Struktur- und Funktionsanalyse von Ataxin-2 mittels bioinformatischer Methoden.

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Hohen Medizinischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von: Michael Golatta
aus : Wuppertal

2006

Angefertigt mit Genehmigung der
Medizinischen Fakultät der Universität Bonn

1. Gutachter: Prof. Dr. Ullrich Wüllner

2. Gutachter: Prof. Dr. Max P. Baur

Tag der mündlichen Prüfung: 13.09.2006

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert

Aus der Klinik und Poliklinik für Neurologie
der Universitätsklinik Bonn
Direktor: Prof. Dr. med. Thomas Klockgether

Meinen Eltern und Großeltern

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis:

aa	„amino acid“ (Aminosäure)
ADCA	„autosomal dominant cerebellar ataxias“ (autosomal dominant vererbte zerebelläre Ataxie)
A2bp1	Ataxin-2 Bindungsprotein 1 (Maus)
A2BP1	Ataxin-2 Bindungsprotein 1 (human)
A2RP	„Ataxin-2 related protein“ (ein zu Ataxin-2 verwandtes Protein auf Chr. 16)
BLAST	„Basic Local Alignment Search Tool”
CE	„Combinatorial Extension of the Optimal Path“ (Kombinatorische Erweiterung des optimalen Pfads)
DCP	„Decapping complex“ (Entkappungskomplex)
DDBJ	„DNA Data Bank of Japan“ (Japanischer Zweig der internationalen Nukleotidsequenzdatenbank)
DRPLA	„dentatorubral pallidoluysian atrophy” (dentatorubrale pallidoluysiane Atrophie)
DSSP	„Dictionary of Secondary Structure of Proteins“ (Standardprogramm zur Gewinnung von Sekundärstrukturinformationen)
EA	episodische Ataxien
EBI	„European Bioinformatics Institute” (europäisches Institut für Bioinformatik)
EMBL	„European Molecular Biology Laboratory” (Europäische Molekularbiologielaboratorien)
EPO-R	„endogenous erythropoietin receptor“ (endogener Erythropoetinrezeptor)
E-value	„Expectation value“ (Erwartungswert)
ExPASy´s	„**Ex**pert **P**rotein **A**nalysis **Sy**stem“ (Proteinanalysierungsprogramm)
FASTA	„fast all”
Fox	„feminizing locus on X“ (Feminisierungsort auf dem X-Chromosom)
GRAVY	„Grand Average of Hydropathicity”
ExPASy	„Expert Protein Analysis System”
HGNC	„HUGO Gene Nomenclature Committee“

HSA	„hereditary spastic ataxia“ (hereditäre spastische Ataxie)
HUGO	„Human Genome Organisation“
IDCA	idiopathischen zerebellären Ataxie
INSDC	„International Nucleotide Sequence Database Collaboration“ (Internationale Kooperation einer Nukleotidsequenzdatenbank)
IUP	„intrinsically unstructured proteins“ (physiologisch unstrukturierte Proteine)
MJD	Machado Joseph Erkrankung (Machado Joseph disease)
NCBI	„National Center for Biotechnological Information” (Nationales Zentrum für biotechnologische Informationen)
NIs	„intranuclear inclusions“ (intranukleäre Einschlüsse)
NMR	„nuclear magnetic resonance“ (Kernresonanzspektroskopie)
ORF	„open reading frame“ (offene Leseraster)
Pab1p	„poly(A)-binding protein“ (Poly(A) – Bindungsprotein)
PAM	„PABP interacting motif“ (Interaktionsmotiv mit dem Poly(A) - Bindungsprotein
PBP1	„Pab1p-binding protein“ (Pab1p Bindungsprotein) (Hefe)
PDB	„Protein Data Bank“ (Protein Datenbank)
Pfam	„Protein family database” (Proteindomändatenbank)
polyQ	Polyglutaminkette
PONDR	„Predictor Of Natural Disordered Regions“ (Vorhersage von unstrukturierten Regionen)
PSI-BLAST	„Position Sensitive Iterated - Basic Local Alignment Search Tool”
RCSB	„Research Collaboratory for Structural Bioinformatics“
RMSD	„root-mean-square-deviation“ (mittlere quadratische Abweichung)
SAPS	„Statistical analysis of protein sequences“ (statistische Analyse von Proteinsequenzen)
SCA	“spinocerebellar ataxin” (spinozerebelläre Ataxie)
SCOP	„Structural Classification of Proteins”
SD	“standard deviation” (Standardabweichung)
SIB	Schweizerischen Institut für Bioinformatik
SMART	„Simple Modular Architecture Research Tool“
SMBA	„Spinal bulbar muscular atrophy“ (Spinobulbäre Muskelatrophie Typ Kennedy)

snRNPs	„small nuclear ribonucleoprotein particles“ (Ribonucleinproteinkomplexes)
SWISS-PROT	„Swiss Protein” (schweizerische Proteindatenbank)
TrEMBL	„translated European Molecular Biology Laboratory“
UTR	„untranslated region“ (nichttranslatierte Region)

1. Einleitung

1.1. Bioinformatik als Grundlage der Genomanalyse

Die Bioinformatik (englisch „bioinformatics“, oder auch „computational biology“ genannt) ist eine relativ junge Disziplin im Bereich der Naturwissenschaften, die in den letzten Jahren erheblich an Bedeutung gewonnen hat. Die Bioinformatik zeichnet sich besonders durch ihren interdisziplinären Charakter aus und verbindet Bereiche wie die Informatik, die Mathematik und die Statistik mit der Biologie, der Biochemie und der Medizin. Eine einheitliche Definition des Begriffes „Bioinformatik“ existiert nicht, da hierunter ein großes Gebiet zusammengefasst wird, das viele Disziplinen vereinigt und daher je nach Perspektive unterschiedlich gesehen wird. Entstanden ist die Bioinformatik aus der Notwendigkeit heraus, mit großen Mengen genetischer und biochemischer Daten effizient umzugehen.
Eine wichtige Aufgabe der Bioinformatik ist die Datenverwaltung und -speicherung sowie die Bereitstellung des möglichst tagesaktuellen Wissensstandes. Hier bietet das World Wide Web eine ideale Basis und so stehen die entwickelten Datenbanken dort jederzeit zur privaten Nutzung frei zur Verfügung. Aufgrund der Tatsache, dass die Bioinformatik für diese rapide wachsenden Datenbanken Benutzeroberflächen entwickelt hat, die zahlreiche Analysenprogramme, so genannte Tools, miteinander verbinden, bilden diese auch für komplexe biologische und medizinische Fragestellungen eine wertvolle Basis und dienen so der Informationsgewinnung.
Weiterhin finden bioinformatische Methoden beispielsweise Anwendung bei der automatischen Sequenzierung und Annotation pro- und eukaryontischer Genome, bei homologiebasierenden Protein- und Nukleotidsequenzanalysen oder bei der Homologiemodellierung („homology modelling“) und den daraus folgenden Struktur- und Funktionsvorhersagen.
Besonders in der Molekularbiologie spielt die Bioinformatik eine bedeutende Rolle bei der Aufnahme und Verwaltung der Nukleotid- und Aminosäuresequenzen in Datenbanken.
Im Rahmen der Genomprojekte ist es in den letzten Jahren zu einem explosionsartigen Anstieg der Datenmenge und damit zu einem stetigen Wachstum der bioinformatischen

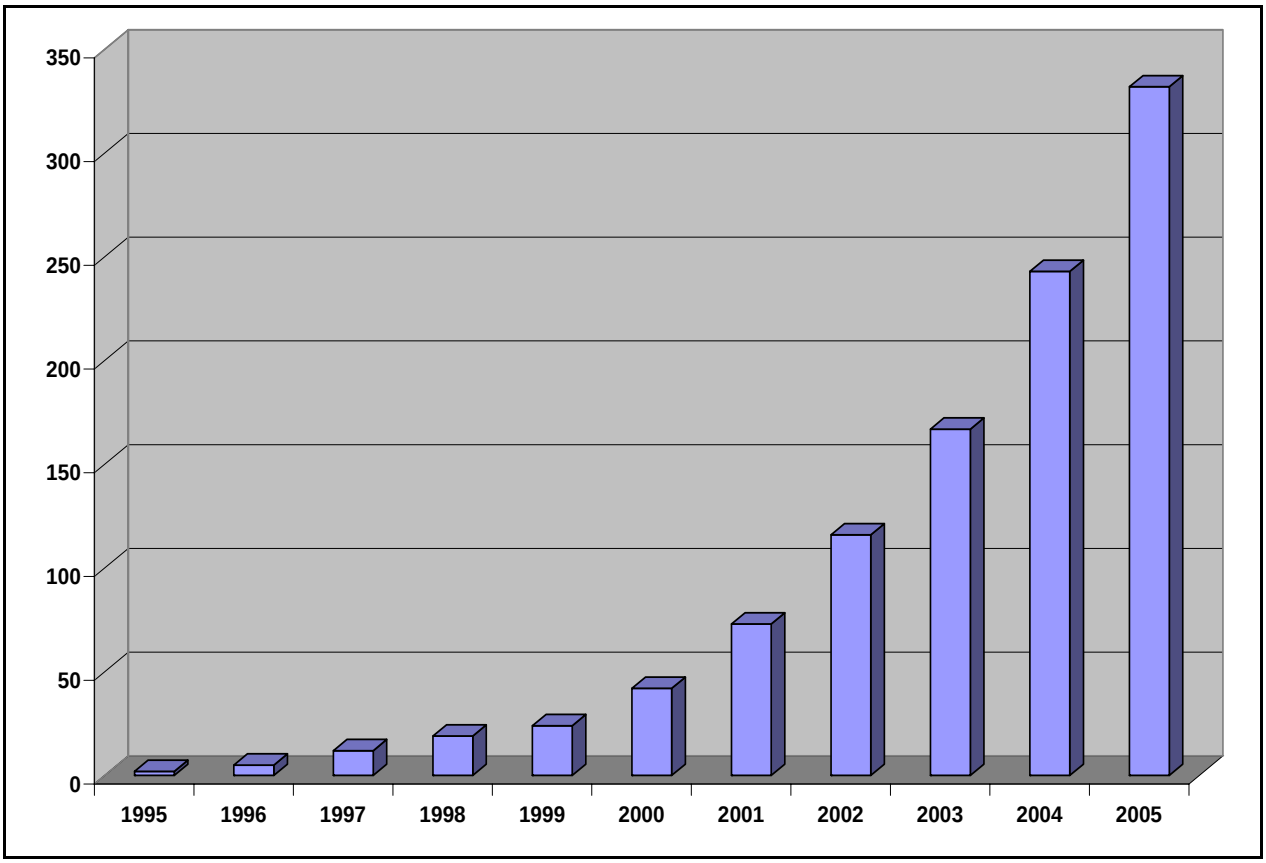
Anwendungsmöglichkeiten gekommen, ohne die dieser Wissenszuwachs nicht zu bewältigen gewesen wäre.

Die Genomprojekte die inzwischen vollständig sind bzw. an denen zurzeit gearbeitet wird, konzentrieren sich hauptsächlich auf evolutionär interessante Organismen, biologische Modellorganismen und pathogene Organismen. Die Ergebnisse aus diesen Forschungsprojekten werden in genomischen Sequenzdatenbanken gespeichert und sind frei verfügbar. Die Sequenzdatenbanken enthalten hauptsächlich einzelne vollständige Gene, Abschnitte mit mehreren Genen und partielle Sequenzen von Genen. Neben den Sequenzen fallen auch noch Annotierungsdaten wie Molekülstrukturen, biologische Funktionen, Eigenschaften von ähnlichen Sequenzen usw. an. Neben den eigentlichen Primärsequenzen werden diese Informationen Teil der Datenbankeinträge. Um die Daten zu verknüpfen werden sie in Datenbanken organisiert, in denen mit entsprechenden Algorithmen nach grundlegenden Mustern und Beziehungen gesucht werden kann. Dadurch sind heutzutage Entdeckungen allein durch Analyse der bereits existierenden Datenmenge mit speziellen Methoden der Bioinformatik möglich.

Besondere Aufmerksamkeit erregte das im Oktober 1990 in den USA begonnene Human-Genomprojekt. Zu Beginn nahmen über 1000 Wissenschaftler in 40 Ländern teil mit dem Ziel, das menschliche Genom bis 2010 zu sequenzieren. Im Juni 1995 schloss sich auch Deutschland dem internationalen Human-Genomprojekt der „Human Genome Organisation“ an. 1998 bekam das Projekt private Konkurrenz durch die neu gegründete US-Firma Celera, die mit der so genannten "whole genome shotgun strategy" eine schnellere Sequenzierung ermöglichte, die aber laut Kritikern auch fehleranfälliger ist. Schließlich wurde schon 2001 eine erste Version des vollständigen genetischen Codes des Homo sapiens veröffentlicht [Lander et al. 2001], die aber noch diverse Lücken und Fehler enthielt. An dieser Version wurde weiter gearbeitet, so dass inzwischen die menschliche Genomsequenz mit 2,85 Milliarden Nukleotiden beschrieben wird, die nur noch von 341 Lücken unterbrochen ist. Dies deckt 99 % des Genoms ab, mit einer Fehlerrate von 1 auf 100.000 Basen. Außerdem wurde festgestellt, dass die ursprünglich auf über 100.000 geschätzte Anzahl an proteincodierenden Genen, die schon 2001 auf nur noch 30.000 – 40.000 geschätzt wurde, tatsächlich nur ca. 20.000 – 25.000 beträgt [2004]. Dies entspricht ungefähr der Anzahl codierender Gene in Caenorhabditis elegans, kurz C. elegans genannt oder Drosophila melanogaster. Allerdings sind die humanen Gene komplexer und weisen mehr Spleißungsalternativen und eine größere Anzahl an Proteinprodukten auf, so dass man nicht unbedingt

Rückschlüsse von der Anzahl der Gene auf den Entwicklungsgrad der jeweiligen Spezies ziehen kann.
Aber die Entschlüsselung des menschlichen Genoms ist nur eines von vielen Projekten. Beginnend mit dem Bakterium Haemophilus influenzae [Fleischmann et al. 1995] sind bis heute (Stand 17. Januar 2006) 335 Genome vollständig sequenziert und publiziert worden (siehe Abbildung 1) (siehe detaillierte Liste unter http://www.genomesonline.org/CompleteGenomesList.html [Kyrpides 1999; Bernal et al. 2001; Liolios et al. 2006] und http://www.TIGR.ORG/tdb/mdb/mdbcomplete.html). Darunter befinden sich zurzeit alleine 40 eukaryontische Genome, wie zum Beispiel Saccaromyces cerevisiae [1997], welches als erstes komplettes eukaryontisches Genom veröffentlicht wurde. Desweiteren findet sich unter diesen Caenorhabditis elegans [1998], Drosophila melanogaster [Adams et al. 2000], Mus musculus [Waterston et al. 2002], Rattus norvegicus [Gibbs et al. 2004], sowie der im Rahmen des Human-Genomprojektes entschlüsselte Homo sapiens [Lander et al. 2001]. An knapp 1500 weiteren Genomen wird zurzeit an der Sequenzierung gearbeitet, unter denen sich ca. 570 eukaryontische Genome befinden.

Abbildung 1: Die Entwicklung vollständig sequenzierter Genome von 1995 – 12/2005



Im Gegensatz zu dieser rapide zunehmenden Anzahl bekannter genomischer Sequenzen ist die Anzahl der experimentell untersuchten und charakterisierten Proteine relativ klein. Diese Differenz zwischen bekannten und experimentell untersuchten Sequenzen wird in Zukunft eher größer als kleiner werden, da weltweit nicht annähernd genügend Laborkapazitäten zur Verfügung stehen, um die im Rahmen der Sequenzierungsverfahren beschriebenen Strukturen weiter zu untersuchen. Beispielsweise sind nur ca. 7% der circa 20.000 C. elegans Proteine Teil laborbiologischer Experimente. Noch gravierender dürfte der Unterschied bei Spezies sein, die generell seltener Verwendung als experimentelle Organismen finden.
Diese Diskrepanz veranschaulicht, welch wertvolle Arbeit die Bioinformatik liefern kann, die es zunehmend ermöglicht, sozusagen virtuell Experimente durchzuführen, um so

schon frühzeitig entscheiden zu können, in welchen Bereichen es sich lohnen würde, tatsächlich experimentell tätig zu werden. Mittels homologiebasierender Sequenzanalysen besteht die Möglichkeit, Proteine in struktureller und speziell funktioneller Hinsicht zu erforschen und so die Lücke zwischen wenigen experimentell untersuchten und zahlreichen neuen unbekannten Proteinen schließen zu können.
Die bioinformatische Untersuchung eines Proteins findet auf unterschiedlichen Ebenen statt, mit dem Ziel eine möglichst exakte Prognose über den Aufbau und die Funktion des Proteins zu bekommen. So existieren zum Beispiel wenige spezifische Aminosäuren, die für die katalytische Aktivität eines Enzyms verantwortlich sind. Diese funktionell spezifischen Aminosäuren sind ihrerseits in die dreidimensionale Struktur eingebunden, die für die zweckmäßige Bereitstellung der katalytischen Reste im räumlichen Kontext notwendig ist. Die Funktion des Proteins in der Zelle hängt nicht allein von seiner Beschaffenheit ab; es interagiert mit anderen Proteinen und Zellkomponenten gewebsspezifisch und abhängig von der Lokalisation (z.B. nukleär, cytoplasmatisch oder membrangebunden). Diese und andere Aspekte müssen bei der Übertragung der Funktion eines experimentell untersuchten Proteins auf unbekannte Homologe berücksichtigt werden.
Mittels bioinformatischer Standardmethoden wird die Signifikanz der Sequenzähnlichkeit ermittelt und damit die Homologie. Diese wiederum erlaubt Rückschlüsse auf die Struktur und bedingt auf die Funktion des Proteins. Dieser Analyse liegt die Idee zugrunde, dass beide Proteine aus einem gemeinsamen Vorläufer entstanden sind. Bei entsprechender Homologie über die gesamte Sequenz kann es sich um Orthologe, also dem funktionell vergleichbaren Gegenstück in anderen Spezies, oder um Paraloge, die eine ähnliche Funktion in einem anderen Kontext in der gleichen Spezies aufweisen, handeln.
Je größer die Datenbanken und das Wissen um die Gene und ihre Funktion werden, um so effizienter werden sich die Methoden der Bioinformatik verwenden lassen, die letztendlich darauf aufbauen, durch gewaltige Rechenleistung neue, unbekannte Proteine mit den schon beschriebenen, bekannten zu vergleichen.

1.2. Ataxie

Ataxie (gr. ataxia „die Unordnung“) bezeichnet in der klinischen Neurologie die gestörte Koordination bei der Ausführung von Bewegungen.
Als Ataxien werden nicht-fokale Erkrankungen des Kleinhirns und seiner Verbindungen bezeichnet, deren Leitsymptom eine progressive Ataxie ist.
Neuropathologisch liegt diesen Erkrankungen eine Atrophie des Kleinhirns zugrunde, die entweder alleine auftritt oder in Kombination mit degenerativen Veränderungen des Rückenmarks, des Hirnstamms, der Basalganglien und / oder der peripheren Nerven vorkommt.
Fokale Erkrankungen des Kleinhirns und Polyneuropathien werden nicht zur Krankheitsgruppe der Ataxien gezählt, auch wenn Ataxie das Hauptsymptom darstellt.

1.2.1. Klassifikation der degenerativen Ataxien

Basierend auf den Arbeiten von [Holmes 1907] und [Greenfield 1954] wurden die degenerativen Ataxien bis Anfang der 80´er Jahre nach den zugrunde liegenden neuropathologischen Veränderungen klassifiziert. Dabei unterschied man im wesentlichen zwischen der rein zerebellären Atrophie, der spinalen Atrophie, der spinozerebellären Atrophie sowie Erkrankungen, die sowohl das Zerebellum als auch den Hirnstamm betreffen und als olivo-ponto-zerebelläre Atrophie bezeichnet wurden.
Diese neuropathologische Klassifikation zeigte sich aber für den klinischen Alltag als ungeeignet, da unter einem einzigen Namen Krankheiten zusammengefasst werden können, die zwar eine ähnliche Pathologie aufweisen, jedoch gravierende Unterschiede in Ätiologie, Klinik und Prognose zeigen können. Außerdem konnte es nach dieser Krankheitseinteilung vorkommen, dass die Erkrankungen bei Personen aus derselben Familie verschiedenen Krankheitsgruppen zugeteilt wurden.
Deshalb schlug die britische Neurologin Anita Harding eine neue Einteilung vor, die zunächst zwischen erblichen und nicht erblichen Ataxien unterschied [Harding 1983]. Die genetisch bedingten Ataxien wurden in autosomal-dominante zerebelläre Ataxien (ADCA) und autosomal-rezessive zerebelläre Ataxien unterteilt. In äußerst seltenen Fällen kommen auch X-chromosomal vererbte Formen vor. Dagegen werden die nicht-genetisch bedingten Ataxien in symptomatische, d.h. die Ataxie ist ein Symptom im Rahmen einer fassbaren Krankheitsursache, und idiopathische zerebelläre Ataxien

(IDCA), deren Ursache unbekannt ist, unterteilt. Letztgenannte machten in einem kürzlich untersuchten Kollektiv von Patienten mit sporadischen Ataxien 58% der Fälle aus [Abele et al. 2002].

Des Weiteren unterteilte Harding die autosomal-dominanten zerebellären Ataxien in drei Hauptgruppen:

ADCA I: progressive Ataxie mit zusätzlichen nicht zerebellären Symptomen

ADCA II: progressive Ataxie mit pigmentärer Retinadegeneration

ADCA III: progressive, rein zerebelläre Ataxie ohne zusätzliche Symptome

Aber auch mit dieser Einteilung der autosomal-dominanten Ataxien zeigte sich in der klinischen Praxis, dass sich teils Personen aus einer Familie mit unterschiedlichen Symptomen präsentierten und so in unterschiedliche Gruppen zugeordnet werden mussten.

Mit der genetischen Entzifferung diverser dominanter Ataxien stellte sich die schon vermutete Heterogenität der ADCA heraus. So wurde die Harding Klassifikation für autosomal-dominant vererbte Ataxien durch eine molekulargenetische Einteilung ersetzt. Nach dieser zurzeit aktuellen Nomenklatur wird unterschieden zwischen der Gruppe der spinozerebellären Ataxien (SCA), in der in den letzten Jahren eine wachsende Anzahl von genetischen Subtypen nachgewiesen wurde, der dentatorubralen pallidoluysianen Atrophie (DRPLA), den episodische Ataxien (EA1 und EA2) sowie der spastischen Ataxie (HSA „hereditary spastic ataxia“).

Tabelle 1. Übersicht über die autosomal-dominanten Ataxien

Erkrankung	Genname	chromosomale Lokalisation	Protein	Quellenangabe
SCA 1	ATXN1	6p23	Ataxin-1	[Orr et al. 1993]
SCA 2	ATXN2	12q24.1	Ataxin-2	[Imbert et al. 1996; Nechiporuk et al. 1996; Pulst et al. 1996; Sanpei et al. 1996]
SCA 3/ MJD	ATXN3	14q24.3 - q31	Ataxin-3/ Machado-Joseph disease protein 1	[Kawaguchi et al. 1994]
SCA 4	Q9H7K4	16q22.1	Puratrophin-1	[Ishikawa et al. 2005]
SCA 5	SCA5	11p11 - q11		[Ranum et al. 1994]
SCA 6	CACNA1A	19p13	P/Q Calciumkanal Protein	[Zhuchenko et al. 1997]
SCA 7	ATXN7	3p21.1 - p12	Ataxin-7	[David et al. 1997]
SCA 8	KLHL1AS	13q21		[Koob et al. 1999]

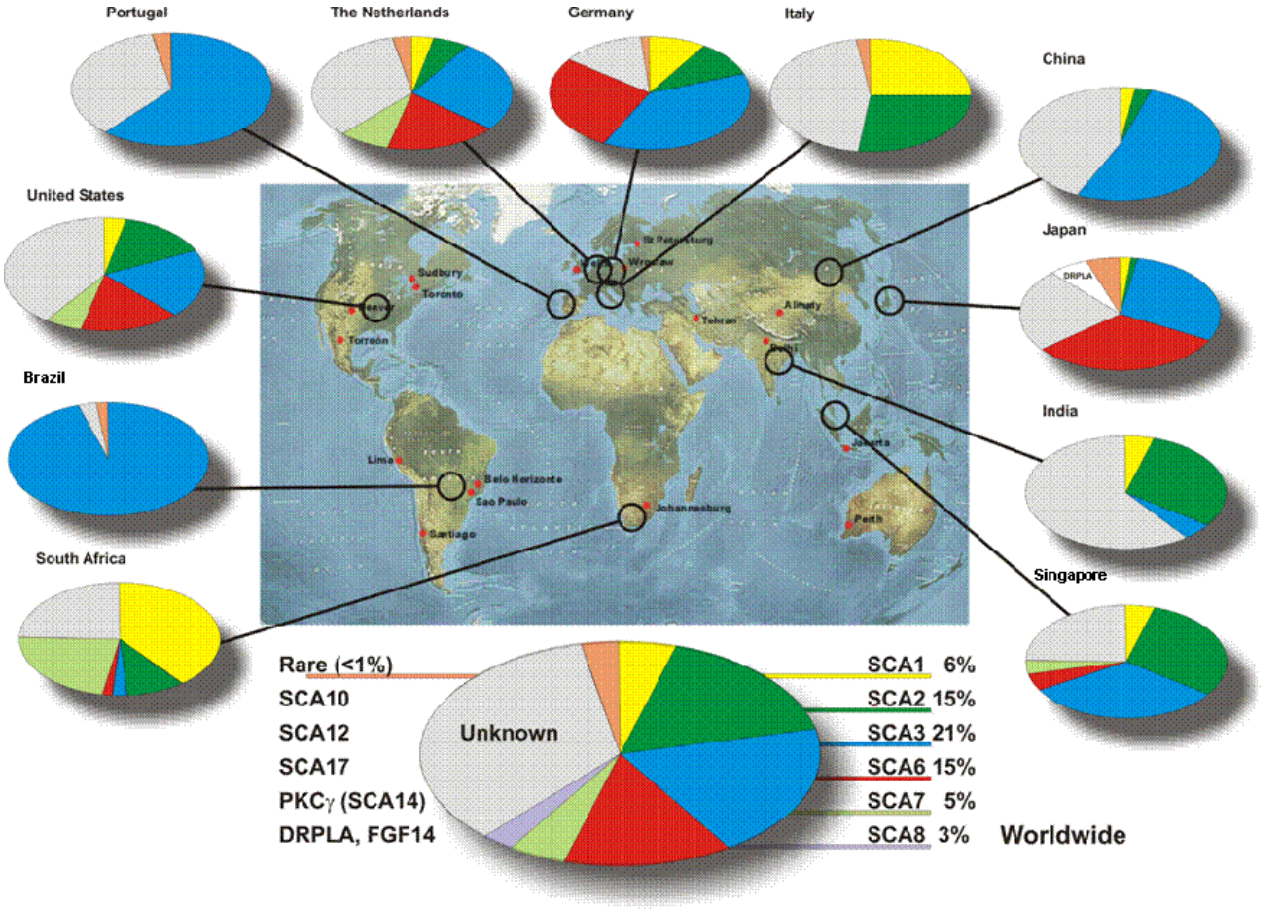
SCA 9		reserviert		
SCA 10	ATXN10	22q13	Ataxin-10	[Matsuura et al. 1999; Matsuura et al. 2000]
SCA 11	SCA11	15q14 - q21.3		[Worth et al. 1999]
SCA 12	PPP2R2B	5q31 - q33	Phosphatase - Untereinheit (PP2A – PR55ß)	Holmes et al 1999 , Fujigasaki et al 2001
SCA 13	SCA13	19q13.3 - q13.4		[Herman-Bert et al. 2000]
SCA 14	PRKCG	19q13.4	Proteinkinase C, Gamma Typ	[Yamashita et al. 2000; Brkanac et al. 2002; Yabe et al. 2003; Chen et al. 2005]
SCA 15	SCA15	3p24.2-3 (Verdacht)		[Knight et al. 2001; Storey et al. 2001; Knight et al. 2003; Gardner et al. 2005]
SCA 16	SCA16	8q22.1 - q24.1		[Miyoshi et al. 2001]
SCA 17/ TBP	TBP	6q27	TATA - Box bindendes Protein/ TFIID	[Koide et al. 1999; Nakamura et al. 2001]
SCA 18	SCA18	7q22-q32 (Verdacht)		[Brkanac et al. 2002]
SCA 19	SCA19	1p21 - q21		[Verbeek et al. 2002; Schelhaas et al. 2003; Schelhaas et al. 2004]
SCA 20	SCA20	11cen		[Knight et al. 2004a; Storey et al. 2005]
SCA 21	SCA21	7p21 – 15		[Devos et al. 2001; Vuillaume et al. 2002]
SCA 22	SCA22	1p21 - q23 (möglicherweise identisch mit SCA 19)		[Chung et al. 2003; Chung und Soong 2004; Schelhaas et al. 2004]
SCA 23	SCA23	20p13 - 12.3		[Verbeek et al. 2004]
SCA 24		bisher nicht vergeben		
SCA 25	SCA25	2p21 - p13		[Stevanin et al. 2004; Stevanin et al. 2005]
SCA 26		19p13.3		[Yu et al. 2005]
SCA 27	FGF14	13q34	Fibroblasten-wachstumsfaktor 14	[van Swieten et al. 2003; Brusse et al. 2005]
SCA 28	SCA28	18p11.22 - q11.2		[Cagnoli et al. 2006]
DRPLA	DRPLA	12p13.31	Atrophin-1	[Smith 1958; Naito und Oyanagi 1982; Koide et al. 1994]
EA1	KCNA1	12p13	spannungsaktivierbare Kaliumkanäle des Kv1.1 Typs	[Browne et al. 1994]
EA2	CACNA1A	19p13	alpha- 1A Untereinheit eines spannungsabhängigen Kalziumkanals	[Ophoff et al. 1996]
	CACNB4	2q22 - q23	beta-4 Untereinheit eines dihydropyridin-empfindlichen Kalzium-kanals des L-Typs	[Escayg et al. 2000]
HSA	SAX1	12p13		[Meijer et al. 2002]

1.2.2. Dominante Ataxien / Spinozerebelläre Ataxien

Die autosomal-dominanten zerebellären Ataxien sind eine heterogene Gruppe dominant vererbter Erkrankungen, deren gemeinsames Merkmal eine progrediente Ataxie ist, meist mit zusätzlichen extrazerebellären Symptomen. Neuropathologisch findet sich fast immer ein Untergang von Purkinje-Zellen, während degenerative Veränderungen von Hirnstamm, Basalganglien, Retina und peripheren Nervensystem unterschiedlich stark ausgeprägt sind [Burk et al. 1999b; Mizusawa et al. 2003].

Die Prävalenz der dominanten Ataxien wird auf 1-10 : 100.000 geschätzt [Klockgether 2000]. Die meisten Untersuchungen wurden an geographisch isolierten Bevölkerungsgruppen durchgeführt, daher schwanken die Häufigkeiten zwischen den einzelnen Untersuchungen, wobei so genannte Gründereffekte (Abstammung der Bevölkerungsgruppe von einem oder von wenigen gemeinsamen Vorfahren) eine große Rolle spielen. Selbst innerhalb Deutschlands gibt es erhebliche regionale Unterschiede.

Abbildung 2: Übersicht über die regional unterschiedlichen Auftretenswahrscheinlichkeiten der autosomal-dominanten Ataxien



[Schols et al. 1997a; Moseley et al. 1998; Saleem et al. 2000; Storey et al. 2000; Tang et al. 2000; Maruyama et al. 2002; Silveira et al. 2002; van de Warrenburg et al. 2002; Bryer et al. 2003; Brusco et al. 2004; Schols et al. 2004; Shimizu et al. 2004; Zortea et al. 2004; Jiang et al. 2005b]

Graphik von L Schöls, P Bauer, T Schmidt, T Schulte, O Reiss; Universität Tübingen und Ruhruniversität Bochum.

Bei den meisten Patienten setzen die Krankheitssymptome im Erwachsenenalter ein, mit einer großen Varianz zwischen den verschiedenen Untergruppen und auch innerhalb der einzelnen Familien. Im Durchschnitt versterben die Patienten frühzeitig, ca. 20-25 Jahre nach dem Krankheitsbeginn. Bis heute wurden 26 mit SCA assoziierte Genorte (SCA 1-8, 10-23 und 25-28) gefunden, die in der humanen Gennomenklaturdatenbank „Genew: Human Gene nomenclature Database [Wain et al. 2002; Wain et al. 2004]“ verzeichnet sind, in der bisher insgesamt 22.000 menschliche

Gene verzeichnet und mit Kürzeln versehen wurden. Von diesen 26 sind SCA 1-3 und 6 am häufigsten anzutreffen [Wüllner 2003]. SCA19 und SCA22 könnten identisch sein, da ihre Genorte nahezu an der gleichen Stelle liegen, allerdings wurden zunächst beide vom HGNC „HUGO Gene Nomenclature Committee“, dem Komitee für Gennomenklatur anerkannt, da in den jeweiligen Erstveröffentlichungen [Verbeek et al. 2002; Chung et al. 2003; Schelhaas et al. 2003] unterschiedliche Symptome beschrieben wurde. Die Frage, ob hier Anpassungsbedarf durch das HGNC dahingehend besteht, dass SCA19 oder SCA22 wieder aus der Datenbank entfernt werden müssen, wurde bisher nicht geklärt [Chung und Soong 2004; Schelhaas et al. 2004].

Die Benennung der SCA geschah in der Reihenfolge der Identifizierung der Gene in einzelnen Familien. Lücken in diesem System lassen sich dadurch erklären, dass Reservierungen für vermutete Genloci vorgenommen wurden, ohne dass bisher genetische Daten dazu publiziert wurden.

Obwohl bei allen SCAs Ataxie das auffallenste Symptom darstellt, präsentieren sie sich klinisch unterschiedlich. Die meisten Formen zeigen multisystemische Störungen, die sowohl das zentrale als auch das periphere Nervensystem betreffen. Patienten mit SCA5, SCA6 und SCA8 zeigen in erster Linie zerebelläre Symptome, während Patienten mit SCA1, -2, -3, -4 und -7 häufig extrapyramidale Auffälligkeiten aufweisen [Maschke et al. 2005].

Die häufigsten Erkrankungen SCA1, -2 und -3 weisen vielfach ein fortgeschrittenes Stadium der Ataxie im Kontext mit zusätzlichen Symptomen wie Verminderung geistiger Fähigkeiten, Basalganglienzeichen, Ophthalmoplegia, Verlust des Sehvermögens, Spastik, Muskelatrophie, Empfindungsstörung und den so genannten Restless legs auf [Burk et al. 1996; Burk et al. 1999a; Burk et al. 2003]. Dementsprechend finden sich Degenerationen nicht nur im spinozerebellären System, sondern auch im Kortex, den Basalganglien sowie im Hirnstamm. Nur wenige Mutationen zeichnen sich durch rein zerebrale Syndrome mit isolierten Degenerationen im Kortex aus, wie es zum Beispiel bei SCA6 der Fall ist [Geschwind et al. 1997b].

Die Mutationen SCA 1-3, 6-8, 10, 12 und 17 weisen verlängerte Trinukleotidsequenzen auf, die sich in sechs von neun Fällen im Exonbereich befinden. Interessanterweise zeigen diese sechs Gene (SCA 1-3, SCA6/CACNA1A, SCA7 und SCA17/TBP) eine bereits physiologisch vorhandene polyglutamincodierende CAG Kette, die im Krankheitsfall über ein kritisches Maß hinaus weiter verlängert ist (siehe Kapitel 1.2.3).

Dagegen weisen die Gene SCA8, SCA10 und SCA12 Verlängerungen im nicht codierenden Bereich der Introne auf. Im Falle von SCA8 liegt eine CTG-Kette im 5´UTR-Bereich (5´nichttranslatierte Region), bei SCA10 existiert in einem Intron ein Pentanukleotid (ATTCT), während SCA12 eine CAG-Kette im 3´UTR-Bereich aufweist.
50-80% der erkrankten Personen lassen sich, je nach Region, einer der vorgenannten Untergruppen zuordnen. Weitere genaue Bestimmungen der Erkrankungen werden möglich sein, wenn für SCA 5, 8, 11, 13, 15-16, 18-23 und 25-26 und 28 auch die entsprechenden Gene, an den schon vorbeschriebenen Stellen gefunden werden, beziehungsweise die daraus gebildeten Proteine. Des Weiteren werden wahrscheinlich weitere Mutationen in der immer noch wachsenden Gruppe der SCA (z.B. SCA25, SCA26, SCA 27 und SCA28 wurden erst in den Jahren 2004 – 2006 beschrieben) entdeckt werden. Für SCA14 wurden Punktmutationen in dem Gen PRKCG in einigen Familien entdeckt aber hier steht eine genauere Klärung noch aus [Chen et al. 2003; Chen et al. 2005]. Außerdem konnte vor kurzem in einem Stammbaum SCA20 identifiziert werden, welches sich nahe dem Genort von SCA5 befindet. Hier gilt es jetzt noch zu erkunden, ob es sich bei den beiden phänotypisch deutlich unterschiedlichen Erkrankungen, um zwei verschiedene Mutationen des selben Gens auf Chromosom 11 handelt, oder ob es zwei unterschiedliche Genorte sind, die nur sehr nah beieinander liegen [Knight et al. 2004b; Storey et al. 2005].

1.2.3. Polyglutamin Erkrankungen

Einen besonderen Schwerpunkt in der Ataxieforschung stellen die so genannten Polyglutamin Erkrankungen dar. Diese zeigen expandierte, instabile Trinukleotidwiederholungen, welche sich aus den Basen Cytosin, Adenin und Guanin, kurz CAG, zusammensetzen. Diese CAG-Ketten liegen im Exonbereich des Gens und werden so als eine ungewöhnlich lange Glutaminsequenz übersetzt und in das entsprechende Genprodukt eingebaut. Daher spricht man auch von Polyglutamin (PolyQ) Erkrankungen.
Neben den oben schon erwähnten SCA Erkrankungen finden sich in dieser immer noch wachsenden Gruppe weitere vererbte neurodegenerative Erkrankungen. Momentan sind insgesamt mindestens neun Formen beschrieben worden: Chorea Huntington [The_Huntington´s_Disease_Collaborative_Research_Group 1993], SMBA [La Spada 1991], Dentatorubro pallidoluysiane Atrophie (DRPLA), SCA 1, SCA 2, SCA 3, SCA 6,

SCA 7 und SCA 17. SCA12 zeichnet sich zwar auch durch eine verlängerte CAG-Kette aus, da diese aber im Intronbereich des Gens liegt, wird sie nicht zu Glutamin transkribiert und zählt deshalb nicht zu der Gruppe der PolyQ-Erkrankungen [Holmes et al. 2003].

Die Länge der CAG-Ketten kann für die Ausprägung der Erkrankung sowohl in Bezug auf den Beginn als auch für den weiteren Verlauf eine Rolle spielen. Häufig besteht eine negative Korrelation zwischen der Länge der CAG-Kette und dem Alter der Patienten beim Eintreten erster Symptome. Längere Ketten sind somit prognostisch ein eher ungünstiges Zeichen, da mit einer früher eintretenden Symptomatik, im Vergleich zu Patienten mit kürzeren Ketten zu rechnen ist. Zusätzlich führen die längeren CAG-Ketten vermutlich zu einem schwereren Krankheitsverlauf.

Tabelle 2. Übersicht über die physiologische und pathologische Länge der Polyglutaminketten bei PolyQ-Erkrankungen.

Erkrankung	Normale Länge des CAG Repeats	Übergangsbereich	pathologische Kettenlänge
SCA1	6 – 44	36 - 38	39 – 91
SCA2	≤ 30		≥ (32)33 - > 500
SCA3	≤ 47	48 - 51	53 – 86
SCA6	≤ 18	19	≥ (19) 20 – 33
SCA7	4 – 35	28 - 35	≥ 36 - > 450
SCA17	24 – 44		45 – 63
Chorea Huntington	≤ 26	27 -35	≥ 36
SMBA	9 – 36	37	38 – 66
DRPLA	≤ 35		48 – 93

Die Tatsache, dass eine Tendenz zu einem früheren Krankheitsbeginn und einem schweren Verlauf in nachfolgenden Generationen vererbbarer Krankheiten beobachtet wird, bezeichnet man als Antizipation. Molekularbiologisch lässt sich dies durch eine Instabilität der expandierten CAG-Ketten und der Tendenz einer zunehmenden Expansion der Trinukleotidwiederholungen in der Gametogenese erklären. Dies führt dazu, dass das an die nachfolgende Generation vererbte Allel oft längere Ketten aufweist, als dies in der Elterngeneration der Fall war, und somit kommt es zu einem früheren Beginn und schwereren Verlauf. Die Antizipation ist bei paternaler Vererbung

stärker ausgeprägt als bei der maternalen, was mit einer größeren Instabilität der CAG-Ketten bei der Spermatogenese erklärt wird [Cancel et al. 1997].
Die PolyQ-Kette ruft vermutlich einen toxischen Funktionsgewinn des Proteins hervor, aufgrund von ungewöhnlichen Proteinfaltungen, die durch die Verlängerung entstehen. So wird es wahrscheinlicher, dass das Protein aggregiert und ungewöhnliche Proteinbindungen eingeht, die dazu führen, dass Proteine inaktiviert werden, die für die normale Zellfunktion von Bedeutung sind [Perutz et al. 1994]. Eine häufige Manifestation dieser Fehlbildung und Aggregationen ist die Bildung von intranukleären Einschlüssen (NIs) im krankheitsverursachenden Protein. In vielen PolyQ-Erkrankungen wurden NIs in den neuronalen Zellen gefunden, die mit der Krankheit assoziiert sind, was zu der Vermutung führt, dass die intranukleären Einschlüsse eine zentrale Rolle in der Pathogenese spielen könnten [Paulson et al. 1997; Scherzinger et al. 1997; Igarashi et al. 1998; Li et al. 1998]. Darüber hinaus sind die meisten PolyQ-Proteine in erster Linie zytoplasmatisch, während im Kontrast dazu das pathologische Korrelat der Einschlüsse intranukleär liegt, was zu der Vermutung führt, dass die Kernumgebung die Aggregation stimuliert [Paulson et al. 1997]. Zusätzlich mögen Interaktionen mit Transkriptionsfaktoren und bestimmten Kernprodukten eine transkriptionelle Dysregulation auslösen und der Funktionsverlust des entsprechenden Proteins kann für den zellspezifischen Degenerationsprozess verantwortlich sein [Evert et al. 2003].

1.2.4. Spinozerebelläre Ataxie Typ II (SCA2)

SCA2 wurde von [Orozco et al. 1989] als erstes in einer großen kubanischen Population in der Provinz von Holguin beschrieben. In dieser Provinz ist unter dem Bevölkerungsteil mit spanischen Vorfahren ein klarer Gründereffekt zu erkennen, der zu einer Prävalenz von bis zu 503 : 100.000 [Gispert et al. 1993; Klockgether 2000; Velazquez Perez et al. 2001] in dieser Gegend führen kann, im Vergleich zu den normalerweise zu erwartenden 1-10 : 100.000. Je nach geographischem und ethnischem Ursprung der Bevölkerungsgruppe, schwankt die regionale Häufigkeit für SCA2, innerhalb der autosomal-dominanten Ataxien zwischen 3 % und 57 %, während in Europa mit ca. 15 % zu rechnen ist (siehe Tabelle 3).

Tabelle 3. Die regionale Häufigkeit von SCA2 im Verhältnis zu den sonstigen autosomal-dominanten Ataxien

Land	Prävalenz	Quelle
Brasilien	9 %	[Lopes-Cendes et al. 1997]
China	6,7 - 7,4 %	[Jiang et al. 2005a; Xie et al. 2005]
Deutschland	10 – 14 %	[Riess et al. 1997; Schols et al. 1997b]
Europa	15 %	[Cancel et al. 1997]
Großbritannien	40 %	[Giunti et al. 1998]
Indien	17,5 – 57 %	[Basu et al. 2000; Sinha et al. 2004]
Italien	29 %	[Pareyson et al. 1999]
Japan	4 - 5,9 %	[Watanabe et al. 1998; Sasaki und Tashiro 1999; Matsumura et al. 2003]
Korea	14,5 %	[Kim et al. 2001; Lee et al. 2003]
Portugal	3 – 4 %	[Silveira et al. 1998; Silveira et al. 2002]
Spanien	15,3 %	[Pujana et al. 1999]
USA	13 – 18 %	[Geschwind et al. 1997a; Lorenzetti et al. 1997; Moseley et al. 1998]

Genetische Analysen zeigten 1993 auf dem Chromosom 12 Marker für die Erkrankung an der Position 12q23 – 24.1 [Gispert et al. 1993]. Intensivere Suchen nach verlängerten CAG-Ketten bei Familien mit Veränderungen im Chromosom 12 führten 1996 zur Identifizierung des SCA2 Gens an der Stelle 12q24.1 [Imbert et al. 1996; Nechiporuk et al. 1996; Pulst et al. 1996; Sanpei et al. 1996]. Nach neuester Nomenklatur läuft das SCA2 Gen, welches zeitweise auch mit ATX2 und TNRC13 bezeichnet wurde, jetzt unter ATXN2.

Die krankheitsverursachende CAG-Verlängerung findet sich in der 5´coding Region des Exons 1 [Sahba et al. 1998]. Das aus dem Gen transkribierte Protein Ataxin-2 ist 1312 Aminosäuren lang und hat ein Molekulargewicht von 140,1 kDal, wenn der PolyQ-Abschnitt 22 Glutamine enthält.

Krankheitsbeginn ist in den meisten Fällen die 3. - 4. Lebensdekade bei einem Durchschnittsalter von 29 +/- 11 Jahren SD (Standardabweichung) mit einer Spannweite von 2 – 65 Jahren [Nance 1997; Manto 2005; Maschke et al. 2005]. Klinisch leiden die

Betroffenen an einer variablen Kombination aus Gang- und Standataxie, im fortgeschrittenen Krankheitsverlauf auch Feinmotorikstörungen der Hände, Dysarthrie, verlangsamte Blicksakkaden, Störungen des pyramidalen und extrapyramidalen Systems, Inkontinenz, peripherer Neuropathie, Chorea und Demenz [Nance 1997; Riess et al. 1997; Robitaille et al. 1997]. Diese Symptome werden durch neuronale Degenerationen verursacht, die in erster Linie das Kleinhirn betreffen, aber auch die Nuclei pontis, die inferiore Olive und die Substantia nigra. Das Striatum scheint ausgespart.

1.3. Aufgabenstellung / Zielsetzung

Das Hauptproblem bei der Suche nach Therapiemöglichkeiten der SCA Erkrankungen liegt darin begründet, dass bisher sehr wenig über die Pathogenese bekannt ist. So weiß man zwar häufig, wo der veränderte Genort liegt, allerdings sind bisher in den meisten Fällen die physiologische Funktion und Struktur des dazugehörigen Proteins weitgehend unbekannt. Entsprechend schwer fällt es zu erklären, wie die beobachteten Krankheitssymptome zustande kommen.

Die klassische Herangehensweise wäre die experimentelle Analyse des Proteins, in der unter anderem die dreidimensionale Struktur mittels Röntgenkristallographie, Kernresonanzspektroskopie oder Kryo-Elektronenmikroskopie bestimmt werden könnte. Diese Verfahren sind allerdings sehr aufwendig und setzen voraus, dass das Protein in ausreichenden Mengen gereinigt zur Verfügung steht (siehe Kapitel 2.1).

Bioinformatische Methoden können dabei eine wertvolle Alternative bzw. Ergänzung darstellen und werden in der vorliegenden Arbeit für ATXN2 und dessen Genprodukt Ataxin-2 angewendet, um Thesen über eine potentielle, physiologische Funktion des Proteins zu generieren.

Hierzu wird Ataxin-2 statistisch analysiert, mögliche Homologe identifiziert und ein Modell für eine 3D-Struktur für Teilbereiche entwickelt.

Außerdem werden einige weitere Polyglutaminerkrankungen mit dem Ziel betrachtet, durch Vergleich eventuell vorhandene Ähnlichkeiten auf der bioinformatischen Analyseebene aufzudecken. Dies könnte Hinweise darauf geben, dass zumindest teilweise vergleichbare Mechanismen zum entsprechenden Krankheitsgeschehen führen.

2. Methoden

Grundlage für diese Arbeit sind sowohl das experimentell erworbene Wissen über Gene und deren Proteinprodukte (siehe Kapitel 2.1) als auch die im Internet frei verfügbaren und rapide wachsenden Datenbanken (siehe Kapitel 2.2), sowie zahlreiche Analyseprogramme, welche die im Folgenden beschriebenen Methoden der Proteinanalyse überhaupt erst ermöglichen.

2.1. Experimentelle Proteinstrukturbestimmung

Die „klassische“ experimentelle Proteinstrukturbestimmung kommt, wie eingangs schon erwähnt, in dieser Arbeit nicht zu Anwendung. Trotzdem soll hier ein kurzer Überblick über die Methodiken gegeben werden, die zur Aufklärung der dreidimensionalen Struktur von Proteinen Verwendung finden, da erst durch das experimentell erworbene Wissen die Grundlage für die theoretische Proteinstrukturvorhersage geschaffen wird. Dies lässt sich dadurch erklären, dass experimentell bestimmte Proteinstrukturen als Modell für die homologiebasierte Proteinstrukturvorhersage dienen.
Drei Verfahren, mit denen experimentell die dreidimensionale Struktur eines Proteins bestimmt werden kann, sollen an dieser Stelle kurz vorgestellt werden.

2.1.1. Röntgenkristallographie

Bei der Röntgenkristallographie wird ein Proteinkristall mit Röntgenstrahlen bestrahlt und das resultierende Diffraktionsmuster gemessen, um so die Koordinaten der Atome berechnen zu können. Für diese Methode ist die Kristallisation der Moleküle Voraussetzung, was bei Proteinkristallen sehr schwierig sein kann. Durch Kristallisation in Gegenwart von Substraten (Liganden) kann versucht werden, verschiedene metabolische Zustände des Proteins zu erfassen. Ein weiteres Problem hierbei besteht darin, dass die Proteinstruktur sich durch die Kristallisation möglicherweise ändert. Solche Strukturänderungen können aber problematisch sein, da von der Struktur auf die Funktion des Proteins, sowie auf mögliche Liganden (z.B. Medikamente) geschlossen werden soll und so falsche Ergebnisse erzielt werden können.

2.1.2. Kernresonanzspektroskopie

Das zweite Verfahren zur Aufklärung von Makromolekülstrukturen ist die Kernresonanzspektroskopie (engl. Nuclear magnetic resonance, NMR). Die NMR – Methode basiert darauf, dass bestimmte Atomkerne (zum Beispiel ^{1}H, ^{13}C , ^{15}N und ^{31}P) ein magnetisches Moment oder Spin haben. Dieser Spin richtet sich an einem starken magnetischen Feld aus. Aus dieser Ausrichtung heraus, können die Atomkerne kurzfristig angeregt werden und beim Zurückfallen in den Gleichgewichtszustand wird eine Radiofrequenz emittiert, die abhängig vom jeweiligen Atomtyp ist. Mit der Erweiterung der Messung auf zwei Dimensionen lassen sich die Abstände zwischen den kovalent gebunden Wasserstoffatomen, sowie der im Raum benachbart liegenden Aminosäuren bestimmen und so werden Rückschlüsse auf die räumliche Struktur des Proteins möglich. Vorteil dieser Methode ist, dass die Messung in einer wässrigen Lösung erfolgt, in der sich die Proteine auch physiologischerweise im Organismus befinden. Außerdem entfällt der aufwendige Kristallisationsschritt. Allerdings ist die Methode etwas ungenauer als die Röntgenkristallographie, da die bestimmten Distanzen unterschiedliche Konformationen zulassen. Der Hauptnachteil ist aber die Einschränkung, dass nur Proteine mit einer Aminosäurensequenz von maximal 200 Aminosäuren untersucht werden können. Ein Protein wie Ataxin-2 mit 1312 Aminosäuren wäre also eindeutig zu groß.

2.1.3. Kryo-Elektronenmikroskopie

Als letzte Methode soll noch die Kryo-Elektronenmikroskopie erwähnt werden, deren Funktionsweise vergleichbar mit der Computer-Tomographie ist. Aus vielen zweidimensionalen Projektionsbildern, aufgenommen aus verschiedenen Richtungen, wird über computergestützte Verfahren der Bildverarbeitung die dreidimensionale Struktur des untersuchten Objektes rekonstruiert. Mit dieser, als Einzelpartikelmethode bezeichneten Technik, können Makromoleküle in einem Größenbereich von ca. 200 kDa bis mehrere MDa strukturell untersucht und dreidimensional visualisiert werden. Die zurzeit erreichbare Auflösung dieser Technik liegt in einem Bereich von 6–25 Å, was hauptsächlich abhängig von der Art der Probe (z.B. Symmetrie, Stabilität) und der Probenqualität ist.

2.2. Sequenzdatenbanken

Die drei größten primären Sequenzdatenbanken weltweit sind GenBank (USA) [Benson et al. 2006] betrieben von der NCBI („National Center for Biotechnology Information“), EMBL (European Molecular Biology Laboratory, Europa) [Cochrane et al. 2006] von der EBI („Eurpean Bioinformatics Institute“) und DDBJ (DNA DataBase of Japan) [Okubo et al. 2006] von der NIG („National Institutes of Genetics“). Diese drei Datenbanken sind seit knapp 20 Jahren vereinigt in der INSCD („International Nucleotide Sequence Database Collaboration“), so das weltweit ein großes Archiv für Nucleinsäuresequenzen in Form eines Gemeinschaftsunternehmens mit drei Partnern existiert. In jede der drei Datenbanken können Wissenschaftler neue Sequenzen eintragen und die Daten werden jeden Tag ausgetauscht. Durch den täglichen Austausch wird gewährleistet, dass die Rohdaten identisch sind, im Format der Speicherung und bei der Annotation gibt es jedoch geringfügige Unterschiede.

Das Problem der drei großen Datenbanken ist ihre Redundanz. Größtenteils werden die Sequenzen von den Forschern selbst eingetragen, beschrieben und mit Sequenznamen versehen. Dies führt dazu, dass viele identische Sequenzen unter verschiedenen Namen doppelt eingetragen werden. Deshalb gibt es Datenbanken, die die Einträge der drei Datenbanken übernehmen und diese einem Filterungsprozess und einer Kontrolle der Annotation unterziehen. So entstehen Datenbanken mit nicht redundantem Inhalt. Ein Beispiel für eine nicht-redundante Datenbank ist Swiss-Prot [Apweiler et al. 2004; Bairoch et al. 2005], welche 1986 von dem Schweizerischen Institut für Bioinformatik (SIB) und dem EBI aufgebaut und bis heute betrieben wird.

Neben der hohen Qualität der Annotationen, die per Hand von Biologen geleistet wird, zeichnet sich Swiss-Prot besonders durch seine zahlreichen Querverweise zu anderen Datenbanken aus.

Als Beispiele für das exponentielle Wachstum von Datenbanken mit genetischen Informationen soll die Entwicklung der Datenbanken EMBL und SWISS-PROT dargestellt werden.

Abbildung 3: **Entwicklung der in der EMBL Datenbank gespeicherten Sequenzen und Basenpaaren von 1982 bis 2005**

(in logarithmischer Darstellung)

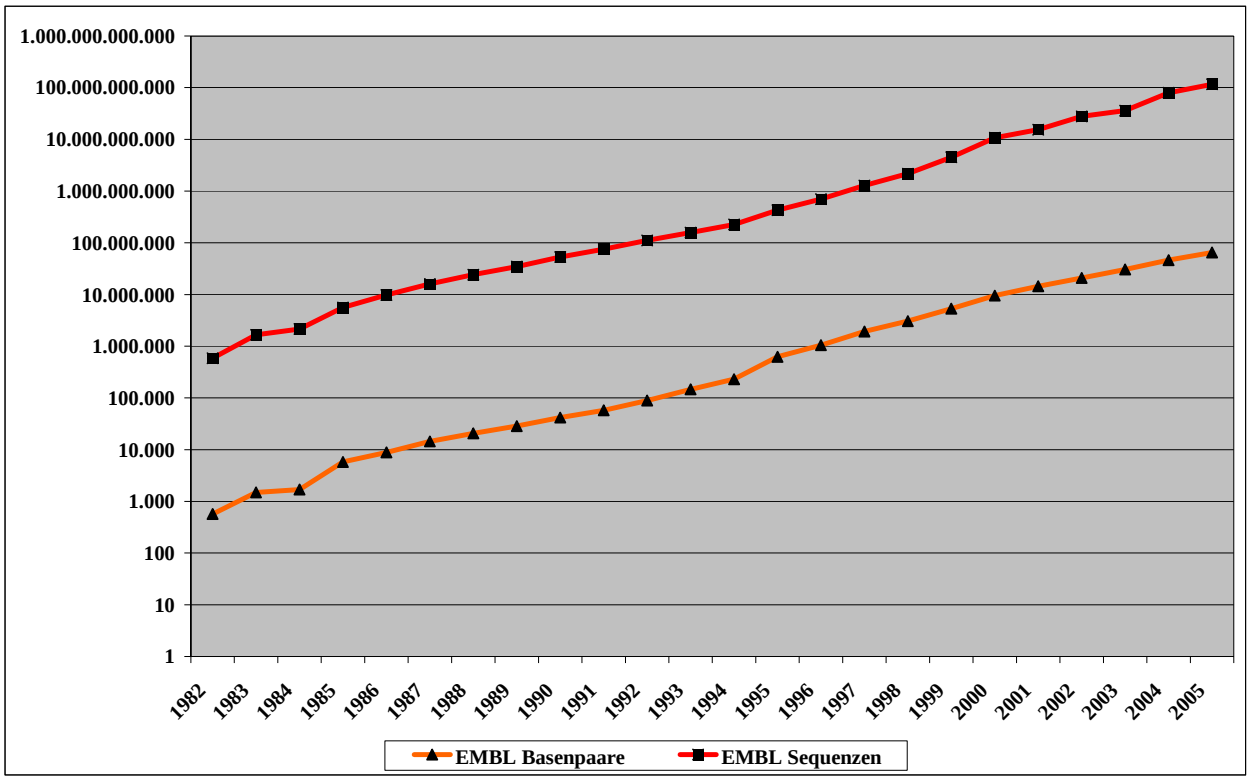
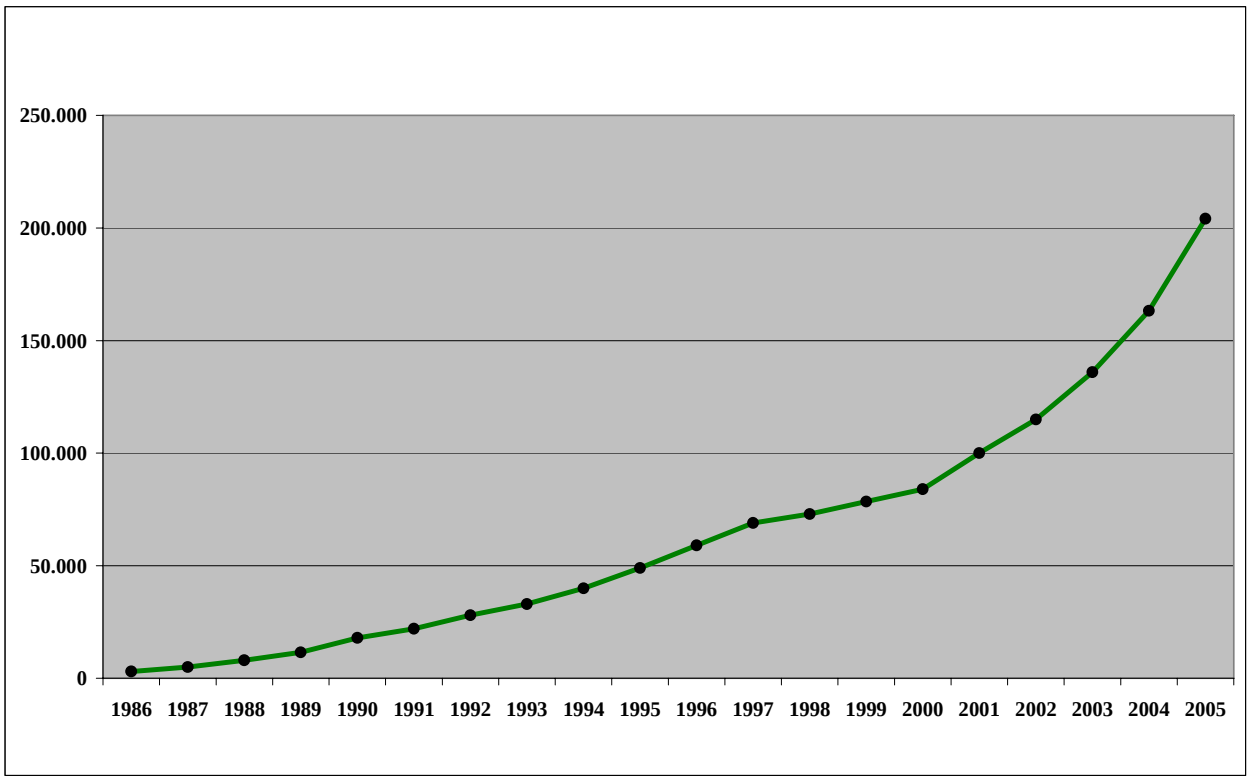


Abbildung 4: Entwicklung der in der SWISS – Prot Datenbank gespeicherten Sequenzen von 1986 bis 2005



Die in der Arbeit verwendeten Proteinsequenzen stammen aus den Datenbanken von NCBI [Wheeler et al. 2005], Ensembl [Birney et al. 2004; Hubbard et al. 2005; Birney et al. 2006] und SWISS-PROT/TrEMBL.

2.3. Gen- / Proteinanalyse

Der erste Ansatz für die Untersuchung kann, je nach Wissensstand über das zu untersuchende Objekt, sehr unterschiedlich sein. Am Beginn steht zum Beispiel der Name eines Gens oder Proteins oder auch die jeweilige Nukleotid- bzw. Aminosäuresequenz entsprechend der möglichen Suchfunktionen der jeweiligen Datenbanken. Folglich sind auch die Möglichkeiten der Informationsgewinnung zahlreich, die über die Sequenz selber, über einen Namen oder eine Codierung erfolgen kann.

Ausgangspunkt der Analyse war in dieser Arbeit sowohl der Name des Gens bzw. des Proteins als auch die Nucleinsäuresequenz bzw. die daraus abgeleitete Aminosäuresequenz. Je nach Datenbank können diese Informationen als Suchkriterium eingegeben werden und von dort aus dann weitere Untersuchungen erfolgen. Meist sind die Datenbanken wie schon für Swiss-Prot beschrieben, untereinander verknüpft, so dass man schnell von der Suche nach dem Protein zum Beispiel auch Informationen über das Gen und den jeweiligen Wissensstand darüber bekommen kann.
Jedes bekannte humane Gen wird in der humanen Gennomenklaturdatenbank „Genew: verzeichnet und mit einem eigenen Gennamen sowie Symbol versehen. Ob das jeweilige zu untersuchende Gen verzeichnet ist, lässt sich auf der HUGO Webseite überprüfen (http://www.gene.ucl.ac.uk/nomenclature/index.html).
Die Sequenzanalyse kann grob in die folgenden relevanten Hauptbereiche untergliedert werden:
1) Identifizierung der Genstruktur, einschließlich Leseraster, Verteilung von Exonen, Intronen, Differenzierung nach verschiedenen Transkriptions- und Translationsvarianten,
2) Vergleich von Gen- oder Proteinsequenzen, um nach ähnlichen Sequenzen zu suchen und Definition von Homologien aus der phylogenetischen Analyse
3) Vorhersage von strukturellen Elementen der Proteine mittels Alignments
4) Bestimmung der 3D-Struktur
5) eine Funktionsbestimmung und dazu passende Ableitungen für mögliche medizinische oder pharmakologische Konsequenzen

Ziel einer ersten Proteinanalyse ist es, in Erfahrung zu bringen, was bisher über das Protein bekannt ist. Zunächst bietet sich hier eine Suche über Swiss-Prot an, da durch die dort vorhandenen zahlreichen Verknüpfungen zu anderen Datenbanken schnell ein Überblick gewonnen werden kann. Zusätzlichen ist die Auswertung von Fachartikeln und von Internetseiten, wie zum Beispiel Genereviews (http://www.geneclinics.org/), hilfreich.
Wenn zunächst nur die Nukleotid- oder die Aminosäuresequenz bekannt ist, kann mittels Übersetzungsprogrammen die Aminosäure- bzw. die Nukleotidsequenz eines Proteins bestimmt werden. Die entsprechenden Programme finden sich unter anderem auf der ExPASy Webseite, auf der diverse Proteomanalyseprogramme gelistet sind (http://www.expasy.org/tools/).

2.3.1. Statistische Analyse/ Physikochemische Eigenschaften von Proteinen

Die statistische Analyse eines Proteins kann mit dem Programm SAPS (Statistical analysis of protein sequences) (http://www.ebi.ac.uk/saps/) [Brendel et al. 1992] erfolgen.

Bei der Auswertung ist zu beachten, dass die Ergebnisse erst ab Proteinen von einer Länge über 200 Aminosäuren als reliabel anzusehen sind.

Die Sequenz kann unter anderem im FASTA Format (fast all Format) eingefügt werden. Zusätzlich wird noch spezifiziert, um was für eine Spezies es sich handelt und ob außer Lysin (K) und Arginin (R) auch noch Histidin (H) als positiv geladen interpretiert werden soll.

Als Ergebnis erhält man eine Datei, in der zunächst die Anzahl der Aminosäuren angegeben wird, sowie das Molekulargewicht des Proteins. Im Folgenden wird der genaue Aufbau des Proteins analysiert, um herauszufinden aus welchen Aminosäuren es zusammengesetzt ist, welche Ladungsverhältnisse vorherrschen und ob es sich eher um ein hydrophobes oder hydrophiles Protein handelt. Das Protein wird weiter dahingehend untersucht, ob es auffällige Aminosäurenblöcke enthält, um nicht punktuelle Ansammlungen, zum Beispiel einer bestimmten Aminosäure, zu übersehen, die vielleicht in der Gesamtbetrachtung wieder nivelliert würden. Außerdem wird nach Wiederholungen einzelner Abschnitte innerhalb des Proteins gesucht.

Ein anderes Programm mit dem die physikochemischen Eigenschaften eines Proteins vorhergesagt werden können ist ProtParam. Dieses Programm steht auf der Seite des ExPASy´s Molekularbiologieservers (Expert Protein Analysis System Molecular Biology Server) zur Verfügung (http://www.expasy.ch/tools/protparam.html) [Gasteiger et al. 2003]. Mittels ProtParam ist es möglich, das Molekulargewicht, den isoelektrischen Punkt (pI), die Aminosäurenzusammensetzung, die atomare Zusammensetzung, einen geschätzten Extinktionswert, die Halbwertszeit, einen Instabilitätsindex und die Hydrophobizität (GRAVY „Grand Average of Hydropathicity“) direkt über die eingegebene Aminosäuresequenz bzw. über jedes in Swiss-Prot/TrEMBL vorhandene Protein bestimmen zu lassen. Der isoelektrische Punkt wird ermittelt gemäß Bjellqvist [Bjellqvist et al. 1993], der geschätzte Extinktionswert nach Pace, Edelhoch und Gill [Edelhoch 1967; Gill und von Hippel 1989; Pace et al. 1995], die Halbwertszeit von Proteinen nach Bachmair [Bachmair et al. 1986], der Instabilitätsindex nach Guruprasad (ein Wert kleiner 40 bezeichnet ein stabiles Protein) [Guruprasad et al. 1990] und die Hydrophobizität nach Kyte und Doolittle. Der GRAVY Index gibt die globale

Hydrophobizität eines Proteins bzw. Peptids an, je größer der Wert, desto hydrophober ist das Protein [Kyte und Doolittle 1982].

2.3.2. Sekundärstruktur

Für das Ziel einer 3D-Strukturvorhersage für ein Protein ist es bisher nicht ausreichend, die Aminosäurensequenz als Grundlage heranzuziehen. Einen wichtigen Zwischenschritt stellt die Vorhersage der Sekundärstruktur dar und vereinfacht die Entwicklung eines 3D-Modells. Die Grundelemente der Sekundärstruktur von Proteinen sind die α-Helixes, die β-Faltblätter, Zufallsstrangkonfirmationen „random coil“ sowie Faltblattstrukturen.
Verwendet wurden die Webserver PSIPRED [McGuffin et al. 2000], SAM-T99 [Karplus et al. 1998], und SSpro2 [Pollastri et al. 2002], um die Sekundärstruktur der Proteine vorherzusagen. Ein Konsensus über die verschiedenen Ergebnisse hinweg wurde durch ein Mehrheitsentscheidverfahren „majority voting“, sowie einer Herausfilterung von zu kurzen Sekundärstrukturelementen erreicht [Albrecht et al. 2003]. Dieses Verfahren erreicht eine höhere Vorhersagequalität als die Anwendung nur einer einzelnen Methode oder der komplexen Kombination von noch mehr Verfahren, wie es zum Beispiel vom Jpred Server (siehe http://www.compbio.dundee.ac.uk/~www-jpred/) praktiziert wird.

2.3.3. Domänen – und Motivdatenbanken

Die Proteindomänenarchitekturen stammen aus der Pfam „Protein family database” (Proteindomändatenbank) [Bateman et al. 2004], Smart „Simple Modular Architecture Research Tool“ [Schultz et al. 1998; Letunic et al. 2006] und SCOP „Structural Classification of Proteins” [Murzin et al. 1995; Lo Conte et al. 2002; Andreeva et al. 2004] Datenbank.

2.3.3.1. Pfam

Pfam ist eine Zusammenstellung von Alignments von mehr als 8000 Proteindomänen-Familien (Stand Dezember 2005) (http://www.sanger.ac.uk/Software/Pfam/). Pfam enthält für jede Familie zwei manuell kontrollierte multiple Sequenzalignments, sowie ein Hidden-Markov-Modell, das die Suche nach diesen Domänen in neuen Sequenzen

ermöglicht. Pfam erlaubt Lücken, so genannte „Gaps“, bei den multiplen Alignments und enthält funktionelle Annotationen, sowie Literaturangaben und Querverweise auf andere Datenbanken. Die Proteinsequenzen basieren auf SWISS-PROT und SPTrEMBL. Die Domänenarchitekturen einer Proteinfamilie können grafisch dargestellt werden.
Zusätzlich zu den oben beschriebenen manuell kontrollierten Proteindomänen der Pfam, welche auch als Pfam A Datenbank bezeichnet werden, gibt es noch eine zweite Datenbank „Pfam B“, die automatisch nach Proteindomänen, die nicht in der Pfam A enthalten sind, in der ProDom Datenbank [Corpet et al. 2000; Servant et al. 2002; Bru et al. 2005] sucht.
75 % aller Proteinsequenzen zeigen wenigstens einen Treffer in Pfam.

2.3.3.2. Smart

Smart basiert auf einer Datenbank von über 650 Domänen-Familien, die in extrazellulären, chromatingebundenen und Signal-Proteinen enthalten sind (Stand Januar 2006) (http://smart.embl-heidelberg.de/). Die Zusammenstellung der Domänen-Familie besteht aus handgefertigten Alignments und den zugehörigen Hidden-Markov-Modellen. Zu den erfassten Domänen erhält der Benutzer Auskunft über die Funktion, wichtige Aminosäuren, die phylogenetische Entwicklung und die Tertiärstruktur.

2.3.3.3. SCOP

SCOP klassifiziert Proteine, deren Struktur bekannt ist (PDB), aufgrund ihrer strukturellen und evolutionären Verwandtschaft. Zurzeit enthält die SCOP Datenbank über 70.000 manuell kontrollierte Domänen (Stand Juli 2005), die sowohl nach Spezies als auch nach drei Verwandtschaftsgraden unterschieden werden.
Der erste Verwandtschaftsgrad ist die „Familie“, bei der mit hoher Sicherheit davon ausgegangen werden kann, dass die Proteine evolutionär verwandt sind. Die Aminosäuresequenzen stimmen paarweise bei mindestens 30 % überein bzw. die ähnliche Funktion und Struktur machen die Verwandtschaft sehr wahrscheinlich.
Der zweite Verwandtschaftsgrad wird als „Superfamilie“ bezeichnet und zeigt an, dass ein gemeinsamer evolutionärer Ursprung wahrscheinlich ist, während der dritte Grad „Faltung“ genannt wird und eine deutliche Strukturähnlichkeit zwischen den Proteinen beschreibt. Dies bedeutet, dass größere Teile der Sekundärstruktur gleichartig

angeordnet sind. In der Peripherie können aber deutliche Unterschiede existieren, so dass ein gemeinsamer evolutionärer Ursprung nicht unbedingt vorliegen muss.

2.4. Homologiesuche

Homologie im engeren Sinne ist die Abstammung von einem gemeinsamen Vorfahren. Eine Homologiesuche im Sinne der Bioinformatik erfolgt in der Regel mit der Absicht, zu einer Aminosäuren- oder Nukleotidsequenz eines Proteins, dessen Funktion teilweise oder vollständig unbekannt ist, Proteine aufzuspüren, die zu der Startsequenz abschnittsweise oder über die Gesamtsequenz Homologien aufzeigen. Informationen über die homologen Sequenzen lassen sich dann mit Einschränkungen auf die Startsequenz übertragen.
Ziel ist eine sowohl möglichst sensitive, das bedeutet auch entfernte Homologien aufdeckende, als auch selektive, das heißt eine idealerweise nur die tatsächlichen Verwandtschaftsbeziehungen anzeigende, Suche. Je nachdem wie die Toleranzschwellen für die Suchverfahren in den Datenbanken vorgegeben werden, wird ein Kompromiss zwischen Sensitivität und Selektivität erreicht. Teilweise wird eine niedrigere Toleranzschwelle bevorzugt, da bei dieser gewährleistet ist, dass keine wichtigen Treffer herausgefiltert werden. Der Nachteil dieser Vorgehensweise ist dann wiederum, dass die Ergebnisse genau zu analysieren sind, um die falschpositiven Treffer, die hierbei vermehrt auftauchen können, auszuschließen.
Mit Methoden wie PSI-BLAST (Position Sensitive Iterated - Basic Local Alignment Search Tool) kann man Verwandtschaftsbeziehungen zwischen Proteinen sowohl innerhalb eines Lebewesens als auch zwischen verschiedenen Arten identifizieren. Wurde die zu einer homologen Sequenz gehörende Struktur bereits experimentell aufgeklärt, kann man zumindest allgemein Rückschlüsse auf das Faltungsmuster der betreffenden Proteinfamilie ziehen [Altschul et al. 1997; Lesk 2003].
Das Programm PSI-BLAST durchsucht eine Datenbank nach Sequenzen, so genannten Homologen, die einer vorgegebenen Sequenz ähneln. Es stellt eine Weiterentwicklung des früheren BLAST Programms dar, welches jeden Datenbankeintrag einzeln im Vergleich zur vorgegebenen Sequenz prüft. PSI-BLAST führt zunächst ebenfalls eine solche Stück-für-Stück-Suche durch, leitet dann aber aus einem multiplen Sequenz-Alignment der ersten Treffer eine Gesetzmäßigkeit ab, anhand derer es die Datenbank

ein zweites Mal durchsucht. Anschließend wiederholt sich der Vorgang, wobei das gefundene Muster in jedem Zyklus weiter verfeinert wird.
Die Suche geschieht mit einem Schwellenwert „E-value cut-off“ von 0,005, was einem guten Kompromiss zwischen der gewünschten Sensitivität und Sensibilität darstellt.

2.5. Vorhersage unstrukturierte Bereiche/ Disordervorhersage

Unter unstrukturierten Bereichen eines kompletten Proteins oder auch nur einer Region eines Proteins, versteht man den Mangel einer definitiven Tertiärstruktur. Unstrukturierte Proteine, die auch als IUP´s („intrinsically unstructured proteins“) bezeichnet werden, und Proteinregionen weisen also keine oder nur eine teilweise vorhandene Faltung auf.
Bis vor einigen Jahren galt, dass eine bestimmte Proteinstruktur die Grundvoraussetzung für die Funktionsfähigkeit eines Proteins darstellt. Dieses Paradigma scheint widerlegt, wenn man die wachsende Zahl von Veröffentlichungen sieht, die sich mit der Funktion unstrukturierter Proteine beschäftigt. Die Erkenntnis, dass zahlreiche Proteine und Proteinregionen erst durch die Tatsache, dass sie ungefaltet vorliegen, ihre Funktion erfüllen können [Dunker et al. 2002; Dyson und Wright 2002; Iakoucheva et al. 2002], führt zu neuen Perspektiven in Hinsicht auf das Verhältnis von Proteinstruktur und Proteinfunktion.
Da die Grundlage zur Bildung einer Faltung in der Aminosäurensequenz begründet liegt, hat man letztere im Vergleich zu der Sekundärstruktur von strukturierten globularen Proteinen sowie Domänen untersucht. Hierbei wurde festgestellt, dass unstrukturierte Proteine und Regionen sich von strukturierten globularen Proteinen sowie Domänen durch divergierende Aminosäurenzusammensetzung, Sequenzkomplexität, Hydrophobizität, Ladung, Flexibilität und der Art und Häufigkeit der im Laufe der Evolution ausgetauschten Aminosäuren unterscheiden. Diese Unterschiede können für die computergestützte Vorhersage von unstrukturierten Bereichen, ausgehend von der Aminosäurensequenz, genutzt werden.
Inzwischen konnte nachgewiesen werden, dass diese unstrukturierten Bereiche unter anderem eine Rolle in der DNA Erkennung, der Proteinbindungsfähigkeit, der molekularen Faltung, der Zellteilung sowie der Kontrolle der Proteinlebenszeit spielen. Auch wenn diese ungeordneten Strukturen in ihrem natürlichen Zustand keine definitive

3-D Struktur aufweisen, wechseln sie doch häufig vom ungeordneten in den geordneten Zustand, wenn sie mit entsprechenden Partnern Bindungen eingehen.
Für die Vorhersage von unstrukturierten und ungeordneten Bereichen innerhalb von Proteinen existierten zum Zeitpunkt der Analyse vier verschiedene Vorhersageserver, die alle im Rahmen dieser Arbeit verwendet wurden. Die entsprechenden Programme heißen DisEMBL 1.3 [Linding et al. 2003a], GlobProt 1.2 [Linding et al. 2003b], NORSp [Liu und Rost 2003] und PONDR [Ichikawa et al. 2001] und arbeiten jeweils mit unterschiedlichen Definitionen. Aus den Ergebnissen der einzelnen Server wurde ein Konsens gebildet, bei dem sich klare Tendenzen zeigen, mit denen sich eine relativ sichere Vorhersage wagen lässt.

2.6. Multiples Alignment

Die multiplen Alignments wurden von dem Alignierungsprogramm T-COFFEE [Poirot et al. 2003] erstellt und danach noch von Hand durch kleine Anpassungen optimiert. Diese Anpassungen erfolgten auf der Grundlage von Strukturvorhersagen und paarweisen Strukturüberlagerungsberechnungen durch das Programm CE [Shindyalov und Bourne 1998]. Die mittlere Summe der Abweichungsquadrate („root-mean-square deviation“ RMSD) wurden aus der CE Superposition entnommen.
Ein Strukturalignment (Vergleich von Proteinstrukturen) dient dazu, Proteinstrukturen zu klassifizieren, um Verwandtschaftsbeziehungen oder die Funktion eines Proteins zu ermitteln. Der gängigste Parameter, der zum Vergleich zweier Proteinstrukturen eingesetzt wird, ist die RMSD, die auf Vergleichen der einzelnen Atome und ihrer Position in den Strukturen beruht. Man kann die RMSD über sämtliche Atome berechnen oder über eine bestimmte Teilmenge, wie z.B. die Atome des Peptidrückgrats oder auch nur die Alpha-Kohlenwasserstoffatome. Bei einem Vergleich von Proteinstrukturen spielt die Orientierung im Raum der Proteinstrukturen zueinander eine wichtige Rolle, sonst wird die RMSD fälschlicherweise zu groß, wenn die Proteinmoleküle zu weit auseinander liegen. Um aussagekräftige RMSD-Werte zu erhalten, wird zunächst ein Sequenzvergleich durchgeführt, um die Eins-zu-eins-Beziehungen zwischen den Paaren analoger Atome zu erkennen, die der RMSD-Berechnung zugrunde gelegt werden. Mit Superpositionsprogrammen kann man eine

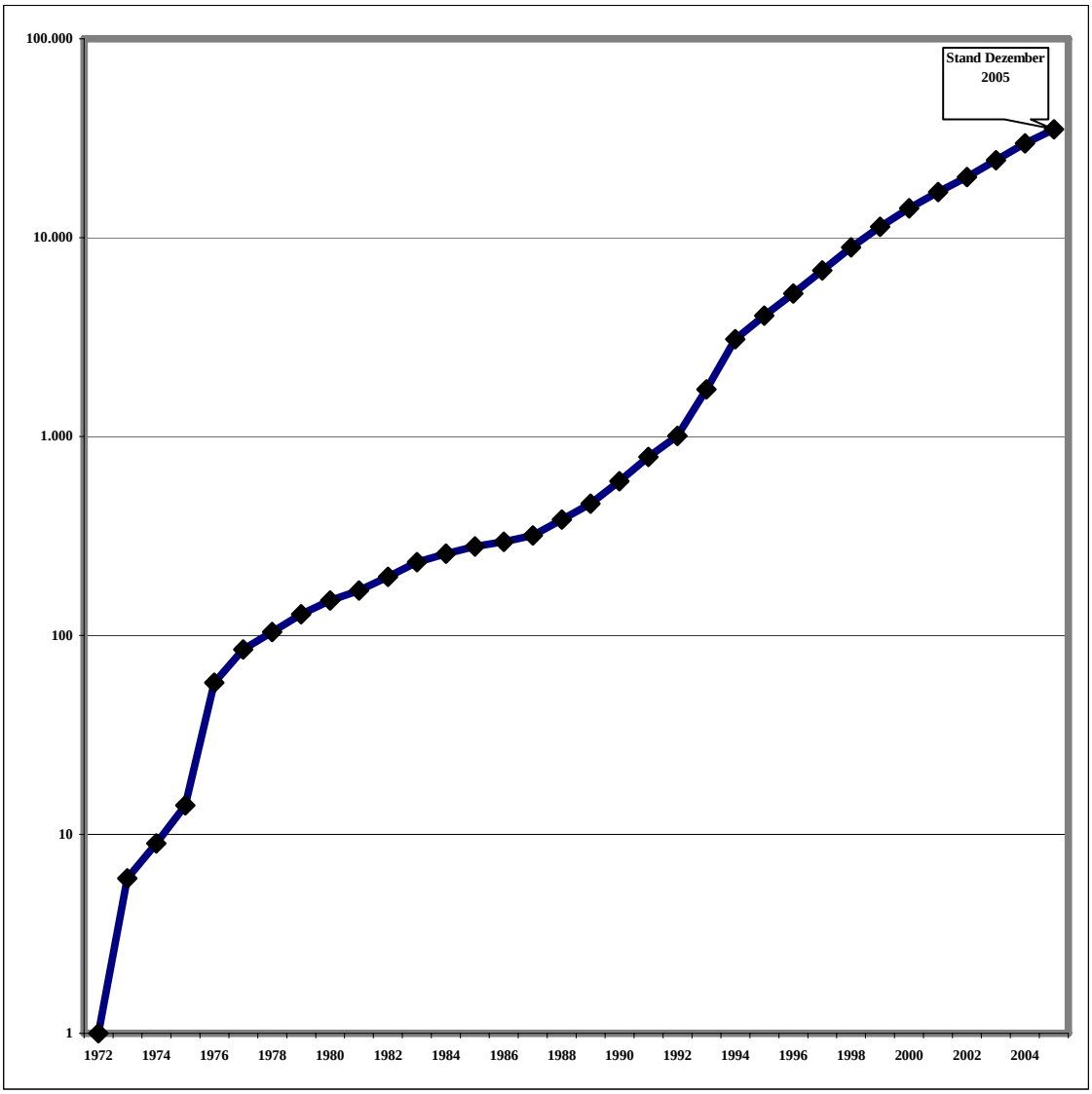
optimale Überlagerung der beiden Strukturen berechnen, wobei ein RMSD Minimum angestrebt wird.
Zur Darstellung können die Sequenzalignments mit dem SEAVIEW Editor [Galtier et al. 1996] bearbeitet werden und zur Veranschaulichung kann das Web basierte Programm ESPript [Gouet et al. 2003] (http://prodes.toulouse.inra.fr/ESPript/) genutzt werden.

2.7. 3D-Modellierung

Experimentell bestimmte 3D-Koordinaten von Proteinmolekülen sind abrufbar im Internet in der Protein Data Bank (PDB) [Berman et al. 2000; Bourne et al. 2004; Kouranov et al. 2006]. Die PDB wird derzeit vom Research Collaboratory for Structural Bioinformatics (RCSB) betrieben und ist die am längsten etablierte Informationsquelle für 3-D Strukturdaten von Proteinen.
Während die oben beschriebenen Sequenzdatenbanken in den letzten 10 Jahren durchschnittlich um ca. 80% pro Jahr gewachsen sind, weist die PDB ein vergleichsweise bescheidenes durchschnittliches Jahreswachstum von ca. 25 % über die letzten 10 Jahre auf (siehe Abbildung 4). Dies lässt sich durch die aufwendigen Methoden erklären, mit denen die dreidimensionalen Strukturen biologischer Makromoleküle bestimmt werden, die in der PDB Datenbank gespeichert werden (siehe auch Kapitel 2.1). Von den zurzeit enthaltenen knapp 35.000 Strukturen (Stand 31.1.2006) wurden ca. 29500 mittels der Röntgenkristallographie, ca. 5300 mittels der Kernresonanzspektroskopie und knapp über 100 mittels der Kryo-Elektronenmikroskopie bestimmt und nur ein kleiner Rest mit anderen Methoden.

Abbildung 5: **Die Entwicklung der in der PDP Datenbank gespeicherten Proteinstrukturen von 1972 - 2005**

(in logarithmischer Darstellung)



Die Proteinstrukturen für die 3D-Modellierung wurden aus der PDB Datenbank entnommen. Der einzelne Buchstabe, der hinter dem PDP Identifikationskürzel angehängt wird, steht für die entsprechende Kette der Struktur.

Das Standardprogramm zur Bestimmung der Sekundärstrukturinformationen aus den in der PDB gespeicherten Proteinstukturen ist DSSP „Dictionary of Secondary Structure of Proteins“ [Kabsch und Sander 1983]. DSSP analysiert die Geometrie und die möglichen Wasserstoffbrücken-Bindungspartner jeder Aminosäure eines Proteins auf der Basis der

in der Protein Datenbank gespeicherten Atomkoordinaten und gibt die entsprechenden Resultate detailliert in tabellarischer Form aus. DSSP darf in der Funktion nicht mit den in Kapitel 2.3.2. beschriebenen Programmen verwechselt werden, da es keine Proteinstruktur vorhersagt.

Eine Alternative zu DSSP wäre das Programm STRIDE, dessen Quellcode auf der EMBL Homepage erhältlich ist. Es kam aber im Rahmen dieser Arbeit nicht zum Einsatz, da DSSP den Standard in der Sekundärstrukturanalyse darstellt.

In die Untersuchung sind alle zurzeit verfügbaren Methoden zur Faltungserkennung die online über den Metaserver BioInfo.PL [Bujnicki et al. 2001] genutzt werden können, eingeflossen. Der Metaserver kontaktiert ein gutes Dutzend Vorhersageserver (die Namen der einzelnen Server finden sich auf der Webseite http://Bioinfo.PL/Meta/).

Das assoziierte 3D-Jury System erlaubt einen Vergleich und eine Beurteilung der vorhergesagten 3D-Modelle in einer Konsensusübersicht [Ginalski und Rychlewski 2003]. Um die Proteinstruktur von Ataxin-2 zu modellieren, wurde die erstellten Sequenzstrukturalignments an den WHAT IF 3D „modeling“ Server [Rodriguez et al. 1998] geschickt. Das WHAT IF Programm berechnet eine Reihe von stereochemischen Parametern, mit deren Hilfe die Qualität von Proteinstrukturen beurteilt werden kann.

Die Abbildungen der Proteinstrukturen wurden mit dem Accelrys Discovery Studio ViewerLight erstellt.

3. Ergebnisse

Die nachfolgenden Ergebnisse spiegeln den Stand bis einschließlich Februar 2006 wider. Dies zu erwähnen ist deshalb wichtig, da aufgrund der oben schon beschriebenen exponentiell wachsenden Datenmengen und den ständigen Bemühungen die Datenbanken zu optimieren, es durchaus sein kann, dass das eine oder andere Resultat nicht mehr replizierbar ist. Alleine während der drei Jahre in denen ich Ataxin-2 untersucht habe, musste ich häufiger, bereits durchgeführte Analysen wiederholen, um auch wirklich den aktuellsten Wissensstand zu nutzen. Dies gilt sowohl für die Datenbanken selbst, die inzwischen alleine aufgrund ihres Wachstums andere bzw. mehr Ergebnisse ausgeben, aber auch für die Methoden, die verfeinert oder gar neu entwickelt wurden. Auch die Nomenklatur wurde teils geändert, wie zum Beispiel das SCA2 Gen ja inzwischen mit dem Kürzel ATXN2 bezeichnet wird.

3.1. Gen-/ Proteinanalyse

Das Gen ATXN2 aus dem Ataxin-2 transkribiert wird, befindet sich im menschlichen Genom an der Stelle 12q24.1 und es besteht aus 25 Exonen [Sahba et al. 1998] (siehe Anhang).

3.1.1. Statistische Betrachtung/ Physikochemische Eigenschaften von Ataxin-2

Ataxin-2 besteht aus 1312 Aminosäuren und hat ein Molekulargewicht von 140 kDal. Dabei ist die physiologischerweise vorhandene Kette von 22 (oder 23) Glutaminen (Q) (siehe Abbildung 6), die sich im Krankheitsfall auf bis zu 200 verlängern kann, mit eingeschlossen.

Abbildung 6: Ataxin-2 und die Häufigkeitsverteilung der einzelnen Aminosäuren

```
   1  MRSAAAAPRS PAVATESRRF AAARWPGWRS LQRPARRSGR GGGGAAPGPY PSAAPPPPGP
  61  GPPPSRQSSP PSASDCFGSN GNGGGAFRPG SRRLLGLGGP PRPFVVVLLP LASPGAPPAA
 121  PTRASPLGAR ASPPRSGVSL ARPAPGCPRP ACEPVYGPLT MSLKPQQQQQ QQQQQQQQQQ
 181  QQQQQQQPPP AAANVRKPGG SGLLASPAAA PSPSSSSVSS SSATAPSSVV AATSGGGRPG
 241  LGRGRNSNKG LPQSTISFDG IYANMRMVHI LTSVVGSKCE VQVKNGGIYE GVFKTYSPKC
 301  DLVLDAAHEK STESSSGPKR EEIMESILFK CSDFVVVQFK DMDSSYAKRD AFTDSAISAK
 361  VNGEHKEKDL EPWDAGELTA NEELEALEND VSNGWDPNDM FRYNEENYGV VSTYDSSLSS
 421  YTVPLERDNS EEFLKREARA NQLAEEIESS AQYKARVALE NDDRSEEEKY TAVQRNSSER
 481  EGHSINTREN KYIPPGQRNR EVISWGSGRQ NSPRMGQPGS GSMPSRSTSH TSDFNPNSGS
 541  DQRVVNGGVP WPSPCPSPSS RPPSRYQSGP NSLPPRAATP TRPPSRPPSR PSRPPSHPSA
 601  HGSPAPVSTM PKRMSSEGPP RMSPKAQRHP RNHRVSAGRG SISSGLEFVS HNPPSEAATP
 661  PVARTSPSGG TWSSVVSGVP RLSPKTHRPR SPRQNSIGNT PSGPVLASPQ AGIIPTEAVA
 721  MPIPAASPTP ASPASNRAVT PSSEAKDSRL QDQRQNSPAG NKENIKPNET SPSFSKAENK
 781  GISPVVSEHR KQIDDLKKFK NDFRLQPSST SESMDQLLNK NREGEKSRDL IKDKIEPSAK
 841  DSFIENSSSN CTSGSSKPNS PSISPSILSN TEHKRGPEVT SQGVQTSSPA CKQEKDDKEE
 901  KKDAAEQVRK STLNPNAKEF NPRSFSQPKP STTPTSPRPQ AQPSPSMVGH QQPTPVYTQP
 961  VCFAPNMMYP VPVSPGVQPL YPIPMTPMPV NQAKTYRAVP NMPQQRQDQH HQSAMMHPAS
1021  AAGPPIAATP PAYSTQYVAY SPQQFPNQPL VQHVPHYQSQ HPHVYSPVIQ GNARMMAPPT
1081  HAQPGLVSSS ATQYGAHEQT HAMYACPKLP YNKETSPSFY FAISTGSLAQ QYAHPNATLH
1141  PHTPHPQPSA TPTGQQQSQH GGSHPAPSPV QHHQHQAAQA LHLASPQQQS AIYHAGLAPT
1201  PPSMTPASNT QSPQNSFPAA QQTVFTIHPS HVQPAYTNPP HMAHVPQAHV QSGMVPSHPT
1261  AHAPMMLMTT QPPGGPQAAL AQSALQPIPV STTAHFPYMT HPSVQAHHQQ QL
```

Hieraus ergibt sich mittels der automatisierten Auswertung durch das SAPS oder ProtParam Programm, folgende Verteilung der Aminosäuren:

```
Ala (A): 126 ( 9.6%); Arg (R):  74 ( 5.6%); Asn (N):  56 ( 4.3%);
Asp (D):  33 ( 2.5%); Cys (C):  11 ( 0.8%); Gln+(Q): 101 ( 7.7%);
Glu (E):  57 ( 4.3%); Gly (G):  85 ( 6.5%); His (H):  46 ( 3.5%);
Ile (I):  31 ( 2.4%); Leu-(L):  53 ( 4.0%); Lys (K):  49 ( 3.7%);
Met (M):  32 ( 2.4%); Phe (F):  27 ( 2.1%); Pro+(P): 184 (14.0%);
Ser+(S): 176 (13.4%); Thr (T):  65 ( 5.0%); Trp (W):   7 ( 0.5%);
Tyr (Y):  31 ( 2.4%); Val (V):  68 ( 5.2%);
```

Hervorgehoben werden die Aminosäuren, die durch die Häufigkeit ihres Auftretens, statistisch gesehen, auffällig sind. Als Vergleich dient die Auftretenswahrscheinlichkeit der einzelnen Aminosäuren in der jeweiligen Spezies, hier also dem Menschen. Die Pluszeichen hinter Glutamin, Prolin und Serin bedeuten, dass diese drei Aminosäuren in

der Häufigkeit ihres Auftretens über dem 95% Quantil liegen, während das Minuszeichen hinter Leucin anzeigt, dass sich die Häufigkeit unter dem 5% Quantil bewegt.

Mittels ProtParam wurde der theoretische isoelektrische Punkt für Ataxin-2 mit 9,6 bestimmt. Desweiteren wurden 90 negativ geladene Aminosäuren (Asparaginsäure (Asp) 33 und Glutaminsäure (Glu) 57) sowie 123 positiv geladene Aminosäuren (Arginin (Arg) 74 und Lysin 49) gezählt. Hierbei wurden die 46 Histidinaminosäuren in der Gruppe der basischen Aminosäuren nicht berücksichtigt, da der pK - Wert von Histidin nahe dem physiologischen pH - Wert liegt und Histidin somit sowohl basische als auch neutrale Zustände annehmen kann.

Die atomare Zusammensetzung von Ataxin-2 sieht wie folgt aus:

Kohlenstoff	(C)	6052
Wasserstoff	(H)	9487
Stickstoff	(N)	1839
Sauerstoff	(O)	1922
Schwefel	(S)	43

mit der Summenformel: $C_{6052}H_{9487}N_{1839}O_{1922}S_{43}$

und einer Gesamtzahl von 19343 Atomen.

Der geschätzte molare Extinktionskoeffizient für Ataxin-2 beträgt $\varepsilon = 85315\ M^{-1}cm^{-1}$ bei 280 nm. Die geschätzte Halbwertzeit, in der die Hälfte des in der Zelle synthetisierten Ataxin-2 wieder abgebaut wird, beträgt 30 Stunden.

Ein Instabilitätsindex der von ProtParam kalkuliert wird, prognostiziert mit einem Wert von 78,29, dass Ataxin-2 im Reagenzglas instabil ist (Werte ab 40 werden als instabil interpretiert).

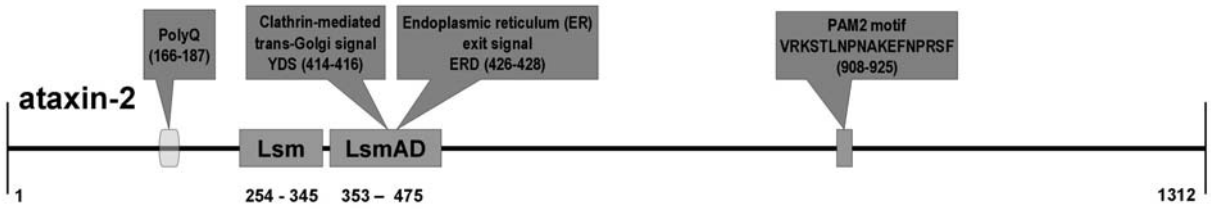
Zuletzt wird noch ein negativer GRAVY Index von -0,823 berechnet, was dafür spricht, dass Ataxin-2 hydrophil ist.

3.1.2. Proteinarchitektur von Ataxin-2

Ataxin-2 ist ein basisches Protein, abgesehen von einer aziden Region von Position 254 bis 475, in der sich 46 saure Aminosäuren befinden (siehe Abbildung 7). Diese Region besteht grob gesagt aus den Abschnitten der Exonen 2 bis 7 und laut Vorhersage befinden sich hier zwei globuläre Domänen namens Lsm (Like Sm, aa 254-345) [Neuwald und Koonin 1998] und LsmAD (Lsm-assozierte Domäne, aa 353-475).

Die LsmAD von Ataxin-2 enthält sowohl ein clathrin-mediated trans-Golgi Signal (YDS, aa 414-416) und ein endoplasmatisches Retikulum (ER) Ausgangssignal (ERD, aa 426-428) [Shibata et al. 2000; Huynh et al. 2003]. Laut der Sekundärstrukturvorhersage mehrerer Server besteht die LsmAD in erster Linie aus α-Helixes.

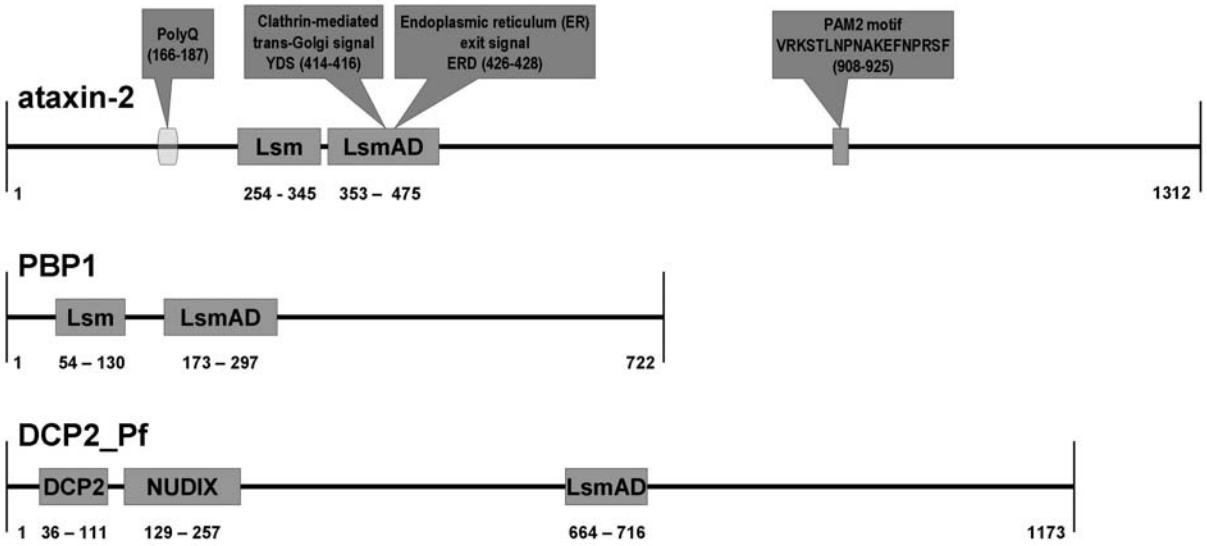
Abbildung 7: Proteinarchitektur des humanen Ataxin-2.



3.2. Homologe zu Ataxin-2

Die Lsm Domäne von Ataxin-2 ist typisch für die RNA bindenden Proteine Sm und Sm-like, welche häufig zyklische 6-, 7- oder sogar 14- Oligomere bilden [He und Parker 2000; Mura et al. 2003; Sauter et al. 2003]. Im Allgemeinen sind Proteine mit einer Lsm Domäne in verschiedene Prozesse der RNA Verarbeitung involviert. Darunter befinden sich Funktionen wie RNA Modifikationen, prä-mRNA Spleißung und Entkappung und Degeneration der mRNA. Einige von diesen sind wichtige Komponenten des Ribonucleinproteinkomplexes (snRNPs „small nuclear ribonucleoprotein particles“).
Die LsmAD Domäne firmiert in der Pfam Datenbank unter dem Namen Ataxin-2_N und taucht auch noch in anderen bisher uncharakterisierten Plasmodium falciparum / yoelii yoelii Genprodukten PF13_ 0048 / PY07327 auf. Allerdings ist in diesen Fällen keine zusätzliche Lsm Domäne vorhanden (siehe Abbildung 8). Beide Plasmodium Genprodukte haben eine zusätzliche N-terminale DCP2 Domäne (auch BoxA genannt), welche in allen bekannten DCP2 Homologen immer von einer NUDIX Domäne gefolgt wird. Diese NUDIX Domäne bildet eine katalytische Untereinheit des mRNA Entkappungsholoenzyms DCP1-DCP2 [Piccirillo et al. 2003; She et al. 2004].

Abbildung 8: **Proteinarchitektur des humanen Ataxin-2 im Vergleich zu seinem Hefe Homolog PBP1 und dem Plasmodium falciparum Homolog PF13_0048 des Entkappungsenzyms „Decapping Enzyms“ DCP2 (DCP2_Pf).**



Die physiologische Funktion von Ataxin-2 und eng verwandten eukaryontischen Homologen in der RNA Verarbeitung ist bist jetzt mehr oder weniger unerforscht [Kiehl et al. 2000; Meunier et al. 2002; Satterfield et al. 2002; Figueroa und Pulst 2003]. Interessanterweise wurde beobachtet, dass Ataxin-2 mit dem A2BP1 (Ataxin-2 Bindungsprotein 1) [Shibata et al. 2000] interagiert, dessen Caenorhabditis elegans Homolog Fox-1 (Feminisierungsort auf dem X-Chromosom) RNA bindet und ein gewebsspezifisches, alternatives Spleißen ermöglicht [Jin et al. 2003]. Außerdem zeigt Ataxin-2 eine auffallende Homologie zu dem Hefeprotein PBP1 (PAP1/PABP-Bindungsprotein 1), welches ebenfalls eine Lsm und LsmAD Domäne enthält. Bereiche außerhalb dieser beiden globulären Domänen werden in PBP1 ebenso wie bei Ataxin-2 als hauptsächlich unstrukturiert vorhergesagt.

Auch wenn der C-terminale Schwanz von PBP1 kein PAM2 Motive enthält [Albrecht und Lengauer 2004], reguliert dieses Hefeprotein die Polyadenylation nach der prä-mRNA Verarbeitung und interagiert mit der PABC Domäne im Hefe Ortholog PAB1 des humanen Poly(A) Bindungsprotein PABP [Mangus et al. 1998]. Sowohl A2BP1 als auch

PABP sind evolutionär miteinander verwandt und besitzen RNA Erkennungsmotive (RRMs „RNA recognition motifs) [Shibata et al. 2000].
Diese Beobachtungen deuten stark darauf hin, das Ataxin-2 in ähnliche mRNA Verarbeitungsprozesse involviert ist.

3.3. Vorhersage unstrukturierte Bereiche/ Disordervorhersage

3.3.1. Strukturierte und unstrukturierte Bereiche in Ataxin-2

Die Bereiche von Ataxin-2, die außerhalb der Lsm und LsmAD Domänen liegen, sind in den eukaryonten Homologen von Ataxin-2 nur schwach konserviert. Laut der übereinstimmenden Vorhersage der Server DisEMBL, GlobProt, NORSp und PONDR handelt es sich im Wesentlichen um unstrukturierte Bereiche. Diese nicht-globulären, flexiblen N- und C- terminalen Bereiche (aa 1 - 253 und 476 - 1312) enthalten die PolyQ-Region (aa 166 - 187), einige hoch konservierte Kurzsequenzmotive, hinter denen sich möglicherweise einige Proteininteraktionsstellen verbergen und einige auffällige (R)RG Peptide am C-terminus der LsmAD Domäne, welche bekanntermaßen RNA in anderen Proteinen binden [Dreyfuss et al. 1993]. Eines dieser Sequenzmotive bildet ein mögliches PAPB (poly(A)-Bindungsprotein), welches mit dem PAM2 Motiv (aa 908 - 925) [Albrecht und Lengauer 2004] interagiert. Der N- und C-terminale Schwanz von Ataxin-2 zeigt außerdem noch einen ungewöhnlich hohen Gehalt an Prolin (179 Proline auf 1090 Aminosäuren, 16,4%).

3.3.2. Fragliche Treffer bei der Homologiesuche

Der hohe Prolingehalt und die niedrige Komplexität der unstrukturierten Sequenzregionen mag zu einigen signifikanten, aber möglicherweise falsch-positiven Treffern während der PSI-BLAST Suche nach Homologen von Ataxin-2 führen. Zum Beispiel finden sich trotz der Nutzung des Standardfilters gegen niedrige Komplexität in den PSI-BLAST Suchen für humanes Ataxin-2 zahlreiche fragliche Treffer zu Homologen des DRPLA (dentatorubrale pallidoluysiane Atrophie) Genprodukts Atrophin. Dies ist besonders bemerkenswert, da es sich ja bei DRPLA ebenfalls um eine Krankheit der PolyQ-Gruppe handelt, obwohl diese Treffer höchstwahrscheinlich über die niedrige Komplexität zu erklären sind, könnte dies ein Hinweis auf Wechselwirkung

verschiedener PolyQ-Erkrankungen sein, zumal diese Treffer mehrfach auftauchten, hier einige Ergebnisse: bei einer PSI-BLAST Suche mit einem Ataxin-2 Orthologen der Arabidopsis thaliana (SPTrEMBL: Q94AM9) wurde Atrophin in der dritten Iteration mit einem E-value von $5 \cdot 10^{-11}$ angezeigt. Umgekehrt zeigte sich bei der Suche mit dem Ortholog von Atrophin in der Ratte (SPTrEMBL: Q62901) als Startsequenz, unter anderem humanes Ataxin-2 in der zweiten Iteration mit einem E-value von $8 \cdot 10^{-04}$.

3.3.3. Vergleich mit anderen PolyQ-Proteinen

Im Verlauf der Arbeit zeigte es sich, wie oben bereits beschrieben, dass der PolyQ-Abschnitt von Ataxin-2 sich im Vergleich zu den globulären Domänen in einem Bereich mit einem sehr niedrigen Konservierungsgrad befindet. Ebenfalls zeigten Vorhersagen, dass diese Bereiche unstrukturiert sind. Ähnliche Vorhersagen wurde auch für den PolyQ-Bereich in Ataxin-3 getroffen und experimentell bestätigt [Masino et al. 2003]. Außerdem ist bekannt, dass der PolyQ-Trakt selber eine Zufallsstrang Konfirmation „random coil“ annimmt [Masino et al. 2002]. Dies könnte auch einer der Gründe sein, warum eine Verlängerung der PolyQ-Kette zu einer Funktionseinschränkung des Proteins führt.

Um herauszufinden, ob auch bei anderen Polyglutaminerkrankungen die PolyQ-Region von ungeordneten Strukturen umgeben ist, wurden die Proteine Ataxin-1 (SCA1), Ataxin-7 (SCA7), Atrophin (DRPLA) und Huntingtin (Morbus Huntington) ebenfalls untersucht. Zu diesem Zweck wurden die Proteinsequenzen an die Vorhersageserver PONDR, NORSp, GlobPlot und DisEMBL geschickt, mit dem Resultat, dass die Ergebnisse in der Gesamtbetrachtung darauf hindeuten, dass der PolyQ-Trakt generell in einer unstrukturierten Region liegt (siehe Tabelle 4). Dies entspricht auch den Sekundärstrukturvorhersagen, welche für diesen Bereich keine globulären Faltungen aus α-Helixes oder β-Faltblätter prognostizieren. Diese Ergebnisse sprechen dafür, dass die verlängerten Polyglutaminproteine Aggregate mit der exponierten PolyQ-Region bilden können.

Tabelle 4. Unstrukturierte Abschnitte/ „Disorder“ im Bereich der PolyQ-Region

Vorhersage unstrukturierter Abschnitte in der Nachbarschaft der PolyQ-Region für die Proteine Ataxin-1,-2,-3,-7, Atrophin und Huntingtin durch die Server DisEMBL, GlobProt, NORSp und PONDR. Eintragungen, die dunkelgrau hervorgehoben sind, zeigen eine Überlappung zwischen dem Disorderbereich und der PolyQ-Region an, während die hellgrau markierten Bereiche anzeigen, dass der Beginn oder das Ende eines Disorderabschnitts nahe der PolyQ-Region liegt (maximal 20 Aminosäuren entfernt)

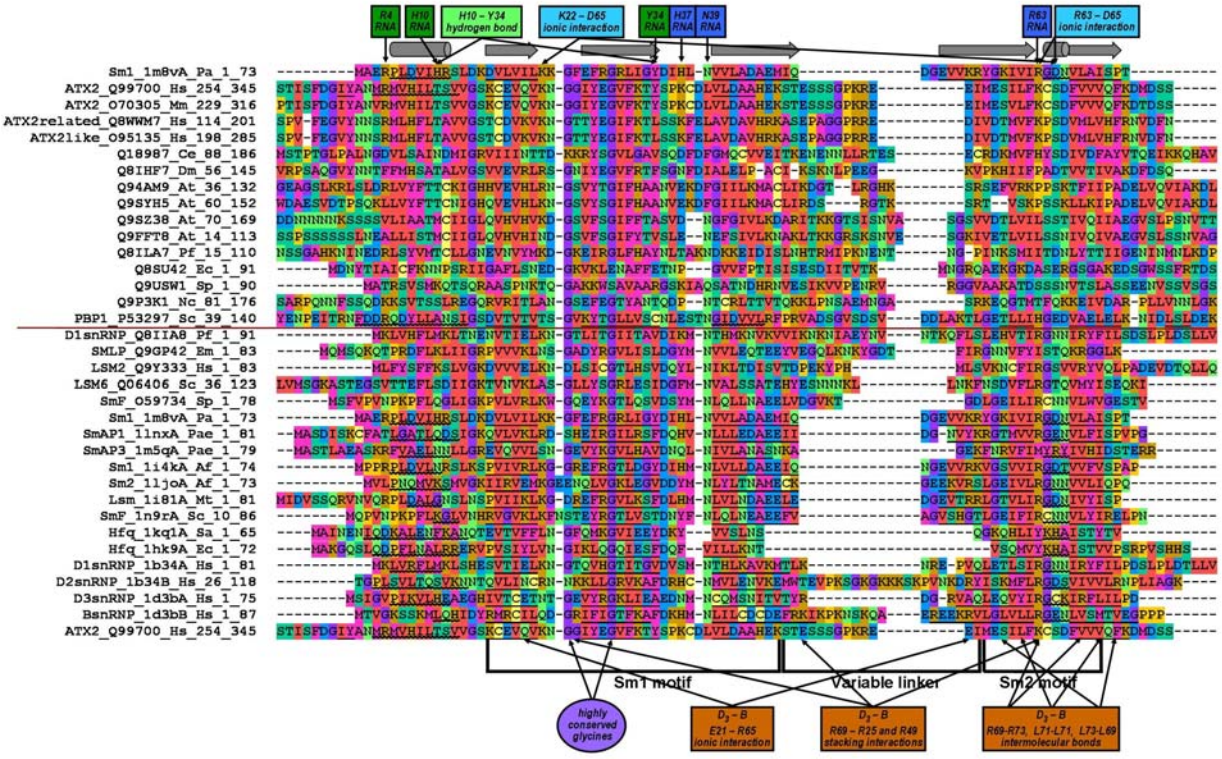
						DIsEMBL		GlobProt	NORSp	PONDR		
Gen	Protein	Protein-länge	polyQ Position	polyQ Länge	Lokalisation im Bezug auf polyQ	Disordered nach COILS Definition	Disordered nach REMARK465 Definition	putativ unstrukturierte Bereiche (disorder)	NORS Region Vorhersage	VLXT	XL1_XT	CAN_XT
			Start - Ende			Start - Ende	Start - Ende	Start - Ende	Start - Ende	Start - Ende	Start - Ende	Start - Ende
					vor polyQ	82 – 172	154 - 171	**185 - 192**		80 - 118	**179 - 197**	**178 – 187**
SCA 1	**Ataxin-1**	**816**	**197 - 226**	**29**	**Überschneidung**	**184 - 195**	**188 - 229**			**140 - 294**	**207 - 255**	**220 – 229**
					nach polyQ	**228 - 289**	252 - 262	**231 - 245**		302 - 351	289 - 348	**241 – 247**
					vor polyQ	**109 - 166**	114 - 145	**112 - 160**		20 - 76	108 - 131	83 – 132
SCA 2	**Ataxin-2**	**1312**	**166 -187**	**21**	**Überschneidung**	**184 - 259**	**163 - 197**		**184 – 256**	**106 - 248**	**156 - 216**	**185 – 229**
					nach polyQ	275 - 302	**203 - 232**	**188 - 223**	469 – 783	308 - 326	223 - 245	300 – 317
					vor polyQ	**258 - 278**	221 - 229	63 - 72		**187 - 294**	147 - 149	223 – 268
SCA 3	**Ataxin-3**	**376**	**292 - 317**	**25**	**Überschneidung**	**317 - 350**	**288 - 336**			**299 - 337**	**183 - 323**	**315 – 325**
					nach polyQ			**320 - 344**				
					vor polyQ	**1 – 12**					**1 - 12**	**13 – 17**
SCA 7	**Ataxin-7**	**892**	**30 - 39**	**9**	**Überschneidung**	**35 – 70**	**1 - 74**	**37 - 69**		**1 - 87**	**17 - 59**	
					nach polyQ	78 – 95	163 - 178	81 - 89	166 – 250	116 - 125	69 - 80	**57 – 61**
					vor polyQ	**60 – 484**	374 - 402	**434 - 479**		1 - 113	**323 - 464**	450 – 463
DRPLA	**Atrophin**	**1185**	**484 - 497**	**13**	**Überschneidung**	**497 - 607**	**464 - 504**	**497 - 604**	**51 – 771**	**126 - 753**	**480 - 507**	**491 – 505**
					nach polyQ	617 - 758	544 - 591	620 - 717		787 - 889	519 - 529	671 – 711
					vor polyQ					**1 - 9**		**1 – 9**
Huntington	**Huntingtin**	**3144**	**18 - 40**	**22**	**Überschneidung**	**37 – 101**	**14 - 56**	**37 - 84**	**5 – 114**	**23 - 99**	**1 - 87**	
					nach polyQ	198 – 225	219 - 229	427 - 436	498 – 580	149 - 150	110 - 117	79 – 88

3.4. Multiples Alignment für Ataxin-2 und ausgewählter Homologe

In der folgenden Abbildung ist ein strukturbasiertes multiples Sequenzalignment der Lsm Domäne von Ataxin-2 und Homologen zu sehen [Albrecht et al. 2004]. Unter den Homologen befindet sich das Hefehomolog PBP1 (in der oberen Hälfte) und Sm/ Sm-like Proteinen (untere Hälfte). Die von DSSP für das Sm1 Protein von P. abyssi (Pyrococcus abyssi) zugewiesene Sekundärstruktur wird am oberen Rand des

Alignments dargestellt (Zylinder stehen für die α-Helix und Pfeile für die β-Faltblattstruktur). Ebenso wird die Aminosäuresequenz, soweit bekannt, entsprechend den in PDB festgehaltenen crystallographisch festgestellten Sekundärstrukturen unterstrichen (eine gewellte Linie symbolisiert die α-Helix, während eine einfache Unterstreichung für die β-Faltblattstruktur steht). Die korrespondierenden Sekundärstrukturvorhersagen für die Lsm-Domänen von Ataxin-2 und PBP1 sind ebenfalls angegeben. Physikochemisch ähnliche Aminosäuren sind identisch eingefärbt worden. Die hochkonservierten Glycine, die charakteristisch für die Lsm Domäne sind, wurden hervorgehoben. Im oberen Abschnitt verweisen die blauen Textboxen auf die funktionell relevanten Aminosäuren, die eine interne Bindungsstelle für ein Uridinheptamer bilden, welches an das Sm1 Protein von P. abyssi gebunden ist, während die grauen Textboxen auf Aminosäuren weisen, die eine externe RNA-Bindungsstelle bilden. Im unteren Teil verweisen die brauen Textlabels auf intermolekulare Interaktionen, die die Dimerisation der snRNPs D_3 und B stabilisieren. Die PDB-Kennungen und die korrespondierenden SPTrEMBL-Nummern für die Lsm-Proteine werden im Anhang in Tabelle 5 wiedergegeben.

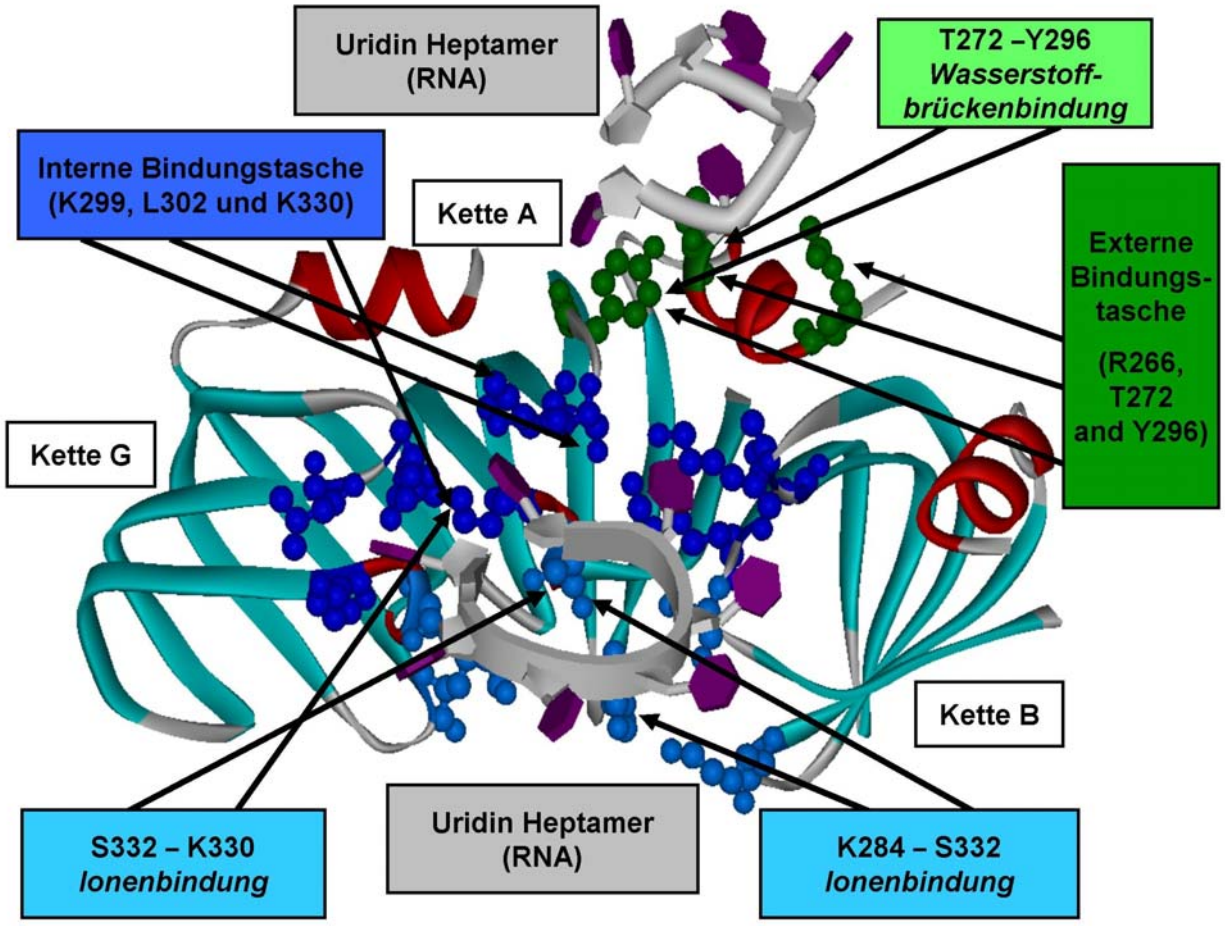
Abbildung 9: **Multiples Sequenzalignment der Lsm Domäne von Ataxin-2 und ausgewählter Homologe**



3.5. 3D-Modell der Lsm-Domäne für Ataxin-2

Um das 3D-Modell zu bilden, wurden als Templat / Vorlage die drei benachbarten Untereinheiten des Sm1 Proteins aus P. abyssi (PDB Kennung 1m8v, Kette A, B und G) genutzt.

Abbildung 10: 3D-Modell der Lsm-Domäne für Ataxin-2



Das Modell illustriert die vorhergesagten internen (blau) und externen (grün) Bindungstaschen in Ataxin-2 für RNA (grau). Die α-Helixes sind rot und die β-Faltblätter in Cyano gefärbt. Von den Seitenketten von Ataxin-2 sind hier nur die funktionell relevanten Reste in der zentralen Untereinheit markiert worden: Die dunkelblauen Kästchen zeigen auf die Reste, die die interne Bindungstasche bilden, während die hellblauen Kästchen auf die Aminosäuren verweisen, die die RNA Bindungsstelle

stabilisieren; die dunkelgrünen Kästchen markieren die Seitenketten, die Teil der externen RNA-Bindungsstelle sind und die hellgrünen deuten auf die stabilisierenden Wasserstoffbrückenbindungen hin.

3.6. Funktionsanalyse der Lsm-Domäne

Die funktionell relevanten Aminosäuren werden sowohl im multiplen Sequenzalignment als auch im 3D-Modell von Ataxin-2 mit dem gleichen Farbschema hervorgehoben, um so einen optisch besseren Überblick über deren Position und Funktion zu bekommen (siehe Abbildung 9 und 10). Basierend auf der Kristallstruktur von Sm1 von P. abyssi, welches an ein Uridinhepatemer (U_7) gebunden ist, wurden einige Aminosäuren von Sm1 markiert, die in die RNA Bindung involviert [Thore et al. 2003] und größtenteils physiko-chemisch in Ataxin-2 konserviert sind. Im folgenden Text stehen jeweils in Klammern die entsprechenden Stellen der Aminosäuren in Sm1 und Ataxin-2. Die Reste, die interne Bindungstasche von U_7 bilden, sind H37/K299, N39/L302 und R63/K330, während Ionenbindungen zwischen K22/K284, R63/K330 und D65/S332 die RNA Bindungsstelle stabilisieren. Die Reste, die Einfluss auf die externe Bindungsstelle für RNA nehmen sind R4/R266, H10/T272 und Y34/Y296. Zusätzlich Stabilität entsteht durch die Wasserstoffbrückenbindungen zwischen H10/T272 und Y34/Y296. Die Relevanz dieser Stellen wird noch dadurch untermauert, dass diese RNA-Bindungsreste in Sm1 in P. abyssi und in Archaeoglobus fulgidus (PDB Kennung 1i4k, Kette A) identisch sind mit der Ausnahme von Histidin an der Stelle 10 (H10), welches durch Asparagin ersetzt ist [Törö et al. 2001; Thore et al. 2003].
Außerdem wurde untersucht, ob Ataxin-2 möglicherweise auch mittels der Lsm Domäne Oligomere formen kann. Hier konnten als Vergleichsgrundlage die sehr detaillierten Kristallstrukturanalysen der äußerst ähnlichen snRNP Heterodimere D_1-D_2 und D_3-B [Kambach et al. 1999] herangezogen werden.
Da die intermolekularen Interaktionen in beiden Dimeren fast identisch sind, soll hier exemplarisch der D_3 und B Komplex betrachtet werden. Dieser Komplex wird in erster Linie durch eine Verbindung zwischen dem fünften β-Strang (β5) von D_3 mit dem vierten β-Strang (β4) von B ((D_3/Ataxin-2 – B/Ataxin-2): R69/V335 - R73/K330, L71/V337 - L71/L328, und L73/F339 - L69/S326 stabilisiert. Zusätzlich festigen zwei hydrophobe Cluster, geformt durch die Seitenketten von D_3 und B, die Struktur des Dimers. Der erste

Cluster schließt F70/ V336 und I72/Q338 (beide im β5 Strang) von D_3 und F27/Y289 (β2 Strang), L67/M324, V70/I327 und L72/L328 (alle im β4 Strang) von B mit ein. Der zweite Cluster besteht aus P6/M267, L10/L271 (beide in der α-Helix), V18/C279 (β1 Strang), L32/F293 (β2 Strang), I33/K294 (Loop nach dem β2 Strang), I68/F334, L71/V337 und L73/F339 (alle im β5 Strang) von D_3 und I41/L304, C43/A306 (beide in β3), L69/S326 und L71/L328 (beide in β4) von B.

Ein Stapelungsbeitrag „stacking interactions“ zwischen den Guanidinogruppen eines Arginins R69/V335 aus D_3 und zwei Argininen R25/G287 und R49/T312 aus B, sowie eine ionische Interaktion zwischen E21/Q282 aus D_3 und R65/S322 aus B führen zu einer weiteren Stabilisierung des Dimers. Hierbei ist anzumerken, dass die letztbeschriebene Salzbrücke „salt bridge“ im D_1-D_2 Komplex nicht beobachtet werden kann, obwohl die Aminosäuren an den entsprechenden Stellen identisch sind.

Der hohe Konservierungsgrad, der für die Heterodimerisation relevanten Aminosäuren, spricht dafür, dass auch Ataxin-2 über die Lsm-Domäne Oligomere formen kann.

4. Diskussion

Ziel dieser Arbeit war es, mit Hilfe von bioinformatischen Methoden, Thesen über eine potentielle, physiologische Funktion des Ataxin-2 Proteins zu generieren sowie mögliche Homologe zu identifizieren und ein Modell für eine 3D-Struktur für Teilbereiche zu entwickeln.

Hierzu wurde eine detaillierte Analyse von Ataxin-2 und seiner Homologe, darunter das Hefe Homolog PBP1 durchgeführt. Mittels eines strukturbasierten multiplen Sequenz-alignments von Sm und Sm-like Proteinen ist es gelungen, ein 3D-Modell für die Lsm Domäne von Ataxin-2 zu generieren. Die 3D-Struktur von Ataxin-2 bzw. von einzelnen Domänen ist bisher unbekannt und so stellt das hier beschriebene Modell für die Lsm Domäne einen ersten Ansatz für die Bestimmung eines kompletten 3D-Modells für Ataxin-2 dar. Zukünftige experimentelle Ansätze wie z.B. mittels der Röntgenkristallographie können dann Aufschluss geben über die Qualität der Vorhersage. Da aber derartige Versuche sehr zeitaufwendig sind und soweit bekannt zurzeit auch nicht durchgeführt werden, steht im Augenblick nur das hier entwickelte Modell für weitere Funktionsbestimmungen zur Verfügung.

Weitere Vergleiche zeigten einen hohen Grad der Konservierung der chemischen Eigenschaften der RNA bindenden Aminosäuren in den alignierten Lsm Domänen. Besonders auffällig war der Vergleich zwischen P. abyssi und dem menschlichen Ataxin-2. Basierend auf diesen Beobachtungen kann man vermuten, das Ataxin-2 in der Lage ist, mit Hilfe der identifizierten Aminosäuren RNA zu binden. Diese Beobachtung deckt sich gut mit schon veröffentlichten Studien, die erste experimentelle Hinweise darauf liefern, dass Ataxin-2 in der mRNA Übersetzung und/ oder Verteilung eine Rolle spielen könnte. Die physiologische Funktion von Ataxin-2 konnte zwar auch mittels dieser Studien nicht bestimmt werden [Kiehl et al. 2000; Meunier et al. 2002; Satterfield et al. 2002; Figueroa und Pulst 2003], allerdings lässt sich beobachten, dass Ataxin-2 mit dem A2BP1 [Shibata et al. 2000] interagiert. Dieses ist homolog zu dem Caenorhabditis elegans Protein Fox-1. Beide Proteine scheinen eine Rolle in der Embryogenese und im erwachsenen Nervensystem zu spielen sowie RNA binden zu können [Kiehl et al. 2000]. Für Fox-1 wurde außerdem die Fähigkeit für ein alternatives Spleißen beschrieben [Jin et al. 2003].

Diese Beobachtungen deuten stark darauf hin, dass Ataxin-2 in ähnliche mRNA Verarbeitungsprozesse involviert ist.
Deshalb wäre es jetzt sinnvoll, in weiteren Schritten experimentell zu erforschen, inwieweit Ataxin-2 und seine Homologe eine Rolle im Rahmen der RNA-Verarbeitung spielen. Die Funktion könnte im Bereich der Regulation der Polyadenylation der mRNA liegen, wie es auch bereits für PBP1 beschrieben wurde.
Zusätzlich lässt die beobachtete Übereinstimmung von Aminosäuren in den D_1-D_2 und D_3-B Heterodimeren, die in den Proteinen mit Lsm Domänen für die Oligomerisierung verantwortlich sind, mit den entsprechenden Aminosäuren in Ataxin-2 vermuten, dass auch Ataxin-2 in der Lage ist, solche Komplexe zu bilden. Dies ist eine neue These, die im Rahmen der vorliegenden Arbeit entstanden ist und bisher experimentell nicht untersucht wurde.
Einige weitere potentielle Funktionen von Ataxin-2 wurden in der Vergangenheit beschrieben und konnten hier weder bestätigt noch widerlegt werden. Hierzu zählen die Beobachtungen, dass z.B. das Drosophila Protein Datx2, welches zwei homologe Domänen zu Ataxin-2 aufweist, die auch in anderen Organismen gut konserviert sind, ein Regulator für den Aktinfilamentaufbau ist. Hieraus lässt sich die Vermutung ableiten, dass durch eine veränderte Ataxin-2 Aktivität eine Dysregulation der Aktinzytoskelettstruktur entsteht, die verantwortlich sein könnte für die Neurodegeneration in SCA2 [Satterfield et al. 2002]. Die Interaktion von Thrombopoetin mit dem EPO-R (endogener Erythropoetin Rezeptor) und einem Ataxin-2 Homolog namens A2D, weist darauf hin, dass eine in Funktion im Bereich der Zytokinsignalkette vorhanden sein könnte [Meunier et al. 2002]. Desweiteren wurde bei der Suche nach einem Zelltod verursachendem Gen bei Neuroblastomen entdeckt, dass Ataxin-2 die Prädispositon zur Apoptose *in vitro* und *in vivo* der Neuroblastomzellen beeinflusst [Wiedemeyer et al. 2003]. Eine weitere Studie zeigte, dass Ataxin-2 physiologischerweise primär im Golgi-Apparat anzufinden ist. Ataxin-2 mit einer verlängerten PolyQ-Kette ist dagegen nicht im Golgi-Apparat anzutreffen und führt zu einer veränderten Morphologie des Golgi-Apparates und könnte daher eine Rolle beim Zelltod spielen [Huynh et al. 2003]. Schließlich konnte in einer kürzlich veröffentlichten Studie eine Interaktion zwischen Ataxin-2 und zwei Mitgliedern aus der Endophilin Familie, Endophilin A1 und Endophilin A3 gezeigt werden [Ralser et al. 2005]. In der gleichen Studie wurde interessanterweise festgestellt, das Huntingtin, welches ja ebenfalls ein Polyglutaminprotein ist, auch mit Endophilin A3 interagiert. Endophilin-A

Proteine haben wahrscheinlich die Funktion zelluläre Ereignisse zu koordinieren wie z.B. die Aktinfunktion und die Signalkaskade für die Endozytose.

Die Grundlage für weitere *in vivo* Experimente zur Erforschung der physiologischen Funktion von Ataxin-2 wurde kürzlich gelegt, in dem eine Mauspopulation geschaffen wurde, die kein ATXN2 Gen aufweist (*„knockout“* Maus) [Kiehl et al. 2006]. Erste Untersuchungen dieser Population zeigten, dass die Mäuse keine größeren makroskopischen oder lichtmikroskopischen nachweisbaren Auffälligkeiten zeigten; auch das Verhalten war weitgehend ungestört; allerdings zeigen ATXN2 *„knockout“* Mäuse eine gesteigerte Gewichtszunahme. Daraus lässt sich folgern, dass, obwohl Ataxin-2 in vielen Gewebearten gebildet wird, es keine existentielle Rolle in der Entwicklung bzw. für das Überleben der Mäuse spielt. Weitere Versuche hinsichtlich der verstärkten Gewichtszunahme im Erwachsenenalter und daraus hervorgehende Erkenntnisse werden wohl demnächst zu erwarten sein.
Abzuwarten bleibt es für die Zukunft, wenn mehr über die physiologische Funktion von Ataxin-2 bekannt ist, inwieweit der verlängert Polyglutamintrakt in mutierten Proteinen die normale Funktion von Ataxin-2 beeinflusst.

5. Zusammenfassung

Die Spinozerebelläre Ataxie Typ 2 (SCA2) ist eine autosomal-dominant vererbte, neurodegenerative Erkrankung. Verursacht wird sie durch eine verlängerte Glutaminkette (PolyQ-Kette) im 5´ Ende (coding Region) des Exons 1, des ATXN2 Gens welches zu Ataxin-2 transkribiert wird.

Ziel der vorliegenden Arbeit war die Analyse des Proteins mittels bioinformatischer Methoden, um eine Grundlage für weitere Experimente was die Funktion von Ataxin-2 angeht, zu erstellen. Hierfür wurde das Protein einer statistischen Analyse unterzogen, die zur Bestimmung der physikochemischen Eigenschaften dient. Desweiteren wurden multiple Sequenzalignments mit homologen Proteinen gebildet, sowie Domänen, Sequenzmotive und mögliche Interaktionspartner bestimmt. Durch diese Untersuchungen sollten Hypothesen zu möglichen zellulären Funktionen von Ataxin-2 generiert werden, die experimentell überprüfbar sind. Daher habe ich mich auf die Untersuchung eines potentiell funktionellen Abschnittes konzentriert, der die RNA bindende Lsm Domäne enthält. In den Sequenzalignments zeigte sich ein hoher Konservierungsgrad, der für die RNA-Bindung chemisch relevanten Aminosäuren in diesem Bereich.

Für die Lsm Domäne wurde, unter der Nutzung des multiplen Sequenzalignments mit Sm und Sm-like Proteinen ein 3D–Modell entwickelt.

Im Weiteren wäre es jetzt interessant, biochemische Experimente durchzuführen, mit deren Hilfe man feststellen könnte, ob Ataxin-2 und seine Homologe tatsächlich eine Rolle in der Regulation der Polyadenylation der mRNA spielen. Außerdem sollte man überprüfen, ob Ataxin-2 mittels der Lsm Domänen Oligomere bilden kann.

6. Literaturverzeichnis

1. The yeast genome directory. Nature 1997;387(6632 Suppl):5.

2. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998;282(5396):2012-2018.

3. Finishing the euchromatic sequence of the human genome. Nature 2004;431(7011):931-945.

4. Abele M, Burk K, Schols L, Schwartz S, Besenthal I, Dichgans J, Zuhlke C, Riess O, Klockgether T. The aetiology of sporadic adult-onset ataxia. Brain 2002;125(Pt 5):961-968.

5. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, WoodageT, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. The genome sequence of Drosophila melanogaster. Science 2000;287(5461):2185-2195.

6. Albrecht M, Golatta M, Wullner U, Lengauer T. Structural and functional analysis of ataxin-2 and ataxin-3. Eur J Biochem 2004;271(15):3155-3170.

7. Albrecht M, Lengauer T. Survey on the PABC recognition motif PAM2. Biochem Biophys Res Commun 2004;316(1):129-138.

8. Albrecht M, Tosatto SC, Lengauer T, Valle G. Simple consensus procedures are effective and sufficient in secondary structure prediction. Protein Eng 2003;16(7):459-462.

9. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389-3402.

10. Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res 2004;32 Database issue:D226-229.

11. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 2004;32 Database issue:D115-119.

12. Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. Science 1986;234(4773):179-186.

13. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. The Universal Protein Resource (UniProt). Nucleic Acids Res 2005;33(Database issue):D154-159.

14. Basu P, Chattopadhyay B, Gangopadhaya PK, Mukherjee SC, Sinha KK, Das SK, Roychoudhury S, Majumder PP, Bhattacharyya NP. Analysis of CAG repeats in SCA1, SCA2, SCA3, SCA6, SCA7 and DRPLA loci in spinocerebellar ataxia patients and distribution of CAG repeats at the SCA1, SCA2 and SCA6 loci in nine ethnic populations of eastern India. Hum Genet 2000;106(6):597-604.

15. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32 Database issue:D138-141.

16. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res 2006;34(Database issue):D16-20.

17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235-242.

18. Bernal A, Ear U, Kyrpides N. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res 2001;29(1):126-127.

19. Birney E, Andrews D, Bevan P, Caccamo M, Cameron G, Chen Y, Clarke L, Coates G, Cox T, Cuff J, Curwen V, Cutts T, Down T, Durbin R, Eyras E, Fernandez-Suarez XM,

Gane P, Gibbins B, Gilbert J, Hammond M, Hotz H, Iyer V, Kahari A, Jekosch K, Kasprzyk A, Keefe D, Keenan S, Lehvaslaiho H, McVicker G, Melsopp C, Meidl P, Mongin E, Pettett R, Potter S, Proctor G, Rae M, Searle S, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Ureta-Vidal A, Woodwark C, Clamp M, Hubbard T. Ensembl 2004. Nucleic Acids Res 2004;32 Database issue:D468-470.

20. Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Flicek P, Graf S, Hammond M, Herrero J, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Kokocinski F, Kulesha E, London D, Longden I, Melsopp C, Meidl P, Overduin B, Parker A, Proctor G, Prlic A, Rae M, Rios D, Redmond S, Schuster M, Sealy I, Searle S, Severin J, Slater G, Smedley D, Smith J, Stabenau A, Stalker J, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Hubbard TJ. Ensembl 2006. Nucleic Acids Res 2006;34(Database issue):D556-561.

21. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis 1993;14(10):1023-1031.

22. Bourne PE, Addess KJ, Bluhm WF, Chen L, Deshpande N, Feng Z, Fleri W, Green R, Merino-Ott JC, Townsend-Merino W, Weissig H, Westbrook J, Berman HM. The distribution and query systems of the RCSB Protein Data Bank. Nucleic Acids Res 2004;32 Database issue:D223-225.

23. Brendel V, Bucher P, Nourbakhsh IR, Blaisdell BE, Karlin S. Methods and algorithms for statistical analysis of protein sequences. Proc Natl Acad Sci U S A 1992;89(6):2002-2006.

24. Brkanac Z, Fernandez M, Matsushita M, Lipe H, Wolff J, Bird TD, Raskind WH. Autosomal dominant sensory/motor neuropathy with Ataxia (SMNA): Linkage to chromosome 7q22-q32. Am J Med Genet 2002;114(4):450-457.

25. Browne DL, Gancher ST, Nutt JG, Brunt ER, Smith EA, Kramer P, Litt M. Episodic ataxia/myokymia syndrome is associated with point mutations in the human potassium channel gene, KCNA1. Nat Genet 1994;8(2):136-140.

26. Bru C, Courcelle E, Carrere S, Beausse Y, Dalmar S, Kahn D. The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res 2005;33(Database issue):D212-215.

27. Brusco A, Gellera C, Cagnoli C, Saluto A, Castucci A, Michielotto C, Fetoni V, Mariotti C, Migone N, Di Donato S, Taroni F. Molecular genetics of hereditary spinocerebellar ataxia: mutation analysis of spinocerebellar ataxia genes and CAG/CTG repeat expansion detection in 225 Italian families. Arch Neurol 2004;61(5):727-733.

28. Brusse E, de Koning I, Maat-Kievit A, Oostra BA, Heutink P, van Swieten JC. Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene (SCA27): A new phenotype. Mov Disord 2005.

29. Bryer A, Krause A, Bill P, Davids V, Bryant D, Butler J, Heckmann J, Ramesar R, Greenberg J. The hereditary adult-onset ataxias in South Africa. J Neurol Sci 2003;216(1):47-54.

30. Bujnicki JM, Elofsson A, Fischer D, Rychlewski L. Structure prediction meta server. Bioinformatics 2001;17(8):750-751.

31. Burk K, Abele M, Fetter M, Dichgans J, Skalej M, Laccone F, Didierjean O, Brice A, Klockgether T. Autosomal dominant cerebellar ataxia type I clinical features and MRI in families with SCA1, SCA2 and SCA3. Brain 1996;119 (Pt 5):1497-1505.

32. Burk K, Fetter M, Abele M, Laccone F, Brice A, Dichgans J, Klockgether T. Autosomal dominant cerebellar ataxia type I: oculomotor abnormalities in families with SCA1, SCA2, and SCA3. J Neurol 1999a;246(9):789-797.

33. Burk K, Globas C, Bosch S, Klockgether T, Zuhlke C, Daum I, Dichgans J. Cognitive deficits in spinocerebellar ataxia type 1, 2, and 3. J Neurol 2003;250(2):207-211.

34. Burk K, Klockgether T, Dichgans J. [New insights in the molecular genetics and pathophysiology of hereditary ataxias]. Nervenarzt 1999b;70(6):491-495.

35. Cagnoli C, Mariotti C, Taroni F, Seri M, Brussino A, Michielotto C, Grisoli M, Di Bella D, Migone N, Gellera C, Di Donato S, Brusco A. SCA28, a novel form of autosomal dominant cerebellar ataxia on chromosome 18p11.22-q11.2. Brain 2006;129(Pt 1):235-242.

36. Cancel G, Durr A, Didierjean O, Imbert G, Burk K, Lezin A, Belal S, Benomar A, Abada-Bendib M, Vial C, Guimaraes J, Chneiweiss H, Stevanin G, Yvert G, Abbas N, Saudou F, Lebre AS, Yahyaoui M, Hentati F, Vernant JC, Klockgether T, Mandel JL, Agid Y, Brice A. Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. Hum Mol Genet 1997;6(5):709-715.

37. Chen DH, Brkanac Z, Verlinde CL, Tan XJ, Bylenok L, Nochlin D, Matsushita M, Lipe H, Wolff J, Fernandez M, Cimino PJ, Bird TD, Raskind WH. Missense mutations in the regulatory domain of PKC gamma: a new mechanism for dominant nonepisodic cerebellar ataxia. Am J Hum Genet 2003;72(4):839-849.

38. Chen DH, Cimino PJ, Ranum LP, Zoghbi HY, Yabe I, Schut L, Margolis RL, Lipe HP, Feleke A, Matsushita M, Wolff J, Morgan C, Lau D, Fernandez M, Sasaki H, Raskind WH, Bird TD. The clinical and genetic spectrum of spinocerebellar ataxia 14. Neurology 2005;64(7):1258-1260.

39. Chung M, Soong BW. Reply to: SCA-19 and SCA-22: evidence for one locus with a worldwide distribution. Brain 2004;127:E6; author reply E7.

40. Chung MY, Lu YC, Cheng NC, Soong BW. A novel autosomal dominant spinocerebellar ataxia (SCA22) linked to chromosome 1p21-q23. Brain 2003;126(Pt 6):1293-1299.

41. Cochrane G, Aldebert P, Althorpe N, Andersson M, Baker W, Baldwin A, Bates K, Bhattacharyya S, Browne P, van den Broek A, Castro M, Duggan K, Eberhardt R, Faruque N, Gamble J, Kanz C, Kulikova T, Lee C, Leinonen R, Lin Q, Lombard V, Lopez R, McHale M, McWilliam H, Mukherjee G, Nardone F, Pastor MP, Sobhany S, Stoehr P, Tzouvara K, Vaughan R, Wu D, Zhu W, Apweiler R. EMBL Nucleotide Sequence Database: developments in 2005. Nucleic Acids Res 2006;34(Database issue):D10-15.

42. Corpet F, Servant F, Gouzy J, Kahn D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res 2000;28(1):267-269.

43. David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou E, Drabkin H, Gemmill R, Giunti P, Benomar A, Wood N, Ruberg M, Agid Y, Mandel JL, Brice A. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat Genet 1997;17(1):65-70.

44. Devos D, Schraen-Maschke S, Vuillaume I, Dujardin K, Naze P, Willoteaux C, Destee A, Sablonniere B. Clinical features and genetic analysis of a new form of spinocerebellar ataxia. Neurology 2001;56(2):234-238.

45. Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem 1993;62:289-321.

46. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41(21):6573-6582.

47. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12(1):54-60.

48. Edelhoch H. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 1967;6(7):1948-1954.

49. Escayg A, De Waard M, Lee DD, Bichet D, Wolf P, Mayer T, Johnston J, Baloh R, Sander T, Meisler MH. Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. Am J Hum Genet 2000;66(5):1531-1539.

50. Evert BO, Vogt IR, Vieira-Saecker AM, Ozimek L, de Vos RA, Brunt ER, Klockgether T, Wullner U. Gene expression profiling in ataxin-3 expressing cell lines reveals distinct effects of normal and mutant ataxin-3. J Neuropathol Exp Neurol 2003;62(10):1006-1018.

51. Figueroa KP, Pulst SM. Identification and expression of the gene for human ataxin-2-related protein on chromosome 16. Exp Neurol 2003;184(2):669-678.

52. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995;269(5223):496-512.

53. Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 1996;12(6):543-548.

54. Gardner RJ, Knight MA, Hara K, Tsuji S, Forrest SM, Storey E. Spinocerebellar ataxia type 15. Cerebellum 2005;4(1):47-50.

55. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 2003;31(13):3784-3788.

56. Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM. The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. Am J Hum Genet 1997a;60(4):842-850.

57. Geschwind DH, Perlman S, Figueroa KP, Karrim J, Baloh RW, Pulst SM. Spinocerebellar ataxia type 6. Frequency of the mutation and genotype-phenotype correlations. Neurology 1997b;49(5):1247-1251.

58. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A,

Old S, Mockrin S, Collins F. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 2004;428(6982):493-521.

59. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 1989;182(2):319-326.

60. Ginalski K, Rychlewski L. Detection of reliable and unexpected protein fold predictions using 3D-Jury. Nucleic Acids Res 2003;31(13):3291-3292.

61. Gispert S, Twells R, Orozco G, Brice A, Weber J, Heredero L, Scheufler K, Riley B, Allotey R, Nothers C, et al. Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23-24.1. Nat Genet 1993;4(3):295-299.

62. Giunti P, Sabbadini G, Sweeney MG, Davis MB, Veneziano L, Mantuano E, Federico A, Plasmati R, Frontali M, Wood NW. The role of the SCA2 trinucleotide repeat expansion in 89 autosomal dominant cerebellar ataxia families. Frequency, clinical and genetic correlates. Brain 1998;121 (Pt 3):459-467.

63. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 2003;31(13):3320-3323.

64. Greenfield J. The Spino-cerebellar Degenerations. Springfield, IL; 1954.

65. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 1990;4(2):155-161.

66. Harding AE. Classification of the hereditary ataxias and paraplegias. Lancet 1983;1(8334):1151-1155.

67. He W, Parker R. Functions of Lsm proteins in mRNA degradation and splicing. Curr Opin Cell Biol 2000;12(3):346-350.

68. Herman-Bert A, Stevanin G, Netter JC, Rascol O, Brassat D, Calvas P, Camuzat A, Yuan Q, Schalling M, Durr A, Brice A. Mapping of spinocerebellar ataxia 13 to chromosome 19q13.3-q13.4 in a family with autosomal dominant cerebellar ataxia and mental retardation. Am J Hum Genet 2000;67(1):229-235.

69. Holmes G. An attempt to classify cerebellar disease with a note on Marie´s hereditary ataxia. Brain 1907;30:545 - 567.

70. Holmes SE, O'Hearn E, Margolis RL. Why is SCA12 different from other SCAs? Cytogenet Genome Res 2003;100(1-4):189-197.

71. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM,

Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinsci F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E. Ensembl 2005. Nucleic Acids Res 2005;33(Database issue):D447-453.

72. Huynh DP, Yang HT, Vakharia H, Nguyen D, Pulst SM. Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death. Hum Mol Genet 2003;12(13):1485-1496.

73. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323(3):573-584.

74. Ichikawa Y, Goto J, Hattori M, Toyoda A, Ishii K, Jeong SY, Hashida H, Masuda N, Ogata K, Kasai F, Hirai M, Maciel P, Rouleau GA, Sakaki Y, Kanazawa I. The genomic structure and expression of MJD, the Machado-Joseph disease gene. J Hum Genet 2001;46(7):413-422.

75. Igarashi S, Koide R, Shimohata T, Yamada M, Hayashi Y, Takano H, Date H, Oyake M, Sato T, Sato A, Egawa S, Ikeuchi T, Tanaka H, Nakano R, Tanaka K, Hozumi I, Inuzuka T, Takahashi H, Tsuji S. Suppression of aggregate formation and apoptosis by transglutaminase inhibitors in cells expressing truncated DRPLA protein with an expanded polyglutamine stretch. Nat Genet 1998;18(2):111-117.

76. Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, Weber C, Mandel JL, Cancel G, Abbas N, Durr A, Didierjean O, Stevanin G, Agid Y, Brice A. Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet 1996;14(3):285-291.

77. Ishikawa K, Toru S, Tsunemi T, Li M, Kobayashi K, Yokota T, Amino T, Owada K, Fujigasaki H, Sakamoto M, Tomimitsu H, Takashima M, Kumagai J, Noguchi Y, Kawashima Y, Ohkoshi N, Ishida G, Gomyoda M, Yoshida M, Hashizume Y, Saito Y, Murayama S, Yamanouchi H, Mizutani T, Kondo I, Toda T, Mizusawa H. An autosomal dominant cerebellar ataxia linked to chromosome 16q22.1 is associated with a single-nucleotide substitution in the 5' untranslated region of the gene encoding a protein with spectrin repeat and Rho guanine-nucleotide exchange-factor domains. Am J Hum Genet 2005;77(2):280-296.

78. Jiang H, Tang B, Xia K, Zhou Y, Xu B, Zhao G, Li H, Shen L, Pan Q, Cai F. Spinocerebellar ataxia type 6 in Mainland China: molecular and clinical features in four families. J Neurol Sci 2005a;236(1-2):25-29.

79. Jiang H, Tang BS, Xu B, Zhao GH, Shen L, Tang JG, Li QH, Xia K. Frequency analysis of autosomal dominant spinocerebellar ataxias in mainland Chinese patients and clinical and molecular characterization of spinocerebellar ataxia type 6. Chin Med J (Engl) 2005b;118(10):837-843.

80. Jin Y, Suzuki H, Maegawa S, Endo H, Sugano S, Hashimoto K, Yasuda K, Inoue K. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. Embo J 2003;22(4):905-912.

81. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22(12):2577-2637.

82. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, Nagai K. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 1999;96(3):375-387.

83. Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. Bioinformatics 1998;14(10):846-856.

84. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 1994;8(3):221-228.

85. Kiehl TR, Nechiporuk A, Figueroa KP, Keating MT, Huynh DP, Pulst SM. Generation and characterization of Sca2 (ataxin-2) knockout mice. Biochem Biophys Res Commun 2006;339(1):17-24.

86. Kiehl TR, Shibata H, Pulst SM. The ortholog of human ataxin-2 is essential for early embryonic patterning in C. elegans. J Mol Neurosci 2000;15(3):231-241.

87. Kim JY, Park SS, Joo SI, Kim JM, Jeon BS. Molecular analysis of Spinocerebellar ataxias in Koreans: frequencies and reference ranges of SCA1, SCA2, SCA3, SCA6, and SCA7. Mol Cells 2001;12(3):336-341.

88. Klockgether T. Handbook of ataxia disorders. New York: Dekker, M.; 2000.

89. Knight MA, Gardner RJ, Bahlo M, Matsuura T, Dixon JA, Forrest SM, Storey E. Dominantly inherited ataxia and dysphonia with dentate calcification: spinocerebellar ataxia type 20. Brain 2004a;127(Pt 5):1172-1181.

90. Knight MA, Kennerson M, Nicholson GA, Gardner RJM, Storey E, Thomas PQ, Forrest SM. A new spinocerebellar ataxia, SCA15. Am J Hum Genet 2001;69:509.

91. Knight MA, Kennerson ML, Anney RJ, Matsuura T, Nicholson GA, Salimi-Tari P, Gardner RJ, Storey E, Forrest SM. Spinocerebellar ataxia type 15 (sca15) maps to 3p24.2-3pter: exclusion of the ITPR1 gene, the human orthologue of an ataxic mouse mutant. Neurobiol Dis 2003;13(2):147-157.

92. Knight MA, McKinlay Gardner RJ, Bahlo M, Matsuura T, Dixon JA, Forrest SM, Storey E. Dominantly inherited ataxia and dysphonia with dentate calcification: spinocerebellar ataxia type 20. Brain 2004b.

93. Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, et al. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 1994;6(1):9-13.

94. Koide R, Kobayashi S, Shimohata T, Ikeuchi T, Maruyama M, Saito M, Yamada M, Takahashi H, Tsuji S. A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? Hum Mol Genet 1999;8(11):2047-2053.

95. Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet 1999;21(4):379-384.

96. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM. The RCSB PDB information portal for structural genomics. Nucleic Acids Res 2006;34(Database issue):D302-305.

97. Kyrpides NC. Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide. Bioinformatics 1999;15(9):773-774.

98. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157(1):105-132.

99. La Spada AR, Wilson, E.M., Lubahn, D.B., Harding, A.E., Fischbeck, K.H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 1991;352:77 - 79.

100. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA,

Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

101. Lee WY, Jin DK, Oh MR, Lee JE, Song SM, Lee EA, Kim GM, Chung JS, Lee KH. Frequency analysis and clinical characterization of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Korean patients. Arch Neurol 2003;60(6):858-863.

102. Lesk AM. Bioinformatik: Eine Einführung. Heidelberg Freiburg: Spektrum Akademischer Verlag; 2003.

103. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res 2006;34(Database issue):D257-260.

104. Li M, Miwa S, Kobayashi Y, Merry DE, Yamamoto M, Tanaka F, Doyu M, Hashizume Y, Fischbeck KH, Sobue G. Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. Ann Neurol 1998;44(2):249-254.

105. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Structure (Camb) 2003a;11(11):1453-1459.

106. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 2003b;31(13):3701-3708.

107. Liolios K, Tavernarakis N, Hugenholtz P, Kyrpides NC. The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. Nucleic Acids Res 2006;34(Database issue):D332-334.

108. Liu J, Rost B. NORSp: Predictions of long regions without regular secondary structure. Nucleic Acids Res 2003;31(13):3833-3835.

109. Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res 2002;30(1):264-267.

110. Lopes-Cendes I, Teive HG, Calcagnotto ME, Da Costa JC, Cardoso F, Viana E, Maciel JA, Radvany J, Arruda WO, Trevisol-Bittencourt PC, Rosa Neto P, Silveira I, Steiner CE, Pinto Junior W, Santos AS, Correa Neto Y, Werneck LC, Araujo AQ, Carakushansky G, Mello LR, Jardim LB, Rouleau GA. Frequency of the different mutations causing spinocerebellar ataxia (SCA1, SCA2, MJD/SCA3 and DRPLA) in a large group of Brazilian patients. Arq Neuropsiquiatr 1997;55(3B):519-529.

111. Lorenzetti D, Bohlega S, Zoghbi HY. The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia. Neurology 1997;49(4):1009-1013.

112. Mangus DA, Amrani N, Jacobson A. Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-binding protein, regulates polyadenylation. Mol Cell Biol 1998;18(12):7383-7396.

113. Manto MU. The wide spectrum of spinocerebellar ataxias (SCAs). Cerebellum 2005;4(1):2-6.

114. Maruyama H, Izumi Y, Morino H, Oda M, Toji H, Nakamura S, Kawakami H. Difference in disease-free survival curve and regional distribution according to subtype of spinocerebellar ataxia: a study of 1,286 Japanese patients. Am J Med Genet 2002;114(5):578-583.

115. Maschke M, Oehlert G, Xie TD, Perlman S, Subramony SH, Kumar N, Ptacek LJ, Gomez CM. Clinical feature profile of spinocerebellar ataxia type 1-8 predicts genetically defined subtypes. Mov Disord 2005;20(11):1405-1412.

116. Masino L, Kelly G, Leonard K, Trottier Y, Pastore A. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. FEBS Lett 2002;513(2-3):267-272.

117. Masino L, Musi V, Menon RP, Fusi P, Kelly G, Frenkiel TA, Trottier Y, Pastore A. Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. FEBS Lett 2003;549(1-3):21-25.

118. Matsumura R, Futamura N, Ando N, Ueno S. Frequency of spinocerebellar ataxia mutations in the Kinki district of Japan. Acta Neurol Scand 2003;107(1):38-41.

119. Matsuura T, Achari M, Khajavi M, Bachinski LL, Zoghbi HY, Ashizawa T. Mapping of the gene for a novel spinocerebellar ataxia with pure cerebellar signs and epilepsy. Ann Neurol 1999;45(3):407-411.

120. Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L, Achari M, Pulst SM, Alonso E, Noebels JL, Nelson DL, Zoghbi HY, Ashizawa T. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet 2000;26(2):191-194.

121. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16(4):404-405.

122. Meijer IA, Hand CK, Grewal KK, Stefanelli MG, Ives EJ, Rouleau GA. A locus for autosomal dominant hereditary spastic ataxia, SAX1, maps to chromosome 12p13. Am J Hum Genet 2002;70(3):763-769.

123. Meunier C, Bordereaux D, Porteu F, Gisselbrecht S, Chretien S, Courtois G. Cloning and characterization of a family of proteins associated with Mpl. J Biol Chem 2002;277(11):9139-9147.

124. Miyoshi Y, Yamada T, Tanimura M, Taniwaki T, Arakawa K, Ohyagi Y, Furuya H, Yamamoto K, Sakai K, Sasazuki T, Kira J. A novel autosomal dominant spinocerebellar ataxia (SCA16) linked to chromosome 8q22.1-24.1. Neurology 2001;57(1):96-100.

125. Mizusawa H, Clark HB, Koeppen AH. Spinocerebellar Ataxias. In: Dickson DW, editor. Neurodegeneration: The Molecular Pathology of Dementia and Movement Disorders. Basel: ISN Neuropath Press; 2003. p. 242-256.

126. Moseley ML, Benzow KA, Schut LJ, Bird TD, Gomez CM, Barkhaus PE, Blindauer KA, Labuda M, Pandolfo M, Koob MD, Ranum LP. Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. Neurology 1998;51(6):1666-1671.

127. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D. Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc Natl Acad Sci U S A 2003;100(8):4539-4544.

128. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247(4):536-540.

129. Naito H, Oyanagi S. Familial myoclonus epilepsy and choreoathetosis: hereditary dentatorubral-pallidoluysian atrophy. Neurology 1982;32(8):798-807.

130. Nakamura K, Jeong SY, Uchihara T, Anno M, Nagashima K, Nagashima T, Ikeda S, Tsuji S, Kanazawa I. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. Hum Mol Genet 2001;10(14):1441-1448.

131. Nance MA. Clinical aspects of CAG repeat diseases. Brain Pathol 1997;7(3):881-900.

132. Nechiporuk A, Lopes-Cendes I, Nechiporuk T, Starkman S, Andermann E, Rouleau GA, Weissenbach JS, Kort E, Pulst SM. Genetic mapping of the spinocerebellar ataxia type 2 gene on human chromosome 12. Neurology 1996;46(6):1731-1735.

133. Neuwald AF, Koonin EV. Ataxin-2, global regulators of bacterial gene expression, and spliceosomal snRNP proteins share a conserved domain. J Mol Med 1998;76(1):3-5.

134. Okubo K, Sugawara H, Gojobori T, Tateno Y. DDBJ in preparation for overview of research activities behind data submissions. Nucleic Acids Res 2006;34(Database issue):D6-9.

135. Ophoff RA, Terwindt GM, Vergouwe MN, van Eijk R, Oefner PJ, Hoffman SM, Lamerdin JE, Mohrenweiser HW, Bulman DE, Ferrari M, Haan J, Lindhout D, van Ommen GJ, Hofker MH, Ferrari MD, Frants RR. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4. Cell 1996;87(3):543-552.

136. Orozco G, Estrada R, Perry TL, Arana J, Fernandez R, Gonzalez-Quevedo A, Galarraga J, Hansen S. Dominantly inherited olivopontocerebellar atrophy from eastern Cuba. Clinical, neuropathological, and biochemical findings. J Neurol Sci 1989;93(1):37-50.

137. Orr HT, Chung MY, Banfi S, Kwiatkowski TJ, Jr., Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, Zoghbi HY. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 1993;4(3):221-226.

138. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995;4(11):2411-2423.

139. Pareyson D, Gellera C, Castellotti B, Antonelli A, Riggio MC, Mazzucchelli F, Girotti F, Pietrini V, Mariotti C, Di Donato S. Clinical and molecular studies of 73 Italian families with autosomal dominant cerebellar ataxia type I: SCA1 and SCA2 are the most common genotypes. J Neurol 1999;246(5):389-393.

140. Paulson HL, Perez MK, Trottier Y, Trojanowski JQ, Subramony SH, Das SS, Vig P, Mandel JL, Fischbeck KH, Pittman RN. Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. Neuron 1997;19(2):333-344.

141. Perutz MF, Johnson T, Suzuki M, Finch JT. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci U S A 1994;91(12):5355-5358.

142. Piccirillo C, Khanna R, Kiledjian M. Functional characterization of the mammalian mRNA decapping enzyme hDcp2. Rna 2003;9(9):1138-1147.

143. Poirot O, O'Toole E, Notredame C. Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res 2003;31(13):3503-3506.

144. Pollastri G, Przybylski D, Rost B, Baldi P. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins 2002;47(2):228-235.

145. Pujana MA, Corral J, Gratacos M, Combarros O, Berciano J, Genis D, Banchs I, Estivill X, Volpini V. Spinocerebellar ataxias in Spanish patients: genetic analysis of familial and sporadic cases. The Ataxia Study Group. Hum Genet 1999;104(6):516-522.

146. Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 1996;14(3):269-276.

147. Ralser M, Nonhoff U, Albrecht M, Lengauer T, Wanker EE, Lehrach H, Krobitsch S. Ataxin-2 and huntingtin interact with endophilin-A complexes to function in plastin-associated pathways. Hum Mol Genet 2005;14(19):2893-2909.

148. Ranum LP, Schut LJ, Lundgren JK, Orr HT, Livingston DM. Spinocerebellar ataxia type 5 in a family descended from the grandparents of President Lincoln maps to chromosome 11. Nat Genet 1994;8(3):280-284.

149. Riess O, Laccone FA, Gispert S, Schols L, Zuhlke C, Vieira-Saecker AM, Herlt S, Wessel K, Epplen JT, Weber BH, Kreuz F, Chahrokh-Zadeh S, Meindl A, Lunkes A, Aguiar J, Macek M, Jr., Krebsova A, Macek M, Sr., Burk K, Tinschert S, Schreyer I, Pulst SM, Auburger G. SCA2 trinucleotide expansion in German SCA patients. Neurogenetics 1997;1(1):59-64.

150. Robitaille Y, Lopes-Cendes I, Becher M, Rouleau G, Clark AW. The neuropathology of CAG repeat diseases: review and update of genetic and molecular features. Brain Pathol 1997;7(3):901-926.

151. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G. Homology modeling, model and software evaluation: three related resources. Bioinformatics 1998;14(6):523-528.

152. Sahba S, Nechiporuk A, Figueroa KP, Nechiporuk T, Pulst SM. Genomic structure of the human gene for spinocerebellar ataxia type 2 (SCA2) on chromosome 12q24.1. Genomics 1998;47(3):359-364.

153. Saleem Q, Choudhry S, Mukerji M, Bashyam L, Padma MV, Chakravarthy A, Maheshwari MC, Jain S, Brahmachari SK. Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. Hum Genet 2000;106(2):179-187.

154. Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, Tashiro K, Ishida Y, Ikeuchi T, Koide R, Saito M, Sato A, Tanaka T, Hanyu S, Takiyama Y, Nishizawa M, Shimizu N, Nomura Y, Segawa M, Iwabuchi K, Eguchi I, Tanaka H, Takahashi H, Tsuji S. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet 1996;14(3):277-284.

155. Sasaki H, Tashiro K. [Frequencies of triplet repeat disorders in dominantly inherited spinocerebellar ataxia (SCA) in the Japanese]. Nippon Rinsho 1999;57(4):787-791.

156. Satterfield TF, Jackson SM, Pallanck LJ. A Drosophila Homolog of the Polyglutamine Disease Gene SCA2 Is a Dosage-Sensitive Regulator of Actin Filament Formation. Genetics 2002;162(4):1687-1702.

157. Sauter C, Basquin J, Suck D. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. Nucleic Acids Res 2003;31(14):4091-4098.

158. Schelhaas HJ, van de Warrenburg BP, Hageman G, Ippel EE, van Hout M, Kremer B. Cognitive impairment in SCA-19. Acta Neurol Belg 2003;103(4):199-205.

159. Schelhaas HJ, Verbeek DS, Van de Warrenburg BP, Sinke RJ. SCA19 and SCA22: evidence for one locus with a worldwide distribution. Brain 2004;127(Pt 1):E6; author reply E7.

160. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, Bates GP, Davies SW, Lehrach H, Wanker EE. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell 1997;90(3):549-558.

161. Schols L, Amoiridis G, Buttner T, Przuntek H, Epplen JT, Riess O. Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? Ann Neurol 1997a;42(6):924-932.

162. Schols L, Bauer P, Schmidt T, Schulte T, Riess O. Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. Lancet Neurol 2004;3(5):291-304.

163. Schols L, Gispert S, Vorgerd M, Menezes Vieira-Saecker AM, Blanke P, Auburger G, Amoiridis G, Meves S, Epplen JT, Przuntek H, Pulst SM, Riess O. Spinocerebellar ataxia type 2. Genotype and phenotype in German kindreds. Arch Neurol 1997b;54(9):1073-1080.

164. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 1998;95(11):5857-5864.

165. Servant F, Bru C, Carrere S, Courcelle E, Gouzy J, Peyruc D, Kahn D. ProDom: automated clustering of homologous domains. Brief Bioinform 2002;3(3):246-251.

166. She M, Decker CJ, Sundramurthy K, Liu Y, Chen N, Parker R, Song H. Crystal structure of Dcp1p and its functional implications in mRNA decapping. Nat Struct Mol Biol 2004;11(3):249-256.

167. Shibata H, Huynh DP, Pulst SM. A novel protein with RNA-binding motifs interacts with ataxin-2. Hum Mol Genet 2000;9(9):1303-1313.

168. Shimizu Y, Yoshida K, Okano T, Ohara S, Hashimoto T, Fukushima Y, Ikeda S. Regional features of autosomal-dominant cerebellar ataxia in Nagano: clinical and molecular genetic analysis of 86 families. J Hum Genet 2004;49(11):610-616.

169. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11(9):739-747.

170. Silveira I, Coutinho P, Maciel P, Gaspar C, Hayes S, Dias A, Guimaraes J, Loureiro L, Sequeiros J, Rouleau GA. Analysis of SCA1, DRPLA, MJD, SCA2, and SCA6 CAG repeats in 48 Portuguese ataxia families. Am J Med Genet 1998;81(2):134-138.

171. Silveira I, Miranda C, Guimaraes L, Moreira MC, Alonso I, Mendonca P, Ferro A, Pinto-Basto J, Coelho J, Ferreirinha F, Poirier J, Parreira E, Vale J, Januario C, Barbot C, Tuna A, Barros J, Koide R, Tsuji S, Holmes SE, Margolis RL, Jardim L, Pandolfo M, Coutinho P, Sequeiros J. Trinucleotide repeats in 202 families with ataxia: a small expanded (CAG)n allele at the SCA17 locus. Arch Neurol 2002;59(4):623-629.

172. Sinha KK, Worth PF, Jha DK, Sinha S, Stinton VJ, Davis MB, Wood NW, Sweeney MG, Bhatia KP. Autosomal dominant cerebellar ataxia: SCA2 is the most frequent mutation in eastern India. J Neurol Neurosurg Psychiatry 2004;75(3):448-452.

173. Smith JK, Gonda V.E. and Malamud N. Unusual form of cerebellar ataxia: combined dentato-rubral and pallido-lysian degeneration. Neurology 1958;8:205 - 209.

174. Stevanin G, Bouslam N, Thobois S, Azzedine H, Ravaux L, Boland A, Schalling M, Broussolle E, Durr A, Brice A. Spinocerebellar ataxia with sensory neuropathy (SCA25) maps to chromosome 2p. Ann Neurol 2004;55(1):97-104.

175. Stevanin G, Broussolle E, Streichenberger N, Kopp N, Brice A, Durr A. Spinocerebellar ataxia with sensory neuropathy (SCA25). Cerebellum 2005;4(1):58-61.

176. Storey E, du Sart D, Shaw JH, Lorentzos P, Kelly L, McKinley Gardner RJ, Forrest SM, Biros I, Nicholson GA. Frequency of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Australian patients with spinocerebellar ataxia. Am J Med Genet 2000;95(4):351-357.

177. Storey E, Gardner RJ, Knight MA, Kennerson ML, Tuck RR, Forrest SM, Nicholson GA. A new autosomal dominant pure cerebellar ataxia. Neurology 2001;57(10):1913-1915.

178. Storey E, Knight MA, Forrest SM, Gardner RJ. Spinocerebellar ataxia type 20. Cerebellum 2005;4(1):55-57.

179. Tang B, Liu C, Shen L, Dai H, Pan Q, Jing L, Ouyang S, Xia J. Frequency of SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA CAG trinucleotide repeat expansion in patients with hereditary spinocerebellar ataxia from Chinese kindreds. Arch Neurol 2000;57(4):540-544.

180. The_Huntington´s_Disease_Collaborative_Research_Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell 1993;72(6):971-983.

181. Thore S, Mayer C, Sauter C, Weeks S, Suck D. Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J Biol Chem 2003;278(2):1239-1247.

182. Törö I, Thore S, Mayer C, Basquin J, Séraphin B, Suck D. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. Embo J 2001;20(9):2293-2303.

183. van de Warrenburg BP, Sinke RJ, Verschuuren-Bemelmans CC, Scheffer H, Brunt ER, Ippel PF, Maat-Kievit JA, Dooijes D, Notermans NC, Lindhout D, Knoers NV, Kremer HP. Spinocerebellar ataxias in the Netherlands: prevalence and age at onset variance analysis. Neurology 2002;58(5):702-708.

184. van Swieten JC, Brusse E, de Graaf BM, Krieger E, van de Graaf R, de Koning I, Maat-Kievit A, Leegwater P, Dooijes D, Oostra BA, Heutink P. A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]. Am J Hum Genet 2003;72(1):191-199.

185. Velazquez Perez L, Santos Falcon N, Garcia Zaldivar R, Paneque Herrera M, Hechevarria Pupo RR. [Epidemiology of cuban hereditary ataxia]. Rev Neurol 2001;32(7):606-611.

186. Verbeek DS, Schelhaas JH, Ippel EF, Beemer FA, Pearson PL, Sinke RJ. Identification of a novel SCA locus (SCA19) in a Dutch autosomal dominant cerebellar ataxia family on chromosome region 1p21-q21. Hum Genet 2002;111(4-5):388-393.

187. Verbeek DS, van de Warrenburg BP, Wesseling P, Pearson PL, Kremer HP, Sinke RJ. Mapping of the SCA23 locus involved in autosomal dominant cerebellar ataxia to chromosome region 20p13-12.3. Brain 2004;127(Pt 11):2551-2557.

188. Vuillaume I, Devos D, Schraen-Maschke S, Dina C, Lemainque A, Vasseur F, Bocquillon G, Devos P, Kocinski C, Marzys C, Destee A, Sablonniere B. A new locus for spinocerebellar ataxia (SCA21) maps to chromosome 7p21.3-p15.1. Ann Neurol 2002;52(5):666-670.

189. Wain HM, Lush M, Ducluzeau F, Povey S. Genew: the human gene nomenclature database. Nucleic Acids Res 2002;30(1):169-171.

190. Wain HM, Lush MJ, Ducluzeau F, Khodiyar VK, Povey S. Genew: the Human Gene Nomenclature Database, 2004 updates. Nucleic Acids Res 2004;32(Database issue):D255-257.

191. Watanabe H, Tanaka F, Matsumoto M, Doyu M, Ando T, Mitsuma T, Sobue G. Frequency analysis of autosomal dominant cerebellar ataxias in Japanese patients and clinical characterization of spinocerebellar ataxia type 6. Clin Genet 1998;53(1):13-19.

192. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M,

Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420(6915):520-562.

193. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Helmberg W, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pontius JU, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2005;33(Database issue):D39-45.

194. Wiedemeyer R, Westermann F, Wittke I, Nowock J, Schwab M. Ataxin-2 promotes apoptosis of human neuroblastoma cells. Oncogene 2003;22(3):401-411.

195. Worth PF, Giunti P, Gardner-Thorpe C, Dixon PH, Davis MB, Wood NW. Autosomal dominant cerebellar ataxia type III: linkage in a large British family to a 7.6-cM region on chromosome 15q14-21.3. Am J Hum Genet 1999;65(2):420-426.

196. Wüllner U. Genes implicated in the pathogenesis of spinocerebellar ataxias. Drugs Today (Barc) 2003;39(12):927-937.

197. Xie QY, Liang XL, Li XH. [Molecular genetics and its clinical application in the diagnosis of spinocerebellar ataxias]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 2005;22(1):71-73.

198. Yabe I, Sasaki H, Chen DH, Raskind WH, Bird TD, Yamashita I, Tsuji S, Kikuchi S, Tashiro K. Spinocerebellar ataxia type 14 caused by a mutation in protein kinase C gamma. Arch Neurol 2003;60(12):1749-1751.

199. Yamashita I, Sasaki H, Yabe I, Fukazawa T, Nogoshi S, Komeichi K, Takada A, Shiraishi K, Takiyama Y, Nishizawa M, Kaneko J, Tanaka H, Tsuji S, Tashiro K. A novel locus for dominant cerebellar ataxia (SCA14) maps to a 10.2-cM interval flanked by D19S206 and D19S605 on chromosome 19q13.4-qter. Ann Neurol 2000;48(2):156-163.

200. Yu GY, Howell MJ, Roller MJ, Xie TD, Gomez CM. Spinocerebellar ataxia type 26 maps to chromosome 19p13.3 adjacent to SCA6. Ann Neurol 2005;57(3):349-354.

201. Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, Lee CC. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet 1997;15(1):62-69.

202. Zortea M, Armani M, Pastorello E, Nunez GF, Lombardi S, Tonello S, Rigoni MT, Zuliani L, Mostacciuolo ML, Gellera C, Di Donato S, Trevisan CP. Prevalence of inherited ataxias in the province of Padua, Italy. Neuroepidemiology 2004;23(6):275-280.

7. Glossar

Ähnlichkeit
Ein Maß für Übereinstimmungen und Unterschiede unabhängig von ihrer Ursache (siehe auch Homologie).

Alignment
Ein Alignment wird dadurch gebildet, dass zwei zu vergleichende Sequenzen direkt untereinander geschrieben werden, mit dem Ziel einzelne Positionen der Sequenz so anzuordnen, dass identische oder ähnliche Aminosäuren oder Nukleotide untereinander stehen. Ein Alignment ist also eine Hypothese für die positionelle Homologie zwischen Basenpaaren bzw. Aminosäuren.

Decapping (~ Entkappung)
Entfernung der Poly(A)-Struktur am 5´-Ende der mRNA. Durch diesen Deadenylationsprozeß kann die mRNA nicht mehr translatiert werden und wird deshalb auch kurz nach dem Decapping abgebaut. Einen wichtigen Schutzmechanismus stellen die Poly(A)-Bindungsproteine dar, die diesen Deadenylationsprozeß verhindern, indem sie den Poly(A)-Schwanz durch ihre Bindung daran stabilisieren.

Domäne (engl. domain)
Viele Proteine enthalten im Faltungsmuster einer einzigen Molekülkette mehrere kompakte Bereiche, bei denen es den Anschein hat, als könnten sie auch unabhängig von den restlichen Proteinabschnitten stabil sein. Solche Bereiche bezeichnet man als Domänen.

E- Wert „E-value“
Erwartungswert/ Expectation value.
Statistische Signifikanz für den gefundenen Treffer bei Datenbanksuchen. Je niedriger der jeweilige E-value ist, desto höher ist die Wahrscheinlichkeit, dass es sich nicht um einen zufälligen Treffer handelt.

Gap
Eine Lücke in einem Alignment. Sie entsteht z.B. durch Insertionen oder Deletionen in Sequenzen.

Genom
Als Genom oder auch Erbgut wird eine Gesamtheit der vererbbaren Nucleinsäure einer mehr oder weniger autonomen Struktur bezeichnet. Diese autonome Struktur kann ein Virus, eine Zelle, ein Organell oder ein Organismus sein.

Homologie
Ähnlichkeit in der DNA oder Proteinsequenz zwischen Individuen derselben Spezies oder zwischen verschiedenen Spezies. Die Ähnlichkeit lässt sich durch die Abstammung von einem gemeinsamen Ursprungsgen erklären. Die Identifizierung und Analyse von Homologien ist eine zentrale Aufgabe der Phylogenie.

Loop (engl. für Schleife)
Die allermeisten Proteine sind aus Kombinationen von regelmäßigen Sekundärstrukturelementen – α-Helixes und β-Strängen – aufgebaut, die durch Loopregionen variabler Länge und irregulärer Form verbunden sind. Viele dieser Loops befinden sich an der Proteinoberfläche. Insertionen und Deletionen in den Aminosäuresequenzen homologer Proteine treten fast ausschließlich in Loopregionen auf.

Motiv
Ein Muster oder ein Profil, das charakteristisch für die Aminosäuresequenz ist.

Multiples Sequenzalignment
Ein Alignment aus drei oder mehr Sequenzen, in welches Gaps „Lücken“ so eingefügt werden, dass ähnliche bzw. identische Abschnitte in der gleichen Spalte angeordnet werfen und so Homologien aufgedeckt werden können. Clustal W ist eines der am häufigsten genutzten multiplen Sequenzalignmentprogrammen.

Messenger RNA (mRNA)
Englische Bezeichnung für Boten-RNA. Die mRNA dient als Muster (engl. template) für die Proteinsynthese.

Offenes Leseraster (engl. „open reading frames, ORFs)
Ein ORF ist ein Abschnitt in der DNA-Sequenz, der mit einem Startcodon (ATG) beginnt und mit einem Stopcodon endet. ORFs sind also potenziell proteincodierende Abschnitte. Computerprogramme zur Genomanalyse identifizieren offene Leseraster.

Polymorphismus
Eine häufig vorkommende Variation in der DNA-Sequenz. Polymorphismen können durch Nukleotidsubstitution, Insertion (Verlängerung der ursprünglichen DNA-Sequenz), Deletion (Verkürzung) oder Mikrosatelliten definiert werden. Polymorphismen können funktionell (Lage im Exon) oder stumm sein (Lage im Intron).

Proteom
bezeichnet die Gesamtheit aller verschiedenen Proteine, die in einem Organismus auftreten (engl. proteome = PROTEin complement of the genOME). Auch posttranslational veränderte Proteine, zum Beispiel durch Glykosylierung oder Phosphorylierung, gehören zum Proteom.

Zufallsstränge „random coils“
sind statistische Knäule, die unter anderem bei denaturierten Proteinen auftreten, bei denen über Peptidbindungen verbundene Aminosäurereste in einem ungeordneten Zustand vorliegen und so die native Funktion des Polypeptids/ Proteins verhindern.

8. Anhang

Tabelle 5. Alternative PDB identifiers and corresponding SPTrEMBL accession numbers for Lsm proteins.

PDB	SPTrEMBL
1b34A	P13641
1b34B	P43330
1d3bA	P43331
1d3bB	P14678
1h64A	Q9V0Y8
1m8vA	Q9V0Y8
1hk9A	P25521
1i4kA	O29386
1i5lA	O29386
1i81A	O26745
1mgqA	O26745
1jriA	O26745
1kq1A	Q99UG9
1kq2A	Q99UG9
1ljoA	O29885
1lnxA	Q8ZYG5
1i8f	Q8ZYG5
1m5qA	Q8ZVU2
1n9rA	P54999
1n9sA	P54999

Exonstruktur von Ataxin-2

1. Exon: 1 - 244
MRSAAAAPRSPAVATESRRFAAARWPGWRSLQRPARRSGRGGGGAAPGPYPSAAPPPPGPGPPPSRQSSPPSASDCFG
SNGNGGGAFRPGSRRLLGLGGPPRPFVVVLLPLASPGAPPAAPTRASPLGARASPPRSGVSLARPAPGCPRPACEPVY
GPLTMSLKPQQQQQQQQQQQQQQQQQQQQQQPPPAAANVRKPGGSGLLASPAAAPSPSSSSVSSSSATAPSSVVAATS
GGGRPGLGRG

2. Exon: 245 - 256
RNSNKGLPQSTI

3. Exon: 257 - 276
SFDGIYANMRMVHILTSVVG

4. Exon: 277 - 300
SKCEVQVKNGGIYEGVFKTYSPKC

5. Exon: 301 - 350
DLVLDAAHEKSTESSSGPKREEIMESILFKCSDFVVVQFKDMDSSYAKRD

6. Exon: 351 - 392
AFTDSAISAKVNGEHKEKDLEPWDAGELTANEELEALENDVS

7. Exon: 393 - 423
NGWDPNDMFRYNEENYGVVSTYDSSLSSYTV

8. Exon: 424 - 489
PLERDNSEEFLKREARANQLAEEIESSAQYKARVALENDDRSEEEKYTAVQRNSSEREGHSINTRE

9. Exon: 490 - 548
NKYIPPGQRNREVISWGSGRQNSPRMGQPGSGSMPSRSTSHTSDFNPNSGSDQRVVNGG

10. Exon: 549 - 618
VPWPSPCPSPSSRPPSRYQSGPNSLPPRAATPTRPPSRPPSRPSRPPSHPSAHGSPAPVSTMPKRMSSEG

11. Exon: 619 - 679
PPRMSPKAQRHPRNHRVSAGRGSISSGLEFVSHNPPSEAATPPVARTSPSGGTWSSVVSGV

12. Exon: 680 - 745
PRLSPKTHRPRSPRQNSIGNTPSGPVLASPQAGIIPTEAVAMPIPAASPTPASPASNRAVTPSSEA

13. Exon: 746 - 781
KDSRLQDQRQNSPAGNKENIKPNETSPSFSKAENKG

14. Exon: 782 - 805
ISPVVSEHRKQIDDLKKFKNDFRL

15. Exon: 806 - 907
QPSSTSESMDQLLNKNREGEKSRDLIKDKIEPSAKDSFIENSSSNCTSGSSKPNSPSISPSILSNTEHKRGPEVTSQG
VQTSSPACKQEKDDKEEKKDAAEQ

16. Exon: 908 - 928
VRKSTLNPNAKEFNPRSFSQP

17. Exon: 929 - 979
KPSTTPTSPRPQAQPSPSMVGHQQPTPVYTQPVCFAPNMMYPVPVSPGVQP

18. Exon: 980 - 999
LYPIPMTPMPVNQAKTYRAV

19. Exon: 1000 - 1061
PNMPQQRQDQHHQSAMMHPASAAGPPIAATPPAYSTQYVAYSPQQFPNQPLVQHVPHYQSQH

20. Exon: 1062 - 1105
PHVYSPVIQGNARMMAPPTHAQPGLVSSSATQYGAHEQTHAMYA

21. Exon: 1106 - 1123
CPKLPYNKETSPSFYFAI

22. Exon: 1124 - 1172
STGSLAQQYAHPNATLHPHTPHPQPSATPTGQQQSQHGGSHPAPSPVQH

23. Exon: 1173 - 1248
HQHQAAQALHLASPQQQSAIYHAGLAPTPPSMTPASNTQSPQNSFPAAQQTVFTIHPSHVQPAYTNPPHMAHVPQA

24. Exon: 1249 - 1304
HVQSGMVPSHPTAHAPMMLMTTQPPGGPQAALAQSALQPIPVSTTAHFPYMTHPSV

25. Exon: 1305 - 1312
QAHHQQQL

9. Danksagung

Hiermit bedanke ich mich bei Prof. Dr. med. U. Wüllner für die Überlassung des interessanten Themas und die Unterstützung meiner Arbeit.
Für die Einführung in die Methodik und die fachliche Betreuung im Bereich der Bioinformatik gilt mein besonderer Dank Mario Albrecht, der Mitarbeiter am Max-Planck-Institut für Informatik in Saarbrücken ist.
Außerdem möchte ich mich bei meiner Familie und meinen Freunden bedanken, die mich im Laufe dieser Arbeit moralisch unterstützt und mit Anregungen geholfen haben.