff, M., O., 1980.

² Marshall, C., J., 1993.

³ Irving, J., A., Pike, R., N., Lesk, A., M., Whisstock, J., C., 2000.

⁴ Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

Diese diagnostischen Aminosäuresequenz-Positionen dienten als „aufbereitete“ phylogenetische Marker und ermöglichen eine Klassifikation von Serpinen, deren Exon-Intron-Struktur noch nicht bekannt ist.

Interessanterweise führen beide Ansätze zu einer ähnlichen Einteilung der Serpine. Trotz unterschiedlicher Informationsquellen (Aminosäuresequenzen bzw. Exon-Intron-Strukturen) wurden größtenteils identische Gruppen formuliert (Tabelle 2: 1/b; 2/a; 5/c; 6/h). Die Hauptunterschiede der Serpin-Klassifikationen beschränken sich dabei auf Zusammenfassungen von Familien bzw. auf Bildung von Subfamilien (Tabelle 2: 3/e,i ; 4/f,g). Tabelle 2 stellt Gemeinsamkeiten sowie Unterschiede der beiden Serpin-Klassifikationssysteme dar.

Gruppe	Serpine
1	Ovalbumin (b), Gen Y Protein (b), PAI-2 (b), SCCA-1 (b), SCCA-2 (b), PI-2 (b), Pi-6 (b), Pi-9 (b)
2	α_1 -Antitrypsin (a), CBG (a), HCII (d), Protein C Inhibitor (a), Angiotensinogen (a), α_1 -Antichymotrypsin (a), Protease Inhibitor 4 (a), THBG (a)
3	PAI-1 (e), Nexin-1 (e), Neuroserpin (i)
4	α_2 -Antiplasmin (f), PEDF (f), C1-Inhibitor (g)
5	Antithrombin III (c)
6	HSP47 (h)

Tabelle 2: Vergleich der Serpin-Klassifikationssysteme

Dargestellt ist die Einteilung der Vertebraten-Serpine in sechs Genfamilien (Gruppen 1-6) nach Ragg *et al.* (2001), basierend auf Exon-Intron-Strukturen. Den Serpinen ist zum Vergleich die Serpin-Klassifikation nach Irving *et al.* (2000) zugeordnet (16 Familien, a – p, hier sind nur die Serpine der Vertebraten gezeigt, a- h). Die Einteilung der Serpine (Vertebraten und Invertebraten) basiert hier ausschließlich auf den Aminosäuresequenzen.

Der Einfluss von verschiedenen Baumkonstruktions-Algorithmen (Neighbor-Joining, Abschnitt 4.7.3; Maximum-Likelihood, Abschnitt 4.7.2) und Substitutionsmatrizen (BLOSUM 62, Jones-Taylor-Thornton, Abschnitt 2.3.1) auf die Konstruktion der Serpin-Phylogenie wurde von Atchley *et al.* untersucht¹. Die Untersuchung basierte auf Aminosäuresequenz-Daten von 110 Serpinen. Um die auf diese Weise erhaltenen Serpin-Stammbäume zu bewerten, wurden auch hier Exon-Intron-Strukturen und diagnostische Aminosäuresequenz-Positionen als Referenz verwendet.

¹ Atchley, W.,R., Lokot, T., Wollenberg, K., Dress, A., Ragg, H., 2001.

2.3 Molekulare Evolution

Das Arbeitsfeld der „Molekularen Evolution“ widmet sich der Aufklärung der Evolution von Makromolekülen, vor allem der von DNA- und Aminosäuresequenzen.

Biologische Sequenzen (OTUs, *operational taxonomic units*) tragen an jeder Sequenzposition ein Merkmal (DNA: 1 von 4 Basen; Proteine: 1 von 20 Aminosäuren). Um diese Merkmale zu vergleichen, werden verwandte Sequenzen miteinander aligniert, wobei vermutete Insertionen bzw. Deletionen durch Gaps angezeigt werden. Das Gap-Zeichen „-“ kann somit als ein weiteres Merkmal gewertet werden.

2.3.1 Methoden der molekularen Phylogenetik

Das Ziel der molekularen Phylogenetik ist es, einen Stammbaum zu (re-)konstruieren, der die Verwandtschaftsverhältnisse der paralogen und orthologen Mitglieder einer Genfamilie repräsentiert. In Tabelle 3 sind ausgewählte Methoden der Baumkonstruktion aufgeführt.

Verfahren	Daten	Evolutionsmodelle (bzw. Aminosäure-Substitutionsmatrizen)	Abschnitt
Maximum-Parsimony (MP)	Merkmale	-	4.7.4
Maximum-Likelihood (ML)	Merkmale	+	4.7.2
Neighbor-Joining (NJ)	Distanzen	+	4.7.3

Tabelle 3: Methoden der Baumkonstruktion

Die Methoden können nach Art der verwendeten Daten unterschieden werden. Bei Distanzverfahren werden Merkmale (Charaktere) zunächst in evolutionäre Distanzen konvertiert. Als Ausgangspunkt dient heute in der Regel ein multiples Alignment, mit dessen Hilfe, gegebenenfalls unter Verwendung von Evolutionsmodellen, die Distanzen für jedes Sequenzpaar bestimmt werden. Aus den ermittelten Distanzen wird eine Distanzmatrix zusammengestellt, die als Ausgangspunkt für die Baumkonstruktion dient. Bei anderen Verfahren (MP, ML) werden die Merkmale (Charaktere) direkt in Form eines multiplen Alignments zur Baumkonstruktion verwendet.

Um die Vielzahl der unterschiedlichen Mutationsereignisse auf DNA-Ebene spezifisch gewichten zu können, verwenden die meisten Baumkonstruktionsmethoden Evolutionsmodelle.

Tabelle 4 fasst ausgewählte Evolutionsmodelle zusammen.

Evolutionsmodell	Anzahl der Substitutionsraten-Klassen (Transition / Transversion)	Parameter
JC, „1-ParameterModell“ Jukes und Cantor, 1969.	1 (keine Unterscheidung)	1
K2P, „2-Parameter-Modell“ Kimura, 1980.	2 (1/1)	2
K3ST, „3-Substitutionen-Modell“ Kimura, 1981.	3 (1/2)	3
„Generelles zeitirreversibles Modell“ (berücksichtigt Substitutionsrichtung)	6 (4/2)	12

Tabelle 4: Evolutionsmodelle für DNA-Sequenzen

Alle Modelle setzen voneinander unabhängige Substitutionen und eine im wesentlichen konstante Substitutionswahrscheinlichkeit, oft bei gleicher Basenfrequenz (1:1:1:1), voraus.

Diese Evolutionsmodelle unterscheiden sich im wesentlichen durch ihren Differenzierungsgrad, d.h. durch die Anzahl der verwendeten Substitutions-Parameter. Analog zu den theoretischen DNA-Evolutionsmodellen werden auch auf Proteinebene Substitutionsmatrizen für Aminosäuren verwendet. Die Matrizen basieren auf empirischen Daten und spiegeln physikochemische Eigenschaften der Aminosäuren sowie deren Codon-Ähnlichkeit wider (Tabelle 5).

Aminosäuren-Substitutionsmatrix	Kurzbeschreibung
PAM (Point Accepted Mutations) Dayhoff, M., O., <i>et al.</i> 1978.	Die Substitutionsraten sind aus Alignments abgeleitet worden (As-Identität $\geq 85\%$, d.h. gesicherte Phylogenien)
BLOSUM (Blocks Substitution Matrix) Henikoff, S., Henikoff, J., G., 1992.	Die Substitutionsraten sind aus Alignments von entfernt verwandten Sequenzen abgeleitet worden.
JTT (Jones-Taylor-Thornton) Jones, D., T., <i>et al.</i> 1992.	Basiert methodisch auf PAM, berücksichtigt aktualisierte Datenbanken.

Tabelle 5: Aminosäure-Substitutionsmatrizen

Die molekulare Phylogenetik konzentriert sich zur Zeit auf Analysen von DNA- oder Aminosäuresequenzen. Die Stammbaum-Konstruktion mit biologischen Datensätzen kann durch eine Reihe von unterschiedlichen Prozessen erschwert werden. Beispiele für diese Prozesse sind:

- variable Substitutionsraten,
- Mutations-Sättigung,
- konvergente Evolution,
- von einander abhängige Substitutionsereignisse,
- funktionelle Beschränkungen der Proteine.

Aufgrund der Genomprojekte stehen neben einer Vielzahl von neuen Gensequenzen auch die Exon-Intron-Strukturen und chromosomalen Lokalisationsdaten der Gene als zusätzliche phylogenetische Informationsquellen¹ zur Verfügung. Aufgrund der seltenen Veränderung der Exon-Intron-Strukturen, tritt eine Mutations-Sättigung deutlich langsamer auf. Die in der Genstruktur enthaltene phylogenetische Information ist besonders für frühe Evolutionsereignisse aussagekräftig.

2.4 Struktur eukaryotischer Gene

Die weit überwiegende Zahl eukaryotischer Gene setzt sich mosaikartig aus Exons zusammen, die durch Introns getrennt sind. Introns sind Sequenzen, die durch Spleißen aus der prä-mRNA ausgeschnitten werden, während die Exons diejenigen Sequenzen sind, die in der reifen mRNA verbleiben. Die Anzahl der Introns pro eukaryotischem Gen variiert sehr stark. Tabelle 6 gibt die durchschnittliche Intron-Anzahl für ausgewählte eukaryotische Genome an. Introns, die im Protein-kodierenden Bereich eines Gens lokalisiert sind, wird eine Intronphasen zugeordnet. Je nachdem, ob das Intron nach dem ersten, zweiten oder dritten Nukleotid des Codons beginnt, handelt es sich um die Intronphase a, b oder c.

¹ Rokas, A., Holland, W., H., 2000.

Organismus	Durchschnittliche Intron-Anzahl pro Gen
<i>Saccharomyces cerevisiae</i> , Hefe	0
<i>Caenorhabditis elegans</i> , Nematode	3
<i>Drosophila melanogaster</i> , Insekt	3
<i>Arabidopsis thaliana</i> , Pflanze	4
<i>Homo sapiens sapiens</i> , Mensch	6

Tabelle 6: Durchschnittliche Intron-Anzahl eukaryotischer Genome

Neben den bekannten spleißosomalen Introns gibt es weitere Intronotypen, die andere Spleißmechanismen aufweisen. Diese Intronotypen wurden bisher nur selten gefunden. Tabelle 7 führt die bekanntesten Intronotypen auf und beschreibt sie kurz.

Intronotyp	Vorkommen	Kurzbeschreibung
Spleißosomale Introns Klasse GT-AG-Intron (Nukleäre prä-mRNA Introns) (Intron Klasse III)	Eukaryoten	Benötigt Spleißosom, vermutlich aus Klasse II Introns entwickelt. U1/U2-RNA-abhängige Introns, die die kanonische Spleißsignale (GT-AG) erkennen.
Klasse AT-AC-Intron ¹ (benötigt Spleißosom)	Eukaryoten	Benötigt Spleißosom, seltene Variante der U1/U2-abhängigen spleißosomalen Introns. Die Spleißsignale AT-AC (anstatt GT-AG) werden unter Beteiligung der U11/U12-RNA-Spleißfaktoren erkannt.
Klasse I-Intron (selbstspleißend, potentiell mobiles Element)	Mitochondrien, Chloroplasten, Bakteriophagen, Cyanobakterien, frühe Eukaryoten	<i>In vitro</i> selbstspleißend bei Anwesenheit eines Guanosins als Cofaktor ² . <i>In vivo</i> wird die Spleißreaktion durch Spleißfaktoren unterstützt. Einige Introns (ca. 30 %) enthalten einen ORF für diesen Spleißfaktor (mit DNA-Endonuklease-Aktivität).
Klasse II-Intron (selbstspleißend, potentiell mobiles Element)	Bakterien, Chloroplasten, Mitochondrien	Verwandtschaft zu non-LTR Elementen (z.B. LINEs) ³ . <i>In vitro</i> selbstspleißend, <i>in vivo</i> wird die Spleißreaktion durch Spleißfaktoren unterstützt. Einige Introns (ca. 25 %) enthalten einen ORF für diesen Spleißfaktor (mit Reverse-Transkriptase-Aktivität).

Tabelle 7: Intronotypen

Abk.: LINE (long interspersed nuclear element), LTR (long terminal repeats), ORF (open reading frame).

¹ Mount, S., M., 1996.

² Grabowski P., J., Zaug A., J., Cech T., R., 1981.

³ Zimmerly *et al.* 2001.

Die Exon-Intron-Struktur der Gene ist nicht konstant, sondern unterliegt Veränderungen. So können Punktmutationen in Intron-Spleißstellen zum Verlust eines Introns führen oder durch Bilden von neuen alternativen Spleißstellen die Intron Grenzen verschieben.

Die Entstehung neuer Introns wird zur Zeit durch *Retrohoming*- oder *Retrotransposition*-Prozesse erklärt¹. Hier spielen selbstspleißende Klasse II-Introns eine wesentliche Rolle, die einen ORF für eine Reverse Transkriptase enthalten. Es wird angenommen, dass die Lariat-Form des (selbst-)gespleißten Klasse II-Introns durch Reverse-Spleißen in doppelsträngige DNA (oder in mRNAs) inserieren kann. Ein weiterer wichtiger Faktor, der zur Veränderung der Genstruktur führen kann, ist die Rekombination. Durch (homologe) Rekombination kann es zur Homogenisierung der Genstrukturen von benachbarten, verwandten Genen kommen.

2.4.1 Alternatives Spleißen

Alternatives Spleißen ermöglicht es, aus einem Gen unterschiedliche Transkripte und somit unterschiedliche Protein-Varianten (Protein-Isoformen) zu bilden. Beim Prozess des alternativen Spleißens kombiniert das Spleißosom 5'-Spleißstellen (Donorspleißstelle) mit unterschiedlichen 3'-Spleißstellen (Akzeptorspleißstelle) bzw. vice versa. In Tabelle 8 sind die möglichen alternativen Spleißvorgänge aufgeführt.

In aktuellen Schätzungen² wird angenommen, dass ca. 30 % der menschlichen Gene alternativ gespleißt werden. Das Spleißmuster eines Gens kann sich dabei in Abhängigkeit des Gewebetyps (spatial) und des Entwicklungsstadiums (temporal) verändern. Bekannte Beispiele für alternatives Spleißen sind die Sex-Determination³ bei *Drosophila melanogaster* und das Dscam-Gen⁴ (*down syndrome cell adhesion molecule*) von *Drosophila melanogaster*, das theoretisch die Transkription von 38016 mRNAs bzw. Protein-Varianten ermöglicht.

¹ Eickbush, T., H., 2000.

² Gelfand, M., S., *et al.* 1999.

³ Graveley, B., R., 2002.

⁴ Celotto, A., M., Graveley, B., R., 2001.

Alternativer Spleißvorgang	Kurzbeschreibung
Alternative 5'-Spleißstelle	Verkleinert bzw. vergrößert das Intron. Kann in Abhängigkeit des veränderten Exons zu einer Rastermutation führen.
Alternative 3'-Spleißstelle	Verkleinert bzw. vergrößert das Intron. Kann in Abhängigkeit des veränderten Exons zu einer Rastermutation führen.
Intron-Retention	Beide Spleißstellen werden nicht genutzt. Kann in Abhängigkeit der Introngröße zu einer Rastermutation führen.
Alternative 5'-Exons	(Oft in Verbindung mit alternativen Transkriptionsstartpunkt)
Alternative 3'-Exons	(Oft in Verbindung mit alternativen Polyadenylierungsstellen)
Exon-Skipping	Die 5'-Spleißstelle des ersten Introns wird mit der 3'-Spleißstelle eines zweiten Introns (liegt weiter in 3'-Richtung) kombiniert. Kann in Abhängigkeit der Intron/Exongröße zu einer Rastermutation führen.
Multiple Exon-Varianten	Ein Exon liegt in mehreren leicht unterschiedlichen Exon-Varianten vor. Es wird selektiv eine dieser Exon-Varianten ausgewählt.

Tabelle 8: Alternative Spleißvorgänge

Der Prozess des alternativen Spleißens¹ wird über DNA-Erkennungssequenzen (*cis-acting* Sequenzelemente im Exon oder Intron) und durch Spleißfaktoren (*trans-acting* Proteinfaktoren) reguliert. So ermöglichen ESE-Elemente (*exonic splicing enhancer*) die Bindung eines so genannten SR-Proteins, das die Aktivierung von benachbarten, suboptimalen 3'-Spleißstellen fördert. Aber auch die Konzentration bzw. das Verhältnis von verschiedenen Spleißfaktoren zueinander beeinflusst die Auswahl von 5'-Spleißstellen. So wird bei einem hohen Verhältnis der Spleißfaktoren SR / hnRNPA/B die intron-proximale 5'-Spleißstelle ausgewählt, während ein niedriges SR / hnRNPA/B-Verhältnis die Auswahl der intron-distalen 5'-Spleißstelle fördert. Spatial-temporale Unterschiede in Konzentration und Aktivität der einzelnen *trans-acting* Spleißfaktoren können somit zu unterschiedlichen Spleißmustern führen.

¹ Cáceres, J., F., Kornblihtt, A., R., 2002.

2.5 Zielsetzung der Arbeit

Phylogenetische Klassifikationen von Proteinfamilien basieren häufig auf Vergleichen ihrer Aminosäuresequenzen. Solche Analysen führen in Abhängigkeit der eingesetzten Datenanalysetechniken und den ihnen zugrundeliegenden Evolutionsmodellen bzw. Aminosäure-Substitutionsmatrizen zu unterschiedlichen Ergebnissen.

Die Superfamilie der Serpine zeichnet sich durch eine starke funktionelle Radiation, eine variable Gencluster-Organisation und eine ungewöhnlich große Vielfalt der Genstrukturen aus. Damit bieten sich die Serpine als geeignetes Untersuchungsobjekt an, um Einblicke in die Evolution der Proteine zu erhalten¹. In einer Analyse der Serpine von Vertebraten konnte gezeigt werden, dass sich neben den Aminosäuresequenzen auch die Gencluster-Organisationen, Indels und die Exon-Intron-Strukturen der Serpin-Gene als Kriterien zur phylogenetischen Klassifikation der Serpine eignen².

Das Ziel dieser Arbeit ist die Erstellung einer verlässlichen phylogenetischen Klassifikation der Serpine von *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana* und *Ciona intestinalis*. Dazu sollten zunächst die Kriterien „chromosomale Lokalisationen“, „Exon-Intron-Strukturen“ und „Aminosäuresequenzen“ der Serpin-Gene identifiziert bzw. überprüft werden. Diese Kriterien sollten anschließend zur phylogenetischen Klassifikation der Serpine verwendet werden. Mit dieser Untersuchung sollte ein Beitrag zur Aufklärung der Evolution der Serpin-Superfamilie erbracht werden.

¹ Atchley, W., R., Lokot, T., Wollenberg, K., Dress, A., Ragg, H., 2001.

² Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

3. Material

3.1 Verwendete Geräte

Geräte	Hersteller
CCD-Videokamera	Biometra
Heizblock, Thermostat 5320	Eppendorf
Heizblock, TB1 Thermoblock	Biometra
LI-COR DNA-Sequenzer 4200	MWG-Biotech AG
Pipetten	Eppendorf, Gilson
Spectrophotometer UV-1602	Shimadzu
Thermocycler Progene	Thermo-Dux
Thermocycler T-Gradient	Biometra
Transilluminator (312 nm)	Biometra
UV-Stratalinker 2400	Stratagene
Vortex Genie 2, G-560 E	Scientific Industries
Analytische Waage A120S	Sartorius
Universal Waage U4800 P	Sartorius
Wasser-Schüttelbad 1083	GFL
Biofuge 28 RS	Heraeus
RC5C Sorvall-Zentrifuge, Rotor: SS-34	Sorvall Instruments, Du Pont
Tisch-Zentrifuge 5415 C	Eppendorf
Software	Hersteller
Base ImagIR 4.1	LI-COR-Biotech Division
WinCam Version 2.2 Software	Cybertech Berlin
Photo Finish ZSoft 3.0 Software	WordStar Atlanta Technology Center

3.2 Chemikalien und Verbrauchsmaterialien

Das jeweils zur Herstellung von Medien und Puffern benötigte Wasser wird mittels einer Milli-Q[®]-Anlage (Millipore, Eschborn) gereinigt. Das demineralisierte und deionisierte, filtrierte Wasser wird im Text mit dH₂O abgekürzt. In besonderen Fällen (PCR) wird dieses dH₂O noch zusätzlich autoklaviert. Alle anderen Chemikalien sind von z.A. Qualität und stammen von Merck, Darmstadt oder Boehringer Mannheim. Für alle Versuche werden ausschließlich sterile, autoklavierte Gefäße und sterile, autoklavierte Pipettenspitzen verwendet.

Chemikalien	Hersteller / Bezugsquelle
Agarose	Life Technologies
Ammoniumpersulfat (APS)	SERVA
Ampicillin	SIGMA
Bromphenolblau	SIGMA
Carbenicillin	Novagen
dATP	Genecraft
dCTP	Genecraft
dGTP	Genecraft
dTTP	Genecraft
Ethylendiaminotetraessigsäure (EDTA)	Merck
Ethidiumbromid	SIGMA
Glukose	Merck
Glycerin	SIGMA
Glycin	Baker
1 kb DNA Leiter	Genecraft
100 bp DNA Leiter	Genecraft
Lennox L Agar	Life Technologies
Lennox L Broth Base	Life Technologies
β-Mercaptoethanol	SIGMA
Natrium-Dodecylsulfat (SDS)	FLUKA
Tetramethylethylendiamin (TEMED)	SERVA
Tris[hydroxymethyl]aminomethan (Tris-Base)	USB
Tris[hydroxymethyl]aminomethanhydrochlorid (Tris-HCl)	USB
Xylencyanol	SIGMA

3.3 Verwendete Kits

Kit	Hersteller, Bestellnummer
GeneRacer™ Kit	Invitrogen, L150201
Omniscript™ Reverse Transcriptase Kit	Qiagen, 205111
QIAquick Gel Extraction Kit	Qiagen, 28704
QIAquick PCR Purification Kit	Qiagen, 28104
TOPO TA Cloning® Kit for Sequencing	Invitrogen, K4575J10
QIAprep Spin Miniprep Kit	Qiagen, 27106
QIAGEN Plasmid Midi Kit	Qiagen, 12143
DYEnamic Direct™ Cycle Sequencing Kit	Amersham Biosciences, US79525

3.4 Verwendete Enzyme

Enzym	Hersteller, Bestellnummer
RNase A	Amresco, 0675
KlenTherm DNA-Polymerase	Genecraft, GC-001-0100
<i>Taq</i> DNA-Polymerase	Qiagen, 201203
<i>Pfu</i> DNA-Polymerase	Promega, M7741

3.4.1 KlenTherm DNA-Polymerase (Genecraft)

Es handelt sich hier um ein Fragment der *Taq* DNA-Polymerase (aus *Thermus aquaticus*), das auf Temperaturstabilität optimiert ist. Der KlenTherm DNA-Polymerase fehlen die 5'→3'-Exonuklease- und die 3'→5'-Exonuklease-Aktivitäten. Eine Besonderheit ist der 3'-Überhang von einem Nukleotid dAMP, den die Polymerase Template-unabhängig an die ursprüngliche Sequenz des PCR-Produktes anhängt. Die KlenTherm DNA-Polymerase wird zu analytischen Zwecken (Kolonie-PCR) eingesetzt.

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

3.4.2 *Taq* DNA-Polymerase¹ (Qiagen)

Die *Taq* DNA-Polymerase verfügt über eine 5'→3'-Exonuklease-Aktivität. Eine Besonderheit ist der 3'-Überhang von einem Nukleotid dAMP, den die *Taq* DNA-Polymerase Template-unabhängig an die ursprüngliche Sequenz des PCR-Produktes anhängt. Die *Taq* DNA-Polymerase wird zur präparativen DNA-Amplifikation verwendet. Es wird hier mit einer Elongationsgeschwindigkeit von ca. 1000 Basen/min bei ca. 72 °C gerechnet.

<i>Taq</i> DNA-Polymerase (5 U/μL, Lagerung bei -20 °C)	Stammlösung (Herstellerangaben): 20 mM Tris-Cl, pH 8,0, 100 mM KCl, 0,1 mM EDTA, 1,0 mM DTT, 0,5 % (v/v) Nonidet P-40, 0,5 % (v/v) Tween 20, 50 % (v/v) Glycerin
---	---

3.4.3 *Pfu* DNA-Polymerase² (Promega)

Die *Pfu* DNA-Polymerase verfügt über die korrekturlesende 3'→5'-Exonuklease-Aktivität. Damit wird eine um den Faktor 12 niedrigere Fehlerrate der DNA-Synthese (Fehlerrate ca. $1,3 \cdot 10^{-6}$) gegenüber der *Taq* DNA-Polymerase erreicht. Die *Pfu* DNA-Polymerase erzeugt PCR-Produkte mit glatten Enden. Die *Pfu* DNA-Polymerase wird zur präparativen DNA-Amplifikation verwendet. Es wird hier mit einer Elongationsgeschwindigkeit von ca. 500 Basen/min bei ca. 72 °C gerechnet.

Klonierte <i>Pfu</i> DNA-Polymerase (2,5 U/μL, Lagerung bei -20 °C)	Stammlösung (Herstellerangaben): 50 mM Tris-HCl, pH 8,2, 0,1 mM EDTA, 1,0 mM DTT, 50 % (v/v) Glycerin, 0,05 % CHAPS
---	--

3.4.4 *Taq/Pfu* DNA-Polymerasen Mix³

Der *Taq/Pfu*-Mix (40:1 (v:v) bzw. 80:1 U) wird verwendet, um die hohe Prozessivität der *Taq* DNA-Polymerase mit der Korrekturlesefunktion (3'→5'-Exonukleaseaktivität) der *Pfu* DNA-Polymerase zu kombinieren. Es wird hier mit einer Elongationsgeschwindigkeit von ca. 1000 Basen/min bei ca. 68 - 72 °C gerechnet.

¹ *Taq* PCR-Handbook (Qiagen), 2002.

² Lundberg, K., S., *et al.* 1991.

³ Barnes, W., M., 1994.

3.4.5 Restriktionsenzyme

Restriktionsendonuklease	Restriktionsschnittstelle	Hersteller
Restriktionsendonuklease <i>EcoRI</i> (<i>E. coli</i> RY 13)	5'...G▼AATTC...3'	New England Biolabs

<i>EcoRI</i> Restriktionsendonuklease (Lagerung bei -20 °C)	Stammlösung (Herstellerangaben): 10 mM Kaliumphosphat-Puffer, pH 7,5, 300 mM NaCl, 0,1 mM EDTA, 200 µg/mL BSA, 10 mM β-Mercaptoethanol, 0,15 % Triton X-100, 50 % Glycerin.
--	---

3.5 Verwendete Puffer und Lösungen

10-fach KlenTherm Reaktionspuffer (Lagerung bei -20 °C)	670 mM Tris-HCl, pH 9,1, 160 mM (NH ₄) ₂ SO ₄ , 35 mM MgCl ₂ , 0,15 % (w/v) BSA
10-fach Taq Reaktionspuffer (Lagerung bei -20 °C)	Tris-HCl, pH 8,7, 15 mM MgSO ₄ , KCl, (NH ₄) ₂ SO ₄ (Herstellerangaben sind unvollständig)
10-fach Pfu Reaktionspuffer (Lagerung bei -20 °C)	200 mM Tris-HCl, pH 8,8, 20 mM MgSO ₄ , 100 mM KCl, 100 mM (NH ₄) ₂ SO ₄ , 1 % (v/v) Triton X-100, 1 mg/mL Nuklease-freies BSA
10-fach EcoRI Restriktionspuffer (Lagerung bei -20 °C)	100 mM Tris-HCl, pH 7,5, 50 mM NaCl, 10 mM MgCl ₂ , 0,025 % (v/v) Triton X-100.
TAE-Puffer (Agarose-Gelelektrophorese)	40 mM Tris, pH 8,0, 40 mM Essigsäure (Eisessig), 1 mM EDTA
10-fach DNA-Probenpuffer (Agarose-Gelelektrophorese)	50 % (v/v) Glycerin, 0,1 % (w/v) SDS, 0,1 % (w/v) Bromphenolblau, 0,1 % (w/v) Xylencyanol, 0,1 M EDTA in 10x TAE Puffer.
Ethidiumbromid-Färbebad (Agarose-Gelelektrophorese)	40 µg/L Ethidiumbromid in dH ₂ O.
Ladepuffer (pH8,0) (Plasmid-Schnell-Charakterisierung)	20 % (w/v) Ficoll 400, 100 mM EDTA, 1 % (w/v) SDS, 0,5 mg/mL RNase A, 0,25 % (w/v) Bromphenolblau.
Lyse/Ladepuffer (Plasmid-Schnell-Charakterisierung)	1:1:1 (v/v/v) Phenol / Chloroform / Ladepuffer
10-fach TBE-Puffer (DNA-Sequenzierung)	0,89 M Tris, 0,89 M Borsäure, 0,02 M EDTA, pH 8,3
Gellösung, 6 % Acrylamid (DNA-Sequenzierung)	10,50 g Harnstoff, 14,00 mL dH ₂ O, 2,50 mL 10-fach TBE-Puffer, 3,75 mL Gellösung (RapidGel-XL- 40 %, Fa. USB), 250 µL DMSO filtrieren und entgasen (Filter: Sartorius, Cellulose Acetat, 0,45 µm) Zugabe von 37,50 µL TEMED, 175 µL Ammoniumpersulfat-Lsg. (w/v = 10 %)

3.6 Primer

Die in Tabelle 9 aufgeführten Primer wurden von der Firma MWG Biotech AG (Ebersberg) bezogen. Alle Primer-Stammlösungen werden auf eine Konzentration von 100 μM (in dH_2O) eingestellt und bei -80°C gelagert. Die Primer-Bindestellen im Gen *Dm-Spn4* bzw. in den Transkripten von *Dm-Spn4* sind in Abbildung 5 symbolisch dargestellt.

Primer	Orientierung	Lokalisation	Sequenz (5' → 3')
DM_SP4_1	Plusstrang	1384 - 1406	CCTCACGACCCTCTCGCAGATCA
DM_SP4_2	Minusstrang	1962 - 1984	CAGGTACCATCACAGCTCATCAT
DM_SP4_3	Minusstrang	2317 - 2340	GATGGAGTGCTAACTTCAAACTT
DM_SP4_4B	Minusstrang	1695 - 1701, 2396 - 2412*	AAGGCTCATAAACATACCCGTGGC
DM_SP4_5	Minusstrang	2913 - 2937	GCTTCCTATTGCTTAACCCAAGTAG
DM_SP4_6	Plusstrang	1804 - 1830	GGCGGTGCGTAGGAAGCGCGCTATTAT
DM_SP4_7	Plusstrang	2193 - 2222	GGTCATGTGCTACGCCTCCATGCTCACGTT
DM_SP4_8	Plusstrang	2419 - 2445	CTGCCCATGCCGAAGCCGGATCCGATA
DM_SP4_9B	Plusstrang	2811 - 2835	GGTGATGGCCGTAGCTGCCTTCAGT
DM_SP4_10	Minusstrang	863 - 866	GTAGGCGGCCAGCACCTGGTGAAA
DM_SP4_11	Plusstrang	314 - 332	AGGGCGATAGCCGGTTTAT
DM_SP4_12	Plusstrang	67 - 77, 564 - 571*	GTGCCATGTGGTTGCTGG
DM_SP4_13	Minusstrang	1802 - 1818	TTCCTACGCACCGCCAT
DM_SP4_14	Minusstrang	2321 - 2326, 2396 - 2414*	GTAAGGCTCATAAACATACTTCAAA
DM_SP4_15	Minusstrang	2390 - 2395, 2490 - 2503*	GAGCGGTTGAGTCCTGTGA
DM_SP4_16	Minusstrang	1697 - 1701, 2396 - 2410*	GGCTCATAAACATACCCGTG
DM_SP4_17	Minusstrang	1697 - 1701, 2804 - 2815*	TCACCCGCCATACCGTG

Tabelle 9: Verwendete Primer für *Drosophila melanogaster* Serpin-Gen *Dm-Spn4*

Die Primer werden verwendet, um die Expression des Gens *Dm-Spn4* (Abschnitt 5.6) zu untersuchen. Die unter Lokalisation angegebene Nummerierung bezieht sich auf die genomische Sequenz des Gens *Dm-Spn4*. Die 3018 Bp lange Sequenz beginnt mit dem ersten Nukleotid des am weitesten 5'-gelegenen Exons (siehe Anhang 8.5.2). Primer, die über eine Introngrenze binden, sind mit einem Stern markiert. Siehe Abbildung 5 für eine symbolischen Darstellung der Primer-Bindestellen.

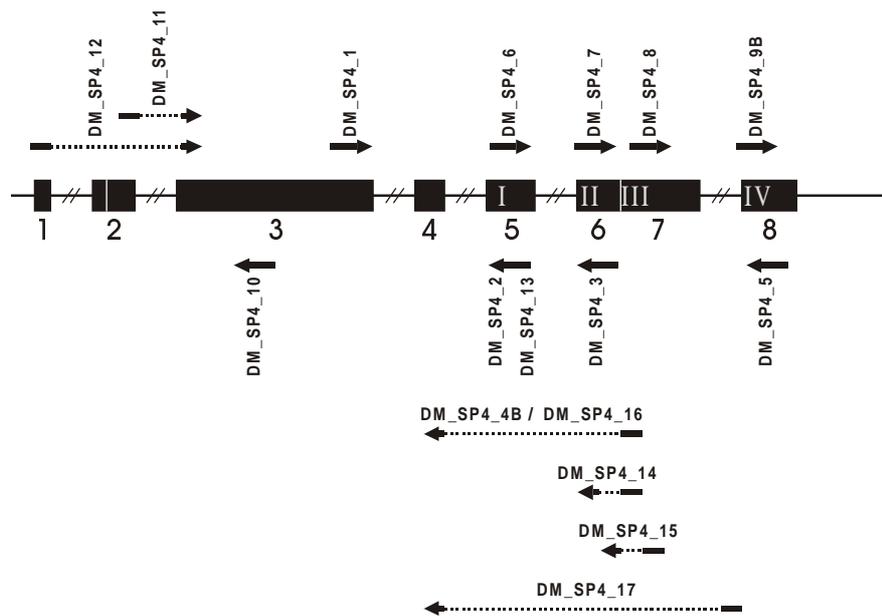


Abbildung 5: Symbolische Darstellung der Bindestellen der DM_SP4-Primer

Dargestellt ist die Exon-Intron-Struktur des *Dm-Spn4*-Gens. Die nummerierten Exons sind als schwarze Kästchen dargestellt. Die Introns sind als Linien dargestellt (für eine ausführlichere Beschreibung siehe Abbildung 35). Die Plusstrang-Primer sind oberhalb und die Minusstrang-Primer unterhalb der Genstruktur dargestellt. Primer-Darstellungen die eine unterbrochene Linie enthalten binden über Exon-Intron-Grenzen und sind somit transkript-spezifisch. Die gemeinsam dargestellten Primer DM_SP4_4B / DM_Sp4_16 und DM_Sp4_2 / DM_Sp4_13 binden im selben Genbereich unterscheiden sich jedoch in ihrer Länge (Tabelle 9).

Die Primer-Schmelztemperatur T_M kann mit folgender Regel¹ abgeschätzt werden:

$$T_M = 2 \cdot (A + T) + 4 \cdot (G + C) \text{ } ^\circ\text{C}$$

dabei ist: T_M : Schmelztemperatur

A + T : Anzahl der Basen A und T im komplementären Bereich.

G + C : Anzahl der Basen G und C im komplementären Bereich.

Bei der Durchführung der PCR-Experimente wird die von MWG Biotech AG angegebene Primer-Schmelztemperatur (T_M) als Richtwert für die Annealingtemperatur (T_A) verwendet ($T_A \approx T_M - 3 \text{ } ^\circ\text{C}$). In seltenen Fällen ist eine Optimierung der Primer-Annealingtemperatur durch PCR-Temperaturgradienten notwendig.

¹ Newton, C., R., Graham, A., 1994.

Zur Sequenzierung von TOPO-klonierten PCR-Fragmenten (Vektorsystem pCR4-TOPO, Invitrogen) werden die IRD800 (*infra-red dye*) markierten Standardsequenzierprimer T7-Promotor und T3-Promotor verwendet. Die zehn amplifizierten *Dm-Spn4* Transkripte wurden nicht kloniert, sondern direkt als PCR-Produkt sequenziert. Um die ca. 1300 Bp großen PCR-Fragmente vollständig zu sequenzieren, wurden die in Tabelle 10 aufgeführten IRD800-markierten *Dm-Spn4*-spezifischen Primer verwendet.

Primer	Orientierung	Lokalisation	Sequenz (5' → 3')
IRD800-Dm_SP4_1+	Plusstrang	985 - 1004	CTCGGCGGCTCAGAGCGTGG
IRD800-Dm_SP4_3+	Plusstrang	1384 - 1406	CCTCACGACCCTCTCGCAGATCA
IRD800-Dm_SP4_1-	Minusstrang	1050 - 1071	AGATGGTTGGTGCGCTGCTCAA

Tabelle 10: Verwendete *Dm-Spn4*-spezifische Sequenzierprimer

Die Primer werden zur direkten Sequenzierung von *Dm-Spn4*-spezifischen PCR-Fragmenten verwendet. Für die Beschreibung der Orientierung und Lokalisation siehe Beschriftung von Tabelle 9.

3.7 Bakterienstämme und Medien

Top10 (Invitrogen):

Genotyp: F⁻ *mcrA*Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74* *recA1* *deoR* *araD139* Δ(*ara-leu*)7697 *galU* *galK* *rpsL* (Str^R) *endA1* *nupG*

LB-Flüssigmedium	1,0 % (w/v) SELECT Peptone 140 0,5 % (w/v) SELECT Hefeextrakt 0,5 % (w/v) NaCl oder als Fertigmischung: 2,0 % (w/v) Lennox LB (Luria-Bertani) Powder
SOB-Flüssigmedium	2,0 % (w/v)SELECT Tryptone 0,5 % (w/v)SELECT Hefeextrakt 10 mM NaCl
SOC-Flüssigmedium	10 mM MgCl ₂ 10 mM MgSO ₄ 20 mM Glukose in SOB-Flüssigmedium

Alle Medien werden bei 121 °C und 2 bar für 30 – 60 min autoklaviert und sind bei 4 °C im Kühlschrank für mehrere Wochen lagerbar.

LB-Agar-Plattenmedium	1,0 % (w/v) SELECT Peptone 140 0,5 % (w/v) SELECT Hefe Extrakt 0,5 % (w/v) NaCl 1,2 % (w/v) SELECT Agar oder als Fertigmischung: 3,2 % (w/v) Lennox LB-Agar Powder (Luria-Bertani) + Antibiotika: 100 µg/mL Ampicillin oder 100 µg/mL Carbenicillin
------------------------------	---

LB-Agar wird bei 121 °C und 2 bar für 30 min autoklaviert. Nach Antibiotika-Zugabe (unter ~ 40 °C) wird das Medium anschließend unter sterilen Bedingungen unter der Reinraumwerkbank in Petrischalen gegossen und kann mehrere Monate bei 4 °C gelagert werden.

4. Methoden

4.1 Polymerase-Kettenreaktion

Die von Kary Mullis 1983 entwickelte Polymerase-Kettenreaktion (*polymerase chain reaction*, PCR) ist eine *in vitro* DNA-Amplifikationstechnik. Mit dieser Technik können DNA-Abschnitte gezielt vervielfältigen werden, die von zwei bekannten DNA-Sequenzen eingerahmt werden^{1,2,3}.

4.1.1 Standard PCR-Ansatz

Der folgende PCR-Ansatz^{4,5} wird zur Bestimmung von Primer-Schmelztemperaturen und für präparatorische Zwecke verwendet. Je nach Anforderung wird das Volumen des PCR-Ansatzes (Tabelle 11) zwischen 20 µL und 50 µL variiert. Die PCR-Bedingungen sind in Tabelle 12 aufgeführt. Aufgrund des hier verwendeten Polymerasen-Mix verfügen die PCR-Produkte über einen 3'-dAMP-Überhang.

20 µL PCR-Ansatz	Volumen (µL)	Konz. der Stamm-Lsg.	Konz. im PCR-Ansatz
dH ₂ O	10,8		
10-fach Qiagen PCR-Puffer	2,0	10-fach	1-fach
5-fach Q-Solution	4,0	5-fach	1-fach
dNTP-Mix (jeweils 10 mM)	0,4	je 10 mM	je 200 µM
Plusstrangprimer	0,8	10 µM	0,4 µM
Minusstrangprimer	0,8	10 µM	0,4 µM
Template	~1,0	10 - 1000 ng	
<i>Taq/Pfu</i> -Polymerasen Mix (siehe Abschnitt 3.4.4)	0,2	5 U/µL	0,05 U/µL

Tabelle 11: Ansatz für eine Standard-PCR von 20 µL

¹ Mullis, K., B., Faloona, F., A., 1987.

² Saiki, R., K., *et al.* 1988.

³ Kwok, S., Higuchi, R., 1989.

⁴ QIAGEN GmbH, Taq PCR Handbook 03/2002.

⁵ Sambrook, J., Russel, D., W., 2001.

Zyklen	Temperatur	Zeit
1	94 C°	2 min
25 - 30	94 C°	30 s
	T _A C°	30 s
	72 C°	1 min / kb
1	72 C°	5 min
	5 C°	

Tabelle 12: Standard PCR-Bedingungen¹, mit Pseudo-Hot Start

4.1.2 Kolonie-PCR

Die Methode der Kolonie-PCR² ermöglicht ein template-spezifisches Klon-Screening. Durch die Wahl der Primer kann gezielt nach Insertgröße oder Insertorientierung gesucht werden. Mit Hilfe einer sterilen Plastikspitze wird Koloniematerial in 20 µL dH₂O resuspendiert. Zum Aufschluss der *E. coli*-Zellen wird der resuspendierte Kolonie-Ansatz für 10 min bei 94 °C inkubiert und anschließend für 2 min auf Eis gekühlt. Je nach Reaktionsvolumen der Kolonie-PCR verwendet man 0,25 - 1,0 µL der aufgeschlossenen Kolonie-Lösung als Template. Die PCR-Bedingungen sind von den verwendeten Primern und der erwarteten Insertgröße abhängig. Es wird die DNA-Polymerase KlenThermTM (Genecraft) verwendet. In Tabelle 13 ist der Ansatz für eine Kolonie-PCR aufgeführt. Die PCR-Bedingungen entsprechen den Bedingungen der Standard-PCR (Tabelle 12).

20 µL PCR-Ansatz	Volumen (µL)	Konz. der Stamm-Lsg.	Konz. im PCR-Ansatz
dH ₂ O	14,8		
10-fach KlenTherm-Puffer	2,0	10-fach	1-fach
dNTP-Mix (jeweils 10 mM)	0,4	je 10 mM	je 200 µM
Plusstrangprimer	0,8	10 µM	0,4 µM
Minusstrangprimer	0,8	10 µM	0,4 µM
Template (aufgeschlossenes Koloniematerial)	1,0	10 - 1000 ng	
KlenTherm-Polymerase (siehe Abschnitt 3.4.4)	0,2	5 U/µL	0,05 U/µL

Tabelle 13: Ansatz für die Kolonie-PCR, 20 µL

¹ QIAGEN GmbH, Taq PCR Handbook 03/2002.

² Zon, L., I., *et al.* 1989.

4.2 Reverse Transkription

Das Enzym „Reverse Transkriptase“ ermöglicht das Umschreiben von mRNA (*messenger RNA*) in cDNA (*complementary DNA*). Dabei wird die „RNA-abhängige DNA-Polymerase“-Aktivität der Reversen Transkriptase (RT) genutzt. Reverse Transkriptasen stammen ursprünglich aus Retroviren.

Für die cDNA-Synthese wird standardmäßig das „Omniscript RT“-Kit (Qiagen, Hilden) verwendet (Abschnitt 4.2.1). Die vollständigen 5'-Enden der cDNAs wird mit Hilfe des „GeneRacerTM“-Kits (Invitrogen, Karlsruhe) identifiziert (Abschnitt 4.2.2). Als Ausgangsmaterial dient adulte sowie embryonale Poly(A⁺)-RNA von *Drosophila melanogaster* (BD Biosciences Clontech, Heidelberg).

4.2.1 Reverse Transkription mittels „Omniscript RT“-Kit

Über die Herkunft der Reversen Transkriptase „Omniscript“ macht Qiagen keine Angaben. Es werden die im Kit enthaltenen Puffer und Reagenzien verwendet. Zur Durchführung bzw. Pufferzusammensetzung siehe Herstellerangaben¹.

4.2.2 Reverse Transkription mittels „GeneRacerTM“-Kit

Das GeneRacerTM-Kit² basiert auf der RLM-RACE-Methode (*RNA ligase-mediated rapid amplification of 5' and 3' cDNA ends*), die es ermöglicht vollständige 5'-cDNA-Enden zu erhalten (Abbildung 6). Zunächst werden mit Hilfe des Enzyms CIP (*calf intestinal phosphatase*) die 5'-Phosphatgruppen derjenigen mRNAs abgespalten, die über keine schützende CAP-Struktur verfügen (5'-trunkierte RNAs und DNA-Verunreinigungen). Im zweiten Schritt wird mit dem Enzym TAP (*tobacco acid pyrophosphatase*) die CAP-Struktur von intakten mRNAs spezifisch abgespalten. Durch die Abspaltung wird eine 5'-Phosphatgruppe zugänglich, an die durch das Enzym T4-RNA-Ligase ein RNA-Oligonukleotid ligiert wird. Nach der folgenden Reversen Transkription befindet sich am 3'-Ende des cDNA-Erststranges (Minusstrang) die bekannte Oligonukleotid-Sequenz.

¹ OmniscriptTM Reverse Transcriptase Handbook, Qiagen, 2002.

² Instruction Manual „GeneRacerTM“ Kit, Invitrogen 2002.

Durch Kombination eines genspezifischen Plusstrang-Primers mit einem zur Oligonukleotid-Sequenz komplementären Minusstrang-Primer, kann eine exponentielle Amplifikation des vollständigen 5'-Endes erreicht werden. Es wird die im Kit enthaltene Reverse Transkriptase „SuperScript II“ (MMLV-RT, *moloney murine leukemia virus*) verwendet. Zur Durchführung bzw. Pufferzusammensetzung siehe Herstellerangaben¹.

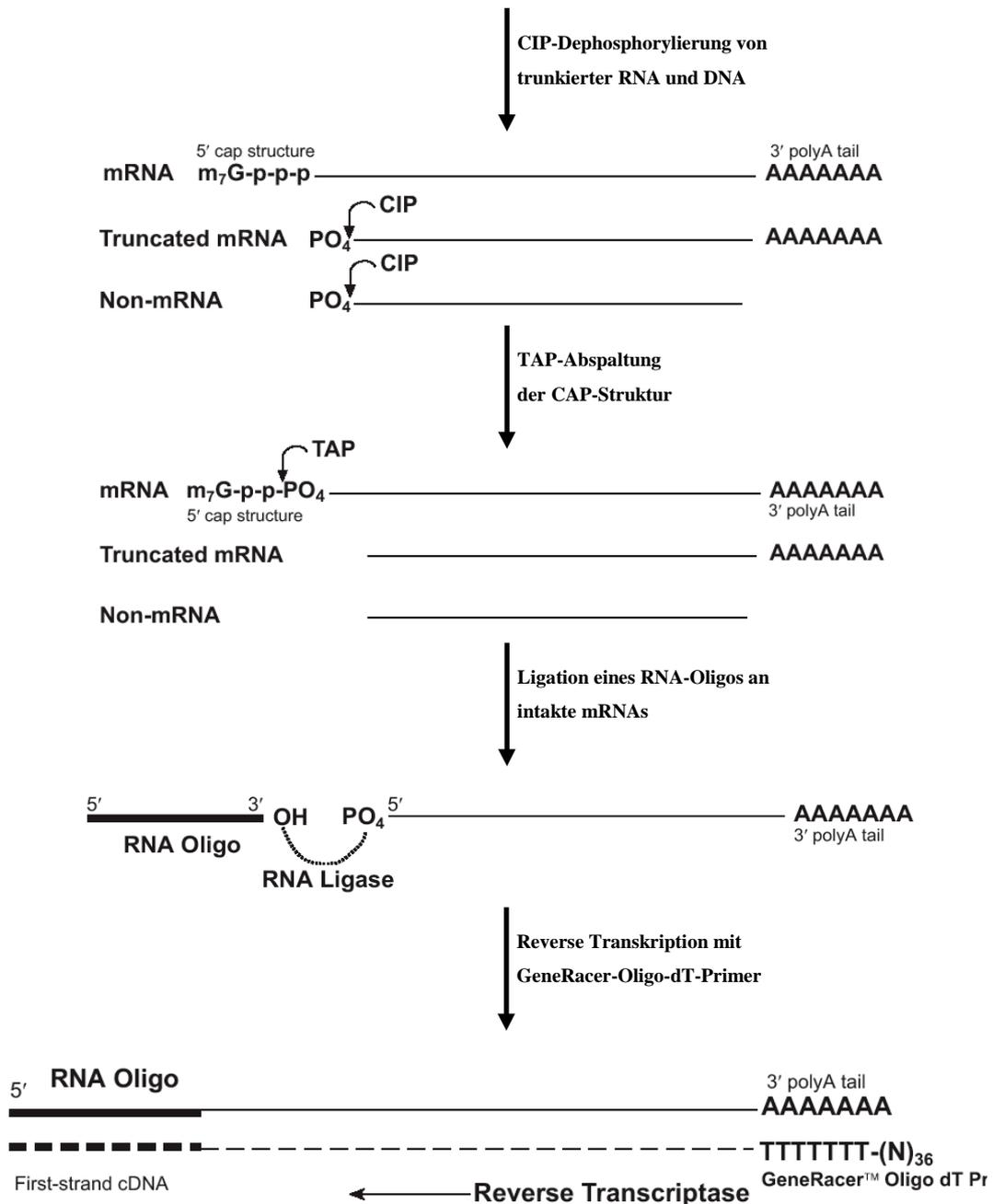


Abbildung 6: RLM-RACE-Methode (Invitrogen GeneRacer™ Kit)¹

¹ Instruction Manual „GeneRacer™ Kit, Invitrogen 2002.

4.3 Analyse und Präparation von DNA

4.3.1 Analyse der PCR-Produkte mittels Agarose-Gelelektrophorese

Bei der Agarose-Gelelektrophorese werden DNA-Fragmente elektrophoretisch nach ihren Größen aufgetrennt und durch Anfärbung mit Ethidiumbromid sichtbar gemacht. In der Praxis wird beobachtet, dass die Mobilität der DNA-Fragmente über weite Bereiche proportional zum negativ dekadischen Logarithmus der molaren Masse bzw. der Größe ist^{1,2}. Der optimalen Trenngrößenbereich wird durch Wahl der geeigneten Gel-Agarosekonzentration erzielt.

Die Agarose wird in TAE-Puffer aufgekocht. Nach Gießen und Aushärten des Agarose-Gels wird mit TAE-Puffer überschichtet. Zu den zu analysierenden DNA-Proben gibt man 1/10 Volumen des 10-fach konzentrierten Probenpuffers und pipettiert die Proben in die Probenaschen. Um den Banden später eine Größe zuordnen zu können, trägt man parallel DNA-Längenstandard-Marker (1 kb-Leiter oder 100 bp-Leiter) als Größenvergleich auf. Die Agarose-Gelelektrophorese wird bei 20 V/cm für ca. 30 - 60 min durchgeführt. Nach Auftrennung der DNA-Fragmente wird das Agarose-Gel in einem Ethidiumbromid-Färbebad angefärbt. Durch Bestrahlung mit UV-Licht (312 nm) werden die fluoreszierenden DNA-Banden sichtbar, mit Hilfe der Software Photofinish fotografiert und gegebenenfalls mit WinCam 2.2 ausgewertet. Die zur Durchführung der Agarose-Gelelektrophorese benötigten Puffer sind im Materialteil (Abschnitt 3.5) aufgeführt.

4.3.2 Aufreinigung der PCR-Produkte mittels „QIAquick Gel Extraction“-Kit

Zur Aufreinigung der PCR-Produkte wird zunächst eine Agarose-Gelelektrophorese durchgeführt (Abschnitt 4.3.1). Nach Ethidiumbromid-Färbung des Agarose-Gels werden die gewünschten DNA-Banden mit einem Skalpell bei möglichst kurzer UV-Bestrahlung aus dem Gel geschnitten. Die Gelextraktion wird mit Hilfe des „QIAquick Gel Extraction“-Kits³ (Qiagen GmbH, Hilden) durchgeführt. Die DNA/Gel-Lösung wird hierbei in Gegenwart von chaotropen Salzen bei einem pH-Wert von $\leq 7,5$ an eine Silicagel-Membran einer Spin-Säule

¹ McDonell, M., W., *et al.*, 1977.

² Kozulic, B., 1995.

³ QIAGEN GmbH, Gel Extraction Kit Protocol, QIAquick Spin Handbook, 2002.

gebunden. Nach einem Waschschrift wird die gebundene DNA mit einem kleinen Volumen Tris-HCl (10 mM, pH 8,5) durch eine pH-Änderung von der Silicagel-Membran der Säule eluiert. Zur Durchführung bzw. Pufferzusammensetzung siehe Herstellerangaben.

4.3.3 Aufreinigung der PCR-Produkte mittels „QIAquick PCR Purification“-Kit

Die Aufreinigung eines PCR-Ansatzes wird mit Hilfe des „QIAquick PCR Purification“-Kits¹ (Qiagen GmbH, Hilden) durchgeführt. Gemäß des Hersteller-Protokolls wird der PCR-Ansatz in Gegenwart von chaotropen Salzen bei einem pH-Wert von $\leq 7,5$ an eine Silicagel-Membran einer Spin-Säule gebunden. Nach einem Waschschrift wird die gebundene DNA mit einem kleinen Volumen Tris-HCl (10 mM, pH 8,5) durch eine pH-Änderung von der Silicagel-Membran der Säule eluiert. Zur Durchführung bzw. Pufferzusammensetzung siehe Herstellerangaben.

4.3.4 Plasmidisolierung mittels „QIAprep[®] Miniprep“-Kit

Die Plasmidisolierung wird mit Hilfe des „QIAprep[®] Miniprep“-Kits der Firma Qiagen durchgeführt. Hierbei handelt es sich um eine modifizierte Form der Methode der alkalischen Lyse². Nach alkalischer Lyse der Bakterienzellen und Neutralisation der Lösung wird die Plasmid-DNA in Gegenwart von chaotropen Salzen bei einem pH-Wert von $\leq 7,5$ an eine Silicagel-Membran einer Spin-Säule gebunden. Nach Waschschriften zur Entfernung von Nukleasen und Salzen wird die gebundene DNA mit einem kleinen Volumen Tris-HCl (10 mM, pH 8,5) durch eine pH-Änderung von der Silicagel-Membran der Säule eluiert. Zur Durchführung bzw. Pufferzusammensetzung siehe Herstellerangaben. Die Plasmidlösung dient als Ausgangsmaterial für die Restriktionsspaltung (Abschnitt 4.4.3) und DNA-Sequenzierung (Abschnitt 4.4.4).

¹ QIAGEN GmbH, PCR Purification Kit Protocol, QIAquick Spin Handbook, 2002.

² Birnboim, H., C., Doly, J., 1979.

4.3.5 Anlegen von Glycerinkulturen

Zur Langzeitlagerung werden Glycerinkulturen¹ der sequenzierten Klone angelegt. Zu 415 μL einer „über Nacht“ Kultur werden 85 μL 87 %iges, autoklaviertes Glycerin (14,8 % v/v Endkonzentration) pipettiert. Die Glycerinkulturen werden bei $-80\text{ }^{\circ}\text{C}$ gelagert.

4.3.6 Bestimmung der DNA-Konzentration

Die Konzentration einer DNA-Lösung lässt sich durch Messung der Absorption bei 260 nm im Vergleich zum reinen Lösungsmittel bestimmen. Bei der Berechnung der Konzentration (c) aus der gemessenen optischen Dichte (oD) einer 1 cm Quarz-Küvette wird unterschieden, ob es sich um doppelsträngige (ds) oder single stranded, (ss) DNA oder um Oligonukleotide (z.B. Primer) handelt².

dsDNA :	$c = 50 \cdot \text{oD}_{260\text{ nm}} \mu\text{g/mL}$
ssDNA :	$c = 40 \cdot \text{oD}_{260\text{ nm}} \mu\text{g/mL}$
Oligonukleotide :	$c = 30 \cdot \text{oD}_{260\text{ nm}} \mu\text{g/mL}$

Um bei Plasmidisolierungen die Verunreinigung der DNA durch Proteine (Absorptionsmaximum bei 280 nm) abschätzen zu können, wird der Quotient $\text{OD}_{260\text{nm}}/\text{OD}_{280\text{nm}}$ bestimmt. Bei reiner DNA liegt dieser Quotient bei $\geq 1,8$. Die Konzentration einer DNA-Lösung lässt sich auch mittels Agarose-Gelelektrophorese densitometrisch bestimmen. Eine Verdünnungsreihe der Probe wird mit einer Referenz (bekannte DNA-Konzentration) elektrophoretisch aufgetrennt (Abschnitt 4.3.1). Das angefärbte Agarose-Gel wird fotografiert und mit Hilfe der Software WinCam 2.2 densitometrisch ausgewertet.

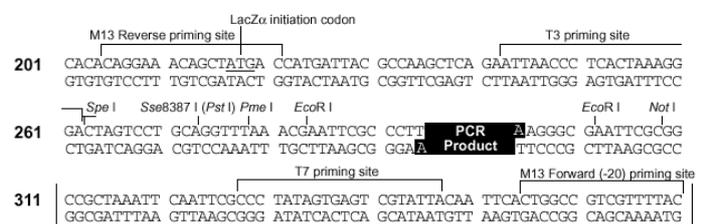
¹ Sambrook, J., Russell, D., W., 2001.

² Sambrook, J., Russell, D., W., 2001.

4.4 Klonierung und Klonanalyse

4.4.1 TA-Klonierung mittels „TOPO TA Cloning® Kit for Sequencing“

Das verwendete „TOPO TA Cloning® Kit for Sequencing“ (Invitrogen, Karlsruhe) erlaubt eine schnelle Klonierung von PCR-Fragmenten mit 3'-dAMP-Überhang. Das klonierte PCR-Fragment wird im Vektor von Bindestellen für Standardprimer (T3-Promotor, T7-Promotor, M13-reverse und M13-forward) flankiert, wodurch eine anschließende Sequenzierung des Inserts erleichtert wird.



Vektorelemente:

pUC ori	starker Replikationsursprung
P _{lac}	Promotor für das LacZ α -Gen
lacZ α -ccdB	Fusionsprotein aus dem α -Fragment der β -Galaktosidase und dem toxischen ccdB Protein.
MCS	Ausschnitt der „multiple cloning site“ im LacZ α -ccdB-Gen
Kanamycin	Kanamycin-Resistenz-Gen
Ampicillin	Ampicillin-Resistenz-Gen

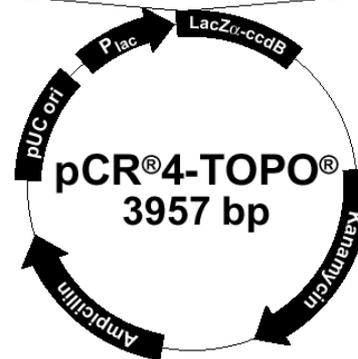


Abbildung 7: Vektorsystem pCR4-TOPO mit Vektorelementen¹

Der Klonierungsvektor verfügt als Selektionsmarker über Ampicillin- und Kanamycin-Resistenz-Gene, einen starken Replikationsursprung (pUC ori) und über eine multiple Klonierungsstelle (*multiple cloning site*, MCS) im lacZ α -ccdB-Reporter-gen. Ein Insert-Einbau in die MCS des pCR4-TOPO wird durch die Inaktivierung der β -Galaktosidase angezeigt (Blau/Weiß-Screening). Zusätzlich wird die Expression des toxischen ccdB-Fusionsproteins² gestoppt. Beim ccdB-Protein handelt es sich um einen Gyrase-Inhibitor, der das Wachstum von Transformanten stoppt, die einen Klonierungsvektor ohne Insert (Religation) aufgenommen haben¹.

¹ Invitrogen Manual, TOPO TA Cloning Kit for Sequencing, 2002.

² Bernard P., et al. 1994.

Das TOPO-System¹ nutzt die Ligations-Aktivität der Topoisomerase I. Der linearisierte pCR4-TOPO-Vektor verfügt an beiden Vektorenden über 3'-dTMP-Überhänge. An der 3'-Phosphatgruppe der Desoxythymidin-Überhänge liegen kovalent gebundene Topoisomerasen I (*Vaccinia* Virus) vor. Topoisomerase I ist für das Aufwinden und Entspannen von DNA verantwortlich und muss dazu DNA binden, schneiden und wieder ligieren. Diese DNA-Bindungs-fähigkeit und die Ligase-Aktivität der Topoisomerase I werden ausgenutzt, um PCR-Produkte mit 3'-dAMP-Überhängen in den Vektor zu ligieren. Dabei verhindern die voluminösen Topoisomerasen, dass sich die Vektorenden gegenseitig annähern, wodurch Religationen ohne Inserteinbau vermieden werden. Abbildung 8 stellt das Funktionsprinzip der „Topo TA Klonierung“ dar.

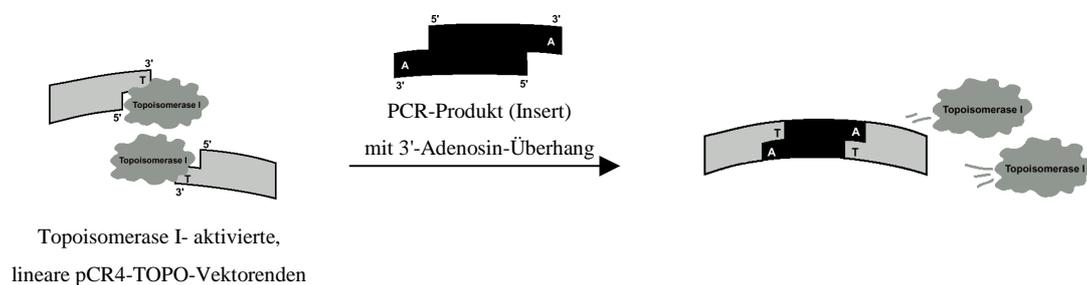


Abbildung 8: Funktionsprinzip der „TOPO TA Klonierung“

Verändert nach Invitrogen, Instruction Manual - TOPO TA Cloning Kit, 2002.

Die Ligation und die Transformation werden nach Herstellerangaben durchgeführt. Nach Abschluss der Transformation werden mit einem sterilen Glasstab 10 µL, 50 µL und 150 µL vom Transformationsansatz auf LB-Agarplatten (100 µg/mL Ampicillin) ausplattiert und über Nacht im Brutschrank bei 37 C° inkubiert. Es können nur Klone wachsen, die eine Plasmid-vermittelte Ampicillin-Resistenz aufweisen. Die Plasmide der gewachsenen Klone werden nach Übertragen auf eine neue Ampicillin-LB-Agarplatte (Klon-Bibliothek) auf Insert-Charakteristika untersucht (Plasmid-Schnell-Charakterisierung, Abschnitt 4.4.2; Kolonie-PCR, Abschnitt 4.1.2; Restriktionsspaltung, Abschnitt 4.4.3).

¹ Invitrogen Manual, TOPO TA Cloning Kit for Sequencing, 2002.

4.4.2 Plasmid-Schnell-Charakterisierung

Die Methode der Plasmid-Schnell-Charakterisierung¹, auch als „Toothpick“-Assay bezeichnet, ermöglicht ein Screening der gewachsenen Klone aufgrund ihrer Plasmidgröße. Mit Hilfe eines sterilen Zahnstochers wird eine Kolonie in 30 µL LB-Flüssigmedium resuspendiert. Durch Zugabe von 30 µL eines frisch gemischtem Lyse/Ladepuffers und nach anschließendem Mischen durch 1 min Vortexen kommt es zur Lyse der resuspendierten Zellen. Nach 3 min Zentrifugieren bei 14000 x g erhält man einen klaren Überstand, der chromosomale DNA und Plasmid-DNA enthält. In einer Agarose-Gelelektrophorese (Abschnitt 4.3.1) werden 12 µL des klaren Überstandes aufgetragen. Die Größe der ccc-Plasmidform (*covalently closed circular*) läßt auf die ungefähre Größe des Plasmid-Inserts schließen. Die Methode wird für ein schnelles Plasmidgrößen-Screening eingesetzt. Klone mit erwarteter Plasmidgröße werden anschließend mit aussagekräftigeren Methoden genauer charakterisiert (Restriktionsspaltung, Abschnitt 4.4.3; Sequenzierung, Abschnitt 4.4.4). Die zur Durchführung benötigten Puffer sind im Materialteil (Abschnitt 3.5) aufgeführt.

4.4.3 Restriktionsspaltung

Die Methode der Restriktionsspaltung² ermöglicht ein Klon-Screening aufgrund der Insertgröße bzw. Insertorientierung. Die pCR4-TOPO-basierenden Plasmide (Abschnitt 4.4.1) werden mit der Restriktionsendonuklease *EcoRI* gespalten. Der Ausgangsvektor pCR4-TOPO enthält zwei *EcoRI*-Restriktionsschnittstellen, die das eingebaute Insert flankieren (Abbildung 7). Eine *EcoRI*-Restriktionsspaltung führt zum Insertausschnitt, der eine leichte Bestimmung der Insertgröße ermöglicht. Als Ausgangsmaterial für die Restriktionsspaltung dient eine Plasmid-Präparation des jeweiligen Klons (Abschnitt 4.3.4). Der Restriktionsansatz (Tabelle 14) wird für 1 h bei 37 °C inkubiert.

Anschließend wird der Restriktionsansatz mittels Agarose-Gelelektrophorese (Abschnitt 4.3.1) analysiert. Bei Klonen, die die erwartete Insertgröße aufweisen, folgt eine Sequenzierung des Plasmid-Inserts (Abschnitt 4.4.4).

¹ Barnes, W., M., 1977.

² Sambrook, J., Russell, D., W., 2001.

Komponenten	Volumen (in μL)	Konzentration der Stamm-Lsg.	Konzentration im Ansatz
dH ₂ O	Restvolumen		
Plasmid-Präparation	2,0 - 8,5	~ 0,2 $\mu\text{g}/\mu\text{L}$	-
<i>Eco</i> RI Puffer (Abschnitt 3.5)	1,0	10-fach	1-fach
<i>Eco</i> RI Restriktionsendonuklease	0,5	-	-

Tabelle 14: Allgemeiner Restriktionsansatz, 10 μL

4.4.4 DNA-Sequenzierung

Die DNA-Sequenzierung erfolgt nach der Sanger-Methode¹ und wird mit einem LI-COR 4200 (MWG Biotech AG, Ebersberg) durchgeführt. Die isolierten und charakterisierten Plasmide werden mit Hilfe des „DYEnamic Direct Cycle Sequencing“-Kits (Amersham Biosciences Europe GmbH, Freiburg) sequenziert. Pro Sequenzierung wird ein Mastermix (Tabelle 15) angesetzt, der auf die 4 ddNTP-Reaktionen aufgeteilt wird (Tabelle 16). Die Bedingungen der Sequenzierreaktionen sind in Tabelle 17 angegeben.

Komponenten	Menge
Plasmid-DNA (Abschnitt 4.3.4)	200 - 500 fmol
IRD800-Primer, 2 μM	1,0 -2,0 μL
dH ₂ O	Restvolumen

Tabelle 15: Mastermix, 21 μL

Bei der Direkt-Sequenzierung von PCR-Produkten werden im Mastermix anstelle der Plasmid-DNA 200 - 300 fmol gelextrahiertes PCR-Produkt (Abschnitt 4.3.2) eingesetzt.

Komponenten	Menge in μL
ddNTP-Reagenz	1,0
Mastermix (Tabelle 15)	4,0

Tabelle 16: ddNTP-Reaktionsansatz, 5 μL (4 Reaktionen pro Sequenzierung)

¹ Sanger, F., Nicklen, S., Coulson, A., R., 1977.

Zyklen	Temperatur in °C	Zeit
1	95	2 min
35	95	30 s
	T _A	30 s
	70	30 s
1	70	1 min
	4	Pause

Tabelle 17: Sequenzier-Programm

Die Annealing-Temperatur T_A ist vom verwendeten Primer abhängig. Zur direkten Sequenzierung von PCR-Produkten wird zusätzlich ein Polishing-Schritt (20 x (95 °C, 30 s ; 70 °C, 60 s)) durchgeführt. Die Sequenzierreaktion wird nach Beendigung des Programms durch Zugabe von 5 µL Loading-Dye gestoppt. Vor dem Auftragen werden die Proben für 2 min bei 70 °C denaturiert und anschließend auf Eis gelagert.

Die benötigten Puffer sind im Materialteil angegeben (Abschnitt 3.5). Die Analyse der erhaltenen DNA-Sequenzen ist im Methodenteil 4.5 beschrieben.

4.5 Auswertung von DNA-Sequenzen

Der DNA-Sequenzierlauf wird mit der Software „*Base ImagIR Data Collection*“ (LI-COR Biotechnology Division) aufgezeichnet. Die automatische Sequenzauswertung (sowie manuelles Korrekturlesen) erfolgt mit der Software „*Base ImagIR Analysis*“ (LI-COR Biotechnology Division). DNA-Sequenzabschnitte einer Probe (z.B. Plasmid-Insert) werden mit AlignIR (LI-COR Biotechnology Division) aligniert bzw. assembliert. Zur Bearbeitung von DNA-Sequenzen werden die „*BCM Sequence Utilities*“ verwendet (<http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html>). Um schließlich die erhaltenen DNA-Sequenzen mit Genbank-Einträgen zu vergleichen, werden geeignete BLAST-Varianten¹ verwendet (NCBI, <http://www.ncbi.nlm.nih.gov/BLAST/>).

¹ Altschul, S., F., *et al.* 1990.

4.6 Identifikation und Analyse von Serpin-Genen

Ein Genomprojekt gliedert sich in die Arbeitsschritte „shotgun“-Sequenzierung, Contig-Assemblierung und Annotation der assemblierten Chromosomen-Sequenzen. Die Annotation umfasst neben den Angaben zu den eigentlichen Genstrukturen (Exon-Intron-Struktur) auch eine Vielzahl von weiteren Informationen wie z.B. Promotor-Elemente und repetitive Elemente.

Um eine belastbare Phylogenie für die Serpin-Proteinfamilie zu erstellen, ist eine Überprüfung der automatisch annotierten Serpin-Genstrukturen notwendig. Die Ziele dieser Überprüfung sind:

- Identifikation von falsch positiven Serpin-Genen,
- Kontrolle und Vervollständigung der annotierten Exon-Intron-Strukturen,
- Identifikation von doppelt annotierten Serpin-Genen,
- Identifikation von bisher nicht annotierten Serpin-Genen.

Für die Überprüfung der Serpin-Gene sind Entscheidungskriterien notwendig, die eine Verifikation oder Falsifikation eines Serpin-Gens ermöglichen.

Die Datenbank **Prosite**¹ erstellt für eine Vielzahl von unterschiedlichen Proteinfamilien kurze Aminosäuresequenzen, so genannte Signaturen oder Muster, die eine schnelle Entscheidung erlauben, ob es sich bei einem Protein um ein Mitglied einer bekannten Proteinfamilie handeln könnte. Die Prosite-Konsensussequenz der Signatur für Serpin-Gene (Serpinsignatur) ist folgendermaßen aufgebaut (Abbildung 9):

[LIVMFY]-x-[LIVMFYAC]-[DNQ]-[RKHQS]-[PST]-F-[LIVMFY]-[LIVMFYC]-x-[LIVMFAH]

Abbildung 9: Konsensussequenz der Serpin-Signatur

Die Serpin-Signatur (*Prosite Signature*, PS00284) ist 11 As lang. Die Aminosäuren sind im Einbuchstabencode angegeben. Eckige Klammern beinhalten diejenigen Aminosäuren, die an dieser Position erlaubt sind. Ein x steht an Positionen, an denen alle Aminosäuren erlaubt sind.

Bei der Serpin-Signatur von Prosite handelt es sich um eine stark konservierte *Turn*-Struktur, die den carboxyterminalen Teil des RCL auf der Serpin-„Rückseite“ ins Proteingerüst

¹ Hofmann, K., *et al.* 1999. (<http://us.expasy.org/prosite/>).

zurückführt und arretiert (Abschnitt 2.1.2, Abbildung 1). Trotz ihrer kritischen biologischen Funktion weist die Region bei ca. 500 bekannten Serpin-Sequenzen eine beachtliche Variabilität auf, die sich in einer relativ unstringenten Serpin-Signatur widerspiegelt. Als Folge kommt es häufig zu falsch positiven Treffern, d.h. ein Gen wird irrtümlich als Serpin annotiert.

In der Datenbank **Pfam**¹ (*Protein families database of alignments and HMMs*) wird zunächst für jede Proteinfamilie ein multiples Aminosäuresequenz-Alignment der konservierten Domänen durchgeführt. Diese Alignments dienen wiederum als Basis für die Konstruktion von „Hidden Markov Models“ (HMMs). HMMs stellen statistische Modelle einer Proteinfamilie dar und können in Programmen wie HMMER² als Sucheinträge für Datenbanken dienen. Ein Vorteil dieser statistischen Methode ist die Bewertung der „Treffer“ mit Erwartungswerten (E-value). Da die Methode fast die vollständige Serpin-Sequenz berücksichtigt, ist ein Pfam-Treffer deutlich aussagekräftiger und stellt ein geeignetes Mittel zur Verifikation eines Serpins dar.

4.6.1 Serpin-Suche mit AGREP

Das „*approximate pattern matching tool*“ AGREP³ 3.37 (*approximate grep*) ermöglicht die Suche nach unscharfen Mustern in Datenbanken, wie z.B. in einer Genomsequenz. AGREP bietet dazu eine Vielzahl von Optionen wie z.B. das Setzen von Wildcards (an diesen Positionen des Suchmusters sind alle Zeichen erlaubt) und das Zulassen von Fehlern (es sind Mismatches zwischen Suchmuster und Datenbankeintrag erlaubt). AGREP beinhaltet verschiedene Suchalgorithmen. Zur Optimierung der Suche wird für jeden der unterschiedlich komplexen Suchmuster-Typen der entsprechende Suchalgorithmus ausgewählt.

AGREP 3.37 wird hier verwendet, um Serpin-Gene in verschiedenen Genomen zu suchen und ihre Exon-Intron-Struktur gegebenenfalls zu vervollständigen. Das Arbeiten mit AGREP erfordert eine Formatierung der Genomsequenz (bzw. ihrer Teilsequenzen). Die chromosomale DNA-Sequenz wird mit dem Programm DNA2AA in die sechs Aminosäure-Leseraster übersetzt und für die AGREP-Suche formatiert⁴. Der aufbereitete Datensatz wird

¹ Bateman, A., et al. 2002. (<http://www.sanger.ac.uk/Software/Pfam/index.shtml>).

² Eddy, S., R., 1998.

³ Wu, S., Manber, U., 1992.

⁴ Krüger, O., unveröffentlicht.

mit Serpin-typischen Suchmustern (Aminosäuresequenz) durchsucht. Dazu wurden basierend auf multiplen Serpin-Alignments^{1,2} drei stark konservierte Bereiche identifiziert und als Serpin-Suchmuster verwendet (Tabelle 18).

Bezeichnung	Position	Suchmuster
s3A-Breach-s4C	186 - 208	[NS]..[HYF]F[KR][GA].W...F...T...F
s5A-s4A	334 - 351	[HQ][KR]A...[DN][DE][DE]G[TS]EAA..[TS]
s1C-Turn-s4B	364 - 386	F..[DN][HRK]PF.[FLV]#F.G

Tabelle 18: Serpin-Suchmuster

Die Positionen der Suchmuster sind in Bezug auf die α_1 -Antitrypsin-Aminosäuresequenz angegeben. Die Suchmuster (Aminosäuren-Einbuchstabencode) sind in der AGREP-Notation angegeben: „.“ steht für ein Wildcardsymbol; „#“ steht für eine variable Anzahl von Positionen. Eckige Klammern beinhalten diejenigen Aminosäuren, die an dieser Position erlaubt sind.

Das Muster „s3A-Breach-s4C“ beinhaltet die *Breach*-Region und die β -Faltblattstränge s3A und s4C. Das Muster „s5A-s4A“ umfasst die β -Faltblattstränge s5A, s4A und die *Hinge*-Region. Bei der *Turn*-Struktur „s1C-Turn-s4B“ handelt es sich um eine Variante der Serpin-Signatur von Prosite. Es ist teilweise notwendig, die in Tabelle 18 gezeigten Suchmuster an die einzelnen Genome anzupassen. So sind z.B. die Suchmuster „s3A-Breach-s4C“ und „s5A-s4A“ im Genom von *Caenorhabditis elegans* oft durch Introns zertrennt.

¹ Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

² Irving, J.,A., Pike, R., N., Lesk, A., M., Whisstock, J., C., 2000.

4.7 Phylogenetische Analyse

4.7.1 Alignments

Zur Durchführung der phylogenetischen Analyse werden die Aminosäuresequenzen der Serpine zunächst aligniert (*to align*: anreihen, angleichen). Die Aminosäure-Alignments dienen als Ausgangspunkt zur Konstruktion von Serpin-Stammbäumen und werden auch zur Identifikation von Intronpositionen verwendet. Die Alignments werden mit Hilfe der Programme DIALIGN 2.2.1^{1,2} (*diagonal alignment*) und ClustalX 1.83³ (*cluster alignment*) erstellt.

DIALIGN 2.2.1 basiert auf einem so genannten „*segment-to-segment*“-Ansatz. Zunächst werden alle Segmentpaare (Matrix-Diagonalen) identifiziert, die ein definiertes Konsistenz-Kriterium erfüllen. Nach Gewichtung der Segmentpaare wird schließlich die optimale Segmentpaar-Abfolge gesucht, die das Alignment repräsentiert. Die Stärken von DIALIGN liegen daher besonders im lokalen Alignment.

ClustalX 1.83 führt zunächst ein paarweises Alignment aller Sequenzkombinationen durch. Anhand der ermittelten Distanzen wird für den eigentlichen Alignierungsprozess ein Guide-Tree nach der Neighbor-Joining-Methode (Abschnitt 4.7.3) erstellt. Zur Erstellung des multiplen Alignments wird mit der Alignierung der ähnlichsten Sequenzen begonnen und auf diese Weise ein Kern-Alignment gebildet. Zu diesem werden sukzessiv Sequenzen hinzugenommen, wobei die Reihenfolge durch die Verzweigungsordnung des Guide-Tree bestimmt wird. Die Stärken von ClustalX liegen besonders im globalen Alignment⁴.

Die erhaltenen Alignments werden mit Hilfe von GeneDoc⁵ überprüft und gegebenenfalls optimiert. Zur Darstellung der Alignments wird ESPript 2.0⁶ verwendet.

¹ Morgenstern, B., Atchley, W., R., Hahn, K., Dress, A., 1998.

² Morgenstern, B., 1999.

³ Higgins, D., Thompson, J., Gibson, T., Thompson, J., D., Higgins, D., G., Gibson T., J., 1994.

⁴ Lassmann, T., Sonnhammer, E., L., 2002.

⁵ Nicholas, K., B., Nicholas Jr., H., B., Deerfield II., D., W., 1997.

⁶ Gouet, P., Courcelle, E., Stuart, D., I., Metoz, F., 1999.

4.7.2 Maximum-Likelihood-Analyse

Das Ziel des „character state“-basierten Maximum-Likelihood-Verfahrens (ML) ist es, unter Berücksichtigung von Evolutionsmodellen (bzw. Aminosäure-Substitutionsmatrizen) die Parameter Kantenlängen bzw. Substitutionsraten für eine gewählte Topologie zu optimieren, bis der Sequenz-Datensatz mit maximaler Wahrscheinlichkeit durch die gewählte Topologie beschrieben wird^{1,2,3}.

Die Parameter-Optimierung und Berechnung dieser Wahrscheinlichkeit wird für alle Topologien durchgeführt. Die Topologie für welche der Sequenz-Datensatz die höchste Wahrscheinlichkeit aufweist, wird als Maximum-Likelihood-Baum ausgewählt. Da bei ML alle (hinreichend vernünftigen) Topologien betrachtet werden, handelt es sich um ein sehr rechenintensives Verfahren, das in dieser Form nur für kleine Datensätzen oder wenige Baumtopologien anwendbar ist.

Der „*Quartett Puzzling*“-Algorithmus (QP) ermöglicht die Anwendung des ML-Verfahrens für größere Datensätze⁴. Beim QP wird der Datensatz zunächst in Quartette zerlegt, deren „beste“ ML-Topologie berechnet wird (ML-Schritt). Im *Puzzling*-Schritt wird ein Quartett zufällig als Kerntopologie ausgewählt, zu dem sukzessiv weitere OTUs hinzugefügt werden, bis schließlich eine Gesamt-Topologie konstruiert ist. Dieser *Puzzling*-Schritt wird mindestens 1000 Mal wiederholt, wobei die Anzahl der *Puzzling*-Schritte in Abhängigkeit von der Größe des Datensatzes gewählt werden sollte. Im dritten Schritt wird durch Anwenden der Mehrheitsregel ein Konsensusbaum (QP-Baum, *majority-rule consensus* > 50 %) gebildet. Durch die wiederholten *Puzzling*-Schritt kann ein Zuverlässigkeitswert (*reliability value*) für eine OTU-Gruppierung bestimmt werden, der angibt, in wieviel Prozent der ermittelten Topologien die OTU-Gruppierung wiedergefunden wird. Der *reliability value* kann als methoden-interner *Bootstrap*-Wert (Abschnitt 4.7.5) betrachtet werden.

Die Maximum-Likelihood-Analyse wird mit Hilfe von Tree Puzzle 5.0⁴ durchgeführt. Folgende Parameter weichen von der Standardeinstellung ab:

- Exakte Parameterschätzung (*Quartet sampling* + *NJ tree*)
- JTT-Substitutionsmatrix (Jones *et al.* 1992)

¹ Wägele, J., W., 2000.

² Nei, M., Kumar, S., 2000.

³ Li, W., H., 1997.

⁴ Strimmer, K., von Haeseler, A., 1996.

4.7.3 Neighbor-Joining-Analyse

Beim Neighbor-Joining-Verfahren (NJ) handelt es sich um eine vereinfachte, heuristische Form des „Minimum Evolution“-Ansatzes¹. Das Optimierungskriterium „minimale Evolution“ bzw. „minimale Evolutionskosten“ kann beschrieben werden, als Suche nach derjenigen Topologie, die die kürzeste Baumlänge (Summe alle Baumkanten) aufweist^{2,3,4}.

Bei NJ handelt es sich um ein Distanzverfahren, das auf Evolutionsmodellen bzw. Aminosäure-Substitutionsmatrizen (Abschnitt 2.3.1) basiert. Bei Distanzverfahren werden unter Zuhilfenahme von Evolutionsmodellen die evolutionären Distanzen zwischen allen Sequenzpaaren (OTUs) eines multiplen Alignments ermittelt. Die erhaltene Distanzmatrix bildet die Grundlage der Baumkonstruktion.

Der NJ-Algorithmus beginnt mit einem so genannten stern-förmigen Baum (bzw. Busch). Er besteht aus einem zentralen Vertex *, an welchen alle OTUs durch jeweils eine Kante angehängt sind. Es wird unter allen von dem zentralen Vertex * ausgehenden Kanten nach demjenigen OTU-Paar (Neighbors) gesucht, deren Gruppierung als ein gemeinsamer Cluster nach entsprechender Berechnung zu einer minimalen Baumlänge führt. Der neue Cluster ersetzt die beiden enthaltenen OTUs in der Distanzmatrix, die anschließend aktualisiert wird. Dieser Prozess wird iteriert, bis von dem zentralen Vertex * nur noch drei Kanten ausgehen und damit ein binäre Baum entstanden ist.

Die Neighbor-Joining-Analyse wird mit PAUP*4.0b10⁵ (*Phylogenetic Analysis Using Parsimony, *and other methods*) durchgeführt. Folgende Parameter weichen von der PAUP-StandardEinstellung ab:

- Minimum Evolution, Distanz-Kriterium
- PAM250 bzw. BLOSUM62 als Aminosäure-Substitutionsmatrix⁶
- 1000 *Bootstrap*-Replikate

¹ Saitou, N., Nei, M., 1987.

² Wägele, J., W., 2000.

³ Nei, M., Kumar, S., 2000.

⁴ Li, W., H., 1997.

⁵ Swofford, D., L., 2003.

⁶ Dayhoff, M., O., *et al.* 1978, konvertiert von Kuzoff, R., K.

4.7.4 Maximum-Parsimony-Analyse

Das Maximum-Parsimony-Verfahren (MP) basiert auf dem Sparsamkeitsprinzip, gemäß dem diejenige Erklärung (Topologie) der Wahrheit am nächsten kommt, die die wenigsten Annahmen (evolutionäre Schritte) benötigt¹. Das „character state“-basierte MP-Verfahren berücksichtigt keine Evolutionsmodelle^{2,3,4}.

Wie beim Maximum-Likelihood-Verfahren wird bei MP jede Position eines multiplen Alignments isoliert betrachtet. Für jede Position des Alignments werden die Baumängen (Anzahl an notwendigen Substitutionen) für alle denkbaren Topologien berechnet. Anschließend wird für jede Topologie die Substitutionsanzahl über alle Alignment-Positionen addiert. Die Topologie(n) mit der kleinsten Anzahl an Substitutionen gibt/geben den MP-Baum wieder. Da bei MP alle Topologien betrachtet werden, handelt es sich um ein sehr rechenintensives Verfahren, das in dieser Form (*exhaustive search*) nur für kleine Datensätzen ($n < 10$) anwendbar ist. Für größere Datensätze existieren verschiedene Suchstrategien⁵, um den optimalen MP-Baum zu identifizieren, wie zum Beispiel auch die „*branch-and-bound*“-Variante⁶ oder die verschiedenen „*heuristic search, branch swapping*“-Varianten. Die Maximum-Parsimony-Analyse wird mit PAUP*4.0b10⁷ (*Phylogenetic Analysis Using Parsimony, *and other methods*) durchgeführt. Im Rahmen der vorliegenden Arbeit wird der *heuristic search*-Algorithmus in Verbindung mit der *branch swapping*-Variante „*tree bisection reconnection*“ (TBR) angewendet. Bei TBR wird der mittels *heuristic search* provisorisch konstruierte Baum an einer zufällig ausgewählten Kante in zwei Teilbäume zerlegt (*tree bisection*). Die zwei Teilbäume werden anschließend wieder über zwei zufällig ausgewählte Kanten verbunden (*reconnection*). Ausgehend vom provisorischen Baum werden mit dem TBR-Algorithmus viele alternative Topologien erzeugt. Der Baum mit der kürzesten Baumlänge wird als MP-Baum gewählt.

¹ Ockham's Razor, „Don't multiply entities beyond necessity“, William of Ockham (1287-1347).

² Wägele, J., W., 2000.

³ Nei, M., Kumar, S., 2000.

⁴ Li, W., H., 1997.

⁵ Nixon, K., C., Davis, J., I., Goloboff, P., A., 1998.

⁶ Hendy, M., D., Penny, D., 1982.

⁷ Swofford, D., L., 2003.

Folgende Parameter weichen von der PAUP-StandardEinstellung ab:

- Maximum-Parsimony, Parsimony-Kriterium
- PAM250 bzw. BLOSUM62 als Aminosäure-Substitutionsmatrix¹
- 100 *Bootstrap*-Replikate

4.7.5 *Bootstrap*-Testverfahren

Das *Bootstrap*-Testverfahren² (*to bootstrap*: sich selbst an den Riemen seiner Stiefel aus dem Sumpf hochziehen) ist der am häufigsten verwendete Wiederfindungswahrscheinlichkeits-Test. *Bootstrap*-Verfahren werden verwendet, um die Belastbarkeit des konstruierten Stammbaums zu überprüfen. Die Basis des *Bootstraps* ist der originale Datensatz, von dem 100 bis 1000 Pseudoreplikate angelegt werden. Ein Pseudoreplikat wird generiert, indem Spalten des originalen Alignments zufällig ausgewählt und neu zusammengestellt werden, bis der ursprüngliche Datensatzumfang erreicht ist. Dadurch können in Pseudoreplikaten einige Merkmale (Alignmentspalten) häufiger vorliegen oder aber ganz fehlen. Die Stammbäume, die aufgrund dieser Pseudoreplikate konstruiert werden, fasst man schließlich in einem Konsensusbaum (*strict consensus* oder *majority-rule consensus*) zusammen. Für jede OTU-Gruppierung wird ein prozentualer *Bootstrap*-Wert (*bootstrap confidence value*, P_B) ermittelt. Dieser gibt an, wie häufig die Gruppierung in den Replikatbäumen wiedergefunden wird. Im allgemeinen können Gruppierungen mit einem *Bootstrap*-Wert von $> 95\%$ als hoch signifikant betrachtet werden.

Da für jeden der erzeugten Replikat-Datensätze ein phylogenetischer Baum konstruiert wird, kann das *Bootstrap*-Testverfahren für die „character state“-basierten Baumkonstruktions-Methoden (ML und MP) sehr rechenintensiv werden.

¹ Dayhoff, M., O., *et al.* 1978, konvertiert von Kuzoff, R., K.

² Felsenstein, J., 1985.

5. Ergebnisse

Zur Analyse der Phylogenie der Serpin-Gene in unterschiedlichen eukaryotischen Organismen wurden Genom-Analysen mehrerer eukaryotischer Modellorganismen durchgeführt. Die vorliegende Arbeit konzentrierte sich auf *Caenorhabditis elegans* (Abschnitt 5.1), *Drosophila melanogaster* (Abschnitt 5.2), *Arabidopsis thaliana* (Abschnitt 5.3) und *Ciona intestinalis* (Abschnitt 5.4). Für jedes Genom wurden die folgenden Schritte durchgeführt:

- Überprüfung der Genstrukturen der annotierten Serpin-Gene (Abschnitt 4.6)
- Suche bzw. Lokalisation der Serpin-Gene im Genom (Abschnitt 4.6.1)
- Identifikation von Serpin-Pseudogenen bzw. falsch positiven Serpin-Genen (Abschnitt 4.6)
- Phylogenetische Analyse der Serpin-Gene und ihrer abgeleiteten Aminosäuresequenzen (Abschnitt 4.7)

In Abschnitt 5.5 sind die Interspezies-Serpin-Stammbäume dargestellt, die jeweils mit Neighbor-Joining-, Maximum-Likelihood- und Maximum-Parsimony-Verfahren rekonstruiert wurden.

Im Rahmen dieser Arbeit konnten Serpin-Gene identifiziert werden, die über multiple RCL-kodierende Exon-Varianten verfügen (Abschnitt 5.6). Durch Experimente wurde anhand des Gens *Dm-Spn4* gezeigt, dass durch alternatives Spleißen der RCL-kodierenden Exon-Varianten die funktionelle Vielfalt dieser Serpin-Gene erhöht wird. Als Vergleich dienten die Serpin-Gene *Ag-SRPN10* und *Ms-Serpin-Gene-1*, für die auch ein alternatives Spleißen gezeigt werden konnte^{1,2}.

- *Dm-Spn4* von *Drosophila melanogaster* (Abschnitt 5.6.1)
- *Ce-Srp-7* und *Ce-Srp-9* von *Caenorhabditis elegans* (Abschnitt 5.6.2)
- *Ci-Spn-10* von *Ciona intestinalis* (Abschnitt 5.6.3)
- *SRPN10* von *Anopheles gambia* (Abschnitt 5.6.4)
- *Serpin-Gene-1* von *Manduca sexta* (Abschnitt 5.6.4)

¹ Danielli, A., Kafatos, F., C., Loukeris, T., G., 2003.

² Jiang, H., Wang, Y., Huang, Y., Mulnix, A., B., Kadel, J., Cole, K., Kanost, M., R., 1996.

5.1 Analyse der Serpin-Gene von *Caenorhabditis elegans*

5.1.1 Modellorganismus *Caenorhabditis elegans*

Caenorhabditis elegans (*C. elegans*) ist ein ca. 1 mm langer Fadenwurm (Nematoda, Abbildung 10). Bis heute konnte die Position der Nematoden im Reich der Animalia noch nicht eindeutig geklärt werden. Vermutlich spalteten sich die Nematoden vor ca. 600 bis 1200 Millionen Jahren von den Deuterostomia ab¹. *C. elegans* lebt im Boden und ernährt sich von Mikroorganismen. Ein adultes Exemplar besteht aus exakt 959 somatischen Zellkernen, deren Entwicklung aufgrund der Transparenz des Wurms sehr gut zu verfolgen ist. Aufgrund der leichten Kultivierbarkeit und eines relativ kurzen Lebenszyklus (ca. 2 bis 3 Wochen) wird *C. elegans* seit mehr als 30 Jahren als Modellorganismus verwendet¹. Das diploide Genom besteht aus sechs Chromosomenpaaren (fünf autosomale Chromosomenpaare I, II, III, IV und V) und einem Paar Geschlechtschromosomen (XX). Man kann bei *C. elegans* zwei



Geschlechter unterscheiden, Hermaphroditen (Zwitter, XX) und die weitaus selteneren Männchen (X0). Die Sequenz des *C. elegans*-Genoms wurde 1998 veröffentlicht². Das Genom umfasst ca. 97 Millionen Basenpaare und kodiert ca. 19000 Gene. Derzeit wird das *C. elegans*-Genom durch die WormBase-Datenbank³ betreut und aktualisiert.

Abbildung 10: Fadenwurm *Caenorhabditis elegans*

(Quelle: <http://www.wormbase.org/about/mission.html>)

¹ Blaxter, M., 1998.

² The *C. elegans* Sequencing Consortium, 1998.

³ <http://www.wormbase.org/>, Todd, W., H., et al. 2003.

5.1.2 Die Serpin-Gene von *Caenorhabditis elegans*

Als Basis dieser Untersuchungen diente die WormBase-Datenbank¹. Die chromosomale Genomsequenzen sind über das Sanger Centrum² zugänglich. Tabelle 19 führt die Serpin-Gene von *C. elegans* auf.

Genname (CGC)	Synonyme (Zang <i>et al.</i> 2001)	Genmodell / Transkript	Protein- Größe (As)	cDNA- Größe (Nt)	Gen- Größe (Bp)	WORF DB
Ce-Srp-1	<i>Ce</i> -SPN-05	C05E4.3	366	>1152	>1677	cDNA
Ce-Srp-2	<i>Ce</i> -SPN-06	C05E4.1	359	>1158	>1729	cDNA
Ce-Srp-3	<i>Ce</i> -SPN-09	Y32G9A.4	362	>1172	>1765	cDNA
Ce-Srp-5	<i>Ce</i> -SPN-03	C03G6.18	329	>1048	>1347	- keine -
Ce-Srp-6	<i>Ce</i> -SPN-04	C03G6.19	375	>1186	>1821	cDNA
Ce-Srp-7	<i>Ce</i> -SPN-01	Variante 7A	366	>1209	>2496	cDNA
		Variante 7B	366	>1102	>2496	- keine -
		Variante 7C	365	>1578	>2496	- keine -
Ce-Srp-8	<i>Ce</i> -SPN-02	F20D6.3	366	>1125	>1479	cDNA
Ce-Srp-9	<i>Ce</i> -SPN-07	Variante 9A	356	>1198	>3726	cDNA
		Variante 9B	365	>1730	>3726	- keine -
Ce-Srp-10	<i>Ce</i> -SPN-08	F09C6.4	346	>1612	>1941	- keine -

Tabelle 19: Serpin-Gene von *Caenorhabditis elegans*

Die Gennamen basieren auf der Nomenklatur des *Caenorhabditis* Genetics Centers (CGC). In der Spalte „Synonyme“ ist die Serpin-Nomenklatur nach Zang *et al.* 2001. aufgeführt. Ein Serpin mit der Bezeichnung Srp-4 existiert nicht. Unter Protein-Größe ist die Größe des Präproteins in Aminosäuren (As) angegeben. Die cDNA- und Gen-Größen sind in Nukleotiden (Nt) bzw. in Basenpaaren (Bp) angegeben. Die Größenangaben umfassen die DNA-Sequenz beginnend mit dem Startcodon bis einschließlich Poly(A)-Signal. Unter WORFDB³ sind cDNA- und PCR-Experimente aufgeführt, die die Genmodelle stützen. Die Serpin-Gene *Ce*-Srp-7 und *Ce*-Srp-9 verfügen über multiple RCL-kodierende Exon-Varianten, die alternatives Spleißen ermöglichen (Abschnitt 5.6.2).

Die chromosomalen DNA-Sequenzen des Genoms von *C. elegans* wurden wie im Abschnitt 4.6.1 beschrieben, in die sechs Aminosäure-Raster translatiert und für eine AGREP-Suche

¹ WormBase, <http://www.wormbase.org>, release WS99, Mai 2003

² ftp://ftp.sanger.ac.uk/pub/C.elegans_sequences/CHROMOSOMES/CURRENT_RELEASE/, Mai 2003.

³ The *C. elegans* ORFeome cloning project, Reboul, J., *et al* 2001.

formatiert. Im Genom von *C. elegans* wurden neun Serpin-Gene annotiert¹. Diese Gene wurden zunächst lokalisiert und ihre Exon-Intron-Strukturen auf Vollständigkeit überprüft. Aufgrund der eigenen durchgeführten Untersuchungen konnten drei der neun Serpin-Genen (*Ce-Srp-7*, *Ce-Srp-8* und *Ce-Srp-9*) in ihren Genstrukturen korrigiert bzw. ergänzt werden. Hier wurden zusätzliche Exons mit konservierten, für Serpine charakteristischen Aminosäuresequenzen identifiziert. Die durchgeführten Korrekturen der Exon-Intron-Strukturen (siehe unten) sind bereits in Tabelle 19 enthalten. Die aktualisierten Genstrukturen sind in Abbildung 12 dargestellt und die abgeleiteten Aminosäuresequenzen befinden sich im Anhang (Abschnitt 8.5.1).

Ce-Srp-7

Eine eigene Überprüfung (Abschnitt 4.6.1) des WormBase-Genmodells *Ce-Srp-7* führte zur Identifikation von zwei weiteren Varianten des RCL-kodierenden Exons. Somit verfügt *Ce-Srp-7* über insgesamt drei RCL-kodierende Exon-Varianten, die mittels eines alternativen Spleißprozesses selektiv an ein gemeinsames Serpingrundgerüst angefügt werden könnten (siehe Abbildung 36).

Ce-Srp-8

Eine Überprüfung (Abschnitt 4.6.1) des WormBase-Genmodells *Ce-Srp-8* führte unter Berücksichtigung eines zusätzlich identifizierten Exons (kodiert eine für Serpine charakteristische Aminosäuresequenz) zur Erweiterung des Genmodells in 3'-Richtung. Aufgrund der Korrektur konnten die vorher fehlenden Sekundärstruktur-Elemente s4A, s1C, s4B und s5B ergänzt werden (siehe Abbildung 45).

Ce-Srp-9

Eine Überprüfung (Abschnitt 4.6.1) des WormBase-Genmodells *Ce-Srp-9* führte unter Berücksichtigung eines zusätzlich identifizierten Introns (enthält 5'- und 3'-Spleißsignale) zur Korrektur des Genmodells. Das zusätzliche, 48 Bp große Intron an Position 339c verkürzt den vorher ungewöhnlich langen RCL um 16 As. Darüber hinaus führte die Überprüfung des Genmodells zur Identifikation eines zusätzlichen, RCL-kodierenden Exons. Vermutlich können die insgesamt zwei RCL-kodierenden Exon-Varianten mittels eines alternativen Spleißprozesses selektiv an das gemeinsame Serpingrundgerüst angefügt werden (siehe Abbildung 36 und Diskussion 6.2).

¹ WormBase, <http://www.wormbase.org>, release WS99, Mai 2003

5.1.3 Chromosomale Lokalisation der Serpin-Gene von *C. elegans*

Das haploide Genom von *C. elegans* besteht aus fünf autosomalen Chromosomen (I bis V) und einem Geschlechtschromosom (X). Die autosomalen Chromosomen lassen sich in drei Bereiche gliedern. Im Zentrum befindet sich eine „zentrale Cluster-Region“, die durch eine hohe Gendichte und durch Abwesenheit von repetitiven Elementen charakterisiert wird. An diese Cluster-Region schließen sich die linke und rechte Armregionen an. Die Armregionen der Chromosomen zeichnen sich durch eine geringere Gendichte und durch häufigeres Vorkommen repetitiver Elemente aus.

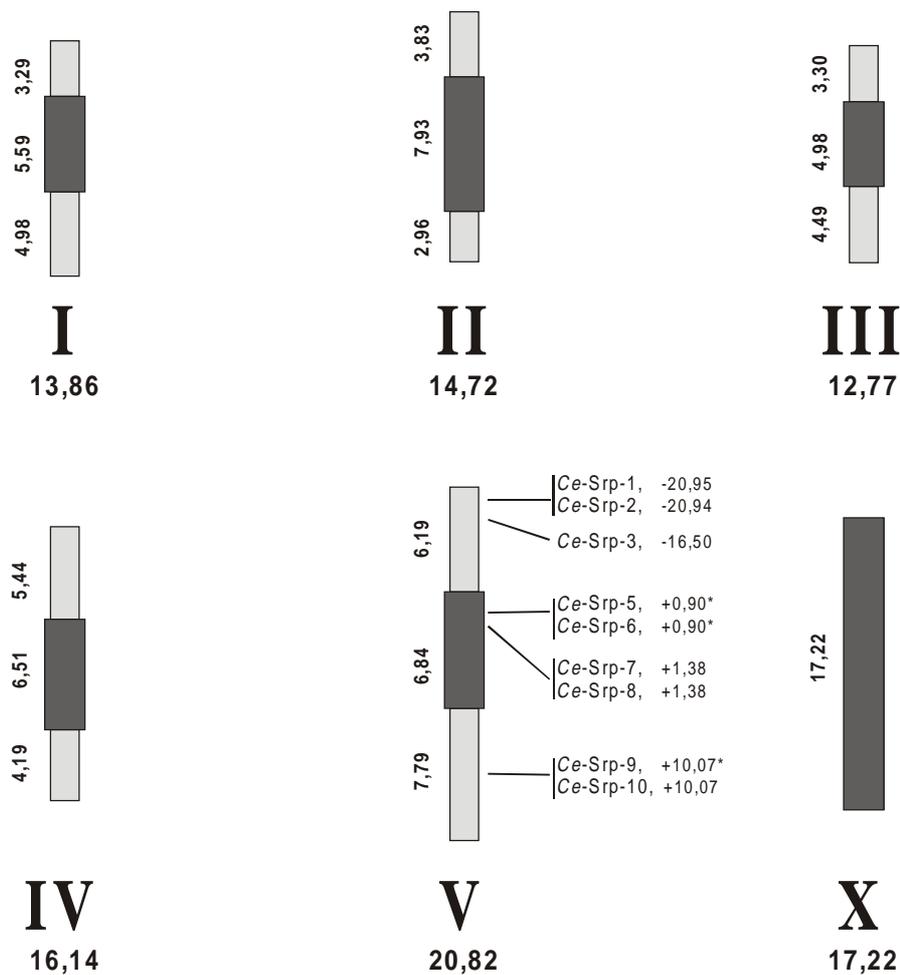


Abbildung 11: Chromosomale Lokalisation der Serpin-Gene von *C. elegans*

Die Abbildung basiert auf Angaben des *C. elegans* Sequencing Consortiums¹. Die zentrale Cluster-Regionen der Chromosomen sind schwarz markiert. Die Armregionen sind grau markiert. Die linke Armregion zeigt jeweils nach oben. Chromosom X weist keine Armregionen auf. Die Größen der Chromosomen-Bereiche und die Gesamtgröße der Chromosomen sind in Millionen Basenpaaren (MBp) angegeben. Auf dem Minus-Strang lokalisierte Serpin-Gene sind mit einem Stern markiert.

¹ WormBase, <http://www.wormbase.org>, release WS99, Mai 2003

In Abbildung 11 ist die Lage der Serpin-Gene auf den Chromosomen von *C. elegans* dargestellt¹. Alle neun Serpin-Gene sind auf Chromosom V lokalisiert. Bis auf *Ce-Srp-3* sind die Serpin-Gene in Genclustern (*Ce-Srp-1* mit *Ce-Srp-2*, *Ce-Srp-5* mit *Ce-Srp-6*, *Ce-Srp-7* mit *Ce-Srp-8* und *Ce-Srp-9* mit *Ce-Srp-10*) organisiert. Die Positionsangabe (in cMorgan, cM) ist relativ zur jeweiligen Chromosomenmitte angegeben.

¹ The *C. elegans* Sequencing Consortium, 1998.

5.1.4 Genstrukturen der Serpin-Gene von *Caenorhabditis elegans*

In Abbildung 12 sind die Genstrukturen der Serpine von *C. elegans* dargestellt. Obwohl von einigen Serpinen cDNA-Daten existieren (Tabelle 19, siehe Spalte „WORFDB“¹) sind bisher noch keine 5'- und 3'-UTR-Sequenzen zugänglich. Die Intronphasen (a, b, c) und die Intronpositionen sind der Tabelle 20 zu entnehmen. Die von den Genstrukturen abgeleiteten Aminosäuresequenzen befinden sich im Anhang (Abschnitt 8.5.1).

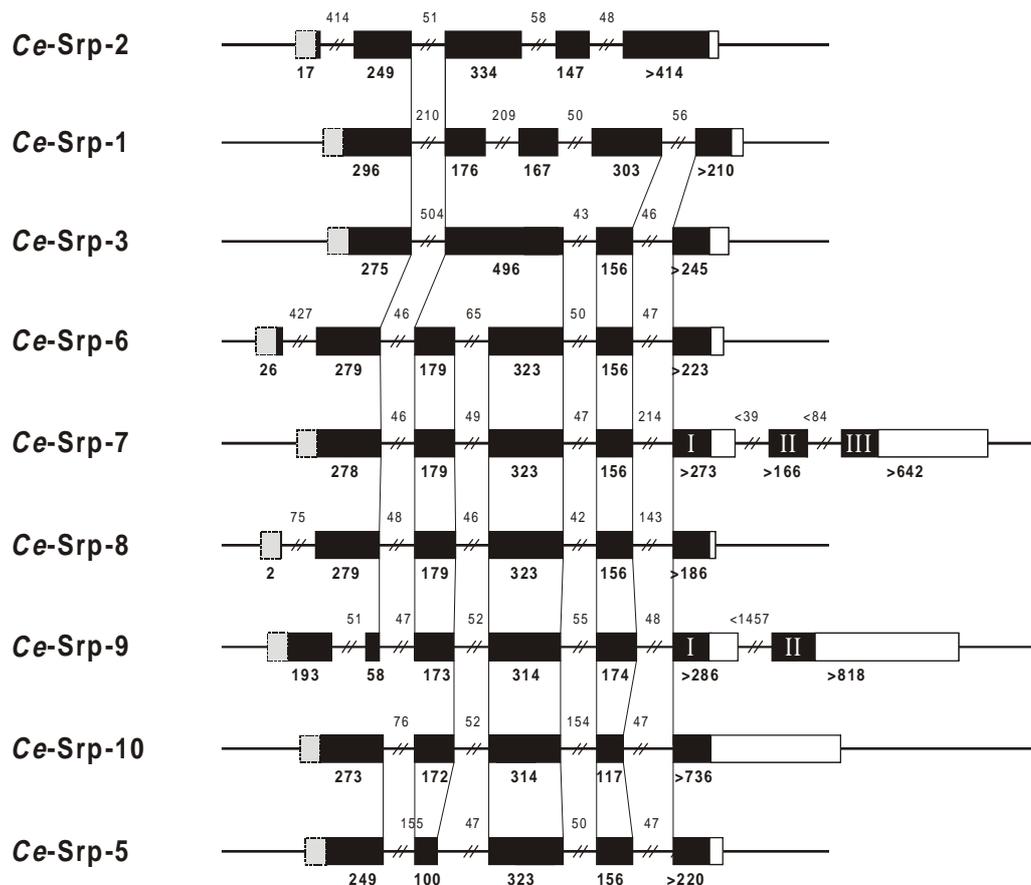


Abbildung 12: Genstrukturen der Serpin-Gene von *Caenorhabditis elegans*

Protein-kodierende Exons sind als schwarze Kästchen dargestellt. Die nicht bekannten 5'-UTRs sind durch graue Kästchen dargestellt. Die 3'-UTRs (weiße Kästchen) umfassen die 3'-untranslatierte Region einschließlich des Poly(A)-Signals (AATAAA oder ATTAAA). Die Exon-Größenangaben (Bp) stehen unterhalb und die Intron-Größenangabe oberhalb der Genstruktur. Das jeweils 3'-terminale Exon endet mit dem Poly(A)-Signal und ist dementsprechend ca. 10 - 40 Bp größer als angegeben. Die Serpin-Gene *Ce-Srp-7* und *Ce-Srp-9* verfügen über multiple Kopien des RCL-kodierenden Exons (Abschnitt 5.6.2). Diese RCL-Exon-Varianten sind mit römischen Ziffern markiert. Die gemeinsamen Intronpositionen bzw. 5'- und 3'-Spleißstellen sind durch Linien verbunden.

¹ The *C. elegans* ORFeome cloning project, Reboul, J., et al 2001.

5.1.5 Intronpositionen der Serpin-Gene von *Caenorhabditis elegans*

Die Intronpositionen und Intronphasen (a, b, c) der Serpin-Gene von *Caenorhabditis elegans* wurden mit Hilfe des Programms Genewise¹ bestimmt. Auf der Basis dieser genspezifischen Intronpositionen wurden anschließend mit Hilfe eines Aminosäuresequenz-Alignments die relativen Intronpositionen in Bezug auf das Referenz-Serpin reifes, humanes α_1 -Antitrypsin (hum-a1AT) ermittelt. Um die Intronpositionen gesichert bestimmen zu können, wurden die Alignments mit zwei verschiedenen Alignment-Programmen (DIALIGN 2.2.1, ClustalX 1.83, siehe Abschnitt 4.7.1) erstellt. In Tabelle 20 sind nur die Intronpositionen innerhalb der konservierten Serpinstruktur (ab Position 52 in Bezug auf hum-a1AT) aufgeführt.

Genname (CGC)	Chromosomale Lokalisation (Release WS85) (Chromosom / cM / Start (Bp) / *Orientierung)	Intronpositionen (in Bezug auf reifes α_1 -Antitrypsin)								
		82a/85a	108c/111	121b	179a/180	217c	236c	266c/265	285c	339c
Ce-Srp-1	V / -20.95 / 753193			+	+		+			+
Ce-Srp-2	V / -20.94 / 763115			+		+		+		
Ce-Srp-3	V / -16.50 / 1908724			+					+	+
Ce-Srp-5	V / +0.90 / 7384623*		+		+				+	+
Ce-Srp-6	V / +0.90 / 7386844*			+	+				+	+
Ce-Srp-7	V / +1.38 / 8184024			+	+				+	+
Ce-Srp-8	V / +1.38 / 8186916			+	+				+	+
Ce-Srp-9	V / +10.07 / 16896820*	+		+	+				+	+
Ce-Srp-10	V / +10.07 / 16898946		+		+				+	+

Tabelle 20: Intronpositionen der Serpin-Gene von *Caenorhabditis elegans*

Intronpositionen sind grau markiert, wenn ihre Position aufgrund von Indels nicht eindeutig bestimmbar ist. Sind für ein Intron zwei Positionen aufgeführt, so steht die erste Zahl für die Position im DIALIGN-Alignment und die zweite Zahl für die Position im ClustalX-Alignment. In der Spalte „Chromosomale Lokalisation“ ist die Lage der Serpin-Gene in cM angegeben. Serpin-Gene, die auf dem Minus-Strang lokalisiert sind, werden mit einem Stern markiert.

¹ Genewise, <http://www.ebi.ac.uk/Wise2/>

Folgende Intronpositionen der Serpine von *C. elegans* stimmen mit Intronpositionen von Serpin-Genen anderer Organismen überein, Beispiele sind in Klammern angegeben:

- Ce-Srp-2:** 217c (*Bm-Spn-2*, *Brugia malayi*)¹,
Ce-Srp-3-10: 285c (*Ci-Spn-5*; *Bm-Spn-2*),
Ce-Srp-1, 3-10: 339c (*Ci-Spn-5*, *Bm-Spn-2*, humanes ATIII)².

5.1.6 Phylogenetische Analyse der Serpine von *Caenorhabditis elegans*

Die korrigierten Aminosäuresequenzen der Serpine von *C. elegans* wurden mit DIALIGN aligniert. Das Alignment wurde mit GeneDoc überprüft und optimiert. Die hoch divergente RCL-Region wurde strikt linksbündig ausgerichtet. Zur Darstellung des Alignments wurde ESPript 2.0 verwendet (siehe Anhang, Abbildung 45). Alle Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (hum-a1AT).

Die Serpine *Ce-Srp-5*, *Ce-Srp-9* und *Ce-Srp-10* weisen zum Teil größere Deletionen auf. Im Falle des Serpins *Ce-Srp-9* wird die Deletion (Positionen 86-102) von 17 As durch das 51 Bp große 82/85a-Intron verursacht. Interessanterweise kodiert die Intron-Sequenz eine stark konservierte, für Serpine charakteristische Aminosäuresequenz („SGSTDSQM*ISDLCEE“, vergleiche Abbildung 45), die jedoch durch ein Stopp-Codon („*“) unterbrochen wird. Auf diesen Befund wird im Diskussionsteil (Abschnitt 6.2.2) eingegangen.

Das Serpin *Ce-Srp-5* weist eine 34 As große Deletion (Position 141-174) auf, die hier aber nicht auf ein Intron zurückzuführen ist. Durch die Deletion fehlen die zwei wichtigen Sekundärstruktur-Elementen s1A (β -Faltblattstrang) und hF (α -Helix).

Das Serpin *Ce-Srp-10* weist zwei größere Deletionen auf. Die erste Deletion (Position 112-121) grenzt an das 76 Bp großes 108c/111c-Intron an. Dieses Intron enthält eine stark konservierte, für Serpine charakteristische Aminosäuresequenz („EVNVANRLFIKY“, vergleiche Abbildung 45). Diese Deletion scheint zum Verlust des Sekundärstruktur-Elements s2A zu führen. Die zweite Deletion (Position 286-304) grenzt an das 154 Bp große 285c-Intron. Auch dieses Intron enthält eine stark konservierte, für Serpine charakteristische Aminosäuresequenz („IAIPKWSIETDLDIS*TLQ“), die jedoch ein Stopp-Codon (symbolisiert durch ein Stern, „*“) enthält. PCR-Experimente (WORFDB³), die die

¹ Zang, X., Yazdanbakhsh, M., Jiang, H., Kanost, M., R., Maizels, R., M., 1999.

² Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

³ The *C. elegans* ORFeome cloning project, Reboul, J., et al 2001.

Transkription der Serpin-Gene *Ce-Srp-5* und *Ce-Srp-10* nachweisen sollten, schlugen bisher fehl (siehe Diskussionsteil 6.2).

Der Aufbau des aminoterminalen Teils des RCLs erlaubt eine Abschätzung der Protease-Inhibierungsaktivität eines Serpins (Abschnitt 2.1.4). Für den β -Faltblatt-Extensions-Mechanismus der Serpine sind die Positionen P15-P8 besonders wichtig.

Fast alle Serpine von *C. elegans* weisen im Bereich von Position P15-P8 eine für inhibitorische Serpine charakteristische Aminosäuresequenz auf, die sich durch kleine Aminosäurereste auszeichnet, um eine Insertion des β -Faltblattstranges s4A (Teil des RCL) ins β -Faltblatt sA zu ermöglichen. Das Serpin *Ce-Srp-8* bildet hier die Ausnahme. An Position P12 befindet sich ein Prolin (P), das auf eine zumindest eingeschränkte Protease-Inhibierungsaktivität schließen läßt.

Das Aminosäuresequenz-Alignment der Serpine von *C. elegans* (Abbildung 45) diente zur Konstruktion eines Intraspezies-Stammbaums (Abbildung 13), um mit dessen Hilfe die Verwandtschaftsverhältnisse der Serpine von *C. elegans* untereinander aufzuklären. Hierbei wurde die so genannte Innengruppe von den *C. elegans*-Serpinen gebildet. Als Außengruppe eignet sich humanes α_1 -Antitrypsin, das bereits für das Alignment als Referenz-Serpin diente. Als Baum-Rekonstruktionsmethode wurde das Maximum-Likelihood-Verfahren verwendet (Abschnitt 4.7.2). Die ML-Analyse wurde mit Tree Puzzle 5.0¹ durchgeführt. Anstatt des klassischen, externen Datensatz-*Bootstraps* wird der methoden-interne „*Bootstrap*“-Wert, der *reliability value*, als Unterstützungswert für eine OTU-Gruppierung angegeben (Abschnitt 4.7.5).

Der analysierte Datensatz weist ca. 1 % ungelöste Quartette auf (*unresolved quartets*), so dass keine *Multifurcation* (*furcation*: Verzweigung) in der Topologie auftritt. Im dargestellten ML-Baum (Abbildung 13) treten zwei deutliche Cluster auf. Die Serpine *Ce-Srp-9A* und *Ce-Srp-10* und die Serpine *Ce-Srp-5* und *Ce-Srp-6* bilden jeweils einen Cluster, die beide mit einem maximalen Unterstützungswert von 100 bewertet werden. In einem hier nicht gezeigten Neighbor-Joining-Baum wird zusätzlich ein Cluster durch die Serpine *Ce-Srp-7A* und *Ce-Srp-8* gebildet (Abschnitt 6.2). Wird bei der Konstruktion des ML-Baums mit Tree Puzzle die Aminosäure-Substitutionsmatrix (VT, PAM) variiert (Ergebnisse nicht dargestellt), so erhält man identische bzw. sehr ähnliche Topologien, die jedoch leichte Abweichungen der Unterstützungswerte (*reliability values*) zeigen.

¹ Strimmer, K., von Haeseler, A., 1996.

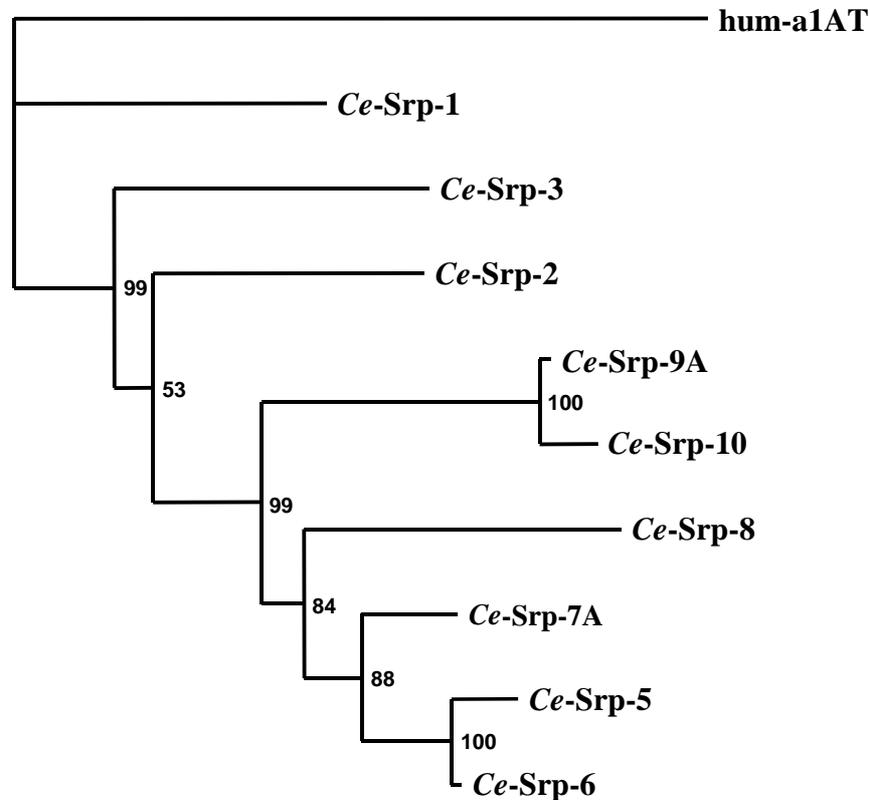


Abbildung 13: Maximum-Likelihood-Baum der Serpine von *C. elegans*

Das Phylogramm wurde mit Tree Puzzle 5.0 erstellt und basiert auf der JTT-Substitutionsmatrix (Jones-Taylor-Thornton)¹. Abweichend von den Voreinstellungen wurde eine exakte Parameterschätzung (*quartet sampling + NJ tree*) gewählt. Der Baum wurde mit humanem α_1 -Antitrypsin (hum-a1AT) gewurzelt. Die Unterstützungswerte (*reliability values, 1000 puzzle steps*) sind jeweils rechts vom zugehörigen Knoten angegeben. Zur Darstellung des Baums wurde das Programm TreeView² verwendet.

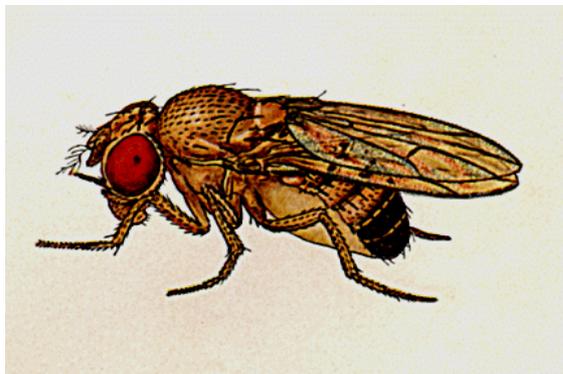
¹ Jones, D., T., *et al.* 1992

² Page, R., D., M., 1996.

5.2 Analyse der Serpin-Gene von *Drosophila melanogaster*

5.2.1 Modellorganismus *Drosophila melanogaster*

Drosophila melanogaster (*Drosophila*, Fruchtfliege) ist ein ca. 2 - 3 mm langes, zweiflügeliges Insekt (Diptera, Abbildung 14). Phylogenetisch wird die Fruchtfliege *Drosophila melanogaster* dem Stamm der Arthropoda zugeordnet, der vermutlich wie die Nematoda vor ca. 500 bis 1200 Millionen Jahren vom Stamm der Deuterostomia abzweigte¹. Die Fruchtfliege ernährt sich von verrottenden organischen Materialien (Obst). Aufgrund der leichten Kultivierbarkeit und eines relativ kurzen Lebenszyklus (ca. 2 Wochen) wird *Drosophila melanogaster* seit 1907 als Versuchsobjekt² bzw. Modellorganismus der Genetik verwendet. Das diploide Genom besteht aus vier Chromosomenpaaren (den drei autosomalen Chromosomenpaaren 2, 3, 4 und einem Paar Geschlechtschromosomen X, Y). Man kennt bei *Drosophila melanogaster* zwei Geschlechter, Männchen (XY) und Weibchen (XX). Die Sequenz des *Drosophila*-Genoms wurde im Jahr 2000 veröffentlicht³. Das Genom umfasst ca. 180 Millionen Basenpaare (MBp), wobei ca. 60 MBp aus schwer sequenzierbarem



Heterochromatin bestehen, das sich um die Centromer-Struktur der Chromosomen anordnet. Die sequenzierte, 120 MBp große euchromatische Sequenz kodiert ca. 13600 Gene. Derzeit wird das *Drosophila*-Genom durch die FlyBase-Datenbank⁴ betreut und aktualisiert.

Abbildung 14: *Drosophila melanogaster*

(Quelle: Patterson, J., T., 1943.)

¹ Tree of Life - Web Project, <http://tolweb.org/tree/phylogeny.html>

² Morgan, T., H., 1919.

³ Adams, M., D., *et al.* 2000.

⁴ FlyBase, A Database of the *Drosophila* Genome, <http://www.flybase.org/>, The FlyBase Consortium, 2003.

5.2.2 Die Serpin-Gene von *Drosophila melanogaster*

Als Basis der vorliegenden Untersuchungen diente die FlyBase-Datenbank. Die Genomsequenz wurde über das „*Berkeley Drosophila Genome Project*“ (BDGP)¹ bezogen. Im *Drosophila*-Genom wurden ursprünglich 29 Serpin-Gene annotiert². Zur Überprüfung der Serpin-Gene wurden zunächst die chromosomalen DNA-Sequenzen, wie im Abschnitt 4.6.1 beschrieben, in die sechs Aminosäureraster translatiert und für eine AGREP-Suche formatiert. Anschließend wurden die Serpin-Gene lokalisiert (Abschnitt 4.6.1) und ihre Exon-Intron-Strukturen auf Vollständigkeit überprüft. In Tabelle 21 sind die durch eigene Datenbank-Recherchen ergänzten Serpin-Gene von *Drosophila melanogaster* aufgeführt.

Genmodell (FlyBase)	Synonyme	TrEMBL- Eintrag	Protein- Größe (As)	cDNA- Größe (Nt)	Gen- Größe (Bp)	Exp. Daten
<i>Dm-Spn27a</i>	CG11331	Q9V3N1	447	1525	1525	cDNA
<i>Dm-CG6717</i>		Q9VLZ8	397	>1279	>1384	-
<i>Dm-CG31902</i>	CG12318 (b)	Q9VLV3	384	>1236	>1401	-
<i>Dm-CG33121</i>	CG12318 (a)	Q9VLV3	377	>1188	>1376	-
<i>Dm-CG7219</i>		Q9VLU4	536	>1681	>2552	-
<i>Dm-Spn2</i>	CG8137, Sp2	Q9U1I7	375	1547	1721	cDNA
<i>Dm-CG4804</i>		Q9VL44	382	>1330	>3012	5'-EST
<i>Dm-Spn3</i>	CG9334, Sp3	Q9U1I6	372	1405	1572	cDNA
<i>Dm-CG14470</i>		Q9V9I8	1976	>6507	>6507	3'-EST
<i>Dm-Spn4</i>	CG9453, Sp4	Q9V991, Q9U1I5	424	(Abschnitt 5.6.1)	3018	cDNA
<i>Dm-CG9454</i>		Q9V990	388	>1179	>1410	-
<i>Dm-CG9455</i>		Q9V989	403	1557	1696	cDNA
<i>Dm-Spn1</i>	CG9456, Sp1	Q9U1I8	372	1472	1871	cDNA
<i>Dm-CG9460</i>		Q9V987	386	1303	1862	cDNA
<i>Dm-Spn43Aa</i>	CG12172	Q9V3R7	390	1251	1555	cDNA
<i>Dm-Spn43Ab</i>	CG1865	Q9V4I3, Q9U5W8	393	1373	2025	cDNA
<i>Dm-CG1859</i>	Spn43Ad	Q9V4I4	407	1431	1686	cDNA

¹ <http://www.fruitfly.org/sequence/release3genomic.shtml>, 2003.

² FlyBase, <http://www.flybase.org/>, Release 3.1, Februar 2003

<i>Dm-Nec</i>	CG1857, Spn43Ac	Q9V4I5, Q9U5W7	476	1621	1753	cDNA
<i>Dm-CG7722</i>		Q9V5S3	382	1356	1482	cDNA
<i>Dm-CG10956</i>		Q9V7Y9	379	>1250	>1311	-
<i>Dm-Spn6</i>	CG10913, Sp6	Q9V3L3	374	1348	1915	cDNA
<i>Dm-Acp76A</i>	CG3801	Q9V VW1, O46224	388	>1193	>1254	-
<i>Dm-CG6680</i>		Q9VWB3	442	>1369	>1540	cDNA
<i>Dm-CG6663</i>	CG6266	Q9VWB5	362	>1198	>1389	5'-EST
<i>Dm-CG6289</i>		Q9VPH9	416	>1389	>1580	5'-EST
<i>Dm-Spn5</i>	CG18525, Sp5	Q9VFC2, Q9U1I4	427	1507	3365	cDNA
<i>Dm-CG6687</i>		Q9VFC1	426	1438	2067	cDNA
<i>Dm-CG12807</i>		Q9VH46	555	>2046	>2619	-
<i>Dm-CG1342</i>		Q9VA48	649	2036	2125	cDNA

Tabelle 21: Die Serpin-Gene von *Drosophila melanogaster*

Die Namen der Genmodelle basieren auf der Nomenklatur von FlyBase/Celera Genomics (CG). Um in Interspezies-Stammbäumen eine klare Identifikation der Serpine zu ermöglichen, werden die FlyBase-Namen mit dem Präfix *Dm* (für *Drosophila melanogaster*) erweitert. Die Reihenfolge der Serpine orientiert sich an der chromosomalen Lokalisation (2L, 2R, 3L, 3R, siehe auch Tabelle 22). In der Spalte „Synonyme“ sind die Serpin-Namen aufgeführt, die aus experimentellen Arbeiten stammen, oder es handelt sich um die Nomenklatur älterer Genmodelle. Die TrEMBL-Einträge weisen teilweise auf die Doppelannotationen hin. Unter „Protein-Größe“ ist die Größe des Präproteins in Aminosäuren (As) angegeben. Die cDNA- und Gen-Größen sind in Nukleotiden (Nt) bzw. in Basenpaaren (Bp) angegeben. In den Fällen, in denen keine experimentelle Daten zur Verfügung stehen, umfassen die Größenangaben die DNA-Sequenz zwischen Startcodon und Poly(A)-Signal. Das Serpin-Gen *Spn4* (*Dm-Spn4*) verfügt über multiple RCL-kodierende Exons, die alternatives Spleißen ermöglichen (Abschnitt 5.6.1).

Aufgrund der eigenen Untersuchungen (Abschnitt 4.6.1) konnten fünf der oben aufgeführten 29 Serpin-Gene (*Dm-CG31902*, *Dm-CG33121*, *Dm-Spn4*, *Dm-CG9454*, *Dm-CG10956*) in ihren Genstrukturen erweitert bzw. korrigiert werden. Hier wurden zusätzliche Exons identifiziert, die für Serpine charakteristische Aminosäuresequenzen kodieren. Die Korrekturen der Exon-Intron-Strukturen (siehe unten) sind in Tabelle 21 bereits berücksichtigt.

Die aktualisierten Genstrukturen sind in Abbildung 16 - Abbildung 19 dargestellt und die abgeleiteten Aminosäuresequenzen sind im Abschnitt 8.5.1 aufgeführt. Angaben zu Intronposition oder Aminosäure-Indels beziehen sich auf die Nummerierung des Referenz-Serpins reifes, humanes α_1 -Antitrypsin (hum-a1AT).

***Dm*-CG31902 und *Dm*-CG33121**

Die Serpin-Gene *Dm*-CG31902 und *Dm*-CG33121 bilden einen Gencluster (Abbildung 15). In der ersten veröffentlichten Version des *Drosophila*-Genoms, wurde ein Genmodell formuliert, in dem beide Sequenzen zu einem ungewöhnlich großen Serpin (Genmodell *Dm*-CG12318) verbunden wurden. Im aktuellen Release 3 des Genoms¹ wurde das Genmodell berichtigt und die beiden Serpin-Gene *Dm*-CG31902 und *Dm*-CG33121 getrennt annotiert.

Eine eigene Überprüfung (Abschnitt 4.6.1) der beiden FlyBase-Genmodelle *Dm*-CG31902 und *Dm*-CG33121 führte unter Berücksichtigung von zusätzlich identifizierten Exons (kodieren jeweils eine für Serpine charakteristische Aminosäuresequenz) zu weiteren Korrekturen der Genmodelle:

Das Serpin *Dm*-CG33121 konnte in 5'- und 3'-Richtung erweitert werden. Durch Zuordnung von zwei zusätzlichen Introns wurde die abgeleitete Proteinsequenz sowohl amino- wie carboxyterminal erweitert (Abbildung 16). Das erste Intron ist 69 Bp groß und im mutmaßlichem Signalpeptid lokalisiert, während das zweite 66 Bp große Intron an Position 358a lokalisiert ist. Die beiden Introns verfügen jeweils über kanonische 5'- und 3'-Spleißsignale. Bei der Intronposition 358a handelt es sich vermutlich um eine Variante von 352a (Tabelle 22). Aufgrund dieser Korrekturen verfügt die abgeleitete Proteinsequenz über ein potentielles Signalpeptid (PSORT II², TargetP v1.1³) und die vorher fehlenden Sekundärstruktur-Elemente s1C, s4B und s5B wurden ergänzt.

Das Serpin *Dm*-CG31902 konnte in 5'- und 3'-Richtung erweitert werden. Durch Setzen von drei Introns wurde die abgeleitete Proteinsequenz aminoterminal erweitert (Abbildung 16). Das erste 56 Bp große Intron ist im mutmaßlichem Signalpeptid lokalisiert. Das zweite Intron (an Position 305c) ist 54 Bp groß. Das dritte 57 Bp große Intron ist an Position 360a lokalisiert. Alle Introns verfügen über 5'- und 3'-Spleißsignale. Die Intronpositionen 305c und 360a (als Variante von 352a) treten auch bei anderen Serpin-Genen von *Drosophila* auf

¹ FlyBase, <http://www.flybase.org/>, <http://www.wormbase.org/>, Release 3.1, Februar 2003

² Nakai, K., Kanehisa, M., 1992.

³ Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000.

(Tabelle 22). Aufgrund dieser Korrekturen verfügt die abgeleitete Proteinsequenz über ein potentiell Signalpeptid (PSORT II, TargetP v1.1). und durch Setzen der beiden C-terminalen Introns (Positionen 305c und 360a) konnten zwei ungewöhnliche Insertionen (18 As und 19 As) entfernt werden.

Dm-Spn4

Eine eigene Überprüfung (Abschnitt 4.6.1) des FlyBase-Genmodells *Dm-Spn4* (*Dm-Sp4*) führte zur Identifikation von vier potentiellen RCL-kodierenden Exon-Varianten. Die Vermutung, dass die Exon-Varianten mittels eines alternativen Spleißprozesses selektiv an das gemeinsame Serpingrundgerüst angefügt werden können, wurde im Rahmen dieser Arbeit experimentell verifiziert. Für das Gen *Dm-Spn4* konnten mindestens zehn Transkripte (mRNAs) nachgewiesen werden, die acht unterschiedliche Protein-Varianten kodieren. Zusätzlich konnten alternative Transkriptionsstartpunkte identifiziert werden (vergleiche Abschnitt 5.6.1 und Abbildung 35).

Dm-CG9454

Eine Überprüfung (Abschnitt 4.6.1) des FlyBase-Genmodells *Dm-CG9454* führte unter Berücksichtigung eines zusätzlich identifizierten Exons (kodiert eine für Serpine charakteristische Aminosäuresequenz) zur Erweiterung des Genmodells in 3'-Richtung.

Es wurde ein zusätzliches 173 Bp großes Intron an Position 352a postuliert, das über kanonische 5'- und 3'-Spleißsignale verfügt. Die Intronposition 352a tritt auch bei anderen Serpin-Genen von *Drosophila* auf. Aufgrund der Korrektur wurde die abgeleitete Proteinsequenz carboxyterminal erweitert, wodurch die vorher fehlenden Sekundärstrukturelemente s1C, s4B und s5B ergänzt wurden (Anhang, Abbildung 46).

Dm-CG10956

Eine Überprüfung (Abschnitt 4.6.1) des FlyBase-Genmodells *Dm-CG10956* führte unter Berücksichtigung eines zusätzlich identifizierten Exons (kodiert eine für Serpine charakteristische Aminosäuresequenz) zur Erweiterung des Genmodells in 3'-Richtung.

Das ursprünglich annotierte 65 Bp große 351c-Intron nutzte eine suboptimale 5'-Spleißstelle („GTTGGT“). In 3'-Richtung konnte eine (weitere) potentielle 5'-Spleißstelle lokalisiert („GTAAAT“) werden, die nur eine anstatt zwei Abweichungen von der Konsensussequenz einer 5'-Spleißstelle aufweist (in DNA-Notation „GTRAGT“, mit R= A oder G). Durch Verwenden der optimaleren 5'-Spleißstelle erhält man ein 61 Bp großes Intron an Position

353a. Aufgrund der Korrektur wurde die abgeleitete Proteinsequenz um die vorher fehlenden Sekundärstruktur-Elemente s1C, s4B und s5B ergänzt (Abbildung 46).

Dm-CG12807* und *Dm-CG14470

Obwohl die Genmodelle *Dm-CG12807* und *Dm-CG14470* eindeutig kurze für Serpine charakteristische Sequenzabschnitte aufweisen, zeigen sie keine strukturellen bzw. globalen Ähnlichkeiten mit einem „Standard“-Serpine. Einerseits weisen die postulierten Genprodukte große Insertionen auf, andererseits fehlen wichtige Strukturelemente (z.B. *Breach*-Region). Aufgrund der Größen (555 As bzw. 1976 As) und ungewöhnlichen Exon-Intron-Strukturen wurden diese *Drosophila*-Serpine nicht in die weiteren Untersuchungen einbezogen.

Serpine-Pseudogene von *Drosophila melanogaster*

Es konnten keine eindeutigen Serpine-Pseudogene im Genom von *Drosophila melanogaster* annotiert werden. Potentielle Kandidaten wären die oben genannten Serpine-Gene *Dm-CG12807* und *Dm-CG14470*. Die Expression von *Dm-CG14470* wird jedoch durch EST-Sequenzdaten¹ gestützt.

Ursprünglich wiesen die Serpine-Gene *Dm-CG6663* und *Dm-CG6289* jeweils Rastermutationen im Protein-kodierenden Bereich auf, die zu vorzeitigen Terminierungen führten. In der aktuellen 3. Version des *Drosophila*-Genoms wurden die Rastermutationen in beiden Serpine-Sequenzen entfernt¹. Die Serpine-Gene *Dm-CG6663* und *Dm-CG6289* werden daher vollständige Genmodelle annotiert.

¹ FlyBase, <http://www.flybase.org/>, Release 3.1, Februar 2003

5.2.3 Chromosomale Lokalisation der Serpin-Gene von *Drosophila*

Das haploide Genom von *Drosophila melanogaster* besteht aus drei autosomalen Chromosomen (2, 3, 4) und den Geschlechtschromosomen (X, Y). In Abbildung 15 ist die Lage der Serpin-Gene auf den *Drosophila*-Chromosomen 2 und 3 dargestellt.

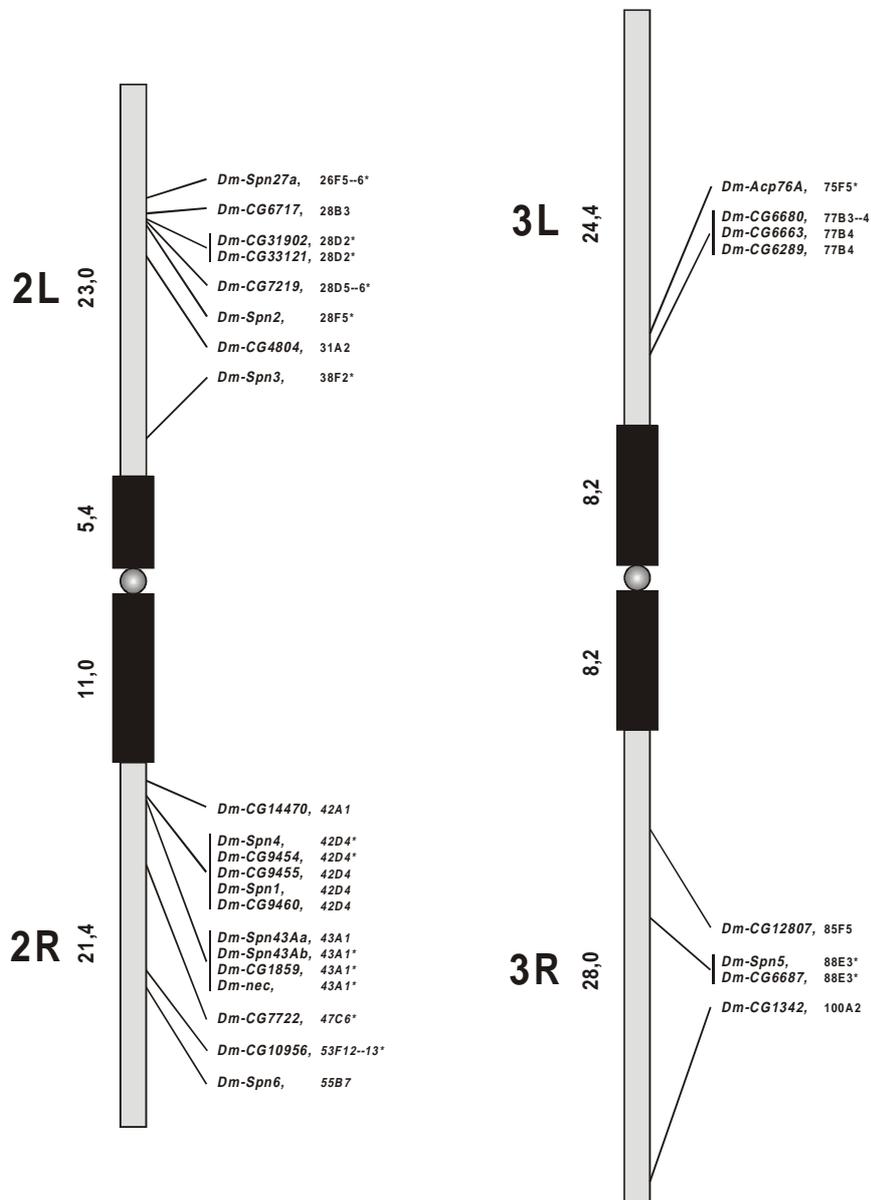


Abbildung 15: Lage der Serpin-Gene auf den *Drosophila*-Chromosomen 2 und 3

In der Abbildung¹ sind die Centromere der Chromosomen als Kreise dargestellt. Das bisher noch nicht sequenzierte Heterochromatin (flankiert das Centromer) ist schwarz markiert. Das sequenzierte Euchromatin (Armregionen 2L, 2R, 3L, 3R) ist hellgrau markiert. Die Größen der Chromosomen-Bereiche sind jeweils auf der linken Seite der Chromosomendarstellung in Millionen Basenpaaren angegeben. Auf dem Minus-Strang lokalisierte Serpin-Gene, sind mit einem Stern gekennzeichnet.

¹ basiert auf Angaben von Adams, M., D., et al 2000.

Die Chromosomen 4, X und Y enthalten keine Serpin-Gene. Im Gegensatz zu der Chromosomenstruktur von *C. elegans* zeigen die *Drosophila*-Chromosomen einen Metazoatypischen Aufbau. Die Heterochromatin-Bereiche, die das jeweilige Centromer flankieren, bestehen zum größten Teil aus kurzen repetitiven Elementen, die nur schwer oder nicht stabil klonierbar sind. Auf den Chromosomarmen 2L und 2R sind insgesamt 21 Serpin-Gene lokalisiert, während sich auf den Chromosomarmen 3L und 3R insgesamt acht Serpin-Gene befinden. Im Vergleich zu *Caenorhabditis elegans* ist bei *Drosophila* die Organisation der Serpin-Gene in größeren Genclustern (teilweise vier oder fünf Serpin-Gene) hervorzuheben. Die Lokalisations-Angaben basieren auf der cytologischen Kartierung (Banden der polytären Chromosomen, nach FlyBase, Release 3.1).

5.2.4 Genstrukturen der Serpin-Gene von *Drosophila melanogaster*

In den folgenden Abbildungen sind die Genstrukturen der Serpin-Gene von *Drosophila melanogaster* dargestellt. Soweit experimentelle Daten zur Verfügung stehen, wurden sie in den Exon-Intron-Strukturen berücksichtigt (Tabelle 21). Die Phasen (a, b, c) und die Positionen der Introns sind in Tabelle 22 angegeben.

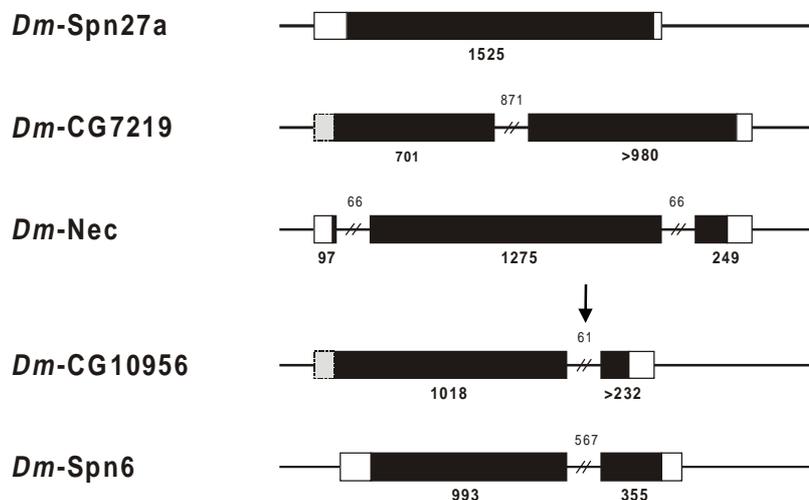


Abbildung 16: Strukturen der Serpin-Gene von *Drosophila* auf Chromosom 2 (Teil 1)

Protein-kodierende Exons sind als schwarze Kästchen dargestellt. Die Größenangaben (Bp) für Exons und Introns stehen unterhalb bzw. oberhalb der Genstrukturen. 5'-UTRs für die keine EST- oder cDNA-Daten zur Verfügung stehen, werden als graue Kästchen symbolisiert. Bei unbekanntem 3'-UTRs (weiße Kästchen) ist die Größe angegeben, die die 3'-untranslatierte Region einschließlich des Poly(A)-Signals (AATAAA oder ATTTAAA) umfasst. Die Abbildung stellt fünf Serpin-Gene von Chromosom 2 dar, die sich in ihrer Exon-Intron-Struktur von den übrigen 15 „Chromosom 2“-Serpinen unterscheiden. Der Pfeil kennzeichnet die gemeinsame Intronposition 305c von *Dm-CG10956* und *Dm-Spn6*.

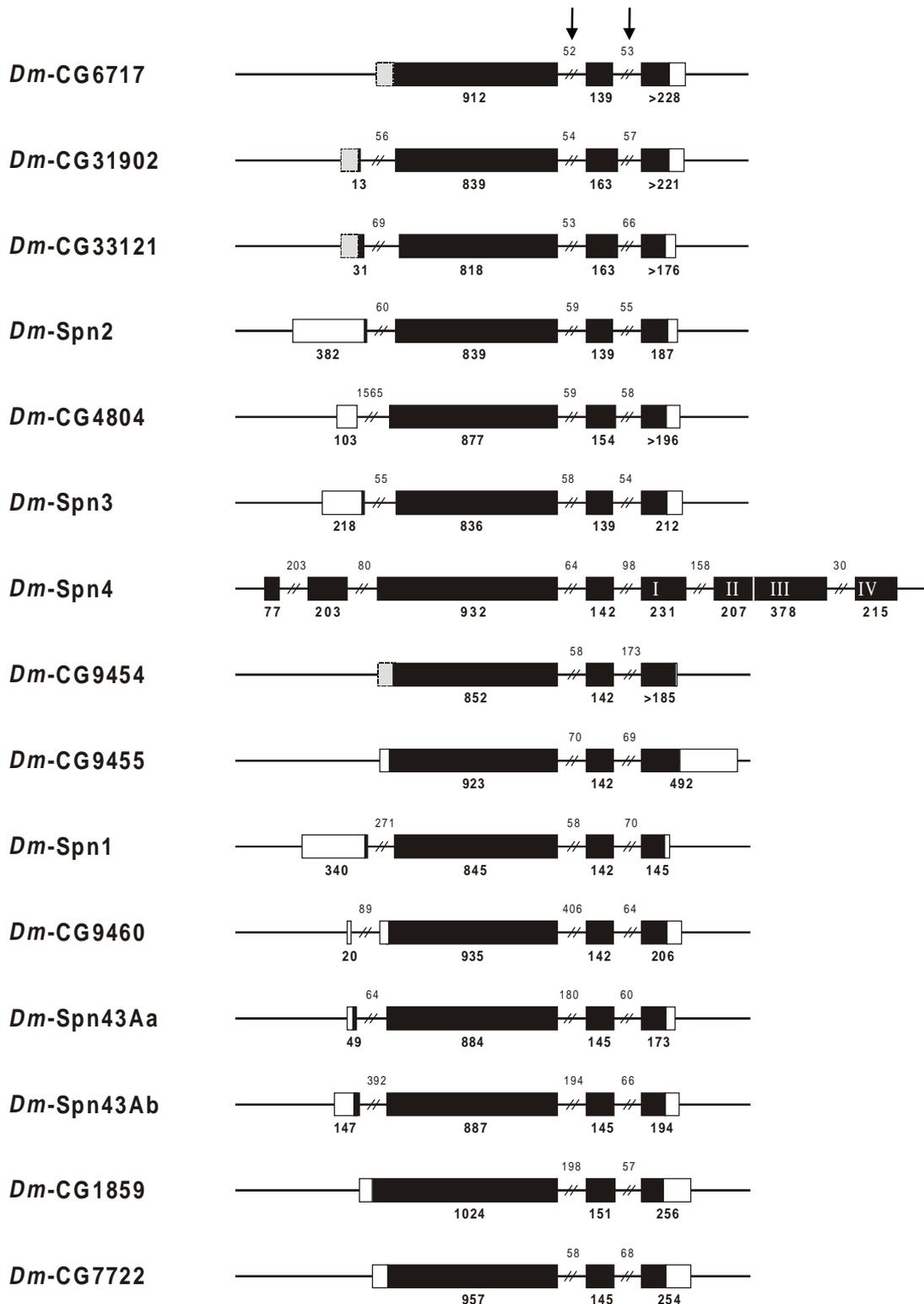


Abbildung 17: Strukturen der Serpin-Gene von *Drosophila* auf Chromosom 2 (Teil 2)

Zur Erklärung der Darstellung siehe Abbildung 16. Die 15 dargestellten Serpin-Gene sind auf Chromosom 2 lokalisiert und weisen die gemeinsamen Intronpositionen 305c und 352a (durch Pfeile markiert) auf. Das Serpin-Gen *Dm-Spn4* verfügt über multiple Kopien des RCL-kodierenden Exons (Abschnitt 5.6.1). Diese RCL-kodierenden Exon-Varianten sind mit römischen Ziffern markiert.

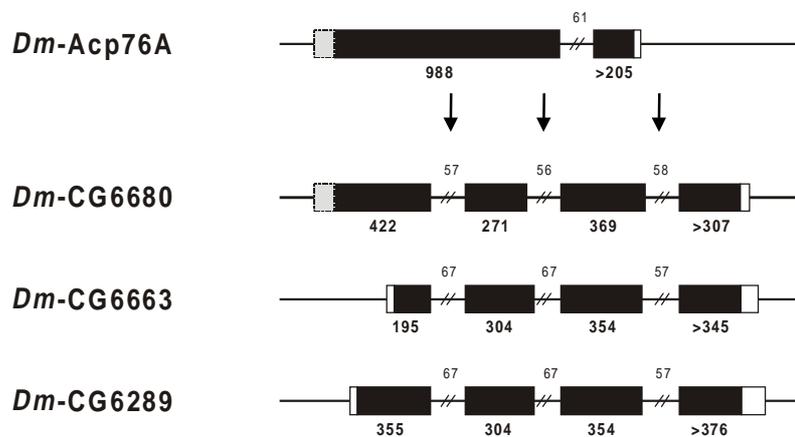


Abbildung 18: Strukturen der Serpin-Gene von *Drosophila* auf Chromosomenarm 3L

Zur Erklärung der Darstellung siehe Abbildung 16. Die vier dargestellten Serpin-Gene sind auf Chromosomenarm 3L lokalisiert. Die Serpin-Gene *Dm-CG6680*, *Dm-CG6663* und *Dm-CG6289* weisen die gemeinsamen Intronpositionen 102b, 195c und 305c (durch Pfeile markiert) auf.

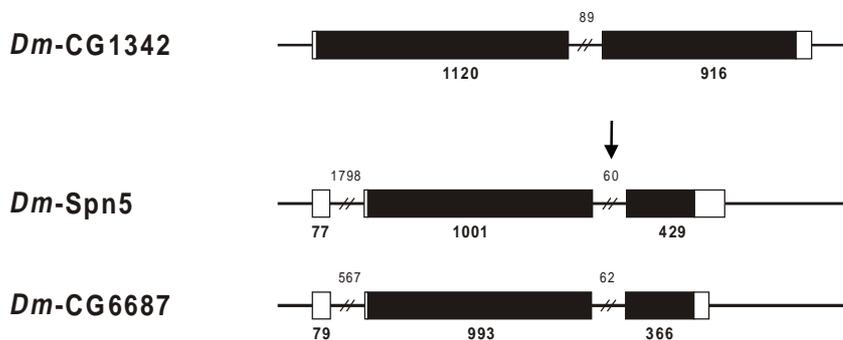


Abbildung 19: Strukturen der Serpin-Gene von *Drosophila* auf Chromosomenarm 3R

Zur Erklärung der Darstellung siehe Abbildung 16. Die drei dargestellten Serpin-Gene sind auf Chromosomenarm 3R lokalisiert. Die beiden Serpin-Gene *Dm-Spn5* und *Dm-CG6687* weisen die gemeinsamen Intronpositionen 301c (durch einen Pfeil markiert) auf.

5.2.5 Intronpositionen der Serpin-Gene von *Drosophila melanogaster*

In Tabelle 22 sind die Intronpositionen (ab Position 52 in Bezug auf reifes, humanes α_1 -Antitrypsin) der Serpin-Gene von *Drosophila melanogaster* aufgeführt.

Genmodell (FlyBase)	Chromosomale Lokalisation (Arm, Band, Startposition in Bp)	Intronpositionen (in Bezug auf reifes humanes α_1 -Antitrypsin)						
		102b	131b	-c/156c	195c	301c	305c	340a
<i>Dm-Spn27a</i>	2L, 26F5--6*, 6665560							
<i>Dm-CG6717</i>	2L, 28B3, 7574166						+	+
<i>Dm-CG31902</i>	2L, 28D2*, 7881274						+	360a
<i>Dm-CG33121</i>	2L, 28D2*, 7882761						+	358a
<i>Dm-CG7219</i>	2L, 28D5--6*, 7999363		+					
<i>Dm-Spn2</i>	2L, 28F5*, 8234366						+	+
<i>Dm-CG4804</i>	2L, 31A2, 10053306						+	357a
<i>Dm-Spn3</i>	2L, 38F2*, 20807921						+	+
<i>Dm-Spn4</i>	2R, 42D4*, 1939825						+	+
<i>Dm-CG9454</i>	2R, 42D4*, 1941617						+	+
<i>Dm-CG9455</i>	2R, 42D4, 1942618						+	+
<i>Dm-Spn1</i>	2R, 42D4, 1943905						+	+
<i>Dm-CG9460</i>	2R, 42D4, 1947307						+	+
<i>Dm-Spn43Aa</i>	2R, 43A1, 2208600						+	+
<i>Dm-Spn43Ab</i>	2R, 43A1*, 2214585						+	353a
<i>Dm-CG1859</i>	2R, 43A1*, 2216324						+	+
<i>Dm-Nec</i>	2R, 43A1*, 2219130							+
<i>Dm-CG7722</i>	2R, 47C6*, 6006395						+	+
<i>Dm-CG10956</i>	2R, 53F12--13*, 12187952							353a
<i>Dm-Spn6</i>	2R, 55B7, 13190039						+	
<i>Dm-Acp76A</i>	3L, 75F5*, 19000905						+	
<i>Dm-CG6680</i>	3L, 77B3--4, 20235766	+			+		+	
<i>Dm-CG6663</i>	3L, 77B4, 20239555	+			+		+	
<i>Dm-CG6289</i>	3L, 77B4, 20261244	+			+		+	
<i>Dm-Spn5</i>	3R, 88E3*, 11032165					+		
<i>Dm-CG6687</i>	3R, 88E3*, 11034502					+		
<i>Dm-CG1342</i>	3R, 100A2, 26505732			+				

Tabelle 22: Intronpositionen der Serpin-Gene von *Drosophila melanogaster*

Sind für ein Intron zwei Positionen aufgeführt, so steht die erste Zahl für die Position im DIALIGN-Alignment und die zweite Zahl für die Position im ClustalX-Alignment. Ein Bindestrich steht für Intronpositionen, die

aufgrund einer Insertion im Serpin (im Vergleich zum reifen human α_1 -Antitrypsin) von einem der beiden verwendeten Alignment-Programme als nicht alignierbar dargestellt werden. In der Spalte „Chromosomale Lokalisation“ wird der Chromosomenarm (2L, 2R, 3L oder 3R), die cytologische Kartierung und die absolute chromosomale Startposition in Bp angegeben¹. Auf dem Minus-Strang lokalisierte Serpin-Gene, sind mit einem Stern markiert. Die Intronposition 353a, 357a, 358a und 360a werden aufgrund der identischen Intronphase und der Alignment-Unsicherheiten im RCL als Varianten der Intronposition 352a betrachtet (siehe Diskussion Abschnitt 6.3.1).

Zur Bestimmung der Intronpositionen und Intronphasen (a, b, c) wurde das Programm Genewise verwendet. Auf der Basis dieser genspezifischen Intronpositionen wurden anschließend mit Hilfe eines Aminosäuresequenz-Alignments die relativen Intronpositionen in Bezug auf das Referenz-Serpin α_1 -Antitrypsin (reifes hum-a1AT) ermittelt. Um die Intronpositionen gesichert bestimmen zu können, wurden die Alignments mit zwei verschiedenen Alignment-Programmen (DIALIGN 2.2.1, ClustalX 1.83, siehe Abschnitt 4.7.1) erstellt.

Neben der weitverbreiteten Intronposition 352a treten vereinzelt die Intronpositionen 353a, 357a, 358a und 360a auf. Diese Intronpositionen befinden sich im hoch divergenten Bereich des RCL, der nicht eindeutig alignierbar ist. Aufgrund der identischen Intronphase (a) und der Alignment-Unsicherheiten im RCL werden diese Intronpositionen als Varianten von Position 352a betrachtet (siehe Diskussion Abschnitt 6.3.1).

Die auf Chromosom 2 lokalisierten Serpin-Gene von *Drosophila* weisen untereinander sehr ähnliche Exon-Intron-Strukturen auf. Mit nur wenigen Ausnahmen (*Dm-Spn27a*, *Dm-CG7219*, *Dm-Nec*, *Dm-CG10956* und *Dm-Spn6*) enthalten diese Serpin-Gene im alignierten Bereich zwei gemeinsame Introns (Positionen 305c und 352a, Abbildung 17).

Dagegen weisen die auf Chromosom drei lokalisierten Serpin-Gene teilweise sehr unterschiedliche Exon-Intron-Strukturen auf. Interessant ist das Auftreten von zusätzlichen Intronpositionen, während die auf Chromosom 2 weitverbreitete Intronposition 352a nicht auftritt. Damit korrelieren die Exon-Intron-Strukturdaten deutlich mit den chromosomalen Lokalisationen der Serpin-Gene. Folgende Intronpositionen der Serpin-Gene von *Drosophila* stimmen mit Intronpositionen von Serpin-Genen anderer Organismen überein, Beispiele sind in Klammern angegeben:

305c (*Ms-Alaserpin*², *Manduca sexta*),

352a (*Ms-Alaserpin*, humaner PAI-1, humanes Nexin, murines Neuroserpin)³

¹ FlyBase, <http://www.flybase.org/>, Release 3.1, Februar 2003

² Jiang, H., Wang, Y., Huang, Y., Mulnix, A., B., Kadel, J., Cole, K., Kanost, M., R., 1996.

³ Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

5.2.6 Phylogenetische Analyse der Serpin-Gene von *Drosophila melanogaster*

Die korrigierten Aminosäuresequenzen der *Drosophila*-Serpine wurden mit DIALIGN aligniert. Das Alignment wurde mit GeneDoc überprüft und optimiert. Die hoch divergente RCL-Region wurde strikt linksbündig ausgerichtet. Zur Darstellung des Alignments (siehe Anhang, Abbildung 46) wurde ESPrift 2.0 verwendet. Alle Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (hum-a1AT). Im Vergleich zu den Serpinen von *C. elegans* weisen die *Drosophila*-Serpine keine Deletionen von Sekundärstruktur-Elemente auf. Auffällig ist hier jedoch das Auftreten von Insertionen, die größtenteils zwischen den konservierten Sekundärstruktur-Elementen lokalisiert sind.

Das Serpin *Dm*-CG1342 ist mit 649 As ungewöhnlich groß. *Dm*-CG1342 verfügt über eine aminoterminal Extension von ca. 200 As und eine 21 As große Insertion zwischen den Strukturelementen hD und s2A. Der relativ lange RCL (38 As) weist auf eine geringe Protease-Inhibierungsaktivität hin. Interessant ist der Aufbau des carboxyterminalen Teils des RCL. Er enthält ungewöhnlich viele Prolin-Reste („PVPEPEPELP“), die eine vorzeitige Insertion der RCL-Struktur ins β -Faltblatt A (latente Inhibitorform) verhindern könnten.

Das Serpin *Dm*-CG7219 weist zwei Insertionen auf. Die erste Insertion, ungefähr 21 As lang, befindet sich wie im Serpin *Dm*-CG1342 zwischen den Sekundärstruktur-Elementen hD und s2A und die zweite ungefähr 27 As große Insertion ist zwischen den Sekundärstruktur-Elementen hI und s5A lokalisiert.

Die in einem Gencluster organisierten Serpine *Dm*-Spn5 und *Dm*-CG6687 enthalten beide eine 15 As große Insertion zwischen den Sekundärstruktur-Elementen s2B und s3B.

Der Aufbau des aminoterminalen Teils des RCL erlaubt eine Abschätzung der Protease-Inhibierungsaktivität eines Serpins (Abschnitt 2.1.4).

Der RCL des Serpins *Dm*-Acp76A ist ungewöhnlich kurz. Ein für Serpine charakteristisches Glycin an Position P15 ist hier durch ein voluminöseres Valin ersetzt. Für Serpine uncharakteristisch sind auch die voluminösen Aminosäureresten in den Positionen P14 - P11 („DDID“, mit D für Aspartat und I für Isoleucin) die darauf schließen lassen, dass es sich bei *Dm*-Acp76A um ein nicht-inhibitorisches Serpin handelt.

Der aminoterminal Teil des RCL (Position P17 - P4) von *Dm*-CG6663 und *Dm*-CG6289 ist interessanterweise identisch. Die Positionen P13 und P9 sind jeweils mit einem Prolin besetzt. Im Zusammenhang mit der relativ geringen Größe des RCL scheint dies darauf hinzuweisen, dass es sich bei *Dm*-CG6663 und *Dm*-CG6289 um nicht-inhibitorische Serpine handelt.

Eine Beurteilung der inhibitorischen Eigenschaften der Serpine *Dm*-CG44804, *Dm*-Spn43Ab, *Dm*-CG1859, *Dm*-CG7722 und *Dm*-CG10956 fällt dagegen schwerer. Obwohl in diesen

Fällen der Bereich P15-P8 überwiegend der Konsensussequenz für inhibitorische Serpine entspricht, weisen doch einzelne Prolin-Aminosäurereste oder voluminöse Aminosäurereste auf eine zumindest eingeschränkte Protease-Inhibierungsaktivität hin.

Das in Abbildung 46 dargestellte Aminosäuresequenz-Alignment diente zur Konstruktion eines Intraspezies-Stammbaums (Abbildung 20). Für die Innengruppe der *Drosophila*-Serpine wird humanes α_1 -Antitrypsin als Außengruppe verwendet (diente bereits als Referenz-Serpin für das Alignment).

Größere OTU-Datensätze (hier 28 Sequenzen) führen bei Verwendung von Tree Puzzle oft zu kleinen Unterstützungswerten (*reliability values*), die als *Multifurcation* dargestellt werden (Ergebnisse nicht gezeigt). Daher wurde zur Konstruktion des Intraspezies-Stammbaums von *Drosophila* das Neighbor-Joining-Verfahren (Minimum Evolution, Abschnitt 4.7.3) eingesetzt. Zur Konstruktion des NJ-Serpin-Baums (Abbildung 20) und zur Durchführung der *Bootstrap*-Tests wurde PAUP*4.0b10¹ verwendet. Die gewählten Parameter sind der Beschriftung der Abbildung 20 zu entnehmen.

Es werden nur *Bootstrap*-Werte > 50 % aufgeführt. Der NJ-Baum kann mit einer *Bootstrap*-Unterstützung von 57 % in zwei große Cluster eingeteilt werden (gestrichelte Linie). Die erste Cluster-Gruppierung wird von fünf Serpin gebildet, die ausschließlich auf Chromosom 3 lokalisiert sind (*Dm*-CG6680, *Dm*-CG6663, *Dm*-CG6289, *Dm*-Spn5 und *Dm*-CG6687). Die zweite Cluster-Gruppierung ist wesentlich heterogener und setzt sich hauptsächlich aus „Chromosom 2“-Serpinen zusammen, beinhaltet aber auch die auf Chromosom 3 lokalisierten Serpine *Dm*-Acp76A und *Dm*-CG1342. Das Serpin *Dm*-CG7219 nimmt im NJ-Baum eine Außenseiterposition ein (siehe Diskussion Abschnitt 6.3.2). Deutlich höhere *Bootstrap*-Unterstützungswerte (> 75 %), die auch Monophylieaussagen ermöglichen, erhält man für die folgenden kleineren Gruppierungen:

Die Serpine *Dm*-Spn5 und *Dm*-CG6687 bilden einen Gencluster und werden auch in 100 % der *Bootstrap*-Replikatbäume gemeinsam gruppiert. Das gleiche gilt für die Gruppierungen der Serpine *Dm*-CG31902/*Dm*-CG33121 und der Gruppierung „*Dm*-CG6680/*Dm*-CG6663/*Dm*-CG6289“.

Mit einem *Bootstrap*-Wert von 77 % wird die Gruppierung von *Dm*-Spn43Aa und *Dm*-Spn43Ab unterstützt. Auch diese beiden Serpine sind Teil eines Serpin-Genclusters (aus insgesamt vier Serpinen). Dagegen weisen die Gruppierungen *Dm*-Spn2/*Dm*-Spn3 (100 %) und *Dm*-CG9455/*Dm*-CG7722 (98 %) zwar hohe Unterstützungswerte-Werte auf, aber die

¹ Swofford, D., L., 1998.

Serpin-Gene liegen nicht in gemeinsamen Genclustern vor (zum Vergleich siehe Abbildung 15).

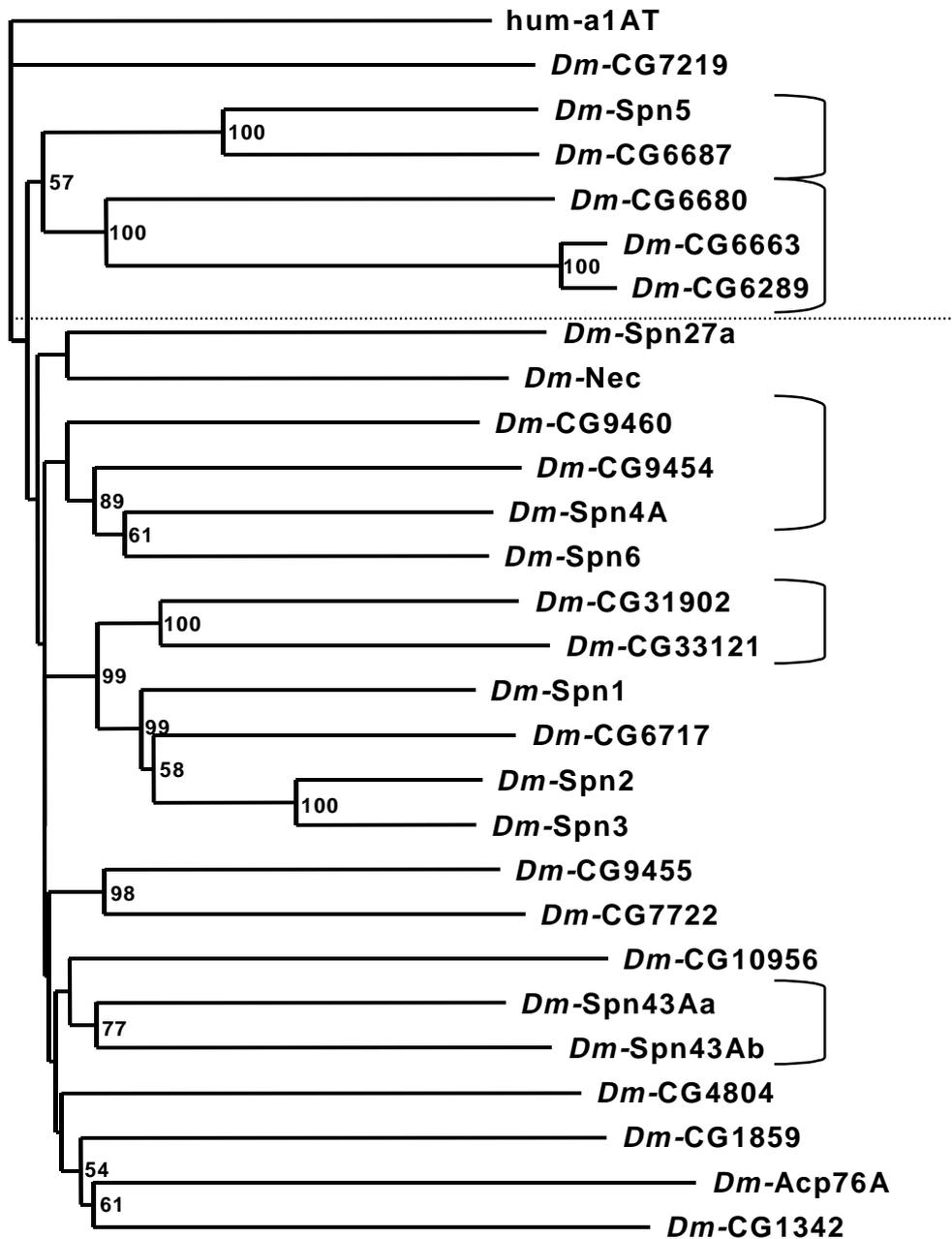


Abbildung 20: Neighbor-Joining-Baum der Serpine von *Drosophila melanogaster*

Das Phylogramm wurde mit PAUP*4.0b10 erstellt (Minimum Evolution, NJ-Distanzverfahren) und basiert auf der PAM250-Substitutionsmatrix¹. Der Baum wurde mit humanem α_1 -Antitrypsin (hum-a1AT) gewurzelt. Die *Bootstrap*-Werte (> 50 %, 1000 Replikate) sind jeweils rechts vom zugehörigen Knoten angegeben. Serpin-Gene, die zusammen in Genclustern vorliegen, sind durch eine Klammer markiert.

¹ Dayhoff, M., O., *et al.* 1978, konvertiert von Kuzoff, R., K.

5.3 Analyse der Serpin-Gene von *Arabidopsis thaliana*

5.3.1 Modellorganismus *Arabidopsis thaliana*

Arabidopsis thaliana (Ackerschmalwand) gehört zur Familie der Kreuzblütengewächse (Brassicaceae) und wurde als Modellorganismus¹ für Pflanzen ausgewählt. Phylogenetisch ordnet man die Ackerschmalwand *Arabidopsis thaliana* in die Abteilung der Spermatophyta (Samenpflanzen) ein. Vermutlich spaltete sich das Pflanzenreich (Plantae) vor ca. 700 - 1500 Millionen Jahren vom Tierreich (Animalia) ab^{2,3}. Ausschlaggebend für die Wahl von



Arabidopsis thaliana als Modellorganismus waren neben einem kleinen Genom⁴ auch die Etablierung von Agrobacterium-vermittelten Transformationstechniken⁵. Die Pflanze ist in Europa, Asien und Nordamerika verbreitet. Ein vollständiger Lebenszyklus wird in ca. sechs Wochen durchlaufen. Als weiteren Vorteile bietet *Arabidopsis* (Abbildung 21) eine einfache Kultivierbarkeit bei geringem Platzbedarf (durchschnittliche Größe ca. 15 bis 20 cm). Für die genetische Analyse ist hervorzuheben, dass das *Arabidopsis*-Genom vor ca. 112 Millionen Jahren eine Genomverdopplung durchlaufen hat⁶.

Abbildung 21: *Arabidopsis thaliana*

(Quelle: <http://www.mpimp-golm.mpg.de/arabidopsis/thaliana-e.html>)

Der Genomverdopplung folgten wiederum Genverluste, aber auch weitere Verdopplungen von einzelnen Genabschnitten (mit einer Größe von bis zu 100 kBp)⁶. Die Sequenz des *Arabidopsis*-Genoms wurde im Jahr 2000 veröffentlicht. Das Genom umfasst ca. 125 Millionen Basenpaare (MBp). Bisher konnten ca. 115 MBp sequenziert werden, die

¹ Meinke, D., W., Cherry, J., M., Dean, C., Rounsley, S., D., Koornneef, M., 1998.

² Seilacher A., Bose P., K., Pflüger F, 1998.

³ Morris, S., C., 1998.

⁴ Meyerowitz, E., M., Pruitt, R., E., 1985.

⁵ An, G., Watson, B., D., Chiang, C., C., 1986.

⁶ The Arabidopsis Genome Initiative, 2000.

ca. 25500 Gene kodieren. Die verbleibenden, bisher nicht sequenzierbaren Genomabschnitte (ca. 10 MBp) beinhalten repetitive bzw. centromere Sequenzabschnitte. Der haploide Chromosomensatz umfasst fünf Chromosomen (I - V), wobei die Chromosomengrößen im Bereich von ca. 17 - 30 MBp variieren. Derzeit wird das Genom von *Arabidopsis thaliana* durch die TAIR-Datenbank (*The Arabidopsis Information Resource*)¹ betreut und aktualisiert.

5.3.2 Die Serpin-Gene von *Arabidopsis thaliana*

Als Basis der eigenen Untersuchungen diente die TAIR-Datenbank (*The Arabidopsis Information Resource*). Die chromosomale Genomsequenzen wurden über die TAIR-Datenbank bezogen². Tabelle 23 führt die Serpin-Gene von *Arabidopsis thaliana* auf.

Name	Genmodell (TAIR-Nomenklatur)	Synonym	Protein- Größe (As)	cDNA- Größe (Nt)	Gen- Größe (Bp)
At-Spn-1	AT1G47710.1	T2E6.22	377	1176	1791
At-Spn-2	AT1G62170.1	F19K23.10	433	>1445	>1731
At-Spn-3	AT1G64030.1	F22C12.22	385	>1714	>2032
At-Spn-4	AT1G64035.1	F22C12.21	393	>1299	>1412
At-Spn-5	AT2G14540.1	T13P21.8	406	1451	1699
At-Spn-6	AT2G25240.1		385	>1279	>1384
At-Spn-7	AT2G26390.1	T9J22.6	389	>1890	>2121
At-Spn-8	AT2G35580.1	T32F12.4	388	>1466	>1984
At-Spn-9	AT3G45220.1	T14D3.160	393	>1275	>1543

Tabelle 23: Serpin-Gene von *Arabidopsis thaliana*

Die Namen der Serpin-Gene von *Arabidopsis thaliana* wurden in Anlehnung an die Nomenklatur der *C. elegans*-Serpine gewählt. Die Nummerierung der Serpine orientiert sich an der Reihenfolge ihrer chromosomalen Lokalisation (Chromosom I bis V, siehe Tabelle 25). Um in Interspezies-Stammbäumen eine eindeutige Identifikation der *Arabidopsis*-Serpine zu ermöglichen, wurden die Genbezeichnungen mit dem Präfix *At* für *Arabidopsis thaliana* erweitert. In der Spalte „Genmodell“ sind die Namen nach der aktuellen TAIR-Nomenklatur aufgeführt. In der Spalte „Synonyme“ sind die ursprünglichen Genbezeichnungen aufgeführt, die auf der BAC-Annotation basieren. Unter Protein-Größe ist die Größe des Präproteins in Aminosäuren (As)

¹ Garcia-Hernandez, M., et al. 2002.

² ftp://tairpub:tairpub@ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes/, June 2001.

angegeben. Die cDNA- und Gen-Größen sind in Nukleotiden (Nt) bzw. in Basenpaaren (Bp) angegeben. Die Genmodelle *At-Spn-1* und *At-Spn-5* basieren auf experimentellen Daten (TAIR-Klon „2313696“: 5'-UTR von *At-Spn-1*; TAIR-Klon „3009771“: 3'-UTR von *At-Spn-1*; TAIR-Klon „4721028“: mRNA von *At-Spn-5*). In den übrigen Fällen stehen zur Zeit keine experimentellen Daten zur Verfügung. Hier umfassen die Größenangaben die DNA-Sequenz zwischen Startcodon und Poly(A)-Signal.

Wie im Abschnitt 4.6.1 beschrieben, wurden die chromosomalen DNA-Sequenzen des *Arabidopsis*-Genoms in die sechs Aminosäureraster translatiert und für eine AGREP-Suche formatiert. Die Serpin-Gene wurden zunächst lokalisiert und ihre Exon-Intron-Strukturen auf Vollständigkeit überprüft.

Ursprünglich wurden im *Arabidopsis*-Genom 11 Serpin-Gene, zwei Serpin-Pseudogene und fünf Serpin-verwandte Gene (TAIR-Klassifikation: *serpin-related* bzw. *hypothetical protein, similar to putative serpin*) annotiert¹.

Aufgrund der eigenen Untersuchungen (Abschnitt 4.6.1) konnten von den 11 annotierten Serpin-Genen drei als Serpin-Pseudogene (AT1G62160.1, AT1G64010.1 und AT2G35590.1) identifiziert werden. Ein zuvor als Serpin-Pseudogen annotiertes Genmodell (AT1G64035) konnte dagegen als Serpin-Gen (*At-Spn-4*, siehe Tabelle 23) identifiziert werden. Von den fünf Serpin-verwandten Genen konnten drei (AT1G64020.1, AT2G35560.1 und AT2G35570.1) als Serpin-Pseudogene bzw. als deren Teilfragmente identifiziert werden. Bei den zwei verbleibenden Serpin-verwandten Genen (AT1G51330.1 und AT1G63280.1) handelt es sich um kleine Proteine (< 200 As), die eindeutig über konservierte, für Serpine charakteristische Aminosäuresequenzen verfügen. Aufgrund der Protein-Größe fehlen bedeutende Sekundärstruktur-Elemente, so dass es sich vermutlich nicht um funktionsfähige Serpine handelt.

Als Resultat der Überarbeitung konnten im Genom von *Arabidopsis thaliana* neun Serpin-Gene (*At-Spn-1* bis *At-Spn-9*), fünf Serpin-Pseudogene (AT1G62160.1, „AT1G64010.1/AT1G64020.1“, AT1G65385.1, AT2G35590.1 und „AT2G35560.1/AT2G35570.1“) und zwei Serpin-verwandte Gene (AT1G51330.1 und AT1G63280.1) identifiziert werden. Die identifizierten Serpin-Pseudogene bzw. Serpin-verwandten Gene sind in Tabelle 24 aufgeführt und wurden in den weiteren Untersuchungen ausgeschlossen.

Die durchgeführten Korrekturen der Serpin-Genstrukturen (siehe unten) sind bereits in Tabelle 23 und Abbildung 25 berücksichtigt. Die abgeleiteten Aminosäuresequenzen befinden sich im Anhang (Abschnitt 8.5.1). Angaben zu Intronposition oder Aminosäuren-Indels beziehen sich auf die Nummerierung des Referenz-Serpins reifes, humanes α_1 -Antitrypsin.

¹ TAIR, <http://www.arabidopsis.org/home.html>, Garcia-Hernandez, M., et al. 2002.

Genmodell der Pseudo-Serpingene (TAIR ¹ -Nomenklatur)	TAIR-Klassifizierung	Beschreibung
AT1G51330.1	<i>serpin-related</i>	AT1G51330.1 ist ein 193 As großes Protein, das eine hoch konservierte, für Serpine charakteristische Aminosäuresequenz enthält (Position ca. 158 –224 in Bezug auf hum-a1AT).
AT1G62160.1	<i>serpin, putative</i>	Serpin-Pseudogen AT1G62160.1 enthält eine Rastermutation, eine Deletion und eine Insertion (siehe Abbildung 23A).
AT1G63280.1	<i>serpin-related</i>	AT1G63280.1 ist ein 120 As großes Protein, das eine hoch konservierte, für Serpine charakteristische Aminosäuresequenz enthält (Position ca. 158 –224 in Bezug auf hum-a1AT).
AT1G64010.1	<i>serpin, putative</i>	AT1G64020.1 und AT1G64010.1 bilden zusammen ein Serpin-Pseudogen (siehe Abbildung 22B).
AT1G64020.1	<i>hypothetical protein</i>	
AT1G65385.1	<i>pseudogene</i>	Serpin-Pseudogen AT1G65385.1 enthält einen Rastermutation, die zur Termination führt (siehe Abbildung 23B).
AT2G35590.1	<i>serpin, putative</i>	Serpin-Pseudogen AT2G35590.1 enthält zwei Deletionen (siehe Abbildung 22A).
AT2G35560.1	<i>serpin-related</i>	AT2G35570.1 und AT2G35560.1 bilden zusammen ein Serpin-Pseudogen (siehe Abbildung 22A).
AT2G35570.1	<i>serpin-related</i>	

Tabelle 24: Die Serpin-Pseudogene von *Arabidopsis thaliana*

In der Tabelle sind fünf Serpin-Pseudogene und zwei Serpin-verwandte Gene aufgeführt. Die Serpin-verwandten Gene enthalten zwar Aminosäuresequenzen, die für Serpine charakteristisch sind, jedoch sind die abgeleiteten Proteinsequenzen für ein funktionsfähiges Serpin zu klein (< 200 As).

At-Spn-4

Das Serpin *At-Spn-4* (AT1G64035 bzw. F22C12.21) wurde ursprünglich von TAIR¹ als Serpin-Pseudogen annotiert. Eine eigene Untersuchung der Genstruktur (Abschnitt 4.6.1) zeigte, dass offensichtlich ein intaktes Serpin-Gen mit einer 5'-gelegenen Serpin-verwandten Sequenz zu einem Genmodell zusammengefasst wurde. Durch Entfernen der 5'-gelegenen Sequenz konnte die Genstruktur des Serpin-Gens *At-Spn-4* korrigiert werden. Das Serpin-Gen *At-Spn-4* ist Teil eines Genclusters (siehe Abbildung 22B).

¹ TAIR, <http://www.arabidopsis.org/home.html>, Garcia-Hernandez, M., et al. 2002.

At-Spn-8

Eine eigene Überprüfung (Abschnitt 4.6.1) des annotierten TAIR-Genmodells von *At-Spn-8* (AT2G35580.1) führte unter Berücksichtigung eines zusätzlich identifizierten Exons (kodiert eine für Serpine charakteristische Aminosäuresequenz) zur Erweiterung des Genmodells in 5'-Richtung. Durch Veränderung des 5'-gelegenen Introns (Position 20c in Bezug auf *At-Spn-8*) wird die abgeleitete Proteinsequenz des Serpins in Richtung Aminoterminus erweitert (siehe Abbildung 25).

Serpin-Pseudogen AT1G62160.1

Eine eigene Überprüfung (Abschnitt 4.6.1) des annotierten TAIR-Genmodells AT1G62160.1 führte zur Identifikation einer Rastermutation (Position ca. 40), einer ca. 31 As große Deletion (Positionen ca. 122 – 152) und einer ca. 24 As großen Insertion (Position ca. 213), die zu einer weiteren Rastermutation führt. Da beide Rastermutationen zur Termination der abgeleiteten Proteinsequenz führen, wird das Genmodell AT1G62160.1 als Serpin-Pseudogen eingestuft. Das Pseudogen bildet mit *At-Spn-2* einen Gencluster (Abbildung 23A).

Serpin-Pseudogen „AT1G64010.1 / AT1G64020.1“

Eine Überprüfung (Abschnitt 4.6.1) der annotierten TAIR-Genmodelle AT1G64010.1 und AT1G64020.1 zeigte, dass es sich um das erste Exon (AT1G64020.1) und zweite Exon (AT1G64010.1) eines Serpin-Pseudogens (Abbildung 22B) handelt. Das Pseudogen „AT1G64010.1/AT1G64020.1“ weist eine ca. 10 As große Deletion (Position ca. 140 – 149) auf, die zum Verlust der 5'-Spleißstelle des 149c-Introns führt. Zusätzlich kann eine Rastermutation an Position ca. 186 identifiziert werden, die zur vorzeitigen Termination führt. Das Pseudogen ist Teil eines Genclusters.

Serpin-Pseudogen AT1G65385.1

Eine Überprüfung (Abschnitt 4.6.1) des annotierten TAIR-Genmodells AT1G65385.1 führte zur Identifikation einer Rastermutation (Position ca. 238). Die Rastermutation befindet sich im Protein-kodierenden Bereich des Serpin-Gens und führt zur vorzeitigen Termination (Abbildung 23B). Die TAIR-Klassifikation als Serpin-Pseudogen wird somit beibehalten.

Serpin-Pseudogen AT2G35590.1

Eine Überprüfung (Abschnitt 4.6.1) des annotierten TAIR-Genmodells AT2G35590.1 führte zur Identifikation von zwei Deletionen (Position ca. 49 – 72 und 138 – 149, Abbildung 22A). Die zweite Deletion führt zum Verlust der 5'-Spleißstelle des 149c-Introns. Das Serpin-Pseudogen ist Teil eines Genclusters.

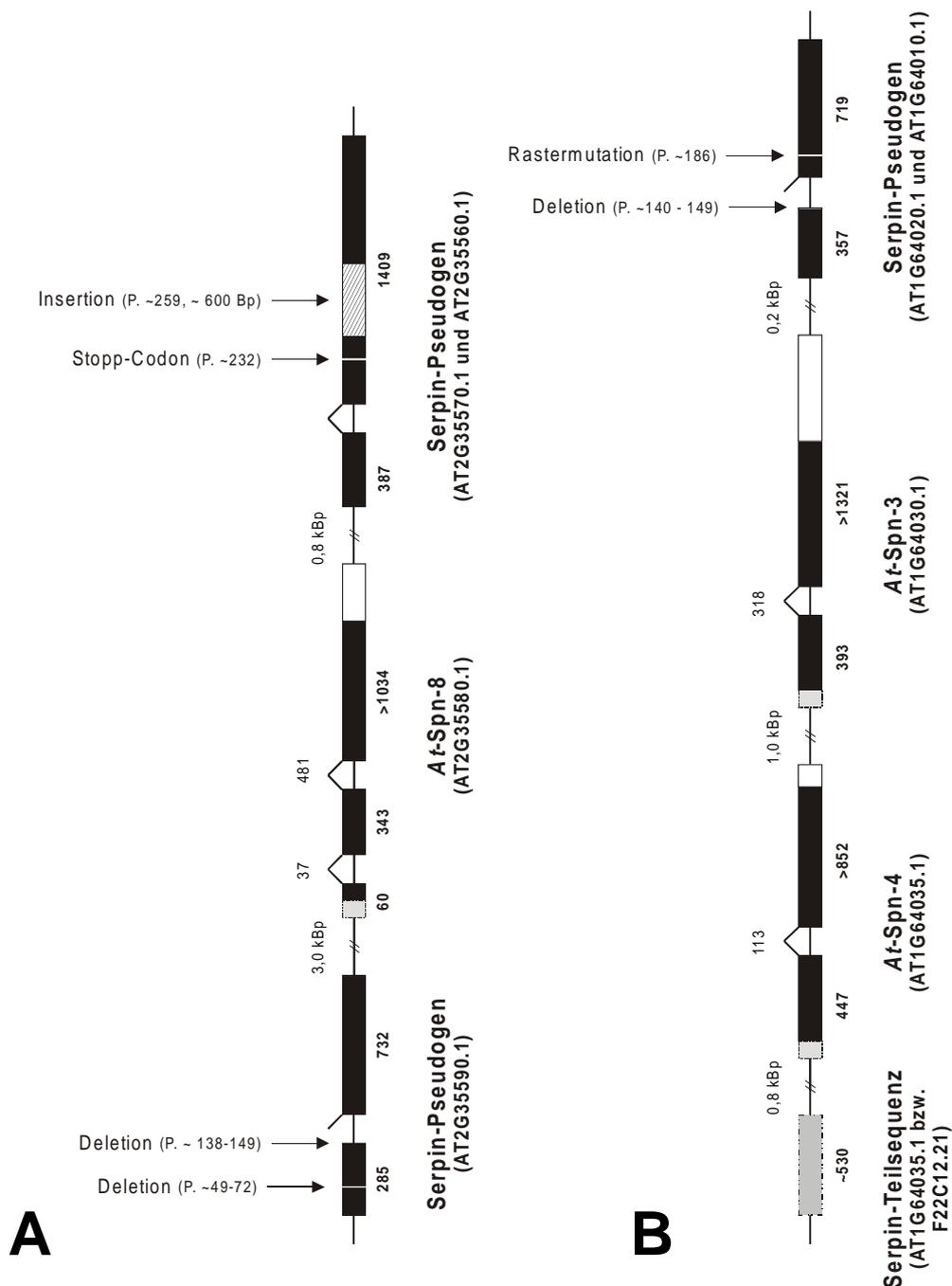


Abbildung 22: Organisation der Serpin-Gencluster in *Arabidopsis thaliana*

Abbildung 22A zeigt den Aufbau des Serpin-Genclusters auf Chromosom II (Minus-Strang, Start ca. 14879500 Bp). Der Cluster setzt sich zusammen aus dem Serpin-Pseudogen AT2G35590.1, *At-Spn-8* (AT2G35580.1) und dem Serpin-Pseudogen „AT2G35570.1/AT2G35560.1“. Abbildung 22B zeigt den Aufbau des Serpin-Genclusters auf Chromosom I (Minus-Strang, Start ca. 23350900 Bp). Der Cluster setzt sich zusammen aus einer Serpin-verwandten Teilsequenz (F22C12.21), *At-Spn-4* (AT1G64035.1), *At-Spn-3* (AT1G64030.1) und dem Serpin-Pseudogen „AT1G64020.1/AT1G64010.1“. Die durchgehenden schwarzen Linien stellen die chromosomale Sequenz dar. Schwarze Kästen symbolisieren die Exons der Serpin-Gene. Graue Kästen symbolisieren die 5'-UTRs der Serpin-Gene, während die weißen Kästchen die 3'-UTRs (umfassen

die DNA-Sequenz zwischen Startcodon und Poly(A)-Signal) darstellen. Die 5'- und 3'-Spleißstellen werden oberhalb der Gene durch abgeknickte Linien verbunden.

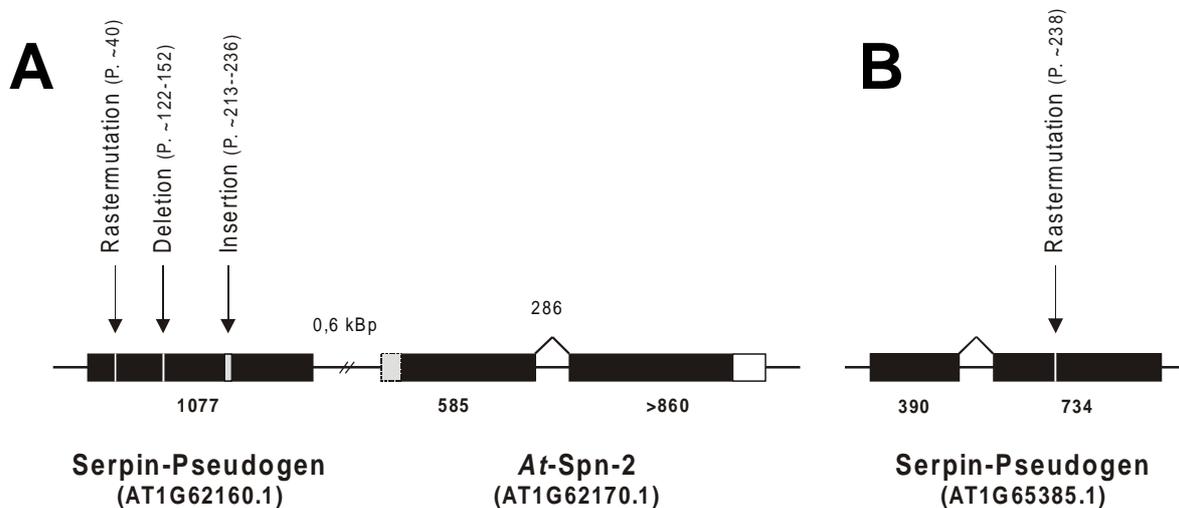


Abbildung 23: Struktur der Serpin-Pseudogene AT1G62160.1 und AT1G65385.1

Abbildung 23A zeigt den Aufbau des Serpin-Genclusters auf Chromosom I (Plus-Strang, Start ca. 22572000 Bp). Der Cluster setzt sich zusammen aus dem Serpin-Pseudogen AT1G62160.1 und *At-Spn-2* (AT1G62170.1). Abbildung 23B zeigt den Aufbau des Serpin-Pseudogens AT1G65385.1 auf Chromosom I (Minus-Strang, Start ca. 24356549 Bp).

Serpin-Pseudogen „AT2G35560.1 / AT2G35570.1“

Eine Überprüfung (Abschnitt 4.6.1) der annotierten TAIR-Genmodelle AT2G35560.1 und AT2G35570.1 zeigte, dass es sich um das erste Exon (AT2G35570.1) und zweite Exon (AT2G35560.1) eines Serpin-Pseudogens (Abbildung 22A) handelt. Das Pseudogen „AT2G35570.1/AT2G35560.1“ weist ein Stopp-Codon in Position 232 auf, das eine frühzeitige Termination verursacht. Zusätzlich befindet sich eine ca. 660 Bp große Insertion innerhalb des zweiten Exons (an Position 259). Das Pseudogen ist Teil eines Genclusters.

Es kann nicht ausgeschlossen werden, dass die als Serpin-Pseudogene bzw. Serpin-verwandte Gene eingestuft Sequenzen neue Aufgaben in *Arabidopsis thaliana* übernommen haben, und es sich somit um transkribierbare bzw. translatierbare Gene handelt. Das Fehlen von experimentellen Daten, weist zur Zeit eher auf eine Klassifikation als Serpin-Pseudogene hin. Die Genomverdopplung von *Arabidopsis thaliana* könnte die Ursache für das Auftreten dieser großen Zahl von Pseudogenen sein (siehe Diskussion, Abschnitt 6.4.1).

5.3.3 Chromosomale Lokalisation der Serpin-Gene von *Arabidopsis*

Das haploide Genom von *Arabidopsis thaliana* besteht aus fünf Chromosomen (I, II, III, IV, V). In Abbildung 24 ist die Lage der neun intakten Serpin-Gene dargestellt.

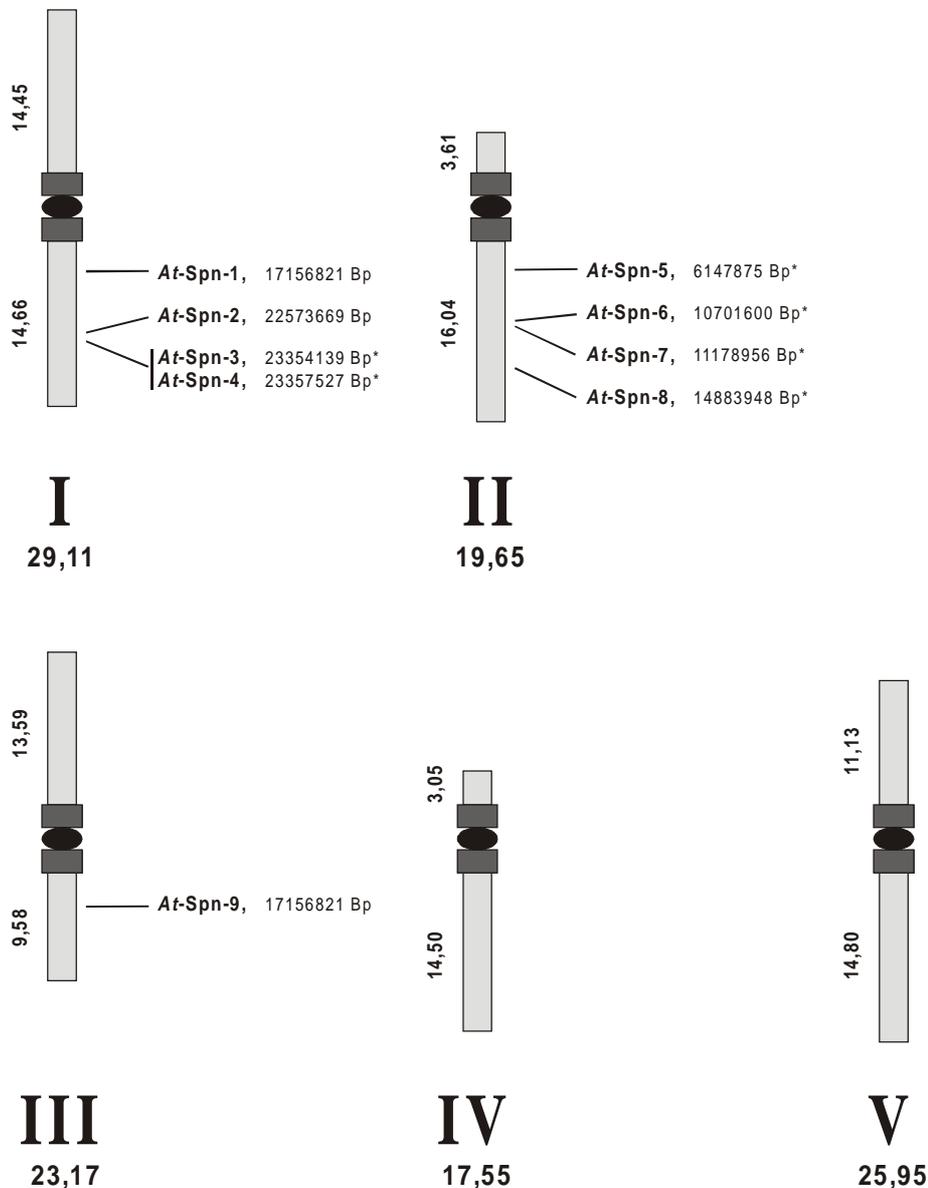


Abbildung 24: Chromosomale Lokalisation der Serpin-Gene von *Arabidopsis*

In der Abbildung¹ sind die Centromere der Chromosomen als schwarze Ovale symbolisiert. Die größtenteils noch nicht sequenzierten, centromeren Regionen (flankieren das Centromer) sind dunkelgrau markiert. Das sequenzierte Euchromatin ist hellgrau markiert, wobei der „Top“-Arm nach oben und der „Bottom“-Arm nach unten weist. Die Größen der Chromosomen-Bereiche sind jeweils auf der linken Seite der Chromosomen in MBp angegeben. Auf dem Minus-Strang lokalisierte Serpin-Gene sind mit einem Stern markiert.

¹ Basiert auf Angaben von „The Arabidopsis Genome Initiative“, 2000.

Von den neun Serpin-Genen von *Arabidopsis thaliana* sind vier Gene auf Chromosom I, vier Gene auf Chromosom II und ein Serpin-Gen auf Chromosom III lokalisiert. Im Vergleich zu *Caenorhabditis elegans* und *Drosophila melanogaster* liegen die Serpin-Gene im *Arabidopsis*-Genom größtenteils vereinzelt vor und nur *At-Spn-4/At-Spn-3* bilden einen Serpin-Gencluster. Auffällig ist jedoch die hohe Anzahl an Serpin-Pseudogenen. In vielen Fällen befinden sich die Serpin-Pseudogene in unmittelbarer Nähe der intakten Serpin-Gene (siehe Abbildung 22 und Abbildung 23).

5.3.4 Genstrukturen der Serpin-Gene von *Arabidopsis thaliana*

In Abbildung 25 sind die Genstrukturen der Serpin-Gene von *Arabidopsis thaliana* dargestellt. Soweit experimentelle Daten zur Verfügung stehen, wurden sie in die Abbildungen integriert (*At-Spn-1* und *At-Spn-5*). Die Phasen (a, b, c) und die Positionen der Introns sind in Tabelle 25 angegeben. Die von den Genstrukturen abgeleiteten Aminosäuresequenzen sind im Anhang (Abschnitt 8.5.1) aufgeführt.

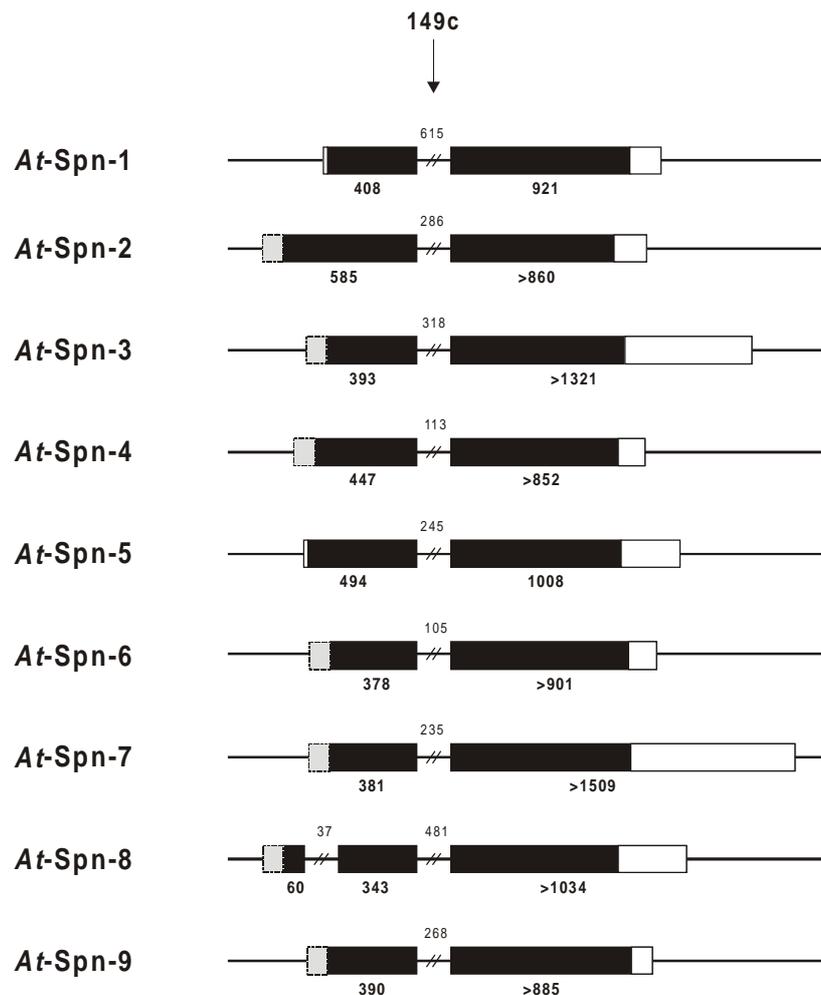


Abbildung 25: Genstrukturen der Serpin-Gene von *Arabidopsis thaliana*

Protein-kodierende Exons sind als schwarze Kästchen dargestellt. Die 5'-UTRs mit unbekannter Größe sind durch graue Kästchen dargestellt. Die 3'-UTRs (weiße Kästchen) umfassen die 3'-untranslatierte Region einschließlich des Poly(A)-Signals (AATAAA oder ATTAAA). Die Größenangaben (Bp) für Exons und Introns stehen unterhalb bzw. oberhalb der Genstrukturen. Das 3'-terminale Exon endet jeweils mit einem Poly(A)-Signal und ist dementsprechend um ca. 10 - 40 Bp größer als angegeben.

5.3.5 Intronpositionen der Serpin-Gene von *Arabidopsis thaliana*

Die Intronpositionen und Intronphasen (a, b, c) der Serpin-Gene von *Arabidopsis thaliana* wurden mit Hilfe des Programms Genewise bestimmt. Auf der Basis dieser genspezifischen Intronpositionen wurden anschließend mit Hilfe eines Aminosäuresequenz-Alignments die relativen Intronpositionen in Bezug auf das Referenz-Serpin reifes, humanes α_1 -Antitrypsin (hum-a1AT) ermittelt. Um die Intronpositionen gesichert bestimmen zu können, wurden die Alignments mit zwei verschiedenen Alignment-Programmen (DIALIGN 2.2.1, ClustalX 1.83, siehe Abschnitt 4.7.1) erstellt. In Tabelle 25 sind die Intronpositionen in Bezug auf reifes, humanes α_1 -Antitrypsin aufgeführt (Alignment siehe Abbildung 47).

Name	Genmodell (TAIR-Nomenklatur)	Chromosomale Lokalisation (Chromosom, Startposition in Bp, Orientierung*)	Intronpositionen	
			20c	149c
At-Spn-1	AT1G47710.1	I, 17156821		+
At-Spn-2	AT1G62170.1	I, 22573669		+
At-Spn-3	AT1G64030.1	I, 23354139*		+
At-Spn-4	AT1G64035.1	I, 23357527*		+
At-Spn-5	AT2G14540.1	II, 6147875*		+
At-Spn-6	AT2G25240.1	II, 10701600*		+
At-Spn-7	AT2G26390.1	II, 11178956*		+
At-Spn-8	AT2G35580.1	II, 14883948*	+	+
At-Spn-9	AT3G45220.1	III, 16574325*		+

Tabelle 25: Intronpositionen der Serpin-Gene von *Arabidopsis thaliana*

Die grau markierte Intronposition 20c bezieht sich auf die Sequenz von *At-Spn-8* und ist im mutmaßlichen Signalpeptid lokalisiert. DIALIGN- und ClustalX-Alignments ergeben übereinstimmend eine Intronposition von 149c. In der Spalte „Chromosomale Lokalisation¹“, wird die Lage der Serpin-Gene und die absolute chromosomalen Startpositionen in Bp angegeben. Auf dem Minus-Strang lokalisierte Serpin-Gene, sind mit einem Stern markiert.

Alle neun Serpin-Gene von *Arabidopsis thaliana* weisen das 149c-Intron auf. Aufgrund der Gleichförmigkeit der Exon-Intron-Strukturen handelt es sich hier um ein nicht-informatives Merkmal. Die Intronposition 149c kann ebenfalls beim Gerste-Serpin BSZ4 identifiziert werden².

¹ TAIR, <http://www.arabidopsis.org/home.html>, Garcia-Hernandez, M., et al. 2002.

² Rasmussen, S., K., Dahl, S., W., Norgard, A., Hejgaard, J., 1996.

5.3.6 Phylogenetische Analyse der Serpine von *Arabidopsis thaliana*

Die korrigierten Aminosäuresequenzen der *Arabidopsis*-Serpine wurden mit DIALIGN aligniert. Das Alignment wurde mit GeneDoc überprüft und optimiert. Die hoch divergente RCL-Region wurde strikt linksbündig ausgerichtet. Zur Darstellung des Alignments wurde ESPript 2.0 verwendet (siehe Anhang Abbildung 47). Alle Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (reifes hum-a1AT).

Die *Arabidopsis*-Serpine weisen untereinander eine hohe Aminosäuresequenz-Identität auf und enthalten im Vergleich zum α_1 -Antitrypsin nur kleinere Indels. Diese Indels sind hauptsächlich zwischen den Sekundärstruktur-Elementen lokalisiert, ähnlich wie bei den Serpinen von *Caenorhabditis elegans* und *Drosophila melanogaster*. Ein größerer Indel-Bereich kann nur zwischen den Sekundärstruktur-Elementen hI und s5A identifiziert werden. Hier kommt es im Vergleich zum α_1 -Antitrypsin teilweise zu Deletionen (*At-Spn-2*, *At-Spn-3*, *At-Spn-4* und *At-Spn-5*), aber auch zu Insertionen (*At-Spn-9*). Das Serpin *At-Spn-2* (AT1G62170.1) weist im Vergleich zu den anderen *Arabidopsis*-Serpinen eine aminoterminalen Extension von ca. 60 Aminosäuren auf.

Der Aufbau des aminoterminalen Teils des RCL erlaubt eine Abschätzung der Protease-Inhibierungsaktivität eines Serpins (Abschnitt 2.1.4). Bei alle *Arabidopsis*-Serpine zeigt der aminoterminalen Teil des RCL (Positionen P17-P7) einen hoch konservierten, für Serpine charakteristischen Aufbau. Die Aminosäuresequenzen wie auch die Länge der RCLs weisen darauf hin, dass alle *Arabidopsis*-Serpine (*At-Spn-1* - *At-Spn-9*) über eine potentielle Protease-Inhibierungsaktivität verfügen.

Das Alignment der *Arabidopsis*-Serpine (Abbildung 47) diente zur Konstruktion eines Intraspezies-Stammbaums (Abbildung 26). Die Serpin-Pseudogene wurden bei der Stammbaumrekonstruktion nicht berücksichtigt. Als Außengruppe wurde reifes, humanes α_1 -Antitrypsin verwendet, das bereits für das Alignment als Referenz-Serpin diente.

Als Baum-Rekonstruktionsmethode wurde das Maximum-Likelihood-Verfahren (Abschnitt 4.7.2) eingesetzt. Die ML-Analyse wurde mit Tree Puzzle 5.0¹ durchgeführt. Zur Bewertung der OTU-Gruppierungen wird der methoden-interne „*Bootstrap*“-Wert, der *reliability value*, als Unterstützungswert angegeben.

Der analysierte Datensatz weist ca. 2,4 % ungelöste Quartette auf (*unresolved quartets*), so dass keine *Multifurcation* in der Topologie auftritt. Im dargestellte ML-Baum (Abbildung 26) gruppieren sich die *Arabidopsis*-Serpine in zwei Cluster (siehe auch Diskussion 6.4):

¹ Strimmer, K., von Haeseler., A., 1996.

Die erste Gruppierung von drei Serpinen (*At-Spn-6*, *At-Spn-7* und *At-Spn-9*) setzt sich deutlich von der zweiten Gruppierung (*At-Spn-1*, *At-Spn-2*, *At-Spn-3*, *At-Spn-4*, *At-Spn-5* und *At-Spn-8*) ab. Im Unterschied zu den Serpinen von *Caenorhabditis elegans* und *Drosophila melanogaster* korreliert der Serpin-Stammbaum von *Arabidopsis thaliana* nicht mit der chromosomalen Lokalisation der Serpin-Gene. Die gezeigte Baum-Topologie erhält man auch bei Anwendung anderer Baumrekonstruktions-Methoden (z.B. NJ) oder bei Variation der Aminosäure-Substitutionsmatrix.

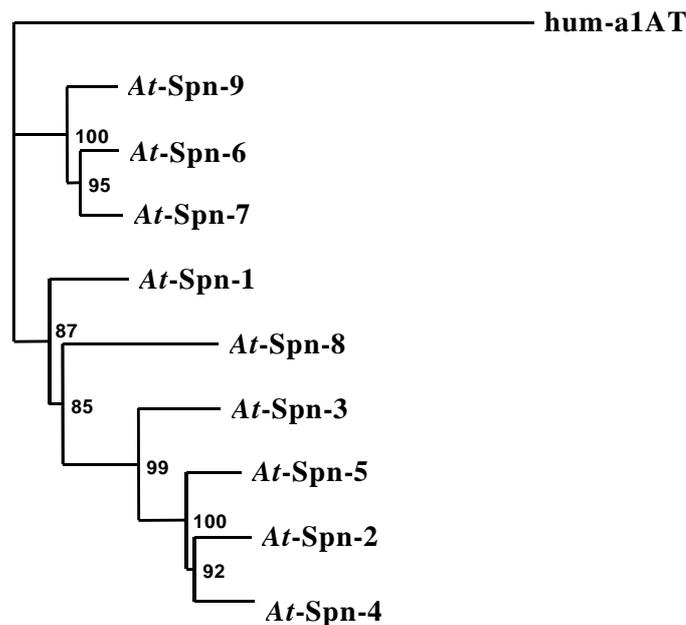


Abbildung 26: Maximum-Likelihood-Baum der Serpine von *Arabidopsis thaliana*

Das Phylogramm wurde mit Tree Puzzle 5.0 erstellt und basiert auf der JTT-Substitutionsmatrix (Jones-Taylor-Thornton)¹. Abweichend von den Voreinstellungen wurde eine exakte Parameterschätzung (*quartet sampling + NJ tree*) gewählt. Der Baum wurde mit humanem α 1-Antitrypsin (hum-a1AT) gewurzelt. Die Unterstützungswerte (*reliability values*, 1000 puzzle steps) sind jeweils rechts vom zugehörigen Knoten angegeben. Zur Darstellung des Baums wurde das Programm TreeView² verwendet.

¹ Jones, D., T., *et al.* 1992

² Page, R., D., M., 1996.

5.4 Analyse der Serpin-Gene von *Ciona intestinalis*

5.4.1 Modellorganismus *Ciona intestinalis*

Ciona intestinalis (Seescheide, Seevase) gehört zu den Manteltieren (Tunicata) und wurde als Modellorganismus¹ für einfach strukturierte Chordata bzw. als Vertreter eines Ur-Chordaten ausgewählt. Phylogenetisch ordnet man die Seescheide *Ciona intestinalis* in den Unterstamm der Urochordata (Tunicata) ein. Vermutlich hat sich der Stamm der Chordata vor ca. 550 Millionen Jahren in die drei Unterstämme Urochordata, Cephalochordata und Vertebrata aufgespalten¹.

Ausschlaggebend für die Wahl von *Ciona* waren neben der geringen Größe des Genoms auch die leicht zu verfolgende Zellentwicklung und der relativ kurze Lebenszyklus von ungefähr drei Monaten. Seescheiden sind marine Invertebraten, die in Flachwassergebieten der Ozeane leben und zwei Lebensstadien durchlaufen. Im freischwimmenden Kaulquappen-Stadium (Tadpole, Abbildung 27A) verfügen die ca. 1 mm großen Larven über ein dorsal-zentrales Nervensystem und eine *Chorda dorsalis* und zeigen somit die einfache Version eines Chordaten-Bauplans. Im Kaulquappen-ähnlichen Stadium ist *Ciona* aus ca. 2500 Zellen aufgebaut, deren Entwicklung sich aufgrund der Transparenz der Larve leicht unter dem Mikroskop beobachten läßt.



Abbildung 27: *Ciona intestinalis*, Larven-Stadium (A) und adultes Stadium (B)

JGI, *Ciona intestinalis* v1.0 , (http://www.jgi.doe.gov/xpress/ciona_pics/).

Nach einer Metamorphose heftet sich *Ciona* dauerhaft an Felsen bzw. den Meeresuntergrund und bildet einen unsegmentierten Mantel, der aus der Stützsubstanz Tunicin (Glukose-

¹ Meinke, D., W., Cherry, J., M., Dean, C., Rounsley, S., D., Koornneef, M., 1998.

Polysaccharid, „tierische Zellulose“^{1,2} besteht. Im adulten Stadium wird *Ciona* ca. 10 - 20 cm groß (Abbildung 27B) und ernährt sich als Nahrungsstrudler. Die Chordata-Merkmale sind in der adulten Form zurückgebildet.

Die erste Draft-Version des *Ciona*-Genoms wurde im Jahre 2002 veröffentlicht³. Die Größe des Genoms wird auf ca. 153 bis 159 Millionen Basenpaare geschätzt⁴. Das derzeitige *Ciona*-Assembly (*Ciona intestinalis* v1.0) umfasst 116,7 MBp und kodiert ungefähr 16000 Gene. Das *Ciona*-Genom wird durch die JGI-Datenbank (*Joint Genome Institute*)⁵ betreut und aktualisiert.

5.4.2 Die Serpin-Gene von *Ciona intestinalis*

Als Basis der vorliegenden Untersuchung diente die „*Ghost Database: a Ciona intestinalis cDNA resource*“⁶ und die JGI-Datenbank „*Ciona intestinalis v1.0*“¹. In Tabelle 26 sind die Serpin-Gene von *Ciona intestinalis* aufgeführt.

Die chromosomalen DNA-Sequenzen wurden, wie im Abschnitt 4.6.1 beschrieben, in die sechs Aminosäureraster translatiert und für eine AGREP-Suche formatiert. Im *Ciona*-Genom konnten zehn Serpin-Gene annotiert werden. Die Serpin-Gene wurden zunächst lokalisiert und ihre Exon-Intron-Strukturen auf Vollständigkeit überprüft. Aufgrund eigener Untersuchungen (Abschnitt 4.6.1) konnten vier der in Tabelle 26 aufgeführten Serpin-Genmodelle (ci0100143209, ci0100146394, ci0100148346 und ci0100154072) korrigiert bzw. ergänzt werden. Es konnten zusätzliche Exons identifiziert werden, die für Serpine charakteristische Aminosäuresequenzen kodieren.

Die durchgeführten Korrekturen der Exon-Intron-Strukturen (siehe unten) sind bereits in Tabelle 26 berücksichtigt. Die aktualisierten Strukturen sind in Abbildung 28 dargestellt und die abgeleiteten Aminosäuresequenzen sind im Anhang Abschnitt 8.5.1 aufgeführt. Angaben zu Intronpositionen oder Aminosäuren-Indels beziehen sich auf die Nummerierung des Referenz-Serpins reifes, humanes α_1 -Antitrypsin.

¹ Hagenbuchner, K., 1999.

² Bracher, F., 1994.

³ Dehal, P., et al. 2002. (<http://genome.jgi-psf.org/ciona4/ciona4.home.html>).

⁴ Simmen, M., W., Leitgeb, S., Clark, V., H., Jones, S., J., Bird A., 1998.

⁵ Garcia-Hernandez, M., et al. 2002.

⁶ Satou, Y., et al. 2002.

Genname	JGI Protein Id / Genmodell	Synonym	Protein- Größe (As)	cDNA- Größe (Nt)	Gen- Größe (Bp)
<i>Ci-Spn-1</i>	ci0100132788		449	1574	4712
<i>Ci-Spn-2</i>	ci0100132818		412	1741	3185
<i>Ci-Spn-3</i>	ci0100134682		402	1348	3804
<i>Ci-Spn-4</i>	ci0100141118	Serpin 2	441	1657	5361
<i>Ci-Spn-5</i>	ci0100143209		413	2079	6240
<i>Ci-Spn-6</i>	ci0100146394		377	1287	1287
<i>Ci-Spn-7</i>	ci0100146394		380	1798	1798
<i>Ci-Spn-8</i>	ci0100146394		379	1688	1688
<i>Ci-Spn-9</i>	ci0100148346		409	1470	3579
<i>Ci-Spn-10A</i>	ci0100154072		408	1403	3393
<i>Ci-Spn-10B</i>	ci0100154072	Serpin 1	407	1507	3393

Tabelle 26: Die Serpin-Gene von *Ciona intestinalis*

Da die Untersuchung des *Ciona*-Genoms auf der ersten Draft-Version basiert, steht zur Zeit keine beständige Nomenklatur zur Verfügung. In der Spalte „Genname“ wurde schließlich eine Bezeichnung gewählt, die mit der Nummerierung der Genmodell/ProteinId-Nomenklatur (JGI *Ciona* v1.0) korreliert. Um in Interspezies-Stammbäumen (Abschnitt 5.5) eine klare Identifikation der Serpine zu ermöglichen, werden die Gennamen mit dem Präfix *Ci* (für *Ciona intestinalis*) erweitert. Unter „Protein-Größe“ ist die Größe des Präproteins in Aminosäuren (As) angegeben. Die Größen von cDNA und Gen sind in Nukleotiden (Nt) bzw. in Basenpaaren (Bp) angegeben. Bei allen Genmodellen wurden die zur Verfügung stehenden experimentellen Daten (cDNA- bzw. EST-Sequenzen) zur Ableitung der Exon-Intron-Strukturen berücksichtigt. Das Serpin-Gen *Ci-Spn-10* weist zwei Varianten des RCL-kodierenden Exons auf, die alternatives Spleißen ermöglichen (Abbildung 37).

***Ci-Spn-5* (ci0100143209)**

Eine eigene Überprüfung (Abschnitt 4.6.1) des annotierten JGI-Genmodells für *Ci-Spn-5* führte unter Berücksichtigung von cDNA-Daten¹ („ciad038b20_20“) zur Erweiterung der Genstruktur in 5‘- und 3‘-Richtung. Aufgrund der durchgeführten Korrektur wurde der Protein-kodierende Bereich am Aminoterminus um 33 As und am Carboxyterminus um 2 As erweitert.

¹ Dehal, P., et al. 2002. (<http://genome.jgi-psf.org/ciona4/ciona4.home.html>).

Ci-Spn-6, Ci-Spn-7 und Ci-Spn-8 (ci0100146394)

Eine eigene Überprüfung (Abschnitt 4.6.1) des annotierten JGI-Genmodells ci0100146394 führte unter Berücksichtigung von EST-Daten¹ zu dem Ergebnis, dass sich das annotierte Serpin-Genmodell aus drei intronfreien Serpin-Genen zusammensetzt. Die Serpin-Gene bilden zusammen einen Gencluster. Mit Hilfe der EST-Daten gelang es, die Genstrukturen der drei Serpin-Gene zu vervollständigen:

Serpin-Gen *Ci-Spn-6* (ci0100146394**A**) wird durch die EST-Sequenzen¹ „cinc006o22“ und „rcinc006o22“ begrenzt und ist in Bezug auf den Gencluster am 5'-Ende lokalisiert.

Serpin-Gen *Ci-Spn-7* (ci0100146394**B**) ist das mittlere Serpin-Gen im Gencluster und wird durch die EST-Sequenzen¹ „rcibd043j06“ und „citb069b04“ begrenzt.

Serpin-Gen *Ci-Spn-8* (ci0100146394**C**) ist am 3'-Ende des Genclusters lokalisiert und wird durch die EST-Sequenzen¹ „rcibd080f02“ und „cibd002g03“ begrenzt.

Ci-Spn-9 (ci0100148346)

Eine Untersuchung (Abschnitt 4.6.1) des annotierten JGI-Genmodells für *Ci-Spn-9* führte unter Berücksichtigung von EST-Daten¹ („rcicl06l02“ und „citb049d07“) zur Erweiterung der Genstruktur in 5'- und 3'-Richtung. Aufgrund der durchgeführten Korrektur wurde der Protein-kodierende Bereich am Aminoterminus um 8 As und am Carboxyterminus um 10 As erweitert (Abschnitt 8.5.1, grau markiert Sequenz). Das resultierende Protein verfügt am carboxyterminalen Ende über die Aminosäuresequenz „HDEF“, die eine große Ähnlichkeit zum Rückhaltesignal für das endoplasmatischen Retikulum zeigt („HDEL“ oder „KDEL“).

Ci-Spn-10 (ci0100154072)

Eine eigene Untersuchung (Abschnitt 4.6.1) des annotierten JGI-Genmodells für *Ci-Spn-10* führte unter Berücksichtigung von EST-Daten zur Identifikation von insgesamt zwei Varianten des RCL-kodierenden Exons (Abbildung 28 bzw. Abbildung 37). Die Expression der Spleißvariante *Ci-Spn-10A* wird durch die EST-Sequenz¹ „rcibd064c17“ gestützt, die die Verwendung der (5'-gelegenen) Variante I der beiden RCL-kodierenden Exons zeigt. Die Expression der zweiten Spleißvariante *Ci-Spn-10B* (ursprüngliches Genmodell ci0100154072) wird ebenfalls durch EST-Daten¹ („rciht030l17“) gestützt.

Zur Zeit kann das Assembly der Draft-Version des *Ciona*-Genoms noch nicht den 14 *Ciona*-Chromosomen zugeordnet¹ werden. Eine genomische Positionierung der Serpin-Gene ist daher nicht möglich.

¹ Dehal, P., *et al.* 2002. (<http://genome.jgi-psf.org/ciona4/ciona4.home.html>).

5.4.3 Genstrukturen der Serpin-Gene von *Ciona intestinalis*

In Abbildung 28 sind die Genstrukturen der Serpin-Gene von *Ciona* dargestellt. Alle Genmodelle basieren auf experimentellen Daten (EST- oder cDNA-Daten¹). Die Phasen (a, b, c) und die Positionen der Introns sind in Tabelle 27 angegeben.

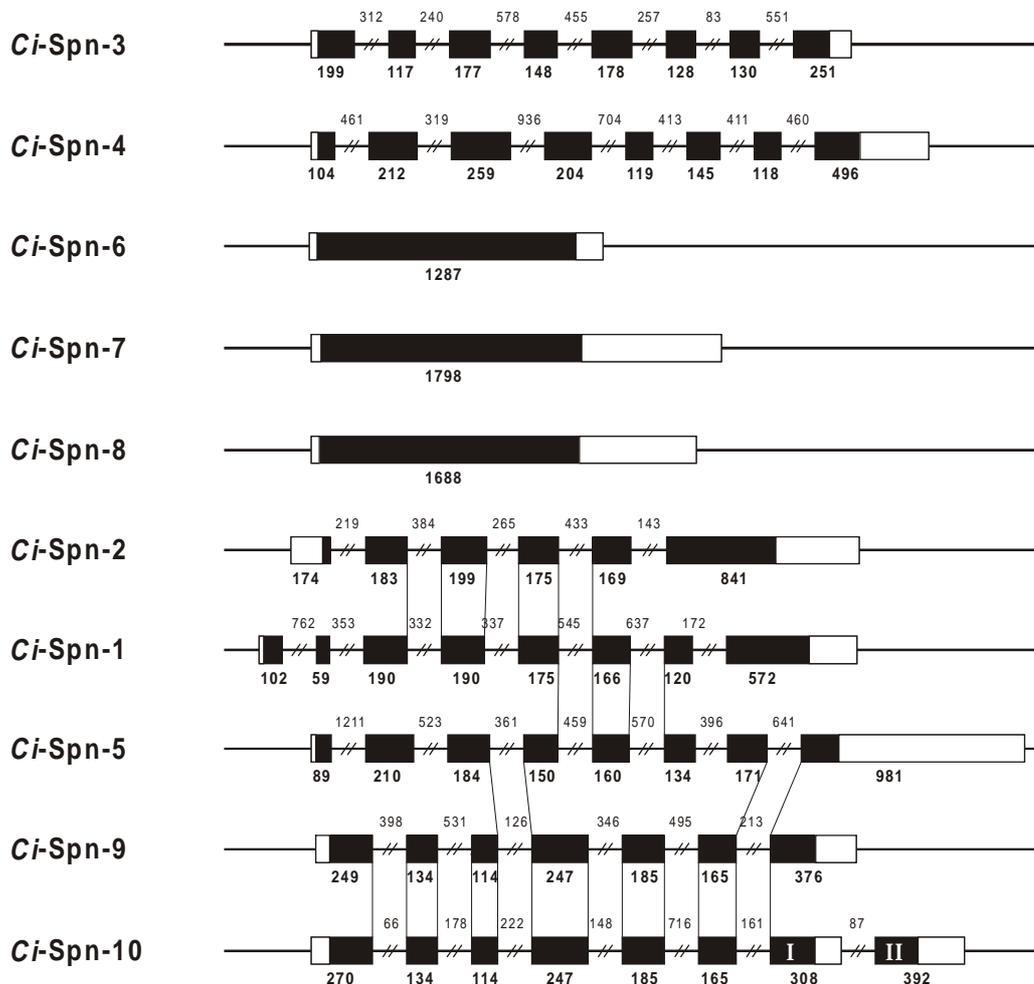


Abbildung 28: Genstrukturen der Serpin-Gene von *Ciona intestinalis*

Protein-kodierende Exonbereiche sind als schwarze Kästchen und 5'- und 3'-UTR-Exonbereiche sind als weiße Kästchen dargestellt. Die Größenangaben (Bp) für Exons und Introns stehen unterhalb bzw. oberhalb der Genstrukturen. Das Serpin-Gen *Ci-Spn-10* enthält zwei Varianten des RCL-kodierenden Exons (markiert mit römischen Ziffern; siehe auch Abbildung 37). Die gemeinsamen Intronpositionen bzw. 5'- und 3'-Spleißstellen sind durch Linien verbunden.

¹ Dehal, P., et al. 2002. (<http://genome.jgi-psf.org/ciona4/ciona4.home.html>).

5.4.4 Intronpositionen der Serpin-Gene von *Ciona intestinalis*

Die Intronpositionen und Intronphasen (a, b, c) der Serpin-Gene von *Ciona intestinalis* wurden mit Hilfe des Programms Genewise bestimmt. Auf der Basis dieser genspezifischen Intronpositionen wurden anschließend mit Hilfe eines Aminosäuresequenz-Alignments die relativen Intronpositionen in Bezug auf das Referenz-Serpin α_1 -Antitrypsin (reifes hum-a1AT) ermittelt. Um die Intronpositionen gesichert bestimmen zu können, wurden die Alignments mit zwei verschiedenen Alignment-Programmen (DIALIGN 2.2.1, ClustalX 1.83, siehe Abschnitt 4.7.1) erstellt. In Tabelle 27 sind nur die Intronpositionen innerhalb der konservierten Serpinstruktur (ab Position 52 in Bezug auf hum-a1AT) aufgeführt.

Serpin-Gen	Intronpositionen (in Bezug auf reifes α_1 -Antitrypsin, hum-a1AT)																										
	56b	60a	67a	73b	-b / 83 b	97b	-c / 106c	136b	144c	151c	156b	191c	204c / 203c	217c	225a	245b	246a	247a / -a	259a	274a / -a	283c	285c	292c	301c	332a	339c	344a
<i>Ci-Spn-1</i> (ci0100132788)			+				+				+							+		+							
<i>Ci-Spn-2</i> (ci0100132818)			+				+				+					+											
<i>Ci-Spn-3</i> (ci0100134682)	+					+					+		+						+					+			+
<i>Ci-Spn-4</i> (ci0100141118)				+					+					+		+							+		+		
<i>Ci-Spn-5</i> (ci0100143209)					+				+			+						+			+						+
<i>Ci-Spn-6</i> (ci0100146394A)																											
<i>Ci-Spn-7</i> (ci0100146394B)																											
<i>Ci-Spn-8</i> (ci0100146394C)																											
<i>Ci-Spn-9</i> (ci0100148346)		+					+		+							+						+				+	
<i>Ci-Spn-10</i> (ci0100154072)		+					+		+							+						+				+	

Tabelle 27: Intronpositionen der Serpin-Gene von *Ciona intestinalis*

Intronpositionen, die aufgrund von Indels nicht eindeutig bestimmbar waren, sind hellgrau markiert. Sind für ein Intron zwei Positionen aufgeführt, so steht die erste Zahl für die Position im DIALIGN-Alignment und die zweite Zahl für die Position im ClustalX-Alignment. Ein Bindestrich steht für Intronpositionen, die aufgrund einer Insertion im Serpin (im Vergleich zum reifen hum-a1AT) von einem der beiden verwendeten Alignment-Programmen als nicht alignierbar dargestellt wurde.

Die Serpin-Gene von *Ciona intestinalis* weisen sehr heterogene Exon-Intron-Strukturen auf. Eine Ausnahme bilden die Gene *Ci-Spn-9* (ci0100148346) und *Ci-Spn-10* (ci0100154072), die eine identische Exon-Intron-Struktur zeigen, sowie die drei intronfreien Gene *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8*, die einen Gencluster bilden (siehe Diskussion 6.5.1). Von den 27 Intronpositionen der *Ciona*-Serpine sind nur vier Positionen identisch mit den Serpin-Genen der Vertebraten. Folgende Intronpositionen der *Ciona*-Serpine stimmen mit Intronpositionen von Serpin-Genen anderer Organismen überein, Beispiele sind in Klammern angegeben:

- Ci-Spn-1, 2:*** **67a** (humanes α_2 -Antiplasmin, humanes PEDF, humaner C1-Inhibitor)¹,
Ci-Spn-1, 2, 5: **191c** (*Ms*-Alaserpin, *Manduca sexta*; humanes ATIII)¹,
Ci-Spn-4: **217c** (*Bm*-SPN-2, *Brugia malayi*)²,
Ci-Spn-9, 10: **225a** (murines HSP47),
Ci-Spn-9, 10: **285c** (*Ce*-Srp-7, *Ce*-Srp-8; *Bm*-Spn-2),
Ci-Spn-5, 9, 10: **339c** (*Ce*-Srp-7; *Bm*-Spn-2, humanes ATIII).

5.4.5 Phylogenetische Analyse der Serpine von *Ciona intestinalis*

Die korrigierten Aminosäuresequenzen der *Ciona*-Serpine wurden mit DIALIGN aligniert. Das Alignment wurde mit GeneDoc überprüft und optimiert. Die hoch divergente RCL-Region wurde strikt linksbündig ausgerichtet. Zur Darstellung des Alignments wurde ESPript 2.0 verwendet (siehe Anhang, Abbildung 48). Alle Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (reifes hum-a1AT).

Die *Ciona*-Serpine weisen im Vergleich zum α_1 -Antitrypsin nur kleinere Indels (< 11 As) auf. Diese Indels sind hauptsächlich zwischen den Sekundärstruktur-Elementen lokalisiert. Die Serpine *Ci-Spn-1* und *Ci-Spn-2* enthalten jeweils zwischen den Sekundärstruktur-Elementen hH und s2C (Position ca. 279) eine Insertion von 10 bzw. 9 As. Das Serpin *Ci-Spn-4* weist zwei Insertionen auf. Die erste, 8 As große Insertion ist zwischen den Sekundärstruktur-Elementen hD und s2A lokalisiert (Position 109) und die zweite Insertion ist innerhalb der β -Faltblattstruktur s3C lokalisiert (Position 224).

Die Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* (ci0100146394) weisen einen sehr hohen Grad an Aminosäuresequenz-Identität untereinander auf, der besonders im aminoterminalen Teil

¹ Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

² Zang, X., Yazdanbakhsh, M., Jiang, H., Kanost, M., R., Maizels, R., M., 1999.

der Proteine ausgeprägt ist. Interessanterweise zeigen diese drei Serpine eine deutliche Variabilität im spezifitäts-bestimmenden, carboxyterminalen Teil der RCLs (Positionen P6-P1, siehe Diskussion 6.5).

Alle *Ciona*-Serpine weisen RCL-Größen auf, die für inhibitorische Serpine charakteristisch sind (Abschnitt 2.1.4). Die RCL-Positionen P15-P8 der Serpine *Ci-Spn-1* - *Ci-Spn-10A* zeigen eine für inhibitorische Serpine charakteristische Aminosäuresequenz, die auf eine Protease-Inhibierungsaktivität hinweist (Abschnitt 2.1.4).

Eine Ausnahme stellt die Protein-Variante *Ci-Spn-10B* dar (Abbildung 40). *Ci-Spn-10B* enthält an den Positionen P10-P7 sperrige Aminosäurereste („PQE“ für Prolin, Glutamin, Glutamat), die auf eine zumindest eingeschränkte Protease-Inhibierungsaktivität hinweisen.

Das *Ciona*-Alignment (Abbildung 48) diente zur Konstruktion eines Intraspezies-Stammbaums (Abbildung 29). Als Außengruppe wurde humanes α_1 -Antitrypsin gewählt, das bereits im Alignment als Referenz-Serpin diente.

Als Baum-Rekonstruktionsmethode wurde hier das Maximum-Likelihood-Verfahren (Abschnitt 4.7.2) eingesetzt. Die ML-Analyse wurde mit Tree Puzzle 5.0¹ durchgeführt. Anstatt des klassischen, externen Datensatz-*Bootstraps* wurde der methoden-interne „*Bootstrap*“-Wert, der *reliability value*, als Unterstützungswert der OTU-Gruppierungen angegeben (Abschnitt 4.7.5). Zur Konstruktion des ML-Stammbaums mit Tree Puzzle wurden abweichend von der Grundeinstellung folgende Parameter gewählt.

- 1000 Puzzle-Schritte
- Exakte Parameterschätzung (*quartet sampling* + *NJ tree*)
- JTT-Substitutionsmatrix (Jones *et al.* 1992)

Der analysierte Datensatz weist ca. 5,5 % ungelöste Quartette auf (*unresolved quartets*), wodurch eine *Multifurcation* in der Topologie auftritt. Im dargestellten ML-Baum (siehe Abbildung 29) können folgende Gruppierungen der *Ciona*-Serpine beobachtet werden (siehe Diskussion, Abschnitt 6.5):

Der größte Cluster von sechs Serpinen wird durch *Ci-Spn-5*, *Ci-Spn-6*, *Ci-Spn-7*, *Ci-Spn-8*, *Ci-Spn-9* und *Ci-Spn-10A* gebildet. Innerhalb dieses Clusters treten zwei Gruppierungen auf, die hohe Unterstützungswerte (> 95 %) aufweisen. Die Serpine *Ci-Spn-9* und *Ci-Spn-10A* werden mit einem Unterstützungswert von 100 % zusammen gruppiert und verfügen auch über identische Exon-Intron-Strukturen. Die drei Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8*

¹ Strimmer, K., von Haeseler, A., 1996.

sind zusammen in einem Gencluster organisiert, enthalten keine Introns und werden mit Unterstützungswerten von 96 % bzw. 100 % im Aminosäuresequenz-Stammbaum zusammen gruppiert.

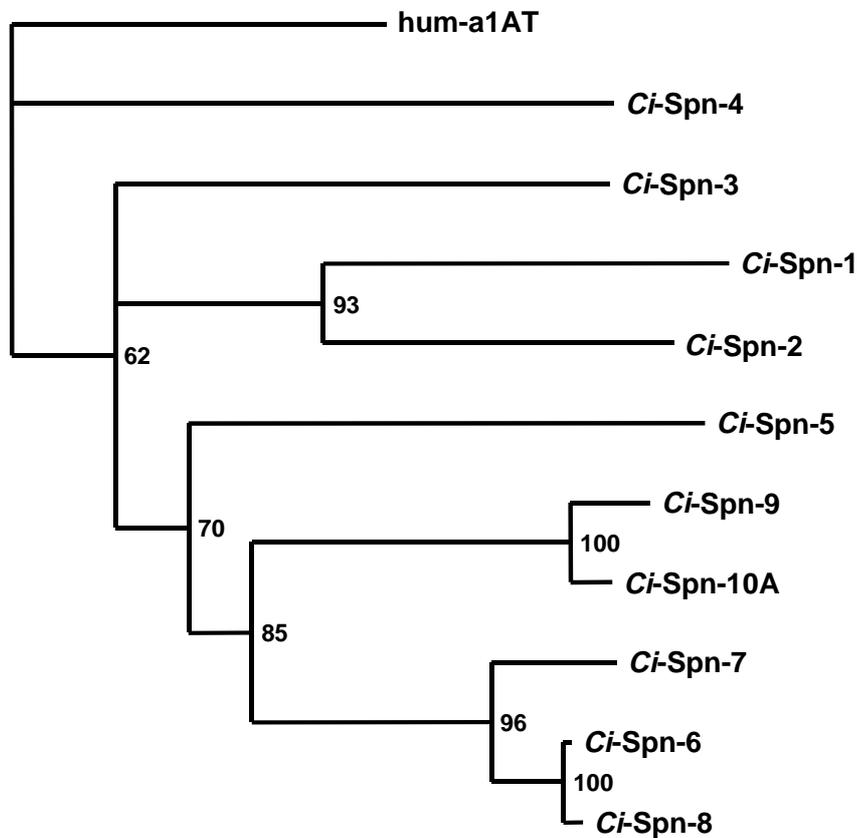


Abbildung 29: Maximum-Likelihood-Baum der Serpine von *Ciona intestinalis*

Das Phylogramm wurde mit Tree Puzzle 5.0 erstellt und basiert auf der JTT-Substitutionsmatrix (Jones-Taylor-Thornton)¹. Der Baum wurde mit humanem α 1-Antitrypsin (hum-a1AT) gewurzelt. Die Unterstützungswerte (*reliability values*, 1000 puzzle steps) sind jeweils rechts vom zugehörigen Knoten angegeben. Zur Darstellung des Baums wurde TreeView² verwendet.

Außerhalb dieses großen Clusters von sechs Serpinen, werden *Ci-Spn-1* und *Ci-Spn-2* mit einem Unterstützungswert von 93 % zusammen gruppiert. Beide Serpine weisen eine gemeinsame Insertion (Position 279) auf und enthalten jeweils die Intronpositionen 67a, 136b und 191c. Die dargestellte Baum-Topologie erhält man auch bei Anwendung anderer Baumrekonstruktions-Methoden (z.B. NJ, Ergebnisse nicht gezeigt) oder Variationen der Aminosäure-Substitutionsmatrix.

¹ Jones, D., T., *et al.* 1992

² Page, R., D., M., 1996.

5.5 Interspezies-Stammbäume

Anhand der Interspezies-Stammbäume sollte der Frage nachgegangen werden, inwieweit die Gruppierungen der einzelnen Spezies-Datensätze im Gesamt-Datensatz erhalten bleiben.

Die überprüften und korrigierten Aminosäuresequenzen von Serpinen aus den vier analysierten Genomen (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana* und *Ciona intestinalis*) wurden mit DIALIGN 2.2.1 aligniert. Als Referenz-Serpine dienten die in Tabelle 28 aufgeführten Vertebraten-Serpine, die die sechs Serpin-Genfamilien der Vertebraten vertreten¹ (siehe Abschnitt 2.2).

Gruppe	Vertreter der Serpin-Genfamilie
1	<i>plasminogen activator inhibitor-2</i> (hum-PAI-2)
2	<i>α_1-antitrypsin</i> (hum-a1AT) <i>heparin cofactor II</i> (hum-HCII)
3	<i>plasminogen activator inhibitor-1</i> (hum-PAI-1) <i>nexin-1</i> (hum-NEXIN1) <i>neuroserpin</i> (hum-NEUROS)
4	<i>pigment epithelium-derived factor</i> (hum-PEDF) <i>α_2-antiplasmin</i> (hum-A2ANTIPLAS)
5	<i>antithrombin</i> (hum-ATIII)
6	<i>heat shock protein 47</i> (hum-HSP47)

Tabelle 28: Referenz-Serpine für die Interspezies-Stammbäume

Das Alignment umfasst insgesamt 65 Serpin-Sequenzen. Insertionen und Extensionen, die länger als 20 As sind und nur in einer Sequenz auftreten, wurden mit GeneDoc entfernt. Das Alignment (nicht dargestellt) diente zur Konstruktion der Interspezies-Stammbäume (Abbildung 30 - Abbildung 32). Es wurde keine formale Außengruppe gewählt. Die Interspezies-Stammbäume wurden mit Hilfe der Verfahren Neighbor-Joining (Abschnitt 4.7.3, Abbildung 30), Maximum-Likelihood (Abschnitt 4.7.2, Abbildung 31) und Maximum-Parsimony (Abschnitt 4.7.4, Abbildung 32) erstellt. Die gewählten Parameter sind den Beschriftungen der Abbildungen zu entnehmen.

¹ Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

Im **Neighbor-Joining-Baum** (Abbildung 30) können vier Serpin-Cluster identifiziert werden. Die Serpin-Sequenzen von *Arabidopsis thaliana* und *Caenorhabditis elegans* bilden jeweils gut abtrennbare Gruppierungen (durch Klammern hervorgehoben), die innerhalb dieser Cluster sehr hohe Unterstützungswerte aufweisen. Die sehr heterogenen Serpin-Sequenzen von *Drosophila* bilden ebenfalls eine Gruppierung. Jedoch werden die drei *Drosophila*-Serpine *Dm-Nec*, *Dm-SPN27a* und *Dm-CG7219* außerhalb dieses Clusters positioniert. *Dm-Nec* und *Dm-SPN27a* werden zusammen mit dem *Ciona*-Serpine *Ci-Spn-4* gruppiert (Unterstützungswert $< 50\%$). Diese Gruppe nimmt eine Zwischenstellung ein. Sie kann entweder den *Ciona*/Vertebraten-Serpinen oder den Serpinen der Invertebraten zugerechnet werden. Dagegen wird das *Drosophila*-Serpine *Dm-CG7219* (markiert durch einen Pfeil) mit einer *Bootstrap*-Unterstützung von 58% zusammen mit Serpinen *Ciona intestinalis* gruppiert (siehe Diskussion Abschnitt 6.6).

Die *Ciona*-Serpine bilden keine klar abtrennbare Gruppierung, sondern werden zusammen mit Vertebraten-Serpinen (Referenz-Serpine für die sechs Genfamilien der Vertebraten) positioniert (oberhalb der gestrichelten Linie). So werden die intronlosen Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* zusammen mit *hum-PAI-2* (Gruppe 1 der Vertebraten-Genfamilie) und *Ci-Spn-9* und *Ci-Spn-10A* zusammen mit *hum-ATIII* (Gruppe 5 der Vertebraten-Genfamilie) gruppiert (siehe Diskussion 6.6). Es muss einschränkend erwähnt werden, dass ein Großteil der Knoten im NJ-Baum mit *Bootstrap*-Werten $< 50\%$ (nicht beschriftete Knoten) keine belastbaren Aussagen erlaubt.

Im **Maximum-Likelihood-Baum** (Abbildung 31) tritt eine starke *Multifurcation* auf (20% *unresolved quartets*). Es können nur zwei klare Gruppierungen identifiziert werden. Die Serpin-Sequenzen von *Arabidopsis thaliana* und *Caenorhabditis elegans* bilden jeweils scharf abtrennbare Gruppierungen (durch Klammern hervorgehoben), die innerhalb der jeweiligen Gruppierung hohe Unterstützungswerte aufweisen. Neben diesen beiden Hauptclustern können noch kleinere Gruppen identifiziert werden, die auch in den jeweiligen Intraspezies-Serpine-Stammbäumen auftreten. Interessant ist, dass die gemeinsame Gruppierung der *Ciona*-Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* zusammen mit dem Vertebraten-Serpine *hum-PAI-2* (siehe NJ-Baum, Abbildung 30) im ML-Baum ebenfalls auftritt. Die *Multifurcation*, die durch *reliability values* $< 50\%$ bedingt wird, schränken die Aussage des ML-Baums auf die belastbaren Anteile der Topologie ein.

Im **Maximum-Parsimony-Baum** (Abbildung 32) können die Serpine von *Caenorhabditis elegans* und *Arabidopsis thaliana* jeweils als Hauptgruppierung identifiziert werden, die größtenteils durch hohe *Bootstrap*-Werte unterstützt werden (durch Klammern hervorgehoben). Die Serpine von *Drosophila melanogaster* werden dagegen in zwei Cluster getrennt. Der kleinere Cluster (*Dm-Spn27a*, *Dm-Spn5*, *Dm-CG6687*, *Dm-CG6680*, *Dm-CG6663* und *Dm-CG6289*) wird dabei in der Nähe der *Arabidopsis*- und *C. elegans*-Cluster positioniert (siehe Diskussion 6.6). Wie auch schon beim NJ-Baum beobachtet werden konnte, wird *Dm-CG7219* (markiert durch einen Pfeil) zusammen mit Serpinen von Vertebraten bzw. *Ciona intestinalis* gruppiert. Die Serpine von *Ciona intestinalis* werden im MP-Baum ebenfalls in zwei Gruppen getrennt, wobei sie in beiden Fällen zusammen mit Vertebraten-Serpinen positioniert werden. Wie im NJ-Baum werden die Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* zusammen mit *hum-PAI-2* und *Ci-Spn-9* und *Ci-Spn-10A* zusammen mit *hum-ATIII* gruppiert. Jedoch sind die Unterstützungswerte für diese beiden Gruppierungen mit 38 und 33 deutlich geringer als im NJ-Baum (siehe Diskussion 6.6). Unter den *Ciona*-Serpinen nimmt *Ci-Spn-4*, das zusammen mit den Serpinen von *Arabidopsis thaliana* lokalisiert wird, eine Außenseiterposition ein. Auch hier muss einschränkend erwähnt werden, dass ein Großteil der Knoten im MP-Baum mit *Bootstrap*-Werten $< 50\%$ nicht belastbar ist.

Die hier erhaltenen Gruppierungen basieren auf den Aminosäuresequenzen der Serpine. Im Abschnitt 6.6 (Diskussion) werden die Serpin-Gruppierungen aller verwendeten phylogenetischen Kriterien, „chromosomalen Lokalisationen“, „Genstrukturen“ und „Aminosäuresequenzen“ bzw. die davon abgeleiteten Stammbäume miteinander verglichen.

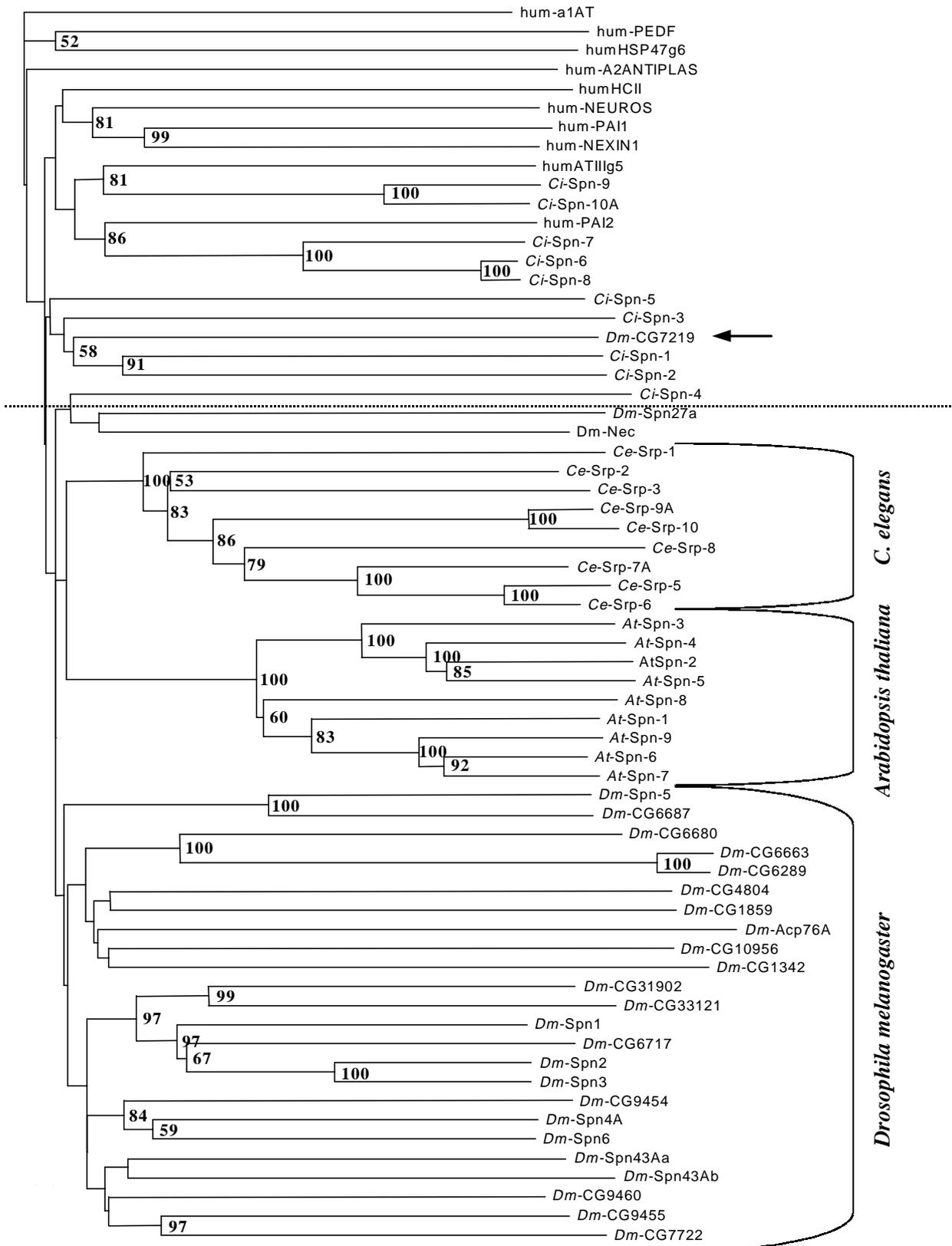


Abbildung 30: Neighbor-Joining-Baum von 65 Serpin-Sequenzen

Das Phylogramm wurde mit PAUP*4.0b10 (Minimum Evolution, Distanz-Kriterium) erstellt und basiert auf der PAM-Substitutionsmatrix. Die *Bootstrap*-Werte (> 50 %, 1000 Replikate) sind jeweils rechts vom zugehörigen Knoten angegeben. Zur Darstellung des Baums wurde TreeView verwendet.

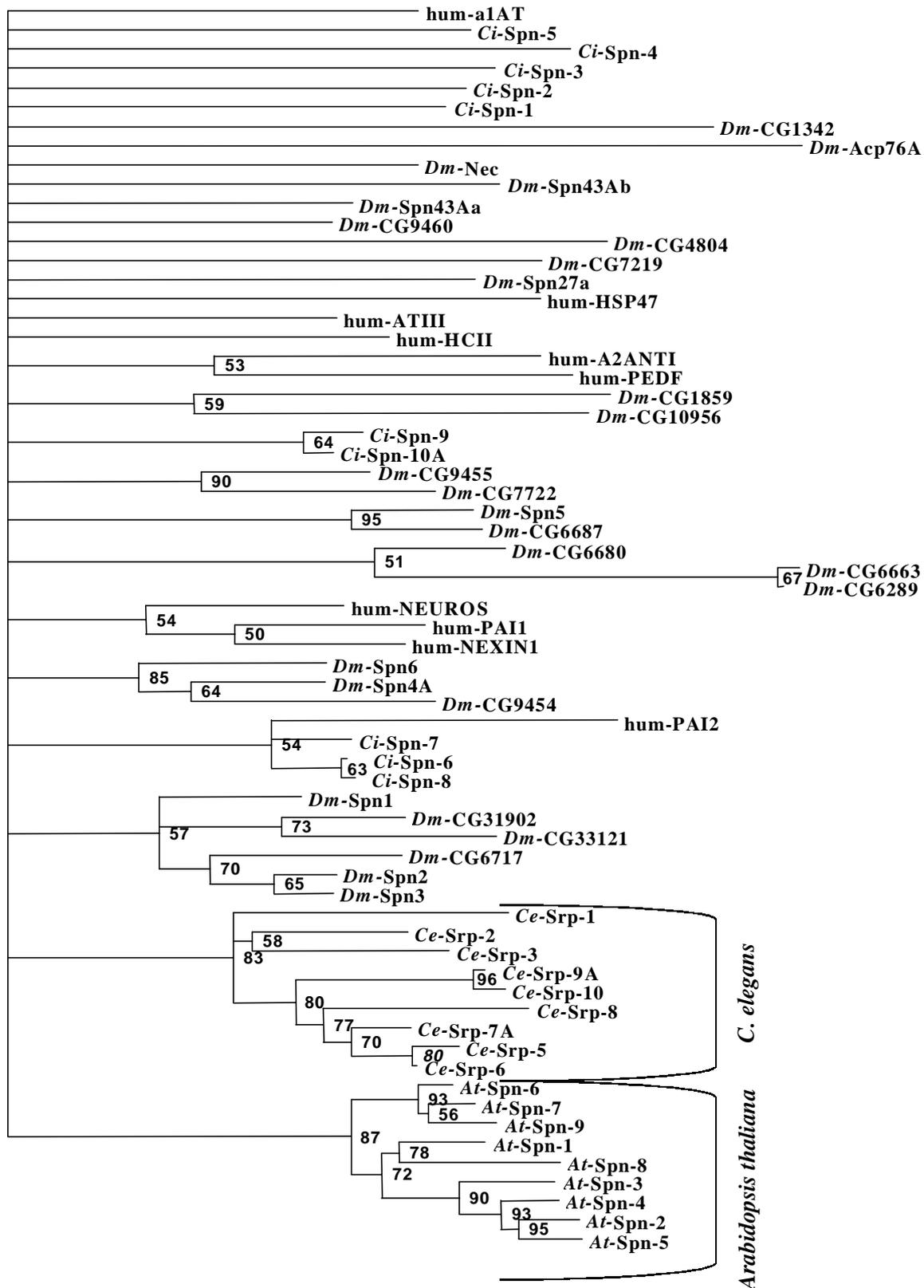


Abbildung 31: Maximum-Likelihood-Baum von 65 Serpin-Sequenzen

Das Phylogramm wurde mit Tree Puzzle erstellt und basiert auf der PAM-Substitutionsmatrix. Die *reliability values* (> 50 %, 10000 *puzzle steps*) sind jeweils rechts vom zugehörigen Knoten angegeben. Zur Darstellung des Baums wurde TreeView verwendet.

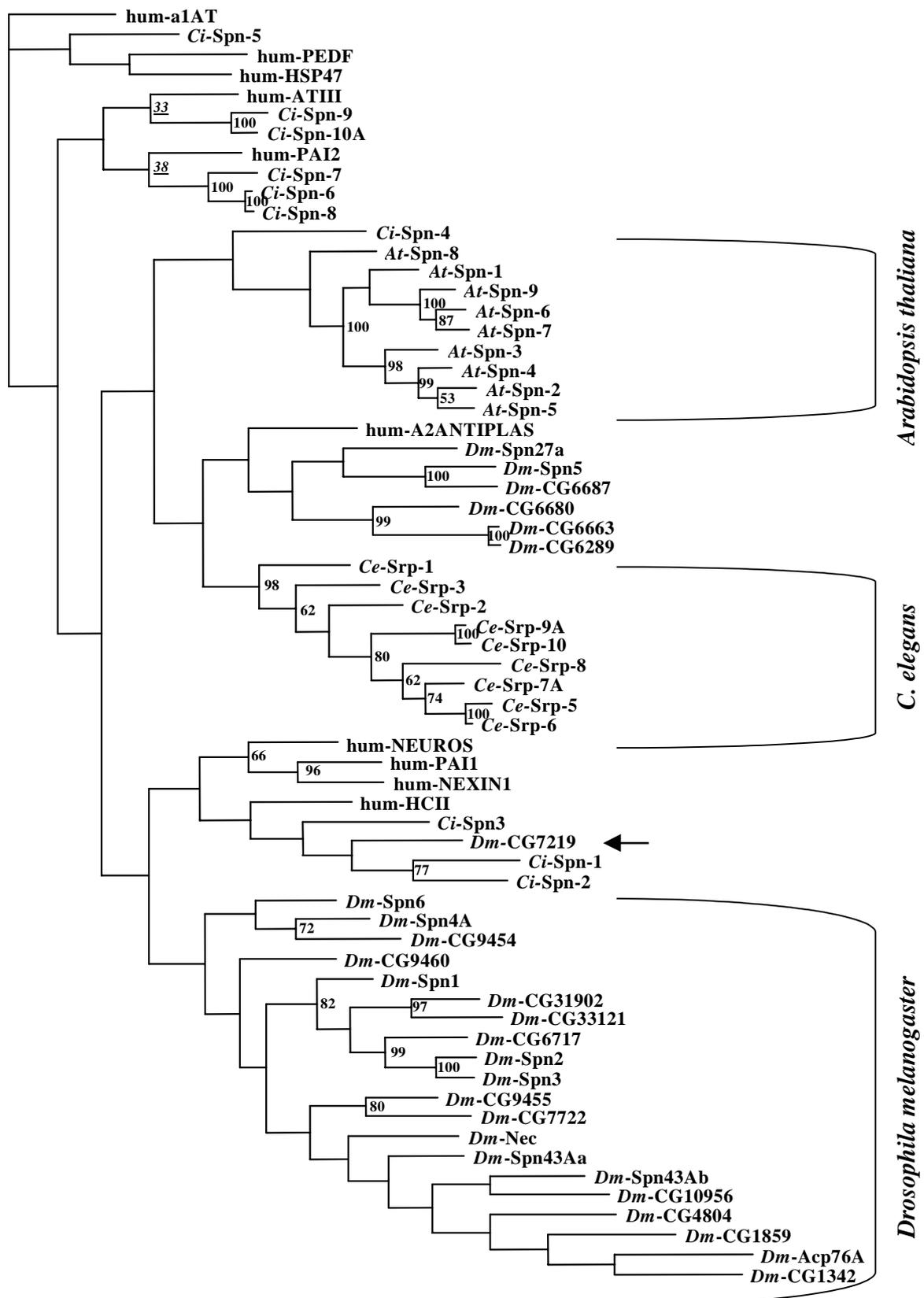


Abbildung 32: Maximum-Parsimony-Baum von 65 Serpin-Sequenzen

Das Phylogramm wurde mit PAUP*4.0b10 (Parsimony-Kriterium, heuristischen Suche, TBR Branch-swapping Algorithmus) erstellt und basiert auf der PAM-Substitutionsmatrix. Die *Bootstrap*-Werte (> 50 %, 100 Replikate) sind jeweils rechts vom zugehörigen Knoten angegeben. Zur Darstellung des Baums wurde TreeView verwendet.

5.6 Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten

Um die große Zahl der Serin- und Cystein-Proteasen zu kontrollieren, scheint die Natur zwei Strategien zur Steigerung der funktionellen Vielfalt der Serpine zu verfolgen. Die am weitesten verbreitete evolutive Strategie besteht aus Gen-Duplikation und anschließender Gen-Diversifikation. Eine alternative, konzentriertere Strategie ermöglicht der Mechanismus des alternativen Spleißens. So konnten Jiang *et al.* in *Manduca sexta* (Tabakswärmer) das Serpin-Gen *Ms-Serpin-Gene-1* identifizieren, das über 12 RCL-kodierende Exon-Varianten verfügt (Abschnitt 5.6.4)¹. Durch alternatives Spleißen wird hierbei eine der 12 Exon-Varianten an das allen Varianten gemeinsame Serpingrundgerüst angefügt (Abbildung 39). Anstatt auf das gesamte Gen beschränkt sich der Duplikations- und Diversifikationsprozess auf das RCL-kodierende Exon, das die Target-Spezifität des Serpins bestimmt. Als Konsequenz dieser Gensequenz-sparenden Strategie kodiert das Gen *Ms-Serpin-Gene-1* potentiell 12 unterschiedliche Protease-Spezifitäten.

Die in dieser Arbeit beschriebenen Analysen der Serpin-Gene in den eukaryotischen Genomprojekten führten zur Identifizierung von weiteren Serpin-Genen mit multiplen Varianten des RCL-kodierenden Exons. Mit der AGREP-Suchstrategie (Abschnitt 4.6.1) konnten im Genom von *Drosophila melanogaster* für das Serpin-Gen *Dm-Spn4* insgesamt vier RCL-kodierende Exon-Varianten identifiziert werden. Im Rahmen dieser Arbeit wurde die Expression des Serpin-Gens *Dm-Spn4* experimentell untersucht² (Abschnitt 5.6.1). Auch im Genom von *Caenorhabditis elegans* (Nematode) gelang die Identifikation von zwei Serpin-Genen *Ce-Srp-7* und *Ce-Srp-9*, die beide über multiple RCL-kodierende Exons verfügen (Abschnitt 5.6.2). Im Genom von *Ciona intestinalis* (Chordata) konnte für Serpin-Gen *Ci-Spn-10* (ci0100154072) eine zweite RCL-kodierende Exon-Variante identifiziert werden (Abschnitt 5.6.3). Zum Vergleich sind weitere bekannte Insekten-Serpine, die über multiple RCL-kodierende Exon-Varianten verfügen, in Abbildung 38 (*Anopheles gambia* Serpin SRPN-10) und Abbildung 39 (*Manduca sexta* Serpin-Gene-1) aufgeführt. Das mutmaßlich erste Serpin-Gen mit multiplen RCL-Exons stellt vermutlich sw-AchyI/II dar („Antichymotrypsin“ der Seidenspinnerraupe *Bombyx mori*)³. Ein experimenteller Beweis steht jedoch noch aus.

¹ Jiang, H., Wang, Y., Kanost, M., R., 1994.

² Krueger O., Ladewig J., Koester K., Ragg H., 2002.

³ Sasaki, T., 1991.

5.6.1 Experimentelle Untersuchungen zur Expression von *Dm-Spn4*

Han¹ *et al.* identifizierten die cDNA für das *Drosophila*-Serpine *Dm-Spn4* (sp4) im Rahmen einer Suche nach Serpinen, die potentiell in *Drosophila melanogaster* an den Prozessen Oogenese und Embryogenese beteiligt sind. Für *Dm-Spn4* wurde eine Exon-Intron-Struktur aus drei Exons und 2 Introns postuliert (Exons 3, 4 und 5, siehe Abbildung 35).

Eine eigene Überprüfung (Abschnitt 5.2.2) der annotierten Exon-Intron-Struktur von *Dm-Spn4* führte zur Identifizierung von drei weiteren potentiellen Varianten des RCL-kodierenden Exons (Exons 6, 7 und 8, siehe Abbildung 35).

Die *open reading frames* (ORF) dieser drei Exon-Varianten weisen Aminosäuresequenzen auf, die für Serpine charakteristisch sind. Eine Suche nach zugehörigen 3'-Spleißsignalen auf DNA-Ebene führte in allen drei Fällen zur Identifizierung einer pyrimidin-reichen DNA-Sequenz, die mit einer potentiellen kanonischen 3'-Spleißstelle („AG“) endet. Die vorhandenen Spleißsignale und die stark konservierten, für Serpine charakteristischen Aminosäuresequenzen weisen auf funktionsfähige RCL-kodierende Exon-Varianten hin. Die Schlussfolgerungen, die sich aus der Analyse des Serpin-Gens *Dm-Spn4* ergaben, wurden anschließend experimentell überprüft. Die genomische DNA-Sequenz des Gens *Dm-Spn4* ist im Anhang (Abschnitt 8.5.2) und die Exon-Intron-Struktur in Abbildung 35 aufgeführt.

Nachweis der Transkription der vier RCL-kodierenden Exon-Varianten

Zum Nachweis der Expression bzw. Transkription der vier RCL-kodierenden Exon-Varianten wurde adulte sowie embryonale Poly(A⁺)-RNA von *Drosophila melanogaster* (BD Biosciences Clontech, Heidelberg) als Ausgangsmaterial verwendet.

Zur Durchführung der Reversen Transkription (Abschnitt 4.2.1) wurden jeweils 0,5 µg Template-RNA (adulte / embryonale Poly(A⁺)-RNA) bei einem Reaktionsvolumen von 20 µL eingesetzt. Damit standen die adulten bzw. embryonalen cDNA-Pools als Template für die sich anschließenden analytischen PCR-Experimente zur Verfügung.

Die Ansätze der PCR-Experimente (Abschnitt 4.1.1) sind in Tabelle 29 aufgeführt. Ziel ist es, die *Dm-Spn4*-cDNAs (bzw. mRNAs) nachzuweisen, die die verschiedenen RCL-kodierenden Exon-Varianten beinhalten. Die PCR-Ergebnisse sind in Abbildung 33 dargestellt. Es konnten alle vier PCR-Fragmente mit den erwarteten Größen amplifiziert werden (Tabelle 29). Damit konnte die Expression der vier RCL-kodierenden Exon-Varianten auf Ebene der

¹ Han, J., Zhang, H., Min, G., Kemler, D., Hashimoto, C., 2000.

Transkription bewiesen werden. Die vier RCL-kodierenden Exons sind in Abbildung 35 mit römischen Ziffern markiert (Exons 5 - Exon 8).

Ansatz	Primerkombination Plusstrang / Minusstrang	Beschreibung des PCR-Fragments
1	DM_SP4_1 / DM_SP4_2	DM_SP4_1 bindet im Exon 3 und DM_SP4_2 bindet im Exon 5. Größe des erwarteten PCR-Fragments: 439 Bp
2	DM_SP4_1 / DM_SP4_3	DM_SP4_1 bindet im Exon 3 und DM_SP4_3 bindet im Exon 6. Größe des erwarteten PCR-Fragments: 406 Bp
3	DM_SP4_1 / DM_SP4_4B	DM_SP4_1 bindet im Exon 3 und DM_SP4_4B bindet über die Grenzen von Exon 4 / Exon 7. Größe des erwarteten PCR-Fragments: 271 Bp
4	DM_SP4_1 / DM_SP4_5	DM_SP4_1 bindet im Exon 3 und DM_SP4_5 bindet im Exon 8. Größe des erwarteten PCR-Fragments: 388 Bp

Tabelle 29: Transkriptionsnachweis der vier RCL-kodierenden Exon-Varianten

Die exakten Positionen der Primer im Gen *Dm-Spn4* sind im Materialteil Abschnitt 3.6 angegeben (siehe auch Abbildung 5).

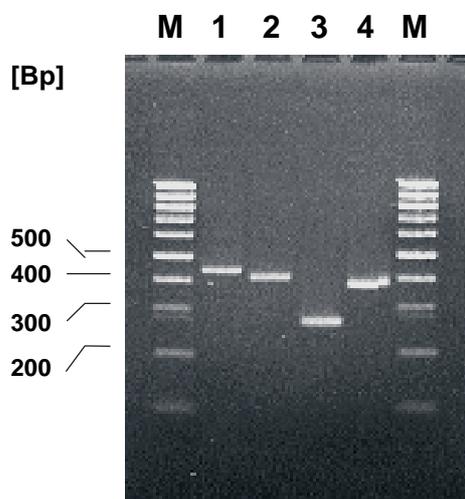


Abbildung 33: Transkriptionsnachweis der vier RCL-kodierenden Exon-Varianten

Gelelektrophoretische Analyse der PCR-Fragmente (2,5 %iges Agarose-Gel). In den Spuren 1 bis 4 sind die PCR-Ansätze 1 bis 4 aufgetragen (Tabelle 29). Als Größenvergleich wurde der Marker „100 Bp DNA Leiter“ aufgetragen (M).

Klonierung der unbekanntenen 3'-UTRs von *Dm-Spn4*

Zur Aufklärung der vollständigen Exon-Intron-Struktur des *Dm-Spn4*-Gens wurden zunächst die unbekanntenen 3'-UTRs per PCR amplifiziert und anschließend in den Vektor pCR-TOPO kloniert.

Als Template stand der bereits präparierte embryonale cDNA-Pool zur Verfügung (Abschnitt 4.2.1). Zur Herstellung des cDNA-Pools wurde der Primer „GeneRacer-Oligo-dT“ von Invitrogen verwendet. Damit stand am 5'-Ende des cDNA-Minusstranges eine bekannte DNA-Sequenz zur Verfügung (Abbildung 6), die eine Amplifikation der unbekanntenen 3'-UTRs durch Kombination eines genspezifischen Primers mit einem genunspezifischen Primer (GeneRacerTM3') erlaubt.

Die Ansätze zur Amplifikation der 3'-UTRs (Abschnitt 4.1.1) sind in Tabelle 30 aufgeführt. Um ein *Dm-Spn4*-spezifisches Fragment zu amplifizieren, wurde anschließend eine seminested PCR (mit dem genunspezifischen Primer GeneRacerTM3'Nested) durchgeführt. Die Ergebnisse der PCRs sind nicht dargestellt. Die amplifizierten Fragmente der seminested-PCR wurden mittels Gelextraktion isoliert (Abschnitt 4.3.2) und mit Hilfe des „TOPO TA Cloning“-Kits (Abschnitt 4.4.1) kloniert.

Ansatz	Primerkombination Plusstrang / Minusstrang	Beschreibung des PCR-Fragments
1	DM_SP4_6 / GR-3'	Plusstrang-Primer DM_SP4_6 (bindet im Exon 5) erlaubt die Amplifikation der 3'-UTR von Variante I (Exon 5).
2	DM_SP4_7 / GR-3'	Plusstrang-Primer DM_SP4_7 (bindet im Exon 6) erlaubt die Amplifikation der 3'-UTR von Variante II (Exon 6).
3	DM_SP4_8 / GR-3'	Plusstrang-Primer DM_SP4_8 (bindet im Exon 7) erlaubt die Amplifikation der 3'-UTR von Variante III (Exon 7).
4	DM_SP4_9B / GR-3'	Plusstrang-Primer DM_SP4_9B (bindet im Exon 8) erlaubt die Amplifikation der 3'-UTR von Variante IV (Exon 8).
5	DM_SP4_6 / GR-3'-nes	seminested-PCR mit Ansatz 1 als Template.
6	DM_SP4_7 / GR-3'-nes	seminested-PCR mit Ansatz 2 als Template.
7	DM_SP4_8 / GR-3'-nes	seminested-PCR mit Ansatz 3 als Template.
8	DM_SP4_9B / GR-3'-nes	seminested-PCR mit Ansatz 4 als Template.

Tabelle 30: Klonierung der unbekanntenen 3'-UTRs von *Dm-Spn4*

Die exakten Positionen der Primer im Gen bzw. in den Transkripten von *Dm-Spn4* sind im Materialteil 3.6 angegeben (siehe Abbildung 5). Als genunspezifische Minusstrang-Primer wurden die Invitrogen-Primer GeneRacerTM3' (GR-3') und GeneRacerTM3'Nested (GR-3'-nes) verwendet.

Die gewachsenen Klone wurden in Klon-Bibliotheken übertragen und anschließend durch Plasmid-Schnell-Charakterisierung (Toothpick-Assay, Abschnitt 4.4.2), Restriktionsspaltung (*EcoRI*, Abschnitt 4.4.3) und Kolonie-PCR (Abschnitt 4.1.2) analysiert. Klone mit erwarteter bzw. interessanter Insertgröße wurden anschließend sequenziert (Abschnitt 4.4.4). Zur vollständigen Sequenzierung der Klon-Inserts wurden IRD800-markierte Standardprimer (T7-Promotor und T3-Promotor) eingesetzt.

Als Resultat konnten fünf *Dm-Spn4*-spezifische 3'-UTR-Sequenzen kloniert und sequenziert werden. Für die RCL-Exon-Variante II konnten zwei 3'-UTRs identifiziert werden, die sich im untranslatierten Bereich durch die Lage eines Introns unterscheiden (Abbildung 35). Ein Vergleich der Sequenzdaten der 3'-UTR-Kloninserts von *Dm-Spn4* mit der veröffentlichten genomischen Sequenz¹ ergab keine Abweichungen.

Klonierung der unbekannt 5'-UTRs von *Dm-Spn4*

Zur Vervollständigung der Exon-Intron-Struktur des *Dm-Spn4*-Gens wurden die bislang unbekannt 5'-UTRs mittels PCR amplifiziert und ebenfalls in den Vektor pCR-TOPO kloniert.

Als Template wurde ein embryonaler cDNA-Pool verwendet, der mit Hilfe des GeneRacerTM-Kits (RLM-RACE-Methode, Abschnitt 4.2.2) hergestellt wurde. Durch die Ligation eines RNA-Oligonukleotids (Sequenz ist bekannt) an die 5'-Enden intakter mRNAs steht nach Reverser Transkription an den 3'-Enden der cDNA-Minusstränge eine bekannte DNA-Sequenz (komplementär zum RNA-Oligonukleotid) zur Verfügung. Mit Hilfe dieser bekannten Sequenz ist eine exponentielle Amplifikation der unbekannt(n) 5'-UTR(s) durch Kombination von genspezifischen Primern mit einem genunspezifischen Primer (Plusstrang-Primer GeneRacerTM5') möglich.

Die Ansätze zur Amplifikation der 5'-UTRs (Abschnitt 4.1.1) sind in Tabelle 31 aufgeführt. Um ein *Dm-Spn4*-spezifisches Fragment zu amplifizieren, wurde anschließend eine seminested-PCR (mit dem genunspezifischen Plusstrang-Primer GeneRacerTM5'Nested) durchgeführt. Die Ergebnisse der PCRs sind in Abbildung 34 dargestellt. Die amplifizierten Fragmente (ca. 310 Bp, 450 Bp, 500 Bp und 570 Bp groß) der seminested-PCRs wurden mittels Gelextraktion isoliert (Abschnitt 4.3.2) und mit Hilfe des „TOPO TA Cloning“-Kits (Abschnitt 4.4.1) kloniert.

¹ FlyBase, <http://www.flybase.org/>, <http://www.wormbase.org>, Release 3.1, Februar 2003

Ansatz	Primerkombination	Beschreibung des PCR-Fragments
	Plusstrang / Minusstrang	
1	DM_SP4_10 / GR-5'	Der Minusstrang-Primer DM_SP4_10 (bindet im Exon 3) erlaubt in Kombination mit dem Plusstrang-Primer GR-5' die Amplifikation der 5'-UTRs.
2	DM_SP4_10 / GR-5'-nes	Der Minusstrang-Primer DM_SP4_10 (bindet im Exon 3) erlaubt in Kombination mit dem Plusstrang-Primer GR-5'-nes die seminested Amplifikation der 5'-UTRs.

Tabelle 31: Klonierung der unbekannt 5'-UTRs von *Dm-Spn4*

Die exakten Positionen der Primer im *Dm-Spn4*-Gen sind im Materialteil 3.6 angegeben (siehe Abbildung 5). Als genunspezifische Minusstrang-Primer wurden die Invitrogen-Primer GeneRacerTM5' (GR-5') und GeneRacerTM5'Nested (GR-5'-nes) verwendet.

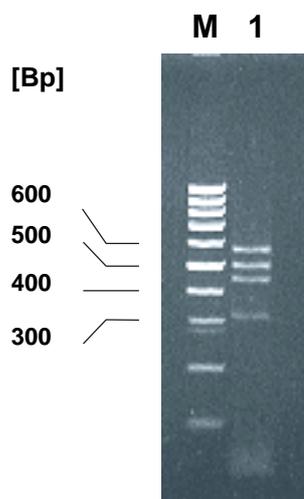


Abbildung 34: PCR-Amplifikation der unbekannt 5'-UTRs von *Dm-Spn4*

Ergebnisse der PCR nach Gelelektrophorese (2,5 %iges Agarose-Gel, Abschnitt 4.3.1). In Spur 1 sind 5 μ L des seminested PCR-Ansatzes zur Amplifikation der 5'-UTRs (Ansatz 2, Tabelle 31) aufgetragen. Als Größenvergleich wurde der Marker „100 Bp DNA Leiter“ aufgetragen (M).

Die gewachsenen Klone wurden in Klon-Bibliotheken übertragen und anschließend durch Plasmid-Schnell-Charakterisierung (Toothpick-Assay, Abschnitt 4.4.2), Restriktionsspaltung (*Eco*RI, Abschnitt 4.4.3) und Kolonie-PCR (Abschnitt 4.1.2) analysiert. Klone mit erwarteter bzw. interessanter Insertgröße wurden anschließend sequenziert (Abschnitt 4.4.4). Zur vollständigen Sequenzierung der Klon-Inserts wurden IRD800-markierte Standardprimer (T7-Promotor und T3-Promotor) eingesetzt.

Als Resultat konnten drei *Dm-Spn4*-spezifische 5'-UTR-Sequenzen kloniert und sequenziert werden (die ca. 450 Bp, 500 Bp und 570 Bp großen PCR-Fragmente). Beim vierten klonierten, ca. 310 Bp großen PCR-Fragment handelte es sich um ein unspezifisches Nebenprodukt. Ein Vergleich der eigenen Sequenzdaten der 3'-UTR-Kloninserts von *Dm-Spn4* mit der veröffentlichten genomischen Sequenz¹ ergab keine Abweichungen in Form von Punkt- oder Rastermutationen.

Transkriptionsanalyse des Gens *Dm-Spn4*

Nach Identifikation von drei unterschiedlichen 5'-cDNA-Enden und fünf unterschiedlichen 3'-cDNA-Enden, wurden in einer PCR-basierten Transkriptionsanalyse die vorhandenen *Spn4*-Transkripte nachgewiesen. Als PCR-Template stand jeweils ein embryonaler und adulter cDNA-Pool zur Verfügung (Abschnitt 4.2.1).

Zum Nachweis eines individuellen *Dm-Spn4*-Transkripts (cDNA-Variante) wurde eine PCR mit transkript-spezifischen Primern durchgeführt. Um eine Amplifikation der genomischen *Dm-Spn4*-Sequenz (eine potentielle Verunreinigung des cDNA-Pools) zu vermeiden, wurden cDNA-spezifische Primer konstruiert, die über Exon-Exongrenzen hinweg reichen (siehe Abbildung 5). Ausnahmen stellen der Minusstrang-Primer DM_SP_4_13 und der Plusstrang-Primer DM_SP4_11 dar. Das Exon 2 liegt in zwei Varianten (2a und 2b) vor. Das Exon 2a umfasst das Exon 2b vollständig und ist zusätzlich um 70 Nt in 5'-Richtung verlängert (siehe Abbildung 35). Dadurch ist es nicht möglich, die Exon-Variante 2b durch PCR-Experimente individuell zu erfassen. Die Transkriptionsanalyse beschränkt sich daher auf die 5'-cDNA-Enden, die die Exons 1 oder 2a beinhalten. Die PCR-Ansätze zur Transkriptionsanalyse (Abschnitt 4.1.1) sind in Tabelle 32 aufgeführt. Die beiden verwendeten PCR-Templates (embryonaler und adulter cDNA-Pool) führten zu nahezu identischen Transkriptionsmustern, die qualitativ gleich sind (Ergebnisse nicht dargestellt).

Die amplifizierten PCR-Fragmente (embryonaler cDNA-Pool) wurden mittels Gelextraktion isoliert (Abschnitt 4.3.2) und anschließend ohne Klonierschritt direkt sequenziert (Abschnitt 4.4.4). Zur vollständigen Direkt-Sequenzierung der jeweils ca. 1300 Bp großen PCR-Fragmente wurden IRD800-markierte, *Dm-Spn4*-spezifische Primer (Materialteil, Abschnitt 3.6; IRD800-Dm_SP4_1+, IRD800-Dm_SP4_3+ und IRD800-Dm_SP4_1-) eingesetzt. Ein Vergleich der erhaltenen Sequenzdaten der PCR-Fragmente mit der veröffentlichten genomischen Sequenz ergab keine Abweichungen.

¹ FlyBase, <http://www.flybase.org/>, <http://www.wormbase.org/>, Release 3.1, Februar 2003

Ansatz	Primer-kombination	Beschreibung der Primer und des PCR-Fragments
1	DM_SP4_12 / DM_SP4_13	Der Plusstrang-Primer DM_SP4_12 (bindet im Exon 1) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_13 (bindet im Exon 5) die Amplifikation eines Fragments von Transkript 1.
2	DM_SP4_12 / DM_SP4_14	Der Plusstrang-Primer DM_SP4_12 (bindet im Exon 1) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_14 (bindet über Exongrenze 6a-7a) die Amplifikation eines Fragments von Transkript 2.
3	DM_SP4_12 / DM_SP4_15	Der Plusstrang-Primer DM_SP4_12 (bindet im Exon 1) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_15 (bindet über Exongrenze 6b-7b) die Amplifikation eines Fragments von Transkript 3.
4	DM_SP4_12 / DM_SP4_16	Der Plusstrang-Primer DM_SP4_12 (bindet im Exon 1) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_16 (bindet über Exongrenze 4-7a) die Amplifikation eines Fragments von Transkript 4.
5	DM_SP4_12 / DM_SP4_17	Der Plusstrang-Primer DM_SP4_12 (bindet im Exon 1) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_17 (bindet über Exongrenze 4-8) die Amplifikation eines Fragments von Transkript 5.
6	DM_SP4_11 / DM_SP4_13	Der Plusstrang-Primer DM_SP4_11 (bindet im Exon 2a) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_13 (bindet im Exon 5) die Amplifikation eines Fragments von Transkript 6.
7	DM_SP4_11 / DM_SP4_14	Der Plusstrang-Primer DM_SP4_11 (bindet im Exon 2a) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_14 (bindet über Exongrenze 6a-7a) die Amplifikation eines Fragments von Transkript 7.
8	DM_SP4_11 / DM_SP4_15	Der Plusstrang-Primer DM_SP4_11 (bindet im Exon 2a) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_15 (bindet über Exongrenze 6b-7b) die Amplifikation eines Fragments von Transkript 8.
9	DM_SP4_11 / DM_SP4_16	Der Plusstrang-Primer DM_SP4_11 (bindet im Exon 2a) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_16 (bindet über Exongrenze 4-7a) die Amplifikation eines Fragments von Transkript 9.
10	DM_SP4_11 / DM_SP4_17	Der Plusstrang-Primer DM_SP4_11 (bindet im Exon 2a) erlaubt in Kombination mit dem Minusstrang-Primer DM_SP4_17 (bindet über Exongrenze 4-8) die Amplifikation eines Fragments von Transkript 10.

Tabelle 32: Transkriptionsanalyse des Gens *Dm-Spn4*

Die exakten Positionen der Primer im *Dm-Spn4*-Gen sind im Materialteil 3.6 angegeben (siehe Abbildung 5).

Durch die Transkriptionsanalyse des Gens *Dm-Spn4* konnten zehn unterschiedliche mRNAs (Transkripte) identifiziert werden, die potentiell acht unterschiedliche Protein-Isoformen kodieren. Die Exon-Intron-Struktur des Gens und die Strukturen der identifizierten Transkripte¹ sind in Abbildung 35 dargestellt.

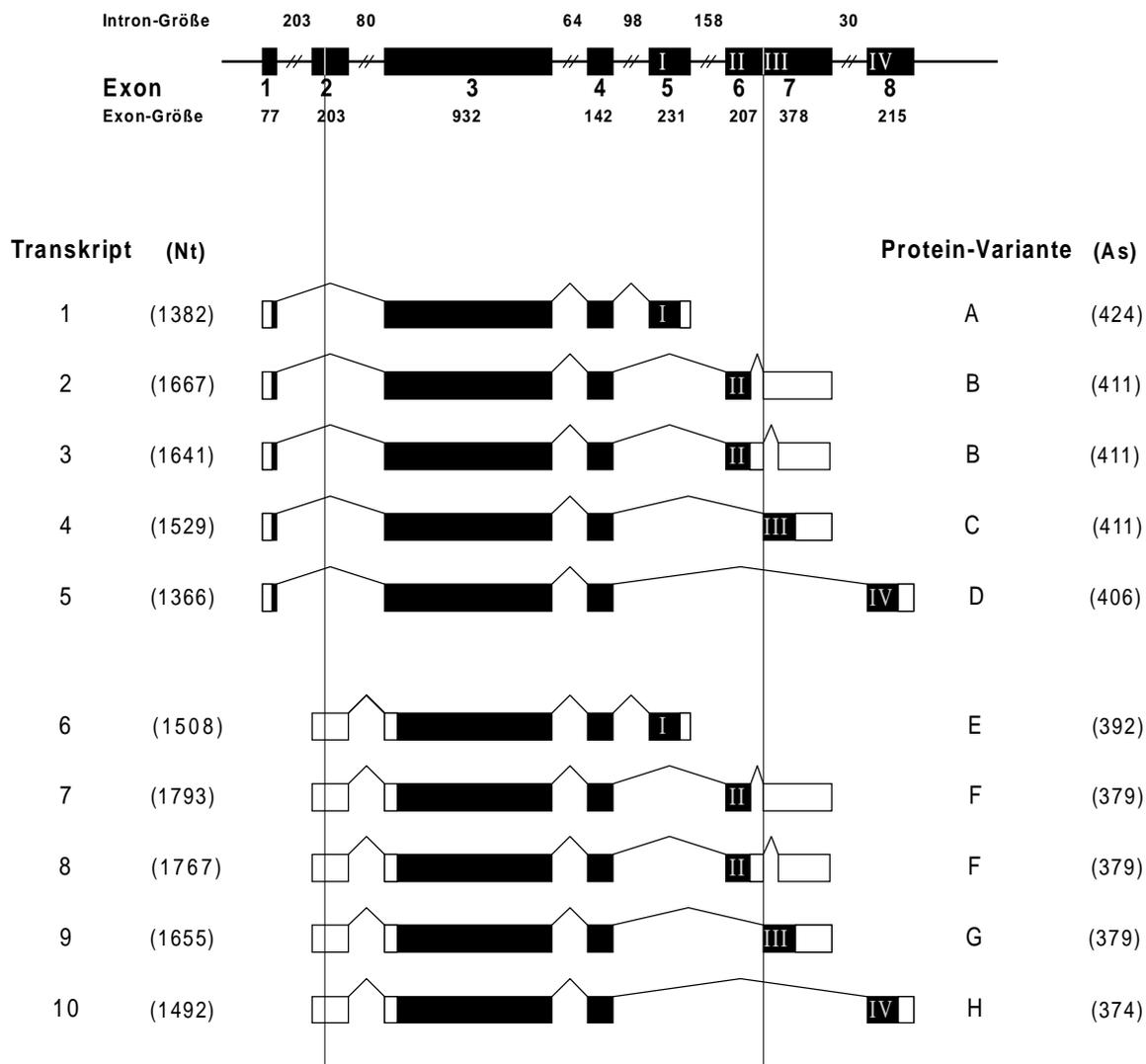


Abbildung 35: Organisation des Gens *Dm-Spn4* von *Drosophila melanogaster*

In der Exon-Intron-Struktur des *Dm-Spn4*-Gens (oben) sind Exons als schwarze Kästchen dargestellt. Die Exon-Nummerierung und Exon-Größenangaben (Nt) stehen unterhalb der Genstruktur. Die vier RCL-kodierenden Exon-Varianten sind mit römischen Ziffern markiert. Die Introns sind als Linien dargestellt. Die Intron-Größenangaben (Nt) sind oberhalb der Genstruktur angegeben. Die untranslatierten Sequenzen der Transkripte sind als weiße Kästchen (5'- und 3'-UTRs) und die Protein-kodierenden Sequenzen als schwarze Kästchen dargestellt. Die Größen der Transkripte und der resultierenden Protein-Varianten sind in Klammern angegeben.

¹ Krueger, O., Ladewig, J., Koester, K., Ragg, H., 2002.

Zelluläre Lokalisation und reaktive Zentren der acht Spn4-Protein-Isoformen

Aus den Sequenzen der zehn identifizierten *Dm*-Spn4-Transkripte können acht unterschiedliche Protein-Isoformen abgeleitet werden (Abbildung 40). In Bezug auf die potentielle zelluläre Lokalisation können die Isoformen in drei Gruppen eingeteilt werden. Die Varianten A-D verfügen über ein potentielles N-terminales Signalpeptid (identifiziert mit Hilfe der Programme PSORT II¹ und TargetP v1.1²). In der Variante A ist dieses Signalpeptid mit einem C-terminalen HDEL-Motiv (Aminosäuren: Histidin, Aspartat, Glutamat, Leucin) kombiniert, das als Rückhaltesignal für das endoplasmatische Retikulum (ER) fungieren kann^{3,4}. Diese Kombination weist darauf hin, dass die Variante A im ER-Zellkompartiment lokalisiert ist, während die Varianten B-D vermutlich aus der Zelle exportiert werden. Dagegen fehlt den Varianten E-H ein N-terminales Signalpeptid (PSORT II, TargetP v1.1), die daher vermutlich im Cytoplasma zu lokalisieren sind.

Hinweise zur Funktion und Protease-Spezifitäten der vier Protein-Varianten liefern die RCL-Strukturen. Die RCL-Struktur inhibitorischer Serpine umfasst gewöhnlich 16 oder 17 As (siehe Alignment in Abbildung 40)⁵. Die für die Inhibierungsreaktion wichtige *Hinge*-Region (P17 - P8) wird durch das gemeinsame Exon 4 kodiert. Die Struktur der *Hinge*-Region zeigt somit einen für inhibitorische Serpine charakteristischen Aufbau. Ein konserviertes Glycin (P15) und eine Abfolge von vier Alaninen sind eine wichtige Voraussetzung, um den durch Protease-Spaltung eingeleiteten β -Faltblatt-Extensionsmechanismus (Abschnitt 2.1.4) zu ermöglichen. Die in Bezug auf die Exon-Varianten (I - IV) unterschiedlichen Bereiche des RCL beginnen mit P7 und umfassen auch die Positionen P1-P1', die einen starken Einfluss auf die Protease-Spezifität von Serpinen ausüben. Die Position P1 der Protein-Variante I ist vermutlich mit Lysin oder Arginin besetzt. Bei den Positionen P1-P1' der Variante II handelt es sich um die Sequenz Alanin-Serin und für die Variante III ist Threonin-Serin als P1-P1'-Bereich wahrscheinlich. Im Falle der Variante IV ist keine eindeutige Bestimmung der Position P1 möglich, vermutlich ist P1 mit einem Alanin oder Valin besetzt.

¹ Nakai, K., Kanehisa, M., 1992.

² Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000.

³ Harter, C., Wieland, F., 1996.

⁴ Denecke, J., De Rycke, R., Botterman, J., 1992.

⁵ Zhou, A., Carrell, R., W., Huntington, J., A., 2001.

5.6.2 Die Serpine *Ce-Srp-7* und *Ce-Srp-9* von *Caenorhabditis elegans*

Im Rahmen der Analyse des Genoms von *Caenorhabditis elegans* (Abschnitt 5.1) konnten zwei Serpin-Gene identifiziert werden, die beide über multiple Varianten des RCL-kodierenden Exons verfügen könnten. Im Falle der Serpine von *Caenorhabditis elegans* muss einschränkend erwähnt werden, dass die „*in silico*“ abgeleiteten Genstrukturen nicht experimentell überprüft wurden. Abbildung 36 stellt die Organisation der Serpin-Gene *Ce-Srp-7* und *Ce-Srp-9* dar.

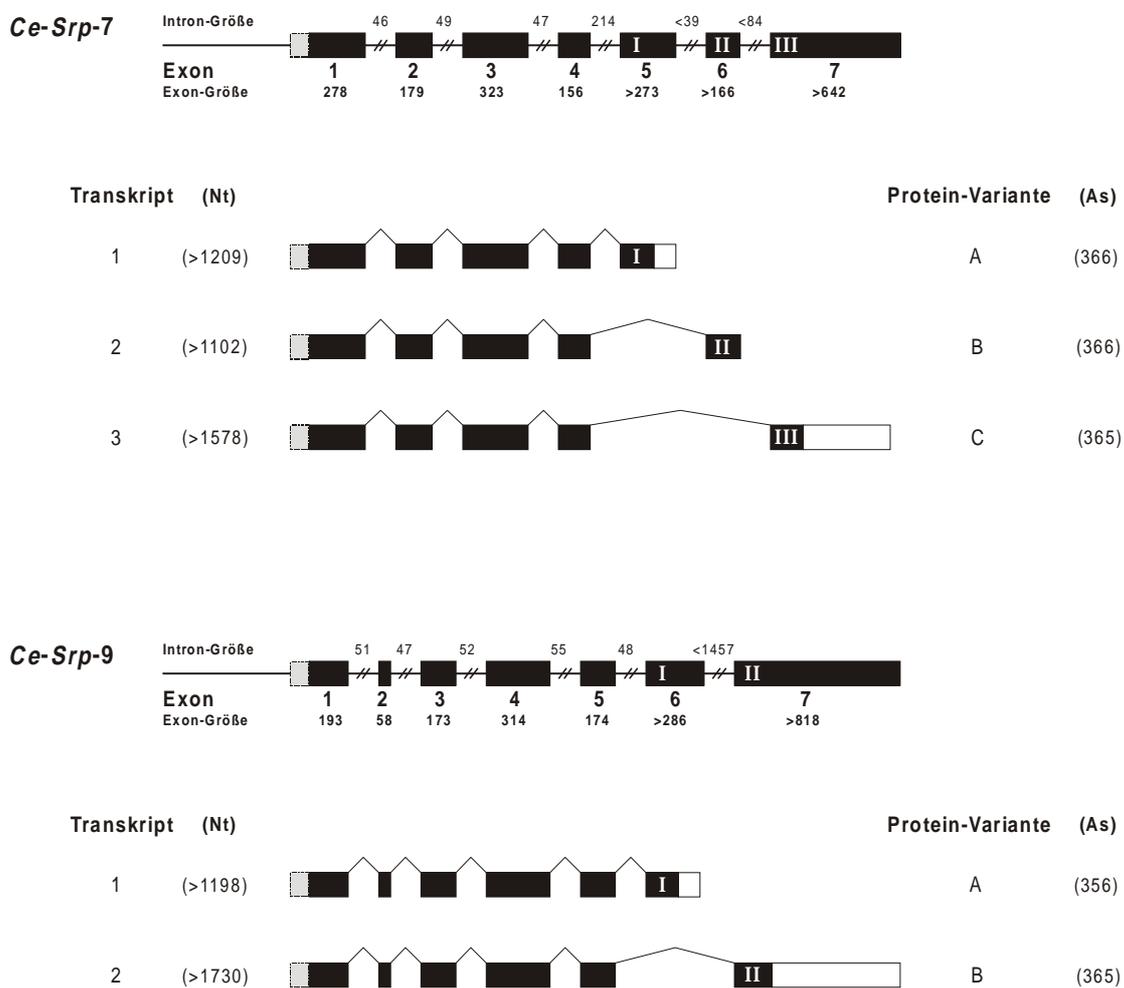


Abbildung 36: Organisation der Serpin-Gene *Ce-Srp-7* und *Ce-Srp-9*

Für die Beschreibung der Darstellung siehe Abbildung 35. Die nicht bekannten 5'-UTRs sind durch graue Kästchen dargestellt. Die 3'-UTRs (weiße Kästchen) umfassen die 3'-untranslatierte Region einschließlich Poly(A)-Signal (AATAAA, ATTTAA). Aufgrund der nicht bekannten Enden der 3'- und 5'-UTRs sind die Mindestgrößen der Transkripte angegeben.

Die Gene *Ce-Srp-7* und *Ce-Srp-9* zeigen eine ähnliche Exon-Intron-Struktur¹. So weisen beide Gene die Intronpositionen 179a/180a, 121b und 285c auf und verwenden die Position 339c als 5'-Spleißstelle zum Anfügen der potentiellen RCL-kodierenden Exon-Varianten (alle Positionen sind in Bezug auf α_1 -Antitrypsin angegeben).

Eine eigene Überprüfung des WormBase-Genmodells von *Ce-Srp-7* führte zur Identifizierung von insgesamt drei potentiellen RCL-kodierenden Exon-Varianten. Die ORFs der Exon-Varianten weisen jeweils Aminosäuresequenzen auf, die für Serpine charakteristisch sind. Eine Suche auf DNA-Ebene führte in allen drei Fällen zur Identifizierung von zugehörigen 3'-Spleißsignalen (pyrimidin-reiche DNA-Sequenz mit einer kanonischen 3'-Spleißstelle). Die ORFs und die vorhandenen Spleißsignale weisen auf funktionsfähige RCL-kodierende Exon-Varianten hin. Jede Exon-Variante verfügt über ein eigenes Polyadenylierungssignal. Im Exon 6 überlappt das potentielle Polyadenylierungssignal mit dem Stopp-Codon und führt so zu einer relativ kurzen 3'-UTR. Exon-Variante III verfügt über insgesamt vier Polyadenylierungssignale, die in 3'-Richtung vom Stopp-Codon in einem 900 Bp großen Bereich lokalisiert werden können. Die durch die Exon-Varianten I und II kodierten RCLs weisen eine hohe Aminosäuresequenz-Identität zueinander auf und scheinen auch über identische P1-P1'-Positionen (Lysin-Serin) zu verfügen. Der durch Exon-Variante III kodierte RCL weist vermutlich die P1-P1'-Sequenz Cystein-Arginin auf (Abbildung 40).

Die Überprüfung des WormBase-Genmodells von *Ce-Srp-9* führte zur Identifizierung von insgesamt zwei potentiellen RCL-kodierenden Exon-Varianten. Auch hier weisen die ORFs der Exon-Varianten jeweils Aminosäuresequenzen auf, die für Serpine charakteristisch sind. Eine Suche auf DNA-Ebene führte in beiden Fällen zur Identifizierung von zugehörigen 3'-Spleißsignalen (pyrimidin-reiche DNA-Sequenz mit einer kanonischen 3'-Spleißstelle). Auch hier weisen die Aminosäuresequenzen der ORFs und die vorhandenen Spleißsignale auf funktionsfähige RCL-kodierende Exon-Varianten hin. Aufgrund der Überprüfung des Genmodells *Ce-Srp-9* konnte auch ein zusätzliches Intron postuliert (ein 48 Bp große Intron an Position 339c, siehe Abschnitt 5.1.2) werden. Beim Gen *Ce-Srp-9* weisen die durch die Exon-Varianten I und II kodierten RCLs eine hohe Aminosäuresequenz-Identität zueinander auf (siehe Abbildung 40). Interessanterweise zeigen nur die spezifitäts-bestimmenden Positionen P5-P1 eine erhöhte Variabilität. Potentielle N-terminale Signalpeptide können für die Protein-Varianten der Gene *Ce-Srp-7* und *Ce-Srp-9* nicht identifiziert werden (TargetP, PSORT II). Es wird daher eine cytoplasmatische Lokalisation angenommen.

¹ Krueger, O., Ladewig, J., Koester, K., Ragg, H., 2002.

5.6.3 Das Serpin-Gen *Ci-Spn-10* von *Ciona intestinalis*

Im Rahmen der Analyse des Genoms von *Ciona intestinalis* (Abschnitt 5.4) konnte für das Serpin-Gen *Ci-Spn-10* (JGI Genmodell ci0100154072) eine zweite RCL-kodierende Exon-Variante identifiziert werden. Mit Hilfe der Ghost Database „*Ciona intestinalis* cDNA resource“¹ bzw. JGI-Datenbank „*Ciona intestinalis* v1.0“² war es möglich EST-Daten zu finden, die die Transkription der verschiedenen Spleißvarianten unterstützen. In Abbildung 37 ist die Organisation des Gens *Ci-Spn-10* dargestellt.

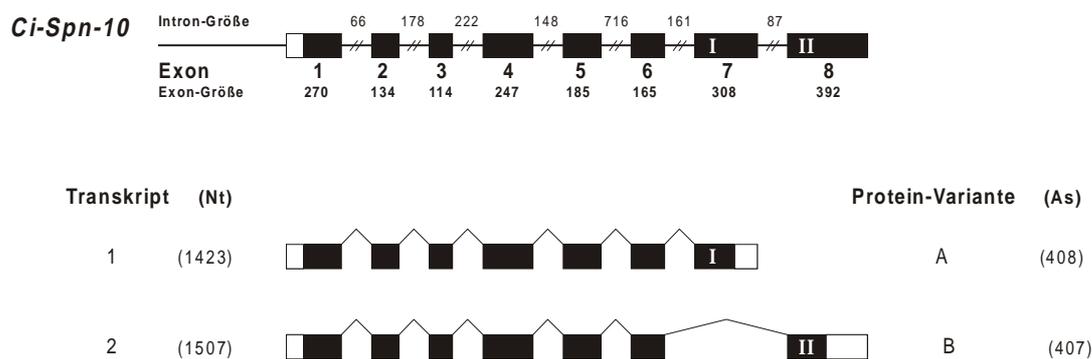


Abbildung 37: Organisation des Gens *Ci-Spn-10* von *Ciona intestinalis*

Die Abbildung basiert auf den EST-Daten² „rcibd064c17“ (Protein-Variante A) und „rciht030i117“ (Protein-Variante B). Für die Beschreibung der Darstellung siehe Abbildung 35.

Die Protein-Varianten *Ci-Spn-10A* und *Ci-Spn-10B* verfügen jeweils über ein potentielles N-terminales Signalpeptid (identifiziert mit Hilfe der Programme PSORT II³ und TargetP v1.1⁴). In der Variante B ist dieses Signalpeptid mit einem C-terminalen HDEL-Motiv (Aminosäuresequenz Histidin-Aspartat-Glutamat-Leucin) kombiniert, das als Rückhaltesignal für das endoplasmatische Retikulum (ER) fungieren kann⁵. Diese Kombination weist darauf hin, dass *Ci-Spn-10B* im ER-Zellkompartiment lokalisiert ist. Die Variante *Ci-Spn-10A* wird vermutlich aufgrund des Signalpeptids sezerniert und liegt somit extrazellulär vor. Möglicherweise handelt es sich aber auch bei der C-terminalen Sequenz HDEF (Histidin-Aspartat-Glutamat-Phenylalanin) von *Ci-Spn-10A* um eine Variation des

¹ Satou, Y., *et al.* 2002.

² Dehal, P., *et al.* 2002. (<http://genome.jgi-psf.org/ciona4/ciona4.home.html>).

³ Nakai, K., Kanehisa, M., 1992.

⁴ Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000.

⁵ Denecke, J., De Rycke, R., Botterman, J., 1992.

ER-Rückhaltesignals. Ob das tatsächlich der Fall ist, kann nur experimentell überprüft werden. Ein Vergleich der durch die Exons I und II kodierten RCLs zeigt eine nur geringe Aminosäuresequenz-Identität (ca. 30 %) zwischen den Varianten *Ci-Spn-10A* und *Ci-Spn-10B*. Interessanterweise ist die Aminosäuresequenz-Identität zwischen den RCLs von *Ci-Spn-9* und *Ci-Spn-10A* mit ca. 80 % sehr hoch (siehe Diskussion Abschnitt 6.7). Die P1-P1'-Positionen sind bei *Ci-Spn-10A* vermutlich mit Arginin-Serin und bei *Ci-Spn-10B* mit Prolin-Alanin besetzt. Die Sequenz der *Hinge*-Regionen läßt im Falle der Variante *Ci-Spn-10A* eine inhibitorische Funktion vermuten, während im Fall von *Ci-Spn-10B* eine protease-inhibierende Funktion nicht wahrscheinlich ist.

5.6.4 Weitere Insekten-Serpine mit multiplen RCL-kodierenden Exon-Varianten

Neben dem *Drosophila melanogaster* Serpin *Dm-Spn4* konnten auch in den Insekten *Anopheles gambia* (SRPN10)² und *Manduca sexta* (Serpine-Gene-1)¹ Serpine identifiziert werden, die über multiple Varianten des RCL-kodierenden Exons verfügen.

Das Serpin-Gen SRPN10 von *Anopheles gambia*

Danielli *et al.* konnten für das *Anopheles gambia* Serpin SRPN10 (*Ag-SRPN10*, GenBank-Eintrag AJ420785 bzw. Spi21F) vier alternative RCL-kodierende Exon-Varianten identifizieren^{2,3}. *Ag-SRPN10* weist eine hohe Aminosäuresequenz-Identität zu *Dm-Spn4* auf, das ebenfalls über vier RCL-kodierenden Exon-Varianten verfügt. In Abbildung 38 werden die Exon-Intron-Strukturen von *Ag-SRPN10* und *Dm-Spn4* verglichen. Die Genstrukturen weisen große Ähnlichkeiten auf, so verfügen beide Gene über die Intronpositionen 305c und 352a (in Bezug auf reifes, humanes α_1 -Antitrypsin).

Die Unterschiede in der Organisation der alternativen RCL-kodierenden Exon-Varianten, resultieren größtenteils aus der Lage der Polyadenylierungssignale. Während im Gen *Ag-SRPN10* die Exon-Varianten I und II ein gemeinsames Polyadenylierungssignal verwenden, wird im Gen *Dm-Spn4* von den Exon-Varianten II und III ein gemeinsames Polyadenylierungssignal genutzt.

¹ Jiang, H., Wang, Y., Huang, Y., Mulnix, A., B., Kadel, J., Cole, K., Kanost, M., R., 1996.

² Danielli, A., Kafatos, F., C., Loukeris, T., G., 2003.

³ Christophides, G., K., *et al.* 2002.

Aufgrund des Fehlens eines Signalpeptid-kodierenden Exons im *Ag-SRPN10* wird derzeit angenommen, dass die vier kodierten Protein-Varianten intrazellulär vorliegen, während die Protein-Varianten von *Dm-SPN4* als Folge der unterschiedlichen Transkriptionsstartpunkte (siehe Abschnitt 5.6.1) im Cytoplasma und im endoplasmatischen Retikulum lokalisiert sind bzw. sezerniert werden.

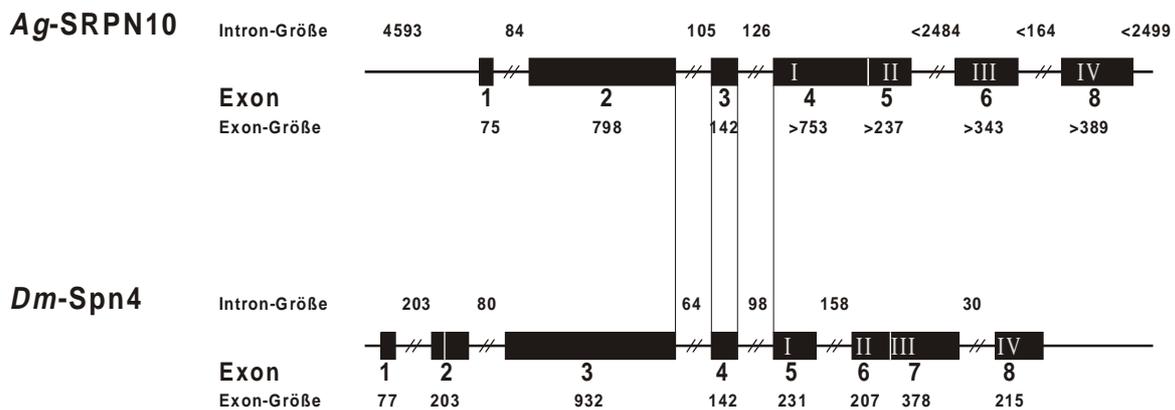


Abbildung 38: Vergleich der Exon-Intron-Struktur von *Ag-SRPN10* und *Dm-Spn4*

Die Abbildung von *Ag-SRPN10* basiert auf dem GenBank-Eintrag AJ420785. Für die Beschreibung der Darstellung siehe Abbildung 35. Die übereinstimmenden Intronpositionen 305c und 352a sind durch Linien dargestellt.

Ein Vergleich der Aminosäuresequenzen der RCL-Varianten zeigt, dass sich die jeweiligen spezifitäts-bestimmenden RCL-Bereiche der Varianten II, III und IV der beiden Gene deutlich voneinander unterscheiden (siehe Alignment Abbildung 40). Dagegen weisen die RCL-Strukturen der jeweiligen Variante I interessante Gemeinsamkeiten auf. So enthalten beide eine potentielle Furin-Spaltstelle (R-X-X-R), wodurch auf eine ähnliche Protease-Spezifität und eine ähnliche physiologische Funktion geschlossen werden könnte. Eine weitere Gemeinsamkeit, sind die jeweiligen carboxyterminalen Sequenz-Motive „HEEL“ bzw. „HDEL“, die als Rückhaltesignale für das ER-Zellkompartiment dienen können.

Die Organisation des Gens Serpin-Gene-1 von *Manduca sexta*

Jiang *et al.* gelang die Aufklärung der Organisation des Serpin-Gene-1 von *Manduca sexta* (*Ms*-Serpin-Gene-1, GenBank-Eintrag U58361)^{1,2}. In Abbildung 39 ist die Organisation des Gens *Ms*-Serpin-Gene-1 dargestellt. Es verfügt über 12 RCL-kodierende Exon-Varianten.

Ms-Serpin-Gene-1

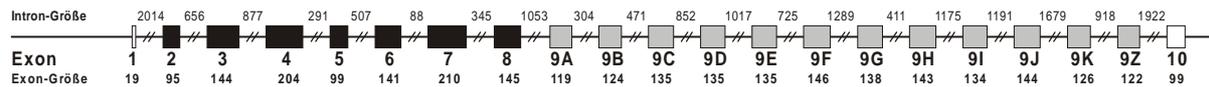


Abbildung 39: Organisation des Gens *Ms*-Serpin-Gene-1

Die Abbildung basiert auf dem GenBank-Eintrag U58361. In Bezug auf die Sequenz U58361 fehlen genflankierende Bereiche (5'-flankierend 862 Bp und 3'-flankierend 132 Bp). Für die Beschreibung der Abbildungselemente siehe Abbildung 35. Die 12 RCL-kodierenden Exon-Varianten sind als graue Kästchen dargestellt (Exons 9A - 9Z).

5.6.5 Alignment der Serpine mit multiplen RCL-kodierenden Exon-Varianten

In Abbildung 40 wurden die Protein-Varianten der Serpin-Gene *Dm*-SPN4, *Ag*-SPRN10, *Ci*-SPN-10, *Ce*-Srp-7 und *Ce*-Srp-9 in einem Alignment zusammengefasst. Als Referenz-Serpin diente reifes, humanes α_1 -Antitrypsin (hum-a1AT). Aus Gründen der Übersichtlichkeit wurde nur eine Variante des *Manduca sexta* Serpins *Ms*-Serpin-Gene-1 (*Ms*-Alaserpin-9B) im Alignment berücksichtigt. Das Alignment wurde mit DIALIGN 2.2.1 durchgeführt und mit GeneDoc überprüft und optimiert. Die hoch divergente RCL-Region wurde strikt linksbündig ausgerichtet. Um Redundanzen zu vermeiden, wurden die Protein-Varianten in konstanten Bereichen des Alignments zusammengefasst. Die Protein-Varianten E-H des Gens *Dm*-Spn4 unterscheiden sich von den Varianten A-D durch Fehlen einer 32 As langen N-terminalen Sequenz (kodiert ein potentielles Signalpeptid). In Abbildung 41 sind die Intronpositionen der Serpin-Gene *Dm*-Spn4, *Ag*-SPRN10, *Ci*-SPN-10, *Ce*-Srp-7 und *Ce*-Srp-9 dargestellt (siehe Diskussion Abschnitt 6.7).

¹ Jiang, H., Wang, Y., Kanost, M., R., 1994.

² Jiang, H., Wang, Y., Huang, Y., Mulnix, A., B., Kadel, J., Cole, K., Kanost, M., R., 1996.

hum-a1AT	-----EDPQGDAAQKTDTSHHDDQHPTFNKITPNLAEFAFSLYRQLAHQS	45
Ms-Alaserpin-9B	-----MKIIMCIFGLAALAMAGETDLQKILRESNDQFTAQMFSEVVKAN	44
Dm-Spn4A, D	MDYRLVPCGCWLLPLLLGLALFPFPVHTADVTMADAAHQEFARRLALFINSVYKGLSGQK	60
Dm-Spn4E, H	-----MADAAHQEFARRLALFINSVYKGLSGQK	28
Ag-SRPN10A, B, C, D	-----MADNSSSLDAQFVSQSNFSATKLYQRISAKH	31
Ci-Spn-10A, B	-----MQFLYAIVMLLVLDANAKIIDTSEHVEKLESEANIEFTLNLYKNLIEGD	48
Ce-Srp-7A, B, C	-----MALLLQSETFDFGLGLLRQQN	20
Ce-Srp-9A, B	-----MAVTPVTRCTYLLRDELDFALKLLHQQD	28
hum-a1AT	NSTNIIFFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNLQ	105
Ms-Alaserpin-9B	PGQNVVLSAFSVLPPLGLQALASVGHSHDELLRALALPNDNVT---KDVFADLNRRGVRA	100
Dm-Spn4A-H	PGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDP----EQIAHSFHQVFLAA	115, 83
Ag-SRPN10A, B, C, D	AGENVVISPFISISACLSLAAMGAGGLTAEQMSVLEFGAPDRK---QTVDNRYRRLMER	87
Ci-Spn-10A, B	PMKNVMFSPVSI SAALAMTHLGAFAKGTAKQIDDAFMFSKIEDGRF-HSAFGLHGLLFDK	107
Ce-Srp-7A, B, C	ISESLAFSPLSIALALSIVHVAAGETRDQIREALVKGSTDEQ---LEQHFANISAAALLA	77
Ce-Srp-9A, B	LTKSFVFSPMISIMAMALAHAGAKDETREEIRNVLLK-----LRE	68
hum-a1AT	PDSQLQLTTGNGLFSLSEGLKLVDFLEDVKKLYHSEAFVNFVGDTEEAKKQINDYVEKGT	165
Ms-Alaserpin-9B	VKG-VDLKMASKIYVAKGLELNDFFAAVSRDVFVSEVQNVDFVKVEAAGAINKWVEDQT	159
Dm-Spn4A-H	YQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRT	175, 143
Ag-SRPN10A, B, C, D	LATDSTVNANKIYVMQNYAVKGAFAIATGSRFRSEAESVNFASAAAANKINGWVEEKT	147
Ci-Spn-10A, B	ASDNVTVKSSNRVFAADKKRKLVEDYKNALT-VYGAKLENVDFKTPSNAVKQINDWASDAT	166
Ce-Srp-7A, B, C	AERGTEVKLANHVFTFRAGFKIKQSYLDDVKKLYNAGASSLDFDNKEATAEAINNFVRENT	137
Ce-Srp-9A, B	TEHGAENVVANRLFIKNGYPKKS-YLKEVQKLYNASASNLNFDDTKRSKAVINKFASDNT	127
hum-a1AT	QGKIVDLVK--ELDRDTVFALVNYIFFFKGWERPFVEVKDTEEDFHVDQVTTVVKVPMMKR	223
Ms-Alaserpin-9B	NNRIKNLVDPDALDETTTRSVLVNAIYFKGSWKDKFVKERTMDRDFHVSVDKTIKVPMTIG	219
Dm-Spn4A-H	NHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQFAKHLTRPDTFHLDGERTVQVPMMSL	235, 203
Ag-SRPN10A, B, C, D	NNKIKDLISPDALDELSSRMVNVAVHFKGTWTYQFDPSLTRPFPFVLSSETESRDVPMMI	207
Ci-Spn-10A, B	NGKISNMLQDDAVDSNTALIVANAVYFRGDWHSKFNEMQTERRAFVYSHYKIVETPFMFQ	226
Ce-Srp-7A, B, C	GDHIKKIIGSDSINSDLVAVLTNALYFKADWQNKFKKDSFTKSEFFSSADSKREIDFLHA	197
Ce-Srp-9A, B	KGKINKIISPEALKGIK-VVICNAIYFKAKWLKIFEDS--WEVSFFSAKKTSRQVEFLHM	184
hum-a1AT	LGMFNIQHCKKLSWVLLMKYLGNTAIFFLPDEGKQLHLENELTHD---IITKFLNED	280
Ms-Alaserpin-9B	KKDVRVADVPELDAMKIMESYEGDQASMIILPNQVDGITALEQKLDKPKALSRAEERLY	279
Dm-Spn4A-H	KERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEKLR-LTTLQSITQSLY	294, 262
Ag-SRPN10A, B, C, D	KKHFAFNFEELGFSALELTYGSDMTMMLLLPNERMGLAALEKLP-TLNLAELAGKMH	266
Ci-Spn-10A, B	RGHFYAYISELTLQVLEMDYAGKDYSMVLLMPEN-FDLAKVEANLN-HANLTNWSALK	284
Ce-Srp-7A, B, C	S-SVSRDYAENDQFVLSLPYKDNFTALTIPLPKTRFGLTESLTKLTD-SATIQHLLSNVS	255
Ce-Srp-9A, B	C-EMT-KFNENAHFKVVISMPYKDENFALAIIFLPKRFALKKALKSLD-SASILQLLTSFS	241
hum-a1AT	RRSASLHLPKLSITGTYDLKSVLQGLGITKVFVSNAGADLSG---VTEEAPLKLKSAVHKA	336
Ms-Alaserpin-9B	NTEVEIYLPKFKIETTTDLKEVLSNMNKKLFTPGAARLE--NLLKTKESLYVDAAIQKA	337
Dm-Spn4A-H	ETKVALKLRPRFKAQFVELSEVFKLGMSSRMFS-DQAEFG--KMLQSPEPLKVSIAIHKA	351, 319
Ag-SRPN10A, B, C, D	KQEVEVFLPKFKIETFTDLNEDLQALGMERMFS-DSAEFP--DLLEQNEPMKVSQVVKHA	323
Ci-Spn-10A, B	YKSVDLVSPKFKLEETLQLEVLQKSLGVTDLFDRQACDLT--GISKSKD-LNVDQIVHKT	341
Ce-Srp-7A, B, C	STSVNVQIPKWKIETKLGLEALQSLGKKAFFDNAD-----LGNMADGLYVSKVTHKA	309
Ce-Srp-9A, B	ECFVKIAIPKWSIETDMDIGAALQDLGIKDAFDNCAANFEDMYTSAGGDPLYISKALHKA	301
P17 P13		
hum-a1AT	VLTIDEKGTEAAGAMFLEAIPMS-----IPPEVKFNKPFVFLMIEQNTKSPLFMGKVVN	390
Ms-Alaserpin-9B	FIEVNEEGAEAAAAANAFGIVPASLI---LYPEVHIDRPFYFELKIDGIPMFNGKVIIEP-	392
Dm-Spn4A, E	FIEVNEEGTEAAAAATGMVRRKRAIMSPPEPIEFFADHPFTYVVLVHQKDLPLFWGSSVRL	411, 379
Dm-Spn4B, F	FIEVNEEGTEAAAAATGMVMCYASMLTFEPQVQFHVQHPFNYYIINKDSTILFAGRINKF	411, 379
Dm-Spn4C, G	FIEVNEEGTEAAAAATGMFMSLTSPLMPKPDPIRFNVDHPFTFYILNKDSTALFAGS IKKL	411, 379
Dm-Spn4D, H	FIEVNEEGTEAAAAATVVRVMAVAAF---SRKHFIANHPFAFYVVKTHYDLPIFTGRYLG-	406, 374
Ag-SRPN10A	FIEVNEEGTEAAAAATAVVRVKRALI---NRLKVRLDHHPFLYALMMGSSSRQTAFVGRLVK	380
Ag-SRPN10B	FIEVNEEGTEAAAAATGMIMMRCMP---MHPYFTVDHHPFLYVLRHQQMVYFVGRVAKID	379
Ag-SRPN10C	FIEVNEEGTEAAAAATAMIAVSFCMIS--EPVKFTVDHHPFIYALMSPEKSVFFIGKTCKM	381
Ag-SRPN10D	FIEVNEEGTEAAAAATGMVMMTCAMI---MYPMFTADHHPFLYALKDSQGNILF IGKILKV	379
Ci-Spn-10A	VLEVEQGEAAAAATTVRIQARSLN---SRPSFVADHPFLWAIRHRQSELLIFMGRLSR	397
Ci-Spn-10B	VLEVEENGGAQVQERADANQTPAL----DRPVVYVDHPFIIIVRGRANNAFHLFGAYKR	396
Ce-Srp-7A	LIEVDEEGTKAAAATTVSISLKSAMFVMEEPKDFADHPFFVFLSKDNHPLFVGLHH---	366
Ce-Srp-7B	LIEVDEEGTTAAAATTVSFVRK SARPSAIKPIVFOADHPFLFMLTQTNHPIFLGIHQ---	366
Ce-Srp-7C	LIEVDEGTVAVAATRCISIERCKKMK---ENIEFHAEHPPFFILHGGTSYIFLGVFTG---	365
Ce-Srp-9A	IIEVDKMGTTAAAVTSLMYTDGVPQLP--EHVEFWADHPFMFVLTNNHPMFMGIYN---	356
Ce-Srp-9B	IIEVDKLGTTAAAVTSFDYEEELGVPQLP-KYVEFWADHPFMFVLTNNHQMFMGIYNFIS	360

hum- α 1AT	PTQK-----	394
Ms-Alaserpin-9B	-----	
Dm-Spn4A, E	E ENTFASSEHDEL-----	424, 392
Dm-Spn4B, F	-----	
Dm-Spn4C, G	-----	
Dm-Spn4D, H	-----	
Ag-SRPN10A	P DQSDVAQANRHEEL-----	395
Ag-SRPN10B	A -----	380
Ag-SRPN10C	P -----	382
Ag-SRPN10D	-----	
Ci-Spn-10A	P EGPLLDHDEF-----	408
Ci-Spn-10B	P AGKIRSHDEL-----	407
Ce-Srp-7A	-----	
Ce-Srp-7B	-----	
Ce-Srp-7C	-----	
Ce-Srp-9A	-----	
Ce-Srp-9B	Y PSFS-----	365

Abbildung 40: Alignment der Protein-Varianten der alternativ-spleißbaren Serpine

Die alignierten Aminosäuresequenzen wurden von den jeweiligen Transkripten der Serpin-Gene *Ms-Serpin-Gene-1*, *Dm-Spn4*, *Ag-SRPN10*, *Ci-SPN-10*, *Ce-Srp-7* und *Ce-Srp-9* abgeleitet. Das Alignment wurde mit DIALIGN 2.2.1 durchgeführt und anschließend mit GeneDoc editiert. Aminosäuresequenzen die durch RCL-Exon-Varianten kodiert werden, sind grau unterlegt. Zur Orientierung wurden im RCL die Reste P17 und P13 markiert und die Positionen P1-P1' im humanen α 1-Antitrypsin durch Invertieren hervorgehoben..

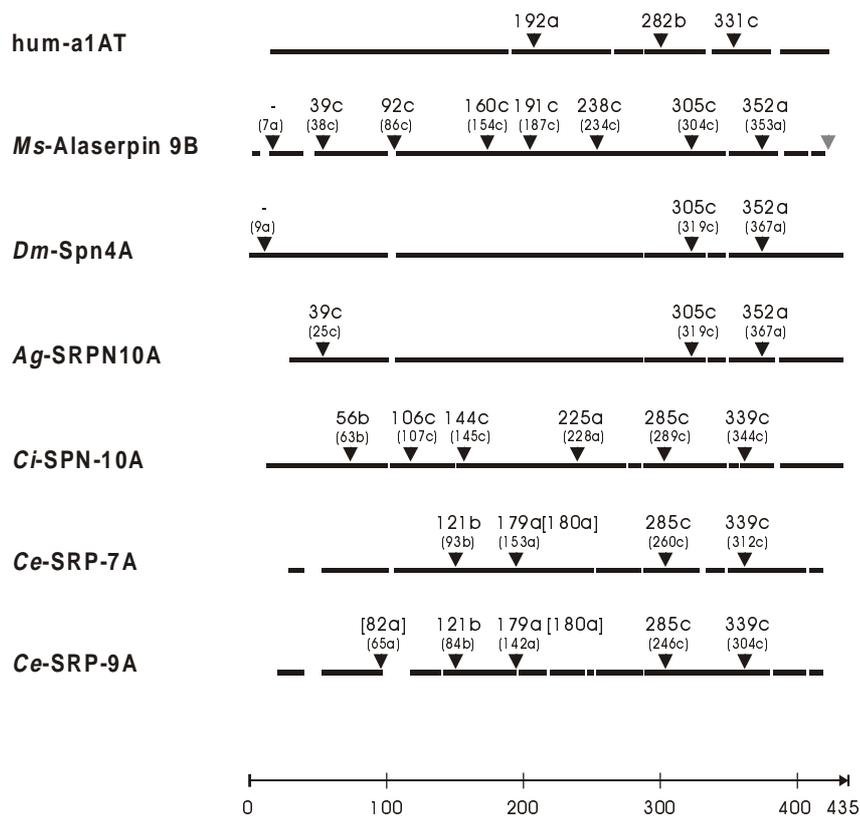


Abbildung 41: Exon-Intron-Strukturen der alternativ-spleißbaren Serpin-Gene

Darstellung der Intronpositionen der Serpin-Gene *Ms-Serpin-Gene-1*, *Dm-Spn4*, *Ag-SRPN10*, *Ci-Spn-10*, *Ce-Srp-7* und *Ce-Srp-9*. Die Pfeile markieren die Intronpositionen. Die obere Angabe der Intronpositionen bezieht sich auf reifes, humanes α 1-Antitrypsin (hum-a1AT). Die unteren Angaben geben die genspezifische Intronpositionen wieder. Die unten aufgeführte Skala repräsentiert das gesamte Alignment aus Abbildung 40.

6. Diskussion

Die Genomprojekte zu *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana* und *Ciona intestinalis* gewähren interessante Einblicke in die molekulare Evolution von Genen. Am Beispiel der Serpin-Superfamilie läßt sich eine Vielzahl von molekularen Evolutionsprozessen beobachten.

- Entstehung von Serpin-Pseudogenen (*Arabidopsis*, *C. elegans*, Abschnitt 6.4.1)
- Hotspot-Diversifikation (*Ciona*, Abschnitt 6.5.1)
- Alternatives Spleißen (*C. elegans*, *Drosophila*, *Ciona*, Abschnitt 6.7)

Die Hauptaufgabe der Genom-Analysen bestand in der Bestimmung bzw. Überprüfung der Gencluster-Organisationen, chromosomalen Lokalisationen, der Exon-Intron-Strukturen der Serpin-Gene und ihrer Aminosäuresequenzen. Die Überprüfungen (Abschnitt 4.6.1) der annotierten Serpin-Gene führten in einer Vielzahl der Fälle (ca. 17 - 60 %) zu Korrekturen bzw. Erweiterungen der Genmodelle (Tabelle 33).

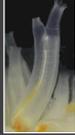
	<i>Drosophila melanogaster</i> 	<i>Caenorhabditis elegans</i> 	<i>Arabidopsis thaliana</i> 	<i>Ciona intestinalis</i> 
Anzahl der Serpin-Gene	29	9	9	10
Anteil der korrigierten Genmodelle	~ 17 %	~ 33 %	~ 22 %	60 %
Serpin-Gene mit multiplen, RCL-kodierenden Exons	1	2	-	1

Tabelle 33: Übersicht zu den Serpin-Genen aus den analysierten Genomprojekten

Die aufgrund der Überprüfungen vorgenommenen Veränderungen der Serpin-Genmodelle basieren auf der Identifizierung von zusätzlichen Aminosäuresequenzen, die für die Serpin-Superfamilie charakteristisch sind (Abschnitt 4.6).

Die Veränderungen werden dabei von vorhandenen EST- und cDNA-Daten¹ oder neu identifizierten kanonischen Spleißsignalen (5'-Spleißstelle / pyrimidin-reiche Sequenz mit anschließender 3'-Spleißstelle) gestützt. Die neu identifizierten Spleißsignale stehen dabei im Einklang mit den Spleiß-Konsensus-Signalen und die resultierenden Intronpositionen stimmen in allen Fällen mit bereits bekannten Serpin-Intronpositionen des jeweiligen Genoms überein.

6.1.1 Vergleich der phylogenetischen Kriterien

Die überprüften, phylogenetischen Kriterien (chromosomale Lokalisation, Gencluster-Organisation, Exon-Intron-Struktur und Aminosäuresequenz) wurden zur phylogenetischen Klassifikation der jeweiligen Serpin-Gene verwendet. Dabei wurde auch der Fragestellung nachgegangen, inwieweit die einzelnen Kriterien unabhängig voneinander zu identischen Gruppierungen der Serpin-Gene führen.

Es zeigte sich, dass bei *Drosophila melanogaster* und *Ciona intestinalis*, (siehe auch Abschnitte 6.3.2 und 6.5.2) die verschiedenen phylogenetischen Kriterien zu nahezu identischen Gruppierungen der Serpin-Gene führten.

Die Beobachtungen gelten im eingeschränkten Maße auch für die Serpine von *Caenorhabditis elegans* (Abschnitt 6.2.3). Da alle Serpin-Gene von *C. elegans* auf Chromosom V lokalisiert sind (Abschnitt 5.1.3), handelt es beim Merkmal „chromosomale Lokalisation“ um ein nicht-informatives Merkmal, das nicht zur Klassifikation der Serpine eingesetzt werden kann. Dagegen ist der Informationsgehalt der Gencluster-Organisation der Serpin-Gene von *C. elegans* sehr hoch. Je zwei Serpin-Gene bilden einen Gencluster (*Ce-Srp-1/Ce-Srp-2*, *Ce-Srp-5/Ce-Srp-6*, *Ce-Srp-7/Ce-Srp-8* und *Ce-Srp-9/Ce-Srp-10*) und diese Gruppierungen spiegeln sich teilweise im Aminosäuresequenz-basierten ML-Stammbaum (Abbildung 13) wider (*Ce-Srp-5/Ce-Srp-6* und *Ce-Srp-9A/Ce-Srp-10*).

Im Falle der Serpin-Gene von *Arabidopsis thaliana* (Abschnitt 6.4.2) ist der Nutzen der Exon-Intron-Strukturen und der Gencluster-Organisation für die Klassifikation der Serpine gering. So verfügen alle Serpine von *Arabidopsis thaliana* über ein Intron an Position 149c. Somit

¹ Satou, Y., et. al. 2002.

handelt es sich um ein nicht-informatives Merkmal. Die Klassifikation der Serpine aufgrund der chromosomalen Lokalisationen und der Serpin-Aminosäuresequenzen führt zu widersprüchlichen Gruppierungen. Vermutlich wurden infolge der Genomverdopplung (vor ca. 120 Millionen Jahren) große Genabschnitte umgeordnet (siehe auch Diskussion 6.4.2). Damit ging vermutlich auch ein großer Teil der ursprünglichen phylogenetischen Information verloren.

Es wird deutlich, dass immer mehrere Kriterien für die Erstellung einer verlässlichen Phylogenie verwendet werden sollte. Die Gewichtung der Kriterien hängt jedoch von den verschiedenen Organismen ab.

6.1.2 Indels als phylogenetische Marker

Neben den oben aufgeführten phylogenetischen Informationsquellen können auch seltene Indels zur Gruppierung von Serpinen verwendet werden. Eigene Beobachtungen zeigten, dass gemeinsame Indels häufig bei Genen mit identischer Exon-Intron-Struktur auftreten (*Ce-Srp9/Ce-Srp-10*, Insertion an Position 313, siehe Abbildung 45; *Dm-Spn-5/Dm-CG6687*, Insertion an Position 245, siehe Abbildung 46).

Die Entstehung von Indels wird im Falle der Serpine durch die hohen strukturellen Anforderungen, die die inhibitorische Proteinfunktion stellt, stark eingeschränkt. Deletion von Sekundärstruktur-Elementen führen häufig zum Verlust bzw. zur Einschränkung der Proteinfunktion (eventuell Bildung von Serpin-Pseudogenen, siehe Abschnitte 6.2.1 und 6.4.1). Dagegen zeigten Untersuchungen anhand der Aminosäuresequenz-Alignments (siehe Anhang 8.5) von *Drosophila*-, *Arabidopsis*- und *Ciona*-Serpinen, dass Insertionen bevorzugt zwischen Sekundärstruktur-Elementen lokalisiert sind. Eine Insertion, die Sekundärstruktur-Elemente zerstört, führt vermutlich eher zum Verlust der Proteinfunktion und wird infolge eines hohen Selektionsdrucks vermieden. Aufgrund der geringen Anzahl von klar begrenzten Indels, wurden Indels im Rahmen dieser Arbeit nicht als phylogenetische Marker verwendet. Einige Serpine von *Caenorhabditis elegans* enthalten größere Deletionen innerhalb der konservierten Serpin-Struktur (*Ce-Srp-5*, *Ce-Srp-9* und *Ce-Srp-10*, siehe Abschnitte 6.2.1 und 6.2.2). Insertionen treten bei *C. elegans*-Serpinen nicht auf.

Dagegen enthalten die Serpine von *Drosophila melanogaster*, *Arabidopsis thaliana* und *Ciona intestinalis* vorrangig Insertionen, die größtenteils zwischen Sekundärstruktur-Elementen lokalisiert sind (*Dm-CG-1342*, *Dm-Spn5*, *Dm-CG6687* und *Dm-CG7219*, Abschnitt 5.2.6; *At-Spn-2* (AT1G62170.1), Abschnitt 5.3.6; *Ci-Spn-1*, *Ci-Spn-2* und

Ci-Spn-4, Abschnitt 5.4.5). Diese Extensionen und Insertionen der Serpin-Struktur wurden mit Hilfe von ScanProsite¹ auf charakteristische Aminosäure-Sequenzmotive untersucht. In vielen Fällen konnten potentielle Phosphorylierungsstellen (für Casein-Kinase II, cAMP- und cGMP-abhängige Protein-Kinase und Protein-Kinase C) und potentielle N-Myristoylierungsstellen identifiziert werden. Jedoch ist es wichtig, darauf hinzuweisen, dass die entsprechenden Sequenzmotive relativ häufig vorkommen und die Ergebnisse einer solchen ScanProsite-Analyse daher vorsichtig zu bewerten sind.

6.2 Die Evolution der Serpin-Gene von *Caenorhabditis elegans*

Die in dieser Arbeit beschriebenen Untersuchungen des *C. elegans*-Genoms (Abschnitt 5.1) basierten auf der publizierten Genomsequenz von *C. elegans* (Sanger Centrum) und der WormBase-Datenbank². Alle Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (reifes hum-a1AT).

Die Prüfung der annotierten Genmodelle (Abschnitt 5.1.2) ergab, dass drei von neun *C. elegans*-Genmodellen korrigiert bzw. erweitert werden konnten. Die bei der Veränderung der Exon-Intron-Struktur neu zugeordneten Intronpositionen stimmen in allen Fällen mit bereits bekannten Intronpositionen anderer Serpin-Gene von *C. elegans* überein (Position 339c, vergleiche Tabelle 20).

Durch die Überprüfungen und Korrekturen der Exon-Intron-Strukturen konnten fehlende Strukturelemente ergänzt (*Ce-Srp-8*) und eine Insertion entfernt werden (*Ce-Srp-9*). Darüber hinaus war es möglich, Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten (*Ce-Srp-7* und *Ce-Srp-9*) zu identifizieren.

6.2.1 Die potentiellen Pseudogene *Ce-Srp-5* und *Ce-Srp-10*

Die Serpine *Ce-Srp-5* und *Ce-Srp-10* weisen Deletionen innerhalb der konservierten Serpin-Struktur auf (siehe Abschnitt 5.1.6 und Abbildung 45). Diese Deletionen sind zum Teil sehr umfangreich und führen zum Verlust von Sekundärstruktur-Elementen, die vermutlich für die

¹ ScanProsite, <http://ca.expasy.org/cgi-bin/scanprosite>; Gattiker A., Gasteiger E., Bairoch A., 2002.

² WormBase, <http://www.wormbase.org>, release WS99, Mai 2003

inhibitorische Funktion notwendig sind. Durch eine 34 As große Deletion fehlen im *Ce-Srp-5* die Strukturelemente s1A und hF. Im Serpin *Ce-Srp-10* treten zwei Deletionen auf (10 As bzw. 19 As), die zum Verlust bzw. zur Verkürzung der Strukturelemente s2A, s2C, s6A und hI führen. Interessant ist in diesem Zusammenhang, dass in beiden Fällen ein experimenteller Nachweis der jeweiligen cDNAs nicht möglich war (WORFDB¹). Die PCR-Experimente könnten daher ein Hinweis dafür sein, dass es sich bei *Ce-Srp-5* und *Ce-Srp-10* um Serpin-Pseudogene handelt.

6.2.2 Das Serpin-Gen *Ce-Srp-9*

Das ursprüngliche annotierte Genmodell von *Ce-Srp-9* enthält ein 51 Bp-großes Intron (Position 82a/85a), das eine Deletion von 17 As (Positionen 86-102) verursacht (siehe Abbildung 12 und Abbildung 45). Die Deletion führt zum Verlust des Sekundärstruktur-Elements hD. Durch PCR-Experimente (WORFDB¹, Primerpaar FOR-179@H4/REV-179@H4) konnte eine Expression des Gens nachgewiesen werden. Eine eigene Überprüfung der zugehörigen OST-Daten (*open reading frame sequence tags*, Sequenzdaten des zugehörigen Klon: FOR_plate179_H4.SCF) zeigte jedoch, dass das amplifizierte *Ce-Srp-9*-Transkript die Sequenz des postulierten „82a/85a-Introns“ enthält, und es sich somit nicht um ein Intron handelt. Abbildung 42 vergleicht die 51 Bp-große „Intronsequenz“ mit dem zugehörigen ORF von *Ce-Srp-9* und *Ce-Srp-10*. Besonders interessant ist, dass das „Intron“ eine hoch konservierte, für Serpine charakteristische Aminosäuresequenz kodiert, die jedoch ein Stopp-Codon enthält (rot markiert im Vergleich zu *Ce-Srp-10*).

```

„Intronsequenz“ Ce-Srp-9:  gt ggc tcc acg gat tct caa atg caa aaa taa ata agt gat ctc tgt gaa g
As-Sequenz Ce-Srp-9:      G  S  T  D  S  Q  M  Q  K  I  S  D  L  C  E
As-Sequenz Ce-Srp-10:   G  S  T  D  S  Q  L  Q  K  Y  I  S  D  L  C  E
DNA-Sequenz Ce-Srp-10:  gt ggc tcc acg gat tcc cag ctt caa aaa tac atc agt gat ctc tgt gaa g

```

Abbildung 42: „82a/85a-Intron“-Sequenz des Serpin-Gens *Ce-Srp-9*

Alignment der DNA-Sequenz des postulierten „82a/85a-Introns“ von *Ce-Srp-9* mit der abgeleiteten Aminosäuresequenz im Einbuchstabencode. Zum Vergleich ist die entsprechende Aminosäure- und DNA-Sequenz von *Ce-Srp-10* gezeigt. Die postulierten 5'- und 3'-Spleißsignale des „82a/85a-Introns“ sind grün markiert. Die Mutation Y → Stopp-Codon ist rot markiert (siehe auch Abbildung 45).

¹ The *C. elegans* ORFeome cloning project, Reboul, J., et al 2001.

Die naheliegendste Erklärungsmöglichkeit ist, dass es sich hier nicht um ein Intron handelt, und es erst kürzlich zu einer Punktmutation an As-Position 95 gekommen ist, die zu einem Stopp-Codon im ORF führt. *Ce-Srp-9* müsste damit als Serpin-Pseudogen eingestuft werden. Eine weitere Erklärung wäre, dass es sich um ein Intron handeln, das aufgrund der suboptimalen 5'- und 3'-Spleißstellen (in Abbildung 42 grün markiert) nicht immer gespleißt wird (Intron-Retention als „Nebenprodukt“).

6.2.3 Die phylogenetische Klassifikation der Serpin-Gene von *C. elegans*

Alle neun Serpin-Gene von *C. elegans* sind auf Chromosom V lokalisiert, so dass die chromosomalen Lokalisationen nicht-informativ sind und nicht zur phylogenetischen Klassifikation der Serpin-Gene dienen können. Interessant ist die Gruppierung der *C. elegans*-Serpine in Gencluster. Bis auf *Ce-Srp-3* liegen alle Serpin-Gene in Genclustern aus jeweils zwei Genen vor (*Ce-Srp-1/Ce-Srp-2*, *Ce-Srp-5/Ce-Srp-6*, *Ce-Srp-7A/Ce-Srp-8* und *Ce-Srp-9/Ce-Srp-10*, Abbildung 11). Diese Gruppierungen spiegeln sich zum Teil im Maximum-Likelihood-Baum (Abbildung 13) wider. Hier werden die Gruppierungen *Ce-Srp-5/Ce-Srp-6* und *Ce-Srp-9/Ce-Srp-10* gebildet. Das NJ-Verfahren (Ergebnisse nicht dargestellt) bestätigt diese Gruppierungen der Serpine und positioniert zusätzlich *Ce-Srp-7/Ce-Srp-8* zusammen.

Die Exon-Intron-Strukturen der Serpin-Gene von *C. elegans* (Tabelle 20) erscheinen sehr homogen und sind phylogenetisch wenig informativ. Aufgrund der Genstrukturen lassen sich die Serpin-Gene in zwei Gruppen einteilen. Gruppe I wird von den Serpin-Genen *Ce-Srp-6*, *Ce-Srp-7* und *Ce-Srp-8* gebildet, die eine gemeinsame Genstruktur aufweisen, die durch Introns an den Positionen 121b, 179a/180a, 285c und 339c gekennzeichnet ist. Gruppe II umfasst die Serpin-Gene *Ce-Srp-1*, *Ce-Srp-3*, *Ce-Srp-5*, *Ce-Srp-9* und *Ce-Srp-10*, die jeweils mindestens drei von vier Intronpositionen der Gruppe I aufweisen und teilweise über eine zusätzliche Intronposition verfügen. Eine Außenseiterposition nimmt das Serpin-Gen *Ce-Srp-2* ein, das einerseits nur über eine charakteristische Intronposition der *C. elegans*-Serpine verfügt (Position 121b) und andererseits zwei zusätzliche „*Ce-Srp-2*“-spezifische Introns aufweist (Positionen 217c und 266c/265c).

Besonders hervorzuheben ist, dass die Kriterien „Gencluster-Organisation“ und „Aminosäuresequenz“ zu ähnlichen Gruppierungen der Serpine von *C. elegans* führen. Die Außenseiterpositionen werden von *Ce-Srp-1* und in *Ce-Srp-2* in Bezug auf die Genstruktur eingenommen. Beide Gene bilden zusammen einen Gencluster.

6.3 Evolution der Serpin-Gene von *Drosophila melanogaster*

Die in dieser Arbeit durchgeführte Analyse des Genoms von *Drosophila melanogaster* (Abschnitt 5.2) basiert auf der publizierten Genomsequenz des *Berkeley Drosophila Genome Project* (BDGP) und der FlyBase-Datenbank¹. Alle Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (reifes, hum-a1AT).

Eine Überprüfung der annotierten Genmodelle (Abschnitt 5.2.2) ergab, dass fünf von 29 *Drosophila*-Genmodellen (*Dm*-CG31902, *Dm*-CG3312, *Dm*-Spn4, *Dm*-CG9454 und *Dm*-CG10956) korrigiert bzw. erweitert werden konnten. Die bei der Korrektur der Exon-Intron-Strukturen neu zugeordneten Intronpositionen stimmen in allen Fällen mit bereits bekannten Intronpositionen anderer Serpin-Gene von *Drosophila melanogaster* überein (Position 305c sowie Variationen der Position 352a, siehe unten).

Durch die Überprüfung der Exon-Intron-Strukturen (Abschnitt 5.2.2) konnten fehlende Strukturelemente ergänzt (*Dm*-CG31902, *Dm*-CG9454 und *Dm*-CG10956) und Insertionen entfernt werden (*Dm*-CG3312). Darüber hinaus war es möglich, ein Serpin-Gen mit vier RCL-kodierenden Exon-Varianten (*Dm*-Spn4, Abschnitt 5.6.1) zu identifizieren.

Die von FlyBase annotierten Serpin-Gene *Dm*-CG14470 (1976 As) und *Dm*-CG12807 (555 As) wurden aufgrund ihrer für Serpine untypischen Größen, Genstrukturen und Aminosäuresequenzen von der weiteren Analyse ausgeschlossen.

6.3.1 Variationen der Intronposition 352a

Das Intron an Position 352a ist bei Serpin-Genen von *Drosophila melanogaster* weit verbreitet (siehe Tabelle 22). Dieses Intron ist am Ende eines konservierten Bereiches (β -Faltblattstrang s4A) innerhalb des RCL lokalisiert. Neben dieser, durch zwei Alignierungsmethoden gut abgesicherten Position, scheinen noch Varianten des Introns an eng benachbarten Positionen aufzutreten (Positionen 353a, 357a, 358a und 360a). Es stellt sich die Frage, ob es sich hier um unabhängig voneinander entstandene Intronpositionen handelt, oder ob ein vorher vorhandenes 352a-Intron durch Insertions- oder Deletionsprozesse verschiedene, benachbarte Positionen eingenommen hat.

¹ FlyBase, A Database of the *Drosophila* Genome, <http://www.flybase.org/>, The FlyBase Consortium, 2003.

Im Falle einer mehrfachen, voneinander unabhängigen Intron-Insertion (*intron gains*) sind aus statistischen Gründen auch unterschiedliche Intronphasen zu erwarten. Die identischen Intronphasen (jeweils a) der vier „Varianten“ des 352a-Intron sprechen somit gegen ein mehrfaches voneinander unabhängiges Entstehen der Introns.

Das Auftreten der Varianten des 352a-Introns könnte auch durch Intron-Sliding¹-Vorgänge erklärt werden. Unter Intron-Sliding werden verschiedene Prozesse zusammengefasst, die zur Verschiebung einer Intronposition in der Genstruktur führen. In diesem Fall befinden sich die Varianten des Intron jedoch im hoch diversen Teil des RCL (Position 352 – 360, siehe Abbildung 46). Eine Alignierung dieses Sequenzbereichs wird durch Indels erschwert und erscheint in Bezug auf die Protein-Funktion nicht sinnvoll. Aufgrund des unsicheren Alignments im Bereich 353 – 360 können keine verlässlichen Aussagen über die Entstehung der Intronvarianten gemacht werden.

6.3.2 Die phylogenetische Klassifikation der Serpin-Gene von *Drosophila*

Am Beispiel der Serpin-Gene des Genoms von *Drosophila melanogaster* zeigt sich eindrucksvoll wie die drei unabhängigen Kriterien, „chromosomale Lokalisationen“, „Exon-Intron-Strukturen“ und „Aminosäuresequenz-Daten“ zu einer nahezu identischen phylogenetischen Klassifikation der Serpin-Gene führen können.

Von insgesamt 27 betrachteten Serpin-Genen von *Drosophila melanogaster* sind 20 Gene auf Chromosom 2, und sieben auf Chromosom 3 lokalisiert (Abschnitt 5.2.3). Interessanterweise korrelieren die Exon-Intron-Strukturen der Serpin-Gene stark mit ihren chromosomalen Lokalisationen (Tabelle 22). Auf den Chromosomen-Armen 2L und 2R sind vorwiegend Serpin-Gene lokalisiert, die über die Intronpositionen 305c und 352a (bzw. 352a-Varianten) verfügen, während die auf den Chromosomen-Armen 3L und 3R lokalisierten Serpin-Gene abweichende Genstrukturen zeigen. Dieser Befund weist darauf hin, dass duplizierte Gene wahrscheinlich infolge von Rekombinationsereignissen, vorwiegend intrachromosomal umgeordnet werden bzw. dass interchromosomale Gen-Umordnungen ein sehr seltenes Ereignis darstellen.

Die Serpin-Gene von *Drosophila melanogaster* können aufgrund ihrer Exon-Intron-Strukturen in drei Gruppen gegliedert werden. Die Gruppe I wird von 18 Serpin-Genen gebildet, die alle auf Chromosom 2 lokalisiert sind (*Dm*-CG6717, *Dm*-CG31902,

¹ Stoltzfus, A., Logsdon, J., M., Jr, Palmer, J., D., Doolittle, W., F., 1997.

Dm-CG33121, *Dm-Spn2*, *Dm-CG4804*, *Dm-Spn3*, *Dm-Spn4*, *Dm-CG9454*, *Dm-CG9455*, *Dm-Spn1*, *Dm-CG9460*, *Dm-Spn43Aa*, *Dm-Spn43Ab*, *Dm-CG1859*, *Dm-Nec*, *Dm-CG7722*, *Dm-CG10956* und *Dm-Spn6*, Tabelle 22). Von diesen 18 Serpin-Genen verfügen 15 im konservierten Bereich der Serpinstruktur über Introns an den Positionen 305c und 352a (bzw. 352a-Varianten), während die Serpin-Gene *Dm-Nec*, *Dm-CG10956* und *Dm-Spn6* eine Untergruppe bilden und nur jeweils eine der beiden „Chromosom 2“-typischen Intronpositionen (305c oder 352a) aufweisen.

Gruppe II umfasst die auf Chromosomarm 3L liegenden Serpin-Gene *Dm-CG6680*, *Dm-CG6663* und *Dm-CG6289*, die identische Genstrukturen mit Introns an den Positionen 102b, 195c und 305c zeigen.

Die Gruppe III wird durch die Serpine *Dm-Spn5* und *Dm-CG6687* gebildet, die über die gemeinsame Intronposition 301c verfügen und zusammen einen Gencluster auf dem Chromosomenarm 3R bilden. Die Serpin-Gene *Dm-Spn27a* (ohne Intron), *Dm-CG7219* (131b-Intron), *Dm-Acp76A* (340a-Intron) und *Dm-CG1342* (-c/156c-Intron) weisen kein bzw. nur ein Intron an einer nicht konservierten Position auf und nehmen in Bezug auf die Genstruktur Außenseiterpositionen ein.

Die durch chromosomale Lokalisationen und Genstrukturen bedingten Gruppierungen spiegeln sich zum großen Teil im Neighbor-Joining-Baum wider (Abbildung 20). Die unbeschrifteten Baumknoten werden nur schwach unterstützt und sollten mit Vorsicht interpretiert werden. Interessant ist der Split von „Chromosom 2“-Serpinen, die sich mit einer *Bootstrap*-Unterstützung von 57 % von einem Großteil der „Chromosom 3“-Serpine absetzen. Der einzige Ausreißer ist hier das Serpin *Dm-CG7219*, das aber auch aufgrund seiner Genstruktur sowie charakteristischen Insertionen eine separate Stellung einnimmt. Möglicherweise ist *Dm-CG7219* als Kandidat für ein horizontal transferiertes Gen anzusehen (z.B. von *Baculo-Viren*). Es sollte aber auch hervorgehoben werden, dass die *Bootstrap*-Unterstützung von 57 % keine klare Monophylieaussage (*Bootstrap* > 75 %) erlaubt.

6.4 Evolution der Serpin-Gene von *Arabidopsis thaliana*

Die in der vorliegenden Arbeit beschriebene Analyse des Genoms von *Arabidopsis thaliana* (Abschnitt 5.3) basierte auf der publizierten Genomsequenz der TAIR-Datenbank (*The Arabidopsis Information Resource*)¹.

Die Überprüfung der annotierten Genmodelle (Abschnitt 5.3.2) führte zur Identifikation von fünf Serpin-Pseudogenen (AT1G62160.1, „AT1G64010.1/AT1G64020.1“, AT1G65385.1, AT2G35590.1 und „AT2G35560.1/AT2G35570.1“, Tabelle 24), zwei Serpin-verwandte Genen (AT1G51330.1 und AT1G63280.1) und neun Serpin-Genen (*At-Spn-1* bis *At-Spn-9*, Tabelle 23).

Durch die eingeführten Korrekturen der Exon-Intron-Strukturen konnte eine aminoternale Extension bei *At-Spn-4* entfernt und das aminoternale Ende von *At-Spn-8* durch Setzen eines Intron (Position 20c in Bezug auf *At-Spn-8*) erweitert werden. Es konnten keine Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten im Genom von *Arabidopsis thaliana* identifiziert werden.

Die Serpin-Gene von *Arabidopsis thaliana* sind hinsichtlich ihrer Genstrukturen und Aminosäuresequenzen sehr homogen und weisen im Alignment im Vergleich zum α_1 -Antitrypsin nur kleinere Indels auf, die überwiegend zwischen den Sekundärstruktur-Elementen lokalisiert sind. Eine Ausnahme stellt *At-Spn-2* (AT1G62170.1) dar, das im Vergleich zu den anderen *Arabidopsis*-Serpinen eine aminoternale Extension von ca. 60 Aminosäuren aufweist.

6.4.1 Die Serpin-Pseudogene von *Arabidopsis thaliana*

Im Falle des Genoms von *Arabidopsis thaliana* bestand die Hauptaufgabe in der Identifikation der Serpin-Pseudogene. Zur Zeit werden in der TAIR-Datenbank fünf Pseudogene als Serpin-Genmodelle annotiert (AT1G62160.1, „AT1G64010.1/AT1G64020.1“, AT1G65385.1, AT2G35590.1 und „AT2G35560.1/AT2G35570.1“).

Die Serpin-Pseudogene spiegeln vermutlich die vor ca. 112 Millionen Jahren durchlaufene Genomverdopplung² wider. Die dabei duplizierten Serpin-Gene unterstanden vermutlich nicht

¹ Garcia-Hernandez, M., *et al.* 2002.

² The Arabidopsis Genome Initiative, 2000.

mehr dem evolutionären Druck zur Aufrechterhaltung ihrer Funktionen und konnten ohne funktionelle Beschränkungen Gen-Diversifikationsprozesse durchlaufen. Gene, die eine neue, für den Organismus vorteilhafte Funktion übernahmen, blieben erhalten, während sich die funktionslosen Serpin-Gene möglicherweise zu Pseudogenen entwickelten (Akkumulation von Deletionen und Rastermutationen, siehe dazu Abbildung 22 und Abbildung 23).

6.4.2 Die phylogenetische Klassifikation der Serpin-Gene von *Arabidopsis*

Die neun Serpin-Gene von *Arabidopsis thaliana* verteilen sich auf die Chromosomen I (*At-Spn-1*, *At-Spn-2*, *At-Spn-3*, *At-Spn-4*), II (*At-Spn-5*, *At-Spn-6*, *At-Spn-7*, *At-Spn-8*) und III (*At-Spn-9*). Aufgrund der vor ca. 112 Millionen Jahren durchlaufenen Genomverdopplung (siehe Abschnitt 6.1.1) ist damit zu rechnen, dass es zu intra- und interchromosomalen Gen-Umordnungen gekommen ist. Vermutlich wird die Information der chromosomalen Lokalisationen nur noch einen eingeschränkten phylogenetischen Wert enthalten. Interessant ist die häufige „Vergesellschaftung“ von Serpin-Genen mit Serpin-Pseudogenen in Genclustern (siehe Abbildung 22 und Abbildung 23).

Alle neun Serpin-Gene von *Arabidopsis thaliana* verfügen im konservierten Kernbereich über ein Intron an der Position 149c. Damit ist die Genstruktur bei *Arabidopsis thaliana* ein nicht-informatives Merkmal. Für die Klassifikation der *Arabidopsis*-Serpine ist man somit auf die Aminosäuresequenz-Daten der Serpine angewiesen, die als Ausgangspunkt zur Konstruktion eines Serpin-Stammbaums dienen. Im Maximum-Likelihood-Stammbaum (Abbildung 26) gruppieren sich die *Arabidopsis*-Serpine in zwei Cluster. Alle Unterstützungswerte sind sehr hoch (> 85 %) und weisen auf eine Evolution ohne Mutations-Sättigung hin bzw. auf eine späte Radiation der Serpin-Gene.

Die vom ML-Stammbaum abgeleiteten Gruppierung in zwei Cluster (*At-Spn-6/At-Spn-7* und *At-Spn-9* gegenüber *At-Spn-1/At-Spn-2/At-Spn-3/At-Spn-4/At-Spn-5* und *At-Spn-8*) zeigt keine erkennbare Ähnlichkeit zur Gruppierung aufgrund der chromosomalen Genlokalisierungen (Abbildung 24). Der einzige Serpin-Gencluster im *Arabidopsis*-Genom wird von den Serpin-Genen *At-Spn-3/At-Spn-4* gebildet (Abbildung 22B). Im Gegensatz zu einigen Serpin-Genclustern anderer Organismen (*Drosophila melanogaster*) werden die beiden Serpine im ML-Stammbaum nicht zusammen gruppiert.

6.5 Evolution der Serpin-Gene von *Ciona intestinalis*

Die in dieser Arbeit beschriebene Analyse des *Ciona*-Genoms (Abschnitt 5.4) basierte auf Daten aus der „Ghost Database: a *Ciona intestinalis* cDNA resource“¹ und der JGI-Datenbank „*Ciona intestinalis* v1.0“¹. Alle (Intron)-Positionsangaben beziehen sich auf das Referenz-Serpin α_1 -Antitrypsin (reifes, hum-a1AT).

Eine eigene Untersuchung der annotierten Genmodelle (Abschnitt 5.4.2) führte zur Identifikation von insgesamt zehn Serpin-Genen (*Ci-Spn-1* - *Ci-Spn-10*, Tabelle 26). Die bei der Veränderung der Exon-Intron-Strukturen neu zugeordneten Intronpositionen stimmen in allen Fällen mit bereits bekannten Intronpositionen anderer Serpin-Gene von *Ciona intestinalis* überein (Position 339c, Tabelle 27).

Durch die Korrekturen der Exon-Intron-Strukturen der Serpin-Gene von *Ciona intestinalis* konnten Protein-kodierende Serpin-Sequenzen amino- und carboxyterminal ergänzt werden (*Ci-Spn-5* und *Ci-Spn-9*). Die Genstrukturen von *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8*, die zuvor in einem Genmodell annotiert waren, konnten korrigiert werden. Darüber hinaus war es möglich, ein Serpin-Gen mit zwei RCL-kodierenden Exon-Varianten (*Ci-Spn-10*) zu identifizieren.

6.5.1 Hotspot-Diversifikation bei Serpin-Genen von *Ciona intestinalis*

Die Serpin-Gene *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* waren ursprünglich im JGI-Genmodell ci0100146394 (ci0100146394) als ein Serpin annotiert. Eigene Untersuchungen von EST-Daten (JGI, Abschnitt 5.4.2) führten zur Identifikation von drei intronfreien Serpin-Genen, die zusammen einen Gencluster bilden. Ein Vergleich ihrer Aminosäuresequenzen (Abbildung 48) zeigte über die gesamte kodierende Region eine Sequenz-Identität² von ca. 68 %. Interessant ist, dass der ca. 180 As lange aminoterminal Bereich hoch konserviert ist (Identität² > 93 %), während die Anzahl der Substitutionen in der carboxyterminalen Hälfte der Proteine zunimmt. Der Bereich der größten Variabilität kann im RCL lokalisiert werden. Der spezifitäts-bestimmende Sequenzbereich des RCLs von Position P7 bis zur β -Faltblattstruktur s1C ist nicht zuverlässig alignierbar (Identität² < 25 %) und enthält als einziger Bereich Indels (< 2 As). Der sich anschließende, carboxyterminale

¹ Satou, Y., et. al. 2002.

² Aminosäure-Identität bezogen auf das multiplen Alignment (von *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8*).

Bereich (umfasst die Sekundärstruktur-Elemente s1C, S4B, s5B und den Carboxyterminus) weist wiederum eine Aminosäuresequenz-Identität² von ca. 60 % auf. Damit stellen die Serpin-Gene *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* ein eindrucksvolles Beispiel für einen Hotspot-Diversifikationsprozess dar, der sich auf eine funktionale Domäne, hier der RCL, konzentriert. Eine mögliche Ursache für die Hotspot-Diversifikation ist zum Beispiel eine Selektion durch Parasiten-Proteasen. Ähnliche Beobachtungen konnten auch bei Serpin-Genen von Maus und Ratte gemacht werden¹.

6.5.2 Die phylogenetische Klassifikation der Serpin-Gene von *Ciona*

Da das genomische Sequenz-Assembly von *Ciona intestinalis* noch nicht den 14 Chromosomen zugeordnet wurde², stehen zur Zeit keine Informationen über die chromosomalen Lokalisationen der Serpin-Gene zur Verfügung.

Die Serpin-Gene von *Ciona intestinalis* weisen sehr unterschiedliche Exon-Intron-Strukturen auf (Tabelle 27). Aufgrund der Genstruktur können die Serpin-Gene in drei Gruppen eingeordnet werden. Gruppe I umfasst die intronfreien Gene *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8*. Die Gruppe II wird durch die Serpine *Ci-Spn-9* und *Ci-Spn-10* gebildet, die identische Exon-Intron-Strukturen (Introns an den Positionen 60a, -c/106c, 144c, 225a, 285c und 339c) aufweisen.

Die Serpin-Gene *Ci-Spn-1* und *Ci-Spn-2* können zusammen in Gruppe III eingeordnet werden. Sie verfügen über drei gemeinsame Introns (Positionen 67a, 136b und 191c), aber zusätzlich auch über individuelle Intronpositionen. *Ci-Spn-5* nimmt eine interessante Stellung zwischen Gruppe II (durch die zwei gemeinsamen Introns an den Positionen 144c und 339c) und Gruppe III ein (durch Introns an den Positionen 191c und 247a/-a). Die Serpin-Gene *Ci-Spn-3* und *Ci-Spn-4* weisen keine gemeinsame Intronpositionen mit anderen Serpin-Genen von *Ciona intestinalis* auf und nehmen daher Außenseiterpositionen ein.

Im Maximum-Likelihood-Stammbaum (Abbildung 29) spiegelt sich die von der Genstruktur abgeleiteten Gruppierungen vollständig wider. Die drei Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* (Gruppe I) bilden einen Gencluster und werden mit hohen Unterstützungswerten von 96 % bzw. 100 % im ML-Baum zusammen gruppiert.

Auch *Ci-Spn-9* und *Ci-Spn-10A* (Gruppe II) und *Ci-Spn-1* und *Ci-Spn-2* (Gruppe III) werden mit einem Unterstützungswert von 100 % bzw. 93 % zusammen gruppiert. Besonders

¹ Hill, R., E., Hastie, N., D., 1987.

² Dehal, P., et al. 2002. (<http://genome.jgi-psf.org/ciona4/ciona4.home.html>).

interessant ist, dass *Ci-Spn-5* aufgrund der Genstruktur und der Aminosäuresequenz (bzw. ML-Baum) eine Stellung zwischen Gruppe II und Gruppe III einnimmt. Im ML-Baum nehmen *Ci-Spn-3* und *Ci-Spn-4* eine separate Stellung ein. Auch diese Positionierung ist mit der aus der Genstruktur abgeleiteten Klassifizierung kongruent. Damit sind die Serpin-Gene von *Ciona intestinalis* ein gutes Beispiel für die Kongruenz der phylogenetischen Kriterien „Aminosäuresequenz“ und „Genstruktur“.

6.6 Interspezies-Stammbäume der Invertebraten-Serpine

Um die Frage zu beantworten, inwieweit die Gruppierungen der Intraspezies-Stammbäume in einem Interspezies-Stammbaum erhalten bleiben, wurden mit verschiedenen Verfahren (Neighbor-Joining, Maximum-Likelihood und Maximum-Parsimony) Stammbäume für den gesamten Datensatz von 65 Serpin-Sequenzen konstruiert (siehe Abbildung 30 - Abbildung 32). Für PAUP*4.0b10 standen nur die Substitutionsmatrizen PAM und BLOSUM zur Verfügung. In der vorliegenden Arbeit wurde die PAM-Substitutionsmatrix¹ eingesetzt, die als Vorläufer der sonst in dieser Arbeit verwendeten JTT-Matrix bezeichnet werden kann.

Ein Vergleich der drei Interspezies-Serpin-Stammbäume (NJ, ML, MP) zeigt, dass eine große Anzahl der Gruppierungen jeweils nur über kleine Unterstützungswerte (< 50 %) verfügt bzw. einen hohen Grad an *Multifurcation* aufweist (MP-Baum). Es wird deutlich, dass ein Datensatz, der Serpine von Pflanzen, Insekten, Nematoden und Vertebraten enthält, zu heterogen für die Konstruktion einer verlässlichen Serpin-Phylogenie ist. Interessanterweise zeigen die drei Phylogramme dennoch eine Reihe von gemeinsamen Gruppierungen, mit zum Teil hohen Unterstützungswerten.

Hervorzuheben sind die klar umrissenen Gruppierungen der Serpine von *Arabidopsis thaliana* und *Caenorhabditis elegans*, die im NJ-, ML- und im MP-Baum hohe Unterstützungswerte aufweisen. Die Gruppierungen sind nahezu identisch mit den jeweiligen Intraspezies-Stammbäumen.

Besonders interessant sind die gemeinsamen Gruppierungen von *Ciona intestinalis*- und Vertebraten-Serpinen. So werden in den mit NJ-, ML- und MP-Verfahren erstellten Phylogrammen die Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* zusammen mit hum-PAI-2

¹ Dayhoff, M., O., Schwartz, R., and Orcutt, B., C., 1978.

(Gruppe 1 der Vertebraten-Genfamilie) gruppiert. Die Unterstützungswerte sind jedoch sehr unterschiedlich. So wird diese Gruppierung im NJ-Baum in 86 % der *Bootstrap*-Replikat wiedergefunden, während die Gruppierung im ML-Baum und MP-Baum mit 54 % bzw. 38 % deutlich schwächer unterstützt wird. Eine *Diagnostic Sites*-Analyse der Serpine *Ci-Spn-6*, *Ci-Spn-7* und *Ci-Spn-8* erbrachte keine signifikanten Gemeinsamkeiten mit dem für Gruppe I-Serpine spezifischen Aminosäuremuster¹. Da die drei Serpine keine Introns enthalten, kann eine gemeinsame Gruppierung mit hum-PAI-2 auch nicht mit Hilfe der Genstruktur unterstützt werden.

In einer zweiten Gruppe von *Ciona*- und Vertebraten-Serpinen werden *Ci-Spn-9* und *Ci-Spn-10A* zusammen mit hum-ATIII (Gruppe 5 der Vertebraten-Genfamilie) positioniert. Interessanterweise verfügen diese drei Serpin-Gene auch über die gemeinsame Intronposition 339c. Der Cluster „*Ci-Spn-9/Ci-Spn-10A/hum-ATIII*“ wird im NJ-Baum mit einem *Bootstrap*-Wert von 81 % unterstützt. Dagegen ist die Unterstützung im MP-Baum mit 33 % deutlich geringer. Im ML-Baum tritt diese Gruppierung nicht auf. Eine *Diagnostic Sites*-Analyse der Serpine *Ci-Spn-9* und *Ci-Spn-10A* zeigte deutliche Abweichungen vom Aminosäuremuster, das für Serpine der Gruppe 5 erwartet wird.

Hervorzuheben sind auch die separaten Stellungen der Serpine *Ci-Spn-4* und *Dm-CG7219*. *Ci-Spn-4* aus *Ciona intestinalis* nimmt bereits aufgrund der Genstruktur eine Ausreißerposition unter den *Ciona*-Serpinen ein. In Abhängigkeit des gewählten Baumrekonstruktions-Verfahrens wird *Ci-Spn-4* mit *Drosophila*-Serpinen (NJ-Baum, mit *Dm-Spn27a/Dm-Nec*) oder in die Nähe von *Arabidopsis*-Serpinen (MP-Baum) positioniert. Auch *Dm-CG7219* (durch einen Pfeil hervorgehoben) besitzt eine für *Drosophila*-Serpine ungewöhnliche Genstruktur (Tabelle 22) und wird in den NJ- und MP-Interspeziesbäumen zusammen mit *Ciona*-Serpinen gruppiert.

¹ Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.

6.7 Serpin-Gene mit multiplen RCL-kodierenden Exon-Varianten

Mit der in Abschnitt 4.6.1 beschriebenen AGREP-Suchstrategie konnten in den Genomen von *Caenorhabditis elegans*, *Drosophila melanogaster* und *Ciona intestinalis* Serpin-Gene identifiziert werden, die über multiple RCL-kodierende Exon-Varianten verfügen. Die durch Datenbank-Recherchen identifizierten Serpin-Gene *Dm-Spn4*, *Ce-Srp-7*, *Ce-Srp-9* und *Ci-Spn-10* (ci0100154072) können vermutlich die verschiedenen RCL-kodierenden Exon-Varianten durch einen alternativen Spleißprozess an ein gemeinsames Serpin-Grundgerüst anfügen¹. Im Rahmen der vorliegenden Arbeit konnte dieser Spleißvorgang für *Dm-Spn4* experimentell nachgewiesen werden² (Abschnitt 5.6.1). Die Ergebnisse dieser Arbeiten zeigten, dass durch eine Kombination von alternativen Spleißprozessen und Verwendung alternativer Transkriptionsstarts mindestens zehn *Dm-Spn4*-Transkripte gebildet werden, die acht unterschiedliche Protein-Isoformen kodieren. Nach der Sequenz der RCL-Region und entsprechenden Signalen am N- und C-Terminus unterscheiden sich die *Dm-Spn4*-Isoformen in ihren Protease-Spezifitäten und ihren Zelllokalisationen (Abschnitt 5.6.1). Das Vorliegen von multiplen RCL-kodierenden Exon-Varianten kann somit als ein genetischer Sparmechanismus angesehen werden, der zur Erhöhung der Target-Diversität von Serpinen dient.

Bei Serpin-Genen mit multiplen RCL-kodierenden Exon-Varianten kam es wahrscheinlich durch Rekombinationen zur Duplikation der RCL-kodierenden Exons. Anschließend durchlaufen die zusätzlichen Exon-Varianten ohne Selektionsdruck Diversifikationsprozesse und übernehmen entweder eine neue Funktion (z.B. Inhibition einer anderen Protease) oder degenerieren durch Ansammeln von Deletionen und Rastermutationen zum RCL-Pseudoexon.

Einen Hinweis auf eine erst „kürzlich“ erfolgte Duplikation eines RCL-Exons könnten die Serpin-Gene *Ci-Spn-9* und *Ci-Spn-10* geben (Abschnitt 5.6.3). Ein Vergleich zeigte hier, dass die RCL-Sequenzen von *Ci-Spn-9* und der Variante A von *Ci-Spn-10* eine Aminosäuresequenz-Identität von ca. 80 % aufweisen (Abbildung 48). Aufgrund der Organisation der Gene in einem Gencluster, erscheint eine Exon-Duplikation infolge eines Rekombinationsprozesses wahrscheinlich.

¹ Jiang, H., Wang, Y., Kanost, M., R., 1994.

² Krueger O., Ladewig J., Koester K., Ragg H., 2002.

Ein Beispiel für das Entstehen von RCL-Pseudoexons könnte das Serpin-Gen *Ce-Srp-9* von *C. elegans* sein (Abschnitt 6.2.2). Hier konnte im 1457 Bp großen Intron, das die beiden RCL-kodierenden Exon-Varianten trennt, eine für RCLs charakteristische aber bruchstückhafte Aminosäuresequenz identifiziert werden (Abbildung 43):

```
Ce-Srp-9 Pseudo-RCL: VVTDGEKAGPIVYFLKCQSLLTW-----RKAD-PFRMLGVLTRVNM MYFEAVIVSRAVFNHPSFTMC
Ce-Srp-9A          VDKMGTAAAVTSLMYTDGVPQLPEHVEFWADHPP-ME-VLTHNNHPMFMGIYN
```

Abbildung 43: RCL-Pseudoexon von *Ce-Srp-9*

Alignment der ORF-Sequenz des RCL-Pseudoexons (*Ce-Srp-9* Pseudo-RCL, oben) mit der ORF-Sequenz der RCL-Exon-Variante I (*Ce-Srp-9A*, unten).

Trotz der teilweisen konservierten Aminosäuresequenz konnten in der Nähe des ORFs keine Spleißsignale (fehlende pyrimidin-reiche Sequenz und 3'-Spleißstelle) für ein Exon identifiziert werden. Damit handelt es sich wahrscheinlich um ein RCL-Pseudoexon.

Um eine Entstehung von Serpin-Genen mit RCL-kodierenden Exon-Varianten zu ermöglichen, scheint auch ein Intron in der unmittelbaren Nähe der RCL-kodierenden DNA-Sequenz notwendig zu sein. So nutzen die Serpin-Gene *Dm-Spn4*, *Ag-SRPN-10* und *Ms-Serpin-gene-1* jeweils die Intronposition 352a zum Spleißen der vorhandenen RCL-kodierenden Exon-Varianten. Die Serpin-Gene *Ce-Srp-7* und *Ce-Srp-9* von *Caenorhabditis elegans* sowie *Ci-Spn-10* von *Ciona intestinalis* nutzen das Intron an Position 339c zum Spleißen der alternativen Exon-Varianten.

Das Fehlen eines geeigneten Introns kann als Begründung dafür dienen, warum keines der Serpin-Gene von *Arabidopsis thaliana* multiple RCL-kodierende Exon-Varianten aufweist. Diese Serpin-Gene verfügen in der Regel über ein Intron (Position 149c), das aber in der Mitte der Gene lokalisiert ist. Für eine Bildung von alternativen RCL-kodierenden Exons müsste hier das zweite, mindestens 800 Bp große Exon dupliziert werden. Die Vorteile des alternativen Spleißens, Sequenz-sparend eine hohe Diversität zu ermöglichen, wären in diesem Falle deutlich geringer.

Einige Serpin-Gene von Vertebraten, wie etwa das humane Antithrombin III-Gen, verfügen ebenfalls über eine geeignete Intronposition (339c), die ein alternatives Spleißen von RCL-Exons ermöglichen könnte. Dennoch konnten bisher im menschlichen Genom keine Serpin-Gene identifiziert werden, die über multiple RCL-kodierende Exon-Varianten verfügen.

Vermutlich ist die Nutzung des alternativen Spleißens auch mit Nachteilen verbunden. So erscheint eine exakte Kontrolle der Expression der einzelnen Protein-Isoformen sehr aufwendig. Besonders wenn die Protein-Isoformen eines Gens an verschiedenen

physiologischen Prozessen beteiligt sind, kann sich die temporal-spatiale Expression der Isoformen als schwierig gestalten. Neben Duplikation und Diversifikation der RCL-Exons wäre auch eine Etablierung von *cis*- und *trans-acting* Spleiß-Kontrollelementen notwendig, die erst eine gezielte Auswahl einer der vorliegenden RCL-kodierenden Exon-Varianten ermöglicht (Abschnitt 2.4.1).

Wenn alle Protein-Isoformen am selben physiologischen Prozess beteiligt wären, z.B. Abwehr von Parasiten-Proteasen oder eine Schutzfunktion vor eigenen Proteasen(kaskaden), ist das Problem der Expressionskontrolle zu vernachlässigen. Hier könnte die Expression des Gen bzw. die Expression aller Genprodukte über die Promotorelemente des Gens kontrolliert werden.

7. Literaturverzeichnis

Adams, M., D., *et al.* 2000.

The genome sequence of *Drosophila melanogaster*.
Science 287, 2185-2195.

Altschul, S., F., Gish, W., Miller, W., Myers, E., W., Lipman, D., J., 1990.

„Basic local alignment search tool.“
J. Mol. Biol. 215, 403-410.

An, G., Watson, B., D., Chiang, C., C., 1986.

Transformation of Tobacco *Nicotiana ssp. tomato Lycopersicon esculentum*, potato *Solanum tuberosum* and *Arabidopsis thaliana* using binary Ti vector system.
Plant Physiol. 81, 301.

Atchley, W.,R., Lokot, T., Wollenberg, K., Dress, A., Ragg, H., 2001.

Phylogenetic analyses of amino acid variation in the serpin proteins.
Mol. Biol. Evol. 18, 1502-11.

Barnes, W., M., 1992.

The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion.
Gene 112, 29-35.

Barnes, W., M., 1994.

PCR amplification of up to 35-kb DNA with high fidelity and high yield from bacteriophage templates.
Proc. Natl. Acad. Sci. USA 91, 2216-2220.

Barnes, W., M., 1977.

Plasmid detection and sizing in single colony lysates.
Science 195, 393-394.

Bateman, A., *et al.* 2002.

The Pfam Protein Families Database
Nucleic Acids Research 30, 276-280.

Bernard, P., Gabant, P., Bahassi, E., M., Couturier, M., 1994.

Positive-selection vectors using the F plasmid *ccdB* killer gene.
Gene 148, 71-74.

Birnboim, H., C., Doly, J., 1979.

A rapid alkaline extraction procedure for screening recombinant plasmid DNA.
Nucleic Acids Research. 7, 1513-1523.

Blaxter, M., 1998.

Caenorhabditis elegans is a nematode.
Science 282, 2041-2006.

Bracher, F., 1994.

Seescheiden (Ascidien) - eine neue Quelle pharmakologisch aktiver Substanzen,
Pharmazie in unserer Zeit, 23. Jahrgang, Juni 1994.

Bruce, D., Perry, D., J., Borg, J., Y., Carrell, R., W., Wardell, M., R., 1994.

Thromboembolic disease due to thermolabile conformational changes of antithrombin Rouen-VI (187 Asn-->Asp)
J. Clin. Invest., 94, 2265-2274.

Brown, J., W., Smith, P., Simpson, C., G., 1996.

Arabidopsis consensus intron sequences.
Plant Mol. Biol. 32, 531-535.

Caceres, J., F., Kornblihtt, A., R., 2002.

Alternative splicing: multiple control mechanisms and involvement in human disease.
Trends Genet. 18, 186-193.

Carrell, C., and J. Travis, 1985.

α_1 -Antitrypsin and the serpins: variation and countervariation.
Trends Biochem. Sci. 10, 20-24.

Carrell, R., W., Lomas, D., A., 2002.

Alpha1-antitrypsin deficiency--a model for conformational diseases.
N. Engl. J. Med., 346, 45-53.

Carrell, R., W., Stein, P., E., 1996.

The biostructural pathology of the serpins: critical function of sheet opening mechanism.
Biol. Chem. Hoppe Seyler., 377, 1-17.

Celniker, S., E., *et al.* 2002.

Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence.
Genome Biol. 3(12):RESEARCH0079-9.

Celotto, A., M., Graveley, B., R., 2001.

Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated.
Genetics 159, 599-608.

Chester, A., Scott, J., Anant, S., Navaratnam, N., 2000.

RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA.
Biochim. Biophys. Acta. 1494, 1-13.

Christophides, G., K., *et al.* 2002.

Immunity-related genes and gene families in *Anopheles gambiae*.
Science 298, 159-165.

Danielli, A., Kafatos, F., C., Loukeris, T., G., 2003.

Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion.
J. Biol. Chem. 278, 4184-93.

- Dayhoff, M., O., Schwartz, R., M., Orcutt, B., C., 1978.
A model of evolutionary change in proteins.
In Atlas of Protein Sequence and Structure Vol. 5 suppl. 2 (ed. M.O. Dayhoff), pp. 345-352. National Biomedical Research Foundation, Washington DC.
- Dehal, P., *et al.* 2002.
The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins.
Science 298, 2157-2167.
- Denecke, J., De Rycke, R., Botterman, J., 1992.
Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope.
EMBO J. 11, 2345-2355.
- Eddy, S., R., 1998.
Profile hidden Markov models.
Bioinformatics 14, 755-763.
- Eickbush, T., H., 2000.
Molecular biology. Introns gain ground.
Nature, 404, 940-941.
- Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000.
Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.
J. Mol. Biol. 300, 1005-1016.
- Frey, P., A., Whitt, S., A., Tobin, J., B., 1994.
A low-barrier hydrogen bond in the catalytic triad of serine proteases.
Science 264, 1927-1930.
- Garcia-Hernandez, M., *et al.* 2002.
TAIR: a resource for integrated Arabidopsis data.
Funct. Integr. Genomics. 2, 239-253.
- Gattiker, A., Gasteiger, E., Bairoch, A., 2002.
ScanProsite: a reference implementation of a PROSITE scanning tool.
Applied Bioinformatics 1, 107-108.
- Gelfand, M., S., Dubchak, I., Dralyuk, I., Zorn, M., 1990.
ASDB: database of alternatively spliced genes.
Nucleic Acids Research 27, 301-302.
- Genest, M., Ptak, M., 1982.
Interactions between Asp, His, Ser residues within models of the active site of serine proteases. A theoretical empirical study.
Int. J. Pept. Protein Res. 19, 420-431.
- Gettins, P., G., W., Patston, P., A., Olson, S., T., 1996.
Serpins: structure, function and biology.
Springer-Verlag, Heidelberg .
- Giegerich, R., Meyer, F., Schleiermacher, C., 1996.
GeneFisher - Software Support for the Detection of Postulated Genes.
Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology, AAAI Press (ISSN 57735-002-2).

Gouet, P., Courcelle, E., Stuart, D., I., Metoz, F., 1999.

ESPrInt: multiple sequence alignments in PostScript.

Bioinformatics 15, 305-308.

Grabowski, P., J., Zaug, A., J., Cech, T., R., 1981.

The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of *Tetrahymena*.

Cell 23, 467-476.

Graveley, B., R., 2002.

Sex, AGility, and the regulation of alternative splicing.

Cell 109, 409-412.

Guex, N., Peitsch, M., C., 1997.

SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling.

Electrophoresis 18, 2714-2723.

Hagenbuchner, K., 1999.

Marine Lebewesen als Pharmaquelle - Neue Tendenzen in der Naturstoffchemie

Chemieportal, <http://www.eduhi.at>.

Han, J., Zhang, H., Min, G., Kemler, D., Hashimoto, C., 2000.

A novel *Drosophila* serpin that inhibits serine proteases.

FEBS Lett. 468, 194-198.

Harter, C., Wieland, F., 1996.

The secretory pathway: mechanisms of protein sorting and transport.

Biochim. Biophys. Acta. 1286, 75-93.

Henikoff, S., Henikoff, J., G., 1992.

Amino acid substitution matrices from protein blocks.

Proc. Natl. Acad. Sci. USA 89, 10915-10919.

Hill, R., E., Hastie, N., D., 1987.

Accelerated evolution in the reactive centre regions of serine protease inhibitors.

Nature 326, 96-9.

Hofmann, K., Bucher P., Falquet L., Bairoch A., 1999.

The PROSITE database, its status in 1999.

Nucleic Acids Research 27, 215-219.

Holt, R., A., *et al.* 2002.

The genome sequence of the malaria mosquito *Anopheles gambiae*.

Science 298, 129-149.

Hunt, L., T., Dayhoff, M., O., 1980.

A surprising new protein superfamily containing ovalbumin, antithrombin-III, and alpha 1-proteinase inhibitor.

Biochem. Biophys. Res. Commun. 95, 864-871

Huntington, J., A., McCoy, A., Belzar, K., J., Pei, X., Y., Gettins, P., G., Carrell, R., W., 2000a

The conformational activation of antithrombin. A 2.85-Å structure of a fluorescein derivative reveals an electrostatic link between the hinge and heparin binding regions.

J. Biol. Chem., 275, 15377-15383.

- Huntington, J., A., Read, R., J., Carrell, R., W., 2000b.
Structure of a serpin-protease complex shows inhibition by deformation.
Nature, 407, 923-926.
- Invitrogen, 2002.
Instruction Manual GeneRacer™ Kit.
Invitrogen GmbH, Karlsruhe.
- Invitrogen, 2002.
Invitrogen Manual, TOPO TA Cloning Kit for Sequencing.
Invitrogen GmbH, Karlsruhe.
- Irving, J., A., Pike, R., N., Lesk, A., M., Whisstock, J., C., 2000.
Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function.
Genome Res. 10, 1845-1864.
- Irving, J., A., Steenbakkens, P., J., Lesk, A., M., Op den Camp, H., J., Pike, R., N., Whisstock, J., C., 2002.
Serpins in prokaryotes.
Mol. Biol. Evol. 19, 1881-1890.
- Jiang, H., Wang, Y., Huang, Y., Mulnix, A., B., Kadel, J., Cole, K., Kanost, M., R., 1996.
Organization of serpin gene-1 from *Manduca sexta*. Evolution of a family of alternate exons encoding the reactive site loop.
J. Biol. Chem. 271, 28017-28023.
- Jiang, H., Wang, Y., Kanost, M., R., 1994.
Mutually exclusive exon use and reactive center diversity in insect serpins.
J. Biol. Chem. 269, 55-58.
- Jones, D., T., Taylor, W., R., Thornton., J., M., 1992.
The rapid generation of mutation data matrices from protein sequences.
Comput. Appl. Biosci. 8, 275-282.
- Kim, S., Woo, J., Seo, E., J., Yu, M., Ryu, S., 2001.
A 2.1 Å resolution structure of an uncleaved alpha(1)-antitrypsin shows variability of the reactive center and other loops.
J. Mol. Biol. 306, 109-119.
- Kozulic, B., 1995.
Models of gel electrophoresis.
Anal. Biochem. 231, 1-12.
- Krueger, O., Ladewig, J., Koester, K., Ragg, H., 2002.
Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes.
Gene 293, 97-105.
- Kwok, S., Higuchi, R., 1989.
Avoiding false positives with PCR.
Nature 339, 237-238.
- Lassmann, T., Sonnhammer, E., L., 2002.
Quality assessment of multiple alignment programs.
FEBS Lett. 529, 126-130.

- Lawrence, D., A., Ginsburg, D., Day, D., E., Berkenpas, M., B., Verhamme, I., M., Kvassman, J., O., Shore, J., D., 1995.
Serpine-protease complexes are trapped as stable acyl-enzyme intermediates.
J. Biol. Chem., 270, 25309-25312.
- Lawrence, D., A., Olson, S., T., Muhammad, S., Day, D., E., Kvassman, J., O., Ginsburg, D., Shore, J., D., 2000.
Partitioning of serpin-proteinase reactions between stable inhibition and substrate cleavage is regulated by the rate of serpin reactive center loop insertion into beta-sheet A.
J. Biol. Chem., 275, 5839-5844.
- Lehninger, A., L., Nelson, D., L., Cox, M., M., 1994.
Prinzipien der Biochemie.
Spektrum, Akad. Verl., 2. Aufl.
- Li, W., H., 1997.
Molecular evolution.
Sinauer Associates, Inc., Publishers.
- Lomas, D., A., Carrell, R., W., 2002.
Serpinoopathies and the conformational dementias.
Nat. Rev. Genet., 3, 759-768.
- Lomas, D., A., Mahadeva, R., 2002
Alpha1-antitrypsin polymerization and the serpinopathies: pathobiology and prospects for therapy.
J. Clin. Invest. 110, 1585-1590.
- Lundberg, K., S., Shoemaker, D., D., Adams, M., W., Short, J., M., Sorge, J., A., Mathur, E., J., 1991.
High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*.
Gene 108, 1-6.
- Marshall, C., J., 1993.
Evolutionary relationships among the serpins.
Philos. Trans. R. Soc. Lond. B. Biol. Sci. 342, 101-119.
- McDonell, M., W., Simon, M., N., Studier, F., W., 1977.
Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels.
J. Mol. Biol. 110, 119-146.
- Meinke, D., W., Cherry, J., M., Dean, C., Rounsley, S., D., Koornneef, M., 1998.
Arabidopsis thaliana: a model plant for genome analysis.
Science 282, 662, 679-682.
- Morgenstern, B., 1999.
DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment.
Bioinformatics 15, 211 - 218.
- Morgenstern, B., Atchley, W., R., Hahn, K., Dress, A., 1998.
Segment-based scores for pairwise and multiple sequence alignments.
Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 115-121.
- Morgenstern, B., Frech, K., Dress, A., Werner, T., 1998
DIALIGN: finding local similarities by multiple sequence alignment.
Bioinformatics 14, 290-294.

- Morgan, T., H., 1919.
The Physical Basis of Heredity
J. B. Lippincott Company, <http://www.esp.org/books/morgan/physical-basis/facsimile/title3.html>
- Morris, S., C., 1998.
Metazoan phylogenies: falling into place or falling to pieces? A palaeontological perspective.
Current Opinion in Genetics & Development 8, 662-667.
- Mount, S., M., 1996.
AT-AC introns: an ATtACk on dogma.
Science 271, 1690-1692.
- Muller, T., Vingron, M., 2000.
Modeling amino acid replacement.
J. Comput. Biol. 7, 761-776.
- Mullis, K., B., Faloona, F., A., 1987.
Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.
Methods Enzymol. 155, 335-350.
- Nakai, K., Kanehisa, M., 1992.
A knowledge base for predicting protein localization sites in eukaryotic cells.
Genomics 14, 897-911.
- Nei, M., Kumar, S., 2000.
Molecular evolution and phylogenetics.
Oxford University Press.
- Newton, C., R., Graham, A., 1994.
PCR
Spektrum Akademischer Verlag GmbH Heidelberg, Berlin, Oxford.
- Nicholas, K., B., Nicholas, H., B., Jr., Deerfield, D., W., II., 1997.
GeneDoc: Analysis and Visualization of Genetic Variation.
EMBNEW.NEWS 4, 14.
- Nixon, K., C., Davis, J., I., Goloboff, P., A., 1998.
Abstract: Search strategies for large datasets: an example using rbcL.
1998 Annual Meeting of the Botanical Society of America 2-6 August, 1998 Baltimore, MD.
- Page, R., D., M., 1996.
TREEVIEW: An application to display phylogenetic trees on personal computers.
Computer Applications in the Biosciences 12, 357-358.
- Patterson, J.T. 1943
Studies in the genetics of Drosophila.
Univ. Texas Publ. 4313. 7-216
- Plotnick, M., I., Mayne, L., Schechter, N., M., Rubin, H., 1996.
Distortion of the active site of chymotrypsin complexed with a serpin.
Biochemistry 35, 7586-7890.

- Pruitt R., E., Meyerowitz, E.,M., 1986.
Characterization of the genome of *Arabidopsis thaliana*.
J. Mol. Biol. 187, 169-83.
- QIAGEN, 2002.
Gel Extraction Kit Protocol, QIAquick Spin Handbook, 2002.
QIAGEN GmbH, Hilden.
- QIAGEN, 2002.
Omniscript Reverse Transcriptase Handbook, 2002.
QIAGEN GmbH, Hilden.
- QIAGEN, 2002.
PCR Purification Kit Protocol, QIAquick Spin Handbook, 2002.
QIAGEN GmbH, Hilden.
- QIAGEN, 2002.
Taq PCR Handbook, 2002.
QIAGEN GmbH, Hilden.
- Ragg, H., Lokot, T., Kamp, P., B., Atchley, W., R., Dress, A., 2001.
Vertebrate serpins: construction of a conflict-free phylogeny by combining exon-intron and diagnostic site analyses.
Mol. Biol. Evol. 18, 577-584.
- Rasmussen, S., K., Dahl, S., W., Norgard, A., Hejgaard, J., 1996.
A recombinant wheat serpin with inhibitory activity.
Plant Mol. Biol. 30, 673-677.
- Reboul, J., *et al.* 2001.
Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*.
Nat. Genet. 27, 332-336.
- Risler, J., L., Delorme, M., O., Delacroix, H., Henaut, A., 1988.
Amino acid substitutions in structurally related proteins.
J. Mol. Biol. 204, 1019-1029.
- Robertson, A., S., Belorgey, D., Lilley, K., S., Lomas, D., A., Gubb, D., Dafforn, T., R., 2003.
Characterization of the necrotic protein that regulates the Toll-mediated immune response in *Drosophila*.
J. Biol. Chem. 278, 6175-6180.
- Rokas, A., Holland, P., W., 2000.
Rare genomic changes as a tool for phylogenetics.
Trends Ecol. Evol. 15, 454-459.
- Rose, T., M., Schultz, E., R., Henikoff, J., G., Pietrovski, S., McCallum, C.,M., Henikoff, S., 1998.
Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences.
Nucleic Acids Research 26, 1628-1635.
- Saiki, R., K., Gelfand, D., H., Stoffel, S., Scharf, S., J., Higuchi, R., Horn, G., T., Mullis, K., B., Erlich, H., A., 1988.
Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.
Science 239, 487-491.

- Sambrook, J., Russell, D., W., 2001.
Molecular cloning : A Laboratory Manual, 3rd Edition.
New York: Cold Spring Harbor Laboratory Press, 2001.
- Sanger, F., Nicklen, S., Coulson, A., R., 1977.
DNA sequencing with chain-terminating inhibitors.
Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Sasaki, T., 1991.
Patchwork-structure serpins from silkworm (*Bombyx mori*) larval hemolymph.
Eur. J. Biochem. 202, 255-261.
- Satou, Y., *et. al.* 2002.
A cDNA resource from the basal chordate *Ciona intestinalis*.
Genesis 33, 153-154.
- Schechter, I., Berger, A., 1967.
On the size of the active site in proteases. I. Papain.
Biochem. Biophys. Res. Commun. 27, 157-162.
- Seilacher, A., Bose, P., K., Pflüger, F., 1998.
Triploblastic animals more than 1 billion years ago: trace fossil evidence from India.
Science 282, 80-83.
- Silverman, G., A., *et al.* 2001.
The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature.
J. Biol. Chem. 276, 33293-33296.
- Simmen, M., W., Leitgeb, S., Clark, V., H., Jones, S., J., Bird A., 1998.
Gene number in an invertebrate chordate, *Ciona intestinalis*.
Proc. Natl. Acad. Sci. USA 95, 4437-4440.
- Sokal, R., R., Michener, C., D., 1958.
A statistical method for evaluating systematic relationships.
University of Kansas Science Bulletin, 38, 1409-1438.
- Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J., Spieth, J., 2001.
WormBase: network access to the genome and biology of *Caenorhabditis elegans*.
Nucleic Acids Research 29, 82-86.
- Stoltzfus, A., Logsdon, J., M., Jr, Palmer, J., D., Doolittle, W., F., 1997.
Intron „sliding“ and the diversity of intron positions.
Proc. Natl. Acad. Sci. USA 94, 10739-10744.
- Stone, S., R., Le Bonniec, B., F., 1997.
Inhibitory mechanism of serpins. Identification of steps involving the active-site serine residue of the protease.
J. Mol. Biol., 265, 344-362.
- Swofford, D., L., 1998.
PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
Version 4. Sinauer Associates, Sunderland, Massachusetts.

- The Arabidopsis Genome Initiative, 2000.
Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.
Nature 408, 796 - 815.
- The *C. elegans* Sequencing Consortium, 1998.
Genome Sequence of the Nematode *Caenorhabditis elegans*. A Platform for Investigating Biology.
Science 282, 2012-2018.
- The FlyBase Consortium, 2003.
The FlyBase database of the *Drosophila* genome projects and community literature.
Nucleic Acids Research 31, 172-175.
- The International Human Genome Mapping Consortium, 2001.
A physical map of the human genome.
Nature 409, 934 - 941.
- Thompson, J., D., Higgins, D., G., Gibson, T., J., 1994.
CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.
Nucleic Acids Research 22, 4673-4680.
- Todd, W., H., *et al.* 2003.
WormBase: a cross-species database for comparative genomics
Nucleic Acids Research 31, 133-137.
- Vaglio, P., Lamesch, P., Reboul, J., Rual, J., F., Martinez, M., Hill, D., Vidal, M., 2003.
WormDB: the *Caenorhabditis elegans* ORFeome Database.
Nucleic Acids Research 31, 237-240.
- van Gent, D., Sharp, P., Morgan, K., Kalsheker, N., 2003.
Serpins: structure, function and molecular evolution.
Int. J. Biochem. Cell Biol. 35, 1536-1547.
- Venter, J., C., *et al.* 2001.
The Sequence of the Human Genome.
Science 292, 1304-1351.
- Wägele, J., W., 2000.
Grundlagen der phylogenetischen Systematik.
Verlag Dr. Friedrich Pfeil, München.
- Whisstock, J., C., Irving, J., A., Bottomley, S., P., Pike, R., N., Lesk, A., M., 1999.
Serpins in the *Caenorhabditis elegans* genome.
Proteins 36, 31-41.
- Wu, S., Manber, U., 1992.
Agrep – a fast approximate pattern-matching tool.
Proceedings of USENIX Technical Conference, San Francisco, CA, Usenix Association, 1991, pp. 153–162.
- Yoo, B., C., *et al.* 2000.
Characterization of *cucurbita maxima* phloem serpin-1 (CmPS-1). A developmentally regulated elastase inhibitor.
J. Biol. Chem. 275, 35122-35128.

Zang, X., Maizels, R., M., 2001.

Serine proteinase inhibitors from nematodes and the arms race between host and pathogen.

Trends Biochem. Sci. 26, 191-197.

Zang, X., Yazdanbakhsh, M., Jiang, H., Kanost, M., R., Maizels, R., M., 1999.

A novel serpin expressed by blood-borne microfilariae of the parasitic nematode *Brugia malayi* inhibits human neutrophil serine proteinases.

Blood 94, 1418-1428.

Zhou, A., Carrell, R., W., Huntington, J., A., 2001.

The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop.

J. Biol. Chem. 276, 27541-27547.

Zon, L., I., Dorfman, D., M., Orkin, S., H., 1989.

The polymerase chain reaction colony miniprep.

Biotechniques 7, 696-698.

8. Anhang

8.1 Veröffentlichungen

Teile der vorliegenden Arbeit wurden publiziert:

Krueger, O., Ladewig, J., Koester, K., Ragg, H., 2002.

Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes.

Gene 293, 97-105.

8.2 Abkürzungsverzeichnis

Ag	Abkürzung für <i>Anopheles gambia</i> .
AGREP	Suchprogramm, Abkürzung für approximate grep.
As	Abkürzung für Aminosäuren.
At	Abkürzung für <i>Arabidopsis thaliana</i> .
Bp	Abkürzung für Basenpaare.
Bm	Abkürzung für <i>Brugia malayi</i> .
Breach	Bezeichnung für eine Region in der Serpinstruktur, die den RCL „breschenartig“ auf die Serpin-Oberfläche leitet.
cDNA	Abkürzung für complementary DNA, entspricht der transkribierten mRNA.
Ce	Abkürzung für <i>Caenorhabditis elegans</i> .
Ci	Abkürzung für <i>Ciona intestinalis</i> .
ECM	Abkürzung für extrazelluläre Matrix.
ER	Abkürzung für endoplasmatisches Retikulum.
EST	Abkürzung für „enhanced sequence tag“.
Dm	Abkürzung für <i>Drosophila melanogaster</i> .
Hinge	Bezeichnung für eine Region in der Serpinstruktur, die als „Scharnier“ das Einklappen des β -Faltblattstranges s4A nach Spaltung des RCL erlaubt.
Indels	Abkürzung für Insertion / Deletion.

IRD800	Abkürzung für einen „infra-red dye“-Farbstoff der zur Markierung von Sequenzierprimern verwendet wird.
NJ	Abkürzung für Neighbor-Joining.
Nt	Abkürzung für Nukleotide.
ML	Abkürzung für Maximum-Likelihood.
MP	Abkürzung für Maximum-Parsimony.
Ms	Abkürzung für <i>Manduca sexta</i> .
Multifurcation	Eine Knoten mit mehr als drei Kanten bzw. Ästen.
ORF	Abkürzung für „open reading frame“.
OTU	Abkürzung für „operational taxonomic units“.
PCR	Abkürzung für „polymerase chain reaction“.
Serpine	Abkürzung für „serine protease inhibitor“.
RCL	Abkürzung für „reactive centre loop“.

<u>Aminosäure</u>	<u>Abkürzungen</u>		<u>DNA Codons</u>
Alanin	Ala	A	GCA, GCC, GCG, GCT
Cystein	Cys	C	TGC, TGT
Aspartat	Asp	D	GAC, GAT
Glutamat	Glu	E	GAA, GAG
Phenylalanin	Phe	F	TTC, TTT
Glycin	Gly	G	GGA, GGC, GGG, GGT
Histidin	His	H	CAC, CAT
Isoleucin	Ile	I	ATA, ATC, ATT
Lysin	Lys	K	AAA, AAG
Leucin	Leu	L	TTA, TTG, CTA, CTC, CTG, CTT
Methionin	Met	M	ATG
Asparagin	Asn	N	AAC, AAT
Prolin	Pro	P	CCA, CCC, CCG, CCT
Glutamin	Gln	Q	CAA, CAG
Arginin	Arg	R	CGA, CGC, CGG, CGT
Serin	Ser	S	TCA, TCC, TCG, TCT, AGC, AGT
Threonin	Thr	T	ACA, ACC, ACG, ACT
Valin	Val	V	GTA, GTC, GTG, GTT
Tryptophan	Trp	W	TGG
Tyrosin	Tyr	Y	TAC, TAT
Stopp-Codons		* oder .	TAA, TAG, TGA

8.3 Abbildungsverzeichnis

Abbildung 1: Proteinstruktur von α_1 -Antitrypsin (gespannt, nativ).....	4
Abbildung 2: Reaktionsmechanismus der katalytischen Triade.....	5
Abbildung 3: „ <i>branched-pathway</i> “-Modell.....	7
Abbildung 4: Physiologische Funktionen ausgewählter human-Serpine	9
Abbildung 5: Symbolische Darstellung der Bindestellen der DM_SP4-Primer	26
Abbildung 6: RLM-RACE-Methode (Invitrogen GeneRacer TM Kit) ¹	32
Abbildung 7: Vektorsystem pCR4-TOPO mit Vektorelementen.....	36
Abbildung 8: Funktionsprinzip der „TOPO TA Klonierung“	37
Abbildung 9: Konsensussequenz der Serpin-Signatur	41
Abbildung 10: Fadenwurm <i>Caenorhabditis elegans</i>	50
Abbildung 11: Chromosomale Lokalisation der Serpin-Gene von <i>C. elegans</i>	53
Abbildung 12: Genstrukturen der Serpin-Gene von <i>Caenorhabditis elegans</i>	55
Abbildung 13: Maximum-Likelihood-Baum der Serpine von <i>C. elegans</i>	59
Abbildung 14: <i>Drosophila melanogaster</i>	60
Abbildung 15: Lage der Serpin-Gene auf den <i>Drosophila</i> -Chromosomen 2 und 3.....	66
Abbildung 16: Strukturen der Serpin-Gene von <i>Drosophila</i> auf Chromosom 2 (Teil 1).....	67
Abbildung 17: Strukturen der Serpin-Gene von <i>Drosophila</i> auf Chromosom 2 (Teil 2).....	68
Abbildung 18: Strukturen der Serpin-Gene von <i>Drosophila</i> auf Chromosomenarm 3L	69
Abbildung 19: Strukturen der Serpin-Gene von <i>Drosophila</i> auf Chromosomenarm 3R	69
Abbildung 20: Neighbor-Joining-Baum der Serpine von <i>Drosophila</i>	74
Abbildung 21: <i>Arabidopsis thaliana</i>	75
Abbildung 22: Organisation der Serpin-Gencluster in <i>Arabidopsis thaliana</i>	80
Abbildung 23: Struktur der Serpin-Pseudogene AT1G62160.1 und AT1G65385.1	81
Abbildung 24: Chromosomale Lokalisation der Serpin-Gene von <i>Arabidopsis</i>	82
Abbildung 25: Genstrukturen der Serpin-Gene von <i>Arabidopsis thaliana</i>	84
Abbildung 26: Maximum-Likelihood-Baum der Serpine von <i>Arabidopsis thaliana</i>	87
Abbildung 27: <i>Ciona intestinalis</i> , Larven Stadium (A) und adultes Stadium (B)	88
Abbildung 28: Genstrukturen der Serpin-Gene von <i>Ciona intestinalis</i>	92
Abbildung 29: Maximum-Likelihood-Baum der Serpine von <i>Ciona intestinalis</i>	96
Abbildung 30: Neighbor-Joining-Baum von 65 Serpin-Sequenzen.....	100
Abbildung 31: Maximum-Likelihood-Baum von 65 Serpin-Sequenzen	101

Abbildung 32: Maximum-Parsimony-Baum von 65 Serpin-Sequenzen	102
Abbildung 33: Transkriptionsnachweis der vier RCL-kodierenden Exon-Varianten	105
Abbildung 34: PCR-Amplifikation der unbekanntenen 5'-UTRs von <i>Dm-Spn4</i>	108
Abbildung 35: Organisation des Gens <i>Dm-Spn4</i> von <i>Drosophila melanogaster</i>	111
Abbildung 36: Organisation der Serpin-Gene <i>Ce-Srp-7</i> und <i>Ce-Srp-9</i>	113
Abbildung 37: Organisation des Gens <i>Ci-Spn-10</i> von <i>Ciona intestinalis</i>	115
Abbildung 38: Vergleich der Exon-Intron-Struktur von <i>Ag-SRPN10</i> und <i>Dm-Spn4</i>	117
Abbildung 39: Organisation des Gens <i>Ms-Serpin-Gene-1</i>	118
Abbildung 40: Alignment der Protein-Varianten der alternativ-spleißbaren Serpine	120
Abbildung 41: Exon-Intron-Strukturen der alternativ-spleißbaren Serpin-Gene	120
Abbildung 42: „82a/85a-Intron“-Sequenz des Serpin-Gens <i>Ce-Srp-9</i>	125
Abbildung 43: RCL-Pseudoexon von <i>Ce-Srp-9</i>	137
Abbildung 44: DNA-Sequenz des Serpin-Gens <i>Dm-Spn4</i>	161
Abbildung 45: Alignment der neun Serpine von <i>Caenorhabditis elegans</i>	162
Abbildung 46: Alignment der 27 Serpine von <i>Drosophila melanogaster</i>	166
Abbildung 47: Alignment der neun Serpine von <i>Arabidopsis thaliana</i>	167
Abbildung 48: Alignment der zehn Serpine von <i>Ciona intestinalis</i>	168

8.4 Tabellenverzeichnis

Tabelle 1: Genomprojekte - Anzahl der Serpin-Gene pro Genom	2
Tabelle 2: Vergleich der Serpin-Klassifikationssysteme	12
Tabelle 3: Methoden der Baumkonstruktion	13
Tabelle 4: Evolutionsmodelle für DNA-Sequenzen	14
Tabelle 5: Aminosäure-Substitutionsmatrizen	14
Tabelle 6: Durchschnittliche Intron-Anzahl eukaryotischer Genome	16
Tabelle 7: Intronotypen	16
Tabelle 8: Alternative Spleißvorgänge	18
Tabelle 9: Verwendete Primer für <i>Drosophila melanogaster</i> Serpin-Gen <i>Dm-Spn4</i>	25
Tabelle 10: Verwendete <i>Dm-Spn4</i> -spezifische Sequenzierprimer	27
Tabelle 11: Ansatz für eine Standard-PCR von 20 µL	29
Tabelle 12: Standard PCR-Bedingungen, mit Pseudo-Hot Start	30

Tabelle 13: Ansatz für die Kolonie-PCR, 20 µL.....	30
Tabelle 14: Allgemeiner Restriktionsansatz, 10 µL.....	39
Tabelle 15: Mastermix, 21 µL.....	39
Tabelle 16: ddNTP-Reaktionsansatz, 5 µL (4 Reaktionen pro Sequenzierung).....	39
Tabelle 17: Sequenzier-Programm.....	40
Tabelle 18: Serpin-Suchmuster.....	43
Tabelle 19: Serpin-Gene von <i>Caenorhabditis elegans</i>	51
Tabelle 20: Intronpositionen der Serpin-Gene von <i>Caenorhabditis elegans</i>	56
Tabelle 21: Die Serpin-Gene von <i>Drosophila melanogaster</i>	62
Tabelle 22: Intronpositionen der Serpin-Gene von <i>Drosophila melanogaster</i>	70
Tabelle 23: Serpin-Gene von <i>Arabidopsis thaliana</i>	76
Tabelle 24: Die Serpin-Pseudogene von <i>Arabidopsis thaliana</i>	78
Tabelle 25: Intronpositionen der Serpin-Gene von <i>Arabidopsis thaliana</i>	85
Tabelle 26: Die Serpin-Gene von <i>Ciona intestinalis</i>	90
Tabelle 27: Intronpositionen der Serpin-Gene von <i>Ciona intestinalis</i>	93
Tabelle 28: Referenz-Serpine für die Interspezies-Stammbäume.....	97
Tabelle 29: Transkriptionsnachweis der vier RCL-kodierenden Exon-Varianten.....	105
Tabelle 30: Klonierung der unbekannt 3'-UTRs von <i>Dm-Spn4</i>	106
Tabelle 31: Klonierung der unbekannt 5'-UTRs von <i>Dm-Spn4</i>	108
Tabelle 32: Transkriptionsanalyse des Gens <i>Dm-Spn4</i>	110
Tabelle 33: Übersicht zu den Serpin-Genen aus den analysierten Genomprojekten.....	121

8.5 Sequenzen und Alignments

8.5.1 Aminosäuresequenzen der Serpine im FASTA-Format

Die aufgrund der Überprüfung der Serpin-Gene vorgenommenen Änderungen der Aminosäuresequenzen sind grau hervorgehoben.

Das Referenz-Serpin reifes, humanes α_1 -Antitrypsin

```
>hum-a1AT
EDPQGDAAQKTDTSHTDQDHTFNKITPNLAEFASFSLYRQLAHQSNSTNIFFSVPSIATAFAMLSLGTKADTHDEILEGLNFNLTETIPEAQIHE
GFQELRLTLNQPDSQLQTLTGNLFLSEGLKLVKDFLEDVKKLYHSEAFVFNFGDTEEAKKQINDYVEKGTQGVKIDVLKELDRDTVFALVNYI
FFKKGWERPFVEKDTTEEDFHVDQVTTVKVPMKRLGMFNIQCHKLSSWVLLMKYLGDATAIFFLPDEGKQLHLENELTHDITTKFLENEDRR
SASLHLPKLSITGTDLKSVLGLGKITKVSNGADLSGVTEEAPLKLKSAVHKAVLTIDEKGTAAAGAMFLEAIPMSIPPEVKFNKPFVFLMIE
QNTKSPLFMGKVVNPTQK
```

Die Serpine von *Caenorhabditis elegans* (Abschnitt 5.1.2)

```
>Ce-Srp-1
MSFSLINSTFTEAQFGIKLLSDLTSDQLTPCVFSPVSIILSLALVHLGAKGTRHDIRNSVNVNGSTDEQFIEHFSFINKLLNSSVNDVETLIAN
RLFVSPEQAIRKAFTDELREHYNAETATIDFKKSQEAAKIMNQFISESTKGKIPDMIKPDNLKDVDAILINAIFFQGDWRKFGPEAESNFSIS
ATENRLVPLRETRDYFNKDDDEWQVIGIPFKDKSAWFAIFLPTRRFALAENLKSLSNAAKFHNLINNVEYQYEIFLTFPKFKMDYKINLKTALAK
FGLAELFTEQADLSGIGPLQLASATHQALIEVDQVGTAAAAATEAKIFFTSASDEPLHIRVDHPFLFAIKDNSPLFLGTYYT
>Ce-Srp-2
MSDNATSQTDFAKLLLATLPHSGSVVLSPLSISLGLALIHAGACGSTQKLELDVGGSRIFEEFSGLMFAVGDTDNGVETKIVNRVFNQAYTI
HQDYLETVEKLYKASGESLDFSQTEQAAKTMTNFVENHTNGKIKDLIPADSANNAFAFLVNAVYFKADWQSKFAKESTTGREFFTSEAESRQIP
FLTELDEHRDYTEDVLFQVLSLKYADPKFTLAIIFLPKQRFGLVDALEKINGEYIQNLLNDLKSYSVSVQIPKFKIEKELDLKETLEAIGIKEIF
AEGADLSGIADKVFISSGIHKALIEVDEEDGTTAAAASAFKVLQEMMIMAEPTQFVADHPFLFAVLFENHTLFLGVHA
>Ce-Srp-3
MFDVAERRFALNLTLPVHNESELVFSPLSIALVLSLVHTGVRGSSRDQIRNTLLSGATDEQLEVEHFSFVSKEVKNKGTGVEVYLVANKVYLLKKG
FTVNPTFLSTALKNYGADAKSLDLTPPAVQVEINSFVNTATNGKIKNIATQDSIKDAIALLINSIYFKADWDDKFDGMSVSEQDFTLHTGEGKK
IKFMKEFMNDRSFSDDVDFDLVAVSDQRYQFSVFLPKLRNSLKEALKLNKRFNDLLKTKRRTFMNTQLPKFTIEKDLNLKSHLQTLGITD
IFSDSADLSGLAENLKIIEGVHKAIEVNEEGTTAAAVTMMKAVPMSARMEQPVNFIADHPFFFTITFLNHPFIVGVFNG
>Ce-Srp-5
MALFLQSETDFGLSLLRQQLNTESEFVSPFSIALALS LVHVAAGGETRDEIRKALLNGATDEELEGHFNSISAGLLAAEKGTETIKKHLKRLTI
SSVKIPKVTSKQLSTRVSVNVYLVAVLTNALYFKANWQTKFKKESTYKREFFSSENSKRETEFLHSRNSNRKYSENGQFQVLSLPHYKDTSFAL
SIFLPKTRFGLSEALQNLDSVTIQQLMSNTSNTLVNIAAMPKWKIETALGLNRLMAVGIIEKAFDTSADLSNFADGIYISQAHHKALIEVDEEDGT
VAAAATTISSVGSVQRTEKPIEFETADHPFLFLLSKDNHPLFVIGVHN
>Ce-Srp-6
MSDSSSDEKMLLLHSETDFGLSLLRQQLNTESEFVSPFSIALALS LVHVAAGGETRDEIRKALLNGATDEELEGHFNSISAGLLVAEKGTETV
VANHIFSRKFTTIKLYLNDVKKLYNAGASQLNPFEDQEASAEAINNFVSENTKGHIKKIINPDSISEELVAVLTNALYFKANWQTKFKKESTYK
REFFSENSKRETEFLHSRNSNRKYSENGQFQVLSLPHYKDTSFALSIFLPKTRFGLSEALQNLDSVTIQQLMSNTSNTLVNIAAMPKWKIETALG
LNRALMAVGIIEKAFDTSADLSNFADGIYISQAHHKALIEVDEEDGTVAAAATTISFSLTSVFI PAEPIEFETADHPFLFLLSKDNHPLFVIGIHN
>Ce-Srp-7A
MALLQSETDFGLGLLRQQLNTESESLAFSPLSIALALS LVHVAAGGETRQIREALVKGSTDEQLEQHFANISAALLAAERGTEVLANHVFTRA
GFKIKQSYLDDVKKLYNAGASSLDFDNKEATAEAINNFVRENTGDHIKKIIGSDSINSDLVAVLTNALYFKADWQNKFKKDSTFKSEFFSSADS
KREIDFLHASSVSRDYAENDQFQVLSLPHYKDNFTALTIIFLPKTRFGLTESLKTLDSATIQHLLSNVSSTSVNVQIPKWKIETKLGLEALQSLG
IKKAFDNDADLGNMADGLYVSKVTHKALIEVDEEGTKAAAATTVSISLKSAMFVMEEPKDFETADHPFFVLSKDNHPLFVGLHH
>Ce-Srp-7B (alternatives Spleißen)
MALLQSETDFGLGLLRQQLNTESESLAFSPLSIALALS LVHVAAGGETRQIREALVKGSTDEQLEQHFANISAALLAAERGTEVLANHVFTRA
GFKIKQSYLDDVKKLYNAGASSLDFDNKEATAEAINNFVRENTGDHIKKIIGSDSINSDLVAVLTNALYFKADWQNKFKKDSTFKSEFFSSADS
KREIDFLHASSVSRDYAENDQFQVLSLPHYKDNFTALTIIFLPKTRFGLTESLKTLDSATIQHLLSNVSSTSVNVQIPKWKIETKLGLEALQSLG
IKKAFDNDADLGNMADGLYVSKVTHKALIEVDEEGTTAAAATTVSVFVRKSARPSAIKPIVQADHPFLFMLTQTNHPFIFLGIHQ
>Ce-Srp-7C (alternatives Spleißen)
MALLQSETDFGLGLLRQQLNTESESLAFSPLSIALALS LVHVAAGGETRQIREALVKGSTDEQLEQHFANISAALLAAERGTEVLANHVFTRA
GFKIKQSYLDDVKKLYNAGASSLDFDNKEATAEAINNFVRENTGDHIKKIIGSDSINSDLVAVLTNALYFKADWQNKFKKDSTFKSEFFSSADS
KREIDFLHASSVSRDYAENDQFQVLSLPHYKDNFTALTIIFLPKTRFGLTESLKTLDSATIQHLLSNVSSTSVNVQIPKWKIETKLGLEALQSLG
IKKAFDNDADLGNMADGLYVSKVTHKALIEVDEEDGTVAVAATRCSEIERCRKMKKENIEFHAHEPFFFFILHHGTSYIFLGVFTG
>Ce-Srp-8 (carboxyterminales Exon ergänzt)
MHEFLSKSEAEFALKLLHQDHSQSFVFSPIAISLALYSLYEAARGETRSQIHDCFLKFNATDGMVTHFSNVSTALSAMKHKDPQVKICNHIICR
NGIKTESLCLDKVKKLYNCEVSSHLDNEKETIKTINKLIRENTDGHIKKINTKISIDKDQVVVFLSAACFEAQWQNKFKLDSTSKKEFFCTAN
SHRIKFLHATGANRKYCENDKQVLSLPHYTDSSEFMTIFLPKQFGLAEALKTLDGTSTIQELKSNVSNYLVNVQIP IWRNETEIDLNSTLQAI
GITKILNESAYIGNFAENVHISEFIHKALIEVNEENGTSPTANTKVIKTRWREETGEIRDFIADHPFLYTIHYKNSILFMGVFAG
>Ce-Srp-9A (ein zusätzliches Intron identifiziert)
MAVTPVTRCTYLLRDELDFALKLLHQDQLTKSFVSPMSISMAMALAHAGAKDETREEIRNVLLKLTRETEHGAENVVANRLFIKNGYPKKSYLK
EVQKLYNASASNLNFDFTKRSKAVINKFASDNTKGGINKIISPEALKGKIVVICNAIYFKAKWLIKIFEDSWEVSVFSAKKTSRQVEFLHMCEMT
KFNENAHFKVISMPYKDNFALAIIFLPKRFALKKALKSLDSASILQLLTSFSECFVKIAIPKWSIETDMDIGALQDLGIKDAFDNCAANFED
MYTSAGGDPLYISKALHKALIEVDKMGTTAAAVTSLMYTDGVPQLPEHVEFWADHPFMFVLTNNHPMFMGIYN
```

>Ce-Srp-9B (alternatives Spleißen)

MAVTPVTRCTYLLRDELDFALKLLHQDLTKSFVSPMSISMAMALAHAGAKDETREEIRNVLLKRETEHGAENVVANRLFIKNGYPKKSYLE
EVQKLYNASASNLNFDKRSKAVINKFASDNTKGGINKIISPEALKGKVVICNAIYFKAKWLKIFEDSWEVSFFSAKTSRQVEFLHMCCEMT
KFNENAHFKVISMYPKDNFALAIPLPKRFRALKKALKSLDSASILQLLTSFSECFVKIAPKWSIETDMDIGAALQDLGKDAFDNCAANFED
MYTSAGGDPLYISKALHKAIIEVDKLGTTAAAVTSFDYEEELGVLPKVEFWADHPFVLTHTNNHQMFMIYNFISYPSFS

>Ce-Srp-10

MAVTPVTRCTYLLRDELDFALKLLHQDLTKSFVSPMSISMAMALAHAGANDETREEIRNVLLSGSTDSQLQKYISDLCEELRETEHGAEKRI
SKKILLKEVQKLYNASASNLNFDKRSKAEVINKFASDNTKGGINKIISPEALEGKVVICNAIYFKAKWLKIFEDSWEVPPFSSKNSSRRIEF
LHKSEMTKFNENAHFKVISMYPKDDNFALAIPLPKRFRALKKALKSLDSASILRLTSFSECFVKDLGKDAFDNCTANFEDLYTSAGGAPLYI
NKALHKAVIEVDKMGTTAAAVTSHNYVTLGAPQLPDHVEFWADHPFVLTHTNNHPMFMIH

Die Serpine von *Drosophila melanogaster* (Abschnitt 5.2.2)

>Dm-Spn27a

MTKMGGNLAVMLLSLFLSALATGNGNSIPTTTTTQGVFETRTDKLPGGAASVPSGAGIYDDIDTFVPPFRSDSHDPFVSHLLKTVLQNETADKNV
IISPFVSKLVALLAEAGAGTQTQVELANTQTDIRSQNNVREFYRKTLSNFKKENQLHETLSVRTKLFDSFIEQQKFTATLKHFDSEVEA
LDFTNPEAAADAINAWAANITQGRQLQVLAPDNVRSSVMLLTNLIYFNGLWRRQFATTFQGSFRRSKDDQSRAEFMEQTDYFYTTSEKLAQI
LRLPYKGNLSVFLPYALNGIHDLVKLENDLKSQAQWAMEEVKVKVTLPKFHFYDQQNLKETLRSLVGREIFEDSASLPLGLTRGADVAGVKV
VSNILQKAGINVNEKTEAYAATVVEIENKFGGSTAIEEFVNRPFVFFIEESTGNILFAGKVVHSPTTQN

>Dm-CG6717

MKYLCKRSTKSYFTLQFSLIPLDLLLATSVESEGFWEFYRILASQNAKRNLIYSPISAEIIMSMVYMASGGKTFEELRNVLKFSENKTLVAN
NYSRLSLLKRRRTFIIILHMANRIYVKNKYCLVPEFNQLARKAFKAKAKSIRLDDPVSAIVNSWILNRTGRMIRNIVLPKDFNSDTS AFLVN
AIYFKGQWLYNFKADQTHIADFYVSANEIIPVKMMTSLASLLSGYIDDIIDAKIIELPYWNSTLSMRIILPNSVDGLRKLKEKVGFDYHLEKKS
VNVKLPKFKIESKALGKIFENLGLDVFKPSADLNGLVLESGAKIDKIQKAFKLIDEKGEASAATGVLTRRKKSIDNLIQPPMEFIADHPF
FYVIHDNKVIYFQGHIVEPRW

>Dm-CG31902

(zwei zusätzliche Introns identifiziert)

MKYLWILVTTSVLQGTQKLYRSFLQDNKQYNIASPLCIEIGSMILMGADGNTANELRTALNLPEDKKNVATYDCLKLTKLERGKVAIILH
LANRFLVNETIGVNRKYNKLVNKHFRAEAEAIKLADRLKAAWAINDWLDQTLDNVKDIIIPDLTPDESAVMINAAFFKGYWKRFDKMTKP
KVFYVSKSYQVNVNMQSVGRFKMRTSTIDQIIELPFAYSNLSMIVLPKDNGSLTQAEATIESYPIVLTMDVHVQLPKFKIDFRMELVETL
KSMGIQDLFNSSSDISVLLNQSSTRISQVHVHKAFFIEIDEEGGSAGSASASPIRGLSDYATSVVTFVNSPFVFMIRDDDNIIYFRGRVVDPLKKS
NPIDRITI

>Dm-CG33121

(zwei Introns postuliert)

MQGNKIKYLVLLLIATSVLQKFKLNLELVMDKAESNFIASPLCIEIGSMILMGAGTAAEELRSVLDLPVDVTEMAKKEYERIMSNFQKHNG
LRFNTWLYVNETYEVQRDYNTLMKSTFMAEGKDPKSRKASNSISFSIHRKSHKGMRTISNDHNLQINESAVLVNTVYYSGAWKTRFSKDKTKL
KVFHGDHKKVYVRMMSHVGRFRADHSYGQIIEMPFDNSDLMIIGLPLHNTYLSSEIEKILRTLSESLVENNVHVELPKFKIKYQTELVESLK
KLGILHIFSNSTDLGGLTNGTGAKINHVVHKSFEINERGAESTGEASDAESIQKTRASTSFKVNRPFVFLIRDKHTVYFRGRVVRPLNELH

>Dm-CG7219

MWRLLALLLVSSVCESELFRDDLRTPETMAYINGLMQRRHQMQEAAQHQIQAIPPAVPLQSPGLVNGLGNQNDPALNRIISGTSVKPSNLPAA
YNSGYVDLATSDRIANSVLNFANILGQHLANGKTQIYSPLSIVHSLALLLGAAGRSYEEELSTVFDIPDTSRHLHQFGLMLQDLQOPTREAIISA
GRPLTDWRASSAMRSNRRAQRPGAHEVHLANGLFTQTGYTLNPDYRRVIVEVYASDLQIQDFEGSPATARYNINAYVAQHTKNHNIENIASDIP
QTRMILANALYFKAFWETDFIE SATRPDNFYPNEGTEPVMRVQMMATGGAYPYHEDHELKCKIIGLPYRGNLSTMYIIPFKSSVRELMAIQ
KRLTADKIESMISRMRYRAALVAFPKMHLTESVNLKTMQRMGLGGIFSAVQNDLSLIATNEATRINALGGNSLQNLQEAQRAGTGGARSDLVV
DDIVHKVDFTVNEQGTAAASSVYLLKSGPDVLFGRDTPFMVLRHDPKLVLFYGLINEPPAAA

>Dm-Spn2 CG8137 sp2

MKHLCLLLLATSVSRCRFADDFYQLLAKENANLIISSPLSVEIALSMA YGARAKTAQEMRNVLKLPPDDKKEVAAYKDLLSKLEGREKVATLS
LANRIYVNNKFKLVPSYNQMVKDSFMAEAEAEIDIVDPNKASSIVNWNVDNQRGKIKDLVSSNDMSKMEILVNLAIYFKGQWYKFNPKLTKKR
NFRVSDQKSVPPVEMSLFQSFRAAHDSELGAKIIELPYRNSLMLIFLPDQVLDGLSELEKIVGFKPKLSKMDVTLRLPKFKIEFFAQLNKL
VAMGIQDAFEKSAADKDLVENSNVHVKVIVHKAFFIEVNEEGAEAAAATALLFVRYSMPPSSQMVFNADHPFAVYIRDRETIYFQGHFVKPNE

>Dm-CG4804

MSEANPTPYDCHIGAGIYHSIATSFAEQNVVSPLLLEATLSLLFLGSDGATAEELQKQLRLKQRFASNAKMANFYAAELGNIITDADTFLQL
QNRMLSESGVADDFQKIAQTYFHATAECVDLEQTEKLRRIHQIILASVGGGSWKDIHVAGGSSANTLLLLLANLQSKWFLPFSAYRTGLY
EFHSGSQVKSVPMLFDDDMFVKFAELRDLARAIELEPYEHAEEELSMILLPNQRGGLQLEKQLHDLDLGALQRRMQMEGVQLLPKFSIDFEC
SLRQPLKQLGFEEIFAASANFKHLHASANLPIADVQLKLRINLNESSGSGPELKNATEYKPIVISNSSRQKFRADHPFFFAIRSENVTYLM
GHVVEF

>Dm-Spn3 CG9334 sp3

MKYLCLLLLATSVSRCRFADDFYQLLAKENADKNLITSPLSVEIALSMA YGARGKTAQEMRDVCLKLPDDKKEVAAYKDLLSKLEGRESVAIIS
LANRIYVNNKFKLVPEYNQMVKDSFMAEAEAEISANNPKITASIVNKWVDQTSGKIRDLVMPSDVANLVLVILNAIYFKGQWQKFNTEQTKSD
FHSIDQKSVPPVQMMSLVRPFGVSYDRELGANVIELPYRNSLMSVIFLPDKVDGLPELEKMMVGFTPKLIINVHLRLPKFKIEFSARLEQVLI
AMGIQDAFKTSADFNDLVANSAGHVGVVHKAFFIEVNEEGEAAAATAVVFYRYSIRSPMDFNVNHPFAVYIRDAENIYFQGHFVNPEL

>Dm-Spn4A (alternatives Spleißen)

MDYRLVPCGCWLLPLGLALFPFPVHTADVTMADAAHQEFARRLALFISINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQG
LGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVL
NSESRLLVNAIHFKGTWQHQAFAKHLTRPDTFHLDGERTVQVPMMSLKERFRYADLPALDAMALELPYKDSLSMLIVLPNTKTGLPALEKLR
LTTLSQITQSLYETKVALKLPKFAEFQVELSEVFKLGMRSRMSQAEFGKMLQSPPELKVSAI IHKAFIEVNEEGTEAAAATGMVRRKRAI
MSPEEPIEFFADHPFTYLVHQDLPLFWGSVVRLEENTFASSEHDEL

>Dm-Spn4B (alternatives Spleißen)

MDYRLVPCGCWLLPLGLALFPFPVHTADVTMADAAHQEFARRLALFISINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQG
LGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVL
NSESRLLVNAIHFKGTWQHQAFAKHLTRPDTFHLDGERTVQVPMMSLKERFRYADLPALDAMALELPYKDSLSMLIVLPNTKTGLPALEKLR
LTTLSQITQSLYETKVALKLPKFAEFQVELSEVFKLGMRSRMSQAEFGKMLQSPPELKVSAI IHKAFIEVNEEGTEAAAATGMVMCYASML
TFEPQPVQFHVQHPFNYYIINKDSTILFAGRINKF

>Dm-Spn4C (alternatives Spleißen)

MDYRLVPCGCWLLPPLLGLALFFPPVHTADVTMADAAHQEFARRLALFSINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQAFAKHLTRPDTFHLGDERTVQVPMMSLKERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEEKLR LTTLSQITQSLYETKVALKLRPFKAQEFQVELSEVFKLGM SRMFSQAEFGKMLQSPPEPLKVSAL IHKAFIEVNEEGTEAAAATGMFMSLTSLPMPKPDPIRFNVDPHPTFFYILNKDSTALFAGSIKKL

>Dm-Spn4D (alternatives Spleißen)

MDYRLVPCGCWLLPPLLGLALFFPPVHTADVTMADAAHQEFARRLALFSINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQAFAKHLTRPDTFHLGDERTVQVPMMSLKERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEEKLR LTTLSQITQSLYETKVALKLRPFKAQEFQVELSEVFKLGM SRMFSQAEFGKMLQSPPEPLKVSAL IHKAFIEVNEEGTEAAAATVWRVMVAFAA SRKHFIANHPFAFYVVKTHYDLPIFTGRYLG

>Dm-Spn4E CG9453 Sp4

MADAAHQEFARRLALFSINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQAFAKHLTRPDTFHLGDERTVQVPMMSLKERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEEKLR LTTLSQITQSLYETKVALKLRPFKAQEFQVELSEVFKLGM SRMFSQAEFGKMLQSPPEPLKVSAL IHKAFIEVNEEGTEAAAATGMVRRKRAIMSPPEPIEFFADHPFTYVLVHQDLPLFWGSVVRLEENTFASSEHDEL

>Dm-Spn4F (alternatives Spleißen)

MADAAHQEFARRLALFSINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQAFAKHLTRPDTFHLGDERTVQVPMMSLKERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEEKLR LTTLSQITQSLYETKVALKLRPFKAQEFQVELSEVFKLGM SRMFSQAEFGKMLQSPPEPLKVSAL IHKAFIEVNEEGTEAAAATGMVMCYASMLTFEPQPVQFHVQHPFNYYIINKDSTILFAGRI NKF

>Dm-Spn4G (alternatives Spleißen)

MADAAHQEFARRLALFSINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQAFAKHLTRPDTFHLGDERTVQVPMMSLKERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEEKLR LTTLSQITQSLYETKVALKLRPFKAQEFQVELSEVFKLGM SRMFSQAEFGKMLQSPPEPLKVSAL IHKAFIEVNEEGTEAAAATGMFMSLTSLPMPKPDPIRFNVDPHPTFFYILNKDSTALFAGSI KKL

>Dm-Spn4H (alternatives Spleißen)

MADAAHQEFARRLALFSINVYKLSGQKPGENIVFSPFSIQTCAAMARLGAENETATQLDQGLGLASSDPEQIAHSFHQVLAAYQDSQILRIANKIFVMDGYQLRQEFDQLLSKQFLSAAQSVDFSKNVQAAATINNWEQRTNHLIKDLVPADVLNSESRLVLVNAIHFKGTWQHQAFAKHLTRPDTFHLGDERTVQVPMMSLKERFRYADLPALDAMALELPYKSDLSMLIVLPNTKTGLPALEEKLR LTTLSQITQSLYETKVALKLRPFKAQEFQVELSEVFKLGM SRMFSQAEFGKMLQSPPEPLKVSAL IHKAFIEVNEEGTEAAAATVWRVMVAFAA SRKHFIANHPFAFYVVKTHYDLPIFTGRYLG

>Dm-CG9454 (ein zusätzliches Intron identifiziert)

MKDEEFAQGLEQFALCLHDHLCRASAGLNI IYSPLSIHI SAAMLRMGTSEGSATAKEMDEGLRFGGLEAQVAESFGVVLKSYEQCVLKMANG LYVMKGLQVDEQFGHILBQKFRSKPMEIDFGSEQAASI INKWESQTNLKI DI IGRVLTCKDSRLCLVNGIHFKEGWSISFNEKETREEDFFG SDRPTRVRMNYCENF7AVLPMFEATLRMNY SACNLAMI ILLPDEKSNLTSLEKCLSDI SLEVVSSAMNLEKVDVKIPSTAEFQQLSOLV MLMGMNRI FSGQAEELGMLQSEESL FVSQIVHKAFIEINEVGT EAAAAATAAVATFRSMPARQGPVKVFHANRPPFYAIKDNTHGLLFAGHFITT KVEQSEKNTCT

>Dm-CG9455

MSEPEQGRNQFARNLIDVITKDALQSKDPHINTVFPSPASVQSALTFLAFMGASGSTAEELRNLQLGPGDRHHIALNFGEFWRTSCNYGDRGPV LKSVNRLYVNDSELELLETFNEIAVDFFQSKAEATRFADSEGATQLINDVVEQTEHKITNLLQSDAVNNETSALLINVLYFKGWKQKPFMPETT SIDHFHVDTRDTHVQVNMNYQEDKFRFAELPQLKARAVQLPYDYSNIHMLILLPNEVNGLQEQLEQLNTVDLADIDAALTLQDVEIFLPRMCIEY DVDLKQVLNQLGITEVFSKAKLDGLFTSQSQQKISAAARHGYIDVNEAGSEAAAVSFMKIVPMLNMMNKKLFKADHPFVFYIRNPQAVFFAGR FSNPKSGSGSGEELSRGFDANMYNV

>Dm-Spn1 CG9456 sp1

MYYLICIFLWVTSVACQTSKEIYQLLSKSHNTQNLVVPVSIETILSMVFMGAEGSTAKELQSALGLPSEDKEA VAARYGALLNDLQGEQEPIL KLANRIYVNDQYSLNQYNLAVREPFKSEAESI SLTNGPVAAERINQWVLDQTSQGIKGMIDPGSMTSDVKALLVNAIYFKGQWESKFPDPAKTR ASTFQVTANKSVPVQMMQMGTFRANYFRDLDAQVIELPYLNSNLSMTIFLPREVEGLSALEEKIVGFARPLVAKEVYLKLPKFKIEFRDELKE TLEKLGIRELFTDKSDLSGLFADKSGGKVSQVSHKAFLEVNEEGEAAGATSVAVTNRAGFSTFLMADHPFAFVIRDANTYFQGRVVSP

>Dm-CG9460

MANTLNYSKSPAGEAQFASQLFGQLAKSQSGRNIVFSPSSIRTGLALAYLGAEGSTADELKLGLGLEGAGKTEVAEKLDQLLAKGQWEKASGDE DVPKLYANRIFVTRFKLTQTYQDLVSKNFAAAAENVNTQKADTAKHINSWVEEQTHQIKDLIAPESLDADTSAILVNAIYFKADWQSSFP DYATYASDFVNHGGRKVSVDTMWTLQNFNYAEVNSLDAKVVELPVTGTDIVFLIILPQEEQGLAIVEEKLMLIDLNEISSQLRRRKRVRVQLPKF KFEFDVPLQAAL EELGIKKLFS PANLSSLYQGSEPLRISSEVKKHAIIEVNEKGTASGATFIKVSVESLTI GEEVFEFIADHPFFFAIKDAQN TFLFLGHVSQL

>Dm-Spn43Aa CG12172

MNHWLSIILLGVWISAPFGLGNTIKDRNLFATELFTLATDRQDENVIISPVSIQLALGLAYYGAEGRTAAELQKTLHASAKESKDLAESYHN LLHSYIKSKTVLEIANKVYTRQNLTVSSHFREVAQKYFDSEVEPLDFSRETEAVEQINRWVKQQTENKIERVVESELEPDTNVALVNAIYFKARW ARPFNDETRDREFWLSESRSIQVPTMFADNWWYADYPELDAKAIELFFENINLTMWFLPNQRSGLQALEQKLGKGVDFNLLDRWQWQSVSV YLPKFKFEFDTLRPTLHKMGISAMFSDAADFSNIQDSPIGTRITKVQHKTFIDVNEIGCEAAGASYAAGVPMPSLPLDKPTFVADHPFAFIR DKHAVYFTGHIVKF

>Dm-Spn43Ab CG1865

MAVILISCLLLLLLATVVSQKTVGDAADRNLAAADLYNAVAADHLNENNVISPATIQSSMALAFVGAQGTASELQQGLRGLPGDADAVSQRSG SYQQALTRDNNFRLANNIYINENLEFKGSFRDVAQRQFDSNIDKLDHPPYKNRATADGINRAVATKTNGKITDILRAELLNDRTEGVIVNGVSY SAAWQKAFRLDKTEKRSFRTGSGQSVKVDTMWTLQNFNYAEVNSLDAKVVELPVTGTDIVFLIILPQEEQGLAIVEEKLMLIDLNEISSQLRRRKRVRVQLPKF KVEVLLPKFSVTFGLGLEGPFKLGVHTMFSRDGDFGNMYRMFVSHFINAVEHKANVEVTEAGVDQPLETGLLKGFLSRSKKFEADHPFFFAIK YKDSIAFIGHIANAYV

>Dm-CG1859

MIHWRLSALLVGLAIALTLPVDGELLARSPASVSNRFLRLTTKLGLTQPDANVVVSPLLIQAALSLLYAESSSEYGSQLRQALELTHASHP KLAVQDFETLLTDLQSAAGICRLRLSLLDYAQRFTFNFRNEFETLAARMGVGCHRLSWESASNAQDINYAFLSRSNFSLGEIVSAPQLESLE AEHNTPFHLVSGVTFRAPWAFDPTETQSINFAGGNRPLVDAMFGQHRYAEVPAALDAQLIEVPPATADLRMLIVFPNRPDGLAQLERKLAQSDLHLQRSQLEERKVALTPLKLRVLVHSDLKHVLEELGLAKLFTSEVHLSEVFSIILSSSAPPLGAVVQSGLLELQEDGGNADDSFSFGDLF RRALPLVINHPFFYIAGNKLTLLSGHIVDI

>Dm-Nec CG1857

MASKVSI LLLLVHLLAAQTFAQELIAWQRQQQQQQQQQLQLQQQLLQQQQHQRNRPPELGLRSLPGNPWTQNNQEAISDVVAVDLTKREPVT
PPPNRPPVFSYMDRFSSSELFKEI IKSQSQQNVVFSVSVHALLALIYGASDGKTFRELQKAGEFSKNAMAVAQDFESVIKYKKHLEGADLTLA
TKVYYNRELGGVNHYSYDEYAKFYFSAGTEAVDMQNAKDTAAKINAWVMDTTRNKIRDVLTPTDVPDQTQALLVNAVYFQGRWEHEFATMDTSPY
DFQHTNGRISKVAMFNDVYGLAELPELGATALELAYKDSATSMILLPNETTGLGKMLQQLSRPEFDLNRVAHRLRRQSVAVRLPKFQFEFE
QDMTEPLKNLGVHQMFTPNQVTKLMDQPVRVSKILQKAYINVGEAGTEASAASYAKFVPLSLPPKPTFVANRPFVFAVRTPASVLFIGHVEY
PTPMSV

>Dm-CG7722

MASGGTTAPSLASPIVFNRLFRALNDEVPPVNMVSPAGARSAMTLVFMGAGKSADELRSKILIGVSNKSEVAKQHAESWTDECSCAKKGV
ALRLVTRLYVNEEEKIRTDNFDMALEFFNAEAYSLNYLNPEDSVKKNKWELEKHTFYTVRNLFTPEVFNSSSVILVNSLFFRAKWNKIFPQQL
TQIDDFWLNPRQRMEVSMRQIGQFRYGESKLLKSQILQLPFRSNLTMMIILPTAIDGLPELEKLGQDMNEVAAKSLMKEVDVTIPKFRIE
CTVDLKVPLQKMGINSVFDAGQADLSDFEMKTPQKISEARHKVFLNVTEFGCEVAPEAEVQPEVLKKNPDRKFFKADRPVFAIRDKNVYFV
GHFVKP

>Dm-CG10956

MVYLLILLGISRYRAQKNSQLPELYAAIVNSFSNRNIMFSTEMIRSSMLFIYVGVEEDESEQIRKAMHYRGTHLSEYKPKTQKIFAMSVKKA
PVAKSLTRFYVRQNMKSTERYVFMRHTEGRARNIAFAREQLEVNFTFYSHEMGEQIGQVVKESWVKPNSQGLLVNAIFNLSWERTFNPEATY
PREFRVNAATKSMIPMMHEDSKFAFGILGNLKAATAVLVPFSGDLRMLLTKPDQDGLAALQMKLQAMNILLSVARNLTMMDFVFGIPKFKIHSD
LELSPAFKEMGIKIDIFKPSKSFSTLLHRNTNFRIDGVIHVVTFEFQEQGIGTPTDVG**NGSLTHTFNGVKYFLATHPFAFYI**IDNTSIYFAGH
TSE

>Dm-Spn6 CG10913 Sp6

MFDNFLEFARGGARFTSELFQLLSAGGLKENVVFSPFSIQTCIALAFAGSQGETADEIAKALHFVSNFPPEVAQTFQFVLEKYRNSNLLRVANK
LYVQEGKQLKPAYQSAIKEQYHSEAESINFALNDAAAQAINAWVNAKTQKGITELVSADSFSDNTRLVLLNALHFKGSAWAKHFEERTEEDIFW
VGEHEQVKINYMNQKAKFNYPEDLGCCTALEMPYQDSLSMFLPQERTGIYALAEKLTVNLVDLADKLTVEEVHVFKPKFKVDYSLELAE
KLKQLGITKMFDTQAEFNSLLESPEGVFVSKVLHKATIEVNEEGTEAAAATGMIMMTRMMTFPLQFQADRPFVYIWNKNILFAGAFVKAA

>Dm-Acp76A CG3801

MGNHQVIFLVLCTSLFQNTIQNVSFQIUREIDRYTPENFVLSVLNIEMLFEIHAAKAVESSNDLERSLIINFGYSEARQEVLDWGLRYKKA
SSAKFQMANKVAVSQKLPKLSQKLRVNEVLMTSAKKYDVTKDVRSKLMDEWLSHLDGVLANFVQEKKNLAGENIVAIISGMTVTPWASHFQS
EINRYFVNNPGTYASKDPTCVPMHSLSSFETMSTDEAKGIYIPFSSANLGMILLPRKGVTCCKIDLNLNNQINVEYNDHKDVHLLLPIFKE
KFDYNIKFFNGINIEDTFKDSAFKSKAKIKINNFRVNHGIRFQPIRLREVVDIDTGTETFEVNRPFVVIKDKINVYAVGRINLGLTDTK
VNCSKYYADLKS

>Dm-CG6680

MGKIADCLMLWIPLLLGAIFLCSADPQNNQAPLQLSVGNPLTAFTAFTAQSGVSHIQSMRSNFDTDVLSVISQGVQDFALDLLQRISVEVEKA
NKDFMISPFVSVSLLVLLYEGSEGETRNLKSLRINVEDEKLRGAYKVVSSFLNITTTSTIEVATLQAIYTGKGYPIKNNYRDAIQNYNVQPM
VDFYSPDSVIQINEDTNRTRGLIPYITLIPQDVYAGKMFLLSLLYFKAQGWKFPFNKTLTREETPFSESEGEVIGKIPMMVQEANFAYVSNVEGLD
GYVLELPYGTQDLRINAMIVLPLKRGVLEKLNVDVANNLALGLRPIILQRLAFAFRNASELDENEVEVMMPKFVTATDFTLKGVLQIMGIRDLDENTANL
DRMSSGLFAKLVVHSTKIIVDEQGTAGAVTEAALANKATPPKFLNRPFYQYIMEKATGLLLFAGQVRNPKAA

>Dm-CG6663

MKLGFLGLFGMVLIMIFVEGAEGYTVNELRDVLGIYVDYPTLRRWYEDVRAHYLNSENTKFLSLRYAYYDDVGDLELVKGYNSVVLEGVGEGN
VVLREGRPRGVDFDQASIIINDDIDKASHAKIFSSYSRRSFNSTVTVLGITVSYFKAQWYFPDKSQTKEQFYNDDGGSFAGKVVEMMVQTKY
AYVNNVKGLQADVLELPPFGEHELVMIVLPPKSSQGVNVLVYQLKLNGLHRLLEKLEASKNETDVEVKLPKFDTRS SVLSLEDTVYDAGLTDLRND
FADLDKLLIAIGHRGACLTLYHQFARIIVDEEGLPNAVPQKSSGKNNIKFHVNRPFAYLVLQKKHLLIHSVGFREGEIQ

>Dm-CG6289

MKLGFLGLFGMVLISLRAHICANTVDTKSLQKMTDARLQFALNLLQMESTHLNLENFAMTFFSTWLSMIMIFVEGAEGTTLKQIRDVLAIYVDYP
TLRRWYEDVRAHYLNSENTKFLSLRYAYYDDVGDLELVKGYNSVVLEGVGEGNVVLEGRPRGVDFDQASIIINDDIDKASHAKIFSSYSRR
SFNSTVTVLGITVSYFKAQWYFPDKSQTKEQFYNAGGSPAGKVVEMMVQTKYAYVNNVKGLQADVLELPPFGEHELVMIVLPPKPSQRVSLV
KQLKLNGLHRLLEKLEASKNESDVEVKLPKFDTRS SVLSLEDTVYEAGLTDLRNEFADLGRMLIPTGDRGAYLSLHQFARIVVDEEGLPNAVPQ
KSSGTNNIKFHINRPFAYLVLRTHKLLIHSVGFREGEIQ

>Dm-Spn5 CG18525 Sp5

MHILSISLMAVLPALALAGLGVPEPDAGLLDQRLNLYKQQNFVAVSMLNVRQSTPNENVFFSPYSTYHALLLAYFGSSGDTEKELAKVLHLDW
ADSKEVVR SAYILEKMNKQSKMPLEFSSADRIFFANDLHVTECARNLAEEVQIDFKSQTEESRQINDWIAKQTHDQIRNMLSADEITP
RTRLVLANAAYLKQWLSQFKTEKTVMPFFYTPSNYSLSVMMQKQFTFLNVDQQLRAHVLPYRTPVESQEKEDSSPDENSISMLVILPP
FNSNSLEDVLSRLNADSLDSSLKQAMPREIEVSLPKFQFEQRLELNPILAKMGVSKMFDESATFDLDTSETISIGDSKHVAKIKVDEEGSTAA
AATVLFYRSARPVEPAKFECHPFLFVIYDRTSRSILFTGIYRDPKTIKQ

>Dm-CG6687

MHTFSLVLLALLPVVTTAALDKPELSFLNEFSQIFKGERDFSLALMKQIREIYPSGNLFFSPFSTYNALLLAYFSSSEQTERELAQALNLGVAL
NKQQVLSYVTLAQRQDEFWRQSPMELSSANRIFVDRITNVSNKFNLTLYGATKELDFKNDPETGLKEINDWIADKTHNQIRDMLSSSEITPHT
MLVLANAAYMKGQWLSQFVKEETALKPFFINEREQEMVMMHKTGAFKMTIDEGLQSQI IKLPYRTIYKSKETHISTPESKSDISMI IILPNSN
KISLNRVLSRLNADSVKWFERALPQKIELSLPKFQFEQRLELNPILSLMGVNTMFRNATFGDLTADPISLVIDDAQHLAKIKVDEEGSTAAA
ATILLVRSRQPDPTKFNCHPFLFVIYDEKVDITLIFAGVYSDPRQMQH

>Dm-CG1342

MKAVLCAIFVTTLLAAGQLPTQLEDENHGSFAGQVSQLIALQLLKFNKIDIDANQVHSPGLGVALAMLAEAESGDTYSEFEQVFGYPKDRTKL
RDAYKRILGSYQNRDAVALPSFQTWLYIYRNNSAREEFKDLLEKHYYVDVKDISRQYDWNENPTSLQLSEEKNDTSTEGSTEQSNKSDVIG
FETLKRKLDDEDIPAIIPSDGYQEIIDKEASKFDRDVKQYVEKPAQAEAEELLQKEQKQATTESELESQPEETTTLSVEKQEKPDMAAEE
NPAEKQQNRSDQESQIKNLENETVQEEELAKIMAAPALTAGPEKVRLLPQKLENVAVKTAADGGADEIMLALESHPVSVRVNGARSLFQ
QDDITSALSANSITGRSAGSKMMLFNGLYYRGSWANPFYQLRDSGSEFFMTNEDAVKAPMMHARGKQVADLPQVKARVLSLPYETSRYAL
CIVLPDETEGLSDVLSQLQTSDFLLAKKQFQMKELHISMPKQVEETSRESEAMLKQMGKLVSRTEAQLSLLSEDPDVHVEIVQFVNVVRVDE
GGSSANSLSAATMQARTPSVESTVLPVPEPEPELPGVERFEVNRPFAYFIVDCQEQVFLASGKIYTPFEFKEDLPSVSEVELEQS

Die Serpine von *Arabidopsis thaliana* (Abschnitt 5.3.2)

>At-Spn-1

MDVRESISLQNVQSMNLAKHVITTVSQQNSNVIFSPASINVVLSIIAAGSAGATKDQILSFLKFSSTQDLSNFSSEIVSAVLADGSANGGPKLSV
ANGAWIDKSLSFKPSFKQLLEDSYKAASNQADFQSKAVEVIAEVNSWAETKNGLITVLEPEGSADSMTKLIFANALYFKGTWNEKFDSELTQE
GEFHLLDGNKVTAPFMTSKKKQYVSAYDGFVKVGLPYLQGGQDKRQFSMYFYLPDANGLSDLLDKIVSTPGFLDNHIPPQVVKVREFKIPKFKF
SFGFDASNVLKGGLTSPFSGEEGLTEMVESPEMGNKLCVSNIFHKACIEVNEEGTEAAAAASAGVIKLRGLLMEEDEIDFVADHPFLVVTENI
TGVVLFIGQVVDPLH

>At-Spn-2 AT1G62170.1

MEPKEKKQKLDTEVASPSSKTHLKKKTKKQKIRKSQEITSPSLSKNTDLVIAASPSLSNIDVGEAMKKQNDVAIFLTGIVISSVAKNSNFVF
SPASINAALTMVAASSGGEQGEELRSFILLSFLKSSSTDELNAIFREIASVVLVDGSKKGGPKIAVVNGMMDQSLSVNPLSKDLFKNFFSAAFA
QVDFRSKAEVRETEVNAWASSHTNGLIKDLLPRGSVTSLTDRVYGSALYFKGTWEKYSKSMTKCKPFYLLNGTSVSVFPMSSFEKQYIAAYDG
FKVLRRLPYRQGRDNTNRNFAMYIYLPDKKGGELDDLERMTSTPGFLDSHNPERRVVKVGFRIPIPKFIEFGFEASSAFSDFELDVSYFQKTLIEI
DEKGEAVTFTAFRSAYLGCALVKPIDFVADHPFLFLIREEQTGTVLFAGQIFDPSA

>At-Spn-3 AT1G64030.1

MDVREAMKNQTHVAMILSGHVLSSAPKDSNVIFSPASINSAITMHAAGPGDVLVSGQILSFLRSSSIDELKTVFRELASVYADRATGGPKIT
AANGLWIDKSLPTDPKFKDLFENFFKAVYVPVDFRSEAEVREKENVNSVVEHHTNLIKDLLPDGVSSTLNKIYANALSFKGAWKRPFEKYYTR
DNDFYLVNGTSVSVFPMSSYENQYVRAYDGFVKVLRRLPYRQGSDDTNRKFSMYFYLPDKKDGDDLEKMASTPGFLDSHIPYRDELEKFRIPK
FKIEFGFSVTSVLDRLGLRSMYHKACVEIDEEGAEAAAATADGDCGCSLDFVEPPKIDFVADHPFLFLIREEKTGTVLFVGGQIFDPSGPCS
GSNSDSDDY

>At-Spn-4 AT1G64035.1

(aminoterminal Extension entfernt)

MDSLQELLDKQYDEEAKRQCQEAAMKKQNDVALFLAGEVISAADKNSNVVFPASINSVLTMAAATSDEALKSCILSFLRSSSTDELNDIFRE
IASVVLVDGSKTGGPKITVVNGVWREQSLPCSPESKDLFENFYKSAFAQVDFRSEAEVREKENVNSVVEHHTNLIKDLLPDGVSSTLNKIYANALSFKGAWKRPFEKYYTR
ALYFKGAWEDKFYKSMTHKHPFYLNVNGQVHVPFMSSQSYVKAADGFKVLRRLPYRQGVNDTSRQFSMCTYLPDEKDGDDLEKMASTPGFLDSHIPYRDELEKFRIPK
DSHIPSWRVEVGEFRIPKFKIEFGFEASSVFNDFALDVSILYQKAMIEIDEEGTEAAAATALVGAGCSCSLYRPPPIDFVADHPFFFIREDKTG
TVLFAGQIFDPSLSSA

>At-Spn-5 AT2G14540.1

MDSKRKNQELSTSETADPSSKTNKKQKIDMQEAMKNQNEVSLLVGKVI SAVAKNSNCFVFPASINAVLTVTAANTDNKTLRSFILLSFLKSSS
TEETNAIFHELASVVFQDSETGGPKIAAVNGVWMEQSLSCNPDWEDLFLNFFKASFAKVDFRHKAEEVRLDVTWASRHTNDLIKEILPRGSV
TSLTNWIYGNALYFKGAWEKAFDKSMTRDKPFHLLNGKSVSVPFMRSEYKQFIEAYDGFVKVLRRLPYRQGRDNTNREFSMYLYLPDKKGGELNDLL
ERITSNPGFLDSHIPYRVDVDFRIPKFKIEFGFEASSVFNDFALDVSILYQKAMIEIDEEGTEAAAATVTVVVTGSCLEWPKKKIDFVADHPF
LFLIREDKTGTLLFAGQIFDPSLSSALDRA

>At-Spn-6 AT2G25240.1

MELGKSIENHNDVVRLTKHVIAVANGSNLVFSPISINVLLSLIAAGSVCVTKEQILSFLMLPSTDHNLNLVLAQIIDGGTEKSDLRLSIANGV
WIDKFFSLKLSFKDLLENSYKATCSQVDFASKPSEVIDEVNTWAEVHTNGLIKQILSRDSIDTIRSSSTLVLANAVYFKGAWSSKFDANMTKKND
FHLLDGTSVKVPFMTNYEDQYLRSDYDGFVKVLRRLPYIEDQRQFSMYIYLPNDKEGLAPLLEKIGSEPSFFDNHIIPLHCISVGAFRIPKFKFSFEF
NASEVLKDMGLTSPFNNGGLTEMVDSPSNGDDLVSILHKACIEVDEEGTEAAAASVGVVSCTSFRRNPDFVADRPFLFTVREDKSGVILFM
GQVLDPSKH

>At-Spn-7 AT2G26390.1

MELGKSIENQNNVVARLAKKVIETDVANGSNVFPSPMSINVLLSLIAAGSNPVTKEIILSFLMSPSTDHNLNAVLAKIADGGTERSDLCSTAHG
VWIDKSSYLKPSFKDLLENSYKATCSQVDFASKPSEVIDEVNTWAEVHTNGLIKQILSRDSIDTIRSSSTLVLANAVYFKGAWSSKFDANMTKKND
KDNDFHLLDGTNTKVPFMTNYEDQYLRSDYDGFVKVLRRLPYIEDQRQFSMYIYLPNDKEGLAPLLEKIGSEPSFFDNHIIPLHCISVGAFRIPKFKFSFEF
SFEFKASEVLKDMGLTSPFTSKGNLTEMVDSPSNGDKLVHSSIHKACIEVDEEGTEAAAASVVAIMMPQCLMRNPDFVADHPFLFTVREDNSGV
ILFIGQVLDPSKH

>At-Spn-8 AT2G35580.1

(Introngrenzen verändert)

MDLEESIEKQYKAMMDLKESHVIATTSRKTNSLVFSPALINVILSIIAASSPGDTDADKIVSLLQASSTDKLHAVSSEIVTTVLADSTASGGP
TISAANGLWIEKTLNVEPSFKDLLLLNSYKAAFNRVDFRTKADENVREVNNSWVEKQTNGLITNLLPSNPKSAPLTDHIFANALFFNGRWDSQFDP
SLTKDSDFHLLDGTKVRVPFMTGASCRTYHYEGFKVINLQYRRGREDRSFSMQIYLPDEKDGGLPSMLERLASTRGLKDNVLPSSHSAVIKE
LKLPRFKDFAFEAASEALKGFGVLVPLSMIMHKSCIEVDEEVGSKAAAAAFRIGICRRPPEKHDFVADHPFLFIVKEYRSGVLVFLGQVMDPS
KH

>At-Spn-9 AT3G45220.1

MELGKSMENQTDVMVLLAKHVIPVANGSNLVFSPMSINVLLCLIAAGSNPVTKEIILSFLMSPSTDHNLNAVLAKIADGGTERSDLCSTAHG
AYGVWIDKSLSFKPSFKDLLENSYKATCSQVDFASKPSEVIDEVNTWAEVHTNGLIKQILSRDSIDTIRSSSTLVLANAVYFKGAWSSKFDANMTKKND
KSYDFHLLDGTMTKVPFMTNYKQYLEYDGFVKVLRRLPYVEDQRQFAMYIYLPNDRDGLPTLLEEISSKPRFLDNHIPPQRILTEAFKIPKFKF
SFEFKASDVLKEMGLTLPFTHGSLTEMVESPSIPENLCAENLFSVSNVFKACIEVDEEGTEAAAASVVAIMMPQCLMRNPDFVADHPFLFTVREE
KSGVILFMGQVLDPSIH

Die Serpine von *Ciona intestinalis* (Abschnitt 5.4.2)

>Ci-Spn-1 (ci0100132788)

MYTASAYGVLFLCLAIYQVGATKHLQAEFDYGEYEDDANSWDPRQASITKIGKMDGLTIDQVELPPFEKPAARLVNNAFVAVKLLNEIASDNEDNV
VFSPLSIFTSLATLRPALNGTSLQNLNDVTGLDITRESMDNDYDGIFFKSSSYKLLQASRIYVDRGIRLSRSYRTDLYRMKISRARRLDFRRA
PEESRNKINKYVKKRTRKLIKELVPGAISSATMMYLVNAIYKAKWDIPFQKSLTRMRRFRVSNNESIRVETMISKNTFCTRVNRRDLQASVT
VLSLGGFSFVIMSPHSAGNFSRYDGGVTTMQEKMTAFNKIWRTRGNRQQQLCSVKLPKFKVDYAENIKEVVKGLGIRDIFSINADFSRLSV
RNNRELYVSEARSHAVLSADEAGVEAAGATAFGISLRSTSLQVTNPKPFI FALRHDPGALIFVGGKIVRPSVG

>Ci-Spn-2 (ci0100132818)

MLLVIAACMLSAFNGAVGEPYAPTNAFREPVAHALYDFGMDMYNQLPEWSRPTENIVISPMSMYAIIISILLPGLNGASHDQVYNALRMTNLPRN
GVDAESAMCSKIFQINPNYDLTRANRIFGDRTLTFKKSYPNETSWHHKAAHKKVDFQHYPNRARRKMNRVSKMTDGEIQQILPREAVTTDTRI
FLVNAIAFKAAWKSSFIKDATTLTNFHVSPTKVKQADLMTYSSAVCFHQSHDAQLESDLIVLPFKGAKTMMVFIPIVAGNFGPLKGAVGASKI
SQALDRWYTYRNMPIPMRVCEVRMPKFKIITHSVDDLMGAMRAMNVTDIFSTEADFSMPTELVYVTDMRHKAVIKVNEQGVKATAATSIGLTI
RSLPIRVEINRPFMYMIRHEPTGALLFLGRVVDPTK

>Ci-Spn-3 (ci0100134682)
 MKLLICSLLLLVIATGYCQNRWRINHTFDNQNTFSGSLYWAISKEKPNKNVLFSPISVSQTLGMVLAGAMGNTYDEITRALQMTDLTPSRIHTL
 MRKTRNNVVRPNQGTVKLANSVFIGSNYPVVQYIDLLRQNYKSSVFPVNFHNSNAAANMINWVSNMTEDEKIRELVDPSSITAFTRMILVNA
 VYFQADWAISFRRIPTKQNFLLSNGTTVQVPPMVRWEAVVKSYNRDKIEFFFIKYKTTSNQNTYFVVGLPGDNYLQQFSREAQQILSRFRTI
 NKMFRITHFKMPLIELSHKTDVKEVQLTGLGVVLDLFDGASNLGTISTVEQLYVSEFTQKAYINVNENGTVAASAAATVQGRSLSIPRQVTVDR
 PFFIGVYQEKSNFSLFLGKVENPLEN

>Ci-Spn-4 (ci0100141118)
 MFLKQVLVLCVFFMFTSSAFYMPMVRTHPPQMDMPAYCAEVVNATRVFSGFILNAATHANSATDEYVAERNVFFSPFGAANVVGILRLASAGRT
 REQFDGLPLFSSILQHNDRFMRGFQTLRSLILSVTSFPPTNPGQSDKLLNSGVFTSRWLYLQTRFISDARNFYKAVVASVDFSDPELASSHIN
 MWINARTQRKIKIVSPSDLSPITLVTVFNTLFFALWKPFTTGRTSNSTFVLANGTPVLTMPMEVTANHFLHYSGEFCQLFSMRRCHPNTPD
 IVLLPYKGERRQMIVLIPNQITLREIERQFGTNLEKWRSSLDVGNVELHIPKFKELKSNLDLKSVLRSGLTEPFNRRTADYSTMTSRQLAISK
 LFQATASISMDETGVRATSTTAAFFELRSFFFRTRINANKPFLFIIEDIHTRTPLFLGRVTDPRPL

>Ci-Spn-5 (ci0100143209) (cDNA „ciad038b20_20“ berücksichtigt)
 MRFIFLCFVLLVSAGFNEAKRTRVISKWRLTATANKLFAHRLFMEVARTTPEQENFFISPYAVSAGLSMTLYGAHSTTAREIMDTLGYTQLSTS
 GNFNQAKVPRLYQKMLHQVHQDHGFELTSVNRMFGESRNIFVPSYVKVEHFYGAKLKVPFRNPERARQEINTWVEVTNGTIREALPPNS
 VTAETLLVLMSTLYFKGLWEKPFIEINLRSTFYTTNNEQYQTDVFVQQTMFALHSFSEQFAHIVELPFKTSSSRYKMVMQILILPESRGADNLNLI
 EDQFDEENFDATEDQENISVTIRLPKFRLEYETDLKETLYNMGIQSLFSRGEADLSGISTNGDLSLGAHKKTFIQVDESGTTAGASAYAQQGF
 RSVSDLDLVFNHFFIIVIREKYTQMPMFGRVAPMY

>Ci-Spn-6 (ci0100146394) (ESTs „cinc006o22“ und „rcinc006o22“ berücksichtigt)
 MAFCKVAAAKTDFALGLYKELSKQKEDGNLFFSPYSISTALMMLLGSKEKTRREMLDVLGLKDLNESDINSGLFQLIHLHRRSSRGDVVLEMANK
 LFPEAIYKLEKDFLSKCKEFYETEIQALDFKGNPDASREAINWAEEKETSGKIKDLLPSGIDSLVRLVLANAVYFKGSLWHLKFKEQQTMMKDF
 HIRENKVEKVNMMFMKRFRFNFDQSLGLQVVEIPYIGNKLSMVVFLPTEFALNKIENALTTEKLHGLLAGLWEETLMLSLPRMKFEQDFDLG
 GVLKMGMMDAFDERAANFEAISGSRDLVISKVVHKAFIEVNEEGSEAAAATAVVMMLRSMPPVPMVNCDDHPFLFLIRHNQTKTILFLGRFSG
 P

>Ci-Spn-7 (ci0100146394) (ESTs „rcibd043j06“ und „citb069b04“ berücksichtigt)
 MAFCKVAAAKTDFALGLYKELSKQKEDGNLFFSPYSISTALMMLLGSKEKTRREMLDVLGLKDLNESDINSGLFQLIHLHRRSSRGDVVLEMANK
 LFPEAIYKLEKDFLSKCKEFYETEIQALDFKGNPDASREAINWAEEKETSGKIKDLLPSGIDSLVRLVLANAVYFKGSLWHLKFKEQQTMMKDF
 HIRENKVEKVNMMFMKRFRFNFDQSLGLQVVEIPYIGNKLSMVVFLPTEFALNKIENALTTEKLHGLLAGLWEETLMLSLPRMKFEQDFDLG
 GVLKMGMMDAFDERAANFEAISGSRDLVISKVVHKAFIEVNEEGSEAAAATAVVMMLRSMPPVPMVNCDDHPFLFLIRHNQTKTILFLGRFSG
 SDPS

>Ci-Spn-8 (ci0100146394) (ESTs „rcibd080f02“ und „cibd002g03“ berücksichtigt)
 MAFCKVAAAKTDFALGLYKELSKQKEDGNLFFSPYSISTALMMLLGSKEKTRREMLDVLGLKDLNESDINSGLFQLIHLHRRSSRGDVVLEMANK
 LFPEAIYKLEKDFLSKCKEFYETEIQALDFKGNPDASREAINWAEEKETSGKIKDLLPSGIDSLVRLVLANAVYFKGSLWHLKFKEQQTMMKDF
 HIRENKVEKVNMMFMKRFRFNFDQSLGLQVVEIPYIGNKLSMVVFLPTEFALNKIENALTTEKLHGLLAGLWEETLMLSLPRMKFEQDFDLG
 GVLKMGMMDAFDERAANFEAISGSRDLVISKVVHKAFIEVNEEGSEAAAATAVVMMLRSMPPVPMVNCDDHPFLFLIRHNQTKTILFLGRFSG
 PSI

>Ci-Spn-9 (ci0100148346) (ESTs „rcicl06102“ und „citb049d07“ berücksichtigt)
 MQFLYAIVMILVLDANAKIIDTSEHVEKLESEANIEFTLNLYKNLIEGDPKKNVMSFVSPVSIITLALIAHLGAKGNTAKQIDDAFMFSKIEDGRFH
 SAFGELHGLLFDKASEKVTAKSSNRVFAADKHITVFEDYQDSLVSYSATVESVDFKMPKSAVKKINDWSSDATNGVIKSMLEEDGVNNDTALLII
 NALYFRGNWDYEFDEGRTKRRPFYVSKDKAVETSFMFQNEHFYAYINELTLQVLEMDYAGTDYSMVLLMPENFDLAKVEANLNHANLTKWLSA
 LEHESVDLTIPKFKLEETLQLEVLPKMGVTDLFDQRACDLTGIANRNNLFDVQIVHKTVDLVNEQGSEAAATTVRVTQCDQSVAFNPISFVADH
 PFLWAIRHRQSELLIFMGRFSRPEGPLLGHDEF

>Ci-Spn-10A (ci0100154072) (EST „rcibd064c17“ berücksichtigt)
 MQFLYAIVMILVLDANAKIIDTSEHVEKLESEANIEFTLNLYKNLIEGDPKKNVMSFVSPVSIISAAALAMTHLGAKGKTAQIDDAFMFSKIEDGRFH
 SAFGELHGLLFDKASDNVTVKSSNRVFAADKRRKVLVDYKALTYGAKLENVDFKTPSNAVVKQINDWASDATNGKISNMLQDDAVDSNTALIVA
 NAVYFRGDWHSKFNEMQTERRAFYVSHYKIVETPFMFQRGHFKYAYISELTLQVLEMDYAGKDYSMVLLMPENFDLAKVEANLNHANLTKWLSA
 LKYKSVLDLSPKFKLEETLQLEVLPKMGVTDLFDQRACDLTGISKSKDLNVDQIVHKTVDLVNEQGSEAAATTVRVTQCDQSVAFNPISFVADH
 FLWAIRHRQSELLIFMGRFSRPEGPLLGHDEF

>Ci-Spn-10B (ci0100154072) (EST „rciht030117“ berücksichtigt)
 MQFLYAIVMILVLDANAKIIDTSEHVEKLESEANIEFTLNLYKNLIEGDPKKNVMSFVSPVSIISAAALAMTHLGAKGKTAQIDDAFMFSKIEDGRFH
 SAFGELHGLLFDKASDNVTVKSSNRVFAADKRRKVLVDYKALTYGAKLENVDFKTPSNAVVKQINDWASDATNGKISNMLQDDAVDSNTALIVA
 NAVYFRGDWHSKFNEMQTERRAFYVSHYKIVETPFMFQRGHFKYAYISELTLQVLEMDYAGKDYSMVLLMPENFDLAKVEANLNHANLTKWLSA
 LKYKSVLDLSPKFKLEETLQLEVLPKMGVTDLFDQRACDLTGISKSKDLNVDQIVHKTVDLVNEQGSEAAATTVRVTQCDQSVAFNPISFVADH
 IIVRGRANNAFLHFGAYKRPAGKIRSHDEL

8.5.2 DNA-Sequenz des Serpin-Gens *Dm-Spn4* von *Drosophila melanogaster*

```

1   TCTATTACGTACGCTTTGCGAGACTAGGTTTACTGTAAGCAAAATTAAGTATGGATTA
61  TCGTTTGGTGCCATGTGAGTTATCCGGCAGTTTACTTTAAAAAGCCAGTGAAAGAA
121 AACGATTATGGATGTAAAGAGTGAATCAATGAGAACACCGAAGGCAGGCAGAGAAAA
181 ACCCGTTTGGGAGCACGAAGTGAAGAGAAATCATTGTTTTGTGGGGGAATACATATGCAT
241 ATGCGAGTATAAAAAACAATGTACGCTCGCTGAGAGAATTCAGTTGTTGCTCAAAGGCCAC
301 ATTATAAAAGTTCGAGGGCGATAGCCGGTTTATAAAAAAGAAATTAATTTACACACTCGAAG
361 CCGCTGGTGAACACAGTACCAGCAGTCCGAGCAGGCAGTTAATTACCCTTTCCGATAAT
421 CAAGCTAGTGAATCCCGGGCAAACCGGCAACCGGAGAGATATAGTGAACACCGTTTGTG
481 GAGGTGACTCATCAGTACACTGCAATTACATAAATTTATGAAAACGAGCACACCTCATCC
541 ACGGTCGATTTGGATACTTACAGGTTGCTGGTTGCTCCCCCTGCTCGGCCAGCCCTCTT
601 CCCGTTTCCACCGGTCCACACGGCTGACGTCACCATGCTGACGCCGCCACCAGGAGTT
661 CGCCCGCCGGCTGGCCCTTTTCTCTATCAACGTGTACGGCAAGCTGTGCGGGCAGAAGCC
721 CGGCGAGAACATCGTCTTCTCGCCCTTTTCCATCCAGACTTGTGCGGGCAGTGGCCAGGCT
781 GGGTGCAGAAAACGAGACGGCTACCCAGCTGGACCAGGGACTGGGCTTGGCCTCTAGTGA
841 CCCAGAGCAAATCGCACACAGTTTTTACCAGGTGCTGGCCGCCTACCAGGACAGTCAGAT
901 CTTGCGCATCGCCAACAAGATTTTTCGTCATGGATGGCTACCAGCTGCGCCAGGAGTTTGA
961 TCAGCTGTTATCCAAGCAGTTTCTCTCGGCGGCTCAGAGCGTGGACTTTTCCAAAAATGT
1021 ACAAGCAGCGGCCACCATCAACAATTTGGGTTGAGCAGCGCACCAACCATCTGATCAAGGA
1081 CCTTGTGCGGGCTGATGTATTGAACTCCGAATCGCGACTCGTCTTGTGAACGCCATTCA
1141 CTTCAAGGGCACCTGGCAGCACCAGTTCCGCCAAACACTTAACCCGCCAGACCTTTCA
1201 CCTGGATGGGGAGCGAACAGTTTCAAGTTCCGATGATGAGTTTAAAGGAGCGGTTCCGCTA
1261 CGCCGATCTGCCGGCGCTAGATGCCATGGCTCTGGAAGTGCCTACAAGGACTCAGACCT
1321 CTCTATGCTGATCGTACTGCCAACACCAAGACGGGTCTGCCCGCTCTGGAGGAGAAAGTT
1381 GCGCCTCACGACCCTCTCGCAGATCACGCGAGTCGTTGTACGAAAACAAAGGTAGCGCTCAA
1441 ATTGCCAGGTTCAAGGCAGAGTTCCAAGTGAATTTGTCTGAGGTTTTTCCAGAAGGTGAG
1501 CTTTACTTTAAACACATAAGCAATGATATGGACTTCATGTTAATATAAACCTCCTCCAGC
1561 TGGGCATGTCAAGGATGTTCTCCGATCAGGCTGAATTTGGAAAAATGCTACAAAAGTCCGG
1621 AGCCATTGAAAGTGTGAGCCATCATAACAAGGCCTTCATTGAAGTCAACGAGGAGGGAA
1681 CGGAGGCTGCGGCCGCCACGGGTAAGTGGTCAAAGTCAATGGTATTTTTTGGCATCCATC
1741 ATCCTGTGCATGTCCATTTGTTTATCCGTTGCTGTCTTGTCTTCTTCCCATCCTGCAGC
1801 CATGGCGGTGCGTAGGAAGCGCGCTATTATGTCGCCTGAGGAACCAATTGAGTTCTTTGC
1861 CGACCATCCTTTACCTATGTCTTGTGTCATCAGAAGGATCTGCCATTTGTTTGGGGCTC
1921 AGTTGTGCGGCTCGAGGAAAATACCTTCCGCTCCAGCGAGCATGATGAGCTGFGATGGTA
1981 CCTGATTCTTTGGCAAAAACAATAAAGACAGCTATTTATTTACAATGGAAATTAAGCATTT
2041 GTCTGAAGCTTTGGGATAATCCATTTATCGGCATACCTGGGTGGCATAACATTACGAGAAA
2101 TAAATTGCATCTGCTCCATTACCAGTTTGGCCACTTACTAATTCGAGTTATATAACCTT
2161 GAGCTACGACTTTGATGTTTCAATTTGCAGGAATGGTCAATGTGCTACGCTCCATGCTCAG
2221 TTCGAACCGCAGCCCGTCCAATTCATGTCCAGCATCCATTTAACTATTACATTATCAAT
2281 AAGGATTCTACTATTTTGTGTTGCGGGAAGAATCAATAAGTTTTGAGAGTTAGCACTCCATC
2341 AAAATCCACCAATCACAACCATTGAACTGAAATGCAATGTAATCCCAATTCACAGGTATG
2401 TTTATGAGCCTTACTTCCGCTGCCCATGCCGAAGCCGGATCCGATACGTTTCAATGTCGAC
2461 CATCCATTTACATTTTACATCCTTAAACAAGGACTCAACCGCTCTCTTCCGCTGGCAGCATA
2521 AAGAACTCTAAATAGGCAAACATACTAACCAATCCACAATGTTACTCGAGGGTCTCGTA
2581 TTCATTTAACGTGCAAGCGGGTCCAGTCTATATTGCTGGCTTACGAGCATATTCAGCTT
2641 TTACATTGTCTATAAGGATTCACAAAATTATAGTCTGTGCGCAATTGATGATCTTATATA
2701 TACACTTAATATTTTATACTAACCAAGCATGTAGCCGTAGCCTCAATAAAGTCTAAAAG
2761 CAAATGTCTAAAAGTCTGCTTCCGTTTTTCAATGGTTTCTCCTCAGTATGGCGGGTATGGCC
2821 GTAGCTGCCTTCCAGTCCGAAACACTTCATTGCCAACCATCCTTTTGTCTTCTACGTGAAG
2881 ACCCATTACGATCTGCCGATATTTCACTGGGCGCTACTTGGGTAAAGCAATAGGAAGCTTC
2941 CTATGAAAAAAAATTTGAAGTTCAATAAAAATAAAAATAAACCGGCAGCCACTTAATAGCTG
3001 TGATTTTCGTCAAGTCGAA

```

Abbildung 44: DNA-Sequenz des Serpin-Gens *Dm-Spn4*

Die DNA-Sequenzen der Exons 1, 2, 3 und 4 sind gelb und die vier RCL-kodierenden Exons grün markiert. Start-Codons und Stopp-Codons sind hellblau bzw. rot hervorgehoben.

8.5.3 Aminosäure-Alignment der Serpine von *Caenorhabditis elegans*

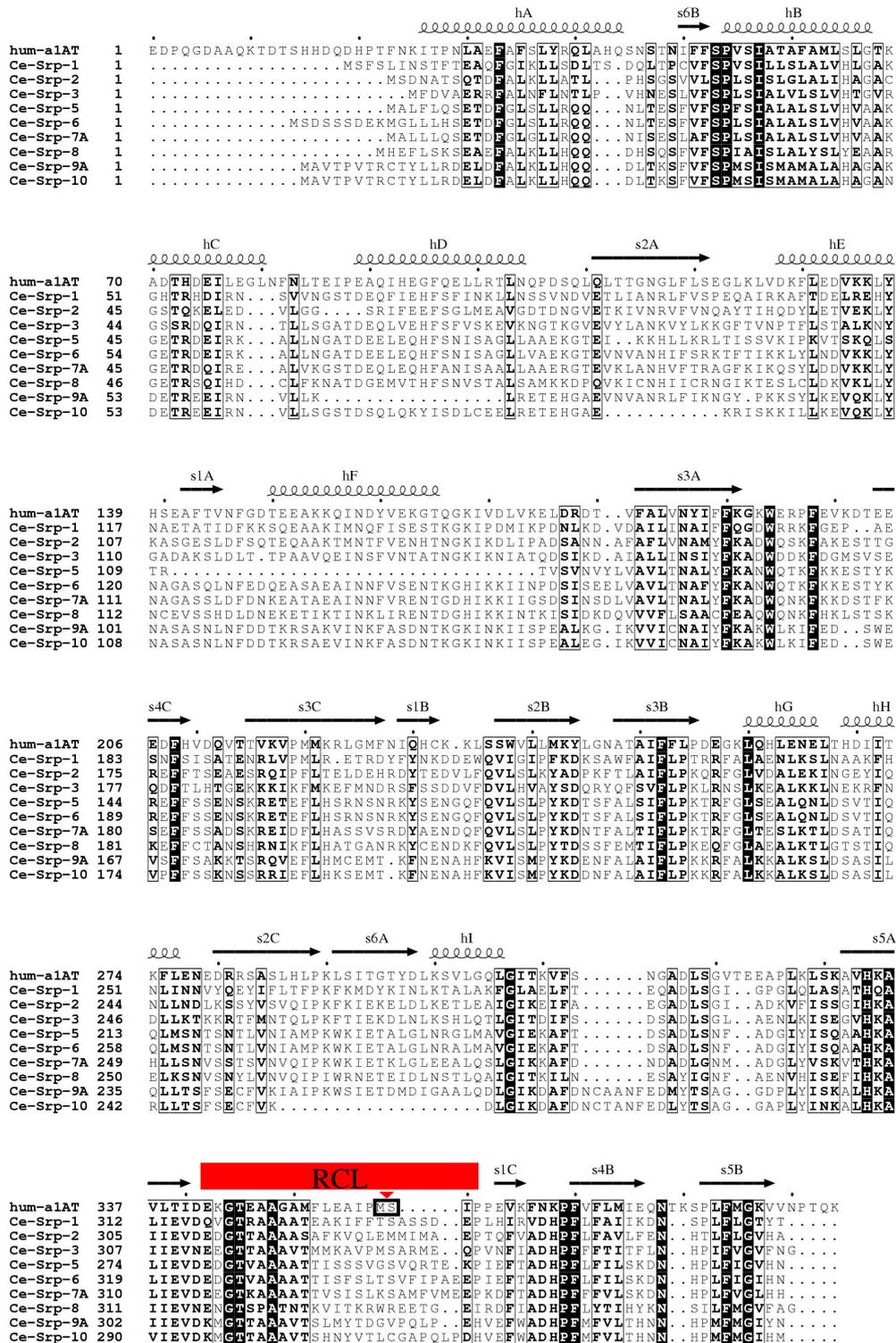


Abbildung 45: Alignment der neun Serpine von *Caenorhabditis elegans*

Das Alignment wurde mit Hilfe des Programms ESPrnt 2.0 dargestellt. Oberhalb des Alignments sind die Sekundärstrukturelemente (Irving *et al.* 2000, abgeleitet von α_1 -Antitrypsin) dargestellt. α -Helices werden durch Kringel und β -Faltblätter durch Pfeile symbolisiert. Der RCL ist durch einen roten Balken hervorgehoben. Strikt konservierte Alignment-Spalten sind invertiert dargestellt. Alignment-Spalten mit hoher Konservierung sind fett markiert und eingerahmt (Risler-Substitutionsmatrix (Risler *et al.* 1998), *global score* 0,6). Die RCL-Spaltstelle P1-P1' von α_1 -Antitrypsin durch Einrahmen markiert (MS, Positionen 358-359).

8.5.4 Aminosäure-Alignment der Serpine von *Drosophila melanogaster* (1/4)

```

hum-a1AT .....
Dm-Spn27a .....
Dm-CG6717 .....
Dm-CG31902 .....
Dm-CG33121 .....
Dm-CG7219 .....
Dm-Spn2 .....
Dm-CG4804 .....
Dm-Spn3 .....
Dm-Spn4-V1 .....
Dm-CG9454 .....
Dm-CG9455 .....
Dm-Spn1 .....
Dm-CG9460 .....
Dm-Spn43Aa .....
Dm-Spn43Ab .....
Dm-CG1859 .....
Dm-Nec .....
Dm-CG7722 .....
Dm-CG10956 .....
Dm-Spn6 .....
Dm-Acp76A .....
Dm-CG6680 .....
Dm-CG6663 .....
Dm-CG6289 .....
Dm-Spn5 .....
Dm-CG6687 .....
Dm-CG1342 1 MKA VLCAIFV TLLAA IGG LPTQL EDENHG SFAGQVS QLI ALQL LKFNKD IDANQV HSP LGVASILA
    
```

```

hum-a1AT .....
Dm-Spn27a .....
Dm-CG6717 .....
Dm-CG31902 .....
Dm-CG33121 .....
Dm-CG7219 1 ..... MW RLL L L L L L V S V C C E S E L F R D D L R T P E T M A
Dm-Spn2 .....
Dm-CG4804 .....
Dm-Spn3 .....
Dm-Spn4-V1 .....
Dm-CG9454 .....
Dm-CG9455 .....
Dm-Spn1 .....
Dm-CG9460 .....
Dm-Spn43Aa .....
Dm-Spn43Ab .....
Dm-CG1859 .....
Dm-Nec 1 ..... M A S K V S I L L L L T V H L L A A Q T F A Q E L I A W Q R Q Q Q Q
Dm-CG7722 .....
Dm-CG10956 .....
Dm-Spn6 .....
Dm-Acp76A .....
Dm-CG6680 .....
Dm-CG6663 .....
Dm-CG6289 .....
Dm-Spn5 .....
Dm-CG6687 .....
Dm-CG1342 68 M L A E A S E G D T Y S E F E Q V F G Y P K D R T K L R D A Y K R I L G S Y Q N R D A A V A L P S F Q T W L Y I Y R N N S A R E E F K
    
```

```

hum-a1AT 1 ..... E D P Q G D A A Q K T D T S H H D Q D H P T F N
Dm-Spn27a 1 M T K M G G N L A V M L L S L F L S A L A T G N G N S I P T T T T P Q G V F E T R T D K L P G G A A S V P S G A G I Y D D I D T F V P
Dm-CG6717 1 ..... M K Y L C K R S T K S Y F T L Q Q F S L I P L D L L L L
Dm-CG31902 1 ..... M K Y L C ..... W I L
Dm-CG33121 1 ..... M Q G N N K I K Y L V ..... L L L
Dm-CG7219 33 Y I N G L M Q R R H Q M Q Q E A Q Q H I Q A I P P A V P L Q S P G L V N G L G N Q N D P A L N R I S G T S V K F S N L P A A Y S N G Y
Dm-Spn2 1 ..... M K H L C ..... L L L
Dm-CG4804 1 ..... M S E A A N .....
Dm-Spn3 1 ..... M K Y L C ..... L L L
Dm-Spn4-V1 1 ..... M D Y R L V P C G C W L L P L L G L A L F P F P P V H T A D V T M A D A A H ..... Q
Dm-CG9454 1 ..... M K D ..... E
Dm-CG9455 1 ..... M S E P Q E G ..... R N Q
Dm-Spn1 1 ..... M Y Y L C ..... I F L
Dm-CG9460 1 ..... M A N T L ..... N Y S
Dm-Spn43Aa 1 ..... M N H W L S I L L G V W I S A ..... P E G L G
Dm-Spn43Ab 1 ..... M A V I I S C L L L L L A T V S Q ..... S K T V G Y
Dm-CG1859 1 ..... M I H W R L L S A L L V G L A I A L T L P V D G ..... E L L A R S
Dm-Nec 35 Q Q Q Q Q L Q L Q Q Q L L L Q Q Q Q H Q R N P R P E L G L R S L P G N P W T Q N N Q E A I S D V V A V D L T K R E P V T P P P N R P P
Dm-CG7722 1 ..... M A S G G T T ..... A P S
Dm-CG10956 1 ..... M Y V L L L I L L G ..... I S R
Dm-Spn6 1 ..... M F D F N ..... L
Dm-Acp76A 1 ..... M G N H Q V I F L V L C T ..... S L L
Dm-CG6680 1 ..... M L W I P L L L G A I F L C S A D P Q N N Q A P L Q L S V G N P L T A P T A P Q S G V S H I Q S M R S N F D T D V L V
Dm-CG6663 1 ..... M K L G F L G L F G M V L .....
Dm-CG6289 1 ..... M K L G F L G L F G M V L S L R A H I C A N T V D T K S L L Q
Dm-Spn5 1 ..... M H I L S I S L M A V L P A I A L A G L C G V E P D A G L L D Q R L
Dm-CG6687 1 ..... M H T F S L V L L A L L P V V T I A A L D ..... K P E L S F L N E F S
Dm-CG1342 135 D L L E K H Y V D V K D I S R Q E Y D W N E P N T S L Q L S E E K D N T D S T E G S T E Q S N S K D V I G F E T L K R I K L D D E D
    
```

Aminosäure-Alignment der Serpine von *Drosophila melanogaster* (2/4)

		hA	s6B	hB	hC
hum-a1AT	25	KITPNLAEFAFSLYR...QLAHQSNST	NIFFSPVSIATA	FAMLSLGT	KA...DTHDEILEGLN
Dm-Spn27a	68	FRSDSHDFPSWHLLKT...VLQNETADKN	NVIISFVSKLV	LALLAEAA	GAGT...QTQVELANTQ
Dm-CG6717	28	LATSVEGFWEDFYR...ILASQNAKRN	LIYISFISAEI	MSMVYMAS	GG...KTFFELRNVL
Dm-CG31902	9	VTTSVLQGFTKQLYR...SFLQDNKQY	NIIASF	LCVEIGMSM	LMGADG...NTANELRRTAL
Dm-CG33121	15	IATSVLGGFKLNL...ELVMDKAES	NFIASF	LCIEIGISMI	LMGAKG...TTAEELRSVLD
Dm-CG7219	100	VDLATSDRIANSVLFNFANILGQHLANG	TQIYISF	LSIVHSLALL	LLGAKG...RSYEELRSVLD
Dm-Spn2	9	LATSVCRFADDFYQ...LLAKENAAN	NLISISF	LSVEIALSMA	YMGARA...KTAQEMRNVL
Dm-CG4804	7	.PTPYDCHIGAGIYH...SIATSFAEQ	NVVVVSF	LLLEATLSL	LFLGSDG...ATAEELQKQL
Dm-Spn3	9	LATSVCRFDDLYQ...LLAKENADKN	NLITISF	LSVEIALSMA	YMGARA...KTAQEMRNVL
Dm-Spn4-V1	40	EFARRLALFSINVIY...KLSGQKPGEN	NIVFISF	SIQCAAMA	RLGAEN...ETATQLDQGL
Dm-CG9454	5	EFAQGLEQFALCLHD...HLGRASAGI	NIIISF	LSIHISAA	MRLMGTSEGS...ATAKEMDEGL
Dm-CG9455	11	FARNLIDVITKDALQ...QSKD.PHINTV	FSPASVQSA	LTLAFMG	GASG...STAEELRNGL
Dm-Spn1	9	WVTSVACQTSKEIYQ...LLSKSHTNQ	NLVVVSF	VSIEITLSM	VFMGAE...STAKELQSA
Dm-CG9460	9	KSPAGEAQFASQLFG...QLAKSQSGRN	NIVFISF	SIIRTLGALA	LYLGAEG...STADELKLGL
Dm-Spn43Aa	22	NTIKDRNFLFATELFO...TLATDRQDEN	NVIISF	VSIIQALGL	LAAGAEG...RTAEELQKTL
Dm-Spn43Ab	24	DAADRNLAAADLYN...AVAADHLENE	NVVIISF	ATIQQSSMA	LAFVGAEG...KTAQEMRNVL
Dm-CG1859	31	PASVSNRFRGLRITT...KLGTLQPDAN	NVVVVSF	LLIQAA	LSLYAES...EYGSQLRQAL
Dm-Nec	102	PVFSYMDRFSSELFK...EIKSQSQQN	NVVFISF	SVHALLAL	IYGAESD...KTRFRELQKAG
Dm-CG7722	11	LSASPI.VFARNLFR...ALNDEVPPV	NMMVVSF	AGARSA	MTLVFMGAG...KSADELRSKL
Dm-CG10956	14	YRAQKNSQLPDELYA...AIVNSFSNRN	NIMFST	EMIRSS	MLFYVGV...ESEDEQIRDM
Dm-Spn6	7	EFARGGARFTSELFQ...LLSAGFKEN	NVVFISF	SIQCAALA	FAGSGG...TADAEIAKAL
Dm-Acp76A	17	FQNTIQQNVSFQLIR...EIDRYPTE	NFVLSV	LNIEML	LFETHAKAV...ESNNDLERS
Dm-CG6663	63	SISQGVQDFALDQLQRI...SVEVEKANK	DFMISF	FSVWSL	LVLVEG...ETRNQLKGL
Dm-CG6663	14	MIMFYE	GAEG...YTVNELRDV
Dm-CG6289	32	KMTDARLQFALNLLQ...MESTHLNLE	NFAMTF	FSTWSL	MIMFYEGAEG...TTLKQIRDV
Dm-Spn5	35	NLYKGGQNFVAVSMLN...VIRSTPNE	NVVFISF	YSTYHAL	LLAYFG...DTEKELAKVL
Dm-CG6687	33	QIFKGERDFSLAMK...QIREIYPSG	NLFFISF	STYNALA	LLAYFSS...QTERELAQAL
Dm-CG1342	202	TPAIPSDGYGQELIDK...EASKFDRDVK	QYVEKF	VAAQAE	ELLOKRE...KQKQQA

		hD	s2A
hum-a1AT	86	EIPPEAQIH...EGFQELLRTLNQPDSQLO
Dm-Spn27a	131	..SONNVR...EFYRKTLSNFKKENQLHETLS
Dm-CG6717	87	.ENKT...LVANNYRSLSDLKRRETFI
Dm-CG31902	68	.EDKK...NVAIYDKLLTKLERGKKVAI
Dm-CG33121	73	.VDVT...EMA...KKYERIMSNFQHN...GLR
Dm-CG7219	165	SRLHE...QFG...LMLQDLQOPTREAT	SAGRPLTDWRASSAMRNSRRARQ
Dm-Spn2	68	.DDK...EVA...AKYKDLLSKLEGREKVA
Dm-CG4804	67	FASNA...KMA...NFYAAELGNITTDADT
Dm-Spn3	68	.DDK...EVA...AKYKDLLSKLEGREVAI
Dm-Spn4-V1	100	.SDPE...QIA...HSHFHQVLAYQDSQ
Dm-CG9454	67	.LEAQ...QVA...EESFGVVLSKYEQCVL
Dm-CG9455	69	.GDRH...HIA...LNFGEFWRTS...CNYGDRGP
Dm-Spn1	69	.EDKE...AVA...ARYGALLNDLQGEFEGP
Dm-CG9460	69	.AGRT...EVA...EKLDQLLAKQWEKASGDE
Dm-Spn43Aa	82	.KESK...DGL...AESYHNLHLSYIKSKT
Dm-Spn43Ab	84	.GDAD...AVS...QRSGS.YQQALTRDNN
Dm-CG1859	91	ASHPK...LAV...QDFETLLTDLKQSAAIGCR
Dm-Nec	161	.KNAM...AVA...QDFESVIKYKHLLEGAD
Dm-CG7722	70	.SNKS...EVA...KQHAESWTDE...CSCAKK
Dm-CG10956	74	.THLS...EYK...PKTKIFAMSVKKAP
Dm-Spn6	67	.NFPPE...EVA...QTFQVLEKYRNSN
Dm-Acp76A	77	YSEAR...EVE...LDWGLRYKKASSAK
Dm-CG6680	126	.DEKL...RGA...YKVVSSFLNITSTI
Dm-CG6663	38	.DYPT...LRR...WYEDVRYAYHYNSENTK
Dm-CG6289	92	.DYPT...LRR...WYEDVRYAYHYNSENTK
Dm-Spn5	95	.ADSK...EVR...RSAYILEKMRNKERQSKM
Dm-CG6687	93	.ALNK...QV...LVSYTLAQQRDEFRWRQSP
Dm-CG1342	268	LSVEK...QEK...PDMAAEEENPAEK	QONKRSQDEESQIKNLEENETVQEE

		hE	s1A	hF
hum-a1AT	126	LVDFK...LE...DVK...KLYH...SEAF...TVNFGDTEE	AKKQINDYV
Dm-Spn27a	170	TQQR...FAT...TLK...HFYD...SEVEAL...DFTNPEAAA	DAINAWAANI
Dm-CG6717	126	LVPE...FNQ...LAR...KAFKAKAKS	IRLDD...PVSAS	AINVSWILNRT
Dm-CG31902	107	VNKR...YKLVN...KHFRA...EAEAI	IKLAD...RLKAA	AWAINDWVLDQT
Dm-CG33121	109	VRQD...YNT...LMK...STFMA...EGKDP	LSQR...KASNS	ISFSIHRKSHKGM
Dm-Spn2	229	LNPD...YRR...VIV...EYAS...DQLIQ	DFEGS...PATAR	YNINAYVAOHT
Dm-Spn2	107	LVPS...YVQ...MVK...DSFMA...EAEAI	DIVD...PNKAS	SIVNWNVNDQ
Dm-CG4804	106	VADF...QKIAQ...TYFHA...TAEACV	DLEQT...EKLRR	HISEQILASV
Dm-Spn3	107	LRPE...YVQ...MVK...DSFMA...EAEAI	ISANN...PKITA	SIVNWNVNDQ
Dm-Spn4-V1	136	LRQEF...DQL...LSK...QFLS...AAQSV	DFSKN...VQAA	ATINNVEQRT
Dm-CG9454	103	VD...EQ...GHILE...QKFR...SKPME	IDFGS...EQAA	SINKWVESQ
Dm-CG9455	109	LLTE...FNE...IAV...DFFQ...KAAEA	IRFAD...SEGAT	QLINQWVVE
Dm-Spn1	108	LNQNY...NLAVR...EPPFK...SEAES	ISLNT...GPAAR	ERINQWVLDQT
Dm-CG9460	113	LTQTY...QD...LVSK...NFAA...AENVF	TQ...KADTA	KHINQWVVE
Dm-Spn43Aa	120	VSSH...FR...EVA...QK...YFDSE	VEPLD...ETE	AVEQINRWV
Dm-Spn43Ab	120	FKGS...FR...VAQR...QDFG...NIDK	LDHPPY...NKR	TADGINRAVAT
Dm-CG1859	133	NFRN...EFT...LAARM...GVCN	IRLWS...ASNAA	QDINYAFVLS
Dm-Nec	200	NVHS...YDE...YAR...FYFSA...GTEA	VMQN...AKDTA	AKINAWVMDT
Dm-CG7722	110	IRTF...DFN...MALE...FFNA...EAYS	LNLYLN...PEDSV	KKVNKLEKHT
Dm-CG10956	111	MS...TEYR...VFM...RHT...EGRARNI	IAPAR...PQLE	DEVNTFY
Dm-Spn6	103	LKPAY...QS...AIK...EYH...SEAES	INFALN...DAAA	QAINAWVNAKT
Dm-Acp76A	112	PLSQ...KRL...RLVNE...VLM	TSAKKYDVTKD...VRS	KLMDVLS
Dm-CG6680	163	IKNN...YRD...A...IQN...YVQ	PMVDFYS...PDSVI	IQINEDTNR
Dm-CG6663	78	LVKGY...NS...VVLE...GVG...ENNV	LDREGRPRGVDFDQCAS	IIINDIDKAS
Dm-CG6289	132	LVKGY...NS...VVLE...GVG...ENNV	LDREGRPRGVDFDQCAS	IIINDIDKAS
Dm-Spn5	137	VT...E...CAR...RLA...E...VQ...Q	DFDKS...S...SRKQ	INDYIAKQ
Dm-CG6687	135	VS...NK...F...NT...LLY...GAT...K	ELDFKN...PET...GLKE	INDYIAKQ
Dm-CG1342	335	PLQR...L...EN...AVK...TAAK	DGADELMLALES...HLP	SVSRVNGARSL

Aminosäure-Alignment der Serpine von *Drosophila melanogaster* (3/4)

		s3A	s4C	s3C	s1B	s2B
hum-a1AT	180	T V F A L V N Y I F F K G K W E R P F E V K D T E E E D F H V D Q . . V T T V . K V P M M K R L G M F N I Q H . C K K L S S W V L L				
Dm-Spn27a	226	. V M L L T N L I Y F N G L W R R Q F A T . . T F Q G S F F R S K . . D D Q S . R A E F M E Q T D Y F Y Y T T . S E K L K A Q I L R				
Dm-CG6717	182	T S A F L V N A I Y F K G W L Y N F K A D Q T H I A D F Y V S A . . N E I I . P V K M M T L S A S L L S G Y . I D D I D A K I I E				
Dm-CG31902	163	E S A V M I N A A F F K G Y W K T R F D K M N T K P K V F Y V S K . . S Y Q V . N V N M M S Q V G R F K M R . . T S T I D Q I I E				
Dm-CG33121	163	E S A V L V N T V Y S G A W K T R F S K K D T K L K V F H G D H . . N K K V . Y V R M M S H V G R F R I A . . D H S Y G Q I I E				
Dm-CG7219	285	T R M I L A N A L Y F K A F W E T D F I E S A T R P D N F Y P N G E . G T E P V M R V Q M M A T G G A Y P Y H E . D H E L G C K I I G				
Dm-Spn2	162	M E L I V L N A I Y F K G W E Y K F N P K L T K K R N F R V S D . . Q K S V . P V E M M S L F Q S F R A A H . D S E L G A K I I E				
Dm-CG4804	162	N T L L L L L A A N L Q S K W L P F F S A Y R T G L Y E F H S G S . . Q V K S . V P M L F D D M D F V K F A E L R D L D A R A I E				
Dm-Spn3	162	L V L V I L N A I Y F K G W Q K K F N T E Q T K S . D F H I S D . . Q K S V . P V Q M M S L V R F F C V S Y . D R E L G A N A V I E				
Dm-Spn4-V1	192	S R L V L V N A I F F K G T W Q H Q F A K H L T R P D T F H L D G . . E R T V . Q V P M M S L K E R F R Y A D . L P A L D A M A L E				
Dm-CG9454	158	S R L C L V N G I F F K G E W S I S F N E K E T T R E E D F F G S . . D R E P T . R V R M M H V C E N F F F A V . L P M F E A T A L R				
Dm-CG9455	165	T S A L L I N V L Y F K G K W Q K F M P E T T S I D H F H V D R . . D T H V . Q V N M M Y Q E D K F R F A E . L P Q L K A R A V Q				
Dm-Spn1	164	V K A L L V N A I Y F K G Q W E S K F D P A K T R A S T F Q V T A . . N K S V . P V Q M M A Q M G T F R A N Y . F R D L D A Q V I E				
Dm-CG9460	169	T S A I L V N A I Y F K A D W Q S S F P D Y A T Y A S D F V N H G . . G R K V . S V D T M S Q E D Y F R F G E . L T E L K A K V V E				
Dm-Spn43Aa	174	T N V A L V N A I Y F K A R W A R P F N E D E T R D R E F W L S E . . S R S I . Q V P T M F A D N W Y Y A D . Y P E L D A K A I E				
Dm-Spn43Ab	178	T E G V I V N G V Y S A A W Q K A F R L D K T E K R S F R T G S . . G Q S V . K V D T M W T L Q N F N Y A E . V N S L D A K A V V E				
Dm-CG1859	193	T P F L H V S G V I F R A P W A W A F D P T E T Q S Y D F F A G G . . N R P R . L V D A M F G Q H R Y R Y A E . V P A L D A Q L I E				
Dm-Nec	256	T Q A L L V N A V Y F Q G R W E H E F A T M D T S P Y D F Q H T N . . G R I S . K V A M M F N D V A F Y G L A E . L P E L G A T A L E				
Dm-CG7722	166	S S V I L V N S L F F R A K W N K I F P Q O L T Q I D D F W I N P . . R Q R M . E V S M M R Q I G Q F R Y G E . S K K L K S Q I L O				
Dm-CG10956	164	S Q G L V N A I F F N L S W E R T F N P E A T Y P R E F R V N A . . T K S V . M I P M M H E D S K F A F G T . L G N L K A T A V L				
Dm-Spn6	159	T R L V L L N A L F F K G S W A H K F S E E R T E E D I F W V G E . . E E Q V . K I N Y M N Q A K F N Y G F . L E D L G C T A L E				
Dm-Acp76A	168	E N I V A I S G M V T P L W A S H F Q S E I N R Y E V N N P G T G Y A S K D P T C V P M M H S L S S F E T M S . . T E A K G I E				
Dm-CG6680	217	K M F L L S . S L Y F K G Q W K F P F N K T L T R E E P F F S E S . . G E V I G K I P M M V Q E A N F A Y V N V G E L D G Y V L E				
Dm-CG6663	140	V T V L G I T V S Y F K A K W K Y P F D K S Q T K V E Q F Y N D G . . G S P A G K V E M M V Q T G K Y A Y V N N V K G L Q A D V L E				
Dm-CG6289	194	T M L V L G I T V S Y F K A K W K Y P F D K S Q T K V E Q F Y N A G . . G S P A G K V E M M V Q T G K Y A Y V N N V K G L Q A D V L E				
Dm-Spn5	190	T R L V L A N A A Y L K G Q W L S Q F K T E K T Y P M F F Y T S P . . S N Y S . L V S M M Q Q K G T F L M V . D E Q L S A H V L Q				
Dm-CG6687	188	T M L V L A N A A Y M K G W L S Q F K V E E T A L K E F F I N E . . R E Q E . M V Y M M H K T G A F K M T I . D E G L Q S A I I K				
Dm-CG1342	398	S K M L L F N G L Y R G S W A N F P Y Q L R D G S D E F F F M T . . N E D A V K A P M M H A R G K F E Q V A D . L P Q V R A R V L S				

		s3B	hG	hH
hum-a1AT	242	M K Y L G N A T A I F F I P D . . E G K L Q H L E N E L T . . . H D I I T K F L E N S D E R R S . .	000000	00000000
Dm-Spn27a	285	L P Y K G K N S L F V L L P Y . . A L N G I H D L V K N L E . . . N D E L K S A Q W A M E E V K . .		
Dm-CG6717	244	L P Y W N S T L S M V I I L P N . . S V D G L R K L K E K V G . . . F I D Y H . L E K K S		
Dm-CG31902	223	L P F A Y S N L S M V I V L P K . . D N G S L T Q A E A T I E . . . S Y P Q I V L T E M D		
Dm-CG33121	223	M P F D N S D L S M I G L P L . . H N T Y L S S I E K I L R . . . T L S E S L V E N N		
Dm-CG7219	350	L P Y R G N I S T M Y I T Q P F K S S V R E L M A L Q K R I T . . . A D K I E S M I R M Y R R A		
Dm-Spn2	224	L P Y R N S S L S M L I F L P D . . Q V D G L S E L E K K I V . . . G F K . P K L S K M D		
Dm-CG4804	224	L P Y E H A E E L S M L I L P N . . Q R G G L Q E L E K Q L H . . . D L D L G A L Q R M Q M E G		
Dm-Spn3	223	L P Y R N S N L S M V I F L P D . . K V D G L P E L E K K M V . . . G F T . P K L I N		
Dm-Spn4-V1	254	M N Y S D S D L S M L I V L P N . . T K T G L P A L E E K L R . . . L T T L S Q I T Q S L Y E T K		
Dm-CG9454	219	L P Y S A C N L A M I T I L L P D . . E K S N L T S L E K K L S . . . D I S L E V S V S A M N L E K		
Dm-CG9455	227	L P Y D Y S N I H M L I L L P N . . E V N G L Q E L E Q Q L N . . . T V D L A D I D A L T L Q D		
Dm-Spn1	226	L P Y L N S N L S M T I F L P R . . E V E G L S A L E E K I V . . . G F A R P . L V A K E		
Dm-CG9460	231	L P Y T G T D I V F L I L P Q . . E E Q G L A I V E E K L M . . . G I D L N E I S S Q L R R R K		
Dm-Spn43Aa	236	L F F E N I N L T M W F I L P N . . Q R S G L Q A L E Q K L K . . . G V D F N L E D R W Q W Q S		
Dm-Spn43Ab	240	L P Y Q N P D F S M L I V L P N . . R K D G L R S L Q Q S L S . . . G K N L L A E I G A M S Q Q K		
Dm-CG1859	255	V P F A T A D R M L I V F P N . . R P D G L A Q L E R K L A . . . Q S D L H Q L R S L R E R K		
Dm-Nec	318	L A Y K D S A T S M L I L L P N . . E T T G L G K M L Q Q L S R P . . E F D L N R V A H R L E R R K S		
Dm-CG7722	228	L P F E R S N L T M I L L P D . . A I D G L P E L E E K L G . . . Q L D M N E V A N K S L M K E		
Dm-CG10956	226	V P F S H G D L R M L I L L P D . . Q P D G L A A L Q M K L O . . . A M N I L S V A R N L T M M D		
Dm-Spn6	221	M P Y Q D S D L S M F I L L P Q . . E R T G I Y A L A E K L K . . . T V N L V D L A K L T V E E		
Dm-CG6680	232	I P F S S A N L G M L I L L P R . . K G V T C K D I L D N L N N O I N V E Y N D H K		
Dm-CG6680	280	L P Y G T Q D R L A M I V V L P K . . R G F K L N D V A N N L K A L G L R P I L Q L A A F R N R A S E D		
Dm-CG6663	204	L P F G E H E L V M I V L L P K . . S S Q G V N L V L Y Q L K . . . N L G L H R L L E K L E A S K N E		
Dm-CG6289	258	L P F G E H E L V M I V L L P K . . P S Q R V S L V L K O L K . . . N L G L H R L L E K L E A S K N E		
Dm-Spn5	252	L P Y R T V F E S Q E K E D S S P D E N S D I S M V L I L P P F . . N S N S L E D V L S R L N . . . A D S L D D L E Q A M P R E		
Dm-CG6687	250	L P Y R T I Y K S K E T H I S T P E S K S D I S M I I L P N S . . N K I S L N R V I S R L N . . . A D S V K K W F E R A L P Q K		
Dm-CG1342	461	L P Y E T S R Y A L C I V L P D . . E T E G L S D V I S Q I Q . . . T S D F L L A K Q F Q M K E		

		s2C	s6A	hI
hum-a1AT	284	. . A S L H L P K F S I T G T Y D L K S V L G O L G I T R K V F S N . G A D L S G V T E E A	0000000	
Dm-Spn27a	328	. . V K V T L P K F H F D Y Q O N L K E T L R S L G V R E I F E D . S A S L P G L T R G A D V A		
Dm-CG6717	283	. . V N V K L P K F K I E S K A Q L K G I F E N L G I L D V F K P . S A D L N G L V L E		
Dm-CG31902	263	. . V H V Q L P K F K I D F R M E L V E T L K S M G I Q D L F N S . S S D I S . V L L N Q		
Dm-CG33121	262	. . V H V E L P K F K I K Y Q T E L V E S L K K L G I H L F S N . T S D L S G L L I N G		
Dm-CG7219	396	. . A L V A F P K M H L T E S V N L K T V M O R M G L G G I F S A V Q N D L S I I A I N E A T R N A L G G N S L Q N L E A Q R R A G		
Dm-Spn2	263	. . V T L R L P K F K I E F F A Q L N K V L V A M G I Q D A F E K . S A D F K D L V E N		
Dm-CG4804	269	. . V Q V L L P K F S I D F E C S L R Q P L K O L G F E E I F A A . S A N F K H L H A S		
Dm-Spn3	262	. . V H L R L P K F K I E F S A R L E Q V L I A M G I Q D A F K T . S A D F N D L V A N		
Dm-Spn4-V1	298	. . V A L K L P R F K A E F Q V E L S E V F Q K L G M S R M F S D . Q A E F G K M L Q S P		
Dm-CG9454	263	. . V D V K I P S F T A E F Q Q E L S Q V L M L M G M N R I F S G . Q A E L G G M L Q S E		
Dm-CG9455	271	. . V E I F L P R M C I E Y D V D L K Q V L N Q L G T E V F S D . K A K L D G L F T S Q		
Dm-Spn1	265	. . V Y L K L P K F K I E F R D E L K E T L E K L G I R L F T D . K S D L S G L F A D K		
Dm-CG9460	275	. . V R V Q L P K F K F E F D V P L Q A A L E E L G I K R L F S P . G A N L S S L Y Q G S		
Dm-Spn43Aa	280	. . V S V Y L P K F K F E F D T D L R P T L R K M G I S A M F S D A A D F S N I F Q D S P		
Dm-Spn43Ab	284	. . V E V L L P K F S V T F G L G L E G P P K L G V H T M F S R D G D F G N M Y R M F		
Dm-CG1859	299	. . V A L T L P K L R V L V H S D L K H V L E E L G L A K L F T S . E V H L S E F S S I L S S		
Dm-Nec	364	. . V A V R L P K F Q F E F E Q D M T E P L K N L G V H Q M F T P . N S Q V T K L M D Q P		
Dm-CG7722	272	. . V D V T I P K F R I E C T V D L K V P L Q C M G N S V F D A G Q A D L S D L F E M K		
Dm-CG10956	270	. . V F V G I P K F K I H S D L E L S P A F E K M G K I K D I F K P . S K S F S T L L H R N		
Dm-Spn6	265	. . V H V K F P K F K V D Y S L E L A E K L K Q L G T K M F T D . Q A E F S N L L E S P		
Dm-Acp76A	272	. . D V H L L L P I F K E K F D Y N I A K F F N G I N E D T F K D S A F K S K A K I K I N		
Dm-CG6680	331	. . N E V E V M P K F V T A T D F T L K G V L I O M G I R D L F E N T A N L D R M S S G		
Dm-CG6663	250	. . T D V E V K L P K F D T R S V L S L E D T V Y D A G L T D L R N D F A D L D K L I A I G H		
Dm-CG6289	304	. . S D V E V K L P K F D T R S V L S L E D T V Y E A G L T D L R N E F A D L G R M L I P T G D		
Dm-Spn5	312	. . I E V S L P K F E F E Q R L E L N P I L A K M G V S M F D E S V A T F D D L T S E T		
Dm-CG6687	310	. . I E L S L P K F Q F E Q R L E L T P I L S L M G V N T M F T R N . A T F G D L T A D P		
Dm-CG1342	505	. . L H I S M P K F Q V E E T S R S E A M L K O M G L K K V F S R T E A Q L S L L S E D P		

Aminosäure-Alignment der Serpine von *Drosophila melanogaster* (4/4)

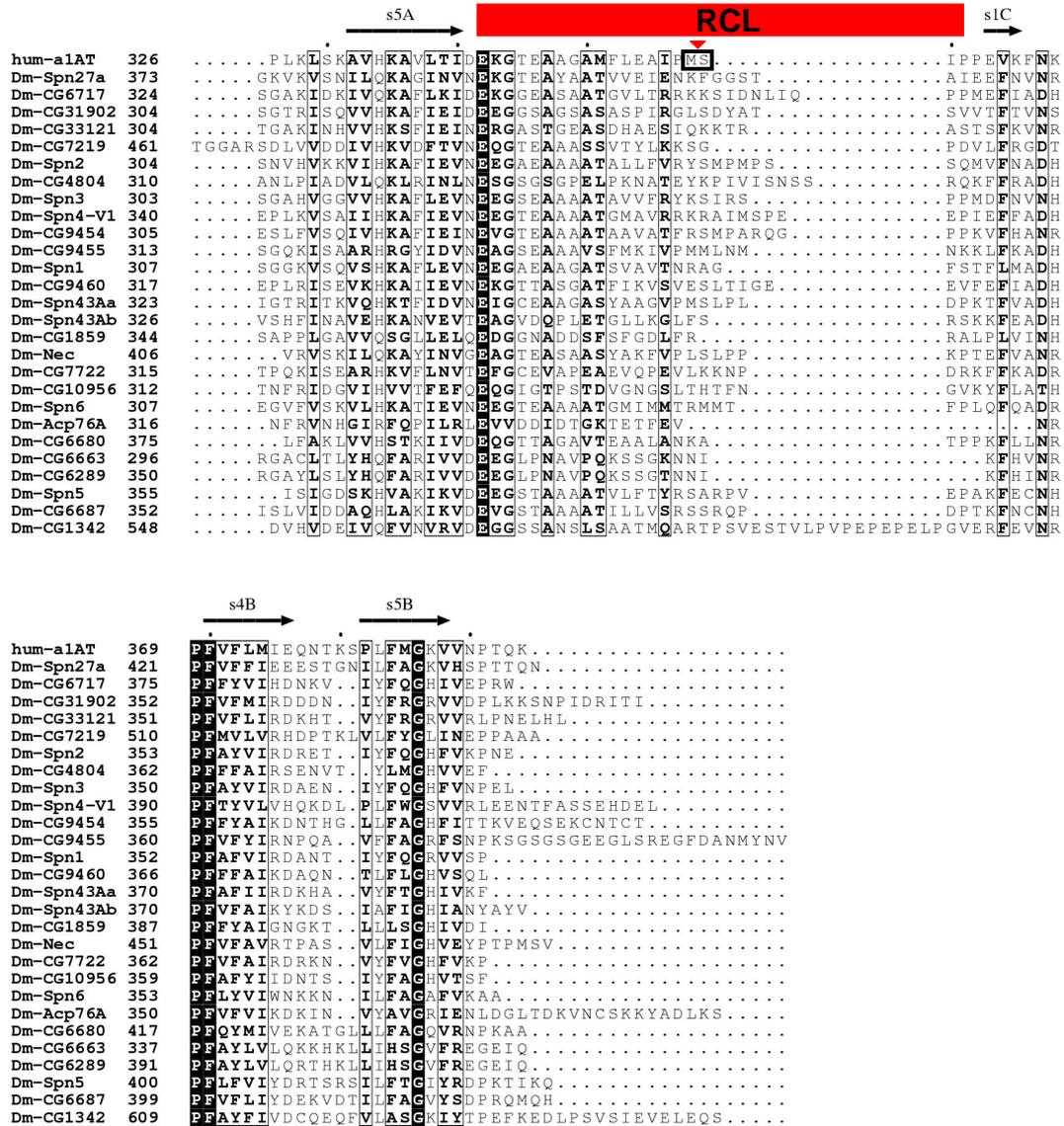


Abbildung 46: Alignment der 27 Serpine von *Drosophila melanogaster*

Beschriftung siehe Abbildung 45.

8.5.5 Aminosäure-Alignment der Serpine von *Arabidopsis thaliana*

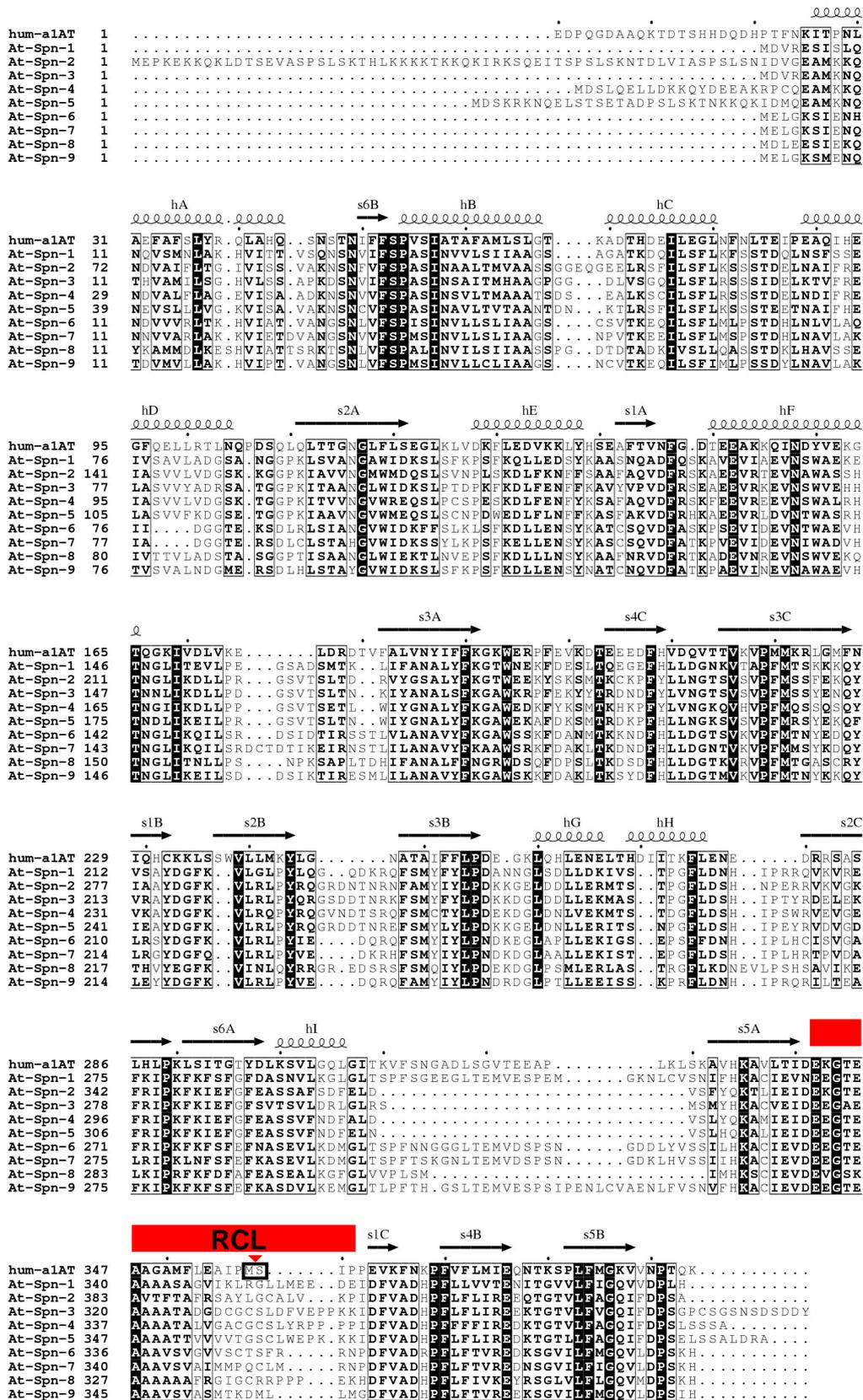


Abbildung 47: Alignment der neun Serpine von *Arabidopsis thaliana*

Beschriftung siehe Abbildung 45.

Danksagung

An erster Stelle möchte ich Prof. Dr. H. Ragg für das interessante Thema, die hervorragende wissenschaftliche Betreuung und die vielen anregenden Diskussionen danken.

Bei Prof. Dr. A. Dress von der Fakultät für Mathematik möchte ich mich ganz herzlich für die Übernahme des Zweitgutachtens bedanken.

Bei allen aktiven und ehemaligen Mitgliedern der Arbeitsgruppe „Zelluläre Genetik“ möchte ich mich für die überaus angenehme Arbeitsatmosphäre und gute Zusammenarbeit bedanken. Namentlich erwähnen will ich hier besonders Dr. Stephan Brinkmeyer für die Einführung in die Arbeitstechniken und Dipl. Biochem. Mareke Oley und Dipl. Biol. Ulf Tödttmann, die mit ihren Korrekturvorschlägen die „Lesbarkeit“ der vorliegenden Arbeit gesteigert haben. Auch meinen beiden Lieblingsdiplomandinnen Dipl. Biol. Hülya Nazlican und Dipl. Biol. Julia Ladewig möchte ich für die tolle Zusammenarbeit danken.

Mein Dank gilt auch meinen Mit-Stipendiaten, den Zellkulturtechnikern und den BiBiTec'ern für so manchen Blick über den Tellerrand.

Bei der Deutschen Forschungsgemeinschaft (DFG) möchte ich mich für das Stipendium im Rahmen der Graduiertenkollegs „Bioinformatik“ und „Strukturbildungsprozesse“ bedanken.

Zu guter Letzt möchte ich meinen Eltern herzlich für ihre tolle Unterstützung danken.