

1 Einleitung

Die Entschlüsselung des humanen Genoms, erstmalig der Öffentlichkeit präsentiert in einer Rohfassung im Juni 2000 durch Celera Genomics (Venter et al., 2001) und den international zusammenschlossenen akademischen Gruppen (Lander et al., 2001), stellte einen Meilenstein in der Geschichte der Naturwissenschaften dar. In Analogie zu den erfolgreichen Hochdurchsatz-Sequenzierungen des humanen Genoms sowie weiterer Genome entstand die Idee der Hochdurchsatz-Strukturaufklärung aller Proteine. Weltweit etablierten sich in den letzten Jahren *structural genomics initiatives* mit dem Ziel der systematischen Hochdurchsatz-Strukturaufklärung zur Bestimmung aller Proteinfaltungen. Ein Teilschritt zur Entwicklung eines „Proteinfaltungskataloges“ ist die Aufklärung der Sequenz-Struktur-Funktions-Beziehungen durch die Charakterisierung einer weitverbreiteten Organisationsstufe von Proteinen, den Proteindomänen. Viele Proteine können als Abfolge von Domänen beschrieben werden. Man geht insgesamt von etwa 1500 Domänenfamilien aus und nimmt an, dass in absehbarer Zeit für jede Familie mindestens eine NMR- oder Kristallstruktur verfügbar sein wird (Blundell & Mizuguchi, 2000, Burley, 2000, Lan et al., 2003, Liu & Rost, 2003, Muller, 2003, Zhang & Kim, 2003).

Im Zentrum der vorliegenden Arbeit steht die Strukturaufklärung von Proteindomänen mittels NMR-Spektroskopie im Kontext von *structural and functional genomics*. Dies beinhaltet die systematische Generierung und Analyse großer Datenmengen und führte zur Entwicklung von Hochdurchsatzmethoden auf den Gebieten der Expression, Datenaufnahme, Interpretation und Strukturrechnung auf jeder Stufe der Strukturbestimmung. In dieser Arbeit wurden modernste Expressionskonzepte zur Herstellung von Domänenkonstrukten etabliert. Diese Methoden und Konzepte wurden im Zusammenhang der Strukturbestimmung von WW-, RUN- und SEP-Domänen angewendet. Weiterhin wurden NMR-Strukturen der humanen Gas7 WW-Domäne und der humanen p47 SEP-Domäne bestimmt.

1.1 Proteindomänen

Proteine werden unter anderem dadurch systematisch analysiert, indem man sie in ihre einzelnen Domänen zerlegt. Diese reduktionistische Herangehensweise ist vergleichbar mit der Generierung eines Sprachcodes unter Verwendung des Alphabetes. Einzelne Buchstaben (Domänen) bilden zusammengesetzt ein Wort (Protein), aus denen sinnvoll aneinandergesetzt ganze Sätze (Proteinkomplexe) entstehen. Somit ist es möglich,

komplexe Sachverhalte in einzelne Schritte aufzuteilen und experimentell ausgehend von der einfacheren Stufe der Proteindomäne zu analysieren (M. Gimona, EURESCO conference on Functional Protein Modules, Seefeld 2001). Erstmals beschrieb Wetlauber 1973 für globuläre Proteine distinkte Regionen, die sich unabhängig vom Restprotein in kompakte Einheiten falten (Wetlauber, 1973). Zugleich erkannte man, dass spezifische Abschnitte eines Proteins wichtig sind für die Ausführung einer spezifischen Funktion. Das Auftreten von ähnlichen funktionalen Segmenten in verschiedenen Proteinen führte zum Konzept der „*modular building blocks*“ (Kong & Ranganathan, 2004, Wetlauber, 1973).

1.1.1 Verwendung des Domänenbegriffes in der Strukturbiologie

Domänen sind die vorherrschenden strukturellen und oft auch funktionellen Einheiten eines Proteins (Copley et al., 2002). Der Begriff Domäne wird in der Strukturbiologie je nach Zusammenhang als strukturelle, funktionelle und evolutionäre Domäne verwendet, wobei Überschneidungen dieser Unterbegriffe möglich sind.

Unter einer strukturellen Domäne versteht man eine kompakte, sich unabhängig faltende Organisationseinheit eines Proteins, welche typischerweise einen hydrophoben Kern enthält (Kyte, 1995). In der hierarchischen Proteinstrukturklassifikation lassen sich Domänen zwischen der Supersekundär- und Tertiärstrukturebene einordnen. Sie sind in der Regel 40-200 Aminosäuren groß und autonom, da sie unabhängig vom Proteinkontext in isolierter Form ihre Struktur, Funktion und spezifischen Eigenschaften wie beispielsweise Ligandenbindung beibehalten. Aufgrund dieser Eigenschaften sind sie hervorragende Modellobjekte zur Untersuchung komplexer biologischer Sachverhalte. Proteindomänen sind auf der Spielwiese der Evolution im übertragenen Sinne vergleichbar mit „LEGO-Bausteinen“. Durch Transfer, Neukombination oder Aneinanderreihung von Modulen lassen sich neue Strukturen und damit erweiterte oder neue Funktionen und Eigenschaften kreieren (Aasland et al., 2002, Copley et al., 2002, Doolittle & Bork, 1993, Pawson, 1995). Beispielsweise besteht das humane Protein Pin1 aus einer katalytischen Peptidylprolyl-cis-trans-isomerase-Domäne (PPIase) und einer nicht-katalytischen WW-Domäne, die spezifische Protein-Protein-Interaktionen vermittelt (Verdecia et al., 2000). Die Variabilität und Flexibilität in der Verwendung konservierter Domänen ermöglicht es, komplexe biologische Prozesse mit einem effizientem ökonomischen Repertoire an Proteinfaltungen zu steuern.

1.1.2 Direkte und indirekte Verfahren zur Vorhersage von Domänen

Ohne die Kenntnis der dreidimensionalen Struktur ist es nicht möglich, die Domänengrenzen eindeutig zu bestimmen. Neue, nicht strukturell klassifizierte Domänen werden zunächst aufgrund theoretischer Sequenzhomologievergleiche identifiziert und später durch ihre Strukturbestimmung nachgewiesen. Die Resultate der Sequenzhomologievergleiche werden zur Zeit hauptsächlich durch die beiden sequenzbasierten Datenbanken SMART (*a Simple Modular Architecture Research Tool*, (Schultz et al., 1998) und Pfam (Bateman et al., 2000, Bateman et al., 2004) öffentlich zugänglich gemacht. Sie enthalten eine große Anzahl an strukturell und funktionell nicht aufgeklärten putativen Proteindomänen. Derzeitig sind in der SMART-Datenbank (Version 4.0, Stand 01/05; (Letunic et al., 2004)) mehr als 600 Proteindomänenfamilien annotiert. Während die Sequenzhomologievergleiche nur kontinuierliche Bereiche hervorheben, die durch einen hohen Konservierungsgrad gekennzeichnet sind, kann die eigentliche Domäne erheblich größer sein und sich auch über einen diskontinuierlichen Sequenzabschnitt erstrecken. Für die thermodynamische Stabilität und korrekte Faltung können flankierende, weniger gut konservierte Bereiche wichtig sein, als auch das Vorliegen von Fusionspartnern und essentiellen Kofaktoren, wie beispielsweise Zink-Ionen. Zur Stabilität trägt weiter die Ausbildung von Disulfidbrücken bei (Muller, 2003). Neben der Möglichkeit, Domänengrenzen aufgrund von Sequenzhomologievergleichen direkt vorherzusagen, gibt es indirekte Methoden zur Eingrenzung von Domänen durch die Bestimmung von *Linker*-Sequenzen zwischen zwei Domänen. Dazu gehört zum Beispiel die LinkerDB Datenbank, (George & Heringa, 2002). Bisher wurden Domänen-*Linker* zum Großteil als unstrukturierte, flexible Elemente betrachtet. Weitere Erkenntnisse zeigen jedoch, dass Linker die Kooperativität und Interaktion von benachbarten Domänen mit beeinflussen und insgesamt die Struktur und Funktionalität von Multidomänenproteinen modulieren können (Kong & Ranganathan, 2004).

Zusammenfassend lässt sich sagen, dass es bis heute keine verlässliche ab initio Methode gibt, die Domänengrenzen exakt vorherzusagen. Zur Aufklärung von neuen, unbekanntem Domänen ist es daher unerlässlich, auf experimentellem Wege mehrere unterschiedliche Domänenkonstrukte herzustellen, die dann mittels NMR-Spektroskopie oder Röntgenkristallographie weiter untersucht werden.

1.2 Vermittlung von Protein-Protein-Interaktionen durch nicht-katalytische Domänen

Biologische Prozesse in der Zelle sind komplexe, dynamische Ereignisse, die über ein Netzwerk aus unterschiedlichen Komponenten, bestehend aus Proteinen, Lipiden und Nukleinsäuren, reguliert werden. Intrazelluläre Proteine wechselwirken zum Teil über konservierte Mechanismen, die durch spezialisierte Domänen, z.B. nicht-katalytische Domänen vermittelt werden. Ein Verständnis der Struktur, Funktion und Spezifität von nicht-katalytischen Domänen ist damit essentiell, um insbesondere die komplexe Funktionsweise von Proteinen aus dem Bereich der Signaltransduktion zu verstehen. Über komplementäre Regionen (Epitope) zwischen den Interaktionspartnern werden sowohl zeitlich als auch räumlich präzise funktionelle Proteinkomplexe gebildet. Für eine Reihe von modularen nicht-katalytischen Domänen wurden die Erkennungsmechanismen in den letzten Jahren detailliert beschrieben. Sie basieren meist darauf, dass kurze, lineare oder räumlich benachbarte, durch konservierte Reste geprägte Kernmotive des Zielproteins von einer Domäne erkannt werden. Bekannte Vertreter sind zum Beispiel SH3-, EVH1-, GYF- und WW-Domänen, die prolinreiche Sequenzen erkennen und dadurch spezifische Protein-Protein-Kontakte vermitteln (Freund et al., 1999, Zarrinpar et al., 2003). Andere Erkennungsmechanismen finden sich beispielsweise bei PDZ-Domänen, die spezifisch den C-Terminus ihres Interaktionspartners identifizieren (van Ham & Hendriks, 2003). Diese Kernmotive können durch flankierende Schlüsselreste weiter verfeinert und spezifischer gestaltet werden, wodurch sich innerhalb einer Domänenfamilie weitere Bindungsspezifitäten ausbilden können.

1.2.1 WW-Domäne

Die WW-Domäne, benannt nach einem hochkonservierten Tryptophan-Paar, ist ein Proteinmodul mit einer Größe von 35-40 Aminosäuren und damit die bisher kleinste gefundene stabile Domäne. Abbildung 1.1 zeigt den Prototyp einer WW-Domäne (Macias et al., 2000). Strukturell bildet die WW-Domäne ein kompaktes, antiparalleles dreisträngiges β -Faltblatt, welches sich ohne Kofaktoren und ohne Ausbildung eines echten hydrophoben Kernes stabilisiert. Alle bisher bestimmten NMR- und Kristallstrukturen der WW-Domäne zeigen die gleiche übereinstimmende Faltung (Huang et al., 2000, Kanelis et al., 2001, Macias et al., 1996, Pires et al., 2005, Pires et al., 2001, Ranganathan et al., 1997). Die namensgebenden Tryptophanreste W17 und W39 (zur Notation der Nummerierung siehe Abbildung 1.1/Abbildung 1.2; Seite 5/7) sind in der Aminosäuresequenz 20-22 Reste voneinander entfernt und für die Stabilität, Faltung und

Funktion der Domäne entscheidend verantwortlich. Stabilitäts- und Strukturanalysen ergaben, dass die Reste W17, P42 und F/Y29 einen hydrophoben Cluster auf der Unterseite des β -Faltblattes bilden, der für die Stabilität der Domäne wichtig ist (Macias et al., 2000). Während das Tryptophan W17 maßgeblich zur Stabilität der Domäne beiträgt und somit ein strukturell konservierter Rest ist, ist das zweite Tryptophan W39 ein funktionell konservierter Rest. Es bildet zusammen mit weiteren hydrophoben Resten oberhalb des Faltblattes des zweiten β -Stranges eine exponierte hydrophobe Tasche, die spezifisch prolinreiche Sequenzen erkennt.

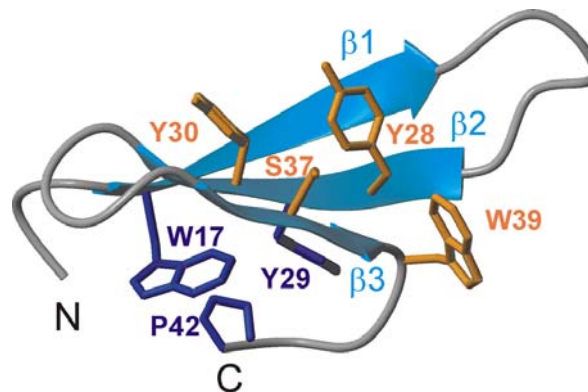


Abbildung 1.1: Bänderdarstellung des WW-Domänen Prototyps (PDB: 1EOM) nach Macias et al. (2000). Die WW-Domäne bildet ein antiparalleles dreisträngiges β -Faltblatt (hellblau). Die konservierten Reste W17, Y29 und P42 (blau) bilden unterhalb des Faltblattes einen hydrophoben Cluster und tragen zur Stabilität der WW-Domäne bei. Die konservierten Reste Y28, Y30, S37 und W39 (gelb) vermitteln spezifische Interaktionen zu prolinreichen Sequenzen.

Komplexstrukturen zeigten, dass eine spezifische Bindungstasche durch die aromatischen Reste W39 und Y28 gebildet wird, welche vornehmlich zwei konsekutiv aufeinanderfolgende Prolinreste aufnimmt. Diese Proline liegen in einer PolyProlin Typ II (PPII)-Helix-Konformation vor.

Die linksgängige PPII-Helix ist ein entscheidendes, für die Interaktion wichtiges ausgestrecktes Sekundärstrukturmotiv, welches durch eine Abfolge von mehr als 4 Prolinen gebildet wird, wobei auch Nicht-Prolin-Reste toleriert werden. Sie enthält drei Reste pro Windung und ist durch die Absättigung von Carbonylgruppen des Peptidrückgrates durch Wassermoleküle gut solvatisiert. Durch die flache, hydrophobe und exponierte Oberfläche der Prolinreste ist die PPII-Helix in der Lage parallel, von der Waals Kontakte zu aromatischen Ringen des Bindungspartners einzugehen (*stacking*). Über die in Wasser zuvor abgesättigten Carbonylgruppen der PPII-Helix können Wasserstoffbrückenbindungen zum indolischen NH des konservierten Tryptophanrestes

W39 der WW-Domäne ausgebildet werden (Adzhubei & Sternberg, 1993, Kelly et al., 2001).

So entstehen hochkomplementäre Bindungstaschen zwischen der Helixoberfläche und den konservierten aromatischen Resten der WW-Domäne. Die Bindung prolinreicher Liganden durch WW-Domänen sind durch geringe reversible, jedoch spezifische Affinitäten gekennzeichnet. Wechselwirkungen von Signaltransduktions-Proteinen erfordern moderate aber hochspezifische Affinitäten zur Aufrechterhaltung des Informationsflusses. Sehr starke Affinitäten würden zu einer Blockade des Systems führen. Die bestimmten Dissoziationskonstanten liegen im μ -molaren Bereich (Kato et al., 2004, Pires et al., 2001) und verdeutlichen den transienten, also durch einen schnellen Wechsel von Assoziation und Dissoziation des Liganden, geprägten Charakter der WW-Domänen-Interaktionen.

WW-Domänen werden mit den molekularen Ursachen verschiedener Krankheiten in Zusammenhang gebracht, wie z.B. dem Rett-Syndrom, der Duchenne-Muskeldystrophie, der Alzheimer'schen Krankheit sowie der Huntington's-Krankheit und Krebs (Buschdorf & Stratling, 2004, Faber et al., 1998, Russo et al., 1998). In dieser Arbeit wurden weitere Erkenntnisse über die molekularen Erkennungsmechanismen der WW-Domäne aus dem humanen Gas 7 (*Growth-arrest-specific-protein /hGas7*) Protein gewonnen, die als Grundlage für das weitere Verständnis der Ursachen und der Bekämpfung der aufgeführten Krankheiten dienen. Das *hGas7* Protein wird vornehmlich im Gehirn exprimiert und ist wichtig für die Bildung von neuritischen Zellen (Chang et al., 2005). Es enthält eine WW-Domäne und wird mit der Krebserkrankung *acute myeloid leukemia* in Verbindung gebracht Fusionspartner des Onkoproteins MLL (mixed lineage leukemia) zu sein (So et al., 2003).

1.2.2 Klassifizierung von WW-Domänen aufgrund ihrer spezifischen Bindungsmotive

WW-Domänen erkennen prolinreiche Abschnitte ihrer Interaktionspartner. Anhand ihrer Bindungspräferenzen lassen sich WW-Domänen in 4-6 Klassen einteilen, welche z.T. kontrovers diskutiert werden (Hu et al., 2004, Otte et al., 2003, Sudol & Hunter, 2000).

		15	20	25	30	35	40
Klasse I	Rn NEDD4-3	PPG	WEERTHTD	GRV	FFINHN	IKKT	QWEDPRM
	Hs Dystrophin	QGP	WERAI	SPN	KVP	YYIN	HETQTT
	Hs YAP65	PAG	WEMAKTSS	GQR	YFLNH	IDQTT	WQDPRK
Klasse II	Hs FBP11-1	KSM	WTEHKSPD	GRT	YYNT	TETK	QSTW
	Mm FBP28	VSE	WTEYKTAD	GKT	YYNN	RTE	LESTW
	Hs FE65	PAG	WMRVQDTS	G	TY	WHIPT	GTTQW
	Hs GAS7	PPG	WQSYLSPQ	GRR	YYVNT	TTNET	TWERPSS
Klasse III	Hs FBP21-1	KGR	WVEGITSE	GYH	YYDL	LISGAS	QW
	Hs FBP21-2	KTW	VEGLSED	GFT	YYNT	TETGES	SRW
	Hs HYP109-1	MGD	WQEVWD	ENTGSY	YYWNT	QTNEV	TWELPQY
Klasse IV	Hs PIN1	PPG	WEKRMS	RSSGRV	YYFN	HITNAS	QW
	Sc ESS1	PTW	TVRYS	KSKKRE	YFF	NEETKHS	QW

Abbildung 1.2: repräsentative Auswahl von WW-Domänensequenzen klassifiziert nach ihren Bindungspartnern. Klasse I: PPxY-Motiv (blau); Klasse II: PPLPp-Motiv (magenta) und PolyP-Motiv Gas 7 (schwarz); Klasse III: PPR-Motiv (orange) und Klasse IV: poS/poTP-Motiv (grün). Konservierte Reste sind in schwarz und spezifitäts-vermittelnde Aminosäurereste farbig hervorgehoben. Exponierte Reste, oberhalb des β -Faltblattes (in Abbildung 1.1), die Teil der potentiellen Interaktionsfläche zwischen Ligand und WW-Domäne darstellen, sind grau unterlegt. Schwarz eingerahmt sind die β -Stränge. Die NMR-Komplexstruktur der *hGas7* WW-Domäne mit dem prolinreichen Liganden LIPPPPPPL wurde in dieser Arbeit bestimmt. Rn (*Rattus norvegicus*), Hs (*Homo sapiens*), Mm (*Mus musculus*), Sc (*Saccharomyces cerevisiae*). SwissProt-Zugangsnummern von oben: P46934, P11532, P46937, O75400, Q8CGF7, O00213, O60861, O75554, Q9NT61, Q13526 und P226696.

In dieser Arbeit wurde die Klassifizierung der WW-Domänen nach Sudol und Hunter (2000) bevorzugt (siehe Abbildung 1.2). Diese Untersuchung teilt sie in 4 Gruppen auf. Die Arbeit von Otte und Wiedemann (2003) unterteilt Klasse II und Klasse III zusätzlich in jeweils zwei weitere Unterklassen. Für einige der gefundenen Gruppen wurden die molekularen Zusammenhänge zwischen Sequenz und Bindungsspezifität der WW-Domänen durch die Aufklärung von Komplexstrukturen und der Ermittlung von Bindungskonstanten belegt. So wurde die molekulare Basis der WW-Domänen-Erkennungsspezifität für die Klasse I (Y-Gruppe) durch die Arbeiten an den WW-Domänen aus den Proteinen *hYAP*, *Dystrophin* und *Nedd4* (Huang et al., 2000, Kanelis et al., 2001, Macias et al., 1996, Pires et al., 2001) geklärt. Für die Klasse IV (*poS/poT*-Gruppe) sind ebenso molekulare Daten durch die Komplexstrukturbestimmung verfügbar (Verdecia et al., 2000). In der kürzlich erschienenen Arbeit (Pires et al., 2005) wurde die Bindungsspezifität für einen Vertreter der Klasse II, *FBP11-WW1*, mit dem Liganden

APPTPPPLPP strukturell analysiert. Die Basis der Bindungsspezifität für Klasse III (PPR-Gruppe) ist nicht eindeutig verstanden und auf der molekularen Ebene bisher nicht beschrieben worden.

Für die Poly-P-Gruppe aus der Klasse II zeigten die Arbeiten von Otte (2003), dass das Bindungsmotiv für die *hGas7* WW-Domäne durch mindestens 5 aufeinanderfolgende Proline bestimmt ist (siehe Substitutionsanalyse in Abbildung 1.3), denen ein weiteres Prolin oder ein hydrophober Rest vorangeht. Diese Arbeit zeigte, dass keines dieser 5 Proline durch einen anderen Rest ausgetauscht werden konnte.

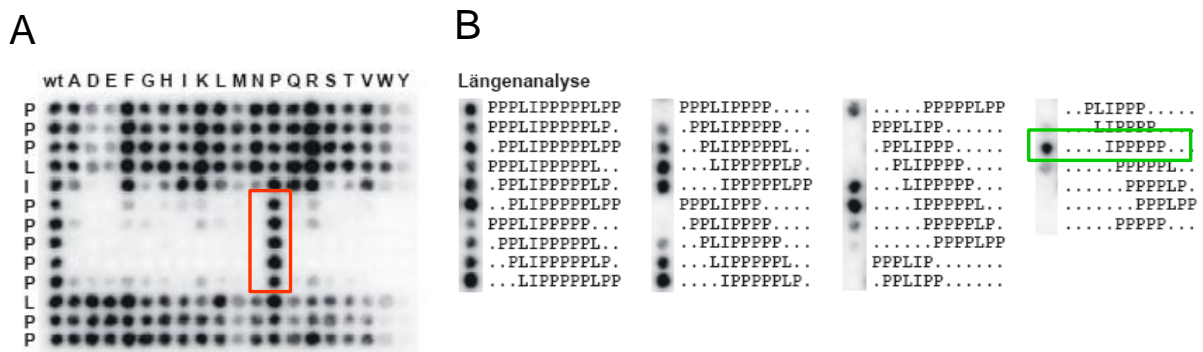


Abbildung 1.3: Substitutionsanalyse (A) und Längenanalysen (B) des prolinreichen Peptides PPPLIPPPPLPP (linker Rand) mit der *hGas7*-WW Domäne. Das Bindungsmotiv des Liganden ist durch 5 aufeinanderfolgende Proline bestimmt (rot eingerahmt). Die Längenanalyse des Liganden ergab die minimal bindende Sequenz IPPPPP (grün eingerahmt). Längere prolinreiche Peptide zeigen zum Teil eine höhere oder gleichstarke Affinität zur *hGas7* WW Domäne, ersichtlich an der stärkeren Intensität der Schwarzfärbung (Otte, 2003).

Diese Reihung von 5 Prolinen bildet mit hoher Wahrscheinlichkeit eine PPII-Helix aus, wobei aber nicht alle Proline mit der Oberfläche der Domänen in Kontakt stehen können. Vielmehr ist zu vermuten, dass die Ausbildung einer PPII-Helixkonformation für die Interaktion essentiell ist. Im Rahmen dieser Arbeit konnte die molekulare Bindungsspezifität für das Erkennungsmotiv der Poly-P-Gruppe der Klasse II durch die Aufklärung des Strukturkomplexes aus der *hGas 7* WW-Domäne und dem prolinreichen Liganden (LIPPPPL) geklärt werden.

1.3 Strukturell und funktionell nicht klassifizierte Domänen

Die Strukturanalyse von unbekanntem Domänen aus humanen Proteinen erfordert die Herstellung zahlreicher Expressionklone zur Identifizierung geeigneter löslicher Konstrukte für die weitere Strukturbestimmung. In diesem Rahmen wurden modernste Expressionsmethoden zur Herstellung von geeigneten Domänenkonstrukten für die strukturelle und funktionelle Klassifizierung getestet. Diese Methoden sind automatisierbar und hochdurchsatzfähig. Sie lassen sich in ein fabrikähnliches Konzept zur Hochdurchsatz-Strukturaufklärung, wie das der Proteinstrukturfabrik Berlin, integrieren. In dieser Arbeit wurde das Gateway[®]-System (Invitrogen) zur Herstellung von Proteindomänenkonstrukten am Beispiel der humanen RUN Domäne etabliert.

In Kooperation mit der Proteinstrukturfabrik Berlin wurden 88 unterschiedliche Expressionskonstrukte von 17 verschiedenen humanen Proteindomänen hergestellt (Scheich et al., 2004) und auf ihre Faltung mit der 1D ¹H-NMR Spektroskopie untersucht. Von den 17 Proteindomänen ergaben sich insgesamt zwei Proteindomänen als geeignete Kandidaten für die NMR-Strukturanalyse. Dies waren die strukturell bestimmte DCX-Domäne (Kim et al., 2003) und die SEP-Domäne aus dem humanen Protein p47.

1.3.1 RUN-Domäne

Die RUN-Domäne ist eine bisher nicht strukturell bestimmte Proteindomäne, benannt nach den Proteinen **R**PIP8 (*RaP2 interacting protein 8*), **U**NC-14 und **N**ESCA (*new molecule containing SH3 at the carboxyl-terminus*). Ihre Domänengrenzen werden in den Datenbanken SMART und Pfam mit stark unterschiedlicher Größe angegeben. Während in der SMART-Datenbank die vorraussichtliche Größe für die RUN-Domäne mit etwa 60 Aminosäureresten vorhergesagt wird, wird in der Pfam-Datenbank für die Run-Domäne eine deutlich größere Länge von etwa 130 Aminosäuren angezeigt. In der Abbildung 1.4 ist die Größenvorhersage nach SMART und Pfam für die RUN-Domäne des humanen Proteins RUFY1 gezeigt. Das C-terminale Ende der Domäne bei Rest 161 (Smart) bzw. 163 (Pfam) wird von beiden Datenbanken ähnlich vorhergesagt, während die Pfam-Datenbank die N-terminale Domänengrenze um 70 Reste erweitert angibt. Beide genannten Datenbanken detektieren die RUN-Domäne in etwa 130 eukaryontischen Proteinen mit wechselnder Domänenarchitektur (Stand 03/2005).

Protein: RUFY1

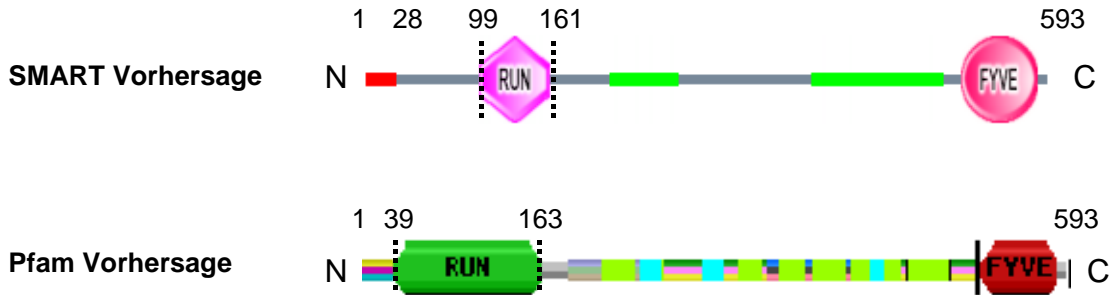


Abbildung 1.4: Unterschiedliche RUN-Domänen Größenvorhersage für das humane Protein RUFY1 nach SMART (oben) und Pfam (unten). In der SMART-Datenbank wird die RUN-Domäne in den Grenzen 99-161 und in der Pfam-Datenbank mit der Größe 39-163 angegeben. Der Proteinabschnitt 1-28 wird in der SMART-Datenbank als Signalpeptid (rot markiert) vorhergesagt.

Callebaut et al. (2001) finden in der RUN-Domäne, auf Grundlage eines manuell erstellten Sequenzhomologie-Alignments, sechs konservierte Blöcke (A-F). Die Verbindungssequenzen dieser Blöcke tolerieren unterschiedlich große Insertionen mit teilweise mehr als 40 Aminosäuren Länge. Die Sekundärstrukturvorhersage geht von einem überwiegend α -helikalen Faltungstyp aus. Bemerkenswert ist die Tatsache, dass Block A und D konservierte basische Reste enthalten, wie es bei RasGAP- und RhoGAP-Domänen der Fall ist. Dies liegt den Schluss nahe, dass die RUN-Domäne auch eine katalytische Funktion im Zusammenspiel mit GTPasen (Callebaut et al., 2001, Rittinger et al., 1997, Scheffzek et al., 1997) haben könnte.

Die RUN-Domäne findet sich gehäuft in Proteinen wieder, die mit GTPasen der Rap- und Rab-Familie assoziiert sind. GTPasen spielen eine entscheidende Rolle in der Signaltransduktion und kontrollieren wichtige biologische Funktionen wie Zellproliferation und Differenzierung. Für das N-terminale Fragment (1-298) des RPIP8 Proteins, welches eine RUN-Domäne beinhaltet, konnte gezeigt werden, dass es spezifisch mit Rap2 und GTP interagiert (Janoueix-Lerosey et al., 1998). Das gemeinsame Auftreten von RUN- und TBC-Domänen in zahlreichen Proteinen lässt den Schluss zu, dass RUN-Domänen spezifische Funktionen im Rab-Signaltransduktionsweg erfüllen. TBC-Domänen sind GTPase Aktivatoren von Rab ähnlichen Proteinen (Albert et al., 1999). Bisher gibt es allerdings keine Erkenntnisse, dass RUN-Domänen mit Proteinen der Rab-Familie direkt interagieren (Callebaut et al., 2001).

1.3.2 SEP-Domäne

Ebenso wie die RUN-Domäne ist die SEP-Domäne (*Saccharomyces cerevisiae* **S**hp1, *Drosophila melanogaster* eyes closed gene (**e**yc) and *vertebrate* **p**47) nicht näher klassifiziert, sowie strukturell und funktionell unbeschrieben. Sie wird in der SMART-Datenbank mit einer durchschnittlichen Größe von 100 Aminosäuren für gegenwärtig 32 eukaryontische Proteine vorhergesagt. Die namensgebenden Proteine Shp1 und Eyes closed gene sind homologe Proteine des humanen Adaptorproteins p47. Bemerkenswerterweise tritt die Domänenkombination SEP-UBX gehäuft auf. Diese Domänen sind dabei über einen konstanten *Linker*-Abschnitt miteinander verknüpft.

1.3.2.1 Modularer Aufbau des Adaptorproteins p47

Wie die meisten Adaptorproteine ist der 370 Aminosäure große eukaryontische Kofaktor p47 modular aus bisher drei annotierten Domänen aufgebaut. Am N-Terminus befindet sich die strukturell gelöste und 45 Reste umfassende UBA-(**U**biquitin-**a**ssociated domain) Domäne. Daran anschließend folgt über einen nichtstrukturierten, etwa 130 Reste langen Proteinabschnitt die bisher unklassifizierte SEP-Domäne. Am C-Terminus befindet sich die strukturell gelöste UBX-(**U**biquitin **r**egulatory **X** domain) Domäne.

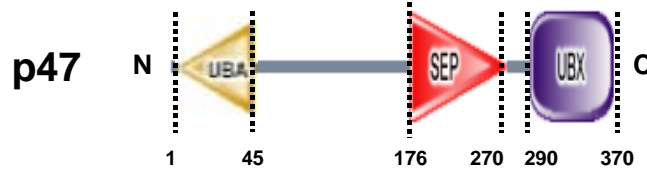


Abbildung 1.5: Domänenarchitektur des p47 Adaptorproteins (modifiziert nach SMART)

Die UBA-Domäne ist ein dreigliedriges, kompaktes Helixbündel (Dieckmann et al., 1998) und funktionell in der Ubiquitinierung, der Signaltransduktion, in Mechanismen der DNA-Reparatur und im Vesikeltransport involviert (Buchberger, 2002, Buchberger et al., 2001). Die UBA-Domäne bindet an Ubiquitin und Multi-Ubiquitinketten, während sie im Komplex p97/p47 bevorzugt Mono-Ubiquitin bindet und außerdem eine weitere Anlagerung von Ubiquitin mit der Ausbildung einer Poly-Ubiquitinkette verhindert (Withers-Ward et al., 2000).

Die C-Terminale UBX-Domäne besteht aus etwa 80 Aminosäureresten und nimmt eine Ubiquitin-ähnliche Faltung ein, gekennzeichnet durch ein gemischtes 5-strängiges β -Faltblatt und zwei α -Helices in einer $\beta\beta\alpha\beta\beta\alpha\beta$ Sekundärstruktur-Topologie (Buchberger et al., 2001, Yuan et al., 2001). Für die UBX-Domäne wurde noch keine generelle Funktion

beschrieben. Es wird aber angenommen, dass sie in mit Ubiquitin verknüpften Regulationsmechanismen eine Rolle spielt (Buchberger, 2002). Für die p47 UBX-Domäne wurde eine Interaktion mit dem Protein p97 festgestellt (Kondo et al., 1997). Die Kristallstruktur des Komplexes aus dem Hexamer p97 und Monomer p47 beschreibt den strukturellen Mechanismus der Bindung. Dabei bindet der S3/S4 Loop der p47 UBX-Domäne an eine hydrophobe Tasche, die durch die beiden N-terminalen Subdomänen von p97 (N-Subdomänen) gebildet wird.

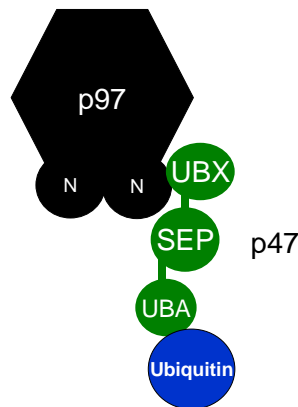


Abbildung 1.6: Schematischer Überblick der Domänenkomposition des p47 Adaptorproteins (grün) mit den drei Domänen UBA, SEP und UBX hier im Komplex mit den jeweiligen Interaktionspartnern p97 (schwarz) und Ubiquitin (blau).

Interessanterweise wurden keine einfachen Bindungsverhältnisse von p47 an die 6 potentiellen N-Subdomänentaschen des p97 Hexamers festgestellt. In der Wissenschaft wird das stöchiometrische Verhältnis des Komplexes p47/p97 von 6/3 bzw. 6/6 diskutiert (persönliche Kommunikation mit Prof. S. Matthews). Das Adaptorprotein p47 wird in Zusammenhang mit einer regulatorischen Funktion bei der Membranfusion gebracht, insbesondere während der postmitotischen Restrukturierung des Golgi-Apparates. Es soll außerdem weiterhin an proteolytischen Ubiquitin-abhängigen Prozessen beteiligt sein. Möglicherweise auch am Transport von Substraten zum 26 S Proteasom (Dreveny et al., 2004, Uchiyama et al., 2002, Uchiyama et al., 2003).

1.3.2.2 Biologischer Kontext p47/p97

Das durch p47 gebundene p97 Protein gehört zu der Superfamilie der AAA-ATPasen (*ATPasen associated with various cellular activities*; (Ogura & Wilkinson, 2001)) und setzt sich zusammen aus den ATPasen-Domänen D1/D2 (auch bezeichnet als AAA-1/AAA-2-Domänen) sowie einer N-Terminalen variablen N-Subdomäne, die hauptsächlich der Substrat- und/oder Adaptorerkennung dient. Zu dieser Superfamilie gehören weiter die

Clp- und Hsp-Proteinfamilien (Dougan et al., 2002) und das Hefehomologe Cdc48 Protein. Die ubiquitär verbreiteten AAA-ATPasen und ihre Interaktionspartner formen Proteinnetzwerke, die an verschiedenen zellulären Prozessen beteiligt sind, wie an der Zellteilung, Membranfusion, Transkription, Proteindegradation, ATP-abhängigen Entfaltung/Auflösung von Protein-Protein- und Protein-DNA-Komplexen. Weitere assoziierte Moleküle sind Chaperone und regulatorische Komponenten aus zellulären proteolytischen Maschinerien. Entsprechend spielen die AAA-ATPasen eine lebenswichtige zellerhaltende Rolle unter normalen Wachstumsbedingungen und auch unter Stressbedingungen (Dougan et al., 2002). Ein Funktionsausfall von p97 führt zu einer Akkumulation von ubiquitinierten Proteinen durch die Stilllegung des Ubiquitin vermittelten Proteinabbaus (Dai & Li, 2001).

1.4 NMR-Spektroskopie

1.4.1 Prinzipien der NMR-Spektroskopie

In der hochauflösenden NMR-Spektroskopie von Proteinen sind die Atomkerne ^1H , ^{13}C und ^{15}N am bedeutendsten. Sie besitzen eine Spinquantenzahl $I_S = \frac{1}{2}$ und weisen daher kein Quadrupolmoment auf. In einem statischen Magnetfeld präzedieren die Kernspins mit dem magnetischen Moment μ um die z-Achse. Folglich ist μ_z nach Gleichung 1.1,

$$\mu_z = \gamma I_z = \gamma \hbar \Delta m \quad (1.1)$$

mit $\hbar = \frac{h}{2\pi}$ ein Vielfaches des Planck'schen Wirkungsquantum, wobei die magnetische Quantenzahl Δm Werte von $+1$ und -1 annimmt. Kerne mit Spin $\frac{1}{2}$ nehmen zwei Spinzustände entweder parallel ($+\frac{1}{2}$ oder α) oder antiparallel ($-\frac{1}{2}$ oder β) zum äußeren Magnetfeld an. Diese Zeeman Aufspaltung ist von der jeweiligen Stärke des angelegten Magnetfeldes B_0 und dem gyromagnetischem Verhältnis γ des betrachteten Kernes (Abbildung 1.7) abhängig.

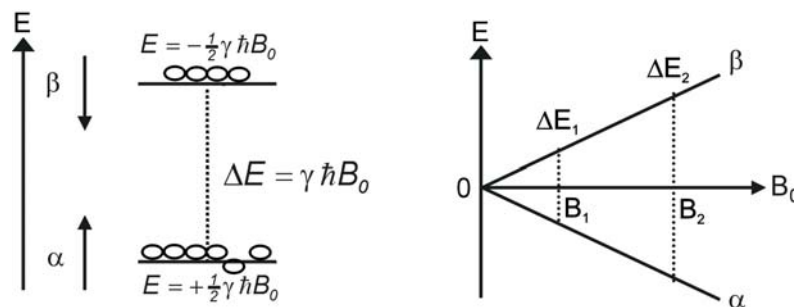


Abbildung 1.7: (links) Aufspaltung der Energieniveaus eines Kernspins ($i = \frac{1}{2}$) im Magnetfeld, (rechts) Abhängigkeit der Energiedifferenz der aufgespaltenen Zustände ΔE von der Magnetfeldstärke B_0

Durch elektromagnetische Strahlung, z.B. Radiowellen, wird ein Übergang zwischen den beiden Energiezuständen angeregt. Um einen Übergang zwischen den beiden Energieniveaus herbeizuführen, wird elektromagnetische Strahlung mit der Energie $E = h \nu_0$ eingestrahlt. Die Frequenz, bei der diese Resonanzbedingung erfüllt ist, wird als Larmor-Frequenz bezeichnet.

$$\omega_0 = -\gamma \times B_0 \quad (1.2)$$

Man beobachtet für ^1H bei einer Magnetfeldstärke von 14,1 Tesla eine Resonanzfrequenz von 600 MHz.

Die Population N des α - und β -Zustandes kann mit Hilfe der Boltzmann Verteilung

$$N_{\beta} / N_{\alpha} = \exp^{(-\hbar B_0 / 2 \pi kT)} \quad (1.3)$$

beschrieben werden. Die Populationsdifferenz ist bei einer Resonanzfrequenz von 600 MHz verhältnismäßig klein ($N(\beta/\alpha) = 0.999904$) und der Grund für die relative Unempfindlichkeit der Methode. Das Signal-Rausch Verhältnis S/R (siehe Gleichung 1.4) kann durch Verwendung von höheren Magnetfeldstärken, konzentrierteren Proben, Verwendung eines Kryoprobenkopfes und durch Erhöhung der Scananzahl verbessert werden:

$$S / R \approx N \gamma_I \gamma_D^{(3/2)} B_0^{(3/2)} (NS) T_2^{(1/2)} \quad (1.4)$$

- S/R: Signal- Rausch-Verhältnis
- N: Anzahl der Spins (Probenkonzentration)
- γ_I : gyromagnetisches Verhältnis des angeregten Kernes
- γ_D : gyromagnetisches Verhältnis des detektierten Kernes
- B_0 : externes Magnetfeld
- NS: Anzahl der Scans
- T_2 : transversale Relaxationszeit

1.4.1.1 Chemische Verschiebung

Um Resonanzfrequenzen untereinander vergleichen zu können, werden Sie nicht absolut bestimmt, sondern mit einem internen Standard ν_{ref} abgeglichen und in ppm (parts per million) angegeben. Als „Nullpunkt“ für Protonen wird bei Proteinlösungen das Methylsignal von DSS (2,2-Dimethyl-2-silapentan-5-sulfonsäure) als Standard verwendet. Die chemische Verschiebung δ ist wie folgt definiert:

$$\delta = 10^6 (\nu - \nu_{ref}) / \nu_{ref} \quad (1.5)$$

Sie ist ein grundlegender Parameter für die Resonanzzuordnung des Makromoleküls. Die chemische Verschiebung hängt von der chemischen Umgebung des Kernes ab. Elektronen rufen lokale Magnetfelder hervor und schirmen den Kern vom angelegten Magnetfeld ab. Somit können strukturelle Unterschiede in der Umgebung des Kernes

unterschiedliche Larmor-Frequenzen für Kerne gleicher Art bedingen. Dadurch ist es letztendlich möglich, eine vollständige Resonanzzuordnung aller chemisch nicht äquivalenten Kerne eines Moleküls zu erhalten.

1.4.1.2 Skalare Kopplungen und Dipol-Dipol-Wechselwirkungen

Die magnetischen Dipole der Kerne können indirekt über chemische Bindungen und direkt über dipolare Wechselwirkungen durch den Raum miteinander interagieren. Skalare Kopplungen (J-Kopplung) beobachtet man in der Regel zwischen Kernen, die ein (¹J), zwei (²J) oder (³J) drei Bindungen voneinander separiert sind. Sie bilden die Basis für COSY- (*correlation spectroscopy*) und TOCSY- (*total correlation spectroscopy*)-Experimente. Für die Strukturbestimmung ist die Bestimmung der vicinalen ³J_{H^NH_α - Kopplungskonstante interessant, da Ihr Wert vom Torsionswinkel zwischen H^N und H_α abhängt. Durch die semiempirische Karplus-Beziehung ist es möglich, Torsionswinkel aus experimentell ermittelten ³J_{H^NH_α-Kopplungskonstanten zu erhalten. Dipol-Dipol-Wechselwirkungen zwischen Kernen durch den Raum erzeugen den für die Strukturbestimmung wichtigsten Effekt, den Kern Overhauser Effekt (*Nuclear Overhauser effect*/NOE). Die Intensität des NOEs korreliert mit der inversen sechsten Potenz des Kern-Kern-Abstandes r und einer Korrelationsfunktion f (τ_c), die die Modulation der Dipol-Dipol-Wechselwirkung durch stochastische Prozesse mit einer effektiven Korrelationszeit τ_c beschreibt .}}

$$NOE \propto \frac{1}{r^6} f(\tau_c) \quad (1.6)$$

Für die Strukturermittlung ist das ¹H-¹H-NOE Signal von großer Bedeutung, da Proteine über ein dichtes Netzwerk an Wasserstoffatomen verfügen.

In der NMR-Spektroskopie wird die Molekülstruktur u.a. durch einen Satz gefundener ¹H-¹H-Abstandslimits ermittelt. Diese Protonenabstände werden durch Umformung des experimentellen NOE-Peakvolumens nach Gleichung 1.6 in einen definierten Abstand r unter Annahme, dass für alle ¹H-¹H-Kombinationen die Korrelationsfunktion f (τ_c) den gleichen Wert annimmt, erhalten. Unter Zuhilfenahme bekannter Abstände aus Sekundärstrukturelementen werden, die durch den Raum gefundenen unbekanntes Abstände kalibriert (Lottspeich, 1998).

1.4.2 Klassische und moderne Verfahren zur Strukturbestimmung von Proteinen

Für die Resonanzzuordnung des *hGas7* WW-Domänenkomplexes mit dem Liganden LIPPPPPL wurden die von Will Gibbons und Kurt Wüthrich maßgeblich entwickelten klassischen Standardverfahren zur sequenzspezifischen Signalzuordnung verwendet (Gibbons et al., 1972, Wuthrich, 2003, Wüthrich, 1986). Durch die Annotation charakteristischer Spinsysteme der Aminosäuren wurde versucht, alle konformationsrelevanten Parameter zu erhalten. Dies erfolgte durch die Aufnahme von klassischen 2D- ^1H -NMR Experimenten wie ROESY (*Rotating frame Overhauser effect spectroscopy*), COSY, TOCSY und NOESY (*Nuclear Overhauser effect spectroscopy*) Experimenten.

Mit der Möglichkeit der Herstellung rekombinanter stabiler $^{13}\text{C}/^{15}\text{N}$ -markierter Proteine wurden die Methoden der klassischen 2D NMR-Resonanzzuordnung erweitert. Man verwendet heute im wesentlichen 3D [$^1\text{H}, ^{13}\text{C}, ^{15}\text{N}$]-Tripelresonanz- und ^{13}C oder ^{15}N editierte 3D [$^1\text{H}, ^1\text{H}$]-NOESY-Experimente. Die manuelle Resonanzzuordnung der humanen p47 SEP-Domäne erfolgte unter Verwendung dieser Standardmethoden (Bax & Grzesiek, 1993, Clore & Gronenborn, 1991, Grzesiek & Bax, 1993, Montelione & Wagner, 1990, Sattler et al., 1999).

In den letzten Jahren gibt es große Bestrebungen, die bisher manuell durchgeführten Einzelschritte der Resonanzzuordnung und die sich anschließende Strukturrechnung weitestgehend zu automatisieren. In diesem Zusammenhang wurden in dieser Arbeit neue Methoden und Experimente verwendet, die prinzipiell einen automatischen Prozeß der Datenakquisition, deren Auswertung und eine automatische Strukturrechnung ermöglichen. Dazu wurde ein Standardkanon, bestehend aus einem minimal notwendigen Satz an NMR-Experimenten mit Hilfe von PASTE/PAPST (Labudde, 2002), aufgenommen. Damit ist es möglich, die Datenaufnahme ganzer NMR-Sätze für alle erforderlichen Parameter und Experimente einheitlich zu gestalten.

Um eine Automatisierung der Datenauswertung zu erreichen, wurden kürzlich Aminosäure-selektive Experimente entwickelt, um die konventionelle Rückgrat-Zuordnung zu vereinfachen. Diese sogenannten 2D (i+1) und (i,i+1)-HSQC Experimente (Schmieder et al., 1998, Schubert et al., 2000, Schubert et al., 2001, 2001, 2001, Schubert et al., 1999) sind Tripelresonanzexperimente, welche ^1H - ^{15}N Korrelationen ergeben. Sie enthalten zusätzlich die Möglichkeit, eindeutige ^1H - ^{15}N Zuordnungen durch die geeignete Kombination von Experimenten zu treffen.

1.5 Aufgabenstellung

Die Strukturaufklärung von medizinisch relevanten humanen Proteinen und ihren Domänen, die in Protein-Protein-Interaktionen involviert sind, kann dabei vor allem einem Verständnis ihrer Funktion im allgemeinen und ihrer Spezifität im besonderen dienen. Das Hauptziel dieser Arbeit bestand darin, diese Domänen in ihrer dreidimensionalen Struktur mit Hilfe der NMR-Spektroskopie aufzuklären, „neue *folds*“ zu finden, und darauf aufbauend Struktur-Funktions-Beziehungen durch Untersuchung relevanter Komplexe zu entwickeln. Für bislang wenig charakterisierte Domänen sollten die gefundenen strukturbasierten Hypothesen zur Funktion anhand von geeigneten Methoden zur Untersuchung von Protein-Protein-Interaktionen näher beschrieben und überprüft werden. Dazu sollten in dieser Arbeit systematische automatisierbare Expressionsverfahren zur Herstellung von geeigneten Konstrukten von unbekanntem Domänen etabliert werden. Diese Verfahren sollten sich in ein „fabrikähnliches Konzept“ zur automatischen Strukturbestimmung von Proteindomänen integrieren können. Am Beispiel der unbekanntem RUN- und SEP-Domänen sollten diese Verfahren in Zusammenarbeit mit der Proteinstrukturfabrik Berlin eingeführt werden. Für die weitere Strukturaufklärung sollten lösliche, stabile und markierte Domänen exprimiert und gereinigt werden.

Das Domänenprojekt zur Bestimmung von WW-Domänen-Ligand-Komplexen dieser Arbeit hatte das Ziel, die Erkennung prolinreicher Sequenzen durch WW-Domänen der Klasse II (Poly-Prolin-Gruppe) am Beispiel der *hGas7* WW-Domäne im Komplex mit dem spezifischen Peptidliganden LIPPPPPL durch die Bestimmung der NMR-Komplexstruktur zu charakterisieren. Das *hGas7* Protein enthält eine WW-Domäne und ist ein Fusionspartner des Onkoproteins MLL (*mixed lineage leukemia*). Es wird daher mit den molekularen Ursachen der Krebserkrankung *acute myeloid leukemia* im Zusammenhang gesehen. Die Klärung der strukturellen Zusammenhänge zwischen der Sequenz- und Bindungsspezifität der *hGas7* WW-Domäne und des prolinreichen Liganden LIPPPPPL soll dabei als Ausgangspunkt für die weitere Entwicklung von peptidischen Inhibitoren dienen.

1.6 Referenzen

- Aasland, R., Abrams, C., et al. (2002) Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett* 513, 141-144
- Adzhubei, A. A. & Sternberg, M. J. (1993) Left-handed polyproline II helices commonly occur in globular proteins. *J Mol Biol* 229, 472-493
- Albert, S., Will, E. & Gallwitz, D. (1999) Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J* 18, 5216-5225
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000) The Pfam protein families database. *Nucleic Acids Res* 28, 263-266
- Bateman, A., Coin, L., et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32, D138-141
- Bax, A. & Grzesiek, S. (1993) Methodological Advances in Protein NMR. *Accounts of Chemical Research* 26, 131-138
- Blundell, T. L. & Mizuguchi, K. (2000) Structural genomics: an overview. *Prog Biophys Mol Biol* 73, 289-295
- Buchberger, A. (2002) From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell Biol* 12, 216-221
- Buchberger, A., Howard, M. J., Proctor, M. & Bycroft, M. (2001) The UBX domain: a widespread ubiquitin-like module. *J Mol Biol* 307, 17-24
- Burley, S. K. (2000) An overview of structural genomics. *Nat Struct Biol* 7 Suppl, 932-934
- Buschdorf, J. P. & Stratling, W. H. (2004) A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. *J Mol Med* 82, 135-143
- Callebaut, I., de Gunzburg, J., Goud, B. & Mornon, J. P. (2001) RUN domains: a new family of domains involved in Ras-like GTPase signaling. *Trends Biochem Sci* 26, 79-83
- Chang, P. Y., Kuo, J. T., Lin-Chao, S. & Chao, C. C. (2005) Identification of rat Gas7 isoforms differentially expressed in brain and regulated following kainate-induced neuronal injury. *J Neurosci Res* 79, 788-797
- Clore, G. M. & Gronenborn, A. M. (1991) Applications of 3-Dimensional and 4-Dimensional Heteronuclear NMR-Spectroscopy to Protein-Structure Determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* 23, 43-92
- Copley, R. R., Doerks, T., Letunic, I. & Bork, P. (2002) Protein domain analysis in the era of complete genomes. *FEBS Lett* 513, 129-134

Dai, R. M. & Li, C. C. (2001) Valosin-containing protein is a multi-ubiquitin chain-targeting factor required in ubiquitin-proteasome degradation. *Nat Cell Biol* 3, 740-744

Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S. Y. & Feigon, J. (1998) Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nature Structural Biology* 5, 1042-1047

Doolittle, R. F. & Bork, P. (1993) Evolutionarily mobile modules in proteins. *Sci Am* 269, 50-56

Dougan, D. A., Mogk, A., Zeth, K., Turgay, K. & Bukau, B. (2002) AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett* 529, 6-10

Dreveny, I., Kondo, H., Uchiyama, K., Shaw, A., Zhang, X. & Freemont, P. S. (2004) Structural basis of the interaction between the AAA ATPase p97/VCP and its adaptor protein p47. *Embo J* 23, 1030-1039

Faber, P. W., Barnes, G. T., Srinidhi, J., Chen, J., Gusella, J. F. & MacDonald, M. E. (1998) Huntingtin interacts with a family of WW domain proteins. *Hum Mol Genet* 7, 1463-1474

Freund, C., Dotsch, V., Nishizawa, K., Reinherz, E. L. & Wagner, G. (1999) The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nat Struct Biol* 6, 656-660

George, R. A. & Heringa, J. (2002) An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng* 15, 871-879

Gibbons, W. A., Alms, H., Bockman, R. S. & Wyssbrod, H. R. (1972) Homonuclear indor spectroscopy as a means of simplifying and analyzing proton magnetic resonance spectra of peptides and as a basis for determining secondary and tertiary conformations of complex peptides. *Biochemistry* 11, 1721-1725

Grzesiek, S. & Bax, A. (1993) Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins. *Journal of Biomolecular Nmr* 3, 185-204

Hu, H., Columbus, J., et al. (2004) A map of WW domain family interactions. *Proteomics* 4, 643-655

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M. & Eck, M. J. (2000) Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nat Struct Biol* 7, 634-638

Janoueix-Lerosey, I., Pasheva, E., de Tand, M. F., Tavitian, A. & de Gunzburg, J. (1998) Identification of a specific effector of the small GTP-binding protein Rap2. *Eur J Biochem* 252, 290-298

Kanelis, V., Rotin, D. & Forman-Kay, J. D. (2001) Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat Struct Biol* 8, 407-412

- Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J. J., Sudol, M. & Tanokura, M. (2004) Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J Biol Chem* 279, 31833-31841
- Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F. & Creamer, T. P. (2001) Host-guest study of left-handed polyproline II helix formation. *Biochemistry* 40, 14376-14383
- Kim, M. H., Cierpicki, T., et al. (2003) The DCX-domain tandems of doublecortin and doublecortin-like kinase. *Nat Struct Biol* 10, 324-333
- Kondo, H., Rabouille, C., Newman, R., Levine, T. P., Pappin, D., Freemont, P. & Warren, G. (1997) p47 is a cofactor for p97-mediated membrane fusion. *Nature* 388, 75-78
- Kong, L. & Ranganathan, S. (2004) Delineation of modular proteins: domain boundary prediction from sequence information. *Brief Bioinform* 5, 179-192
- Kyte, J. (1995) *Structure in protein chemistry*, Garland Publishing, New York, USA
- Labudde, D., Leitner, D., Schmieder, P., Oschkinat, H. (2002) Bruker Report 150. In pp. 8-11
- Lan, N., Montelione, G. T. & Gerstein, M. (2003) Ontologies for proteomics: towards a systematic definition of structure and function that scales to the genome level. *Curr Opin Chem Biol* 7, 44-54
- Lander, E. S., Linton, L. M., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921
- Letunic, I., Copley, R. R., et al. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 32 Database issue, D142-144
- Liu, J. & Rost, B. (2003) Domains, motifs and clusters in the protein universe. *Curr Opin Chem Biol* 7, 5-11
- Lottspeich, F., Zorbas, H. (1998) *Bioanalytik*, Spektrum akademischer Verlag, Heidelberg, Berlin
- Macias, M. J., Gervais, V., Civera, C. & Oschkinat, H. (2000) Structural analysis of WW domains and design of a WW prototype. *Nat Struct Biol* 7, 375-379
- Macias, M. J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. & Oschkinat, H. (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* 382, 646-649
- Montelione, G. T. & Wagner, G. (1990) Conformation-Independent Sequential Nmr Connections in Isotope-Enriched Polypeptides by H-1-C-13-N-15 Triple-Resonance Experiments. *Journal of Magnetic Resonance* 87, 183-188
- Muller, Y. H., U. (2003) *Molekülmodelle und Modellmoleküle: Strukturanalyse großer biologischer Moleküle für die Medizin* Vol. 2. Auflage, Springer-Verlag, Berlin Heidelberg

- Ogura, T. & Wilkinson, A. J. (2001) AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* 6, 575-597
- Otte, L. (2003) Charakterisierung von WW-Domänen-Ligand-Wechselwirkungen mit Hilfe synthetischer Protein- und Peptidbibliotheken. In Mathematisch-Naturwissenschaftliche Fakultät I, Humboldt-Universität zu Berlin, Berlin
- Otte, L., Wiedemann, U., et al. (2003) WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci* 12, 491-500
- Pawson, T. (1995) Protein modules and signalling networks. *Nature* 373, 573-580
- Pires, J. R., Parthier, C., et al. (2005) Structural basis for APPTPPPLPP peptide recognition by the FBP11WW1 domain. *J Mol Biol* 348, 399-408
- Pires, J. R., Taha-Nejad, F., et al. (2001) Solution structures of the YAP65 WW domain and the variant L30 K in complex with the peptides GTPPPPYTVG, N-(n-octyl)-GPPPY and PLPPY and the application of peptide libraries reveal a minimal binding epitope. *J Mol Biol* 314, 1147-1156
- Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. (1997) Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* 89, 875-886
- Rittinger, K., Walker, P. A., et al. (1997) Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP. *Nature* 388, 693-697
- Russo, T., Faraonio, R., Minopoli, G., De Candia, P., De Renzis, S. & Zambrano, N. (1998) Fe65 and the protein network centered around the cytosolic domain of the Alzheimer's beta-amyloid precursor protein. *FEBS Lett* 434, 1-7
- Sattler, M., Schleucher, J. & Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34, 93-158
- Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F. & Wittinghofer, A. (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* 277, 333-338
- Scheich, C., Leitner, D., et al. (2004) Fast identification of folded human protein domains expressed in E. coli suitable for structural analysis. *BMC Struct Biol* 4, 4
- Schmieder, P., Leidert, M., Kelly, M. & Oschkinat, H. (1998) Multiplicity-selective coherence transfer steps for the design of amino acid-selective experiments - A triple-resonance experiment selective for ash and gin. *Journal of Magnetic Resonance* 131, 199-202
- Schubert, M., Ball, L. J., Oschkinat, H. & Schmieder, P. (2000) Bridging the gap: A set of selective H-1-N-15-correlations to link sequential neighbors of prolines. *Journal of Biomolecular NMR* 17, 331-335

- Schubert, M., Oschkinat, H. & Schmieder, P. (2001) Amino acid type-selective backbone H-1-N-15-correlations for Arg and Lys. *Journal of Biomolecular NMR* 20, 379-384
- Schubert, M., Oschkinat, H. & Schmieder, P. (2001) MUSIC and aromatic residues: Amino acid type-selective H-1-N-15 correlations, III. *Journal of Magnetic Resonance* 153, 186-192
- Schubert, M., Oschkinat, H. & Schmieder, P. (2001) MUSIC, selective pulses, and tuned delays: Amino acid type-selective H-1-N-15 correlations, II. *Journal of Magnetic Resonance* 148, 61-72
- Schubert, M., Smalla, M., Schmieder, P. & Oschkinat, H. (1999) MUSIC in triple-resonance experiments: Amino acid type-selective H-1-N-15 correlations. *Journal of Magnetic Resonance* 141, 34-43
- Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95, 5857-5864
- So, C. W., Karsunky, H., Passegue, E., Cozzio, A., Weissman, I. L. & Cleary, M. L. (2003) MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* 3, 161-171
- Sudol, M. & Hunter, T. (2000) NeW wrinkles for an old domain. *Cell* 103, 1001-1004
- Uchiyama, K., Jokitalo, E., et al. (2002) VCIP135, a novel essential factor for p97/p47-mediated membrane fusion, is required for Golgi and ER assembly in vivo. *J Cell Biol* 159, 855-866
- Uchiyama, K., Jokitalo, E., et al. (2003) The localization and phosphorylation of p47 are important for Golgi disassembly-assembly during the cell cycle. *J Cell Biol* 161, 1067-1079
- van Ham, M. & Hendriks, W. (2003) PDZ domains-glue and guide. *Mol Biol Rep* 30, 69-82
- Venter, J. C., Adams, M. D., et al. (2001) The sequence of the human genome. *Science* 291, 1304-1351
- Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol* 7, 639-643
- Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 70, 697-701
- Withers-Ward, E. S., Mueller, T. D., Chen, I. S. Y. & Feigon, J. (2000) Biochemical and structural analysis of the interaction between the UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr. *Biochemistry* 39, 14103-14112

Wuthrich, K. (2003) NMR studies of structure and function of biological macromolecules (Nobel Lecture). *Angewandte Chemie-International Edition* 42, 3340-3363

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York

Yuan, X., Shaw, A., Zhang, X., Kondo, H., Lally, J., Freemont, P. S. & Matthews, S. (2001) Solution structure and interaction surface of the C-terminal domain from p47: a major p97-cofactor involved in SNARE disassembly. *J Mol Biol* 311, 255-263

Zarrinpar, A., Bhattacharyya, R. P. & Lim, W. A. (2003) The structure and function of proline recognition domains. *Sci STKE* 2003, RE8

Zhang, C. & Kim, S. H. (2003) Overview of structural genomics: from structure to function. *Curr Opin Chem Biol* 7, 28-32