

Kapitel 4

Diskussion

Im Rahmen der Aufklärung von Protein-Protein-Interaktionsnetzwerken sind Protein-Interaktionsdomänen hoch interessante Untersuchungsobjekte, da sich die Erkenntnisse für einzelne Vertreter einer Domänenfamilie potentiell auf die restlichen Vertreter übertragen lassen.²¹⁰ Auf diese Weise könnte mit geringem experimentellen Aufwand eine große Anzahl an potentiell, d.h. *in vitro*, möglichen Protein-Protein-Wechselwirkungen charakterisiert werden. Gleichzeitig sind Protein-Interaktionsdomänen hervorragende Modellsysteme zur Analyse der Grundlagen spezifischer Protein-Protein-Interaktionen.

Die in dieser Arbeit entwickelten und experimentell verifizierten Wechselwirkungsmodelle zeigen, dass durch das Training statistischer Modelle auf der Basis experimenteller Daten für wenige Vertreter einer Domänenfamilie die Vorhersage von neuen potentiellen Wechselwirkungen für experimentell nicht untersuchte Vertreter möglich ist. Durch die quantitativen Modelle werden dabei nicht nur die Wechselwirkungspartner, sondern vor allem auch die Affinitäten der Wechselwirkungen vorhergesagt. Um solche quantitative Modelle auch auf Sequenzebene erstellen zu können, wurde die Methode der *Quantitativen Spezifitäts-Profile* entwickelt. Durch all diese Modelle wird damit eine Umorientierung, weg von den einfachen Sequenzmotiven und den mehrdeutigen, teilweise schlecht definierten traditionellen Klassifizierungssystemen,^{91,189} hin zu einer quantitativen und umfassenderen Betrachtung der Spezifität vollzogen.

Die untersuchten WW- und PDZ-Domänen sind aber nicht nur interessante Modellsysteme für die Analyse und Vorhersage von Protein-Protein-Wechselwirkungen, sondern die Interaktionen einiger Vertreter sind auch direkt in die Entstehung einer Reihe klinisch relevanter Krankheiten involviert (siehe 1.3.3 und 1.4.2). Für das Verständnis der Rolle, welche diese Domänen bei der Entstehung von Krankheiten im Speziellen und in zellulären Prozessen im Allgemeinen spielen, ist das Verständnis der Bindungs- und Spezifitäts-Mechanismen essentiell. Dies ist damit auch ein Schritt in Richtung einer gezielten Therapie dieser Erkrank-

ungen, unter anderem durch die Intervention an den Domänen-vermittelten Wechselwirkungen.¹⁵⁷

Neben dem speziellen Verständnis der Wechselwirkungs- und Spezifitätsmechanismen für die untersuchten WW- und PDZ-Domänen wird im Rahmen dieser Arbeit aber auch allgemein das Potential der verschiedenen Vorhersagemodelle evaluiert. Zusammen mit der Definition der Anforderungen zum Training der einzelnen Modelle soll diese Arbeit als Leitfaden für die Anwendung der hier vorgestellten Methoden auf andere Protein-Interaktionsdomänenfamilien dienen.

4.1 Wechselwirkungsmodelle auf unterschiedlichen Informationsebenen

Am Beispiel der WW-Domänen wurden Wechselwirkungsmodelle auf der Sequenz-, der Tertiärstruktur- und der Quartärstrukturebene erstellt. Die verschiedenen Methoden zur Erstellung dieser Modelle stellen unterschiedliche Anforderungen an die zum Training zur Verfügung stehenden experimentellen Daten und setzen unterschiedlich komplexe Ausgangsdaten für die Vorhersage voraus. Gleichzeitig liefern sie aber auch Vorhersagen unterschiedlicher Qualität und Aussage. Im Folgenden werden die Wechselwirkungsmodelle unter anderem auch unter diesem Aspekt diskutiert und verglichen.

4.1.1 Sequenzbasierte Klassifizierung

Das Ziel eines idealen sequenzbasierten Klassifizierungssystems ist es, sowohl von der Sequenz der Domäne als auch der des Liganden ausgehend die Klasse der interagierenden Liganden bzw. Domänen vorherzusagen. Dabei fasst eine Klasse Wechselwirkungen gleicher Spezifität zusammen. Aus diesem Grund wurden solche Klassen in dieser Arbeit *Spezifitätsgruppen* genannt.

Von praktischem Nutzen ist ein derartiges Klassifizierungssystem aber erst dann, wenn für beide Seiten – Liganden und Domänen – jeder Spezifitätsgruppe die charakteristischen, für die spezifische Wechselwirkung verantwortlichen Sequenzmerkmale bekannt sind. Unter diesen Umständen kann die Spezifitätsgruppe des Wechselwirkungspartners vorhergesagt werden, wenn die Sequenzmerkmale einer Spezifitätsgruppe in der Sequenz einer zu untersuchenden Domäne bzw. eines Liganden auftauchen.

Für die untersuchten WW- und PDZ-Domänenfamilien existierten bereits ansatzweise solche Klassifizierungssysteme.^{170,177} Für beide Domänenfamilien sind die Klassen also fast ausschließlich durch die Erkennungsmotive in den Liganden definiert und die Zuordnung der Domänen ist nicht immer eindeutig.^{88,91,189} Die Sequenzmerkmale der Domänen, die für die Spezifität verantwortlich sind, sind in den meisten Fällen jedoch unklar. In der Vergangenheit wurden die Domänen daher vorrangig an Hand der oft widersprüchlichen und unvollständigen experimentellen Ergebnissen einer Klasse zugeordnet.^{19,48,91}

Klasseneinteilung der WW- und PDZ-Domänen

In dieser Arbeit wurde das Klassifizierungssystem der WW-Domänen erweitert und vor allem durch charakteristische Sequenzmotive der Domänen ergänzt (siehe 3.2.3.1-3). Hierfür wurde eine repräsentative Auswahl an 42 WW-Domänen mit einer repräsentativen Auswahl an 9 potentiellen Liganden auf Wechselwirkungen untersucht (WW-Screening, siehe 3.1). Die identifizierten Spezifitätsgruppen wurden anschließend durch Erkennungsmotive in den Liganden bekräftigt, die für die einzelnen Spezifitätsgruppen charakteristisch und konserviert sind (siehe 3.2.1).

Experimentell basierte Spezifitätsgruppen der WW-Domänen

Die Y- bzw. poS/poT-Spezifitätsgruppen dieses Klassifizierungssystems entsprechen eindeutig den beiden gut definierten Gruppen I und IV der traditionellen Klassifizierung der WW-Domänen.¹⁷⁷ Die WW-Domänen der weit weniger gut definierten Gruppen II und III⁹¹ wurden auf insgesamt vier, die R_a-, R_b-, L- und poly-P-Spezifitätsgruppen aufgeteilt. Dadurch konnten WW-Domänen, wie beispielsweise hFE65-WW, die in einer Studie in Gruppe II⁴⁸ und in einer anderen in Gruppe III¹⁹ klassifiziert wurde, eindeutig einer Spezifitätsgruppe zugeordnet werden.

Verhältnis der traditionellen WW-Domänen-Klassen zu den Spezifitätsgruppen

Anschließend wurde nach den spezifitätsbestimmenden Resten in den Domänensequenzen einer jeden Spezifitätsgruppe mit Hilfe von Sequenzlogos, Konsensussequenzen und *Profile-Hidden-Markov-Modellen* (pHMM) gesucht. Dadurch konnten plausible Motive für die Y-, poS/poT- sowie für die beiden neuen R-Spezifitätsgruppen bestimmt werden (siehe 3.2.3.1). Für die Y- und poS/poT-Spezifitätsgruppen ließen sich die identifizierten Domänen-Epitope an Hand der Strukturen der jeweiligen WW-Domänen/Liganden-Komplexe erklären.^{113,193} Damit konnte gezeigt werden, dass durch diese Herangehensweise die spezifitätsbestimmenden Motive prinzipiell identifiziert werden können. Unterstützt wird die Herangehensweise auch dadurch, dass die innerhalb der Spezifitätsgruppen konservierten Reste vorwiegend auf der zum Liganden exponierten Seite des β -Faltblatts liegen und damit vermutlich eine funktionelle Bedeutung besitzen (Tabelle 3-6).

Spezifitätsbestimmende Reste in den Domänen

Erkennungsmotive in argininspezifischen WW-Domänen – Indizien für zwei verschiedene Spezifitätsgruppen

Über die Domänen-Epitope der argininspezifischen WW-Domänen war bislang nichts bekannt. Für diese WW-Domänen konnten in dieser Arbeit ebenfalls in sich konsistente und plausible Erkennungsmotive vorgeschlagen werden: demnach zeichnen sich argininspezifische WW-Domänen vor allem durch konservierte negativ geladene Aminosäuren in der $\beta\alpha\beta$ -Schleife aus (Abbildung 3-12). Auf Grund dieser zum positiv geladenen Arginin komplementären Ladung in der Domäne erscheinen die Epitope plausibel. Gleichzeitig korreliert die konservierte aromatische Aminosäure an $\omega 30$, welche die hydrophobe Oberfläche zur Wechselwirkung mit dem Liganden vergrößert, mit den im Vergleich zur Y-Spezifitätsgruppe längeren Erkennungsmotiven in den argininhaltigen Liganden (7-8 Reste). Bemerkenswerterweise wurden auch zwei prinzipiell ähnliche aber im Detail unterschiedliche Erkennungsmotiven in den Domänen identifiziert, die mit den beiden experimentell definierten argininspezifischen Spezifitätsgruppen korrelieren. Dies untermauert erneut den Befund zwei verschiedener R-Spezifitätsgruppen.

Bestätigung der Y-, R_a- und R_b-Spezifitätsgruppen durch Sequenzverwandtschaftsbaum

Unabhängig von den experimentell bestimmten Spezifitätsgruppen wurden die WW-Domänen auch ausschließlich an Hand ihrer Sequenzen mit Hilfe eines Sequenzverwandtschaftsbaums gruppiert (siehe 3.2.3.2). Sowohl die Y- als auch die beiden R-Spezifitätsgruppen spiegelten sich in klar abgegrenzten Ästen des Baums wieder (Abbildung 3-13). Diese drei Gruppen konnten auch mit anderen Berechnungsverfahren für Sequenzverwandtschaftsbäume reproduziert werden (Daten nicht gezeigt). Dadurch wird belegt, dass sich die Sequenzen der Y- und der beiden R-Spezifitätsgruppen in der Tat signifikant voneinander und von den restlichen WW-Domänen unterscheiden. Damit wird erneut die Aufteilung in zwei argininspezifische Gruppen bestätigt und gleichzeitig die Konsensussequenzen der beiden R-Spezifitätsgruppen, die nur auf wenigen Sequenzen basieren, unterstützt.

Vorhersage des Wechselwirkungsmodells der Tyrosin-Spezifität

Die Vorhersageleistung des sequenzbasierten Klassifizierungssystems der WW-Domänen wurde an Hand der Tyrosin-Spezifität untersucht. Die Sequenzmerkmale der Y-Spezifitätsgruppe wurden dazu in Form eines einfachen Konsensusmotiv, eines pHMMs der Y-Spezifitätsgruppe sowie eines pHMMs des Y-Asts zusammengefasst (Abbildung 3-14). Die Klassifizierung der 42 experimentell untersuchten WW-Domänen bestätigte, dass diejenigen Sequenzen, mit denen die pHMMs trainiert worden waren, auch die höchsten Homologiewerte zeigten, und damit als tyrosinspezifisch klassifiziert wurden (siehe Anhang A, Tabelle A-1).

Vorhersageleistung des Modells der Tyrosin-Spezifität

Mit Hilfe des Konsensusmotivs und den beiden pHMMs wurden anschließend alle 482 zu diesem Zeitpunkt bekannten WW-Domänensequenzen in tyrosinspezifische und nicht tyrosinspezifische WW-Domänen klassifiziert (Tabelle 3-7). Erstaunlicherweise wurden

zwischen einem Drittel und über der Hälfte aller WW-Domänen als tyrosinspezifisch klassifiziert. Dabei korrelierte die vorhergesagte Tyrosin-Spezifität stark mit den Proteinfamilien, in denen diese WW-Domänen vorkommen. Demnach wird Tyrosin-Spezifität für beinahe alle WW-Domänen aus Proteinen der NEDD4-Ubiquitin-Ligase Familie, der Membran-Assoziierten-Guanylat-Kinase Familie, der Dystrophin-Familie und der YAP-Familie vorhergesagt. Da für einzelne WW-Domänen aus diesen Familien schon experimentell die Tyrosin-Spezifität gezeigt worden war (Tabelle 3-6),¹⁷⁷ kann dies als Indiz für eine korrekte Klassifizierung gewertet werden.

Folglich scheint die Wechselwirkung mit tyrosinhaltigen prolinreichen Liganden auch eine konservierte Eigenschaft der WW-Domänen dieser Proteinfamilien zu sein. Diese Vorhersage wird für die Familie der NEDD4-Ubiquitin-Ligasen durch eine ganze Reihe von *in vitro* und *in vivo* Analysen der Wechselwirkungen dieser Proteine bzw. der individuellen WW-Domänen bestätigt: demnach sind für die Interaktion mit ENaC,⁸⁷ SMADs,⁹² RNA-Polymerase II,³⁵ NF-E2,⁵⁸ Atrophin-1²⁰⁴ und Occludin¹⁸⁵ prolin- und tyrosinhaltige PY-Motive essentiell. Dies belegt somit die *in vivo* Relevanz der vorhergesagten Tyrosin-Spezifität. Interessant ist hierbei auch, dass in Proteinen mit tandemwiederholten multiplen WW-Domänen für alle gleichzeitig entweder Tyrosin-Spezifität oder Nicht-Tyrosin-Spezifität vorhergesagt wurde. Damit scheint eine Aufgabe der tandemwiederholten Domänen die Erhöhung der makroskopischen Affinität⁶² bzw. die Selektion verschiedener Liganden innerhalb der Y-Spezifitätsgruppe und nicht die Interaktion mit grundsätzlich verschiedenen Liganden aus unterschiedlichen Spezifitätsgruppen zu sein. Schließlich verdeutlichen diese Ergebnisse auch, dass sich die scheinbare Überrepräsentation tyrosinspezifischer WW-Domänen unter den 42 WW-Domänen auf eine hohe natürliche Häufigkeit der tyrosinspezifischen WW-Domänen zurückzuführen lässt.

In vivo Relevanz der vorhergesagten Tyrosin-Spezifität

Im Vergleich zum pHMM des Y-Asts scheint das pHMM der Y-Spezifitätsgruppe besser zwischen tyrosinspezifischen und nicht tyrosinspezifischen WW-Domänen zu unterscheiden (Tabelle 3-8). Darauf weisen sowohl die höhere Korrelation mit den Proteinfamilien als auch die höhere Korrelation mit dem Auftreten des Kernkonsensusmotivs der Y-Spezifitätsgruppe hin. Nichts desto trotz werden auch durch das pHMM der Y-Spezifitätsgruppe WW-Domänen erfasst, die den tyrosinspezifischen Domänen sehr nahe stehen.

Vergleich der verschiedenen sequenzbasierten Modelle

Die Bedeutung dieser Domänen, die durch das pHMM der Y-Spezifitätsgruppe in den Übergangsbereich zwischen Tyrosin- und Nicht-Tyrosin-Spezifität klassifiziert werden (Tabelle 3-7A), wird unter anderem durch die Familie der WW-Domänen enthaltenden

Übergangsbereich zwischen Tyrosin-Spezifität und Nicht-Tyrosin-Spezifität

Oxidoreduktasen (WWOX) verdeutlicht. Alle WW-Domänen dieser Familie weichen in mindestens einer Position vom Kernkonsensusmotiv der Y-Spezifitätsgruppe ab (Tabelle 3-6), weshalb sie auch in den Übergangsbereich klassifiziert werden. Es gibt jedoch eine ganze Reihe von experimentellen Hinweisen darauf, dass sie *in vivo* für die Interaktion mit Zielproteinen über **xPPxY**-Motive verantwortlich sind.^{11,111} *tbZFP2*-WW ist das umgekehrte Beispiel: die Domäne erfüllt zwar das Kernkonsensusmotiv der Y-Spezifitätsgruppe (Tabelle 2-1), wird aber dennoch durch das pHMM der Y-Spezifitätsgruppe – auf Grund der Sequenzabweichungen an anderen Positionen – in den Übergangsbereich klassifiziert. In Substitutionsanalysen konnte dann auch für diese WW-Domäne die Wechselwirkung mit dem Erkennungsmotiv **xPPxY** im Liganden nachgewiesen werden (Abbildung 3-15E). Diese Substitutionsanalyse zeigt aber zugleich, dass *tbZFP2*-WW auch mit der Tyrosin→Arginin-Mutante des Y-Liganden mit gleicher, wenn nicht höherer Affinität interagieren kann. Damit ist diese Domäne exemplarisch für den Übergang von einer Spezifitätsgruppe zu einer anderen. Die Analyse der Domänensequenz macht sofort klar, warum *tbZFP2*-WW sowohl tyrosin- als auch argininhaltige Liganden bindet: parallel zum Kernkonsensusmotiv der Y-Spezifitätsgruppe besitzt sie auch eine negativ geladene Aminosäure in der $\beta\text{A}\beta\text{B}$ -Schleife und kein Arginin, sondern im Gegenteil ein Glutamat, an Position $\omega 26$. Im Vergleich zur Konsensussequenz der WW-Domänen der R-Spezifitätsgruppen fehlt also nur eine aromatische Aminosäure an $\omega 30$.

Limitierung der
Klassifizierungssysteme

Damit wird zugleich eine maßgebliche Limitierung solcher Klassifizierungssysteme deutlich: eine Klassifizierung ist immer ein künstliches System, das „harte“ Grenzen zwischen den Klassen definiert. Wo die Grenzen gezogen werden, ist dabei relativ willkürlich. So wurden in dieser Arbeit WW-Domänen, die nur mit tyrosinhaltigen Liganden interagieren, mit solchen, die sowohl mit tyrosin- als auch phosphotyrosinhaltigen Liganden interagieren können, in eine Spezifitätsgruppe klassifiziert (siehe 3.2.2). Im Gegensatz dazu wurden WW-Domänen die mit dem R-Liganden interagieren zwei verschiedenen R-Spezifitätsgruppen zugeordnet (siehe 3.2.3.1). Letztendlich muss also immer ein Kompromiss zwischen Generalisierung (alle WW-Domänen interagieren mit prolinhaltigen Liganden) und Individualisierung (z.B. *hNEDD4*-WW1 zeigt Unterschiede zu *hNEDD4*-WW3) gewählt werden. Dabei sollte das Hauptaugenmerk auf der Anwendung des Klassifizierungssystems zur Vorhersage der Spezifität im Vordergrund stehen. In diesem Zusammenhang wurde in letzter Zeit in der Literatur über eine Zusammenführung der Gruppe II und III diskutiert.^{91,202} Die vorgeschlagene Zusammenführung wird unter anderem damit begründet, dass die Unter-

schiede in den Ligandenpräferenzen gering und die beiden Gruppen bislang nur unzureichend charakterisiert sind. Beide Gründe sind natürlich keine zulässigen Argumente dafür, dass auf eine Unterscheidung zwischen arginin-, leuzin- und polyprolinspezifischen WW-Domänen verzichtet werden sollte, auch wenn die Unterschiede nur gering sind. Tatsächlich konnten in dieser Arbeit auch relativ klare Ligandenpräferenzen für die WW-Domänen der Gruppe II und III identifiziert werden. Dabei demonstriert die Wechselwirkung des poly-P-Liganden mit WW-Domänen der R_{α} -Spezifitätsgruppe nur, dass sich eine Spezifitätsgruppe nicht immer durch ein einziges Liganden-Erkennungsmotiv beschreiben lässt. Schließlich unterscheiden sich die Domänen der poly-P-Spezifitätsgruppe von denen der R_{α} -Spezifitätsgruppe klar durch ihre Unfähigkeit zur Wechselwirkung mit argininhaltigen Liganden. Letzten Endes spricht die Möglichkeit, auch geringe – möglicherweise *in vivo* relevante – Präferenz zu unterscheiden, eindeutig gegen eine Zusammenführung der Gruppen II und III. Die Problematik der Klassifizierung der WW-Domänen auf Grund publizierter Bindungsdaten macht zugleich auch deutlich, dass die Kriterien zur Klassifizierung konsistent und umfassend sein müssen. Es ist sehr wahrscheinlich, dass sich das Klassifizierungssystem in Folge der Entdeckung zusätzlicher Ligandentypen weiter verändern wird.

Unabhängig von diesen formalen Problemen der Klassifizierung, existieren für eine Prote-

Affinität der Wechselwirkung als Selektivitätskriterium

in-Interaktionsdomänenfamilie nicht zwangsläufig nicht-überlappende Spezifitätsgruppen. Die Wechselwirkung des poly-P-Liganden sowohl mit der poly-P- als auch der R_{α} -Spezifitätsgruppe ist ein Beispiel dafür. Die „Dual-Spezifität“ vieler PDZ-Domänen, d.h. die Wechselwirkung mit Liganden der Spezifitätsgruppen I und II, ist ein weiteres Beispiel (siehe 1.5.1 und 3.4). Die „weichen“ Übergänge zwischen den „Spezifitätsgruppen“ können erst durch die Modellierung des quantitativen Zusammenhangs zwischen Sequenz und Affinität hinlänglich beschrieben werden. Dies wurde mit der Entwicklung der Methode der *Quantitativen Spezifitäts-Profile* verfolgt und am Beispiel der PDZ-Domänen demonstriert (siehe 4.3 und 3.5).

Das bedeutet gleichzeitig, dass die einfachen Sequenzmotive zur Beschreibung der Spezifität auf Grund ihrer alles oder nichts Aussage nicht besonders geeignet sind. Besser sind Profilmodelle, wie Sequenzlogos und pHMMs. Schließlich zeigte das pHMM der Y-Spezifitätsgruppe sogar eine gewisse Unterscheidung der „stark“ tyrosinspezifischen und der „schwach“ tyrosinspezifischen WW-Domänen (siehe 3.2.3.3). Jedoch werden auch durch Sequenzprofile und pHMMs einige Sequenzmerkmale nicht repräsentiert. Vor allem Aminosäuren, die an einer Position *nicht* auftreten dürfen, werden nicht bzw. nur sehr schlecht er-

Beschreibung der spezifitätsbestimmenden Sequenzmerkmale

Sequenzmotive versus Sequenzprofile(pHMM)

fasst. Deutlich wurde diese Limitierung an der Unfähigkeit, auf Sequenzebene die Position $\omega 26$ als ausschlaggebend für WW-Domänen der R-Spezifitätsgruppen zu identifizieren.

Neben der Klassifizierung bzw. Vorhersage der Spezifität können die Erkennungsmotive in den Domänen und den Liganden auch einen Einblick in die Wechselwirkungs- und Spezifitäts-Mechanismen geben (4.2). Erst auf Grund der Kenntnis der Erkennungsmotive konnten in dieser Arbeit Strukturmodelle der WW-Domänen/Liganden-Komplexe modelliert werden (siehe 3.3.1). Letztendlich bauen also auch die tertiär- und quartärstrukturbasierten Methoden auf den sequenzbasierten Methoden auf.

4.1.2 Strukturbasierte Wechselwirkungsmodelle

Mit strukturbasierten Wechselwirkungsmodellen wird die Vorhersage der Spezifität im Vergleich zu den sequenzbasierten Modellen in zwei Richtungen erweitert: zum einen wird die relative räumliche Anordnung der Aminosäuren im dreidimensionalen Raum berücksichtigt. Diese ist für die Form und Eigenschaften der Moleküloberfläche verantwortlich, welche letztendlich die Spezifität der Wechselwirkung bestimmen. Auf diese Weise wird mit Hilfe strukturbasierter Wechselwirkungsmodelle ein strukturelles Verständnis der Spezifitäts-Mechanismen entwickelt. Zum anderen wird ein quantitativer Zusammenhang zwischen der Struktur und der Aktivität bzw. Affinität der Wechselwirkung modelliert. Durch die daraus resultierende Affinitätsvorhersage kann letztendlich die Problemstellung des „weichen Übergangs“ zwischen Spezifitätsgruppen oder der Selektivität innerhalb einer Spezifitätsgruppe gelöst werden.

Zur Modellierung des Struktur-Aktivitätszusammenhangs mit Hilfe der *Comparative-Molecular-Field-Analysis* (CoMFA)⁴² werden die Raumstrukturen der zu analysierenden Moleküle durch sogenannte molekulare Eigenschaftsfelder beschrieben. Für das in dieser Arbeit entwickelte CoMFA-Modell waren dies sterische und elektrostatische Felder, welche die sterischen und elektrostatischen Eigenschaften der Strukturen an definierten Punkten eines 3D-Raumgitters repräsentieren. Diese multivariate Beschreibung der Strukturen wird dann mittels einer PLS-Regression mit der quantitativen Aktivität in Beziehung gesetzt (Anhang D).

In klassischen quantitativen Struktur-Aktivitäts-Modellen werden die Strukturmerkmale unterschiedlicher niedermolekularer Liganden in Beziehung zur Aktivität, meist der Bindungsstärke gegenüber einem Rezeptor, gesetzt.⁴ Mit dem Struktur-Aktivitäts-Modell dieser Arbeit wird die Sichtweise umgekehrt: mit seiner Hilfe wird der Zusammenhang zwischen den Strukturmerkmalen unterschiedlicher „Rezeptoren“ respektive WW-Domänen und der

Verständnis der Wechselwirkungs- und Spezifitäts-Mechanismen

Ziele der strukturbasierten Vorhersage

Comparative-Molecular-Field-Analysis (CoMFA)

Umkehrung der klassischen Sichtweise: QSAR des Rezeptor

Wechselwirkung mit einem definierten Liganden (Y-Modellliganden) modelliert (siehe 3.3.2).

Allerdings sind bislang nur für die wenigsten WW-Domänen Raumstrukturen experimentell bestimmt worden. Aus diesem Grund basieren die CoMFA-Modelle dieser Arbeit ausschließlich auf modellierten Strukturen der WW-Domänen (siehe 3.3.1). Damit können die CoMFA-Modelle im Prinzip als „dreidimensionale Sequenzalignments“ angesehen werden, die zusätzlich zur Domänensequenz Informationen über die ungefähre relative räumliche Anordnung der Aminosäuren enthalten. Bei der Interpretation der vorgestellten CoMFA-Modelle müssen daher die Ungenauigkeiten der Strukturmodelle mit berücksichtigt werden, um eine „Überinterpretation“ zu vermeiden.

Strukturmodelle als Basis für struktur-basierte Wechselwirkungsmodelle

3D-Sequenzalignment

Gegen die alleinige Analyse der vorhandenen experimentell bestimmten Strukturen sprach zum einen die zu geringe Anzahl an verfügbaren Strukturen. Zum anderen hatten Voruntersuchungen zu quantitativen Struktur-Aktivitäts-Modellen auf der Basis der wenigen experimentell bestimmten Strukturen erhebliche Probleme gezeigt. Inhomogenitäten, die auf Grund der unterschiedlichen Strukturbestimmungsmethoden und der unterschiedlichen Qualität der verwendeten Strukturen auftraten, führten dazu, dass lediglich die „Herkunft“ der Struktur und nicht ihre Aktivität modelliert wurde (Daten nicht gezeigt). Dieselben Argumente sprachen auch gegen eine Mischung von experimentell bestimmten und modellierten Domänenstrukturen.

Alternative Analyse der experimentell bestimmten Strukturen

Die Raumstrukturen der 42 WW-Domänen wurden auf der Basis einer einzigen Strukturvorlage [PDB: 1PIN] modelliert (siehe 3.3.1.1). Diese Strukturvorlage wurde ausgewählt, da sie die höchste durchschnittliche Sequenzidentität zu den Sequenzen aller zu modellierenden WW-Domänen besaß und gleichzeitig die am besten aufgelöste Röntgenkristallstruktur einer WW-Domäne war. Gegen die Verwendung unterschiedlicher Strukturvorlagen sprachen dieselben Argumente, die auch schon gegen die Mischung experimenteller und modellierter Strukturen angeführt wurden. Diese Herangehensweise wird durch eine Studie von Winn *et al.*, 2005, zur Hochdurchsatzmodellierung homologer Strukturen bestätigt. In ihrer Studie demonstrieren Winn *et al.*, dass die Verwendung mehrerer Strukturvorlagen nicht zwangsläufig zu besseren Strukturmodellen führt. Im Gegenteil, sie stellten dieselben, im letzten Absatz beschriebenen Probleme beim statistischen Vergleich der Strukturmodelle fest, die mit verschiedenen Strukturvorlagen modelliert worden waren. Die Autoren schlagen schließlich dieselben Kriterien zur Auswahl einer einzelnen Strukturvorlage vor, die auch in dieser Arbeit verwendet wurden.

Modellierungs-prozedur der Domänenstrukturen

Qualität der
Strukturmodelle

Die Qualität der Strukturmodelle wurde an Hand der Abweichungen der modellierten von experimentell bestimmten Strukturen evaluiert (siehe 3.3.1.2). Es stellte sich heraus, dass diese nicht deutlich größer sind als die Abweichungen der experimentellen Strukturen untereinander (Tabelle 3-9), die auf methodische Ungenauigkeiten zurückzuführen sind. Bemerkenswert ist insbesondere die gute Qualität der Strukturmodelle der WW-Domänen der Y-Spezifitätsgruppe, trotz der Modellierung auf der Basis einer Strukturvorlage [PDB: 1PIN] aus einer anderen Spezifitätsgruppe ($\rho\text{S}/\rho\text{T}$). Dies bestärkt die Annahme, dass auf diese Weise auch für diejenigen Spezifitätsgruppen verwertbare Strukturmodelle erhalten werden konnten, für die zur Überprüfung der Qualität keine experimentell bestimmten Strukturen zur Verfügung standen.

CoMFA-Modell der
Tyrosin-Spezifität

Auf der Basis dieser Strukturmodelle und des WW-Screenings der 42 WW-Domänen wurde ein CoMFA-Modell der Tyrosin-Spezifität erstellt (siehe 3.3.2). Dazu wurden sowohl die Strukturmodelle der WW-Domänen der Y-Spezifitätsgruppe als auch repräsentative Vertreter der anderen Spezifitätsgruppen ausgewählt (Tabelle 3-6) und mit Hilfe der CoMFA-Methode beschrieben. Durch diese Auswahl wurde das gesamte Spektrum der experimentell bestimmten Aktivitäten abgedeckt. Als Aktivität wurde die Veränderung der chemischen Verschiebung (CSP) des Wasserstoffatoms am Indolstickstoff von $\text{W}\omega\text{39}$ nach Bindung des Y-Modellliganden, wie sie durch das NMR-Screening bestimmt worden war (Tabelle 3-3), verwendet. Für die Interpretation des CoMFA-Modells wurde ein Zusammenhang zwischen dieser gemessenen Aktivität der WW-Domänen und der zugrundeliegenden Affinität der Wechselwirkung mit tyrosinhaltigen Peptiden angenommen. Diese Annahme wird sowohl durch die Korrelation zwischen den CSPs und den Signalintensitäten der Screenings des WW-Arrays als auch durch die Korrelation der CSPs und den durch UV-Fluoreszenzspektroskopie bestimmten Dissoziationskonstanten unterstützt (siehe Anhang C).

Fehlerquellen der
QSAR-Modelle, die
auf modellierten
Strukturen basieren

Das resultierende CoMFA-Modell zeigte ein im Vergleich zu gewöhnlichen CoMFA-Modellen für niedermolekulare Liganden stärkeres Rauschen (Abbildung 3-24). Dies ist im wesentlichen auf zwei Faktoren zurückzuführen: (i) Das CoMFA-Modell basiert auf Strukturmodellen der WW-Domänen, die einer modellierungsbedingten Ungenauigkeit unterliegen. (ii) Die Strukturen der WW-Domänen sind wesentlich größer als gewöhnliche niedermolekulare Liganden. Da die Anzahl der Deskriptoren mit der 3. Potenz der Steigerung der Molekülgröße ansteigt, müssen eine Vielzahl mehr Deskriptoren durch die statistischen Wechselwirkungsmodelle bewältigt werden.

Trotz der vielen potentiellen Fehlerquellen ergab die CoMFA-Methode ein aussagekräftiges Modell der Tyrosin-Spezifität (Abbildung 3-23). Bestätigt wird dies unter anderem dadurch, dass sich ein Großteil der Strukturmerkmale mit hohem Beitrag im CoMFA-Modell mit den konservierten Sequenzmerkmalen der Y-Spezifitätsgruppe deckt (Abbildung 3-24). Über die reine Wiedergabe der konservierten Sequenzmerkmale hinaus gestattet das CoMFA-Modell aber auch deren strukturelle Interpretation. Demnach bilden die Seitenketten von H ω 32 und R/K/Q ω 35 den Rand sowie V/I/L ω 30 den hydrophoben Boden der Tyrosin-Bindungsstelle (Abbildung 3-24, Region A). Ein ausgeprägter Rand und keine unvorteilhaften Interaktionen mit der Aminosäure in ω 30 sind die Voraussetzungen für eine hoch affine Bindung tyrosinhaltiger Liganden. Erwartungsgemäß wurden durch das elektrostatische Feld in diesem Bereich sowohl eine vorteilhafte positive Ladung um R/K/Q ω 35 als auch eine vorteilhafte negative Ladung um E ω 18 identifiziert (Abbildung 3-24, Region A und C). Der Einfluss einer möglichen Ladung von H ω 32 konnte mit diesem CoMFA-Modell nicht analysiert werden, da der Protonierungszustand und damit die Ladung dieses Histidin unbekannt waren. Auf Grund der Entfernung bleibt aber auch der direkte Zusammenhang mit der Ligandenbindung für E ω 18 unklar. Und obwohl für die positive Ladung an ω 35 eine vorteilhafte elektrostatische Wechselwirkung mit der negativen Partialladung der Hydroxylgruppe des Tyrosins denkbar ist, lässt sich die Unterscheidung zwischen reiner Korrelation und Kausalität allein an Hand des CoMFA-Modells nicht treffen. In einer Analyse zu Varianten der YAP65-WW Domäne zeigten Toepert *et al.*, 2001, dass an ω 35 neben R/K/Q auch M/H/A/G eine Wechselwirkung mit tyrosinhaltigen Liganden ermöglichen, wenn auch teilweise mit geringerer Affinität.¹⁸³ Damit scheint eine basische Gruppe an dieser Position vorteilhaft, aber eben nicht zwingend notwendig für eine Wechselwirkung mit tyrosinhaltigen Liganden zu sein. Daneben ist aber auch denkbar, dass sich R/K/Q durch die hydrophile funktionale Gruppe besser in Richtung des wässrigen Lösungsmittels „aufrichten“ können als hydrophobe Aminosäuren, und dadurch den für die Bindung des Tyrosins nötigen hydrophoben Teil ihrer Seitenkette zugänglich machen.

Spezifitätsbestimmende
Strukturmerkmale der
Domänen mit direktem
Bezug zur Sequenz

Darüber hinaus wird durch das CoMFA-Modell ein Zusammenhang zwischen der Aktivität und Unterschieden in den Strukturen vorgeschlagen, die sich nicht auf einfache Unterschiede in den Sequenzen zurückführen lassen. Dazu zählen vor allem der Einfluss der Konformation der konservierten Prolin-Bindungsstelle, sowie der Konformation der β A β B-Schleife (Abbildung 3-24, Region B). Allerdings zeigen die meisten NMR-Strukturen von WW-Domänen, dass die β A β B-Schleife relativ flexibel ist, und somit keine Aussage über eine kon-

Spezifitätsbestimmende
Strukturmerkmale der
Domänen ohne
direkten Bezug zu
Sequenz

krete Konformation getroffen werden kann. Für *hFBP11-WW1* wurde sogar gezeigt, dass sich die Konformation der $\beta\text{A}\beta\text{B}$ -Schleife zwischen freier und gebundener Domäne unterscheidet.¹³⁷ Dynamische Eigenschaften werden aber in einem CoMFA-Modell nicht repräsentiert. Zusätzlich basiert dieses CoMFA-Modell nur auf modellierten Strukturen, und es könnte sich damit auch um Modellierungsartefakte handeln. Daher muss ungeachtet der Plausibilität der Einfluss der $\beta\text{A}\beta\text{B}$ -Schleife und der Konformation der konservierten Prolin-Bindungsstelle durch weitere experimentelle Analysen bestätigt werden.

Vorhersage der
Aktivität

Die Vorhersageleistung des CoMFA-Modells der Tyrosin-Spezifität wurde an Hand der drei bislang experimentell nicht untersuchten WW-Domänen *hBAG3-WW*, *hSYNIP-WW* und *tbZFP2-WW* beurteilt (siehe 3.3.2.2). Die vorhergesagte Wechselwirkung mit tyrosinhaltigen Liganden konnte für alle drei WW-Domänen experimentell durch Substitutionsanalysen bestätigt werden (Abbildung 3-15). Die Tatsache, dass die *tbZFP2-WW* Domäne, für welche das CoMFA-Modell die schwächste Aktivität gegenüber tyrosinhaltigen Liganden vorher-sagt, auch mit argininhaltigen Liganden interagiert, bestätigt die Aussagekraft der vorhergesagten Wechselwirkungsaktivität.

Limitierung der
CoMFA-Modelle

Allerdings erlaubt ein CoMFA-Modell nur die Vorhersage der Affinität gegenüber dem-jenigen Liganden, auf dessen experimentellen Daten das CoMFA-Modell trainiert wurde. Um also für eine experimentell nicht untersuchte WW-Domäne ein umfassendes Spektrum der interagierenden Liganden vorherzusagen, wird – in Analogie zu den Spezifitätsgruppen – für jeden potentiellen Liganden(-typ) jeweils ein CoMFA-Modell benötigt.

Fallstudie für Protein-
Interaktionsdomänen-
familien mit wenig
struktureller
Information

Darüber hinaus stellt die Forderung nach der Kenntnis der Raumstrukturen ein großes Hindernis für die Erstellung und Anwendung eines CoMFA-Modells dar. Das hier vorge-stellte CoMFA-Modell demonstriert aber, dass sowohl für das Training als auch für die Vorhersage Strukturmodelle ausreichen können. Voraussetzung dafür sind allerdings eine hoch konservierte Struktur und ein hoch konservierter Liganden-Bindungsmechanismus. Somit können die Vorteile dieser strukturbasierten Methode auch für Domänenfamilien ausgenutzt werden, für die nur wenig strukturelle Information zur Verfügung steht.

4.1.3 Komplexstrukturbasierte Wechselwirkungsmodelle

Die komplexstrukturbasierten *Comparative-Binding-Energy-Modelle* (COMBINE) erweitern die Analyse der Wechselwirkungen um die strukturelle Information über den Liganden im Komplex mit der Domäne. Durch COMBINE-Modelle können folglich essentielle Domä-nen/Liganden-Interaktionen identifiziert werden, die nur an Hand der Komplexstruktur

evident werden, im Gegensatz zu strukturbasierten CoMFA-Modellen, die nur Informationen über die Struktur der Domänen enthalten. Mit der Einbeziehung des Liganden in das Wechselwirkungsmodell geht gleichzeitig eine erweiterte Vorhersagefähigkeit einher. So kann mit Hilfe von COMBINE-Modellen die Affinität der Wechselwirkung auch für WW-Domänen/Liganden-Komplexe vorhergesagt werden, die weder in der Domäne noch im Liganden mit den Komplexen übereinstimmen, auf denen das Modell trainiert worden war. Im Gegensatz dazu kann mit einem einzelnen CoMFA-Modell immer nur die Affinität einer Domäne gegenüber demjenigen Liganden vorhergesagt werden, auf dessen experimentellen Daten das CoMFA-Modell trainiert wurde. Damit generalisieren COMBINE-Modelle die Trainingsdaten prinzipiell stärker als CoMFA-Modelle.

Ziele der strukturbasierten Vorhersage

Zur Modellierung des Zusammenhangs zwischen Komplexstruktur und Aktivität mit Hilfe der COMBINE-Methode¹²⁷ werden die Komplexe durch paarweise Interaktionsenergien zwischen allen Liganden- und allen Domänenresten beschrieben. Dabei wird für jede paarweise Interaktion zwischen elektrostatischen und van-der-Waals Interaktionsenergietermen unterschieden. Diese multivariate Beschreibung der Interaktionsenergien der Domänen/Liganden-Komplexe wird anschließend mittels einer PLS-Regression (siehe Anhang D) mit der quantitativen Aktivität in Beziehung gesetzt. Dadurch werden diejenigen paarweisen Interaktionen identifiziert, die mit einer erhöhten bzw. erniedrigten Aktivität korrelieren.

Comparative-Binding-Energy-Analyse (COMBINE)

Allerdings sind experimentell bestimmte Strukturen der WW-Domänen/Liganden-Komplexe noch rarer als Strukturen der Domänen alleine. Nur für die Y- und die poS/poT-Spezifitätsgruppen sind bislang Komplexstrukturen bestimmt worden. Daher wurden die COMBINE-Modelle, wie auch schon das CoMFA-Modell, ausschließlich auf modellierten Strukturen trainiert.

Strukturmodelle der WW-Domänen/Liganden-Komplexe als Basis für komplexstruktur-basierte Wechselwirkungsmodelle

Die Komplexstrukturen wurden dazu auf der Basis der Analyse aller experimentell bestimmter WW-Domänen/Liganden-Komplexstrukturen sowie den experimentellen Daten des WW-Screenings und der Substitutionsanalysen modelliert (siehe VW3.1 und 3.2.1). Als Strukturen der WW-Domänen wurden dieselben Strukturmodelle wie für die CoMFA-Modelle verwendet (siehe 3.3.1.1). Für die prolinreichen Liganden wurde eine PPII-Helix-Konformation angenommen, da bekannt ist, dass prolinreiche Proteinabschnitte tendenziell eine PPII-Helix-Konformation annehmen.²⁰³ Dies konnte zumindest für den Kernbereich der Liganden ($\lambda-2$ bis $\lambda 2$) im Komplex mit WW-Domänen durch die Analyse der experimentell bestimmten Komplexstrukturen bestätigt werden (Abbildung 3-17). Für die flankierenden Bereiche zeigte diese Analyse allerdings keine bevorzugte Konformation. Selbst ähnliche

Domänen- und Ligandenstruktur-Modellierungs-prozedur

Liganden unterschieden sich in der Konformation der flankierenden Bereiche teilweise deutlich. Aus diesem Grund wurde auch für die flankierenden Bereiche vereinfachend von einer PPII-Helix ausgegangen.

Manuelles Docking der Strukturmodelle der Liganden auf die Strukturmodelle der WW-Domänen

Die Analyse der experimentell bestimmten Komplexstrukturen hat ebenfalls eine hoch konservierte Positionierung der Ligandenreste im Kernbereich relativ zur konservierten Prolin-Bindungsstelle der WW-Domänen ergeben (Tabelle 3-10). Die modellierten Liganden wurden an Hand dieser durchschnittlichen relativen Positionierung an die konservierte Prolin-Bindungsstelle der Strukturmodelle der WW-Domänen platziert. Die verschiedenen, entlang der Helixachse verschobenen Positionierungen der Liganden wurden anschließend mit den entsprechenden Substitutionsanalysen verglichen. Dabei wurde versucht die in den Substitutionsanalysen als Schlüsselreste identifizierten Ligandenreste mit der Domäne in Kontakt zu bringen (siehe 3.3.1.4). Eines der Kriterien zur Positionierung des Liganden wurde allerdings, zumindest für *hFBP11-WW1*, durch die experimentell bestimmte Struktur [PDB: 1YWJ]¹³⁷ des Komplexes mit dem Peptid aus Formin widerlegt:¹³⁷ in diesem Komplex sind nämlich nicht die beiden konsekutiven, hochspezifisch erkannten Proline des Liganden (Abbildung 3-7), sondern das **LP**-Motiv an der konservierten Prolin-Bindungsstelle lokalisiert. Damit besetzt das Leuzin die Ligandenposition $\lambda 1$, analog zur Position des Phosphoserins/Phosphothreonins im Komplex von *hPIN1-WW*. Eine analoge Platzierung ($\lambda 1$) des Arginins in den Komplexen der WW-Domänen der beiden R-Spezifitätsgruppen war ebenfalls getestet worden, allerdings konnten damit nicht alle aus den Substitutionsanalysen abgeleiteten Kontakte erfüllt werden. Letztendlich muss die Richtigkeit der vorgeschlagenen Komplexe durch experimentelle Strukturbestimmung überprüft werden.

Limitierung der Komplexstruktur-Modellierungsprozedur

Limitierung der COMBINE-Modelle

Die Modellierung der Komplexstrukturen ist gleichzeitig der größte Schwachpunkt und die größte Stärke der hier vorgestellten COMBINE-Modelle. Neben den Ungenauigkeiten, welche die Modellierung der Komplexe mit sich bringt, ist das größte Problem vor allem die Frage, wie ein „nicht bindender“ Komplex strukturell repräsentiert werden kann. Die Strukturmodelle derjenigen Komplexe, für welche die experimentellen Daten nur sehr schwache oder keine Wechselwirkungen anzeigen, werden jedoch benötigt, um das gesamte Affinitätsspektrum von stark bindend bis nicht bindend im Wechselwirkungsmodell abzudecken. In dieser Arbeit wurde davon ausgegangen, dass der Ligand in den „nicht bindenden“ Kombinationen in der gleichen Weise wie in den für den Liganden charakteristischen Komplexen über der Domäne positioniert ist (siehe VW3.3.1.4). Damit wurde für die „nicht bindenden“ Kombinationen implizit eine sehr schwache, unterhalb der Sensitivität der

experimentellen Messung liegende Wechselwirkung angenommen. Für die Wechselwirkungen zwischen WW-Domänen und prolinreichen Liganden ist diese Annahme sicherlich berechtigt,⁹¹ da die konservierte Prolin-Bindungsstelle prinzipiell immer eine Wechselwirkung mit zwei konsekutiven Prolinen ermöglicht. Darüber hinaus deuten auch die QSP-Modelle (4.3) auf einen mehr oder weniger kontinuierlichen Übergang zwischen „Bindung“ und „Nicht-Bindung“ hin. Dennoch besteht die Gefahr, dass in Folge der Modellierung Ligand und Domäne aufeinander abgestimmt werden und somit Artefakte in der Affinitätsvorhersage entstehen. Aus diesem Grund müssen die Komplexstrukturen durch unabhängige experimentelle Ergebnisse fundiert sein, sowohl für das Training des COMBINE-Modells als auch für die Affinitätsvorhersage. Dazu dienten in dieser Arbeit die experimentell definierten Spezifitätsgruppen und Substitutionsanalysen. Ferner muss für eine Vorhersage der Affinität einer unbekanntenen Wechselwirkung der erwartete Komplex den Komplexen ähnlich sein, auf denen das COMBINE-Modell trainiert wurde. Die folgenden Ergebnisse demonstrieren, dass die COMBINE-Methode trotz dieser Limitierungen ein großes Potential für die Analyse von Komplexstrukturmodellen besitzt.

Für die R_a- und die R_b-Spezifitätsgruppen, für welche keine experimentellen Komplexstrukturen verfügbar waren, konnten in sich konsistente Komplexstrukturmodelle erstellt werden (Abbildung 3-19 und Abbildung 3-20). Diese Modelle vereinen dabei die Informationen aus den Substitutionsanalysen (Abbildung 3-3 und Abbildung 3-4) und den für die Spezifitätsgruppen charakteristischen Konsensussequenzen (Abbildung 3-12). Es bleibt abzuwarten, ob die vorgeschlagenen Komplexstrukturen durch experimentell bestimmte Strukturen bestätigt werden können. Unabhängig davon konnte die Plausibilität dieser Komplexstrukturmodelle durch die Hauptkomponenten-Analyse (PCA) der paarweisen Interaktionsenergien weiter bestärkt werden (Abbildung 3-25). Am interessantesten ist dabei die Bedeutung der Position $\omega 26$ für die Bindung argininhaltiger Liganden (siehe 3.3.3.1). Für diese bislang unbeachtete Position wurden in der PCA signifikante Interaktionen mit dem Ligandenrest $\lambda - 2$ nachgewiesen. Eine anschließende Analyse der WW-Domänensequenzen offenbarte dann, dass der Großteil der WW-Domänen in dieser Position ein Arginin, Lysin oder selten auch ein Glutamin besitzt (Tabelle 3-6). Im Gegensatz dazu besitzen die WW-Domänen der beiden R-Gruppen nie ein Arginin, Lysin oder Glutamin an dieser Position: die WW-Domänen der R_a-Gruppe haben vorrangig kleine Aminosäuren in $\omega 26$, die der R_b-Gruppe aromatische Aminosäuren (Abbildung 3-11). Die sonst unter den WW-Domänen übliche Konservierung von R/K/Q an $\omega 26$ lässt sich dadurch erklären, dass Arginine oder Lysine häufig

PCA der Komplexstrukturmodelle der R_a- und R_b-Spezifitätsgruppen

Bedeutung der WW-Domänenposition $\omega 26$ für die Bindung argininhaltiger Liganden

mit dem aromatischen Tryptophan-Ring über eine Kationen/ π -Elektronen-Wechselwirkung interagieren.⁵⁶ Dadurch wird in den WW-Domänen die eine Seite des hydrophoben Tryptophans vom wässrigen Lösungsmittel abgeschirmt. Diese intramolekulare Wechselwirkung wird eindrucksvoll in der Röntgenkristallstruktur von *hPIN1*-WW [PDB: 1F8A]¹⁹³ verdeutlicht. In den beiden vorgeschlagenen Komplexstrukturen für die R_a - und die R_b -Spezifitätsgruppen kontaktiert aber das Arginin des Liganden die konservierten negativ geladenen Reste in der $\beta A\beta B$ -Schleife von den Ligandenpositionen $\lambda-2$ bzw. $\lambda-1$ aus (Abbildung 3-19 und Abbildung 3-20). Damit wäre ein Arginin oder Lysin in $\omega 26$ elektrostatisch unvorteilhaft und würde im Falle der R_a -Gruppe sogar zu einem Zusammenstoß der Seitenketten führen. Interessanterweise haben auch genau die beiden WW-Domänen der Y-Spezifitätsgruppe (*hWWP3*-WW1 und *hWWP3*-WW2), die im WW-Screening eine schwache Wechselwirkung mit dem doppelten R-Liganden zeigten (Tabelle 3-3), ein Glutamat bzw. Glyzin an $\omega 26$ (Tabelle 3-6). Darüber hinaus können auch die beiden für die R_a - und die R_b -Spezifitätsgruppen vorgeschlagenen unterschiedlichen Positionen des Arginins im Liganden durch die PCA erklärt werden. Die kleinen Seitenketten in $\omega 26$ der Domänen der R_a -Spezifitätsgruppe erlauben einem Arginin aus $\lambda-2$ die negativ geladenen Reste in der $\beta A\beta B$ -Schleife zu kontaktieren. Im Falle der R_b -Spezifitätsgruppe wird ein Zusammenstoß mit den großen aromatischen Seitenketten in $\omega 26$ vermieden, da hier das Arginin von $\lambda-1$ die $\beta A\beta B$ -Schleife kontaktiert. Ferner kann spekuliert werden, dass der konservierte aromatische Rest in $\omega 26$ weitere Kontakte mit Ligandenresten ausbildet, die C-terminal zum Arginin liegen. Die Substitutionsanalysen für WW-Domänen der R_b -Spezifitätsgruppe hatten eine spezifische Erkennung einiger Reste C-terminal zum Arginin gezeigt (Abbildung 3-4), ganz im Gegensatz zur R_a -Spezifitätsgruppe. Damit die Ligandenpositionen $\lambda-5$ bis $\lambda-3$ die aromatische Seitenkette in $\omega 26$ überhaupt erreichen können, müsste die PPII-Helix geknickt sein (Abbildung 3-20). Dies könnte damit eine plausible Erklärung für die ungewöhnlich hohe Präferenz für Glyzin an $\lambda-2$ sein, die sich in den Substitutionsanalysen der R_b -Spezifitätsgruppe zeigte. Damit konnte durch die COMBINE-Methode die Plausibilität der Strukturmodelle für die Komplexe der beiden R-Spezifitätsgruppen weiter gestützt werden. Dies demonstriert, dass die COMBINE-Methode auch erfolgreich zur Analyse und Validierung von Komplexstrukturmodellen verwendet werden kann.

Positionierung der Liganden in den R_a - und R_b -Spezifitätsgruppen

Rolle des Glyzins in $\lambda-2$ für die Erkennung des R-Liganden durch WW-Domänen der R_b -Spezifitätsgruppe

COMBINE-Modell der Tyrosin-Spezifität

Für die Strukturmodelle der WW-Domänen/Liganden-Komplexe der Y-Spezifitätsgruppe konnte zusammen mit den CSPs für die Bindung des Y-Modellliganden (Tabelle 3-6) ein prädiktives COMBINE-Modell der Tyrosin-Spezifität erstellt werden (siehe 3.3.3.2). In

Analogie zum CoMFA-Modell wurden alle WW-Domänen der Y-Spezifitätsgruppe mit quantifizierbaren CSPs und je ein Vertreter aus den übrigen Spezifitätsgruppen ausgewählt, um das COMBINE-Modell zu trainieren. Auf Grund des Trainings auf der Basis von Komplexstrukturmodellen muss auch bei der Interpretation des COMBINE-Modells der Tyrosin-Spezifität die Ungenauigkeit der Strukturmodelle mit einbezogen werden.

Grundsätzlich spiegelte das COMBINE-Modell der Tyrosin-Spezifität die wesentlichen konservierten Sequenzmerkmale der Y-Spezifitätsgruppe wieder (Abbildung 3-12). So stammen die größten Beiträge von den Wechselwirkungen zwischen Tyrosin in $\lambda 3$ und den Domänenresten $\omega 30$, $\omega 32$ und $\omega 35$ (Abbildung 3-27A-B). Dabei macht das COMBINE-Modell deutlich, dass die Präferenz für V/I/L an $\omega 30$ eine Folge der Vermeidung unvorteilhafter Kontakte zwischen $\gamma \lambda 3$ und größeren Seitenketten in $\omega 30$ ist.

Vergleich des COMBINE-Modells der Tyrosin-Spezifität mit dem CoMFA-Modell und der Konsensussequenz

Im Vergleich zum CoMFA-Modell und der Konsensussequenz liegt jedoch einer der wesentlichen Vorteile des COMBINE-Modells darin, dass durch die Einbeziehung des Liganden nicht nur eine Korrelation sondern vielmehr ein kausaler Zusammenhang zwischen Sequenz/Struktur und der Wechselwirkungsaffinität hergestellt wird. So scheint das innerhalb der Y-Spezifitätsgruppe hoch konservierte Glutamat an $\omega 18$ keine direkte Rolle bei der Ligandenbindung zu spielen (Abbildung 3-27A–B). Die Beteiligung dieses Rests an der Ligandenbindung wurde auch schon im Zuge der Interpretation der Konsensussequenz und der Ergebnisse des CoMFA-Modells (siehe 4.1.2) in Frage gestellt. Jedoch erlaubte keine der beiden Methoden allein eine Beteiligung auszuschließen. Angesichts der Tatsache, dass $\omega 18$ an keiner signifikant zum COMBINE-Modell beitragenden Interaktion beteiligt ist, handelt es sich bei $\omega 18$ und der korrespondierenden Region C des CoMFA-Modells (Abbildung 3-24) eher um eine Scheinkorrelation. Allerdings kann auch nicht ausgeschlossen werden, dass das Glutamat einen übergeordneten elektrostatischen Effekt auf die gesamte WW-Domäne, H $\omega 32$ oder R/K/Q $\omega 35$ ausübt. Aber auch für die intramolekularen Interaktionen, die für das COMBINE-Modell der Tyrosin-Spezifität bestimmt wurden, spielt $\omega 18$ keine signifikante Rolle (Daten nicht gezeigt).

Scheinkorrelation der WW-Domänenposition $\omega 18$ und der Bindung tyrosinhaltiger Liganden

Umgekehrt werden auch einige der durch das CoMFA-Modell vorhergesagten spezifitätsbestimmenden Strukturmerkmale durch das COMBINE-Modell bestätigt. So deutet der signifikant positive Beitrag der elektrostatischen Wechselwirkung zwischen $\gamma \lambda 3$ und $\omega 35$ an (Abbildung 3-27B), dass die Präferenz für eine positiv geladene Seitenkette in dieser Position tatsächlich im Zusammenhang mit der Wechselwirkungsaffinität steht. Ebenso bestätigt wird der Einfluss der Konformation und Umgebung der konservierten Tyrosin-Bindungsstelle

Bestätigung der Bedeutung von R/K/Q $\omega 35$ und der Konformation der Prolin-Bindungs tasche

auf die Affinität gegenüber dem Y-Liganden. So haben die van-der-Waals-Kontakte zwischen $\lambda 0$ und $\omega 39$, $\lambda 1$ und $\omega 28$ und überraschenderweise auch $\lambda 1$ und $\omega 37$, sowie die Wasserstoffbrücke zwischen $\lambda -2$ und $\omega 39$ einen signifikanten positiven Einfluss auf die Wechselwirkungsaktivität (Abbildung 3-27A–B).

Vorhersage der Aktivität für ähnliche Komplexe

Der weitaus größte Vorteil des COMBINE-Modells liegt aber auf der Anwendungsseite. Durch die Betrachtung des gesamten Komplexes kann die Wechselwirkungsaktivität auch für Domänen/Liganden-Kombinationen vorhergesagt werden, von denen weder die Domäne noch der Ligand im Trainingsdatensatz enthalten war. Im Gegensatz dazu ist die Vorhersage des CoMFA-Modells auf denjenigen Liganden beschränkt, mit dem das Modell trainiert wurde (siehe 3.3.2.2). Natürlich hat die Generalisierung des COMBINE-Modells auch ihre Grenzen: die Struktur und der Wechselwirkungsmechanismus des Komplexes, für den die Affinität vorhergesagt werden soll, müssen den Komplexen ähneln, auf denen das Modell trainiert wurde. Trotz dieser Einschränkungen ermöglicht das COMBINE-Modell damit das *de novo* Design von ähnlichen Komplexen, wobei theoretisch gleichzeitig die Domäne und der Ligand verändert werden können.

Design höher affiner peptidischer Liganden

Auf Grund seiner hervorragenden Vorhersageleistung (Abbildung 3-26) wurde das COMBINE-Modell zum Design eines höher affinen peptidischen Liganden für die γ RSP5-WW1 Domäne aus *Saccharomyces cerevisiae* verwendet (siehe 3.3.3.3). Die Bäckerhefe ist ein interessanter Modellorganismus für die Aufklärung des WW-Domänen/Liganden-Interaktionsnetzwerks, da für sie nur 6 Proteine mit insgesamt 9 WW-Domänen identifiziert wurden (SMART-Datenbank^{160,161}). Trotz umfangreicher experimenteller Daten konnte bislang jedoch noch kein klares Bild des WW-Domänen-vermittelten Interaktionsnetzwerks gewonnen werden.^{57,70} Für die Aufklärung des Interaktionsnetzwerks ist die *in silico* Vorhersage von Liganden in zweierlei Hinsicht nützlich: zum Einen bildet sie eine fundierte Arbeitshypothese für die experimentelle Suche nach weiteren Wechselwirkungspartnern für WW-Domänen. Zum Anderen können mit dem Wechselwirkungsmodell hochaffine peptidische Liganden designiert werden, die als kompetitive Inhibitoren wertvolle Werkzeuge in zellbiologischen Experimenten darstellen.

Anwendung der Vorhersage von Liganden

Design eines peptidischen Liganden für γ RSP5-WW1

Für das Design eines höher affinen peptidischen Liganden für γ RSP5-WW1 (siehe 3.3.3.3) wurde vom Erkennungsmotiv **xPPxY** ausgegangen, da dieses für WW-Domänen der Y-Spezifitätsgruppe nicht weiter optimierbar ist (Abbildung 3-2). Aus diesem Grund, wurde mit Hilfe der COMBINE-Analyse nach affinitätssteigernden Interaktionen mit den flankierenden Bereichen des Liganden gesucht (Abbildung 3-27). Ein vielversprechender

Kandidat war die Interaktion zwischen der WW-Domänenposition $\omega 26$ und der Ligandenposition $\lambda - 2$ (Abbildung 3-28). Experimentell konnte bestätigt werden, dass die vorgeschlagene Mutation von $T\lambda - 2$ zu $D\lambda - 2$ im Y-Liganden die Affinität fast verdreifacht (Tabelle 3-12). Auch die Veränderung der Affinität war in derselben Größenordnung, wie die durch das COMBINE-Modell vorhergesagte Veränderung der Wechselwirkungsaktivität. Dies deutet auf einen guten quantitativen Zusammenhang zwischen vorhergesagter Wechselwirkungsaktivität und Affinität hin. Interessanterweise konnte unter den bekannten, physiologisch relevanten Wechselwirkungspartnern für $\gamma RSP5^{82}$ das Hefe-Homolog zu DIA1 [UP: YM95_YEAST] identifiziert werden, das tatsächlich ein Aspartat in der Ligandenposition $\lambda - 2$ besitzt. Dies bestätigt damit die *in vivo* Relevanz der vorhergesagten Wechselwirkung.

Experimentelle Bestätigung der vorhergesagten Wechselwirkungsaktivität

In vivo Relevanz der vorhergesagten Wechselwirkung

Mit den Erkenntnissen, die durch das Design peptidischer Liganden mit höherer Affinität gewonnen werden, kann das COMBINE-Modell prinzipiell iterativ optimiert werden. Das COMBINE-Modell mag damit auch die Richtung für das Design therapeutisch relevanter, nicht peptidischer Substanzen vorgeben. Gerade in dieser Anwendung liegt eine weitere Stärke des COMBINE-Modells. So kann durch das COMBINE-Modell die Affinität nicht nur für peptidische Substanzen vorhergesagt werden, sondern auch für Peptidomimetika.¹⁹⁹

Transfer der Erkenntnisse für peptische Liganden auf nicht peptidische Liganden

4.2 WW-Domänen Spezifitätsmechanismen

Jedes der hier vorgestellten Wechselwirkungsmodelle hat eigene, sich ergänzende Einblicke in die Spezifitätsmechanismen der WW-Domänen gegeben (siehe 4.1). Gemeinsam mit den experimentellen Ergebnissen (siehe 3.1 und 3.2.1 und Pires *et al.*, 2005) kann daraus im Folgenden ein Überblick über die Spezifitätsmechanismen für die Familie der WW-Domänen entwickelt werden.

Die prinzipielle Wechselwirkung mit prolinreichen Liganden konnte für den Großteil der 42 untersuchten WW-Domänen bestätigt werden (siehe 3.1). Ausnahmen bildeten ungefaltete WW-Domänen und solche, in denen die Prolin-Bindungsstelle (**Y/F $\omega 28$** und **W $\omega 39$**) nicht konserviert ist (Tabelle 3-6). Für einige der WW-Domänen mit einem Phenylalanin oder Tyrosin anstatt des Tryptophans in $\omega 39$ wurden durch das NMR-Screening schwache Wechselwirkungen festgestellt, die aber durch die festphasegebundenen Bibliotheken nicht reproduziert werden konnten (Tabelle 3-3). Der Einfluss von Y/F $\omega 39$ auf die Spezifität bleibt damit und im Hinblick auf Berichte über Wechselwirkungen dieser Domänen mit prolinreichen Liganden weiterhin unklar.¹³⁹

Konservierte Prolin-Bindungsstelle: **Y/F $\omega 28$** und **W $\omega 39$**

Position ω 30:
Bedeutung für
Tyrosin-Spezifität

Die Position ω 30 spielt eine essentielle Rolle bei der Unterscheidung der tyrosinspezifischen WW-Domänen von den restlichen, nicht tyrosinspezifischen WW-Domänen (4.1). Zwar ist V/I/L ω 30 nicht ausreichend für die Wechselwirkung mit tyrosinhaltigen Liganden, jedoch verhindert eine aromatische Aminosäure in dieser Position die Wechselwirkung mit tyrosinhaltigen Liganden wie auch Sudol¹⁷⁶ spekulierte. Die meisten nicht tyrosinspezifischen WW-Domänen besitzen daher auch eine aromatische Aminosäure in ω 30 (Abbildung 3-11). Dennoch wurde auch für die WW-Domänen von γ PRP40 (Tabelle 3-6), insbesondere für γ PRP40-WW2, eine schwache, aber reproduzierbare Wechselwirkung mit **xPPxY**-Motiven beschrieben.²⁰² Damit übereinstimmend ergab die Analyse der WW-Domänenstrukturen von γ PRP40 [PDB: 1O6W],²⁰² dass die Tyrosin-Seitenkette von Y ω 30, vor allem in γ PRP40-WW2, auf Grund einer etwas anderen Konformation die $\beta\beta$ C-Schleife weniger „abschirmt“. ¹³⁷

Position ω 30:
Bedeutung für nicht
tyrosinspezifische
WW-Domänen

Für die nicht tyrosinspezifischen WW-Domänen vergrößert eine aromatische Aminosäure in ω 30 zugleich die hydrophobe Oberfläche zur Wechselwirkung mit dem Liganden (siehe 3.3.1.4). Dies korreliert sowohl mit der Tendenz zu längeren Erkennungsmotiven, als auch mit der Tendenz zur Bindung reiner Polyprolin-Motive (siehe 3.2.1). Die einzigen Ausnahmen hiervon sind die WW-Domänen der *poS/poT*-Spezifitätsgruppe, bei denen F ω 30 keine essentielle Rolle bei der Ligandenbindung spielt.¹⁹³ Damit erfüllt der aromatische Rest in ω 30 für die *poS/poT*-Spezifitätsgruppe wahrscheinlich vor allem die Aufgabe der Vermeidung der Wechselwirkung mit tyrosinhaltigen Liganden. Im Gegensatz dazu ist das konservierte Tyrosin in ω 30 für die Bindung des L-Liganden durch WW-Domänen der L-Spezifitätsgruppe essentiell, auch wenn es nicht direkt für die Leuzin-Spezifität verantwortlich ist.¹³⁷ Auch bei den beiden R-Spezifitätsgruppen scheint die aromatische Aminosäure in ω 30 eine große Bedeutung für die Ligandenbindung zu haben (Abbildung 3-11). Die R_o-Spezifitätsgruppe scheint an dieser Position ein Tryptophan zu benötigen, um möglicherweise eine weitere hydrophobe Vertiefung zusammen mit V ω 20 und Y ω 28 auszubilden (Abbildung 3-19), analog zu der zweiten hydrophoben Vertiefung in *hFBP11*-WW1.¹³⁷ Dies würde auch die spezifische Interaktion mit den beiden Resten in –5 und –6 Position zum Arginin im R-Liganden erklären (Abbildung 3-3). In den WW-Domänen der R_b-Spezifitätsgruppe scheint diese Vertiefung zu fehlen (Abbildung 3-20). Dies würde die geringere Anzahl an spezifisch erkannten Ligandenresten N-terminal zum Arginin erklären (Abbildung 3-4). Dennoch scheint auch Y ω 30 essentielle Kontakte zum Liganden auszubilden (Abbildung 3-25).

Eine wichtige Rolle bei der Spezifität für argininhaltigen Liganden spielt die Position $\omega 26$ (Abbildung 3-25). Sie dient, ähnlich wie die Position $\omega 30$ für tyrosinhaltige Liganden, vor allem der selektiven Hemmung der Wechselwirkung gegenüber argininhaltigen Liganden. So besitzen beinahe alle WW-Domänen, die nicht den beiden R-Spezifitätsgruppen angehören, ein Arginin oder Lysin, selten auch ein Glutamin in dieser Position (Tabelle 3-6). In den vorgeschlagenen Komplexen für die Wechselwirkung mit argininhaltigen Liganden (Abbildung 3-19 und Abbildung 3-20) würden diese Aminosäuren in $\omega 26$ zu unvorteilhaften elektrostatischen und sterischen Interaktionen mit dem Arginin des Liganden führen (Abbildung 3-25). Unterstützt wird diese Hypothese auch dadurch, dass die zwei WW-Domänen der Y-Spezifitätsgruppe, die kein R/K/Q in $\omega 26$ besitzen, als einzige eine, wenn auch schwache, Wechselwirkung mit dem doppelten R-Liganden zeigen (Tabelle 3-3 und Tabelle 3-6). Darüber hinaus könnte in den WW-Domänen der R_b -Spezifitätsgruppe die konservierte aromatische Aminosäure in $\omega 26$ zusammen mit $W\omega 39$ eine weitere hydrophobe Vertiefung ausbilden (Abbildung 3-20). Diese würde eine logische Erklärung für die spezifische Erkennung des **gPp**-Motivs liefern, welches C-terminal zum Arginin im R-Liganden liegt und nur von WW-Domänen der R_b -Spezifitätsgruppe erkannt wird (Abbildung 3-4).

Position $\omega 26$:
Bedeutung für
Arginin-Spezifität

Zusätzlich zu $\omega 26$ zeichnen sich die WW-Domänen der R-Spezifitätsgruppen vor allem durch konservierte negativ geladene Aminosäuren in der **$\beta A\beta B$ -Schleife** aus (Abbildung 3-11). Diese können, entsprechend der vorgeschlagenen Komplexe, elektrostatisch mit dem Arginin des Liganden interagieren. Die Unterschiede in der Position der negativ geladenen Reste (R_a : **D $\omega 22$** ; R_b : **D/E $\omega 24$** ; siehe Abbildung 3-12) können mit der vorgeschlagenen, unterschiedlichen „Positionierung“ des Arginins des Liganden zusammenhängen. Auch in den WW-Domänen der *poS/poT*-Spezifitätsgruppe sind die spezifitätsbestimmenden Reste in der $\beta A\beta B$ -Schleife lokalisiert (Abbildung 3-12). Dabei spielen vor allem **S $\omega 22$** und **R/K $\omega 24$** eine essentielle Rolle bei der Erkennung der Phosphatgruppe des Phosphoserins bzw. Phosphothreonins.¹⁹³ Im Falle der Wechselwirkung von *hFBP11*-WW1 und dem L-Liganden konnte eine Reihe von Kontakten zwischen dem Leuzin des Liganden und den Resten der $\beta A\beta B$ -Schleife identifiziert werden.¹³⁷ Allerdings zeigen die Domänensequenzen der L-Spezifitätsgruppe für die $\beta A\beta B$ -Schleife keine Konservierung bestimmter Aminosäuren (Tabelle 3-6). Ebenso wenig konserviert ist die $\beta A\beta B$ -Schleife in der Y-Spezifitätsgruppe. Gleichwohl ist für die tyrosinspezifischen WW-Domänen gesichert, dass dieser Teil der WW-Domäne für die Ligandenbindung keine Rolle spielt.¹³⁸

$\beta A\beta B$ -Schleife

D $\omega 22$ und D/E $\omega 24$:
Bedeutung für
Arginin-Spezifität

S $\omega 22$ und R/K $\omega 24$:
Bedeutung für
poS/poT-Spezifität

$\beta\beta\beta$ C-Schleife

Die Domänen der Y-Spezifitätsgruppe besitzen dagegen in der **$\beta\beta\beta$ C-Schleife** ein hochkonserviertes Erkennungsmotiv zur Bindung des Tyrosins im Liganden: dazu gehören neben **V/I/L ω 30** auch **H ω 32** und **R/K/Q ω 35** (Abbildung 3-24). Die nicht tyrosinspezifischen WW-Domänen zeigen in der $\beta\beta\beta$ C-Schleife eine weitaus geringere Konservierung der Sequenz (Tabelle 3-6). Dies ist einleuchtend angesichts der „Abschirmung“ dieses Bereichs durch aromatische Aminosäuren in ω 30 (siehe 3.3.1.4).

Weniger klare Beiträge zur Spezifität kommen von den Positionen ω 18, ω 20 und ω 37.

Position ω 18 Die starke Korrelation zwischen E ω 18 mit der Erkennung tyrosinhaltiger Liganden ist wohl nur eine Scheinkorrelation (siehe 4.1.3). Dennoch kann auch ω 18 zur Ausbildung einer weiteren hydrophoben Vertiefung beitragen, wie der FBP11-WW1/L-Liganden-Komplex

Position ω 20 zeigt. Gleichwohl kommt der Position ω 20 dabei eine weitaus bedeutendere Rolle zu.¹³⁷

Position ω 37 Möglicherweise ist dieser Rest auch in WW-Domänen der R α -Spezifitätsgruppe an der Ausbildung einer weiteren Vertiefung beteiligt (Abbildung 3-12 und Abbildung 3-19). Für ω 37 wurde zwar unter anderem ein Einfluss auf die Bindung tyrosinhaltiger Liganden vorgeschlagen (Abbildung 3-24 und Abbildung 3-27A-B), allerdings spiegelt sich dies nicht klar in den Sequenzen wieder (Tabelle 3-6).

Zusammenfassung des Klassifizierungssystems der WW-Domänen

Tabelle 4-1 fasst die spezifitätsbestimmenden Sequenzmerkmale der Domänen und der Liganden zu einem einheitlichen Klassifizierungssystem der WW-Domänen zusammen. Diese basieren auf den experimentell definierten Spezifitätsgruppen aus Abschnitt 3.2.2, wurden aber grundlegend durch die Erkenntnisse aus den sequenz- und strukturbasierten Wechselwirkungsmodellen, sowie Ergebnissen aus der Literatur^{137,193} ergänzt.

Weitere Erkennungsmotive in den Liganden

Es ist wichtig anzumerken, dass die in Tabelle 4-1 angegebenen Erkennungsmotive allein auf den in dieser Arbeit untersuchten Liganden basieren (siehe 3.2.1). Nicht für jede Spezifitätsgruppe muss aber zwangsläufig gelten, dass sie ausschließlich mit diesem Erkennungsmotiv interagiert, wie durch die Wechselwirkung der WW-Domänen der R α -Spezifitätsgruppe mit dem poly-P-Liganden verdeutlicht wird. Auf Grund der Beschränkung der Substitutionsanalysen auf Einzelsubstitutionsanaloge lassen sich eben nicht alle möglichen Liganden von einem einzigen Ausgangspeptid ableiten.

Multiple Epitope in den Zielproteinen

Die Stärke der Wechselwirkung einer WW-Domäne mit einem Zielprotein wird neben der Zielsequenz durch einen weiteren Faktor beeinflusst: die erkannten Epitope kommen in den Zielproteinen häufig mehrfach vor. Durch die Wiederholung eines oder mehrerer Motive kann auch für nicht optimale Motive eine relevante makroskopische Affinität erreicht werden, wobei die Motive auch überlappen dürfen.^{115,186} Besonders augenfällig ist dieser Effekt

Tabelle 4-1 Zusammenfassung der spezifitätsbestimmenden Reste der WW-Domänen und ihrer Liganden für die einzelnen Spezifitätsgruppen

Für alle experimentell definierten Spezifitätsgruppen, außer für die poly-P-Spezifitätsgruppe, konnten mit Hilfe der experimentellen Analysen und der verschiedenen Wechselwirkungsmodelle charakteristische spezifitätsbestimmende Reste identifiziert werden.

Spezifitätsgruppe ^{a)}	Minimalmotiv der spezifitätsbestimmenden Reste ^{b)}	Zusätzliche spezifitätsbestimmende Reste ^{c)}	Liganden-Erkennungsmotiv ^{d)}
Y	V/I/L ω 30, H ω 32, R/K/Q ω 35	R/K/Q ω 26, T ω 37, (E ω 18)	xPPxY xPPxpoY
R_a	D/E ω 22, ^ (R/K/Q/Y/F/W) ω 26, W ω 30	V ω 20, D/E ω 23, (V/T ω 37)	(p/Φ) P (p/g) PPpR (p/Φ) PFFFF
R_b	D/E ω 24, Y/F ω 26, Y ω 30	(V ω 18), (G ω 20), T/S ω 22, D/E ω 23,	pPPRgpPp
L	H/Y ω 20, Y ω 30, β A β B-Schleife	R/K/Q ω 26	PPLPp
poS/poS	S ω 22, R/K ω 23	R ω 20, S ω 24, R/K/Q ω 26, F ω 30	(poT/poS) P
poly-P	?	?	(p/Φ) PFFFF

^{a)} Experimentell definierte Spezifitätsgruppe (siehe 3.2.2).

^{b)} Minimal notwendige Sequenzmerkmale für die Wechselwirkung mit dem charakteristischen Liganden der Spezifitätsgruppe. Ausgeschlossene Aminosäuren sind durch ein vorangestelltes „^“ gekennzeichnet.

^{c)} Sequenzmerkmale, die nicht unbedingt notwendig sind für die Bindung des charakteristischen Liganden aber dennoch mit der Spezifitätsgruppe korrelieren. Sie können zur Vermeidung der Wechselwirkung mit anderen Liganden dienen (z.B. R/K/Q ω 26), sie können die Wechselwirkung mit dem charakteristischen Liganden verstärken oder ihre Funktion kann unklar sein, obwohl sie mit der Spezifitätsgruppe korrelieren (in Klammern).

^{d)} Für die charakteristischen Liganden sind die Erkennungsmotive angegeben. Dies muss nicht zwangsläufig bedeuten, dass nur diese Liganden gebunden werden (siehe Text). Für Nomenklatur der Sequenzmotive siehe 2.1.3.

für die Wechselwirkung mit reinen Polyprolin-Abschnitten, wie sie für die WW-Domänen der R_a- und poly-P-Spezifitätsgruppen (siehe 3.2.1.4), aber auch der L-Spezifitätsgruppe¹³⁷ gefunden wurde. Dies hängt damit zusammen, dass kein spezifitätsbestimmender Rest die Anzahl der durch die Länge des Polyprolin-Abschnitts bestimmten Bindungsmöglichkeiten einschränkt. Interessanterweise sind viele WW-Domänen-Bindungsstellen in den natürlichen Proteinen innerhalb verhältnismäßig langer prolinreicher Abschnitte.^{3,93} Umgekehrt kann die Stärke der Wechselwirkung auch durch mehrfaches Auftreten der Domänen erhöht werden. Dies kann zum Beispiel für die tandemwiederholten WW-Domänen der NEDD4-Ubiquitin-Ligasen, die alle der Y-Spezifitätsgruppe angehören, vermutet werden. Kommen sowohl die Domänen als auch die Zielmotive wiederholt in den interagierenden Proteinen vor, wird die Wechselwirkung zusätzlich durch Aviditätseffekte verstärkt, da mehrere Interaktionen gleichzeitig getrennt werden müssen.

Ist die poly-P-Gruppe eine Spezifitätsgruppe?

Daneben fällt in Tabelle 4-1 auf, dass für die poly-P-Spezifitätsgruppe keine spezifitätsbestimmenden Reste identifiziert werden konnten, auch nicht unter Einbeziehung der Erkenntnisse für die anderen Spezifitätsgruppen. Damit stellt sich die Frage, ob die beiden in die poly-P-Spezifitätsgruppe klassifizierten WW-Domänen *hGAS7-WW* und *yPRP40-WW2* tatsächlich den poly-P-Liganden über den gleichen Bindungsmechanismus erkennen. Oder handelt es sich bei dieser Wechselwirkung eher um eine grundlegende Eigenschaft von WW-Domänen mit möglicherweise mehreren hydrophoben Vertiefungen, wie es die Wechselwirkung der WW-Domänen der R_α-Spezifitätsgruppe mit dem poly-P-Liganden andeutet. Diese Fragen können erst durch die noch ausstehende strukturelle Aufklärung der entsprechenden Komplexe geklärt werden.

Klassifizierung der bekannten WW-Domänen in die 5 definierten Spezifitätsgruppen

Mit Hilfe der spezifitätsbestimmenden Reste aus Tabelle 4-1 können nun die WW-Domänensequenzen, zusätzlich zur Y-Spezifitätsgruppe (Tabelle 3-7), auch in die R_α-, R_β-, L- und poS/poT-Spezifitätsgruppen klassifiziert werden (Tabelle 4-2). Damit können insgesamt 66% aller WW-Domänen einer der 5 definierten Spezifitätsgruppen zugeordnet werden. Werden nur diejenigen WW-Domänen betrachtet, die konservierten Reste W/F/Y ω 17, Y/F ω 28, Y/F ω 29 und W ω 39 gleichzeitig erfüllen (Tabelle 4-2, Spalte WW), kann durch diese Klassifizierung für 72% der WW-Domänen die Spezifitätsgruppe vorhergesagt werden.

Nutzen der Strukturmodelle

Eine ganze Reihe der hier zusammengefassten Einsichten in die Spezifitätsmechanismen der WW-Domänen wären ohne die strukturbasierten Analysen und damit ohne die Modellierung der Komplexstrukturen nicht möglich gewesen (siehe beispielsweise ω 26, 4.1.3). Besonders interessante und plausible Hypothesen konnten an Hand der Komplexstruktur-

Validität der Komplexstrukturmodelle der argininspezifischen WW-Domänen

modelle der beiden R-Spezifitätsgruppen ausgearbeitet werden. Unterstützung für diese Modelle kommt überraschenderweise von der Struktur [PDB: 1EG4]⁷⁶ der *hDYSTRO-WW* Domäne im Komplex mit einem Peptid aus β -Dystroglycan (NMTPYR _{λ -2}SP _{λ 0}PPYVP). Der tyrosinhaltige Ligand enthält nämlich interessanterweise auch ein Arginin das die β A β B-Schleife kontaktiert, und zwar von λ -2 aus. Noch überzeugender wird dieser Vergleich dadurch, dass auch *hDYSTRO-WW* kein R/K/Q in ω 26 besitzt. Dass es sich bei *hDYSTRO-WW* jedoch nicht um eine argininspezifische WW-Domäne handelt liegt sowohl an den fehlenden negativ geladenen Resten in der β A β B-Schleife als auch an Isoleuzin in ω 30.

Mechanismen der „Feinspezifität“

Darüber hinaus verdeutlicht dies auch einen Mechanismus, mit dem die „Feinspezifität“ innerhalb der Spezifitätsgruppen erreicht wird. So würde der β -Dystroglycan-Ligand sehr wahrscheinlich nicht mit einer hohen Affinität von einer WW-Domäne wie *yRSP5-WW1*

Tabelle 4-2 Klassifizierung der WW-Domänen in die Y-, R_a-, R_b-, L- und poS/poT-Spezifitätsgruppen

66 % aller 482 WW-Domänen aus der SMART-Datenbank^{160,161} (Stand: 15.11.2001) wurden einer der 5 definierten Spezifitätsgruppen zugeordnet (Tabelle 4-1). Werden die 113 WW-Domänen nicht miteingerechnet, welche nicht gleichzeitig alle konservierten Reste von WW-Domänen erfüllen (Spalte WW), können 72 % der WW-Domänen einer der Spezifitätsgruppen zugeordnet werden.

Proteinfamilie	Total	Y ^{a)}	Y ^{b)}	R _a ^{c)}	R _b ^{d)}	L ^{e)}	poS ^{f)}	WW ^{g)}
NEDD4-Ubiquitin-Ligasen	155	144	112					32
MAGI und homologe	36	19	17					17
Dystrophin und homologe	21	19	14					1
YAP und homologe	9	9	9					
Synip	1	1	1					
BAG und homologe	5	5	5					
fbZFP2	1	1	1					
WVVOX und homologe	24	24						12
HYP109 und homologe	8			4				
FE65 und homologe	11			11				
PQBP1 und homologe	9			9				
YFB0	2							1
FBP21 und homologe	6				3			2
FCA-γ und homologe	2				2	2		
FBP11 und homologe	27					20		
CA150 und homologe	8					2		1
WAC	1					1		
Dinap11P1	2					1		
PIN1 und homologe	10						7	
GAS7 und homologe	6							
Small-GTPase-activating und homol.	7							
YPR152C	1							
WW45 und homologe	10							5
PEPP2	2							1
RNA Helikase und homologe	8							8
YJQ8	1							1
unklassifiziert	109	17	9		1	10	6	32
Σ	482	239	168	24	6	36	13	113
%	100%	50%	35%	5%	1%	7%	3%	23%

Die Sequenzen der WW-Domänen wurden klassifiziert durch das pHMM der Y-Spezifitätsgruppe und die Minimalmotive aus Tabelle 4-1. Die Motive wurden durch die konservierten Reste einer WW-Domäne (W/F/Y_{ω17}, Y/F_{ω28}, Y/F_{ω29} und W_{ω39}) erweitert. Für Nomenklatur der Sequenzmotive siehe 2.1.3.

^a Y-Spezifitätsgruppe: pHMM der Y-Spezifitätsgruppe. Die WW-Domänen des Übergangsbereichs wurden ebenfalls zu den tyrosinspezifischen Domänen gezählt (siehe 3.2.3.3).

^b Motiv der Y-Spezifitätsgruppe: **(W/F/Y) x₉₋₁₁ (Y/F) (Y/F) (V/I/L) xHx₂ (R/K/Q) x₃W**

^c Motiv der R_a-Spezifitätsgruppe: **(W/F/Y) x₄ (D/E) x₂₋₄ (R/K/Q/Y/F/W) x (Y/F) (Y/F) Wx₈W**

^d Motiv der R_b-Spezifitätsgruppe: **(W/F/Y) x₆ (D/E) x₁₋₂ (Y/F/W) x (Y/F) (Y/F) Yx₈W**

^e Motiv der L-Spezifitätsgruppe: **(W/F/Y) x₂ (H/Y) x₆₋₈ (Y/F) (Y/F) Yx₈W**

^f Motiv der poS/poT-Spezifitätsgruppe: **(W/F/Y) x₄ (S/T) (R/K) x₃₋₅ (Y/F) (Y/F) (Y/F/W) x₈W**

^g WW-Domänen, die nicht gleichzeitig alle konservierten Reste einer WW-Domäne aufweisen, d.h. folgendes Motiv nicht erfüllen: **Wx₉₋₁₁ (Y/F) (Y/F) x₉W**

Die angegebenen Proteinbezeichnungen sind der UniProt-Datenbank¹⁰ entnommen. Die wichtigsten Proteinfamilien sind im Text und der Einleitung besprochen (siehe 1.3.3).

gebunden werden, da das COMBINE-Modell für einen Liganden mit einem Arginin in λ-2 eine deutlich erniedrigte Wechselwirkungsaktivität vorhersagt (Tabelle 3-12). Die vorhergesagte niedrige Wechselwirkungsaktivität für Liganden mit Arginin in λ-2 ist vor allem eine Folge des Auftretens eines Arginins in ω26 von γRSP5-WW1 (siehe 3.3.3.3). Interessanterweise haben auch alle WW-Domänen der humanen Homologe von γRSP5 (*hNEDD4* und *hITCHY*) ein Arginin an ω26 (Tabelle 3-6). Somit kann vorausgesagt werden, dass das β-

Dystroglycan-Motiv nicht stark von NEDD4-ähnlichen Ubiquitin-Ligasen gebunden wird.

In vivo Relevanz der
vorhergesagten
Feinspezifität

Diese Vorhersage korreliert gut mit der Funktion der beiden Wechselwirkungen, denn die Dystrophin-Dystroglycan-Wechselwirkung erfüllt eine strukturelle Aufgabe,⁷⁶ während über die Wechselwirkung mit den NEDD4-ähnlichen Ubiquitin-Ligasen die Stabilität der gebundenen Proteine über den Ubiquitin-vermittelte Proteinabbau reguliert wird.^{172,173}

Phosphorylierung der
Erkennungsmotive im
Liganden

Zusätzlich zur Aminosäuresequenz des Erkennungsmotivs im Zielprotein spielen auch sekundäre Modifikationen, wie z.B. die Phosphorylierung dieser Motive, für die Spezifität der WW-Domänen-vermittelten Wechselwirkungen eine Rolle (siehe 3.2.1.5, *poS/poT*-Spezifitätsgruppe). Im Falle der Phosphorylierung des Tyrosins in **xPPxY**-Motiven wurde bislang angenommen, dass dies die Wechselwirkung mit WW-Domänen ausschließlich unterbindet.^{171,178} Überraschenderweise wurde aber in dieser Arbeit festgestellt, dass einige WW-Domänen der Y-Spezifitätsgruppe mit der phosphorylierten Variante ihrer tyrosinhaltigen Liganden interagieren können, zum Teil sogar mit höherer Affinität als mit der nicht phosphorylierten Variante (Tabelle 3-4 und Abbildung 3-1). Damit ist die Hypothese einer ausschließlich negativen Regulation der WW-Domänen/Y-Liganden-Wechselwirkung nicht mehr haltbar. Dennoch konnten auch in dieser Arbeit nicht die Sequenzmerkmale identifiziert werden, welche einzelne natürliche WW-Domänen zur Bindung phosphotyrosinhaltiger Liganden befähigen.¹⁸² Für eine detailliertere Diskussion der Bedeutung der Phosphorylierung für WW-Domänen-vermittelte Wechselwirkungen siehe Otte *et al.*, 2003 und Otte, 2004.

4.3 Quantitative Spezifitäts-Profile: eine quantitative Beschreibung der Spezifität

Klassischerweise wird die Spezifität einer Protein-Interaktionsdomäne durch ein Konsensussequenzmotiv der bekannten Liganden beschrieben (siehe 3.2.2 und 4.2).^{170,177} Meist lassen sich dadurch aber nicht alle experimentellen Ergebnisse erklären (siehe 3.4).⁸⁸ Darüber hinaus beinhalten Sequenzmotive auch keine Informationen über die Affinitäten der Wechselwirkungen (siehe 4.1.1). Zwar trifft die letzte Einschränkung nicht auf die bisher vorgestellten quantitativen strukturbasierten Wechselwirkungsmodelle zu (siehe 4.1.2 und 4.1.3), dennoch können diese die Affinität auch immer nur für ähnliche Liganden (COMBINE-Modell) bzw. denjenigen Liganden vorhersagen (CoMFA-Modell), mit welchem die Modelle trainiert wurden.

Klassische Beschreibung der Spezifität

Im Gegensatz zu dieser ligandenzentrierten Sicht (ein Ligand – viele Domänen) wird mit Hilfe der im Rahmen dieser Arbeit entwickelten *Quantitativen Spezifitäts-Profile* (QSP) eine domänenzentrierte Sicht eingenommen (eine Domäne – viele Liganden). QSPs sind dabei unabhängig von einer artifiziell qualitativen Einteilung in „Ligand“ bzw. „Nicht-Ligand“ (siehe 4.1.1). Vielmehr beschreiben sie die Spezifität quantitativ in Form der Wechselwirkungsaffinitäten für jedes Mitglied aus der Menge aller potentiellen peptidischen Liganden. Dadurch ergibt sich ein umfassenderes Bild der Spezifität der analysierten Domäne.

Ziele der Quantitativen Spezifitäts-Profile

Die Grundlage für die *Quantitativen Spezifitäts-Profile* bildet also eine Beschreibung der Spezifität in Form einer Ligandensequenz-abhängigen Affinitätsfunktion (siehe 3.5). Hierfür wird eine konservierte und damit weitgehend sequenzunabhängige Struktur der peptidischen Liganden im Komplex mit der Domäne vorausgesetzt. Diese kann auf Grund des konservierten Liganden-Bindungsmechanismus für viele Protein-Interaktionsdomänenfamilien angenommen werden. Infolgedessen reduziert sich die Spezifität auf eine Funktion der Aminosäuresequenz desjenigen Peptidabschnitts der Zielproteine, welcher während der Wechselwirkung im direkten Kontakt mit der Domäne steht, und damit dem klassischen Erkennungsmotiv entspricht. Die relativen Affinitätsbeiträge $\Delta \ln(K_d)$ der einzelnen Aminosäuren an den einzelnen Ligandenpositionen können damit als sogenannte *Quantitative Spezifitäts-Profile* zusammengefasst werden. Über die Gleichung $\Delta G_0 = -RT \ln(K_d^{-1})$ repräsentieren sie zugleich die relativen Bindungsenergiebeiträge der einzelnen Aminosäuren des Liganden: $\Delta \Delta G_0 = -RT \ln(K_d^{\text{Peptid}}/K_d^{\text{Referenz}}) = -RT \Delta \ln(K_d)$.

Konzept der Quantitativen Spezifitäts-Profile

Konservierter Bindungsmechanismus als Voraussetzung

Quantitative
Spezifitäts-Profile der
PDZ-Domänen aus
hAF6, *hERBIN* und
mSNA1

Die Methode der *Quantitativen Spezifitäts-Profile* wurde in dieser Arbeit beispielhaft an den PDZ-Domänen aus den drei Proteinen *hAF6*, *hERBIN* und *mSNA1* getestet (siehe 3.5). Die Annahme einer konservierten Struktur der Liganden im Komplex mit den Domänen konnte sowohl für PDZ-Domänen im Allgemeinen durch die Analyse der bekannten experimentell bestimmten Komplexstrukturen (Daten nicht gezeigt), als auch im Speziellen für die drei untersuchten PDZ-Domänen an Hand der Substitutionsanalysen bestätigt werden (Abbildung 3-31). Beide Analysen zeigten, dass die Spezifität vorwiegend durch die letzten 4 C-terminalen Aminosäuren des Liganden bestimmt wird. Die Menge der potentiellen peptidischen Liganden für jede dieser PDZ Domänen reduziert sich damit auf den theoretischen Liganden-Sequenzraum, gebildet durch die 20 proteinogenen Aminosäuren an den vier konservierten Ligandenpositionen (λ_0 bis $\lambda-3$). Dieser Liganden-Sequenzraum wird als theoretisch bezeichnet, da im Vergleich dazu der reale Liganden-Sequenzraum durch die Verfügbarkeit der C-Termini real existierender Proteine in der Zelle stark beeinflusst werden kann (siehe auch 4.5).

Konzept des Liganden-
Sequenzraums

Design der
Profilbibliotheken

Das experimentelle Training der QSP-Modelle für einen so großen Sequenzraum ($19^4 = 130\,321$ Tetrapeptide, da von den 20 proteinogenen Aminosäuren Cystein experimentell nicht untersucht werden konnte) erforderte eine effiziente experimentelle Strategie. Hierfür wurden die sogenannten Profilbibliotheken entwickelt (siehe 3.5.1), die auf den Ergebnissen des experimentellen PDZ-Screenings basierten (siehe 3.4). Für alle drei Domänen konnten aus den Substitutionsanalysen (Abbildung 3-31), unabhängig von der Sequenz der Ausgangs-peptide, verhältnismäßig konstante Aminosäurepräferenzen für die einzelnen Ligandenpositionen abgeleitet werden (Tabelle 3-16). Dies demonstrierte gleichzeitig, dass die Affinitätsbeiträge der einzelnen Ligandenpositionen weitgehend additiv zur Gesamtaffinität beitragen. Dadurch wurde der relevante Sequenzraum, der zum Training des QSP-Modells experimentell untersucht werden musste, für die jeweilige PDZ-Domäne auf etwa 500 bis 2000 Peptide eingegrenzt. Unter einer domänenspezifischen Profilbibliothek wird nun diejenige festphasengebundene Peptidbibliothek verstanden, die alle Peptide dieses Sequenzraums enthält (siehe 3.5.1). Über das Massenwirkungsgesetz konnten die Signalintensitäten der Inkubationen der festphasegebundenen Profilbibliotheken mit den jeweiligen PDZ-Domänen erfolgreich mit den unabhängig davon durch SPR bestimmten Affinitäten für einige Kontrollpeptide kalibriert werden (siehe 2.6.3.2 und 3.5.1).

Experimentelles
Training

An Hand der Profilbibliotheken wurden anschließend QSP-Modelle für diesen reduzierten Liganden-Sequenzraum trainiert. Technisch gesehen handelt es sich dabei um multifakto-

rielle Varianzanalysen (ANOVA), die eine lineare Regression zwischen den faktoriellen Variablen der Aminosäuresequenz der Liganden und den verhältnisskalierten Affinitäten erlauben (siehe 3.5.2).

Abhängig von der Formulierung der zugrundeliegenden ANOVA-Modelle – mit bzw. ohne statistische Interaktionsterme (Gleichung 3:1 bzw. Gleichung 3:2) – kann mit Hilfe der QSP-Methode die Kooperativität zwischen den einzelnen Ligandenpositionen zumindest innerhalb des reduzierten Liganden-Sequenzraums bestimmt werden. Erwartungsgemäß wurde für die *hAF6*-PDZ und die *hERBIN*-PDZ Domänen nur ein geringer kooperativer Beitrag (etwa 10%) zur Gesamtaffinität festgestellt (siehe 3.5.2 und Tabelle 3-17), der im Einklang mit den Ergebnissen der Substitutionsanalysen steht (Abbildung 3-31). Strukturell lässt sich dies durch die relativ unabhängigen Wechselwirkungen der Seitenketten der vier C-terminalen Reste mit konstanten Bereichen (Interaktionsflächen) auf der Domäne erklären (siehe 1.4.1 und 4.4). Nichts desto trotz können unter anderem geringe Veränderungen der Position des Liganden innerhalb der Bindungstasche oder elektrostatische Wechselwirkungen der Seitenketten des Liganden untereinander zu einer geringen Kooperativität führen. Obwohl die Methode der QSPs zwar hauptsächlich additive Affinitätsbeiträge für die Betrachtung des gesamten theoretischen Liganden-Sequenzraum voraussetzt, kann sie innerhalb des reduzierten Liganden-Sequenzraum durchaus zur Bestimmung der Kooperativität eingesetzt werden. Sie ermöglicht damit die Überprüfung ihrer eigenen Voraussetzung.

Kooperativität

Auf Grund der Feststellung der fast ausschließlich additiven Affinitätsbeiträge war es legitim, die QSP-Modelle der *hAF6*-PDZ und der *hERBIN*-PDZ Domänen durch die Integration der Signalintensitäten der Substitutionsanalysen zu erweitern. Durch diese finalen QSP-Modelle wird der komplette theoretische Liganden-Sequenzraum von 130321 Tetrapeptiden abgedeckt. Die Tatsache, dass sich die aminosäurespezifischen Affinitätsbeiträge der QSP-Modelle für den gesamten theoretischen Liganden-Sequenzraum nicht von denen der QSP-Modelle für den reduzierten Sequenzraum unterscheiden, bestätigt die Gültigkeit dieser Vorgehensweise (siehe Anhang B, Abbildung B-2 und Abbildung B-6). Mit dem QSP-Modell für die *mSNA1*-PDZ Domäne wurde schließlich überprüft, in wie weit ein solches Wechselwirkungsmodell auch ausschließlich auf den Daten von Substitutionsanalysen trainiert werden kann. Da für *mSNA1*-PDZ die experimentellen Ergebnisse (siehe 3.4) und die experimentell bestimmte Struktur [PDB: 2PDZ]¹⁶⁰ zeigen, dass auch $\lambda-4$ einen Beitrag zur Spezifität der Domäne liefert, wurde diese Position in das QSP-Modell mitaufgenommen. Damit deckt das QSP-Modell der *mSNA1*-PDZ Domäne den theoretischen Liganden-

Erweiterung des abgedeckten Liganden-Sequenzraums

Sequenzraum von $19^5 = 2\,476\,099$ Pentapeptiden ab. Die Qualität dieses Modells (Tabelle 3-17) ermöglicht zumindest eine semi-quantitative Interpretation der Spezifität der mSNA1-PDZ Domäne.

Zur Visualisierung der QSP-Modelle wurden – in Anlehnung an Termschemata der Energieniveaus von Molekülen – Termschemata der Energie- bzw. Affinitätsbeiträge der einzelnen Aminosäuren an den konservierten Ligandenpositionen entwickelt (Abbildung 3-33). Sie ermöglichen eine intuitive Interpretation der Spezifität einer Domäne. Jede Ligandenposition ist dabei durch eine vertikal angeordnete Verteilung der Aminosäuren repräsentiert. Diejenigen Aminosäuren, die in einer Ligandenposition für eine hohe Bindungsaffinität verantwortlich sind, finden sich am unteren Ende der Verteilung wieder (niedrige/negative $\Delta\ln(K_d)$), Aminosäuren mit einem negativen Beitrag zur Affinität am oberen Ende (hohe/positive $\Delta\ln(K_d)$). Die klassischen Sequenzmotive lassen sich aus den Aminosäuren am unteren Ende der einzelnen Verteilungen rekonstruieren, z.B. **(E/D) (T/S/V) x (V/L/I)**_{COOH} für hERBIN-PDZ (Abbildung 3-33C). Aus der Art der Verteilung der Aminosäuren kann die Selektivität für die jeweilige Ligandenposition abgelesen werden. So sind hochselektive Positionen durch zwei wohl separierte Cluster mit niedrigen bzw. hohen $\Delta\ln(K_d)$ -Werten charakterisiert (z.B. Position $\lambda-2$ in Abbildung 3-33E). Für unselektive Positionen würden sich die $\Delta\ln(K_d)$ -Werte aller Aminosäuren nicht signifikant von 0 unterscheiden. Eine wirklich unselektive Position konnte aber durch keines der QSP-Modelle identifiziert werden. Dagegen zeigen die Termschemata eine dritte Variante, die durch eine gleichmäßige Verteilung der Aminosäuren über einen größeren $\Delta\ln(K_d)$ -Bereich gekennzeichnet ist (z.B. Position $\lambda-1$ in Abbildung 3-33A). Solch eine Position ist zwar nicht deutlich selektiv, dennoch tragen die Aminosäuren signifikant unterschiedlich zur Affinität bei. $\lambda-1$ wurde dagegen in den klassischen Sequenzmotiven nur durch **x** repräsentiert und damit unpräzise als unselektiv eingestuft (Tabelle 3-15). Für hERBIN-PDZ wird die Selektivität, insbesondere für Tryptophan, an $\lambda-1$ durch die *Phage-Display-Studie* von Skelton *et al.*, 2003, unabhängig experimentell bestätigt.

Zusätzlich zur Beschreibung der Spezifität ermöglichen die QSP-Modelle vor allem auch die Affinitätsvorhersage für alle Tetrapeptide des 130321 großen Liganden-Sequenzraums. Der Vergleich der vorhergesagten mit den unabhängig durch SPR experimentell bestimmten Affinitäten (Tabelle 3-18) bestätigt den QSP-Modellen der hAF6-PDZ und hERBIN-PDZ Domänen eine hervorragende Vorhersageleistung (Tabelle 3-17). Die Abweichung der Modellvorhersagen liegen in derselben Größenordnung wie die Unterschiede der mit ver-

schiedenen experimentellen Methoden bestimmten Dissoziationskonstanten. Für das *hERBIN*-PDZ QSP-Modell wurde im Vergleich zum QSP-Modell für *hAF6*-PDZ eine stärkere Abweichung der Affinitätsvorhersagen festgestellt. Dies kann auf zusätzliche, durch das QSP-Modell nicht repräsentierte Wechselwirkungen von Aminosäuren in Positionen N-terminal zu $\lambda-3$ zurückgeführt werden. Für *hERBIN*-PDZ identifizierten Skelton *et al.*, 2003, und Birrane *et al.*, 2003, unabhängig voneinander eine zusätzliche hydrophobe Bindungsstelle in der $\beta\beta\beta\text{C}$ -Schleife. In den beiden Komplexstrukturen [PDB: 1N7T]¹⁶⁸ bzw. [PDB: 1MFG]²³ interagiert ein Tryptophan in $\lambda-4$ bzw. ein Tyrosin in $\lambda-7$ mit dieser Bindungsstelle. In der Tat korrelieren die stärker abweichenden Affinitätsvorhersagen des *hERBIN*-PDZ QSP-Modells auch fast ausnahmslos mit dem Vorhandensein aromatischer Aminosäuren in $\lambda-4$ bis $\lambda-7$ (Tabelle 3-18). Darüber hinaus wird ein Affinitätsbeitrag der Ligandensequenz N-terminal zu $\lambda-3$ auch dadurch demonstriert, dass für unterschiedliche Liganden mit identischer Sequenz von $\lambda-3$ bis $\lambda 0$ experimentell unterschiedliche Affinitäten bestimmt wurden. Am deutlichsten wird dies für die Wechselwirkung der *hERBIN*-PDZ Domäne mit ATB1, ATB2, CYG4 und CYG4k (Abbildung 3-33D und Tabelle 3-18). In die Beurteilung der Qualität der Affinitätsvorhersage muss auch der Fehler der experimentellen SPR-Messungen miteinbezogen werden. Dieser liegt für *hAF6*-PDZ und *hERBIN*-PDZ zwischen einem Faktor 1,2 und 1,3 (Tabelle 3-18). Der große Unterschied der experimentell bestimmten Dissoziationskonstanten für CYG4 und CYG4k, der sich nur schwer an Hand der unterschiedlichen Länge der Peptide erklären lässt, weist aber auf einen möglicherweise größeren experimentellen Fehler hin. Für das QSP-Modell der *mSNA1*-PDZ Domäne sind die Abweichungen der Affinitätsvorhersage zwar noch größer (Tabelle 3-18), aber angesichts der wenigen experimentellen Daten, die zum Training des Modells verwendet wurden, ist die Vorhersageleistung beachtenswert. Die Gültigkeit des Modells wird durch die Übereinstimmung des Termschemas (Abbildung 3-33E) mit den bisher bekannten Wechselwirkungsmotiven^{74,160} unterstützt. Im Gegensatz zu Sequenzmotiven erlaubt das QSP-Modell aber eine Aussage über die Stärke des Affinitätsbeitrags einzelner Aminosäuren und die Stärke der Wechselwirkung insgesamt.

Domänenspezifische Interaktionen mit Ligandenreste N-terminal zu $\lambda-3$

Beurteilung der Vorhersagequalität im Verhältnis zum Fehler der experimentell bestimmten K_d -Werte

Qualität des QSP-Modells der *mSNA1*-PDZ Domäne

Mit Hilfe dieser QSP-Modelle ist das Design der optimalen peptidischen Liganden trivial. Die Sequenz der sogenannten Superbinder lässt sich direkt aus den Termschemata ablesen, indem die Aminosäuren mit den höchsten Affinitätsbeiträgen (niedrigsten $\Delta\ln(K_d)$ -Werten) kombiniert werden (Abbildung 3-33A/C/E). Die vorhergesagten Superbinder konnten experimentell bestätigt werden, denn sie zeigten im Vergleich zu allen anderen experimentell

Design der optimalen peptidischen Liganden (Superbinder)

untersuchten Liganden die höchste Affinität gegenüber ihrer jeweiligen PDZ-Domäne (Tabelle 3-18). Der Superbinder SB-ERBIN zeigte dabei sogar eine vierfach höhere Affinität im Vergleich zum nächstbesten Liganden AVRC. Die letzten 4 C-terminalen Reste von SB-ERBIN stimmten auch mit dem Ergebnis einer *Phage-Display-Studie* für hERBIN-PDZ überein.¹⁶⁸

Die Größe des abgedeckten Liganden-Sequenzraums darf aber nicht darüber hinwegtäuschen, dass auch die QSP-Modelle die Spezifität einer Domäne nur dann sinnvoll beschreiben können, wenn alle durch das Modell abgedeckten Liganden über einen einzigen konservierten Liganden-Bindungsmechanismus erkannt werden. Das Screening der *Bibliothek humaner C-Termini* macht deutlich, wie wichtig es ist, die verschiedenen Bindungsmodi zu unterscheiden und durch getrennte Modelle zu beschreiben: erst in Folge der Gruppierung der mutmaßlich klassisch bzw. verschoben gebundenen Liganden konnten klare Sequenzmotive erhalten werden (Abbildung 3-29). Für das Training der QSP-Modelle wurden daher die Profilbibliotheken auf die letzten 4 C-terminalen Reste beschränkt, um Ligandensequenzen zu vermeiden, die ein verschobenes Motiv enthalten und daher auch über den alternativen Bindungsmechanismus gebunden werden können (Abbildung 3-30). Prinzipiell trifft diese Limitierung aber auch auf alle anderen in dieser Arbeit vorgestellten Wechselwirkungsmodelle sowie die Substitutionsanalysen zu. Wenn versehentlich unterschiedliche Bindungsmodi zusammengefasst werden, verlieren die Modelle ihre Aussagekraft.⁷³ Sollten sich in einem Experiment verschiedene Bindungsmodi verbergen, kann dies aber häufig an den „verrauschten“ Ergebnissen erkannt werden. Letztendlich müssen also für jeden Bindungsmechanismus separate Wechselwirkungsmodelle erstellt werden.

4.4 PDZ-Domänen Spezifitätsmechanismen

Mit Hilfe der Strukturen der drei untersuchten PDZ-Domänen (Abbildung 4-1) konnten die durch die QSP-Modelle quantifizierten Aminosäurepräferenzen an den einzelnen Ligandenpositionen erklärt werden. Dazu wurde die Definition der Interaktionsflächen (IF), die von Songyang *et al.*, 1997, eingeführt worden war, auf der Basis der Analyse aller experimentell bestimmten Strukturen der PDZ-Domänen/Liganden-Komplexe erweitert (siehe 1.4.1, Tabelle 1-4 und Abbildung 1-4).

Strukturbasierte
Erklärung der
Spezifitäten

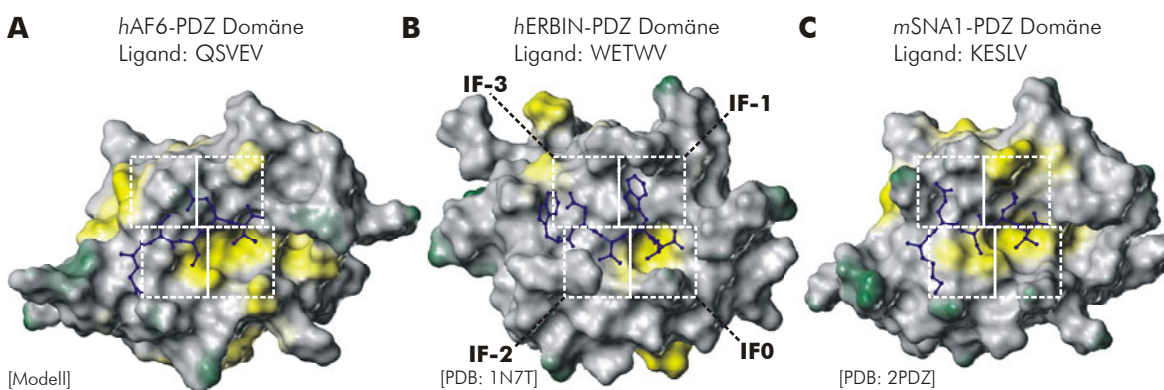


Abbildung 4-1 Interaktionsflächen der *hAF6*-PDZ, *hERBIN*-PDZ und *mSNA1*-PDZ Domänen
Oberflächendarstellung³⁹ (A) der *hAF6*-PDZ Domäne im Komplex mit dem Peptid **QSVEV** (modelliert),
(B) der *hERBIN*-PDZ Domäne im Komplex mit dem Peptid **WETWV** [PDB: 1N7T]¹⁶⁸ und (C) der *mSNA1*-
PDZ Domäne im Komplex mit dem Peptid **KESLV** [PDB: 2PDZ].¹⁶⁰ Hydrophobe Bereiche der Oberfläche
sind gelb, hydrophile Bereiche grün gefärbt.⁶⁰ Die Liganden sind in Blau dargestellt. Die konservierten
Interaktionsflächen (IF0 bis IF-3) sind durch gestrichelte Kästen angedeutet.

Die Präferenz für aliphatische Aminosäuren an Ligandenposition $\lambda 0$ lässt sich auf die ausgeprägte hydrophobe Tasche in der Interaktionsfläche **IF0** zurückführen. Darüber hinaus aber ließen sich die Unterschiede in den Präferenzen für V, I oder L nicht anhand der Unterschiede in der Ausformung der Tasche erklären.

Interaktion
 $\lambda 0$ und **IF0**

Dagegen korreliert die Fähigkeit, große aromatische Aminosäuren in $\lambda-1$ spezifisch zu binden, klar mit der Stärke der Ausbildung einer Tasche in **IF-1**. Für *hERBIN*-PDZ wird die Bedeutung dieser Tasche, speziell für die Erkennung eines Tryptophans, unabhängig von dieser Analyse durch eine experimentell bestimmte Komplexstruktur demonstriert.¹⁶⁸ Im Falle von *hAF6*-PDZ ist die Tasche weniger stark ausgeprägt, während sie im Falle von *mSNA1*-PDZ durch ein Phenylalanin in $\beta C7$ blockiert ist. Dies erklärt die von *hERBIN*-PDZ zu *mSNA1*-PDZ hin abnehmende Wichtigkeit der Ligandenposition $\lambda-1$ und die abnehmenden Beiträge aromatischer Aminosäuren in dieser Position (Abbildung 3-33A/C/E).

Interaktion
 $\lambda-1$ und **IF-1**

Interaktion
 $\lambda-2$ und **IF-2**

Im Gegensatz zu *mSNA1*-PDZ und *hERBIN*-PDZ zeigt die *hAF6*-PDZ Domäne eine ausgeprägte Dual-Spezifität (siehe 1.4.1), d.h. sie toleriert sowohl hydrophobe Aminosäuren als auch Threonin/Serin an Ligandenposition $\lambda-2$. Die Interaktion von *hAF6*-PDZ mit $\lambda-2$ beruht daher vermutlich vorrangig auf van-der-Waals-Kontakten mit der hydrophoben Oberfläche in **IF-2**. Darauf deutet auch die Präferenz für Threonin gegenüber Serin hin. Klasse I PDZ-Domänen erkennen aber klassischerweise S/T in $\lambda-2$ über die Ausbildung einer Wasserstoffbrücke mit Position $\alpha B1$ in **IF-2**, in der sich ein Histidin befindet. Klasse II PDZ-Domänen erkennen dagegen hydrophobe Aminosäuren an $\lambda-2$ über van-der-Waals-Kontakte mit **IF-2**. *hAF6*-PDZ, mit einem Glutamin an Position $\alpha B1$, stellt damit eine Übergangsform dieser beiden Interaktionsmechanismen dar. Die abnehmende Präferenz für Serin bzw. Threonin an $\lambda-2$ (*mSNA1*-PDZ > *hERBIN*-PDZ > *hAF6*-PDZ) deutet auf eine abnehmende Bedeutung der Wasserstoffbrücke mit $\alpha B1$ hin (Abbildung 3-33A/C/E). Dieser Verlust kann scheinbar, wie durch *hAF6*-PDZ verdeutlicht, über van-der-Waals-Kontakte mit **IF-2** kompensiert werden, so dass sowohl hydrophobe Aminosäuren als auch Serin bzw. Threonin toleriert werden.

Interaktion
 $\lambda-3$ und **IF-3**

Im Fall von *hAF6*-PDZ ist der Zugang zur Interaktionsfläche **IF-3** für Aminosäuren in Ligandenposition $\lambda-3$ größtenteils durch ein Valin in Position $\beta B4$ blockiert. Dies erklärt die Präferenz für die kleinen Aminosäuren G, A, P und S (π) in $\lambda-3$, denn dadurch werden abstoßende Kontakte mit einem Valin in $\beta B4$ vermieden. Glutamat in $\lambda-3$ wird dagegen toleriert, weil die vorteilhafte elektrostatische Interaktion mit Lysin in $\beta C4$ die abstoßenden Kontakte mit $V\beta B4$ kompensiert. Im Gegensatz dazu haben sowohl *hERBIN*-PDZ als auch *mSNA1*-PDZ eine ausgeprägte Tasche in **IF-3**, welche die Seitenkette von $\lambda-3$ aufnehmen kann. Das Arginin (*hERBIN*-PDZ) bzw. das Lysin (*mSNA1*-PDZ) in $\beta C5$ bedingt dabei die Präferenz für Glutamat und in geringerem Maße auch für Aspartat an $\lambda-3$.

Gleichberechtigung
der Ligandenreste $\lambda 0$
bis $\lambda-3$

Zusammen bestätigt diese strukturelle Analyse den Befund der QSP-Modelle, dass die letzten 4 C-terminalen Ligandenpositionen im Prinzip als gleichberechtigt für die Spezifität betrachtet werden müssen. Die scheinbar unselektive Position $\lambda-1$ trägt sehr wohl zur Feinspezifität der Domänen bei. Allein die Annahme eines „Reißverschlussmechanismus“ bei der Ligandenbindung könnte eine absteigende Priorisierung der Ligandenpositionen vom zuerst fixierten C-Terminus aus bedingen. Allerdings gibt es für solch eine Hypothese keine experimentelle Unterstützung.

Zusätzliche optionale
Interaktionen von
Ligandenreste N-
terminal $\lambda-3$

Zusätzlich zu den in der gesamten Familie der PDZ-Domänen konservierten Interaktionsflächen **IF0** bis **IF-3** können *mSNA1*-PDZ und *hERBIN*-PDZ auch spezifisch mit Resten in

$\lambda-4$ bzw. $\lambda-5$ interagieren. Im Falle von *mSNA1*-PDZ wird ein Arginin oder Lysin in Ligandenposition $\lambda-4$ spezifisch durch das Aspartat in $\alpha B2$ gebunden. Im Falle von *hERBIN*-PDZ können aromatische Aminosäuren N-terminal von $\lambda-3$ im Liganden mit einer hydrophoben Tasche in der $\beta\beta C$ -Schleife der Domäne interagieren.^{23,168} Vor allem die Substitutionsanalysen für *hERBIN*-PDZ demonstrieren, dass diese Interaktionen vorrangig optional und häufig nur auf bestimmte Domänen/Liganden-Paare beschränkt sind (Abbildung 3-31).

Ein weiteres erstaunliches Ergebnis ist, dass alle drei PDZ-Domänen im Screening der *Bibliothek humaner C-Termini* mit einer relativ großen Anzahl an Liganden, die durch ein genau um eine Ligandenposition N-terminal verschobenes Erkennungsmotiv gekennzeichnet sind, interagierten (Tabelle 3-14). Unterstützt durch eine Moleküldynamik-Simulation und experimentelle Untersuchungen konnte ein plausibler alternativer Liganden-Bindungsmechanismus abgeleitet werden (Abbildung 3-30). Die experimentelle Strukturaufklärung dieses Komplexes muss nun endgültig klären, welche Rolle die noch notwendige C-terminale Carboxylgruppe in Komplexen mit „verschobenen Liganden“ spielt und welchen Platz die Seitenkette der Aminosäure in $\lambda 1$ einnimmt. Dass PDZ-Domänen auch generell interne Motive binden können, die nicht eine β -Finger-Konformation einnehmen,⁶⁹ konnte durch die experimentelle Aufklärung der Komplexstruktur von Par6-PDZ und Pals1 gezeigt werden [PDB: 1X8S].¹³⁶ Auch in dieser Struktur befindet sich zusätzlich eine Aminosäure in einer Ligandenposition $\lambda 1$ innerhalb der konservierten Bindungstasche. Im Gegensatz zu dem in dieser Arbeit vorgeschlagenen Wechselwirkungsmechanismus der verschobenen Motive wird in der Par6-PDZ Komplexstruktur scheinbar kein freier C-Terminus an $\lambda 1$ benötigt, da weitere Ligandenreste in C-terminaler Richtung folgen. Die Wasserstoffbrücken der C-terminalen Carboxylgruppe des Liganden werden allerdings von der Carboxylgruppe der Seitenketten des Aspartats in $\lambda 1$ imitiert. Obwohl diese Struktur nicht genau den in dieser Arbeit identifizierten Bindungsmechanismus wiedergibt, verdeutlicht sie dennoch, dass eine zusätzliche Aminosäure in der Liganden-Bindungstasche einer PDZ-Domäne Platz finden kann. Im Gegensatz zur Bindung des „verschobenen“ Motivs, wurde keine signifikante Wechselwirkung mit internen Motiven für die drei in dieser Arbeit untersuchten PDZ-Domänen festgestellt. Damit bleibt zu klären, welche strukturellen Determinanten eine PDZ-Domäne zu einer solchen Wechselwirkung befähigen. Die Tatsache, dass alle drei PDZ-Domänen mit verschobenen Motiven interagieren können, deutet an, dass es sich hierbei um eine prinzipielle Eigenschaft von PDZ-Domänen handeln könnte. Damit müssen dem Wechselwirkungs-Repertoire der PDZ-Domänen (C-Termini, β -Finger, Oligomerisierung

Alternativer Bindungsmechanismus für „verschobene“ Motive

Vergleich mit Par6-PDZ/Pals1-Komplex

Erweiterung des Wechselwirkungs-Repertoires

und Lipide) zwei weitere Mechanismen hinzugefügt werden. Das Besondere an der Wechselwirkung mit verschobenen Motiven (Abbildung 3-30) bzw. internen Motiven¹³⁶ ist, dass sie mit dem Paradigma der Erkennung C-terminaler Motive brechen. Damit müssen zukünftig bei der sequenzbasierten Suche von PDZ-Domänen-Liganden eine größere Anzahl an Möglichkeiten in Betracht gezogen werden.

4.5 Selektivität von Protein-Interaktionsdomänen

Quantifizierung der Selektivität

Die *Quantitativen Spezifitäts-Profile* eröffnen eine weitere sehr interessante Anwendung, die Quantifizierung der Selektivität einer Protein-Interaktionsdomäne. Die Selektivität bezeichnet dabei die Trennschärfe einer Domäne zwischen „gewünschten“ und „ungewünschten“ Wechselwirkungspartnern. Sie kann entweder durch die Steilheit der Affinitätsfunktion in Abhängigkeit von der Anzahl an gebundenen Liganden beschrieben werden, oder durch die Größe des Sequenzraums, der von denjenigen Liganden gebildet wird, die mit mindestens einer als Schwellenwert gewählten Affinität gebunden werden (Abbildung 3-35). Dadurch wird die Konkurrenz zwischen den Liganden um die Bindung einer einzelnen Domäne beschrieben. Umgekehrt kann die Konkurrenz mehrerer Domänen um einen Liganden analysiert werden, wenn QSP-Modelle die zu vergleichenden Domänen vorliegen. Letzten Endes liefern die QSPs also einen wertvollen Beitrag zur Systembiologie, die sich mit der Simulation und dem quantitativen Verständnis von komplexen Prozessen einer Zelle beschäftigt.^{72,174,216} Hierbei spielt natürlich die Konkurrenz um Wechselwirkungspartner eine wesentliche Rolle.

Ziele der Quantifizierung der Selektivität

Quantifizierung der Selektivität der einzelnen Domänen

Mit Hilfe der QSP-Modelle für die drei PDZ-Domänen aus *hAF6*, *hERBIN* und *mSNA1* wurde die Selektivitäten dieser Domänen innerhalb des theoretischen Sequenzraums für die 4 C-terminalen Reste analysiert. Dabei zeigen die drei PDZ-Domänen wesentliche Unterschiede in der Größe der erkannten Liganden-Sequenzräume. Den größten Sequenzraum erkennt dabei *mSNA1*-PDZ, während *hERBIN*-PDZ nur mit einem relativ kleinen Liganden-Sequenzraum interagiert (Abbildung 3-35). Es ist bemerkenswert, dass sich dies nicht aus den klassischen Erkennungsmotiven erschließt (Tabelle 3-15). Interessanterweise korreliert dieser Befund mit der zellulären Funktion der beiden Proteine: so bindet *hSNA1* an das Dystrophin/Utrophin-Netzwerk und agiert als Vermittler zwischen diesem cytoskelett-assoziierten Komplex und einer vermutlich großen Anzahl unterschiedlicher Proteine.⁷⁴ Dagegen

Relevanz der quantifizierten Selektivität

besitzt *hERBIN* eine spezifischere Rolle als Suppressor des Ras/Raf-Signalwegs⁹⁵ und interagiert dementsprechend mit nur wenigen Proteinen.

Überraschend ist auch, dass die Spezifitäten der drei untersuchten PDZ-Domänen im Bereich physiologisch relevanter Affinitäten verhältnismäßig stark überlappen (Abbildung 3-35). Dieser Befund legt die Vermutung nahe, dass ein Protein über ein einzelnes, C-terminales PDZ-Domänen-Erkennungsmotiv eine Reihe unterschiedlicher PDZ-Domänen-Proteine binden kann, um damit unterschiedliche Funktionen zu rekrutieren.

Überlappung der Spezifität der drei PDZ-Domänen

Mit den QSP-Modellen wird allerdings nur der Einfluss der Spezifität auf die Selektivität betrachtet. *In vivo* wird die Selektivität einer Protein-Interaktionsdomäne aber auch durch die prinzipielle Verfügbarkeit der potentiellen Liganden bestimmt. Die Verfügbarkeit wiederum wird durch eine Reihe von Faktoren beeinflusst, die im Rahmen dieser Arbeit nicht untersucht wurden: so müssen die beiden theoretisch wechselwirkenden Proteine zur selben Zeit in der Zelle exprimiert werden, damit die Wechselwirkung stattfinden kann (Expression). Beide Proteine müssen sich im selben Kompartiment innerhalb der Zelle befinden (Kompartimentalisierung). Das Bindungsmotiv muss strukturell zugänglich sein (Konformation). Die Wechselwirkung kann ferner durch sekundäre Modifikation des Erkennungsmotivs verhindert bzw. ermöglicht werden (sekundäre Modifikation). Und schließlich gibt es in der Zelle zusätzliche Konkurrenz durch nicht verwandte Protein-Interaktionsdomänen¹⁷⁸ und andere Wechselwirkungspartner, die in die theoretischen Überlegungen nicht miteingegangen sind (Konkurrenz). Diese zusätzlichen Ebenen der Selektivität erklären, wie trotz eines scheinbar „unselektiven“ Verhaltens – wie es für einige PDZ- und WW-Domänen festgestellt wurde – *in vivo* hoch selektive Wechselwirkungen zustande kommen können. Nichts desto trotz kann, wie weiter oben beschrieben, auch ein relativ unselektives Verhalten eine definierte Funktion innerhalb eines feinabgestimmten Wechselwirkungsnetzwerks spielen. Gleichzeitig erlaubt jede dieser Ebenen auch die gezielte Regulation der Wechselwirkung. Wie schon im zweiten Absatz diskutiert, kann aber auch eine verhältnismäßig unselektive Ligandenbindung eine Funktion *in vivo* erfüllen. Dabei muss beachtet werden, dass die funktionale Selektivität auch von der Verteilung der Erkennungsmotive in den Zielproteinen abhängt. So ist durchaus ein Szenario vorstellbar, in dem eine Protein-Interaktionsdomäne zwar hochselektiv mit einem gut definierten Erkennungsmotiv interagiert, dieses aber in einer Vielzahl von Proteinen mit unterschiedlichsten Funktionen vorkommt (siehe beispielsweise die NEDD4-Ubiquitin-Ligasen). In diesem Fall kann die Wechselwirkung zwar mechanistisch als

In vivo Selektivität

Weitere Selektivitätsfaktoren

Funktionale versus mechanistische Selektivität

selektiv bezeichnet werden, funktional aber eher als unselektiv. Dies macht deutlich, dass zwischen diesen beiden Ebenen der Selektivität ebenfalls unterschieden werden muss.

4.6 Schlussfolgerungen

Die Ergebnisse für die WW- und die PDZ-Domänen zeigen zusammen, dass sich die Bindungseigenschaften von Protein-Interaktionsdomänen effizient durch die gezielte Kombination von Hochdurchsatz-Screeningmethoden mit theoretischen Modellen charakterisieren lassen. Die resultierenden Wechselwirkungsmodelle vermitteln nicht nur ein Verständnis der Wechselwirkungs- und Spezifitätsmechanismen, sondern sie ermöglichen auch die Vorhersage potentieller neuer Wechselwirkungspartner. Die quantitativen Modelle liefern dabei sogar Informationen über die Affinität der Wechselwirkung. Zusammen unterstützten die entwickelten Wechselwirkungsmodelle die theoretische und experimentelle Analyse von Protein-Protein-Wechselwirkungsnetzwerken auf vielfältige Weise:

- (i) Mit Hilfe der Vorhersage potentieller Wechselwirkungspartner kann die Zahl der aufwendigen, zellbiologischen Experimente reduziert werden.
- (ii) Das umfassende Bild der Spezifität in Form der QSP-Modelle erlaubt, alle potentiell möglichen Wechselwirkungen in das Experimentdesign und die Netzmodelle einzubeziehen.
- (iii) Die Vorhersage der Wechselwirkungen ermöglicht die kritische Beurteilung der Wechselwirkungsdaten, die mit Hilfe von Hochdurchsatzverfahren gewonnen wurden.
- (iv) Die Affinitätsvorhersage, insbesondere die der QSP-Modelle, liefert die Grundlage für die Modellierung von Protein-Protein-Interaktionsnetzwerken im Rahmen der Systembiologie.
- (v) Die COMBINE- und QSP-Modelle ermöglichen das Design peptidischer Liganden mit gewünschter Affinität; das COMBINE-Modells auch für Peptidomimetika.

Schließlich wurde mit der Methode der *Quantitativen Spezifitäts-Profile* eine allgemein anwendbare und effiziente Strategie zur Charakterisierung von Protein-Interaktionsdomänen entwickelt. Die resultierenden QSP-Modelle liefern dabei ein umfassendes Bild der Spezifität und Selektivität einer Protein-Interaktionsdomäne basierend auf den biophysikalischen Eigenschaften der Wechselwirkung.