

Kapitel 2

Material und Methoden

2.1 Nomenklatur-Regeln

2.1.1 Nomenklatur der Proteine und Domänen

Proteine werden durch ihren UniProt¹⁰ *Entry Name* oder den gängigen Proteinnamen benannt. Diesem Proteinnamen wird ein Präfix vorangestellt, der den Organismus bezeichnet, aus dem das Protein stammt: *h* = Mensch (*Homo sapiens*), *m* = Maus (*Mus musculus*), *y* = Bäckerhefe (*Saccharomyces cerevisiae*), *tb* = eukaryotischer Erreger der Schlafkrankheit (*Trypanosoma brucei*), *a* = Acker-Schmalwand (*Arabidopsis thaliana*), *d* = Fruchtfliege (*Drosophila melanogaster*), *c* = Fadenwurm (*Caenorhabditis elegans*), *r* = Ratte (*Rattus norvegicus*). Beispiele: *hYAP65*, *hRSP5*.

Domänen innerhalb eines Proteins werden durch den Proteinnamen und ein Suffix zur Bezeichnung der Domänenfamilie (-WW bzw. -PDZ), sowie der sequentiellen Nummer der Domänen im Falle mehrerer Domänen einer Familie benannt (-WW2 bzw. -PDZ3). Beispiele: *hYAP65-WW* (einzige WW-Domäne aus *hYAP65*), *hRSP5-WW3* (dritte WW-Domäne aus *hRSP5*).

2.1.2 Nomenklatur der homologen Aminosäurereste in Protein-Interaktionsdomänen/Liganden-Komplexen

Für die vergleichende Analyse der Vertreter einer Protein-Interaktionsdomänenfamilie wird eine einheitliche Nomenklatur auf der Basis der homologen strukturellen Positionen der Aminosäurereste, die diese in den konservierten Raumstrukturen einnehmen, verwendet. Für konservierte Bindungsmechanismen wird dieses Nomenklatur-Prinzip auch auf die Reste des Liganden angewandt. Dafür werden die Ligandenreste nach den homologen virtuellen Positionen benannt, die sie im Komplex mit der Domäne einnehmen. Zur Bezeichnung eines

spezifischen Aminosäuretyps wird der einheitlichen Positionsbezeichnung zusätzlich der IUPAC Ein-Buchstaben-Aminosäure-Code vorangestellt. Alternative Aminosäuretypen werden durch einen Schrägstrich getrennt.

2.1.2.1 Nomenklatur der WW-Domänen

Die konservierten strukturellen Positionen der Aminosäurereste von WW-Domänen werden analog zu *hYAP65*-WW (die erste entdeckte WW-Domäne)³⁰ durchnummeriert und durch ein vorangestelltes ω als Positionen der WW-Domäne gekennzeichnet (siehe Einleitung, Tabelle 1-2 und Abbildung 1-2).^{130,157} Insertionen im Vergleich zur Sequenz von *hYAP65*-WW (z.B. zwischen $\omega24$ und $\omega25$) werden mit der Nummer des vorangehenden Rests und einem Insertions-Suffix (b, c, d, ...) durchnummeriert (z.B. $\omega24b$).

Die konservierten virtuellen Ligandenpositionen λ werden in Relation zur konservierten Prolin-Bindungsstelle der WW-Domänen benannt: der Ligandenrest, dessen Seitenkette sich an die Seitenkette von $W\omega39$ innerhalb der konservierten Bindungsstelle anlagert, wird $\lambda0$ genannt (siehe Einleitung, Abbildung 1-2). Die Position des benachbarten Ligandenrests, der vor allem Y/F $\omega28$ kontaktiert, wird mit $\lambda1$ bezeichnet. Alle weiteren Reste werden entsprechend, unabhängig von der N \rightarrow C-terminalen bzw. C \rightarrow N-terminalen Orientierung des Liganden, durchnummeriert. Die beiden Ligandenorientierungen N \rightarrow C bzw. C \rightarrow N entsprechen dabei einer etwa parallelen bzw. antiparallelen Orientierung des Liganden im Verhältnis zum ersten β -Strang der WW-Domäne.

Die β -Stränge der WW-Domänen werden entsprechend ihrer N- nach C-terminalen Abfolge βA -, βB - und βC -Strang genannt. Die Schleifen zwischen diesen β -Strängen werden entsprechend als $\beta A\beta B$ - bzw. $\beta B\beta C$ -Schleife bezeichnet (siehe Einleitung, Abbildung 1-2).

2.1.2.2 Nomenklatur der PDZ-Domänen

Die konservierten strukturellen Positionen der Aminosäurereste von PDZ-Domänen werden an Hand ihrer Position relativ zum Beginn des vorausgehenden konservierten Sekundärstrukturelements bezeichnet, z.B. $\alpha B8$ für den 8. Rest nach Beginn der zweiten α -Helix αB (modifiziert nach Songyang *et al.*, 1997, siehe Einleitung, Tabelle 1-4 und Abbildung 1-4). Hochkonservierte Sequenzmotive, wie z.B. die GLGF-Schleife,⁴⁵ werden explizit bezeichnet (z.B. GLGF2 für den zweiten Rest der GLGF-Schleife). Die Reste in den Schleifen zwischen zwei Sekundärstrukturelementen werden vereinfachend zum vorausgehenden Sekundärstrukturelement gezählt.

Die konservierten virtuellen Ligandenpositionen λ werden in Relation zur C-terminalen Aminosäure (λ_0) des Liganden bzw. dem dazu homologen Rest des Liganden durchnummeriert (N-terminal benachbarten Reste heißen entsprechend $\lambda-1$, $\lambda-2$, usw.; siehe Einleitung, Abbildung 1-4).

Die konservierten Sekundärstrukturelemente der PDZ-Domänen werden entsprechend ihrer N- nach C-terminalen Abfolge bezeichnet: die 6 β -Stränge βA bis βF und die zwei α -Helices αA und αB . Der Beginn und das Ende des Sekundärstrukturelemente-Konsensus wurde an Hand der Klassifizierung mittels DSSP⁸⁶ für alle bekannten PDZ-Domänen-Strukturen ermittelt (siehe Einleitung, Tabelle 1-4 und Abbildung 1-4). Die Schleifen zwischen den Sekundärstrukturelementen werden entsprechen der beiden benachbarten Sekundärstrukturelemente benannt (z.B. $\beta B\beta C$ -Schleife zwischen βB und βC).

2.1.3 Nomenklatur der Sequenzmotive

Die Nomenklatur der Sequenzmotive folgt der modifizierten Seefeld-Convention-2001-Nomenklatur.¹ Alle L-Aminosäuren werden durch ihren IUPAC Ein-Buchstaben-Code repräsentiert. **x** steht stellvertretend für alle proteinogenen L-Aminosäuren, Φ für hydrophobe, Ψ für aliphatische und π für kleine Aminosäuren. Im Falle der festphasegebundenen Peptidbibliotheken steht **b** für ein Gemisch aus 17 L-Aminosäuren (alle proteinogenen Aminosäuren außer C, M und W). Weniger stark konservierte Aminosäuren werden durch Kleinbuchstaben repräsentiert. Alternative Aminosäuren stehen in runden Klammern in der Reihenfolge absteigender Präferenz durch Schrägstriche getrennt (z.B. **(V/I/L)**). Ausgeschlossene Aminosäuren werden durch ein vorangestelltes „^“ gekennzeichnet. Phosphorylierte Aminosäurederivate werden durch das **po**-Präfix gekennzeichnet (z.B. **poY** für Phosphotyrosin). Abweichend von der Seefeld-Convention-2001-Nomenklatur wird auf die *fn/ fc*-Nomenklatur verzichtet. Stattdessen wird eine freie C-terminale Carboxylgruppe durch ein **COOH**-Suffix gekennzeichnet. Wiederholungen einer Aminosäure werden durch einen tiefgestellten Index bezeichnet (z.B. **x₂₋₃** für **xx** bzw. **xxx**).

2.2 Herstellung und Reinigung der Protein-Interaktionsdomänen und Peptidliganden

Die Synthese, Expression und Reinigung der Domänen sowie der Liganden wurde im Rahmen zweier Kooperationsprojekte von Dr. Livia Otte (WW-Domänen) und Dr. Prisca Boisguerin (PDZ-Domänen) aus der Molecular Libraries and Recognition Group des Instituts für Medizinische Immunologie, Universitätsklinikum Charité, durchgeführt.

2.2.1 Synthese und Reinigung der WW-Domänen und der prolinreichen Liganden

Die WW-Domänen (Tabelle 2-1) und die prolinreichen Peptide (Tabelle 2-2) wurden mittels Festphasen-Peptid-Synthese nach R.B. Merrifield¹¹⁶ und Standard Fmoc-Protokoll in einem AMS 422 Peptidsyntheseautomaten (Abimed, Langenfeld, Deutschland) hergestellt. Alle Domänen und Peptide wurden über Reverse-Phase-Hochleistungsflüssigkeitschromatographie (RP-HPLC) aufgereinigt und durch Massenspektrometrie auf Reinheit und korrekte Masse hin untersucht.

Tabelle 2-1 Synthetisierte WW-Domänen: Herkunft, Position im Gesamtprotein und Sequenz

Domäne ^{a)} Aminosäure-Sequenz der synthetisierten Domäne ^{e)}	UP-ID ^{b)}	Protein ^{c)}	Beginn ^{d)}	Ende ^{d)}
<i>h</i> CA150-WW1 CJ-GTPALPPTTEEIIVWENKTPDGKVYYNARTRESAWTKPDGVKV	O14776	<i>Putative transcription factor CA150</i>	126	168
<i>h</i> CA150-WW3 CJ-PVATAPIPGTPWSVVWTGDERVFFYNPTTRLSMWDPRDDLIGR	O14776	<i>Putative transcription factor CA150</i>	523	564
<i>h</i> DYSTRO-WW CJ-QHFLSTSVQGPWERAI SPNKVPYYINHETQTTCDWHPKMTELY	P11532	<i>Dystrophin</i>	3050	3092
<i>h</i> FBP11-WW1 CJ-AAGTASGAKSMWTEHKSPDGRTYYYNTEKQSTWEKPDLLKTP	O75400	<i>Huntingtin-interacting protein HYPA/FBP11</i>	110	152
<i>h</i> FBP11-WW2 CJ-TPAEQLLSKSPWKEYKSDSGKPYYYNSQTKESRWAKPKELEDL	O75400	<i>Huntingtin-interacting protein HYPA/FBP11</i>	151	193
<i>h</i> FBP21-WW1 CJ-KKRKDPKGRWVEGITSEGYHYYYDLISGASQWEKPEGFQGD	O75554	<i>Formin binding protein 21</i>	117	159
<i>h</i> FBP21-WW2 CJ-GDLKKTAVKTVWVEGLSEDGFTYYNTEKTSRWEKPDDEFIPH	O75554	<i>Formin binding protein 21</i>	158	200
<i>h</i> FE65-WW CJ-AFETDSDLPAGWMRVQDTSGTYYWHIPTGTTQWEPGRASPSQ	O00213	<i>FE65 protein</i>	248	290
<i>h</i> GAS7-WW CJ-EESQTVILPPGWQSYLSFQGRYYVNTTTNETTWERPSSSPGI	O60861	<i>Growth-arrest-specific protein 7</i>	8	50
<i>h</i> HYP109-WW1 CJ-SLAGVGIEMGDWQEVWIDENTGSYYWNTQTNEVTWELPQYLATQ	Q9NT81	<i>Hypothetical 109.9 kDa protein</i>	205	248
<i>h</i> HYP109-WW2 CJ-KQYEINATPKGWSSHWDRDHRRYFYVNEQSGESQWEPDGEEDD	Q9NT81	<i>Hypothetical 109.9 kDa protein</i>	586	629
<i>h</i> IQGAP1-WW CJ-KKLAVGDNNKSWVKHWVKGYYYYHNLETQEGGWDEPPNFVQN	P46940	<i>Ras GTPase-activating-like protein IQGAP1</i>	674	716
<i>m</i> ITCHY-WW1 CJ-NTVSQAPLPPGWQRVDQHRVYVDHVEKRTTWDREPEPLPPG	O54971	<i>E3 Ubiquitin protein ligase ITCHY*</i>	305	346
<i>m</i> ITCHY-WW3 CJ-EFDPLGLPLPPGWKERTDSNGRVYFVNHNTRIQWEDPRSQGL	O54971	<i>E3 Ubiquitin protein ligase ITCHY*</i>	384	426
<i>h</i> KIAA1052-WW CJ-REGIVAPLPGWEKPSQDITGDIYYFNFANGQSMWDHPSPDEHYR	Q9UPV0	<i>KIAA1052 protein</i>	51	93

Domäne ^{a)}	UP-ID ^{b)}	Protein ^{c)}	Beginn ^{d)}	Ende ^{d)}
Aminosäure-Sequenz der synthetisierten Domäne ^{e)}				
hNEDD4-WW1	P46934	<i>E3 Ubiquitin protein ligase NEDD4</i>	213	255
CJ-QQEPSPPLPPGWEERQDILGRTYVYNHESRRTQWKRPDQDNL				
hNEDD4-WW2	P46934	<i>E3 Ubiquitin protein ligase NEDD4</i>	370	412
CJ-LLPTSSGLPPGWEEKQDERGRSYYVDHNSRRTTWTWKPTVQATV				
hNEDD4-WW3	P46934	<i>E3 Ubiquitin protein ligase NEDD4</i>	443	485
CJ-SEIEQGFPLPGWEVRHAPNGRPFIDHNTKTTTWEDPRLKIPA				
hNEDD4-WW4	P46934	<i>E3 Ubiquitin protein ligase NEDD4</i>	495	537
CJ-TSNDLGLPPGWEERTHTDGRIIFYINHNIKRTQWEDPRLNVA				
hPEPP2-WW1	Q9HAU0	<i>Phosphoinositol 3-phosphate-binding protein-2</i>	5	47
CJ-LNLEWISLPRSWTYGIRGGRVFFINEEAKSTTWLHPVTGEAV				
hPEPP2-WW2	Q9HAU0	<i>Phosphoinositol 3-phosphate-binding protein-2</i>	51	93
CJ-HRRQSTDLPTGWEEAYTFEGARYYINHNERKVTSSKHPVTGQPS				
hPIN1-WW	Q13526	<i>Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1</i>	1	43
CJ-MADEEKLPPGWKMRSSGRVYFNHITNASQWERPSGNSSS				
hPQBPI-WW	O60828	<i>Polyglutamine binding protein 1</i>	41	84
CJ-EATRLEGLPPSWYKVFDPSSGLPYWNADTLVSWLSPHPDPSNV				
hSMURF1-WW1	Q9UJT8	<i>E3 Ubiquitin protein ligase SMURF1</i>	220	262
CJ-HGHQSPLEPEGYEQRTTVQGQVYFLHTQTGVSTWHDPRIPRDL				
hSMURF1-WW2	Q9UJT8	<i>E3 Ubiquitin protein ligase SMURF1</i>	266	308
CJ-NSDELGLPLPGWEVRSTVSGRIYFVDHNNRRTQFTDPRLLHHIM				
hSMURF2-WW1	Q9H260	<i>E3 Ubiquitin protein ligase SMURF2</i>	152	194
CJ-SRLFNDLDPGWEEERTASGRIQYLNHITRTTQWERPTRPASE				
hWWOX-WW1	Q9NZC7	<i>WW domain-containing oxidoreductase</i>	11	53
CJ-DTSEDELPPGWEERTTKDGVVYANHTTEKTQWEHPKTGKRK				
hWWOX-WW2	Q9NZC7	<i>WW domain-containing oxidoreductase</i>	52	94
CJ-RKRVAGDLPYGWEQETDENGQVFFVDHINKRRTTYLDPRLAFT				
hWWP3-WW1	O00309	<i>Membrane associated guanylate kinase WWP3</i>	144	186
CJ-AEDNLGLPENWEMAYTENGEVYFIDHNTKTTSWLDRSLNKQ				
hWWP3-WW2	O00309	<i>Membrane associated guanylate kinase WWP3</i>	205	247
CJ-ELDSELELPAGWEKIEDPVYGIYYVDHINRKTQYENPVLEAKR				
hYAP65-WW	P46937	<i>65 kDa YES-associated protein</i>	166	208
CJ-EIPDDVPLPAGWEMAKTSSGQRYFLNHIDQTTTWDQPRKAMLS				
yESS1-WW	P22696	<i>Processing/termination factor 1 (ESS1)</i>	24	66
CJ-DVASRTGLPTPWTVRYSKSKKREYFFNPETKHSQWEEPEGTNGK				
yPRP40-WW1	P33203	<i>Pre-mRNA processing protein PRP40</i>	1	35
CJ-MSIWKEAKDASGRIYYNTLTKKSTWEKPKELISQ				
yPRP40-WW2	P33203	<i>Pre-mRNA processing protein PRP40</i>	34	76
CJ-SQEELLLRENGWKAARTADGKVVYYNPTTRETSTWTIPAFEKKV				
yRSP5-WW1	P39940	<i>E3 Ubiquitin protein ligase RSP5</i>	224	266
CJ-FEDQYGRLLPPGWERRTDNFGRTYYVDHNRTRTTWKRPTLDQTE				
yRSP5-WW2	P39940	<i>E3 Ubiquitin protein ligase RSP5</i>	325	368
CJ-TTSGLGELPSGWQRFTEPGRAYFVDHNRTRTTWVDPRRQQYI				
yRSP5-WW3	P39940	<i>E3 Ubiquitin protein ligase RSP5</i>	382	424
CJ-PVSQLGLPSGWEMRLTNTARVYFVDHNTKTTTWDDPRLPSSL				
ySSM4-WW	P40318	<i>SSM4 protein</i>	770	811
CJ-PEAYKPTSIISWKFNILLTLYFTKRILESSYVKPLLERYW				
yTIIN1-WW	P53076	<i>Hypothetical 108.2 kDa protein</i>	249	291
CJ-ANIETESRNLWDTSDKNSGLQYPPDQSPSSSFSSPRVSSGN				
yYFB0-WW	P43582	<i>Hypothetical 22.7 kDa protein (YFL010C)</i>	4	47
CJ-SKSNPPQVPSPGWKAVFDDEYQTWYVVDLSTNSSQWEPFRGTTWP				
yYJQ8-WW	P46995	<i>Hypothetical 84.5 kDa protein (YJL168C)</i>	470	512
CJ-IDLRRVRLPPGWEIIHENGRLYYNAEQTKLHYPPSGSSKVF				
yYPR152C-WW	Q06525	<i>Similar to S. cerevisiae hypothetical protein YKL012P</i>	1	36
CJ-MRGEWQEFKTPAGKKYYNKNKTKQSRWEKPNLKKGS				
hBAG3-WW	O95817	<i>BCL2-associated athanogene 3</i>	15	58
CJ-GNGDRDPLPPGWEIKIDPQTGWPFVVDHNSRRTTWNDRVPVSEG				
mSYNIP-WW	Q9WV89	<i>Syntaxin4-interacting protein SYNIP</i>	494	537
CJ-AILDMDSLPYGWEEAYTADGIKYFINHVTQTTSWIHPVMSALN				
tbZFP2-WW	Q967Y4	<i>Zinc-finger protein 2</i>	13	55
CJ-PSYAAVQLPPGWQMAYSVEGEVYIDHNRTRTHWKIPQEVLNNE				

^{a)} Domänennamen gebildet aus Organismus-Präfix, Proteinnamen und Domänen-Suffix (siehe 2.1.1).

^{b)} Primärer UniProt¹⁰ Zugriffscode des Proteins, aus dem die Domänensequenz entnommen worden war.

^{c)} Kurzbeschreibung des Gesamtproteins aus UniProt.

^{d)} Position der ersten und letzten Aminosäure der Domänensequenz innerhalb des Gesamtproteins.

^{e)} Aminosäuresequenz der synthetisierten Domäne (J = 6-Aminohexansäure, s = Cysteine in der Originalsequenz wurden durch Serine in den synthetisierten WW-Domänen ersetzt).

Zur späteren Kopplung an Cellulosemembranen bzw. an Peroxidase wurde allen WW-Domänen an ihrem N-Terminus über eine 6-Aminohexansäure ein Cystein angehängt (**CJ-**). Um eine Kopplung an unerwünschte Reste zu vermeiden, wurde jedes Cystein in der Domänensequenz durch Serin (**S**) ausgetauscht. Zusätzlich zu den anfänglich für diese Studie ausgewählten WW-Domänen sind in Tabelle 2-1 auch die drei WW-Domänen angegeben, die zur experimentellen Überprüfung der vorhergesagten Spezifität synthetisiert wurden.

Alle Peptide wurden als C-terminales Amid (**-NH₂**) verwendet; die N-terminalen Aminogruppen der Peptide für die NMR-Spektroskopie wurden mittels Acetanhydrid acetyliert (**ac-**, Tabelle 2-2). Die übrigen Peptide wurde zur direkten Kopplung an Peroxidase mit einem N-terminalen Cystein, welches über zwei β -Alaningruppen an die eigentliche Peptidsequenz gekoppelt wurde, synthetisiert.

Tabelle 2-2 Synthetisierte prolinreiche Peptidliganden

Ligand ^{a)}	UP-ID ^{b)}	Protein ^{c)}	Beginn ^{d)}	Ende ^{d)}	Aminosäuresequenz ^{e)}
Y-Modellligand		Modellligand			<i>ac</i> -GPPPPYG-NH ₂
poY-Modellligand		Modellligand			<i>ac</i> -GPPPPpoYG-NH ₂
poT-Modellligand†	P05129	hPKC γ	653	– 660	<i>ac</i> -VL poT PPDRL-NH ₂
Y-Ligand	P97764	mWBP1	170	– 179	<i>ac</i> - GT PPPPYTVG-NH ₂
Y-Ligand-T λ -2D		vorhergesagt			<i>ac</i> -GDPPPPYTVG-NH ₂
Y ₂ -Ligand*	P97764	mWBP1	170	– 179	C β β - GT PPPPYTVGTPPPPYTVG-NH ₂
poY ₂ -Ligand*	P97764	mWBP1	170	– 179	C β β - GT PPPPpoYTVGTPPPPpoYTVG-NH ₂
R-Ligand	Q9Y2W2	hNpwBP	479	– 490	C β β - PP GGPPRGPPPR-NH ₂
R ₂ -Ligand*	Q9Y2W2	hNpwBP	479	– 490	C β β - PP GGPPRGPPPRPEGPPPRGPPPR-NH ₂
L-Ligand	Q05859	mFMN1	872	– 881	C β β - AP PTPPLPP-NH ₂
L/P-Ligand	Q05859	mFMN1	872	– 892	C β β - AP PTPPLPPPLIPPPPLPP-NH ₂

† Originalsequenz aus hPKC γ ALTPDRL.

* Tandemwiederholung (2fach) der Originalsequenz im synthetisierten Peptid: Y₂ = doppelter Y-Ligand, poY₂ = doppelter poY-Ligand und R₂ = doppelter R-Ligand.

^{a)} Bezeichnung des Liganden.

^{b)} Primärer UniProt¹⁰ Zugriffscod des Gesamtproteins.

^{c)} Kurzbeschreibung des Gesamtproteins: hPKC γ = Protein Kinase C γ , mWBP1 = WW domain binding protein 1, hNpwBP = Npw38-binding protein hSNP70, mFMN1 = Formin 1.

^{d)} Position der ersten und letzten Aminosäure der Ligandensequenz (fett) innerhalb des Gesamtproteins.

^{e)} Aminosäuresequenz des synthetisierten Liganden (**ac-** = acetylierte N-terminale Aminogruppe, **-NH₂** = amidierter C-terminale Carboxylgruppe, **C β β -** = N-terminale Cystein- β -Alanin- β -Alanin-Kopplungsgruppe).

2.2.2 Expression und Reinigung der PDZ-Domänen

Die PDZ-Domänen aus hAF6, hERBIN und mSNA1 (Tabelle 2-3) wurden wie in Boisguerin *et al.*, 2004,^{28,200} beschrieben kloniert, in *Escherichia coli* BL21(DE3) exprimiert und aufgereinigt.

Tabelle 2-3 PDZ-Domänen: Herkunft, Position im Gesamtprotein, Sequenz und Vektoren

Domäne ^{a)}	UP-ID ^{b)}	Protein ^{c)}	Beginn ^{d)}	Ende ^{d)}	Vektor ^{e)}
<i>h</i> AF6-PDZ LRKEPEIITVTLKKQNGMGLSIVAAGKAGQDKLGIYVKSUVKGGAAVDGRLAAGDQLLSVDGRSLVGLSQERAAELMTRTSSVVTLEVAQGGAIYHGLATLL	P55196	<i>all-1 fusion partner on chromosome 6</i>		985 – 1086	pGEX6p2
<i>h</i> ERBIN-PDZ EIRVRVEKDPPELGFESISGGVGGGRGNPFRPDDDGIFVTRVQPEGPASKLLQPGDKIIQANGYSFINIEHQAVSLLKTFQNTVELIIVREVSS	Q9NR18	<i>ErbB2-interacting protein</i>		1321 – 1412	pGEX6p2
<i>m</i> SNA1-PDZ RVTVRKADAGGLGISIKGGRENKMPILISKIFKGLAADQTEALFVGDAILSVNGEDLSSATHDEAVQALKKTGKEVVLEVKYMK	Q61234	<i>α1-Syntrophin</i>		81 – 164	pGAT2

^{a)} Domänennamen gebildet aus Organismus-Präfix, Proteinennamen und Domänen-Suffix (siehe 2.1.1).

^{b)} Primärer UniProt¹⁰ Zugriffscode, aus dem die Domänensequenz entnommen worden war.

^{c)} Kurzbeschreibung des Gesamtproteins aus UniProt.

^{d)} Position der ersten und letzten Aminosäure der Domänensequenz innerhalb des Gesamtproteins.

^{e)} Verwendeter Expressionsvektor.

^{f)} Aminosäuresequenz der exprimierten Domäne.

2.2.3 Synthese und Reinigung C-terminaler Peptide

Die C-terminalen Peptide (Tabelle 2-4) wurden mittels Festphasen-Peptid-Synthese nach R.B. Merrifield¹¹⁶ und Standard Fmoc-Protokoll in einem AMS 422 Peptidsyntheseautomaten (Abimed, Langenfeld, Deutschland) hergestellt. Alle Peptide wurden über Reverse-

Tabelle 2-4 Synthetisierte C-terminale Peptidliganden

Ligand ^{a)}	UP-ID ^{b)}	Protein ^{c)}	Aminosäuresequenz ^{d)}
A1AD	P25100	<i>α-1D adrenergic receptor</i>	ADYSNLRRETDI _{COOH}
APC	P25054	<i>Adenomatous polyposis coli protein</i>	HSGSYLVTSV _{COOH}
ARVC	O00192	<i>Aemadillo repeat protein deleted in velo-cardio-facial syndrome</i>	DAKQPVDWSV _{COOH}
ATB1	P20020	<i>Calcium-transporting ATPase plasma membrane, isoform 1B</i>	GSPLHSLETSL _{COOH}
ATB2	Q01814	<i>Calcium-transporting ATPase plasma membrane, isoform 1B</i>	GSPIHSLLETSL _{COOH}
AVR2	P27037	<i>Activin receptor type II precursor</i>	NVDFPPKESL _{COOH}
BCR	P11274	<i>Breakpoint cluster region protein</i>	RQSILFSTEV _{COOH}
CIK5	P22459	<i>Voltage-gated potassium channel protein KV1.5</i>	LCLDTSRETDL _{COOH}
CIN4	P35499	<i>Sodium channel protein, skeletal muscle α-subunit.</i>	TVRPGVKESLV _{COOH}
CIN4k	P35499	<i>Sodium channel protein, skeletal muscle α-subunit.</i>	GVKESLV _{COOH}
CTNB	P35222	<i>β-catenin</i>	SNQLAWFDTDL _{COOH}
CYG4	P33402	<i>Guanylate cyclase soluble, α-2 chain</i>	IGTMFLRETS _{COOH}
CYG4k	P33402	<i>Guanylate cyclase soluble, α-2 chain</i>	FLRETS _{COOH}
EPA7	Q15375	<i>Ephrin type-A receptor 7 precursor</i>	MLHLHGTTGIQV _{COOH}
EPB2	P29323	<i>Ephrin type-B receptor 2 precursor</i>	IQSVEV _{COOH}
ERB2	P04626	<i>Receptor protein-tyrosine kinase ERBB-2 precursor</i>	PEYLGLDVVPV _{COOH}
FXI1	Q12951	<i>Forkhead box protein 11</i>	GVLYPPREGTEV _{COOH}
MCM7	P33993	<i>DNA replication licensing factor MCM7</i>	VNASRTRITFV _{COOH}
NME2	Q13224	<i>Glutamate [NMDA] receptor subunit epsilon 2</i>	YKKLSSIESDV _{COOH}
PTPZ	P23471	<i>Protein-tyrosine phosphatase ζ precursor</i>	GNIAESLES _{COOH}
REL	Q04864	<i>C-REL proto-oncogene protein</i>	DSFPYEFFQV _{COOH}
RGSC	O14924	<i>Regulator of G-protein signaling 12</i>	PKTSAHHATFV _{COOH}
TAT	P03409	<i>Trans-activating transcriptional regulatory protein</i>	SEKHFE _{COOH}
SB-AF6		Vorhergesagter Superbinder für hAF6-PDZ	LEGIFV _{COOH}
SB-ERBIN		Vorhergesagter Superbinder für hERBIN-PDZ	WLETWV _{COOH}
SB-SNA1		Vorhergesagter Superbinder für mSNA1-PDZ	IRETIV _{COOH}

^{a)} Bezeichnung des Liganden (SB- = vorhergesagter Superbinder der entsprechenden PDZ-Domäne). Der Suffix k kennzeichnet die verkürzte Version eines doppelt auftretenden Peptids.

^{b)} Primärer UniProt¹⁰ Zugriffscode des Proteins, aus dem die Ligandensequenz entnommen worden war.

^{c)} Kurzbeschreibung des Gesamtproteins aus UniProt.

^{d)} Aminosäuresequenz des synthetisierten Liganden (_{COOH} = freie C-terminale Carboxylgruppe).

Phase-Hochleistungsflüssigkeitschromatographie (RP-HPLC) aufgereinigt und durch Massenspektrometrie auf Reinheit und korrekte Masse hin untersucht.

Die Peptide lagen nach der Synthese mit einer freien N-terminalen Aminogruppe und einer freien C-terminalen Carboxylgruppe (COOH) vor. Zusätzlich wurden als Negativkontrollen von einigen Peptiden C-terminale Carboxamid-Derivate (CONH_2) synthetisiert. Zur experimentellen Verifikation wurden die vorhergesagten Superbinder (SB-) ebenfalls synthetisiert.

2.3 Liganden-Screening der Protein-Interaktionsdomänen

2.3.1 NMR-Spektroskopie

Das experimentelle Screening der WW-Domänen mit Hilfe der NMR-Spektroskopie wurde von Brigitte Schlegel, Dr. Peter Schmieder und Dr. Ricardo Pires durchgeführt.

Alle 1D- ^1H -NMR-Messungen wurden in einem Puffer bestehend aus 10 mM Phosphat, 100 mM NaCl, 0,1 mM DTT, 0,1 mM EDTA und 0,02% NaN_3 bei einem pH von 6,0 durchgeführt. Zu jeder Probe (500 μl) wurden im Anschluss zusätzlich 50 μl $^2\text{H}_2\text{O}$ hinzugegeben. WW-Domänen wurden in einer Konzentration von etwa 400 bzw. 200 μM eingesetzt. Für jeden Peptidliganden wurde eine Stammlösung mit jeweils 16 mg Peptid in 400 μl Puffer vorbereitet (etwa 40 mM Konzentration der Liganden). Zur Bestimmung der Domänen/Liganden-Wechselwirkung wurde ein Referenzspektrum jeder Domäne ohne Ligand mit den jeweiligen Spektren der Domänen mit Ligand verglichen. Der Ligand wurde in einem molekularen Verhältnis von 2:1 zur Domäne hinzugegeben.

Die Spektren wurden an einem DRX600 Spektrometer (Bruker, Karlsruhe, Deutschland) in Standardkonfiguration aufgenommen. Ein inverser Tripel-Resonanz-Probenkopf mit drei aktiv abgeschirmten Gradienten-Spulen kam dabei zum Einsatz. Die Messungen wurden bei 285 K durchgeführt. Die Spektren wurden mit dem Programm XWIN-NMR der Firma Bruker aufgenommen und prozessiert.

2.3.1.1 Bestimmung des Faltungszustands der Domänen

Zur Bestimmung des Faltungszustands wurden die Signalmuster der 1D- ^1H -NMR-Spektren im Bereich zwischen 9,0 und 9,6 ppm und im Bereich um 10,0 ppm bewertet. Scharfe Signale zwischen 9,0 und 9,6 ppm wurden als Indiz für das Vorliegen des WW-Domänen β -Faltblatts interpretiert. Im Bereich um 10 ppm konnten die Signale der Wasserstoffatome

am Indolstickstoff der Tryptophane identifiziert werden. Im Falle von nur zwei Tryptophanen, sind die Signale der zwei konservierten Tryptophane $W_{\omega 17}$ und $W_{\omega 39}$ für gefaltete Domänen auf Grund der verschiedenen chemischen Umgebung (hydrophober Kern für $W_{\omega 17}$ und wässrige Umgebung für $W_{\omega 39}$) scharf und unterscheidbar. In ungefalteten Domänen ergibt sich dagegen ein einzelnes weniger scharfes Signal.

2.3.1.2 Quantifizierung der Wechselwirkungsaktivität der WW-Domänen

Zur Quantifizierung der Stärke der Wechselwirkung wurde die Tatsache ausgenutzt, dass sich bei der klassischen Bindung des Liganden die chemische Umgebung des $W_{\omega 39}$ (siehe Einleitung, Abbildung 1-2) stark und bei gleichem Liganden in konstanter Weise (hydrophobere Umgebung) ändert. Diese Änderung konnte an Hand der Veränderung der chemischen Verschiebung (CSP in ppm) des Wasserstoffatoms am Indolstickstoff von $W_{\omega 39}$ nach Zugabe des Liganden quantifiziert werden. Dazu wurden die Signale um 10 ppm des $^1\text{H-NMR}$ -Referenzspektrums (ohne Ligand) mit den Signalen des Spektrums der jeweiligen WW-Domänen/Liganden-Kombinationen verglichen. Die Resonanzzuordnung basierte auf der Annahme, dass sich nur das Signal des Wasserstoffatoms am Indolstickstoff von $W_{\omega 39}$ signifikant bei Ligandenbindung ändert. Dagegen sollte das Signal des nicht an der Ligandenbindung beteiligten $W_{\omega 17}$ keine bzw. nur geringe Veränderungen zeigen.

Die Wechselwirkungen wurden an Hand der Größe der Veränderung der chemischen Verschiebung als stark ($\text{CSP} \geq 0,150$) bzw. als schwach ($0,030 \leq \text{CSP} < 0,150$) eingestuft. Zusätzlich wurden auch die Veränderungen der chemischen Verschiebungen in anderen Bereichen (siehe 2.3.1.1) des NMR-Spektrums qualitativ beurteilt. Im Falle von WW-Domänen ohne Tryptophan an $\omega 39$ wurden signifikante Veränderungen in anderen Bereichen des Spektrums als schwache Bindung eingestuft und mit „*shift*“ bezeichnet. Dagegen wurden für WW-Domänen mit Tryptophan an $\omega 39$ signifikante Veränderungen in anderen Bereichen des Spektrums aber ohne quantifizierbare CSP des Wasserstoffatoms am Indolstickstoff von $W_{\omega 39}$ als keine Bindung gewertet (ebenfalls mit „*shift*“ gekennzeichnet).

Durch „*keine*“ wurden diejenigen WW-Domänen ohne $W_{\omega 39}$ gekennzeichnet, die nach Zugabe des Liganden auch in den anderen Bereichen des Spektrums keine Veränderungen zeigten. Die Ergebnisse des NMR-Screenings stimmten qualitativ mit denen des WW-Array-Screenings (siehe Ergebnisse, Tabelle 3-3) überein. Der Nachweis von Interaktionen mittels NMR-Spektroskopie schien aber sensitiver zu sein.

2.3.2 Herstellung und Inkubation festphasegebundener Protein- und Peptidbibliotheken

Das Screening der Domänen mit Hilfe festphasegebundener Peptid- bzw. Proteinbibliotheken wurde im Rahmen zweier Kooperationsprojekte von Dr. Livia Otte (WW-Domänen) und Dr. Prisca Boisguerin (PDZ-Domänen) aus der Molecular Libraries and Recognition Group des Instituts für Medizinische Immunologie, Universitätsklinikum Charité, durchgeführt.

2.3.2.1 WW-Array: Bibliothek 42 synthetisch hergestellter WW-Domänen

Zur Herstellung des „WW-Arrays“ wurden die 42 synthetisierten WW-Domänen über ihr N-terminales Cystein an eine Maleimid-aktivierte Cellulosemembran gekoppelt.¹³⁰

2.3.2.2 Peptidbibliotheken: SPOT-Synthese und Inkubation

Die festphasegebundenen Peptidbibliotheken wurden mittels SPOT-Synthese auf Cellulosemembranen hergestellt.^{54,100,130} Während der SPOT-Synthese wurden die Peptide schrittweise an definierten, lokalisierbaren Positionen direkt auf der Membran in C- nach N-terminaler Richtung synthetisiert. Um einen freien C-Terminus für die Wechselwirkung mit PDZ-Domänen zu erhalten wurde die SPOT-Synthese der invertierten Peptide²⁸ entwickelt. Das Peptid wurde dabei zuerst mittels Standard SPOT-Synthese hergestellt. In einem anschließenden Zyklisierungsschritt wurde das N-terminale Ende des Peptids, das mit einer speziellen funktionalen Gruppe versehen worden war, mit einer weiteren komplementären funktionalen Gruppe innerhalb des Linkers zwischen Membran und C-terminalen Ende des Peptids verknüpft. Schließlich wurde in einem Spaltungsschritt die speziell eingeführte labile Bindung zwischen dem C-Terminus des Peptids und dem Linker hydrolysiert und somit das Peptid „direkt an der Membran invertiert“.²⁷

Substitutionsanalysen sind systematische Peptidbibliotheken aller Einzelsubstitutionsanaloge eines gegebenen Ausgangspeptids. Die Bibliothek ist so organisiert, dass die Zeile bestimmt welche Position des Ausgangspeptids substituiert wird, während die Spalte den Aminosäuretyp bestimmt, durch den substituiert wird. Substitutionsanalysen erlauben es, lineare Bindungsmotive direkt visuell zu interpretieren.

Längenanalysen sind Peptidbibliotheken der systematisch vom N- und C-Terminus um je eine Aminosäure verkürzten Peptidvarianten. Sie erlauben die Bestimmung der minimalen Länge des Bindungsmotivs.

Profilbibliotheken wurde speziell im Rahmen der neu entwickelten Methode der *Quantitativen Spezifitäts-Profile* (2.6.3) für die Quantifizierung der aminosäurespezifischen Affinitätsbeiträge an den einzelnen Ligandenpositionen designt. In Vorstudien (z.B. Substitutionsanalysen) wurden für jede Domäne die präferierten Aminosäuren an den einzelnen für die Spezifität relevanten Ligandenpositionen identifiziert. Alle Kombinationen dieser reduzierten Aminosäure-Sätze ergaben zusammen die domänenspezifische Profilbibliothek. Diese deckt damit den Subraum des potentiellen Liganden-Sequenzraums ab, in dem eine Wechselwirkung im relevanten Affinitätsbereich stattfinden kann.

Die **Bibliothek humaner C-Termini** enthielt 6223 repräsentative C-Termini (11 mere, die letzten 11 Aminosäurereste) der Proteine des humanen Genoms. Sie wurde speziell für das Liganden-Screening von PDZ-Domänen designt. Dazu wurden aus der Swiss-Prot-Datenbank (Release 40) alle 6584 humanen Proteine extrahiert und die C-Termini mit redundanten Sequenzen entfernt.

2.3.2.3 Direkte Markierung von Peptiden und Proteinen mit Peroxidase

Sowohl die prolinreichen Liganden als auch die WW-Domänen wurden über das am N-Terminus angehängte Cystein direkt mit Maleimid-aktivierter Meerrettichperoxidase markiert.¹²⁹

2.3.2.4 Bindungsstudien mit festphasegebundenen Bibliotheken

Die festphasegebundenen Bibliotheken wurden mit Ethanol (1x10 min) und Tris-gepufferter-Saline pH 8,0 (3x10 min) vorgewaschen und dann für 2-4 h mit Blockierungspuffer (Blockierungsreagenz 10-fach Konzentrat, Sigma-Genosys, Cambridge, USA) behandelt, um unspezifische Bindungen zu unterbinden. Die Bibliotheken wurden anschließend mit direkt Peroxidase-markierten Peptidliganden (2–10 mg/ml, *WW-Array*), mit direkt Peroxidase-markierten WW-Domänen (10–50 mg/ml, Substitutions- und Längenanalysen für WW-Domänen) oder GST-markierten PDZ-Domänen (10 µg/ml, *Bibliothek humaner C-Termini*, Substitutionsanalysen und Profilbibliotheken) über Nacht in Blockierungspuffer bei 4 °C inkubiert.^{27,129}

Die Bibliotheken, die mit direkt Peroxidase-markierten Peptiden und Protein inkubiert worden waren, wurden schließlich mit Tris-gepufferter-Saline pH 8,0 (4x10 min) gewaschen.¹²⁹ Dagegen wurden die Bibliotheken, die mit GST-markierten PDZ-Domänen inkubiert worden waren, zuerst mit Tris-gepufferter-Saline pH 8,0 (3x10 min) gewaschen, dann

mit anti-GST-Meerrettichperoxidase (1 $\mu\text{g/ml}$, 3 h) in Blockierungspuffer inkubiert und schließlich nochmals mit Tris-gepufferter-Saline pH 8,0 (3x10 min) gewaschen.²⁷

Die gebundenen Peptide bzw. Domänen wurden über die direkt oder indirekt gekoppelte Peroxidase durch das Chemilumineszenz-System SuperSignal West Pico (Pierce, Rockford, Illinois, USA) nachgewiesen. Die Signalintensität wurde mit Hilfe eines Lumineszenzdetektor (Lumilmager, Boehringer Mannheim GmbH, Deutschland) und der Software LumiAnalyst in Form von *Boehringer-Light-Units* (BLU) quantifiziert.^{27,129}

2.4 Bestimmung der Domänen/Liganden-Affinitäten

2.4.1 UV-Fluoreszenzspektroskopie an WW-Domänen

Die Veränderung des UV-Fluoreszenz-Emissions-Spektrums von W ω 39 bei Bindung eines prolinreichen Liganden wurde dazu verwendet die Menge an gebildeten WW-Domänen/Liganden-Komplexen in Abhängigkeit von der freien Konzentration des Liganden zu bestimmen. Unter Annahme einer 1:1 Stöchiometrie zwischen Domäne und Ligand im Komplex konnte aus diesen Daten über das Massenwirkungsgesetz die Dissoziationskonstante K_d der Wechselwirkung durch Kleinste-Quadrate-Anpassung des Modells in Gleichung 2:1 bestimmt werden:

$$F = F_f + (F_g - F_f) [L_f] (K_d + [L_f])^{-1} \quad (2:1)$$

Modell des Zusammenhangs zwischen der Veränderung der Fluoreszenzemission und der K_d

F repräsentiert die gemessene Fluoreszenz bei 298 nm, F_f die Fluoreszenz der freien Domäne ohne Ligand und F_g die Fluoreszenz der gebundenen Domäne. L_f ist der freie Ligand im Gleichgewicht, K_d die Dissoziationskonstante.

Die Proben wurde durch Licht mit einer Wellenlänge von 298 nm bei einer Spaltbreite von 5 nm angeregt. Für jeden der mindestens 8 Titrationsschritte (bis zu etwa 1 mM Gesamtkonzentration des Liganden) wurde jeweils ein Emissionsspektrum zwischen 320 nm und 420 nm bei einer Spaltbreite von 10 nm aufgenommen und die Fluoreszenzintensität bei 340 nm bestimmt. Alle Messungen wurden bei 298 K an einem LS50B (Perkin-Elmer, Wellesley, Massachusetts, USA) Lumineszenz-Spektrometer durchgeführt. Die WW-Domänen wurde in einer Konzentration von 10 μM eingesetzt, für die Liganden wurde jeweils eine 32 mM Stammlösung vorbereitet. Die WW-Domäne und die Liganden waren in einem 40 mM Phosphatpuffer pH 7,2, 100 mM NaCl und 50 mM Dithiothreitol gelöst.

2.4.2 Oberflächen-Plasmon-Resonanz-Messungen (SPR) an PDZ-Domänen

Die Oberflächen-Plasmon-Resonanz-Messungen (SPR) wurden im Rahmen eines Kooperationsprojekts von Dr. Prisca Boisguerin (PDZ-Domänen) aus der Molecular Libraries and Recognition Group des Instituts für Medizinische Immunologie, Universitätsklinikum Charité, durchgeführt.

Die Messungen wurden an einem BIACORE X System (Biacore, Uppsala, Schweden) mit einem CM5-Chip, an den die PDZ-Domänen über das fusionierte GST (Glutathion-S-Transferase) gekoppelt worden waren, durchgeführt. Als Referenz diente GST. Die Wechselwirkungen wurden bei 20 °C und einer Flussrate von 5 $\mu\text{l}/\text{min}$ (10 μl Injektionsvolumen) in einem HEPES-gepufferte-Saline-Puffer pH 7,4 (10 mM HEPES, 150 mM NaCl, 3,4 mM EDTA, 0,005% Detergenz P20) gemessen. Die Datenanalyse und die Anpassung des Wechselwirkungsmodells (analog zu Gleichung 2:1) wurde mit der Software BIA-EVALUATION 3.0 durchgeführt.

2.5 Allgemeine statistische Methoden

Für die statistische Datenanalyse wurde R 1.8.1¹⁴⁸ eingesetzt (<http://www.R-project.org>).

2.5.1 MAD-Z Standardisierung

MAD-Z-Werte beschreiben in standardisierter Weise die Richtung und das Ausmaß der Abweichung eines Werts vom Median – als robustem Schätzwert für den Mittelwert – der Verteilung (Gleichung 2:2). Die Abweichung wird dabei in Einheiten des Medians der absoluten Abweichungen – als robuster Streuungsschätzwert – angegeben (Gleichung 2:3).

$$MAD-Z_i = (x_i - \text{median}(x)) MAD^{-1} \quad (2:2)$$

$$MAD = \text{median}(|x_i - \text{median}(x)|) \quad (2:3)$$

MAD-Z Transformation von Messdaten

$MAD-Z_i$ repräsentiert den MAD-Z-transformierten Wert des Messwerts x_i . MAD ist der Median der absoluten Abweichung aller Messwerte x_i , mit $i = 1, \dots, l$ vom Median aller Messwerte.

2.5.2 Validierung statistischer Modelle: Determinationskoeffizient und SDEP

Die Anpassung der Vorhersagen eines Modells an die experimentellen Daten wird durch den Determinationskoeffizienten R^2 quantifiziert (Gleichungen 2:4 - 2:6).

$$R^2 = \frac{\sum (y_{Mod,i} - \langle y_{Exp} \rangle)^2}{\sum (y_{Exp,i} - \langle y_{Exp} \rangle)^2} \quad (2:4)$$

$$R^2 = 1 - \frac{\sum (y_{Exp,i} - y_{Mod,i})^2}{\sum (y_{Exp,i} - \langle y_{Exp} \rangle)^2} \quad (2:5)$$

$$R^2 = r^2 \quad (2:6)$$

Determinationskoeffizient R^2

$y_{Exp,i}$ repräsentiert den experimentell bestimmten Wert für den Fall i , $y_{Mod,i}$ repräsentiert die entsprechende Vorhersage des Modells für den Fall i und $\langle y_{Exp} \rangle$ den Mittelwert der experimentellen Werte. r entspricht dem Korrelationskoeffizient nach Bravais-Pearson (Produkt-Moment-Korrelationskoeffizient).

R^2 repräsentiert dabei den Anteil der durch das Modell erklärten Variation ($\sum (y_{Mod,i} - \langle y_{Exp} \rangle)^2$) an der Gesamtvariation der Daten ($\sum (y_{Exp,i} - \langle y_{Exp} \rangle)^2$), wobei ($\sum (y_{Exp,i} - y_{Mod,i})^2$) die nicht erklärte Residualstreuung darstellt. R^2 kann direkt aus dem Korrelationskoeffizient r nach Bravais-Pearson berechnet werden. Zur Berechnung des kreuzvalidierten Determinationskoeffizienten Q^2 werden die Vorhersagen mit kreuzvalidierten Modellen berechnet. Dazu werden aus den Trainingsdatensätzen genau diejenigen Fälle weggelassen, für die das Modell den entsprechenden Wert vorhersagt. Für die Kreuzvalidierung durch *leave-one-out* wird für die Vorhersage für jeden einzelnen Fall ein Modell erstellt, in dessen Training die experimentellen Daten des vorherzusagenden Falls nicht eingegangen sind.

Die Standardabweichung der Fehler der Vorhersage *SDEP* repräsentiert die durch die Anzahl der Fälle n im Trainingsdatensatz normierte Residualstreuung.

$$SDEP = \left(\sum (y_{Exp,i} - y_{Mod,i})^2 n^{-1} \right)^{1/2} \quad (2:7)$$

Standardabweichung der Fehler der Vorhersage *SDEP*

$y_{Exp,i}$ repräsentiert den experimentell bestimmten Wert für den Fall i , $y_{Mod,i}$ repräsentiert die entsprechende Vorhersage des Modells, für n Fälle.

2.6 Quantitative Auswertung der Screenings mittels festphasegebundener Bibliotheken

2.6.1 WW-Array

Die quantifizierten Signalintensitäten (in BLU) von 3 unabhängigen WW-Array-Screenings pro Ligand (für den L-Liganden nur 2 Screenings) wurden Logarithmus-transformiert. Daraufhin wurden diese Werte für jedes unabhängige Screening in Form von MAD-Z-Werten standardisiert (2.5.1). Schließlich wurden die Mittelwerte der MAD-Z-Werte der 3 unabhängigen Screenings in folgende Klassen eingeteilt:

„–“ keine signifikante Bindung ($\text{MAD-Z} \leq 4$), „+/-“ unklare Bindung ($4 < \text{MAD-Z} \leq 8$), „+“ schwache Bindung ($8 < \text{MAD-Z} \leq 12$) und „++“ starke Bindung ($\text{MAD-Z} > 12$).

2.6.2 Bibliothek humaner C-Termini

Die Signalintensitäten (in BLU) der Screenings der *Bibliothek humaner C-Termini* mit PDZ-Domänen wurden in Form von MAD-Z-Werten standardisiert (2.5.1). Für eine vorläufige Analyse wurden die C-Termini mit den 100 höchsten Signalintensitäten ausgewählt (Top100 Peptide).

2.6.3 Quantitative Spezifitäts-Profile

Zur Quantifizierung der Spezifität einer Protein-Interaktionsdomäne wurde die Methode der *Quantitativen Spezifitäts-Profile* (QSP) entwickelt. Im Wesentlichen basiert die Methode auf dem Design einer festphasegebundenen Profilbibliothek (2.3.2.2), der Kalibrierung der Signalintensitäten (2.6.3.2) an Hand experimentell bestimmter Affinitäten (2.4.2) und der Modellierung des quantitativen Zusammenhangs zwischen Ligandensequenz und Affinität (2.6.3.3). Die theoretische Grundlage der QSPs wird in 2.6.3.1 entwickelt.

2.6.3.1 Ableitung des Zusammenhangs zwischen der Signalintensität (in BLU) und den zugrundeliegenden Affinitäten

In den Screenings festphasegebundener Bibliotheken werden die in Lösung befindlichen Domänen (bzw. Liganden) im Falle einer Wechselwirkung mit den festphasegebundenen Liganden (bzw. Domänen) an die Membran gebunden. Die Menge an Domäne D_g (bzw. Liganden), die dadurch aus der Lösung an die Membran gebunden wurde, wird über eine

direkt ($\kappa = 1$) oder indirekt über Antikörper mit der Domäne (bzw. dem Liganden) gekoppelten Peroxidase E_g (Enzym) nachgewiesen

$$E_g = \kappa D_g \quad (2:8)$$

Die Menge an gebundener Peroxidase E_g wird durch die Signalintensität I der durch sie katalysierten Chemilumineszenz-Reaktion in *Böhringer-Light-Units* (BLU) quantifiziert

$$I = \varepsilon E_g + \alpha \quad (2:9)$$

Unter der Annahme eines Überschusses an Chemilumineszenz-Substrat besteht damit ein direkter Zusammenhang zwischen der Signalintensität I und der Menge an gebundener Domäne D_g

$$I = \lambda D_g + \alpha \quad (2:10)$$

Dabei wird die Proportionalitätskonstante ε (Einfluss der enzyspezifischen Michaelis-Menten-Konstante, der Expositionszeit, der Größe der Detektionsfläche, usw.) und die Proportionalitätskonstante κ (Einfluss der Affinität des indirekten Detektionssystems, usw.) in eine einzige Proportionalitätskonstante λ zusammengefasst. Die Hintergrund-Signalintensität wird durch den Offset-Parameter α repräsentiert.

Über das Massenwirkungsgesetz und unter Annahme eines 1:1 Bindungsmodells steht die Menge an gebundener PDZ-Domäne D_g zusammen mit der Menge an freier Domäne D_f bzw. freiem Liganden L_f in direktem Zusammenhang mit der Dissoziationskonstante K_d

$$K_d = [D_f L_f] D_g^{-1} \quad (2:11)$$

$$K_d = [(D_f - D_g) (L_f - D_g)] D_g^{-1} \quad (2:12)$$

Dabei wird die für das Wechselwirkungs-Gleichgewicht zur Verfügung stehende Gesamtmenge der PDZ-Domäne D_f in Lösung und die Gesamtmenge des peptidischen Liganden L_f , die pro Flächeneinheit an der Membran fixiert vorliegt, als konstant für alle Bereiche einer Membran („Spots“) und einer Inkubation angesehen. Über die Umkehrfunktion von Gleichung 2:12

$$D_g = \frac{1}{2} \{ K_d + D_f + L_f - [(-K_d - D_f - L_f)^2 - 4D_f L_f]^{-1/2} \} \quad (2:13)$$

lässt sich zusammen mit Gleichung 2:10 schließlich das Modell des Zusammenhangs zwischen den Logarithmus-transformierten Signalintensitäten $\ln(I)$ – wegen der annähernd log-normal-verteilten Intensität – und den zugrundeliegenden K_d -Werten formulieren

$$\ln(I - \alpha) = \ln(\lambda \frac{1}{2} \{ K_d + D_f + L_f - [(-K_d - D_f - L_f)^2 - 4D_f L_f]^{-1/2} \}) \quad (2:14)$$

Für große Dissoziationskonstanten – in Abhängigkeit von D_f und L_f – ist der Zusammenhang annähernd linear ($\ln(I - \alpha) \sim \ln(K_d)$).

2.6.3.2 Kalibrierung der Signalintensität der Profilbibliotheken

Zur Schätzung der zugrundeliegenden Dissoziationskonstanten auf Basis der Signalintensitäten der Profilbibliotheken (Kalibrierung der Signalintensitäten) enthielt jede Bibliothek eine Reihe von Kontrollpeptiden in 5 Replikaten. Für diese Kontrollpeptide waren zuvor die Dissoziationskonstanten per Oberflächen-Plasmon-Resonanz bestimmt worden (2.4.2). Dadurch ermöglichten sie eine Schätzung der Parameter (α , λ , D_i und L_i) des Zusammenhangs zwischen den Signalintensitäten und den Dissoziationskonstanten (2.6.3.1).

Der Offset-Parameter α wurde durch ein Maximum-Likelihood-Verfahren mittels der Funktion LOGTRANS¹⁹¹ in R¹⁴⁸ geschätzt. Die Parameter λ , D_i und L_i wurden durch nicht-lineare kleinste-Quadrate Anpassung des Modells (Gleichung 2:14) an die Intensitäts- und K_d -Daten der Kontrollpeptide bestimmt. Für die Minimierung wurde der L-BFGS-B-Algorithmus³⁴ inklusive Randbedingungen für die Parameter verwendet. Insbesondere wurde auf Grund der Waschschriffe davon ausgegangen, dass $D_i \leq L_i$ ist.

Mittels dieser Parameter wurden ausgehend von den Signalintensitäten die $\ln(K_d)$ -Werte (kalibrierte Signalintensitäten) geschätzt. Der Determinationskoeffizient dieser Schätzung betrug für die hAF6-PDZ Profilbibliothek $R^2 = 0,57$ (siehe Anhang B, Abbildung B-1) und für die hERBIN-PDZ Profilbibliothek $R^2 = 0,49$ (siehe Anhang B, Abbildung B-5).

2.6.3.3 Quantifizierung der Sequenz-Affinitäts-Beziehung: Quantitative Spezifitäts-Profile

Zur Modellierung des quantitativen Zusammenhangs zwischen der Sequenz (faktorielle Variablen) und der Affinität (verhältnisskalierte Variable) wurde im Rahmen der Methode der Quantitativen Spezifitäts-Profile die multifaktorielle Varianzanalyse verwendet (ANOVA = engl. *analysis of variance*).⁵³ Alle Analysen wurden in R 1.8.1 (<http://www.R-project.org>) durchgeführt.¹⁴⁸

Als Schätzwerte für die Affinitäten ($\ln(K_d)$) der einzelnen Peptide der untersuchten Profilbibliothek wurden die kalibrierten Signalintensitäten verwendet (2.6.3.2). Zur Beurteilung der Kooperativität zwischen den einzelnen Ligandenpositionen wurden ANOVA-Modelle mit und ohne paarweise statistische Interaktionsterme verglichen. Für die Erstellung der kreuzvalidierten ANOVA-Modelle wurden diejenigen Peptide aus dem Trainingsdatensatz entfernt, für welche die Affinität der Wechselwirkung mit der jeweiligen PDZ-Domäne vorhergesagt und mit der durch SPR bestimmten Affinität verglichen werden sollte.

Um den abgedeckten Sequenzraum zu erweitern wurden die logarithmus-transformierte Signalintensitäten $\ln(I)$ der Substitutionsanalysen mit in die Trainingsdatensätze der ANOVA-Modelle mitaufgenommen. Vorher mussten die Signalintensitäten der Substitutionsanalysen an die der Profilbibliothek angepasst werden. Dazu wurden für diejenigen Peptide, die in beiden Bibliotheken enthalten waren, lineare Regressionen zwischen den Signalintensitäten der Substitutionsanalyse und denen der Profilbibliothek (Referenz) erstellt. Auf Basis dieser linearen Regressionsmodelle wurden die Signalintensitäten transformiert. Zur Kompensation der korrelierten Fehler und expliziten Modellierung der Zwischen-Experiment-Fehler wurden Gemischte-Effekte-ANOVA-Modelle¹⁰² erstellt. Für die Modellierung wurde die LME Funktion des R Pakets NLME¹⁷ verwendet. Zur Anpassung der Gemischte-Effekte-ANOVA-Modelle wurde das eingeschränkte Maximum-Likelihood-Verfahren angewandt.

2.7 Methoden der Bioinformatik

2.7.1 Sequenzdatenbanken

2.7.1.1 WW-Domänensequenzen aus der SMART-Datenbank

Die SMART-Datenbank¹⁶¹ (Stand: 15.11.2001) enthielt 482 WW-Domänensequenzen, von denen 158 Sequenzen einmalig waren, während 96 jeweils mehrfach vorkamen, allerdings häufig im Kontext unterschiedlicher Proteine bzw. Organismen. Diese Tatsache machte eine klare Identifizierung redundanter Sequenzen unmöglich, weshalb mit der gesamten Datenbank gearbeitet wurde.

2.7.2 Sequenzmethoden

Zur Verarbeitung von Aminosäuresequenzen wurden vorrangig das *GCG Wisconsin Package 10.2* (Accelrys, San Diego, Kalifornien, USA), CLUSTAL-X¹⁸¹ und GenDoc¹²¹ verwendet.

2.7.2.1 Konsensusfärbung von Sequenzalignments

Sequenzalignments wurden mit der in CLUSTAL-X¹⁸¹ implementierten Konsensusfärbemethode eingefärbt. Dabei werden nur diejenigen Aminosäuren mit einer Farbe hinterlegt, die in der jeweiligen Position eine konservierte Eigenschaft repräsentieren: negativ geladene

Aminosäuren (D/E: Rot), positiv geladenen Aminosäuren (K/R: Blau), polare Aminosäuren (Q/N/T/S: Violett), hydrophobe Aminosäuren (I/L/V/M/A/C: Grün), aromatische Aminosäuren (Y/F/W/H: Gelb), Glyzin (G: Dunkelgrün), Prolin (P: Braun).

2.7.2.2 Konsensussequenz

Einfache Konsensussequenzen wurden mit SeqLab (GCG Wisconsin Package) berechnet. Dazu wurde die Peptid-Identitätsmatrix mit einem Schwellenwert von 2,1 und einem Prozentsatz von 75 % verwendet. Im Falle konservierter Eigenschaften, aber nicht konservierter Aminosäuretypen, wurde die repräsentative Aminosäure manuell angepasst.

2.7.2.3 Sequenzlogos

Sequenzlogos¹⁵⁸ wurden über Weblogo (<http://weblogo.berkeley.edu>) erstellt.⁴³ Die Aminosäuren wurden analog zum CLUSTAL-X-Farbschema eingefärbt (2.7.2.1). Für genügend umfangreiche Sequenzalignments wurde eine Korrektur für kleine Stichproben durchgeführt, um den Einfluss der *a-priori*-Wahrscheinlichkeitsverteilung der Aminosäuren und der Stichprobengröße zu berücksichtigen.

2.7.2.4 Sequenzverwandtschaftsbaum

Zur Berechnung des Sequenzverwandtschaftsbaums wurde das PHYLIP 3.6a Packet⁵⁰ verwendet. Die paarweise Verwandtschaft der Sequenzen (Sequenzdistanzmatrize) wurde mittels PROTDIST auf Basis des PAM-Mutationsmodells (*engl. percent accepted mutations*)⁴⁴ berechnet. Ausgehend von dieser Sequenzdistanzmatrize wurde der Baum mittels NEIGHBOR auf Basis des *neighbor-joining* Algorithmus¹⁵⁵ berechnet. In die Berechnung des Verwandtschaftsbaums wurden nur diejenigen Positionen miteinbezogen, deren Seitenketten potentiell mit dem Liganden interagieren können ($\omega 18$, $\omega 20$, $\omega 22$ – $\omega 26$, $\omega 28$, $\omega 30$, $\omega 32$ – $\omega 35$, $\omega 37$, $\omega 39$, sowie $\omega 17$ als Fixpunkt, siehe Einleitung, Tabelle 1-2). Für die Darstellung des Baums wurde TREEVIEW¹³¹ verwendet.

2.7.2.5 Profile-Hidden-Markov-Modelle

Die *Profile-Hidden-Markov-Modelle* (pHMM) wurden mit Hilfe des HMMer 2.2 Pakets⁴⁷ berechnet. Zur Klassifizierung mittels der pHMMs wurde der Roh-Homologiewert verwendet, der eine Schwellenwertdefinition unabhängig von der durchsuchten Datenbankgröße ermöglichte. Die Schwellenwerte für die Klassifizierung wurden durch den Vergleich der Se-

quenzen des Trainings-Alignments mit dem pHMM bestimmt (siehe Anhang A, Tabelle A-1). Die pHMMs wurden durch HMMLogo (<http://logos.molgen.mpg.de>)¹⁶² in einer den Sequenzlogos analogen Form visualisiert.

2.7.3 Strukturmethode

Zur Bearbeitung der Strukturen wurde SYBYL 6.9.1 (Tripos, St. Louise, USA) und zur Strukturvisualisierung MolMol⁹⁷ verwendet. Für SYBYL wurde zusätzlich ein Skript-Paket entwickelt, das die effiziente Verwaltung und Verarbeitung vieler Repräsentanten einer Domänenfamilie in einer Strukturdatenbank ermöglicht. Dieses Paket erlaubt darüber hinaus die Integration eines Sequenzalignments mit der entsprechenden Strukturdatenbank. Damit können Sequenzmotive in den Strukturen und Struktur motive im Sequenzalignment visualisiert werden.

2.7.3.1 Hochdurchsatz-Modellierung der 42 WW-Domänen und der WW-Domänen/Liganden-Komplexe

Die 42 WW-Domänen wurden ausgehend von der Strukturvorlage von *hPIN1*-WW [PDB: 1PIN] und dem Sequenzalignment der WW-Domänensequenzen mit Hilfe eines dafür entwickelten automatischen Skript-Pakets in Tripos SYBYL modelliert. Die Strukturen wurden in Tripos SYBYL unter Verwendung des Kraftfelds von Cornell *et al.*, 1995, im Vakuum minimiert (mit geladenen N- und C-Termini, maximal 200 Schritte Simplex-Minimierung,¹⁴⁵ maximal 2000 Schritte Powell-Minimierung¹⁴⁴). Während der Minimierung wurden die Atompositionen des konservierten hydrophoben Cluster ($W\omega 17$, $Y/F\omega 29$ und $P\omega 42$) fixiert. Liganden wurden als ideale Polyprolin-Typ-II-Helices (Φ -Winkel = -78° und Ψ -Winkel = 146°) modelliert.

Alle 504 Kombinationen der 42 WW-Domänen und der 12 Ligandenorientierungen (Y-Ligand in 2 Positionen, R-Ligand in 4 Positionen, L-Ligand in 4 Positionen und poly-P-Ligand in 2 Positionen) wurden semi-automatisch durch manuelles Docking der Liganden an Hand der konservierten Ligandenpositionen an der konservierten Prolin-Bindungsstelle erstellt. Die konservierten Ligandenpositionen waren aus der Überlagerung aller experimentell bestimmten Komplexstrukturen abgeleitet worden. Die 504 Komplexe wurden mit dem AMBER 7 Paket⁸ unter Verwendung des Kraftfelds von Cornell *et al.*, 1995, minimiert (maximal 5000 Schritte). Ein harmonisches Potential mit einer Kraftkonstante von

32 kcal mol⁻¹ Å⁻² fixierte die Atompositionen des konservierten hydrophoben Cluster (W ω 17, Y/F ω 29 und P ω 42) während der Minimierung.

An Hand der aus den Substitutionsanalysen abgeleiteten Kontakten wurde die „Korrektheit“ der Komplexe bestimmt und die Komplexe somit in „bindende“ (reale) und „nicht-bindende“ Komplexe (in Realität kommt solch eine Komplexstruktur nicht vor) eingeteilt.

2.7.3.2 CoMFA – Comparative-Molecular-Field-Analyse

Die *Comparative-Molecular-Field-Analyse* (CoMFA)⁴² wurde mit Hilfe von Tripos SYBYL durchgeführt (siehe auch Anhang D). Die sterischen und elektrostatischen molekularen Felder wurden nur für die dem Liganden zugewandte Seite des WW-Domänen β -Faltblatts mit einer Raster-Schrittweite von 2 Å berechnet. Modelle, die ausgehend von molekularen Feldern mit einer Rasterschrittweite von 1 Å berechnet worden waren, ergaben vergleichbare Ergebnisse. Das PLS-Regressions-Modell (*engl. partial least squares*) korrelierte diese molekularen Felder mit der Stärke der Veränderung der chemischen Verschiebung (CSP) des Wasserstoffatoms am Indolstickstoff von W ω 39 (siehe Ergebnisse, Tabelle 3-6). CoMFA-Gitterpunkte mit weniger als 2,0 kcal mol⁻¹ Energievariation gingen nicht in das PLS-Modell ein. Zur Validierung des Modells und zur Bestimmung der optimalen Anzahl an Regressions-Komponenten wurden kreuzvalidierte Modelle durch Auslassung jeder einzelnen WW-Domäne erzeugt (*leave-one-out*). Die optimale Anzahl an Regressions-Komponenten ist daran zu erkennen, dass sie den kreuzvalidierten Determinationskoeffizienten maximiert und SDEP minimiert (siehe 2.5.1).

Für die Darstellung der Beiträge der molekularen Felder zur Vorhersage des PLS-Modells wurden für das elektrostatische Feld die 10 % stärksten Beiträge und für das sterische Feld die 20 % stärksten Beiträge im dreidimensionalen Raum konturiert.

2.7.3.3 COMBINE – Comparative-Binding-Energy-Analyse

Die COMBINE-Analyse basiert auf der Beschreibung der zu vergleichenden WW-Domänen/Liganden-Komplexstrukturen durch alle paarweisen Interaktionsenergien – sowohl intra- als auch intermolekular – zwischen allen Aminosäureresten der Domäne und des Liganden (siehe auch Anhang D). Dazu wurden sowohl die elektrostatischen (Coulomb-Potential) als auch die van-der-Waals (Lennard-Jones-Potential) Interaktionsenergien zwischen allen Paaren von Aminosäureresten mit dem ANAL Modul und dem Kraftfeld von Cornell *et al.*, 1995, aus dem AMBER 7 Paket⁸ bestimmt. Um ein „Alignment“ der paar-

weisen Interaktionsenergien für Strukturen unterschiedlicher Längen zu ermöglichen wurden durch ein speziell dafür entwickeltes Python-Skript im Falle von fehlenden homologen Positionen neutrale Energierme (0,00) als „Platzhalter“ eingefügt. Bei entlang der Helixachse verschobenen Liganden wurden nur die in allen Komplexen überlappenden 9 virtuellen Kern-Ligandenpositionen in die Analyse mitaufgenommen.

Für die Hauptkomponenten-Analyse (PCA) der R-Gruppen Komplexe wurde jeder Komplex durch die intermolekularen Interaktionsenergieterme (je 2) zwischen 9 Liganden- und 36 Domänenresten beschrieben ($2 \times 9 \times 36 = 648$ Interaktions-Energierme). Die unterschiedlichen absoluten Variationsbereiche der Beiträge der elektrostatischen und der van-der-Waals Energierme wurden durch eine *block unscaled weights* Transformation der Energierme ausgeglichen. Die Hauptkomponenten-Analyse wurde mit GOLPE 4.5.1¹⁵ durchgeführt.

Für das prädiktive PLS-Regressions-Modell der Y-Gruppen Komplexe wurde jeder Komplex sowohl durch die intermolekularen Interaktionsenergieterme (je 2) zwischen 12 Liganden- und 36 Domänenresten ($2 \times 12 \times 36 = 864$) als auch durch die intramolekularen Interaktionsenergieterme zwischen Resten innerhalb der WW-Domänen ($36 \times 35/2 = 630$) beschrieben. Die letzteren sollten die Faltungstabilität der Domänen in dem COMBINE-Modell repräsentieren. Zusätzlich wurden die elektrostatischen Desolvatisierungs-Energierme der Domäne und des Liganden mit Hilfe des UHBD 6.1 Programms¹⁴ bestimmt. Dazu wurde die Differenz der elektrostatischen Wechselwirkung zwischen dem wässrigen Lösungsmittel und der freien Domäne (dem freien Liganden) bzw. der gebundenen Domäne (dem gebundenen Liganden) als Schätzwert für die Desolvatisierungs-Energie berechnet. Die unterschiedlichen absoluten Variationsbereiche der verschiedenen Beiträge wurden durch eine *block unscaled weights* Transformation ausgeglichen. Als Aktivität wurde die Stärke der Veränderung der chemischen Verschiebung des Wasserstoffatoms am Indolstickstoff von W ω 39 (CSP – siehe Ergebnisse, Tabelle 3-6) verwendet. Die PLS-Regressions-Analyse wurde mit GOLPE 4.5.1¹⁵ durchgeführt. Die optimale Modell-Dimension wurde durch Kreuzvalidierung bestimmt (*leave-one-out*, siehe 2.5.2). Durch *D-Optimal-Pre-Selection*- und *Fractional-Factorial-Design*-Variablen-Selektion wurden Variablen ohne signifikanten Beitrag zur Vorhersage aus dem finalen PLS-Modell entfernt.

Für die Kreuzvalidierung unter gleichzeitiger Auslassung von 4 Komplexen wurden zufällig je 2 Komplexe mit einer CSP im Bereich von 0,00–0,09 ppm bzw. 0,10–0,27 ppm aus dem Trainingsdatensatz entfernt. Diese Kreuzvalidierungsprozedur wurde 20-mal wiederholt.

2.7.3.4 Modellierung der hAF6-PDZ Domäne

Das Strukturmodell der hAF6-PDZ Domäne wurde auf der Strukturvorlage von hSNA1-PDZ [PDB: 2PDZ]¹⁶⁰ erstellt. Die beiden PDZ-Domänen hatten eine etwa 40 %ige Sequenzidentität gemeinsam. Das Modell wurde in SYBYL 6.6 (Tripos, St. Louise, USA) modelliert und einer Moleküldynamik Simulation bei 300 K für 500 ps im Kraftfeld von Cornell *et al.*, 1995, unterzogen. Die Qualität der Struktur wurde mit PROCHECK¹⁰⁵ verifiziert.

2.7.3.5 Moleküldynamik-Simulation des hERBIN-PDZ/METFVS_{COOH}-Komplexes

Die hERBIN-PDZ Domäne im Komplex mit dem C-Terminus (**METFVS_{COOH}**) des Angiotensin II Typ 2 Rezeptors AG22 [UP: P50052] wurde durch Homologie-Modellierung auf der Basis des hERBIN-PDZ/ERB2-Komplexes [PDB: 1MFG]²³ erstellt. Die Moleküldynamik-Simulation (MD) wurde mit AMBER 7⁸ und dem Kraftfeld von Cornell *et al.*, 1995, durchgeführt. Nach einer kurzen Energieminimierung im Vakuum (1000 Schritte) wurde die Ladung des Systems durch 13 Na⁺ und 10 Cl⁻ Ionen neutralisiert und der Komplex in eine octaedrische Wasserbox eingebettet. Ein harmonisches Potential mit einer Kraftkonstante von 30 kcal mol⁻¹ Å² fixierte die Atome des Komplexes zu Beginn der Equilibrierungsphase. Während dieser Phase wurde das System sukzessive von 100 K auf 300 K innerhalb von 25 ps bei konstantem Druck erhitzt. Danach wurde während kurzer Minimierungsphasen (jede 600 Schritte) das harmonische Potential auf die Atome des Komplexes in Schritten von 5 kcal mol⁻¹ Å² reduziert, bis die Atome des Komplexes komplett frei waren. Die Equilibrierung wurde mit einer 20 ps MD-Simulation abgeschlossen. Schließlich wurde der Komplex einer MD-Simulation für 1 ns bei konstanter Temperatur (300 K) und konstantem Druck (1013 mbar) unterzogen. Die Atomkoordinaten wurden alle 1 ps aufgezeichnet. Als „Endstruktur“ wurde die Struktur zwischen 700 und 900 ps der Trajektorie ausgewählt, die zur Durchschnittsstruktur die geringste mittlere quadratische Abweichung (RMSD) aufwies.

