

Forschungsinstitut für Molekulare Pharmakologie, Berlin

Analyse und Vorhersage der Wechselwirkungen von Protein-Interaktionsdomänen

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im
Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von

Urs Wiedemann
aus München

April, 2005

1. Gutachter: Prof. Dr. Hartmut Oschkinat
2. Gutachter: Prof. Dr. Jens Schneider-Mergener

Disputation am 16.6.2005

Meinem Großvater
Hubert Wiedemann

Das Ganze ist mehr als die Summe seiner Teile.
– *Aristoteles*

Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise.

– *John W. Tukey Ann. Math. Stat.* **33** (1962)

Danksagung

Prof. Dr. Hartmut Oshkinat danke ich für die Vergabe des interessanten Promotionsthemas und für die Möglichkeit diese Arbeit in der Gruppe NMR-unterstützte Strukturbiologie am Forschungsinstitut für molekulare Pharmakologie durchführen zu können.

Dr. Gerd Krause danke ich für die zurückhaltende aber immer zielgerichtete Betreuung meiner Arbeit. Seine wertvollen Anregungen und Ratschläge habe ich immer geschätzt.

Prof. Dr. Jens Schneider-Mergener und Dr. Rudolf Volkmer-Engert danke ich dafür, dass sie die Kooperationsprojekte zu WW- und PDZ-Domänen ermöglicht haben.

Mein ganz besonderer Dank gilt Dr. Livia Otte und Dr. Prisca Boisguerin für die immer ausgesprochen angenehme Atmosphäre, die wertvollen Diskussionen und die produktive wissenschaftliche Zusammenarbeit im Rahmen der WW- und PDZ-Domänen-Projekte. Livia, ich danke Dir von ganzem Herzen für die unvergessliche Zeit unserer Zusammenarbeit und die daraus entstandene, für mich sehr wichtige Freundschaft.

Für die Durchführung des NMR-Screenings danke ich Brigitte Schlegel, Dr. Peter Schmieder und Dr. Ricardo Pires. Henrik Holtmann danke ich sehr für die Durchführung der UV-Fluoreszenzmessungen zu dem durch das COMBINE-Modell vorhergesagten Peptid.

Dr. Rebecca Wade danke ich dafür, dass sie das Kooperationsprojekt zu 3D-QSAR-Analysen der WW-Domänen ermöglicht hat. Besonders danken möchte ich Dr. Karin Schleinkofer für die angenehme und produktive Zusammenarbeit in diesem Projekt.

Daneben möchte ich allen ehemaligen und aktuellen Mitgliedern der Gruppe NMR-unterstützte Strukturbiologie für die stets kollegiale Atmosphäre danken. Insbesondere möchte ich Andrea Steuer für ihre Unterstützung danken. Ganz herzlich danke ich Dr. Dirk Labudde und Dr. Dietmar Leitner für ihre Ratschläge und Diskussionen sowie für Ihre freundschaftliche Aufnahme am FMP. Dr. Ricardo Pires danke ich herzlich für die hilfreichen wissenschaftlichen Diskussionen, die angenehme Atmosphäre in unserem Büro und die daraus erwachsene Freundschaft.

Dr. Claudia Gutjahr danke ich ganz herzlich für das Zuhause auf Zeit in Berlin, das sie mir völlig unkompliziert geboten hat.

Meinen Eltern danke herzlich ich für ihre fortwährende Unterstützung und ihr Interesse an meiner Arbeit.

Meiner Frau Dr. Ricarda Bauer danke ich von ganzem Herzen für ihre unermüdliche Unterstützung, ihre Liebe und Motivation.

Inhaltsverzeichnis

Abkürzungen und Konventionen	vi
Kooperationen.....	ix
1 Einleitung	1
1.1 Protein-Protein-Wechselwirkungsnetzwerke und Protein-Interaktionsdomänen	1
1.2 Prinzipien der Protein/Liganden-Wechselwirkung.....	4
1.3 WW-Domänen.....	7
1.3.1 Struktur und Funktion prolinreicher Sequenzmotive.....	8
1.3.2 WW-Domänen/Liganden-Bindungsmechanismus.....	9
1.3.3 Funktioneller Kontext der WW-Domänen.....	12
1.4 PDZ-Domänen.....	13
1.4.1 PDZ-Domänen/Liganden-Bindungsmechanismus.....	14
1.4.2 Funktioneller Kontext der PDZ-Domänen.....	17
1.4.3 Untersuchte PDZ-Domänen.....	18
1.5 Analyse und Vorhersage von Protein-Protein-Wechselwirkungen.....	20
1.5.1 Experimentelle Analyse von Protein-Protein-Wechselwirkungen.....	20
1.5.2 Theoretische Vorhersage von Protein-Protein-Wechselwirkungen	21
1.6 Zielsetzung und Strategie	23
2 Material und Methoden	25
2.1 Nomenklatur-Regeln	25
2.1.1 Nomenklatur der Proteine und Domänen	25
2.1.2 Nomenklatur der homologen Aminosäurereste in Protein-Interaktionsdomänen/Liganden-Komplexen	25
2.1.2.1 Nomenklatur der WW-Domänen	26

2.1.2.2	Nomenklatur der PDZ-Domänen	26
2.1.3	Nomenklatur der Sequenzmotive	27
2.2	Herstellung und Reinigung der Protein-Interaktionsdomänen und Peptidliganden	28
2.2.1	Synthese und Reinigung der WW-Domänen und der prolinreichen Liganden	28
2.2.2	Expression und Reinigung der PDZ-Domänen	30
2.2.3	Synthese und Reinigung C-terminaler Peptide	31
2.3	Liganden-Screening der Protein-Interaktionsdomänen	32
2.3.1	NMR-Spektroskopie	32
2.3.1.1	Bestimmung des Faltungszustands der Domänen	32
2.3.1.2	Quantifizierung der Wechselwirkungsaktivität der WW-Domänen	33
2.3.2	Herstellung und Inkubation festphasegebundener Protein- und Peptidbibliotheken	34
2.3.2.1	WW-Array: Bibliothek 42 synthetisch hergestellter WW-Domänen	34
2.3.2.2	Peptidbibliotheken: SPOT-Synthese und Inkubation	34
2.3.2.3	Direkte Markierung von Peptiden und Proteinen mit Peroxidase	35
2.3.2.4	Bindungsstudien mit festphasegebundenen Bibliotheken	35
2.4	Bestimmung der Domänen/Liganden-Affinitäten	36
2.4.1	UV-Fluoreszenzspektroskopie an WW-Domänen	36
2.4.2	Oberflächen-Plasmon-Resonanz-Messungen (SPR) an PDZ-Domänen	37
2.5	Allgemeine statistische Methoden	37
2.5.1	MAD-Z Standardisierung	37
2.5.2	Validierung statistischer Modelle: Determinationskoeffizient und SDEP	38
2.6	Quantitative Auswertung der Screenings mittels festphasegebundener Bibliotheken	39
2.6.1	WW-Array	39
2.6.2	Bibliothek humaner C-Termini	39
2.6.3	Quantitative Spezifitäts-Profile	39
2.6.3.1	Ableitung des Zusammenhangs zwischen der Signalintensität (in BLU) und den zugrundeliegenden Affinitäten	39
2.6.3.2	Kalibrierung der Signalintensität der Profilbibliotheken	41
2.6.3.3	Quantifizierung der Sequenz-Affinitäts-Beziehung: Quantitative Spezifitäts-Profile	41

2.7	Methoden der Bioinformatik.....	42
2.7.1	Sequenzdatenbanken	42
2.7.1.1	WW-Domänensequenzen aus der SMART-Datenbank	42
2.7.2	Sequenzmethoden.....	42
2.7.2.1	Konsensusfärbung von Sequenzalignments.....	42
2.7.2.2	Konsensussequenz.....	43
2.7.2.3	Sequenzlogos	43
2.7.2.4	Sequenzverwandtschaftsbaum	43
2.7.2.5	Profile-Hidden-Markov-Modelle	43
2.7.3	Strukturmethoden.....	44
2.7.3.1	Hochdurchsatz-Modellierung der 42 WW-Domänen und der WW-Domänen/ Liganden-Komplexe.....	44
2.7.3.2	CoMFA – Comparative-Molecular-Field-Analyse	45
2.7.3.3	COMBINE – Comparative-Binding-Energy-Analyse.....	45
2.7.3.4	Modellierung der hAF6-PDZ Domäne.....	47
2.7.3.5	Moleküldynamik-Simulation des hERBIN-PDZ/METFVSCOOH-Komplexes	47
3	Ergebnisse	49
3.1	Experimentelle Charakterisierung der Wechselwirkungen von 42 WW-Domänen	51
3.2	Qualitative Sequenz-Aktivitäts-Beziehungen	56
3.2.1	Experimentelle Bestimmung der Liganden-Epitope.....	56
3.2.1.1	Y-Spezifitätsgruppe: tyrosinspezifische WW-Domänen	56
3.2.1.2	R-Spezifitätsgruppen: argininspezifische WW-Domänen.....	57
3.2.1.3	L-Spezifitätsgruppe: leuzinspezifische WW-Domänen.....	59
3.2.1.4	poly-P-Spezifitätsgruppe: polyprolinspezifische WW-Domänen	60
3.2.1.5	poS/poT-Spezifitätsgruppe: phosphoserin-/phosphothreonin- spezifische WW-Domänen	60
3.2.2	WW-Domänen-Spezifitätsgruppen auf der Basis experimentell bestimmter Ligandenpräferenzen	61
3.2.3	Theoretische Bestimmung der Domänen-Epitope	62
3.2.3.1	Konsensussequenzen der Spezifitätsgruppen	62
3.2.3.2	Funktionaler Verwandtschaftsbaum der WW-Domänensequenzen	66

3.2.3.3	Sequenzbasierte Vorhersage der Spezifität: Wechselwirkungen der WW-Domänen mit tyrosinhaltigen Liganden	68
3.3	Quantitative Struktur-Aktivitäts-Beziehungen	74
3.3.1	Modellierung der Strukturen der 42 WW-Domänen und ihrer Komplexe	74
3.3.1.1	Homologie-Modellierung von WW-Domänen.....	75
3.3.1.2	Qualität der modellierten Domänenstrukturen	75
3.3.1.3	Analyse der Konformation und Positionierung der von WW-Domänen gebundenen Liganden	76
3.3.1.4	Homologie-Modellierung der Strukturen der WW-Domänen/Liganden-Komplexe.....	78
3.3.1.5	Qualität der modellierten Komplexstrukturen	82
3.3.2	3D-QSAR der Domänenstrukturen zur Vorhersage der Affinität gegenüber tyrosinhaltigen Liganden	83
3.3.2.1	CoMFA-Modell der Tyrosin-Spezifität für WW-Domänen	84
3.3.2.2	Strukturbasierte Vorhersage der Tyrosin-Spezifität	87
3.3.3	3D-QSAR der Komplexstrukturen	88
3.3.3.1	Spezifitätsbestimmende Interaktionen der WW-Domänen/R-Liganden-Wechselwirkung	89
3.3.3.2	COMBINE-Modell der WW-Domänen/Y-Liganden-Wechselwirkung.....	91
3.3.3.3	Design peptidischer Liganden mit gesteigerter Affinität.....	94
3.4	Analyse von experimentellen Wechselwirkungsdaten für PDZ-Domänen	97
3.5	Quantitative Sequenz-Aktivitäts-Beziehungen	103
3.5.1	Profilbibliotheken.....	104
3.5.2	Quantitative Spezifitäts-Profile der PDZ-Domänen.....	105
3.5.3	Design peptidischer Liganden mit maximaler Affinität: superbindende Peptide ..	113
3.5.4	Quantifizierung der PDZ-Domänen/Liganden-Selektivität.....	115
4	Diskussion	119
4.1	Wechselwirkungsmodelle auf unterschiedlichen Informationsebenen.....	120
4.1.1	Sequenzbasierte Klassifizierung	120
4.1.2	Strukturbasierte Wechselwirkungsmodelle	126
4.1.3	Komplexstrukturbasierte Wechselwirkungsmodelle	130
4.2	WW-Domänen Spezifitätsmechanismen	137

4.3	Quantitative Spezifitäts-Profile: eine quantitative Beschreibung der Spezifität	145
4.4	PDZ-Domänen Spezifitätsmechanismen	151
4.5	Selektivität von Protein-Interaktionsdomänen	154
4.6	Schlussfolgerungen.....	156
Zusammenfassung		157
Summary		159
Literatur		161
Anhang		172
A	Profile-Hidden-Markov-Modelle	172
B	Quantitative Spezifitäts-Profile (QSP).....	173
C	Veränderung der chemischen Verschiebung.....	183
D	Grundlagen der 3D-QSAR Methoden	184

Abkürzungen und Konventionen

Δ	Differenz
ΔG	freie Gibbs Energie
ΔH	Reaktionsenthalpie
ΔS	Entropie
Φ -Winkel	Torsionswinkel des Peptidrückgrats durch Rotation um die N-C α -Bindung
Ψ -Winkel	Torsionswinkel des Peptidrückgrats durch Rotation um die C α -C-Bindung
σ	Standardabweichung
1D/3D	Ein-/Dreidimensional
AF6	<i>all-1 Fusion Partner on Chromosome 6</i> [UP: P55196]
ANOVA	Varianzanalyse (<i>engl. analysis of variance</i>)
BLU	<i>Boehringer-Light-Units</i>
COMBINE	<i>Comparative-Binding-Energy-Analyse</i>
CoMFA	<i>Comparative-Molecular-Field-Analyse</i>
CSP	Veränderung der chemischen Verschiebung (<i>engl. chemical shift perturbation</i>)
CT	Carboxyterminus eines Peptids/Protein (C-Terminus)
C α	C α -Atom einer Aminosäure
DNA	Desoxyribonukleinsäure (<i>engl. deoxyribonucleic acid</i>)
DTT	1,4-Dithiothreitol
SH3	<i>Src homology 3</i> Domäne
EDTA	Ethylendiamin-N,N,N',N'-tetraessigsäure
Evh1	<i>Ena/Vasp-homology 1</i> Domäne
ERBIN	<i>ErbB2 interacting protein</i> [UP: Q9NR18]
Fmoc	9-Fluorenylmethyloxycarbonyl
GST	Glutathion-S-Transferase
<i>I</i>	Signalintensität
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
K_a	Assoziationskonstante
K_d	Dissoziationskonstante
ln	natürlicher Logarithmus
MAD	Median der absoluten Abweichungen

MALDI-TOF	<i>Matrix-assisted laser desorption and ionisation - time of flight</i>
MD	Moleküldynamik
M_{Fehler}	Multiplikativer Fehler geschätzt an Hand der SDEP
n.b.	nicht bestimmt
NMR	Kernmagnetische Resonanz (<i>engl. nuclear magnetic resonance</i>)
NT	Aminoterminus eines Peptids/Proteins (N-Terminus)
PAM	Mutationsdatenmatrix von Margret Dayhoff ⁴⁴ (<i>engl. percent accepted mutations</i>)
PCA	Hauptkomponenten-Analyse (<i>engl. principal component analysis</i>)
PDB	<i>Protein Data Bank</i> ²⁰ (http://www.rcsb.org/pdb)
ppm	<i>parts per million</i>
PDZ	PDZ-Domäne (Akronym aus den Proteinnamen <u>P</u> SD95, <u>D</u> LG und <u>Z</u> O1)
PFAM	<i>Protein families database of alignments and HMMs</i> ¹⁶ (http://www.sanger.ac.uk/Pfam)
pHMM	<i>Profile-Hidden-Markov-Modell</i>
PLS	multivariates Regressionsverfahren (<i>engl. partial least squares</i>)
PPII	Polyprolin-Typ-II-Helix
Q^2	kreuzvalidierter Determinationskoeffizient
QSAR	quantitative Struktur-Aktivitäts-Beziehung (<i>engl. quantitative structure-activity relationship</i>)
QSP	<i>Quantitative Spezifitäts-Profile</i>
R	Allgemeine Gaskonstante
R^2	Determinationskoeffizient
RG	Peptidrückgrat
RMSD	Mittlere quadratische Abweichung (<i>engl. root mean square deviation</i>)
RNA	Ribonukleinsäure (<i>engl. ribonucleic acid</i>)
RP-HPLC	Reverse-Phase-Hochleistungsflüssigkeitschromatographie (<i>engl. reversed phase – high performance liquid chromatography</i>)
SA	Substitutionsanalysen
SB-	Präfix für den Superbinder einer Domäne
SDEP	Standardabweichung der Fehler der Vorhersage (<i>engl. standard deviation of errors of prediction</i>)
SMART	<i>Simple Modular Architecture Research Tool</i> ¹⁶¹ (http://smart.embl-heidelberg.de)
SNA1	α 1-Syntrophin [UP: Q13424]

SPR	Oberflächen-Plasmon-Resonanz-Verfahren (engl. <i>surface plasmon resonance</i>)
T	Temperatur
UP	UniProt ¹⁰ Protein Datenbank (http://www.expasy.uniprot.org)
UP-ID	Zugriffscod eines Proteins in der UniProt ¹⁰ Protein Datenbank
UV	Ultraviolett
νR^2	validierter Determinationskoeffizient
wt	Wildtyp/Ausgangspeptid in Substitutionsanalysen
WW	WW-Domäne (Akronym aus zwei konservierten Tryptophanresten (W))

Darüber hinaus werden übliche Abkürzungen verwendet.

Bezeichnung konservierter Elemente/Positionen innerhalb homologer Strukturen:

(siehe auch Material und Methoden Abschnitt 2.1.2)

$\alpha A, \alpha B \dots$	erste, zweite, ... α -Helix
$\beta A, \beta B \dots$	erster, zweiter, ... β -Strang
λ	Ligandenposition
ω	WW-Domänenposition

Nomenklatur und Symbole für Aminosäuresequenzen und Sequenzmotive:

L-Aminosäuren werden nach dem Einbuchstabencode bezeichnet.

(siehe auch Material und Methoden Abschnitt 2.1.3)

ac-	acetylierte N-terminale Aminogruppe
-NH₂	amidierte C-terminale Carboxylgruppe
Φ	hydrophobe Aminosäuren ¹
Ψ	aliphatische Aminosäuren ¹
π	kleine Aminosäuren ¹
β	β -Alanin
\mathbf{x}	alle proteinogenen L-Aminosäuren
\mathbf{J}	6-Aminohexansäure
COOH	C-terminale Carboxylgruppe
\mathbf{b}	alle proteinogenen L-Aminosäuren außer C , M und W
po	Präfix für phosphorylierte Aminosäuren

Nomenklatur der Proteine, Domänen und Abkürzungen der Organismen siehe Material und Methoden Abschnitt 2.1.1

Kooperationen

Dr. Livia Otte

Molecular Libraries and Recognition Group

Institut für Medizinische Immunologie

Universitätsklinikum Charité

Hessische Str. 3-4

10115 Berlin

Synthese der WW-Domänen und Screening mittels festphasegebundener Bibliotheken

Dr. Prisca Boisguerin

Molecular Libraries and Recognition Group

Institut für Medizinische Immunologie

Universitätsklinikum Charité

Hessische Str. 3-4

10115 Berlin

Synthese der PDZ-Domänen und Screening mittels festphasegebundener Bibliotheken sowie

Oberflächen-Plasmon-Resonanz-Messungen

Dr. Karin Schleinkofer

Molecular and Cellular Modeling Group

EML Research gGmbH

Villa Bosch

Schloss-Wolfsbrunnenweg 33

69118 Heidelberg

Gemeinsame Durchführung der Comparative-Binding-Energy-Analyse der WW-Domänen

