

# Kapitel 3

## Ergebnisse

Ziel der vorliegenden Arbeit ist die theoretische Vorhersage der Wechselwirkungen von Protein-Interaktionsdomänen. Zu diesem Zweck wurden statistische Wechselwirkungsmodelle auf der Sequenz-, Tertiärstruktur- und Quartärstrukturebene erstellt (Tabelle 3-1). Trainiert wurden die Wechselwirkungsmodelle an Hand experimenteller Wechselwirkungsdaten, die exemplarisch für die WW- und PDZ-Domänenfamilien erhoben wurden. Die experimentellen Untersuchungen dazu wurden im Rahmen zweier Kooperationsprojekte mit der Molecular Libraries and Recognition Group des Instituts für Medizinische Immunologie, Universitätsklinikum Charité, von Dr. Livia Otte (WW-Domänen, siehe 3.1) und Dr. Prisca Boisguerin (PDZ-Domänen, siehe 3.4) durchgeführt. Das NMR-Screening der WW-Domänen wurde von Brigitte Schlegel, Dr. Peter Schmieder und Dr. Ricardo Pires durchgeführt (siehe 3.1).

In dieser Arbeit müssen zwei Arten von Modellen unterschieden werden: (i) Wechselwirkungsmodelle repräsentieren eine meist quantitative Beziehung zwischen Sequenz bzw. Struktur und Aktivität. (ii) Strukturmodelle repräsentieren eine Hypothese über die Raumstruktur eines Moleküls.

**Tabelle 3-1** Generierte Wechselwirkungsmodelle auf den unterschiedlichen Informationsebenen

Ebene <sup>a)</sup>	Beziehung <sup>b)</sup>	Vorhersage für	Vorhersage von	Methode <sup>c)</sup>	Dom. <sup>d)</sup>	Abs. <sup>e)</sup>
Sequenz	qualitativ	jede Ligandensequenz	Zugehörigkeit zur Y-, R <sub>α</sub> -, R <sub>β</sub> -, L-, poly-P- und poS/poT-Gruppe	Sequenzmotiv	WW	3.2.1 4.2
		jede Domänensequenz	Zugehörigkeit zur Y-Spezifitätsgruppe	pHMM <sup>47</sup>	WW	3.2.3 4.2
Tertiärstruktur	quantitativ	jede Domänenstruktur	Affinität gegenüber Y-Liganden	CoMFA <sup>42</sup>	WW	3.3.2
Quartärstruktur	quantitativ	ähnliche Komplexstruktur	Affinität ähnlicher Komplexe	COMBINE <sup>127</sup>	WW	3.3.3
Sequenz	quantitativ	jede Ligandensequenz	Affinität gegenüber hAF6-PDZ, hERBIN-PDZ und mSNA1-PDZ	QSP <b>neu entwickelt</b>	PDZ	3.5

<sup>a)</sup> Sequenz (Primärstruktur), Tertiärstruktur (Raumstruktur), Quartärstruktur (Komplexstruktur).

<sup>b)</sup> Art der Sequenz/Struktur-Aktivitäts-Beziehungen: qualitativ bzw. quantitativ.

<sup>c)</sup> pHMM: Profile-Hidden-Markov-Modell; CoMFA: Comparative-Molecular-Field-Analyse; COMBINE: Comparative-Binding-Energy-Analyse; QSP: Quantitative Spezifitäts-Profile.

<sup>d)</sup> Domänenfamilie, für welche die Wechselwirkungsmodelle exemplarisch entwickelt wurden.

<sup>e)</sup> Abschnitt, in dem die Wechselwirkungsmodelle entwickelt und die Vorhersagen verifiziert werden.

Auf der **Sequenzebene** wurde ein **qualitatives Sequenz-Aktivitäts-Modell** in Form eines Klassifizierungssystems für WW-Domänen erstellt (Tabelle 3-1, 1. und 2. Zeile). Dazu wurden Spezifitätsgruppen an Hand experimenteller Wechselwirkungsdaten definiert (siehe 3.2.2) und sowohl auf der Seite der Liganden (siehe 3.2.1) als auch auf der Seite der Domänen Erkennungsmotive spezifiziert (siehe 3.2.3). Prinzipiell ermöglicht solch ein Wechselwirkungsmodell für jede Liganden- bzw. Domänensequenz die Zugehörigkeit zu denjenigen Spezifitätsgruppen vorherzusagen, die vorher experimentell bestimmt wurden. Im Falle der WW-Domänen ist dies für die Y-, R<sub>α</sub>-, R<sub>β</sub>- und poS/poT-Spezifitätsgruppen gelungen. Die Zugehörigkeit einer experimentell nicht untersuchten WW-Domäne zu einer dieser Spezifitätsgruppen wurde an Hand der Ähnlichkeit ihrer Sequenz gegenüber den gruppenspezifischen *Profile-Hidden-Markov-Modellen* (pHMM)<sup>47</sup> bzw. Konsensusmotiven beurteilt. Die vorhergesagten tyrosinspezifischen WW-Domänen konnten anschließend experimentell bestätigt werden (siehe 3.2.3.3).

Auf der **Strukturebene** wurde ein **quantitatives Struktur-Aktivitäts-Modell** der Tyrosin-Spezifität von WW-Domänen (siehe 3.3.2.1) durch die *Comparative-Molecular-Field-Analyse* (CoMFA)<sup>42</sup> erstellt (Tabelle 3-1, 3. Zeile). Dieses Modell erlaubt ausgehend von der Struktur der WW-Domäne die Vorhersage der Wechselwirkungsaktivität gegenüber dem experimentell untersuchten, tyrosinhaltigen Liganden (siehe 3.3.2.2). Folglich wurden in dieser Arbeit im Gegensatz zur klassischen Anwendung der CoMFA-Methode nicht die Strukturen der Liganden, sondern die Strukturen der „Rezeptoren“ respektive WW-Domänen analysiert. Da nicht genügend experimentell bestimmte WW-Domänenstrukturen zur Verfügung standen, wurde das CoMFA-Modell ausschließlich auf modellierten Domänenstrukturen trainiert. Dadurch demonstriert diese Arbeit zugleich, dass die strukturbasierten Wechselwirkungsmodelle – über Homologie-Modellierung der Domänenstruktur (siehe 3.3.1.1) – auch für diejenigen Domänen zugänglich sind, für die keine experimentell bestimmten Strukturen existieren.

Auf der **Komplexstrukturebene** wurde ein **quantitatives Struktur-Aktivitäts-Modell** der Tyrosin-Spezifität von WW-Domänen (siehe 3.3.3.2) durch die *Comparative-Binding-Energy-Analyse* (COMBINE)<sup>127</sup> erstellt (Tabelle 3-1, 4. Zeile). Für ähnliche WW-Domänen erlaubte dieses Modell die Vorhersage der Stärke der Wechselwirkung nicht nur im Komplex mit den experimentell untersuchten Liganden, sondern auch mit ähnlichen Liganden. Folglich ermöglichte das COMBINE-Modell das Design eines höher affinen Liganden. Dieser konnte anschließend experimentell bestätigt werden (siehe 3.3.3.3). Da das

COMBINE-Modell ebenfalls ausschließlich auf modellierten Strukturen der WW-Domänen/Liganden-Komplexe basierte (siehe 3.3.1.4), demonstriert diese Arbeit, dass die komplexstrukturbasierte Vorhersage – über die Homologie-Modellierung der Komplexe – auch für Domänen/Liganden-Kombinationen zugänglich ist, für die nur die Sequenz bekannt ist.

Am Beispiel der PDZ-Domänen wurden schließlich **quantitative Sequenz-Aktivitäts-Modelle** erstellt (Tabelle 3-1, 5. Zeile). Dazu wurde die Methode der *Quantitativen Spezifitäts-Profile* entwickelt, um die Spezifität der Domänen/Liganden-Wechselwirkung in Form einer sequenzabhängigen Affinitätsfunktion quantitativ zu beschreiben (siehe 3.5). Prinzipiell erlauben QSP-Modelle die Affinität für jeden Liganden gegenüber den experimentell untersuchten Domänen vorherzusagen. In dieser Arbeit konnten QSP-Modelle für die PDZ-Domänen aus den drei Proteinen *hAF6*, *hERBIN* sowie *mSNA1* erstellt werden. Sie ermöglichen die Vorhersage der Affinität gegenüber einem potentiellen Liganden-Sequenzraum von 130 321 Tetrapeptiden (siehe 3.5.2). Mit Hilfe dieser QSP-Modelle konnten für diese PDZ-Domänen die optimalen peptidischen Liganden, die sogenannten *Superbinder*, designiert werden. Die vorhergesagten Superbinder wurden anschließend experimentell bestätigt (siehe 3.5.3). Schließlich erlaubten die QSP-Modelle auch die quantitative Analyse der Selektivität der untersuchten PDZ-Domänen (siehe 3.5.4).

### 3.1 Experimentelle Charakterisierung der Wechselwirkungen von 42 WW-Domänen

Um ein repräsentatives Spektrum der allgemein von WW-Domänen gebundenen Liganden zu erhalten (WW-Screening, Tabelle 3-3), wurden die Wechselwirkungen von 42 synthetisch hergestellten WW-Domänen (Tabelle 2-1) und 9 verschiedenen prolinreichen Liganden (Tabelle 3-2, siehe auch Tabelle 2-2) untersucht.

Kooperationsprojekt mit Dr. Livia Otte, Molecular Libraries and Recognition Group, Institut für Medizinische Immunologie, Universitätsklinikum Charité.

Für dieses Screening wurden zwei komplementäre Methoden verwendet. Zum einen wurden mittels kernmagnetischer Resonanzspektroskopie (NMR-Spektroskopie) die Wechselwirkungen der WW-Domänen mit 3 prolinreichen Modellliganden in Lösung analysiert (Tabelle 3-3, Y-, *poY*- und *poT*-Modellliganden; siehe auch Tabelle 3-2). Gleichzeitig erlaubte diese Methode auch die Bestimmung des Faltungszustands der synthetisch hergestellten WW-Domänen. Zum anderen wurden die WW-Domänen/Liganden-Wechselwirkungen durch Screening festphasegebundener Proteinbibliotheken untersucht. Die 42 an einen Cellulosechip gekoppelten WW-Domänen (WW-Array) wurden hierfür mit 6 verschiedenen

Liganden werden nach dem charakteristischsten Aminosäurerest benannt (Tabelle 3-2).

**Tabelle 3-2** Übersicht der verwendeten peptidischen Liganden (siehe auch Tabelle 2-2)

	Ligand <sup>a)</sup>	Aminosäuresequenz <sup>b)</sup>	Anwendung <sup>c)</sup>	UP-ID <sup>d)</sup>	Beginn <sup>e)</sup>	Ende <sup>e)</sup>
NMR	Y-Modellligand	GPPPPYG	NMR/UV	Modellligand		
	poY-Modellligand	GPPPPpoYG	NMR/UV	Modellligand		
	poT-Modellligand†	VLpoTTPDRL	NMR	Modellligand		
WW-Array	Y <sub>2</sub> -Ligand‡	GTPPPPYTVGTTPPPYTVG	WW-Array	P97764	170 - 179	
	poY <sub>2</sub> -Ligand‡	GTPPPPpoYTVGTTPPPpoYTVG	WW-Array	P97764	170 - 179	
	R-Ligand	PPGPPPRGPPPR	WW-Array	Q9Y2W2	479 - 490	
	R <sub>2</sub> -Ligand‡	PPGPPPRGPPPRPPGPPPRGPPPR	WW-Array	Q9Y2W2	479 - 490	
	L-Ligand	APPTPPPLPP	WW-Array/SA/Modell	Q05859	872 - 881	
	L/P-Ligand	APPTPPPLPPPLIPPPPLPP	WW-Array	Q05859	872 - 892	
andere	Y-Ligand	GTPPPPYTVG	SA/Modell	P97764	170 - 179	
	R <sub>k</sub> -Ligand*	PPGPPPRGPPP	SA/Modell	Q9Y2W2	479 - 489	
	L/P <sub>k</sub> -Ligand*	PPPLIPPPPLPP	SA	Q05859	880 - 892	
	poly-P-Ligand	PLIPPPPLPP	Modell	Q05859	882 - 891	

† Originalsequenz **ALTPPDRL** aus hPKC $\gamma$  653 - 660 [UP: P05129].

‡ Tandemwiederholung der Originalsequenz im synthetisierten Peptid: Y<sub>2</sub> = doppelter Y-Ligand, poY<sub>2</sub> = doppelter poY-Ligand und R<sub>2</sub> = doppelter R-Ligand.

\* Verkürzte Varianten eines Liganden: R<sub>k</sub>-Ligand = kurzer R-Ligand und L/P<sub>k</sub>-Ligand = kurzer L/P-Ligand.

<sup>a)</sup> Bezeichnung des Liganden nach dem charakteristischsten Aminosäurerest. So enthält beispielsweise der Y-Ligand ein Tyrosin (Y).

<sup>b)</sup> Aminosäuresequenz des Liganden. poY = Phosphotyrosin und poT = Phosphothreonin.

<sup>c)</sup> Verwendung des Liganden für das NMR-Screening (NMR), für das WW-Array Screening (WW-Array), für die UV-Fluoreszenzspektroskopie-Messungen (UV), für die Substitutionsanalysen (SA) und für die Modellierung der Komplexstrukturen.

<sup>d)</sup> Primärer UniProt<sup>10</sup> Zugriffscode des Gesamtproteins.

<sup>e)</sup> Position der ersten und letzten Aminosäure der Ligandensequenz innerhalb des Gesamtproteins.

prolinreichen Peptiden aus natürlichen Proteinen inkubiert (Tabelle 3-3, Spalten Y<sub>2</sub>, poY<sub>2</sub>, R, R<sub>2</sub>, L und L/P; siehe auch Tabelle 3-2).

Insgesamt lagen 31 der 42 WW-Domänen in Lösung ohne Ligand gefaltet vor (siehe 2.3.1.1). hPQBP1-WW faltete sich nach Zugabe eines Gemischs prolinreicher Modellliganden, mit denen auch hSMURF2-WW1 Wechselwirkungen zeigte (Daten nicht gezeigt). Bis auf yYJQ8-WW interagierten alle 30 gefalteten WW-Domänen, sowie hPQBP1-WW, mit mindestens einem der getesteten prolinreichen Liganden. Dies stützte die Hypothese, dass WW-Domänen generell spezifisch mit prolinreichen Liganden interagieren.

Vor allem aber zeigte das WW-Screening, dass bestimmte Gruppen von WW-Domänen ausschließlich bestimmte Gruppen von Liganden erkennen. Auf Basis der exklusiv erkannten Liganden konnten die 42 WW-Domänen größtenteils in vorläufige Spezifitätsgruppen eingeteilt werden (Tabelle 3-3, Spalte Vorläufige Klassifizierung).

Die größte Gruppe wird durch Domänen gebildet, die fast ausschließlich mit tyrosin(Y)- bzw. phosphotyrosin(poY)-haltigen Liganden interagieren (Y-Spezifitätsgruppe). Sie entspricht der Gruppe I der klassischen WW-Domänen-Klassifizierung (siehe Tabelle 1-3).<sup>177</sup>

Neben dieser Gruppe wurden WW-Domänen identifiziert, die vorrangig argininhaltige Liganden erkennen (R-Spezifitätsgruppe). Sie entspricht teilweise der klassischen Gruppe III. Drei WW-Domänen dieser Gruppe interagierten indessen auch mit dem L/P-Liganden.

Peptidsequenzen werden nach der modifizierten Seefeld-Convention-2001-Nomenklatur<sup>1</sup> angegeben (siehe 2.1.3). Dabei bezeichnet poY Phosphotyrosin, poS Phosphoserin und poT Phosphothreonin.

Spezifitätsgruppen werden nach dem charakteristischsten spezifitätsbestimmenden Aminosäurerest des Liganden benannt. WW-Domänen der Y-Spezifitätsgruppe binden beispielsweise vorrangig Tyrosin(Y)-haltige Liganden.

Tabelle 3-3 WW-Screening: Analyse der WW-Domänen/Liganden-Interaktionen

Die WW-Domänen wurden in 5 vorläufige Spezifitätsgruppen (siehe Spalte Vorläufige Klassifizierung) auf der Basis der experimentell identifizierten Wechselwirkungen eingeteilt.

Domäne <sup>a)</sup>	UP-ID <sup>b)</sup>	NMR-Screening <sup>c)</sup>			WW-Array-Screening <sup>d)</sup>						Vorläufige <sup>e)</sup> Klassifizierung
		CSP [ppm] <sup>c)</sup>			relative Signalintensität						
	Peptide <sup>b)</sup>	Y	poY	poT	Y <sub>2</sub>	poY <sub>2</sub>	R	R <sub>2</sub>	L	L/P	Gruppe
hWWP3-WW1	O00309	0,27	0,23	0,00	+	+/-	-	+	-	-	Y
yRSP5-WW3	P39940	0,26	0,24	0,00	+	+	-	-	-	-	Y
hNEDD4-WW2	P46934	0,24	0,13	0,00	++	++	-	-	-	-	Y
yRSP5-WW2	P39940	0,24	0,16	0,00	++	++	-	-	-	-	Y
hNEDD4-WW3	P46934	0,23	0,25	0,00	++	++	-	-	-	-	Y
hYAP65-WW	P46937	0,21	0,13	0,00	+	-	-	-	-	-	Y
yRSP5-WW1	P39940	0,20	0,14	0,00	++	+/-	-	-	-	-	Y
rrITCHY-WW1	O54971	0,18	0,21	0,00	++	++	-	-	-	-	Y
rrITCHY-WW3	O54971	0,10	0,06	0,00	++	-	-	-	-	+/-	Y
hNEDD4-WW4	P46934	0,03	0,05	0,00	+	-	-	-	-	-	Y
hNEDD4-WW1	P46934	shift	0,03	0,00	+	-	-	-	-	-	Y
hWWP3-WW2	O00309	shift	shift	keine	-	-	-	+/-	-	+/-	Y?
hSMURF1-WW2	Q9UJT8	shift	shift	keine	-	-	-	-	-	-	Y?
hWVOX-WW2	Q9NZC7	keine	shift	keine	-	-	-	-	-	-	Y?
hHYP109-WW1	Q9NT81	shift	shift	n.b.	-	-	++	++	-	++	R
hFE65-WW	O00213	0,02	0,00	n.b.	-	-	++	++	-	++	R
hPQBP1-WW	O60828	0,00	0,00	0,00	-	-	++	++	-	++	R
hFBP21-WW1	O75554	0,00	0,00	0,31	-	-	++	++	-	+/-	R
hFBP21-WW2	O75554	0,01	shift	0,03	-	-	++	++	-	+	R
yYFB0-WW	P43582	0,03	0,01	0,00	-	-	+	++	-	+	R?
hHYP109-WW2	Q9NT81	0,02	0,04	0,00	-	-	-	+	-	-	R?
hFBP11-WW1	O75400	0,02	0,02	0,00	-	-	-	-	++	++	L
hFBP11-WW2	O75400	0,05	0,08	0,01	-	++	+/-	+/-	++	++	L
yPRP40-WW2	P33203	0,03	shift	0,00	-	-	-	-	-	++	poly-P
hGAS7-WW	O60861	0,02	0,03	0,00	-	-	-	+/-	-	++	poly-P
hPIN1-WW	Q13526	0,01	0,01	shift	-	-	-	-	-	-	poS/poT
yESS1-WW	P22696	0,00	0,00	0,01	-	-	-	-	-	-	poS/poT
yPRP40-WW1	P33203	0,00	0,01	0,29	-	-	-	-	-	-	poS/poT?
yYPR152C-WW	Q06525	0,08	0,09	0,32	-	-	-	-	-	-	poS/poT?
hCA150-WW1	O14776	0,01	0,02	0,03	-	-	-	-	-	-	poS/poT?
hSMURF2-WW1	Q9H260	0,00	0,00	0,00	-	-	-	-	-	-	?
yYJQ8-WW	P46995	keine	keine	n.b.	-	-	-	-	-	-	?
hDYSTRO-WW	P11532	0,00	0,01	0,00	-	-	-	-	-	-	ungefaltet
hPEPP2-WW2	Q9HAU0	keine	keine	keine	-	-	-	-	-	-	ungefaltet
hPEPP2-WW1	Q9HAU0	0,00	0,00	0,00	-	-	-	-	-	-	ungefaltet
hWVOX-WW1	Q9NZC7	0,00	0,00	0,00	-	-	-	-	-	-	ungefaltet
hIQGAP1-WW	P46940	0,00	0,00	0,00	-	-	-	-	-	-	ungefaltet
hCA150-WW3	O14776	0,00	0,00	0,00	-	-	-	-	-	-	ungefaltet
hSMURF1-WW1	Q9UJT8	0,00	0,00	0,00	-	-	-	-	-	-	ungefaltet
hKIAA1052-WW	Q9UPV0	0,00	0,00	0,00	-	-	-	-	-	-	ungefaltet
ySSM4-WW	P40318	keine	keine	keine	-	-	-	-	-	-	ungefaltet
yTIN1-WW	P53076	keine	keine	keine	-	-	-	-	-	-	ungefaltet

Die Stärke der identifizierten Wechselwirkung ist durch die Intensität der Hintergrundfarbe kodiert: starke Wechselwirkung durch intensive Färbung und schwache durch blasse Färbung (n.b. = nicht bestimmt).

<sup>a)</sup> Domänennamen gebildet aus Organismus-Präfix, Proteinnamen und Domänen-Suffix (siehe 2.1.1).

Grau hinterlegte Domänen lagen ohne Ligand ungefaltet vor (siehe 2.3.1.1).

<sup>b)</sup> Primärer UniProt<sup>10</sup> Zugriffscode des die WW-Domäne enthaltenden Proteins.

<sup>c)</sup> Die Stärke der Wechselwirkungen wurde an Hand der Größe der Veränderung der chemischen Verschiebung (CSP) des Wasserstoffatoms am Indolstickstoff von W $\omega$ 39 in den 1D-<sup>1</sup>H-NMR-Spektren nach Zugabe des Y-, poY- oder poT-Modellliganden bewertet (siehe 2.3.1.2 und Anhang C). Die Wechselwirkungen wurden in starke (CSP  $\geq$  0,150) und schwache (0,030  $\leq$  CSP < 0,150) eingeteilt. Parallel dazu wurden die Veränderungen der chemischen Verschiebungen in anderen Bereichen der NMR-Spektren qualitativ bewertet. Für WW-Domänen ohne Tryptophan an  $\omega$ 39 (siehe 1.3) bedeutet keine (keine Bindung) bzw. shift (schwache Bindung), dass keine bzw. nur Veränderungen in anderen

(fortgesetzt)

(Fortsetzung der Legende zu Tabelle 3-3)

Bereichen des Spektrums identifiziert wurden. Für den letzteren Fall werden WW-Domänen mit Tryptophan an  $\omega 39$  aber ohne eine quantifizierbare CSP des Wasserstoffatoms am Indolstickstoff von  $W\omega 39$  ebenfalls durch *shift* (keine Bindung) gekennzeichnet.

<sup>d</sup> Die Stärke der im WW-Array-Screening identifizierten Wechselwirkungen wurde an Hand der Signalintensität nach der Inkubation des WW-Arrays mit den  $Y_2^-$ ,  $poY_2^-$ , R-,  $R_2^-$ , L- und L/P-Liganden bewertet. Die Stärke wurde in 4 Klassen, ++ starke Wechselwirkung, + schwache Wechselwirkung, +/- nicht eindeutige Wechselwirkung und – keine Wechselwirkung eingeteilt (siehe 2.6.1).

<sup>e</sup> Zuordnung der WW-Domänen zu vorläufigen Spezifitätsgruppen auf Basis des WW-Screenings.

<sup>f</sup> Untersuchte peptidische Liganden siehe Tabelle 3-2, Seite 52.

Der leuzinhaltige L-Ligand zeigte dagegen ein verhältnismäßig selektives Verhalten: nur zwei WW-Domänen erkannten diesen Liganden (L-Spezifitätsgruppe). Die L-Spezifitätsgruppe entspricht teilweise der klassischen Gruppe II.

Zwei weitere Domänen interagierten fast ausschließlich mit dem L/P-Liganden, aber nicht mit seiner verkürzten Variante, dem L-Liganden. Da sich der L/P-Ligand vom L-Liganden vor allem durch einen kontinuierlichen Abschnitt aus 5 Prolinen unterscheidet, wurden diese beiden WW-Domänen vorläufig zur poly-P-Spezifitätsgruppe zusammengefasst.

Für die beiden Vertreter der klassischen Gruppe IV, *hPIN1*-WW und *yESS1*-WW, ergab das WW-Screening kein charakteristisches Wechselwirkungsprofil, was auf das Fehlen geeigneter Liganden zurückzuführen ist. Überraschenderweise zeigten beide Domänen mit dem phosphothreonin(poT)-haltigen poT-Modellligand nur recht schwache Wechselwirkungen. Daher wurde für diese beiden WW-Domänen zusätzlich ein Screening mit Phosphopeptid-Bibliotheken durchgeführt (Daten nicht gezeigt, siehe Otte und Wiedemann *et al.*, 2003). Für beide WW-Domänen konnte dadurch die ausschließliche Erkennung von phosphothreonin(poT)- und phosphoserin(poS)-haltigen Peptiden gezeigt werden.<sup>130,193</sup> Deshalb wurden die beiden WW-Domänen zur poS/poT-Spezifitätsgruppe zusammengefasst.

Interessanterweise zeigten auch einige WW-Domänen der Y-Spezifitätsgruppe im WW-Screening die Fähigkeit zur Wechselwirkung mit der Phosphotyrosin(poY)-Variante des Y-Liganden.<sup>36,178</sup> Um diesen Befund näher zu analysieren, wurde die Bindungsaffinität repräsentativer WW-Domänen der Y-Spezifitätsgruppe gegenüber den phosphorylierten bzw. nicht-phosphorylierten Y-Modellliganden mittels UV-Fluoreszenzspektroskopie bestimmt (Tabelle 3-4).

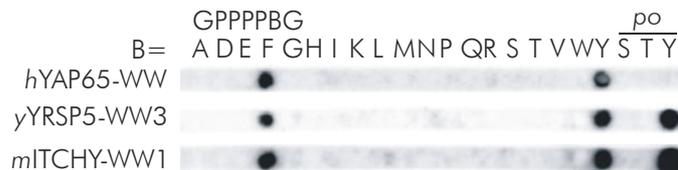
Die gemessenen Dissoziationskonstanten lagen im typischen Bereich zwischen 1  $\mu\text{M}$  und 100  $\mu\text{M}$ .<sup>177</sup> Einige WW-Domänen (*hNEDD4*-WW2 und *hWWP3*-WW1) waren damit in der Lage, die Phosphotyrosin-Variante des Modellliganden zu binden, wenn auch mit einer um einen Faktor 2 bis 7 geringeren Affinität. *mITCHY*-WW1 interagierte mit der phosphorylier-

**Tabelle 3-4 Bindungsaffinität bestimmt durch UV-Fluoreszenzspektroskopie**

Einige WW-Domänen binden phosphotyrosin(poY)-haltige Liganden mit ähnlicher Affinität wie die entsprechenden tyrosinhaltigen Liganden.

Domäne	$K_d$ [ $\mu$ M]	
	GPPPPYG	GPPPPpoYG
<i>m</i> ITCHY-WW1	57 ( 53 - 61 )	32 ( 28 - 36 )
<i>h</i> NEDD4-WW2	30 ( 28 - 33 )	293 ( 280 - 306 )
<i>y</i> RSP5-WW2	9 ( 5 - 17 )	n.b.
<i>h</i> WWP3-WW1	47 ( 36 - 62 )	114 ( 82 - 160 )

In Klammern ist das  $\pm 1\sigma$  Intervall der Messwiederholungen auf logarithmischer Skala angegeben (n.b. nicht bestimmt).

**Abbildung 3-1 Modellpeptidbibliothek GPPPPBG inkubiert mit WW-Domänen der Y-Spezifitätsgruppe**

Die Modellpeptidbibliothek zeigt, dass einigen WW-Domänen der vorläufigen Y-Spezifitätsgruppe auch mit phosphotyrosin(poY)-haltigen Liganden interagieren können. Starke Signalintensitäten (Schwärzung) zeigen starke Wechselwirkung zwischen dem festphasegebundenen Peptid und der jeweiligen WW-Domäne an. Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003.

ten Variante sogar um einen Faktor 2 stärker als mit der nicht-phosphorylierten. Diese Ergebnisse spiegeln sich auch in den Substitutionsanalysen eines Modellliganden wieder (Abbildung 3-1). Zusammengenommen deuten diese Ergebnisse darauf hin, dass eine Untergruppe der Y-Spezifitätsgruppe die Fähigkeit besitzt, mit phosphotyrosinhaltigen Liganden zu interagieren (z.B. *m*ITCHY-WW1, *h*WWP3-WW1 und *h*NEDD4-WW2). Für andere Domänen der Y-Spezifitätsgruppe führt die Phosphorylierung ihrer Liganden dagegen zum Verlust der Wechselwirkung (z.B. *h*YAP65-WW).

Damit zeigte das WW-Screening, dass unterschiedliche WW-Domänen auch unterschiedliche Ligandenpräferenzen besitzen. Daher sollten durch die Erstellung theoretischer Wechselwirkungsmodelle diejenigen Sequenz- bzw. Strukturmerkmale identifiziert werden, die für diese Unterschiede in den Ligandenpräferenzen verantwortlich sind.

## 3.2 Qualitative Sequenz-Aktivitäts-Beziehungen

Um eine qualitative Beziehung zwischen der Sequenz des einen Wechselwirkungspartners und dessen Aktivität, nämlich der Bindung eines definierten anderen Wechselwirkungspartners, herzustellen, wurde nach Aminosäureresten gesucht, die mit der Erkennung eines definierten Liganden bzw. einer definierten Domäne korrelieren. Die Gesamtheit dieser spezifitätsbestimmenden Reste/Schlüsselreste ergibt schließlich das Erkennungsmotiv bzw. Epitop.

Zunächst wurden die Erkennungsmotive auf der Seite der Liganden experimentell durch Substitutionsanalysen bestimmt (siehe 3.2.1), um damit die Domänen in experimentell definierten Spezifitätsgruppen zusammenzufassen (siehe 3.2.2). Auf dieser Basis wurden anschließend die Epitope in Domänen identifiziert (siehe 3.2.3). Schließlich konnte mit Hilfe der erzeugten Domänensequenzprofile die Spezifität für nicht-untersuchte WW-Domänen vorhergesagt werden (siehe 3.2.3.3).

### 3.2.1 Experimentelle Bestimmung der Liganden-Epitope

Mit Hilfe von Substitutionsanalysen wurden die für die Wechselwirkung mit einer bestimmten Domäne essentiellen Schlüsselreste in den Liganden identifiziert. Die Schlüsselreste treten in Substitutionsanalysen als diejenigen Aminosäurereste hervor, die nicht oder nur durch definierte Aminosäuretypen austauschbar sind. Die sequentielle Abfolge der Schlüsselreste ergibt schließlich das für die Wechselwirkung mit einer bestimmten Domäne spezifische Erkennungsmotiv (Epitop) im Liganden.

Basierend auf den Ergebnissen des WW-Screenings wurden für jede WW-Domäne der vorläufigen R-, L-, poly-P- und *poS/poT*-Spezifitätsgruppen sowie für einige ausgewählte Domänen der Y-Spezifitätsgruppe Substitutionsanalysen der entsprechenden Liganden durchgeführt.

#### 3.2.1.1 Y-Spezifitätsgruppe: tyrosinspezifische WW-Domänen

Stellvertretend für die WW-Domänen der Y-Spezifitätsgruppe ist in Abbildung 3-2 die Substitutionsanalyse des Y-Liganden (Tabelle 3-2, Seite 52) für die *hYAP65*-WW Domäne dargestellt. Sie zeigt das für die Y-Spezifitätsgruppe charakteristische Erkennungsmotiv **xPPxY**.<sup>138</sup> Für einige WW-Domänen der Y-Spezifitätsgruppe muss das Erkennungsmotiv allerdings auf **xPPx(Y/poY)** erweitert werden (siehe Abbildung 3-1 und Tabelle 3-4).

Kooperationsprojekt mit Dr. Livia Otte, Molecular Libraries and Recognition Group, Institut für Medizinische Immunologie, Universitätsklinikum Charité.

Die Nomenklatur der Sequenzmotive folgt der modifizierten Seefeld-Convention-2001-Nomenklatur<sup>1</sup> (siehe 2.1.3).

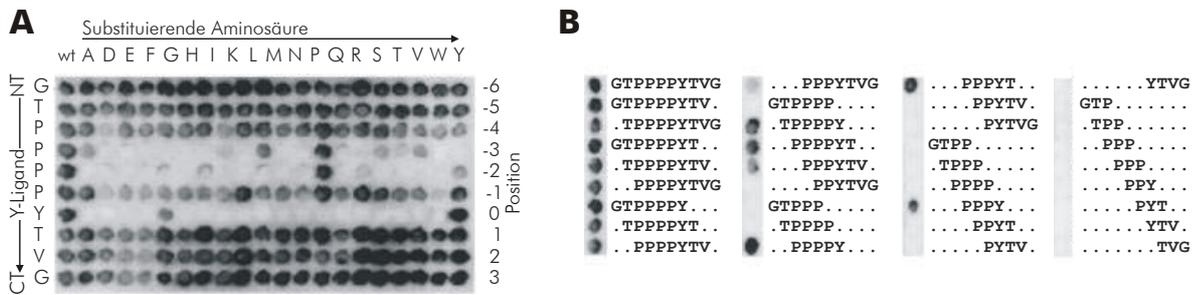


Abbildung 3-2 Substitutions- und Längenanalyse des Y-Liganden für hYAP65-WW

Für WW-Domänen der Y-Spezifitätsgruppe wurde das Erkennungsmotiv **xPPxY** identifiziert.

(A) Substitutionsanalyse des Y-Liganden **GTTPPPYTVG**. Eine Substitutionsanalyse ist eine „tabellenartige“ Peptidbibliothek. Jede „Tabellenzelle“ entspricht einem definierten Einzelsubstitutionsanalogon des Ausgangs-peptids (wt = Wildtyp). Die Zeile bestimmt dabei, welche Position des Peptids substituiert wird, und die Spalte, welcher Aminosäuretyp damit an der jeweiligen Position zu finden ist. Die Aminosäuresequenz des Ausgangs-peptids ist auf der linken Seite den Zeilen entsprechend vom N-Terminus (oben) zum C-Terminus (unten) dargestellt. Die Positionen (rechts) sind ausgehend vom charakteristischsten spezifitätsbestimmenden Rest (hier Y0) durchnummeriert. Der Aminosäuretyp (alle proteinogenen Aminosäuren außer Cystein), durch den die jeweilige Position substituiert wird, ist oberhalb der entsprechenden Spalten angegeben. In allen Zellen der ersten Spalte (wt) befindet sich das unveränderte Ausgangs-peptid. Gebundene WW-Domänen werden durch ein Chemilumineszenzsystem nachgewiesen: eine hohe Dichte an gebundenen WW-Domänen führt zu einem intensiven Signal, welches durch die Intensität der Schwärzung dargestellt ist. Schlüsselreste sind daran zu erkennen, dass nur bestimmte Aminosäuren in der jeweiligen Position zur spezifischen Bindung durch die Domäne führen (nur wenige Zellen der Zeile sind geschwärzt). Positionen, bei denen alle Aminosäuretypen ein etwa gleich starkes Signal liefern, sind nicht an der spezifischen Wechselwirkung beteiligt. Das spezifische Erkennungsmotiv ergibt sich aus der Kombination aller Schlüsselreste. Abbildung modifiziert nach Pires *et al.*, 2001.

(B) Längenanalyse des Y-Liganden **GTTPPPYTVG**. Abwechselnd vom N- und C-Terminus verkürzte Varianten des Ausgangs-peptids werden auf Bindung mit der Domäne untersucht. Damit kann das Kernmotiv (das kleinste mögliche Erkennungsmotiv) identifiziert und der Einfluss flankierender Aminosäuren bestimmt werden. Die Inkubation und Analyse gleicht der von Substitutionsanalysen. Abbildung modifiziert nach Pires *et al.*, 2001.

### 3.2.1.2 R-Spezifitätsgruppen: argininspezifische WW-Domänen

Anders als die WW-Domänen der Y-Spezifitätsgruppe zeigten die WW-Domänen der vorläufigen R-Spezifitätsgruppe kein einheitliches Erkennungsmotiv in den Substitutionsanalysen des R<sub>k</sub>-Liganden (siehe Tabelle 3-2).

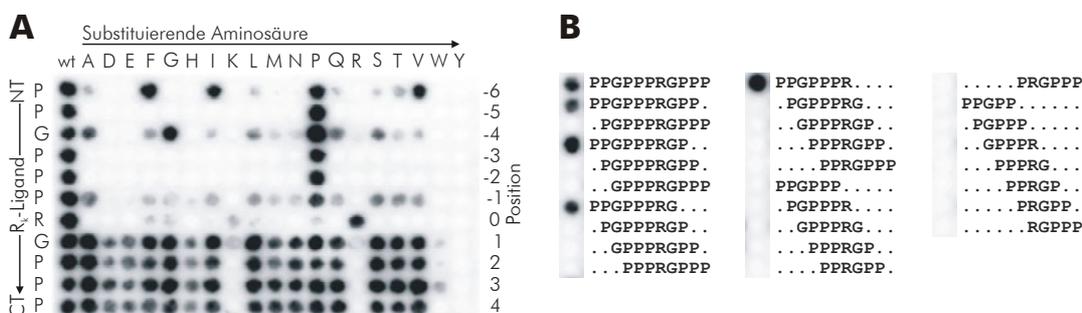


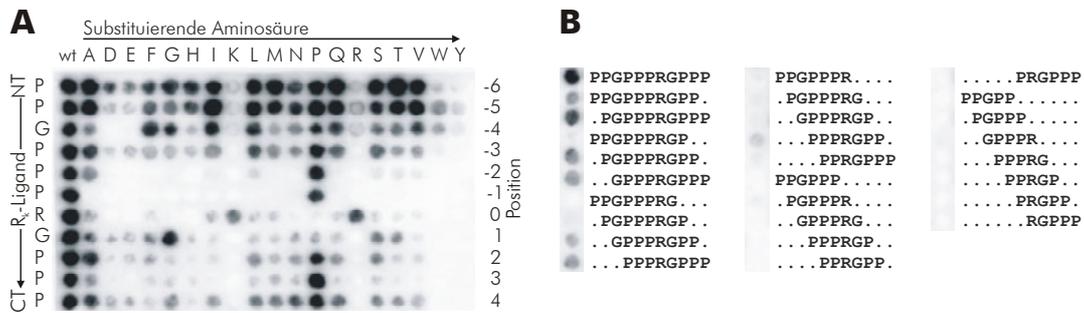
Abbildung 3-3 Substitutions- und Längenanalyse des R<sub>k</sub>-Liganden für hFE65-WW

Für WW-Domänen der R<sub>g</sub>-Spezifitätsgruppe wurde das Erkennungsmotiv **(p/Φ)P(p/g)PPpR** identifiziert.

(A) Substitutionsanalyse des R<sub>k</sub>-Liganden **PPGPPPRGPPP** (siehe auch Legende zu Abbildung 3-2).

Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003.

(B) Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.



**Abbildung 3-4** Substitutions- und Längenanalyse des R<sub>k</sub>-Liganden für hFBP21-WW2

Für WW-Domänen der R<sub>b</sub>-Spezifitätsgruppe wurde das Erkennungsmotiv **pPPRgpPp** identifiziert.

(A) Substitutionsanalyse des R<sub>k</sub>-Liganden **PPGPPPRGPPP** (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003.

(B) Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

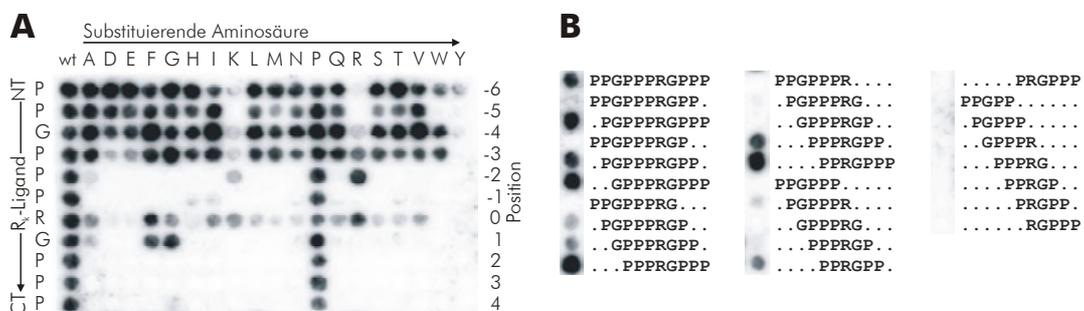
Φ hydrophobe Aminosäuren<sup>1</sup> (siehe 2.1.3).

Für die WW-Domänen hFE65-WW, hPQBP1-WW und hHYP109-WW1 ergaben die Substitutionsanalysen das Erkennungsmotiv **(p/Φ) P (p/g) PpPpR** (Abbildung 3-3). Im Gegensatz dazu waren für die beiden WW-Domänen aus hFBP21 (Abbildung 3-4) sowie für yYFB0-WW (Abbildung 3-5) auch Reste C-terminal des Arginins für eine Bindung nötig.

Für hFBP21-WW1 und hFBP21-WW2 war zusätzlich ein Glyzin an Position 1 essentiell, während yYFB0-WW an dieser Position auch ein Prolin oder Phenylalanin tolerierte. Somit ergab sich für hFBP21-WW1 und hFBP21-WW2 das Erkennungsmotiv **pPPRgpPp**, wobei die Längenanalyse eine deutliche Präferenz für längere Peptide zeigte. Für yYFB0-WW ergab sich das Motiv **(P/R) P (r/f) (P/G/F) PPP**.

hHYP109-WW2 zeigte im WW-Screening nur eine schwache Wechselwirkung mit dem R<sub>2</sub>-Liganden (Tabelle 3-3). In der Substitutionsanalyse des R<sub>k</sub>-Liganden interagierte hHYP109-WW2 nur mit Substitutionsanaloga, bei denen das Glyzin an Position -4 durch eine Reihe anderer Aminosäuren ausgetauscht wurde (Abbildung 3-6).

Auf Basis dieser Ergebnisse kann die vorläufige R-Spezifitätsgruppe in zwei Untergruppen, R<sub>a</sub> (hFE65-WW, hPQBP1-WW und hHYP109-WW1) und R<sub>b</sub> (hFBP21-WW1 und hFBP21-

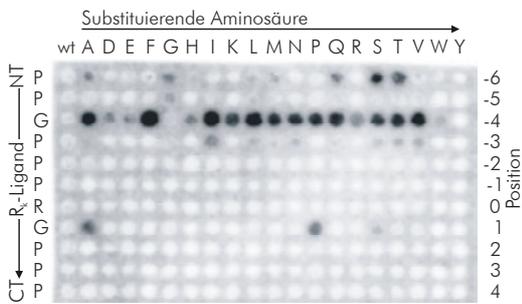


**Abbildung 3-5** Substitutions- und Längenanalyse des R<sub>k</sub>-Liganden für yYFB0-WW

Für die yYFB0-WW Domäne wurde das Erkennungsmotiv **(P/R) P (r/f) (P/G/F) PPP** identifiziert.

(A) Substitutionsanalyse des R<sub>k</sub>-Liganden **PPGPPPRGPPP** (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

(B) Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.



**Abbildung 3-6 Substitutionsanalyse des  $R_k$ -Liganden für hHYP109-WW2**

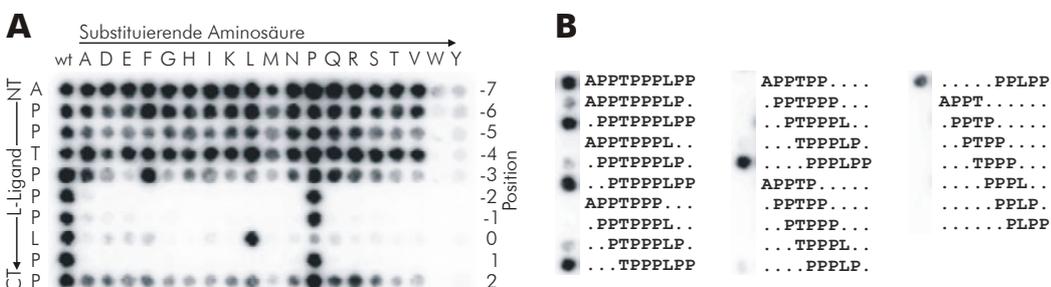
Die hHYP109-WW2 Domäne zeigte nur mit einigen wenigen Substitutionsanaloga, jedoch nicht mit dem  $R_k$ -Liganden Wechselwirkungen. Substitutionsanalyse des  $R_k$ -Liganden **PPGPPPRGPP** (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

WW2), aufgeteilt werden. Der  $R_b$ -Spezifitätsgruppe sehr nahe steht zusätzlich  $\gamma$ YFB0-WW. Dagegen kann hHYP109-WW2 auf Grund der nur schwachen Wechselwirkung mit dem  $R_2$ -Liganden und des unterschiedlichen Erkennungsmotivs keiner dieser beiden Gruppen zugeordnet werden.

Diese Einteilung wird auch durch die Ergebnisse des WW-Screenings unterstützt: die WW-Domänen der  $R_\alpha$ -Spezifitätsgruppe interagieren sowohl mit dem R-Liganden als auch mit dem L/P-Liganden. Dagegen zeigen die WW-Domänen der  $R_b$ -Spezifitätsgruppe sowie  $\gamma$ YFB0-WW nur sehr schwache Wechselwirkungen mit dem L/P-Liganden. Dabei können die WW-Domänen der  $R_\alpha$ -Spezifitätsgruppe sowohl an das Erkennungsmotiv der poly-P-Spezifitätsgruppe (3.2.1.4) im L/P-Liganden binden, als auch an Arginin-Varianten dieses Motivs, welche erwartungsgemäß stärker gebunden werden (siehe Abbildung 3-9, Seite 60).

### 3.2.1.3 L-Spezifitätsgruppe: leuzinspezifische WW-Domänen

Für beide WW-Domänen der L-Spezifitätsgruppe, hFBP11-WW1 und hFBP11-WW2, ergaben die Substitutions- und Längenanalysen des L-Liganden das Erkennungsmotiv **PPLPp** (Abbildung 3-7).

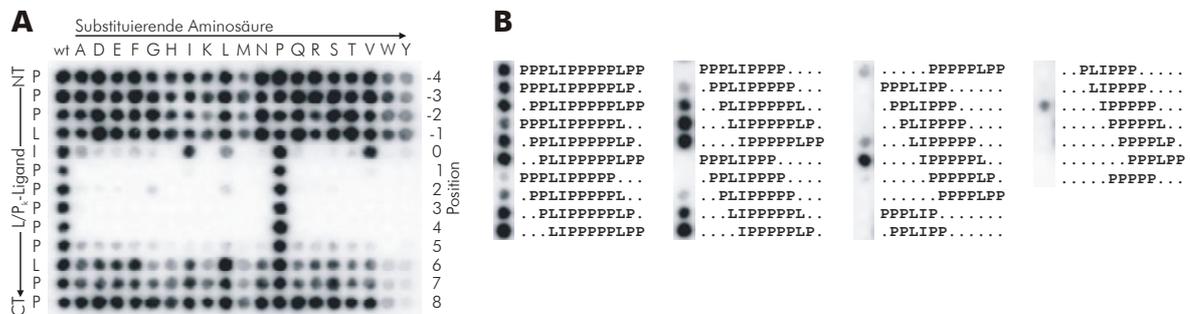


**Abbildung 3-7 Substitutions- und Längenanalyse des L-Liganden für hFBP11-WW1**

Für WW-Domänen der L-Spezifitätsgruppe wurde das Erkennungsmotiv **PPLPp** identifiziert. (A) Substitutionsanalyse des L-Liganden **APPTPPPLPP** (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003. (B) Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

### 3.2.1.4 poly-P-Spezifitätsgruppe: polyprolinspezifische WW-Domänen

Die WW-Domänen der poly-P-Spezifitätsgruppe, *yPRP40-WW2* und *hGAS7-WW*, demonstrierten ebenfalls ein sehr einheitliches Erkennungsmotiv in den Substitutionsanalysen des L/P<sub>k</sub>-Liganden (**PPPLI PPPPLPP**, verkürzte Variante des L/P-Liganden). Beide erkannten spezifisch das Motiv **(p/Φ) P P P P P** (Abbildung 3-8).

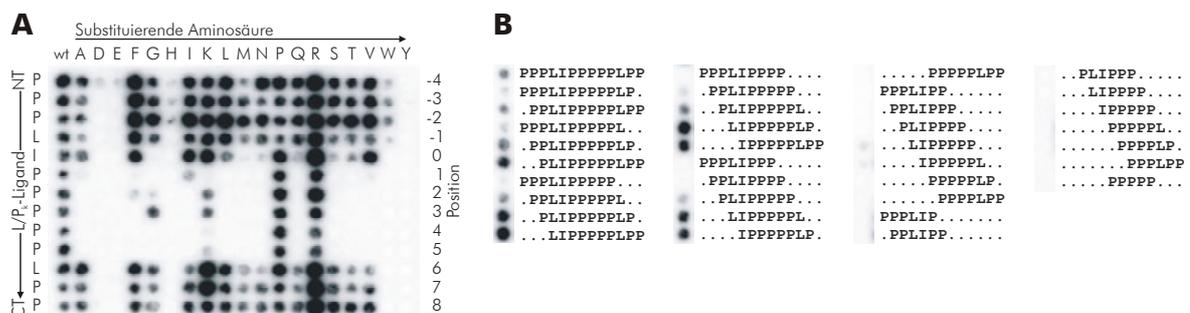


**Abbildung 3-8** Substitutions- und Längenanalyse des L/P<sub>k</sub>-Liganden für *yPRP40-WW2*

Für WW-Domänen der poly-P-Spezifitätsgruppe wurde das Erkennungsmotiv **(p/Φ) P P P P P** identifiziert. **(A)** Substitutionsanalyse des L/P<sub>k</sub>-Liganden **PPPLI PPPPLPP** (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003.

**(B)** Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

Auch die WW-Domänen der R<sub>α</sub>-Spezifitätsgruppe erkannten dieses Erkennungsmotiv im L/P<sub>k</sub>-Liganden (Abbildung 3-9). Allerdings bevorzugen sie Arginin-Varianten des L/P<sub>k</sub>-Liganden und präferieren zusätzliche C-terminale Reste.



**Abbildung 3-9** Substitutions- und Längenanalyse des L/P<sub>k</sub>-Liganden für *hFE65-WW*

Die *hFE65-WW* Domänen zeigte neben dem für die poly-P-Spezifitätsgruppe charakteristischen Erkennungsmotiv vor allem eine Präferenz für verschiedene Arginin-Substitutionen.

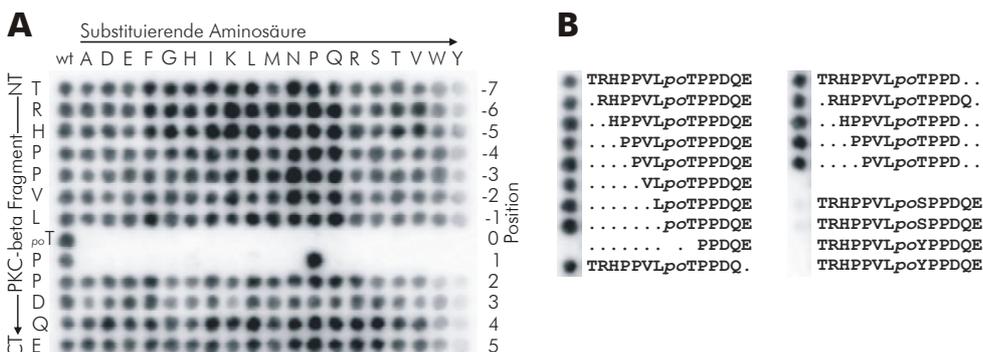
**(A)** Substitutionsanalyse des L/P<sub>k</sub>-Liganden **PPPLI PPPPLPP** (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

**(B)** Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

### 3.2.1.5 poS/poT-Spezifitätsgruppe: phosphoserin-/phosphothreonin-spezifische WW-Domänen

Beide WW-Domänen der *poS/poT*-Spezifitätsgruppe (*hPIN1-WW* und *yESS1-WW*) zeigten in den Substitutionsanalysen verschiedener phosphoserin(*poS*)/phosphothreonin(*poT*)-halti-

ger Peptide beinahe ausschließlich eine Spezifität für das Motiv **(poT/poS) P**. Die Längenanalysen demonstrierte aber meist die Notwendigkeit für einige zusätzliche N- und C-terminale Aminosäurereste, wenn auch unabhängig vom Aminosäuretyp. Daneben ergaben sowohl das Screening der Phosphopeptid-Bibliotheken als auch die Substitutions- und Längenanalysen eine klare Präferenz für Phosphothreonin gegenüber Phosphoserin. Ein Austausch durch Phosphotyrosin war nie möglich.



**Abbildung 3-10** Substitutions- und Längenanalyse eines Fragments aus PKC $\beta$  für hPIN1-WW

Für WW-Domänen der poS/poT-Spezifitätsgruppe wurde das Erkennungsmotiv **(poT/poS) P** identifiziert. **(A)** Substitutionsanalyse des Fragments **TRHPPVLpoTTPPDQE** aus PKC $\beta$  (siehe auch Legende zu Abbildung 3-2). Das Fragment aus der Protein-Kinase-C $\beta$  (PKC $\beta$ , [UP: P05127]) mit einem poTP-Motiv (Position 641) wurde durch das Screening einer Phosphopeptid-Bibliothek als potentieller Wechselwirkungspartner für hPIN1-WW und  $\gamma$ ESS1-WW identifiziert (Daten nicht gezeigt, siehe Otte und Wiedemann *et al.*, 2003). Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003. **(B)** Längenanalyse (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

### 3.2.2 WW-Domänen-Spezifitätsgruppen auf der Basis experimentell bestimmter Ligandenpräferenzen

Die Ergebnisse des WW-Screenings sowie der Substitutions- und Längenanalysen ergaben ein in sich konsistentes Klassifizierungssystem für WW-Domänen in Form von Spezifitätsgruppen (Tabelle 3-5). Jede Spezifitätsgruppe ist jeweils durch ein gruppenspezifisches Ligandenspektrum sowie ein gruppenspezifisches Erkennungsmotiv in den Liganden charakterisiert. Dies weist auf innerhalb der Spezifitätsgruppen konservierte Spezifitätsmechanismen hin, welche im folgenden Abschnitt durch die Untersuchung der spezifitätsbestimmenden Reste der Domänen (Domänen-Epitope) beleuchtet werden.

Zusätzlich wurde die hDYSTRO-WW Domäne aus Dystrophin in die Y-Spezifitätsgruppe aufgenommen, da aus der Literatur bekannt war, dass diese Domäne tyrosinhaltige Liganden bindet, aber ohne benachbarte Domänen nicht korrekt faltet,<sup>76</sup> was die fehlenden Wechselwirkungen im WW-Screening erklärt (Tabelle 3-3).

**Tabelle 3-5** Zusammenfassung der experimentell identifizierten WW-Domänen-Spezifitätsgruppen und der zugehörigen Liganden-Erkennungsmotive

Spezifitätsgruppe	Repräsentative Domänen	Liganden-Erkennungsmotiv <sup>a)</sup>
Y-Spezifitätsgruppe	<i>hYAP65-WW, hNEDD4-WW1-4, hDYSTRO-WW, ...</i>	<b>xPPx (Y/poY)</b>
R <sub>a</sub> -Spezifitätsgruppe	<i>hHYP109-WW2, hFE65-WW, hPQBP1-WW</i>	<b>(p/Φ) P (p/g) PpPR</b>
R <sub>b</sub> -Spezifitätsgruppe	<i>hFBP21-WW1, hFBP21-WW2</i>	<b>PPPRgpPp</b>
L-Spezifitätsgruppe	<i>hFBP11-WW1, hFBP11-WW2</i>	<b>PPLPp</b>
poly-P-Spezifitätsgruppe	<i>γPRP40-WW2, hGAS7-WW</i>	<b>(p/Φ) PFFFF</b>
poS/poT-Spezifitätsgruppe	<i>hPIN1-WW, γESS1-WW</i>	<b>(poT/poS) P</b>

<sup>a)</sup> Für Nomenklatur der Sequenzmotive siehe 2.1.3.

### 3.2.3 Theoretische Bestimmung der Domänen-Epitope

Die Schlüsselreste der Domänen, die für die Erkennung eines bestimmten Ligandentyps verantwortlich sind (Domänen-Epitope), wurden auf Basis der experimentell bestimmten Spezifitätsgruppen (siehe 3.1 und 3.2.1) identifiziert. Dazu wurden die Sequenzen der WW-Domänen an Hand der Spezifitäten gruppiert und gruppenspezifische Konsensussequenzen, Sequenzprofile (siehe 3.2.3.1) und *Profile-Hidden-Markov-Modelle* (pHMM – siehe 3.2.3.3) bestimmt. Als potentiell spezifitätsbestimmende Reste wurden anschließend diejenigen Sequenzmerkmale gedeutet, die innerhalb einer Spezifitätsgruppe konserviert sind, aber durch die sie sich von den anderen unterscheiden. Mittels dieser qualitativen Sequenz-Aktivitäts-Beziehungen wurde schließlich die Zugehörigkeit zu einer dieser Spezifitätsgruppen für experimentell nicht untersuchte WW-Domänen vorhergesagt und experimentell verifiziert (siehe 3.2.3.3). Zusätzlich wurde die funktionale Ähnlichkeit der WW-Domänen in Form eines Sequenz-Verwandtschaftsbaums visualisiert (siehe 3.2.3.2). Darüber hinaus ermöglichte erst die Kenntnis der spezifitätsbestimmenden Reste die Modellierung der WW-Domänen/Liganden-Komplexstrukturen (siehe 3.3.1) für die Analyse von Struktur-Aktivitäts-Beziehungen auf der Ebene der Komplexstrukturen (siehe 3.3.3).

#### 3.2.3.1 Konsensussequenzen der Spezifitätsgruppen

In Tabelle 3-6 sind die Sequenzen der 42 untersuchten WW-Domänen mit ausgewählten Ergebnissen des WW-Screenings zusammengefasst und entsprechend der Spezifitäten gruppiert (siehe 3.2.2). Für alle Gruppen wurden zur Veranschaulichung der konservierten Sequenzelemente Sequenzlogos<sup>43,158</sup> erstellt (Abbildung 3-11) und die Konsensussequenzen auf schematische WW-Domänen projiziert (Abbildung 3-12).

**Tabelle 3-6 WW-Domänen Spezifitätsgruppen: Sequenzalignment, ausgewählte experimentelle Ergebnisse und charakteristische Konsensussequenzen der Spezifitätsgruppen**

Die WW-Domänen sind in 6 Spezifitätsgruppen auf der Basis des Ligandenspektrums und der konservierten Liganden-Erkennungsmotive eingeteilt. Domänen, die einzelnen Gruppen sehr nahe stehen, sich aber dennoch nicht eindeutig zuordnen lassen, sind unterhalb der Gruppe aufgeführt. Domänen, die sich keiner der 6 Gruppen zuordnen lassen, sind als nicht klassifiziert zusammengefasst.

Domäne <sup>a)</sup>	UP-ID <sup>b)</sup>	Aminosäure-Sequenz <sup>d)</sup>										3D-QSAR <sup>f)</sup>					
												COMBINE					
											WW-Screening <sup>e)</sup>			CoMFA	PCA	PLS	
Domänenposition <sup>g)</sup> ω											Y	R	L	L/P	ohne	R-Lig.	Y-Lig.
h WWP3-WW1	O00309	PLPENWEMAYTEN-GEVYFIDHNTKTTTSWLDPRS	0,27	-	-	-	■	■	■	■	■						
y RSP5-WW3	P39940	PLPSGWEMRLTNT-ARVYFVDHNTKTTTWDDPRL	0,26	-	-	-	■	■	■	■							
h NEDD4-WW2	P46934	GLPPGWEEKQDER-GRSYYVDHNSRTTITWTKPTV	0,24	-	-	-	■	■	■	■							
y RSP5-WW2	P39940	ELPSGWEQRFTPE-GRAYFVDHNTKTTTWVDPRL	0,24	-	-	-	■	■	■	■							
h NEDD4-WW3	P46934	FLPKGWEVRHAPN-GRPFYFIDHNTKTTTWEDPRL	0,23	-	-	-	■	■	■	■							
h YAP65-WW	P46937	PLPAGWEMAKTSS-GORYFLNHIDQTTTWQDPRK	0,21	-	-	-	■	■	■	■							
y RSP5-WW1	P39940	RLPPGWERRTDNF-GRTYFVDHNTKTTTWKRPTL	0,20	-	-	-	■	■	■	■							
m ITCHY-WW1	O54971	PLPPGWEQRVDQH-GRVYFVDHVEKRTTWDRPEP	0,18	-	-	-	■	■	■	■							
m ITCHY-WW3	O54971	PLPPGWEKRTDSN-GRVYFVNHNTTITQWEDPRS	0,10	-	-	+/-	■	■	■	■							
h NEDD4-WW4	P46934	PLPPGWEERTHTD-GRIFYINHNKRTQWEDPRL	0,03	-	-	-	■	■	■	■							
h NEDD4-WW1	P46934	PLPPGWEERQDIL-GRTYFVNHESRTQWKRPTP	shift	-	-	-	■	■	■	■							
h DYSTRO-WW	P11532	SVQGPWERAISPN-KVPPYYINHETQTTTCWDHPKM	0,00	-	-	-	■	■	■	■							
<b>Y-Spezifitätsgruppe</b>																	
h WWP3-WW2	O00309	ELPAGWEKIEDPVYG-IYYVDHINKTQYENPVL	shift	-	-	+/-	■	■	■	■							
h WWP3-WW2	Q9UJT8	PLPPGWEVRSTVS-GRIFYFVDHNRITQFTDPRL	shift	-	-	-	■	■	■	■							
h WWOX-WW2	Q9NZC7	DLPGWEQETDEN-GQVFFVDHINKRTTYLDPRL	keine	-	-	-	■	■	■	■							
h HYP109-WW1	Q9NT81	IEMGDWQEVVDENTGSYYYWNQTQNEVTWELPQY	shift	++	-	++	■	■	■	■							
h FE65-WW	O00213	DLPAGWMRVQDTS-G-TYYWHIPTIGTTQWEPGR	0,02	++	-	++	■	■	■	■							
h PQBP1-WW	O60828	GLPPSWYKVFDPSSGLPYYNADTLVSWLSPHD	0,00	++	-	++	■	■	■	■							
<b>R<sub>o</sub>-Spezifitätsgruppe</b>																	
y YFB0-WW	P43582	QVPSGWKAVFDDEYQTWYYVDLSTNSSQWEPGR	0,03	+	-	+	■	■	■	■							
h FBP21-WW1	O75554	PSKGRWVEGITSE-GYHYYYDLISGASQWEKPEG	0,00	++	-	+/-	■	■	■	■							
h FBP21-WW2	O75554	AVKTIVWVEGLSED-GFTYYNTTETGESRWEKPD	0,01	++	-	+	■	■	■	■							
<b>R<sub>b</sub>-Spezifitätsgruppe</b>																	
h HYP109-WW2	Q9NT81	ATPKGWSHWRDRHRRYFYVNEQSGESQWEPD	0,02	-	-	-	■	■	■	■							
h FBP11-WW1	O75400	GAKSMWTEHKSPD-GRTYYYNTETKQSTWEPD	0,02	-	++	++	■	■	■	■							
h FBP11-WW2	O75400	LSKSPWKEYKSDS-GKPYYYNSQTKESRWAKPKE	0,05	+/-	++	++	■	■	■	■							
<b>L-Spezifitätsgruppe</b>																	
y PRP40-WW2	P33203	LRENGWKAAKTAD-GKVYYYNPFTRETTSWTIPAF	0,03	-	-	++	■	■	■	■							
h GAS7-WW	O60861	ILPPGWQSYLSPQ-GRYYVNTTNETTWERPSS	0,02	-	-	++	■	■	■	■							
<b>poly-P-Spezifitätsgruppe</b>																	
h PIN1-WW	Q13526	KLPPGWEKMRSSGRVYYFNEITNASQWERPSG	0,01	-	-	-	■	■	■	■							
y ESS1-WW	P22696	GLPTPWIVRYSKSKKREYFFNPETKHSQWEEPEG	0,00	-	-	-	■	■	■	■							
<b>po S/po T-Spezifitätsgruppe</b>																	
y PRP40-WW1	P33203	--MSIWKEAKDAS-GRIYYNTLTKKSTWEPKPE	0,00	-	-	-	■	■	■	■							
y YPR152C-WW	Q06525	-MRGEWQEFKTPA-GKKYYYNKTKQSRWEKPNL	0,08	-	-	-	■	■	■	■							
h CA150-WW1	O14776	PTEIIVENKTPD-GKVYYYNARTRESAWTKPDG	0,01	-	-	-	■	■	■	■							
h SMURF2-WW1	Q9H260	DLDPGWEERTAS-GRIQYLNHITRTTOWERPTR	0,00	-	-	-	■	■	■	■							
y YJQ8-WW	P46995	RLPPGWEIITHEN--GRPLYNAEQTKLHYPPSG	keine	-	-	-	■	■	■	■							
h SMURF1-WW1	Q9UJT8	ELPEGYEQRTTQV-GQVYFLHTQGTGVSTWHPRI	0,00	-	-	-	■	■	■	■							
h PEPP2-WW1	Q9HAU0	SLPRSWTYGITRG-GRVFFINEEAKSTTWLHPVT	0,00	-	-	-	■	■	■	■							
h PEPP2-WW2	Q9HAU0	DLPTGWEEAYTFE-GARYYINHNKRVTSKHPVT	keine	-	-	-	■	■	■	■							
h WWOX-WW1	Q9NZC7	ELPPGWEERTTKD-GWVYYANHTEEKTQWEHPKT	0,00	-	-	-	■	■	■	■							
h IQGAP1-WW	P46940	DNSKWKVHKHWVKG-GYVYYHNLETQEGGWDEPPN	0,00	-	-	-	■	■	■	■							
h CA150-WW3	O14776	IPGTPWSVVWTGD-ERVFFYNPTRLSMWDRPDD	0,00	-	-	-	■	■	■	■							
h KIAA1052-WW	Q9UPV0	PLPGWPKSQDIT-GDIYFNFANGQSMWDHPSD	0,00	-	-	-	■	■	■	■							
y TIN1-WW	P53076	ESRNLWDTSDKNS-GLOYYPDQSPSSSFSSPRV	keine	-	-	-	■	■	■	■							
y SSM4-WW	P40318	TSIISWKFNITILL--TLYETKRILLESSYVKPLL	keine	-	-	-	■	■	■	■							
<b>nicht klassifiziert</b>																	

<sup>a)</sup> Domänennamen gebildet aus Organismus-Präfix, Proteinnamen und Domänen-Suffix (siehe 2.1.1). Domänen mit grau hinterlegtem Domänennamen sind ohne Ligand ungefaltet.

<sup>b)</sup> Primärer UniProt<sup>10</sup> Zugriffscode des die WW-Domäne enthaltenden Proteins.

<sup>c)</sup> Die konservierten strukturellen Positionen ω der Aminosäurereste der WW-Domänen sind analog zu hYAP65-WW durchnummeriert (siehe 2.1.2.1).

<sup>d)</sup> Sequenzalignment der Sequenzen der WW-Domänen. Die Einfärbung konservierter Aminosäure- (fortgesetzt)

(Fortsetzung der Legende zu Tabelle 3-6)

eigenschaften basiert auf der Konsensussequenz-Färbemethode von CLUSTAL-X.<sup>181</sup> Damit wurden die einzelnen Gruppen jeweils getrennt eingefärbt. Für diejenigen Spezifitätsgruppen, für die sich ein plausibler gruppenspezifischer Sequenzkonsensus ergab, ist dieser unterhalb der Gruppe angegeben. In Fällen, in denen nur eine Aminosäureeigenschaft konserviert ist, ist die häufigste Aminosäure angegeben, bei unvollständiger Konservierung als Kleinbuchstabe. Im Kopf des Alignments sind diejenigen Positionen, die auf der zum Liganden exponierten Seite der WW-Domänen liegen, dunkelgrau und die Positionen in den Schleifen hellgrau hinterlegt (siehe 1.5). Oberhalb des Sequenzalignments ist der Sekundärstrukturkonsensus dargestellt.

<sup>e</sup> Ausgewählte Ergebnisse des WW-Screenings (siehe Legende zu Tabelle 3-3).

<sup>f</sup> Diejenigen WW-Domänen, welche für 3D-QSAR-Analysen (quantitative Struktur-Aktivitäts-Beziehungen) ausgewählt wurden (siehe 3.3.2 und 3.3.3), sind durch ■ markiert. Die Färbung der Spalten symbolisiert die experimentell bestimmte Wechselwirkungsaktivität, die für die einzelnen QSAR-Modelle verwendet wurde. Für die COMBINE-Analysen gibt der Spaltentitel den Liganden an, der zur Modellierung der Komplexe verwendet wurde (R-Lig = R-Ligand und Y-Lig. = Y-Ligand, siehe Tabelle 3-2, Seite 52).

Die strukturell konservierten Positionen der Aminosäurereste sind analog zu hYAP65-WW durchnummeriert und durch ein vorangestelltes  $\omega$  gekennzeichnet (Tabelle 3-6, siehe 2.1.2.1).

Die Sequenzlogos veranschaulichen, dass in allen Gruppen stets das charakteristische Sequenzmuster der WW-Domänen konserviert ist: die strukturellen Reste  $W\omega 17$ ,  $Y/F\omega 29$  und  $P\omega 42$  sowie die funktionalen Reste  $Y\omega 28$ ,  $T/S\omega 37$  und  $W\omega 39$ . Darüber hinaus unterscheiden sich die Gruppen in charakteristischen Sequenzmerkmalen, die interessanterweise vorwiegend auf der dem Liganden zugewandten Seite der WW-Domänen liegen (Abbildung 3-12 und Tabelle 3-6). Dies unterstützt die funktionale Interpretation dieser Sequenzmotive.

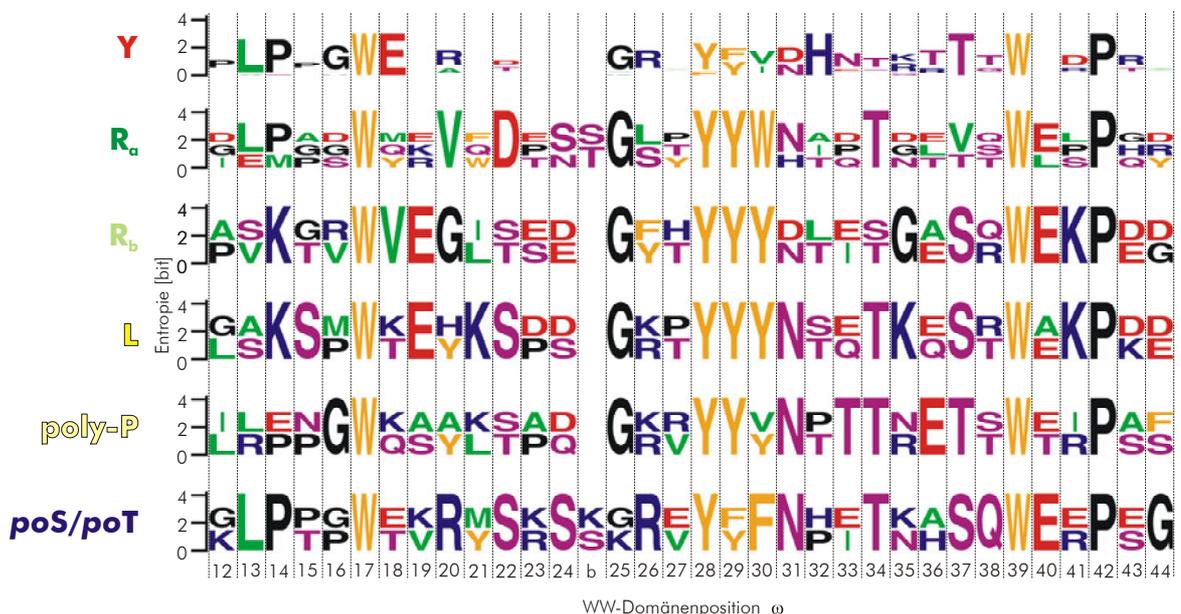
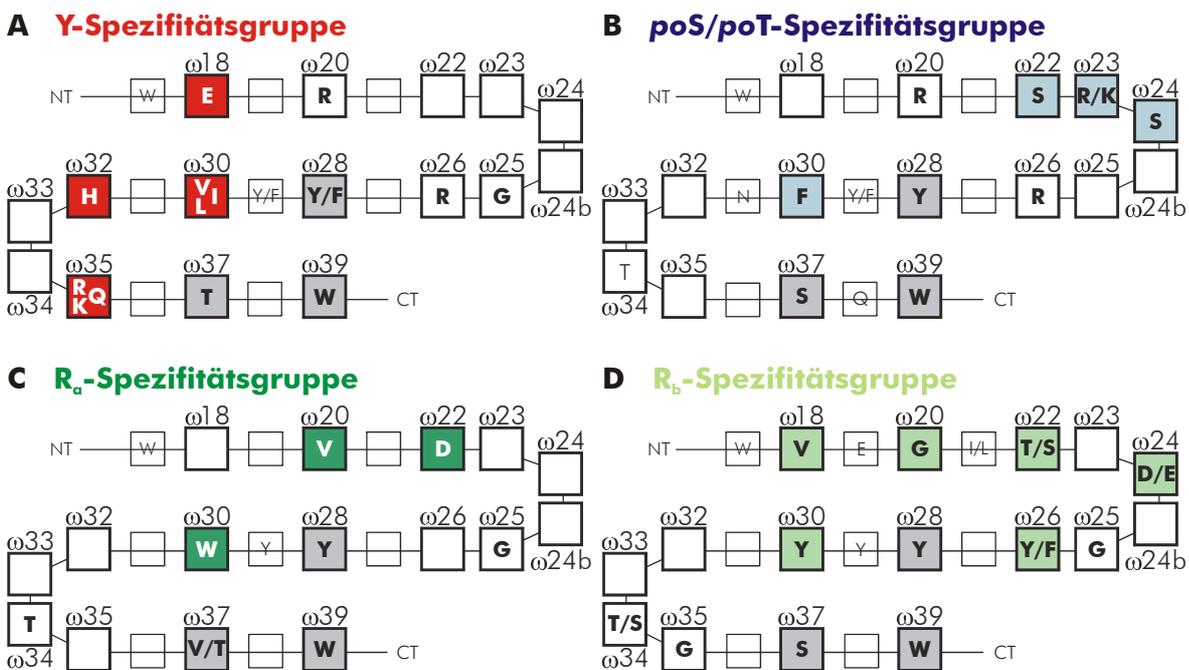


Abbildung 3-11 Sequenzlogos der WW-Domänen-Spezifitätsgruppen

Die Stärke der Sequenzkonservierung und die konservierten Aminosäuren sind für die Y-,  $R_a$ -,  $R_b$ -, L-, poly-P- und poS/poT-Spezifitätsgruppen in Form von Sequenzlogos dargestellt.<sup>43,158</sup> In diesen Sequenzlogos wird jede Position des Sequenzalignments der jeweiligen Spezifitätsgruppe durch einen Stapel derjenigen Aminosäuren repräsentiert, die an der jeweiligen Sequenzposition vorkommen. Die Höhe des Stapels symbolisiert die Konservierung an dieser Position (gemessen als Entropie in bit), während die Höhe der einzelnen Aminosäuren ihre relativen Häufigkeiten versinnbildlicht. Geht die Entropie gegen 0, lässt sich der Stapel nicht mehr auflösen und verschwindet daher an diesen Positionen. Für das Sequenzlogo der Y-Spezifitätsgruppe konnte eine Korrektur für kleine Stichproben vorgenommen werden. Dies war für die anderen Gruppen nicht möglich wegen der zu geringen Stichprobengröße. Die Aminosäuren sind analog zum Sequenzalignment in Tabelle 3-6 eingefärbt.



**Abbildung 3-12** Konsensussequenzen der experimentell bestimmten WW-Domänen-Spezifitätsgruppen

Die Konsensussequenzen der **(A)** Y-, **(B)** poS/poT-, **(C)** R<sub>α</sub>- und **(D)** R<sub>β</sub>-Spezifitätsgruppen sind in Form schematischer WW-Domänen aus der Sicht des Liganden dargestellt. Die funktionalen Reste, die für alle WW-Domänen konserviert sind, sind grau unterlegt. Die Reste, die in jeweils einer Spezifitätsgruppe konserviert sind, sind in der Farbe der jeweiligen Gruppe unterlegt. Diejenigen Aminosäuren, die auf der dem Liganden zugewandten Seite des β-Faltblatts liegen (bzw. in den Schleifen), sind als große Quadrate vor dem schematisierten Peptidrückgrat dargestellt. Diejenigen Aminosäuren, die auf der abgewandten Seite liegen, sind als kleine Quadrate hinter dem Peptidrückgrat dargestellt. Die Domänenpositionen ω sind oberhalb der Quadrate angegeben (siehe 2.1.2.1). Die N- und C-terminalen Enden des β-Faltblatts sind mit NT und CT gekennzeichnet.

Im Falle der **Y-Spezifitätsgruppe** umfasst der gruppenspezifische Sequenzkonsensus die Reste E<sub>ω18</sub>, V/I/L<sub>ω30</sub>, H<sub>ω32</sub> und R/K/Q<sub>ω35</sub> (Tabelle 3-6 und Abbildung 3-12A). Insbesondere das Fehlen einer aromatischen Aminosäure an Position ω30 unterscheidet die WW-Domänen der Y-Spezifitätsgruppe von denen der anderen Gruppen. Außer für E<sub>ω18</sub> wurde für diese Reste, die alle in der ββC-Schleife lokalisiert sind, auch strukturell eine Beteiligung an der Erkennung des Tyrosins im Liganden gezeigt (siehe 1.3.2).<sup>113</sup> Dies unterstützt die in dieser Arbeit gewählte Herangehensweise zur Identifizierung der Schlüsselreste.

Im Gegensatz dazu sind die konservierten Reste der **poS/poT-Spezifitätsgruppe** (S<sub>ω22</sub>, R/K<sub>ω23</sub>, S<sub>ω24</sub> und F<sub>ω30</sub>) vor allem in der βαβB-Schleife lokalisiert (Tabelle 3-6 und Abbildung 3-12B). Auch für diese ist eine Beteiligung an der Erkennung von Phosphoserin bzw. Phosphothreonin strukturell nachgewiesen.<sup>193</sup>

Die Unterschiede der beiden R-Spezifitätsgruppen in ihrem Ligandenspektrum und den Erkennungsmotiven in den Liganden spiegeln sich in den unterschiedlichen spezifitätsbestimmenden Resten in den Domänen wieder. Im Falle der **R<sub>α</sub>-Spezifitätsgruppe** sind die

Reste V $\omega$ 20, D $\omega$ 22 und W $\omega$ 30 und im Falle der **R<sub>b</sub>-Spezifitätsgruppe** die Reste V $\omega$ 18, G $\omega$ 20, T/S $\omega$ 22, D/E $\omega$ 24, Y/F $\omega$ 26 und Y $\omega$ 30 konserviert (Tabelle 3-6 und Abbildung 3-12C-D). Trotz dieser Unterschiede gibt es aber auch Gemeinsamkeiten. Die konservierten Reste beider Gruppen sind in der  $\beta$ A $\beta$ B-Schleife lokalisiert. In beiden Fällen befindet sich dort eine konservierte negativ geladene Aminosäure und beide Gruppen besitzen eine aromatische Aminosäure an Position  $\omega$ 30. Die Sequenz von  $\gamma$ YFB0-WW enthält dabei Merkmale beider Gruppen (D $\omega$ 22 und E $\omega$ 24). Ähnliches trifft auch auf die Sequenz von hHYP109-WW2 zu. In beiden Fällen fehlt allerdings eine aromatische Aminosäure an Position  $\omega$ 30. Für beide Domänen ist aber auch ein alternatives Sequenzalignment denkbar, in dem der zentrale Cluster bestehend aus drei aromatischen Aminosäuren um eine Position in C-terminaler Richtung verschoben ist. Dadurch würden dann auch diese WW-Domänen an Position  $\omega$ 30 eine aromatische Aminosäure besitzen.

Auf Grund der geringen Anzahl an Sequenzen und ihrer hohen Ähnlichkeit konnte keine plausible Konsensussequenz für die L- und poly-P-Spezifitätsgruppe berechnet werden. Dennoch sind für die **L-Spezifitätsgruppe** Y/H $\omega$ 20 und Y $\omega$ 30 vielversprechende Kandidaten für spezifitätsbestimmende Reste.

### 3.2.3.2 Funktionaler Verwandtschaftsbaum der WW-Domänensequenzen

Um die 42 WW-Domänen allein auf der Basis ihrer Sequenzmerkmale zu gruppieren, wurde ein Sequenzverwandtschaftsbaum berechnet. Für die Berechnung des Baums wurden nur diejenigen Aminosäurereste verwendet, die auf der dem Liganden zugewandten Seite der WW-Domänen liegen (hell und dunkelgrau hinterlegte Positionen im Kopf des Sequenzalignments in Tabelle 3-6). Dadurch wurde der Baum darauf ausgerichtet, mehr die funktionale Verwandtschaft als die phylogenetische Verwandtschaft zu repräsentieren (Abbildung 3-13).

Um eine Korrelation der experimentell bestimmten Ligandenpräferenzen mit dieser rein sequenzbasierten Gruppierung der Domänen zu identifizieren, wurden die Ergebnisse des WW-Screenings farbkodiert auf den Baum abgebildet. Dabei werden die WW-Domänen der Y-, R<sub>a</sub>- und R<sub>b</sub>-Spezifitätsgruppen auch im Verwandtschaftsbaum in klar abgegrenzte Äste gruppiert.

Der **Y-Ast** (Abbildung 3-13, rot umrandet), der größte und kompakteste Teil des Baums, enthält alle Domänen der **Y-Spezifitätsgruppe** und 7 nah verwandte Domänen, die meist nur in einer Position vom Sequenzkonsensus der Y-Spezifitätsgruppe abweichen,



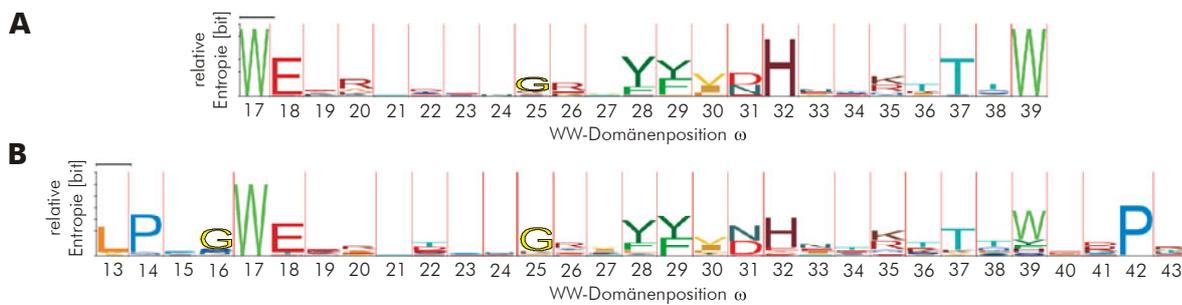
dunkelgrün umrandet) und den **R<sub>b</sub>-Ast** (Abbildung 3-13, hellgrün umrandet). Gleichzeitig ordnet sich yYFB0-WW in der Nähe des R<sub>a</sub>-Asts ein, wohingegen hHYP109-WW2 weit von beiden R-Spezifitätsgruppen entfernt liegt. Dies unterstützt erneut die vorausgegangene Klassifizierung der argininspezifischen WW-Domänen in zwei Subgruppen.

Im Gegensatz dazu sind die WW-Domänen der L-, poly-P- und poS/poT-Spezifitätsgruppen uneinheitlich über den weniger kompakten Teil des Baums verstreut. Damit bestätigt der Sequenzverwandtschaftsbaum zumindest die Definition der Y-Spezifitätsgruppe sowie die Aufteilung der argininspezifischen WW-Domänen auf die R<sub>a</sub>- und R<sub>b</sub>-Spezifitätsgruppen. Gleichzeitig konnte durch den Baum auch gezeigt werden, dass es Sequenzmerkmale gibt, die direkt mit der Aktivität der WW-Domänen korrelieren und somit eine Vorhersage basierend auf diesen Merkmalen möglich sein sollte.

### 3.2.3.3 Sequenzbasierte Vorhersage der Spezifität: Wechselwirkungen der WW-Domänen mit tyrosinhaltigen Liganden

Zur Vorhersage tyrosinspezifischer WW-Domänen wurden *Profile-Hidden-Markov-Modelle* (pHMM) der Sequenzen der Y-Spezifitätsgruppe (Abbildung 3-14A) und der Sequenzen des Y-Asts (Abbildung 3-14B) berechnet. An Hand der Homologiewerte einer gegebenen WW-Domänensequenz gegenüber diesen pHMMs kann die Tyrosin-Spezifität für diese WW-Domäne vorhergesagt werden. Die Vorhersageleistung dieser pHMMs wurde mit der des Kernkonsensusmotivs der Y-Spezifitätsgruppe (**[YF] [YF] [VIL] xHxx [RKQ] x [T] xW**, siehe Abbildung 3-12A) verglichen.

Um die Klassifizierungsleistung der pHMMs zu analysieren und Schwellenwerte für die spätere Vorhersage der Spezifität zu bestimmen, wurde zuerst ein Ranking der 42 experimentell untersuchten WW-Domänen an Hand ihrer Homologie gegenüber den beiden pHMMs durchgeführt (siehe Anhang A, Tabelle A-1). Erwartungsgemäß erhielten mit beiden pHMMs die Mitglieder der Y-Spezifitätsgruppe die höchsten Homologiewerte. Ebenso wurden durch das pHMM des Y-Asts auch die restlichen Mitglieder des Y-Asts einheitlich mit hohen Homologiewerten bewertet. Einzig yYJQ8-WW erhielt schlechtere Homologiewerte als der WW-Domänen Prototyp und hPIN1-WW. Damit wurde gezeigt, dass das pHMM des Y-Asts innerhalb der 42 WW-Domänen erfolgreich zwischen Domänen des Y-Asts und solchen, die ihm nicht angehören, unterscheiden kann. Ebenso kann das pHMM der Y-Spezifitätsgruppe erfolgreich zwischen WW-Domänen der Y-Spezifitätsgruppe und Domänen, die nicht zu dieser Gruppe gehören, differenzieren.



**Abbildung 3-14** Visualisierung der Profile-Hidden-Markov-Modelle der Y-Spezifitätsgruppe und des Y-Asts als HMMLogos

Das HMMLogo<sup>162</sup> stellt die wichtigsten Aspekte des **(A)** Profile-Hidden-Markov-Modells (pHMMs) der Y-Spezifitätsgruppe und des **(B)** pHMMs des Y-Asts in einer zu den Sequenzlogos analogen Form dar. Jeder emittierende Zustand eines pHMMs wird durch eine Spalte mit einem Stapel von Aminosäuren repräsentiert. Die Breite der Spalten repräsentiert die Wahrscheinlichkeit den jeweiligen Zustand zu erreichen. In Analogie zu Sequenzlogos stellt die Höhe des Stapels die Abweichung der Aminosäureverteilung von der Hintergrundverteilung dar (relative Entropie) und die Höhe der einzelnen Aminosäuren ihre relativen Emissionswahrscheinlichkeiten. Die emittierenden *Match*-Zustände – sie entsprechen den Domänenpositionen  $\omega$  im WW-Domänen Sequenzalignment (Tabelle 3-6) – sind durch einen weißen Hintergrund gekennzeichnet. Zwischen diesen liegen die emittierenden *Insert*-Zustände (rote Balken), die auf Grund der hier äußert geringen Wahrscheinlichkeit von Insertionen sehr schmal ausfallen, bis auf die Insertion in Position  $\omega$ 24b (siehe **B**).

Auch die Schwellenwerte der Homologiewerte, ab denen objektiv Tyrosin-Spezifität vorhergesagt wird, wurden an Hand des Rankings der 42 WW-Domänen bestimmt. Der obere Schwellenwert (harte Klassifizierung) wurde hierfür als der größte Homologiewert einer WW-Domäne definiert, welche nicht zur Y-Spezifitätsgruppe ( $>39,7$  für *hSMURF1*-WW2) bzw. zum Y-Ast ( $>52,2$  für WW-Prototyp) gehört. Der untere Schwellenwert (weiche Klassifizierung) wurde dagegen als der kleinste Homologiewert einer der Y-Spezifitätsgruppe nahestehenden WW-Domäne ( $\geq 37,7$  für *hWWOX*-WW2, siehe auch Tabelle 3-6) bzw. des Y-Asts ( $\geq 44,2$  für *yYJQ8*-WW) definiert (siehe Anhang A, Tabelle A-1). Der Bereich zwischen dem unteren und dem oberen Schwellenwert wurde als Übergangsbereich zwischen vorhergesagter Tyrosin-Spezifität bzw. Nicht-Tyrosin-Spezifität definiert.

Mit Hilfe der beiden pHMMs und dem Kernkonsensusmotiv der Y-Spezifitätsgruppe wurden schließlich die 482 WW-Domänen der SMART-Datenbank<sup>161</sup> (Stand: 15.11.2001) in tyrosinspezifische und nicht tyrosinspezifische WW-Domänen eingeteilt (Tabelle 3-7). Um eine potentielle Korrelation zwischen der Funktion eines Proteins und der Spezifität der erhaltenen WW-Domänen zu identifizieren, wurden die Proteine der 482 WW-Domänen manuell in funktionale Proteinfamilien eingeteilt. Als Kriterien für diese Einteilung wurden Datenbank-Annotationen und die innerhalb einer Proteinfamilie konservierte Domänenarchitektur verwendet (Tabelle 3-7, Spalte Proteinfamilie).

Tabelle 3-7 Klassifizierung der Tyrosin-Spezifität der WW-Domänen verschiedener Proteinfamilien

**A** pHMM der Y-Spezifitätsgruppe

Proteinfamilie	Y	Y	ØY	Σ
	> 39,7	≥ 37,7	< 37,7	
NEDD4-Ubiquitin-Ligasen	120	24	11	155
MAGI und homologe	17	2	17	36
Dystrophin und homologe	7	12	2	21
YAP und homologe	6	3		9
Synip	1			1
BAG und homologe		5		5
tbZFP2		1		1
WWOX und homologe		24		24
unklassifiziert	6	11		17
PEPP2			2	2
PIN1 und homologe			10	10
YJQ8			1	1
FBP11 und homologe			27	27
FE65 und homologe			11	11
WW45 und homologe			10	10
PQBP1 und homologe			9	9
HYP109 und homologe			8	8
CA150 und homologe			8	8
RNA Helikase und homol.			8	8
Small-GTPase-activating			7	7
FBP21 und homologe			6	6
GAS7 und homologe			6	6
YFB0			2	2
YPR152C			1	1
unklassifiziert			97	97
Σ	157	82	243	482

**B** pHMM des Y-Asts

Proteinfamilie	Y	Y	ØY	Σ
	> 52,2	≥ 44,2	< 44,2	
NEDD4-Ubiquitin-Ligasen	138	6	11	155
MAGI und homologe	24	11	1	36
Dystrophin und homologe	5	2	14	21
YAP und homologe	9			9
Synip	1			1
BAG und homologe	5			5
tbZFP2	1			1
WWOX und homologe	24			24
unklassifiziert	14	4		18
PEPP2	2			2
PIN1 und homologe		3	7	10
YJQ8		1		1
FBP11 und homologe			27	27
FE65 und homologe			11	11
WW45 und homologe			10	10
PQBP1 und homologe			9	9
HYP109 und homologe			8	8
CA150 und homologe			8	8
RNA Helikase und homol.			8	8
Small-GTPase-activating			7	7
FBP21 und homologe			6	6
GAS7 und homologe			6	6
YFB0			2	2
YPR152C			1	1
unklassifiziert			96	96
Σ	223	27	232	482

**C** Kernkonsensusmotiv der Y-Spezifitätsgruppe

Proteinfamilie	Y	ØY	Σ
NEDD4-Ubiquitin-Ligasen	112	43	155
MAGI und homologe	17	19	36
Dystrophin und homologe	15	6	21
YAP und homologe	9		9
Synip	1		1
BAG und homologe	5		5
tbZFP2	1		1
WWOX und homologe		24	24
unklassifiziert	10		10
PEPP2		2	2
PIN1 und homologe		10	10
YJQ8		1	1
FBP11 und homologe		27	27
FE65 und homologe		11	11
WW45 und homologe		10	10
PQBP1 und homologe		9	9
HYP109 und homologe		8	8
CA150 und homologe		8	8
RNA Helikase und homol.		8	8
Small-GTPase-activating		7	7
FBP21 und homologe		6	6
GAS7 und homologe		6	6
YFB0		2	2
YPR152C		1	1
unklassifiziert		104	104
Σ	170	312	482

482 WW-Domänen der SMART-Datenbank<sup>160,161</sup> (Stand: 15.11.2001) wurden in tyrosinspezifisch (Spalte Y, dunkelrot hinterlegt, oberer Schwellenwert), potentiell tyrosinspezifisch (Spalte Y, hellrot hinterlegt, unterer Schwellenwert) und nicht tyrosinspezifisch (Spalte ØY) klassifiziert. Die Anzahl der WW-Domänen ist für die einzelnen Proteinfamilien, aus denen die WW-Domänen stammen, aufgegliedert. (A) zeigt die Klassifizierung durch das pHMM der Y-Spezifitätsgruppe, (B) die Klassifizierung durch das pHMM des Y-Asts und (C) die Klassifizierung durch das Kernkonsensusmotiv der Y-Spezifitätsgruppe. Die angegebenen Proteinbezeichnungen sind der UniProt-Datenbank<sup>10</sup> entnommen. Die wichtigsten Proteinfamilien sind im Text und der Einleitung besprochen (siehe 1.3.3).

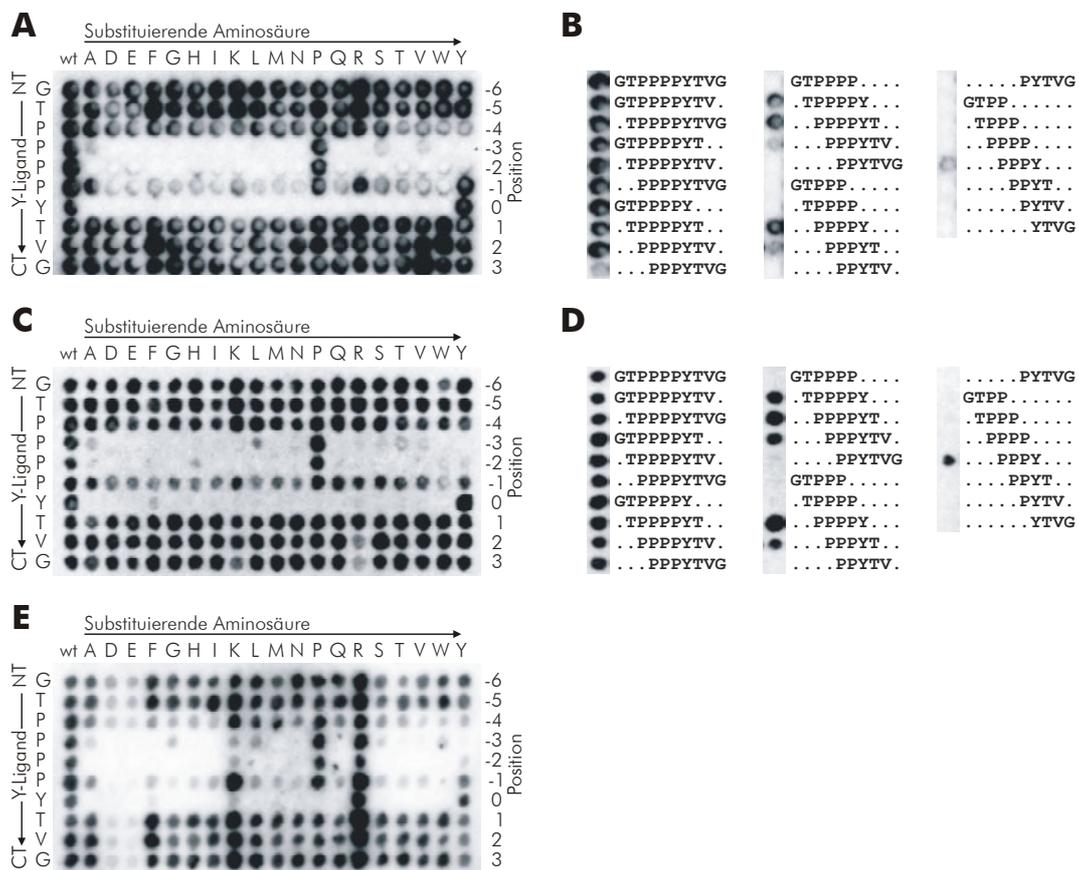
Für alle drei Klassifizierungsmethoden korreliert die vorhergesagte Tyrosin-Spezifität der WW-Domänen mit der Funktion bzw. Familie der sie enthaltenden Proteine. So werden die WW-Domänen der großen Familien der NEDD4-ähnlichen Ubiquitin-Ligasen,<sup>204</sup> der Membran-Assoziierten-Guanylat-Kinasen (MAGI),<sup>133</sup> der Dystrophin-ähnlichen<sup>76</sup> und der YAP-ähnlichen Proteine<sup>138</sup> fast ausschließlich als tyrosinspezifisch klassifiziert. Dies steht im Einklang mit in der Literatur beschriebenen Vertretern dieser Familien, für welche die Tyrosin-Spezifität der WW-Domänen nachgewiesen wurde. Umgekehrt gilt unter anderem für die Familien der PIN1-, FBP11-, FE65-, WW45- und der PQBP1-ähnlichen Proteine, dass ihre WW-Domänen generell als nicht tyrosinspezifisch klassifiziert werden.

WW-Domänen aus generell tyrosinspezifisch klassifizierten Proteinfamilien, die jedoch als nicht tyrosinspezifisch klassifiziert wurden, weichen meist nur in einer Position vom Kernkonsensusmotiv der Y-Spezifitätsgruppe ab. Dies gilt grundsätzlich auch für die Familie der WW-Domänen enthaltenden Oxidoreduktasen (WWOX).<sup>111</sup> Daneben zeigte sich, dass nur 11 WW-Domänen, die das Kernkonsensusmotiv der Y-Spezifitätsgruppe enthalten, ein Leuzin an  $\omega 30$  besitzen. Davon gehören allein 7 zur Familie der YAP-ähnlichen Proteine.

Neben den Proteinfamilien, für welche schon einzelne Vertreter mit tyrosinspezifischen WW-Domänen bekannt waren, wurden auch einige neue tyrosinspezifische Kandidaten vorhergesagt (Tabelle 3-7). Zur experimentellen Überprüfung der vorhergesagten Tyrosin-Spezifität wurden *hBAG3*-WW (BCL2-associated athanogene 3 [UP: Q95817]), *mSYNIP*-WW (Syntaxin4-interacting protein SYNIP [UP: Q9WW89]) und *tbZFP2*-WW (*Trypanosoma brucei* Zinc-finger protein 2 [UP: Q967Y4]) ausgewählt und synthetisiert. Anschließend wurden die WW-Domänen dieser drei Proteine durch Substitutions- und Längenanalysen des Y-Liganden auf Tyrosin-Spezifität getestet (Abbildung 3-15).

Die Bindungsstudien bestätigten für alle drei WW-Domänen die vorhergesagte Tyrosin-Spezifität. Insbesondere zeigen sie das für die Y-Spezifitätsgruppe charakteristische Erkennungsmotiv (**xPPxY**, siehe Abbildung 3-2). Einzig *tbZFP2*-WW zeigte die Tendenz, zusätzlich auch argininhaltige Peptide zu binden. Diese Ergebnisse demonstrieren damit, dass auch die WW-Domänen im Übergangsbereich des pHMMs der Y-Spezifitätsgruppe noch tyrosinspezifisch sind.

Der Vergleich der drei Klassifizierungsmethoden zeigt (Tabelle 3-8, Seite 73; siehe auch Tabelle 3-7), dass insgesamt 141 WW-Domänen durch alle drei Methoden konsistent als tyrosinspezifisch (Tabelle 3-8, Seite 73, oberste Zeile) und 216 WW-Domänen als nicht



**Abbildung 3-15** Substitutions- und Längenanalysen des Y-Liganden zur Verifizierung der vorhergesagten Tyrosin-Spezifität

Für **(A)** *hBAG3*-WW, **(C)** *hSYNIP*-WW und **(E)** *tbZFP2*-WW bestätigen die Substitutionsanalysen des Y-Liganden die vorhergesagte Tyrosin-Spezifität. Alle drei WW-Domänen zeigen das für die Y-Spezifitätsgruppe charakteristische Erkennungsmotiv **xPPxY**. *tbZFP2* zeigt zusätzlich eine Präferenz für Arginin-Substitutionen (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte und Wiedemann *et al.*, 2003, sowie Otte, 2004.

Längenanalysen des Y-Liganden mit **(B)** der *hBAG3*-WW und **(D)** der *hSYNIP*-WW Domäne (siehe auch Legende zu Abbildung 3-2). Abbildung modifiziert nach Otte, 2004.

tyrosinspezifisch klassifiziert werden (Tabelle 3-8, unterste Zeile). Die Unterschiede in den Klassifizierungen liegen vor allem im Übergangsbereich zwischen Tyrosin-Spezifität und Nicht-Tyrosin-Spezifität.

Damit konnte erfolgreich die Spezifitätsgruppe für experimentell nicht-untersuchte WW-Domänen ausgehend von der Domänensequenz vorhergesagt werden. Die Einteilung in Spezifitätsgruppen ist jedoch grob und teilweise subjektiv. Auch die Einteilung in „tyrosinspezifisch“ bzw. „nicht tyrosinspezifisch“ ist auf Grund der Wahl eines Schwellenwertes teilweise willkürlich, wie die Ergebnisse der Vorhersage gezeigt haben. Schließlich lässt sich keine Aussage über die Affinität der Wechselwirkung an Hand der qualitativen Sequenz-Aktivitätsmodelle treffen. Daher werden im folgenden quantitative Wechselwirkungsmodelle der Struktur-Aktivitäts-Beziehungen entwickelt, die einige dieser Einschränkungen aufheben.

**Tabelle 3-8 Vergleich der Klassifizierung der Tyrosin-Spezifität durch die pHMMs der Y-Spezifitätsgruppe und des Y-Asts sowie durch das Kernkonsensusmotiv der Y-Spezifitätsgruppe**

Die Klassifizierung durch das pHMM der Y-Spezifitätsgruppe stimmt am besten mit dem Vorkommen des Kernkonsensusmotivs der Y-Spezifitätsgruppe überein

Klassifizierung der Tyrosin-Spezifität durch																		
pHMM der Y-Spezifitätsgruppe	Kern-Konsensusmotiv	pHMM des Y-Asts																
<table border="1"> <tr><td>Y</td><td>157</td></tr> </table>	Y	157	<table border="1"> <tr><td>Y</td><td>146</td></tr> <tr><td>∅Y</td><td>11</td></tr> </table>	Y	146	∅Y	11	<table border="1"> <tr><td>Y</td><td>141</td></tr> <tr><td>Y</td><td>5</td></tr> <tr><td>Y</td><td>10</td></tr> <tr><td>Y</td><td>1</td></tr> </table>	Y	141	Y	5	Y	10	Y	1		
	Y	157																
	Y	146																
∅Y	11																	
Y	141																	
Y	5																	
Y	10																	
Y	1																	
<table border="1"> <tr><td>Y</td><td>82</td></tr> </table>	Y	82	<table border="1"> <tr><td>Y</td><td>22</td></tr> <tr><td>∅Y</td><td>60</td></tr> </table>	Y	22	∅Y	60	<table border="1"> <tr><td>Y</td><td>13</td></tr> <tr><td>∅Y</td><td>9</td></tr> <tr><td>Y</td><td>51</td></tr> <tr><td>Y</td><td>4</td></tr> <tr><td>∅Y</td><td>5</td></tr> </table>	Y	13	∅Y	9	Y	51	Y	4	∅Y	5
	Y	82																
	Y	22																
∅Y	60																	
Y	13																	
∅Y	9																	
Y	51																	
Y	4																	
∅Y	5																	
<table border="1"> <tr><td>∅Y</td><td>243</td></tr> </table>	∅Y	243	<table border="1"> <tr><td>Y</td><td>2</td></tr> <tr><td>∅Y</td><td>241</td></tr> </table>	Y	2	∅Y	241	<table border="1"> <tr><td>∅Y</td><td>2</td></tr> <tr><td>Y</td><td>8</td></tr> <tr><td>Y</td><td>17</td></tr> <tr><td>∅Y</td><td>216</td></tr> </table>	∅Y	2	Y	8	Y	17	∅Y	216		
∅Y	243																	
Y	2																	
∅Y	241																	
∅Y	2																	
Y	8																	
Y	17																	
∅Y	216																	

Die Anzahl der WW-Domänen, die als tyrosinspezifisch (Y, dunkelrot hinterlegt), potentiell tyrosinspezifisch (Y, hellrot hinterlegt) und nicht tyrosinspezifisch (Spalte ∅Y) klassifiziert wurden, ist durch entsprechend gefärbte Flächen dargestellt. In der linken Spalte ist die Klassifizierung durch das pHMM der Y-Spezifitätsgruppe (siehe Abbildung 3-14A), in der rechten Spalte die Klassifizierung durch das pHMM des Y-Asts (siehe Abbildung 3-14B) und in der mittleren Spalte die Klassifizierung durch das Kernkonsensusmotiv gezeigt (siehe Seite 68). Dabei werden die klassifizierten Gruppen jeweils von links nach rechts durch die folgenden Klassifizierungsmethoden weiter aufgeteilt.

### 3.3 Quantitative Struktur-Aktivitäts-Beziehungen

Die in der Raumstruktur einer Domäne beinhaltete Information sollte nach dem „Sequenzbestimmt-Struktur und Struktur-bestimmt-Funktion“-Paradigma<sup>9</sup> eine im Vergleich zur Sequenzinformation verbesserte Vorhersage der Spezifität ermöglichen. Besonders wichtig ist dabei die Kenntnis der Form und Eigenschaften der Oberfläche, denn zusammen sind sie die eigentlichen Faktoren, welche die Spezifität bestimmen. Gleichzeitig ermöglichen die strukturbasierten Wechselwirkungsmodelle auch eine quantitative Aussage über die Stärke der Wechselwirkungsaktivität bzw. Affinität.

Um für die 42 experimentell untersuchten WW-Domänen 3D-Struktur-Aktivitäts-Beziehungen (QSAR = *engl. quantitative structure activity relationships*) bestimmen zu können, wurden ihre Strukturen durch Homologie-Modellierung erstellt (siehe 3.3.1). Anschließend wurde sowohl in den Strukturen der WW-Domänen nach den spezifitätsbestimmenden strukturellen Determinanten gesucht (siehe 3.3.2), als auch in den Modellen der WW-Domänen/Liganden-Komplexstrukturen die charakteristischen Wechselwirkungen identifiziert (siehe 3.3.3). Schließlich wurden mit einem Wechselwirkungsmodell peptidische Liganden mit erhöhter Affinität vorhergesagt und experimentell verifiziert (siehe 3.3.3.3).

#### 3.3.1 Modellierung der Strukturen der 42 WW-Domänen und ihrer Komplexe

Nur für wenige der 42 WW-Domänen standen experimentell bestimmte Strukturen zur Verfügung (siehe Strukturvergleich in Tabelle 3-9). Daher wurden die Raumstrukturen der übrigen Domänen für die Analysen der Struktur-Aktivitäts-Beziehungen modelliert. Um jedoch in den statischen Wechselwirkungsmodellen eine korrelierte Fehlerstruktur auf Grund der unterschiedlichen Strukturgenerierungsmethoden zu vermeiden, wurden die Analysen ausschließlich mit Strukturmodellen durchgeführt, und zwar auch für diejenigen Domänen, für die experimentell bestimmte Strukturen zur Verfügung standen. Damit belegt diese Analyse von Struktur-Aktivitäts-Beziehungen auf der Basis von modellierten Strukturen gleichzeitig auch das Potential der hier vorgestellten Methoden für Fälle, in denen nur wenige Strukturinformationen zur Verfügung stehen.

### 3.3.1.1 Homologie-Modellierung von WW-Domänen

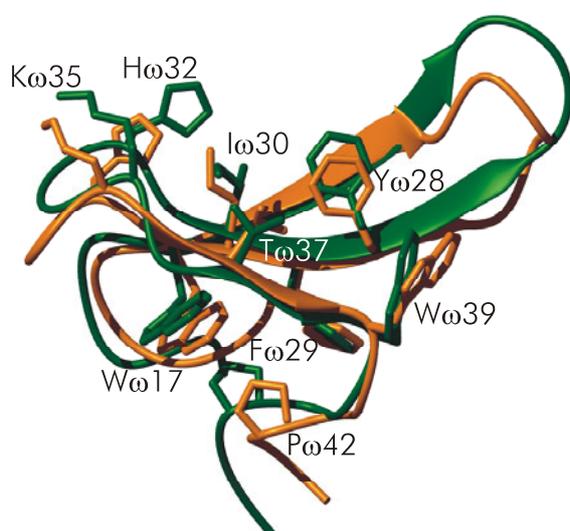
Die Strukturen aller 42 WW-Domänen wurden mittels Homologie-Modellierung basierend auf der experimentell bestimmten Struktur von *hPIN1*-WW [PDB: 1PIN] erstellt. *hPIN1*-WW war als Strukturvorlage auf der Basis folgender Kriterien ausgewählt worden: (i) 1PIN war die zu diesem Zeitpunkt am höchsten aufgelöste Röntgenkristallstruktur einer WW-Domäne (1,35 Å). (ii) Die Sequenz von *hPIN1*-WW hatte die höchste durchschnittliche Sequenzidentität (etwa 37%) gegenüber den Sequenzen der 42 zu modellierenden WW-Domänen. Ausnahmen bildeten die Sequenzen von  $\gamma$ S5M4-WW (15%) und  $\gamma$ TIN1-WW (21%) mit einer Sequenzidentität von deutlich unter 30%. (iii) Von *hPIN1*-WW gab es eine ebenfalls hochaufgelöste Röntgenkristallstruktur (1,84 Å) im Komplex mit einem Liganden [PDB: 1F8A],<sup>193</sup> die sich mit der Struktur 1PIN gut überlagern ließ (Tabelle 3-9B).

*hPIN1* ist die Peptidylprolyl cis-trans Isomerase NIMA-interacting 1 [UP: Q13526]. Die *hPIN1*-WW Domäne gehört zur *poS/poT*-Spezifitätsgruppe bzw. zur Gruppe IV (siehe Tabelle 1.3).

Nach der Modellierung wurde die Energie jeder WW-Domänenstruktur im Vakuum mit dem Kraftfeld von Cornell *et al.*, 1995, minimiert. Alle Schritte der Modellierung wurden automatisch von einem für Hochdurchsatz-Modellierung entwickelten Skript-Paket mit Tripos SYBYL (Tripos, St. Louise, USA) durchgeführt.

### 3.3.1.2 Qualität der modellierten Domänenstrukturen

Zur Bestimmung der Qualität der modellierten Strukturen wurden für diejenigen der 42 WW-Domänen, für die experimentelle Strukturen bekannt waren, die Strukturmodelle mit den experimentell bestimmten Strukturen verglichen. Jeweils zwei zu vergleichende Strukturen wurden dazu an Hand ihrer  $\beta$ -Stränge bzw. des gesamten  $\beta$ -Faltblatts überlagert. Die Ähnlichkeit der Strukturen wurde in Form der mittleren quadratischen Abweichung der Atompositionen (RMSD) quantifiziert (Tabelle 3-9 und Abbildung 3-16).



**Abbildung 3-16 Vergleich der modellierten und der experimentell bestimmten Struktur von *hNEDD4*-WW3**

Die NMR-Struktur [PDB: 115H]<sup>87</sup> (orange) wurde mit der modellierten Struktur von *hNEDD4*-WW3 (grün) an Hand der  $\beta$ -Stränge (Atome des Peptidrückgrats und der Seitenketten) überlagert (RMSD = 1,98 Å). Die Seitenketten der charakteristischen Aminosäuren der WW-Domäne, die zur Y-Spezifitätsgruppe gehört, sind explizit gezeigt und benannt.

**Tabelle 3-9 Abweichung der modellierten von den experimentell bestimmten Strukturen der WW-Domänen**

Die Abweichungen der modellierten von den experimentell bestimmten Strukturen (unten) liegen in derselben Größenordnung, wie die Abweichungen unterschiedlicher experimentell bestimmter Strukturen untereinander (oben).

	Domäne	PDB-ID <sup>a)</sup>	PDB-ID <sup>a)</sup> versus	RMSD <sup>b)</sup> [Å]			
				β-Stränge <sup>c)</sup>		β-Faltblatt <sup>d)</sup>	
				RG <sup>e)</sup>	Gesamt <sup>f)</sup>	RG <sup>e)</sup>	Gesamt <sup>f)</sup>
Experiment versus Experiment	h DYSTRO-WW	1EG4	1EG3*	0,25	0,74	0,55	1,06
	h PIN1-WW	1F8A	1PIN*	0,49	1,39	1,28	2,26
	h YAP65-WW	1K9Q	1JMQ	0,56	1,52	1,14	1,77
	h YAP65-WW	1K9R	1JMQ	0,87	1,45	1,31	1,87
Modell versus Experiment	h DYSTRO-WW	Modell	1EG3	0,89	2,33	1,82	2,99
	h NEDD4-WW3	Modell	1I5H	1,00	1,98	1,67	2,56
	h PIN1-WW	Modell	1PIN	0,39	1,24	0,61	1,16
	γ PRP40-WW1	Modell	1O6W_A1	0,83	1,93	1,18	2,12
	γ PRP40-WW2	Modell	1O6W_A2	0,76	1,91	1,03	2,10
	h YAP65-WW	Modell	1JMQ	1,31	2,12	2,08	2,75
	γ YJQ8-WW	Modell	1E0N	0,82	2,39	1,17	2,71
	Prototyp	Modell	1E0M	1,06	2,28	1,64	2,88

\* Strukturen ohne Ligand. Ansonsten sind die verglichenen Strukturen jeweils im Komplex mit ähnlichen Liganden.

<sup>a)</sup> Für experimentell bestimmte Strukturen ist der Zugriffscodex der Struktur in der *Protein Data Bank*<sup>20</sup> angegeben. Die erstellten Strukturmodelle, die auf der Strukturvorlage 1PIN basieren, sind durch *Modell* bezeichnet.

<sup>b)</sup> RMSD: mittlere quadratische Abweichung derjenigen Atome in Å, die zur Überlagerung verwendet wurden.

<sup>c)</sup> β-Stränge: βA (Reste ω17–ω21), βB (ω27–ω31) und βC (ω36–ω39).

<sup>d)</sup> β-Faltblatt: Positionen ω17–ω39 (βA, βB, βC sowie die Schleifen dazwischen).

<sup>e)</sup> RG: Atome des Peptidrückgrats.

<sup>f)</sup> Gesamt: alle Atome des Peptidrückgrats und der Seitenketten.

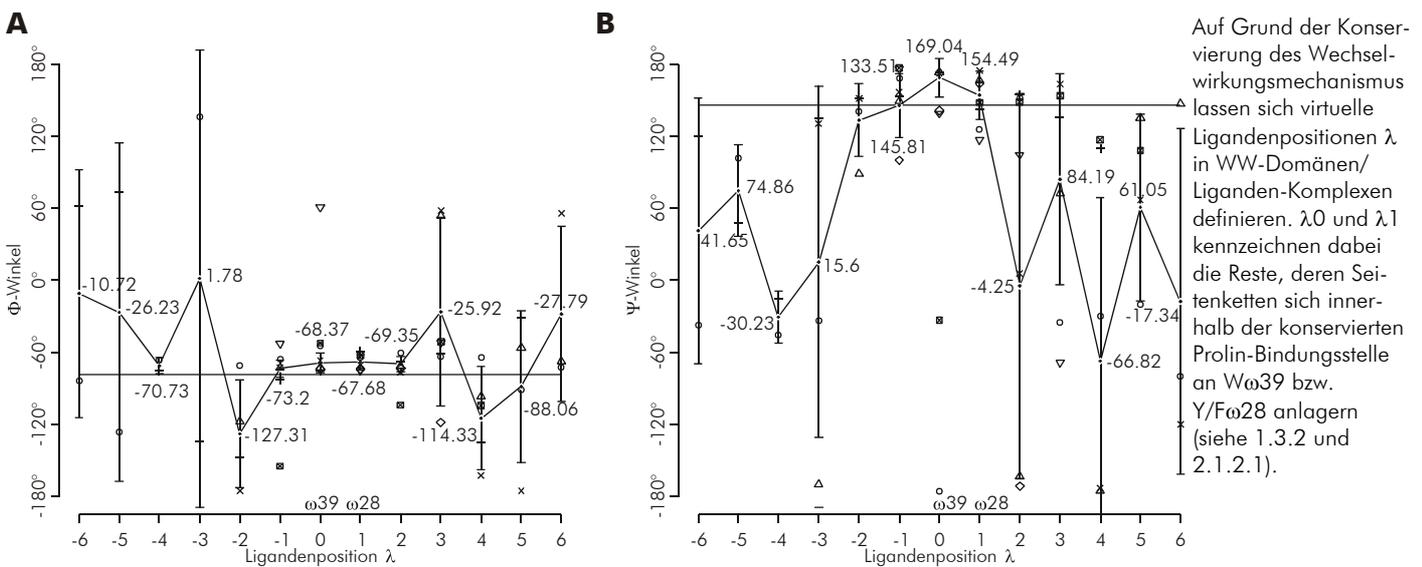
Die Abweichungen der modellierten von den experimentell bestimmten Strukturen (0,39–1,31 Å, RMSD des Peptidrückgrats der β-Stränge) liegen in derselben Größenordnung wie die Abweichungen der experimentell bestimmten Strukturen untereinander (0,25–0,87 Å), für WW-Domänen, für welche mehrere experimentell bestimmte Strukturen zur Verfügung standen (Tabelle 3-9). Damit konnte gezeigt werden, dass die Modellierung der Strukturen mittels Homologie-Modellierung infolge der starken Konservierung der Struktur der WW-Domänen durchführbar ist.

### 3.3.1.3 Analyse der Konformation und Positionierung der von WW-Domänen gebundenen Liganden

Für die Modellierung von WW-Domänen/Liganden-Komplexen war es erforderlich, die Konformation ( $\Phi$ - und  $\Psi$ -Winkel) und Positionierung der peptidischen Liganden an der konservierten Prolin-Bindungsstelle in den experimentell bestimmten Komplexstrukturen zu analysieren. In Abbildung 3-17 ist die Verteilung der  $\Phi$ - und  $\Psi$ -Torsionswinkel des Peptidrückgrats der Liganden in experimentell bestimmten Komplexstrukturen dargestellt.

Die Analyse zeigte, dass der Kernbereich ( $\lambda-2$  bis  $\lambda_2$ ) der von WW-Domänen gebundenen Liganden fast ausschließlich in Polyprolin-Typ-II-Helix-Konformation vorliegt ( $\Phi$ -Winkel  $\approx -78^\circ$  für  $\lambda-1$  bis  $\lambda_2$ ;  $\Psi$ -Winkel  $\approx 146^\circ$  für  $\lambda-2$  bis  $\lambda_1$ ). In den flankierenden Peptidbereichen sind die Torsionswinkel weniger bis gar nicht konserviert. Allein in den beiden ungewöhnlichen Strukturen 1K9R und 1F8A weichen die  $\Phi$ -Winkel bzw.  $\Psi$ -Winkel der Liganden an  $\lambda_0$  deutlich von einer PPII-Helix ab. Auch die Positionierung des Kernbereichs der Liganden an der konservierten Bindungsstelle der WW-Domänen ist stark konserviert. Dies hat der Vergleich der relativen Positionierung der Liganden in verschiedenen experimentell bestimmten WW-Domänen/Liganden-Komplexen gezeigt (siehe Tabelle 3-10, Seite 83). Auf Grund der hohen Konservierung der Konformation und Positionierung der peptidischen Liganden im Komplex mit WW-Domänen können Strukturmodelle der Komplexe ebenfalls mittels Homologie-Modellierung erstellt werden.

Eine Polyprolin-Typ-II-Helix (PPII) ist definiert durch einen  $\Phi$ -Winkel  $\approx -78^\circ$  und  $\Psi$ -Winkel  $\approx 146^\circ$ .



**Abbildung 3-17** Verteilung der  $\Phi$ - und  $\Psi$ -Torsionswinkel des Peptidrückgrats prolinreicher Liganden im Komplex mit WW-Domänen

(A) Die  $\Phi$ -Winkel der peptidischen Liganden sind an den virtuellen Ligandenpositionen  $\lambda-1$  bis  $\lambda_2$ , (B) die  $\Psi$ -Winkel an den Positionen  $\lambda-2$  bis  $\lambda_1$  hoch konserviert. Alle Liganden besitzen an diesen Positionen eine Polyprolin-Typ-II-Helix-Konformation (PPII). Die Diagramme vergleichen die Torsionswinkel peptidischer Liganden in gut aufgelösten Strukturen experimentell bestimmter WW-Domänen/Liganden-Komplexe. Die durchschnittlichen Winkel  $\blacklozenge$  (ohne Berücksichtigung der beiden stärker abweichenden Komplexe 1K9R und 1F8A) sind durch eine Linie und die Standardabweichung durch vertikale Balken angegeben. Die Winkel einer idealen PPII-Helix ( $\Phi = -78^\circ$ ,  $\Psi = 146^\circ$ ) sind in den Diagrammen als horizontale Linie eingezeichnet.

Symbol	PDB-ID	Domäne	Ligandensequenz	Methode
$\bullet$	1I5H	hNEDD4-WW3	GSTLPIPGT <b>PP</b> NYDSL*	NMR
$\triangle$	1K5R	hYAP65-WW(S24M)	GTP <b>PP</b> PYTVG*	NMR
+	1EG4	hDYSTRO-WW	NMTPYRS <b>PP</b> PYVP*	Kristall
$\times$	1JMQ	hYAP65-WW(L30K)	GTP <b>PP</b> PYTVG*	NMR
$\diamond$	1K9Q	hYAP65-WW	G <b>PP</b> PY*	NMR
$\nabla$	1K9R	hYAP65-WW	PL <b>PP</b> PY*	NMR
$\boxtimes$	1F8A	hPIN1-WW	*SP <b>OT</b> POY	Kristall

$\lambda_0\lambda_{-1}$

$\circ = \text{poS}$ ,  $* = \text{C-Terminus}$

Auf Grund der Konservierung des Wechselwirkungsmechanismus lassen sich virtuelle Ligandenpositionen  $\lambda$  in WW-Domänen/Liganden-Komplexen definieren.  $\lambda_0$  und  $\lambda_1$  kennzeichnen dabei die Reste, deren Seitenketten sich innerhalb der konservierten Prolin-Bindungsstelle an  $\omega_{39}$  bzw.  $\omega_{28}$  anlagern (siehe 1.3.2 und 2.1.2.1).

### 3.3.1.4 Homologie-Modellierung der Strukturen der WW-Domänen/Liganden-Komplexe

Basierend auf den charakteristischen Erkennungsmotiven (3.2.1) und den Ergebnissen der Analyse der Konformation und Positionierung der von WW-Domänen gebundenen Liganden (3.3.1.3) wurden Strukturmodelle der WW-Domänen/Liganden-Komplexe für alle WW-Domänen-Spezifitätsgruppen erstellt. Mit Hilfe dieser Komplexstrukturmodelle konnten nicht nur die Bindungs- und Spezifitätsmechanismen der Spezifitätsgruppen veranschaulicht werden, sondern sie waren auch Voraussetzung für die Analyse quantitativer Struktur-Aktivitäts-Beziehungen auf der Ebene der Komplexstrukturen mit Hilfe der COMBINE-Analyse (3.3.3).

Für die Modellierung wurde pro Spezifitätsgruppe je ein charakteristisches Dekapeptid in Polyprolin-Typ-II-Helix-Konformation (PPII) modelliert: für die Y-Spezifitätsgruppe **GTTPPPYTVG** (Y-Ligand), für die R<sub>a</sub>- und R<sub>b</sub>-Spezifitätsgruppen **PPGPPPRGPP** (R-Ligand), für die L-Spezifitätsgruppe **APPTPPPLPP** (L-Ligand) und für die poly-P-Spezifitätsgruppe **PLIPIPPPLP** (poly-P-Ligand). Auf die Erstellung von Komplexstrukturmodellen für die poS/poT-Spezifitätsgruppe wurde verzichtet, da für diese Gruppe nicht genügend experimentelle Ergebnisse für eine quantitative Analyse der Struktur-Aktivitäts-Beziehungen zur Verfügung standen.

Die Ligandenmodelle wurden anschließend manuell an der Bindungsstelle der Strukturmodelle der 42 WW-Domänen positioniert (manuelles Docking). Als „Raster“ für die Positionierung der Aminosäurereste des Liganden dienten die virtuellen Ligandenpositionen  $\lambda-1$  bis  $\lambda 2$  (siehe 1.3.2). Diese waren aus den durchschnittlichen Positionen dieser Ligandenreste in experimentell bestimmten Komplexstrukturen abgeleitet worden. Die Plausibilität der beiden möglichen Orientierungen (N→C- bzw. C→N) sowie der verschiedenen Positionierung des Liganden entlang der PPII-Helixachse wurden an Hand der entsprechenden Substitutionsanalysen bzw. der Erkennungsmotive bestimmt (siehe 3.2.1). Hierfür wurde angenommen, dass die spezifitätsbestimmenden Reste im Liganden mit der Domäne interagieren. Gleichzeitig wurde versucht – in Analogie zu bekannten WW-Domänen/Liganden-Komplexen – zwei aufeinanderfolgende und besonders selektiv erkannte Proline in die virtuellen Ligandenpositionen  $\lambda 0$  und  $\lambda 1$  zu platzieren.

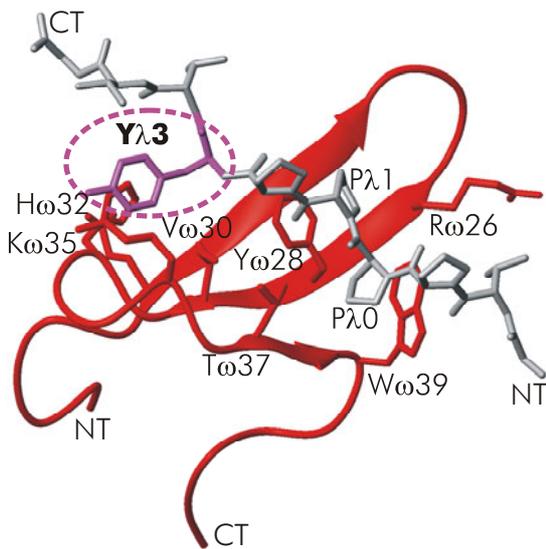


Abbildung 3-18 Y-Spezifitätsgruppe:  
Strukturmodell des hNEDD4-  
WW3/Y-Liganden-Komplexes

Die Struktur basiert auf den Substitutionsanalysen des Y-Liganden, welche das Erkennungsmotiv **xPPxY** ergaben (siehe Abbildung 3-2, Seite 57). Der Komplex aus der hNEDD4-WW3 Domäne und dem Y-Liganden **GTPP<sub>λ0</sub>PPY<sub>λ3</sub>TVG** wurde in C→N-Orientierung modelliert. Das für die Y-Spezifitätsgruppe charakteristische Tyrosin in Ligandenposition **λ3** ist violett eingefärbt und umrandet. Dargestellt sind die für die Ligandenbindung wichtigen Seitenketten. Die N- und C-Termini sind durch NT und CT gekennzeichnet.

Im Falle der **Y-Spezifitätsgruppe** wurden die Komplexe in Übereinstimmung mit den experimentell bestimmten Komplexstrukturen dieser Spezifitätsgruppe modelliert: der Ligand in C→N-Orientierung ist so positioniert, dass die beiden konsekutiven und hoch selektiv erkannten Proline des **xPPxY**-Erkennungsmotivs (siehe Abbildung 3-2, Seite 57) zwischen W<sub>ω</sub>39 und Y<sub>ω</sub>28 gepackt werden. Die Seitenkette des Tyrosins in Ligandenposition λ3 zeigt damit in Richtung der konservierten Tyrosin-Bindungsstelle, gebildet aus V/I/L<sub>ω</sub>30, H<sub>ω</sub>32 und R/K<sub>ω</sub>35 (Abbildung 3-18).

Die Komplexe der **R<sub>α</sub>-Spezifitätsgruppe** wurden mit dem R-Liganden in N→C-Orientierung modelliert. Um die aus den Substitutionsanalysen abgeleiteten potentiellen Kontakte zu erfüllen, wurden die beiden konsekutiven und am stärksten selektiv erkannten Proline (siehe Abbildung 3-3, Seite 57) in λ0 und λ1 positioniert. Dadurch resultierte für das spezifitätsbestimmende Arginin die Position λ-2, von wo aus es potentiell die konservierten negativ geladenen Aminosäuren (u.a. D<sub>ω</sub>22) in der βAβB-Schleife kontaktieren kann (Abbildung 3-19).

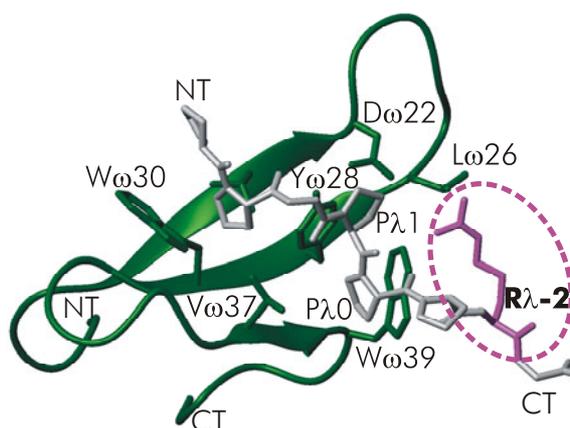
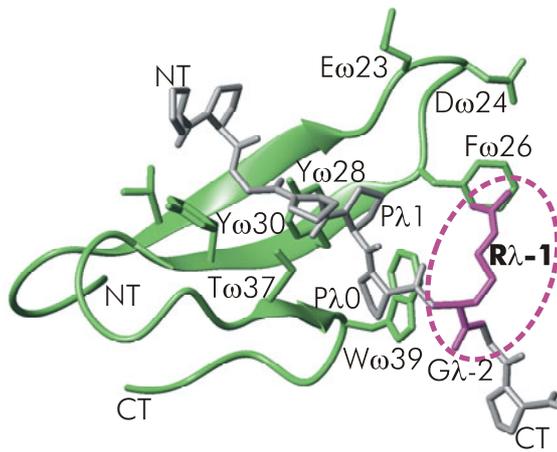


Abbildung 3-19 R<sub>α</sub>-Spezifitätsgruppe:  
Strukturmodell des hPQB1-WW/  
R-Liganden-Komplexes

Die Struktur basiert auf den Substitutionsanalysen des R<sub>α</sub>-Liganden, welche das Erkennungsmotiv **(p/Φ)P(p/g)PPpR** ergaben (siehe Abbildung 3-3, Seite 57). Der Komplex aus der hPQB1-WW Domäne und dem R-Liganden **PPGPP<sub>λ0</sub>PR<sub>λ-2</sub>GPP** wurde in N→C-Orientierung modelliert. Das für die R<sub>α</sub>-Spezifitätsgruppe charakteristische Arginin in Ligandenposition **λ-2** ist violett eingefärbt und umrandet. Dargestellt sind die für die Ligandenbindung wichtigen Seitenketten. Die N- und C-Termini sind durch NT und CT gekennzeichnet.

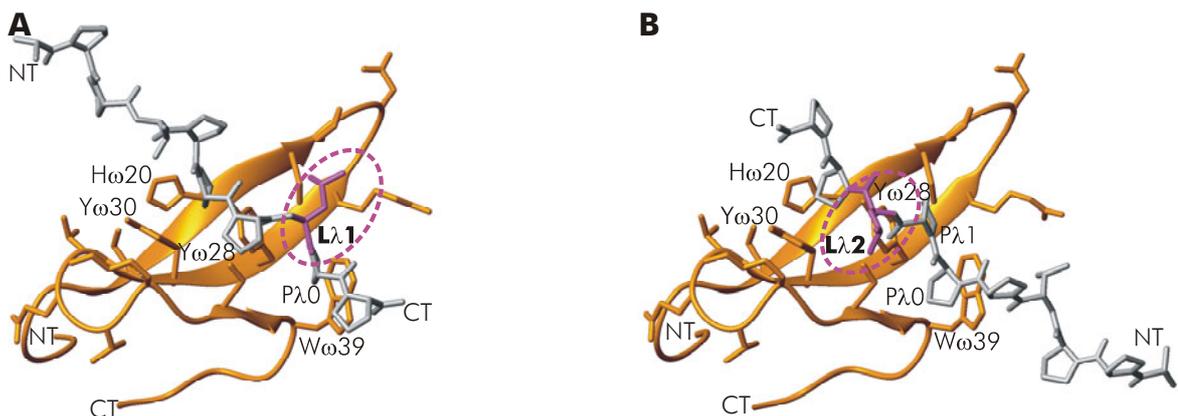


**Abbildung 3-20 R<sub>b</sub>-Spezifitätsgruppe:**  
Strukturmodell des hFBP21-WW2/  
R-Liganden-Komplexes

Die Struktur basiert auf der Substitutionsanalyse des R<sub>k</sub>-Liganden, welche das Erkennungsmotiv **pPFRgpPp** ergab (siehe Abbildung 3-4, Seite 58). Der Komplex aus der hFBP21-WW2 Domäne und dem R-Liganden **PPGPPP<sub>λ0</sub>R<sub>λ-1</sub>GPP** wurde in N→C-Orientierung modelliert. Das für die R<sub>b</sub>-Spezifitätsgruppe charakteristische Arginin in Ligandenposition **λ-1** ist violett eingefärbt und umrandet. Dargestellt sind die für die Ligandenbindung wichtigen Seitenketten. Die N- und C-Termini sind durch NT und CT gekennzeichnet.

Die **R<sub>b</sub>-Spezifitätsgruppe** unterscheidet sich von der R<sub>o</sub>-Spezifitätsgruppe vor allem dadurch, dass das spezifitätsbestimmende Arginin direkt auf die beiden stark substitutions-sensitiven Proline folgt (siehe Abbildung 3-4, Seite 58). Daher befindet sich das Arginin in λ-1 bei einer N→C-Orientierung des R-Liganden und der Positionierung der beiden hoch selektiv erkannten Proline in λ1 und λ0. Auch aus dieser Position ist ein Kontakt mit den konservierten negativ geladenen Aminosäuren (u.a. D/Eω24) in der βαβB-Schleife möglich (Abbildung 3-20).

Für die **L-Spezifitätsgruppe** konnte kein eindeutiges Strukturmodell aus den Substitutionsanalysen abgeleitet werden. Wenn die beiden konsekutiven, hoch selektiv erkannten

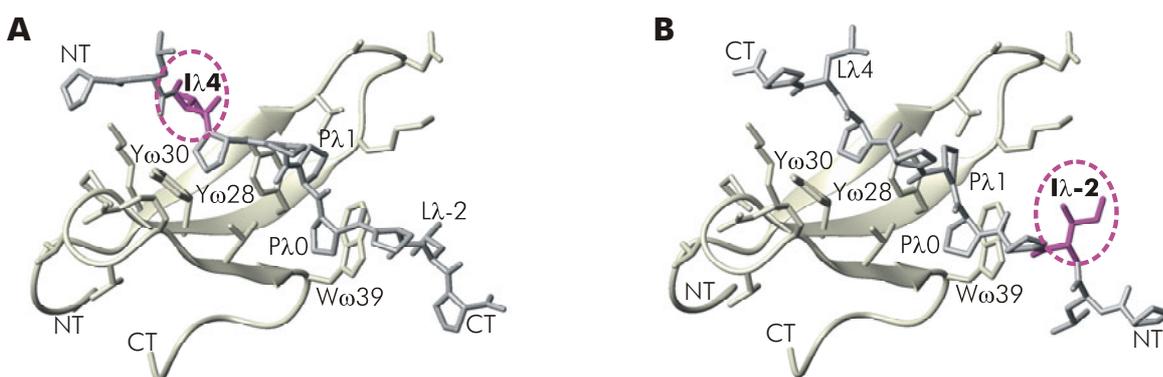


**Abbildung 3-21 L-Spezifitätsgruppe:**  
Vorschläge für Strukturmodelle des hFBP11-WW1/L-Liganden-Komplexes

Die Strukturvorschläge basieren auf den Substitutionsanalysen des L-Liganden, welche das Erkennungsmotiv **PPLPp** ergaben (siehe Abbildung 3-7, Seite 59). **(A)** Der Komplex aus der hFBP11-WW1 Domäne und dem L-Liganden **APPTPPP<sub>λ1</sub>P<sub>λ0</sub>P** wurde in N→C-Orientierung modelliert. Das für die L-Spezifitätsgruppe charakteristische Leucin befindet sich in diesem Vorschlag in Ligandenposition **λ1** (violett eingefärbt und umrandet). **(B)** Der Komplex aus der hFBP11-WW1 Domäne und dem L-Liganden **APPTPP<sub>λ0</sub>PL<sub>λ2</sub>PP** wurde in C→N-Orientierung modelliert. Das für die L-Spezifitätsgruppe charakteristische Leucin befindet sich in diesem Vorschlag in Ligandenposition **λ2** (violett eingefärbt und umrandet). Dargestellt sind die Seitenketten derjenigen Reste, die auf der zum Liganden exponierten Seite der WW-Domäne liegen. Die potentiell an der Ligandenbindung beteiligten Reste sind benannt. Die N- und C-Termini sind durch NT und CT gekennzeichnet.

Proline in die Positionen  $\lambda 0$  und  $\lambda 1$  platziert werden (siehe Abbildung 3-7, Seite 59), zeigt die Seitenkette des spezifitätsbestimmenden Leuzins unabhängig davon, welche Liganden-Orientierung gewählt wurde, in Richtung des hydrophilen Lösungsmittels (Abbildung 3-21B). Dies ist unwahrscheinlich für die hydrophobe Seitenkette von Leuzin und erklärt in keinem Fall die Spezifität für Leuzin. Wird der Ligand in Analogie zur Struktur des *hYAP65*-WW/PLPPY-Komplexes [PDB: 1K9R]<sup>138</sup> so positioniert, dass sich das Leuzin in  $\lambda 0$  befindet, werden nicht alle aus der Substitutionsanalyse abgeleiteten Kontakte erfüllt (Strukturmodell nicht gezeigt). Allein die Positionierung des Leuzins in  $\lambda 1$  bei einer N→C-Orientierung würde die abgeleiteten Kontakte erfüllen (Abbildung 3-21A). Die Spezifität für Leuzin an dieser Position lässt sich jedoch auch in diesem Fall weder aus der Sequenz der Domäne noch aus der Struktur erklären. Nur die Analogie zur Positionierung des Phosphoserins in den Komplexen von *hPIN1*-WW ist ein Indiz für diese mögliche Variante. Auf Grund dieser Unschlüssigkeiten wurde auf eine Analyse der Struktur-Aktivitäts-Beziehungen an den Strukturmodellen für die L-Spezifitätsgruppe verzichtet.

Im Falle der **poly-P-Spezifitätsgruppe** konnte kein eindeutiges Strukturmodell aus den Substitutionsanalysen abgeleitet werden. Eine Bewertung der N→C- bzw. C→N-Orientierung ist auf Grund der Pseudosymmetrie der PPII-Helix mit diesen Strukturmodellen nicht objektivierbar (Abbildung 3-22). Allerdings wird die N→C-Orientierung (Abbildung 3-22A)



**Abbildung 3-22 poly-P-Spezifitätsgruppe:**  
Vorschläge für Strukturmodelle des *hPRP40*-WW2/poly-P-Liganden-Komplexes

Die Strukturvorschläge basieren auf den Substitutionsanalysen des L/P<sub>k</sub>-Liganden, welche das Erkennungsmotiv (p/Φ) P P P P P ergaben (siehe Abbildung 3-8, Seite 60). **(A)** Der Komplex aus der *hPRP40*-WW2 Domäne und dem poly-P-Liganden P L I<sub>λ4</sub> P P P P λ<sub>0</sub> P L P wurde in N→C-Orientierung modelliert. Das für die poly-P-Spezifitätsgruppe charakteristische Motiv (p/Φ) P P P P P beginnt in diesem Vorschlag mit einem Isoleuzin in Ligandenposition **λ4** (violett eingefärbt und umrandet). **(B)** Der Komplex aus der *hPRP40*-WW2 Domäne und dem poly-P-Liganden P L I<sub>λ-2</sub> P P λ<sub>0</sub> P P P L P wurde in C→N-Orientierung modelliert. Das für die poly-P-Spezifitätsgruppe charakteristische Motiv (p/Φ) P P P P P beginnt in diesem Vorschlag mit einem Isoleuzin in Ligandenposition **λ-2** (violett eingefärbt und umrandet). Dargestellt sind die Seitenketten derjenigen Reste, die auf der zum Liganden exponierten Seite der WW-Domäne liegen. Die potentiell an der Ligandenbindung beteiligten Reste sind benannt. Die N- und C-Termini sind durch NT und CT gekennzeichnet.

leicht favorisiert, da in dieser Struktur die beiden konsekutiven und am selektivsten erkannten Proline in den Ligandenpositionen  $\lambda_0$  und  $\lambda_1$  zum Liegen kommen (siehe Abbildung 3-8, Seite 60). Auch in diesem Fall wurde auf eine Analyse der Struktur-Aktivitäts-Beziehungen an den Strukturmodellen für die poly-P-Spezifitätsgruppe verzichtet.

Mit diesem Satz an modellierten Komplexstrukturen war die Voraussetzung für eine vergleichende quantitative Analyse der charakteristischen Interaktionen in WW-Domänen/Liganden-Komplexen mit Hilfe der COMBINE-Methode geschaffen worden (siehe 3.3.3). Allerdings waren für diese Analyse zusätzlich zu den „bindenden“ WW-Domänen/Liganden-Kombinationen, die durch das WW-Screening identifiziert worden waren, auch Strukturmodelle der Komplexe „nicht bindender“ Kombinationen als Referenz nötig. Als „nicht bindend“ werden dabei diejenigen WW-Domänen/Liganden-Kombinationen bezeichnet, die durch das WW-Screening ausgeschlossen worden waren: so zum Beispiel der Komplex einer WW-Domäne der Y-Spezifitätsgruppe mit dem R-Liganden. Für die Modellierung der Komplexe „nicht bindender“ Kombinationen wurde der Ligand in der Orientierung und Positionierung, wie sie für die Spezifitätsgruppe des jeweiligen Liganden charakteristisch ist, an der Bindungsstelle platziert. Insgesamt wurden damit 504 Komplexe „bindender“ und „nicht bindender“ Kombinationen (= 42 x [Y-Ligand in 2 Positionen + R-Ligand in 4 Positionen + L-Ligand in 4 Positionen + poly-P-Ligand in 2 Positionen]) in einem Hochdurchsatz-Modellierungs-Verfahren berechnet.

### 3.3.1.5 Qualität der modellierten Komplexstrukturen

Zur Bewertung der Qualität der modellierten Komplexstrukturen wurde die Positionierung des Liganden an der konservierten Prolin-Bindungsstelle zwischen modellierten und experimentell bestimmten Strukturen verglichen. Dazu wurden die Abweichungen der Ligandenpositionen  $\lambda_{-1}$  bis  $\lambda_3$  zwischen modellierten und experimentell bestimmten Komplexstrukturen der Y-Spezifitätsgruppe ermittelt.  $\lambda_3$  wurde in diesem Fall zum Kernbereich des Liganden gezählt, da auf Grund der spezifischen Erkennung des Tyrosins auch diese Ligandenposition in Komplexen der Y-Spezifitätsgruppe konserviert sein sollte. Jeweils zwei zu vergleichende Komplexe wurden an Hand der  $\beta$ -Stränge der Domäne überlagert und die mittlere quadratische Abweichung (RMSD) der Atompositionen des Liganden bestimmt (Tabelle 3-10).

**Tabelle 3-10** Abweichung der Liganden zwischen modellierten und experimentell bestimmten WW-Domänen/Liganden-Komplexstrukturen

Die Abweichung der Positionierung des Liganden in den modellierten von den experimentell bestimmten Strukturen (unten) liegt in derselben Größenordnung, wie die Abweichung der Positionierung des Liganden in unterschiedlichen experimentell bestimmten Strukturen derselben Domäne (oben).

	Domäne	Ligand <sup>a)</sup>	PDB-ID <sup>b)</sup>		Domäne	Ligand <sup>a)</sup>	RMSD <sup>c)</sup> [Å]	
			versus				Ligand $\lambda-1$ bis $\lambda 3$	$C\alpha$ <sup>d)</sup>
Exp. vs. Exp.	<i>h</i> DYSTRO-WW	... <b>SPPPY</b> ...	1EG4	1JMQ	<i>h</i> YAP65-WW	... <b>P</b> PPPY...	1,30	1,60
	<i>h</i> NEDD4-WW3	... <b>P</b> PPNY...	115H	1EG4	<i>h</i> DYSTRO-WW	... <b>SPPPY</b> ...	1,82	2,43
	<i>h</i> YAP65-WW	... <b>P</b> PPPY...	1JMQ	115H	<i>h</i> NEDD4-WW3	... <b>P</b> PPNY...	2,00	3,13
Modell vs. Exp.	<i>h</i> DYSTRO-WW	... <b>SPPPY</b> ...	Modell	1EG4	<i>h</i> DYSTRO-WW	... <b>SPPPY</b> ...	2,13	2,28
	<i>h</i> NEDD4-WW3	... <b>P</b> PPPY...	Modell	115H	<i>h</i> NEDD4-WW3	... <b>P</b> PPNY...	2,40	2,77
	<i>h</i> YAP65-WW	... <b>P</b> PPPY...	Modell	1JMQ	<i>h</i> YAP65-WW	... <b>P</b> PPPY...	1,61	2,31

<sup>a)</sup> Sequenz des Liganden von Ligandenposition  $\lambda-1$  bis  $\lambda 3$ .  $\lambda 0$  und  $\lambda 1$  sind fett gedruckt.

<sup>b)</sup> Für experimentell bestimmte Strukturen ist der Zugriffscode der Struktur in der *Protein Data Bank*<sup>20</sup> angegeben. Die erstellten Strukturmodelle, die auf der Strukturvorlage 1PIN basieren, sind durch *Modell* bezeichnet.

<sup>c)</sup> RMSD: mittlere quadratische Abweichung der Atome der Ligandenpositionen  $\lambda-1$  bis  $\lambda 3$  in Å. Zur Überlagerung der Komplexe wurden die Atome der drei  $\beta$ -Stränge der Domäne ( $\beta A$  (Reste  $\omega 17-\omega 21$ ),  $\beta B$  ( $\omega 27-\omega 31$ ) und  $\beta C$  ( $\omega 36-\omega 39$ )) verwendet.

<sup>d)</sup>  $C\alpha$ : Berechnung der RMSD für die  $C\alpha$ -Atome der Ligandenpositionen  $\lambda-1$  bis  $\lambda 3$ .

<sup>e)</sup> Gesamt: Berechnung der RMSD für alle schweren Atome des Peptidrückgrats und der Seitenketten der Ligandenpositionen  $\lambda-1$  bis  $\lambda 3$ , welche die zu vergleichenden Aminosäuren gemeinsam haben.

Die Abweichungen der Ligandenpositionierung zwischen modellierten und experimentell bestimmten Strukturen (1,61–2,40 Å RMSD der  $C\alpha$ -Atome der Ligandenpositionen  $\lambda-1$  bis  $\lambda 3$ ) liegen in derselben Größenordnung, wie die Abweichungen zwischen experimentell bestimmten Strukturen untereinander (1,30–2,00 Å, siehe Tabelle 3-10). Dies zeigte, dass – auf Grund der hohen Konservierung der Ligandenpositionen an der konservierten Bindungsstelle – die hier verwendete Modellierungsprozedur qualitativ gute Strukturmodelle für die Y-Spezifitätsgruppe lieferte, welche mittels der COMBINE-Analyse weiter untersucht wurden (siehe 3.3.3).

### 3.3.2 3D-QSAR der Domänenstrukturen zur Vorhersage der Affinität gegenüber tyrosinhaltigen Liganden

Um die strukturellen Determinanten der Ligandenerkennung zu identifizieren und ihren Beitrag zur Ligandenbindung zu quantifizieren, wurden die modellierten Strukturen der WW-Domänen mit Hilfe der *Comparative-Molecular-Field-Analyse* (CoMFA)<sup>42</sup> untersucht. Am Beispiel der spezifischen Erkennung des Y-Liganden sollte dadurch das Potential dieser Methode für die Analyse modellierter Strukturen demonstriert werden.

Im Bereich der Arzneimittelforschung ist die CoMFA-Methode schon lange ein weit verbreitetes Werkzeug für das rationale Design kleiner, pharmakologisch wirksamer Moleküle.

Mit ihrer Hilfe werden 3D-Strukturmerkmale der kleinen Moleküle identifiziert, durch welche ihre Aktivität – meist die Wechselwirkung mit oftmals unbekanntem Proteinen – positiv bzw. negativ beeinflusst wird. Im Gegensatz dazu wurden im Rahmen der hier vorgestellten Arbeit mit Hilfe der CoMFA-Methode die strukturellen Determinanten der bindenden Proteine, genauer der WW-Domänen, analysiert. Meiner Kenntnis nach ist dies eine der ersten Anwendungen dieser Methode zur Identifizierung von Strukturmerkmalen in Proteinen, welche für die Bindung kleiner Moleküle – in diesem Fall Peptide – verantwortlich sind.

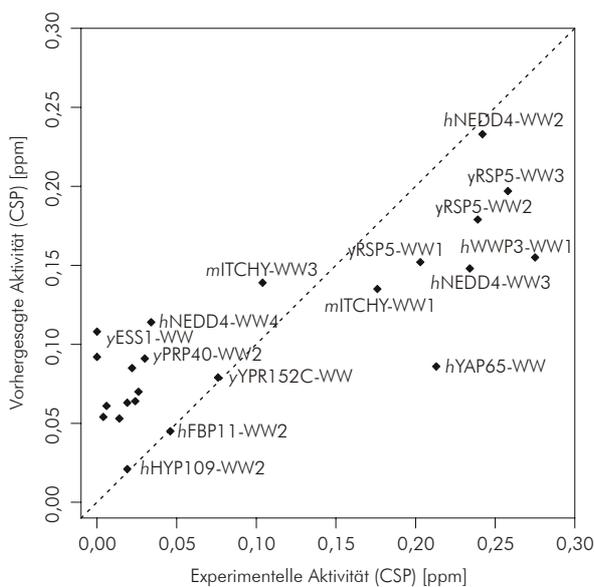
### 3.3.2.1 CoMFA-Modell der Tyrosin-Spezifität für WW-Domänen

Durch das CoMFA-Modell der Tyrosin-Spezifität wurden die Strukturen der WW-Domänen in eine quantitative Beziehung zu ihrer Bindungsaktivität gegenüber tyrosinhaltigen Liganden gesetzt. Als Aktivität wurde die bindungsinduzierte Veränderung der chemischen Verschiebung (CSP = engl. *chemical shift perturbation*) des Wasserstoffatoms am Indolstickstoff von W $\omega$ 39 verwendet, die im NMR-Screening bestimmt worden war (siehe Y-Modellligand Tabelle 3-6, Seite 63). Hierbei wurde davon ausgegangen, dass die Stärke der CSP in einer Beziehung zur Affinität des jeweiligen WW-Domänen/Y-Modellliganden-Komplexes steht (siehe Anhang C).

Für dieses CoMFA-Modell wurden alle 10 WW-Domänen der Y-Spezifitätsgruppe mit quantifizierbaren CSPs ausgewählt (Tabelle 3-6, Spalte CoMFA) sowie weitere 10 WW-Domänen aus den restlichen 5 Spezifitätsgruppen und 3 nicht klassifizierte mit quantifizierbaren CSPs als repräsentative „schwach bindende“ bzw. „nicht bindende“ Kombinationen.

Die CoMFA-Methode basiert darauf, dass jede Molekülstruktur durch je ein sterisches und ein elektrostatisches molekulares Feld beschrieben wird. Dazu werden die Eigenschaften – besetztes Volumen für das sterische und Ladung für das elektrostatische Feld – jeder Struktur an definierten Punkten eines 3D-Gitters bestimmt. Anschließend wird diese multivariate Beschreibung der verschiedenen Strukturen, mit denen das CoMFA-Modell trainiert wird, durch eine *Partial-Least-Squares-Regression* (PLS) in eine quantitative Beziehung zur Aktivität der analysierten Strukturen gesetzt (siehe auch Anhang D).

Das resultierende finale CoMFA-Modell basierte auf 4 Komponenten (latenten Variablen) und zeigte eine exzellente Anpassung an die Trainingsdaten mit einem Determinationskoeffizient  $R^2$  von 0,99. Die Vorhersageleistung dieses Modells wurde durch *leave-one-out* Kreuzvalidierung bestimmt. Hierfür wurde für jede WW-Domäne ein Modell unter Auslassung ihrer eigenen Daten trainiert. Mit diesem Modell wurde anschließend die CSP für die



**Abbildung 3-23 Vorhersageleistung des CoMFA-Modells der Tyrosin-Spezifität**

Die *leave-one-out* kreuzvalidierten Vorhersagen der Veränderungen der chemischen Verschiebung (CSP) durch das CoMFA-Modell der Tyrosin-Spezifität sind gegen die experimentell bestimmten CSPs (siehe Tabelle 3-6) aufgetragen. Der kreuzvalidierte Determinationskoeffizient  $Q^2$  beträgt 0,58.

„ausgelassene“ Domäne vorhergesagt und mit der experimentell bestimmten CSP verglichen. Das CoMFA-Modell zeigt mit einem  $Q^2$  von 0,58 (Determinationskoeffizient des kreuzvalidierten Modells) und einer *SDEP* von 0,07 (Standardabweichung der Fehler der Vorhersage) eine gute Vorhersageleistung. In Abbildung 3-23 sind die durch das kreuzvalidierte Modell vorhergesagten CSPs gegenüber den experimentell bestimmten CSPs aufgetragen.

In Abbildung 3-24 sind diejenigen Bereiche in der Bindungsstelle der WW-Domänen, die einen signifikanten Einfluss auf die vorhergesagte Aktivität zeigten, in Form von unregelmäßigen Sphären visualisiert. Die Farbe der Sphären kennzeichnet dabei, welches molekulare Feld die Vorhersage in welcher Weise beeinflusst. Die drei für die Ligandenbindung charakteristischsten Regionen mit den größten Beiträgen zur Vorhersage der Aktivität sind in Abbildung 3-24 markiert und werden im folgenden besprochen.

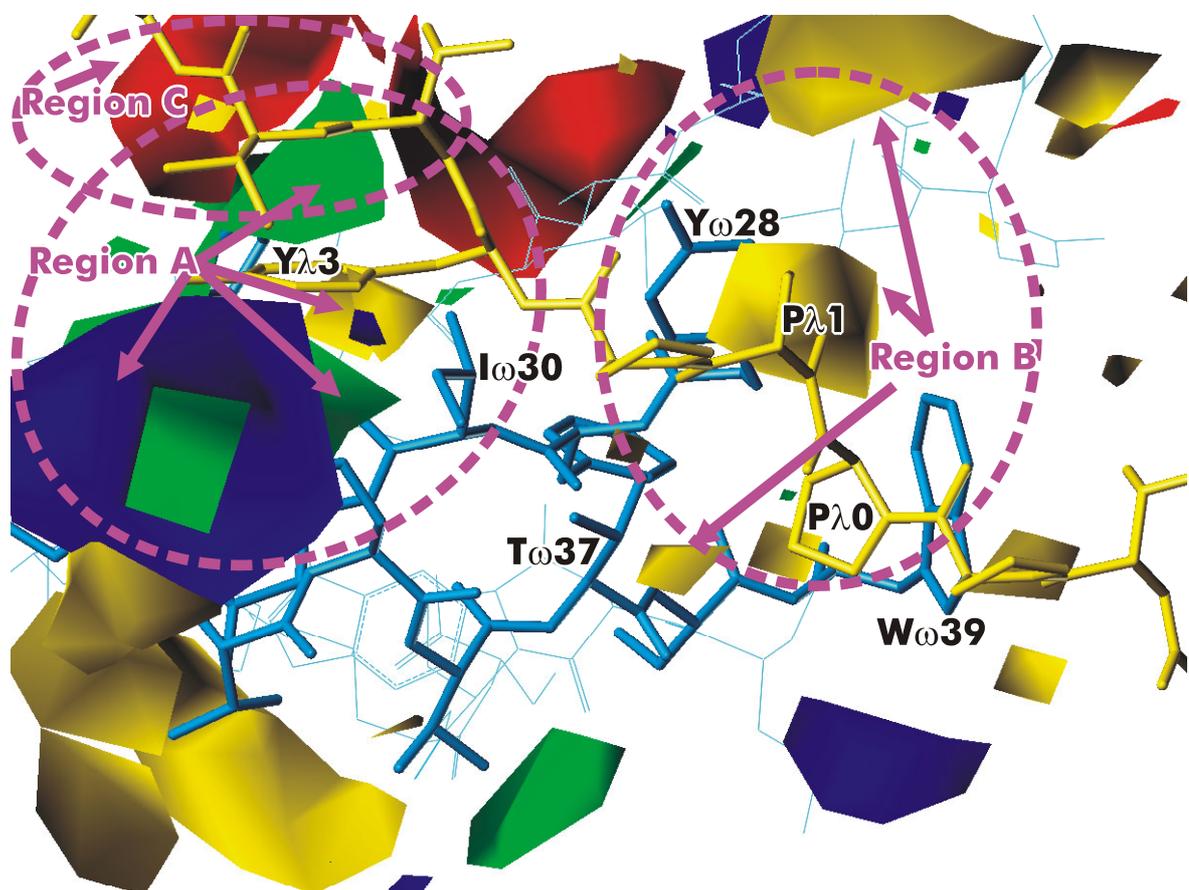
Der Bereich um die  $\beta\beta\beta$ C-Schleife, der für die Erkennung des Tyrosins im Liganden verantwortlich ist, zeigte den größten Einfluss auf die vorhergesagte Aktivität des CoMFA-Modells. Zu einer hohen Aktivität tragen dabei drei Faktoren bei: (i) Ein ausgeprägter Rand entlang der  $\beta\beta\beta$ C-Schleife rund um die zu bildende Tyrosin-Bindungsstelle, welcher durch die hufeisenförmige grüne Sphäre oberhalb der Schleife angedeutet wird. (ii) Eine ausgeprägte Vertiefung zwischen diesen Rändern, in welcher der Phenolring des Tyrosins Platz finden kann (gelbe Sphäre). (iii) Eine positive Ladung im Bereich der Seitenkette von Position  $\omega 35$  (blaue Sphäre).

**Region A**

Die Ausprägung der konservierten Prolin-Bindungsstelle, die durch Y/F $\omega 28$  und W $\omega 39$  gebildet wird, beeinflusst ebenfalls die vorhergesagte Aktivität gegenüber dem Y-Modellliganden. Fällt diese „Tasche“ kleiner aus, sinkt die Aktivität. Die Unterschiede in der Größe

**Region B**

der „Tasche“ hängen vor allem von der Entfernung und Orientierung der beiden Ringe von Y/F $\omega$ 28 und W $\omega$ 39 ab. Insbesondere die Orientierung des Rings von Y/F $\omega$ 28, die wesentlich durch die Konformation der vorangehenden  $\beta\alpha\beta$ -Schleife bedingt wird, trägt zu den Unterschieden in der Taschengröße bei. Dies wird durch eine gelbe Sphäre oberhalb der „Tasche“ auf der Seite von Y/F $\omega$ 28 angezeigt, dort wo der Ring von P $\lambda$ 0 positioniert ist. Daneben sollte die Seitenkette von Position  $\omega$ 37 nicht zu sehr in die „Tasche“ hineinragen, was durch eine gelbe Sphäre zwischen der „Tasche“ und Position  $\omega$ 37 angedeutet wird. Für eine hohe Aktivität sollten auch die Aminosäuren der  $\beta\alpha\beta$ -Schleife die Öffnung der Prolin-Bindungsstelle nicht beeinträchtigen. Dies wird durch eine gelbe Sphäre zwischen der „Tasche“ und der Schleife angedeutet.



**Abbildung 3-24 CoMFA-Modell der Tyrosin-Spezifität: Visualisierung der signifikanten Regionen**

Die Bereiche mit einem großen Einfluss auf die Vorhersage des CoMFA-Modells der Tyrosin-Spezifität sind über der hWWP3-WW1 Domäne (hellblau) als unregelmäßige Sphären visualisiert. Gezeigt ist die Domäne im Komplex mit dem Y-Liganden (gelb). Die wichtigsten Reste für die Bindung des Liganden sind explizit benannt. Die 20 % stärksten Beiträge des sterischen bzw. die 10 % stärksten Beiträge des elektrostatischen molekularen Felds sind durch unregelmäßige Sphären dargestellt. Der Einfluss des sterischen Felds ist durch grüne (durch die Domäne belegtes Volumen ist in diesem Bereich vorteilhaft für die Aktivität) bzw. gelbe Sphären (nachteilig) dargestellt. Der Einfluss des elektrostatischen Felds ist durch rote (negative Ladung der Domäne ist in diesem Bereich vorteilhaft) bzw. blaue Sphären (positive Ladung ist vorteilhaft) dargestellt. Die drei charakteristischsten Regionen **A**, **B** und **C** mit dem größten Einfluss auf das CoMFA-Modell sind hervorgehoben und die wichtigsten Sphären durch Pfeile markiert (siehe Text).

Eine rote Sphäre am Anfang des  $\beta$ A-Strangs zeigt, dass eine negative Ladung in diesem Bereich zu einer erhöhten vorhergesagten Aktivität gegenüber dem Y-Modellliganden führt.

Damit hat das CoMFA-Modell alle Strukturmerkmale identifiziert, die im direkten Zusammenhang zu den konservierten Sequenzmerkmalen der Y-Spezifitätsgruppe stehen. Zusätzlich schlägt diese Analyse auch eine Reihe struktureller Merkmale vor, welche die Wechselwirkung mit dem Liganden beeinflussen, aber nicht einfach auf der Ebene der Sequenz identifizierbar sind.

### 3.3.2.2 Strukturbasierte Vorhersage der Tyrosin-Spezifität

Um die Vorhersage des CoMFA-Modells der Tyrosin-Spezifität zu verifizieren, wurden die CSPs für WW-Domänen, die sich nicht im Trainingsdatensatz des Modells befanden, vorhergesagt. Hierfür wurden die Strukturen der WW-Domänen *hBAG3-WW* [UP: Q95817], *mSYNIP-WW* [UP: Q9WV89] und *tbZFP2-WW* [UP: Q967Y4], wie in Abschnitt 3.3.1.1 beschrieben, modelliert. Anschließend wurden für diese Strukturmodelle mit Hilfe der CoMFA-Methode die sterischen und elektrostatischen molekularen Felder berechnet.

Auf der Basis dieser molekularen Felder sagte das CoMFA-Modell der Tyrosin-Spezifität eine CSP von 0,16 ppm im Falle von *hBAG3-WW*, eine CSP von 0,10 ppm im Falle von *mSYNIP-WW* und eine CSP von 0,09 ppm im Falle von *tbZFP2-WW* voraus. Im Vergleich zu den für die 42 WW-Domänen experimentell bestimmten CSPs (siehe Tabelle 3-3, Seite 53) wurde damit für *hBAG3-WW* eine starke und für *mSYNIP-WW* sowie *tbZFP2-WW* eine schwache Wechselwirkung mit tyrosinhaltigen Liganden vorausgesagt (siehe 2.3.1.2).

Die Vorhersage der Tyrosin-Spezifität konnte für alle drei WW-Domänen experimentell durch Substitutionsanalysen des Y-Liganden (siehe Abbildung 3-15, Seite 72) bestätigt werden. Interessanterweise passt auch die vorhergesagte schwache Tyrosin-Spezifität für *tbZFP2-WW* zu dem Befund, dass diese WW-Domäne nicht mehr nur mit tyrosinhaltigen Liganden, sondern auch mit argininhaltigen Liganden interagieren kann. Damit konnte gezeigt werden, dass das CoMFA-Modell, obwohl es mit modellierten Strukturen trainiert wurde, für ebenfalls modellierte Strukturen experimentell verifizierbare Vorhersagen liefern kann.

### 3.3.3 3D-QSAR der Komplexstrukturen

Um die spezifitätsbestimmenden Interaktionen zwischen WW-Domänen und ihren peptidischen Liganden zu identifizieren, wurden die Strukturmodelle der WW-Domänen/Liganden-Komplexe mit Hilfe der COMBINE-Analyse (*Comparative-Binding-Energy-Analyse*)<sup>127</sup> untersucht. Die Analyse wurde ausschließlich mit modellierten Komplexstrukturen durchgeführt, da zu wenig experimentell bestimmte Komplexstrukturen zur Verfügung standen. Neben den Komplexstrukturen der Y-Spezifitätsgruppe wurden auch die WW-Domänen/Liganden-Komplexe der beiden R-Spezifitätsgruppen analysiert, für die im Rahmen dieser Arbeit in sich schlüssige Komplexstrukturmodelle erarbeitet werden konnten (siehe 3.3.1.4).

Die COMBINE-Analyse der R-Spezifitätsgruppen demonstriert dabei, welche Einblicke diese Methode in die spezifische Erkennung des Liganden liefern kann, und zwar in den Fällen, in denen keine quantitativen Daten über die Ligandenbindung vorliegen (siehe 3.3.3.1). Die COMBINE-Analyse der Y-Spezifitätsgruppe (siehe 3.3.3.2) zeigt darüber hinaus, wie bei Vorliegen quantitativer Aktivitätsdaten (CSP infolge der Bindung des Y-Modellliganden, siehe Tabelle 3-6, Seite 63) ein solches statistisches Wechselwirkungsmodell für das Design spezifischer Liganden mit erhöhter Affinität verwendet werden kann (siehe 3.3.3.3).

Die COMBINE-Methode basiert darauf, dass die Energiebeiträge – je ein elektrostatischer (Coulomb-Potential) und ein van-der-Waals (Lennard-Jones-Potential) Energiebeitrag – aller paarweisen Interaktionen zwischen den Aminosäuren des Liganden und der Domäne aus jeder Komplexstruktur, mit der das COMBINE-Modell trainiert wird, extrahiert werden (siehe auch Anhang D). Dazu werden die paarweisen Interaktionsenergien mit Hilfe eines Kraftfelds, in dieser Arbeit dem von Cornell *et al.*, 1995, in AMBER 7<sup>8</sup> berechnet. Schon ohne Aktivitätsdaten können aus dieser multivariaten Beschreibung der Komplexe die Schlüsselwechselwirkungen für die Bindung bestimmter Liganden mittels Hauptkomponenten-Analyse identifiziert werden (siehe 3.3.3.1 für die R-Spezifitätsgruppen). In einer vollständigen COMBINE-Analyse werden darüber hinaus die paarweisen Interaktionsenergien in eine quantitative Beziehung zu den experimentellen Aktivitätsdaten gesetzt. Dazu wird ein lineares statisches Modells durch PLS-Regression zwischen den paarweisen Interaktionsenergien und der Aktivität berechnet (siehe 3.3.3.2 für die Y-Spezifitätsgruppe).

### 3.3.3.1 Spezifitätsbestimmende Interaktionen der WW-Domänen/ R-Liganden-Wechselwirkung

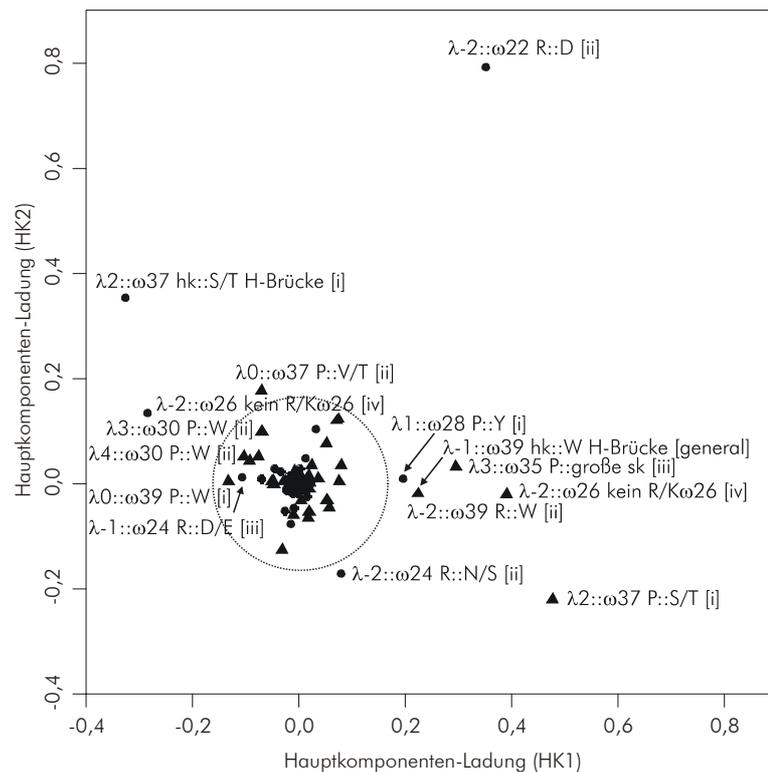
Zur Identifizierung der Schlüsselwechselwirkungen in den Strukturmodellen der WW-Domänen/R-Liganden-Komplexe wurden die mit der COMBINE-Methode extrahierten paarweisen Interaktionsenergien einer Hauptkomponenten-Analyse (PCA = engl. *principal component analysis*) unterzogen. Dafür wurden hFE65-WW und hPQBP1-WW aus der R<sub>α</sub>-Spezifitätsgruppe, hFBP21-WW1 und hFBP21-WW2 aus der R<sub>β</sub>-Spezifitätsgruppe, sowie γRSP5-WW1/2/3 als „nicht den R-Liganden bindende“ Domänen aus der Y-Spezifitätsgruppe ausgewählt (Tabelle 3-6, Spalte COMBINE/PCA). Für jede dieser WW-Domänen im Komplex mit dem R-Liganden wurden die paarweisen intermolekularen Interaktionsenergien mit der COMBINE-Methode berechnet.

Durch die PCA konnte eine Reihe potentiell spezifitätsbestimmender Interaktionen identifiziert werden (Abbildung 3-25, Seite 90). Diese ließen sich in vier Typen einteilen: allgemeine Interaktionen (Typ i), die in allen WW-Domänen/Liganden-Komplexen zu finden sind, und charakteristische Wechselwirkungen, die entweder in der R<sub>α</sub>- (Typ ii), der R<sub>β</sub>-Spezifitätsgruppe (Typ iii) oder in beiden R-Spezifitätsgruppen (Typ iv) zu finden sind.

Unter den **allgemeinen Interaktionen** sind die elektrostatische Wechselwirkung **Typ i** zwischen λ-1 und Wω39 sowie λ2 und ω37 zu finden, die den charakteristischen Wasserstoffbrücken entsprechen. Ebenfalls zu den allgemeinen Interaktionen gehören die van-der-Waals-Kontakte zwischen λ0 und Wω39, λ1 und Y/Fω28 sowie λ2 und ω37.

Charakteristisch für **beide R-Spezifitätsgruppen** ist der van-der-Waals-Kontakt **Typ iv** zwischen λ-2 und ω26. Diese Wechselwirkung kann beiden R-Spezifitätsgruppen zugeordnet werden, da die WW-Domänen der R-Spezifitätsgruppen im Gegensatz zur überwiegenden Zahl der anderen WW-Domänen an ω26 kein Arginin oder Lysin besitzen.

Zu den für die **R<sub>α</sub>-Spezifitätsgruppe** spezifischen Interaktionen zählen die elektrostatischen Wechselwirkungen des Arginins in λ-2 mit dem konservierten negativ geladenen Rest Dω22 sowie dem hydrophilen Rest S/Nω24. Das Arginin in λ-2 bildet auch van-der-Waals-Kontakte mit Wω39 aus. Eine Besonderheit der R<sub>α</sub>-Spezifitätsgruppe ist das Valin anstatt des hochkonservierten Serins/Threonins in ω37. Diese Eigenschaft zeigt sich wahrscheinlich in den van-der-Waals-Kontakten von ω37 mit den Ligandenresten in λ0 und λ2. Schließlich bilden die Ligandenreste in λ3 und λ4 van-der-Waals-Kontakte mit der großen hydrophoben Oberfläche des Tryptophans in ω30 aus. **Typ ii**



**Abbildung 3-25** Charakteristische Interaktionen in Komplexen mit dem R-Liganden

Die charakteristischen Interaktionen wurden durch eine Hauptkomponenten-Analyse der Interaktions-Energierterme der paarweisen intermolekularen elektrostatischen und van-der-Waals Interaktionen für Komplexe der beiden R-Spezifitätsgruppen mit dem R-Liganden identifiziert. Die Beiträge (Ladungen) zu den beiden einflussreichsten Faktoren (Hauptkomponenten HK1 und HK2) sind gegeneinander aufgetragen. Interaktionen, deren Ladung um mehr als 3 Standardabweichungen vom Ursprung abweichen (außerhalb des Kreises), tragen signifikant zum Modell bei. Folgende Typen können dabei unterschieden werden: allgemeine Interaktionen (Typ i), die in allen WW-Domänen/Liganden-Komplexen zu finden sind, und charakteristische Wechselwirkungen, die entweder in der  $R_a$ - (Typ ii), der  $R_b$ -Spezifitätsgruppe (Typ iii) oder in beiden R-Spezifitätsgruppen (Typ iv) zu finden sind. Elektrostatische Interaktionen sind durch ● und van-der-Waals Interaktionen durch ▲ gekennzeichnet. Charakteristische Interaktionen sind mit der interagierenden Liganden- und Domänenposition ( $\lambda::\omega$ ) sowie den charakteristischen Aminosäuren in diesen Positionen gekennzeichnet (hk = Hauptkette, sk = Seitenkette).

**Typ iii** Charakteristisch für die  **$R_b$ -Spezifitätsgruppe** ist vor allem die elektrostatische Wechselwirkung zwischen dem Arginin in  $\lambda-1$  und dem konservierten, negativ geladenen Rest D/E $\omega$ 24. Ebenfalls charakteristisch ist die van-der-Waals-Wechselwirkung zwischen  $\lambda$ 3 und  $\omega$ 35. Durch ein Glyzin in dieser Position werden möglicherweise nachteilige Kontakte mit dem Liganden (vor allem  $\lambda$ 3) vermieden.

Auf Grund der geringen Anzahl an WW-Domänen in den beiden R-Spezifitätsgruppen und den nur semiquantitativen Bindungsdaten konnte kein vorhersagefähiges COMBINE-Modell mittels PLS-Regression berechnet werden. Dennoch zeigen die Ergebnisse der PCA, dass die COMBINE-Methode zusätzliche Einblicke in den vorgeschlagenen Bindungsmechanismus ermöglicht. Die Ergebnisse unterstützen damit nicht nur die vorgeschlagenen Komplexmodelle, sondern sie veranschaulichen auch den Mechanismus der spezifischen Bindung argininhaltiger Liganden.

### 3.3.3.2 COMBINE-Modell der WW-Domänen/Y-Liganden-Wechselwirkung

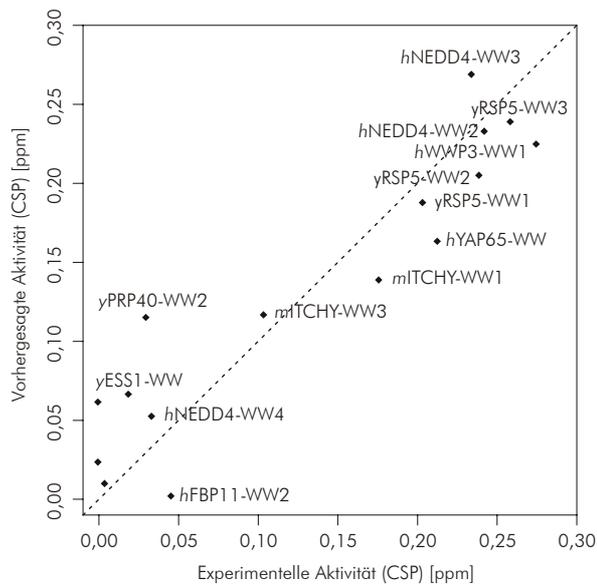
Für die Erkennung tyrosinhaltiger Liganden durch WW-Domänen der Y-Spezifitätsgruppe konnte mit Hilfe der COMBINE-Methode ein vorhersagefähiges Modell erstellt werden. Dies wurde vor allem durch zwei Faktoren ermöglicht: zum einen waren für alle 42 WW-Domänen quantitative Aktivitätsdaten (CSPs – siehe Tabelle 3-6 und Anhang C) vorhanden. Zum anderen waren genügend Vertreter der Y-Spezifitätsgruppe für ein stabiles statistisches Modell experimentell untersucht worden.

Kooperationsprojekt mit Dr. Karin Schleinkofer, Molecular and Cellular Modeling Group, EML Research GmbH, Heidelberg.

Die Strukturmodelle von insgesamt 16 WW-Domänen/Y-Liganden-Komplexe wurden mit der COMBINE-Analyse untersucht (Tabelle 3-6, Spalte COMBINE/PLS). Darunter waren alle 10 WW-Domänen der Y-Spezifitätsgruppe mit quantifizierbaren CSPs vertreten. Zusätzlich wurden aus jeder der restlichen 5 Spezifitätsgruppen sowie aus der Gruppe der nicht klassifizierten WW-Domänen jeweils 1 Vertreter mit quantifizierbaren CSPs als repräsentative „schwach bindende“ bzw. „nicht bindende“ Kombination ausgewählt.

Die 16 Strukturmodelle wurden mit Hilfe der COMBINE-Methode durch die paarweisen Interaktions-Energietermine zwischen allen Aminosäuren der Domäne und des Liganden beschrieben. Zusätzlich zu diesen intermolekularen wurden auch die intramolekularen Interaktions-Energietermen der WW-Domänen in das COMBINE-Modell integriert. Damit sollte ein möglicher Einfluss der Faltungstabilität auf die Aktivität repräsentiert werden. Darüber hinaus wurden Desolvatisierungs-Effekte in Form von je einem elektrostatischen Desolvatisierungs-Energieterm für Domäne bzw. Ligand modelliert. Diese multivariate Beschreibung der Strukturmodelle wurde durch PLS-Regression in Form eines linearen statistischen Modells (COMBINE-Modell) mit der Aktivität der Domänen (CSP infolge der Bindung des Y-Modellliganden, siehe Tabelle 3-6, Seite 63; siehe auch Anhang C) in Beziehung gesetzt.

Das finale COMBINE-Modell basierte auf 2 Komponenten (latenten Variablen) und zeigte eine gute Anpassung an die Trainingsdaten ( $R^2 = 0,97$ ). Auch die durch *leave-one-out* Kreuzvalidierung bestimmte Vorhersageleistung ( $Q^2 = 0,83$  und  $SDEP = 0,04$ ) war ausgezeichnet. Daneben wurde die Robustheit des Modells durch eine Kreuzvalidierung unter der gleichzeitigen Auslassung von 4 Komplexen untersucht. Die dadurch ermittelte mittlere  $SDEP$  von  $0,07 (\pm 0,02)$  demonstrierte, dass das Modell robust gegenüber dem Einfluss einzelner Komplexe ist. In Abbildung 3-26 sind die vorhergesagten CSPs des *leave-one-out* kreuvalidierten Modells gegenüber den experimentell bestimmten CSPs aufgetragen.



**Abbildung 3-26 Vorhersageleistung des COMBINE-Modells**

Die *leave-one-out* kreuzvalidierten Vorhersagen der Veränderung der chemischen Verschiebung (CSP) durch das COMBINE-Modell der Tyrosin-Spezifität sind gegen die experimentell bestimmten CSPs (siehe Tabelle 3-6) aufgetragen. Der kreuzvalidierte Determinationskoeffizient  $Q^2$  beträgt 0,83.

Der Einfluss der einzelnen paarweisen intermolekularen Interaktionen auf die Vorhersage der CSPs sowie exemplarisch die Interaktionsenergien des *hWWP3-WW1/Y*-Liganden-Komplexes sind in Abbildung 3-27 dargestellt. Negative PLS-Koeffizienten kennzeichnen vorteilhafte Beiträge (erhöhte CSP), für den Fall, dass die entsprechende Interaktion eine negative Interaktionsenergie beisteuert. Positive PLS-Koeffizienten zeigen dagegen einen negativen Einfluss der Interaktion auf die Aktivität an.

Die PLS-Koeffizienten (Abbildung 3-27A-B) und die durch die PLS-Koeffizienten gewichteten Beiträge zur Vorhersage des COMBINE-Modells (Abbildung 3-27E-F) zeigen, dass der Großteil dieser Wechselwirkungen einen positiven Effekt auf die vorhergesagte Affinität hat. Dagegen hat vor allem der Kontakt zwischen  $\lambda 3$  und  $\omega 30$  einen negativen Effekt auf die Aktivität. Überraschenderweise zeigt das COMBINE-Modell auch für die elektrostatische Interaktion zwischen  $\lambda 1$  und  $\omega 37$ , die der konservierten Wasserstoffbrücke entspricht, einen negativen Effekt. Dies deutet daraufhin, dass die exakte Positionierung der Hauptketten-Carbonylgruppe des Liganden gegenüber der Hydroxylgruppe von *S/T* $\omega 37$  einen Einfluss auf die Bindungsaffinität hat. Tabelle 3-11 fasst die Interaktionen mit signifikantem Einfluss auf die Vorhersage des COMBINE-Modells zusammen.

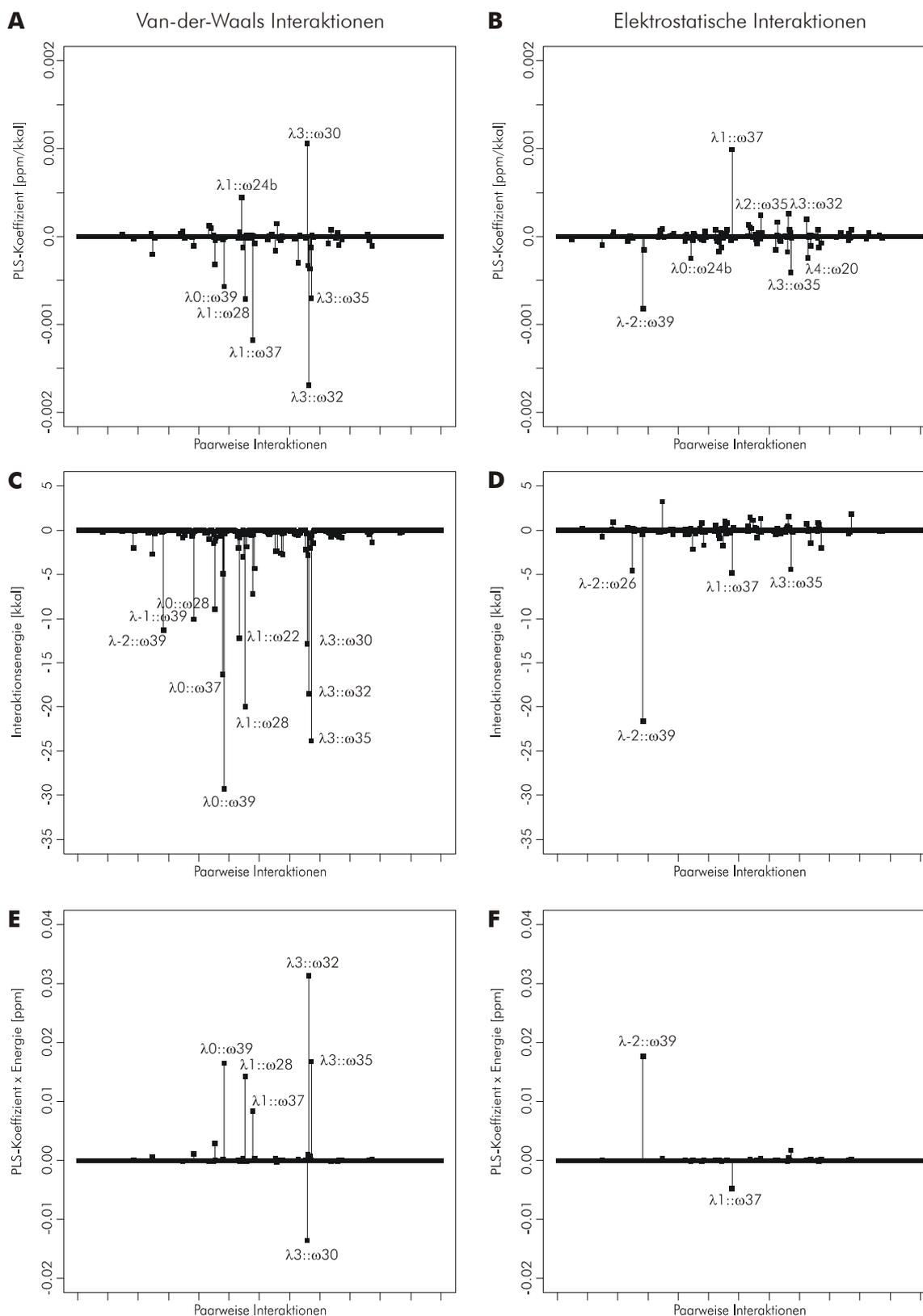


Abbildung 3-27 COMBINE-Modell der Tyrosin-Spezifität: PLS-Koeffizienten und Energiebeiträge im hWWP3-WW1/Y-Liganden-Komplex

Dargestellt sind die PLS-Koeffizienten der (A) van-der-Waals und (B) elektrostatischen Interaktions-Energieerme des COMBINE-Modells der Tyrosin-Spezifität. Exemplarisch für den hWWP3-WW1/Y-Liganden-Komplex (der Komplex mit der größten CSP im NMR-Screening) sind die Energiebeiträge der (C) van-der-Waals und (D) elektrostatischen paarweisen Interaktionen gezeigt. In (E) und (F) werden die van-der-Waals und elektrostatischen Energiebeiträge des hWWP3-WW1/Y-Liganden-Komplexes für die Vorhersage der CSP durch die PLS-Koeffizienten gewichtet. Die Summe dieser gewichteten Beiträge ergibt die COMBINE-Vorhersage der CSP für den entsprechenden Komplex. In allen 6 Teilabbildungen sind diejenigen Werte, die um mehr als 3 Standardabweichungen von 0 abweichen, explizit benannt.

**Tabelle 3-11 Einfluss charakteristischer Interaktionen in WW-Domänen/Y-Liganden-Komplexen auf die COMBINE-Vorhersage**

Ligandenrest	Domänenrest	Beschreibung der Interaktion
van-der-Waals Interaktionen		
$\lambda 0$	$\omega 39, \omega 37, \omega 28$	Positionierung von $P\lambda 0$ auf der „ $\omega 39$ -Seite“ der hydrophoben Bindungsstelle
$\lambda 1$	$\omega 22, \omega 28, \omega 37$	Positionierung von $P\lambda 1$ auf der „ $\omega 28$ -Seite“ der hydrophoben Bindungsstelle
$\lambda 3$	$\omega 30, \omega 32, \omega 35$	Bindung von Y in $\lambda 3$ durch die konservierte Tyrosin-Bindungsstelle
$\lambda -1, \lambda -2$	$\omega 39$	„Umklammerung“ von $W\omega 39$
Elektrostatistische Interaktionen		
$\lambda -2$	$\omega 39$	Konservierte Wasserstoffbrücke
$\lambda 1$	$\omega 37$	Konservierte Wasserstoffbrücke
$\lambda 3$	$\omega 35$	Wechselwirkung der Hydroxylgruppe des $Y\lambda 3$ mit $R/K/Q\omega 35$

Für eine hohe Aktivität der WW-Domänen der Y-Spezifitätsgruppe gegenüber tyrosinhaltigen Liganden sind vor allem die vorteilhaften Kontakte der Proline des Liganden mit  $W\omega 39$ ,  $Y\omega 28$  und  $T\omega 37$ , sowie des Tyrosins mit  $H\omega 32$  und  $R/K/Q\omega 35$  verantwortlich. In  $\omega 30$  sind für die Bindung des Y-Liganden kleine aliphatische Aminosäuren notwendig. Aromatische Aminosäuren, wie sie bei WW-Domänen der anderen Spezifitätsgruppen häufig in dieser Position auftreten, würden die spezifische Tyrosin-Bindungsstelle versperren. Zusätzlich bilden V/I/L an  $\omega 30$  eine für die Interaktion mit Tyrosin vorteilhafte hydrophobe Fläche am „Boden“ der Tyrosin-Bindungsstelle aus.

Damit konnten durch das COMBINE-Modell nicht nur die – durch experimentelle Strukturen nachvollziehbaren – charakteristischen Interaktionen identifiziert werden, sondern auch ihr Beitrag zur Bindung tyrosinhaltiger Liganden quantifiziert werden. Dies erlaubte Unterschiede in der Affinität verschiedener Liganden zu rationalisieren und vorherzusagen.

### 3.3.3.3 Design peptidischer Liganden mit gesteigerter Affinität

Mit Hilfe des in Abschnitt 3.3.3.2 entwickelten COMBINE-Modells der Tyrosin-Spezifität wurde ein Peptid mit gesteigerter Affinität designt. Da das Kernmotiv **xPPxY** schon optimal für die Wechselwirkung mit WW-Domänen der Y-Spezifitätsgruppe gestaltet war, wurde nach affinitätssteigernden Modifikationen in den flankierenden Sequenzen gesucht.

Die Analyse der Wechselwirkungen am Beispiel von *hWWP3-WW1* zeigte eine vorteilhafte elektrostatische Interaktion zwischen  $\lambda -2$  und  $\omega 26$ . Diese bislang wenig beachtete Wechselwirkung spielte auch schon bei den WW-Domänen der R-Spezifitätsgruppen eine essentielle Rolle für deren Spezifität (3.3.3.1). Parallel dazu hatte die Sequenzanalyse eine hohe Konservierung von R/Q an dieser Position in WW-Domänen der Y-Spezifitätsgruppe ergeben. Basierend auf diesen Erkenntnissen wurden verschiedene Mutationen in Position

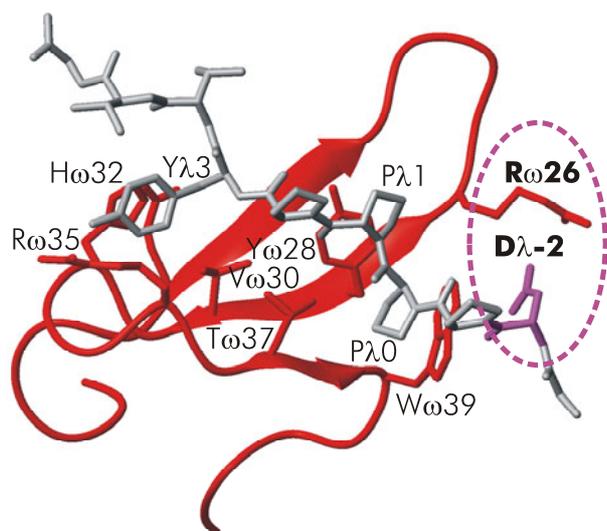


Abbildung 3-28 Design eines Liganden mit gesteigerter Affinität:  $\gamma$ RSP5-WW1/GDPPPPYTVG

Strukturmodell der  $\gamma$ RSP5-WW1 Domäne (rot) im Komplex mit dem designierten Peptid **GDPPPPYTVG** (grau). In Ligandenposition  $\lambda-2$  (violett eingefärbt und umrandet) wurde anstelle des Threonins im Y-Liganden ein Aspartat eingeführt. Die elektrostatische Wechselwirkung des  $D\lambda-2$  mit  $R\omega 26$  steigert die Affinität dieses Komplexes. Die für die Ligandenbindung relevanten Aminosäuren der Domäne und des Liganden sind explizit benannt.

$\lambda-2$  mit dem COMBINE-Modell getestet. Dabei sollte ein negativ geladener Rest (D/E) in  $\lambda-2$  (Abbildung 3-28) durch die elektrostatische Wechselwirkung mit  $R\omega 26$  einen affinitätssteigernden, ein positiv geladener Rest (R/K) im Liganden dagegen einen affinitätssenkenden Effekt haben. Die durch das COMBINE-Modell für verschiedene Liganden im Komplex mit der  $\gamma$ RSP5-WW1 Domäne vorhergesagten CSPs sind in Tabelle 3-12 zusammengefasst.

Zur Verifikation der Vorhersage wurden die Dissoziationskonstanten für die Komplexe von  $\gamma$ RSP5-WW1 mit dem Referenzpeptid **GTPPPPYTVG** bzw. mit der  $T\lambda-2D$ -Mutante **GDPPPPYTVG** durch UV-Fluoreszenzspektroskopie bestimmt. In der Tat bestätigten die experimentellen Messungen die vorhergesagte signifikante Steigerung der Affinität.

Tabelle 3-12 Vergleich der durch das COMBINE-Modell vorhergesagten und der experimentell bestimmten Affinität der designierten Liganden

Die experimentell bestimmten Dissoziationskonstanten bestätigen die vorhergesagte gesteigerte Affinität des Liganden mit einem Aspartat an  $\lambda-2$  im Vergleich zu dem Y-Liganden mit einem Threonin an dieser Position.

Designierte Liganden	COMBINE-Vorhersage		UV-Fluor.-Spektroskopie <sup>c)</sup>	
	CSP <sup>a)</sup> [ppm]	Faktor <sup>b)</sup>	$K_d$ [ $\mu$ M]	Faktor <sup>b)</sup>
<i>ac</i> -G <b>T</b> $\lambda-2$ PPPPYTVG-NH <sub>2</sub>	0,20	1,0	55 (43 - 72)	1,0
<i>ac</i> -G <b>D</b> $\lambda-2$ PPPPYTVG-NH <sub>2</sub>	0,54	2,7	19 (10 - 34)	2,9
<i>ac</i> -G <b>E</b> $\lambda-2$ PPPPYTVG-NH <sub>2</sub>	0,50	2,5	n.b.	
<i>ac</i> -G <b>K</b> $\lambda-2$ PPPPYTVG-NH <sub>2</sub>	0,06	0,3	n.b.	
<i>ac</i> -G <b>R</b> $\lambda-2$ PPPPYTVG-NH <sub>2</sub>	*-0,04	0,0	n.b.	

\* Vorhersagen negativer CSPs sind als Nicht-Bindung zu deuten.

<sup>a)</sup> Durch das COMBINE-Modell vorhergesagte CSPs.

<sup>b)</sup> Multiplikative Veränderung im Vergleich zum Y-Liganden (**GTPPPPYTVG**).

<sup>c)</sup> Experimentell bestimmte  $K_d$  durch UV-Fluoreszenz-Spektroskopie. Angegeben sind die Mittelwerte von drei unabhängigen Messungen. In Klammern ist das  $\pm 1\sigma$  Intervall der Messwiederholungen auf logarithmischer Skala angegeben (n.b. = nicht bestimmt).

Diese Analyse zeigte damit, dass das COMBINE-Modell der Tyrosin-Spezifität – trotz des Trainings durch modellierte Komplexstrukturen und der vagen Beziehung zwischen der gemessenen Aktivität (CSP) und der Affinität der Wechselwirkung – ein mächtiges Werkzeug für die Vorhersage neuer Liganden mit definierter Affinität ist.

Insgesamt haben die CoMFA- und COMBINE-Modelle auf der Ebene der 3D-Struktur nicht nur die strukturellen Merkmale und charakteristischen Interaktionen für die spezifische Ligandenbindung identifiziert, sondern sie haben auch die quantitative Vorhersage der Affinität ermöglicht. Auf der Ebene der Sequenz konnte allerdings mit bekannten Methoden keine quantitative Sequenz-Aktivitäts-Beziehung hergestellt werden. Aus diesem Grund wurde die Methode der *Quantitativen Spezifitäts-Profile* entwickelt (siehe 3.5).

### 3.4 Analyse von experimentellen Wechselwirkungsdaten für PDZ-Domänen

Neben WW-Domänen wurden als repräsentative Vertreter der PDZ-Domänen auch hAF6-PDZ, hERBIN-PDZ und mSNA1-PDZ untersucht. Mit Hilfe dieser Untersuchungen sollte zum einen die Spezifität der PDZ-Domänen besser verstanden und zum anderen Wechselwirkungspartner auf der Basis von quantitativen Sequenz-Aktivitäts-Beziehungen vorhergesagt werden.

PDZ-Domänen interagieren meist mit den letzten 4–7 C-terminalen Resten ihrer Wechselwirkungspartner (siehe 1.4.1); seltener treten andere Wechselwirkungsmechanismen auf.<sup>128</sup> Um einen Überblick über das Ligandenspektrum der drei PDZ-Domänen zu gewinnen, wurden diese mit Hilfe der *Bibliothek humaner C-Termini* auf Wechselwirkungen untersucht. Diese Peptidbibliothek umfasste die 6223 nicht-identischen 11meren C-Termini aller im Release 40 (Oktober, 2001) der Swiss-Prot-Datenbank annotierten humanen Proteine. Dabei wurde die Menge an gebundener PDZ-Domäne – als Maß für die Stärke der Wechselwirkung – über die Signalintensität einer Chemilumineszenz-Reaktion in Form von *Boehringer-Light-Units* (BLU) quantifiziert (siehe 2.3.2.4). In Tabelle 3-13 ist die Verteilung der gemessenen Signalintensitäten zusammengefasst.

Die hohe Anzahl ( $\geq 91$ ) an C-Termini mit MAD-Z-Werten  $\geq 5$  demonstrierte für alle drei Domänen signifikante Wechselwirkungen mit einer Vielzahl von Peptiden (Tabelle 3-13, grau unterlegte Spalte), wenn auch die Anzahl für hERBIN-PDZ deutlich geringer war. Der

Kooperationsprojekt mit Dr. Prisca Boisguerin, Molecular Libraries and Recognition Group, Institut für Medizinische Immunologie, Universitätsklinikum Charité.

mSNA1-PDZ unterscheidet sich von hSNA1-PDZ nur in Position  $\alpha$ B:7. Dort besitzt hSNA1-PDZ ein Valin anstatt eines Alanin.

Die Menge an gebundener Domäne steht über das Massenwirkungsgesetz in direkter Beziehung zur Affinität der Wechselwirkung. Die Menge an gebundener Domäne wird über einen Antikörper mit gekoppelter Peroxidase, welche eine Chemilumineszenz-Reaktion katalysiert, nachgewiesen. Die gemessenen Signalintensitäten  $I$  (in BLU) dieser Reaktion sind annähernd lognormalverteilt, weshalb die logarithmus-transformierten Signalintensitäten ( $\ln(I)$ ) analysiert wurden.

**Tabelle 3-13** Statistik des Screenings der *Bibliothek humaner C-Termini*

Die Verteilung der gemessenen Signalintensitäten ist auf der linearen und logarithmischen Skala durch Kennwerte zusammengefasst (Max, 100ste, Min). Darüber hinaus ist der prozentuale Anteil der Variation der 100 höchsten Signalintensitäten ( $\Delta 100$ ) an der Gesamtvariation und die Anzahl der signifikant vom durchschnitt abweichenden C-Termini ( $\text{MAD-Z} \geq 5$ , grau unterlegt) angegeben.

Domäne	Signalintensität $I$ [BLU] <sup>a)</sup>			$\ln(I)$ <sup>b)</sup>					MAD-Z <sup>c)</sup>		
	Max	100ste	Min	Max	100ste	Min	Med	$\Delta 100$	Max	100ste	$\geq 5^*$
hAF6-PDZ	5,50E+06	7,78E+05	3,58E+04	6,74	5,89	4,55	4,81	38,9%	17,00	9,52	421
hERBIN-PDZ	6,76E+06	1,28E+06	3,09E+04	6,83	6,11	4,49	4,87	30,8%	7,71	4,87	91
mSNA1-PDZ	5,19E+05	1,67E+05	1,32E+04	5,72	5,22	4,12	4,30	30,9%	16,41	10,69	578

\* Anzahl der C-Termini mit einem MAD-Z-Wert  $\geq 5$  (siehe 2.5.1).

<sup>a)</sup> Rohwerte der gemessenen Signalintensitäten in BLU. Maximaler (Max), minimaler (Min) und der hundert-höchste (100ste) Wert.

<sup>b)</sup> Logarithmus-transformierte Signalintensitäten. Maximaler (Max), minimaler (Min), der hundert-höchste (100ste) Wert und Median (Med).  $\Delta 100$  = prozentualer Anteil der Variation der 100 höchsten  $\ln(I)$ -Werte an der Gesamtvariation ( $\Delta 100 = (\ln(I_{\text{Max}}) - \ln(I_{100\text{ste}})) / ((\ln(I_{\text{Max}}) - \ln(I_{\text{Min}})))$ ).

<sup>c)</sup> MAD-Z-Werte sind robuste und standardisierte Beschreibungen der Lage eines  $\ln(I)$ -Werts im Verhältnis zum Median als robuster Schätzwert des Erwartungswerts. Die Abweichung wird in Einheiten des Medians der absoluten Abweichungen als robuster Streuungsschätzwert angegeben (siehe 2.5.1).

Vergleich mit bekannten Wechselwirkungspartnern zeigte (Daten nicht gezeigt), dass diese meist unter den C-Termini mit den 100 höchsten Signalintensitäten (Top100 Peptide) zu finden waren. Daher konnte für diese Auswahl an C-Termini von Affinitäten im mikromolaren Bereich ausgegangen werden.

In allen drei Fällen war die Menge der Top100 Peptide für etwa 30% der Gesamtvariation der Logarithmus-transformierten Signalintensitäten  $\ln(I)$  verantwortlich (Tabelle 3-13, Spalte  $\Delta 100$ ). Dabei wurden die Top100 Peptide keineswegs exklusiv von nur einer PDZ-Domäne gebunden: 9% der Top100 Peptide wurden gleichzeitig von allen drei PDZ-Domänen erkannt. Weitere 6% der Top100 Peptide wurden von *hAF6*-PDZ und *hERBIN*-PDZ, 9% von *hAF6*-PDZ und *mSNA1*-PDZ und 20% von *hERBIN*-PDZ und *mSNA1*-PDZ gebunden.

Um Gemeinsamkeiten der Sequenzen der Top100 Peptide zu identifizieren, wurden die präferierten Aminosäuren an den einzelnen Ligandenpositionen in Form von Sequenzlogos visualisiert (Abbildung 3-29A–C, Oben). Alle drei PDZ-Domänen zeigten dabei fast ausschließlich Aminosäurepräferenzen an den letzten 4 C-terminalen Ligandenpositionen ( $\lambda-3$  bis  $\lambda 0$ ). Nur für *mSNA1*-PDZ spielte zusätzlich  $\lambda-4$  eine Rolle. Daneben zeigten vor allem *hERBIN*-PDZ und *mSNA1*-PDZ eine im Vergleich zu  $\lambda 0$ ,  $\lambda-2$  und  $\lambda-3$  weniger stark ausgeprägte Aminosäurepräferenz an  $\lambda-1$ .

Erstaunlicherweise wurde für alle drei Domänen unter den Top100 C-Termini eine signifikante Anzahl an Peptiden identifiziert (Tabelle 3-14), in denen das eigentlich C-terminale Erkennungsmotiv der jeweiligen Domäne (z.B. **E (T/S) x (V/L)**<sub>COOH</sub> für *mERBIN*-PDZ, siehe Abbildung 3-29, Mitte) genau um eine Aminosäure vom C-Terminus verschoben vorlag (z.B. **E (T/S) x (V/L) x**<sub>COOH</sub> für *mERBIN*-PDZ, siehe Abbildung 3-29, Unten).

Auf Grund der hohen Konservierung der Positionierung der Liganden an der Bindungsstelle in PDZ-Domänen/ Liganden-Komplexen können sogenannte virtuelle Ligandenpositionen  $\lambda$  definiert werden (siehe 1.4.1).  $\lambda 0$  bezeichnet dabei den C-terminalen Ligandenrest. Die N-terminal gelegenen Reste werden entsprechend  $\lambda-1$ ,  $\lambda-2$  usw. durchnummeriert.

**Tabelle 3-14** Ligandenmotive in den Top100 C-Termini

Ligandentyp	<i>hAF6</i> -PDZ	<i>hERBIN</i> -PDZ	<i>mSNA1</i> -PDZ
C-terminal <sup>a)</sup>	80%	75%	67%
Verschoben um 1 AS <sup>b)</sup>	15%	16%	21%
Verschoben um >1 AS <sup>b)</sup>	1%	3%	2%
Unbestimmt <sup>c)</sup>	4%	6%	10%

<sup>a)</sup> C-terminales PDZ-Domänen-Erkennungsmotiv: hohe Sequenzkonservierung an  $\lambda-2$  und  $\lambda 0$ , mit meist aliphatischen Aminosäuren an  $\lambda 0$ .

<sup>b)</sup> Das eigentlich C-terminale Erkennungsmotiv liegt um eine bzw. mehrere Aminosäuren (AS) vom C-Terminus verschoben vor.

<sup>c)</sup> Keine Übereinstimmung mit dem C-terminalen Erkennungsmotiv.

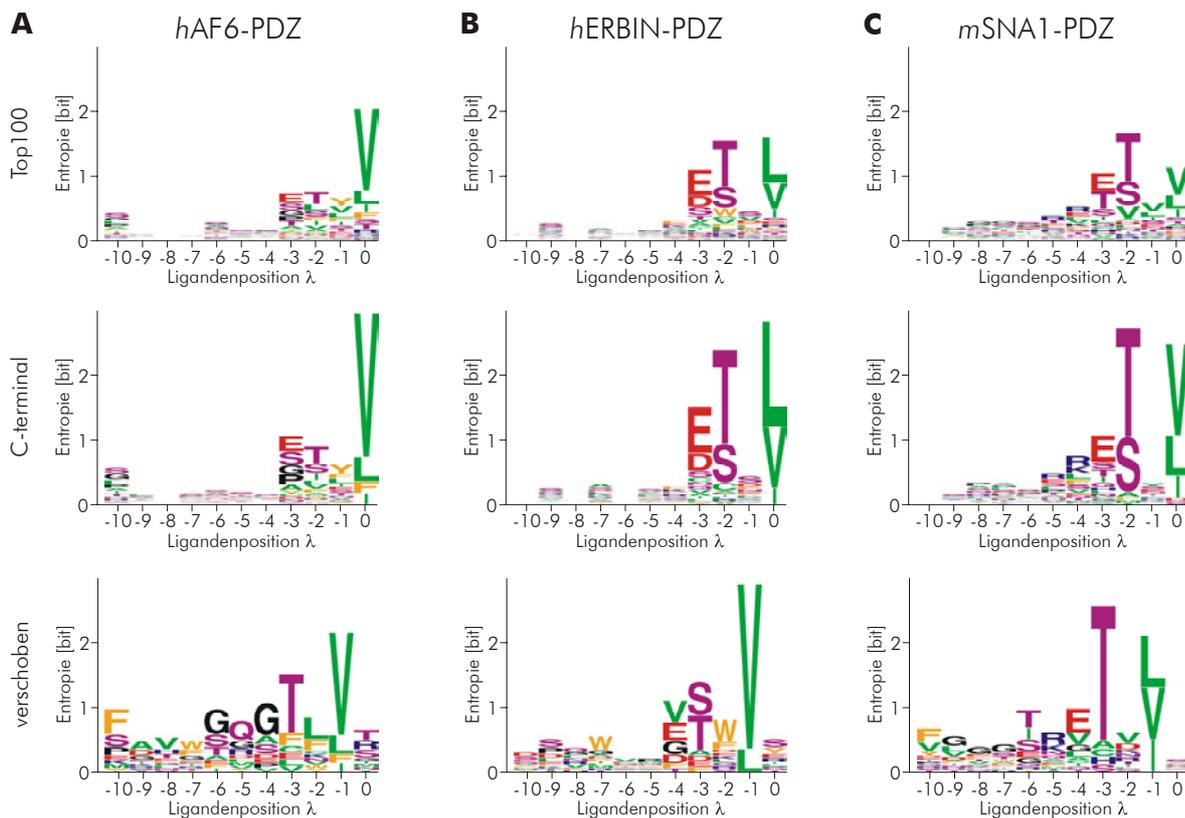


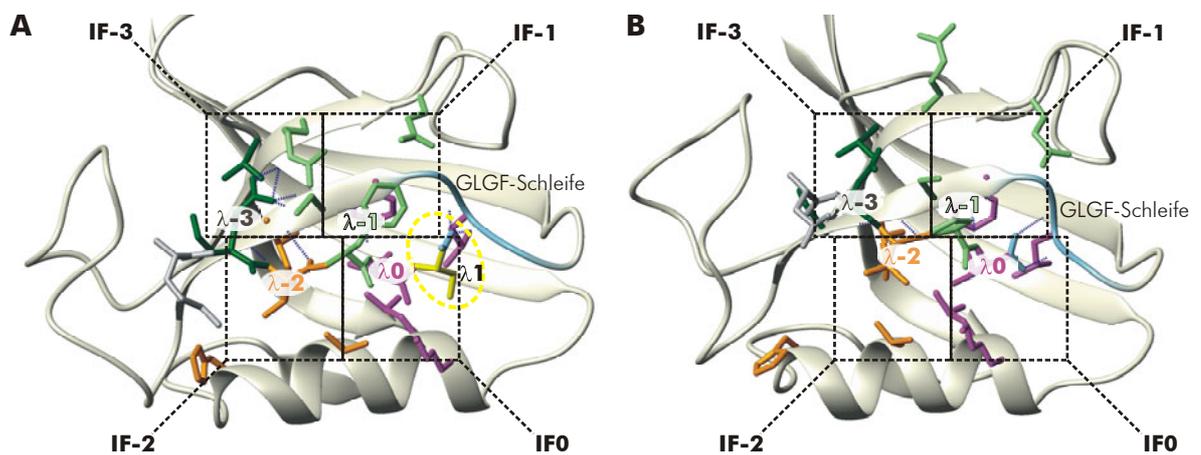
Abbildung 3-29 Sequenzlogos der Top100 C-Termini aus dem Screening der Bibliothek humaner C-Termini

Die oberste Zeile zeigt die Sequenzlogos der gesamten Top100 Peptide jeweils für die (A) hAF6-PDZ, (B) hERBIN-PDZ und (C) mSNA1-PDZ Domäne. Da in einer signifikanten Anzahl an Peptiden das C-terminale Motiv um eine Aminosäure vom C-Terminus verschoben vorliegt, zeigen die mittlere und untere Zeilen die Sequenzlogos getrennt für Peptide mit C-terminalen bzw. verschobenem Motiv. In diesen Sequenzlogos wird jede Ligandenposition  $\lambda$  durch einen Stapel derjenigen Aminosäuren repräsentiert, die an der jeweiligen Sequenzposition vorkommen. Die Höhe des Stapels symbolisiert die Konservierung an dieser Position (gemessen als Entropie in bit), während die Höhe der einzelnen Aminosäuren ihre relativen Häufigkeiten versinnbildlicht. Geht die Entropie gegen 0, lässt sich der Stapel nicht mehr auflösen und verschwindet daher an diesen Positionen. Für alle Sequenzlogos wurde eine Korrektur für kleine Stichproben vorgenommen. Für diese Sequenzlogos wurden die Sequenzen am C-Terminus aligniert. Die Sequenzlogos<sup>158</sup> wurden mit Weblogo<sup>43</sup> erstellt.

Daraufhin wurde eine Moleküldynamik Simulation des Strukturmodells der hERBIN-PDZ Domäne im Komplex mit einem der identifizierten Liganden mit verschobenem Erkennungsmotiv **METFV**<sub>S<sub>COOH</sub></sub> (AG22) durchgeführt. Die Simulation bestätigte, dass das verschobene Erkennungsmotiv (**ETFV**) prinzipiell in der für PDZ-Domänen bekannten Weise gebunden werden kann, auch wenn sich zwischen der C-terminalen Carboxylgruppe und diesem Motiv noch eine zusätzliche Aminosäure befindet (Abbildung 3-30). Damit wurden die virtuellen Ligandenpositionen  $\lambda_0$  bis  $\lambda_{-3}$  auch in diesem Komplex von den Aminosäuren **ETFV** besetzt. Trotz der „eingeschobenen“ Aminosäure an Position  $\lambda_1$  bildete die C-terminale Carboxylgruppe weiterhin Wasserstoffbrücken zur PDZ-Domäne aus. Die Seitenkette der eingeschobenen Aminosäure fand zwischen der GLGF-Schleife und dem C-terminalen Ende der  $\alpha_B$ -Helix Platz.

**Alternativer Bindungsmodus von PDZ-Domänen: Bindung verschobener Motive**

AG22: Angiotensin II  
Typ 2 Rezeptor  
[UP: P50052]



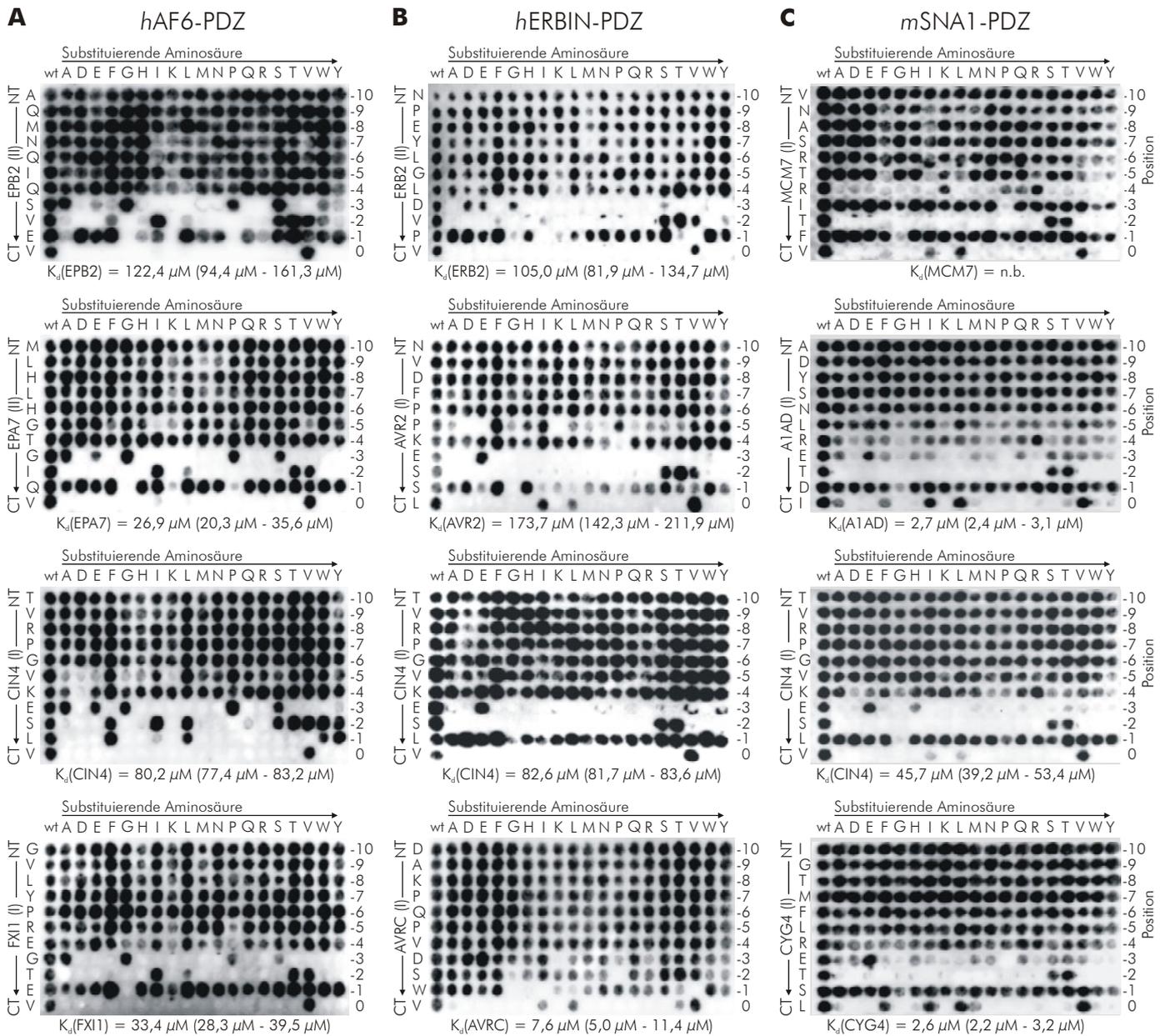
**Abbildung 3-30** Strukturmodell der Bindung verschobener Erkennungsmotive durch PDZ-Domänen im Vergleich mit einer experimentell bestimmten Komplexstruktur derselben PDZ-Domäne

**(A)** Strukturmodell der *hERBIN*-PDZ Domäne im Komplex mit dem Liganden **METFVS**<sub>COOH</sub> aus AG22. Die „eingeschobene“ Aminosäure in  $\lambda 1$ , die das eigentlich C-terminale Erkennungsmotiv von der C-terminalen Carboxylgruppe trennt, ist gelb eingefärbt und umrandet. Die Aminosäuren des ursprünglich C-terminalen Erkennungsmotivs besetzen weiterhin die virtuellen Ligandenpositionen  $\lambda 0$  (violett),  $\lambda -1$  (hellgrün),  $\lambda -2$  (orange) und  $\lambda -3$  (dunkelgrün). Die Seitenketten dieser Ligandenreste kontaktieren die konservierten Interaktionsflächen IF0 bis IF-3 (siehe 1.4.1). Die C-terminale Carboxylgruppe des Liganden (hellblau) bildet weiterhin Wasserstoffbrücken (dunkelblau) zur GLGF-Schleife aus.

**(B)** Zum Vergleich ist die experimentell durch Röntgenkristallanalyse bestimmte Struktur der *hERBIN*-PDZ Domäne im Komplex mit dem Liganden **LDVPV**<sub>COOH</sub> [PDB: 1MFG]<sup>23</sup> dargestellt. Die Liganden- und Domänenreste sind wie in **A** eingefärbt. Ebenso sind die Interaktionsflächen angegeben.

Durch Substitutionsanalysen konnte anschließend experimentell gezeigt werden, dass das eigentlich C-terminale Erkennungsmotiv nur durch genau eine Aminosäure von der C-terminalen Carboxylgruppe „getrennt“ werden darf.<sup>27</sup> Diese bleibt jedoch essentiell für die Bindung, da eine Amidierung der C-terminalen Carboxylgruppe in Liganden mit verschobenem Erkennungsmotiv die Wechselwirkung unterbindet.<sup>27</sup> Damit wurde eine weitere Klasse von Liganden für die PDZ-Domänen identifiziert und ein neuer Bindungsmodus vorgeschlagen.

Neben der Entdeckung des alternativen Bindungsmodus konnte ein Bild der Aminosäurepräferenzen an den einzelnen Ligandenpositionen durch die Analyse der Top100 Peptide gewonnen werden (Abbildung 3-29). Da die bevorzugten Aminosäuren in den unterschiedlichsten Kombinationen unter den Top100 Peptiden auftraten, wurden Substitutionsanalysen von ausgewählten Liganden durchgeführt (23 Peptide für *hAF6*-PDZ, 24 für *hERBIN*-PDZ und 4 für *mSNA1*-PDZ). Pro PDZ-Domäne sind je 4 repräsentative Substitutionsanalysen in Abbildung 3-31 dargestellt. Dadurch sollten die Erkennungsmotive systematisch analysiert und eventuelle Abhängigkeiten zwischen Ligandenpositionen – insbesondere für Liganden der Klasse I bzw. II<sup>170</sup> – bestimmt werden.



**Abbildung 3-31** Substitutionsanalysen natürlicher Liganden für die *hAF6-PDZ*, *hERBIN-PDZ* und *mSNA1-PDZ* Domänen

Die Substitutionsanalysen von je 4 durch die Screenings der *Bibliothek humaner C-Termini* identifizierte Liganden sind für **(A)** *hAF6-PDZ*, **(B)** *hERBIN-PDZ* und **(C)** *mSNA1-PDZ* dargestellt. Für jede Domäne demonstrieren die Substitutionsanalysen ein konserviertes und von der Sequenz des Ausgangs-peptid relativ unabhängiges Erkennungsmotiv. Eine Substitutionsanalyse ist eine „tabellenartige“ Peptidbibliothek. Jede „Tabellenzelle“ entspricht einem definierten Einzelsubstitutionsanalogon des Ausgangs-peptids (wt = wildtyp). Die Zeile bestimmt dabei, welche Position des Peptids substituiert wird, und die Spalte, welcher Aminosäuretyp damit an der jeweiligen Position zuzufinden ist. Die Aminosäuresequenz des Ausgangs-peptids ist auf der linken Seite den Zeilen entsprechend vom N-Terminus (oben) zum C-Terminus (unten) dargestellt. Die Ligandenpositionen  $\lambda$  (rechts) sind ausgehend von der C-terminalen Position  $\lambda_0$  rückwärtszählend durchnummeriert (siehe 2.1.2.2). Der Aminosäuretyp (alle proteinogenen Aminosäuren außer Cystein) durch den die jeweilige Position substituiert wird, ist oberhalb der entsprechenden Spalten angegeben. In allen Zellen der ersten Spalte (wt) befindet sich das unveränderte Ausgangs-peptid. Gebundene PDZ-Domänen werden durch ein Chemilumineszenzsystem nachgewiesen: eine hohe Dichte an gebundenen PDZ-Domänen führt zu einem intensiven Signal, welches durch die Intensität der Schwärzung dargestellt ist. Schlüsselreste sind daran zu erkennen, dass nur bestimmte Aminosäuren in der jeweiligen Position zur spezifischen Bindung durch die Domäne führen (nur wenige Zellen der Zeile sind geschwärzt). Positionen, bei denen alle Aminosäuretypen ein etwa gleich starkes Signal liefern, sind nicht an der spezifischen Wechselwirkung beteiligt. Das spezifische Erkennungsmotiv ergibt sich aus der Kombination aller Schlüsselreste. Für jedes Peptid ist die Klasse nach Songyang *et al.*, 1997, in Klammern neben dem Peptidnamen angegeben. Die experimentell durch das Oberflächen-Plasmon-Resonanz-Verfahren bestimmten Dissoziationskonstanten sind zusammen mit dem  $\pm 1\sigma$  Intervall auf der logarithmischen Skala unter der jeweiligen Substitutionsanalyse angegeben. Abbildung modifiziert nach Wiedemann und Boisguerin *et al.*, 2004.

Tabelle 3-15 PDZ-Domänen Liganden-Erkennungsmotive

Domäne	Liganden-Erkennungsmotiv <sup>a)</sup>				
	$\lambda-4$	$\lambda-3$	$\lambda-2$	$\lambda-1$	$\lambda 0$
<i>h</i> AF6-PDZ		<b>G/A/P/S/E</b>	<b>V/T/S/I/L/F</b>	<b>x</b>	<b>V<sub>COOH</sub></b>
<i>h</i> ERBIN-PDZ		<b>E/D</b>	<b>T/S/V</b>	<b>x</b>	<b>V/L/I<sub>COOH</sub></b>
<i>m</i> SNA1-PDZ	<b>R/K</b>	<b>E</b>	<b>T/S</b>	<b>x</b>	<b>V/I/L<sub>COOH</sub></b>

<sup>a</sup> Für Nomenklatur der Sequenzmotive siehe 2.1.3, für eine Definition der virtuelle Ligandenpositionen  $\lambda$  siehe 1.4.1 und 2.1.2.2.

Die Substitutionsanalysen zeigten – übereinstimmend mit der Analyse der Top100 Peptide – für *h*AF6-PDZ und *h*ERBIN-PDZ nur an den 4 letzten C-terminalen Ligandenpositionen Selektivität für spezifische Aminosäuren. Für *m*SNA1-PDZ bestätigte sich ebenfalls der zusätzliche, wenn auch geringere Beitrag durch  $\lambda-4$ . Darüber hinaus traten jedoch für einzelne PDZ-Domänen/Liganden-Kombinationen auch Selektivitäten an anderen Ligandenpositionen auf. So präferierte beispielsweise die *h*ERBIN-PDZ Domäne in der Substitutionsanalyse des AVR2 Liganden aromatische und aliphatische Aminosäuren an  $\lambda-5$  (Abbildung 3-31B, AVR2).

Insgesamt zeigten alle drei PDZ-Domänen von der Ausgangs-peptidsequenz (z.B. Klasse I bzw. Klasse II) unabhängige, für die Domäne spezifische und konstante Erkennungsmotive (Tabelle 3-15). Diese stimmten auch gut mit den Aminosäurepräferenzen überein, die in den Top100 Peptiden identifiziert wurden (Abbildung 3-29). Daneben konnte vereinzelt auch an  $\lambda-1$  eine Präferenz für bestimmte Aminosäuretypen festgestellt werden: *h*AF6-PDZ tolerierte hier beispielsweise kein Glyzin während *h*ERBIN-PDZ eine schwache Präferenz für aromatische Aminosäuren zeigte. Diese Präferenzen an  $\lambda-1$  erscheinen aber schwach im Verhältnis zu den deutlichen Präferenzen an  $\lambda 0$ ,  $\lambda-2$  und  $\lambda-3$ , weshalb sie in den Erkennungsmotiven (Tabelle 3-15) trotzdem durch **x** repräsentiert werden müssen.

Für ausgewählte Domänen/Liganden-Kombinationen wurde die Affinität der Wechselwirkung durch das Oberflächen-Plasmon-Resonanz-Verfahren (SPR) bestimmt (Tabelle 3-18, Seite 112). Die Dissoziationskonstanten ( $K_d$ ) der gemessenen Wechselwirkungen lagen erwartungsgemäß für alle drei PDZ-Domänen im mikromolaren Bereich.

Die Ergebnisse der experimentellen Charakterisierung der drei PDZ-Domänen mittels Peptidbibliotheken erlaubte folgende Schlussfolgerungen: (i) Die untersuchten PDZ-Domänen interagierten mit einer Vielzahl von Liganden über ein breites Spektrum an Affinitäten. (ii) Das für jede Domäne konstante Erkennungsmotiv deutete darauf hin, dass alle Liganden über einen konservierten Liganden-Bindungsmechanismus erkannt wurden. (iii) Durch

*hAF6*-PDZ und *hERBIN*-PDZ wurden dabei nur die letzten 4 C-terminalen Ligandenpositionen ( $\lambda_0$  bis  $\lambda-3$ ), durch *mSNA1*-PDZ die letzten 5 spezifisch erkannt ( $\lambda_0$  bis  $\lambda-4$ ). (iv) Gleichzeitig schienen die einzelnen Ligandenpositionen relativ unabhängig voneinander zur Bindung beizutragen, da in den Substitutionsanalysen so gut wie keine Abhängigkeiten zwischen dem Aminosäuretyp an einer Ligandenposition und den Präferenzen an einer anderen Positionen beobachtet wurden. (v) Nicht alle Liganden ließen sich durch die identifizierten Erkennungsmotive erklären; viele wichen in mindestens einer Position vom optimalen Erkennungsmotiv ab.

Aus diesem Grund sollten quantitative Sequenz-Aktivitäts-Modelle erstellt werden, welche die Unterschiede in den Ligandenpräferenzen der drei PDZ-Domänen erklären.

### 3.5 Quantitative Sequenz-Aktivitäts-Beziehungen

Die Ergebnisse der experimentellen Charakterisierung der PDZ-Domänen mittels Peptidbibliotheken (siehe 3.4) machte deutlich, dass sich die Spezifität der PDZ-Domänen nur unzureichend durch einfache Sequenzmotive beschreiben lässt. Aus diesem Grund wurde die Methode der *Quantitativen Spezifitäts-Profilen* (QSP) entwickelt. Diese neue Methode ermöglicht es, die Spezifität einer Protein-Interaktionsdomäne als sequenzabhängige Funktion der Affinität für alle potentiell möglichen Ligandensequenzen – den gesamten potentiellen Sequenzraum der Liganden – zu beschreiben (siehe 1.2). Sie basiert im wesentlichen auf dem Design einer sogenannten Profilbibliothek (siehe 3.5.1), die ein effizientes Screening der essentiellen Teile des Liganden-Sequenzraums ermöglicht, sowie einer statistischen multifaktoriellen Varianzanalyse (ANOVA) der Sequenz-Affinitäts-Beziehung (siehe 3.5.2). Die resultierenden QSP-Modelle für die drei PDZ-Domänen wurden anschließend zur Vorhersage des jeweiligen domänenspezifischen *Superbinders* angewandt. Dabei handelt es sich um den optimalen peptidischen Ligand, welcher von der jeweiligen Domänen mit der höchst möglichen Affinität gebunden wird. Die vorhergesagten *Superbinder* wurden anschließend experimentell verifiziert (siehe 3.5.3).

Der potentielle Liganden-Sequenzraum beschreibt formal die Menge aller potentiellen peptidischen Wechselwirkungspartner, die durch eine Protein-Interaktionsdomäne gebunden werden könnten.

Die Dimension des Raums wird durch die Anzahl an variablen Ligandenpositionen  $n$  bestimmt. Die Größe des Sequenzraums berechnet sich folglich aus der Anzahl an Aminosäure-Möglichkeiten in diesen  $n$  Positionen. Werden alle proteinogenen Aminosäuren in Betracht gezogen ergibt sich eine Größe von  $20^n$ .

### 3.5.1 Profilbibliotheken

Für die *hAF6*-PDZ und die *hERBIN*-PDZ Domänen zeigten die Ergebnisse der Screenings mittels Peptidbibliotheken (siehe 3.4), dass fast ausschließlich die letzten 4 C-terminalen Ligandenpositionen ( $\lambda 0$  bis  $\lambda -3$ ) spezifisch von den Domänen erkannt werden. Der daraus resultierende potentielle Sequenzraum der Liganden ( $20^4 = 160\,000$  Peptide) war jedoch zu groß für eine systematische experimentelle Analyse.

Aus diesem Grund wurden zum effizienten Screening der für die Spezifität relevanten Teile des Liganden-Sequenzraums sogenannte Profilbibliotheken designiert. Dabei handelt es sich um systematische Peptidbibliotheken, die nur denjenigen Teil des gesamten potentiellen Sequenzraums abdecken, in welchem potentiell relevante Affinitäten zu erwarten sind. Im Falle der PDZ-Domänen enthielten die domänenspezifischen Profilbibliotheken jeweils alle Kombinationen der Aminosäuren, die an den 4 C-terminalen Ligandenpositionen von der entsprechenden Domäne präferiert wurden (Tabelle 3-16, Oben). Zusätzlich zu den

**Tabelle 3-16 Design der Profilbibliotheken für die *hAF6*-PDZ und *hERBIN*-PDZ Domänen**

Die domänenspezifischen Profilbibliotheken enthalten alle Kombinationen der präferierten Aminosäuren an den einzelnen Ligandenpositionen.

	<i>hAF6</i> -PDZ				<i>hERBIN</i> -PDZ			
	$\lambda -3$	$\lambda -2$	$\lambda -1$	$\lambda 0$	$\lambda -3$	$\lambda -2$	$\lambda -1$	$\lambda 0$
Präferierte <sup>a)</sup> Aminosäuren	G A P S E	V T S I L F	F L T I Y N W D E	V	E D	T S V	F W D A	V L I
Zusätzliche <sup>b)</sup> Aminosäuren	D F		Q P S	L I F	V S G	I	E S L P Y	
Anzahl <sup>c)</sup> Gesamt	7 x 6 x 12 x 4 2016				5 x 4 x 9 x 3 540			

<sup>a)</sup> Präferierte Aminosäuren wurden aus der manuellen Auswertung der Substitutionsanalysen und des Screenings der *Bibliothek humaner C-Termini* abgeleitet.

<sup>b)</sup> Die zusätzlichen Aminosäuren waren entweder nur sehr schwach präferiert oder wurden mit einbezogen, um weitere Aminosäureeigenschaften in den Profilbibliotheken zu repräsentieren.

<sup>c)</sup> Gesamtanzahl der Peptide in der Profilbibliothek durch Multiplikation der Aminosäuremöglichkeiten pro Ligandenposition ( $\lambda 0$  bis  $\lambda -3$ ).

eindeutig präferierten Aminosäuren wurden auch solche mit einbezogen, die in den Substitutionsanalysen nur sehr schwache, aber reproduzierbare Signale lieferten. An den Ligan- denpositionen ohne klare Präferenzen ( $\lambda - 1$ ) wurde für jede Aminosäureeigenschaft je eine repräsentative Aminosäure mit einbezogen (Tabelle 3-16, Mitte). Somit umfasste die Profil- bibliothek für die *hAF6*-PDZ Domäne 2016 und diejenige für die *hERBIN*-PDZ Domäne 540 Peptide (Tabelle 3-16, Unten).

Nach Inkubation der Profilbibliotheken mit der jeweiligen PDZ-Domäne<sup>27</sup> wurde die Menge gebundener Domäne pro Peptid über die Signalintensität  $I$  der Chemilumineszenz- Reaktion in *Boehringer-Light-Units* (BLU) quantifiziert. Aus den Logarithmus-transformierten Signalintensitäten  $\ln(I)$  wurden dann über das Massenwirkungsgesetz Schätzwerte der zu- grundeliegenden Affinitäten abgeleitet ( $\ln(K_d)$ , siehe 2.6.3.1). Zuvor wurden die Parameter der Beziehung zwischen den Signalintensitäten und den zugrundeliegenden Dissoziations- konstanten an Hand von Kontrollpeptiden mit Hilfe der Methode der kleinsten Quadrate geschätzt (siehe Anhang B, Abbildung B-1, B-5 und B-9; siehe auch 2.6.3.2). Für jedes dieser Kontrollpeptide, für die zuvor die  $K_d$  durch SPR bestimmt worden war (Tabelle 3-18, Seite 112), waren jeweils 5 Replikate über jede Profilbibliothek verteilt.

Auf Grund der hohen Variabilität der Signalintensitäten sind die auf diese Weise ge- schätzten Dissoziationskonstanten allein jedoch nicht verlässlich. Daher werden sie im Folgenden als kalibrierte Signalintensitäten bezeichnet. Aus demselben Grund erforderte auch die Quantifizierung der Sequenz-Affinitäts-Beziehung an Hand der Profilbibliotheken eine statistischen Analyse – die Methode der *Quantitativen Spezifitäts-Profile* –, welche es erlaubte, die Effekte auf Grund unterschiedlicher Affinitäten von der Fehlerstreuung der Signalintensitäten zu unterscheiden.

### 3.5.2 Quantitative Spezifitäts-Profile der PDZ-Domänen

Mit Hilfe der in dieser Arbeit entwickelten Methode der *Quantitativen Spezifitäts-Profile* (QSP) kann die Spezifität einer Domäne quantitativ durch die Affinitätsfunktion über den gesamten Ligan- den-Sequenzraum beschrieben werden. Für die hier untersuchten PDZ- Domänen wurde dazu die Beziehung zwischen der Sequenz der letzten 4 C-terminalen Aminosäuren und der Affinität der Wechselwirkung durch multifaktorielle Varianzanalysen (ANOVA, engl. *analysis of variance*)<sup>53</sup> statistisch modelliert. Diese Varianzanalysen stellen Regressionsmodelle dar, die eine lineare Beziehung zwischen den Aminosäuren der 4 C-

terminalen Ligandenpositionen (faktorielle Variablen) und den Affinitäten (verhältnisskalierte Variablen) modellieren (QSP-Modell).

Basierend auf den kalibrierten Signalintensitäten ( $\ln(K_d)$ ) der Profilbibliotheken der hAF6-PDZ und hERBIN-PDZ Domänen wurde jeweils ein QSP-Modell ohne (Gleichung 3:1) und eines mit paarweisen statistischen Interaktionstermen (Gleichung 3:2) erstellt. Dabei handelt es sich, im Gegensatz zu den weiter unten beschriebenen Gemischte-Effekte-ANOVA-Modellen, um sogenannte Feste-Effekte-ANOVA-Modelle, da diese nur eine Quelle der Zufallsstreuung berücksichtigen:

$$Y_{i,k,l,m,n} = \mu + \lambda_{-3,i} + \lambda_{-2,k} + \lambda_{-1,l} + \lambda_{0,m} + \varepsilon_{i,k,l,m,n} \quad (3:1)$$

$$Y_{i,k,l,m,n} = \mu + \lambda_{-3,i} + \lambda_{-2,k} + \lambda_{-1,l} + \lambda_{0,m} + (\lambda_{-3}:\lambda_{-2})_{i,k} + (\lambda_{-3}:\lambda_{-1})_{i,l} + (\lambda_{-3}:\lambda_0)_{i,m} + (\lambda_{-2}:\lambda_{-1})_{k,l} + (\lambda_{-2}:\lambda_0)_{k,m} + (\lambda_{-1}:\lambda_0)_{l,m} + \varepsilon_{i,k,l,m,n} \quad (3:2)$$

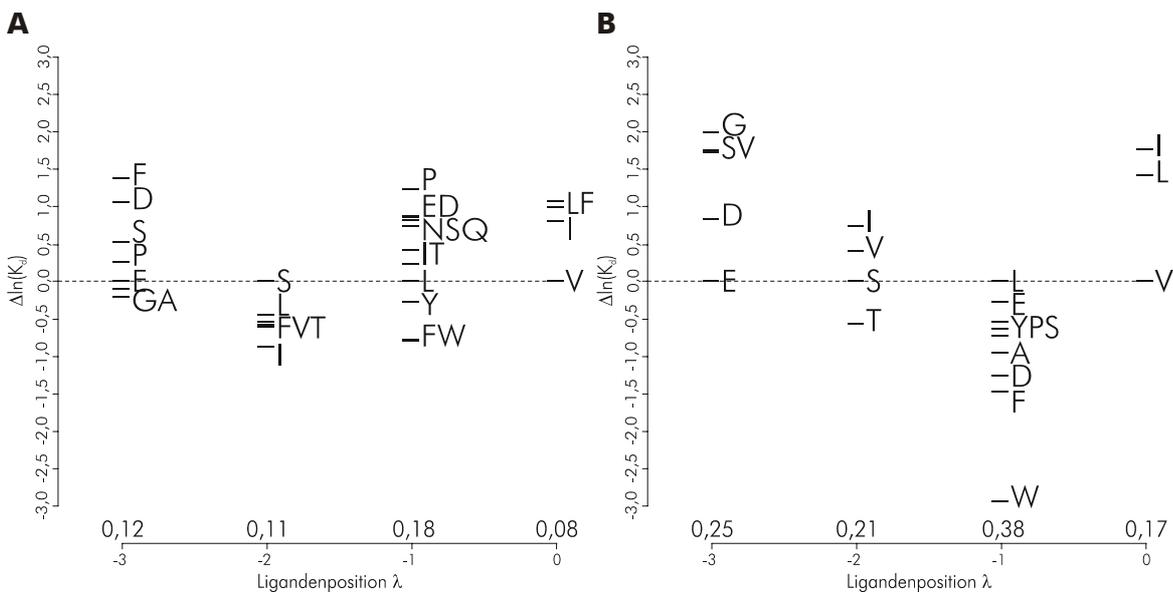
#### QSP-Modelle ohne (3:1) und mit (3:2) paarweiser statistischer Interaktionsterme (Kooperativität)

Bei diesen QSP-Modellen handelt es sich prinzipiell um Feste-Effekte-ANOVA-Modelle. Dabei entspricht  $Y$  den kalibrierten Signalintensitäten (geschätzte  $\ln(K_d)$ ) und  $\mu$  der kalibrierten Signalintensität des gewählten Referenzpeptids **ESLV**. Die  $\lambda_{x,y}$  Terme repräsentieren für jede Ligandenposition  $X$  den Affinitätsbeitrag  $\Delta\ln(K_d)$  der Aminosäure  $Y$  relativ zum Referenzpeptid. Die paarweise Kooperativität zwischen den Aminosäuren  $Y1$  und  $Y2$  in den Ligandenpositionen  $X1$  bzw.  $X2$  wird durch die Terme  $(\lambda_{x1}:\lambda_{x2})_{y1,y2}$  modelliert. Sie entsprechen den paarweisen statistischen Interaktionstermen in ANOVA-Modellen. Die  $\lambda_{x,y}$  Terme werden auch Haupteffekte genannt, um sie von den Interaktionseffekten  $(\lambda_{x1}:\lambda_{x2})_{y1,y2}$  zu unterscheiden.  $\varepsilon$  ist die durch das Modell nicht erklärte Zufallsstreuung.

Die Feste-Effekte-ANOVA-Modelle ohne paarweise statistische Interaktionsterme verglichen die Mittelwerte von Gruppen bestehend aus 168 bis 504 (hAF6-PDZ) bzw. 60 bis 180 (hERBIN-PDZ) einzelnen kalibrierten Signalintensitäten. Für die Feste-Effekte-ANOVA-Modelle mit paarweisen statistischen Interaktionstermen enthielten die analysierten Gruppen noch zwischen 24 und 84 (hAF6-PDZ) bzw. 12 und 45 (hERBIN-PDZ) kalibrierten Signalintensitäten. Auf die Analyse von Modellen mit dreifachen statistischen Interaktionstermen wurde wegen der zu geringen Fallzahl pro Gruppe verzichtet.

Zur Visualisierung der QSP-Modelle wurden – in Anlehnung an die Termschemata der Energieniveaus eines Moleküls – die Termschemata der Spezifität entwickelt. In ihnen werden die relativen Affinitätsbeiträge  $\Delta\ln(K_d)$  der unterschiedlichen Aminosäuren an den einzelnen Ligandenpositionen zusammengefasst. Für die QSP-Modelle ohne Interaktionsterme sind die Termschemata in Abbildung 3-32 dargestellt. Im Anhang B werden in den Abbildungen B-2 und B-6 alle QSP-Modell für hAF6-PDZ und hERBIN-PDZ miteinander verglichen.

Feste-Effekte-ANOVA-Modelle berücksichtigen allein die Zufallsstreuung auf Grund der unabhängigen zufälligen Stichprobenziehungen.



**Abbildung 3-32** Termschemata der QSP-Modelle für hAF6-PDZ und hERBIN-PDZ

Dargestellt sind die Termschemata der QSP-Modelle für den reduzierten Liganden-Sequenzraum der **(A)** hAF6-PDZ und **(B)** hERBIN-PDZ Domänen auf der Basis der ANOVA-Modelle ohne paarweise Interaktionsterme, welche auf den entsprechenden Profilbibliotheken trainiert wurden. Termschemata visualisieren dabei die relativen Affinitätsbeiträge  $\Delta \ln(K_d)$  der verschiedenen Aminosäuren, welche durch die ANOVA-Modelle jeweils für die 4 C-terminalen Ligandenpositionen relativ zur gewählten Referenz **ESLV<sub>COOH</sub>** (CIN4) quantifiziert wurden. Sie stehen im direkten Zusammenhang zu den relativen Energiebeiträgen der einzelnen Interaktionen:  $\Delta \Delta G^0 = -RT \ln(K_d(\text{Peptid})/K_d(\text{Referenz})) = -RT \Delta \ln(K_d)$ . Die Summe der relativen Beiträge  $\Delta \ln(K_d)$  an den einzelnen Ligandenpositionen sowie der  $\ln(K_d)$ -Wert des Referenzpeptids ergibt die vorhergesagte Affinität  $\ln(K_d)$  der jeweiligen Sequenz:  $\ln(K_d) = \ln(K_d(\text{Referenz})) + \sum \Delta \ln(K_d)$ . Die approximierten simultanen 95 % Konfidenzintervall-Bandbreiten ( $\pm$ ) wurden nach Tukey bestimmt<sup>191</sup> und sind oberhalb der Ligandenpositionen angegeben.

Für beide Profilbibliotheken waren die Varianzen der  $\ln(K_d)$ -Werte innerhalb der Gruppen unabhängig von den Mittelwerten. Die übrige Diagnostik zu den QSP-Modellen für hAF6-PDZ und hERBIN-PDZ ist im Anhang in den Abbildungen B-4 bzw. B-8 zusammengefasst. Sie zeigte, dass die Fehlerstreuung unabhängig von den modellierten Effekten und annähernd normalverteilt ist. Damit erfüllten die Profilbibliotheken die wichtigsten Anforderungen für Varianzanalysen.

Um zu überprüfen, in wie weit die Gesamtaffinität der Wechselwirkung durch die Kooperativität zwischen verschiedenen Ligandenpositionen beeinflusst wird, wurden die QSP-Modelle mit und ohne paarweise statistische Interaktionsterme verglichen. Die Kooperativität wird dabei durch die paarweisen statistischen Interaktionsterme repräsentiert. Die durch das Modell erklärte Affinitätsvariation – repräsentiert durch den Determinationskoeffizienten  $R^2$  – ist für die Modelle mit Interaktionstermen maximal 10 % größer als die der Modelle ohne Interaktionsterme (Tabelle 3-17, Seite 109), und zwar sowohl für hAF6-PDZ als auch für hERBIN-PDZ. Der auf diese Weise abgeschätzte Anteil der Kooperativität an der Gesamtvariation der Affinität betrug damit unter 10%. Im Falle von hAF6-PDZ wurde eine geringe

Kooperativität zwischen allen 4 Ligandenpositionen festgestellt, während im Falle von hERBIN-PDZ zwischen  $\lambda-1$  und  $\lambda-2$  keine signifikante Kooperativität festzustellen war.

Dieser geringe Beitrag der Kooperativität wurde durch die Ergebnisse der Substitutionsanalysen bestätigt, welche größtenteils von der Sequenz des Ausgangspeptids unabhängige Aminosäurepräferenzen zeigten. Auf Grund dieser Befunde wurde die Kooperativität in den weitergehenden Analysen vernachlässigt.

Unter dieser Voraussetzung war es zulässig, den durch die QSP-Modelle abgedeckten Sequenzraum zu erweitern, indem die Intensitätsdaten der Substitutionsanalysen, die keine Information über die Kooperativität enthalten, in das Training der Modelle mit einbezogen wurden. Dadurch konnte der komplette potentielle Liganden-Sequenzraum (ohne Cystein) mit einem Umfang von  $19^4 = 130\,321$  Peptiden durch die QSP-Modelle abgedeckt werden. Allein Cystein war auf Grund der Synthesechemie der festphasegebundenen Peptidbibliotheken in keiner der verwendeten Peptidbibliotheken repräsentiert.

Für die QSP-Modelle, die den kompletten potentiellen Liganden-Sequenzraum abdecken, wurden Gemischte-Effekte-ANOVA-Modelle<sup>102</sup> zur Analyse der um die Substitutionsanalysen erweiterten Trainingsdatensätze verwendet. Dadurch wurden die systematischen Fehler zwischen verschiedenen Inkubationen von festphasegebundenen Peptidbibliotheken und die damit verbundene korrelierte Fehlerstruktur berücksichtigt. Dazu modellieren Gemischte-Effekte-ANOVA-Modelle explizit mehrere Ebenen der Fehlerstreuung (Gleichung 3:3), wodurch die gesuchten festen Effekte von den zufälligen Effekten auf Grund verschiedener Experimente unterschieden werden können:

$$Y_{i,k,l,m,n,z} = \mu + \zeta_z + \lambda_{-3,iz} + \lambda_{-2,kz} + \lambda_{-1,lz} + \lambda_{0,mz} + \varepsilon_{i,k,l,m,n,z} \quad (3:3)$$

$$Y_{h,j,k,l,m,n,z} = \mu + \zeta_z + \lambda_{-4,hz} + \lambda_{-3,jz} + \lambda_{-2,kz} + \lambda_{-1,lz} + \lambda_{0,mz} + \varepsilon_{h,j,k,l,m,n,z} \quad (3:4)$$

#### Gemischte-Effekte-QSP-Modelle

Bei diesen QSP-Modellen handelt es sich prinzipiell um Gemischte-Effekte-ANOVA-Modelle.  $\zeta$  repräsentiert die Zufallsstreuung auf Grund der unabhängig voneinander durchgeführten Experimente ( $z = 1, \dots, Z$  unterschiedliche Substitutionsanalysen oder Profilbibliotheken). Gleichung 3:4 enthält für mSNA1-PDZ einen zusätzlichen Term  $\lambda_{-4}$  zur Modellierung des Einflusses von Ligandenposition  $\lambda-4$ .

Die Ergebnisse der finalen Gemischte-Effekte-QSP-Modelle sind in Abbildung 3-33A/C als Termschemata visualisiert. Die Vorhersageleistung dieser Modelle wird in Abbildung 3-33B/D gezeigt. Für eine vergleichende Übersicht aller QSP-Modelle für hAF6-PDZ und hERBIN-PDZ siehe Anhang B, Abbildungen B-2 und B-6.

Gemischte-Effekte-ANOVA-Modelle berücksichtigen zwei Quellen der Zufallsstreuung: die unabhängigen zufälligen Stichprobenziehungen innerhalb einer Gruppe (Experiment) und die zufällig Stichprobenziehen der jeweiligen Gruppe.

Die relativen Affinitätsbeiträge, welche durch die Gemischte-Effekte-QSP-Modelle quantifiziert wurden, unterschieden sich nicht signifikant von denen der entsprechenden Feste-Effekte-QSP-Modelle (vergleiche Abbildung 3-32, Seite 107 und Abbildung 3-33, Seite 111). Demnach konnte für die *mSNA1*-PDZ Domäne ein Gemischte-Effekte-QSP-Modell allein auf den Daten der Substitutionsanalysen trainiert werden (Abbildung 3-33E/F, Seite 111; siehe auch Anhang B, Abbildung B-10 und B-12). Zusätzlich wurde für *mSNA1*-PDZ auch der Einfluss der Ligandenposition  $\lambda-4$  modelliert (Gleichung 3:4).

Die Vorhersageleistung sowie die Anpassung an die Trainingsdaten aller QSP-Modelle vergleicht Tabelle 3-17, die entsprechenden Diagramme sind in Anhang B, Abbildung B-3 (*hAF6*-PDZ), Abbildung B-7 (*hERBIN*-PDZ) und Abbildung B-11 (*mSNA1*-PDZ) dargestellt. Unter Berücksichtigung der hohen Variabilität der Signalintensitäten zeichnen sich die Modelle durchwegs durch gute Anpassung an die Trainingsdaten aus ( $R^2$ ). Mit einem geschätzten multiplikativen Fehler  $M_{\text{Fehler}} < 2$  liegt die geschätzte Fehlerstreuung der Vorhersage im Bereich der Fehlerstreuung experimentell bestimmter Dissoziationskonstanten. Die Anpassung der vorhergesagten Affinitäten an die experimentell bestimmten Affinitäten ( $\nu R^2$ ) bestätigt die hervorragende Vorhersageleistung.

**Tabelle 3-17 Anpassung und Vorhersageleistung der QSP-Modelle**

QSP-Modell <sup>a)</sup>	Trainingsdaten <sup>b)</sup>	Fallzahl <sup>c)</sup>	$R^2$ <sup>d)</sup>	Para. <sup>e)</sup>	SDEP <sup>f)</sup>	$M_{\text{Fehler}}$ <sup>g)</sup>	$\nu R^2$ <sup>h)</sup>	$Q^2$ <sup>i)</sup>
<i>hAF6</i> -PDZ								
Feste-Effekte ohne IT	Profilbibliothek	2016	0,80	26	0,50	1,65	0,69	0,68
Feste-Effekte mit IT	Profilbibliothek	2016	0,90	243	0,35	1,42	0,76	0,75
Gemischte-Effekte ohne IT	Profilbib. + 18 SA.	2016+3280	*0,79	73+19	0,48	1,62	0,72	0,70
<i>hERBIN</i> -PDZ								
Feste-Effekte ohne IT	Profilbibliothek	540	0,82	18	0,67	1,95	0,63	0,63
Feste-Effekte mit IT	Profilbibliothek	540	0,91	116	0,47	1,60	0,62	0,67
Gemischte-Effekte ohne IT	Profilbib. + 25 SA.	540+5020	*0,78	73+26	0,58	1,79	0,55	0,59
<i>mSNA1</i> -PDZ								
Gemischte-Effekte ohne IT	12 SA.	2204	†--	73+12	0,53	1,70	0,76	0,88

<sup>a)</sup> Die QSP-Modelle basieren auf ANOVA-Modellen (IT = paarweise statistische Interaktionsterme).

<sup>b)</sup> Ursprung der experimentellen Intensitätsdaten, mit denen die Modelle trainiert wurden (SA. = Substitutionsanalyse).

<sup>c)</sup> Anzahl unabhängiger Intensitätswerte, die in das Training der Modelle eingegangen sind.

<sup>d)</sup>  $R^2$ : Determinationskoeffizient, quantifiziert die Anpassung des Modells an die Trainingsdaten.

<sup>e)</sup> Para.: Anzahl Modell-Parameter; für Gemischte-Effekte-Modelle Anzahl feste Effekte + Zufallseffekte.

<sup>f)</sup> SDEP: Standardabweichung der Fehler der Vorhersage.

<sup>g)</sup>  $M_{\text{Fehler}} = e^{SDEP}$ : multiplikativer Fehler der  $K_d$ -Vorhersage auf linearer Skala; an Hand SDEP geschätzt.

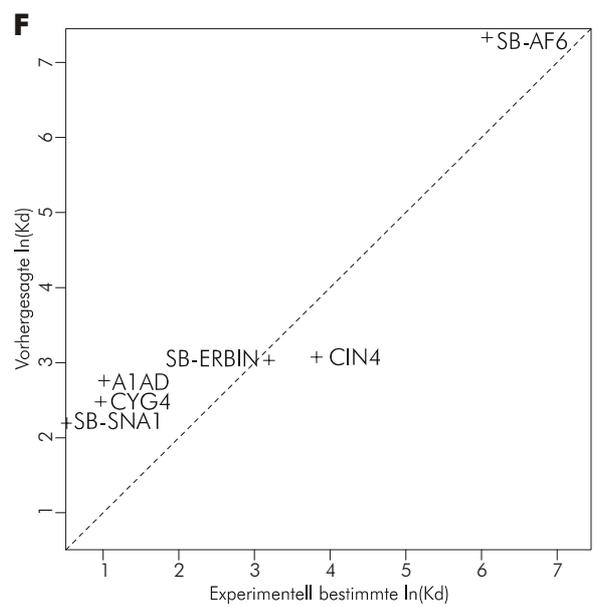
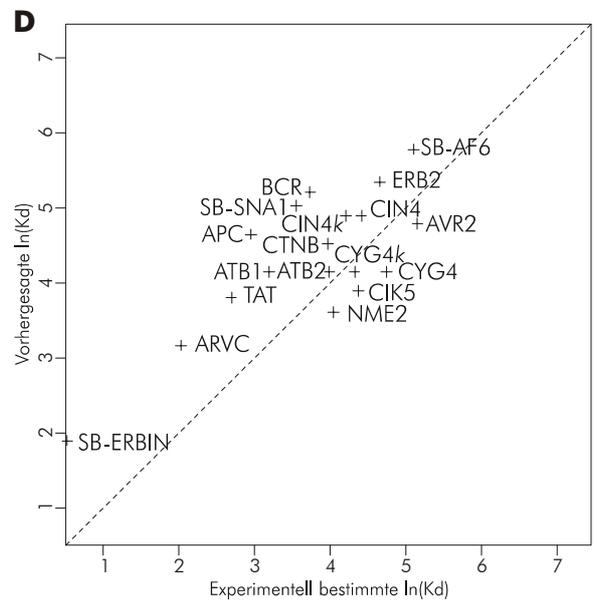
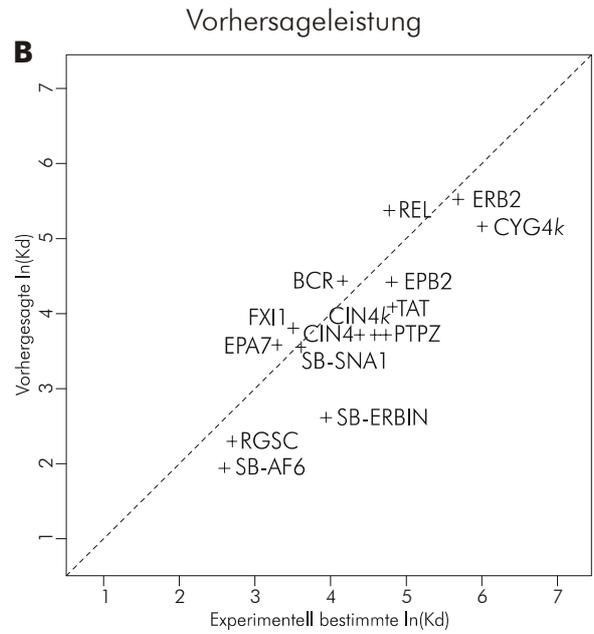
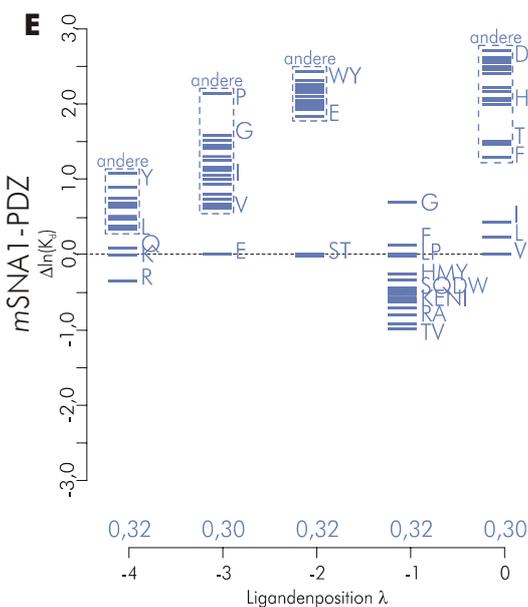
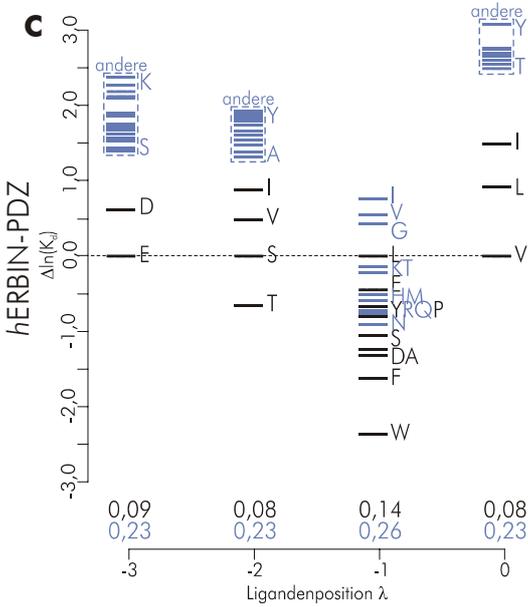
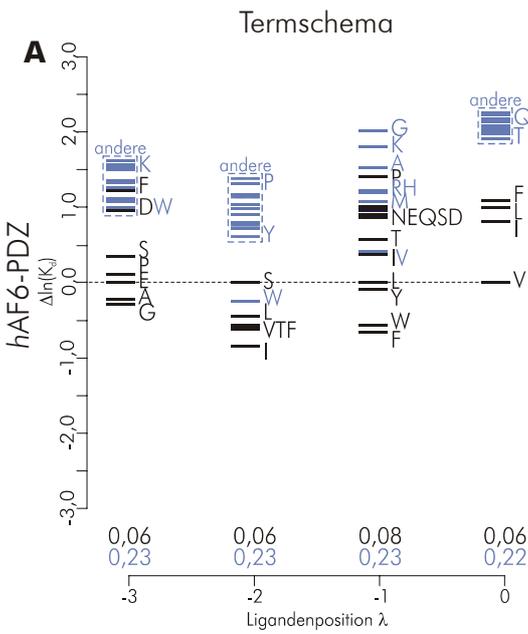
<sup>h)</sup>  $\nu R^2$ : validierter Determinationskoeffizient, bestimmt durch die Anpassung der vorhergesagten Dissoziationskonstanten an die experimentell bestimmten Dissoziationskonstanten.

<sup>i)</sup>  $Q^2$ : kreuzvalidierter Determinationskoeffizient, bestimmt durch die Anpassung der vorhergesagten Dissoziationskonstanten des kreuzvalidierten QSP-Modells an die experimentell bestimmten Dissoziationskonstanten derjenigen Peptide, ohne die das kreuzvalidierte Modell trainiert worden war.

\* Für Gemischte-Effekte-ANOVA-Modelle ist  $R^2$  nicht definiert, daher wurde die Anpassung der Vorhersagen an die Trainingsdaten der Profilbibliothek bestimmt.

† Auf Grund des Fehlens der Profilbibliothek konnte  $R^2$  nicht bestimmt werden.

Die Quantitativen Spezifitäts-Profile der drei PDZ-Domänen sind unter [www.fmp-berlin.de/nmr/pdz](http://www.fmp-berlin.de/nmr/pdz) frei verfügbar. Die dort angebotene Anwendung erlaubt jedem Interessenten die Affinität gegenüber den drei untersuchten PDZ-Domänen vorherzusagen zu lassen. Die Vorhersage kann sowohl für einzelne Ligandensequenzen als auch für ganze Listen von Sequenzen durchgeführt werden.



**Abbildung 3-33 Termschemata und Vorhersageleistung der finalen Gemischte-Effekte-QSP-Modelle**

Die Termschemata (linke Spalte) visualisieren die relativen Affinitätsbeiträge  $\Delta \ln(K_d)$  der verschiedenen Aminosäuren relativ zur gewählten Referenz CIN4 (**K**) **ESLIV**<sub>COOH</sub>, welche durch die QSP-Modelle jeweils für die 4 oder 5 C-terminalen Ligandenpositionen quantifiziert wurden (siehe Legende zu Abbildung 3-32, Seite 107). Durch Schwarz sind dabei Beiträge gekennzeichnet, die vor allem aus der Profilibibliothek abgeleitet wurden, durch Blau solche, die nur aus den Substitutionsanalysen bestimmt wurden. Die approximierten 95 % Konfidenzintervall-Bandbreiten ( $\pm$ ) sind oberhalb der Ligandenpositionen jeweils für die aus den Profilibibliotheken (Schwarz) bzw. für die aus den Substitutionsanalysen (Blau) abgeleiteten Beiträge angegeben. Die Vorhersageleistung der Modelle ist in Form von Diagrammen dargestellt (rechte Spalte), in welchen die vorhergesagten Affinitäten ( $\ln(K_d)$ ) gegen die experimentell bestimmten Affinitäten (Tabelle 3-18) aufgetragen sind. **(A)/(B)** Gemischte-Effekte-QSP-Modell für hAF6-PDZ. **(C)/(D)** Gemischte-Effekte-QSP-Modell für hERBIN-PDZ. **(E)/(F)** Gemischte-Effekte-QSP-Modell für mSNA1-PDZ.

Zusätzlich wurden kreuzvalidierte QSP-Modelle erstellt. Dazu wurden die kalibrierten Signalintensitäten all derjenigen Peptide, für die Dissoziationskonstanten experimentell bestimmt worden waren, aus dem Trainingsdatensatz (Profilibibliothek/Substitutionsanalysen) entfernt. Mit den auf diesem reduzierten Datensatz trainierten kreuzvalidierten QSP-Modellen wurden anschließend die Dissoziationskonstanten für die nicht im Trainingsdatensatz enthaltenen Peptide vorhergesagt und mit den experimentell bestimmten Dissoziationskonstanten verglichen. Die dadurch bestimmten kreuzvalidierten Determinationskoeffizienten ( $Q^2$ ) wichen nicht signifikant von den validierten  $vR^2$ -Werten ab (Tabelle 3-17). Ebenso veränderten sich die aminosäurespezifischen Affinitätsbeiträge zwischen den unterschiedlichen Modellen nicht signifikant (vergleiche in Anhang B, Abbildung B-2 für hAF6-PDZ, Abbildung B-6 für hERBIN-PDZ und Abbildung B-10 für mSNA1-PDZ).

In Tabelle 3-18, Seite 112, sind die experimentell bestimmten Dissoziationskonstanten und die Vorhersagen der QSP-Modelle zusammengefasst. Für hAF6-PDZ und hERBIN-PDZ weichen die Vorhersagen der Feste-Effekte-QSP-Modelle generell nicht um mehr als einen Faktor 2 von den experimentell bestimmten Dissoziationskonstanten ab. Dies entspricht den geschätzten multiplikativen Fehlern (Tabelle 3-17,  $M_{Fehler}$ ). Es gibt jedoch einige Ausnahmen für hAF6-PDZ (BCR, REL, ERB2), die bis um einen Faktor 4 abweichen, sowie einige Ausnahmen für hERBIN-PDZ (TAT, APC, ATB1, BCR, CTNB, NME2), die bis um einen Faktor 4,4 abweichen (Tabelle 3-18). Die Vorhersagen des Gemischte-Effekte-QSP-Modells für mSNA1-PDZ wichen bis zu einem Faktor 5,8 von den experimentell bestimmten  $K_d$ -Werten ab. Die größere Abweichung kann auf den geringeren Umfang und die geringere Qualität (keine Profilibibliothek) der Trainingsdaten zurückgeführt werden.

**Tabelle 3-18 Experimentell bestimmte und vorhergesagte Dissoziationskonstanten**

Die durch die QSP-Modelle vorhergesagten Dissoziationskonstanten weichen bis auf wenige Ausnahmen um nicht mehr als einen Faktor 2 von den experimentell bestimmten  $K_d$ -Werten ab. Die Superbinder (SB-) wurden von der entsprechenden PDZ-Domäne mit der höchsten Affinität im Vergleich zu allen anderen untersuchten Liganden gebunden (siehe grau unterlegte  $K_d$ -Werte). Die experimentell untersuchten Liganden und die vorhergesagten Superbinder sind jeweils nach ihrer Affinität sortiert.

Domäne	Ligand <sup>a)</sup>	UP-ID <sup>b)</sup>	Sequenz <sup>c)</sup>	SPR <sup>d)</sup> (experimentell) $K_d$ [ $\mu$ M]	QSP <sup>e)</sup> (Vorhersage) $K_d$ [ $\mu$ M]
<b>hAF6-PDZ</b>					
	RGSC	O14924	PKTSAHH <b>ATFV</b> <sub>COOH</sub>	<b>14,8</b> ( 12,0 - 18,2 )	<b>10,4</b> ( 8,2 - 13,2 )
	EPA7	Q15375	MLHLHG <b>TGIQV</b> <sub>COOH</sub>	<b>26,9</b> ( 20,3 - 35,6 )	<b>30,0</b> ( 23,6 - 38,2 )
	FXI1	Q12951	GVLYPRE <b>GTEV</b> <sub>COOH</sub>	<b>33,4</b> ( 28,3 - 39,5 )	<b>67,0</b> ( 52,7 - 85,3 )
	BCR	P11274	RQSILF <b>STEV</b> <sub>COOH</sub>	<b>64,1</b> ( 58,4 - 70,3 )	<b>208,1</b> ( 163,6 - 264,8 )
	CIN4	P35499	TVRPGV <b>KESLV</b> <sub>COOH</sub>	<b>80,2</b> ( 77,4 - 83,2 )	<b>82,1</b> ( 64,5 - 104,5 )
	CIN4k	P35499	GV <b>KESLV</b> <sub>COOH</sub>	<b>98,5</b> ( 81,6 - 118,9 )	<b>82,1</b> ( 64,5 - 104,5 )
	PTPZ	P23471	GNIAESL <b>ESLV</b> <sub>COOH</sub>	<b>113,5</b> ( 101,4 - 127,1 )	<b>82,1</b> ( 64,5 - 104,5 )
	REL	Q04864	DSFPYE <b>FFQV</b> <sub>COOH</sub>	<b>118,9</b> ( 112,0 - 126,2 )	<b>474,8</b> ( 373,1 - 604,1 )
	EPB2	P29323	I <b>QSVEV</b> <sub>COOH</sub>	<b>122,8</b> ( 108,9 - 138,3 )	<b>194,6</b> ( 152,9 - 247,6 )
	TAT	P03409	SEKHF <b>RETEV</b> <sub>COOH</sub>	<b>123,4</b> ( 94,4 - 161,3 )	<b>204,0</b> ( 160,3 - 259,6 )
	ERB2	P04626	PEYLGL <b>DVPV</b> <sub>COOH</sub>	<b>297,0</b> ( 261,5 - 337,4 )	<b>625,4</b> ( 491,5 - 795,7 )
	CYG4k	P33402	FL <b>RETSL</b> <sub>COOH</sub>	<b>407,9</b> ( 396,8 - 419,3 )	<b>289,6</b> ( 227,6 - 368,5 )
Superbinder	SB-AF6		LE <b>GIFV</b> <sub>COOH</sub>	<b>13,4</b> ( 12,2 - 14,7 )	<b>7,6</b> ( 5,9 - 9,6 )
	SB-SNA1		IR <b>ETIV</b> <sub>COOH</sub>	<b>36,8</b> ( 32,8 - 41,3 )	<b>46,8</b> ( 36,8 - 59,6 )
	SB-ERBIN		WL <b>ETWV</b> <sub>COOH</sub>	<b>51,3</b> ( 46,8 - 56,3 )	<b>15,5</b> ( 12,1 - 19,7 )
<b>hERBIN-PDZ</b>					
	ARVC	O00192	DAKPQ <b>PVDSWV</b> <sub>COOH</sub>	<b>7,6</b> ( 5,0 - 11,4 )	<b>8,3</b> ( 5,4 - 12,7 )
	TAT	P03409	SEKHF <b>RETEV</b> <sub>COOH</sub>	<b>14,8</b> ( 12,5 - 17,5 )	<b>50,7</b> ( 33,1 - 77,6 )
	APC	P25054	HSGSYL <b>VTSV</b> <sub>COOH</sub>	<b>19,2</b> ( 18,2 - 20,3 )	<b>74,5</b> ( 48,7 - 114,1 )
	ATB1	P20020	GSP <b>LHSL</b> ETSL <sub>COOH</sub>	<b>24,2</b> ( 19,8 - 29,7 )	<b>62,0</b> ( 40,5 - 94,9 )
	BCR	P11274	RQSILF <b>STEV</b> <sub>COOH</sub>	<b>41,9</b> ( 29,9 - 58,6 )	<b>173,7</b> ( 113,4 - 265,9 )
	CTNB	P35222	SNQLAW <b>FD</b> TDL <sub>COOH</sub>	<b>53,0</b> ( 52,6 - 53,5 )	<b>232,0</b> ( 151,5 - 355,3 )
	ATB2	Q01814	GSP <b>IHS</b> LETSL <sub>COOH</sub>	<b>54,1</b> ( 49,5 - 59,1 )	<b>62,0</b> ( 40,5 - 94,9 )
	NME2	Q13244	YK <b>KLSS</b> IESDV <sub>COOH</sub>	<b>57,1</b> ( 55,6 - 58,7 )	<b>23,1</b> ( 15,1 - 35,4 )
	CIN4k	P35499	GV <b>KESLV</b> <sub>COOH</sub>	<b>67,4</b> ( 57,7 - 78,8 )	<b>119,1</b> ( 77,8 - 182,4 )
	CYG4k	P33402	FL <b>RETSL</b> <sub>COOH</sub>	<b>75,4</b> ( 27,8 - 205,1 )	<b>62,0</b> ( 40,5 - 94,9 )
	CIK5	P22459	LCL <b>DS</b> RETDL <sub>COOH</sub>	<b>79,3</b> ( 76,6 - 82,1 )	<b>50,4</b> ( 32,9 - 77,2 )
	CIN4	P35499	TVRPGV <b>KESLV</b> <sub>COOH</sub>	<b>82,6</b> ( 81,7 - 83,6 )	<b>119,1</b> ( 77,8 - 182,4 )
	ERB2	P04626	PEYLGL <b>DVPV</b> <sub>COOH</sub>	<b>105,0</b> ( 81,9 - 134,7 )	<b>181,7</b> ( 118,7 - 278,2 )
	CYG4	P33402	IG <b>TM</b> FL <b>RETSL</b> <sub>COOH</sub>	<b>115,3</b> ( 106,5 - 124,9 )	<b>62,0</b> ( 40,5 - 94,9 )
	AVR2	P27037	NVDF <b>PP</b> KE <b>SSL</b> <sub>COOH</sub>	<b>173,7</b> ( 142,3 - 211,9 )	<b>261,3</b> ( 170,6 - 400,1 )
Superbinder	SB-ERBIN		WL <b>ETWV</b> <sub>COOH</sub>	<b>1,7</b> ( 0,5 - 5,3 )	<b>2,4</b> ( 1,5 - 3,6 )
	SB-SNA1		IR <b>ETIV</b> <sub>COOH</sub>	<b>34,7</b> ( 32,1 - 37,6 )	<b>*153,2</b> ( 121,4 - 193,3 )
	SB-AF6		LE <b>GIFV</b> <sub>COOH</sub>	<b>165,0</b> ( 162,2 - 167,8 )	<b>524,9</b> ( 342,8 - 803,7 )
<b>mSNA1-PDZ</b>					
	CYG4	P33402	IG <b>TM</b> FL <b>RETSL</b> <sub>COOH</sub>	<b>2,6</b> ( 2,2 - 3,2 )	<b>11,9</b> ( 11,3 - 12,6 )
	A1AD	P25100	ADYSNL <b>RETDI</b> <sub>COOH</sub>	<b>2,7</b> ( 2,4 - 3,1 )	<b>15,8</b> ( 15,0 - 16,6 )
	CIN4	P35499	TVRPGV <b>KESLV</b> <sub>COOH</sub>	<b>45,7</b> ( 39,2 - 53,4 )	<b>21,7</b> ( 20,6 - 22,8 )
	MCM7	P33993	VNASR <b>TRITFV</b> <sub>COOH</sub>	<b>n.b.</b>	<b>59,6</b> ( 56,6 - 62,7 )
Superbinder	SB-SNA1		IR <b>ETIV</b> <sub>COOH</sub>	<b>1,7</b> ( 1,0 - 2,6 )	<b>8,9</b> ( 6,5 - 12,2 )
	SB-ERBIN		WL <b>ETWV</b> <sub>COOH</sub>	<b>24,3</b> ( 22,4 - 26,3 )	<b>20,8</b> ( 13,5 - 31,9 )
	SB-AF6		LE <b>GIFV</b> <sub>COOH</sub>	<b>436,2</b> ( 263,4 - 722,2 )	<b>1540,0</b> ( 909,0 - 2610,0 )

<sup>a)</sup> Kurzbezeichnung des Liganden. Der Suffix *k* kennzeichnet die verkürzte Version eines doppelt auftretenden Peptids.

<sup>b)</sup> Primärer UniProt<sup>10</sup> Zugriffscode des Liganden.

<sup>c)</sup> Die für die QSP-Vorhersage verwendeten 4 C-terminalen Reste sind fett gedruckt.

<sup>d)</sup> Durch Oberflächen-Plasmon-Resonanz-Verfahren (SPR) bestimmte  $K_d$  (n.b. nicht bestimmt). In Klammern ist das  $\pm 1\sigma$  Intervall der 2–4 Messwiederholungen auf logarithmischer Skala angegeben. Grau unterlegt ist sowohl die niedrigste Dissoziationskonstante der experimentell untersuchten Liganden als auch die des jeweiligen domänenspezifischen Superbinders.

<sup>e)</sup> QSP-Vorhersage für hAF6-PDZ und hERBIN-PDZ mittels Feste-Effekte-QSP-Modell mit paarweisen statistischen Interaktionstermen; QSP-Vorhersage für mSNA1-PDZ mittels Gemischte-Effekte-QSP-Modell. In Klammern ist das 95 % Konfidenzintervall der Vorhersage auf der logarithmischen Skala angegeben.

\* Für hERBIN-PDZ/SB-SNA1 musste das Gemischte-Effekte-QSP-Modell verwendet werden.

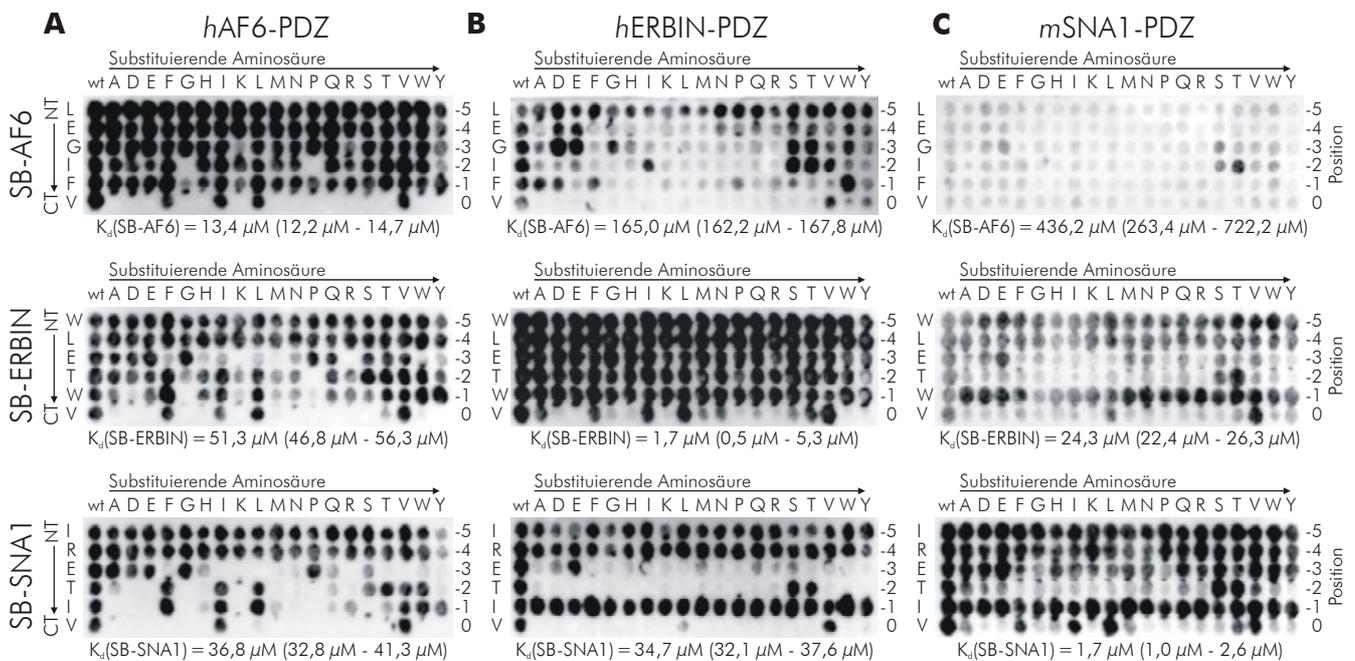
### 3.5.3 Design peptidischer Liganden mit maximaler Affinität: superbindende Peptide

Die in 3.5.2 entwickelten Gemischte-Effekte-QSP-Modelle ermöglichten die Vorhersage der Affinitäten der drei PDZ-Domänen gegenüber allen Peptiden des gesamten potentiellen Sequenzraums. Damit konnte innerhalb des abgedeckten Sequenzraums für jede der drei PDZ-Domänen der peptidische Ligand mit maximaler Affinität vorhergesagt werden. Diese Liganden wurden als domänenspezifische *Superbinder* bezeichnet (Abkürzung *SB-Domäne*).

Aus den Termschemata (Abbildung 3-33A/C/E) können die Superbinder direkt als Kombination derjenigen Aminosäuren abgeleitet werden, die an den einzelnen Ligandenpositionen die niedrigsten  $\Delta\ln(K_d)$  aufweisen. So wird beispielsweise für *hAF6*-PDZ die Ligandensequenz **GIFV**<sub>COOH</sub> als Superbinder vorhergesagt (vergleiche Abbildung 3-33A). Durch alle alternativen QSP-Modelle (Feste-Effekte-QSP-Modelle, usw.) ergab sich für jede Domäne konsistent derselbe Superbinder. Die Aminosäuren N-terminal der Ligandenposition  $\lambda-3$  wurden nach einer semi-quantitativen Analyse der Substitutionsanalysen gewählt. Damit wurden für die *hAF6*-PDZ Domäne der Superbinder SB-AF6 mit der Sequenz **LEGIFV**<sub>COOH</sub>, für *hERBIN*-PDZ der Superbinder SB-ERBIN mit **WLETWV**<sub>COOH</sub> sowie für *mSNA1*-PDZ der Superbinder SB-SNA1 mit **IRETIV**<sub>COOH</sub> vorhergesagt.

Um die Vorhersagen experimentell zu überprüfen, wurden die Superbinder synthetisiert und die Dissoziationskonstanten durch SPR bestimmt. Alle drei Superbinder konnten experimentell bestätigt werden: sie wurden von ihrer entsprechenden PDZ-Domäne – im Vergleich mit allen anderen experimentell untersuchten Liganden – mit der höchsten gemessenen Affinität gebunden. Auch die für die Superbinder vorhergesagten  $K_d$ -Werte wichen um nicht mehr als einen Faktor 2 von den experimentell bestimmten Dissoziationskonstanten ab. Erwartungsgemäß war der Fehler der Vorhersage für SB-SNA1 etwas größer.

Um die Erkennung der Superbinder durch ihre spezifischen Domänen und die Kreuzreaktivität mit den „unspezifischen“ Domänen zu visualisieren, wurden mit allen drei PDZ-Domänen Substitutionsanalysen der drei Superbinder durchgeführt (Abbildung 3-34). Diese bestätigten erneut, dass die Superbinder von ihren spezifischen Domänen am stärksten gebunden werden (Diagonale in Abbildung 3-34). Darüber hinaus konnten durch diese Analyse Unterschiede in der Selektivität der Wechselwirkungen festgestellt werden: so erkennt *hAF6*-PDZ im Vergleich zu den anderen beiden PDZ-Domänen den eigenen Superbinder am wenigsten selektiv (Abbildung 3-34, vertikale Spalten). Umgekehrt zeigte jedoch SB-AF6



**Abbildung 3-34** Substitutionsanalysen der Superbinder

Substitutionsanalysen der drei Superbinder **(A)** SB-AF6, **(B)** SB-ERBIN und **(C)** SB-SNA1 wurden mit den drei PDZ-Domänen *hAF6*-PDZ, *hERBIN*-PDZ und *mSNA1*-PDZ durchgeführt. Eine Substitutionsanalyse ist eine „tabellenartige“ Peptidbibliothek von Einzelsubstitutionsanaloga, wobei die Zeile die Position, die substituiert wird, und die Spalte die Aminosäure, durch die substituiert wird, bestimmt. Die Stärke der Wechselwirkung zwischen Peptid und Domäne wird durch die Schwärzung dargestellt (siehe auch Abbildung 3-31). Die durch Oberflächen-Plasmon-Resonanz experimentell bestimmten Dissoziationskonstanten der Wechselwirkungen sind zusammen mit dem  $\pm 1\sigma$  Intervall auf der logarithmischen Skala unter der jeweiligen Substitutionsanalyse angegeben. Abbildung modifiziert nach Wiedemann und Boisguerin *et al.*, 2004.

die größte Selektivität gegenüber „seiner“ PDZ-Domäne (Abbildung 3-34, horizontale Zeilen). Beides wird auch durch die experimentell bestimmten Dissoziationskonstanten unterstützt. Somit vermittelt diese Analyse einen Eindruck von der Selektivität der drei PDZ-Domänen bzw. der drei Superbinder.

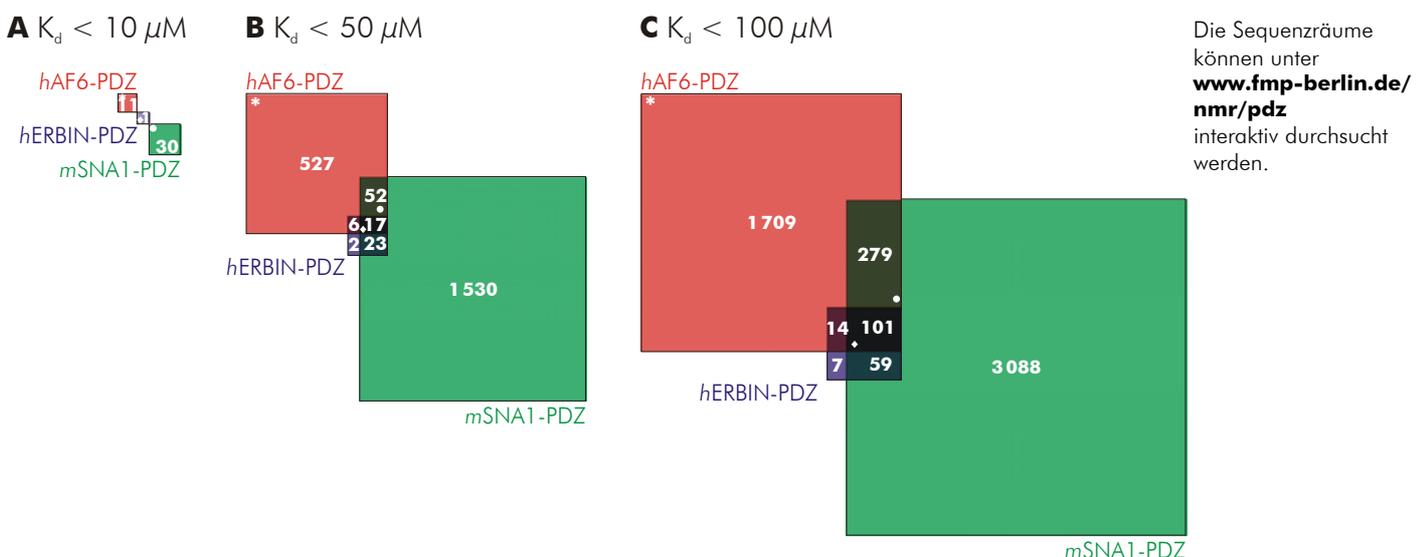
Durch die erfolgreiche Vorhersage der Superbinder wurde die Verlässlichkeit der Vorhersagen der *Quantitativen Spezifitäts-Profile* erneut bestätigt. Gleichzeitig veranschaulicht diese Anwendung das Potential dieser Methode zur Beschreibung der Spezifität von Protein-Interaktionsdomänen und ihren Nutzen für das Design von Liganden mit erhöhter bzw. gewünschter Affinität.

### 3.5.4 Quantifizierung der PDZ-Domänen/Liganden-Selektivität

Neben der quantitativen Beschreibung der Spezifität (siehe 3.5.2) ist die quantitative Beschreibung der Selektivität essentiell für das Verständnis der Wechselwirkungen von Protein-Interaktionsdomänen. Im Kontext der Zelle konkurrieren eine Vielzahl von Domänen um eine Vielzahl von Liganden. Die Quantifizierung der Selektivität, als Maß für die Menge der gebundenen Liganden bzw. die Überschneidung der Liganden-Sequenzräume konkurrierender Domänen, hilft die Konkurrenz-Situation besser zu verstehen.

Selektivität ist dabei immer relativ zu den verfügbaren Liganden zu betrachten. Da jedoch Faktoren, welche die Verfügbarkeit der Liganden beeinflussen, nicht Gegenstand dieser Arbeit waren, wird im Folgenden die Selektivität der Domänen formal, relativ zum gesamten potentiellen Liganden-Sequenzraum analysiert. Dazu wurden durch die Gemischte-Effekte-QSP-Modelle die Affinitäten aller 130 321 potentiell möglicher Ligandensequenzen gegenüber den drei PDZ-Domänen vorhergesagt.

In Abbildung 3-35 sind die Sequenzräume mit Affinitäten bis  $10 \mu\text{M}$ ,  $50 \mu\text{M}$  und  $100 \mu\text{M}$  für die drei PDZ-Domänen schematisch dargestellt. Unterhalb von  $10 \mu\text{M}$  erkennen die drei PDZ-Domänen noch getrennte Sequenzräume. *mSNA1*-PDZ bindet in diesem Affinitätsbe-



**Abbildung 3-35** Überlappung der theoretischen Liganden-Sequenzräume der drei PDZ-Domänen

Die drei PDZ-Domänen zeigen deutliche Unterschiede sowohl in der Größe als auch der Überlappung der erkannten Liganden-Sequenzräume. Die Größe der Liganden-Sequenzräume mit Dissoziationskonstanten unterhalb (A)  $10 \mu\text{M}$ , (B)  $50 \mu\text{M}$  und (C)  $100 \mu\text{M}$  sind durch farbige Flächen dargestellt. Für *hAF6*-PDZ ist die Fläche rot, für *hERBIN*-PDZ blau und für *mSNA1*-PDZ grün gefärbt. Die Überlappungsbereiche der Sequenzräume sind durch die entsprechenden Mischfarben gekennzeichnet. Die Anzahl an Liganden in den einzelnen Subräumen ist innerhalb der Flächen angegeben. Die Position der Superbinde ist durch (\*) für SB-AF6, (♦) für SB-ERBIN und (•) für SB-SNA1 gekennzeichnet.

reich 30, hAF6-PDZ 11 und hERBIN-PDZ 1 Liganden, namentlich SB-ERBIN. Mit steigenden Schwellenwerten für die Dissoziationskonstanten steigt auch die Anzahl an Liganden an, die von den drei PDZ-Domänen erkannt werden, und zwar für mSNA1-PDZ und hAF6-PDZ stärker als für hERBIN-PDZ. Diese Unterschiede in der Größe der Liganden-Sequenzräume werden durch die Ergebnisse des Screenings der *Bibliothek humaner C-Termini* qualitativ bestätigt. Denn die Anzahl an Liganden mit hohen standardisierten Signalintensitäten (Tabelle 3-13,  $MAD-Z \geq 5$ ), die durch das Screening identifiziert worden waren, ist für hERBIN-PDZ am kleinsten und für mSNA1-PDZ am größten.

Gleichzeitig beginnen die Sequenzräume der drei PDZ-Domänen mit steigenden Schwellenwerten für die Dissoziationskonstanten zu überlappen. Auch die hier festgestellte Überlappung stimmt gut mit der Überlappung überein, die für die Top100 Peptide des Screenings der *Bibliothek humaner C-Termini* bestimmt worden war. Umgekehrt wird auch die Auswahl der Top100 Peptide durch die QSP-Modelle gestützt. Denn für 51% der hAF6-PDZ Top100 Peptide, für 29% der hERBIN-PDZ Top100 Peptide sowie für 63% der mSNA1-PDZ Top100 Peptide werden Dissoziationskonstanten unter  $100 \mu\text{M}$  vorhergesagt.

Daneben unterscheiden sich die Liganden-Sequenzräume auch in der Lokalisierung des Affinitätsmaximum (repräsentiert durch die Superbinder). So liegt SB-AF6 weit vom Überlappungsbereich entfernt, während SB-SNA1 über einer vorhergesagten  $K_d$  von etwa 50 mM im Überlappungsbereich zwischen mSNA1-PDZ und hAF6-PDZ liegt. Der Superbinder von hERBIN-PDZ wird sogar bis zu einer Dissoziationskonstante von etwa  $20 \mu\text{M}$  von allen drei PDZ-Domänen erkannt. Dies konnte auch schon durch die Substitutionsanalysen der Superbinder qualitativ festgestellt werden (Abbildung 3-34).

Neben der Größe und Überlappung der Liganden-Sequenzräume ist auch ihr Verhältnis zur traditionellen Klassifizierung nach Songyang *et al.*, 1997, aufschlussreich. In den Sequenzräumen, die in Abbildung 3-35 dargestellt sind, sind sowohl Klasse I als auch Klasse II Liganden enthalten. Dabei bevorzugt hAF6-PDZ Klasse II Peptide, während mSNA1-PDZ und hERBIN-PDZ Klasse I Liganden favorisieren (Tabelle 3-19). Die „nicht favorisierten“ Ligandenmotive machen bei allen drei Domänen meist weit weniger als 20% der gebundenen Liganden aus. Dies gilt auch bei höheren Dissoziationskonstanten. Zusätzlich erkennen alle drei PDZ-Domänen Liganden, die weder in Klasse I noch Klasse II fallen (Nicht-Klasse-I/II). Dies lässt sich durch die QSP-Modelle erklären. Die Sequenzen dieser Liganden zeichnen sich dadurch aus, dass sie eine höhere Frequenz der optimalen Aminosäuren an  $\lambda-1$  und  $\lambda-3$  besitzen und immer entweder an  $\lambda 0$  oder  $\lambda-2$  eine der präferier-

**Tabelle 3-19** Selektivität der PDZ-Domänen: Klassen der Liganden

Mit steigendem Schwellenwert der Dissoziationskonstante nehmen vor allem die Liganden zu, die nicht einer der beiden Ligandenklassen angehören.

Domäne Ligandenklasse <sup>a)</sup>	$K_d < 10 \mu\text{M}$		$K_d < 50 \mu\text{M}$		$K_d < 100 \mu\text{M}$	
	Anzahl	%	Anzahl	%	Anzahl	%
<i>h</i> AF6-PDZ Total	11		602		2103	
Klasse I	2	18%	113	19%	308	15%
Klasse II	9	82%	416	69%	1070	51%
Nicht-Klasse I/II	0	0%	73	34%	725	34%
<i>h</i> ERBIN-PDZ Total	1		48		181	
Klasse I	1	100%	43	90%	131	72%
Klasse II	0	0%	4	8%	25	14%
Nicht-Klasse I/II	0	0%	1	2%	25	14%
<i>h</i> SNA1-PDZ Total	30		1622		3527	
Klasse I	30	100%	1492	92%	2003	57%
Klasse II	0	0%	4	<1%	161	5%
Nicht-Klasse I/II	0	0%	126	8%	1363	39%

<sup>a)</sup> Klassifizierung der Liganden nach Songyang *et al.*, 1997: Klasse I Liganden entsprechen dem Motiv  $\mathbf{x(S/T)x(V/I/L)_{COOH}}$ , Klasse II Liganden dem Motiv  $\mathbf{x\Phi x\Phi_{COOH}}$ , während die Liganden, die keinem der beiden Motive entsprechen als Nicht-Klasse I/II zusammengefasst werden.

ten Aminosäure aufweisen. Der prozentuale Anteil dieser Sequenzen nimmt mit steigender  $K_d$  zu, während der prozentuale Anteil der favorisierten Klasse abnimmt (Tabelle 3-19).

Als weitere Anwendung der *Quantitativen Spezifitäts-Profil*e bestärkt die Quantifizierung der Selektivität damit die Bedeutung dieser neuen Methode. Gleichzeitig ist die Quantifizierung der Selektivität selbst eine Methode, die eine neue, aufschlussreiche Sicht auf die Wechselwirkungen einer Protein-Interaktionsdomäne ermöglicht. Auf ihr aufbauend kann ein Verständnis der Konkurrenz-Situation innerhalb der Zelle aufgebaut werden.

