
Entwicklung von Medikamenten gegen Prionkrankheiten

Thomas Hirschberger



München 2007

Entwicklung von Medikamenten gegen Prionkrankheiten

Thomas Hirschberger

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Thomas Hirschberger
aus München

München, den 24.05.2007

Erstgutachter: Prof. Paul Tavan
Zweitgutachter: Prof. Erwin Frey
Tag der mündlichen Prüfung: 11.07.2007

Zusammenfassung

Die Prionkrankheiten oder Transmissiblen Spongiformen Enzephalopathien (TSE) sind eine Gruppe von übertragbaren, tödlich verlaufenden Erkrankungen des Gehirns, die bislang nicht heilbar sind. Zu diesen Krankheiten zählen die Creutzfeldt-Jakob Krankheit beim Menschen, Scrapie beim Schaf und die Bovine Spongiforme Enzephalopathie (BSE) beim Rind. Auf molekularer Ebene sind Prionkrankheiten gekennzeichnet durch eine Umfaltung des zellulären Prion-Proteins PrP^C in die pathologische Scrapie-Isoform PrP^{Sc} und durch die Bildung von zelltoxischen PrP^{Sc}-Aggregaten, welche die Überträger dieser Krankheiten sind und *Prionen* genannt werden.

Im Jahre 2002 wurde vom Bundesministerium für Bildung und Forschung ein interdisziplinärer Forschungsverbund eingerichtet, der sich zur Aufgabe gestellt hatte, Wirkstoffe zur kausalen Therapie von Prionkrankheiten zu entwickeln, und an dem ich zusammen mit Prof. Paul Tavan als Verbundpartner mitgearbeitet habe. In den Laboren unserer experimentellen Partner aus den Bereichen Humanmedizin, Tiermedizin und Biochemie wurden eine Reihe von Testsystemen erarbeitet und eingesetzt, mit denen umfangreiche Sammlungen von medikamentartigen Substanzen, sogenannte Substanzbibliotheken, systematisch nach neuen Wirkstoffen durchsucht wurden. Meine Aufgabe im Verbund bestand in der Sicherung, Qualitätskontrolle und Analyse der experimentell gewonnenen Daten mit dem Ziel Wirkstoffsubstanzen zu bestimmen. Dazu habe ich eine zentrale Datenbank eingerichtet und diverse Softwarewerkzeuge entwickelt, welche eine klar strukturierte Verbindung von gewonnenen Messdaten mit Informationen über die getesteten Substanzen ermöglichen und so die Voraussetzung zur Identifikation von Wirkstoffklassen schufen.

Eine der experimentellen Methoden, das *Scanning for Intensely Fluorescent Targets* (SIFT), wurde zur Messung der Koaggregation von PrP^C an PrP^{Sc} in einem molekularen Assay entwickelt und zum Hochdurchsatz-Screening (HTS) der Bibliotheken sowie zur Validierung von dabei gefundenen primären Treffern anhand von Verdünnungsreihenmessungen eingesetzt. Zur automatisierten Auswertung und Aufbereitung der rohen SIFT-Messdaten habe ich das Softwarewerkzeug SIFT-Viewer entwickelt, welches die Messdaten von Ausreißern befreit, den gemessenen Substanzen anhand der validen Messdaten anti-aggregatorische Wirksamkeitswerte zuordnet sowie Messdaten, Analyseresultate und Molekülstrukturen der Testsubstanzen in grafischer Form zusammenstellt.

Die durch das SIFT-Screening, die automatisierte Datenauswertung und die anschließende Validierung in einem Zellkultur-Assay realisierte systematische Wirkstoffsuche führte zur Entdeckung einer neuen Klasse von anti-Prion-Wirkstoffen, den N⁷-Benzyliden-Benzohydraziden (NBB). Solche NBB-Substanzen zeigten auch in Tierversuchen Wirksamkeit gegen Prionkrankheiten. Darüber hinaus wurde in einem SIFT-Assay Wirksamkeit gegen

die mit den TSE verwandte Parkinsonsche Krankheit gefunden. Sowohl die Screening-Methode als auch die NBB-Wirkstoffklasse wurden durch ein Patent geschützt [1], veröffentlicht [2] und auf Fachkonferenzen präsentiert [3, 4].

Substanzen der NBB-Klasse, wurden anschließend im Rahmen einer externen Kooperation zusätzlich als Inhibitoren der mit der Huntingtonschen Krankheit einhergehenden Aggregation von poly-Glutamin (Q)-Proteinen gefunden. Simulationen des Dockings dieser Substanzen an ein Modell für polyQ-Protofibrillen gaben Hinweise auf mögliche Mechanismen der generellen anti-aggregatorischen Wirksamkeit dieser Substanzen [5].

Neben der SIFT-Technik wurde im Verbund auch ein Zellkulturmodell für Prionkrankheiten zum HTS der Bibliotheken nach anti-Prion-Wirkstoffen eingesetzt. Die Ergebnisse dieses zweiten Screenings habe ich in Verbindung mit jenen des SIFT-Screenings dazu genutzt, die verwendeten Substanzbibliotheken im Hinblick auf Struktur-Wirkungs-Beziehungen einer Clusteranalyse zu unterziehen. Dabei wurde eine weitere neue Wirkstoffklasse entdeckt, die derzeit patentiert wird.

In einem kleinen Nebenprojekt konnte ich durch Molekulardynamik-Simulationen von bestimmten Mutanten des Prion-Proteins einen Beitrag zur Erforschung der Grundlagen der Prionkrankheiten leisten [6]. Die Simulationen ermöglichten ein Verständnis von Zellkultur-Ergebnissen über die Maturierung des Prion-Proteins und gaben Hinweise auf die Funktion des mutierten Residuums für die korrekte Faltung von PrP.

- [1] U. Bertsch, A. Giese, H. Kretzschmar, P. Tavan, T. Hirschberger, J. Bieschke, P. Weber, K.F. Winklhofer, J. Tatzelt, F.U. Hartl, G. Wunsch, T.J. Högen, PCT/EP2005/005614, WO 2005/116640 A2.
- [2] U. Bertsch, K.F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber, F.U. Hartl, P. Tavan, J. Tatzelt, H.A. Kretzschmar, A. Giese, *J. Virol.* 79, 7785-7791 (2005).
- [3] H.A. Kretzschmar, U. Bertsch, K.F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber, F.U. Hartl, P. Tavan, J. Tatzelt, A. Giese, *J. Neuropath. Exp. Neur.* 64, 441-441 (2005).
- [4] U. Bertsch, K.F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber, F.U. Hartl, P. Tavan, J. Tatzelt, H.A. Kretzschmar, A. Giese, *FEBS J.* 272, 338-339 (2005).
- [5] N.W. Schiffer, S.A. Broadley, T. Hirschberger, P. Tavan, H.A. Kretzschmar, A. Giese, C. Haass, F.U. Hartl, B. Schmid, *J. Biol. Chem.* 282, 9195-9203 (2007).
- [6] T. Hirschberger, M. Stork, B. Schropp, K.F. Winklhofer, J. Tatzelt, P. Tavan, *Biophys. J.* 90, 3908-3918 (2006).

Inhaltsverzeichnis

Abbildungsverzeichnis	xi
Abkürzungsverzeichnis	xiii
1 Einleitung	1
1.1 Prionkrankheiten	1
1.2 Die Prionhypothese	3
1.3 Mechanismen der Prion-Vermehrung	5
1.4 Versuche zur Validierung der Prionhypothese	7
1.5 Die Strukturen des Prion-Proteins	9
1.6 Der Lebenszyklus und die Funktion von PrP	12
1.7 Bekannte anti-Prion-Substanzen	15
1.8 Neue Strategien	16
2 Der Forschungsverbund „TSE-Therapie“	17
2.1 Die DIVERSet-Bibliotheken	19
2.2 Der anti-Prion 2D-SIFT-Assay (TP1)	21
2.2.1 Der <i>in vitro</i> Aggregations-Assay	21
2.2.2 Die 2D-SIFT-Messtechnik	22
2.2.3 2D-SIFT-Daten für den Prion-Aggregations-Assay	24
2.2.4 Verschiedene Einsatzformen des SIFT-Assays	29
2.3 Ein Zellkulturmodell für Prionkrankheiten (TP4)	30
2.4 HTS-fähige Zellkulturen und ein weiteres Modellsystem (TP3)	33
2.5 Therapeutische Behandlung von Scrapie-infizierten Mäusen (TP1)	35
2.6 Medizinal- und kombinatorische Chemie (TP6)	36
3 Methoden zur Haltung, Auswertung und Visualisierung von Daten	37
3.1 Erkennung von Ausreißern	38
3.1.1 Eine Median-basierte Erkennung	39
3.1.2 Eine nach dem Töpfchentyp getrennte Erkennung	40
3.2 SIFT-Wirksamkeitsmaße	41
3.2.1 Observablen aus den 2D-SIFT-Spektren	42
3.2.2 Ein auf die Kontrollen bezogenes Wirksamkeitsmaß	42
3.2.3 Ein auf die Testsubstanzen bezogenes Wirksamkeitsmaß	43
3.3 Elektronische Daten zu den DIVERSet-Bibliotheken	44
3.3.1 Molekülstrukturen	44
3.3.2 Eindeutige Namen für die DIVERSet-Substanzen	46

3.3.3	Physikochemische Eigenschaften	46
3.4	Das Visualisierungspaket SIFT-Viewer I	47
3.4.1	SIFT-Dateien und die Extraktion von Messdaten	47
3.4.2	Grafische Aufbereitung von SIFT-Daten	50
3.4.3	Darstellung von Molekülstrukturen	51
3.4.4	Die SIFT-Viewer I zugrunde liegende Verzeichnishierarchie	51
3.4.5	Webbrowser-basierte Visualisierung mit SIFT-Viewer I	53
3.5	Die Screening-Datenbank TSE-DB	55
3.5.1	Modellierung und Installation der TSE-DB	56
3.5.2	Datenformate und Befüllung der TSE-DB	59
3.5.3	Zugriff mit TSE-DB-Browser	60
3.6	Das Analysewerkzeug SIFT-Viewer II	62
3.6.1	Die Fähigkeiten von SIFT-Viewer II	62
3.6.2	Einblicke in die Bedienung von SIFT-Viewer II	69
3.7	Alternative Datenverwaltung mit ChemFinder	71
4	Methoden zur Untersuchung von Struktur-Wirkungs-Beziehungen	75
4.1	Identifikation und Validierung von Leitstrukturen	77
4.1.1	Die Bildung von Strukturklassen mit MoSS	77
4.1.2	Berücksichtigung von Wirksamkeiten in MoSS	78
4.1.3	Ein verbesserter Wirksamkeitsfilter	79
4.1.4	Aufbereitung und Visualisierung von Strukturklassen	80
4.1.5	Bildung von Strukturhierarchien	82
4.1.6	Visualisierungen von Strukturhierarchien	84
4.2	Strukturelle Clusteranalysen zur Suche nach neuen Leitstrukturen	88
4.2.1	Repräsentation von Substanzstrukturen	88
4.2.2	Strukturelles Clustering	90
4.3	Simulationen des Dockings von Wirkstoffen an polyQ-Protofibrillen	92
5	Ergebnisse und Diskussion	95
5.1	Entdeckung und Entwicklung einer ersten neuen Wirkstoffklasse	95
5.1.1	Das mehrstufige Screening der DIVERSet 1-Bibliothek	96
5.1.2	Struktur-Wirkungs-Beziehungen von NBB-Substanzen	101
5.1.3	Fokussierte NBB-Bibliotheken	103
5.1.4	Therapeutische Behandlung von Scrapie-infizierten Mäusen	110
5.1.5	Studien zur Pharmakokinetik	112
5.1.6	Unterdrückung der Aggregation von α -Synuklein	114
5.1.7	Inhibierung der Aggregation von poly-Glutamin-Proteinen	115
5.2	Eine weitere neue Wirkstoffklasse	119
5.2.1	Das Zellkultur-Screening der DIVERSet-Bibliotheken	119
5.2.2	Die Identifikation einer neuen Leitstruktur	120
5.2.3	Weiterentwicklung der neu identifizierten Substanzklasse	123

6	Ausblick	125
A	Publikationen und Patente	127
A.1	Publikationen	127
A.2	Patente	128
B	Die Java-Bibliothek <code>tsedb</code>	129
C	Eine Methode zur vorurteilsfreien Klassifikation von SIFT-Daten	133
C.1	Das Klassifikationsverfahren	133
C.1.1	Vorverarbeitung	134
C.1.2	Bildung von Klassen	136
C.1.3	Konstruktion eines neuen Wirksamkeitsmaßes	138
C.2	Eine exemplarische Anwendung	138
D	Molekulardynamik-Simulationen von PrP	141
	Literaturverzeichnis	143
	Danksagung	157

Abbildungsverzeichnis

1.1	Histopathologie der CJD und Struktur von Prion-Ablagerungen	2
1.2	Isoformen von PrP im Western-Blot	4
1.3	Qualitative Diskussion der Energetik der Prion-Aggregation	6
1.4	Modell zur Vermehrung von infektiösen Prionen	7
1.5	Strukturelle Merkmale von PrP ^C	9
1.6	NMR-Struktur von PrP ^C	10
1.7	Dreieckiges β -helikales Modell für PrP ^{Sc}	12
1.8	Fibrillen aus β -helikalen PrP ^{Sc} -Molekülen	13
1.9	Lebenszyklus von PrP ^C und Entstehung von PrP ^{Sc}	14
2.1	Teilprojekte des Forschungsverbundes „TSE-Therapie“	18
2.2	Eine 96-Well-Mikrotiterplatte	20
2.3	SIFT-Assay-Mischung	22
2.4	SIFT-Messaufbau	23
2.5	Assay-Mischungen und 2D-FID Verteilungen	25
2.6	2D-SIFT-Spektren für eine DIVERSet-Platte	26
2.7	Plattenvariabilität von 2D-SIFT-Spektren	28
2.8	Der anti-PrP ^{Sc} -Wirkstoff DOSPA	31
2.9	Inhibierung der PrP ^{Sc} -Bildung in ScN2a-Zellen	32
2.10	Ergebnisse des Zellkultur-Assays für eine DIVERSet-Platte	34
3.1	Auszug aus der sdf-Datei zu DIVERSet 1	45
3.2	Zuordnung von SIFT-Plattenmessungen zu DIVERSet-Bibliotheksplatten	48
3.3	Verzeichnishierarchie zu SIFT-Viewer I	52
3.4	Bildschirmschnappschuss von SIFT-Viewer I	54
3.5	ER-Diagramm der TSE-DB	57
3.6	Zugriff auf die TSE-DB mit dem TSE-DB-Browser	61
3.7	Exemplarische Layout-Datei für eine Messplatte	64
3.8	Töpfchenweise Repräsentationen von SIFT-Messplatten	66
3.9	Dosis-Wirkungs- und Dosis-Gesamtintensitäts-Kurven	67
3.10	Von SIFT-Viewer II dargestellte Molekülstruktur	68
3.11	Bildschirmschnappschuss von SIFT-Viewer II	70
3.12	Bildschirmschnappschuss der Software ChemFinder	72
4.1	Systematische Erweiterung von Strukturmotiven	78
4.2	Schnappschuss der Visualisierung eines xminer-getriebenen MoSS-Laufs	81
4.3	Bildung einer Hierarchie von Substrukturklassen	83

4.4	Schnappschuss der Visualisierung mit hMoSS	85
4.5	Box-and-Whisker-Plot für eine Wirksamkeitsverteilung	86
4.6	Bildschirmschnappschuss von hMoSS II	87
5.1	Verteilung der primären SIFT-Wirksamkeiten für DIVERSet 1	97
5.2	Dosis-Wirkungs-Analyse im Zellkulturmodell	99
5.3	Statistik des Screenings von DIVERSet 1	99
5.4	Substanzen mit Zellkulturwirksamkeit	100
5.5	NBB-Struktur	101
5.6	Struktur-Wirkungs-Beziehungen für Klassen von NBB-Derivaten	102
5.7	Substanz 293G02	103
5.8	Synthese von NBB-Substanzen	104
5.9	Belegungsschema der fünf Platten I bis V mit neu synthetisierten NBB-Substanzen	105
5.10	Befüllung einer 384-Töpfchen-Messplatte	105
5.11	SIFT-Ergebnisse für eine 384-Töpfchen-Plattenmessung einer 96-Töpfchen-Platte mit neuen NBB-Substanzen	106
5.12	SAR-Analyse für die NBB-Derivate von Platte II	108
5.13	Mit Wirksamkeit assoziierte Hydrazide und Aldehyde	108
5.14	SIFT-Dosis-Wirkungs-Analysen für neue NBB-Derivate	109
5.15	Mediane der Sterbealter	111
5.16	PrP ^{Sc} -Ablagerungen in der Milz	112
5.17	Konzentrationen der Substanzen 309F02 und 293G02 im Blutplasma von Mäusen	113
5.18	Unterdrückung der α S-Aggregation durch die Substanz 293G02	114
5.19	Strukturen der Substanzen 306H03 und Congo red	116
5.20	Docking von anti-aggregatorischen Substanzen an ein Modell für eine polyQ-Protofibrille	118
5.21	SAR-Karte für die DIVERSet-Bibliotheken	122
B.1	Paketstruktur der Java-Bibliothek <code>t_sedb</code>	130
C.1	Layout der für die Klassifikation verwendeten Plattenmessung	139
C.2	Klassen und Merkmalsvektoren	140
C.3	Wirksamkeitsmaße im Vergleich	140
D.1	Umgebungsabhängige Faltung von Helix 1 in PrP-M205R	141

Abkürzungsverzeichnis

α S	α -Synuklein. Mit der Parkinsonschen Krankheit assoziiertes Protein, Seite 30
ADMET	Absorption, Distribution, Metabolismus, Exkretion und Toxikologie, Seite 19
ADT	AutoDockTools. Werkzeuge für Docking-Simulationen, Seite 93
BMBF	Bundesministerium für Bildung und Forschung, Seite 1
BMO	Lehrstuhl für BioMolekulare Optik, Seite 129
BSE	Bovine Spongiforme Enzephalopathie, Seite 1
CD	Circular Dichroism. Zirkulärer Dichroismus, Seite 10
CDK	Chemistry Development Kit. Eine Java-Bibliothek, Seite 132
CJD	Creutzfeldt-Jakob-Krankheit, Seite 1
cLogP	Calculated LogP. Berechnete Schätzung für LogP, Seite 46
CVS	Concurrent Versions System. Werkzeug zur Quellcode-Verwaltung, Seite 47
DDL	Data Definition Language, Seite 58
DMSO	Dimethylsulfoxid. Ein organisches Lösungsmittel., Seite 20
DOSPA	2,3-Dioleoyloxy- <i>N</i> -[2(sperminecarboxamido)ethyl]- <i>N,N</i> -dimethyl-1-propanaminium trifluoroacetate. Ein kationisches Lipid, das im SIFT-Screening als Positiv-Kontrollsubstanz eingesetzt wurde, Seite 15
EM	Elektronenmikroskopie, Seite 11
ER	Endoplasmatisches Retikulum. Organell, in welchem Proteine posttranslational modifiziert werden., Seite 12
ER	Entity-Relationship. ER-Modelle sind Schemata zur Beschreibung von Objekten und deren Beziehungen, die häufig zum Entwurf von Datenbanken eingesetzt werden, Seite 56
FCS	Fluorescence Correlation Spectroscopy, Seite 21
FK	Foreign Key. Fremdschlüssel. Attribut einer Datenbanktabelle, Seite 56

GFP	Das grün fluoreszierende Protein, Seite 116
GPI	Glycosylphosphatidylinositol. Molekül, über welches Proteine in Membranen verankert werden, Seite 10
GUI	Graphical User Interface. Grafische Benutzerschnittstelle, Seite 131
hMoSS	Hierarchische Erweiterung des Werkzeugs MoSS, Seite 84
HTS	High Throughput Screening. Hochdurchsatz-Screening, Seite 16
IDL	Interactive Data Language. Software-Paket für Datenanalysen, Seite 50
JDBC	Java Database Connectivity. Eine Datenbank-Schnittstelle, Seite 131
LogP	Maß für die Lipophilie einer Substanz, Seite 46
LRZ	Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, Seite 55
MD	Molekulardynamik, Seite 5
MoSS	Molecular Substructure Miner. Werkzeug für systematische Suchen nach chemischen Substrukturen, Seite 77
MPI	Max-Planck-Institut, Seite 30
mRNA	Messenger Ribonucleic Acid. Boten-Ribonukleinsäure, Seite 116
N2a	Eine Zelllinie, Seite 31
NBB	N'-Benzyliden-Benzohydrazide. Im Verbund entdeckte und entwickelte Klasse von Wirkstoffen gegen Proteinaggregationskrankheiten, Seite 18
NMR	Nuclear Magnetic Resonance. Kernspinresonanz, Seite 9
PK	Pharmakokinetik, Seite 112
PK	Primary Key. Primärschlüssel. Attribut einer Datenbanktabelle, Seite 56
PK	Proteinase K. Ein Enzym, das Peptidbindungen in Proteinen spaltet, Seite 3
PMCA	Protein Misfolding Cyclic Amplification. Ein <i>in vitro</i> Modellsystem zur Vermehrung von Prionen, Seite 8
Prn-p	Das für das Prion-Protein kodierende Gen, Seite 3
PrP	Das Prion-Protein, Seite 3
PrP ^C	Zelluläre Isoform des Prion-Proteins, Seite 3

PrP ^{res}	Protease-resistente Isoform des Prion-Proteins, Seite 8
PrP ^{Sc}	Krankheitsassoziierte Scrapie-Isoform des Prion-Proteins, Seite 3
QA	Quinacrine. Malaria-Medikament mit anti-Prion Wirksamkeit, Seite 15
rPrP	Rekombinant hergestelltes Prion-Protein, Seite 21
SAR	Structure-Activity Relationships. Struktur-Wirkungs-Beziehungen, Seite 75
ScN2a	Mit Scrapie infizierte N2a-Zellen, Seite 31
SIFT	Scanning for Intensely Fluorescent Targets, Seite 16
SMB	Scrapie Mouse Brain. Eine Zelllinie, Seite 33
SQL	Structured Query Language, Seite 59
TFE	Trifluoroethanol. Ein organisches Lösungsmittel, Seite 30
TP	Teilprojekte des Forschungsverbundes "TSE-Therapie", Seite 17
tPSA	Total Polar Surface Area. Maß für die Polarität einer Substanz, Seite 46
TSE	Transmissible Spongiforme Enzephalopathien, Seite 1
vCJD	Neue Variante der CJD, Seite 1
wtPrP	Wildtypisches Prion-Protein, Seite 32
ZNP	Zentrum für Neuropathologie und Prionforschung der LMU, Seite 21
2D-FID	Two-Dimensional Fluorescence Intensity Distribution. Zweidimensionale Fluoreszenz-Intensitäts-Verteilung, Seite 24

1 Einleitung

In den frühen 1990er Jahren erkrankten im Vereinigten Königreich jährlich rund 35.000 Rinder an der Bovinen Spongiformen Enzephalopathie (BSE). Wenige Jahre später wurde die Krankheit vCJD, eine neue Variante der Creutzfeldt-Jakob-Krankheit (CJD), entdeckt [1]. Experimentelle Hinweise [2], dass vCJD wahrscheinlich durch neuartige Krankheitserreger von BSE-infizierten Rindern über die Nahrungskette auf den Menschen übertragen worden war, lösten Befürchtungen vor einer vCJD-Epidemie aus. Bis November 2006 starben im Vereinigten Königreich zwar 158 Menschen an vCJD [3], aber die befürchtete Epidemie blieb glücklicherweise aus.

Als Ende des Jahres 2000 die ersten deutschen BSE-Fälle bekannt wurden, fürchtete man auch hierzulande eine Übertragung dieser Krankheit auf den Menschen. Das Bundesministerium für Bildung und Forschung (BMBF) reagierte darauf mit der Errichtung eines Förderschwerpunktes zur Erforschung von Therapiemöglichkeiten für diese Erkrankungen [4]. In diesem Rahmen begann wenig später der Forschungsverbund „Systematische Entwicklung neuer Wirkstoffe für die kausale Therapie von Prionkrankheiten“ seine Arbeit, an dem ich seit Juli 2002 mitgearbeitet habe. Gegenstand der vorliegenden Dissertation ist mein seitdem geleisteter wissenschaftlicher Beitrag zum Forschungsverbund.

In den Teilprojekten dieses interdisziplinären Verbundes wurden eine Reihe von Modellsystemen für Prionkrankheiten zur systematischen Suche nach Wirkstoffen in großen Mengen von Testsubstanzen eingesetzt. Mein Beitrag zum Verbund umfasste die Verwaltung, Aufbereitung und Analyse der dabei gewonnenen Daten mit dem Ziel Wirkstoffklassen zu bestimmen. Zum Verständnis der unserem Forschungsverbund gestellten Aufgabe möchte ich zunächst die grundlegenden Mechanismen von Prionkrankheiten erläutern, soweit diese gegenwärtig bekannt sind.

1.1 Prionkrankheiten

Die Transmissiblen Spongiformen Enzephalopathien (TSE) oder Prionkrankheiten [5] sind eine Gruppe von tödlichen neurodegenerativen Erkrankungen, die bedauerlicherweise noch immer unheilbar sind. Zu den verbreitetsten Prionkrankheiten zählen die BSE beim Rind, die Traberkrankheit (Scrapie) beim Schaf und die CJD beim Menschen. Seltenere auftretende menschliche Prionkrankheiten sind das Gerstmann-Sträussler-Scheinker-Syndrom, die tödliche familiäre Insomnie und die durch Kannibalismus übertragene und inzwischen nahezu ausgerottete Krankheit Kuru. TSE können sporadisch auftreten, über

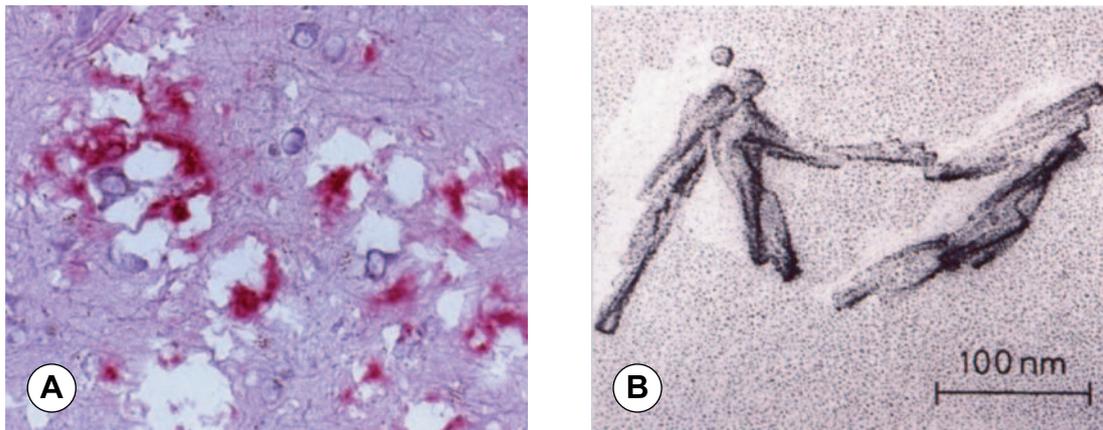


Abbildung 1.1: Histopathologie der CJD und Struktur von Prion-Ablagerungen. **A** Mikroskopische Aufnahme eines eingefärbten Schnittes durch das Gehirn eines CJD-Patienten. Als Merkmale der CJD zeigen sich weißlich erscheinende Vakuolen und violett eingefärbte Ablagerungen, die Prionen genannt werden. Als Größenmaßstab können die als bläuliche Kringel erscheinenden Nervenzellen dienen, die einen Durchmesser von etwa $50 \mu\text{m}$ besitzen. **B** Das elektronenmikroskopische Bild von aufgereinigten Prion-Ablagerungen lässt deren fibrilläre Strukturen erkennen. Mit freundlicher Genehmigung von A. Giese (**A**) bzw. M. Özel und H. Diringer (**B**).

Gene ererbt oder durch Erreger übertragen werden. Die häufigste menschliche Prionkrankheit, die CJD, tritt in allen drei Varianten auf [5].

Patienten der CJD erleiden nach langer, symptomloser Inkubationszeit ab dem ersten Auftreten von Symptomen einen rasch fortschreitenden Abbau ihrer geistigen Fähigkeiten, der – abhängig von der Variante – innerhalb von wenigen Monaten oder höchstens etwa zwei Jahren zum Tode führt. Dem geistigen Abbau liegt ein massives Absterben von Zellen des zentralen Nervensystems zugrunde, dessen Folgen in Abbildung 1.1A zu erkennen sind. Die Abbildung zeigt eine mikroskopische Aufnahme eines eingefärbten Schnittes durch das Gehirn eines CJD-Patienten und macht die histopathologischen Merkmale der CJD deutlich. Dazu zählen die weißlich erscheinenden Vakuolen, welche eine schwammartige Struktur des Gehirns hervorrufen, sowie violett eingefärbte sogenannte Amyloidartige Ablagerungen.

Amyloide entstehen durch Aggregation von abnorm veränderten Proteinen, finden sich zumeist außerhalb der Zellen, sind unlöslich und liegen in Form von kleinen Fasern vor. Abbildung 1.1B zeigt für den Fall der Prionen eine elektronenmikroskopische Aufnahme solcher amyloiden Ablagerungen, die aufgereinigt wurden und dann fibrilläre Strukturen aufweisen. Derartige Amyloide werden in ähnlicher Form auch bei anderen neurodegenerativen Krankheiten wie der Alzheimerschen, der Huntingtonschen oder der Parkinsonschen Krankheit beobachtet.

Die Ähnlichkeiten von lichtmikroskopischen pathologischen Merkmalen in Gehirnschnitten, die bei der übertragbaren Schafkrankheit Scrapie und den menschlichen Krankheiten

Kuru und CJD beobachtet wurden, veranlassten Neuropathologen bereits Ende der 1950er Jahre zu den Vermutungen, dass diese Krankheiten verwandt sind und dass Kuru und CJD, ähnlich wie Scrapie, übertragbar sind [6, 7]. Etwa zehn Jahre später wurde die Übertragbarkeit der menschlichen TSE-Krankheiten Kuru und CJD auf Schimpansen experimentell bestätigt [8, 9].

1.2 Die Prionhypothese

Obwohl man lange vorherrschend davon ausging, dass die TSE-Krankheiten durch „langsame Viren“ übertragen werden, wurde bereits Ende der 1960er Jahre aufgrund des geringen Molekulargewichts der aus erkrankten Gehirnen extrahierten Scrapie-Erreger vermutet, dass diese aus Proteinen bestehen könnten [10, 11]. Erst Anfang der 1980er Jahre gelang es dem amerikanischen Mediziner Stanley Prusiner und seinen Mitarbeitern das Erregermaterial besser aufzureinigen und damit die „protein-only-Hypothese“ experimentell zu untermauern [12–14]. Amyloide Ablagerungen, wie die in Abbildung 1.1 für CJD gezeigt, sind danach die Überträger der TSE-Krankheiten. Prusiner nannte diese **proteinartigen, infektiösen Partikel**, die kein Erbgut in Form von Nukleinsäuren enthalten, „Prionen“ [13]. Für die Entdeckung dieses neuartigen Prinzips der Infektion wurde er 1997 mit dem Nobelpreis für Medizin geehrt.

Prionen bestehen hauptsächlich aus fehlgefalteten Exemplaren eines wirtseigenen Proteins, das Prion-Protein (PrP) genannt wurde [13]. Die Prionhypothese besagt, dass das zentrale Ereignis in der Pathogenese der Prionkrankheiten die Umwandlung von zellulärem Prion-Protein (PrP^C) in die krankheitsassoziierte Scrapie-Isoform PrP^{Sc} ist [13]. Außerdem vermutet man, dass PrP^{Sc} zum einen als Schablone für diese Umwandlung wirkt und zum anderen als toxisches Agens, das die Fehlfunktion und den Tod von Nervenzellen verursacht [5, 15].

Abbildung 1.2 zeigt sogenannte Western-Blots von Hirn-Homogenaten gesunder und mit Prionen infizierter Hamster, in denen Prion-Proteine über Antikörper sichtbar gemacht wurden. Wie anhand der Blots zu erkennen ist, besitzt PrP^{Sc} einen Abschnitt, der resistent gegenüber Verdauung mit der Proteinase K (PK) ist (schwarze Bande in Reihe 4). Die PK ist ein in der Biochemie häufig eingesetztes proteolytisches Enzym, welches Proteine oder Peptide spalten kann, indem es die Hydrolyse von Peptidbindungen katalysiert. Die Anwesenheit des PK-resistenten PrP^{Sc}-Fragmentes, das zwischen 27 und 30 kDa (1 Da = 1 u) schwer ist und deshalb PrP 27-30 genannt wird, ist ein molekulares Kennzeichen der Prionkrankheiten [12, 14]. Bemerkenswerterweise behält PrP 27-30 Infektiosität.

Die protein-only-Hypothese wird eindrucksvoll von Experimenten mit sogenannten PrP-knockout-Mäusen untermauert, deren für das Prion-Protein kodierendes Gen Prn-p durch eingebrachte Veränderungen ausgeschaltet wurde [16–18]. Prn-p0/0 Mäuse, deren Prn-p homozygot in beiden Allelen ausgeschaltet ist, sind immun gegen eine Infektion mit Scrapie-Prionen. Ferner zeigen Prn-p0/+ Mäuse, deren Prn-p heterozygot nur in einem der Al-

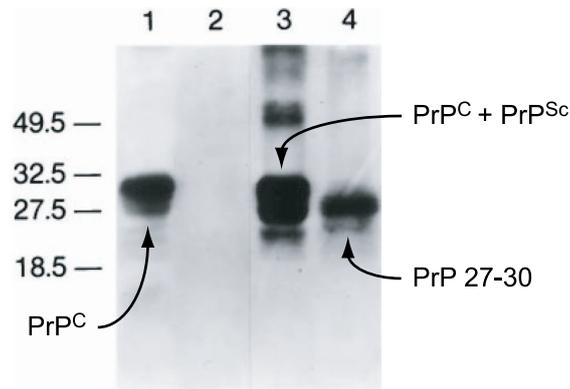


Abbildung 1.2: Isoformen von PrP im Western-Blot. Western- oder Immuno-Blots von Hirn-Homogenaten von Syrischen Hamstern, die nicht (Reihen 1 und 2) oder mit Prionen (Reihen 3 und 4) infiziert worden waren. Nach einer Polyacrylamid Gel-Elektrophorese wurde der Blot mit einem bestimmten anti-PrP Antiserum behandelt, um die enthaltenen Prion-Proteine einzufärben. Zuvor jedoch wurden die Proben der Reihen 2 und 4 mit Proteinase K verdaut. Dabei wurde das PrP^C (Bande bei 27,5 - 32,5 kDa in Reihe 1) in diesen Reihen vollständig hydrolysiert, während von PrP^{Sc} lediglich ungefähr 67 Aminosäuren vom N-Terminus abgespalten wurden, sodass das ca. 27 bis 30 kDa große Fragment PrP 27-30 übrig blieb (Reihe 4). Übernommen aus [5] und zusätzlich beschriftet.

lele ausgeschaltet ist, einen verzögerten Krankheitsverlauf. Maßstab ist dabei der Krankheitsverlauf bei unveränderten Prn-p+/+ Mäusen, die PrP in vollem Umfang produzieren. Diese Ergebnisse zeigen klar, dass PrP für die Entwicklung der Krankheit notwendig ist, und dass der Verlauf der Krankheit von der Menge an verfügbarem PrP abhängt. Interessanterweise entwickelten und verhielten sich die untersuchten PrP-knockout-Mäuse äußerlich völlig normal, d.h. sie zeigten keine Auffälligkeiten gegenüber unveränderten Prn-p+/+ Mäusen.

Ob Prionen, wie nach Prusiners Definition, tatsächlich keine Erbsubstanz in Form von Nucleinsäuren enthalten, wurde in der Gruppe von Detlev Riesner mit sehr empfindlichen physikalisch-chemischen Messungen untersucht [19, 20]. Anhand der Messungen konnte ausgeschlossen werden, dass infektiöse Einheiten von PrP^{Sc} mehr als ein Nucleinsäure-Molekül von maximal 100 oder sogar nur 25 Nucleotiden enthalten. Die in dem Erregermaterial enthaltenen Nucleinsäure-Moleküle sind in jedem Fall zu klein, als dass darin das Prion-Protein verschlüsselt sein könnte. Poly-Nucleotide, die größer als 25 Basen sind, kommen sehr selten vor und können deshalb nicht notwendig für die Infektiosität von Prionen sein [20].

Gegenüber den herkömmlichen Krankheiten, deren Erreger Bakterien, Pilze oder Viren sind, die sich über Nucleinsäuren identisch replizieren, zeichnen sich die Prionkrankheiten also dadurch aus, dass die Erreger aus fehlgefalteten Proteinen zusammengesetzt sind, die sich selbständig vermehren. Es erhebt sich nun die Frage, wie Prionen sich selbst replizieren können.

1.3 Mechanismen der Prion-Vermehrung

Mehrere Mechanismen der Selbstreplikation wurden vorgeschlagen. Cohen et al. [21] schlugen vor, dass sich für die Umwandlung ein einzelnes PrP^{Sc}-Molekül an ein PrP^C anlagert und sich die beiden Moleküle zu einem Hetero-Dimer vereinigen, in dem die strukturelle Umwandlung von PrP^C in PrP^{Sc} autokatalytisch abläuft. Dagegen postulierte Peter Lansbury [22, 23] ein Modell eines im Wesentlichen eindimensionalen Kristallwachstums, nach dem eine Kette aus aneinander gereihten PrP^{Sc}-Molekülen wächst, indem sich einzelne PrP^C-Moleküle an den Enden anlagern und in die PrP^{Sc}-Form umgewandelt werden. Bedingung für das Wachstum einer Kette ist die Bildung eines Kristallisationskeimes. Eine solche Protofibrille muss erst eine kritische Länge erreicht haben, bevor sich schneller neue Moleküle anlagern, als alte wieder abgespalten werden. Das skizzierte Kettenmodell fügt sich zwanglos in den experimentellen Befund ein, dass PrP^{Sc}-Partikel aus amyloiden Fasern bestehen.

Manfred Eigen [24] untersuchte die chemische Kinetik der von Cohen et al. und von Lansbury postulierten Mechanismen und fand, dass das Hetero-Dimer-Modell nicht plausibel ist, wenn realistische Werte für die kinetischen Konstanten angenommen werden. Eigens Untersuchungen ergaben, dass nach dem Hetero-Dimer-Modell die Konzentration von PrP^{Sc} entweder stetig abnehmen oder ansteigen müsste, und dass demzufolge die Prionkrankheiten stets von selbst heilen oder viel häufiger spontan auftreten müssten, als dies bei den sporadischen Varianten der Fall ist [25]. Eigen löste dieses Problem, indem er eine kooperative Wechselwirkung zwischen mehr als zwei PrP-Einheiten postulierte, die für die autokatalytische Strukturumwandlung nötig sein soll. Diese Kooperativität war von dem Prinzip der sogenannten allosterischen Enzyme motiviert, die erst durch Anlagerung eines Effektor-Moleküls in eine wirksame Form umgewandelt werden. Ferner konnte Eigen auch die Keimbildungsphase aus Lansburys Kettenmodell als eine Art kooperative Wechselwirkung interpretieren [25]. Damit wird also die spontane Bildung stabiler Krankheitskeime als sogenannte „seeded aggregation“ aufgefasst.

Abbildung 1.3 zeigt ein Schema der Energetik eines solchen Prozesses, welches von Paul Tavan im Rahmen seiner Vorlesung „Theoretische Grundlagen der molekularen Biophysik“ [26] im Wintersemester 05/06 vorgestellt wurde. Er postulierte dabei den gezeigten qualitativen Verlauf der freien Energie als Funktion der Länge der Prion-Aggregate. Danach muss für die Bildung eines stabilen Aggregationskeimes erst eine energetische Hürde überwunden werden, bevor sich weitere Moleküle energetisch günstig anlagern können.

Dieselbe Fragestellung untersuchte meine Kollegin Martina Stork anhand von Molekulardynamik (MD)-Simulationen von poly-Glutamin (Q)-Peptiden [27], die bei der Huntingtonschen Krankheit und bei weiteren neurodegenerativen Krankheiten als Bausteine von pathologischen Aggregaten vorkommen. Für die untersuchten, 36 Residuen umfassenden poly-Q-Peptide, von denen angenommen wird, dass sie in den amyloiden Aggregaten eine spezielle zweifach gewundene β -helikale Struktur aufweisen, konnte gezeigt werden, dass diese helikale Struktur in Monomeren instabil ist und erst durch Dimerisierung

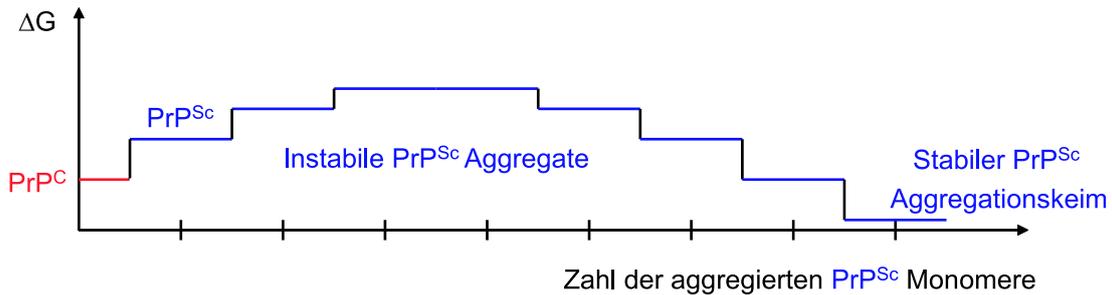


Abbildung 1.3: Qualitative Diskussion der Energetik der Prion-Aggregation. Qualitative Abhängigkeit der freien Energie als Funktion der Länge der Prion-Aggregate. Ausgehend von einem PrP^{C} -Molekül steigt die freie Energie mit der Zahl der gebildeten und aggregierten PrP^{Sc} -Moleküle zunächst bis zu einem Maximalwert an. Überschreitet die Länge der Aggregate einen kritischen Wert, so fällt die freie Energie der Aggregation wieder, sodass sich ein stabiler Aggregationskeim ausbilden kann. Übernommen aus [26].

stabilisiert wird. Damit konnten also MD-Simulationen die mit der Größe der Aggregate zunehmende Stabilität belegen, welche dem Aggregationskeimkonzept zugrunde liegt.

Die Vermehrung der Keime, die für den infektiösen Krankheitsprozess nötig ist und im Falle von Bakterien oder Viren in Form der identischen Replikation auftritt, konnte von dem oben skizzierten Modell von Eigen nicht erklärt werden. Hierzu nahmen Masel et al. [28] den Gedanken des Kettenbruchs in ein verfeinertes mathematisches Modell auf, dessen Parameter aus experimentellen Daten geschätzt wurden. Abbildung 1.4 zeigt schematisch, wie die Verlängerung von Ketten aus PrP^{Sc} -Molekülen und der Bruch von Ketten einer gewissen Länge so zusammenwirken, dass die Zahl der offenen Kettenenden exponentiell zunimmt. Der damit identifizierte physikalisch-chemische Mechanismus erklärt, wie sich Krankheitskeime, die nur aus Proteinen bestehen, vermehren können. Er zeigt ferner, dass im Prinzip schon ein einziges amyloides PrP^{Sc} -Aggregat, welches in die Nähe nativer PrP^{C} -Moleküle gelangt, die Krankheit auslösen kann.

Die Frage nach der infektiösen Einheit, also danach, welche Menge von PrP^{Sc} infektiös ist, versucht man mit Tierversuchen zu beantworten. Dabei werden eine Reihe von unterschiedlich stark verdünnten Prion-Präparationen Gruppen von Versuchstieren verabreicht. Die maximale Verdünnung, bei der noch die Hälfte der Tiere einer Gruppe sterben, wird als infektiöse Dosis ID_{50} bezeichnet. Anhand von solchen Versuchen schätzten verschiedene Experimentatoren die Zahl der PrP^{Sc} -Moleküle pro ID_{50} -Einheit und kamen auf stark unterschiedliche Werte von 10^4 bis 10^5 oder sogar darüber (vgl. die Diskussion von Masel et al. [28]). Masel und Jansen [29] konnten anhand von mathematischen Modellen, ähnlich den oben beschriebenen, erklären, dass die Infektiosität nicht allein durch die Zahl der PrP^{Sc} -Moleküle charakterisiert werden kann, sondern stark von der Größenverteilung der Prion-Aggregate abhängt. Danach sind viele kürzere Aggregate stärker infektiös als wenige lange. Damit können die vorgestellten mathematischen Modelle für die Kinetik der Prion-Vermehrung auch den ungewöhnlichen Verlauf der Prionkrankheiten erklären [25],

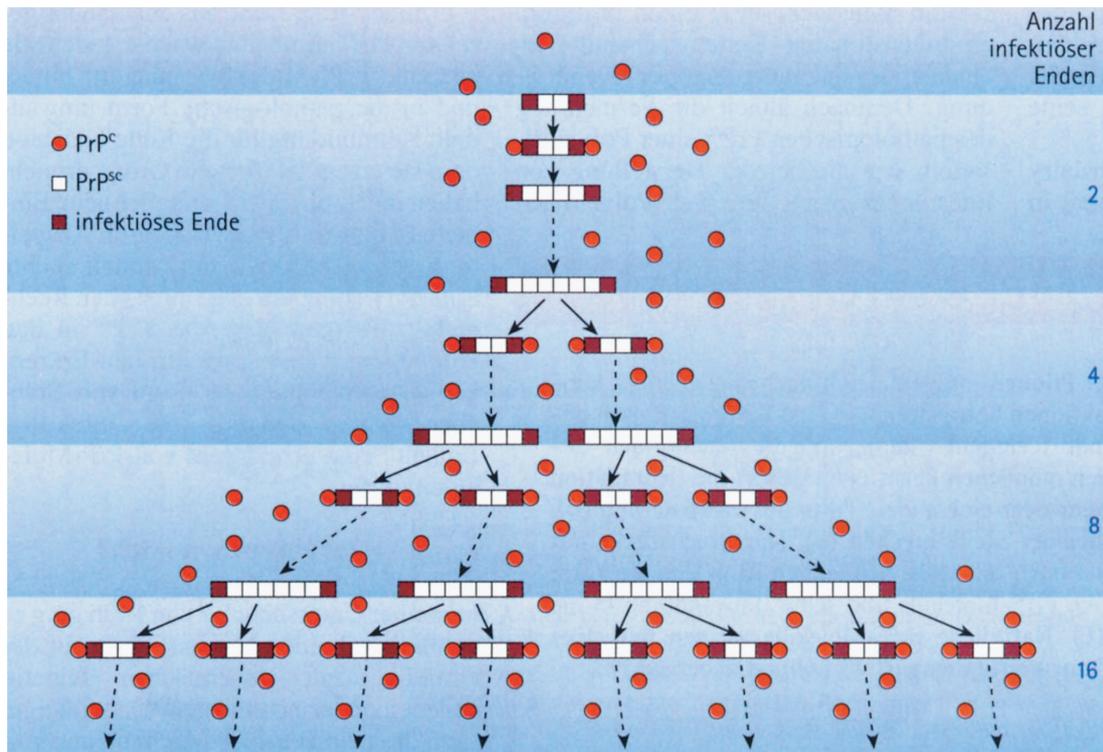


Abbildung 1.4: Modell zur Vermehrung von infektiösen Prionen. Dieses vereinfachte Modell zeigt schematisch, wie innerhalb eines Zyklus Ketten von aneinander gereihten PrP^{Sc}-Molekülen durch Anlagerung von PrP^C-Molekülen an den Kettenenden wachsen. Durch den Bruch von Ketten, deren Länge einen kritischen Wert überschreitet, verdoppelt sich die Zahl der freien Kettenenden von Zyklus zu Zyklus. Übernommen aus [25].

der durch eine lange Inkubationszeit und ein rasches Fortschreiten im klinischen Stadium gekennzeichnet ist.

Die Modelle zur Prion-Vermehrung zeigen exemplarisch, dass die Theorie wichtige Beiträge zur Prion-Forschung leisten kann, indem sie der weitgehend experimentell arbeitenden biologischen und medizinischen Forschungsgemeinschaft Hinweise auf neue Fragestellungen und experimentelle Ansätze gibt. Insbesondere basieren die im folgenden Abschnitt diskutierten Verfahren zur *in vitro* Vermehrung von Prionen auf diesen Modellen.

1.4 Versuche zur Validierung der Prionhypothese

Obwohl die Prionhypothese gemeinhin als richtig angenommen wird, weil sie mittlerweile von sehr vielen experimentellen Ergebnissen untermauert wurde, gibt es nach wie vor Skeptiker, die nicht ausschließen wollen, dass Prionkrankheiten durch Viren übertragen werden könnten [30]. Der direkteste Beweis für die protein-only-Hypothese wäre voll-

bracht, wenn es gelingen würde, infektiöse Prionen *in vitro* zu synthetisieren (eine aktuelle Übersicht zu künstlichen Prionen geben Eiden et al. [31]).

Die strukturelle Umwandlung und Aggregation von PrP^C in fehlgefaltetes, Protease-resistentes PrP^{res} konnte *in vitro* modelliert werden [32], aber die geringe Ausbeute an produziertem PrP^{res} erlaubte keinen direkten Nachweis der Infektiosität. Kürzlich ist es Legname et al. [33] jedoch gelungen, künstliche Prionen aus verkürzten Prion-Proteinen herzustellen und damit Mäuse zu infizieren. Kritiker bemängeln aber mehrere Punkte an der Studie von Legname et al. Zum einen waren die verwendeten Mäuse genetisch so verändert, dass sie die verkürzten Prion-Proteine in stark erhöhter Anzahl bildeten. Damit könnte sich die TSE-Krankheit in einem solchen Milieu auch spontan bilden. Zum anderen erkrankten die mit den künstlichen Prionen infizierten Mäuse erst sehr viel später als Mäuse, denen natürliche Scrapie-Erreger verabreicht worden waren, und starben kaum früher als nicht infizierte Kontrolltiere. Dies deutet darauf hin, dass die Infektiosität der in dieser Studie erzeugten künstlichen Prionen sehr gering ist.

Die zyklische Verstärkung der Fehlfaltung von Proteinen (PMCA, protein misfolding cyclic amplification) ist ein *in vitro* Modellsystem, das es ermöglicht, große Mengen von PrP^C in PrP^{res} umzuwandeln [34, 35]. Nach dem Prinzip der Prion-Vermehrung aus Abbildung 1.4 werden dabei winzige Mengen von authentischem, aus Gehirnen extrahiertem PrP^{Sc} mit großen Mengen von PrP^C, die in Form von Hirnhomogenat zur Verfügung gestellt werden, in einem zyklischen Prozess inkubiert, der aus abwechselnden Schritten der Inkubation und der Ultraschallbehandlung besteht. Man nimmt dabei an, dass während der Inkubation PrP^{Sc}- bzw. PrP^{res}-Aggregate wachsen und dass die anschließende Ultraschallbehandlung die PrP^{res}-Aggregate in kleinere Einheiten zerbricht, welche zusätzliche Keime für weiteres Aggregat-Wachstum bieten. Kürzlich ist es mit der PMCA-Technik gelungen, Aggregate herzustellen, die ebenso infektiös sind wie natürliches PrP^{Sc} [36].

Obwohl diese *in vitro* Studien den Eindruck vermitteln, dass an der Prion-Vermehrung lediglich PrP beteiligt ist, gibt es experimentelle Hinweise aus Studien mit Mäusen, dass im lebenden Organismus bei der Bildung von PrP^{Sc} ein nicht näher bestimmtes Makromolekül teilnehmen könnte, das Protein X genannt wurde [37]. Wie dem auch sei – das Fehlen einer detaillierten Kenntnis der an der *in vivo* Aggregation von Prionen möglicherweise teilnehmenden Kofaktoren ändert nichts an den grundlegenden Resultaten, dass dabei wirtseigene Prion-Proteine infektiöse und toxische Aggregate bilden.

Das geschilderte Wissen über die molekularen Vorgänge der Prion-Vermehrung legt nun einen möglichen Wirkmechanismus für potenzielle anti-Prion-Medikamente nahe. Die zu findenden Wirkstoffe könnten die Prion-Vermehrung stoppen, indem sie die Anlagerung von PrP^C an PrP^{Sc} oder die anlagerungsinduzierte Umwandlung von PrP^C in PrP^{Sc} hemmen oder vollständig blockieren. Das davor bereits gebildete Prion-Material könnte dann vom heilenden Organismus langsam abgebaut werden, sodass auf diese Weise die Krankheit gestoppt und eventuell vollständig geheilt wird. Für die Suche nach derartigen Wirkstoffen, die dazu an PrP^C oder PrP^{Sc} binden müssen, bieten sich *in vitro* Systeme wie das PMCA-Modell an.

Gegenwärtig ist die Suche nach Wirkstoffen, welche an PrP^C und PrP^{Sc} binden und das Wachstum von PrP^{Sc} verhindern können, leider noch ausschließlich auf rein empirische Suchmethoden beschränkt. Ein Grund dafür ist, dass zur molekularen Struktur von PrP^{Sc} nur sehr ungenaue Informationen vorhanden sind. Dennoch möchte ich im Folgenden das vorhandene Wissen zu den Molekülstrukturen von PrP^C und PrP^{Sc} kurz skizzieren.

1.5 Die Strukturen des Prion-Proteins

PrP ist ein Protein, das gut 200 Aminosäuren umfasst und weitgehend homolog in vielen verschiedenen Spezies vorkommt [38]. Es findet sich in großen Mengen an den synaptischen Endknöpfchen von Nervenzellen [39]. Welche Funktion es dort ausübt, ist noch weitgehend unbekannt, wenn auch begründete Vermutungen geäußert wurden, dass es an der Homöostase von Cu(II)-Ionen beteiligt ist [39, 40].

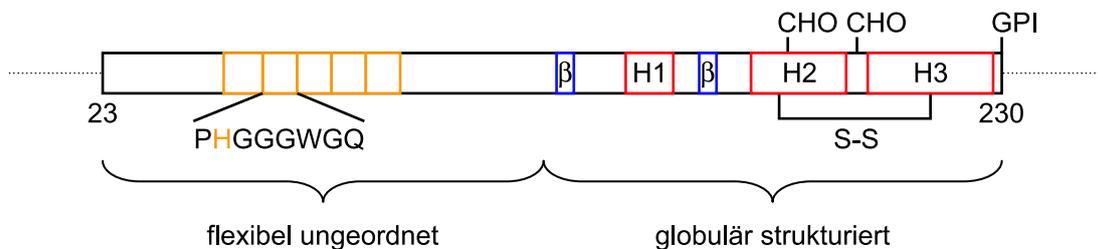


Abbildung 1.5: Strukturelle Merkmale von PrP^C. Das 208 Aminosäuren (23-230) lange menschliche PrP^C besitzt einen flexiblen, ungeordneten N-terminalen Teil (23-124) und einen globulären, strukturierten C-terminalen Teil (125-228). Der flexible N-terminale Abschnitt weist eine Domäne auf, in der sich das Oktapeptid PHGGGWGQ viermal und in leicht modifizierter Form ein fünftes Mal wiederholt (51-91). Diese Region kann über Histidin (H)-Residuen Cu²⁺ Kupfer-Ionen binden [41] und dadurch Struktur annehmen [42]. In der C-terminalen Hälfte von PrP^C befinden sich die drei α -Helizes H1 (144-154), H2 (173-194) und H3 (200-228) sowie zwei kurze β -Stränge (128-131 und 161-164), die zusammen ein anti-paralleles Faltblatt bilden. H2 und H3 sind über eine Disulfidbrücke zwischen den Cystein-Residuen Cys179 und Cys214 kovalent verbunden. Posttranslational werden an die Aminosäurekette zwei komplexe Zuckerseitenketten (CHO) an den Asparagin-Residuen Asn181 und Asn197 und der C-terminalen GPI-Anker an das Serin-Residuum Ser230 angehängt. Die für diese Prozessierung benötigten terminalen Signalsequenzen (1-22 und 231-255; als Punkte dargestellt) werden danach abgespalten und sind im reifen PrP^C nicht mehr vorhanden.

Kurt Wüthrich und seinen Mitarbeitern ist es vor rund zehn Jahren erstmals gelungen, mithilfe von Messungen der kernmagnetischen Resonanz (NMR) die räumliche Struktur von rekombinant hergestelltem PrP^C aufzuklären [43], soweit dieses Protein überhaupt eine geordnete Struktur annimmt. Danach besitzt PrP^C, wie in Abbildung 1.5 schematisch dargestellt ist, einen flexiblen, ungeordneten N-terminalen Abschnitt und einen globulären, strukturierten C-terminalen Teil, die jeweils ungefähr eine Hälfte des Proteins ausmachen.

Man vermutet, dass die N-terminale Hälfte durch Binden von Kupfer-Ionen [41] teilweise Struktur annehmen kann [42]. Aus experimentellen Gründen waren jedoch an das mit NMR untersuchte PrP^C-Material durch Wahl eines niedrigen pH-Wertes keine Cu²⁺-Ionen gebunden. Bei diesen Bedingungen kann Kupfer nicht binden, weil die Cu²⁺-bindenden Histidin-Residuen im N-terminalen Teil von PrP^C protoniert sind.

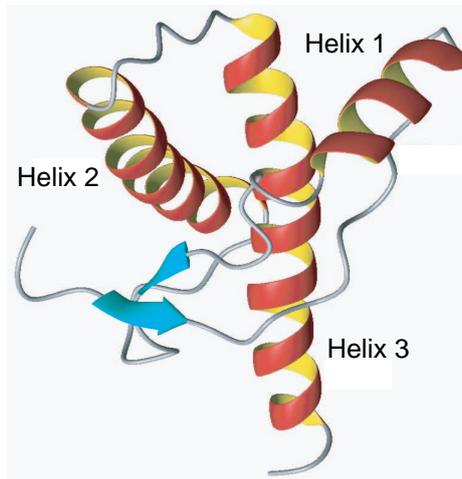


Abbildung 1.6: NMR-Struktur von PrP^C. NMR-Struktur des strukturierten C-terminalen Abschnitts (Residuen 125 bis 228) von rekombinant hergestelltem menschlichen PrP^C [44] in einer vereinfachten Darstellung. Die Struktur umfasst drei α-Helizes (als rote und gelbe Schlangen symbolisiert) und ein anti-paralleles β-Faltblatt (hellblaue Pfeile). Eine Disulfidbrücke (nicht dargestellt) verbindet Helix 2 und Helix 3 (vgl. Abbildung 1.5).

Die Struktur der C-terminalen globulären Hälfte von menschlichem PrP^C ist in Abbildung 1.6 dargestellt und durch drei α-Helizes sowie ein kurzes anti-paralleles β-Faltblatt ausgezeichnet [44]. Aus zahlreichen weiteren NMR-Messungen von rekombinant hergestellten Prion-Proteinen vieler unterschiedlicher Spezies [45–55] weiß man, dass die Prion-Proteine aller untersuchten Spezies sehr ähnliche Strukturen wie das gezeigte menschliche PrP aufweisen.

Weil PrP^C aus natürlichen Quellen in den für NMR-Messungen notwendigen großen Mengen nur schwer zugänglich ist, wurde in den zitierten Studien stattdessen rekombinant von *Escherichia coli* Bakterien hergestelltes PrP verwendet. Den rekombinanten PrP^C-Konstrukten fehlen jedoch die beiden posttranslationalen Modifikationen, die natives PrP^C kennzeichnen (vgl. Abbildung 1.5). Diese Modifikationen der Aminosäurekette sind zwei angehängte Zuckerseitenketten [56, 57] sowie ein C-terminaler Glycosylphosphatidylinositol (GPI)-Anker [58], über den PrP^C von außen in die Zellmembran eingehängt wird. Kürzlich konnten NMR-spektroskopische und Zirkulardichroismus (CD)-Messungen an bovinem PrP^C durchgeführt werden, welches aus gesunden Kälberhirnen behutsam aufgereinigt worden war, sodass die angehängten Zucker-Moleküle und ein Teil des GPI-Ankers erhalten blieben [59]. Nach dieser Studie haben diese Modifikationen keinen Ein-

fluss auf die C-terminale Struktur von PrP^C. Insgesamt ist damit PrP^C ein zum Teil nativ gefaltetes, posttranslational stark prozessiertes und membranverankertes Protein, welches jedoch auch nativ ungefaltete Abschnitte aufweist.

Zur Struktur der krankheitsassoziierten Isoform PrP^{Sc} ist dagegen viel weniger bekannt. Die mangelnde Kenntnis liegt darin begründet, dass PrP^{Sc} bedauerlicherweise für hochauflösende Techniken der strukturellen Analyse unzugänglich ist: Konventionelle Flüssig-NMR-Techniken scheitern, weil die PrP^{Sc}-Fibrillen zu groß und unlöslich sind, und Röntgenbeugung kann nur bedingt eingesetzt werden, weil es kaum möglich zu sein scheint, dreidimensionale Kristalle dieser Fibrillen zu züchten. Damit kann Röntgenbeugung lediglich an der eindimensional geordneten „Kristallstruktur“ der Fibrillen ansetzen. Aus der CD- und Infrarot-Spektroskopie ist aber bekannt, dass PrP^{Sc} im Vergleich zu PrP^C einen stark erhöhten Anteil an β -Faltblattstrukturen und einen leicht verringerten Anteil an α -helikalen Strukturen aufweist [60, 61]. Experimente zur Röntgenbeugung an PrP 27-30-Stäbchen gaben Hinweise darauf, dass die enthaltenen β -Faltblätter parallel zur Fibrillenachse ausgerichtet sind [62].

Aufgrund der unscharfen experimentellen Daten existieren für die Struktur von PrP^{Sc} nur spekulative Modelle. Abbildung 1.7 zeigt ein β -helikales Modell, das ursprünglich basierend auf niedrig auflösenden elektronenmikroskopischen (EM) Daten zu zweidimensionalen Kristallen von PrP^{Sc}-Protofibrillen vorgeschlagen [63, 64] und von meiner Kollegin Martina Stork in Zusammenarbeit mit Armin Giese (Zentrum für Neuropathologie und Prionforschung, LMU) und Paul Tavan hinsichtlich einiger Details verbessert wurde [27]. Dieses Modell geht davon aus, dass bei der Umwandlung von PrP^C in PrP^{Sc} die beiden α -Helizes 2 und 3 erhalten bleiben und dass derjenige Abschnitt, der in PrP^C Helix 1 bildet, in PrP^{Sc} zu einer dreieckigen, linkshändigen β -Helix beiträgt (vgl. Abbildung 1.6). Das β -helikale Sekundärstrukturmotiv ist aus natürlich vorkommenden Proteinen wie dem Enzym UDP-N-Acetylglucosamine Acyltransferase bekannt [65]. Die von Stork et al. [27] vorgeschlagene Detailverbesserung betrifft die Platzierung von polaren oder geladenen Residuen in eine sogenannte Loop-Struktur, welche die reguläre β -Helix unterbricht. Auf diese Weise zeigen keine polaren oder geladenen Residuen ins apolare Helix-Innere, sondern liegen so, dass sie im umgebenden Medium solvatisiert werden können. Eine derartige Anordnung ist bei triangulären parallelen β -Helizes aus Stabilitätsgründen zwingend zu fordern [27].

Abbildung 1.8 zeigt zum einen, wie jeweils drei β -helikale PrP^{Sc}-Moleküle ein Trimer bilden und wie auf diese Weise die aus den EM-Messungen erhaltenen Randbedingungen erfüllt werden können (links). Zum anderen sind dort mehrere derartige Trimer-Scheiben zu einer kurzen Protofibrille gestapelt (rechts). Nach diesem Modell könnten sich mehrere solcher Protofibrillen zu längeren Prion-Stäbchen zusammensetzen. Auf diese Weise versucht das Modell, dem in Abschnitt 1.3 diskutierten Fibrillenwachstum eine strukturelle Basis zu verleihen. Die Zuckerseitenketten und der Membrananker wurden bei der Modellbildung mit berücksichtigt und sind lediglich aus Gründen der Übersichtlichkeit hier nicht mit abgebildet.

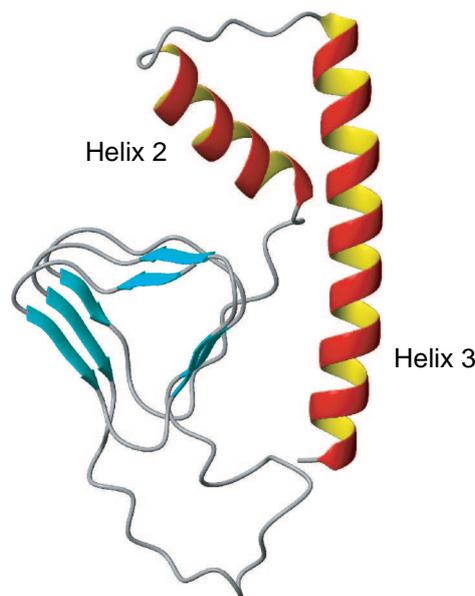


Abbildung 1.7: Dreieckiges β -helikales Modell für PrP^{Sc}. Nach diesem Modell bleiben bei der Umwandlung von PrP^C in PrP^{Sc} Helix 3 vollständig und Helix 2 bis zur Disulfidbrücke erhalten (vgl. Abbildungen 1.5 und 1.6), während der N-terminale Rest der Sequenz beginnend bei Residuum 106 in eine β -helikale Struktur eingebunden ist. Der hier abgebildete Teil des Modells umfasst die Sequenz der Aminosäuren 106 bis 228 und reicht damit N-terminal über den in PrP^C globulär strukturierten Teil hinaus. Abschnitte, die in PrP^C unstrukturiert sind oder Helix 1 bilden, tragen in PrP^{Sc} zu einer linkshändigen β -Helix bei. Die mittlere Windung der regulären dreifach gewundenen β -Helix ist von einer Loop-Region (134-153) unterbrochen, in der geladene und polare Residuen platziert sind. Dies gewährleistet, dass nur apolare Residuen ins Innere der Helix zeigen, was für die Stabilität solcher Strukturen notwendig ist (nach „Model 2“ aus Stork et al. [27]).

Die weiter oben ausgeführten strukturellen Betrachtungen von PrP^C haben offen gelassen, auf welche Weise die posttranslationalen Modifikationen in den Zellen vorgenommen werden. Auch die mögliche Funktion von PrP^C im zellulären Kupferhaushalt wurde hier nur angedeutet. Diese Aspekte sollen im folgenden Abschnitt detaillierter betrachtet werden.

1.6 Der Lebenszyklus und die Funktion von PrP

Abbildung 1.9 zeigt schematisch Abschnitte aus dem Lebenszyklus von PrP^C. Nach der Synthese wird PrP^C im endoplasmatischen Retikulum (ER) und im Golgi-Apparat posttranslational prozessiert und erhält dabei bis zu zwei komplexe Zucker-Seitenketten und den GPI-Membrananker. Anschließend wird das reife PrP^C sekretiert und über den GPI-Anker in Cholesterin-reiche Gebiete der Zellmembran eingehängt (Schritte 1 und 2 in Abbildung 1.9). Dort könnte PrP^C seine mögliche Funktion als Kupfer-bindendes Protein erfüllen (3). Von der Zellmembran wird PrP^C in Vesikel verpackt und wieder in die Zelle

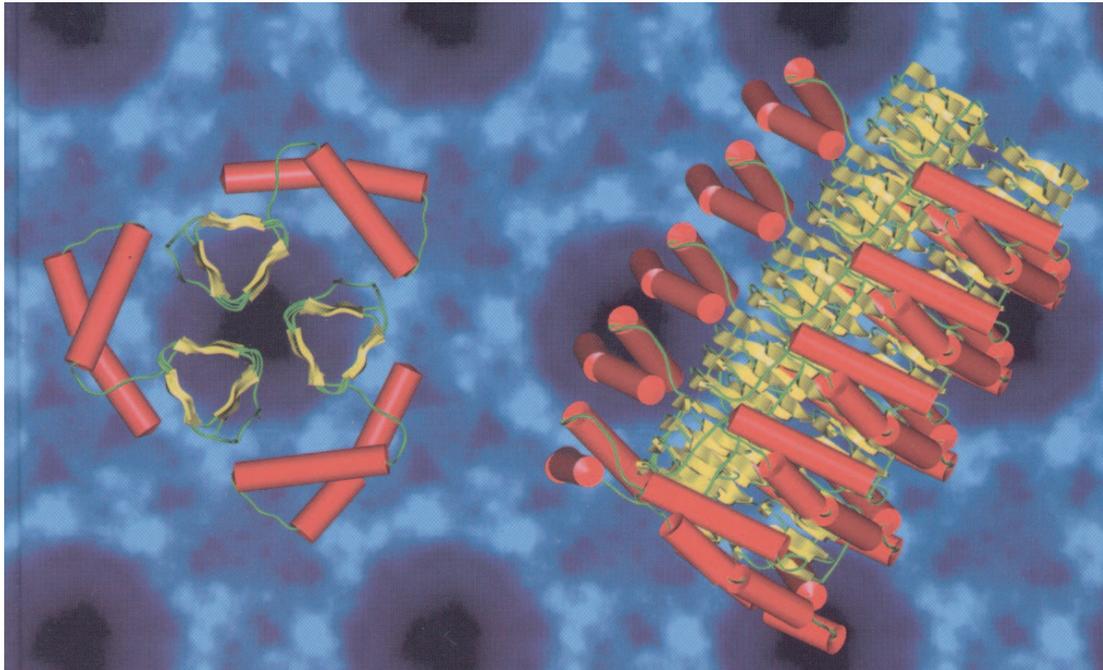


Abbildung 1.8: Fibrillen aus β -helikalen PrP^{Sc}-Molekülen. Im Hintergrund ist eine EM-Karte von PrP 27-30-Kristallen abgebildet, auf die ein Trimer aus β -helikalen Modellen für PrP^{Sc} nach [63, 64] überlagert wurde (im Vordergrund links). Nach diesem Modell bilden mehrere solche Trimere kurze Fibrillen-Abschnitte (rechts). β -Stränge sind als gelbe flache Pfeile und α -Helizes sind als rote Zylinder dargestellt. Übernommen von [66].

zurückgeholt, kann von dort aus abermals an die Zellmembran gebracht werden, oder zu Lysosomen, in denen es von hydrolysierenden Enzymen zerlegt wird (4).

Die zelluläre Funktion von PrP^C ist, obschon sie in zahlreichen Studien untersucht wurde, weitgehend unklar (einen aktuellen Überblick zur Funktion von PrP^C bieten Roucou und LeBlanc [68]). Es gibt aber Hinweise darauf, dass PrP^C Neuronen vor oxidativem Stress schützt und vor der Apoptose, dem programmierten Zelltod, bewahren kann [68]. Weil PrP^C Kupfer-Ionen binden kann, vermutet man außerdem eine Rolle in der Kupfer-Homöostase, d.h. in der Selbstregulierung des Kupferhaushalts der Zellen [40]. Weil PrP^C von den meisten Körperzellen exprimiert wird, wenn auch in besonders großen Mengen nur von Nervenzellen, schließt man auf eine allgemeine Bedeutung von PrP^C, wenn auch andererseits die in Abschnitt 1.2 angesprochenen Experimente mit PrP-knockout-Mäusen andeuteten, dass PrP für das normale Funktionieren von Zellen nicht sonderlich essentiell zu sein scheint.

Im kranken Organismus wird PrP^C an der Zellmembran oder in den Vesikeln in die pathologische Konformation PrP^{Sc} umgefaltet, aus der es nicht oder in beschränktem Umfang abgebaut werden kann (Schritt 5 in Abbildung 1.9). Infolgedessen sammelt sich PrP^{Sc} dort an, verursacht über bislang unbekannte Signalwege neuronale Fehlfunktion und den Zelltod und lagert sich schließlich ab.

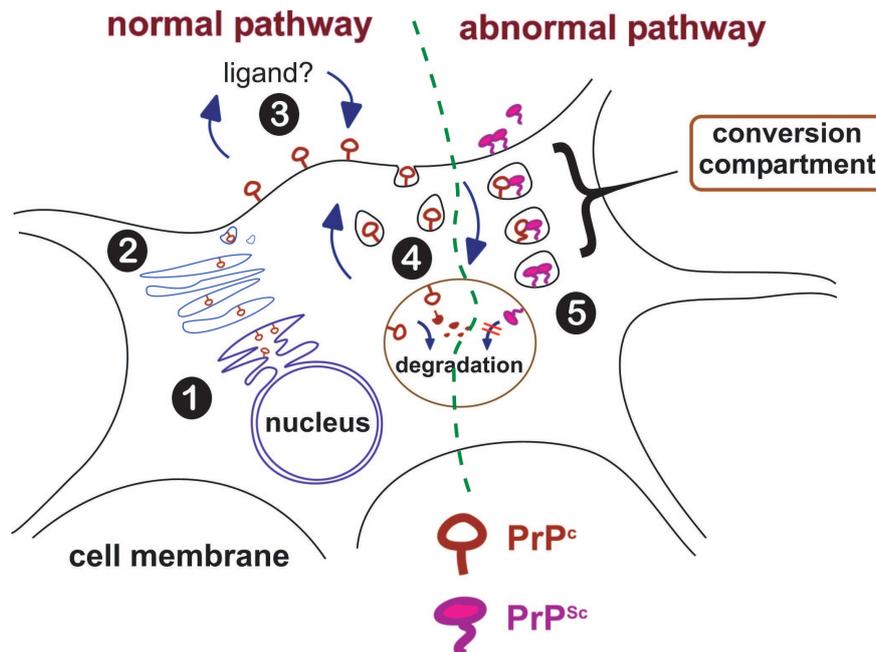


Abbildung 1.9: Lebenszyklus von PrP^C und Entstehung von PrP^{Sc}. PrP^C wird in den Ribosomen hergestellt (1) und anschließend im ER und im Golgi-Apparat posttranslational mit komplexen Zuckerseitenketten und einem GPI-Anker versehen. Über den GPI-Anker wird PrP^C außen in die Plasmamembran eingehängt (2). Dort kann es beispielsweise Liganden wie Kupfer binden (3). Durch Endozytose werden PrP^C-Moleküle in Vesikeln zu den Lysosomen transportiert, wo sie abgebaut werden (4). Die Umwandlung von PrP^C in PrP^{Sc} findet vermutlich an der Zelloberfläche oder in den Vesikeln statt. Im Gegensatz zu PrP^C kann PrP^{Sc} aber nicht abgebaut werden und bildet Ablagerungen (5). Abbildung übernommen von [67] mit freundlicher Genehmigung von Prof. Ian Jones.

Hinsichtlich der Entwicklung von anti-Prion-Therapeutika liefert die Betrachtung des Lebenszyklus von PrP^C zusätzliche Hinweise. Geht man davon aus, dass die Prion-Aggregation teilweise außen an der Zellmembran stattfindet, so könnten mögliche aggregationshindernde Wirkstoffe, sobald sie an der Zelloberfläche sind, ungehindert diesen Wirkort erreichen und müssten die Zellmembran nicht überwinden. Infolgedessen stellen sogenannte Zellkulturmodelle, in denen Zellen frei zugänglich in einem Nährmedium kultiviert werden, ein vielversprechendes Testsystem auch für schlecht membrangängige Wirkstoffkandidaten dar. Derartige Modellsysteme für Prionkrankheiten werden von zwei unserer Partnerprojekte im BMBF-Forschungsverbund zur Suche nach Medikamenten eingesetzt und werden in Kapitel 2 eingehend vorgestellt. Von einem Medikament muss aber verlangt werden, dass es oral oder zumindest parenteral, d.h. direkt in die Blutbahn, verabreicht werden kann. Um in das Zentrale Nervensystem zu gelangen, muss der verabreichte Wirkstoff dazu fähig sein, die Blut-Hirn-Schranke zu überwinden. In Zellkulturmodellen identifizierte Wirkstoffe müssen deshalb hinsichtlich der Hirngängigkeit getestet und gegebenenfalls optimiert werden.

1.7 Bekannte anti-Prion-Substanzen

Vor dem Beginn der Arbeiten des BMBF-Forschungsverbundes waren bereits eine Reihe von chemischen Verbindungen bekannt, die wirksam in die Vermehrung von PrP^{Sc} eingreifen können. Dazu zählen Congo red [69, 70], Porphyrine/Phthalocyanine [71–73], Cp-60 [74], polykationische Lipide wie etwa DOSPA (2,3-Dioleoyloxy-*N*-[2(sperminecarboxamido)ethyl]-*N,N*-dimethyl-1-propanaminium trifluoroacetate) [75], chemische Chaperone [76], Suramin [77] und Akridin-Derivate [78–80]. Die meisten dieser Verbindungen wurden durch empirische und manchmal glückliche Beobachtungen identifiziert oder waren als Medikamente gegen andere Krankheiten bereits bekannt und wurden deshalb auch auf ihre Wirkung gegen die Prion-Aggregation getestet. Leider konnte aber aus diesen Substanzen bislang kein wirksames Medikament entwickelt werden.

Einer der vielversprechendsten Kandidaten ist das seit rund 70 Jahren bekannte Malaria-Medikament Quinacrine (QA) und weitere Akridin-Derivate, für die man herausgefunden hat, dass sie die Anreicherung von Protease-resistentem PrP in kultivierten, mit Scrapie infizierten Zellen hemmen [78–80]. Versuche mit Scrapie-infizierten Mäusen, denen QA verabreicht wurde, zeigten jedoch keine signifikante Verlängerung der Inkubationszeit [81, 82]. Dennoch deuteten vorläufige klinische Versuche, bei denen CJD-Patienten mit QA behandelt wurden, auf eine vorübergehende Verbesserung der Symptome hin [83, 84]. Zur Zeit wird QA in klinischen Studien in London und San Francisco an CJD-Patienten getestet [85–87].

Um herauszufinden, weshalb QA zwar *in vitro* aber nicht oder nur schwach *in vivo* wirksam ist, wurde die Pharmakokinetik von QA untersucht [88, 89], wobei mehreren Versuchstieren QA in therapeutischen Dosen verabreicht wurde und anschließend gemessen wurde, in welchen Konzentrationen der Wirkstoff in verschiedenen Bereichen des Gehirns vorlag. Diese Studien lieferten als mögliche Erklärungen für die mangelnde Wirksamkeit, dass QA in zu geringen Konzentrationen am Wirkort vorliegt. Daher wird derzeit anhand von QA-Analoga getestet, wie sich Änderungen der Molekülstruktur auf die Metabolisierungseigenschaften und die biologische Aktivität der resultierenden Verbindungen gegen die Anreicherung von Prion-Material auswirken [90]. Es wird also versucht, die pharmakokinetischen Eigenschaften von QA zu verbessern.

Das Beispiel QA zeigt, dass es für die Entwicklung eines anti-Prion-Medikaments notwendig sein kann, in jahrelangen Forschungsbemühungen viele verschiedenartige Experimente einzusetzen, ohne dass dabei gewährleistet ist, dass am Ende ein taugliches Medikament zur Verfügung steht. Das entscheidende Problem, das am Beispiel der Entwicklung von QA auch deutlich wird, ist, dass man sich mit QA initial auf eine oder wenige Verbindungen einer strukturellen Klasse festgelegt hatte und erst nachträglich zusätzliche analoge Wirkstoffsubstanzen gesucht hat.

1.8 Neue Strategien

Einen alternativen Ansatz, mit dem üblicherweise in der industriellen Pharmaforschung die Entwicklung von neuen Wirkstoffen gegen eine Krankheit begonnen wird, bietet die systematische Suche nach neuen Wirkstoffen aus riesigen Sammlungen von Wirkstoffkandidaten nach dem Prinzip des Hochdurchsatz-Screenings (HTS). Die dabei identifizierten Leitmoleküle können dann hinsichtlich ihrer therapeutischen Wirksamkeit und ihrer pharmakologischen Eigenschaften optimiert werden. Infolge der Größen der durchsuchten Substanzsammlungen werden üblicherweise eine Reihe von Leitmolekülen gefunden, welche dann parallel weiterverfolgt werden. Auf diese Weise soll verhindert werden, dass man sich in einer frühen Entwicklungsphase auf nur eine oder wenige Substanzklassen festlegt, die sich später als untauglich herausstellen könnten.

Als Voraussetzung für die Durchführung einer derartigen HTS-Kampagne müssen zum einen die für die zu heilende Krankheit wesentlichen Proteine als Target-Moleküle für die zu findenden Wirkstoffe bekannt sein, und zum anderen muss ein *in vitro* Modellsystem existieren, das diese Target-Moleküle enthält und in dem die wesentlichen molekularen Prozesse der Krankheit ablaufen. Im Falle der Prionkrankheiten ist die Prion-Vermehrung wohl der wesentliche Prozess. Damit sind die Isoformen PrP^C und PrP^{Sc} als Target-Moleküle für potenzielle Wirkstoffe identifiziert. Was nun zusätzlich benötigt wird, ist zum einen ein *in vitro* Assay für die Prion-Vermehrung, in dem die Aggregation stabil und reproduzierbar abläuft, und zum anderen ein Messverfahren, mit dem die Größen der gebildeten Aggregate rasch bestimmt werden können.

Ein derartiges Verfahren, das zum Screening von großen Substanzbibliotheken eingesetzt werden kann, wurde im Rahmen des BMBF-Forschungsverbundes von einem unserer Projektpartner, der Gruppe von Prof. Kretzschmar (Zentrum für Neuropathologie und Prionforschung, LMU), entwickelt. Mit dieser auf Fluoreszenz-Korrelations-Spektroskopie basierenden Methode des *Scanning for Intensely Fluorescent Targets* (SIFT) wurde ein *in vitro* Aggregations-Assay etabliert. Im Rahmen des von Prof. Tavan initiierten Teilprojekts zum Forschungsverbund bestand einer meiner Hauptbeiträge in der Entwicklung automatisierter Verfahren zur zuverlässigen Auswertung der mit dem SIFT-Verfahren erzeugten Daten.

Darüber hinaus wurden in den ebenfalls am BMBF-Verbund beteiligten Gruppen von Prof. Tatzelt und Prof. Groschup Zellkulturmodelle für die Medikamentenentwicklung eingesetzt. Das Zellkulturmodell der Gruppe Groschup ist sogar so weit entwickelt, dass damit – wie mit dem *in vitro* SIFT-Assay – Substanzbibliotheken systematisch nach neuen Wirkstoffen durchsucht werden können.

Sowohl das SIFT-Verfahren, die Zellkulturmodelle und weitere im Verbund eingesetzte experimentelle Modellsysteme für Prionkrankheiten, als auch die mit diesen Verfahren gewonnenen Daten, die von mir in einer Datenbank gesammelt wurden, werden im folgenden Kapitel vorgestellt.

2 Der Forschungsverbund „TSE-Therapie“

Wie ich in der Einleitung herausgestellt habe, laufen seit Anfang 2002 die Arbeiten des vom BMBF geförderten Forschungsverbundes „Systematische Entwicklung neuer Wirkstoffe für die kausale Therapie von Prionkrankheiten“ [4], an dem ich seit Juli 2002 mitgearbeitet habe. Dieser interdisziplinäre Verbund gliedert sich in insgesamt sieben Teilprojekte (TP) aus den Bereichen Humanmedizin, Tiermedizin, Virologie, Biochemie, Chemie und Theoretische Physik. Fünf der sieben Teilprojekte (TP1, 2, 3, 4 und 6) verfolgen in enger Zusammenarbeit einen gemeinsamen Ansatz zur Entwicklung von Medikamenten, der eine systematische Suche nach neuen Wirkstoffen in einer großen Sammlung von Testsubstanzen nach dem Prinzip des HTS, sowie die anschließende Optimierung der dabei identifizierten Wirkstoffe hinsichtlich ihrer pharmakologischen Eigenschaften umfasst. Zwei weitere Teilprojekte befassten sich mit dem Einfluss von Kupfer auf die Entwicklung der Prionkrankheiten bzw. mit der Entwicklung einer Immunprophylaxe und spielen Nebenrollen in dem zentralen Screening-Ansatz, von dem hier berichtet wird.

Das Übersichtsdiagramm in Abbildung 2.1 zeigt die fünf am Screening teilnehmenden Teilprojekte des Verbundes, sowie deren Mitarbeiter, die eingesetzten Techniken und die Einsatzzwecke der Techniken für den Screening-Ansatz. Sowohl Teilprojekt TP1 um Prof. Kretzschmar und Dr. Giese als auch TP3 um Prof. Groschup brachten mit dem SIFT-Assay bzw. einem Zellkultur-Assay zwei Modellsysteme für Prionkrankheiten mit in den Verbund, die es erlaubten, große Mengen von Substanzen in hohem Durchsatz nach Wirkstoffen durchzumustern. Demgegenüber diente das von Prof. Tatzelt eingesetzte Zellkulturmodell zur Validierung von mit dem SIFT-Assay gefundenen Wirkstoffen. Prof. Schmidt konnte seine medizinisch-chemische Erfahrung einbringen und gezielt zusätzliche Substanzen aus identifizierten und vielversprechenden Wirkstoffklassen synthetisieren und weiteren Tests zur Verfügung stellen. Das von mir zusammen mit Prof. Tavan bearbeitete Teilprojekt TP2 nahm im Verbund die zentrale Rolle der Sammlung, Auswertung, Aufbereitung und Verteilung der in den Testsystemen des Verbundes aufgenommenen Daten ein.

Einer meiner Hauptbeiträge betraf dabei die automatisierte Auswertung der mit dem SIFT-Verfahren generierten Messdaten. Die Auswertung umfasste die Entwicklung eines Aktivitätsmaßes, welches jeder gemessenen Substanz eine Zahl zuordnet, die kodiert, wie wirksam die gegebene Substanz die Prion-Aggregation in dem mit der SIFT-Technik untersuchten *in vitro* Assay hindern kann. Darüber hinaus richtete ich eine Datenbank ein, in der die rohen SIFT-Daten abgelegt werden. Die Datenbank enthält des Weiteren sowohl die berechneten SIFT-Aktivitäten als auch die in den weiteren Modellsystemen des Verbundes ermittelten Wirksamkeiten der getesteten Substanzen. Die zur automatisierten Auswertung entwickelten Verfahren und die Haltung der Daten werden in Kapitel 3

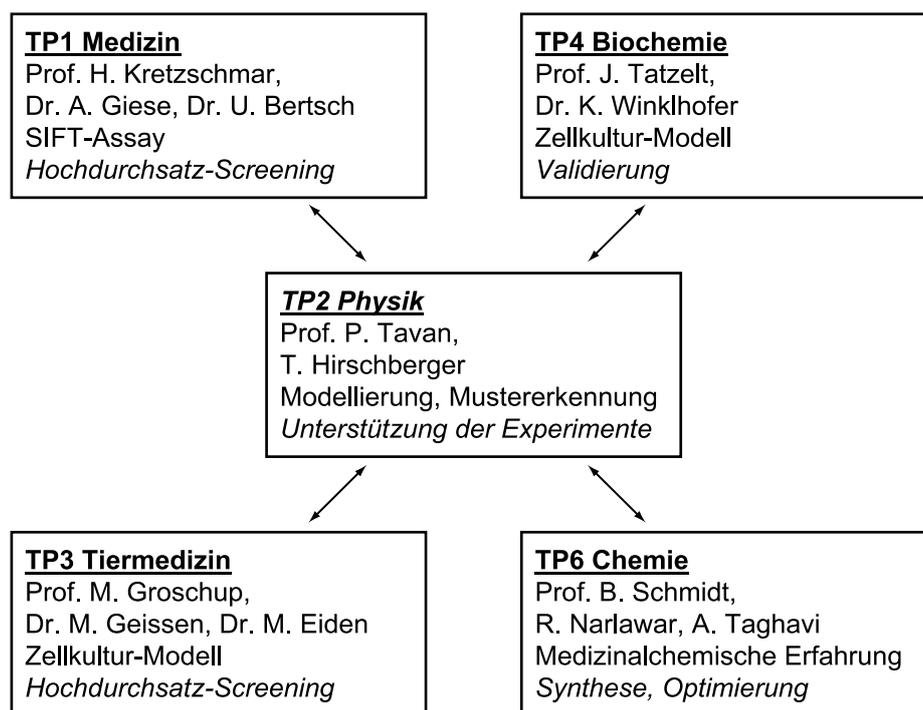


Abbildung 2.1: Teilprojekte des Forschungsverbundes „TSE-Therapie“. Die am Screening beteiligten Teilprojekte des BMBF-Verbundes, die zur Verfügung stehenden experimentellen Techniken und deren Einsatzzwecke für die systematische Suche nach neuen Wirkstoffen. Sowohl der in TP1 eingesetzte SIFT-Assay als auch das Zellkulturmodell aus TP3 sind HTS-fähig und konnten daher zum Testen von großen Anzahlen von Wirkstoffkandidaten eingesetzt werden. Das Zellkulturmodell aus TP4 diente zur Validierung von gefundenen Wirkstoffen. In TP6 wurden Wirkstoffe hinsichtlich ihrer strukturellen Merkmale optimiert und zusätzliche Substanzen synthetisiert, welche dann wiederum mit den Verfahren der anderen Teilprojekte hinsichtlich ihrer Wirksamkeit gegen Prionkrankheiten getestet werden konnten. Unser Teilprojekt TP2 übernahm die Sammlung, Auswertung und Aufbereitung der gewonnenen experimentellen Daten, sowie die Kommunikation der aufbereiteten Daten innerhalb des Verbundes.

skizziert. Auf Basis der für die Substanzen gesammelten Wirksamkeiten sowie der für alle Testsubstanzen zugreifbaren Molekülstrukturen führte ich strukturelle Analysen durch. Die für diesen Zweck von mir entwickelten und eingesetzten Verfahren werden in Kapitel 4 vorgestellt.

Automatisierte Auswertung, zentrale Datenhaltung und strukturelle Analysen dienten der Identifikation von neuen Wirkstoffklassen. So wurde die Klasse der N'-Benzylden-Benzohydrazide (NBB)-Derivate als potenzielle anti-Prion-Therapeutika identifiziert [91–93] und durch ein Patent geschützt [94]. Darüber hinaus konnten wir eine weitere Wirkstoffklasse finden, die derzeit geschützt werden soll und deshalb in dieser Arbeit nicht offenlegt werden kann. Nichtsdestoweniger berichte ich in Kapitel 4 über die Methoden, die zur Entdeckung dieser und weiterer Wirkstoffklassen geführt haben.

Damit ich unsere Beiträge zum Verbund in den Kapiteln 3 und 4 darstellen kann, ist es notwendig, zunächst die experimentellen Verfahren unserer Projektpartner vorzustellen. Dazu sollen vorab die angesprochenen Sammlungen von Substanzen, die in den diversen Assays des Verbundes getestet wurden, kurz charakterisiert werden.

2.1 Die DIVERSet-Bibliotheken

Für die systematische Suche nach neuen Wirkstoffen wurden vom Verbund zwei sogenannte Substanzbibliotheken, d.h. Sammlungen von potenziellen Wirkstoff-Substanzen, beschafft. Wir entschieden uns für die von der Firma ChemBridge [95] angebotene Bibliothek DIVERSet, die insgesamt rund 50.000 Substanzen beinhaltet, und kauften davon zwei jeweils 10.000 Substanzen umfassende Teilbibliotheken, die wir „DIVERSet 1“ und „DIVERSet 2“ genannt haben. DIVERSet wird von ChemBridge [95] als eine vielfältige Sammlung von medikamentartigen kleinen Molekülen beschrieben, die *rational* so ausgewählt wurden, dass sie vielversprechende pharmakologische Eigenschaften besitzen.

Damit eine chemische Verbindung ein Medikament sein kann, muss sie pharmakologische Eigenschaften besitzen, welche in dem häufig verwendeten Kürzel ADMET (Absorption, Distribution, Metabolismus, Exkretion und Toxikologie) zusammengefasst sind. Ob die in DIVERSet enthaltenen Verbindungen in diesem Sinne tatsächlich medikamentartig sind und günstige pharmakologische Eigenschaften besitzen, kann nur durch aufwändige Tierversuche getestet werden, bei denen einzelne Verbindungen verabreicht und deren Verteilung im Organismus untersucht wird. Derartige Messungen können daher nur für einige wenige Substanzen durchgeführt werden, die bereits in verschiedenen Assays als wirksam befunden wurden.

Daher basiert die von ChemBridge getroffene „rationale Auswahl“ auf sogenannten 3D-Pharmakophor-Analysen [96], welche die Eigenschaften von bekannten Medikamenten berücksichtigen. Unter dem Pharmakophor eines Medikamentes versteht man dabei die räumliche Anordnung der wesentlichen chemischen Eigenschaften, die für die biologische Aktivität des Medikamentes verantwortlich sind, ein Konzept das auf Arbeiten von Paul Ehrlich [97] zurückgeht. Die DIVERSet-Bibliothek wurde von ChemBridge auf eine Weise, die nicht offengelegt ist, so zusammengestellt, dass von den enthaltenen Substanzen ein großes Gebiet des Raumes aller möglichen Pharmakophore abgedeckt wird. Dabei versucht ChemBridge zu gewährleisten, dass eine möglichst hohe Vielfalt an strukturell unterschiedlichen Substanzen in der Sammlung enthalten ist und dass diese Substanzen hinsichtlich ihrer biologischen Aktivität vielversprechend sind.

Infolge der von ChemBridge getroffenen Auswahl erfüllen die meisten der DIVERSet-Substanzen die Kriterien von „Lipinski’s rule of five“ [98], einer statistisch begründeten Faustregel, welche die pharmakologischen Eigenschaften einer Substanz aus physikochemischen Eigenschaften der Substanz vorherzusagen sucht. Hierzu hatten Lipinski et al. [98] eine Auswahl von gut 2000 verschiedenartigen Wirkstoffen untersucht, die es in

die klinische Versuchsphase II geschafft hatten und daher durch günstige pharmakologische Eigenschaften ausgezeichnet sind. Statistische Analysen der pharmakologischen und physikochemischen Eigenschaften ergaben die unscharf formulierte Vorhersage, dass die Absorption oder die Permeation einer gegebenen Substanz wahrscheinlich schlecht sind, wenn diese Substanz mehr als 5 H-Brücken-Donatoren, mehr als 10 H-Brücken-Akzeptoren, ein Molekulargewicht größer als 500 oder eine Lipophilie LogP größer als 5 aufweist [98]. Dabei ist das Lipophilie-Maß LogP eine aus der Molekülstruktur empirisch berechenbare Größe. Weil die vier Kriterien als Grenzen Vielfache von 5 enthalten, wird diese Regel „rule of five“ genannt. Auf diese physikochemischen und auf weitere Eigenschaften der DIVERSet-Substanzen werde ich in Abschnitt 3.3 im Zusammenhang mit der von mir vorgenommenen Datenhaltung der Bibliotheken näher eingehen.

In einer kürzlich erschienenen Studie [99] wurden 45 kommerziell verfügbare Bibliotheken hinsichtlich der strukturellen Vielfalt und der Leitstrukturartigkeit („leadlikeness“) [100] der enthaltenen Substanzen systematisch untersucht. Dabei wurden zum einen die enthaltenen Substanzen nach ihrer Molekülstruktur klassifiziert und die Anzahl sowie die Größen der resultierenden Klassen als Kriterien für die strukturelle Vielfalt herangezogen. Zum anderen wurde getestet, wie viele Verbindungen statistisch abgeleitete Faustregeln wie Lipinskis *rule of five* erfüllen und demnach leitstrukturartig sind. Diese Studie bestätigt den von ChemBridge angebotenen Sammlungen eine hohe Vielfalt und Leitstrukturartigkeit.

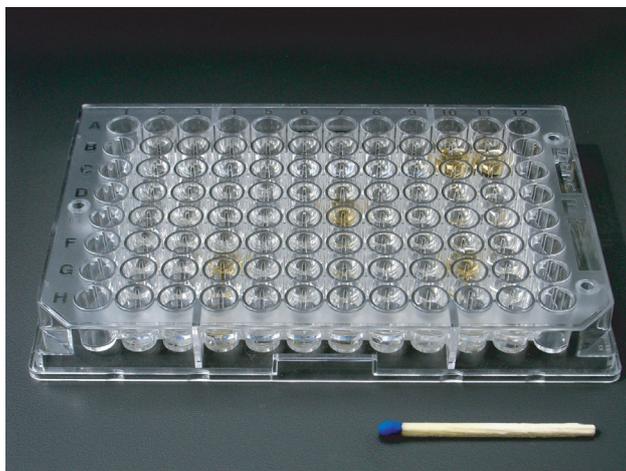


Abbildung 2.2: Eine 96-Well-Mikrotiterplatte. Die chemischen Verbindungen der DIVERSet-Bibliotheken sind in DMSO gelöst und auf 96-Well-Mikrotiterplatten verteilt. Eine solche Mikrotiterplatte umfasst 96 Töpfchen, die „Wells“ genannt werden und in 8 Zeilen und 12 Spalten angeordnet sind. Üblicherweise werden die Zeilen mit den Buchstaben A bis H bezeichnet und die Spalten von 1 bis 12 nummeriert. Auf den Platten der DIVERSet-Bibliotheken sind die erste und die zwölfte Spalte freigelassen und bieten Platz für Kontrollmessungen.

Die Substanzen der DIVERSet-Bibliotheken wurden vom Hersteller in dem organischen Lösungsmittel Dimethylsulfoxid (DMSO) gelöst und auf 96-Well-Mikrotiterplatten ver-

teilt geliefert. DMSO ist als Lösungsmittel weit verbreitet, weil darin viele Substanzen gut lösbar sind und weil DMSO sich gut mit Wasser oder anderen Lösungsmitteln mischen lässt. Abbildung 2.2 zeigt eine solche 96-Well-Mikrotiterplatte. Auf jeder DIVERSet-Platte sind 80 Substanzen so auf die 96 Töpfchen (Wells) verteilt, dass die erste und die letzte Töpfchenspalte (1 und 12) frei bleiben. Dieses Format ermöglicht es, dass alle Substanzen einer 96-Well-Bibliothekspalte (verdünnt) auf eine 96-Well-Messplatte übertragen und die beiden frei gebliebenen Spalten mit Kontrollproben befüllt werden können. Die Platten der Bibliotheken DIVERSet 1 und 2 werden von Projektpartner TP1 gekühlt gelagert. Von dort aus werden Verdünnungen der Substanzen für die Messungen in den verschiedenen Assays an die Verbundpartner verteilt.

2.2 Der anti-Prion 2D-SIFT-Assay (TP1)

Am von Prof. Hans Kretzschmar geleiteten Zentrum für Neuropathologie und Prionforschung (ZNP) der LMU in Großhadern wurde im Rahmen des Forschungsverbundes ein *high throughput screening* (HTS) nach Inhibitoren der Prion-Aggregation durchgeführt. Für das Screening wurde die auf *fluorescence correlation spectroscopy* (FCS) basierende SIFT-Technik eingesetzt, mit der über fluoreszierende Farbstoffe markierte Prion-Aggregate und deren Wachstum detektiert werden können.

2.2.1 Der *in vitro* Aggregations-Assay

Der mit der SIFT-Methode beobachtete Assay basiert auf einem zellfreien *in vitro* Modellsystem für die Bildung von Prion-Aggregaten [32]. Aufgereinigtes PrP^{Sc} wird dabei mit monomeren PrP^C-Molekülen gemischt, welche sich an die PrP^{Sc}-Partikel anlagern, sich dort strukturell umwandeln und so zum Wachstum der PrP^{Sc}-Aggregate beitragen können. Abbildung 2.3 zeigt schematisch die Bestandteile des eingesetzten Modellsystems. Zum einen sind dies PrP^{Sc}-Aggregate, die aus Gehirnen von CJD-Patienten extrahiert wurden – in der Abbildung ist ein solches Aggregat durch eine Kette aus magenta-farbenen Würfeln symbolisiert. Zum anderen beinhaltet die Mischung rekombinant hergestellte Prion-Proteine (rPrP) der Maus (grüne Würfel), welche PrP^C-Monomere modellieren und über Farbstoffe (Alexa 488) markiert sind, die im grünen Spektralbereich fluoreszieren. Des Weiteren befinden sich in der Mischung Exemplare (rote Y) eines speziellen Antikörpers [101], der ausschließlich an die PrP^{Sc}-Aggregate bindet. Die Antikörper sind mit rot fluoreszierenden Farbstoffen (Alexa 647) versehen und markieren daher die PrP^{Sc}-Aggregate.

Nach dem Ansetzen der Mischung binden viele rot markierte Antikörper an die PrP^{Sc}-Aggregate und bilden dadurch intensiv rot fluoreszierende Partikel. Lagern sich anschließend viele grün markierte PrP^C-Moleküle an die PrP^{Sc}-Aggregate an, so entstehen große Aggregate, die starke rote *und* grüne Fluoreszenzsignale emittieren können. Die FCS-basierte

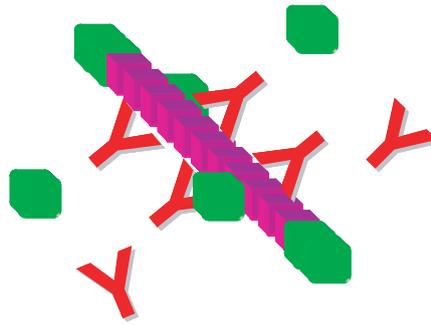


Abbildung 2.3: SIFT-Assay-Mischung. Schematische Darstellung der Zusammensetzung des SIFT-Assays. Ein menschliches PrP^{Sc}-Aggregat ist als Kette von magenta-farbenen Würfeln gezeichnet. Antikörper gegen menschliches PrP sind als rote Y dargestellt. rPrP der Maus ist durch grüne Würfel symbolisiert. Abbildung zur Verfügung gestellt von A. Giese.

SIFT-Technik dient nun dazu, einzelne große rot oder rot-grün fluoreszierende Partikel in der Mischung zu detektieren.

2.2.2 Die 2D-SIFT-Messtechnik

Mit FCS werden üblicherweise zeitliche Zusammenhänge der gemessenen Fluoreszenzintensitäten $F(t)$ untersucht, wobei die Fluktuationen $\delta F(t) = F(t) - \langle F(t) \rangle$ um den zeitlichen Mittelwert $\langle F(t) \rangle$ herangezogen werden, um daraus die normierte Autokorrelationsfunktion $G(\tau) = \langle \delta F(t) \delta F(t + \tau) \rangle / \langle F(t) \rangle^2$ für alle Verzögerungszeiten τ zu berechnen. Aus dem Verlauf von $G(\tau)$ kann die mittlere Diffusionszeit der beobachteten fluoreszierenden Teilchen bestimmt werden.

Bei der konfokalen FCS, auf der die SIFT-Methode basiert, wird der anregende Laserstrahl auf ein kleines Volumenelement (~ 1 fl) innerhalb der Probe fokussiert, und nur die in diesem Volumen erzeugte Fluoreszenz wird gemessen. Auf diese Weise können einzelne durch Fluoreszenzfarbstoffe markierte Partikel detektiert werden, wenn diese sich durch den Laserfokus bewegen [102]. Für die SIFT-Analyse, die auf diesem Prinzip basiert, werden die oben skizzierten zeitlichen Korrelationen und Diffusionszeiten nicht benötigt.

Stattdessen stützt sich die Auswertung auf die seltenen, hochintensiven Signale, die von einzelnen intensiv fluoreszierenden Partikeln im Laserfokus stammen. Um bei Messungen von geringen Konzentrationen großer, langsam diffundierender Partikel die Zahl der detektierten Partikel zu erhöhen, wird bei der SIFT-Technik der Laserfokus durch die Probe bewegt. Auf diese Weise „scannt“ man nach intensiv fluoreszierenden Targets. Ein- und zweifarbige Varianten der SIFT-Technik wurden am Institut von Prof. Kretzschmar schon früher auf dem Gebiet der Diagnose von Prionkrankheiten erfolgreich für die ultrasensitive Detektion von PrP^{Sc}-Aggregaten eingesetzt [103, 104]. Zur Untersuchung des

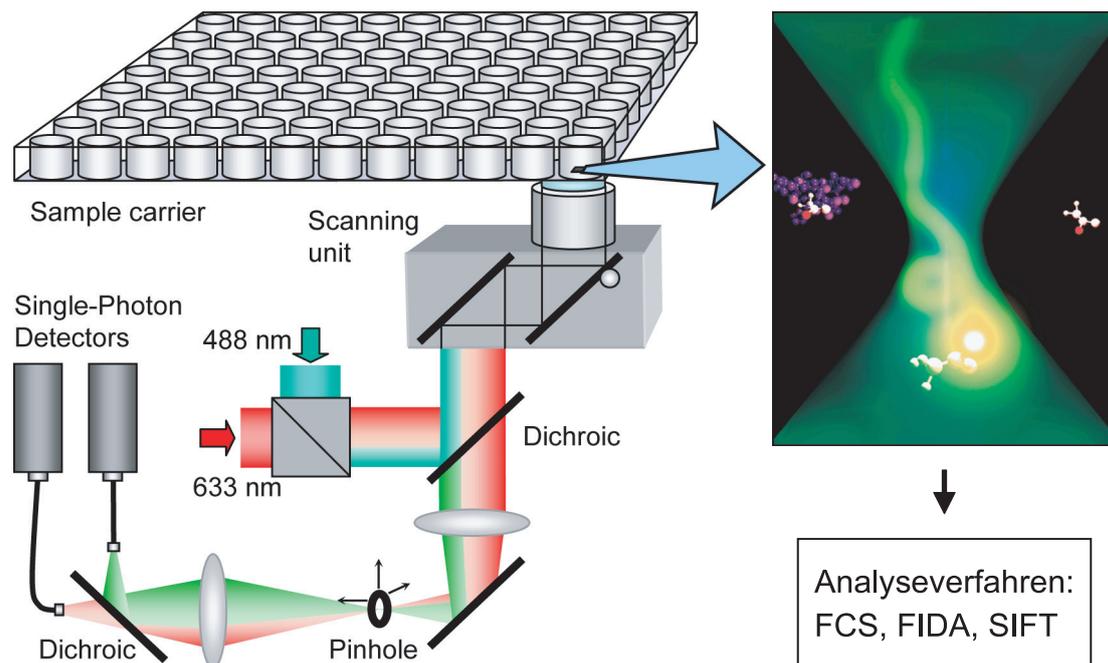


Abbildung 2.4: SIFT-Messaufbau. Mit dem abgebildeten Gerät „Insight Reader“ [105] wurden Mikrotiterplatten, in deren Töpfchen sich Assay-Mischungen mit jeweils unterschiedlichen Testsubstanzen befanden, mit der SIFT-Technik in hohem Durchsatz vermessen. Das Gerät bewegt dazu die einzelnen Töpfchen durch den gemeinsamen Fokus von zwei bei unterschiedlichen Wellenlängen (grün, 488 nm und rot, 633 nm) anregenden Lasern. Die von Fluoreszenzmarkierten Prion-Aggregaten im Laserfokus abgestrahlten grünen und roten Photonen werden von einem dichroitischen Spiegel nach der Farbe sortiert und von zwei getrennten Einzel-Photon-Detektoren aufgefangen. Die aufgenommenen Intensitätszeitreihen werden von einer zugehörigen Software analysiert. Neben SIFT-Daten können mit diesem Gerät auch weitere Messdaten erzeugt werden, die auf FCS-Analyse-Verfahren oder der verbreiteten FIDA-Methode beruhen. Grafik übernommen von A. Giese und modifiziert.

in Abschnitt 2.2.1 beschriebenen PrP^C-PrP^{Sc}-Aggregations-Assays wurde die zweifarbige Variante, die 2D-SIFT-Methode, weiterentwickelt und eingesetzt.

Abbildung 2.4 zeigt schematisch den Aufbau des Messgerätes „Insight Reader“ [105], mit dem die zwei-Farben 2D-SIFT-Messungen der anti-Prion HTS-Kampagne in TP1 durchgeführt wurden. Der Insight Reader prozessiert automatisiert Mikrotiterplatten mit 96 oder 384 einzelnen Töpfchen, in denen sich jeweils kleine Portionen (20 μ l) der in Abschnitt 2.2.1 charakterisierten Assay-Mischung, eventuell kombiniert mit zusätzlichen Substanzen, befinden. Ein Laser, der rotes Licht der Wellenlänge 633 nm ausstrahlt, und ein grüner Laser (488 nm) sind gemeinsam auf ein kleines Volumenelement (\sim 0,5 fl) innerhalb der Probe fokussiert. Von den rot und grün fluoreszierenden Farbstoffen, die sich durch dieses kleine Volumenelement bewegen, werden rote und grüne Lichtquanten emittiert, deren Wellenlängen gegenüber denjenigen der Anregungslaser vergrößert, also rotverschoben, sind. Dadurch ist es möglich, das emittierte Licht aus der „Scanning unit“

über eine Anordnung dichroitischer Spiegel, die das anregende Licht reflektieren und das emittierte Licht durchlassen, sowie über eine weitere optische Anordnung zu zwei getrennten Einzel-Photon-Detektoren zu leiten. Diese optische Anordnung umfasst weitere dichroitische Spiegel, welche die grünen und roten Fluoreszenzphotonen nach der Farbe getrennt in die Detektoren leiten.

Die von den beiden Einzel-Photon-Detektoren während der gesamten Messzeit aufgefangenen Fluoreszenzphotonen werden innerhalb kurzer zeitlicher Abschnitte, sogenannter „Bins“, gezählt. Beim DIVERSet-Screening wurde eine Gesamtmesszeit von fünfmal 15 s also 75 s für jedes Töpfchen gewählt^a, und die Länge der Bins betrug 40 μ s. Aufgrund dieser kurzen Zeitskala ist es nach der Poisson-Statistik sehr unwahrscheinlich, dass Fluoreszenzsignale von mehr als einem Partikel erzeugt werden. Jedes der resultierenden $1,875 \times 10^6$ zeitlichen Bins definiert ein rot/grün-Ereignis, d.h. einen zweidimensionalen Vektor, dessen Einträge durch die Anzahlen der roten bzw. grünen Photonen gegeben sind, die innerhalb des zeitlichen Bins registriert wurden. Jedes rot/grün-Ereignis definiert also einen Punkt im positiven Quadranten einer Ebene. Trägt man diese rot/grün-Ereignisse in ein Diagramm ein, so ergibt sich eine zweidimensionale Fluoreszenz-Intensitäts-Verteilung (2D-FID), welche die Größen und Zusammensetzungen der während der Messung registrierten fluoreszierenden Partikel kodiert. Diese Messdaten sollen nun genauer diskutiert werden.

2.2.3 2D-SIFT-Daten für den Prion-Aggregations-Assay

Die Abbildungen 2.5C und D zeigen 2D-FID Intensitätsverteilungen, die mit der SIFT-Methode gewonnen wurden. Dabei unterscheidet sich das Ergebnis aus Abbildung 2.5D von jenem aus Abbildung 2.5C dadurch, dass dem zu 2.5D gehörenden Modellsystem zusätzlich eine Substanz zugegeben wurde, welche die Anlagerung von PrP^C-Molekülen an PrP^{Sc}-Aggregate hemmt. Um eine Interpretation der beiden 2D-FIDs zu ermöglichen, betrachten wir zunächst die Abbildungen 2.5A und B. Dort sind Prion-Aggregate schematisch dargestellt, die für die beiden Fälle typisch sind und sich im Laserfokus (weißer Kreis) der SIFT-Apparatur befinden sollen. Das in Abbildung A dargestellte Aggregat besteht aus rot markierten PrP^{Sc}-Molekülen, an die grün markierte PrP^C-Moleküle angelagert sind. Dem in Abbildung B dargestellten Aggregat hingegen fehlen weitgehend die angelagerten grünen PrP^C-Moleküle, weil zugefügte Wirkstoffmoleküle (blaue Kugeln) an die PrP^C-Monomere gebunden sind und so die Anlagerung verhindern.

Die in den Abbildungen 2.5C und D gezeigten 2D-FID Intensitätsverteilungen, die für Assay-Mischungen ohne bzw. mit Wirkstoffmolekülen aufgenommenen worden sind, re-

^aInsight Reader kann so eingestellt werden, dass jedes Töpfchen mehrfach gemessen wird, wobei dann die Scanning-Einheit mehrmals Mäander-förmig über die Messplatte bewegt wird. Dadurch soll erreicht werden, dass Veränderungen der Assay-Mischung, die während der Messung stattfinden können, einen weniger ungleichmäßigen Einfluss auf die Messergebnisse haben. Beim Screening der DIVERSet-Bibliotheken wurden zumeist fünf Durchläufe gewählt.

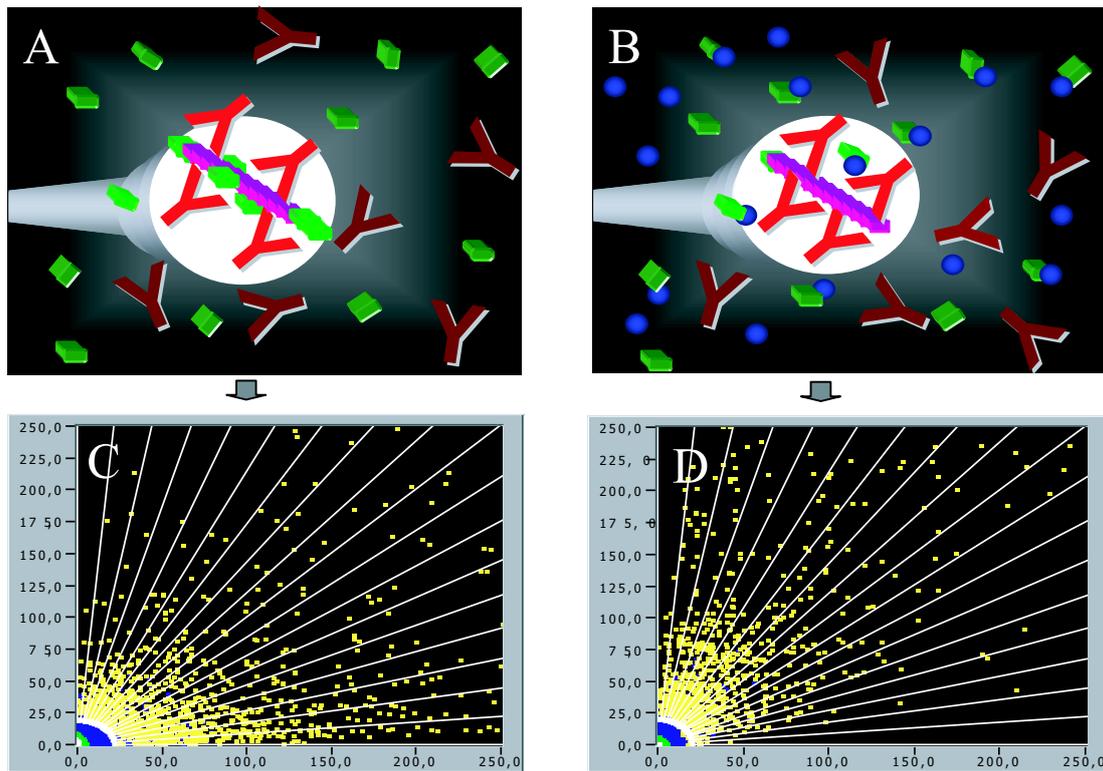


Abbildung 2.5: Assay-Mischungen und 2D-FID Verteilungen. Die schematischen Darstellungen der Mischungen ohne (A) bzw. mit zugegebenem Wirkstoff (B) verwenden die Symbole aus Abbildung 2.3. In B sind Wirkstoffmoleküle durch blaue Kugeln symbolisiert. Die 2D-FID Verteilungen wurden für Mischungen ohne (C) bzw. mit der Wirkstoffsubstanz DOSPA (D) aufgenommen (zu DOSPA siehe Abschnitt 2.3). In den 2D-FIDs ist grüne Fluoreszenzintensität nach rechts und rote Fluoreszenzintensität nach oben aufgetragen, wobei die Einheit jeweils Photonen/Bin ist. Die Häufigkeiten der Ereignisse mit identischen roten und grünen Intensitäten sind zusätzlich farblich kodiert, wobei die Skala für zunehmende Ereigniszahlen von gelb für einzelne Ereignisse über blau und grün bis weiß reicht. Übernommen aus [91].

sultieren aus der in Abschnitt 2.2.2 beschriebenen zeitlichen Diskretisierung der gemessenen Fluoreszenzintensitäten. Die in den Diagrammen aufgetragenen Punkte repräsentieren rot/grün-Ereignisse, die von Prion-Aggregaten, von einzelnen Antikörpern oder von PrP^C-Monomeren hervorgerufen wurden, als diese den Laserfokus passierten. Einzelne Farbstoffe erzeugen dabei rot/grün-Ereignisse niedriger Intensität, die in den 2D-FIDs als Punkte in den dichten Punktwolken in der Nähe des Ursprungs erscheinen. Hingegen werden durch große Aggregate hervorgerufene rot/grün-Ereignisse hoher Intensität auf vom Ursprung weiter entfernt liegende Punkte in den 2D-FIDs abgebildet, die insgesamt seltener sind.

Genau in diesen Teilen der Verteilungen unterscheidet sich nun die für eine Mischung ohne Wirkstoffsubstanz erhaltene 2D-FID aus Abbildung 2.5C von derjenigen für eine Mischung mit einem aggregationshemmenden Wirkstoff (2.5D). Die Anwesenheit des

Aggregationshemmers verschiebt die Ereignisse hoher Intensität weg von den „grünen Sektoren“ unten rechts hin zu den „roten Sektoren“ oben links, da in diesem Fall weniger grün markiertes PrP^C an die rot markierten PrP^{Sc}-Aggregate gebunden ist. Diese visuelle Analyse der 2D-FID Verteilungen zeigt bereits, dass es mit der 2D-SIFT-Technik grundsätzlich möglich ist, die aggregationshemmende Wirkung zugesetzter Substanzen anhand der erhaltenen 2D-FID Verteilungen zu messen.

Um quantitative Analysen zu ermöglichen, wertet die 2D-SIFT-Analyse (Bertsch et al. [91] und [105]) eine gegebene 2D-FID Verteilung weiter aus, indem sie die hochintensiven rot/grün-Ereignisse innerhalb der in den Diagrammen eingezeichneten „Winkel-Sektoren“ zählt, wobei Bereiche geringer Intensität, d.h. Ereignisse in der Nähe des Ursprungs, nicht mitgezählt werden. Es wird also ein „kleine-Partikel-Cutoff“ verwendet. Durch das sektorenweise Zählen intensiver Fluoreszenzereignisse entsteht ein Histogramm, das wir *2D-SIFT-Spektrum* nennen.

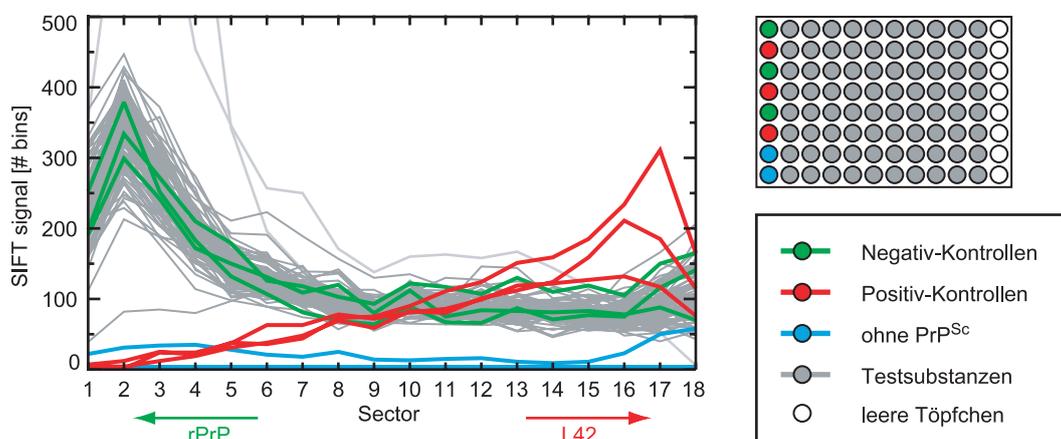


Abbildung 2.6: 2D-SIFT-Spektren für eine DIVERSet-Platte. Die für eine DIVERSet-Platte aufgenommenen 2D-SIFT-Spektren sind gemäß dem Layout der Messplatte farblich kodiert. Auf der Messplatte wurden in 80 Töpfchen Mischungen des Prion-Aggregations-Assays gemessen, in denen sich jeweils unterschiedliche Substanzen einer DIVERSet-Platte befanden. Die Symbole für diese Töpfchen sowie die zugehörigen Spektren sind grau gefärbt. Ferner wurden zur Kontrolle in jeweils drei Töpfchen Mischungen ohne Testsubstanz (grün) bzw. mit der Kontrollsubstanz DOSPA (rot) gemessen. Zwei weitere Töpfchen (cyan) enthielten Mischungen von Antikörpern und monomerem PrP^C und dienten dazu, das Ausbleiben der Aggregation dieser Assay-Bestandteile zu kontrollieren. Zwei der Spektren (hellgrau) wurden von der weiter unten (Kapitel 3) beschriebenen automatisierten Auswertung als Ausreißer erkannt. (Nach Fig. 2A aus Bertsch et al. [91].)

Abbildung 2.6 zeigt 2D-SIFT-Spektren, die bei der Messung einer DIVERSet-Platte erhalten wurden. Wie dem rechts abgebildeten Layout der Messplatte zu entnehmen ist, wurden dabei 80 Testsubstanzen im Aggregations-Assay sowie verschiedene Kontrollmischungen gemessen. Für jedes Töpfchen wurde dabei eine 2D-FID Verteilung aufgenommen und mit der 2D-SIFT-Analyse ausgewertet. Die drei grün gefärbten Spektren, die für Assay-

Mischungen erhalten wurden, denen keine Substanz zugegeben wurde, weisen einen deutlichen Peak im Bereich der Sektoren 1-5 auf. Dies zeigt, dass in diesen Mischungen viele Aggregate vorhanden sind, an die sich grün markiertes PrP^C in großen Mengen angelagert hat. Die drei rot gefärbten Spektren für Mischungen mit der zugegebenen anti-aggregatorischen Kontrollsubstanz DOSPA (vgl. Einleitung und Abschnitt 2.3) unterscheiden sich davon deutlich in ihrer Form. Der Peak für rot-grüne Aggregate in den Sektoren 1-5 fehlt. Hinzugekommen sind erhöhte Werte in den Sektoren 11-18, die anzeigen, dass in der Mischung viele ausschließlich rot markierte PrP^{Sc}-Aggregate enthalten sind, an die wenig grün markiertes PrP^C gebunden ist, weil der Kontrollwirkstoff DOSPA dies verhindern konnte. Die meisten der grau dargestellten Spektren für Mischungen mit Testsubstanzen weisen ähnliche Formen auf wie diejenigen der grünen Negativ-Kontrollen ohne Substanz. Die zugehörigen Testsubstanzen konnten die Aggregation nicht oder nur schwach hemmen oder haben die Aggregation sogar leicht gefördert. Jedoch findet sich eine getestete Substanz, deren Spektrum deutlich reduzierte Werte in den Sektoren 1-5 zeigt. Diese Substanz hemmt also die Prion-Aggregation in unserem *in vitro* System und kann damit als potenzieller anti-Prion-Wirkstoff identifiziert werden.

Neben den Spektren werden bei einer 2D-SIFT-Messung als zusätzliche Observablen die während der gesamten Messzeit aufgenommenen roten und grünen Fluoreszenzintensitäten $I_{\text{tot,r}}$ bzw. $I_{\text{tot,g}}$ ausgegeben (diese Größen werden in kHz gemessen). Im Gegensatz zu den Spektren enthalten diese beiden Größen auch die häufigen Fluoreszenzereignisse niedriger Intensitäten und können Hinweise auf von Testsubstanzen verursachte Fluoreszenzartefakte geben. Testsubstanzen, die Autofluoreszenz zeigen oder Testsubstanzen, die sogenanntes Quenching, d.h. eine Reduktion der Fluoreszenzeffizienz der markierenden Farbstoffe, zeigen, zeichnen sich durch Werte der Größen $I_{\text{tot,r}}$ oder $I_{\text{tot,g}}$ aus, die stark von denjenigen der Negativ-Kontrollmessungen oder der unkritischen Testsubstanzen abweichen. Weil die Spektren solcher Testsubstanzen aufgrund der Artefakte nicht oder nur schlecht interpretiert werden können, müssen derartige Messungen als Ausreißer erkannt und markiert werden. Als Vorgriff auf die Resultate der von mir entwickelten automatisierten Überprüfung der Messdaten sind in Abbildung 2.6 zwei der Spektren hellgrau gefärbt, weil die gemessenen Testsubstanzen Fluoreszenzartefakte zeigten und als Ausreißer erkannt wurden.

Plattenvariabilität der 2D-SIFT-Spektren

Die mit der SIFT-Technik vermessenen Assay-Mischungen werden plattenweise präpariert. Dabei werden im Allgemeinen für verschiedene Platten unterschiedliche Hirnhomogenate verwendet, sodass sich die Größenverteilungen der darin enthaltenen PrP^{Sc}-Aggregate unterscheiden. Zudem unterliegt das Mischungsverhältnis der Farbstoffe und der Antikörper leichten Schwankungen von Platte zu Platte. Daher variieren die Formen der resultierenden Spektren für unterschiedliche Platten.

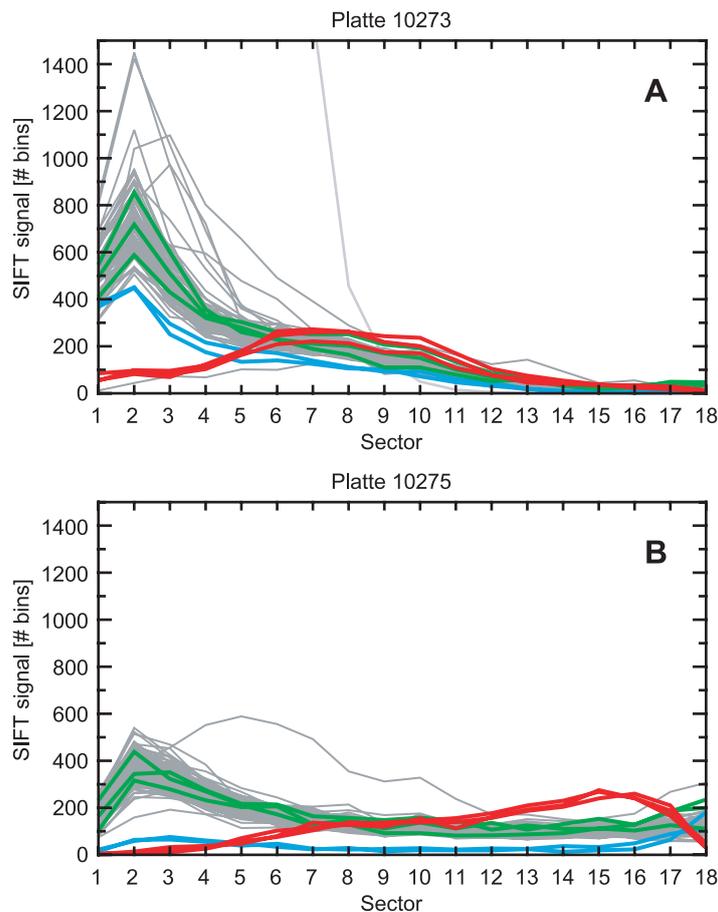


Abbildung 2.7: Plattenvariabilität von 2D-SIFT-Spektren. Die 2D-SIFT-Spektren der beiden gemessenen DIVERSet-Platten unterscheiden sich deutlich. Die Spektren sind wie in Abbildung 2.6 eingefärbt.

Um diese Variabilität der Spektren verschiedener Messplatten zu illustrieren, zeigen die Abbildungen 2.7A und B exemplarisch die Spektren für zwei weitere DIVERSet-Platten (Nummern 10273 und 10275). Wie in Abbildung 2.6 sind die Spektren gemäß der Belegung der Platte eingefärbt. Für einen vom Einfluss der Testsubstanzen unabhängigen Vergleich betrachten wir die Spektren für die Negativ- (grün) und Positiv-Kontrollmischungen (rot) der drei Platten. Die in Abbildung B dargestellten Kontrollspektren zeigen ähnliche Formen wie diejenigen in Abbildung 2.6: Die grünen Negativ-Kontrollspektren weisen den charakteristischen Peak im Sektorenbereich 1-5 auf, und die roten Positiv-Kontrollspektren zeigen erhöhte Signale im Sektorenbereich 11-18. Davon unterscheiden sich die Spektren der beiden Kontrollgruppen in Abbildung A deutlich. Während der Peak der grünen Kontrollspektren im unteren Sektorenbereich verstärkt ist, ist die Erhöhung der roten Kontrollspektren im oberen Sektorenbereich zu den mittleren Sektoren (6-10) verschoben. Derartige Unterschiede in den Spektren sind hauptsächlich auf die unterschiedlichen

Größenverteilungen der PrP^{Sc}-Aggregate zurückzuführen, die in den Assay-Mischungen verwendet wurden.

Es sei aber angemerkt, dass die Formen der Spektren zudem durch die Wahl des weiter oben besprochenen kleine-Partikel-Cutoffs beeinflusst werden, der bei der Extraktion der Spektren aus den 2D-FID-Verteilungen festgelegt wird. Der Cutoff wird für jede Platte einmal und anhand der enthaltenen Kontrollmessungen festgelegt [91], was zunächst manuell und später automatisiert von der 2D-SIFT-Software vorgenommen wurde [91, 105]. Über die Wahl des Cutoffs versucht man zum einen die Plattenvariabilität auszugleichen, was – wie gezeigt – nur in begrenztem Umfang gelingt. Zum anderen versucht man dafür zu sorgen, dass sich die positiven und negativen Kontrollspektren insbesondere im Sektorenbereich 1-5 deutlich unterscheiden, in dem die neu gewachsenen rot-grünen PrP^C-PrP^{Sc}-Aggregate beschrieben sind.

Aus den obigen Erläuterungen ist klar, dass die 2D-SIFT-Spektren die anti-aggregatorischen Eigenschaften jener DIVERSet-Substanzen kodieren, welche den PrP^C-PrP^{Sc}-Mischungen zugesetzt wurden. Für mich ergab sich daraus die Aufgabe, ein verlässliches quantitatives Maß zur Auswertung der Spektren im Hinblick auf die anti-aggregatorischen Eigenschaften der DIVERSet-Substanzen zu entwickeln. Das zu entwickelnde Wirksamkeitsmaß muss der beschriebenen Variabilität der Spektren von unterschiedlichen Platten Rechnung tragen. Angesichts der großen Zahl der DIVERSet-Substanzen und der dazu generierten Spektren musste die Extraktion des Wirksamkeitsmaßes automatisiert erfolgen, und dazu mussten die gemessenen Spektren einer computergestützten Analyse zugänglich gemacht werden. Kapitel 3 beschreibt die dazu von mir entwickelten Software-Tools und Auswerteverfahren.

In mehrfacher Hinsicht sind die Darstellungen der 2D-SIFT-Spektren in den Abbildungen 2.6 und 2.7 Vorgriffe auf die von mir geleistete automatisierte Prozessierung der SIFT-Daten. Zum einen basieren die dargestellten Spektren, wie oben erwähnt, jeweils auf mehreren Einzelmessungen, deren Ergebnisse von der automatisierten Prozessierung zusammengefasst wurden. Zum anderen ist das automatische Einfärben der Spektren gemäß dem Layout der Messplatte ein weiteres Resultat der Prozessierung. Die zugrunde liegende Zuordnung der SIFT-Spektren zu den gemessenen Testsubstanzen und Kontrollen ist ein wesentlicher Bestandteil der von mir entwickelten Software.

2.2.4 Verschiedene Einsatzformen des SIFT-Assays

Das durch die automatisierte Auswertung ermöglichte *primäre* Screening der DIVERSet-Bibliotheken im SIFT-Assay stellte den Ausgangspunkt für die Entwicklung von anti-Prion-Medikamenten im Verbund dar. Die im primären Screening als wirksam gefundenen Substanzen mussten anschließend ebenfalls im SIFT-Assay in Verdünnungsreihen genauer charakterisiert werden. Solche Verdünnungsreihenmessungen liefern sogenannte EC50-Werte, die angeben bei welcher Konzentration eine gegebene Substanz 50% ihrer ma-

ximalen Wirksamkeit erreicht. Für Medikamente werden üblicherweise EC50-Werte im nanomolaren Bereich gefordert.

SIFT-Messung der Aggregation von α -Synuklein

Die SIFT-Technik kann auch dazu eingesetzt werden, die Wirksamkeit von Testsubstanzen gegen die Aggregation eines weiteren wichtigen Proteins, des α -Synukleins (α S), zu untersuchen. Die Aggregation von α S tritt bei den sogenannten Synukleinopathien wie der Parkinsonschen Krankheit auf (für einen Übersichtsartikel siehe Goedert [106]). Wie in Kapitel 1 angesprochen wurde, bilden sich auch hier, ähnlich wie bei den Prionkrankheiten, Aggregate, die hauptsächlich aus Proteinen bestehen und denen zelltoxische Eigenschaften zugeschrieben werden.

Im Gegensatz zu PrP^C ist α S in wässriger Lösung unstrukturiert, nimmt jedoch α -helikale Strukturen an, wenn es an apolare Oberflächen, wie etwa an biologische Membranen oder Mizellen, angelagert ist [107]. *In vitro* kann die von der Umgebung abhängige Faltung von α S durch organische Lösungsmittel [z.B. einfache Alkohole wie Ethanol oder fluoridierte Alkohole wie Trifluoroethanol (TFE)] gesteuert werden [108]. Dabei zeigt sich, dass hohe Konzentrationen von fluoridierten Alkoholen die Faltung von α S in α -helikale Strukturen begünstigen, während niedrige Konzentrationen von einfachen oder fluoridierten Alkoholen die Faltung von α S in β -faltblattreiche Intermediate bewirken, die als Ausgangspunkte für die Bildung von Fibrillen fungieren.

Am ZNP wurde ein *in vitro* Modellsystem etabliert, in welchem anstelle von Alkoholen das Lösungsmittel DMSO bei niedrigen Konzentrationen (0-3%) dazu eingesetzt wird, die Ausbildung β -faltblattreicher α S-Intermediate und deren Aggregation zu induzieren. Ähnlich wie im anti-Prion SIFT-Assay werden dabei α S-Monomere mit Farbstoffen markiert. Dabei wird die Hälfte der α S-Monomere mit grünen und die andere Hälfte mit roten Farbstoffen markiert. Die markierten α S-Monomere werden gemischt und durch Zugabe von DMSO zur Bildung von α S-Multimeren angeregt. Die in solchen Mischungen gebildeten, rot-grün fluoreszierenden α S-Multimere können wiederum unter Einsatz der 2D-SIFT-Technik gemessen und die Messdaten mithilfe der 2D-SIFT-Software [105] ausgewertet werden. Durch Zugabe von Testsubstanzen zur Mischung kann deren Wirksamkeit gegen die Aggregation von α S gemessen werden.

2.3 Ein Zellkulturmodell für Prionkrankheiten (TP4)

Die Arbeitsgruppe um Prof. Jörg Tatzelt und Dr. Konstanze Winklhofer, die zu Beginn der Arbeit des Verbundes am Max-Planck-Institut (MPI) für Biochemie in Martinsried tätig war und inzwischen ins Adolf-Butenandt-Institut der LMU umgezogen ist, brachte ein etabliertes Zellkulturmodell [75, 91] für Prionkrankheiten mit in den Verbund. Dieses

Zellkulturmodell setzt Neuroblastomzellen der Linie N2a ein, die in einem Nährmedium in Petrischalen kultiviert werden und stabil mit Scrapie der Maus infiziert werden können (ScN2a) [109–111]. Die infizierten ScN2a-Zellen produzieren PrP^{Sc}, das sowohl infektiös als auch unlöslich und resistent gegenüber der Verdauung mit Proteinase K ist. Daher gibt das Modellsystem der ScN2a-Zellen zuverlässig die *in vivo*-Kennzeichen der Prionkrankheiten wieder, die bereits in Zusammenhang mit dem in Abbildung 1.2 dargestellten Western-Blot diskutiert wurden.

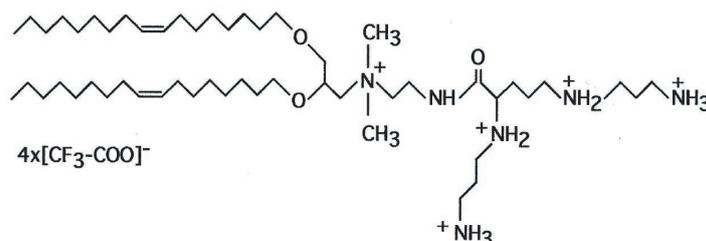


Abbildung 2.8: Der anti-PrP^{Sc}-Wirkstoff DOSPA. Das polykationische Lipid DOSPA verhindert zuverlässig die Anreicherung von PrP^{Sc} in ScN2a-Zellen [75]. DOSPA wurde deshalb für die Zellkulturtests von DIVERSet-Substanzen als Positiv-Kontrolle mitgemessen. Weil DOSPA im molekularen SIFT-Assay die Aggregation von PrP^C und PrP^{Sc} verlässlich hemmt (vgl. Abbildung 2.5), wurde das Lipid auch im SIFT-Screening als Kontrollsubstanz mitgeführt. Übernommen aus [75].

Bereits vor Beginn des Verbundprojektes konnten Winklhofer und Tatzelt durch den Einsatz des ScN2a-Modells eine neue Klasse von anti-Prion-Substanzen, die polykationischen Lipopolyamine, identifizieren [75]. Diese Substanzen beeinflussen die Neubildung von PrP^{Sc} und induzieren den Abbau von PrP^{Sc}-Aggregaten in ScN2a-Zellen. Die effektivste der untersuchten Verbindungen aus dieser Klasse ist das Lipid DOSPA, dessen chemische Struktur in Abbildung 2.8 dargestellt ist. DOSPA erwies sich auch im SIFT-anti-Prion-Assay als wirksam (vgl. Abbildung 2.5D). Weil DOSPA eine starke und verlässliche Abnahme der Bindung von PrP^C an PrP^{Sc}-Aggregate bei niedrigen mikromolaren Konzentrationen induziert, wurde diese Verbindung als Referenz-Substanz für die SIFT-Screening-Kampagne eingesetzt [91].

Im Verbundprojekt wurde das ScN2a-Modell zur *in vivo* Validierung von Substanzen eingesetzt, die im *in vitro* SIFT-Screening als wirksam gefunden worden waren. Die infizierten Zellen wurden dazu zwei Tage mit einer gegebenen Testsubstanz behandelt. Anschließend wurden die Zellen mit einem Puffer gewaschen, von der Schale gekratzt und zentrifugiert, und der entstandene Niederschlag wurde in einem weiteren Puffer gelöst. Das dadurch erzeugte Lysat wurde erneut zentrifugiert und dabei in zwei Fraktionen, den löslichen Überstand (S, Supernatant) und den unlöslichen Niederschlag (P, Pellet) aufgespalten. Dadurch wurden die enthaltenen PrP-Moleküle nach ihrer Isoform sortiert, wobei das lösliche PrP^C hauptsächlich in der S-Fraktion vorliegt und das unlösliche PrP^{Sc} hauptsächlich in der P-Fraktion gefunden wird. Anschließend wurden Western-Blots für die beiden Fraktionen erstellt, wobei die jeweils enthaltenen Moleküle elektrophoretisch nach

der Größe aufgespalten, mittels „Electroblotting“ auf Membranen übertragen und PrP^C und PrP^{Sc} jeweils über ein anti-PrP Antiserum sichtbar gemacht wurden. Für quantitative Bestimmungen des von den Zellen gebildeten PrP^{Sc} wurde in manchen Experimenten das Lysat vor der Western-Blot-Analyse zusätzlich mit Proteinase K verdaut.

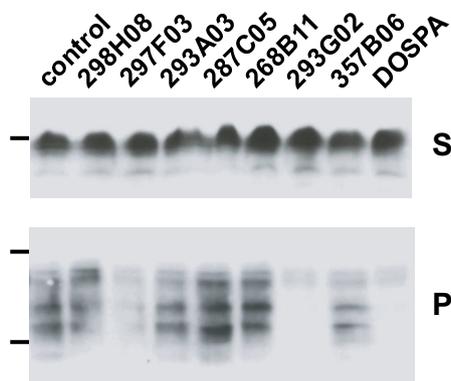


Abbildung 2.9: Inhibition der PrP^{Sc}-Bildung in ScN2a-Zellen. ScN2a-Zellen wurden zwei Tage mit verschiedenen DIVERSet-Substanzen behandelt und anschließend biochemisch analysiert. Die resultierenden Western-Blots zeigen das PrP^C in der löslichen S-Fraktion und das PrP^{Sc} in der unlöslichen P-Fraktion. Unbehandelte Zellen (control) weisen sowohl lösliches PrP^C als auch unlösliches PrP^{Sc} auf. Die Behandlung der Zellen mit dem bekannten Wirkstoff DOSPA diente als Positiv-Kontrolle und führt zum Abbau von PrP^{Sc} in den behandelten Zellen. Übernommen aus [91] (Fig. 3A).

Abbildung 2.9 zeigt Western-Blots der S- und P-Fractionen von Lysaten der ScN2a-Zellkulturen, die mit einer Reihe von DIVERSet-Substanzen und zur Kontrolle mit DOSPA behandelt wurden. Unbehandelte ScN2a-Zellen (control) produzieren sowohl lösliches PrP^C (S-Fraktion) als auch unlösliches PrP^{Sc} (P-Fraktion). Die Behandlung mit dem Kontrollwirkstoff DOSPA befreit die Zellen von PrP^{Sc} (P-Fraktion). Die schwarzen Banden in den S-Fractionen zeigen, dass keine der getesteten Substanzen die Produktion von PrP^C beeinflusst. Die Testsubstanzen 297F03 und 293G02 inhibieren die Bildung von PrP^{Sc}. Mit dem ScN2a-Modell konnte so gezeigt werden, dass eine Reihe jener Substanzen, die *in vitro* durch ihre anti-aggregatorischen Eigenschaften aufgefallen waren, in lebenden Zellen die Vermehrung von PrP^{Sc} inhibieren können. Diese Informationen waren für meine Analyse der Struktur motive von anti-Prion-Wirkstoffen von großer Bedeutung und wurden bei meinen rechnergestützten Analysen dementsprechend berücksichtigt (vgl. Abschnitt 5.1).

Es sei angemerkt, dass das N2a-Zellkulturmodell von Konstanze Winklhofer und weiteren Mitarbeitern der Gruppe Tatzelt auch dazu eingesetzt wurde, den Einfluss von diversen Mutationen des Prion-Protein-Gens Prn-p auf die Faltung und Maturierung von PrP zu untersuchen [112]. Dabei wurden die kultivierten Zellen transient mit veränderten Prn-p Genen transfiziert, sodass sie vorübergehend mutierte Prion-Proteine exprimierten. Im Gegensatz zu nicht mutiertem wildtypischen PrP (wtPrP), das posttranslational mit komple-

nen Zuckerseitenketten und einem GPI-Membrananker versehen wird (vgl. Abschnitt 1.6), wurden einige der untersuchten PrP-Mutanten fehlerhaft prozessiert. Die Fehler bei der Maturierung deuteten auf strukturelle Veränderungen der betroffenen PrP-Mutanten gegenüber wtPrP hin.

Ich habe daraufhin ein Projekt initiiert, in dem zwei der Mutanten und wtPrP mit Molekulardynamik-Simulationen untersucht wurden. Die Simulationen bestätigten den Verdacht, dass die Mutanten eine gegenüber wtPrP veränderte Struktur aufweisen und ergaben zudem Hinweise darauf, dass die Stabilität von Helix 1 durch die Dielektrizitätskonstante der Umgebung gesteuert wird [113]. In wässriger Umgebung ist Helix 1 instabil. Sobald sich diese Helix jedoch an den hydrophoben Kern des Proteins anlagert, kann sie wegen der dort erniedrigten Dielektrizitätskonstante in eine stabile helikale Struktur falten [113]. Anhang D berichtet von diesem Beitrag zu der Grundlagenforschung auf dem Gebiet der Prionkrankheiten, den ich zusammen mit Martina Stork, Bernhard Schropp und Paul Tavan abseits des zentralen Screening-Projektes leisten konnte.

2.4 HTS-fähige Zellkulturen und ein weiteres Modellsystem (TP3)

Prof. Martin Groschup und das auf der Ostseeinsel Riems gelegene Bundesforschungsinstitut für Tiergesundheit, an dem Groschup tätig ist, sind einer breiteren Öffentlichkeit durch die BSE-Pathogenesestudien an Rindern bekannt, die dort durchgeführt werden. Im von Groschup geleiteten „Institut für Neue und Neuartige Tierseuchenerreger“ ist außerdem eine Reihe von Modellsystemen für Prionkrankheiten verfügbar, die im Verbund für die Suche nach Wirkstoffen eingesetzt wurden.

Das für die Wirkstoffsuche bedeutendste auf der Insel Riems eingesetzte System ist ein Zellkulturmodell, mit welchem Testsubstanzen aus den DIVERSet-Bibliotheken in hohem Durchsatz plattenweise prozessiert und nach Wirkstoffen durchsucht werden können. Dieses HTS-fähige Zellkulturmodell unterscheidet sich von jenem aus der Gruppe Tatzelt vor allem hinsichtlich der Kultivierung der Zellen und der anschließenden Analyse der Zell-Lysate. Die Kulturen von Scrapie-infizierten Zellen der hier eingesetzten Linien SMB (scrapie mouse brain) und ScN2a werden nicht in Petrischalen sondern in den Töpfchen von Mikrotiterplatten gehalten. Wegen dieser kompakten Form der Kultivierung können alle Testsubstanzen einer Bibliotheksplatte eins-zu-eins auf eine Zellkulturplatte übertragen und so den einzelnen Zellkulturen zugeführt werden. Nach einer dreitägigen Behandlung der infizierten Zellen werden Lysate der einzelnen Töpfchen in vereinfachter Weise hinsichtlich des von den Zellen produzierten PrP^{Sc} analysiert. Im Gegensatz zum herkömmlichen Western-Blot-Verfahren werden hier die Lysate nicht elektrophoretisch aufgespalten, sondern vollständig als sogenannte „Dots“ von einer gegebenen Zellkulturplatte auf eine Membran übertragen. Auf der Membran wird das produzierte PrP^{Sc} über spezifische Antikörper sichtbar gemacht und kann optisch quantifiziert werden.

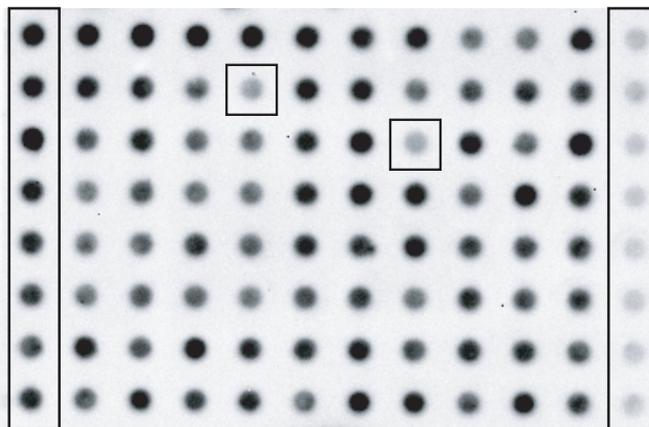


Abbildung 2.10: Ergebnisse des Zellkultur-Assays für eine DIVERSet-Platte. Analysiert wurden acht negative Kontrollen mit der DMSO-Trägerlösung (erste Spalte) und acht positive Kontrollen mit dem Wirkstoff Suramin [77] (letzte Spalte), sowie 80 Testsubstanzen (mittlere Spalten). Jeder Dot stammt von einem Töpfchen mit kultivierten Zellen, die für drei Tage mit DMSO, Suramin oder einer Testsubstanz behandelt wurden. Die Zell-Lysate wurden auf die abgebildete Dot-Blot-Membran übertragen, mit PK verdaut, und die darin enthaltenen PrP^{Sc}-Moleküle wurden über Antikörper angefärbt. Proben, die PrP^{Sc} enthalten, erscheinen als dunkle Dots, sodass wirksame Testsubstanzen durch helle Dots (Quadrate) charakterisiert sind. Mit freundlicher Genehmigung von M. Eiden.

Abbildung 2.10 zeigt exemplarisch das mit diesem Assay für eine DIVERSet-Platte erzielte Ergebnis. Negative Kontrollmessungen mit der DMSO-Trägerlösung sind durch dunkle Dots charakterisiert (erste Spalte), wohingegen mit dem Wirkstoff Suramin [77] behandelte Zellen helle Dots hervorrufen (letzte Spalte). Einzelne helle Dots im mittleren Bereich der Platte markieren Testsubstanzen, welche, wie Suramin, die Anreicherung von PrP^{Sc} in den Zellkulturen wirksam hemmen können.

Mit diesem Zellkultur-Assay wurde ein Großteil der DIVERSet-Bibliotheken nach Inhibitoren durchgemustert. Die Wirksamkeit der dabei gefundenen Substanzen wurde anschließend anhand von Verdünnungsreihen überprüft und genauer charakterisiert. Zur computer-gestützten Auswertung mussten auch die Ergebnisse des primären Zellkultur-Screenings und der Verdünnungsreihen, genau wie die Ergebnisse aus den übrigen Teilprojekten, in eine zentrale Datenbasis aufgenommen werden, um mir eine Suche nach chemischen Kernmotiven potenzieller Wirkstoffe zu ermöglichen (siehe Kapitel 4).

Ein zellfreier PrP^C-PrP^{res}-Konversions-Assay

Neben dem im zentralen Zellkultur-Screening eingesetzten Modellsystem wurde in TP3 ein weiterer Assay eingerichtet. Dieser von Dr. Martin Eiden entwickelte zellfreie Konversions-Assay [114] ist dem mit der SIFT-Technik untersuchten Assay aus TP1 grundsätzlich ähnlich. Auch der Konversions-Assay ist ein *in vitro* Modell, in dem PrP^{Sc} und

rekombinantes PrP^C zusammen gegeben werden. Das PrP^{Sc} induziert eine Umwandlung von PrP^C in eine teilweise PK-resistente Isoform, PrP^{res}. Neu gebildetes PrP^{res} kann über spezifische Antikörper, die nur an PrP^{res} binden, detektiert und quantifiziert werden. Auch diesem Assay können Testsubstanzen zugegeben werden, um deren Einfluss auf die Umwandlung von PrP^C in PrP^{res} zu messen.

Mit dem Konversions-Assay wurde eine Auswahl von DIVERSet-Substanzen, die im SIFT-Screening Wirksamkeit gezeigt hatten, getestet. Für unser Teilprojekt ergab sich daraus die Aufgabe, anhand der experimentellen Ergebnisse aus dem SIFT-Screening und dem Konversions-Assay zu überprüfen, inwiefern die Ergebnisse der beiden Assays einander bedingen, d.h. ob Korrelationen zwischen den Ergebnissen vorliegen. In Abschnitt 5.1.1 wird von diesen Untersuchungen berichtet.

2.5 Therapeutische Behandlung von Scrapie-infizierten Mäusen (TP1)

Zur Überprüfung der therapeutischen Wirksamkeit einzelner ausgewählter Substanzen, die in den vorangegangenen Untersuchungen hohe Wirksamkeiten zeigten, bietet sich das „Mausmodell“ an. Am ZNP werden dabei Mäuse zunächst mit Scrapie infiziert, indem ihnen intrazerebral Erreger injiziert werden, und später mit einzelnen potenziellen Wirkstoffsubstanzen behandelt.

Prionkrankheiten führen üblicherweise erst lange nach der Infektion eines Organismus mit dem Erreger zu klinischen Symptomen und werden daher erst zu diesem Zeitpunkt diagnostiziert. Um die Anwendbarkeit der zu testenden Substanzen in einem für Prionkrankheiten realistischen Szenario zu überprüfen, werden die infizierten Mäuse in einem der gewählten Versuchsabläufe erst mehr als 100 Tage nach der Infektion, also in einem sehr späten Stadium der Inkubationszeit, mit Testsubstanzen behandelt. Dieser Zeitpunkt wird gewählt, weil die Mäuse bei dem gewählten Erregerstamm erst dann subklinische Symptome zeigen. Daher entspricht dies dem frühesten Zeitpunkt, zu dem ein an TSE erkrankter Mensch realistischerweise behandelt werden könnte.

Die zu applizierenden Testsubstanzen sind in DMSO gelöst und werden innerhalb eines festgelegten Zeitraums bei vorab gewählten Konzentrationen verabreicht. Häufig werden die Wirkstoffe intraperitoneal, d.h. in die Bauchhöhle, injiziert, da diese Form der Verabreichung die Tiere schont. Intraperitoneal verabreichte Stoffe werden üblicherweise gut resorbiert, also aus der Bauchhöhle in die Blutbahn aufgenommen. Um aussagekräftige Ergebnisse zu erzielen, wird jeweils eine Gruppe von Mäusen mit einer gegebenen Testsubstanz und zur Kontrolle eine weitere Gruppe mit der DMSO-Trägerlösung behandelt.

Ab dem achtzigsten Tag nach der Infektion werden die Tiere täglich hinsichtlich Anzeichen der Krankheit beobachtet. Mäuse, die das Endstadium der Krankheit erreichen, welches durch das Auftreten von bestimmten klinischen Symptomen festgestellt wird, werden

getötet. Den getöteten Tieren können verschiedene Gewebeproben entnommen werden, die für biochemische und immunohistologische Analysen aufbewahrt werden. So können beispielsweise Gewebeschnitte aus der Milz präpariert und die darin enthaltenen PrP^{Sc}-Ablagerungen über Antikörper eingefärbt werden.

2.6 Medizinal- und kombinatorische Chemie (TP6)

Für die seit 2005 laufende zweite Förderperiode des Verbundprojektes konnte der Medizinalchemiker Prof. Boris Schmidt als Partner gewonnen werden. Schmidt ist derzeit im Bereich Organische Chemie der TU Darmstadt tätig. Er bringt seine mehrjährige Erfahrung aus der industriellen Forschung über das Zentrale Nervensystem und die Alzheimerische Krankheit bei der Novartis Pharma AG mit. Schmidt wurde gegen Ende der ersten Förderperiode des Verbundes zum Projekt hinzugezogen, als von uns bereits eine gewisse Substanzklasse von anti-Prion-Wirkstoffen *in vitro* (TP1) und in Zellkultur (TP4) bestimmt worden war [91] und aus den Ergebnissen des SIFT- (TP1) und des Zellkultur-Screenings (TP3) weitere potenzielle Leitstrukturen identifiziert worden waren.

Seine medizinalchemische Erfahrung erlaubt es Schmidt, eine gegebene Leitstruktur mit den Mitteln der kombinatorischen Chemie hinsichtlich der pharmakologischen Eigenschaften zu variieren. Dabei werden diverse Substituenten an verschiedene Stellen eines Kernmotivs (Leitstruktur) angebracht und beliebig kombiniert. Die Auswahl der Substituenten erfolgt dabei anhand einer Analyse der Substituentenmuster von Substanzen einer Klasse, die in einem vorangegangenen HTS als wirksam identifiziert wurde. Zusätzlich werden Substituenten in Betracht gezogen, welche erfahrungsgemäß Eigenschaften wie Löslichkeit, metabolische Stabilität oder Membrangängigkeit steuern können. Auf diese Weise können erfahrene pharmazeutische Chemiker, aufbauend auf einer schon durchgeführten HTS-Kampagne, große Mengen neuer potenzieller Wirkstoffe erzeugen, welche in einer fokussierten Bibliothek zusammengefasst werden. Derartige fokussierte Bibliotheken können dann in den unterschiedlichen Modellsystemen des Verbundes hinsichtlich der jeweiligen Wirksamkeiten getestet werden. Die Ergebnisse fließen anschließend in die Entwicklung einer nächsten Generation von weiter optimierten Verbindungen ein.

Die Einbindung von Prof. Schmidt in unseren Verbund stellte mich zunächst vor die Aufgabe, ihm die bisher gewonnenen Analyseergebnisse zur Verfügung zu stellen. Anschließend musste für die daraufhin von seiner Gruppe erstellten fokussierten Bibliotheken ein virtuelles Abbild geschaffen werden, um die dazu gewonnenen Messergebnisse rechnergestützt auswerten zu können (siehe Kapitel 4 und 5).

3 Methoden zur Haltung, Auswertung und Visualisierung von Daten

Die im vorangegangenen Kapitel skizzierten experimentellen Verfahren zum *in vitro* und *in vivo* Hochdurchsatz-Screening der DIVERSet-Substanzbibliotheken, zum Validieren von potenziellen Wirkstoffen und zu ihrer Optimierung durch Methoden der kombinatorischen Chemie stellten mich vor die Aufgabe, ein zur Realität der experimentellen Untersuchungen adäquates virtuelles Abbild zu erzeugen, das eine schnelle und sichere Auswertung der experimentell gewonnenen Daten erlaubt. Ich habe dazu eine Reihe von Werkzeugen programmiert, deren Struktur und Funktionsweise im Folgenden erläutert werden sollen.

Der heute erreichte Stand ergab sich in einem evolutionären Entwicklungsprozess, in dessen Verlauf erste Ansätze überarbeitet, zum Teil verworfen und durch Neuentwicklungen ersetzt wurden. Es würde zu weit führen, jede Zwischenstufe im Einzelnen zu erklären und wäre auch für spätere Nutzer der von mir programmierten Software-Tools nicht besonders nützlich. Deshalb sollen hier nur jene Werkzeuge genauer beschrieben werden, die auch künftig nutzbringende Einsatzmöglichkeiten haben werden.

Bei den von mir erzeugten Software-Tools handelt es sich um

- den SIFT-Viewer I, der einen schnellen, Webbrowser-gestützten Zugriff auf 2D-SIFT-Spektren, daraus gewonnene Wirksamkeitsmaße, chemische Strukturen etc. erlaubt,
- den in Java geschriebenen SIFT-Viewer II, der die Funktionalitäten des SIFT-Viewer I im Hinblick auf den Einsatz mustererkennender Verfahren, auf die Variabilität von Mikrotiterplatten-Layouts, auf eine Datenbankbindung etc. erweitert,
- die Oracle-basierte Datenbank TSE-DB zur Haltung möglichst aller im Verbund gewonnener Daten und insbesondere der vollständigen SIFT-Messdaten,
- die dateibasierte Haltung der TSE-DB-Daten mithilfe der Chemie-Software Chem-Finder, welche chemische Strukturen von potenziellen Wirkstoffen mit experimentell in den diversen Assays gewonnenen Wirksamkeitsmaßen kombiniert.

Aus der obigen Auflistung ist zu erkennen, dass die Zuordnung von experimentell bestimmten Wirksamkeiten zu vorgegebenen, durch ihre chemische Struktur definierten Substanzen ein wesentlicher Aspekt der Datenhaltung und -analyse ist. Während bei Zellkulturversuchen Wirksamkeiten direkt als Messergebnisse vorliegen, muss bei SIFT-Mes-

sungen ein geeignetes Wirksamkeitsmaß erst noch aus den 2D-SIFT-Spektren dekodiert werden.

Da das Hochdurchsatz-Screening mit der SIFT-Technik den Ausgangspunkt der Medikamentensuche bildete, musste also zunächst ein einfach zu handhabendes Wirksamkeitsmaß entworfen werden. Noch vor der Schilderung der oben genannten Software-Tools werde ich dieses (und ein weiteres) Wirksamkeitsmaß zur Bestimmung der anti-Prion-Wirksamkeit im SIFT-Assay erläutern, das von mir in Kooperation mit Jan Bieschke, Uwe Bertsch und Armin Giese vom ZNP entworfen wurde.

Damit das zu entwickelnde Maß den Testsubstanzen aus den SIFT-Messdaten zuverlässig Wirksamkeitswerte zuordnen kann, muss zunächst die Validität der Messdaten überprüft werden. Es muss verhindert werden, dass fehlerhafte Messungen die nachfolgenden Schritte der Auswertung negativ beeinflussen könnten. Daher müssen fragwürdige Messungen anhand der Daten erkannt, markiert und aussortiert werden. Vor der Beschreibung der Wirksamkeitsmaße sollen daher Algorithmen zur Erkennung von Ausreißern vorgestellt werden, die von mir entwickelt und in den Werkzeugen SIFT-Viewer I und II eingesetzt wurden.

3.1 Erkennung von Ausreißern

Für die Erkennung von fehlerhaften SIFT-Messungen werden die 2D-SIFT-Spektren und die zugehörigen Gesamtintensitäten $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$ herangezogen. Hier kann es aufgrund von Übertragungsfehlern bei der Messung in seltenen Fällen vorkommen, dass einzelne 2D-SIFT-Spektren exakt verschwinden. Solche Spektren können von einem entsprechenden Filter leicht erkannt und aussortiert werden.

Wie in Abschnitt 2.2.3 bereits angesprochen wurde, kann es aber auch vorkommen, dass eine Testsubstanz Autofluoreszenz zeigt und dadurch die Fluoreszenzsignale der eingesetzten Farbstoffe überlagert werden. Des Weiteren kann eine getestete Substanz durch sogenanntes *Quenching* die Fluoreszenz der eingesetzten Farbstoffe reduzieren. Derartige unerwünschte Einflüsse schlagen sich in den zugehörigen 2D-SIFT-Spektren nieder und beeinträchtigen deren Interpretation. Die Änderungen der Messdaten zeigen sich insbesondere bei den Werten der roten und grünen Gesamtintensitäten $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$. Fluoreszierende oder „quenched“ Testsubstanzen erhöhen bzw. erniedrigen die Werte von $I_{\text{tot,r}}$ oder $I_{\text{tot,g}}$ gegenüber denjenigen von unkritischen Substanzen. Weil sich die I_{tot} -Werte von verfälschenden Substanzen zumeist deutlich abheben, sind diese Messgrößen zur Erkennung von Ausreißern geeignet.

Für die Erkennung von Ausreißern anhand der Gesamtintensitäten wurden von mir zwei Verfahren entwickelt. Beiden Verfahren ist gemein, dass damit die Messdaten plattenweise prozessiert werden, und dass Verteilungen der beiden Größen $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$ für die Töpfchen einer gegebenen Messplatte analysiert werden. Zur Vereinfachung der Notation seien

die beiden Gesamtintensitäten durch $I_{\text{tot},f}$ symbolisiert, wobei $f = r, g$ als Platzhalter für rot oder grün steht.

3.1.1 Eine Median-basierte Erkennung

Das erste Verfahren legt die beiden Verteilungen der für *alle* Töpfchen einer Platte erhaltenen Gesamtintensitäten $I_{\text{tot},f}$ zugrunde und geht von der Annahme aus, dass die meisten $I_{\text{tot},f}$ -Werte eng um zentrale Werte verteilt liegen, sodass Ausreißer durch große Abstände zu diesen zentralen Werten charakterisiert sind. Als zentraler Wert der $I_{\text{tot},f}$ -Verteilung wird dabei der sogenannte Median verwendet, ein dem arithmetischen Mittel ähnliches Lagemaß, das in der deskriptiven Statistik häufig eingesetzt wird [115].

Zur Bildung des Medians der $I_{\text{tot},f}$ -Verteilung ($f = r, g$) werden die N Werte, die hier kurz I_i genannt seien, der Größe nach sortiert, sodass die geordnete Menge (I_1, I_2, \dots, I_N) vorliegt. Der Median ist dann gemäß

$$\tilde{I} = \begin{cases} I_{(N+1)/2} & N \text{ ungerade,} \\ \frac{1}{2}(I_{N/2} + I_{N/2+1}) & N \text{ gerade} \end{cases} \quad (3.1)$$

als der in der Mitte, d.h. an der Stelle $(N + 1)/2$, liegende Wert $I_{(N+1)/2}$ definiert, falls die Länge N der Stichprobe ungerade ist und andernfalls als Mittelwert der beiden der Mitte am nächsten liegenden Werte. Im Gegensatz zum arithmetischen Mittelwert ist der Median robust gegenüber Ausreißern, d.h. sein Wert verändert sich kaum oder nicht, wenn weit vom Zentrum entfernt liegende Werte zur Verteilung hinzukommen.

Die gemäß (3.1) für die beiden Verteilungen $I_{\text{tot},f}$ ($f = r, g$) gebildeten Mediane M_r und M_g werden nun herangezogen, um für jedes Töpfchen (Well) w zu überprüfen, ob die Werte der Gesamtintensitäten gemäß

$$\begin{aligned} I_{\text{tot},r}(w) < g_u M_r & \quad \text{oder} & \quad I_{\text{tot},r}(w) > g_o M_r & \quad \text{oder} \\ I_{\text{tot},g}(w) < g_u M_g & \quad \text{oder} & \quad I_{\text{tot},g}(w) > g_o M_g & \end{aligned} \quad (3.2)$$

weit entfernt vom jeweiligen Median liegen, wobei die Faktoren g_u und g_o , welche die unteren bzw. oberen Grenzwerte definieren, noch zu bestimmen sind. Ist eine der Bedingungen in (3.2) erfüllt, so wird das entsprechende Töpfchen als Ausreißer gekennzeichnet. Für die in Abschnitt 3.4 beschriebene automatisierte Auswertung des DIVERSet-Screenings mit SIFT-Viewer I wurde dieses Verfahren eingesetzt. Dabei wurden für die Faktoren die Werte $g_u = 0,1$ und $g_o = 5,0$ heuristisch gewählt und lieferten im Allgemeinen zuverlässige Resultate.

Bei einigen Platten konnte es jedoch vorkommen, dass durch diese einfache Ausreißererkennung Töpfchen mit Kontrollmischungen als Ausreißer gekennzeichnet wurden, weil die zugehörigen Werte $I_{\text{tot},f}$ gemäß (3.2) zu stark vom jeweiligen Median M_f abwichen.

Dies liegt darin begründet, dass die Mediane M_f durch die Werte der Testsubstanzen definiert sind, die im größten Teil der Töpfchen gemessen wurden. Offensichtlich konnte es also vorkommen, dass sich unterscheidende Assay-Zusammensetzungen für Kontrollen und Testsubstanzen zu deutlich unterschiedlichen $I_{\text{tot},f}$ -Werten führen. Daher war es nötig, die Erkennung dahingehend zu verbessern, dass vor einer statistischen Analyse die Menge der Töpfchen nach den Arten der darin gemessenen Mischungen gruppiert wird.

3.1.2 Eine nach dem Töpfchentyp getrennte Erkennung

Wie aus dem in Abbildung 2.6 dargestellten Layout der Messplatten beim SIFT-Screening hervorgeht, gliedern sich die Töpfchen nach der Zusammensetzung der Assay-Mischung in die vier Gruppen „Positiv-Kontrollen“, „Negativ-Kontrollen“, „Kontrollen ohne PrP^{Sc}“ und „Testsubstanzen“. Nach dem hier vorgestellten verbesserten Verfahren wird jede dieser Töpfchengruppen getrennt hinsichtlich Ausreißern untersucht. Als Ausreißerkriterium soll dabei ein gegenüber dem heuristischen Kriterium (3.2) verbessertes Kriterium eingesetzt werden, das ursprünglich von Grubbs [116] vorgeschlagen wurde. Das Kriterium von Grubbs basiert auf der Annahme, dass die Menge der $I_{\text{tot},f}$ -Werte für eine gegebene Töpfchengruppe durch eine Normalverteilung beschrieben werden kann, die durch den Mittelwert

$$\mu_f = \frac{1}{N} \sum_{w=1}^N I_{\text{tot},f}(w)$$

und die Stichproben-Standardabweichung

$$s_f = \sqrt{\frac{1}{N-1} \sum_{w=1}^N [I_{\text{tot},f}(w) - \mu_f]^2}$$

der Werte $I_{\text{tot},f}(w)$ für die N Töpfchen charakterisiert ist. Um zu entscheiden, ob ein fragwürdiger Wert $I_{\text{tot},f}(w)$ dieser Normalverteilung angehört, wird das Testkriterium

$$T = \frac{|I_{\text{tot},f}(w) - \mu_f|}{s_f} \tag{3.3}$$

verwendet, wobei überprüft wird, ob der Wert T gemäß

$$T > T_c(N, \alpha) \tag{3.4}$$

größer als ein kritischer Wert $T_c(N, \alpha)$ ist, der vom Umfang N der Stichprobe und einem wählbaren Signifikanzniveau α abhängt. In [116] findet sich eine Tabelle (Table 1) mit kritischen Werten $T_c(N, \alpha)$ für verschiedene Stichprobenumfänge N und Signifikanzniveaus α , welcher der kritische Wert $T_c(N, \alpha)$ für den gegebenen Stichprobenumfang N und das

gewählte Signifikanzniveau α entnommen werden kann. Ist T größer als $T_c(N, \alpha)$, also die Ungleichung (3.4) erfüllt, so gehört $I_{\text{tot},f}(w)$ mit der Wahrscheinlichkeit $1 - \alpha$ nicht der zugrunde gelegten Normalverteilung an und wird daher als Ausreißer angesehen.

Bei dieser Ausreißererkenntnis wird zunächst für die anfänglich N Werte der Mittelwert μ_f gebildet. Dann wird der von μ_f am weitesten entfernt liegende Wert $I_{\text{tot},f}(w)$ als fragwürdiger Wert ausgewählt, das Testkriterium (3.3) ausgewertet und der fragliche Wert als Ausreißer aus der Menge entfernt, falls er gemäß (3.4) mit der Wahrscheinlichkeit $1 - \alpha$ nicht der Normalverteilung angehört. Diese Prozedur wird für die verbleibenden Werte solange iteriert, bis kein Ausreißer mehr gefunden wird.

Das skizzierte Verfahren wurde in SIFT-Viewer II implementiert, wobei die kritischen Werte $T_c(N, \alpha)$ für $\alpha = 0,05$ und $0,01$ und für bestimmte $N \leq 100$ aus der Tabelle in [116] übernommen und fehlende Werte geeignet inter- ($N \leq 100$) bzw. extrapoliert ($N > 100$) wurden. Für die Gruppe der Testsubstanzen wurde das Signifikanzniveau $\alpha = 0,01$ und für die kleineren Gruppen von Kontrollmessungen jeweils der großzügigere Wert $\alpha = 0,05$ gewählt. Falls eine Töpfchengruppe jedoch weniger als $N = 3$ Töpfchen umfasst, kann keine Ausreißererkenntnis anhand dieses Tests vorgenommen werden.

Abschließend sei angemerkt, dass in Teilen meiner Auswertung zusätzlich zu den oben geschilderten Ausreißerdetektoren als weitere Prüfgröße die gemäß Gleichung (3.6) gebildete Summe S_1^{18} herangezogen wurde, die weiter unten eingeführt werden wird. Diese aus den 2D-SIFT-Spektren gebildete Größe kann wie die Gesamtintensitäten $I_{\text{tot},f}$ Hinweise auf von Testsubstanzen verursachte Fluoreszenzartefakte geben.

3.2 SIFT-Wirksamkeitsmaße

Wie in Abschnitt 2.2.3 erläutert wurde, liefern die SIFT-Messungen für Mikrotiterplatten mit Testsubstanzen als Ergebnis 2D-SIFT-Spektren, welche die anti-aggregatorischen Wirksamkeiten der getesteten Substanzen kodieren. Um daraus ein Wirksamkeitsmaß zu entwickeln, schreiben wir zunächst das für ein Töpfchen w erhaltene 2D-SIFT-Spektrum als $D = 18$ -dimensionalen Vektor

$$\mathbf{x}_w = (x_{w1}, \dots, x_{wD})^t \in \mathbb{N}^D, \quad (3.5)$$

dessen Einträge $x_{wi} \geq 0$ natürliche Zahlen sind. Gemäß der in Abschnitt 2.2.3 erläuterten Bedeutung der 2D-SIFT-Spektren ist ein solcher Vektor \mathbf{x}_w ein Histogramm, dessen D Einträge x_{wi} durch die Anzahlen von hochintensiven rot/grünen Fluoreszenzereignissen gegeben sind, die von Prion-Aggregaten bestimmter PrP^C-PrP^{Sc}-Zusammensetzungen hervorgerufen wurden. Wie dort erklärt wurde, kommt dabei den Einträgen im Sektorenbereich von etwa $i = 1, \dots, 5$ eine besondere Bedeutung zu: Die zugehörigen Einträge x_{wi} stammen von solchen Aggregaten, die in großen Mengen neu hinzuaggregiertes, grün-

markiertes PrP^C enthalten. Weil daher kleine Einträge in diesem Bereich die anti-aggregatorische Wirksamkeit der Testsubstanzen kodieren, soll das zu entwickelnde Wirksamkeitsmaß auf diesen Einträgen basieren.

3.2.1 Observablen aus den 2D-SIFT-Spektren

Aus den oben angeführten Gründen scheint es also zweckmäßig zu sein, die Einträge x_{wi} verschiedener Sektorenbereiche, wie beispielsweise demjenigen für $i = 1, \dots, 5$, zusammenzufassen. Es wird dazu für ein gegebenes Töpfchen w die Summe

$$S_v^b(w) = \sum_{i=v}^b x_{wi} \quad (3.6)$$

der Einträge x_{wi} in dem von Sektor v bis Sektor b reichenden Bereich gebildet. Diese Definition erfolgt allgemein für Sektorenbereiche von v bis b , weil neben der Summe $S_1^5(w)$ weitere solche bereichsweisen „Vergrößerungen“ der Spektren \mathbf{x}_w für verschiedene Auswertungen verwendet werden, wie beispielsweise S_1^{18} für die Ausreißererkennung (vgl. Abschnitt 3.1). Nach den oben gegebenen Erläuterungen kodieren insbesondere große Werte der Summe $S_1^5(w)$ diejenigen in der Mischung im Töpfchen w enthaltenen Aggregate, die durch Anlagerung von PrP^C-Molekülen gewachsen sind.

3.2.2 Ein auf die Kontrollen bezogenes Wirksamkeitsmaß

Zur Ableitung eines auf die Kontrolltöpfchen bezogenen Wirksamkeitsmaßes wird zunächst – ähnlich wie für die Ausreißererkennung aus Abschnitt 3.1.2 – die Menge \mathcal{W} aller Töpfchen einer gegebenen Platte in die Teilmenge \mathcal{P} der Positiv-Kontrollen, die Teilmenge \mathcal{N} der Negativ-Kontrollen und die Teilmenge \mathcal{S} der Töpfchen mit Testsubstanzen unterteilt. Diese Unterteilung ermöglicht es, als erste Vergleichsgröße den arithmetischen Mittelwert

$$p = \frac{1}{|\mathcal{P}|} \sum_{w \in \mathcal{P}} S_1^5(w) \quad (3.7)$$

der Summen $S_1^5(w)$ für die Menge \mathcal{P} der Positiv-Kontrollen zu definieren. Analog dazu wird als zweite Vergleichsgröße der Mittelwert der Summen $S_1^5(w)$

$$n = \frac{1}{|\mathcal{N}|} \sum_{w \in \mathcal{N}} S_1^5(w) \quad (3.8)$$

für die Menge \mathcal{N} der Negativ-Kontrollen gebildet. Es sei angemerkt, dass in den weiter unten geschilderten Anwendungen in SIFT-Viewer I und II zum Teil anstelle der arithmetischen Mittelwerte (3.7) und (3.8) für p und n die gemäß (3.1) definierten Mediane der Summen S_1^5 innerhalb der beiden Kontrollgruppen verwendet wurden.

Nun kann allgemein für die Töpfchen w einer gegebenen Platte (und insbesondere für die Töpfchen $w \in \mathcal{S}$ mit Testsubstanzen) das Aktivitätsmaß

$$a : \mathcal{W} \rightarrow \mathbb{R}, \quad w \mapsto a(w) = \frac{n - S_1^5(w)}{n - p} \quad (3.9)$$

definiert werden, welches einem gegebenen Töpfchen w eine reelle Zahl $a(w)$ zuordnet. Dabei wird die Differenz zwischen dem Mittelwert n der Negativ-Kontrollen und dem S_1^5 -Wert für Töpfchen w gebildet und mit der Differenz $n - p$ der beiden Kontroll-Mittelwerte ins Verhältnis gesetzt. Für den Fall $w \in \mathcal{S}$, d.h. dass in Töpfchen w eine Testsubstanz gemessen wurde, gibt die zugeordnete Zahl $a(w)$ die anti-aggregatorische Wirksamkeit der Substanz an.

Aus der Definition (3.9) des Aktivitätsmaßes ist leicht zu erkennen, dass den Töpfchen mit Testsubstanzen Werte $a(w) \approx 0$ zugeordnet werden, wenn die Summe $S_1^5(w)$ in der Nähe des Negativ-Mittelwerts n liegt [$S_1^5(w) \approx n$] und daher die Substanz keine oder eine schwache anti- oder pro-aggregatorische Wirkung besitzt. Hingegen werden Töpfchen mit $S_1^5(w) \approx p$, deren Substanzen die Aggregation ähnlich gut hemmen wie die Kontrollwirkstoffe, auf Aktivitätswerte $a(w) \approx 1$ abgebildet.

Durch die in (3.9) vollzogene Normierung auf den Abstand^a $n - p$ der Kontrollenmittelwerte, der von den zugehörigen Spektren abhängt und daher von Platte zu Platte variieren kann, wird also erreicht, dass unabhängig von dieser Variabilität für alle Platten unwirksamen Substanzen Aktivitätswerte $a \approx 0$ und wirksamen Substanzen Werte $a \approx 1$ zugeordnet werden. Dieses auf die plattenlokalen Kontrollmessungen bezogene SIFT-Aktivitätsmaß wurde in SIFT-Viewer I und II implementiert und für die Auswertung des Screenings der DIVERSet-Bibliotheken eingesetzt.

3.2.3 Ein auf die Testsubstanzen bezogenes Wirksamkeitsmaß

Für den Fall, dass auf einer gegebenen Messplatte keine (oder nicht auswertbare) Kontrollmessungen mitgeführt wurden, wurde ein zweites Wirksamkeitsmaß entwickelt, welches die Wirksamkeit einer gegebenen Testsubstanz relativ zu denjenigen der anderen Testsubstanzen beschreibt.

Wie dem kontrollenbezogenen Maß (3.9) werden auch diesem Maß die Summenobservablen S_1^5 zugrunde gelegt. Aufgrund des Fehlens von Kontrollmessungen wird als Vergleichsgröße der gemäß (3.1) gebildete Median \tilde{S} der Werte $S_1^5(w)$ für die Töpfchen $w \in \mathcal{S}$ mit Testsubstanzen herangezogen. Damit kann nun das auf die Testsubstanzen bezogene Wirksamkeitsmaß

$$\tilde{a} : \mathcal{S} \rightarrow \mathbb{R}, \quad w \mapsto \tilde{a}(w) := 1 - \frac{S_1^5(w)}{\tilde{S}} \quad (3.10)$$

^aDie Bezeichnung „Abstand“ ist gerechtfertigt, weil für valide SIFT-Messungen die Differenz $n - p$ positiv ist.

definiert werden, welches, wie das kontrollenbezogene Maß a , jedem gemessenen Töpfchen $w \in S$ eine reelle Zahl $\tilde{a}(w)$ zuordnet. Die Interpretation der für die Testsubstanzen einer Platte resultierenden Aktivitätswerte \tilde{a} ändert sich jedoch gegenüber derjenigen des Maßes a . Werte $\tilde{a} < 0$ beschreiben nun solche Substanzen, deren anti-aggregatorische Wirksamkeit geringer ist als die Wirksamkeit der zum Median gehörigen Substanz(en). Der anderen Hälfte der Substanzen werden Aktivitätswerte $\tilde{a} > 0$ zugeordnet.

Das Wirksamkeitsmaß (3.10) wurde in den SIFT-Viewer II implementiert. Beide Werkzeuge, SIFT-Viewer I und II, wurden für die Auswertung der für die DIVERSet-Bibliotheken gewonnenen Messdaten eingesetzt. Zur Erläuterung ihrer Funktionsweise sollen zunächst die für die DIVERSet-Bibliotheken vom Hersteller mitgelieferten elektronischen Daten beschrieben werden.

3.3 Elektronische Daten zu den DIVERSet-Bibliotheken

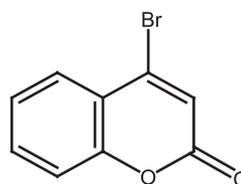
In Abschnitt 2.1 wurden die in unserem Forschungsverbund für die Medikamentensuche beschafften Substanzbibliotheken DIVERSet 1 und 2 hinsichtlich ihrer pharmakologischen Eigenschaften beschrieben. Dort wurde außerdem erklärt, dass die in den Bibliotheken enthaltenen Substanzen in DMSO gelöst und auf Mikrotiterplatten verteilt vom Hersteller ChemBridge [95] geliefert wurden. Neben den realen Substanzen war im Lieferumfang der beiden Bibliotheken jeweils eine Datenbank enthalten, in der zu jeder Substanz die zweidimensionale Molekülstruktur definiert ist sowie eine Reihe von Eigenschaften abgelegt sind. Diese Informationen zu den DIVERSet-Bibliotheken wurden von ChemBridge in Form von zwei großen Dateien im sdf-Format (structure data file, MDL [117]) mitgeliefert. Abbildung 3.1 zeigt exemplarisch einen Auszug der zu DIVERSet 1 gehörigen sdf-Datei, in welchem eine Substanz dieser Bibliothek beschrieben ist.

3.3.1 Molekülstrukturen

Im ersten Teilabschnitt der Substanzbeschreibung ist die Molekülstruktur in dem verbreiteten mol-Format [117] definiert. Dazu sind die in dem ausgewählten Molekül enthaltenen Atome, deren Typen und (hier) zweidimensionale Koordinaten angegeben. Die Bindungen zwischen den Atomen sind durch eine Konnektionstabelle definiert. Das sdf-Format erweitert das mol-Format, welches lediglich zur Definition der Struktur eines einzelnen Moleküls vorgesehen ist, zum einen dadurch, dass im sdf-Format neben der Struktur zusätzliche Eigenschaften angegeben werden können (vgl. Abbildung 3.1 ab „M END“). Zum anderen unterstützt das sdf-Format, dass derartige Substanzdefinitionen aneinander gehängt werden können.

3.3 Elektronische Daten zu den DIVERSet-Bibliotheken

```
-ISIS- 09300311332D
12 13 0 0 0 0 0 0 0 0 0999 V2000
-0.2458 1.1708 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7833 1.4708 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.2458 0.5333 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7833 0.2333 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.3208 1.1708 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.3208 0.5333 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.3167 0.2125 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7833 2.1125 0.0000 Br 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.8625 1.4708 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.8625 0.2333 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.4000 1.1708 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.4000 0.5333 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 1 2 0 0 0 0
3 1 1 0 0 0 0
4 3 1 0 0 0 0
5 2 1 0 0 0 0
6 4 1 0 0 0 0
7 3 2 0 0 0 0
8 2 1 0 0 0 0
9 5 1 0 0 0 0
10 6 1 0 0 0 0
11 9 2 0 0 0 0
12 10 2 0 0 0 0
5 6 2 0 0 0 0
11 12 1 0 0 0 0
M END
> <ID> (5100026)
5100026
> <Plate> (5100026)
10252
> <Col> (5100026)
07
> <Row> (5100026)
A
> <Supplier> (5100026)
ChemBridge
> <Coordinates> (5100026)
A07
> <clogP> (5100026)
2.270000000000000e+000
> <RB> (5100026)
0
> <tPSA> (5100026)
3.021000000000000e+001
> <Hacc> (5100026)
2
> <Hdon> (5100026)
0
```



10252_A07

Abbildung 3.1: Auszug aus der sdf-Datei zu DIVERSet 1. Der gezeigte Abschnitt aus der sdf-Datei zu DIVERSet 1 zeigt die Beschreibung einer exemplarisch ausgewählten Substanz. Im ersten Teil der Beschreibung ist die zweidimensionale Molekülstruktur definiert, welche zudem rechts grafisch dargestellt ist. Im zweiten Teil (nach der Zeile „M END“) sind zusätzliche Eigenschaften der Substanz abgelegt. Nähere Erläuterungen finden sich im Text.

3.3.2 Eindeutige Namen für die DIVERSet-Substanzen

Wie aus Abbildung 3.1 zu erkennen ist, umfassen die in den sdf-Dateien enthaltenen Substanzinformationen die Nummer (Eintrag „Plate“) der Mikrotiterplatte, auf der die Substanz geliefert wurde und die Koordinaten („Coordinates“ bzw. „Row“ und „Col“) des Töpfchens auf dieser Platte. Obschon jede Substanz vom Hersteller über eine Identifikationsnummer („ID“) bereits eindeutig bezeichnet ist, entschlossen wir uns dazu, jeder DIVERSet-Substanz einen weniger abstrakten Namen zu geben. Nach der von uns dazu vereinbarten Konvention wird jede Substanz im Format

$$\text{Substanzname} := \langle \text{Plattennummer} \rangle _ \langle \text{Töpfchenkoordinaten} \rangle \quad (3.11)$$

eindeutig benannt. Dementsprechend trägt die in Abbildung 3.1 dargestellte Substanz den Namen „10252_A07“. Durch diese Namensgebung wird die Zuordnung von Messergebnissen zu Substanzen erleichtert.

3.3.3 Physikochemische Eigenschaften

Neben den soeben beschriebenen Einträgen, über welche die virtuellen Repräsentationen den realen Substanzen zugeordnet werden können, enthalten die Substanzbeschreibungen auch eine Reihe von physikochemischen Eigenschaften. Dazu zählen das Lipophilie-Maß $c\text{LogP}$ (calculated LogP) [118], die Anzahl der drehbaren Bindungen („RB“, rotatable bonds), die sogenannte gesamte polare Oberfläche des Moleküls („tPSA“, total polar surface area) [119], sowie die Anzahlen derjenigen Atome, die als Akzeptoren („Hacc“) bzw. Donoren („Hdon“) von Wasserstoffbrückenbindungen fungieren können. Die aufgezählten Größen sind aus der Molekülstruktur zähl- oder schätzbar und wurden vom Hersteller unter Einsatz entsprechender Softwarewerkzeuge wie denjenigen der Firma Daylight [120] generiert. Die Größe tPSA gibt dabei an, wie groß derjenige Anteil einer das Molekül umgebenden Fläche ist, welcher die polaren Atome umgibt, wobei zumeist Sauerstoff, Stickstoff und eventuell weitere Atomarten als polar betrachtet werden.

Wie in Abschnitt 2.1 hervorgehoben wurde, besitzen die DIVERSet-Substanzen aufgrund der vom Hersteller getroffenen Auswahl vielversprechende pharmakologische Eigenschaften, die gemäß Lipinskis *rule of five*, auf physikochemische Eigenschaften zurückgeführt werden können [98]. Da die Bedeutung des in Lipinskis Faustregel enthaltenen Lipophilie-Maßes ($c\text{LogP}$) bislang nur angedeutet wurde, soll diese Größe hier kurz erläutert werden [121].

Experimentell wird LogP bestimmt, indem die gegebene Substanz in ein zweiphasiges System aus apolarem Oktanol und polarem Wasser gebracht wird und anschließend die Konzentrationen K_O und K_W der Substanz in Oktanol bzw. Wasser gemessen werden. Die Größe LogP ist dann definiert als Logarithmus des Verteilungskoeffizienten $P = K_O/K_W$,

wobei große Werte lipophile Substanzen markieren. Anstelle von durch aufwändige Messungen bestimmten LogP-Werten, sind in DIVERSet *berechnete* Schätzungen cLogP angegeben (siehe [118] für eine Übersicht über Methoden zur cLogP-Berechnung). Das Kriterium „LogP < 5“ aus Lipinskis *rule of five* setzt der Lipophilie eine obere Grenze. Weil aber keine untere Grenze gesetzt wird, sind nach diesem Kriterium beliebig stark polare Substanzen erlaubt. Von bekannten Medikamenten weiß man aber, dass die Polarität nicht beliebig groß sein darf. Um die Polarität zu begrenzen sind daher in DIVERSet die Werte der gesamten polaren Oberfläche tPSA angegeben und wurden wohl bei der Substanzauswahl von ChemBridge berücksichtigt, denn die in DIVERSet enthaltenen Substanzen besitzen tPSA-Werte, die kleiner als 100 Å² sind.

Da die Molekülstrukturen und die physikochemischen Eigenschaften der DIVERSet-Substanzen in verschiedenen Phasen der Medikamentenentwicklung, herangezogen werden, war es notwendig, unseren Projektpartnern einen einfachen Zugriff auf diese Daten zu ermöglichen. Das im Folgenden beschriebene Paket SIFT-Viewer I ist eines der Werkzeuge, die von mir zu diesem Zweck und zur Auswertung der SIFT-Messdaten entwickelt wurden.

3.4 Das Visualisierungspaket SIFT-Viewer I

Zu Beginn des Screenings der DIVERSet-Bibliotheken, als von meinen Partnern in TP1 die ersten Platten mit Testsubstanzen mit der SIFT-Technik gemessen worden waren, wurde rasch klar, dass die generierten Messdaten automatisiert ausgewertet und in einer leicht zugreifbaren Form aufbereitet werden müssen. Die bis dahin in TP1 vorgenommene Auswertung unter Verwendung des Tabellenkalkulationsprogramms Microsoft Excel ist nur für kleinere Datenmengen geeignet, weil dafür einige Handarbeit nötig ist. Daher wurde von mir zur automatisierten Auswertung, Aufbereitung und Visualisierung der SIFT-Daten das Softwarepaket SIFT-Viewer I entwickelt.

Das Paket SIFT-Viewer I umfasst eine Reihe von Softwarewerkzeugen, die in dem Modul „screening“ für das Versionskontrollsystem CVS (Concurrent Versions System) [122] abgelegt und für die Mitarbeiter der Arbeitsgruppe Tavan über den dortigen CVS-Server zugänglich sind. Es sei darauf hingewiesen, dass das Modul „screening“ neben SIFT-Viewer I noch Werkzeuge zur Durchführung von Substruktursuchen und zur Klassifikation von SIFT-Daten (vgl. Abschnitt 4.1 weiter unten bzw. Anhang C) enthält. Die wichtigsten zu SIFT-Viewer I gehörigen Werkzeuge werde ich im Folgenden detailliert vorstellen.

3.4.1 SIFT-Dateien und die Extraktion von Messdaten

Wie in Abschnitt 2.2.2 erläutert wurde, werden von meinen Kollegen vom ZNP mit Hilfe von Insight Reader Mikrotiterplatten töpfchenweise vermessen und die dabei er-

zeugten Messdaten mithilfe der zugehörigen SIFT-Software hinsichtlich des Zählens von rot/grünen Fluoreszenzereignissen ausgewertet. Mit der SIFT-Software können die Ergebnisse einer Plattenmessung in Form einer Textdatei exportiert werden, die wir SIFT-Datei nennen. Auf diese Weise wurde eine Vielzahl von SIFT-Dateien erzeugt und mir zur automatisierten Auswertung zugeschickt. Abbildung 3.2 zeigt einen Auszug aus der Liste der von mir gesammelten SIFT-Dateien für die Platten der DIVERSet 1-Bibliothek. Um die Zuordnung von SIFT-Dateien zu Bibliotheksplatten zu erleichtern, wurde beim Screening der DIVERSet-Bibliotheken jeweils die Nummer der getesteten Platte nach einer vereinbarten Konvention in den Namen der dafür erhaltenen SIFT-Datei aufgenommen.

```
# Zuordnung von Messplatten zu Bibliotheksplatten
#
# vom vereinbarten Format abweichende Messungen sind durch
# die folgenden vorangestellten Buchstaben gekennzeichnet:
#
# u - unvollstaendige Messungen (nicht 5 mal pro Topf)
# f - Verteilung auf Platte falsch
# n - ntp-code in falschem Format
# s - schlechte Platte
# - - nicht da
#
10251 20030514_10251-2_2DSift
10252 20030513_10252-2_2DSift
10253 20030513_10253-2_2DSift
10254 20030513_10254-2_2DSift
10255 20030513_10255-2_2DSift
10256 20030513_10256-2_2DSift
10257 20030509_10257-3_2DSift
10258 20040503_10258n-2_SIFT 20030512_10258-2_2DSift
10259 20040503_10259n-2_SIFT 20030512_10259-2_2DSift uf 20030515_10259-3_2DSift
10260 20030509_10260-2_2DSift
10261 20030509_10261-2_2DSift
10262 20030508_10262-5_2DSift
10263 s 20030508_10263new_SIFT 20030508_10263-2_2DSift
10264 20030507_10264-2_2DSift
```

Abbildung 3.2: Zuordnung von SIFT-Plattenmessungen zu DIVERSet-Bibliotheksplatten. Auszugsweise ist eine von mir gepflegte Datei gezeigt, in der den Plattennummern der DIVERSet 1-Bibliothek die Namen derjenigen SIFT-Dateien zugeordnet sind, in denen die SIFT-Ergebnisse der jeweiligen (teilweise mehrfach gemessenen) Bibliotheksplatte abgelegt sind. Zusätzlich sind in der Liste aufgetretene Messfehler vermerkt. Diese Datei wurde automatisch prozessiert, wobei für jede Platte jeweils die beste (erstgenannte) SIFT-Datei mit den beschriebenen Werkzeugen behandelt wurde.

Eine SIFT-Datei enthält zeilenweise die Ergebnisse von einzelnen Töpfchenmessungen. Dabei sind in einer gegebenen Zeile die Koordinaten des gemessenen Töpfchens auf der Platte und, neben weiteren Messparametern, die sektorenweise gezählten rot/grün-Ereignisse sowie die Gesamtintensitäten $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$ angegeben. Ein Detail der Messung, das in Abschnitt 2.2.2 nur am Rande erwähnt wurde, ist, dass jedes Töpfchen einer Platte mehrfach gemessen wurde. Demzufolge enthält eine SIFT-Datei für jedes Töpfchen mehrere Zeilen mit den Ergebnissen solcher Einzelmessungen. Daraus ergab sich für die von mir entwickelte automatisierte Extraktion der Messdaten die Anforderung, diese Einzelergeb-

nisse töpfchenweise zusammenzufassen. Zu diesem Zweck programmierte ich ein Shell-Skript (`extract_data.sh` im CVS-Modul „screening“), welches unter Verwendung des Linux-Werkzeuges `Gawk` [123] zeilenweise eine SIFT-Datei prozessiert. Dabei werden anhand der in den Zeilen enthaltenen Töpfchenkoordinaten die zu einem Töpfchen gehörigen sektorenweisen rot/grün-Ereignisse der Einzelmessungen aufsummiert und ergeben auf diese Weise 2D-SIFT-Spektren. In der gleichen Weise können mit einem zweiten Skript (`extract_itot.sh`) die Gesamtintensitäten $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$ von Einzelmessungen zusammengefasst werden.

Die beiden Skripten bewerkstelligen zudem die Zuordnung der töpfchenweise extrahierten Messergebnisse zu den getesteten Substanzen. Diese Zuordnung wird dadurch ermöglicht, dass die meisten der im DIVERSet-Screening gemessenen Platten nach dem in Abbildung 2.6 dargestellten Layout „Eins-zu-Eins“ belegt waren. Dadurch kann für jedes Töpfchen, in dem eine Testsubstanz gemessen wurde, aus dem Namen der SIFT-Datei (vgl. Abbildung 3.2) und aus den Koordinaten des Töpfchens der Name der Testsubstanz in dem gemäß (3.11) vereinbarten Format (`<Plattennr.>_<Töpfchenkoord.>`) rekonstruiert werden. Den Töpfchen mit Kontrollmessungen werden dabei „Substanznamen“ der Form `<Plattennr.>_X01` zugeordnet, wobei X01 die Zeile X in Spalte 01 bezeichnet. Weil beim Screening der DIVERSet-Bibliotheksplatten die Kontrollen immer an fest definierten Zeilenpositionen in der Spalte 01 gemessen wurden, kann daraus der Typ der Kontrolle dekodiert werden. Es sei angemerkt, dass die hier und im Folgenden beschriebene Auswertung mit SIFT-Viewer I nur solche Platten unterstützt, die nach dem in Abbildung 2.6 dargestellten Layout belegt sind, welches beim Screening der DIVERSet-Bibliotheken verwendet wurde.

Die mithilfe der Skripten `extract_data.sh` und `extract_itot.sh` extrahierten Daten können in Textdateien abgelegt werden, die zeilenweise Substanznamen und zusammengefasste Messwerte enthalten. Ein weiteres Werkzeug (`filter.c`) dient der plattenweisen Ausreißererkenkung nach den in Abschnitt 3.1.1 definierten Kriterien.

Plattenweise Prozessierung

Da beim Screening der beiden DIVERSet-Bibliotheken in TP1 nach und nach weitere der jeweils 125 Platten durchgemessen und mir kurz darauf die für die neu gemessenen Platten erzeugten SIFT-Dateien zugeschickt wurden, ergab sich für mich die Aufgabe, diese Dateien zu sammeln (vgl. Abbildung 3.2) und den Bestand der mithilfe der oben beschriebenen Werkzeuge extrahierten und bereinigten SIFT-Daten zu aktualisieren. Zu diesem Zweck wurde ein weiteres Skript (`prepre.sh`^a) programmiert, welches die in Abbildung 3.2 auszugsweise gezeigte Liste von SIFT-Dateien zu den jeweiligen Plattennummern abarbeitet und dabei die jeweils beste SIFT-Datei der Datenextraktion und Aus-

^aDer ungewöhnliche Name „`prepre.sh`“ dieses Skriptes erklärt sich durch seine Rolle als Vorverarbeitungsschritt (`pre`) der in Anhang C beschriebenen Klassifikation, welche mit einem weiteren Schritt der Vorverarbeitung (`pre`) beginnt.

reißerererkennung mit den Skripten `extract_data.sh` und `extract_itot.sh` bzw. `filter.c` zufführt. Als Ergebnis dieser Prozessierung liegen die extrahierten, nach Ausreißern gefilterten und mit Substanznamen versehenen 2D-SIFT-Spektren und Gesamtintensitäten $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$ in Form von fünf Dateien teilweise binären Formats vor und stehen somit der weiteren Auswertung zur Verfügung.

3.4.2 Grafische Aufbereitung von SIFT-Daten

Zur automatisierten Erzeugung von Grafikdateien, welche die 2D-SIFT-Spektren, die Summenobservablen $S_v^b(w)$ und die Wirksamkeiten darstellen, wurde von mir ein Softwarepaket für die Auswertungsumgebung IDL (Interactive Data Language, [124]) entwickelt. Die Werkzeuge dieses Pakets sind ebenfalls über das CVS-Modul „screening“ verfügbar. Das Paket umfasst das IDL-Modul `screening_common.pro`, dessen Prozedur `setup` zur Initialisierung von allgemeinen Informationen über die Nummern der gemessenen Platten und die Lage der oben beschriebenen extrahierten Daten im Dateisystem dient. Die vermittels `setup` eingelesenen Informationen werden von den im Modul `screening_visual.pro` befindlichen Prozeduren verwendet, welche der eigentlichen Erzeugung von Grafikdateien dienen. Die Funktionsweisen dieser Prozeduren sollen nun im einzelnen skizziert werden.

Die Prozedur `print_raw` zur plattenweisen Darstellung von 2D-SIFT-Spektren arbeitet die eingelesene Plattenliste ab und wählt dabei anhand der Töpfchennamen nach Konvention (3.11) alle zu dieser Platte gehörigen Spektren aus der Menge aller Spektren aus. Die auf diese Weise ausgewählten Spektren einer Platte werden gemäß dem für das Screening vereinbarten Plattenlayout und wiederum anhand der Töpfchennamen nach Töpfchentypen gruppiert (vgl. Abschnitt 3.2.2), wobei die zusätzliche Gruppe der Töpfchen mit Ausreißerspektren gebildet wird. Die gruppierten Spektren werden dann in unterschiedlichen Farben geplottet (vgl. die in Abbildung 2.6 verwendete Farbkodierung). Für die spätere Visualisierung mit SIFT-Viewer I wird dabei für jedes Töpfchen ein Plot aller Spektren auf der Platte erzeugt, in dem das gegebene Töpfchen hervorgehoben dargestellt ist. Diese Plots werden zunächst in ps-Dateien (PostScript) geschrieben, sogleich in jpeg-Dateien (Joint Photographic Experts Group) umgewandelt und für die Verwendung mit SIFT-Viewer I nach Platten sortiert an (über `setup`) definierten Positionen im Dateisystem abgelegt (vgl. die Beschreibung der Verzeichnishierarchie in Abschnitt 3.4.4 weiter unten). Es sei angemerkt, dass die in Abschnitt 2.2.3 dargestellten 2D-SIFT-Spektren mit einer leicht modifizierten Version dieser Prozedur erstellt wurden.

In ähnlicher Weise werden mit der Prozedur `print_sums` platten- und töpfchenweise einzelne 2D-SIFT-Spektren eingelesen und daraus die gemäß (3.6) definierten Summenobservablen $S_v^b(w)$ für die Sektorenbereiche $[v, b] = [1, 5], [1, 7], [4, 9]$ und $[1, 18]$ berechnet. Diese Prozedur plottet sowohl die für die Töpfchen einer Platte berechneten Summenobservablen als auch die Gesamtfluoreszenzen $I_{\text{tot,r}}$ und $I_{\text{tot,g}}$ und legt den Plot als jpeg-Grafikdatei an einer definierten Stelle ab (vgl. Abschnitt 3.4.4). Analog dazu werden

mit der Prozedur `print_inhibs` die gemäß (3.9) definierten Wirksamkeiten berechnet, geplottet und in weiteren jpeg-Dateien gespeichert.

3.4.3 Darstellung von Molekülstrukturen

Die im Lieferumfang der DIVERSet-Bibliotheken in Form von sdf-Dateien enthaltenen Molekülstrukturen (vgl. Abschnitt 3.3) wurden einer visuellen Repräsentation zugänglich gemacht, indem mit dem Werkzeug `csbr` des Softwarepakets CACTVS [125] für jede Substanz eine Grafikdatei generiert wurde. Mit `csbr` können große Mengen von Substanzen aus sdf-Dateien am Bildschirm dargestellt werden, wobei zudem die in der sdf-Datei enthaltenen Substanzeigenschaften angezeigt werden können. Auf diese Weise wurden mithilfe von `csbr` die Strukturen der DIVERSet-Substanzen zusammen mit den Plattennummern und den Töpfchenkoordinaten dargestellt und einzeln in eps-Dateien (Encapsulated PostScript) abgespeichert. Die unter generischen Namen (etwa „`struct_1.eps`“ bis „`struct_10000.eps`“) abgespeicherten Dateien wurden dann mithilfe eines von mir geschriebenen Skripts (`name_structs.sh` im Modul „screening“) umformatiert und umbenannt. Das Skript extrahiert dabei aus der eps-Datei für eine gegebene Substanz die darin verborgene Plattennummer und die Töpfchenkoordinaten, konvertiert die Datei ins jpeg-Format, benennt die resultierende jpeg-Datei nach den Töpfchenkoordinaten und legt sie in einem nach der Plattennummer benannten Verzeichnis innerhalb der im folgenden Abschnitt beschriebenen Verzeichnishierarchie ab. Dadurch werden auch diese Grafikdateien der Visualisierung mit SIFT-Viewer I zugänglich gemacht.

3.4.4 Die SIFT-Viewer I zugrunde liegende Verzeichnishierarchie

Der Webbrowser-basierten Visualisierung der SIFT-Ergebnisse einer gescreenten Substanzbibliothek mit SIFT-Viewer I liegt intern eine Hierarchie von Dateien und Verzeichnissen zugrunde, auf die in den letzten Abschnitten wiederholt verwiesen wurde. In dieser Hierarchie sind zum einen die oben beschriebenen Grafikdateien für diverse SIFT-Daten und Substanzstrukturen und zum anderen eine Repräsentation der Substanzbibliothek in Form von HTML-Dokumenten (Hypertext Markup Language [126]) abgelegt, welche als Platzhalter für die Grafikdateien dienen und nachfolgend beschrieben werden.

Abbildung 3.3 zeigt auszugsweise die Verzeichnishierarchie, die der Visualisierung des SIFT-Screenings von DIVERSet 2 zugrunde liegt. Die Grafikdateien der 2D-SIFT-Spektren sind unterhalb des Verzeichnisses `sifts` abgelegt, in welchem für jede Platte ein nach der Plattennummer benanntes Unterverzeichnis existiert. In analoger Weise liegen die Grafikdateien der Summenplots, der Wirksamkeiten und der Molekülstrukturen jeweils unterhalb der Verzeichnisse `sums`, `inhibs` bzw. `structs_2d`. Diese Haltung der Daten in getrennten Unterverzeichnishierarchien identischer Struktur erfüllt den Zweck,

```
DIVERSet2
|-- INP
|   |-- index.html
|   |-- plate_drop_down.html
|   [...]
|   `-- platten
|       |-- 30126
|           |-- A01.html
|           [...]
|           |-- H12.html
|           |-- platte_a.html
|           `-- platte_b.html
|       [...]
|       `-- 30250
|-- sifts
|   |-- 30126
|       |-- A01.jpeg
|       [...]
|       `-- H11.jpeg
|   [...]
|   `-- 30250
|-- sums
|   |-- 30126
|       `-- sums.jpeg
|   [...]
|   `-- 30250
|-- inhibs
|   |-- 30126
|       `-- inhibs.jpeg
|   [...]
|   `-- 30250
`-- structs_2d
    |-- 30126
    |   |-- A02.html
    |   |-- A02.jpeg
    |   [...]
    |   `-- H11.jpeg
    [...]
    `-- 30250
```

Abbildung 3.3: Verzeichnishierarchie zu SIFT-Viewer I. Die zur Visualisierung des SIFT-Screenings von DIVERSet 2 generierte Verzeichnishierarchie enthält eine Vielzahl von HTML-Dokumenten und Grafikdateien, die hier nur auszugsweise dargestellt sind. Weggelassene Teile sind durch „[...]“ ersetzt. Die Abbildung wurde mithilfe des Linux-Kommandos `tree` erzeugt.

dass diejenigen Daten, die sich infolge neu hinzugekommener SIFT-Messergebnisse häufig ändern können (Spektren, Summen, Wirksamkeiten) auf diese Weise leicht gelöscht und neu erzeugt werden können, während der statische Bereiche der Molekülstrukturen davon unberührt bleibt.

Ein weiterer statischer Bereich der in Abbildung 3.3 dargestellten Verzeichnishierarchie ist derjenige, der die eingangs erwähnte Repräsentation der Substanzbibliothek enthält und sich unterhalb des Verzeichnisses `INP` befindet. Im Unterverzeichnis `INP/platten` existieren dazu für jede Platte ein nach der Plattennummer benanntes Unterverzeichnis und in jedem dieser Unterverzeichnisse eine Reihe von HTML-Dateien:

Die Datei `platte_b.html` enthält eine Repräsentation der gegebenen Platte in Form einer Tabelle, in der für jedes Töpfchen der realen Bibliotheksplatte ein nach den Töpfchenkoordinaten benannter Eintrag steht. Die Datei `platte_a.html` dient als Platzhalter für die (in `sums` bzw. `inhibs` abgelegten) Grafikdateien der Summen- und Wirksamkeitsplots. Des Weiteren existiert für jedes Töpfchen eine nach den Töpfchenkoordinaten benannte HTML-Datei wie beispielsweise `A01.html`, welche die Grafikdateien der zugehörigen 2D-SIFT-Spektren und Molekülstrukturen (aus `sifts` bzw. `structs_2d`) einbettet. Hinter den Töpfchenrepräsentationen in `platte_b.html` liegen Hyperlinks, welche auf die zu den Töpfchen gehörigen HTML-Dateien (z.B. `A01.html`) verweisen. Auf diese Art ist die plattenweise Verschaltungslogik von SIFT-Viewer I realisiert. Zur Generierung der beschriebenen HTML-Dokumente steht das Skript `gen_html_plates.sh` aus dem CVS-Modul „screening“ zur Verfügung.

Von diesem Skript wird auch die Datei `index.html` im Verzeichnis `INP` generiert, welche den Einstiegspunkt für den Benutzer darstellt^a und mit einem Webbrowser geöffnet werden muss, um die Darstellung von SIFT-Viewer I zu starten. In dieser Datei wird der vom Browser dargestellte Bereich in Unterbereiche aufgeteilt, die Frames genannt werden und deren Inhalt entweder statisch ist oder zur „Laufzeit“ infolge von Benutzerinteraktion dynamisch verändert wird. In einem statischen Frame von `index.html` findet sich ein sogenanntes Drop-Down-Menü zur Auswahl von Platten, das unter Verwendung der Programmiersprache JavaScript [127] implementiert wurde, sowie eine Reihe von variablen Bereichen, in denen die im letzten Abschnitt geschilderten HTML-Dokumente dargestellt werden, welche die Summen- und Aktivitätsplots, die Plattenrepräsentationen, die 2D-SIFT-Spektren und die Molekülstrukturen zu den vom Benutzer ausgewählten Platten und Töpfchen enthalten.

3.4.5 Webbrowser-basierte Visualisierung mit SIFT-Viewer I

Um dem Leser einen Eindruck der Visualisierung mit SIFT-Viewer I zu geben, zeigt Abbildung 3.4 einen Bildschirmschnappschuss dieses Werkzeugs. In diesem Schnappschuss sind die SIFT-Ergebnisse für eine Platte der Bibliothek DIVERSet 2 dargestellt. Die dargestellte Platte (Nummer 30163) wurde vom Benutzer über das Drop-Down-Menü im rechten oberen Bereich aus allen 125 DIVERSet 2-Platten ausgewählt. Daraufhin werden die zu dieser Platte gehörigen Darstellungen in den übrigen Bereichen des Browserfensters angezeigt.

- Die Grafik im oberen zentralen Bereich zeigt in verschiedenen Farben Plots der Summenobservablen S_1^5 (orange), S_1^7 (lila), S_4^9 (blau) und S_1^{18} (cyan), sowie der Gesamtfluoreszenzen $I_{\text{tot,r}}$ (rot) und $I_{\text{tot,g}}$ (grün), welche mit den Werkzeugen aus Abschnitt 3.4.2 erzeugt wurden. In dieser Darstellung durchläuft, wie in der Beschriftung

^aDer Verzeichnisname `INP` wurde gewählt, um meinen Kollegen vom ZNP, das zur Zeit der Erstellung von SIFT-Viewer I noch `INP` (Institut für Neuropathologie) hieß, den Einstiegspunkt zu kennzeichnen.

3 Methoden zur Haltung, Auswertung und Visualisierung von Daten



Abbildung 3.4: Bildschirmschnappschuss von SIFT-Viewer I. Das Visualisierungswerkzeug SIFT-Viewer I ermöglicht dem Benutzer einen schnellen Zugriff auf alle Substanzen der DIVERSet-Bibliotheken, auf deren Molekülstrukturen, diverse SIFT-Messdaten und daraus berechnete Wirksamkeiten. Genauere Erläuterungen finden sich im Text.

ung angegeben ist, die waagrechte Achse zeilenweise die Töpfchen der zugrunde liegenden Platte (A1, A2, ..., B1, B2, ..., H11).

- Im darunter liegenden Plot sind in analoger Weise die gemäß (3.9) berechneten kontrollenbezogenen SIFT-Wirksamkeiten als Kreuze angezeigt. Für die gezeigte Platte werden vier Töpfchen die Wirksamkeiten $a \approx 1$ zugeordnet. Es handelt sich dabei ausschließlich um Kontrollmessungen. Eine der Messungen (E11) zeigt eine recht hohe Aktivität. Aus Gründen des Patentschutzes können zur chemischen Struktur dieser Substanz keine Informationen gegeben werden.
- Im linken unteren Bereich findet sich die Tabellenrepräsentation der Messplatte, in welcher jedes Töpfchen durch einen aus den Töpfchenkoordinaten zusammengesetzten Namen symbolisiert ist. Im dargestellten Schnappschuss wurde durch Kli-

cken auf den Töpfchennamen „F09“ das in Zeile „F“ und Spalte „09“ befindliche Töpfchen ausgewählt.

- Daraufhin wird in dem Plot rechts daneben, der die 2D-SIFT-Spektren aller Töpfchen der ausgewählten Platte 30163 zeigt, das zu F09 gehörige Spektrum gelb hervorgehoben und rechts neben dem Spektren-Plot die Molekülstruktur der ausgewählten Testsubstanz 30163_F09 angezeigt. Wie dem gelb eingefärbten Spektrum und dem Wert der für F09 berechneten SIFT-Aktivität zu entnehmen ist, handelt es sich hierbei um eine kaum aggregationshemmende Substanz.

Durch das Zurverfügungstellen der Software SIFT-Viewer I konnten wir unseren Projektpartnern ermöglichen, die Ergebnisse des SIFT-Screenings von DIVERSet 1 und 2 rasch visuell zu inspizieren. Obschon SIFT-Viewer I erfolgreich eingesetzt wurde, zeigte sich eine Reihe von Schwachpunkten dieses Werkzeugs. Dazu zählt zum einen die in Abschnitt 3.4.1 angesprochene Beschränkung auf das beim primären Screening der DIVERSet-Bibliotheken verwendete Layout der Messplatten. Infolge dieser Beschränkung konnten Messplatten mit abweichendem Layout, wie sie beispielsweise bei den anschließenden SIFT-Validierungsexperimenten in Form von Verdünnungsreihen (vgl. Abschnitt 2.2.4) eingesetzt wurden, mit SIFT-Viewer I nicht ausgewertet werden. Zum anderen gibt es einen grundlegenden Schwachpunkt von SIFT-Viewer I, der in der eingesetzten Prozedur der Datenaufbereitung begründet liegt. Denn dabei werden – wie oben beschrieben – die im Labor von TP1 erzeugten SIFT-Messdaten zunächst zu mir geschickt, von mir ausgewertet, grafisch aufbereitet und in der statischen Verzeichnishierarchie abgelegt, welche dann an die Experimentatoren zurückgeschickt wird, damit diese endlich die Auswertung betrachten können. Mit SIFT-Viewer I können also die Messergebnisse nicht unmittelbar nach der Messung inspiziert werden. Um diese und weitere Nachteile der dateibasierten Auswertung mit SIFT-Viewer I zu beheben, wurde von mir zum einen eine Datenbank zur Haltung der Mess- und Substanzdaten für den Verbund eingerichtet und zum anderen eine zweite Version des Auswertungswerkzeugs, SIFT-Viewer II, entwickelt, mit der auf diese Datenbank zugegriffen werden kann. Daher möchte ich zuerst die Datenbank und weiter unten in Abschnitt 3.6 das Werkzeug SIFT-Viewer II vorstellen.

3.5 Die Screening-Datenbank TSE-DB

Zur Haltung der für die DIVERSet-Bibliotheken (vgl. Abschnitt 3.3) und für einige fokussierte Bibliotheken im Verbund gewonnenen Messdaten und -ergebnisse richtete ich auf einem vom Leibniz-Rechenzentrum (LRZ) in München zur Verfügung gestellten Datenbankserver [128] eine Datenbank ein, die ich TSE-DB genannt habe. Herr Simon Youssef unterstützte mich als studentische Hilfskraft sowohl beim Entwurf und bei der Einrichtung der TSE-DB als auch bei der Entwicklung der zugehörigen Java-Bibliothek `tseadb` und der darauf basierenden Datenbank-Clientprogramme TSE-DB-Browser und SIFT-View-

er II, die dem Zugriff auf die in der TSE-DB abgelegten Daten dienen und weiter unten beschrieben werden.

3.5.1 Modellierung und Installation der TSE-DB

Für den Entwurf und die Installation der TSE-DB wurde das Modellierungswerkzeug „Toad Data Modeler“ der Firma Quest Software [129] eingesetzt^a. Mit Toad Data Modeler wurde zunächst ein sogenanntes Entity-Relationship (ER)-Modell [130] für die zu erstellende TSE-DB entworfen. ER-Modelle dienen der Beschreibung, wie Objekte der Realität (Entitäten) und deren Beziehungen (Relationen) auf Datenstrukturen abgebildet werden können, und werden daher häufig zur Modellierung von Datenbanken eingesetzt.

Abbildung 3.5 zeigt das für die TSE-DB entworfene ER-Modell als Diagramm, in dem die 18 Tabellen der TSE-DB als Kästen dargestellt sind. Die Kästen sind farblich hinterlegt, um inhaltlich eng zusammenhängende Tabellen zu gruppieren. Danach gliedert sich die TSE-DB in vier Gruppen von Tabellen, deren Inhalte substanzbezogene Daten (gelb), SIFT-Daten (türkis), Zellkulturdaten (grau) oder Wirksamkeitsdaten (weiß) sind. In den Kästen sind jeweils der Tabellename und die Namen der Felder oder Attribute angegeben, wobei die sogenannten *Schlüsselattribute* zudem farblich hervorgehoben und durch die Kürzel „PK“ für Primärschlüssel bzw. „FK“ für Fremdschlüssel gekennzeichnet sind. Über diese ausgezeichneten Attribute können Tabellen miteinander verknüpft werden.

Betrachten wir zunächst die gelb hinterlegten Tabellen SUBSTANCE, LIBRARY_PLATE und LIBRARY, die angelegt wurden, um die Organisation der realen Testsubstanzen, die auf Platten verteilt und in Bibliotheken zusammengefasst sind, zu modellieren. Zum einen enthalten diese Tabellen Felder, die rein zur Haltung von Daten eingerichtet wurden – dazu zählen im Fall der SUBSTANCE-Tabelle beispielsweise die Felder C_LOG_P, T_PSA etc. für die Haltung der in Abschnitt 3.3 beschriebenen physikochemischen Eigenschaften. Zum anderen enthalten die Tabellen aber auch Schlüsselattribute wie sie oben angesprochen wurden. Für die SUBSTANCE-Tabelle wurde das Feld SUBSTANCE_ID, über welches Substanzen eindeutig benannt werden, als Primärschlüssel (PK) festgelegt. Ein Fremdschlüssel (FK) der SUBSTANCE-Tabelle ist das Feld LIBRARY_PLATE_ID, welches zudem der Primärschlüssel der Tabelle LIBRARY_PLATE ist. Dadurch wird modelliert, dass mehrere Substanzen auf einer Bibliotheksplatte liegen können und dass jede Substanz auf genau einer Bibliotheksplatte liegt. Zur Darstellung dieser Beziehung im ER-Diagramm sind die beiden Tabellen durch eine Linie verbunden, deren Enden mit den Symbolen „>“ und „-“ versehen sind. Diese Symbole kennzeichnen den Grad der Beziehung, wobei „>“ für „mehrere“ und „-“ für „eines“ steht. Die auf diese Weise realisierte Modellierung der substanzbezogenen Daten in Form von drei separaten, miteinander verknüpften Tabellen soll dafür sorgen, dass Redundanzen verringert und die Integrität der Daten gewährleistet wird.

^aToad Data Modeler hieß damals „CASE Studio 2“ und wurde von der Firma CHARONWARE angeboten.

3.5 Die Screening-Datenbank TSE-DB

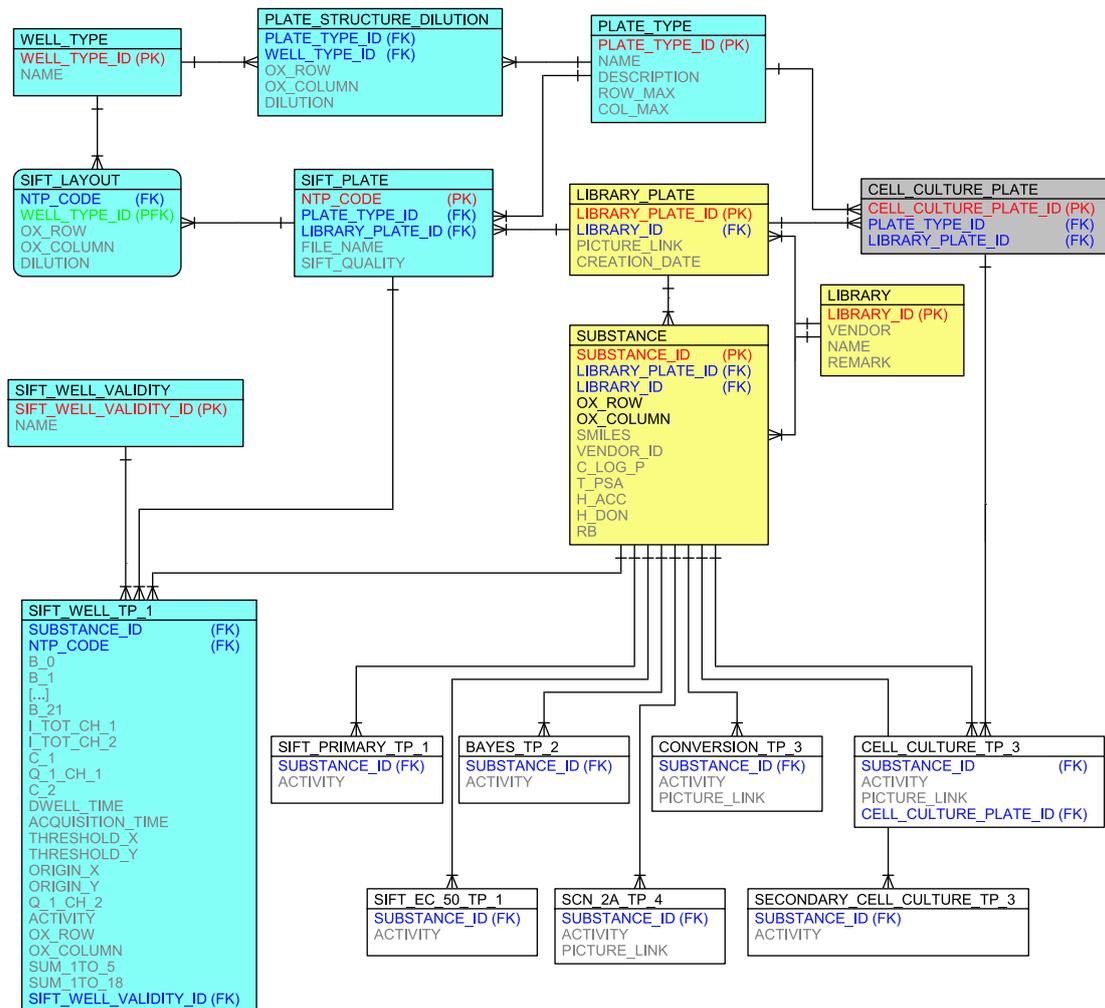


Abbildung 3.5: ER-Diagramm der TSE-DB. Die 18 Tabellen der TSE-DB sind durch Kästen repräsentiert, in denen jeweils der Tabellename, die Primär- (PK) und Fremdschlüssel (FK) sowie die weiteren Felder angegeben sind. Zur Gruppierung nach inhaltlichen Zusammengehörigkeiten sind die Tabellen farblich hinterlegt, wobei nach Substanzdaten (gelb), SIFT-Daten (türkis), Zellkulturdaten (grau) und Wirksamkeitsdaten (weiß) unterschieden wird. Aus Gründen der Übersichtlichkeit wurden einige Felder der Tabelle SIFT_WELL_TP_1 nicht dargestellt („[...]“). Die Abbildung wurde mithilfe von Toad Data Modeler [129] erstellt. Erläuterungen siehe Text.

Unter Berücksichtigung derartiger Modellierungsgesichtspunkte wurden auch die türkis hinterlegten Tabellen entworfen, welche SIFT-Messdaten, SIFT-Messplatten und deren Layout repräsentieren. Dabei ist die Tabelle SIFT_WELL_TP_1 für die Haltung der töpfchenweise gewonnenen SIFT-Daten vorgesehen, zu denen die 2D-SIFT-Spektren, die Gesamtintensitäten und weitere Messparameter zählen (vgl. die Abschnitte 2.2.3 und 3.4.1). Die Tabelle SIFT_PLATE dient der Zuordnung von SIFT-Plattenmessungen zu den dabei gemessenen Bibliotheksplatten und ist für die Haltung der im primären Screening vollstän-

dig gemessenen Bibliotheksplatten vorgesehen. Diese Tabelle korrespondiert zu der von SIFT-Viewer I realisierten dateibasierten Zuordnung anhand der in Abbildung 3.2 dargestellten Liste. Über die Tabellen `PLATE_TYPE`, `WELL_TYPE` und `PLATE_STRUCTURE_DILUTION` sind verschiedene fest vereinbarte Standard-Layouts definiert, welche beim Primärscreening der DIVERSet-Bibliotheken und den anschließenden Validierungsmessungen eingesetzt wurden. Alternativ zu diesen Standard-Layouts können vermittels der Tabelle `SIFT_LAYOUT` frei wählbare Layouts realisiert werden. Hierzu bietet das weiter unten in Abschnitt 3.6 beschriebene Werkzeug SIFT-Viewer II dem Benutzer die Möglichkeit, Messplattenlayouts festzulegen. Die ursprünglich zur Markierung von Ausreißern vorgesehene Tabelle `SIFT_WELL_VALIDITY` blieb ungenutzt, weil Ausreißer stattdessen mit SIFT-Viewer II zur Laufzeit erkannt werden.

Die weiß hinterlegten Tabellen dienen der Haltung der Substanzwirksamkeiten aus allen im Verbund eingesetzten Modellsystemen. Dazu besitzen diese Tabellen jeweils das Feld `SUBSTANCE_ID` der `SUBSTANCE`-Tabelle als Fremdschlüssel und zur Haltung der Wirksamkeiten die Felder `ACTIVITY` vom Typ Gleitkommazahl. Die Tabelle `SIFT_PRIMARY_TP_1` ist für SIFT-Wirksamkeitsmaße, die gemäß den in Abschnitt 3.2 definierten Methoden für das primäre Screening gebildet werden können, vorgesehen und die Tabelle `SIFT_EC_50_TP_1` für die in Abschnitt 2.2.4 angesprochenen SIFT-EC50-Werte. Des Weiteren wurden die Tabelle `SCN_2A_TP_4` für die Wirksamkeiten aus den Zellkulturexperimenten von TP4 (Abschnitt 2.3) und die Tabellen `CELL_CULTURE_TP_3`, `SECONDARY_CELL_CULTURE_TP_3` sowie `CONVERSION_TP_3` für die im Zellkultur- bzw. im Konversions-Assay in TP3 (Abschnitt 2.4) ermittelten Wirksamkeiten angelegt. Die Wirksamkeitsmaße für die Kodierungen der Messergebnisse aus den letztgenannten Modellsystemen werden im folgenden Abschnitt 3.5.2 beschrieben.

Es sei am Rande erwähnt, dass die ebenfalls weiß hinterlegte Tabelle `BAYES_TP_2` für Ergebnisse von versuchsweise von mir durchgeführten Bayesschen Klassifikationen der SIFT-Screening-Daten genutzt wurde (vgl. die ähnlichen Methoden in Anhang C). Ferner sei angemerkt, dass die grau hinterlegte Tabelle `CELL_CULTURE_PLATE`, die für rohe Messdaten des in TP3 durchgeführten Zellkultur-Screening der DIVERSet-Bibliotheken vorgesehen ist, bislang ungenutzt blieb, weil entgegen unserer damaligen Erwartung von unseren Projektpartnern keine derartigen Daten produziert wurden. Stattdessen wurden, wie oben beschrieben ist, lediglich die *Resultate* der Experimente aus TP3 aufgenommen.

Installation der TSE-DB

Um die gemäß dem oben dargestellten ER-Modell für die TSE-DB beschriebenen Tabellen und deren Beziehungen auf den Datenbankserver am LRZ zu übertragen, wurde wiederum das Werkzeug Data Modeler eingesetzt. Mithilfe dieses Tools wurde aus dem ER-Modell Code der Sprache DDL (Data Definition Language) generiert, der zur Erzeugung der Tabellen auf dem Datenbankserver eingesetzt wurde. Die Sprache DDL ist eine

Teilmenge von SQL (Structured Query Language) und für diesen Zweck bestimmt. Auf diese Weise wurde eine leere Datenbankstruktur erzeugt, die noch befüllt werden musste.

3.5.2 Datenformate und Befüllung der TSE-DB

Bevor beschrieben wird, wie die leere TSE-DB mit Substanz-, Mess- und Wirksamkeitsdaten befüllt wurde, sollen zunächst einige Informationen zu den dabei verwendeten Datenformaten gegeben werden.

Wirksamkeitsmaße und weitere Formate

Insbesondere muss noch das Format für die „Wirksamkeitsdaten“ festgelegt werden, die in den dafür vorgesehenen (in Abbildung 3.5 weiß hinterlegten) Tabellen abgelegt werden sollen. Wie im letzten Abschnitt bereits angesprochen wurde, können die SIFT-Wirksamkeiten, für deren Haltung Tabelle SIFT_PRIMARY_TP_1 vorgesehen ist, mit den in Abschnitt 3.2 definierten Maßen aus den SIFT-Messdaten berechnet werden. Für den Fall des primären Screenings der DIVERSet-Bibliotheken geschah dies unter Verwendung des kontrollenbezogenen Wirksamkeitsmaßes gemäß Gleichung (3.9) mithilfe von SIFT-Viewer II.

Während SIFT-Wirksamkeiten berechnet werden müssen, liegen die Ergebnisse der Experimente mit den übrigen Modellsystemen [Zellkultur-Assays (TP3 und TP4) und Konversions-Assay (TP3)] bereits von den Experimentatoren ausgewertet, in einfach zu handhabender Form vor. Dabei wird schlicht angegeben, ob eine getestete Substanz *wirksam* oder *unwirksam* ist. Infolgedessen könnten die Testergebnisse von Booleschen Variablen erfasst werden. Um jedoch die Beschreibung von etwa als *schwach wirksam* eingestuften Substanzen zu ermöglichen, entschloss ich mich dazu, die Ergebnisse auf Gleitkommazahlen abzubilden, wobei 0,0 für „unwirksam“ und 1,0 für „wirksam“ steht, sodass „schwach wirksamen“ Substanzen dazwischenliegende Werte zugeordnet werden können. Es sei angemerkt, dass für die Ergebnisse der Zellkulturexperimente aus TP4 und des Konversions-Assays aus TP3 ausschließlich Werte von 0,0 und 1,0 eingesetzt wurden, während bei den Zellkulturergebnissen aus TP3 auch Zwischenwerte verwendet wurden.

Die Benennung der DIVERSet-Substanzen in der TSE-DB erfolgte nach der Konvention (3.11). Zur Beschreibung der Molekülstrukturen wurde das verbreitete SMILES-Format (Simplified Molecular Input Line Entry System [131]) verwendet, welches die kompakte Darstellung einer Molekülstruktur in Form einer Zeichenkette erlaubt. Beispielsweise kann die in Abbildung 3.1 dargestellte Struktur der DIVERSet-Substanz 10252_A07 durch den SMILES-String „O=C1C=C(Br)C(C=CC=C2)=C2O1“ beschrieben werden. Mithilfe des frei verfügbaren Konvertierungswerkzeugs `babel` [132] und der kommerziellen Chemie-Software ChemFinder der Firma CambridgeSoft [133] wurden die mit

den DIVERSet-Bibliotheken mitgelieferten Molekülstrukturen im sdf-Format (vgl. Abschnitt 3.3) in das SMILES-Format umgewandelt.

Befüllung der TSE-DB

Die Befüllung der Tabellen der TSE-DB wurde teilweise „von Hand“ und teilweise automatisiert vorgenommen. Im Fall der kleineren Tabellen LIBRARY, PLATE_TYPE und WELL_TYPE wurden die Daten händisch mithilfe des Werkzeugs „Datenbank-Pilot“ der kommerziellen Entwicklungsumgebung JBuilder von der Firma Borland [134] in die TSE-DB eingetragen. Zur Befüllung der restlichen Tabellen mit den großen Mengen an Substanz-, SIFT-Mess- und Wirksamkeitsdaten, wurden all diese Daten mithilfe von dazu von mir entwickelten Shell-Skripten zunächst auf Textdateien verteilt, deren Format durch die in Abschnitt 3.5.1 beschriebenen Tabellendefinitionen vorgegeben ist. Nach den vorgegebenen Definitionen wurde zu jeder Tabelle eine Textdatei generiert, deren Spalten durch die Felder der gegebenen Tabelle festgelegt sind. In der ersten Zeile einer solchen Textdatei stehen dabei die Namen der Felder und in den weiteren Zeilen die zu importierenden Datensätze, wobei die Einträge jeweils durch Tabulatorzeichen getrennt werden. Die in dieser Form erzeugten Textdateien wurden schließlich mithilfe von Standarddatenbankfunktionen in die TSE-DB importiert.

Bei der sogenannten Uraladung, dem erstmaligen Befüllen, der TSE-DB wurden die Daten zu den Substanzbibliotheken und die bis zum damaligen Zeitpunkt vorliegenden Messergebnisse und Wirksamkeiten in die TSE-DB übertragen. Spätere Aktualisierungen der Datenbank wurden teils mit den eben genannten Mitteln vollzogen, teils mit Werkzeugen der in Anhang B beschriebenen Java-Bibliothek `tse_db`. Diese Java-Bibliothek enthält eine Vielzahl von Komponenten zur Auswertung und grafischen Darstellung der in der TSE-DB gespeicherten Daten. Insbesondere enthält sie mit dem im folgenden Abschnitt beschriebenen TSE-DB-Browser ein Werkzeug, das unseren Verbundpartnern über eine grafische Oberfläche den Zugriff auf alle in der TSE-DB abgelegten Daten ermöglicht.

3.5.3 Zugriff mit TSE-DB-Browser

Um einen Einblick in die Fähigkeiten des TSE-DB-Browsers zu geben, sind in Abbildung 3.6 zwei Bildschirmschnappschüsse dieses Datenbankclients gezeigt. Wie aus den beiden Schnappschüssen zu erkennen ist, bietet die Oberfläche des TSE-DB-Browsers im linken Bereich Baumstrukturen zur Auswahl von Substanzmengen und im rechten Bereich Tabellen, in welchen die in den ausgewählten Mengen enthaltenen Substanzen aufgelistet sind. Diese Tabellendarstellungen zeigen die Substanznamen, die Wirksamkeiten in den verschiedenen experimentellen Tests und die Molekülstrukturen der ausgewählten Substanzen. Über die beiden Karteikartenreiter im linken unteren Teil kann der Benutzer zwischen zwei verschiedenen Baumstrukturen zur Auswahl von Substanzmengen wählen.

A

Substance	SFT PRIM T	SFT EC 50 T	Bayes TP 2	SCNZA TP4	Conversion	Cell Culture	Secondary	Smiles
10251_D04	-0.0555917		1.24095E-6			0.0		<chem>O=C(O)CC(=O)Nc1ccccc1</chem>
10251_D05	0.107434							
10251_D06	-0.230844							
10251_D07	-0.0107597							
10251_D08	0.244539							

B

Substance	SFT PRIM T	SFT EC 50 T	Bayes TP 2	SCNZA TP4	Conversion	Cell Culture	Secondary	Smiles
10309_F02		30.0		1.0	0.0	0.0		<chem>O=C(O)CC(=O)Nc1ccccc1</chem>
10313_B02	0.570222	3.5	2.13467E-7	1.0	0.0			c1ccc(C=NN=C(N)N)cc1OCc2C
10369_D05	1.04028	0.4	2.72006E-5	1.0	1.0			<chem>O=C(O)CC(=O)Nc1ccccc1</chem>
10283_G02	0.513819	0.3	1.27245E-4	1.0	0.0	0.0		<chem>O=C(O)CC(=O)Nc1ccccc1</chem>
10260_A08	0.772177		3.36374E-4	1.0	0.0	0.0		<chem>O=C(O)CC(=O)Nc1ccccc1</chem>

Abbildung 3.6: Zugriff auf die TSE-DB mit dem TSE-DB-Browser. Der TSE-DB-Browser bietet den Mitarbeitern des Verbundes Zugriff auf die in den verschiedenen Assays gewonnenen Messergebnisse. Dabei können die Messdaten über die Bibliotheksplatten (A) oder über Listen von ausgewählten Substanzen (B) erreicht werden.

Der in Schnappschuss A durch Wahl des Reiters „Libraries Tree“ ausgewählte Auswahlbaum repräsentiert die DIVERSet-Bibliotheken, wobei die Knoten des Baums einzelne Bibliotheksplatten darstellen. Durch Klicken auf solche Knoten wählt der Benutzer eine Platte aus, woraufhin im rechten Bereich des Fensters eine neue Karteikarte geöffnet wird, in der eine Tabelle mit den Substanzen der ausgewählten Platte dargestellt wird.

Der in Schnappschuss B gezeigte Auswahlbaum (Reiter „Substance Lists“) dient dem Zugriff auf die in den verschiedenen experimentellen Untersuchungen als wirksam gefundenen Substanzen. Dazu zeigt dieser Auswahlbaum für jede Untersuchung einen Knoten, über den die zugehörige Menge der wirksamen Substanzen für die Tabellendarstellung ausgewählt werden kann. Die Tabellen (aus A und B) werden dynamisch durch Zugriffe auf die TSE-DB mit Substanzinformationen befüllt, wobei Molekülstrukturdefinitionen aus der Tabelle SUBSTANCE und Wirksamkeitsdaten aus den in Abbildung 3.5 weiß hinterlegten Tabellen der TSE-DB ausgelesen werden. Eine hilfreiche Zusatzfunktion der Tabellendarstellungen bietet sich dem Benutzer beim Klicken auf die Spaltenüberschriften, wodurch die Substanzen nach den in der angeklickten Spalte enthaltenen Wirksamkeitswerten sortiert werden.

Der TSE-DB-Browser ist also ein nützliches Werkzeug für den Zugriff auf die in der TSE-DB abgelegten Daten. An dieser Stelle sei angemerkt, dass die TSE-DB in der Anfangsphase des Verbundprojektes, als in TP1 das SIFT-Screening der DIVERSet-Bibliotheken durchgeführt wurde, für die Haltung der komplizierten SIFT-Messdaten konzipiert wurde. Der ursprüngliche Zweck der TSE-DB war, die rohen SIFT-Messdaten an einer für TP1 und unser Teilprojekt zentralen Stelle zu halten, damit wir gemeinsam die Verfahren zur automatisierten Auswertung und Aufbereitung entwickeln und einsetzen konnten, die von dem im folgenden Abschnitt beschriebenen Werkzeug SIFT-Viewer II zur Verfügung gestellt werden. Erst später, als nach und nach auch weitere Assays in den übrigen Teilprojekten eingesetzt wurden, wurde die TSE-DB auf die Haltung dieser zusätzlichen Wirksamkeitsdaten ausgeweitet und der TSE-DB-Browser programmiert. Die TSE-DB und die beiden zugehörigen Werkzeuge dienten daher vorrangig unseren Münchner Partnern aus TP1 für die Auswertung der komplizierten SIFT-Messdaten und die Inspektion der Ergebnisse aus den übrigen Teilprojekten. Für die Haltung der Ergebnisse aus allen Teilprojekten wurde später zusätzlich ein kommerzielles Softwarepaket eingesetzt, welches ich weiter unten in Abschnitt 3.7 vorstellen werde.

3.6 Das Analysewerkzeug SIFT-Viewer II

Mit SIFT-Viewer II wurden die Fähigkeiten des in Abschnitt 3.4 beschriebenen Werkzeugs SIFT-Viewer I unter Verwendung der Sprache Java in eine saubere Form gegossen und um einige zusätzliche Fähigkeiten erweitert. SIFT-Viewer II ist wie der TSE-DB-Browser ein Bestandteil der Java-Bibliothek `tsedb` (Anhang B) und basiert, wie oben angesprochen wurde, auf darin zur Verfügung gestellten Komponenten.

Zwar existieren kommerzielle Werkzeuge zur Visualisierung und Analyse von Screening-Daten wie beispielsweise diejenigen der Firma Spotfire [135], die möglicherweise für die Darstellung und Auswertung der SIFT-Daten hätten eingesetzt werden können. Jedoch hätte die Integration der komplizierten SIFT-Daten voraussichtlich einen großen Aufwand bedeutet. Ich bevorzugte die Implementierung von SIFT-Viewer II, weil es dadurch möglich wurde, einige spezielle Funktionalitäten umzusetzen. Im Folgenden sollen die wichtigsten Fähigkeiten von SIFT-Viewer II beschrieben werden, wobei die Erweiterungen gegenüber der Vorgängerversion I im Zentrum der Aufmerksamkeit stehen werden.

3.6.1 Die Fähigkeiten von SIFT-Viewer II

Einlesen von SIFT-Dateien

Während beim Vorgängermodell Ergebnisse von SIFT-Messungen, die in Form von SIFT-Dateien vorliegen, erst noch über diverse Shell-Skripten aufgearbeitet werden mussten

(vgl. Abschnitt 3.4.1), können SIFT-Messungen mit SIFT-Viewer II direkt verarbeitet werden. Dabei können die Messdaten entweder aus einer lokalen SIFT-Datei oder aus der TSE-DB eingelesen werden – in der TSE-DB sind u.a. alle SIFT-Plattenmessungen des primären DIVERSet-Screenings zusammen mit ihrem Layout abgelegt. Werden die Messdaten aus einer SIFT-Datei eingelesen, so muss der Benutzer dabei das Layout der gemessenen Platte festlegen, damit SIFT-Viewer II etwa die Typen der Töpfchen kennt und Wirksamkeiten für Testsubstanzen berechnen kann.

Freie Wahl von SIFT-Plattenlayouts

Beim Einlesen einer SIFT-Plattenmessung kann der Benutzer entweder eines der in der TSE-DB definierten Standard-Layouts (vgl. Abschnitt 3.5.1) auswählen oder das Layout über eine Textdatei frei festlegen. In beiden Fällen wird von SIFT-Viewer II für jedes Töpfchen der Name der darin getesteten Substanz, der Typ des Töpfchens („Substanz“, „Negativ-“, „Positiv-Kontrolle“ etc.) und die Substanzkonzentration in der Assay-Mischung verwaltet.

Als Beispiel eines frei gewählten Plattenlayouts zeigt Abbildung 3.7 eine vom Benutzer festgelegte Layout-Datei für eine SIFT-Plattenmessung, bei der sechs Testsubstanzen und diverse Kontrollen in verschiedenen Konzentrationen gemessen wurden. Die gezeigte Datei enthält im ersten Abschnitt einige von SIFT-Viewer II ignorierte Kommentarzeilen, in welchen die verwendbaren Töpfchentypen beschrieben sind. Es folgt der Eintrag „ntp_code“, welcher nach der Messung von der SIFT-Software ausgegeben wird und ein Identifikationsmerkmal für die SIFT-Datei darstellt. Über diesen Code kann überprüft werden, ob SIFT-Datei und Layout-Datei zusammengehören. In der nächsten (nicht-leeren) Zeile ist die Plattengröße auf 96 Wells festgelegt – 384-Well-Platten werden ebenfalls unterstützt. Der nächste Abschnitt „substanceids“ dient der Eingabe von Substanznamen. Wie zu erkennen ist, belegen die hier gemessenen sechs Substanzen jeweils zwei Spalten (1 und 2, 3 und 4 etc.) in den Zeilen B bis H. In der Zeile A wurden Kontrollmessungen mitgeführt. Die Typen der Kontrollmessungen gehen zwar bereits aus den „Substanznamen“ hervor, werden aber tatsächlich über die Einträge des nächsten Abschnitts „welltypes“ festgelegt. Dort sind für alle Töpfchen deren Typen über die im Kommentarabschnitt beschriebenen Zahlen kodiert. Im letzten Abschnitt „dilutions“ sind wiederum töpfchenweise die Substanzkonzentrationen angegeben, welche in den Zeilen B bis H jeweils von 100 μM auf 0,1 μM abfallen.

Eine solche Layout-Datei kann vom Benutzer, beispielsweise unter Verwendung des Tabellenkalkulationsprogramms Microsoft Excel, editiert werden und muss dann als Textdatei, deren Einträge durch Tabulatorzeichen getrennt sind, abgespeichert werden.

3 Methoden zur Haltung, Auswertung und Visualisierung von Daten

```
REMARK Das ist ein Kommentar
REMARK welltype name
REMARK 0 Empty well
REMARK 1 Substance
REMARK 2 Dospa
REMARK 3 Negative (Buffer + CJD Rods + PrP)
REMARK 4 Buffer
REMARK 5 10293_G02
REMARK 6 nur CJD rods
REMARK 7 nur Mix (keine rods, keine Substanz)
REMARK 8 Pipettierfehler

ntp_code 20060518_ntp

platesize 96

substanceids
# 1 2 3 4 5 6 7 8 9 10 11 12
A Negative DOSPA Negative DOSPA Negative DOSPA Negative DOSPA Rods Rods Mix Mix
B Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04
C Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04
D Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04
E Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04
F Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04
G Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04
H Bsc2161 Bsc2161 Bsc2889 Bsc2889 Bsc3119 Bsc3119 Bsc3629 Bsc3629 10353_F11 10353_F11 10339_C04 10339_C04

welltypes
# 1 2 3 4 5 6 7 8 9 10 11 12
A 3 2 3 2 3 2 3 2 6 6 7 7
B 1 1 1 1 1 1 1 1 1 1 1 1
C 1 1 1 1 1 1 1 1 1 1 1 1
D 1 1 1 1 1 1 1 1 1 1 1 1
E 1 1 1 1 1 1 1 1 1 1 1 1
F 1 1 1 1 1 1 1 1 1 1 1 1
G 1 1 1 1 1 1 1 1 1 1 1 1
H 1 1 1 1 1 1 1 1 1 1 1 1

dilutions
# 1 2 3 4 5 6 7 8 9 10 11 12
A 0 17 0 17 0 17 0 17 0 0 0 0
B 100 100 100 100 100 100 100 100 100 100 100 100
C 31,6 31,6 31,6 31,6 31,6 31,6 31,6 31,6 31,6 31,6 31,6 31,6
D 10 10 10 10 10 10 10 10 10 10 10 10
E 3,16 3,16 3,16 3,16 3,16 3,16 3,16 3,16 3,16 3,16 3,16 3,16
F 1 1 1 1 1 1 1 1 1 1 1 1
G 0,316 0,316 0,316 0,316 0,316 0,316 0,316 0,316 0,316 0,316 0,316 0,316
H 0,1 0,1 0,1 0,1 0,1 0,1 0,1 0,1 0,1 0,1 0,1 0,1
```

Abbildung 3.7: Exemplarische Layout-Datei für eine Messplatte. SIFT-Viewer II ermöglicht dem Benutzer, das Layout einer gegebenen Plattenmessung über eine Layout-Datei zu definieren, wobei Substanznamen, Töpfchentypen und Verdünnungen festgelegt werden. Nähere Erläuterungen des Formats finden sich im Text.

Wählbare Methoden zur Ausreißererkenung

Wie sein Vorgänger unterstützt auch SIFT-Viewer II die Erkennung von Ausreißern. Hier ist dem Benutzer jedoch die Möglichkeit gegeben, zur Laufzeit des Programms aus mehreren Methoden zur Ausreißererkenung per Mausklick eine gewünschte auszuwählen. Dabei stehen die in Abschnitt 3.1 beschriebenen Erkennungsmethoden und einige weitere darauf basierende Verfahren zur Auswahl, wobei die Methode nach Grubbs aus Abschnitt 3.1.2 von SIFT-Viewer II vorausgewählt ist. SIFT-Viewer II erkennt Ausreißer gemäß der ausgewählten Methode und markiert die betroffenen Töpfchen. Über die weiter unten beschriebene grafische Benutzerschnittstelle ist dem Benutzer zudem die Möglichkeit gegeben, händisch weitere Töpfchen als Ausreißer zu markieren oder die Markierung von bereits markierten Töpfchen zu entfernen.

Berechnung von Wirksamkeiten

Wie bei der Ausreißererkenkung stehen dem Benutzer in SIFT-Viewer II verschiedene Methoden zur Berechnung von Wirksamkeiten (vgl. Abschnitt 3.2) zur Auswahl. Nach der vorausgewählten oder vom Benutzer per Mausklick ausgewählten Methode werden die Wirksamkeiten von SIFT-Viewer II zur Laufzeit berechnet. Zusammen mit den im letzten Abschnitt angesprochenen Möglichkeiten, kann der Benutzer beispielsweise fragwürdige Kontrollmessungen händisch als Ausreißer markieren und daraufhin die Neuberechnung der Wirksamkeiten anstoßen, falls diese von den Kontrolldaten abhängen.

Töpfchenweise Repräsentationen von Messplatten

Von SIFT-Viewer I wurden Messplatten durch eine Tabelle repräsentiert, deren Einträge über die Töpfchennamen benannt sind und dem Benutzer Zugriff auf die töpfchenweisen Messdaten bieten (vgl. Abbildung 3.4). In SIFT-Viewer II wurde diese einfache Repräsentation von Messplatten durch zwei weiterentwickelte Repräsentationen ersetzt, welche dem Benutzer einige zusätzliche Visualisierungen und Auswahlmöglichkeiten bieten.

Als Vorgriff auf den weiter unten in Abbildung 3.11 gezeigten Bildschirmschnappschuss von SIFT-Viewer II zeigt Abbildung 3.8 zwei Bestandteile der graphischen Benutzerschnittstelle, nämlich die beiden angesprochenen Repräsentationen einer SIFT-Messplatte. Die gezeigten Repräsentationen wurden von SIFT-Viewer II anhand der in Abbildung 3.7 dargestellten Layout-Datei und der zugehörigen SIFT-Datei gebildet. Die 96 Töpfchen der Messplatte sind hier durch quadratische, per Mausklick eindrückbare bzw. rückstellbare Knöpfe repräsentiert. Dabei sind die Knöpfe der Layout-Repräsentation in **A** gemäß den auf der Platte vorhandenen Töpfchentypen eingefärbt. Die Repräsentation in **B** hingegen zeigt zum einen die Ergebnisse der Ausreißererkenkung, wobei diejenigen Knöpfe, deren zugrunde liegende Töpfchen als Ausreißer erkannt wurden, durch einen weißen Querbalken als durchgestrichen markiert sind. Zum anderen sind in **B** die Ergebnisse der Wirksamkeitsberechnung durch die Farben der Knöpfe kodiert. Anhand der hierbei verwendeten Farbskala wurden die Töpfchenrepräsentationen von unwirksamen Substanzen in roten und diejenigen von wirksamen Substanzen in grünen Tönen eingefärbt.

Beide Repräsentationen (**A** und **B**) dienen als Bestandteile von SIFT-Viewer II der Auswahl von einzelnen Töpfchen oder von bestimmten Mengen von Töpfchen per Mausklick. Für die Auswahl von Töpfchenmengen existieren töpfchenabhängige Kontextmenüs, wie das in **B** gezeigte, welche der Benutzer durch Klicken der rechten Maustaste über einem gewählten Töpfchenknopf aufrufen kann. Aus einem solchen Menü können dann verschiedene vordefinierte Töpfchenmengen per Mausklick ausgewählt oder bereits ausgewählte wieder abgewählt werden. Im Falle der gezeigten Platte mit Verdünnungsreihen sind beispielsweise diejenigen Töpfchen auswählbar, in welchen eine gegebene Substanz (in verschiedenen Konzentrationen) gemessen wurde. In **B** wurden auf diese Weise die

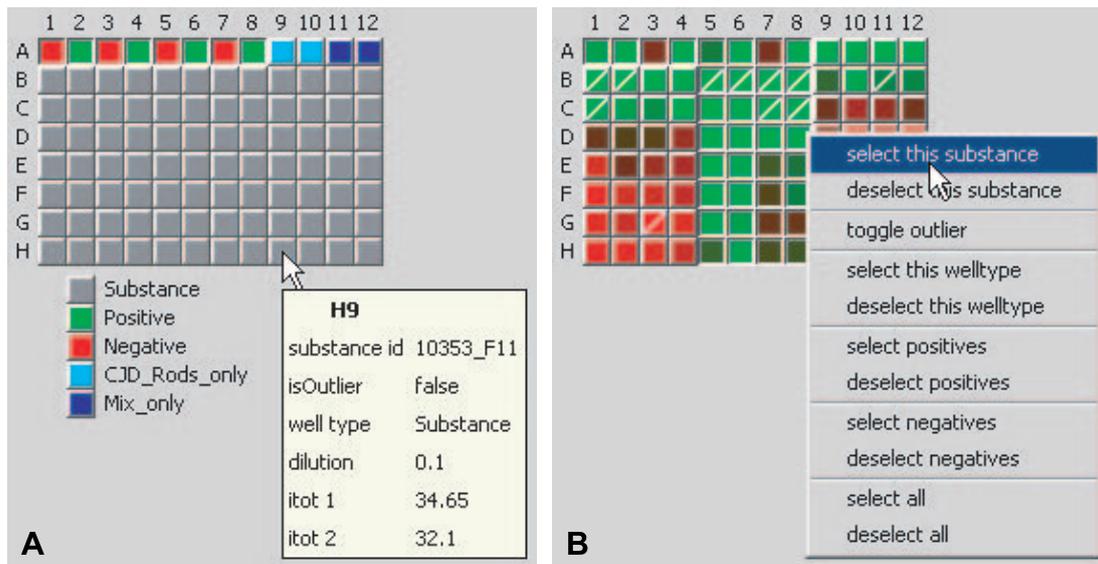


Abbildung 3.8: Töpfchenweise Repräsentationen von SIFT-Messplatten. Dargestellt sind zwei von SIFT-Viewer II verwendete grafische Komponenten zur Repräsentation von SIFT-Messplatten, wobei jeweils die Töpfchen durch anklickbare Knöpfe symbolisiert sind. In **A** sind die Knöpfe nach dem Töpfchentyp eingefärbt, in **B** nach der SIFT-Wirksamkeit. Hierbei sind wirksame Substanzen durch grüne und unwirksame Substanzen durch rote Farbtöne gekennzeichnet. Zudem sind Ausreißer mit einem weißen Balken als durchgestrichen gekennzeichnet. Darüber hinaus bieten beide Plattenrepräsentationen töpfchenweise Tooltips (**A**) zur Anzeige von töpfchenspezifischen Informationen und Kontextmenüs (**B**) zur Auswahl von Substanzmengen.

zu zwei Substanzen gehörigen Töpfchen in den Spalten 5 und 6 bzw. 7 und 8 ausgewählt. Über die Kontextmenüs hinaus enthalten die Töpfchenrepräsentationen sogenannte Tooltips wie den in **A** gezeigten, welche dem Benutzer diverse töpfchenspezifische Informationen anzeigen, sobald der Mauszeiger eine Weile auf ein Töpfchen deutet.

Darstellung von Grafiken

Ähnlich wie der Vorgänger stellt auch SIFT-Viewer II Graphen der Summenobservablen, der Gesamtintensitäten und der Wirksamkeiten sowie der 2D-SIFT-Spektren dar (vgl. den Schnappschuss von SIFT-Viewer I in Abbildung 3.4). Anstelle von statischen Grafikdateien werden hier jedoch dynamische Charts der frei verfügbaren Java-Bibliothek JFreeChart [136] zur Darstellung eingesetzt (vgl. Anhang B). Zudem werden im Falle der Spektren von dem entsprechenden dynamischen Chart nur diejenigen Spektren angezeigt, die zu den vom Benutzer über die Plattenrepräsentationen ausgewählten Töpfchen gehören. Darüber hinaus bieten die Charts dem Benutzer zusätzliche Funktionen wie beispielsweise, per Maus Bereiche eines gegebenen Plots auszuwählen und vergrößert darzustellen.

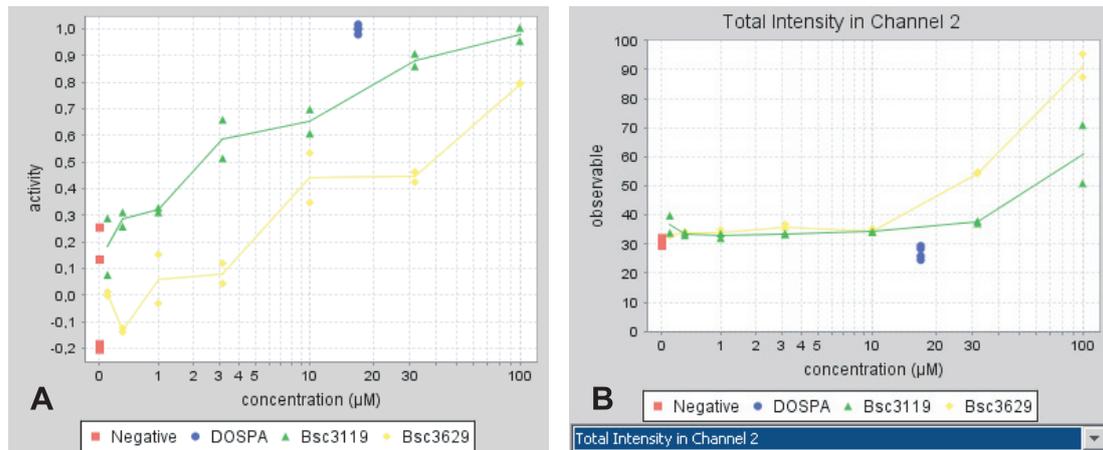


Abbildung 3.9: Dosis-Wirkungs- und Dosis-Gesamtintensitäts-Kurven. Gezeigt sind zwei Charts mit Dosis-Wirkungs-Kurven (**A**) und Dosis-Gesamtintensitäts-Kurven (**B**), die von SIFT-Viewer II für die vom Benutzer ausgewählten Töpfchen generiert wurden. Der Chart in **A** zeigt zum einen die Wirksamkeitswerte der Töpfchen mit negativen und positiven (DOSPA)-Kontrollmessungen, die als rote bzw. blaue Symbole bei den Konzentrationen 0 bzw. 17 μM eingezeichnet sind, und zum anderen die Wirksamkeiten der beiden Substanzen Bsc3119 und Bsc3629 bei den diversen gemessenen Konzentrationen (grüne bzw. gelbe Symbole und Linien). In **B** sind in analoger Weise die Werte der Gesamtintensität $I_{\text{tot,r}}$ (in der Einheit kHz) im roten Spektralbereich („Channel 2“) gegen die Konzentration aufgetragen.

Im Falle von Verdünnungsreihenmessungen wird von SIFT-Viewer II ein zusätzlicher Chart mit Dosis-Wirkungs-Kurven generiert. Abbildung 3.9 **A** zeigt einen solchen Chart, der für die Messplatte aus Abbildung 3.8 anhand derjenigen Töpfchen generiert wurde, die in der Plattenrepräsentation aus Teil **B** ausgewählt sind. In diesem Chart sind die von SIFT-Viewer II berechneten Wirksamkeitswerte für die ausgewählten Töpfchen gegen die Konzentration aufgetragen. Dabei sind zum einen die Wirksamkeitswerte der negativen und positiven (DOSPA)-Kontrollmessungen als rote bzw. blaue Symbole bei den Konzentrationen 0 bzw. 17 μM eingezeichnet. Zum anderen sind die Werte für die beiden ausgewählten Substanzen Bsc3119 und Bsc3629 bei den verschiedenen Konzentrationen durch grüne bzw. gelbe Symbole repräsentiert. Da auf der gewählten Platte die Substanzen bei jeder Konzentration zweifach gemessen wurden, werden von SIFT-Viewer II jeweils die Mittelwerte der Wirksamkeiten gebildet und durch Linien verbunden dargestellt. Anhand solcher Dosis-Wirkungs-Kurven können vom Benutzer Konzentrationsabhängigkeiten der Wirksamkeiten analysiert und EC50-Werte abgelesen werden.

In Abbildung 3.9 **B** ist ein weiterer Chart gezeigt, in welchem in analoger Weise die Gesamtintensitäten $I_{\text{tot,r}}$ im roten Spektralbereich („Channel 2“) der ausgewählten Töpfchen gegen die Konzentration aufgetragen sind. Über das Drop-Down-Menü im unteren Bereich dieses Charts können alternativ die grünen Gesamtintensitäten $I_{\text{tot,g}}$ für die Darstellung ausgewählt werden. Die dargestellten Dosis-Gesamtintensitäts-Kurven dienen dazu, Fluoreszenzartefakte, die etwa durch Autofluoreszenz oder Quenching der Testsubstanzen

verursacht werden können, zu inspizieren. Im gezeigten Beispiel zeigen beide Kurven erhöhte Werte der roten Gesamtintensitäten für hohe Konzentrationen und weisen dadurch auf Autofluoreszenzen der beiden Testsubstanzen hin. Es sei angemerkt, dass diese stark erhöhten Intensitätswerte bereits von der automatischen Detektion als Ausreißer erkannt wurden und demzufolge die zugehörigen Töpfchenrepräsentationen in Abbildung 3.8 B als durchgestrichen markiert sind.

Darstellung von Molekülstrukturen

Auch die Darstellung von Molekülstrukturen geschieht mit SIFT-Viewer II dynamisch. Falls eine Verbindung zur TSE-DB besteht, holt SIFT-Viewer II aus der Tabelle SUBSTANCE anhand der Substanznamen der auf der betrachteten Platte enthaltenen Substanzen die zugehörigen SMILES-Strings und erzeugt daraus Bilder der Molekülstrukturen. Dabei werden Funktionalitäten der frei verfügbaren Java-Bibliothek CDK [137] eingesetzt (vgl. Anhang B).

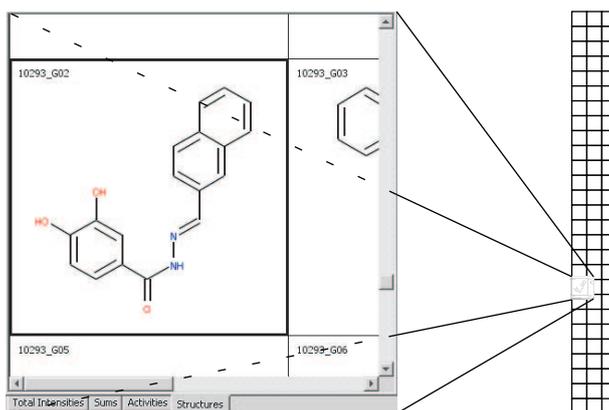


Abbildung 3.10: Von SIFT-Viewer II dargestellte Molekülstruktur. Die Molekülstruktur der vom Benutzer zuletzt ausgewählten Substanz wird von SIFT-Viewer II in den sichtbaren Bereich geholt und fett umrandet. Das Raster im rechten Bereich stellt schematisch die Anordnung aller Substanzstrukturen der betrachteten Messplatte dar. Im hier exemplarisch gezeigten Fall sind dort 80 Strukturen in Zeilen à drei Strukturen angeordnet.

Wie im rechten Bereich von Abbildung 3.10 schematisch dargestellt ist, verteilt SIFT-Viewer II die Bilder der Molekülstrukturen auf ein Raster mit jeweils drei Bildern pro Zeile. Da die Zahl der Substanzen auf einer Platte groß sein kann – im Falle des DIVERSet-Screenings wurden jeweils 80 Substanzen auf einer 96-Well-Platte gemessen – und der Darstellungsbereich begrenzt ist, kann nur ein (kleiner) Ausschnitt des Rasters angezeigt werden (vgl. Abbildung 3.10 links). Wählt der Benutzer über die oben beschriebenen Töpfchenrepräsentationen eine Substanz aus, so wird dieser Ausschnitt so verschoben, dass sich das Bild der zugehörigen Struktur im Ausschnitt befindet. Zudem wird dieses Bild fett umrandet dargestellt.

Generierung von Reports

Um die von SIFT-Viewer II erzeugten Visualisierungen für die spätere Betrachtung und Diskussion aufzubewahren, bietet SIFT-Viewer II die Möglichkeit, einen Report der betrachteten Plattenmessung in Form einer pdf-Datei (Portable Document Format [138]) abzuspeichern. Bei der Generierung eines solchen Reports werden von SIFT-Viewer II Schnappschüsse der diversen oben beschriebenen grafischen Komponenten erzeugt und gemeinsam mit weiteren Informationen in einer pdf-Datei zusammengestellt. SIFT-Viewer II verwendet dabei die frei verfügbare Java-Bibliothek iText [139] (vgl. Anhang B).

Abschließend sei an dieser Stelle noch auf die in SIFT-Viewer II umgesetzten Verfahren zur selbstorganisierten Klassifikation von SIFT-Messdaten verwiesen, welche in Anhang C beschrieben sind. Nachdem damit die wichtigsten Fähigkeiten von SIFT-Viewer II aufgezählt wurden, sollen nun weitere Einblicke in dessen Bedienung gegeben werden.

3.6.2 Einblicke in die Bedienung von SIFT-Viewer II

Abbildung 3.11 zeigt einen Bildschirmschnappschuss von SIFT-Viewer II, in welchem die SIFT-Screening-Ergebnisse für eine Bibliotheksplatte aus DIVERSet 1 angezeigt sind. Im linken Bereich der Darstellung ist eine Baumstruktur zu erkennen, aus welcher einzelne SIFT-Plattenmessungen für die Darstellung im rechten Bereich gewählt werden können. In der Baumstruktur sind zum einen die vom Benutzer aus SIFT-Dateien geladenen Plattenmessungen unterhalb des Knotens „from files“ und zum anderen die in der TSE-DB verfügbaren Plattenmessungen („from database“) zugreifbar.

Über diesen Baum wurde die Platte 10264 aus DIVERSet 1 ausgewählt, wurden daraufhin die zugehörigen Messdaten aus der TSE-DB eingelesen und im rechten Bereich innerhalb einer neu angelegten Karteikarte angezeigt. Im oberen Bereich dieser Karteikarte sind die Summenobservablen dargestellt, weil dieser Plot über den darunter befindlichen Reiter „Sums“ angewählt wurde. Hinter dieser Karteikarte liegen die Plots der Gesamtintensitäten und der Wirksamkeiten sowie Darstellungen der Molekülstrukturen (vgl. Abbildung 3.10) und einer Substanztablette verborgen und können durch Anklicken der übrigen Reiter in den Vordergrund gebracht werden. Im linken unteren Bereich befinden sich die verschiedenen bereits in Abbildung 3.8 erläuterten Plattenrepräsentationen, zwischen denen wiederum über Reiter gewechselt werden kann. Im Schnappschuss ist jene Plattenrepräsentation im Vordergrund, deren Knöpfe nach Wirksamkeiten eingefärbt sind. Durch Drücken der Knöpfe können die zugehörigen Töpfchen ausgewählt werden. Die Knöpfe der positiven und negativen Kontrolltöpfchen (A1, C1 und E1 bzw. B1, D1 und F1) werden von SIFT-Viewer II beim Öffnen der Platte bereits voreingestellt gedrückt. Weil der Mauszeiger über dem vom Benutzer gedrückten Töpfchenknopf „C03“ steht, wird der zugehörige Tooltip mit den töpfchenspezifischen Informationen angezeigt.

3 Methoden zur Haltung, Auswertung und Visualisierung von Daten

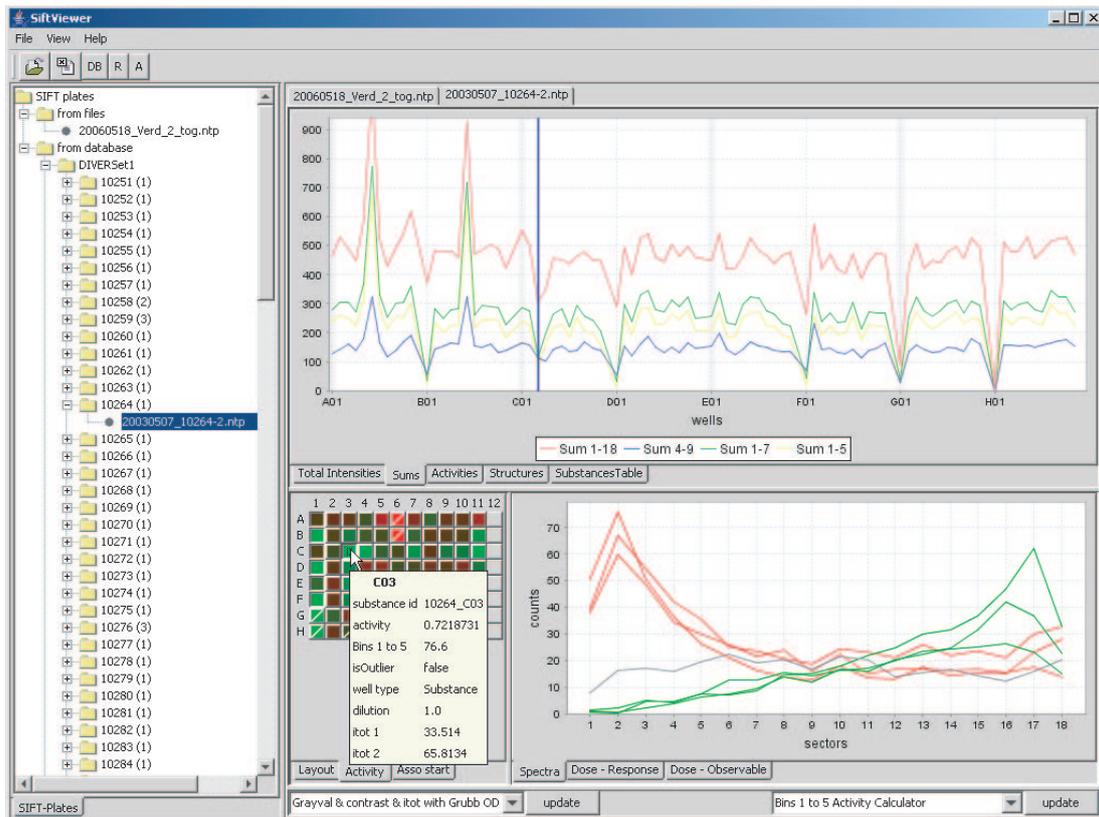


Abbildung 3.11: Bildschrimschnappschuss von SIFT-Viewer II. Wie sein Vorgänger dient SIFT-Viewer II der Analyse von SIFT-Daten. Im linken Bereich sind einzelne Plattenmessungen aufgelistet, die aus dem Dateisystem oder aus der TSE-DB eingelesen werden können. Im rechten Bereich sind die Daten der geöffneten Messplatte (Nr. 10264 aus DIVERSet 1) dargestellt. Dort ist im unteren linken Teil die Platte in Form von einer Matrix aus Knöpfen repräsentiert. Die Knöpfe zeigen durch Farbkodierung die anti-Prion-Aktivität der zugehörigen Substanzen an (grün: aktiv, rot: inaktiv). Ähnlich wie bei SIFT-Viewer I sind im oberen Bereich Summenobservablen und im unteren rechten Bereich 2D-SIFT-Spektren dargestellt.

Im rechten unteren Bereich sind die 2D-SIFT-Spektren für die ausgewählten Töpfchen geplottet, wobei negative Kontrollspektren rot, positive Kontrollspektren grün und Substanzspektren grau eingefärbt sind. Dabei verwendet SIFT-Viewer II dieselbe Farbkodierung wie in der Layout-orientierten Plattenrepräsentation, die hier verborgen ist, aber in Abbildung 3.8A eingesehen werden kann. Über die Karteikartenreiter unterhalb des Spektrplots können Dosis-Wirkungs-Kurven oder Dosis-Gesamtintensitäts-Kurven (Reiter „Dose - Observable“), wie jene aus Abbildung 3.9, angezeigt werden. Weil bei der gezeigten Primärscreeningmessung jede Substanz nur bei *einer* Konzentration gemessen wurde, bieten diese Darstellungen hier nur eingeschränkten Nutzen.

Unterhalb der beschriebenen Darstellungen befinden sich Drop-Down-Menüs, aus welchen, wie oben beschrieben wurde, die Methoden zur Ausreißerererkennung und zur Wirk-

samkeitsberechnung ausgewählt werden können. Über die zugehörigen „update“-Knöpfe wird die gewählte Prozessierung angestoßen.

Die Einträge des Menüs und die Symbolknöpfe (links oben) stellen dem Benutzer zusätzliche Bedienmöglichkeiten zur Verfügung. Dazu gehören ein Dialog zur Auswahl einer SIFT- und der zugehörigen Layout-Datei sowie ein Dialog zur Herstellung der Verbindung mit der TSE-DB (Symbolknopf „DB“). Des Weiteren können die Reportgenerierung (Knopf „R“) oder der Export von Aktivitätswerten (Knopf „A“) aufgerufen werden.

SIFT-Viewer II wurde von meinen Kollegen aus TP1 dazu eingesetzt, die SIFT-Daten zu analysieren, die beim Screening und der Validierung der DIVERSet-Bibliotheken und einiger fokussierter Bibliotheken gewonnen wurden. Wie zum Ende von Abschnitt 3.5 bereits angesprochen wurde, war in der ersten Projektphase des Screenings auch das zugehörige Werkzeug TSE-DB-Browser entstanden, mit welchem unsere Münchner Verbundpartner aus TP1 die Ergebnisse der übrigen Teilprojekte, d.h. die in der TSE-DB abgelegten Wirksamkeitsdaten, inspizieren können. Weil jedoch der Zugriff auf den vom LRZ betriebenen Datenbankserver, auf welchem die TSE-DB installiert ist, auf Angehörige des Münchner Hochschulnetzes beschränkt ist, können unsere auswärtigen Verbundpartner nicht auf die TSE-DB zugreifen. Um dennoch allen Verbundpartnern Zugriff auf die im Verbund erzielten Resultate zu geben, habe ich die Verwaltung der Substanz- und Wirksamkeitsdaten auf die kommerzielle Software „ChemFinder“ umgestellt. Dankenswerterweise hat mich Prof. Schmidt (TP6) zu Beginn der zweiten Förderperiode (Jahresmitte 2005) auf dieses nützliche Softwarepaket aufmerksam gemacht, dessen Einsatz ich im Folgenden vorstellen möchte. Es sei jedoch noch angemerkt, dass die TSE-DB und SIFT-Viewer II nach wie vor von TP1 und mir für die Analyse von SIFT-Messdaten genutzt werden.

3.7 Alternative Datenverwaltung mit ChemFinder

Die Software ChemFinder der Firma CambridgeSoft [133] dient der Verwaltung von Substanzbibliotheken und wurde in der Vollversion bei mir und in einer kostenlosen, auf lesenden Zugriff begrenzten Version bei unseren Verbundpartnern installiert. Gegenüber der Datenverwaltung mit der TSE-DB bietet ChemFinder den Vorteil, dass die damit verwalteten Bibliotheken in Form von Dateien gehalten werden, sodass ich die Bibliotheken regelmäßig aktualisieren und an unsere Verbundpartner verteilen konnte. Zu den von mir mithilfe von ChemFinder verwalteten Substanzbibliotheken zählen die DIVERSet-Sammlungen und einige fokussierte Bibliotheken, die im Labor von Prof. Boris Schmidt (TP6) oder im Rahmen einer externen Kooperation mit der Arbeitsgruppe von Prof. Christian Griesinger (MPI für biophysikalische Chemie, Göttingen) synthetisiert wurden.

Im Falle der DIVERSet-Bibliotheken konnte ich zur Installation eines virtuellen Abbildes dieser Substanzsammlungen die vom Hersteller mitgelieferten sdf-Dateien mit den Substanzbeschreibungen (vgl. Abschnitt 3.3) direkt in ChemFinder importieren und auf

3 Methoden zur Haltung, Auswertung und Visualisierung von Daten

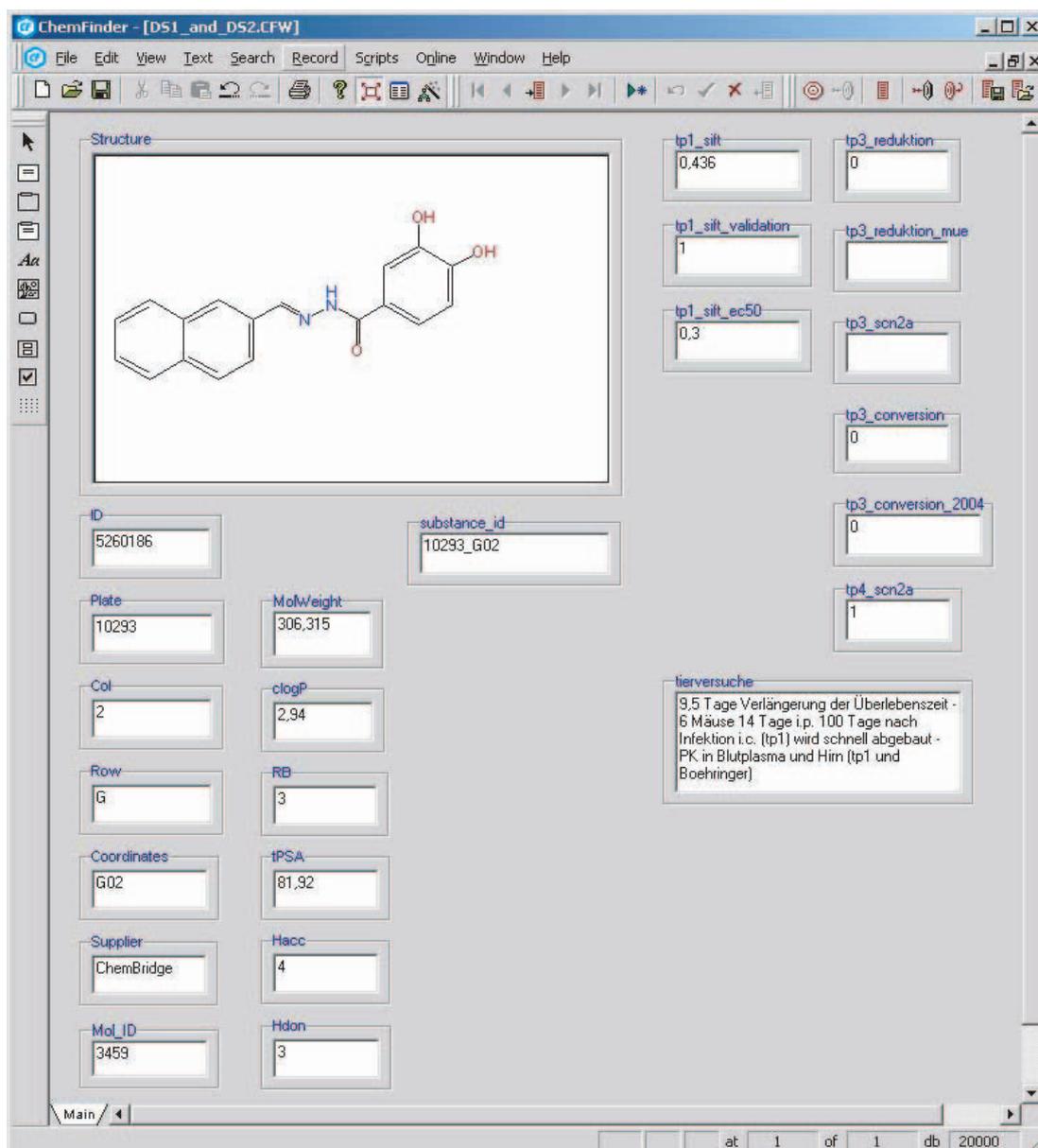


Abbildung 3.12: Bildschirmschnappschuss der Software ChemFinder. Für den gezeigten Schnappschuss wurde die Datenbank mit allen 20.000 Substanzen aus DIVERSet 1 und 2 geladen. Für eine ausgewählte DIVERSet 1-Substanz (10293_G02) sind die Molekülstruktur, physikochemische Eigenschaften und Wirksamkeiten aus den verschiedenen experimentellen Untersuchungen angezeigt. Mit ChemFinder wurden sowohl die DIVERSet-Bibliotheken als auch einige fokussierte Bibliotheken verwaltet.

diese Weise eine Datenbank erzeugen, in welcher für jede Substanz die Molekülstruktur und die weiteren Daten wie die physikochemischen Eigenschaften abgelegt sind. Um die Wirksamkeitsdaten aus den verschiedenen experimentellen Untersuchungen als zusätzli-

che Felder der Datenbank hinzuzufügen, bietet ChemFinder die Möglichkeit solche Daten aus geeignet formatierten Textdateien zu importieren, wobei direkt die zum Befüllen der TSE-DB eingesetzten Textdateien verwendet werden konnten. Es waren dazu lediglich die durch das Format (3.11) definierten Substanznamen aus den sdf-Daten zu generieren, da diese Namen, wie oben erklärt wurde, die zu den Substanzen gehörigen Versuchsergebnisse in den Textdateien markieren.

Die grafische Oberfläche von ChemFinder zeigt dann, wie Abbildung 3.12 ausweist, im linken Bereich die sdf-Daten sowie im Feld „substance_id“ den Substanznamen und im rechten Bereich die aus den Textdateien importierten Versuchsergebnisse^a. Beispielsweise ist dort im Feld „tp1_sift“ die aus dem primären Screening abgeleitete Wirksamkeit (3.9) der abgebildeten Substanz aufgeführt. ChemFinder bietet darüber hinaus die Möglichkeit weitere Textdaten zu Substanzen abzulegen und anzuzeigen. Dies war speziell für die Speicherung der Ergebnisse von Tierversuchen von Vorteil, die im späteren Verlauf des Projekts für einige wenige ausgesuchte Substanzen durchgeführt wurden. In Abbildung 3.12 zeigt das Feld „tierversuche“ ein entsprechendes Versuchsergebnis an, das händisch eingegeben wurde.

Im Falle der fokussierten Bibliotheken mit neu synthetisierten Substanzen mussten die Molekülstrukturen erst noch eingegeben werden. Diese Aufgabe wurde teilweise von den Chemikern und teilweise von mir übernommen. Für die Eingabe der Molekülstrukturen verwendete ich das Programm ChemDraw, welches wie ChemFinder zum ChemOffice-Paket [133] gehört. Mit den geschilderten Mitteln wurden von mir regelmäßig die neu hinzugekommenen Messergebnisse in die ChemFinder-Datenbanken aufgenommen und diese an unsere Partnergruppen verteilt.

ChemFinder bietet einige hilfreiche Fähigkeiten wie die Möglichkeit, Suchen nach chemischen Strukturmotiven durchzuführen oder Listen von Substanzen auszudrucken. ChemFinder löste daher die Verwaltung der Wirksamkeitsdaten mit der TSE-DB ab. Für die Verwaltung und Analyse von SIFT-Messdaten mithilfe von SIFT-Viewer II wurde aber die TSE-DB weiterhin genutzt.

Nachdem nun die von mir zur Verwaltung und Auswertung von Substanz-, Mess- und Wirksamkeitsdaten entwickelten oder eingesetzten Softwarewerkzeuge vorgestellt wurden, sollen im nächsten Kapitel einige weitere Werkzeuge vorgestellt werden, die zur Untersuchung von Beziehungen zwischen chemischen Strukturen und Wirksamkeiten dienen.

^aDie Namen der Felder sind gegenüber denjenigen der entsprechenden TSE-DB-Tabellen leicht abgewandelt.

4 Methoden zur Untersuchung von Struktur-Wirkungs-Beziehungen

Das vorangegangene Kapitel hat die Anstrengungen erläutert, die von mir während der Laufzeit des Verbundprojektes unternommen wurden, um die von meinen Partnern erhobenen Daten zentral zu sammeln, ihre Qualität zu sichern und einen schnellen visuellen Zugriff darauf zu ermöglichen. Da diese Daten Informationen über die anti-Prion-Wirksamkeit einer Vielzahl chemischer Substanzen unterschiedlichen Aufbaus enthalten, können sie genutzt werden, um allgemeine „Struktur-Wirkungs-Beziehungen“ (SAR, structure-activity relationships) [140, 141] abzuleiten. Solchen SAR-Untersuchungen liegt die Vorstellung zugrunde, dass Substanzen, die ähnliche chemische Grundstrukturen besitzen, auch ähnliche biologische Wirksamkeiten zeigen. Daher zielen SAR-Auswertungen von HTS-Daten auf die Identifikation sogenannter chemischer „Leitstrukturmotive“, welche Klassen von wirksamen Substanzen charakterisieren. Sobald eine Leitstruktur identifiziert ist, kann durch systematische Variation von Substituenten versucht werden, die biologische Wirksamkeit und die pharmakologischen Eigenschaften zu optimieren.

Der soeben skizzierte Ansatz liegt häufig den Forschungsanstrengungen zugrunde, die heutzutage von Pharmakonzernen unternommen werden, um neuartige Medikamente zu entwickeln. Entsprechend stellte sich unserem Verbund die Aufgabe, die in Pharmakonzernen üblichen SAR-Analysen von HTS-Daten für unsere Zwecke, d.h. für die Entwicklung von neuen anti-Prion-Wirkstoffen, nachzubilden. Dazu konnte aber nicht auf die in den Pharmakonzernen wie Roche oder Bayer entwickelten und eingesetzten Verfahren zur rechnergestützten SAR-Analyse zurückgegriffen werden, da diese Verfahren geschütztes Eigentum der jeweiligen Firmen sind. Vielmehr mussten wir uns am (keineswegs vollständigen) Stand der Literatur und an kommerziell bzw. im Netz verfügbaren Programmpaketen orientieren, um so eine eigenständige Lösung zur SAR-Auswertung der im Verbund gewonnenen HTS-Daten zu entwickeln.

In publizierten SAR-Untersuchungen (siehe beispielsweise [79] für eine SAR-Studie von Akridin-Derivaten als anti-Prion-Therapeutika) werden in der Regel relativ kleine Anzahlen (~ 10) von Substanzen aus einer *gegebenen* strukturellen Klasse analysiert, die, wie oben dargestellt wurde, durch eine gemeinsame chemische Leitstruktur charakterisiert ist. Dabei werden die Einflüsse von strukturellen Variationen auf die biologischen Wirksamkeiten der betrachteten Substanzen diskutiert, wobei die Wirksamkeiten üblicherweise durch aufwändige Experimente wie beispielsweise Verdünnungsreihen in Zellkulturen oder Tierversuche relativ genau bestimmt sind. Nachdem es uns im Verbund gelungen war, durch visuellen Vergleich der SIFT- und Zellkulturergebnisse aus TP1 und TP4 eine erste

Hypothese für eine Leitstruktur einer anti-Prion-Wirkstoffklasse aufzustellen (siehe Kapitel 5), und sich gezeigt hatte, dass in den DIVERSet-Bibliotheken einige hundert Vertreter dieser potenziellen Wirkstoffklasse vorhanden sind, stellte sich zunächst die Frage, ob es möglich sein würde, SAR-Untersuchungen von solch großen Anzahlen (~ 500) strukturell ähnlicher Substanzen durchzuführen, die durch vergleichsweise ungenau bestimmte SIFT-Wirksamkeiten aus dem primären Screening beschrieben sind. Dabei ging es insbesondere um die Frage, ob die durch die mehr oder weniger zufällig gefundene Leitstruktur definierte Substanzklasse Teilklassen enthält, die zu bestimmten Derivaten der Leitstruktur gehören und durch signifikant große Wirksamkeiten ausgezeichnet sind. Findet man nämlich solche Teilklassen, so ist dadurch gezeigt, dass die ursprünglich gefundene Leitstruktur den Kern eines potenziellen Pharmazeutikums bildet. Man sagt dann, dass durch diesen Prozess der SAR-Analyse von Subklassen die ursprünglich gefundene Leitstruktur erst als solche „validiert“ wird. Die von mir zur Beantwortung der hier aufgeworfenen Fragen eingesetzten bzw. entwickelten Werkzeuge werde ich in Abschnitt 4.1 beschreiben.

Über die oben angesprochenen Fragen hinaus stellt sich hier die allgemeinere Frage, ob das Auffinden neuer potenzieller Leitstrukturen, wie in unserem Fall zunächst geschehen, auf Zufälle und glückliche Umstände angewiesen ist oder ob es nicht doch vielmehr möglich ist, solche Leitstrukturen systematisch aus gesammelten HTS-Daten zu identifizieren. Liest man die Werbeprospekte von kommerziell verfügbaren Softwarepaketen zur pharmakologischen Datenanalyse, so kann man meinen, dass dazu eine Vielzahl von Lösungen einkaufbar sind. Allerdings stellte sich heraus, dass ein erstes für diese Zwecke eingekauftes Paket (Catalyst, Firma Accelrys [142]) die erhoffte Lösung nicht brachte. Erst drei Jahre nach Beginn des Projekts, erhielt ich durch Kontakte zur Firma Evotec Technologies [105], die Pharmaforschung im Auftrag von Firmen wie Roche betreibt, Hinweise auf das neu auf dem Markt erschienene Softwarepaket „Benchmark HTS DataMiner“ der Firma Tripos [143], welches sich tatsächlich als wirksames Instrument zur Identifikation von potenziellen Leitstrukturen aus HTS-Daten erwies. In Abschnitt 4.2 werde ich dieses Werkzeug vorstellen. In jüngster Vergangenheit konnte ich darüber hinaus zeigen, dass eines der in Abschnitt 4.1 beschriebenen Werkzeuge zur Validierung von Leitstrukturen zusätzlich zur Identifikation von neuen Leitstrukturen eingesetzt werden kann.

Hat man schließlich wirksame Substanzklassen identifiziert, so kann man sich die Anschlussfrage stellen, welche molekularen Mechanismen die Wirksamkeit der Substanzen einer gegebenen Klasse vermitteln. Im Falle von anti-aggregatorischen Wirkstoffen ist man dabei allerdings mit dem Problem konfrontiert, dass der Angriffspunkt der Substanzen experimentell bislang nicht bestimmt werden kann. Solche Substanzen können entweder an die PrP^C-Monomere andocken und so deren Koaggregation an die PrP^{Sc}-Fibrillen verhindern oder sie können direkt und mit demselben Resultat an die Fibrillen binden. Trotz dieser Unsicherheit stellte sich mir im Verlauf des Projekts die Aufgabe, *in silico* Docking-Experimente durchzuführen. Die dazu verwendeten Methoden werden in Abschnitt 4.3 skizziert.

4.1 Identifikation und Validierung von Leitstrukturen

Bei Recherchen nach einem brauchbaren Softwarewerkzeug zur Identifikation und Validierung von Leitstrukturen anhand von HTS-Daten stieß ich im Internet auf das Programmpaket „Molecular Substructure Miner“ (MoSS) [144], welches von Informatikern der Universität Magdeburg in Kooperation mit der Firma Tripos [143] für das Auffinden von charakteristischen chemischen Strukturmotiven in Substanzmengen entwickelt wurde [145–149].

4.1.1 Die Bildung von Strukturklassen mit MoSS

Der Ausgangspunkt für die Bildung von Strukturklassen mithilfe von MoSS ist eine Bibliothek

$$\mathcal{B} = \{C_i \mid i = 1, \dots, N\}, \quad (4.1)$$

in welcher die N Molekülstrukturen C_i der Substanzen (compounds) einer zugehörigen realen Sammlung zusammengefasst sind. Optional können die Substanzen zusätzlich durch biologische Wirksamkeiten beschrieben sein. Eine solche Bibliothek \mathcal{B} und ein vom Benutzer gewähltes strukturelles Kernmotiv K bilden die Eingabe für MoSS.

Abbildung 4.1 zeigt schematisch für ein exemplarisches Ausgangskernmotiv K (oben), wie MoSS schrittweise das Kernmotiv erweitert, um strukturelle Klassen zu bilden. Im ersten Schritt versucht MoSS dazu, das Kernmotiv K in die Substanzstrukturen $C_i \in \mathcal{B}$ einzubetten. Falls diese Einbettung möglich ist, sind als Resultat in der Strukturklasse

$$\mathcal{S}(K) = \{C_i \in \mathcal{B} \mid C_i \triangleright K\} \subseteq \mathcal{B}, \quad (4.2)$$

die eine Teilmenge der Bibliothek \mathcal{B} ist, diejenigen Substanzstrukturen C_i zusammenfasst, in die das Motiv K eingebettet werden kann. Diese Einbettungen werden durch die Schreibweise „ $C_i \triangleright K$ “ symbolisiert, die besagen soll, dass die Strukturen C_i das Motiv K konkretisieren. In (möglicherweise zahlreichen) weiteren Schritten, die in Abbildung 4.1 nur angedeutet sind, erweitert MoSS das Kernmotiv K systematisch zu konkretisierten Kernmotiven $K' \triangleright K$, indem zusätzliche Bindungen und Atome hinzugefügt werden. Die Auswahl der Erweiterungen geschieht dabei in sinnvoller Weise anhand der in der Klasse $\mathcal{S}(K)$ enthaltenen Strukturen. MoSS versucht, die erweiterten Kernmotive K' in die verbliebenen Substanzstrukturen einzubetten, wobei die erfolgreichen Einbettungen konkretisierte Substrukturklassen $\mathcal{S}'(K')$ definieren. Die Kernmotive werden wiederum erweitert und eingebettet, bis die resultierenden Klassen nurmehr einzelne Substanzstrukturen enthalten und daher nicht weiter konkretisiert werden können. Als Ergebnis dieser systematischen Substruktursuchen liefert MoSS die Liste

$$\mathcal{S} = \{\mathcal{S}_j(K_j) \mid j = 1, \dots, N_S\} \quad (4.3)$$

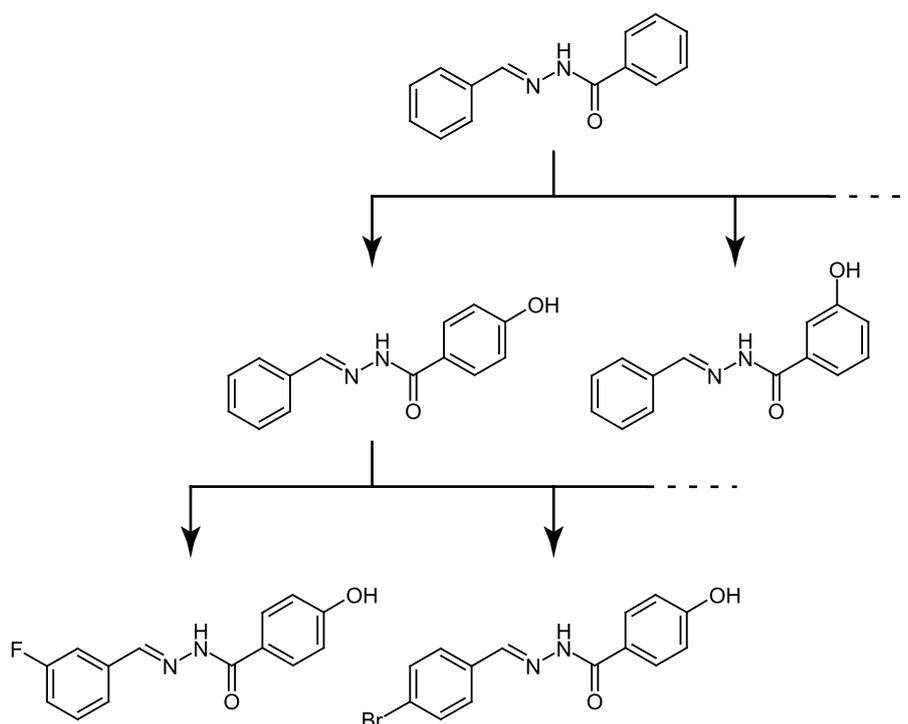


Abbildung 4.1: Systematische Erweiterung von Strukturmotiven. Wie hier schematisch dargestellt ist, erweitert MoSS ein strukturelles Kernmotiv (oben) sukzessive um mögliche Substituenten zu konkretisierten Kernmotiven. Für diese Kernmotive bildet MoSS strukturelle Klassen, in welchen diejenigen Substanzstrukturen zusammengefasst sind, die das jeweilige Kernmotiv als Substruktur enthalten. Die Motive werden dabei schrittweise um Bindungen und Atome erweitert. Es sei angemerkt, dass Wasserstoffatome wie beispielsweise diejenigen der OH-Gruppen aufgrund der von MoSS eingesetzten Repräsentation von Molekülstrukturen nicht als eigenständige Atome betrachtet werden.

aller N_S gebildeten Strukturklassen. Die Klassen S_j sind gemäß (4.2) diejenigen Teilmengen der Bibliothek \mathcal{B} , deren Substanzen die jeweiligen Strukturmotive K_j gemein haben. Die Klassen werden dabei ausschließlich anhand der Strukturen der Bibliothekssubstanzen gebildet; Wirksamkeiten der Substanzen werden zunächst nicht berücksichtigt.

4.1.2 Berücksichtigung von Wirksamkeiten in MoSS

MoSS bietet nun die Möglichkeit, beim Bilden der Strukturklassen auch Wirksamkeiten zu berücksichtigen, wodurch erreicht werden soll, dass „diskriminierende Strukturmotive“ gefunden werden, die *häufig* in den wirksamen und *selten* in den unwirksamen Substanzstrukturen vorkommen [145]. Bereits vor der Bildung der Klassen unterteilt MoSS dazu die zu durchsuchende Bibliothek \mathcal{B} anhand der gegebenen Wirksamkeiten und eines vom Benutzer festgelegten Grenzwertes in die Menge \mathcal{B}_+ der wirksamen und die Menge \mathcal{B}_- der unwirksamen Substanzen. In analoger Weise wird bei der Bildung einer Strukturklas-

se \mathcal{S} diese Klasse in die Menge \mathcal{S}_+ der wirksamen und die Menge \mathcal{S}_- der unwirksamen Substanzen aufgeteilt. Anhand der Größen dieser Mengen werden die Häufigkeiten

$$P_+ = \frac{|\mathcal{S}_+|}{|\mathcal{B}_+|} \quad \text{und} \quad P_- = \frac{|\mathcal{S}_-|}{|\mathcal{B}_-|} \quad (4.4)$$

der wirksamen bzw. der unwirksamen Substanzen bestimmt und gemäß

$$P_+ > P_+^{\min} \quad \text{und} \quad P_- < P_-^{\max} \quad (4.5)$$

überprüft, ob der Anteil P_+ einen Minimalwert P_+^{\min} überschreitet und der Anteil P_- einen Maximalwert P_-^{\max} unterschreitet. Dabei können die Schranken P_+^{\min} und P_-^{\max} vom Benutzer festgelegt werden.

Kritik am verwendeten Wirksamkeitsfilter

Das Stellen der Bedingung „ $P_+ > P_+^{\min}$ “ in (4.5) führt dazu, dass nur solche Klassen \mathcal{S} gebildet werden, die eine gewisse Größe überschreiten und hat daher den Nachteil, dass kleine Substanzklassen übersehen werden. Dies liegt darin begründet, dass der Anteil P_+ der wirksamen Substanzen in einer betrachteten (kleinen) Klasse \mathcal{S} gemäß (4.4) auf die Gesamtzahl $|\mathcal{B}_+|$ der wirksamen Substanzen bezogen wird und nicht auf die Größe $|\mathcal{S}|$ der Klasse. Da aber in vielfältigen Bibliotheken wie DIVERSet zahlreiche kleine Klassen enthalten sind, die dennoch potenzielle Leitstrukturen darstellen können, ist das Fordern dieser Bedingung nachteilig.

4.1.3 Ein verbesserter Wirksamkeitsfilter

Aus diesem Grund habe ich die von MoSS gebotene Möglichkeit, bei der Bildung der Strukturklassen auch Wirksamkeiten zu berücksichtigen, nicht genutzt, was durch die Wahl $P_+^{\min} = 0$ und $P_-^{\max} = 1$ erreicht werden kann. Da die Liste (4.3) der Strukturklassen mithin unabhängig von Wirksamkeiten gebildet wird, umfasst sie auch eine Vielzahl unwirksamer Klassen. Nun ist es aber möglich, Wirksamkeiten im Nachhinein zu berücksichtigen, um aus einer umfangreichen Klassenliste die wirksamen heraus zu filtern. Dabei verwende ich anstelle des im Kriterium (4.5) eingesetzten globalen Anteils P_+ , der die Zahl der wirksamen Substanzen auf alle in der Bibliothek enthaltenen wirksamen Substanzen bezieht, den klassenlokalen Anteil

$$\tilde{P}_+ = \frac{|\mathcal{S}_+|}{|\mathcal{S}|} > \tilde{P}_+^{\min} \quad (4.6)$$

der wirksamen Substanzen innerhalb der fraglichen Klasse \mathcal{S} und überprüfe, ob \tilde{P}_+ einen wählbaren Mindestanteil \tilde{P}_+^{\min} übersteigt. Die Verwendung dieses modifizierten Wirksamkeitsfilters führt dazu, dass nun auch kleine Klassen mit hohen Anteilen wirksamer Substanzen gefunden werden.

Die nachträgliche Berücksichtigung von Wirksamkeiten bei der Auswahl von Klassen bietet den weiteren Vorteil, dass die vollständige Klassenliste (4.3) nur einmal gebildet werden muss und später durch Variation des Grenzwertes \tilde{P}_+^{\min} beliebig gefiltert werden kann. Ferner können auch weitere Filter angewendet werden. So kann beispielsweise eine Mindestgröße der Klassen gefordert werden. Schließlich können experimentell bestimmte Wirksamkeiten aus mehreren Testreihen [z.B. SIFT (TP1) und Zellkultur (TP3)] nacheinander eingesetzt werden, um die Klassenliste nachträglich zu filtern oder um einzelne Klassen zu charakterisieren.

4.1.4 Aufbereitung und Visualisierung von Strukturklassen

Zur Umsetzung des soeben beschriebenen modifizierten Einsatzes von MoSS habe ich das Shell-Skript `xminer.sh` programmiert, welches als Bestandteil des CVS-Moduls „screening“ zur Verfügung steht. Das Skript `xminer.sh` lässt die Klassenbildung mithilfe von MoSS wirksamkeitsunabhängig ablaufen, fügt den Substanzbeschreibungen Wirksamkeiten aus verschiedenen Testreihen hinzu, bildet gefilterte Klassenlisten und bereitet die Ergebnisse für eine Webbrowser-basierte Darstellung grafisch auf. Bei der grafischen Aufbereitung werden, ähnlich wie bei SIFT-Viewer I, eine Vielzahl von Grafikdateien der Struktur motive, wiederum unter Verwendung des Pakets CACTVS [125], erzeugt und einbettende HTML-Dateien angelegt. Die generierte Darstellung bietet einen schnellen Zugriff auf die gebildeten Klassen und die darin enthaltenen Substanzen.

Um den Einsatz von `xminer.sh` zu demonstrieren, soll exemplarisch das Ergebnis einer Klassenbildung für die Substanzen der DIVERSet 1-Bibliothek gezeigt werden. Für die Bildung der Klassen wurde als Ausgangsmotiv ein einzelnes Kohlenstoffatom gewählt und als Randbedingung gefordert, dass die Kern motive der zu bildenden Klassen jeweils mindestens 10 Atome enthalten sollen. MoSS erzeugte daraufhin eine Liste mit insgesamt 5476 Klassen, aus welcher von `xminer.sh` diejenigen 29 Klassen ausgewählt wurden, die mehr als drei Substanzen enthalten ($|S| > 3$), von denen mehr als 50% in Zellkultur wirksam sind [$\tilde{P}_+ > 0,5$ gemäß Gleichung (4.6)].

Abbildung 4.2 zeigt einen Bildschirmschnappschuss der von `xminer.sh` generierten Webbrowser-basierten Visualisierung des beschriebenen Beispiels. Über die beiden Einträge im linken oberen Bereich „Klassenlisten“ kann der Benutzer sowohl die vollständige Klassenliste als auch die Auswahl der großen, überwiegend in Zellkultur wirksamen Strukturklassen für die Darstellung selektieren. Hier wurde die Auswahlliste selektiert, woraufhin im darunter liegenden Bereich eine Tabelle angezeigt wird, in welcher die Klassen, nach generischen Namen („s556“, „s2101“, etc.) benannt und nach absteigender mittlerer Zellkultur-Wirksamkeit sortiert, aufgelistet sind. Die Tabelle zeigt für jede Klasse ihre Größe („Anzahl Substanzen“), die Größe des Kern motifs („Anzahl Atome“) sowie die klassenlokalen Mittelwerte der Zellkultur- und SIFT-Wirksamkeiten. Eine dieser Klassen, die Klasse „s4355“, die hohe mittlere Zellkultur-Aktivität (0,8) aber keine erhöhte SIFT-Aktivität (-0,21) zeigt, wurde durch Klicken auf den Klassennamen für die

4.1 Identifikation und Validierung von Leitstrukturen

The screenshot shows a web browser window with the following content:

Klassenlisten
[alle Strukturklassen](#)
[ausgewählte Klassen](#)

Klasse	Anzahl Substanzen	Anzahl Atome	Mittlere Aktivität Zellkultur	Mittlere Aktivität SIFT
s556	4	19	1.000000	0.478533
s2101	6	14	0.833333	-1.654090
s1185	6	19	0.833333	-0.003136
s4355	5	25	0.800000	-0.211594
s3179	5	21	0.800000	-0.030170
s775	4	17	0.750000	0.385037
s4179	4	19	0.750000	-0.231215
s4132	4	12	0.750000	0.081555
s2098	4	19	0.750000	-1.654090
s2077	4	16	0.750000	-0.451246
s1792	4	16	0.750000	0.082552
s102	4	12	0.750000	0.762915
s4356	6	24	0.666667	-0.216969
s3224	6	22	0.666667	-0.270917
s558	8	18	0.625000	0.616426
s2132	8	11	0.625000	-0.822422
s2105	8	12	0.625000	-0.803946
s779	5	15	0.600000	0.234204

s4355

A03

Compound	Zellkultur-Aktivität	SIFT-Aktivität
10278_A03	1.0	0.194553
10274_E04	1.0	-0.055556
10274_D04	1.0	-0.360269
10274_C04	1.0	-0.685185
10274_H03	0.0	-0.151515
Mittelwert (aus)	0.800000 (5)	-0.211594 (5)

Abbildung 4.2: Schnappschuss der Visualisierung eines `xminer`-getriebenen MoSS-Laufs. Die von MoSS gebildete Liste von Strukturklassen ist im linken Bereich in tabellarischer Form dargestellt. Durch Mausklicks auf Listeneinträge können einzelne Klassen für die Darstellung im mittleren Bereich ausgewählt werden. Dort ist für die ausgewählte Klasse `s4355` oben das Kernmotiv gezeigt, und darunter sind die in der Klasse enthaltenen Substanzen aufgelistet, wobei Substanznamen, Zellkultur- und SIFT-Wirksamkeiten sowie klassenlokal gebildete Mittelwerte der Wirksamkeiten angegeben sind. Durch Anklicken des Substanznamens „10278_A03“ wird rechts die Struktur dieser Substanz gezeigt.

Darstellung im mittleren Bereich ausgewählt. Dort ist oben das Kernmotiv der Klasse und darunter eine Tabelle der darin enthaltenen Substanzen angezeigt. Wie aus der Tabelle entnommen werden kann, waren vier der Substanzen wirksam (1,0) im Zellkultur-Screening, wohingegen im SIFT-Screening lediglich die Substanz 10278_A03 eine leicht anti-aggregatorische Aktivität zeigt. Durch Klicken auf den Substanznamen wurde diese Substanz für die Darstellung im rechten Bereich ausgewählt. Die gezeigte Klasse gehört der Superklasse der Benzothiazole an, die als Wirkstoffe gegen die mit der Huntingtonschen Krankheit assoziierte Proteinaggregation bekannt sind [150].

Das beschriebene Werkzeug `xminer.sh` wurde erst jüngst von mir entwickelt und kann, wie das geschilderte Beispiel demonstriert, auch zur Identifikation von potenziellen Leitstrukturen eingesetzt werden. Infolge des dabei allgemein zu wählenden Ausgangskernmotivs (etwa eines C-Atoms) ist jedoch der Speicherbedarf für die Bildung der umfangreichen Klassenliste beträchtlich und der Einsatz dieses Werkzeugs daher nicht ohne einen ausreichend dimensionierten Rechner zu bewerkstelligen. Aufgrund von damit verbun-

denen Schwierigkeiten wurde MoSS von mir zunächst nur zur Validierung von bereits auf andere Weise gefundenen potenziellen Leitstrukturen eingesetzt, wobei die Zahl der Klassen typischerweise deutlich geringer ausfällt (~ 500). Um die dennoch unübersichtliche Vielzahl von Subklassen einer zu validierenden konkreten Ausgangsstrukturklasse in leichter überschaubarer Form darzustellen, entwickelte ich ein Verfahren, welches die Klassen in einer Hierarchie anordnet.

4.1.5 Bildung von Strukturhierarchien

Weil bei der oben beschriebenen Bildung von Strukturklassen jeweils ein allgemeineres Kernmotiv K_j zu einem konkretisierten Motiv $K_k \triangleright K_j$ erweitert wurde, gilt für die zugehörigen Strukturklassen \mathcal{S}_k und \mathcal{S}_j , dass $\mathcal{S}_k \subset \mathcal{S}_j$ eine Teilmenge von \mathcal{S}_j ist. Die gebildeten Strukturklassen sind also ineinander verschachtelt, und es sollte daher möglich sein, sie in einer Hierarchie anzuordnen, an deren Spitze die zum Ausgangskernmotiv gehörige Strukturklasse steht. In immer tiefer liegenden Ebenen der Hierarchie sollen jeweils immer konkretere Strukturklassen angeordnet sein, ähnlich wie bei der in Abbildung 4.1 dargestellten Hierarchie, die bei der Bildung der Klassenliste (4.3) zu einem gegebenen Ausgangskernmotiv durchlaufen wird.

Zur Konstruktion einer Hierarchie, in welcher die von MoSS gebildeten Klassen und zugehörigen Kernmotive angeordnet werden sollen, ist es notwendig zu wissen, welche Konkretisierungen eines jeden gegebenen Kernmotivs in der Menge aller Kernmotive enthalten sind. Diese Information kann aus der Menge der Kernmotive extrahiert werden, indem MoSS ein zweites Mal eingesetzt und dabei auf die Menge der Kernstrukturmotive K_j angewendet wird. Als Ergebnis liefert MoSS gemäß (4.3) eine Liste der Klassen von Kernmotiven, also für jedes Kernmotiv K_j diejenigen Kernmotive K_k , welche das jeweilige Kernmotiv K_j konkretisieren. Die Konkretisierungsbeziehungen zwischen den Kernmotiven übertragen sich als Hierarchiebeziehungen auf die Strukturklassen $\mathcal{S}_j(K_j)$, sodass jeder Klasse

$$\mathcal{S}_j \rightarrow \{\mathcal{S}_k \in \mathfrak{S} \mid \mathcal{S}_k \subset \mathcal{S}_j\} \quad \text{für } j = 1, \dots, N_{\mathfrak{S}} \quad (4.7)$$

die Menge derjenigen Klassen \mathcal{S}_k zugeordnet wird, welche Teilmengen von \mathcal{S}_j sind.

Um die skizzierte hierarchische Beziehung zwischen den Strukturklassen zu veranschaulichen, ist in Abbildung 4.3 links unten eine beispielhafte Liste mit neun Strukturklassen \mathcal{S}_j ($j = 1, \dots, 9$) gezeigt. Die Liste der Strukturklassen \mathcal{S}_j führt zusätzlich die nach (4.7) in den \mathcal{S}_j jeweils enthaltenen Subklassen \mathcal{S}_k auf. Im oberen Bereich von Abbildung 4.3 werden die zugrunde liegenden Mengenbeziehungen durch Ellipsen und fette Punkte veranschaulicht. Dort bezeichnet eine Ellipse eine Strukturklasse \mathcal{S}_j , welche andere Strukturklassen \mathcal{S}_k als echte Teilmengen enthält. So umfasst beispielsweise die große Ellipse \mathcal{S}_9 alle anderen Klassen \mathcal{S}_1 bis \mathcal{S}_8 . Klassen, die keine anderen Subklassen enthalten, sind als fette Punkte dargestellt (\mathcal{S}_1 bis \mathcal{S}_5). Enthält eine Subklasse (z.B. \mathcal{S}_8) einer übergeordneten Klasse (z.B. \mathcal{S}_9) weitere Subklassen (z.B. $\mathcal{S}_2, \dots, \mathcal{S}_7$), so werden diese Subklassen

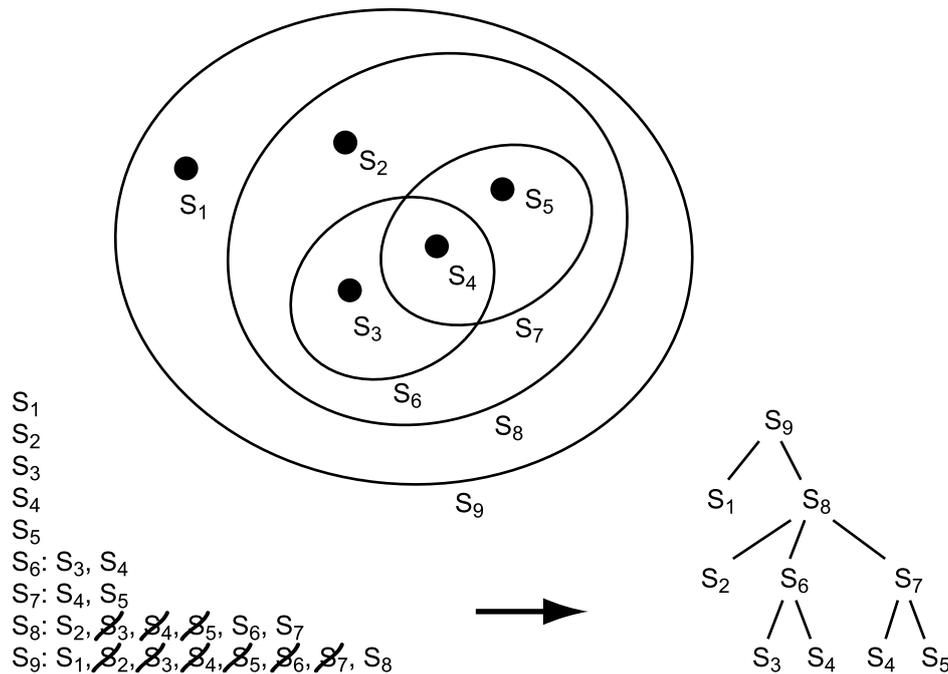


Abbildung 4.3: Bildung einer Hierarchie von Substrukturklassen. Links unten ist eine beispielhafte Liste von neun Strukturklassen S_j dargestellt, die zusätzlich die nach (4.7) jeweils enthaltenen Subklassen S_k aufführt. Zur Veranschaulichung der zugrunde liegenden Mengenbeziehungen sind die Klassen im oberen Bereich als Ellipsen oder fette Punkte dargestellt. Die Liste liegt der rechts unten dargestellten Hierarchie zugrunde. Erläuterungen finden sich im Text.

als *indirekte* Subklassen der übergeordneten Klasse (S_9) aufgefasst und aus der Liste ihrer Subklassen gestrichen (vgl. die zu S_9 gehörige Liste). Damit enthält diese Liste nur noch Subklassen (S_1 und S_8), die selbst nicht Subklassen anderer Klassen aus der ursprünglichen Liste sind. Auf diese Weise wird die rechts unten dargestellte Hierarchie *direkter* Subklassen definiert. Dort besitzt jeder Knoten S_j als Kinder ausschließlich direkte Subklassen.

Ausgangspunkt beim Aufbau der Hierarchie sind also die gemäß (4.7) gebildeten und in Abbildung 4.3 links unten exemplarisch gezeigten Listen von Teilmengen. Unter Verwendung des Skripts `miner_tree.sh` werden aus diesen Listen alle indirekten Teilmengen gestrichen. Da durch die Vorgabe eines übergeordneten Kernmotivs die übergeordnete Klasse von vornherein feststeht, kann ausgehend von der Liste ihrer direkten Subklassen die Hierarchie unmittelbar aufgebaut werden. Im gezeigten Beispiel wird also die Klasse S_9 als Wurzel des Baums definiert und ausgehend von S_9 , in einem sogenannten Tiefendurchlauf, sukzessive die Hierarchie aufgebaut. Dabei wird die erste S_9 -Teilmenge S_1 als Kindknoten von S_9 angelegt, welcher elementar ist und daher ein Blatt im entstehenden Baum ergibt. Die nächste Konkretisierung der Wurzel S_9 ist die Klasse S_8 , welche ihrerseits durch S_2 , S_6 und S_7 konkretisiert wird, etc. Wie an der Klasse S_4 zu erkennen

ist, können Klassen an mehreren Stellen der Hierarchie auftauchen. In dieser Hinsicht unterscheidet sich die auf diese Weise gebildete Hierarchie von derjenigen, die von MoSS ursprünglich bei der Bildung der Kernmotive durchlaufen wurde, denn in jener Hierarchie kommen Kernmotive in der Regel nicht mehrfach vor, um redundante Suchen zu vermeiden.

Details der Implementierung

Der hier vorgestellte Algorithmus zur Hierarchiebildung wurde von mir in Form des Shell-Skripts `miner_tree.sh` implementiert, welches Bestandteil des Softwarepakets „hMoSS“ (hierarchischer MoSS) ist und im gleichnamigen Verzeichnis des CVS-Moduls „screening“ abgelegt ist. Zum Verständnis der Funktionsweise des Skripts `miner_tree.sh` sei angemerkt, dass von MoSS die Liste der gebildeten Strukturklassen in Form von zwei Textdateien ausgegeben wird. Die Datei „sub“ enthält zeilenweise Angaben zu den gebildeten Substrukturklassen, wobei jeweils ein Name für die Klasse, das zur Klasse gehörige Strukturmotiv in Form eines SMILES-Strings sowie u.a. die Anzahlen von wirksamen und unwirksamen Substanzen eingetragen sind. In der Datei „ids“ sind für die Substrukturklassen deren Namen und die darin enthaltenen Substanzen aufgelistet. Diese Dateien werden von dem Skript `miner_tree.sh` unter Verwendung des Werkzeugs Gawk [123] abgearbeitet, um die Teilmengenbeziehungen auszuwerten und daraus die Hierarchie von Strukturklassen zu bilden. Neben der Hierarchiebildung dient das Skript `miner_tree.sh` auch der Generierung einer zugehörigen Webbrowser-basierten grafischen Benutzerschnittstelle.

4.1.6 Visualisierungen von Strukturhierarchien

Die Webbrowser-basierte Visualisierung von Hierarchien von Strukturklassen ähnelt jener von linearen Listen von Klassen, die mithilfe des Werkzeugs `xminer.sh` bewerkstelligt wurde (vgl. Abschnitt 4.1.4). Hier wurde lediglich die tabellarische Darstellung der Klassenliste durch eine JavaScript-basierte Darstellung der Klassenhierarchie ersetzt, wobei das frei verfügbare Paket „Morten’s JavaScript Tree Menu“ [151] eingesetzt wurde.

Abbildung 4.4 zeigt einen Bildschirmschnappschuss der resultierenden Visualisierung einer Substrukturhierarchie, die mithilfe von hMoSS für das Kernmotiv einer potenziellen Leitstruktur (NBB, siehe Kapitel 5) und die im SIFT-Screening vermessene Bibliothek DIVERSet 1 gewonnen wurde. Im linken Bereich befindet sich die JavaScript-basierte Baumdarstellung der Strukturhierarchie. Die Knoten des Baumes sind nach den generischen Klassennamen benannt und zeigen zudem die klassenlokalen Mittelwerte der SIFT-Wirksamkeiten, um das Auffinden von hochwirksamen Klassen zu unterstützen. Durch Klicken in der Baumdarstellung kann der Benutzer, ähnlich wie beim Explorer von Microsoft Windows, durch die Hierarchie navigieren und einzelne Strukturklassen für die

4.1 Identifikation und Validierung von Leitstrukturen

The screenshot shows a web browser window titled "2D-Strukturen - Mozilla Firefox". The address bar contains the URL: `file:///C:/scratch/thomas/queries/DIVERSet1/NBB_diss2/miner_js_tree/index.html`. The page title is "Hierarchie von 2D-Strukturmotiven".

On the left, a tree view shows the hierarchy of structure classes. The selected node is "Struktur s486".

The main content area is divided into two panels. The left panel, labeled "s486", shows the core structure of the selected class. The right panel, labeled "E04", shows the structure of the selected compound, "10305_E04".

Below the structures, a table lists the compounds and their SIFT-activities:

Compound	SIFT-activity
10307_D11	0.552381
10305_E04	0.728636
10305_B08	0.0149925
Mean (3)	0.432003

Abbildung 4.4: Schnappschuss der Visualisierung mit hMoSS. Die von hMoSS gebildete Hierarchie aus Strukturklassen von NBB-Verbindungen der DIVERSet 1-Bibliothek ist im linken Bereich als Baumstruktur dargestellt, deren Knoten Strukturklassen repräsentieren. Durch Mausklicks auf die Knoten können einzelne Klassen für die Darstellung im rechten Bereich ausgewählt werden. Dort ist für die ausgewählte Klasse s486 oben links die Kernstruktur gezeigt, und darunter sind die in der Klasse enthaltenen Substanzen aufgelistet, wobei Substanznamen, SIFT-Wirksamkeiten und die klassenlokal gemittelte Wirksamkeit angegeben sind. Durch Klicken des Namens „10305_E04“ wird rechts die Struktur dieser Substanz gezeigt.

Darstellung auswählen. Wie bei `xminer.sh` wird die ausgewählte Klasse im mittleren Bereich angezeigt, wobei wiederum oben das Kernmotiv und darunter in tabellarischer Form die darin enthaltenen Substanzen, deren Wirksamkeiten und die klassenlokal gemittelte Wirksamkeit angegeben sind. Durch Mausklicks auf Substanznamen können die Strukturen der Substanzen im rechten Bereich dargestellt werden.

Dieses Visualisierungswerkzeug wurde von meinen Verbundpartnern und mir dazu eingesetzt, systematisch SARs für Subklassen von potenziellen Leitstrukturen zu untersuchen. Dadurch konnte verstanden werden, welche Substituenten des Kernmotivs der untersuchten Leitstruktur die Wirksamkeiten der zugehörigen Substanzen erhöhen oder erniedrigen können, sodass damit potenzielle Leitstrukturen validiert werden konnten (siehe Kapitel 5).

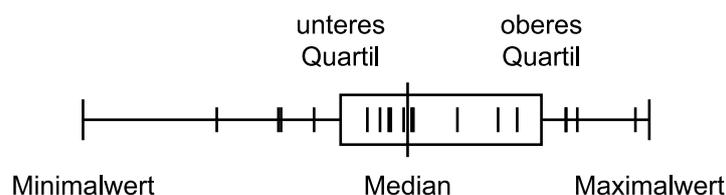


Abbildung 4.5: Box-and-Whisker-Plot für eine Wirksamkeitsverteilung. Die exemplarische Verteilung von Wirksamkeitswerten (kurze senkrechte Striche) wird mithilfe eines Box-and-Whisker-Plots durch Minimal- und Maximalwert, Median (lange senkrechte Striche) und Quartile charakterisiert. Dabei definieren die Quartile eine Box und Minimal- und Maximalwert die Whisker.

Grafische Darstellungen von Wirksamkeitsverteilungen

Damit klassenlokale Wirksamkeitsverteilungen einzelner Substrukturklassen besser beurteilt werden können, wurden zusätzlich zu einfachen Mittelwerten sogenannte *Box-and-Whisker-Plots* eingesetzt. Abbildung 4.5 zeigt einen solchen Box-and-Whisker-Plot für eine exemplarische Verteilung von Wirksamkeitswerten. Diese aus der deskriptiven Statistik stammenden Diagramme repräsentieren eine gegebene Verteilung von Werten durch fünf charakteristische Werte, nämlich durch den Minimal- und den Maximalwert, den Median sowie die beiden sogenannten *Quartile*. Analog zum Median, der gemäß Gleichung (3.1) die Mitte der in der Verteilung enthaltenen Werte angibt, geben das untere und das obere Quartil diejenigen Werte an, unterhalb bzw. oberhalb derer jeweils ein Viertel der Werte der Verteilung liegen.

Zur Darstellung solcher Box-and-Whisker-Plots, die auch Candlestick-Repräsentationen genannt werden, dient die von mir geschriebene Prozedur `candle_inhib` des IDL-Moduls `screening_inhibition.pro`, welches ebenfalls von dem CVS-Modul „screening“ bereitgestellt wird. Box-and-Whisker-Plots wurden nicht in die Webbrowser-basierte Darstellung von hMoSS integriert, sondern mussten für ausgewählte Substrukturklassen mithilfe des IDL-Skripts händisch erzeugt werden.

Dynamische, Java-basierte Bildung von Strukturklassen

Ähnlich wie das Webbrowser-basierte Werkzeug SIFT-Viewer I durch die Java-basierte Variante SIFT-Viewer II ersetzt wurde, so habe ich auch für das statische Werkzeug hMoSS eine dynamische Java-basierte Variante, hMoSS II, entwickelt und implementiert. Die zu diesem Werkzeug gehörigen Java-Klassen werden von dem CVS-Modul `hmoSS` bereitgestellt.

Zur Erläuterung der zusätzlichen Fähigkeiten von hMoSS II zeigt Abbildung 4.6 einen Schnappschuss dieses Werkzeugs. Wie bei hMoSS ist im linken Bereich der Darstellung eine Hierarchie von Strukturklassen und im rechten Bereich eine ausgewählte Klasse zusammen mit dem Kernmotiv (oben links), der Liste der enthaltenen Substanzen und deren

4.1 Identifikation und Validierung von Leitstrukturen



Abbildung 4.6: Bildschirmschnappschuss von hMoSS II. Wie das in Abbildung 4.4 dargestellte HTML-basierte statische Werkzeug hMoSS dient auch diese Java-basierte dynamische Variante der Visualisierung von mit MoSS gebildeten Strukturklassen. Im Gegensatz zur statischen Variante werden hier die Strukturklassen dynamisch infolge von Benutzerinteraktion zur Laufzeit gebildet. Als zusätzliche Beschreibung der gebildeten Strukturklassen ist im rechten mittleren Bereich ein Box-and-Whisker-Plot der klassenlokalen Wirksamkeitsverteilung dargestellt. Die Knoten der Hierarchie (links) tragen als Namen die SMILES-Strings der Kernmotive der repräsentierten Klassen.

Wirksamkeiten (unten) sowie die Struktur einer ausgewählten Substanz (oben rechts) dargestellt. Zusätzlich enthält die Visualisierung mit hMoSS II im mittleren rechten Bereich einen Box-and-Whisker-Plot der klassenlokalen Wirksamkeitsverteilung.

hMoSS II unterscheidet sich von hMoSS in erster Linie dadurch, dass die Strukturklassen dynamisch zur Laufzeit infolge von Benutzerinteraktion gebildet werden. Beim Starten der Anwendung existiert lediglich die zum Ausgangskernmotiv gehörige Klasse, die alle Substanzen mit diesem Motiv umfasst und von hMoSS II als Wurzelknoten der Baumstruktur dargestellt wird. Öffnet der Benutzer durch einen Mausklick den Wurzelknoten, so werden von hMoSS II dynamisch die Strukturklassen der nächst tiefer liegenden Hierarchieebene gebildet, wobei das Wurzelmotiv um ein zusätzliches Atom erweitert wird. In analoger Weise werden dynamisch alle weiteren Knoten erweitert, sobald sie vom Benutzer geöffnet werden. Dabei wird jeweils MoSS dazu eingesetzt, die Subklassen der ausgewählten Klasse zu bilden. MoSS erhält dabei als Eingabe das Kernmotiv der ausgewählten Klasse, die darin enthaltenen Substanzen sowie die Vorgabe, nur solche Subklassen zu bilden, die das Kernmotiv um ein Atom erweitern.

Die beschriebene Integration von MoSS in hMoSS II wurde dadurch erleichtert, dass MoSS in Java geschrieben ist. Es sei angemerkt, dass hMoSS II zur Darstellung von Molekülstrukturen eine weitere Java-Bibliothek einsetzt, die, ebenso wie MoSS, in der Gruppe von Christian Borgelt an der Universität Magdeburg entwickelt wurde und die er mir dankenswerterweise zur Verfügung gestellt hat.

Abschließend sei noch angemerkt, dass auf der Suche nach einem brauchbaren Werkzeug für systematische SAR-Untersuchungen eine kostenlose Testversion des kommerziellen Softwarepakets Leadscope [152] installiert und probeweise eingesetzt wurde. Leadscope unterstützt die Unterteilung einer gegebenen Substanzsammlung in Strukturklassen und deren Charakterisierung durch grafische Darstellungen von klassenlokalen Wirksamkeitsverteilungen. Jedoch waren von der damaligen Version von Leadscope die Kernmotive der Strukturklassen unabhängig von der betrachteten Bibliothek fest vorgegeben und konnten nicht frei gewählt oder etwa anhand der in der Bibliothek enthaltenen Strukturen generiert werden. So war es damals notwendig, die oben geschilderten Werkzeuge für systematische SAR-Untersuchungen selbst zu entwickeln. Inzwischen bietet Leadscope laut Produktbeschreibung eine systematische Suchfunktion in Form einer sogenannten Restgruppen-Analyse (R-group analysis). Dieses Werkzeug wurde jedoch nicht mehr im Projekt beschafft und eingesetzt.

Nachdem ich nun die von mir entwickelten Werkzeuge zur systematischen Untersuchung von SARs für ausgewählte Leitstrukturklassen und zu deren Validierung vorgestellt habe, möchte ich anschließend noch ein kommerzielles Softwarepaket vorstellen, das ich für die Suche nach neuen Leitstrukturen eingesetzt habe.

4.2 Strukturelle Clusteranalysen zur Suche nach neuen Leitstrukturen

Das kommerzielle Softwarepaket Benchware HTS DataMiner der Firma Tripos [143] dient der Suche nach neuen Leitstrukturen und bietet dazu die Möglichkeit, die Substanzen einer gegebenen Sammlung in Cluster strukturell ähnlicher Substanzen zu unterteilen. Anders als von den im letzten Abschnitt beschriebenen Werkzeugen wird dabei strukturelle Ähnlichkeit nicht direkt anhand der Molekülstrukturen bestimmt, sondern anhand von daraus abgeleiteten, abstrakteren Beschreibungen der Substanzen in Form von sogenannten strukturellen Fingerabdrücken.

4.2.1 Repräsentation von Substanzstrukturen

Zur Repräsentation von Substanzstrukturen wird von DataMiner das Fingerabdrucksystem UNITY der Firma Tripos eingesetzt, welches im Detail nicht offengelegt ist. In der Literatur finden sich lediglich Umschreibungen des UNITY-Systems [153–155], aus denen

hervorgeht, dass es dem Fingerabdrucksystem der Firma Daylight [120] ähnelt, welches im hilfreichen „Daylight Theory Manual“ [156] dargestellt ist. An dieser Darstellung orientiert sich daher die Darstellung in diesem Abschnitt.

Fingerabdrucksysteme sind zentrale Bestandteile von chemischen Softwarepaketen wie denjenigen der Firmen Tripos oder Daylight, da sie eine effiziente Umsetzung von Substruktursuchen ermöglichen. Weil Fingerabdrücke Weiterentwicklungen von sogenannten *strukturellen Schlüsseln* sind, sollen letztere zuerst erläutert werden.

Strukturelle Schlüssel

Der strukturelle Schlüssel einer gegebenen Substanz s ist ein Bit-String

$$\mathbf{b}_s = (b_{s1}, \dots, b_{sD}) \quad (4.8)$$

der Länge D , dessen binäre Einträge $b_{si} \in \{0, 1\}$ die Anwesenheit (1) oder Abwesenheit (0) bestimmter vordefinierter struktureller Fragmente in der Struktur der Substanz s kodieren. Für die Erzeugung von strukturellen Schlüsseln für die Substanzen einer Sammlung müssen daher zunächst strukturelle Fragmente wie einzelne Atome, Atomgruppen oder Ringsysteme, die von Interesse sind, ausgewählt und in einem sogenannten Fragmentverzeichnis (*predefined fragment dictionary*) abgelegt werden. Dabei kann die Auswahl der Fragmente händisch oder automatisiert mithilfe von Substruktursuchen in der Substanzsammlung anhand der darin vorkommenden Motive getroffen werden. Ist das Fragmentverzeichnis festgelegt, so werden den darin enthaltenen Fragmenten einzelne Bits in den zu bildenden strukturellen Schlüsseln zugeordnet. Zur Erzeugung des Schlüssels \mathbf{b}_s für eine gegebene Substanz s wird überprüft, welche der Fragmente des Verzeichnisses in der Struktur der Substanz s enthalten sind, und deren An- oder Abwesenheit, wie oben gesagt, über die Werte 1 bzw. 0 der zugehörigen Bits vermerkt. Zusätzlich können für die Fälle, dass Fragmente mehrmals in der Substanzstruktur vorkommen, Bits reserviert sein, welche die Häufigkeiten des Vorkommens kodieren. Da strukturelle Schlüssel inhärent auf das zugrunde gelegte Fragmentverzeichnis bezogen sind, ist ihre Bedeutung nicht allgemein. Dieser Nachteil wird von Fingerabdrücken vermieden.

Fingerabdrücke

Wie strukturelle Schlüssel sind auch Fingerabdrücke Substanzrepräsentationen in Form von Bit-Strings (4.8). Zur Erzeugung des Fingerabdrucks für eine Substanz s wird hier aber nicht das Vorkommen fest definierter Fragmente überprüft, sondern es werden die in der Molekülstruktur vorkommenden Pfade der Gestalt Atom(–Bindung–Atom) $_n$ der Längen $n = 0, 1, 2, \dots$ extrahiert, wobei das von DataMiner eingesetzte UNITY-System Pfadlängen $n < 7$ berücksichtigt [154]. Ein Beispiel, das aus dem Theory Manual von Daylight [156] übernommen ist, zeigt die in dem Molekül OC=CN enthaltenen Pfade der verschiedenen Längen n :

- $n = 0$: O, C, N
- $n = 1$: OC, C=C, CN
- $n = 2$: OC=C, C=CN
- $n = 3$: OC=CN

Um nun das Vorkommen der gefundenen Pfade durch Einträge im Bit-String \mathbf{b}_s zu vermerken, sind im UNITY-System die Bit-Strings in Abschnitte aufgeteilt, die für die Pfade der verschiedenen Längen vorgesehen sind. Dann werden die Pfade einer gegebenen Länge auf Bitmuster abgebildet und diese in den dafür vorgesehenen Bereich des entstehenden Bit-Strings übertragen. Da die Zahl der Pfade, die in Substanzstrukturen vorkommen können, insbesondere für große Pfadlängen n riesig ist, kann dabei nicht für jeden möglichen Pfad ein eigenes Bitmuster reserviert werden. Stattdessen wird der gegebene Pfad über einen Pseudo-Zufallszahlengenerator, dem der Pfad als Keim (*seed*) gegeben wird, eindeutig auf ein Bitmuster, einen sogenannten *Hash*, abgebildet. Dieses Bitmuster wird durch eine logische Oder-Verknüpfung in dem dafür vorgesehenen Bereich dem dortigen Bitmuster hinzugefügt, in welchem bereits die Bitmuster anderer Pfade der gegebenen Länge kodiert sein können.

Die vom Zufallsgenerator vorgenommene *Hash*-Abbildung ist in umgekehrter Richtung nicht eindeutig, d.h. es kann vorkommen, dass ein Bitmuster mehrere Pfade kodiert. Sowohl diese Mehrdeutigkeit als auch die Oder-Verknüpfung der Bitmuster verschiedener Pfade bewirken, dass anhand eines Bitmusters für eine gegebene Substanz nicht mit Sicherheit entschieden werden kann, ob ein fraglicher Pfad in der Struktur der Substanz enthalten ist. Umgekehrt ist es (sehr häufig) möglich, das Vorkommen eines Pfades in der Substanzstruktur mit Sicherheit auszuschließen. Es sei angemerkt, dass Fingerabdrücke häufig für die Durchführung von Substruktursuchen eingesetzt werden und dabei mittels dieser Ausschlussmöglichkeit aufwändigere Einbettungsversuche hinfällig machen können.

Trotz der angesprochenen Mehrdeutigkeiten kodieren Fingerabdrücke Substanzstrukturen in einer gegenüber strukturellen Schlüsseln effizienteren Weise. Während strukturelle Schlüssel üblicherweise spärlich mit Einsen besetzt sind, weil ein typisches Molekül sehr wenige der vielfältigen vorgegebenen Fragmente besitzt, sind Fingerabdrücke vergleichsweise dicht besetzt, ohne dabei Spezifität einzubüßen [156]. Daher stellen Fingerabdrücke Repräsentationen der zugrunde liegenden Substanzstrukturen dar, welche für die Messung struktureller Ähnlichkeiten zwischen Substanzen und daher für das strukturelle Clustering geeignet sind.

4.2.2 Strukturelles Clustering

Der von DataMiner eingesetzte Algorithmus „OptiSim“ [153] bildet Cluster anhand von strukturellen Ähnlichkeiten zwischen den Substanzen, wobei strukturell ähnliche Substan-

zen gemeinsamen Clustern und strukturell unähnliche Substanzen unterschiedlichen Clustern zugeordnet werden. Um die Ähnlichkeit zwischen zwei Substanzen m und n , die durch die Fingerabdrücke \mathbf{b}_m bzw. \mathbf{b}_n repräsentiert sind, zu quantifizieren, wird als Maß der Tanimoto-Koeffizient [157]

$$T_{mn} = \frac{\sum_{i=1}^D b_{mi}b_{ni}}{\sum_{i=1}^D [(b_{mi})^2 + (b_{ni})^2 - b_{mi}b_{ni}]} \quad (4.9)$$

verwendet. Dabei werden die beiden Bit-Strings \mathbf{b}_m und \mathbf{b}_n verglichen, indem einzelne Bits gezählt werden, die auf den Wert 1 gesetzt sind. Im Zähler von (4.9) steht die Anzahl derjenigen Bit-Paare $b_{mi}b_{ni}$ aus den beiden Bit-Strings, deren Bits b_{mi} und b_{ni} gemeinsam (auf 1) gesetzt sind. Über den Ausdruck im Nenner hingegen wird die Anzahl solcher Bit-Paare bestimmt, in welchen das eine oder das andere Bit (oder beide Bits) gesetzt sind. Dazu werden in der Summe über den ersten Term $\sum_{i=1}^D (b_{mi})^2$ die in \mathbf{b}_m gesetzten Bits und in analoger Weise über den zweiten Term $[\sum_{i=1}^D (b_{ni})^2]$ die in \mathbf{b}_n gesetzten Bits gezählt^a. Davon wird schließlich die Anzahl der in beiden Bit-Strings gemeinsam gesetzten Bits abgezogen. Der Tanimoto-Koeffizient T_{mn} gibt also den Anteil der in beiden Bit-Strings gemeinsam gesetzten Bits an. Demzufolge ist der Wertebereich von T_{mn} das Intervall $[0, 1]$, wobei Paaren von Substanzen umso größere Werte zugewiesen werden, je ähnlicher ihre Strukturen sind.

Unter Einsatz des Ähnlichkeitsmaßes (4.9) wird die gegebene Sammlung von Substanzen von OptiSim [153] durch ein zweistufiges Verfahren in Cluster eingeteilt. Im ersten Schritt wählt der Algorithmus aus der Sammlung iterativ einzelne Substanzen aus, die sich gemäß (4.9) paarweise stark unähnlich sind und daher als Repräsentanten der zu bildenden Cluster definiert werden. Im zweiten Schritt werden die restlichen Substanzen einzeln dem jeweils ähnlichsten Repräsentanten zugeordnet und auf diese Weise die Cluster befüllt.

Details der Anwendung auf große Substanzmengen

Soll das beschriebene Clusteringverfahren auf eine große Substanzsammlung angewendet werden, so empfiehlt das Benutzerhandbuch von SARNavigator^b [158], dass die Substanzsammlung zuvor in die Menge der wirksamen und die Menge der unwirksamen Substanzen aufgeteilt und das Clustering nur für die wirksamen Substanzen durchgeführt wird. Die dadurch gebildeten Cluster enthalten dann zunächst ausschließlich wirksame Substanzen. Anschließend können mithilfe von DataMiner die restlichen (unwirksamen) Substanzen nachträglich in die Cluster einsortiert werden. Diese empfohlene Vorgehensweise

^aDas Quadrieren der Einträge b_{mi} (bzw. b_{ni}) in der Definition des Tanimoto-Koeffizienten (4.9) ist nicht notwendig, falls es sich bei den zu vergleichenden Vektoren, wie hier, um Bit-Strings handelt, da dann stets $b_{mi} = (b_{mi})^2$ gilt.

^bDas hier beschriebene Werkzeug DataMiner hieß früher „SARNavigator“.

habe ich für die Durchführung der in Abschnitt 5.2.2 beschriebenen Leitstruktursuche gewählt.

Grafische Darstellung der Cluster in Form einer SAR-Karte

Das Ergebnis eines Clusterings wird von DataMiner in Form einer sogenannten SAR-Karte auf dem Bildschirm dargestellt, in welcher die gebildeten Cluster durch Symbole repräsentiert sind. Die Symbole werden dabei anhand der Ähnlichkeiten gemäß (4.9) zwischen den Repräsentanten der Cluster so in der Ebene verteilt, dass ähnliche Cluster durch nahe beieinander liegende Symbole dargestellt werden und die Symbole für unähnliche Cluster weit entfernt voneinander liegen. Weiter unten ist in Abbildung 5.21 eine solche SAR-Karte dargestellt, und die darin erhaltenen Ergebnisse werden dort diskutiert.

Dem Benutzer von DataMiner ist die Möglichkeit gegeben, Eigenschaften der Cluster wie die Clustergröße oder clusterlokale Anteile von wirksamen Substanzen auszuwählen, anhand derer die repräsentierenden Symbole dargestellt werden sollen. So kann die Form, die Größe und die Farbe der Symbole dazu genutzt werden, charakteristische Eigenschaften der Cluster zu kodieren. Als charakteristische Eigenschaft eines Clusters \mathcal{S} bietet sich neben seiner Größe $|\mathcal{S}|$, d.h. der Anzahl der darin enthaltenen Substanzen, beispielsweise der clusterlokale Anteil

$$P_{\text{AssayA}}(\mathcal{S}) = \frac{| \{ C \in \mathcal{S} \mid C \text{ ist wirksam in Assay A} \} |}{|\mathcal{S}|} \quad (4.10)$$

jener im Cluster \mathcal{S} enthaltenen Substanzen C an, die in einer gewählten Testreihe (Assay A) wirksam waren. Eine auf diese Weise individuell gestaltete SAR-Karte kann vom Benutzer interaktiv hinsichtlich der in den Clustern enthaltenen Substanzen inspiziert und so nach neuen Leitstrukturen durchsucht werden. Das von DataMiner bereitgestellte Clusteringverfahren und die dem Benutzer gegebenen Möglichkeiten, die Clusteringresultate flexibel nach seinen Wünschen grafisch darzustellen, erweisen sich als brauchbare Hilfsmittel für die Identifikation von neuen Leitstrukturen aus einer Vielzahl von durch Screening-Messungen charakterisierten Substanzen.

4.3 Simulationen des Dockings von Wirkstoffen an polyQ-Protofibrillen

Einzelne im Verbund identifizierte anti-Prion-Wirkstoffe wurden in den Gruppen von Prof. F. Ulrich Hartl (Department für zelluläre Biochemie, MPI für Biochemie, Martinsried) und Prof. Christian Haass (Lehrstuhl für Stoffwechselbiochemie, Adolf-Butenandt-Institut, LMU) hinsichtlich ihrer Wirkung gegen die mit der Huntingtonschen Krankheit assoziierte Aggregation von poly-Glutamin (Q)-Peptiden in einem Tiermodell getestet und als

wirksam gefunden. Im Rahmen dieser externen Kooperation führte ich Docking-Simulationen von Modellen dieser Wirkstoffe an ein spezielles Modell für eine polyQ-Protofibrille durch. Unsere Ergebnisse wurden in Schiffer et al. [159] veröffentlicht und werden weiter unten in Abschnitt 5.1.7 beschrieben.

Um zu erkennen, an welchen Stellen eines Rezeptorproteins ein gegebener Wirkstoff binden und dadurch möglicherweise seine Wirksamkeit vermitteln kann, werden üblicherweise experimentelle Methoden der Strukturaufklärung wie NMR- oder Röntgenspektroskopie eingesetzt. Die im Falle der Huntingtonschen Krankheit auftretenden polyQ-Aggregate sind jedoch, wie die PrP^{Sc}-Aggregate der Prionkrankheiten (vgl. Abschnitt 1.5), aufgrund ihrer schlechten Löslichkeit bzw. Kristallisierbarkeit für diese experimentellen Methoden unzugänglich. Um trotz der experimentellen Unzugänglichkeit solcher polyQ-Aggregate zu verstehen, an welchen Stellen dieser Aggregate die untersuchten anti-aggregatorischen Wirkstoffe binden können, habe ich die angesprochenen Docking-Simulationen durchgeführt.

Dabei habe ich als Modell für eine polyQ-Protofibrille eine Struktur verwendet, die von meiner Kollegin Martina Stork gemeinsam mit Paul Tavan [27] entwickelt worden war (vgl. Abschnitt 1.3). Die Struktur ist ein Dimer aus zwei linkshändigen, triangulären β -Helizes, die jeweils 36 Glutamin-Residuen umfassen. Fünf strukturelle Schnappschüsse des in Wasser gelösten Dimers wurden in einem zeitlichen Abstand von 50 ps aus einer Molekulardynamik-Trajektorie entnommen, die in der zitierten Arbeit [27] zur Equilibrierung der modellierten Dimerstruktur mit seiner wässrigen Umgebung bei Zimmertemperatur und Normaldruck gedient hatte. Die Schnappschüsse des polyQ-Dimers wurden für die Docking-Simulationen, die ich mithilfe des Programmpakets AutoDock 3.0 [160] durchführte, unter Verwendung der zugehörigen AutoDockTools (ADT) präpariert, wobei den Atomen der Rezeptorpeptide Partiaalladungen und bestimmte Lösungsmittelparameter zugeordnet wurden. Strukturmodelle für die Ligandmoleküle wurden mithilfe des Quantenchemie-Programms MNDO [161] erzeugt und mit Partiaalladungen versehen. Drehbare Bindungen der Liganden wurden durch ADT definiert.

AutoDock 3.0 modelliert die polyQ-Rezeptorpeptide als starr und die Ligandmoleküle als flexibel und versucht vermittels eines Lamarckschen Genetischen Algorithmus [160] Stellen an der Oberfläche der polyQ-Peptide zu bestimmen, an welchen die Ligandmoleküle binden können. Zur Berechnung der Energie eines Liganden in der Nähe des Rezeptors wendet AutoDock ein Interpolationsschema mit sogenannten Gitterkarten an, die vorab im Bereich des Rezeptors für die in den Liganden vorkommenden Atomarten berechnet werden. Zur Berechnung der Gitterkarten mit $127 \times 127 \times 127$ Punkten im Abstand von 0,375 Å, die um die polyQ-Dimere zentriert waren, habe ich das Werkzeug AutoGrid des Pakets ADT verwendet.

Für jeden der fünf polyQ-Dimer-Schnappschüsse simulierte ich das Docking der ausgewählten Substanzen. Dabei wurden für jedes Rezeptor-Ligand-Paar 100 separate Docking-Läufe durchgeführt und die so gefundenen Dockings nach aufsteigender freier Energie

4 Methoden zur Untersuchung von Struktur-Wirkungs-Beziehungen

sortiert. Die jeweils 20 gedockten Strukturen mit der niedrigsten freien Energie wurden mithilfe des Softwarepakets VMD [162] visualisiert.

Nachdem nun in den beiden vorangegangenen Kapiteln Methoden zur Datenhaltung und -auswertung sowie für SAR-Analysen vorgestellt wurden, kann im nächsten Kapitel über die damit erzielten Resultate berichtet werden.

5 Ergebnisse und Diskussion

Die in den vorangegangenen Kapiteln skizzierten experimentellen Methoden meiner Verbundpartner und die von mir entwickelten Werkzeuge zur Haltung und Auswertung der damit gewonnenen experimentellen Daten sowie zur Untersuchung von Struktur-Wirkungs-Beziehungen konnten im Verbund wohlkoordiniert für die systematische Suche nach neuen anti-Prion-Wirkstoffen eingesetzt werden. Dadurch ist es uns gelungen, zwei neue Wirkstoffklassen zu identifizieren.

Die Entdeckung der ersten dieser beiden neuen Wirkstoffklassen ist das Ergebnis eines mehrstufigen Screenings der Substanzbibliothek DIVERSet 1 unter Einsatz des SIFT-Assays aus TP1, des Zellkultur-Assays aus TP4 und der von mir entwickelten Softwarewerkzeuge zur automatisierten Messdatenauswertung. Gemeinsam mit meinen Kollegen aus TP1 und TP4 haben wir die Screening-Methode und die gefundene Wirkstoffklasse durch ein Patent geschützt [94] und später im *Journal of Virology* veröffentlicht (Bertsch et al. [91]) sowie auf Fachkonferenzen präsentiert [92, 93]. Im folgenden Abschnitt werde ich von der Entdeckung und Weiterentwicklung dieser ersten neuen Wirkstoffklasse berichten.

Die zweite neue Wirkstoffklasse ist ein Resultat des Screenings der DIVERSet-Bibliotheken 1 und 2 im SIFT-Assay und im Zellkultur-Assay von TP3 und wurde von mir bei Leitstruktursuchen in den Daten der beiden Screening-Kampagnen entdeckt. Über die Entdeckung dieser zweiten Wirkstoffklasse, deren Name aus patentrechtlichen Gründen nicht genannt werden kann, werde ich weiter unten in Abschnitt 5.2 in anonymisierter Form berichten.

5.1 Entdeckung und Entwicklung einer ersten neuen Wirkstoffklasse

Die folgende Darstellung der Entdeckung der ersten neuen Wirkstoffklasse, der sogenannten NBB-Klasse (NBB = N'-Benzyliden-Benzohydrazid) orientiert sich an der Publikation Bertsch et al. [91]. Anschließend werden in den Abschnitten 5.1.3 bis 5.1.7 die bisherigen Ergebnisse zur Weiterentwicklung der NBB-Wirkstoffe präsentiert.

5.1.1 Das mehrstufige Screening der DIVERSet 1-Bibliothek

Kurz nachdem die in Abschnitt 2.1 beschriebene Substanzbibliothek DIVERSet 1 vom Verbund beschafft worden war, konnte in TP1 damit begonnen werden, diese Substanzsammlung unter Einsatz des SIFT-Assays (vgl. Abschnitt 2.2) systematisch nach anti-aggregatorischen Verbindungen durchzumustern. Parallel dazu konnten von mir SIFT-Viewer I und die zugehörigen Werkzeuge (vgl. Abschnitt 3.4) rasch entwickelt und eingesetzt werden.

Das primäre Screening im SIFT-Assay

Wie in Abschnitt 2.2 erläutert wurde, wurden im Zuge des primären Screenings der Bibliothek DIVERSet 1 alle Substanzen bei einer einzigen Konzentration (ca. $10 \mu\text{M}^{\text{a}}$) in hohem Durchsatz durchgemustert. Die Messdaten des primären SIFT-Screenings konnten mithilfe der SIFT-Viewer I-Werkzeuge automatisiert aufbereitet und ausgewertet werden. Dabei stellten sich einige der Plattenmessungen als qualitativ schlecht heraus und mussten daher wiederholt werden. Nach dem in Bertsch et al. [91] veröffentlichten Stand des Screenings mussten noch 18 von 125 Platten komplett von der weiteren Auswertung ausgeschlossen werden. Die restlichen 107 Platten mit validen Messdaten wurden der töpfchenweisen Median-basierten Ausreißerererkennung (vgl. Abschnitt 3.1.1) unterzogen, wobei sich nur etwa 7% der Töpfchenmessungen als unbrauchbar erwiesen. Zahlreiche stichprobenartige Überprüfungen der automatisierten Ausreißerererkennung ergaben, dass zumeist Fluoreszenzartefakte (wie Autofluoreszenz oder Quenching, vgl. Abschnitt 3.1) die Ursache der starken Variationen dieser Daten waren. SIFT-Viewer I erwies sich mithin als brauchbares Werkzeug für die automatisierte Aufbereitung und Auswertung der SIFT-Daten sowie für deren Inspektion durch meine Kollegen am ZNP und durch mich. Also konnte mithilfe dieses Werkzeugs die Qualität der SIFT-Daten gesichert werden. Es sei angemerkt, dass die SIFT-Daten zum DIVERSet 1-Screening später in die – damals noch nicht existente – TSE-DB (vgl. Abschnitt 3.5) übertragen wurden.

Anhand der gültigen Töpfchenmessungen ($\sim 93\%$) konnten den zugehörigen DIVERSet 1-Substanzen Primärwirksamkeiten (3.9) zugeordnet werden, welche unter Einsatz der SIFT-Viewer I-Werkzeuge berechnet wurden und welche die anti-aggregatorischen Wirksamkeiten der Testsubstanzen kodieren (vgl. Abschnitt 3.2).

Als Resultat der Wirksamkeitsberechnungen zeigt Abbildung 5.1 die Verteilungen der primären SIFT-Wirksamkeiten für die Mengen der DIVERSet 1-Testsubstanzen (durchgezogene Linie), der Negativ-Kontrollen (gepunktet) und der Positiv-Kontrollen DOSPA (gestrichelt). Wie deutlich zu erkennen ist, liegen die Verteilungen der negativen und positiven Kontrollen klar voneinander getrennt und eng um die Werte 0 bzw. 1 verteilt. Durch

^aAusgehend von der Originalbibliothek, in der die Substanzen bei Konzentrationen von 10 mM vorlagen, wurden über mehrere sequentielle Schritte Verdünnungen von 1:1000 erzeugt [91].

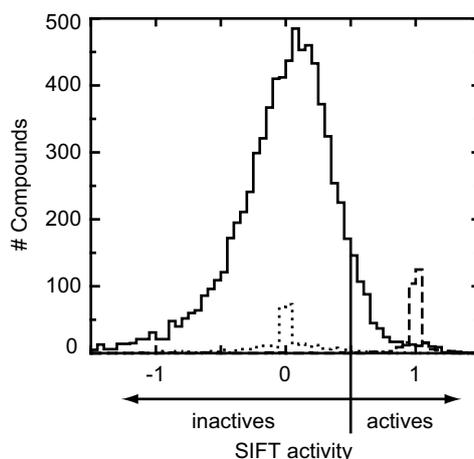


Abbildung 5.1: Verteilung der primären SIFT-Wirksamkeiten für DIVERSet 1. Die Verteilungen der primären SIFT-Wirksamkeiten für die DIVERSet 1-Substanzen (durchgezogene Linie), die Negativ-Kontrollen (gepunktet) und die Positiv-Kontrollen DOSPA (gestrichelt). Die Wirksamkeiten der negativen und positiven Kontrollen liegen eng um die Werte 0 bzw. 1 verteilt und sind deutlich voneinander getrennt. Der Grenzwert zur Definition von wirksamen Substanzen wurde auf 0,5 gesetzt. Übernommen aus Bertsch et al. [91].

die Festlegung eines Grenzwertes von 0,5 wurde die Menge der getesteten Substanzen in die „unwirksamen“ ($< 0,5$) und die „wirksamen“ oder *primären Treffer* ($\geq 0,5$) unterteilt. Gemäß der Definition des Wirksamkeitsmaßes weisen primäre Treffer daher mindestens 50% des anti-aggregatorischen Effektes von DOSPA auf. Aus dieser ersten Runde des SIFT-Screenings erhielten wir 256 primäre Treffer.

Validierung der Primärtreffer im SIFT-Assay

Um die Wirksamkeiten der 256 Primärtreffer zu überprüfen und um Einflüsse der Substanzkonzentrationen auf die Wirksamkeiten zu analysieren, wurden Verdünnungsreihen mit sieben verschiedenen Konzentrationen zwischen $0,1 \mu\text{M}$ und $100 \mu\text{M}$ im SIFT-Assay vermessen. Die Messergebnisse wurden in Form von Dosis-Wirkungs-Kurven, ähnlich den in Abbildung 3.9 A gezeigten, ausgewertet^a. Dabei konnten für 80 Substanzen konzentrationsabhängige anti-aggregatorische Wirksamkeiten bestätigt werden. Vielen der untersuchten Substanzen konnten aus den Dosis-Wirkungs-Kurven zusätzlich EC₅₀-Werte (vgl. Abschnitt 2.2.4) zugeordnet werden, die im Bereich zwischen $0,3 \mu\text{M}$ und $60 \mu\text{M}$ lagen^b.

^aDie von SIFT-Viewer II gebotene Fähigkeit, Dosis-Wirkungs-Kurven zu generieren, stand während der Validierung von DIVERSet 1 noch nicht zur Verfügung und wurde erst beim Screening von DIVERSet 2 genutzt.

^bDie genannten Substanzanzahlen beziehen sich auf den in Bertsch et al. [91] veröffentlichten Stand. Inzwischen konnte die Wirksamkeit von insgesamt 130 Primärtreffern aus DIVERSet 1 bestätigt werden, und für 116 der bestätigten Substanzen konnten EC₅₀-Werte bestimmt werden.

Vergleich von SIFT- und Konversions-Assay

Neben der zentralen SIFT-Screening-Kampagne wurde zusätzlich eine Menge von 70 SIFT-wirksamen Substanzen von unseren Verbundpartnern aus TP3 mithilfe des *in vitro* PrP^C-PrP^{res}-Konversions-Assays überprüft (vgl. Abschnitt 2.4). Im Konversions-Assay zeigten sich 15 der 70 getesteten Substanzen als aktiv. Eine Analyse dieser Daten ergab, dass vom Konversions-Assay tendenziell solche Substanzen als aktiv erkannt wurden, die durch sehr hohe SIFT-Aktivitäten charakterisiert sind, was durch die SIFT-EC50-Werte belegt wird, die für 48 der getesteten 70 Substanzen vorlagen. Während die (elf) im Konversions-Assay aktiven Substanzen einen mittleren SIFT-EC50-Wert von 3,1 μM aufweisen, liegt jener der verbleibenden (37) dort inaktiven Substanzen deutlich höher bei 14,6 μM . Die Übereinstimmung der mit den beiden Methoden erzielten Ergebnisse ist auf die Ähnlichkeit der Assays zurückzuführen, denn in beiden Verfahren wird die anti-aggregatorische Wirksamkeit gemessen.

Zellkulturtest der validierten SIFT-Treffer

Die oben genannten 80 validierten Treffer-Substanzen des *in vitro* SIFT-Assays wurden im ScN2a-Zellkulturmodell von TP4 (vgl. Abschnitt 2.3) hinsichtlich ihrer anti-Prion-Wirksamkeiten untersucht. Dabei wurden in einer ersten Runde alle 80 Substanzen bei einer Konzentration von etwa 15 μM getestet. Die Western-Blots aus Abbildung 2.9 stammen aus dieser ersten Runde und zeigen exemplarisch, dass einige der getesteten Substanzen die Zellkulturen von PrP^{Sc} befreien können, indem sie die Bildung von PrP^{Sc} hemmen oder den Abbau von PrP^{Sc} anregen. In diesem Zellkulturtest wurden acht Substanzen als wirksam gefunden. Sechs dieser acht Substanzen waren in ausreichender Menge vorhanden, sodass sie zusätzlich bei einer Konzentration von 10 μM evaluiert werden konnten. Auch bei dieser reduzierten Konzentration erwiesen sich noch vier dieser sechs Substanzen als wirksam.

Diese vier Substanzen wurden anschließend im Zellkulturtest einer Dosis-Wirkungs-Analyse unterzogen. Abbildung 5.2 zeigt die resultierenden Western-Blots der aus den Zelllysaten präparierten unlöslichen Fraktionen (vgl. Abschnitt 2.3), welche das von den Zellen produzierte PrP^{Sc} enthalten. Die schwachen Banden für die Substanzen 293G02 und 313B02^a zeigen an, dass diese Substanzen die Zellen besonders wirksam von PrP^{Sc} befreien können. Quantitative Analysen der Western-Blots für diese beiden Substanzen ergaben EC50-Werte von ungefähr 2 μM bzw. 6 μM .

Als Zusammenfassung der Ergebnisse des mehrstufigen Screenings im SIFT- und im Zellkultur-Assay zeigt die Tabelle in Abbildung 5.3, wie ausgehend von den 10.000 Substanzen der DIVERSet 1-Bibliothek die Menge der wirksamen Substanzen schrittweise ein-

^aHier wurden Kurzformen der gemäß (3.11) festgelegten Substanznamen gewählt.

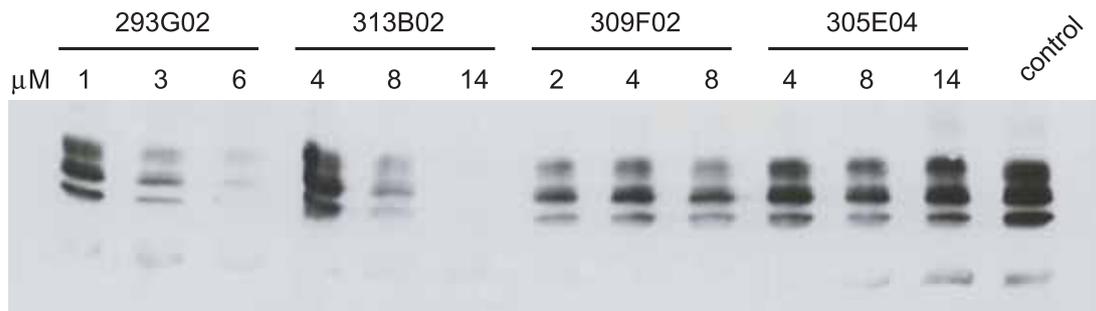


Abbildung 5.2: Dosis-Wirkungs-Analyse im Zellkulturmodell. ScN2a-Zellen wurden für zwei Tage mit den ausgewählten Substanzen 293G02, 313B02, 309F02 und 305E04 in verschiedenen Konzentrationen von einigen μM behandelt und biochemisch analysiert (vgl. Abschnitt 2.3). Dabei wurde vor der Western-Blot-Analyse die unlösliche Fraktion mit der Proteinase K behandelt, um das in den Zelllysaten enthaltene PrP^{Sc} quantifizieren zu können. Im Gegensatz zu unbehandelte Zellen (control), zeigen vor allem die mit den Testsubstanzen 293G02 und 313B02 behandelten Zellen konzentrationsabhängige Reduktionen der PrP^{Sc}-Banden. Übernommen aus Bertsch et al. [91].

Substanzmenge	Anzahl von Substanzen
DIVERSet1 (125 Platten) + Kontrollen (8 pro Platte)	10.000 (+1000 Kontrollen)
SIFT Primärtreffer (TP1)	256
in SIFT-Verdünnungsreihen validierte Treffer (TP1)	80
in Zellkultur validierte Treffer (TP4)	6

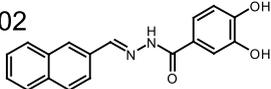
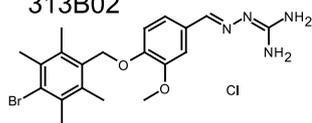
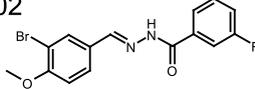
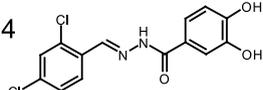
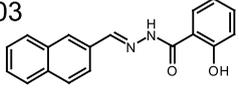
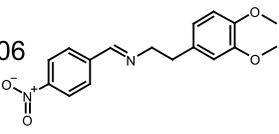
Abbildung 5.3: Statistik des Screenings von DIVERSet 1. Aufgelistet sind die Zahlen der Substanzen, die in den verschiedenen Schritten des Screenings untersucht bzw. als wirksam gefunden wurden. Ausgehend von den 10.000 Substanzen aus DIVERSet 1, wurde die Menge der wirksamen Substanzen unter Einsatz der Assay-Systeme schrittweise eingeschränkt, bis schließlich sechs Substanzen mit validierter SIFT- und Zellkulturwirksamkeit identifiziert waren.

geschränkt wurde bis schließlich nurmehr sechs Substanzen übrig geblieben waren, die in den verwendeten Tests starke anti-Prion-Wirksamkeiten zeigten.

Die Entdeckung der Wirkstoffklasse NBB

Abbildung 5.4 zeigt die Molekülstrukturen der aus dem SIFT-Screening (TP1) und den Zellkulturvalidierungen (TP3) resultierenden sechs Substanzen zusammen mit ihren Wirksamkeiten im primären SIFT-Screening (erste Spalte), in den SIFT-Verdünnungsreihen (zweite Spalte) und im Zellkultursystem (letzte Spalte). Vier dieser Substanzen (293G02, 309F02, 305E04 und 297F03) besitzen das gemeinsame Kernstrukturmotiv NBB, das in Abbildung 5.5 gezeigt ist. Wie eine daraufhin von mir durchgeführte Substruktursuche in

5 Ergebnisse und Diskussion

Compound	Structure	SIFT prim. activ.	SIFT EC50 [μ M]	ScN2a EC50 [μ M]
293G02		0.51	0.3	~2
313B02		0.57	3.5	~6
309F02		n.d.	30	+
305E04		0.73	+	+
297F03		0.43	30	(+)
260D06		0.39	n.d.	(+)

n.d. not determined. + inhibition confirmed by dilution series.
(+) confirmation pending.

Abbildung 5.4: Substanzen mit Zellkulturwirksamkeit. Gezeigt sind die Molekülstrukturen dieser sechs Substanzen und ihre Wirksamkeiten in den drei Schritten des Screenings. Die erste Spalte zeigt die Wirksamkeiten im primären SIFT-Screening (SIFT prim. activ.). Diese Substanzen wurden in SIFT-Verdünnungsreihen validiert, und, falls möglich, wurden EC50-Werte bestimmt (mittlere Spalte). Die letzte Spalte kombiniert die Ergebnisse mehrerer Schritte im Zellkultursystem. Vier dieser Substanzen besitzen das gemeinsame Strukturmotiv NBB. Übernommen aus Bertsch et al. [91].

der DIVERSet 1-Bibliothek ergab, enthalten 413 der 10.000 Substanzen das NBB-Strukturmotiv. Die auffällige Häufung des NBB-Motivs in der Menge der Zellkultur-Treffer legte die Vermutung nahe, dass für die Substanzen der NBB-Klasse relevante Beziehungen zwischen Molekülstruktur und anti-Prion-Wirksamkeit bestehen.

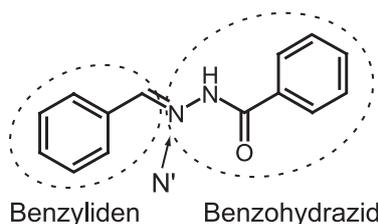


Abbildung 5.5: NBB-Struktur. Die Struktur NBB setzt sich aus einer Benzyliden- und einer Benzohydrazid-Gruppe zusammen, die über ein N'-Atom verbunden sind.

5.1.2 Struktur-Wirkungs-Beziehungen von NBB-Substanzen

Um die Frage nach Struktur-Wirkungs-Beziehungen von NBB-Substanzen anhand der im primären Screening gesammelten SIFT-Wirksamkeiten zu beantworten, habe ich mithilfe der in Abschnitt 4.1 beschriebenen Softwarewerkzeuge (MoSS, hMoSS etc.) systematisch NBB-Substrukturklassen gebildet und auf diese Weise die Einflüsse von verschiedenen Substituenten der NBB-Kernstruktur auf die SIFT-Aktivitäten untersucht. Die Vielzahl der von dem Werkzeug MoSS gebildeten NBB-Substrukturklassen konnte mithilfe des Werkzeugs hMoSS übersichtlich in hierarchischer Form dargestellt werden, wie es schon in Abbildung 4.4 exemplarisch gezeigt wurde. Diese Form der Darstellung ermöglichte die Auswahl einer Reihe von NBB-Substrukturklassen mit ähnlichen Wirksamkeitsprofilen und damit die Identifikation von Struktur-Wirkungs-Beziehungen.

Abbildung 5.6 zeigt eine Gruppe von NBB-Substanzklassen, die auf diese Weise ausgewählt wurden. Das in der ersten Zeile gezeigte allgemeine NBB-Kernmotiv wird von den übrigen Strukturmotiven um spezifische Substituenten erweitert. Gemäß den Erläuterungen in Abschnitt 4.1 definiert jedes dieser Struktur motive eine Klasse von Substanzen aus der DIVERSet 1-Bibliothek, deren Strukturen das jeweilige Motiv beinhalten. Neben den Strukturmotiven sind die klassenlokalen Verteilungen der SIFT-Wirksamkeiten gezeigt, wobei die Verteilungen von stark besetzten Klassen zudem durch Box-and-Whisker-Plots (vgl. Abschnitt 4.1) charakterisiert sind.

Die Klasse aller 413 NBB-Substanzen (erste Zeile) enthält viele unwirksame Substanzen, aber dennoch ist die Wirksamkeitsverteilung auch hier schon leicht in Richtung positiver anti-Prion-Wirksamkeiten verschoben. Im Bezug auf Substitutionen am Benzyl-Ring der Benzohydrazid-Gruppe (vgl. Abbildung 5.5) analysierten wir den Einfluss von Hydroxyl-Substituenten an unterschiedlichen Positionen. Das Hinzufügen einer Hydroxyl-Gruppe in ortho-Position (zweite Zeile) führt zu einer Subklasse mit nur geringfügig besserer mittlerer Wirksamkeit. Hingegen zeigen die Subklassen mit Hydroxyl-Gruppen in meta- oder para-Position (Zeile 3 und 4) erhöhte Wirksamkeiten. Insbesondere weisen diejenigen vier Substanzen aus DIVERSet 1, die eine Kombination von Hydroxyl-Gruppen in der meta- und der para-Position (Zeile 6) besitzen, hohe SIFT-Wirksamkeiten auf. Zudem waren zwei dieser Substanzen (293G02 und 305E04) auch in Zellkultur wirksam (vgl. Abbildung 5.6). Was die N'-Seite des NBB-Kerns anbelangt, fanden wir

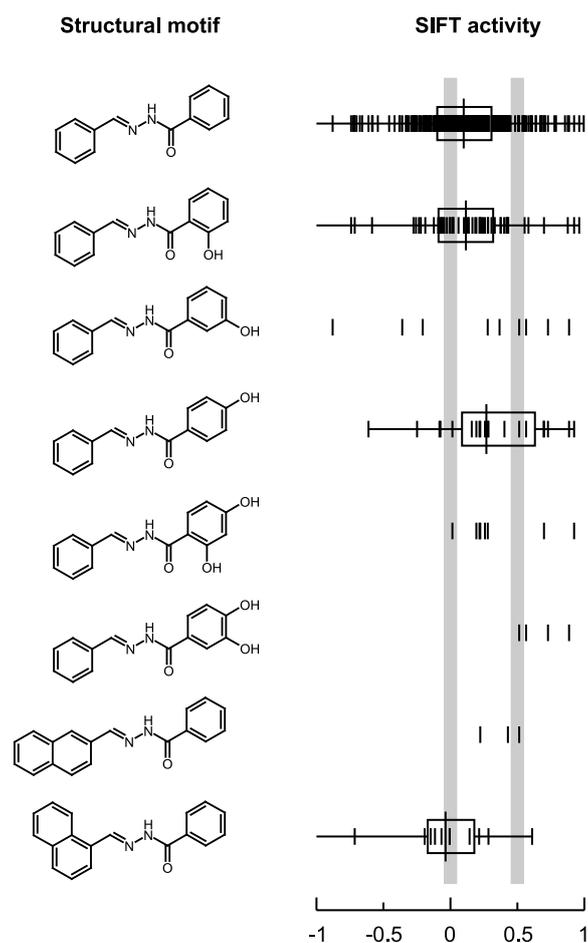


Abbildung 5.6: Struktur-Wirkungs-Beziehungen für Klassen von NBB-Derivaten. Für acht NBB-Substanzklassen sind die jeweils enthaltenen Substrukturmotive und die Verteilungen der primären SIFT-Wirksamkeiten gezeigt. Die Kästen und langen vertikalen Balken innerhalb der stark besetzten Wirksamkeitsverteilungen markieren den Median und die Quartile. Diese Statistik wurde für Klassen, die nur wenige Substanzen enthalten, weggelassen. Die beiden vertikalen grauen Balken markieren das Zentrum des unwirksamen Bereichs bei 0 und die Grenze zum wirksamen Bereich bei 0,5. Übernommen aus Bertsch et al. [91].

einen eindrucksvollen Effekt von Naphthalenylmethylen-Substitutionen. Alle Verbindungen, die diesen Rest in Form eines Naphthalen-2-ylmethylen-Isomers (vorletzte Zeile) enthalten, zeigen erhöhte SIFT-Aktivitäten, während die mittlere Aktivität der Verbindungen mit dem Naphthalen-1-ylmethylen-Isomer (letzte Zeile) nahe bei Null liegt. Eine signifikante Struktur-Wirkungs-Beziehung wird dadurch nahe gelegt, dass zwei Verbindungen (293G02 und 297F03) mit Naphthalen-2-ylmethylen-Gruppen sich unter den Zellkultur-Wirksamen finden. Interessanterweise ist diejenige Verbindung (293G02) mit der stärksten Zellkultur-Wirksamkeit dadurch charakterisiert, dass in ihrer Struktur (siehe Abbildung 5.7) die beiden mit hoher SIFT-Aktivität assoziierten Motive, d.h. die Naphthalen-

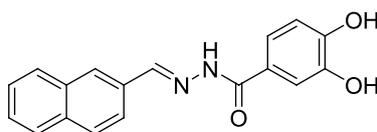


Abbildung 5.7: Substanz 293G02. Diese Substanz zeigte die stärkste Wirksamkeit im Zellkulturtest. Ihre Struktur vereint die beiden mit hoher SIFT-Aktivität assoziierten Motive, nämlich die Naphthalen-2-ylmethylengruppe am N'-Atom und die Hydroxylgruppen in meta- und para-Position am Benzyl-Ring der Benzohydrazid-Gruppe (vgl. Abbildung 5.5).

2-ylmethylengruppe und die Hydroxylgruppen in meta- und para-Position, in Kombination auftreten.

Auf die gleiche Weise wurden eine Reihe von weiteren NBB-Substrukturklassen hinsichtlich ihrer Wirksamkeiten im SIFT-Assay untersucht. Dabei konnte beispielsweise gezeigt werden, dass Substitutionen von bestimmten Halogenen (Fluor, Chlor, Brom oder Jod) an verschiedenen Positionen des Benzyl-Rings der Benzyliden-Gruppe (vgl. Abbildung 5.5) die SIFT-Aktivität des NBB-Motivs erhöhen können. Diese Ergebnisse wurden in unserem Patent veröffentlicht [94].

Die Korrelationen zwischen SIFT- und Zellkultur-Wirksamkeiten für die diskutierten strukturellen Motive zeigen, dass es möglich ist, die Daten aus dem primären Hochdurchsatz-Screening im SIFT-Assay für die Evaluierung von Struktur-Wirkungs-Beziehungen einzusetzen. Des Weiteren weisen diese Korrelationen darauf hin, dass die in Zellkultur gefundene anti-Prion-Wirksamkeit tatsächlich auf eine Inhibierung der Koaggregation von PrP^C an PrP^{Sc} zurückzuführen ist. Die systematischen Substruktursuchen deckten ferner Substituentenkombinationen auf, die mit erhöhter Wirksamkeit verbunden sind. Die Vielzahl der so gefundenen NBB-Subklassen validierte das NBB-Motiv als Leitstruktur, die sich somit zur weiteren Optimierung in Nachfolgeuntersuchungen anbot.

5.1.3 Fokussierte NBB-Bibliotheken

Nachdem die NBBs als neue anti-Prion-Wirkstoffe entdeckt und ihr Charakter als Leitstruktur durch statistische SAR-Analysen belegt worden war, traten wir (TP1 und TP2) gegen Ende der ersten Förderperiode unseres Verbundprojektes in Kontakt mit dem Medizinalchemiker Prof. Boris Schmidt (TU Darmstadt), der schließlich für die zweite Förderperiode als Verbundpartner gewonnen werden konnte (vgl. Abschnitt 2.6). Meine Kollegen Dr. Bertsch und Dr. Giese vom ZNP und ich trafen uns mit Prof. Schmidt und präsentierten ihm die SIFT-Ergebnisse für die NBB-Substanzklasse, deren zahlreiche Subklassen mithilfe des Werkzeugs hMoSS hierarchisch angeordnet worden waren (vgl. z.B. Abbildung 4.4). Dadurch konnte sich Prof. Schmidt rasch einen Überblick darüber verschaffen, welche Subklassen von NBB-Substanzen bereits getestet worden waren und welche Substituenten rund um das NBB-Kernmotiv die aggregationshemmende Wirkung positiv oder negativ beeinflussen können. Darüber hinaus wurden pharmakologische Eigenschaften ein-

zelter Substanzen diskutiert, die auf physikochemische Eigenschaften der zugehörigen Moleküle oder einzelner Atomgruppen zurückgeführt werden, wie etwa die möglichen Schwierigkeiten, die Blut-Hirn-Schranke zu überwinden, welche polaren Substanzen zugeschrieben werden [141]. Beispielsweise gab Prof. Schmidt zu bedenken, dass es bei der Substanz 293G02 (vgl. Abbildung 5.7) aufgrund der beiden polaren Hydroxyl-Gruppen, möglicherweise Probleme mit der Bioverfügbarkeit dieser Substanz im lebenden Organismus geben könnte. Derartige medizinisch-chemische Überlegungen flossen in die Auswahl der Bausteine für die Synthese von neuen NBB-Substanzen ein.

Synthese einer ersten Runde von neuen NBB-Substanzen

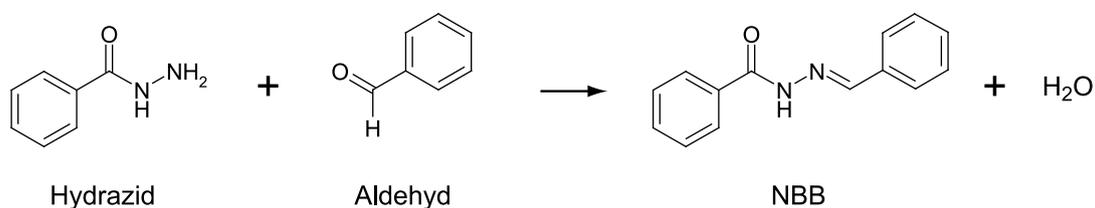


Abbildung 5.8: Synthese von NBB-Substanzen. Aus Benzoesäurehydrazid und Benzaldehyd kann unter Abspaltung von Wasser NBB synthetisiert werden.

Wie in Abbildung 5.8 gezeigt ist, können neue NBB-Substanzen aus Hydraziden und Aldehyden unter Wärmezufuhr einfach hergestellt werden. Durch den Einsatz der kombinatorischen Chemie, bei der spezifische Substituenten am Hydrazid mit spezifischen Substituenten am Aldehyd kombiniert werden, konnte deshalb im Labor von Prof. Schmidt innerhalb von kurzer Zeit eine umfangreiche fokussierte Bibliothek neuer NBB-Derivate synthetisiert werden. Insgesamt wurden dabei auf Mikrotiterplatten zeilenweise zehn unterschiedliche Hydrazide, die mit den Kürzeln H₁ bis H₁₀ bezeichnet seien, mit spaltenweise 33 unterschiedlichen Aldehyden (A₁-A₃₃) zu 330 NBB-Derivaten kombiniert. Die Substituenten der verwendeten Aldehyde und Hydrazide sollten die anti-Prion-Wirksamkeit erhöhen und zugleich die pharmakologischen Eigenschaften der NBB-Substanzen verbessern.

Abbildung 5.9 zeigt schematisch, wie die 330 neu synthetisierten NBB-Substanzen auf fünf 96-Töpfchen-Mikrotiterplatten, Platte I bis Platte V, angeordnet wurden. Wie in der Bildunterschrift beschrieben ist, wurden dabei auf den Platten I, II und III die sieben Hydrazide H₁ bis H₇ und jeweils elf Aldehyde (A₁-A₁₁, A₁₂-A₂₂ und A₂₃-A₃₃) sowie deren 77 NBB-Kombinationen und auf den Platten IV und V die verbliebenen Hydrazide H₈ bis H₁₀ und die jeweiligen Aldehyde und deren Kombinationen angeordnet. Die auf diese Weise befüllten Platten I bis V wurden zur Vermessung im SIFT-Assay ans ZNP geschickt.

5.1 Entdeckung und Entwicklung einer ersten neuen Wirkstoffklasse

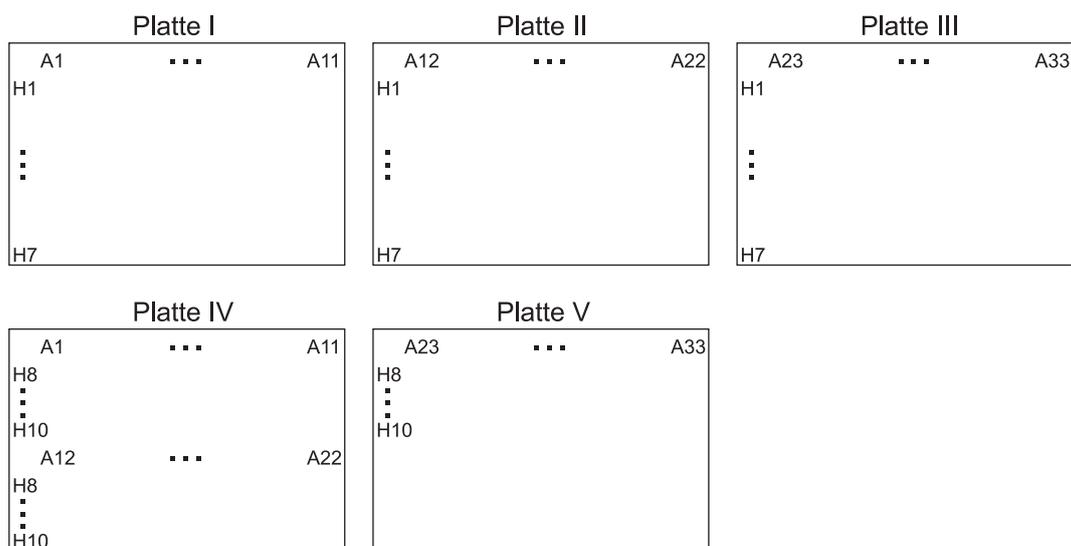


Abbildung 5.9: Belegungsschema der fünf Platten I bis V mit neu synthetisierten NBB-Substanzen. Die Hydrazide H₁ bis H₁₀ und die Aldehyde A₁ bis A₃₃ sowie die daraus kombinierten NBB-Derivate wurden auf fünf Mikrotiterplatten (I bis V) mit je 96 Töpfchen verteilt. Dabei waren jeweils zeilenweise unterschiedliche Hydrazide und spaltenweise unterschiedliche Aldehyde angeordnet. Auf Platte I sind in der ersten Spalte (ab der zweiten Zeile) die Hydrazide H₁ bis H₇, in der ersten Zeile (ab der zweiten Spalte) die Aldehyde A₁ bis A₁₁ und in den restlichen Töpfchen (außer in dem leeren Töpfchen in der ersten Zeile und ersten Spalte) deren Kombinationen angeordnet. Den Substanzen der Platten II und III liegen zeilenweise die selben Hydrazide (H₁-H₇) aber spaltenweise andere Aldehyde (A₁₂-A₂₂ bzw. A₂₃-A₃₃) zugrunde. Die verbliebenen Hydrazide H₈ bis H₁₀ wurden auf den Platten IV und V mit den jeweiligen Aldehyden kombiniert.

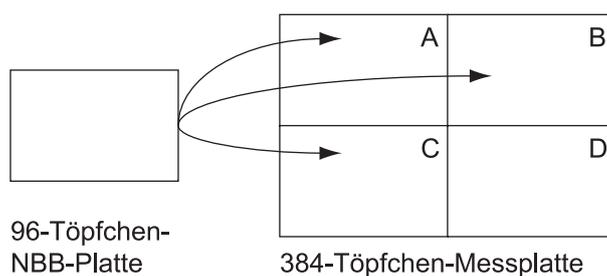


Abbildung 5.10: Befüllung einer 384-Töpfchen-Messplatte. Die NBB-Substanzen einer 96-Töpfchen-Platte wurden bei drei verschiedenen Konzentrationen auf die drei Quadranten A, B und C einer 384-Töpfchen-Messplatte übertragen und mit der SIFT-Technik gemessen. Kontrollen wurden nur in Quadrant D mitgeführt. Dadurch sollte eine Dosis-Wirkungs-Analyse ermöglicht werden.

SIFT-Messung und -auswertung

Um die neu synthetisierten NBB-Substanzen einer Dosis-Wirkungs-Analyse im SIFT-Assay zu unterziehen, wurden, wie in Abbildung 5.10 skizziert ist, die Substanzen einer 96-Töpfchen-Platte jeweils bei drei verschiedenen Konzentrationen ($10\ \mu\text{M}$, $1\ \mu\text{M}$, $0,1\ \mu\text{M}$) auf die drei Quadranten A, B und C einer 384-Töpfchen-Messplatte übertragen. Töpfchen mit positiven und negativen Kontrollmischungen wurden dabei in Quadrant D mitgeführt. Die auf diese Weise befüllten 384-Töpfchen-Platten wurden mit der SIFT-Technik vermessen und die Messdaten mithilfe von SIFT-Viewer II ausgewertet.

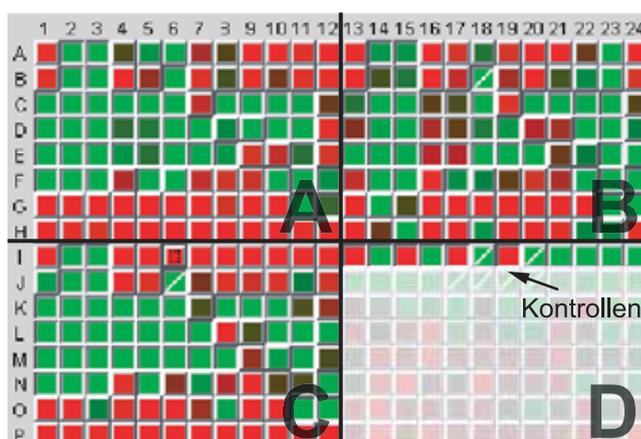


Abbildung 5.11: SIFT-Ergebnisse für eine 384-Töpfchen-Plattenmessung einer 96-Töpfchen-Platte mit neuen NBB-Substanzen. Gezeigt ist die von SIFT-Viewer II generierte Repräsentation einer 384-Töpfchen-Platte, auf der in den Quadranten A, B und C in drei unterschiedlichen Konzentrationen jeweils die NBB-Substanzen einer 96-Töpfchen-Platte (Platte II) im SIFT-Assay gemessen wurden. In der ersten Zeile von Quadrant D wurden Positiv- und Negativ-Kontrollmessungen mitgeführt. Die Töpfchenrepräsentationen sind gemäß den quadrantenlokal Median-basierten Wirksamkeiten (vgl. Abschnitt 3.2.3) eingefärbt, wobei grüne Farbtöne hohe und rote Farbtöne niedrige Wirksamkeiten kennzeichnen. Mithilfe von SIFT-Viewer II wurden die Töpfchen von Substanzen ausgewählt, die in mehreren Quadranten erhöhte Wirksamkeiten zeigen – die zugehörigen Knöpfe erscheinen daher als gedrückt. Im weiß überdeckten Bereich wurden, um die Plattenkapazität zu nutzen, weitere Substanzen getestet, deren Resultate hier nicht relevant sind.

Abbildung 5.11 zeigt die von SIFT-Viewer II erzeugte Repräsentation der Resultate einer solchen Messung der NBB-Platte II und soll zunächst dazu dienen, Probleme zu erläutern, die durch das bei den Messungen eingesetzte Plattenlayout verursacht wurden. Das oben beschriebene quadrantenweise Befüllen der Messplatten führte dazu, dass die Mischungen der unterschiedlichen Quadranten vor der Messung unterschiedlich lange angesetzt waren. Daher war die Wirkdauer der zu testenden Substanzen quadrantenweise unterschiedlich. Ferner bildeten sich in unterschiedlichem Maße beispielsweise Ablagerungen an den Wänden der Töpfe. Dies führte bei der Auswertung dazu, dass die Wirksamkeiten der Substanzen aus den Quadranten A, B und C nicht auf die Kontrollmessungen bezogen

werden konnten, die, wie erwähnt, nur in Quadrant D mitgeführt wurden. Demzufolge konnte das kontrollenbezogene Wirksamkeitsmaß (3.9) zur Auswertung der Messdaten nicht eingesetzt werden.

Um dennoch eine Auswertung zu ermöglichen, wurde das in Abschnitt 3.2.3 beschriebene Median-basierte Wirksamkeitsmaß (3.10) quadrantenweise eingesetzt, wobei jeweils quadrantenlokale Mediane der Summenobservablen S_1^5 für die Testsubstanzen gebildet wurden. Die resultierenden Wirksamkeitswerte geben an, ob und in welchem Maße eine gegebene Testsubstanz in einem gegebenen Quadranten stärker oder schwächer anti-aggregatorisch wirkt als die zum Median gehörige(n) Substanz(en). In Abbildung 5.11 sind die Töpfchenrepräsentationen gemäß diesem quadrantenlokal gebildeten, Median-bezogenen Wirksamkeitsmaß eingefärbt, wobei grüne Farbtöne für höhere und rote Farbtöne für niedrigere Wirksamkeiten stehen. Wie beispielsweise in Quadrant A zu erkennen ist, zeigen die meisten der in den Zeilen C bis F getesteten Substanzen überdurchschnittlich hohe Wirksamkeiten, während die Substanzen aus den Zeilen G und H weniger wirksam waren. In ähnlicher Weise sind in den Spalten 2, 3, 6 und 11 viele Substanzen stärker und in Spalte 12 viele Substanzen schwächer wirksam. Dieses Muster der gefärbten Töpfchen aus Quadrant A wiederholt sich in ähnlicher Weise in den Quadranten B und C für die dort bei den beiden anderen Konzentrationen gemessenen Substanzen. Um dieses Muster hervorzuheben, wurden in der gezeigten Plattenrepräsentation solche Töpfchen ausgewählt, deren Substanzen in mehreren Quadranten erhöhte Wirksamkeiten zeigen. Die zugehörigen Knöpfe erscheinen daher als gedrückt. Auf diese Weise konnten die bei drei Substanzkonzentrationen durchgeführten Messungen zumindest qualitativ ausgewertet werden. Um die Qualität der SIFT-Daten in zukünftigen Screening-Runden zu sichern, wurde das Layout der 384er-Platten modifiziert, sodass nun wieder in ausreichendem Maße Kontrollmessungen mitgeführt werden.

Zur Analyse von Struktur-Wirkungs-Beziehungen für die neu synthetisierten NBB-Substanzen wurden den qualitativen SIFT-Auswertungen, wie den oben diskutierten, die Strukturen der zugrunde liegenden Hydrazide und Aldehyde zugeordnet. Abbildung 5.12 zeigt dazu exemplarisch für die Messplatte aus Abbildung 5.11 die Töpfchenrepräsentationen des Quadranten A, auf welcher die gedrückten Knöpfe die zugehörigen Töpfchen von Substanzen hervorheben, die bei mehreren Konzentrationen wirksam waren. Denjenigen Zeilen und Spalten, die überwiegend als wirksam identifizierte Substanzen enthalten, sind die Strukturen der zugrunde liegenden Hydrazide H_2 , H_3 , H_4 und H_5 sowie der Aldehyde A_{12} , A_{13} , A_{16} und A_{21} zugeordnet. Dabei erscheinen solche Substanzen, die aus diesen Hydraziden und Aldehyden kombiniert sind, als besonders vielversprechend. Dazu zählt beispielsweise die Substanz aus Töpfchen C2, die aus dem Hydrazid H_2 und dem Aldehyd A_{12} zusammengesetzt ist.

In der gleichen Weise wurden die SIFT-Messungen der übrigen NBB-Platten ausgewertet und die Ergebnisse mit den Strukturen der Synthesebausteine zusammengestellt (Daten nicht gezeigt). Dabei wurden die Wirksamkeiten von NBB-Derivaten der Hydrazide H_2 bis H_5 durch die Resultate für die Platten I und III bestätigt, auf welchen weitere Derivate

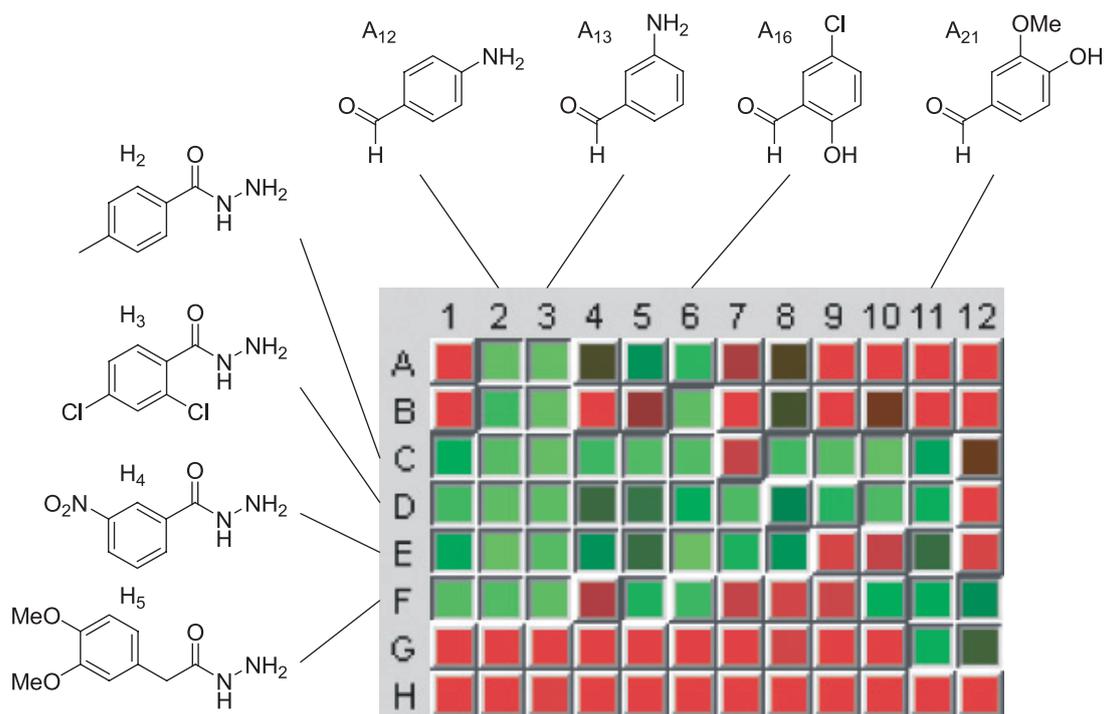


Abbildung 5.12: SAR-Analyse für die NBB-Derivate von Platte II. Die Töpfchenrepräsentationen von Quadrant A für die NBB-Platte II zusammen mit den Strukturen der Hydrazide H₂, H₃, H₄ und H₅ sowie der Aldehyde A₁₂, A₁₃, A₁₆ und A₂₁.

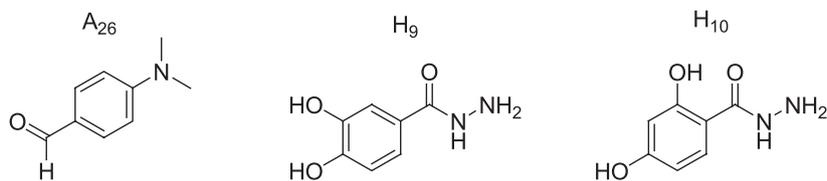


Abbildung 5.13: Mit Wirksamkeit assoziierte Hydrazide und Aldehyde. NBB-Derivate des Aldehyds A₂₆ und der Hydrazide H₉ und H₁₀ zeigten im SIFT-Assay erhöhte Wirksamkeiten.

dieser Hydrazide enthalten waren (vgl. Abbildung 5.9). Darüber hinaus zeigten Derivate des Aldehyds A₂₆, dessen Struktur in Abbildung 5.13 gezeigt ist, vermehrt erhöhte Wirksamkeit. Auf den Platten IV und V waren Derivate der Hydrazide H₉ und H₁₀ (siehe Abbildung 5.13) enthalten, welche jeweils zwei Hydroxyl-Gruppen am Phenyl-Ring der Hydrazid-Gruppe aufweisen, und jenen Klassen angehören, die bereits beim Screening der DIVERSet-Bibliotheken durch hohe Wirksamkeiten aufgefallen waren (vgl. Abbildung 5.6). Auch zahlreiche neu synthetisierte Substanzen dieser beiden Klassen zeigten erhöhte Wirksamkeiten im SIFT-Assay.

Weitere Synthese-Runden

NBB-Substanzen der ersten Runde, die sich, wie oben dargestellt, als wirksam herausgestellt hatten, wurden in einer zweiten Runde erneut synthetisiert, um Verdünnungsreihenmessungen im SIFT-Assay zu ermöglichen. Dabei wurden die Substanzen, ähnlich wie bei den Validierungsmessungen der DIVERSet-Primärtreffer, jeweils zweifach bei sieben Konzentrationen zwischen $0,1 \mu\text{M}$ und $100 \mu\text{M}$ im SIFT-Assay vermessen. Da bei diesen Messungen wieder Kontrollen in ausreichendem Umfang mitgeführt wurden, konnte für die Berechnung der Wirksamkeiten hier wieder das kontrollenbezogene Maß (3.9) eingesetzt werden.

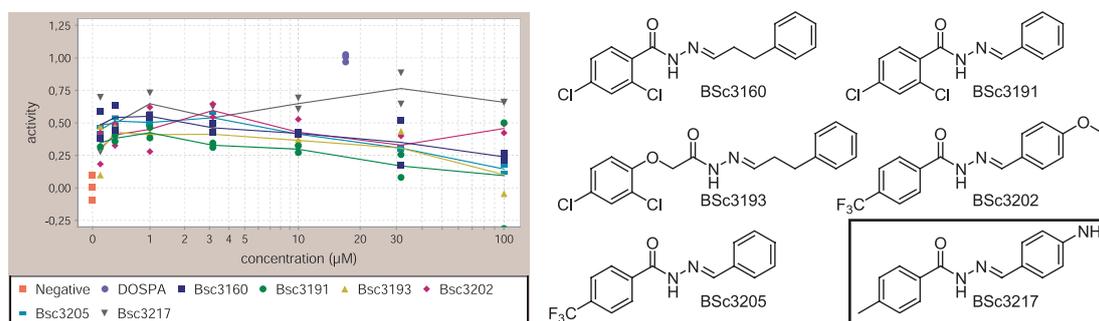


Abbildung 5.14: SIFT-Dosis-Wirkungs-Analysen für neue NBB-Derivate. Gezeigt sind die Dosis-Wirkungs-Kurven, die für sechs NBB-Derivate bei Konzentrationen zwischen $0,1 \mu\text{M}$ und $100 \mu\text{M}$ im SIFT-Assay bestimmt wurden. Wie in der Legende angegeben ist, sind hier die Werte für negative und positive Kontrollen durch rote Quadrate bzw. blaue Kreise und die Werte der Testsubstanzen durch weitere Symbole repräsentiert. Daneben sind die Strukturen der getesteten Substanzen dargestellt. In dieser Versuchsreihe zeigte die Substanz BSc3217 erhöhte Wirksamkeitswerte (graue Dreiecke). Die Struktur dieser Substanz ist daher eingerahmt.

Abbildung 5.14 zeigt exemplarisch die Resultate von solchen Verdünnungsreihenmessungen für sechs NBB-Derivate. Hier sind die mithilfe von SIFT-Viewer II erzeugten Dosis-Wirkungs-Kurven zusammen mit den Strukturen der getesteten Substanzen dargestellt. Wie der Legende des Dosis-Wirkungs-Plots entnommen werden kann, sind die „Wirksamkeiten“ der Negativ-Kontrollen (bei $0 \mu\text{M}$) durch rote Quadrate symbolisiert und liegen um den Wert 0 verteilt. Die positiven DOSPA-Kontrollen (blaue Kreise bei $17 \mu\text{M}$) weisen Werte nahe bei 1 auf. Während fünf der sechs getesteten Substanzen nur moderate Wirksamkeiten in einem breiten Bereich bei etwa 0,5 zeigen, die mit steigenden Konzentrationen leicht abnehmen, zeigte die Substanz BSc3217 deutlich höhere Wirksamkeiten von bis zu 0,9. Diese Substanz, deren Struktur eingerahmt ist, war bereits bei den oben geschilderten SIFT-Messungen der NBB-Substanzen aus der ersten Synthese-Runde aufgefallen (vgl. Abbildung 5.12), denn ihre Struktur kombiniert das Hydrazid H_2 und das Aldehyd A_{12} , welche dort beide mit hohen Wirksamkeiten assoziiert waren. Mithilfe von derartigen Dosis-Wirkungs-Analysen konnten eine Reihe neuer NBB-Substanzen mit starken anti-aggregatorischen Fähigkeiten im SIFT-Assay identifiziert werden.

5.1.4 Therapeutische Behandlung von Scrapie-infizierten Mäusen

Um das therapeutische Potenzial einzelner Substanzen, die in vorangegangenen Testreihen hohe Wirksamkeiten gezeigt hatten, am lebenden Organismus zu überprüfen, wurden von meinen Kollegen am ZNP Versuche mit Mäusen durchgeführt, deren Ablauf in Abschnitt 2.5 skizziert wurde. Wie dort beschrieben ist, waren die Mäuse intrazerebral mit Scrapie infiziert worden und wurden erst mehr als 100 Tage nach der Infektion, in einem späten Stadium der Inkubationszeit, mit den Testsubstanzen behandelt, um die Anwendbarkeit der Substanzen in einem für die Prionkrankheiten realistischen Szenario zu überprüfen.

Für die Behandlung wurden die Substanzen 293G02, 309F02 und 313B02 ausgewählt, die beim Screening von DIVERSet 1 gefundenen worden waren und im Zellkultur-Assay hohe Wirksamkeiten gezeigt hatten (vgl. Abbildung 5.4). Die Substanzen wurden an 14 aufeinander folgenden Tagen täglich in Form von intraperitonealer Injektion von 50 μ l der Substanzen in einer DMSO-Trägerlösung verabreicht. Dabei wurden die Substanzen 293G02 und 309F02 bei Konzentrationen von 10 mM bzw. 5 mM während des gesamten Zeitabschnitts eingesetzt. Die Dosis der Substanz 313B02 betrug zunächst ebenfalls 5 mM, wurde aber nach dem Auftreten von toxischen Anzeichen ab dem vierten Tag auf 2,5 mM reduziert. Die Tiere wurden getötet, sobald sie das Endstadium der Krankheit erreicht hatten. Die im Folgenden geschilderten Ergebnisse wurden in unserem gemeinsamen Patent [94] publiziert.

Verlängerung der Inkubationszeit

Abbildung 5.15 zeigt die Mediane der Überlebenszeiten für die Gruppen von Mäusen, die mit der DMSO-Trägerlösung oder den verschiedenen Testsubstanzen behandelt worden waren. Wie dort zu erkennen ist, starben die $n = 8$ Tiere der mit DMSO behandelten Gruppe knapp 150 Tage nach der Infektion. Während für die Behandlung mit der Substanz 313B02 keine Verlängerung der medianen Überlebenszeit gegenüber derjenigen der DMSO-Kontrollgruppe beobachtet werden konnte, führte die Behandlung mit den NBB-Substanzen 309F02 und 293G02 zu Verlängerungen der Überlebenszeiten um 11,5 bzw. 9,5 Tage.

Die beobachteten Verlängerungen der Überlebenszeiten stimmen in etwa mit der Behandlungsdauer von 14 Tagen überein, was bedeuten könnte, dass die Behandlung mit diesen Substanzen den Krankheitsfortschritt anhalten konnte, solange die Medikamente verabreicht wurden. Obwohl die Krankheit durch die Behandlung nicht geheilt werden konnte, weisen die Resultate darauf hin, dass bei TSE-Patienten der Krankheitsverlauf durch eine Therapie mit diesen Substanzen abgeschwächt werden könnte. In diesem Fall würde die Behandlung mit diesen Medikamenten zu einer Lebensverlängerung und zu einer Stabilisierung des Gesundheitszustandes führen.

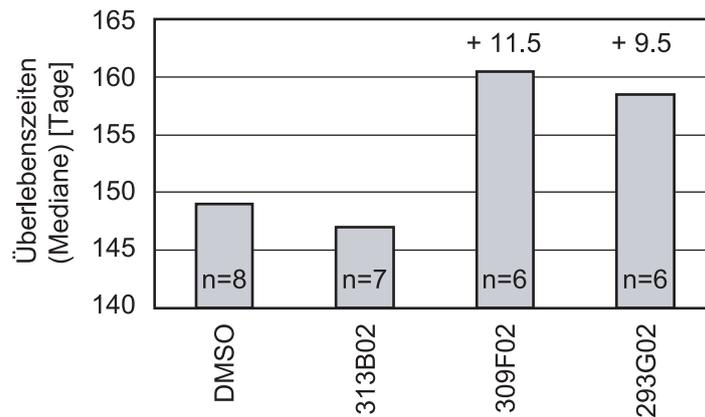


Abbildung 5.15: Mediane der Sterbealter. Für Gruppen von Mäusen, die mit den Substanzen 313B02, 293G02, 309F02 oder zur Kontrolle mit DMSO behandelt wurden, sind die Mediane der Überlebenszeiten in Tagen nach der Infektion dargestellt. Der Median der Überlebenszeit der $n = 8$ mit DMSO behandelten Tiere lag bei knapp 150 Tagen. Während die Behandlung mit der Substanz 313B02 die Überlebenszeit nicht verlängern konnte, führten die Substanzen 309F02 und 293G02 zu deutlichen Verlängerungen der medianen Überlebenszeiten von 11,5 bzw. 9,5 Tagen.

Clearance von PrP^{Sc} in der Milz

Um das Vorkommen von PrP^{Sc}-Ablagerungen in der Milz zu untersuchen, wurden, wie in Abschnitt 2.5 erläutert wurde, aus den Milzen der getöteten Mäuse Gewebeschnitte präpariert und die darin enthaltenen PrP^{Sc}-Ablagerungen über Antikörper eingefärbt.

Abbildung 5.16 zeigt mikroskopische Aufnahmen von solchen Gewebeschnitten durch die Milzen von zwei Mäusen. Der Schnitt in **A** zeigt das Milzgewebe einer mit DMSO behandelten Maus und lässt rötlich eingefärbte PrP^{Sc}-Ablagerungen erkennen, die in **B** weiter vergrößert dargestellt sind. In **C** ist das Milzgewebe einer Maus, die mit der Substanz 293G02 behandelt worden war, gezeigt. Hier, wie auch für die anderen mit der Substanz 293G02 behandelten Mäuse, konnten keine Zeichen von PrP^{Sc} detektiert werden, was zeigt, dass diese in die Bauchhöhle gespritzte Substanz eine sogenannte Clearance von PrP^{Sc} in der Milz bewirkt.

Dieses Ergebnis zeigt, dass die Substanz 293G02 im gewählten Versuchsszenario eine ausgeprägte anti-Prion-Wirksamkeit in der Milz vermittelt. Da die Substanz intraperitoneal verabreicht wurde, ist es wahrscheinlich, dass sie dieses Organ leichter erreichen konnte als das Gehirn. Die Therapie mit der Substanz 293G02 konnte die PrP^{Sc}-Ablagerung und den Krankheitsfortschritt im Gehirn aber wohl deshalb nicht verhindern, weil sie entweder die Blut-Hirn-Schranke nicht effizient überwand oder zu rasch im Organismus abgebaut wurde. Um solche Fragen der Bioverfügbarkeit zu klären, wurden Untersuchungen zur sogenannten Pharmakokinetik durchgeführt.

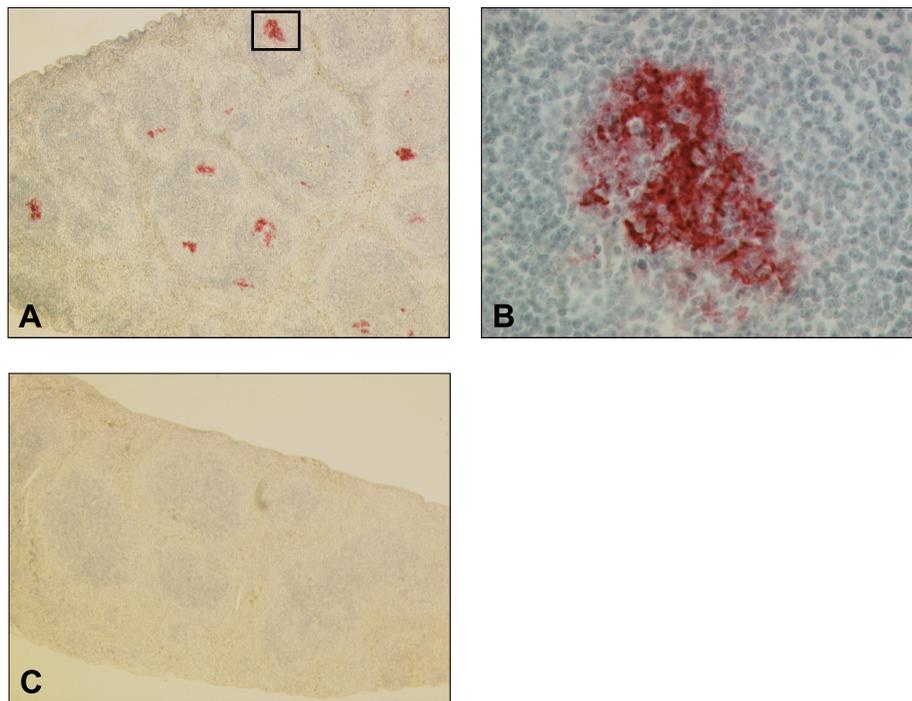


Abbildung 5.16: PrP^{Sc}-Ablagerungen in der Milz. Immunohistochemisch prozessierte Schnitte durch Milzgewebe von einer mit DMSO-Kontroll-Lösung behandelten Maus (**A** und **B**) und einer mit Substanz 293G02 behandelten Maus (**C**). Die Milz der mit DMSO behandelten Maus weist rötlich eingefärbte Ablagerungen von PrP^{Sc} auf, die in **B** vergrößert dargestellt sind. Hingegen ist in der Milz der mit 293G02 behandelten Maus kein PrP^{Sc} detektierbar. Übernommen aus unserem Patent Bertsch et al. [94].

5.1.5 Studien zur Pharmakokinetik

Studien zur Pharmakokinetik (PK) eines untersuchten Wirkstoffs widmen sich den Fragen, auf welche Weise sich der Wirkstoff nach der Verabreichung im Organismus verteilt, wie rasch er ins Blutplasma und in verschiedene Körpergewebe gelangt und wo und in welcher Form er abgebaut oder wieder ausgeschieden wird. Dabei geht es insbesondere um die Frage, ob der untersuchte Wirkstoff tatsächlich in ausreichender Konzentration den Wirkort, d.h. im Falle von anti-Prion-Wirkstoffen in erster Linie das Gehirn, erreicht. Bei PK-Studien werden die zu untersuchenden Substanzen an Versuchstiere verabreicht, zu verschiedenen Zeitpunkten danach jeweils einzelne Tiere getötet und das Vorkommen der Substanzen in verschiedenen Gewebeproben analysiert.

Im Rahmen des Verbundprojektes wurden derartige Studien zur PK der beiden NBB-Substanzen 309F02 und 293G02 durchgeführt, die in den oben geschilderten Therapie-Experimenten vielversprechende Ergebnisse gezeigt hatten. Für die PK-Studien wurden am ZNP die beiden Substanzen einzeln zum Zeitpunkt $t = 0$ jeweils einer Gruppe von zwölf Mäusen intraperitoneal verabreicht, wobei die eingesetzten Dosen von $10 \mu\text{mol/kg}$

(309F02) bzw. 20 $\mu\text{mol/kg}$ (293G02) jenen der Therapie-Experimente entsprachen. Zu den Zeitpunkten $t = 0,5, 1, 2, 4, 8$ und 24 h nach der Verabreichung wurden jeweils zwei Mäuse getötet und Proben von Blutplasma und Hirn entnommen. Das Vorkommen der Testsubstanzen in den entnommenen Proben wurde von der Firma Boehringer Ingelheim [163] massenspektrometrisch analysiert.

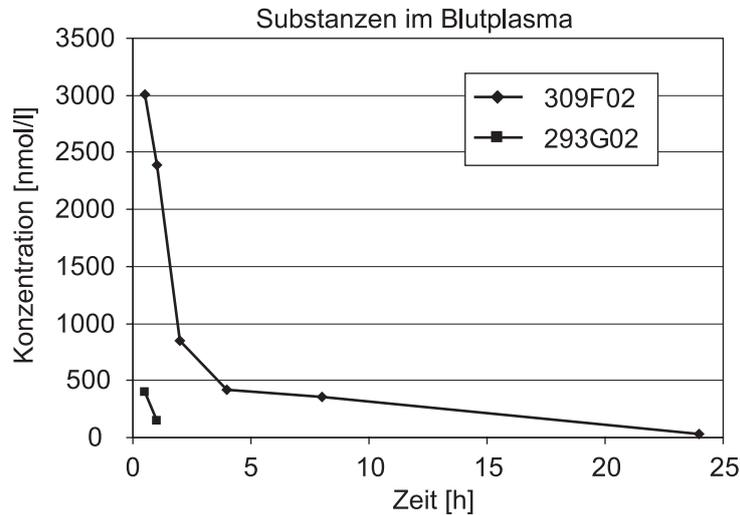


Abbildung 5.17: Konzentrationen der Substanzen 309F02 und 293G02 im Blutplasma von Mäusen. Die mithilfe von Massenspektrometrie ermittelten Konzentrationen der beiden Substanzen im Blutplasma von Mäusen sind gegen die Zeit nach der Verabreichung aufgetragen. Gezeigt sind jeweils die Mittelwerte der Konzentrationen von zwei untersuchten Mäusen. Die Messungen wurden von der Firma Boehringer Ingelheim durchgeführt.

Abbildung 5.17 zeigt die zeitlichen Verläufe der massenspektrometrisch ermittelten Konzentrationen der beiden Substanzen im Blutplasma der untersuchten Mäuse. Gezeigt sind jeweils die Mittelwerte der Werte von zwei Mäusen. Die Konzentration der Substanz 309F02 lag 0,5 h nach der Verabreichung bei etwa 3000 nmol/l und nahm, einem exponentiellen Verlauf folgend, innerhalb der ersten vier Stunden rasch auf Werte unterhalb von 500 nmol/l ab. Die Substanz 293G02 war bereits nach 0,5 h und 1 h nur in Konzentrationen von wenigen hundert nmol/l vorhanden und konnte in den Proben der späteren Zeitpunkte nicht mehr detektiert werden.

Aus den Proben der Gehirne konnten lediglich für die Substanz 309F02 und die ersten beiden Zeitpunkte Werte von einigen hundert nmol/kg ermittelt werden. Ansonsten lagen beide Substanzen unterhalb der Nachweisgrenze.

Diese Ergebnisse zeigen, dass beide Substanzen im Organismus, d.h. höchstwahrscheinlich in dem dafür zuständigen Organ, der Leber, rasch abgebaut werden. Daher konnten bei den Therapieversuchen mit Mäusen nur sehr geringe Wirkstoffdosen in die Gehirne gelangen. Die Tatsache, dass dennoch therapeutische Effekte nachweisbar waren, zeigt, dass dazu schon geringste Konzentrationen dieser Substanzen ausreichen. Man kann daher ver-

muten, dass NBB-Substanzen mit ähnlichen Wirkprofilen aber besserer Bioverfügbarkeit hochpotente anti-Prion-Therapeutika darstellen. Der entscheidende nächste Schritt besteht also in der Identifikation entsprechend modifizierter Substanzen.

Die Entwicklung solcher modifizierter NBB-Substanzen ist Gegenstand der derzeitigen Forschungsbemühungen. Aktuell durchgeführte Tierversuche zeigen vielversprechende Ergebnisse für einzelne neu synthetisierte NBB-Substanzen mit verbesserten pharmakokinetischen Eigenschaften.

5.1.6 Unterdrückung der Aggregation von α -Synuklein

Außer für die Suche nach anti-Prion-Substanzen, wurde die SIFT-Technik am ZNP auch dazu eingesetzt, für einige NBB-Substanzen herauszufinden, ob sie auch die Aggregation des Proteins α -Synuklein (α S) hemmen können, die mit der Parkinsonschen Krankheit assoziiert ist (vgl. Abschnitt 2.2.4). Dabei wurde ein am ZNP entwickelter *in vitro* Aggregations-Assay eingesetzt, in welchem mit rot und grün fluoreszierenden Farbstoffen markierte α S-Monomere bei niedrigen Konzentrationen (0-3%) des Lösungsmittels DMSO zur Aggregation gebracht werden können. Auch die im Folgenden aufgeführten Resultate wurden in unserem gemeinsamen Patent [94] veröffentlicht.

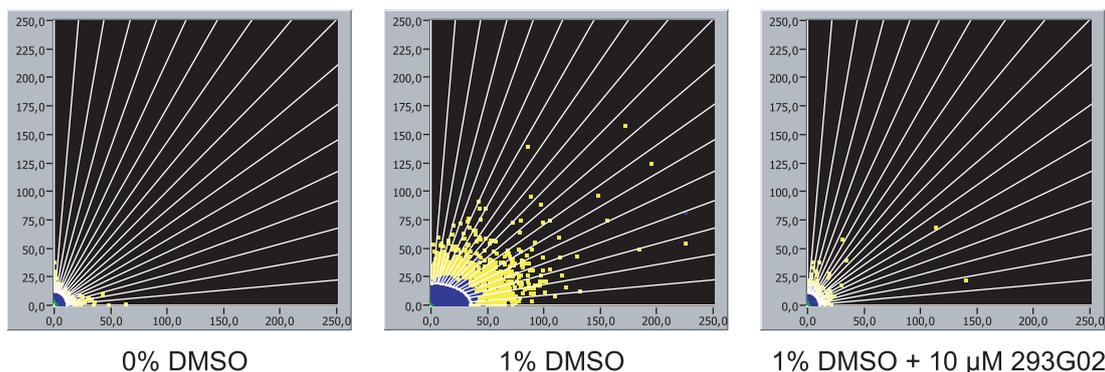


Abbildung 5.18: Unterdrückung der α S-Aggregation durch die Substanz 293G02. Gezeigt sind mit der SIFT-Technik gemessene 2D-Fluoreszenz-Intensitäts-Verteilungen (vgl. Abschnitt 2.2.2) verschiedener Mischungen des α -Synuklein-Aggregations-Assays. Ohne DMSO findet im Assay keine nachweisbare Aggregation statt (links). Bei einer DMSO-Konzentration von 1% bilden sich große rot-grün fluoreszierende Aggregate, die in Form von hochintensiven Fluoreszenzereignissen detektiert werden (Mitte). Die Zugabe der Substanz 293G02 verhindert die Bildung solcher Aggregate (rechts).

Abbildung 5.18 zeigt die zweidimensionalen Fluoreszenz-Intensitäts-Verteilungen (vgl. Abschnitt 2.2.3) für verschiedene Mischungen des α S-Aggregations-Assays. Wie in dem linken Diagramm zu erkennen ist, findet ohne DMSO keine nachweisbare Aggregation statt, was durch das Fehlen von Fluoreszenzereignissen hoher Intensität gekennzeichnet ist. Bei einer DMSO-Konzentration von 1% bilden sich große rot-grün fluoreszierende

α S-Aggregate, die durch hochintensive Fluoreszenzereignisse charakterisiert sind (Mitte). Durch Zugabe der Substanz 293G02 (bei einer finalen Konzentration von 10 μ M) wird die Bildung solcher Aggregate wirksam verhindert (rechts).

Dies ist ein klarer Hinweis darauf, dass die Substanz 293G02 nicht nur als anti-Prion-Substanz wirkt sondern auch das Potenzial besitzt, ein Therapeutikum für Synukleinopathien wie die Parkinsonsche Krankheit zu sein. Dass die Substanz 293G02 sowohl gegen die Aggregation von Prion-Proteinen als auch gegen die α S-Aggregation wirkt, weist auf eine allgemeine anti-aggregatorische Wirksamkeit dieser Substanz bei Protein-Aggregations-Krankheiten hin. Wie bereits mehrfach betont wurde, ist diesen Krankheiten gemein, dass die dabei beteiligten Proteine in β -Faltblatt-Konformationen fehlfalten und zu amyloiden Fibrillen aggregieren. Deshalb könnte es sein, dass die Substanz 293G02 und möglicherweise weitere Substanzen der NBB-Klasse das Potenzial besitzen, als Therapeutika für die kausale Behandlung einer ganzen Palette von neurodegenerativen Protein-Aggregations-Krankheiten eingesetzt werden zu können. Zu diesen Krankheiten zählen neben den Prionkrankheiten und der Parkinsonschen Krankheit auch die Huntingtonsche und die Alzheimersche Krankheit.

5.1.7 Inhibierung der Aggregation von poly-Glutamin-Proteinen

Um die Vermutung der generellen anti-aggregatorischen Wirksamkeit der NBB-Substanzen zu überprüfen, wurden kürzlich, wie es in Abschnitt 4.3 angesprochen wurde, im Rahmen eines gemeinsamen Projekts mit TP1 sowie mit den Gruppen von Prof. F. Ulrich Hartl (MPI für Biochemie, Martinsried) und Prof. Christian Haass (Adolf-Butenandt-Institut, LMU) einige Substanzen, die im Verbund als anti-Prion-Wirkstoffe identifiziert worden waren, hinsichtlich ihrer Wirksamkeit gegen die mit der Huntingtonschen Krankheit assoziierte Aggregation von poly-Glutamin (Q)-Proteinen (vgl. Abschnitt 1.3) in einem Zebrafischmodell untersucht. Mein Beitrag zu diesem Projekt waren Docking-Simulationen von Modellen der experimentell untersuchten Substanzen an ein Modell für polyQ-Protofibrillen. Die Ergebnisse dieses Projekts wurden in einer gemeinsamen Publikation [159] veröffentlicht und sollen hier kurz skizziert werden.

In der Gruppe von Prof. Haass werden Zebrafische seit einiger Zeit als Modelltiere zum Studium der Alzheimerschen und der Parkinsonschen Krankheit eingesetzt. Niclas Schiffer, ein Doktorand von Prof. Hartl, setzte die Zebrafische zum Studium der mit der Huntingtonschen Krankheit verbundenen Aggregation des Proteins Huntingtin (htt) ein. Das Protein htt weist im N-terminalen Bereich einen Abschnitt aufeinander folgender Q-Residuen auf, dessen Länge bereits in gesunden Organismen infolge von zahlreichen Polymorphismen im htt-Gen variabel ist und zwischen sechs und 35 liegt. Überschreitet die Länge dieses polyQ-Abschnitts einen kritischen Wert von etwa 36 Residuen, so bilden sich in Organismen fibrilläre Aggregate des htt-Proteins [164–166], deren zentrale Bestandteile

die polyQ-Abschnitte in β -faltblattartigen Strukturen sind. Um die Aggregation von htt in den Zebrafischen zu untersuchen, wird den Embryonen geeignete Boten-Ribonukleinsäure (messenger ribonucleic acid, mRNA) injiziert, sodass sie htt-Proteine mit verlängerten polyQ-Abschnitten exprimieren, welche Aggregate bilden, die biochemisch detektiert werden können. Es sei angemerkt, dass die den Embryonen injizierten mRNA-Moleküle einen zusätzlichen Abschnitt enthalten, der für das grün fluoreszierende Protein (GFP) kodiert. Weil die Zebrafische transparent sind, können die von ihnen produzierten und mit GFP markierten htt-Proteine sowie deren Aggregation unter dem Mikroskop beobachtet werden (daraus resultierende Bilder und Filme finden sich in [159] und im zugehörigen Begleitmaterial). Das Zebrafischmodell reproduziert also ein wesentliches Merkmal der Pathologie der Huntingtonschen Krankheit. Verabreicht man den Fischen Testsubstanzen, so kann deren Einfluss auf die polyQ-Aggregation analysiert werden.

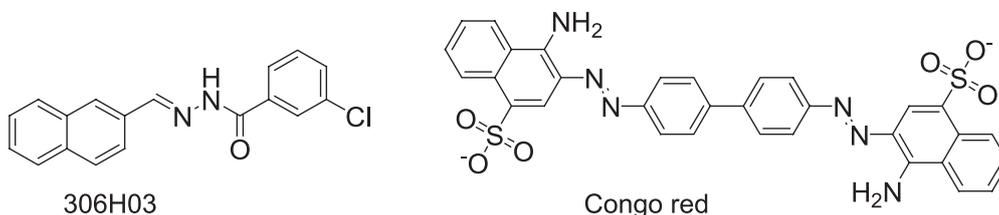


Abbildung 5.19: Strukturen der Substanzen 306H03 und Congo red. Diese Substanzen wurden, wie die Substanzen 293G02 und 313B02, in den experimentellen Testreihen und anhand von Docking-Simulationen untersucht. Die Substanz 306H03, die der Substanz 293G02 strukturell ähnlich ist (vgl. Abbildung 5.7), hatte zuvor im SIFT-Assay und in den Zellkulturmodellen von TP3 und TP4 anti-Prion-Wirksamkeit gezeigt. Congo red ist als amyloid-bindender Farbstoff bekannt.

Für die Untersuchung in dem Zebrafischmodell wählten wir die Substanz 313B02 und die beiden NBB-Substanzen 293G02 und 306H03 (vgl. Abbildung 5.4 bzw. Abbildung 5.19), die in verschiedenen Testreihen (SIFT-Assay, Zellkulturmodell TP4) anti-Prion-Wirksamkeiten gezeigt hatten. Als weitere Testsubstanzen wurde der amyloid-bindende Farbstoff Congo red (siehe beispielsweise [167]) eingesetzt, dessen Struktur in Abbildung 5.19 gezeigt ist, sowie drei Substanzen, die als mögliche Wirkstoffe gegen die Huntingtonsche Krankheit bekannt sind, aber hier nicht weiter diskutiert werden sollen.

Die beiden NBB-Substanzen 293G02 und 306H03 zeigten in den Zebrafischen ähnlich starke Wirksamkeiten gegen die polyQ-Aggregation wie Congo red. Hingegen konnte für die Substanz 313B02 keine aggregationshemmende Wirkung festgestellt werden. Neben dem Zebrafischmodell wurden im Rahmen dieses Projekts zusätzlich ein *in vitro*-Assay und ein Zellkulturmodell zur Untersuchung der polyQ-Aggregation eingesetzt. Auch diese beiden Assays konnten dazu eingesetzt werden, die aggregationshemmende Wirkung von zugegebenen Substanzen zu testen. Dabei zeigten sich alle drei DIVERSet-Substanzen in beiden Assays als wirksam gegen die Aggregation von polyQ-Proteinen. Um Einblicke in die molekularen Mechanismen zu erhalten, durch welche die gefundenen Wirkstoffsub-

stanzen ihre anti-aggregatorische Wirksamkeit vermitteln, führte ich die oben angesprochenen Docking-Simulationen durch.

Docking von NBB-Substanzen an polyQ-Protofibrillen

Bei den Docking-Simulationen setzte ich, wie ich in Abschnitt 4.3 beschrieben habe, als Modell für eine polyQ-Protofibrille eine trianguläre β -helikale Struktur ein, die von meiner Kollegin Martina Stork zusammen mit Paul Tavan entwickelt worden war [27]. Das Modell ist ein Dimer aus zwei polyQ-Peptiden mit jeweils 36 Residuen, die linkshändig gewunden und dadurch zu triangulären β -Helizes angeordnet sind. Fünf verschiedene Schnappschüsse dieser Struktur wurden als Rezeptoren eingesetzt. Modelle für die Wirkstoffsubstanzen (293G02, 306H03, 313B02 und Congo red) wurden, wie in Abschnitt 4.3 beschrieben, berechnet. Mithilfe des Softwarepakets AutoDock [160] wurden für jeden polyQ-Schnappschuss und für jedes Ligandmolekül 100 einzelne Docking-Simulationen durchgeführt.

Abbildung 5.20 zeigt für die vier Wirkstoffsubstanzen jeweils die polyQ-Protofibrillen zusammen mit den Strukturen von gedockten Molekülen. Dabei sind die fünf verschiedenen polyQ-Schnappschüsse und die zugehörigen Dockings farblich kodiert. Für jeden Wirkstoff und jeden Schnappschuss sind die 20 Dockings mit den niedrigsten freien Bindungsenergien dargestellt und zusätzlich die jeweiligen Mittelwerte dieser freien Energien angegeben. Die Simulationen sagen für alle untersuchten Verbindungen relativ große Bindungsenergien zwischen 6,2 und 7,2 kcal/mol vorher, was auf thermisch stabile Dockings hinweist. Alle Substanzen binden bevorzugt an die beiden Enden der polyQ-Helizes. Gut definierte Docking-Positionen werden für die beiden NBB-Verbindungen (293G02 und 306H03) vorhergesagt, welche beide eine Amid-Gruppe besitzen, die hervorragend für eine durch Wasserstoffbrückenbindungen vermittelte Anlagerung an das polyQ-Peptidrückgrat geeignet ist. Die Docking-Positionen von Congo red und 313B02 zeigen eine weniger starke Ordnung. Dass die NBB-Substanzen spezifisch an die Enden der polyQ-Helizes binden, welche die Ansatzpunkte für weitere Aggregation von polyQ-Segmenten darstellen, legt einen Mechanismus der Aggregationshemmung nahe, der ein weiteres Fibrillenwachstum verhindern könnte. Es sei angemerkt, dass auch andere strukturelle Modelle für polyQ-Protofibrillen existieren [168]. Jedoch sind alle diese Modelle, ähnlich wie unser β -helikales Modell, durch freie Anlagerungsstellen für Wasserstoffbrückenbindungen an die C=O- und NH-Gruppen des Proteinerückgrats charakterisiert, welche die ungesättigten Enden der β -Faltblatt-Fibrillen ausmachen.

Allgemeine Bedeutung der Ergebnisse

Insgesamt bestätigen die angeführten Ergebnisse unsere Ausgangshypothese, dass die mit bestimmten neurodegenerativen Krankheiten assoziierten Proteine, PrP^{Sc} und mutiertes

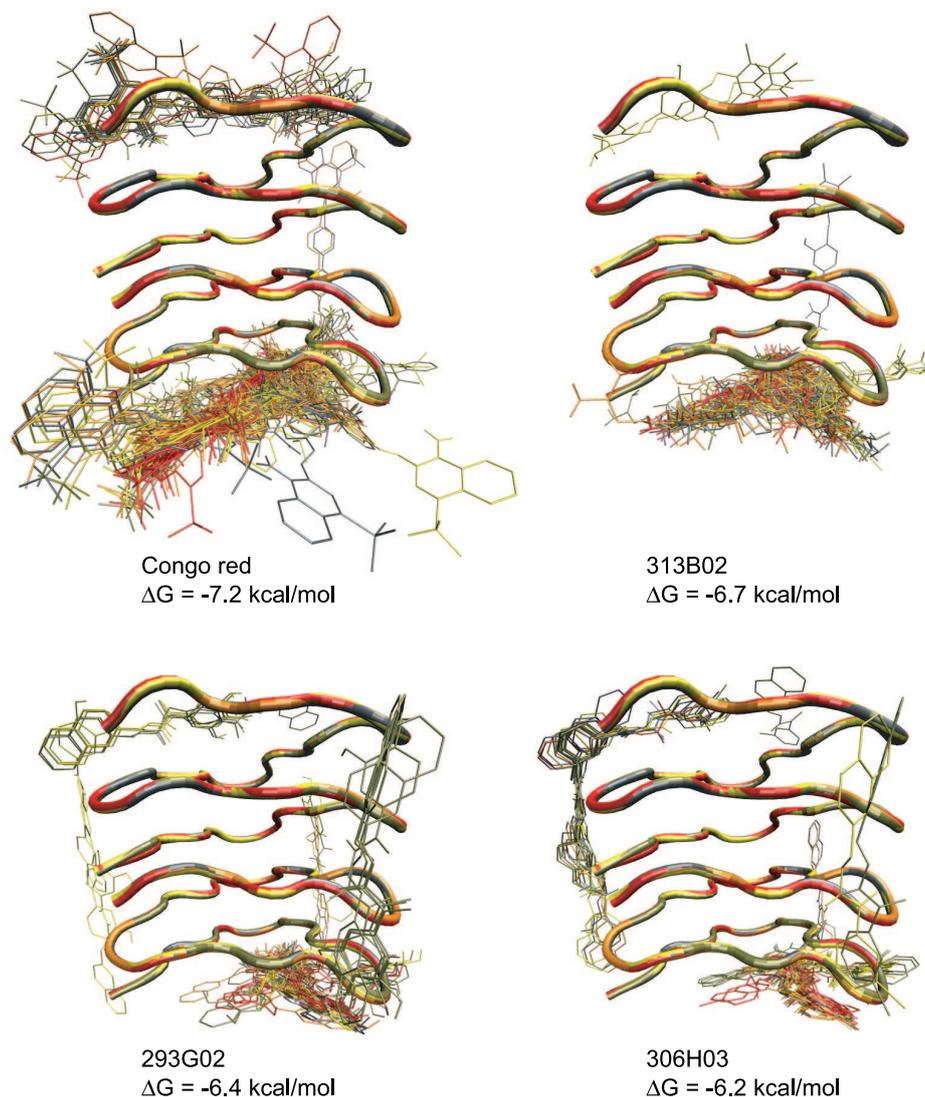


Abbildung 5.20: Docking von anti-aggregatorischen Substanzen an ein Modell für eine polyQ-Protofibrille. Das Modell für eine polyQ-Protofibrille umfasst einen Dimer aus 36Q-Peptiden, die in einer triangulären β -helikalen Struktur angeordnet sind [27]. Nach dem Modell würde die Protofibrille durch Anlagerung weiterer Windungen am oberen und unteren Ende der β -Helix wachsen. Die Peptidrückgrate der polyQ-Dimere sind als Röhren und die gedockten Substanzen in Form von Linien-Repräsentationen dargestellt. Die fünf verschiedenen polyQ-Schnappschüsse und die zugehörigen Dockings werden durch Farbkodierung unterschieden. Für jeden der vier Liganden (Congo red, 313B02, 293G02 und 306H03) und jeden Schnappschuss sind die 20 Dockings mit den niedrigsten freien Energien gezeigt. Zusätzlich sind die Mittelwerte dieser freien Energien für die vier ausgewählten Substanzen angegeben. Übernommen aus Schiffer et al. [159].

ht, strukturell ähnliche fibrilläre Aggregate bilden, wobei die Ähnlichkeit in einer gemeinsamen β -helikalen Kernstruktur bestehen könnte [27, 64]. Deshalb hatten wir vermutet, dass Inhibitoren der PrP^{Sc}-Aggregation auch die Aggregation von polyQ-Proteinen hemmen sollten. Unsere Beobachtung, dass anti-Prion-Substanzen tatsächlich auch wirksame Inhibitoren der polyQ-Aggregation sind, stellt die oben angeführte Bestätigung unserer Arbeitshypothese dar. Diese Ergebnisse bekräftigen also die Sichtweise, dass der Pathogenese verschiedener neurodegenerativer Krankheiten gemeinsame Mechanismen unterliegen und legen nahe, dass gemeinsame therapeutische Targets existieren könnten.

5.2 Eine weitere neue Wirkstoffklasse

Die umfangreichen oben geschilderten Arbeiten, die zur Entwicklung der NBB-Wirkstoffklasse führten, sowie eine Reihe von zusätzlichen Untersuchungen haben schließlich zur Entdeckung einer weiteren Klasse von neuen Wirkstoffen gegen Prion- und verwandte Krankheiten geführt. Da die neue Wirkstoffklasse durch ein Patent geschützt werden soll, kann, wie zu Beginn des Kapitels angesprochen wurde, die chemische Grundstruktur dieser Klasse hier nicht offengelegt werden. Jedoch möchte ich von den Arbeiten, die zur Entdeckung dieser Klasse geführt haben, berichten.

Gegen Ende des Jahres 2003 wurde die zweite Substanzbibliothek DIVERSet 2 beschafft. Wie die erste Bibliothek wurde auch DIVERSet 2 im SIFT-Assay primär durchgemustert, wobei wiederum SIFT-Viewer I zur Datenauswertung eingesetzt wurde. Da dieses Werkzeug, welches während des Screenings von DIVERSet 1 entwickelt worden war, nun in ausgereifter Form zur Verfügung stand, konnten die neuen Messdaten rascher ausgewertet werden. Darüber hinaus wurden die Messdaten und Ergebnisse in die neu eingerichtete TSE-DB übertragen und konnten daher zusätzlich mithilfe des neu entwickelten Nachfolgemodells SIFT-Viewer II analysiert bzw. visuell inspiziert werden. Die beim Primärscreening gefundenen Treffer (gut 200 Substanzen) wurden in Verdünnungsreihen gemessen und anhand der dadurch umgesetzten Dosis-Wirkungs-Analysen validiert (rund 150 Substanzen), wobei sich SIFT-Viewer II als hilfreiches Werkzeug erwies.

5.2.1 Das Zellkultur-Screening der DIVERSet-Bibliotheken

Zusätzlich zum SIFT-Assay aus TP1 wurde ab Ende 2004 der Zellkultur-Assay von TP3 (vgl. Abschnitt 2.4) zum Screening der DIVERSet-Bibliotheken eingesetzt. Dabei wurden die anti-Prion-Wirksamkeiten der DIVERSet-Substanzen zunächst bei einer einzigen Konzentration (20 μ M) getestet (vgl. Abbildung 2.10). Die dabei als wirksam gefundenen Substanzen, die Primärtreffer, wurden anhand von Verdünnungsreihen validiert, wodurch insbesondere nach solchen Substanzen gesucht wurde, die auch noch bei geringeren Konzentrationen wirksam sind. Abgesehen von dieser grundlegenden Vorgehensweise unter-

schieden sich die Screening-Kampagnen von DIVERSet 1 und DIVERSet 2 hinsichtlich einiger Details.

Das primäre Screening von DIVERSet 1 wurde auf SMB-Zellen durchgeführt, und die Primärtreffer wurden in Verdünnungsreihen bei vier Konzentrationen ($1 \mu\text{M}$, $4 \mu\text{M}$, $10 \mu\text{M}$ und $40 \mu\text{M}$) auf SMB-Zellen überprüft. Die Substanzen mit den höchsten Wirksamkeiten wurden zusätzlich auf ScN2a-Zellen getestet. Hingegen werden beim (noch laufenden) Primärscreening von DIVERSet 2 alle Substanzen sowohl auf SMB- als auch auf ScN2a-Zellen getestet, und nur jene Testsubstanzen, die bei beiden Zelllinien Wirksamkeit zeigen, werden als Primärtreffer erachtet. Die Primärtreffer werden in Verdünnungsreihen mit nurmehr drei Konzentrationen ($0,2 \mu\text{M}$, $2 \mu\text{M}$ und $20 \mu\text{M}$) unter Einsatz beider Zelllinien überprüft, wobei die Experimente mit den ScN2a-Zellen lediglich zur Kontrolle mitgeführt werden.

So wurde DIVERSet 1 vollständig durchgemustert, wobei 575 Primärtreffer gefunden wurden (SMB). Die Validierungsmessungen der Primärtreffer ergaben 70 Substanzen, die noch bei einer Konzentration von $1 \mu\text{M}$ wirksam waren (SMB). Bei der anschließenden Überprüfung dieser 70 hochwirksamen Substanzen auf ScN2a-Zellen zeigten sich 52 Substanzen als wirksam. Aus DIVERSet 2 wurden bislang (Stand 11/06) rund die Hälfte der Substanzen im primären Screening getestet, wobei 262 Primärtreffer gefunden wurden (SMB und ScN2a). Bei den Verdünnungsreihenmessungen waren 40 Substanzen auch bei der geringen Konzentration von $0,2 \mu\text{M}$ noch wirksam (SMB).

5.2.2 Die Identifikation einer neuen Leitstruktur

Nachdem also ein Großteil der beiden DIVERSet-Bibliotheken im SIFT- und im Zellkultur-Assay von TP3 durchgemustert worden war, unterzog ich die Bibliotheken einer strukturellen Clusteranalyse^a mithilfe des kommerziellen Softwarewerkzeugs Benchware HTS DataMiner (vgl. Abschnitt 4.2). Diese Analyse unterteilt die Substanzen der gegebenen Sammlung in Cluster von strukturell ähnlichen Substanzen und kann daher dazu genutzt werden, neue potenzielle Leitstrukturen aufzuspüren.

Da, wie ich in Abschnitt 4.2.2 erläutert habe, die Menge aller (20.000) Substanzen der beiden DIVERSet-Bibliotheken als Ausgangsmenge für die Clusterbildung mit DataMiner zu umfangreich gewesen wäre, schränkte ich die Ausgangsmenge auf die Menge der Primärtreffer (837 Substanzen, Stand 11/06) des Zellkultur-Screenings von DIVERSet 1 und 2 ein und ließ so die Cluster ausschließlich anhand der wirksamen Substanzen bilden. Anschließend wurden die restlichen (unwirksamen) Substanzen zusätzlich in die gebildeten Cluster einsortiert. Das Resultat der Clusterbildung wird von DataMiner, wie in Ab-

^aClusteranalysen, wie die im Folgenden geschilderte, waren von mir bereits mehrfach zu verschiedenen früheren Zeitpunkten unter Verwendung des jeweils aktuellen Stands der Screening-Ergebnisse durchgeführt worden. Die hier geschilderte Clusteranalyse basiert auf dem oben zusammengefassten Stand (aus 11/06) des DIVERSet-Screenings.

schnitt 4.2.2 beschrieben ist, in Form einer sogenannten SAR-Karte grafisch dargestellt, in welcher die Substanzcluster \mathcal{S} durch in der Ebene angeordnete Symbole repräsentiert sind. Dabei werden von DataMiner die Symbole so angeordnet, dass sie nahe beieinander liegen, wenn die repräsentierten Cluster strukturell ähnlich sind. Die Größen, die Formen und die Farben der Symbole werden anhand von clusterlokalen Eigenschaften bestimmt, die vom Benutzer ausgewählt werden können.

So legte ich fest, dass die Größen der dargestellten Symbole proportional zu den Clustergrößen $|\mathcal{S}|$, d.h. zu den Anzahlen der jeweils enthaltenen Substanzen C , sein soll. Die Formen der Symbole wurden anhand der gemäß (4.10) definierten Anteile

$$P_{\text{SIFT}}(\mathcal{S}) = \frac{|\{C \in \mathcal{S} \mid a(C) \geq a_{\min}\}|}{|\mathcal{S}|} \quad (5.1)$$

jener in den jeweiligen Clustern \mathcal{S} enthaltenen Substanzen C festgelegt, deren primäre Wirksamkeiten $a(C)$ aus dem SIFT-Screening oberhalb einer gewählten Grenze $a_{\min} = 0,25$ lagen. Anhand dessen werden Cluster, deren Anteil $P_{\text{SIFT}}(\mathcal{S})$ mehr als 50% beträgt, als Sternchen und die restlichen, überwiegend unwirksamen Cluster als Kästchen dargestellt. In analoger Weise kodiert die Farbe der Symbole die clusterlokalen Anteile

$$P_{\text{ZK}}(\mathcal{S}) = \frac{|\{C \in \mathcal{S} \mid C \text{ ist Primärtreffer}\}|}{|\mathcal{S}|} \quad (5.2)$$

der Primärtreffer des Zellkultur-Screenings, wobei die Symbole von Clustern mit mehr als 50% wirksamen Substanzen rot und jene der restlichen Cluster grau gefärbt sind.

Abbildung 5.21 zeigt die aus der beschriebenen Clusteranalyse und der von mir festgelegten Form der Symboldarstellung resultierende SAR-Karte. Dem gewählten Schema entsprechend werden große Cluster, die zudem hohe Anteile an SIFT- und Zellkulturwirksamen Substanzen enthalten, durch große, rote Sternchen symbolisiert. Solche Cluster stellen möglicherweise potenzielle Leitstrukturklassen dar. Mithilfe von DataMiner wurden einzelne solcher interessanter Cluster per Mausklick ausgewählt und so die Strukturen und Wirksamkeiten der darin enthaltenen Substanzen inspiziert. Auf diese Weise konnte eine Gruppe von fünf benachbarten Clustern identifiziert werden, deren Symbole infolge der Auswahl per Mausklick fett umrandet dargestellt sind und die als XXX_1 bis XXX_5 bezeichnet wurden. Da die Symbole benachbart liegen, enthalten die Cluster XXX_1 bis XXX_5 allesamt Substanzen, die einander strukturell ähnlich sind. Die Substanzen (insgesamt 33) dieser Cluster bilden eine gemeinsame chemische Strukturklasse, die mit XXX bezeichnet werden soll. Während die Substanzen der Cluster XXX_1 und XXX_2 lediglich im Zellkultur-Assay überwiegend wirksam waren (rote Kästchen), waren jene der Cluster XXX_3, XXX_4 und XXX_5 in beiden Assays überwiegend wirksam (rote Sternchen). Die in den DIVERSet-Bibliotheken enthaltenen Substanzen der Klasse XXX sind also insgesamt stark Zellkultur- und moderat SIFT-aktiv.

Derartige Clusteranalysen, die ich bereits ab dem Spätsommer 2005 anhand von damals aktuellen Ständen der Screening-Ergebnisse durchgeführt hatte, ließen schon damals die

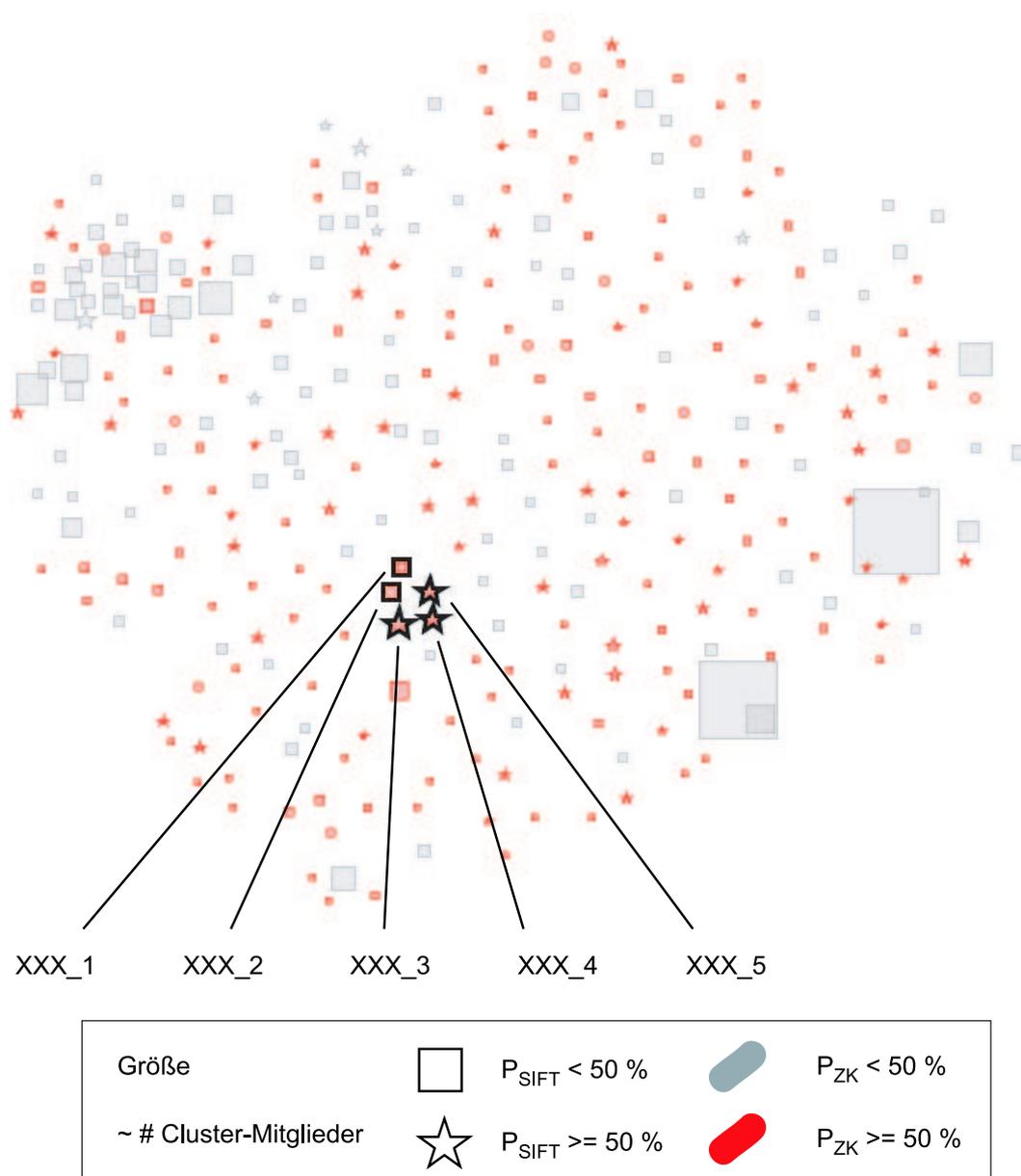


Abbildung 5.21: SAR-Karte für die DIVERSet-Bibliotheken. Die von DataMiner gebildeten Substanzcluster sind durch Symbole repräsentiert, die so in der Ebene verteilt sind, dass jene von strukturell ähnlichen Clustern nahe beieinander liegen. Die Größen der Symbole spiegeln die Zahlen der in den Clustern enthaltenen Substanzen wieder. Die Formen der Symbole sind durch die gemäß Gleichung (5.1) definierten clusterlokalen Anteile P_{SIFT} jener Substanzen C festgelegt, deren Primärwirksamkeit $a(C) > 0,25$ im SIFT-Screening den gewählten Mindestwert überschritt. Dabei werden Cluster, die mehr als 50% wirksame Substanzen enthalten ($P_{\text{SIFT}} \geq 0,5$), durch Sternchen und die übrigen Cluster durch Kästchen repräsentiert. Die Symbole sind rot gefärbt, falls der gemäß (5.2) definierte Anteil $P_{\text{ZK}} \geq 0,5$ der Primärtreffer aus dem Zellkultur-Screening größer als 50% liegt; andernfalls sind die Symbole grau gefärbt. Daraus ergibt sich, dass große, rote Sternchen solche Cluster repräsentieren, die viele, in beiden Assays wirksame Substanzen enthalten. Ausgewählt sind fünf dadurch hervorsteckende Cluster (XXX_1 bis XXX_5), deren Substanzen der gemeinsamen, auf diese Weise entdeckten neuen Wirkstoffklasse XXX angehören.

Wirksamkeit der Substanzklasse XXX erkennen. Daraufhin durchgeführte Recherchen in Literatur- und Patentdatenbanken ergaben, dass die Klasse XXX eine bis dahin unbekannte Klasse von anti-Prion-Wirkstoffen ist. Die neu entdeckte Substanzklasse erwies sich daher als eine neue Leitstrukturklasse, deren Substanzen seither im Verbund in zahlreichen Nachfolgeexperimenten weiterentwickelt werden.

5.2.3 Weiterentwicklung der neu identifizierten Substanzklasse

Im Rahmen einer kurz nach der Entdeckung dieser neuen Leitstruktur initiierten Kooperation mit der Gruppe von Prof. Christian Griesinger (MPI für biophysikalische Chemie, Göttingen) wurde dort eine Vielzahl von weiteren Substanzen (knapp 100) dieser und verwandter Klassen synthetisiert. Die Substanzen dieser fokussierten Bibliothek werden derzeit sowohl im SIFT-Assay gegen die Prion-Aggregation und gegen jene von α -Synuklein als auch im Zellkultur-Assay von TP3 getestet. Von mir wurden virtuelle Repräsentationen der Substanzen erzeugt und zusammen mit den Messergebnissen in die ChemFinder-basierte Datenhaltung (vgl. Abschnitt 3.7) aufgenommen. Bei den SIFT-Experimenten konnten neue XXX-Derivate identifiziert werden, die deutlich stärkere *in vitro*-Wirksamkeiten zeigen als jene XXX-Substanzen aus den DIVERSet-Bibliotheken. Für einige der vielversprechendsten Substanzen dieser Klasse laufen bereits Tierversuche. All diese Experimente sind Gegenstand der laufenden Forschung im Verbund, und die dabei bereits erzielten Ergebnisse können aus den genannten patentrechtlichen Gründen hier nicht gezeigt werden.

Insgesamt zeigt die im Verbund geleistete Entdeckung und Entwicklung von zwei neuen Wirkstoffklassen, dass auch eine kleine Gruppe von eng zusammenarbeitenden Wissenschaftlern Resultate erzielen kann, die sonst nur in Pharmakonzernen erzielt werden.

6 Ausblick

Als unser Forschungsverbund Mitte des Jahres 2002 seine Arbeit zur Entwicklung von Medikamenten gegen Prionkrankheiten aufnahm, war mir in mancherlei Hinsicht unklar, auf welches Abenteuer ich mich mit meinem Vorsatz eingelassen hatte, diese Arbeit im Teilprojekt von Paul Tavan durch Einsatz und, wo nötig, Entwicklung rechnergestützter Methoden zu unterstützen. Die Unklarheiten betrafen insbesondere die Fragen, mit welchen Algorithmen und Softwaremodellen die absehbare Flut an Screening-Daten bewältigt werden kann und wie darauf aufbauend Struktur-Wirkungs-Beziehungen systematisch untersucht und schließlich neue Leitstrukturen gefunden werden können. Wie die vorliegende Arbeit zeigt, ist es mir im Laufe der Zeit zunehmend besser gelungen, so weit wie nötig geeignete Verfahren und Werkzeuge selbst zu entwickeln und, wo immer möglich, vorgefertigte Lösungen zu finden und einzusetzen. Diese Entwicklungsprozesse und die eingesetzten Werkzeuge sind in den Kapiteln 3 und 4 ausführlich dokumentiert.

Der detailgetreue und daher weitgehend trockene Charakter dieses Teils der vorliegenden Arbeit mag nicht jeden Leser dazu motivieren, ihn genau durchzuarbeiten. Die detailgetreue Darstellung dient jedoch dem Zweck, einem potenziellen Nachfolger in einem hoffentlich bald initiierten Folgeprojekt den Einstieg zu erleichtern und die Verwendung der von mir erarbeiteten Lösungen ohne großen Aufwand zu ermöglichen.

Im Rahmen des noch bis Ende 2007 laufenden Projekts konnte der Einsatz dieser Werkzeuge schon bisher wichtige Beiträge dazu leisten, dass wir dem angestrebten Ziel, der Entwicklung von Medikamenten, einen entscheidenden Schritt näher gekommen sind. Es ist uns gelungen, zwei neue Klassen von Wirkstoffen gegen Prionkrankheiten anhand der eingesetzten *in vitro* und Zellkulturmodelle zu identifizieren, weiter zu entwickeln und in zusätzlichen Tiermodellen ihre Tauglichkeit als potenzielle Medikamente aufzuzeigen (vgl. Abschnitte 5.1 und 5.2).

Damit ist die angestrebte Entwicklung einer kausalen Therapie gegen Prionkrankheiten aber noch keineswegs abgeschlossen. Vielmehr müssen weitergehende Fragen der Stabilität im Metabolismus, der Hirngängigkeit, der Toxizität, zu Nebenwirkungen etc. erst noch untersucht und die Wirkstoffe entsprechend optimiert werden. Falls diese Fragen anhand von Tierversuchen befriedigend beantwortet werden können, so stünde im Falle der Creutzfeldt-Jakob-Krankheit der klinische Einsatz unmittelbar auf der Tagesordnung, weil bei derart seltenen, rasch tödlich verlaufenden Krankheiten umfangreiche klinische Studien nicht zwingend erforderlich sind. Anders gestaltet sich die Lage bei häufigen Krankheiten wie Alzheimer oder Parkinson. Hier sind in jedem Fall umfangreiche klinische Studien

der sogenannten Phasen I bis III notwendig, um Wirkstoffe als Medikamente zu etablieren. Solche Studien erfordern einen Partner aus der Industrie.

Die Tatsache, dass Substanzen der identifizierten Wirkstoffklassen (NBB und XXX) Wirksamkeiten nicht nur gegen Prionkrankheiten, sondern auch gegen weitere Proteinaggregationskrankheiten wie Parkinson und Huntington zeigen (vgl. Abschnitte 5.1.6 und 5.1.7), lässt es als möglich erscheinen, dass Wirkstoffe aus diesen Klassen auch die Aggregationsprozesse des τ -Proteins und der A β -Peptide inhibieren können, welche der Pathogenese der Alzheimerschen Krankheit zugrunde liegen. Mithin tragen aus gegenwärtiger Sicht die identifizierten Wirkstoffklassen das Potenzial, Medikamente auch gegen so weit verbreitete Geißeln der Menschheit wie die Parkinsonsche oder die Alzheimersche Krankheit zu liefern. Die damit verbundene Hoffnung steigert den wirtschaftlichen Wert der bisher identifizierten Wirkstoffklassen und die Chance, dass Lizenzinteressenten gefunden werden können.

Es sei an dieser Stelle betont, dass unsere Leitstruktursuchen, neben den ausführlich diskutierten Klassen NBB und XXX, Hinweise auf weitere anti-aggregatorische Wirkstoffklassen geliefert haben, von denen einige bereits aus der Literatur bekannt sind (siehe z.B. die Stoffklasse aus Abbildung 4.2). Andere Klassen wurden jedoch bisher noch nicht beschrieben. Ihre eingehende Analyse stellt eine Aufgabe für eventuelle Nachfolgeprojekte dar. Daher repräsentieren die gesammelten Screening-Daten eine wertvolle Wissensbasis, und daraus identifizierte Leitstrukturen bieten zahlreiche neue Ansatzpunkte für die Entwicklung anti-aggregatorischer Medikamente.

Aus den genannten Gründen ist es wünschenswert, dass ein Nachfolgeprojekt ins Leben gerufen wird, das die Ziele des nun bald auslaufenden Projekts weiterverfolgt und auf die Entwicklung von Medikamenten gegen weitere neurodegenerative Krankheiten erweitert.

A Publikationen und Patente

A.1 Publikationen

Uwe Bertsch, Konstanze F. Winklhofer, Thomas Hirschberger, Jan Bieschke, Petra Weber, F. Ulrich Hartl, Paul Tavan, Jörg Tatzelt, Hans A. Kretzschmar, and Armin Giese: "Systematic Identification of Antiprion Drugs by High-Throughput Screening Based on Scanning for Intensely Fluorescent Targets", *J. Virol.* **79** 7785-7791 (2005).

Hans A. Kretzschmar, Uwe Bertsch, Konstanze F. Winklhofer, Thomas Hirschberger, Jan Bieschke, Petra Weber, F. Ulrich Hartl, Paul Tavan, Jörg Tatzelt, and Armin Giese: "Single Molecule Analysis of Protein Aggregation and Prions by SIFT", *J. Neuropathol. Exp. Neurol.* **64** 441 (2005).

Uwe Bertsch, Konstanze F. Winklhofer, Thomas Hirschberger, Jan Bieschke, Petra Weber, F. Ulrich Hartl, Paul Tavan, Jörg Tatzelt, Hans A. Kretzschmar, and Armin Giese: "Single Molecule Analysis of Protein Aggregation and Prions by SIFT", *FEBS J.* **272** 338-339 Suppl. 1 (2005).

Verena Schultheis, Thomas Hirschberger, Heiko Carstens, and Paul Tavan: "Extracting Markov Models of Peptide Conformational Dynamics from Simulation Data", *J. Chem. Theory Comput.* **1**, 515-526 (2005).

Thomas Hirschberger, Martina Stork, Bernhard Schropp, Konstanze F. Winklhofer, Jörg Tatzelt, and Paul Tavan: "Structural Instability of the Prion Protein upon M205S/R Mutations Revealed by Molecular Dynamics Simulations", *Biophys. J.* **90**, 3908-3918 (2006).

Niclas W. Schiffer, Sarah A. Broadley, Thomas Hirschberger, Paul Tavan, Hans A. Kretzschmar, Armin Giese, Christian Haass, F. Ulrich Hartl, and Bettina Schmid: "Identification of Anti-prion Compounds as Efficient Inhibitors of Polyglutamine Protein Aggregation in a Zebrafish Model", *J. Biol. Chem.* **282**, 9195-9203 (2007).

Thomas Hirschberger, Rudolf Reichold, Heiko Carstens, Martina Stork, and Paul Tavan: "Self-organization of hierarchical classification: Combining density estimation by normal mixtures with scale-space filtering", in preparation.

A.2 Patente

Als wir die Klasse der NBB-Substanzen als neue anti-Prion-Wirkstoffe identifiziert und durch statistische SAR-Untersuchungen ihre Wirksamkeit untermauert hatten (vgl. Abschnitte 5.1.1 und 5.1.2), reichten wir eine Erfindungsmeldung beim Patentbüro der LMU ein. Die LMU nahm unsere Erfindung unbeschränkt in Anspruch. Recherchen in Patentdatenbanken wie dem DEPATIS-System des Deutschen Patent- und Markenamtes [169] ergaben, dass die Klasse der NBB-Substanzen zwar bereits beispielsweise als Apoptose-Aktivatoren geschützt ist [170], jedoch zur Anwendung gegen neurodegenerative Krankheiten noch geschützt werden konnte. Mit Unterstützung einer Patentanwaltskanzlei meldeten wir das Patent

Uwe Bertsch, Armin Giese, Hans Kretzschmar, Paul Tavan, Thomas Hirschberger, Jan Bieschke, Petra Weber, Konstanze F. Winklhofer, Jörg Tatzelt, F. Ulrich Hartl, Gerda Wünsch, and Tobias J. Högen: "*Systematic Identification of New Anti-Prion Drugs by High-Throughput Screening Based on Scanning for Intensely Fluorescent Targets (SIFT)*", PCT/EP2005/005614, WO 2005/116640 A2.

an, welches sowohl die Screening-Methode als auch die Substanzen der NBB-Klasse als Medikamente gegen Prionkrankheiten und weitere neurodegenerative Krankheiten schützt.

Auch im Falle der zweiten neuen Wirkstoffklasse XXX (vgl. Abschnitt 5.2) reichten wir eine Erfindungsmeldung an die LMU ein, die von der LMU unbeschränkt in Anspruch genommen wird. Wiederum wurden umfangreiche Recherchen in Publikations- und Patentdatenbanken durchgeführt, wodurch die Patentierbarkeit dieser Substanzklasse festgestellt wurde. Eine zugehörige Patentmeldung wird derzeit verfasst.

B Die Java-Bibliothek `tseadb`

Unter Verwendung der Programmiersprache Java [171] wurde von meiner studentischen Hilfskraft Herrn Simon Youssef und mir die Bibliothek `tseadb` implementiert. Diese Bibliothek umfasst eine Schnittstelle zur TSE-DB, diverse Ein- und Ausgaberroutinen ins Dateisystem, Verfahren zur Ausreißererkenung und zur Wirksamkeitsberechnung (vgl. Abschnitte 3.1 bzw. 3.2), zahlreiche Komponenten für grafische Benutzerschnittstellen und einige weitere Funktionalitäten und stellt damit ein Software-Framework zur Verfügung, auf dem sowohl die in Abschnitt 3.5.3 beschriebene Datenbank-Client-Software TSE-DB-Browser als auch das Analysewerkzeug SIFT-Viewer II aus Abschnitt 3.6 basieren. Bevor nun einzelne Bestandteile der Bibliothek ausführlicher geschildert werden, sollen einige Werkzeuge und Methoden beschrieben werden, die bei der Entwicklung der Bibliothek von Herrn Youssef und mir eingesetzt wurden.

Für die Programmierung der `tseadb`-Bibliothek und der darauf aufsetzenden Anwendungen wurden zum einen die kommerzielle integrierte Entwicklungsumgebung JBuilder der Firma Borland [134] und zum anderen das frei verfügbare Werkzeug Eclipse [172], welches dem selben Zweck dient, eingesetzt^a. Zur Verwaltung des damit programmierten Quellcodes der `tseadb`-Bibliothek wurde der am Lehrstuhl für BioMolekulare Optik (BMO) eingerichtete CVS-Server verwendet, der bereits in Abschnitt 3.4 im Zusammenhang mit SIFT-Viewer I angesprochen wurde. Der `tseadb`-Quellcode ist im CVS-Modul „TSE_DB“ abgelegt und darüber für die Mitarbeiter des BMO zugänglich. JBuilder und Eclipse unterstützen den Zugriff auf Versionskontrollsysteme wie CVS und erleichtern damit erheblich die Softwareentwicklung in einer Gruppe mit mehreren Entwicklern.

Für eine detaillierte Beschreibung der `tseadb`-Bibliothek sei auf die im Quellcode enthaltenen Kommentare und auf das Java-Werkzeug `javadoc` verwiesen, mit dessen Hilfe aus dem Quellcode der Bibliothek und den darin enthaltenen Kommentaren eine HTML-basierte Dokumentation generiert werden kann. Werkzeuge wie JBuilder oder Eclipse unterstützen die Dokumentationsgenerierung mit `javadoc` und vereinfachen diesen Vorgang erheblich. An dieser Stelle soll lediglich ein Überblick über die wichtigsten in der `tseadb`-Bibliothek umgesetzten Funktionalitäten gegeben werden.

Zur Umsetzung der angesprochenen Funktionalitäten besteht die `tseadb`-Bibliothek aus einer umfangreichen Sammlung von etwa 140 Java-Klassen – das sind Objekte der Sprache Java, die der Datenhaltung und dem Zurverfügungstellen von Methoden dienen. Die Klassen sind nach inhaltlicher Zusammengehörigkeit in sogenannten Paketen gruppiert,

^aJBuilder ist auf dem Windows-Laptop „lisa“ und Eclipse auf dem Linux-Rechner „celeritas“ installiert.

```
tse
|-- generated
|  |-- db
|-- data
|  |-- sift
|-- sql
|-- io
|  |-- imports
|  |  |-- sift
|  |-- exports
|-- algorithm
|  |-- outlier
|  |-- activity
|  |-- classification
|-- gui
|  |-- dbbrowser
|  |  |-- infopane
|  |  |-- substancetree
|  |-- sift
|  |  |-- event
|  |  |-- plate
|  |  |-- selectors
|  |  |-- views
|  |-- pictures
|-- tools
```

Abbildung B.1: Paketstruktur der Java-Bibliothek `tse`. Die Bibliothek `tse` stellt grundlegende Datenstrukturen und Funktionalitäten wie Datenbankankbindung, Dateisystem-Ein- und Ausgabe, verschiedene Analyseverfahren sowie Komponenten der grafischen Benutzerschnittstelle zur Verfügung. Aus diesen Bausteinen sind die ebenfalls in der `tse`-Bibliothek enthaltenen Anwendungen TSE-DB-Browser (Abschnitt 3.5.3) und SIFT-Viewer II (Abschnitt 3.6) zusammengesetzt.

welche zudem in einer hierarchischen Struktur angeordnet sind. Anhand der in Abbildung B.1 gezeigten Paketstruktur der `tse`-Bibliothek sollen nun die in den Paketen umgesetzten Funktionalitäten erläutert werden.

In den Paketen `tse.generated.db`, `tse.data` und `tse.data.sift` werden Datenstrukturen definiert, die ausschließlich zur Haltung von Daten dienen. Dem Gedanken der Objektorientierung folgend wurden dabei die realen Objekte der Medikamentensuche wie Substanzen, Platten, Bibliotheken, Messdaten, Messergebnisse etc. auf virtuelle Objekte in Gestalt von Java-Klassen abgebildet. Diese Abbildung geschah indirekt anhand der in Abschnitt 3.5.1 beschriebenen Tabellen der TSE-DB, deren Einträge die Objekte der realen Welt repräsentieren. Aus den Definitionen der Datenbanktabellen wurden dabei die Java-Klassen des Pakets `tse.generated.db` generiert, wozu die von Herrn Youssef zur Verfügung gestellte Bibliothek `de.simon` eingesetzt wurde, die von ihm bereits vor seiner Tätigkeit am BMO für derartige objektrelationale Abbildungen entwickelt worden war. In den Paketen `tse.data` und `tse.data.sift` sind *handgeschriebene* Klassen enthalten, welche die aus den Datenbanktabellen generierten Klassen zu komplexeren Objekten zusammensetzen. Als Beispiel eines auf diese Weise zusammengesetzten komplexen Objektes sei hier eine Klasse genannt, welche eine mit

der SIFT-Technik gemessene Platte repräsentiert und dazu die auf der Platte enthaltenen Substanzen sowie die zugehörigen SIFT-Messdaten zusammenfügt.

Das angesprochene Paket `de.simon`, das zur statischen objektrelationalen Abbildung von Datenbanktabellen auf Java-Klassen verwendet wurde, bildet ebenfalls die Grundlage der dynamischen Datenbankfunktionalitäten der `tsedb`-Bibliothek. Mithilfe dieses Pakets können Verbindungen zur TSE-DB aufgebaut und verwaltet, SQL-Anfragen an die Datenbank gestellt und die dabei erhaltenen Ergebnisse in Java-Klassen übersetzt werden. Zu diesen Zwecken verwendet das Paket `de.simon` die Schnittstelle „Java Database Connectivity“ (JDBC), die eine einheitliche Behandlung verschiedener relationaler Datenbanksysteme ermöglicht. Für den Zugriff auf den vom LRZ betriebenen Oracle9i-Datenbankserver, auf dem die TSE-DB eingerichtet ist, wurde der von Oracle speziell für diesen Servertyp angebotene JDBC-Treiber [173] („`classes12.zip`“) verwendet.

Das Paket `tsedb.sql` erweitert die allgemeinen Funktionalitäten des Pakets `de.simon` hinsichtlich der speziellen TSE-DB-bezogenen Datenstrukturen. Damit können aus den Ergebnissen von SQL-Anfragen an die TSE-DB entsprechende Instanzen von Java-Klassen erzeugt werden. Dazu sind die meisten der in `tsedb.sql` enthaltenen Klassen nach dem sogenannten Factory-Entwurfsmuster implementiert (siehe Gamma et al. [174] für eine Sammlung von Entwurfsmustern für objektorientiertes Softwaredesign). Durch den Einsatz des Factory-Musters wird eine Trennung der Definition und der Erzeugung von Datenstrukturen vollzogen, die das Framework flexibel hinsichtlich möglicher zukünftiger Änderungen der Datenstrukturen macht.

Die Pakete `tsedb.io.imports` und `tsedb.io.imports.sift` stellen zum einen Methoden für den Import von Daten in die TSE-DB und zum anderen Methoden zum Einlesen von SIFT-Dateien (vgl. Abschnitt 3.4.1) bereit [– Das Einlesen von SIFT-Layout-Dateien (vgl. Abbildung 3.7) ist in der Klasse `PlateLayout` des `tsedb.data.sift` implementiert]. Im Paket `tsedb.io.export` finden sich Exporttools wie beispielsweise ein Werkzeug zum Schreiben von pdf-Reports der mit SIFT-Viewer II betrachteten Messdaten. Die Reportgenerierung setzt dazu die freie Java-Bibliothek `iText` [139] ein.

Auch einige weitere Teile der Bibliothek wie das Paket `tsedb.algorithm` und dessen Unterpakete sind für die Verwendung von SIFT-Viewer II vorgesehen. Im Paket `tsedb.algorithm.outlier` sind die verschiedenen, in Abschnitt 3.1 beschriebenen Algorithmen zur Erkennung von Ausreißern implementiert, das Paket `tsedb.algorithm.activity` dient zur Berechnung von Wirksamkeitsmaßen nach den in Abschnitt 3.2 definierten Methoden, und im Paket `tsedb.algorithm.classification` ist die in Anhang C beschriebene Klassifikation von 2D-SIFT-Spektren umgesetzt.

Im Paket `tsedb.gui` und seinen Unterpaketen werden Komponenten der grafischen Benutzerschnittstelle (GUI, graphical user interface) definiert, welche von den Anwendungen TSE-DB-Browser und SIFT-Viewer II verwendet werden. Zu diesen GUI-Komponenten zählen beispielsweise ein Passwort-Dialog, der zur Herstellung einer Verbindung mit der TSE-DB erscheint sowie eine Komponente zur Darstellung von Molekülstrukturen, wel-

che auf der frei verfügbaren Java-Bibliothek CDK (Chemistry Development Kit [137]) basiert. Des weiteren zählen zu den von beiden Anwendungen genutzten GUI-Komponenten eine Tabellendarstellung von Substanznamen, Molekülstrukturen und Wirksamkeiten sowie die im Paketverzeichnis `tseadb.gui.pictures` abgelegten Symboldateien für die Verzierung von Knöpfen.

Die Komponenten aus dem Paket `tseadb.gui.dbbrowser` und dessen Unterpaketen werden hingegen nur vom TSE-DB-Browser genutzt. Dazu zählen ein navigierbarer Baum zur Auswahl von Bibliotheksplatten und Substanzen sowie einige unsichtbare Komponenten, die zur Verarbeitung von durch Benutzerinteraktionen ausgelösten Ereignissen dienen.

In analoger Weise sind die im Paket `tseadb.gui.sift` und unterhalb davon befindlichen GUI-Komponenten für die Anwendung SIFT-Viewer II vorgesehen. Zu diesen Komponenten zählen die verschiedenen grafischen Töpfchenrepräsentationen von Messplatten aus dem Paket `tseadb.gui.sift.selectors`, die zur Auswahl von Töpfchen dienen. Dynamische Chart-Komponenten wie Plots von SIFT-Spektren, Summenobservablen und Aktivitäten, Dosis-Wirkungs-Kurven und weitere grafische Darstellungen, basieren auf der frei verfügbaren Java-Bibliothek JFreeChart [136] und finden sich im Paket `tseadb.gui.sift.views`. In `tseadb.gui.sift.plate` werden mehrere der genannten Komponenten zur Repräsentation einer Messplatte zusammengestellt. Das Paket `tseadb.gui.sift.event` dient der Verarbeitung von Ereignissen, die zur Laufzeit infolge von Benutzerinteraktionen auftreten. Es sei angemerkt, dass die Funktionalitäten dieser Komponenten auch in Abschnitt 3.6 im Zusammenhang mit den Fähigkeiten von SIFT-Viewer II erläutert werden.

Das Paket `tseadb.tools` enthält einige kleinere Werkzeuge ohne grafische Benutzeroberfläche, beispielsweise eines, das aus den SIFT-Messdaten für alle DIVERSet-Bibliotheken die Wirksamkeiten berechnet und von mir bei der Befüllung der TSE-DB-Tabelle `SIFT_PRIMARY_TP_1` eingesetzt wurde (vgl. Abschnitt 3.5.2).

Als Hinweis für an der Java-Technik Interessierte sei angemerkt, dass der Einsatz von Java ermöglichte, frei verfügbare Programmbibliotheken bei der Entwicklung der `tseadb`-Bibliothek zu verwenden. Dazu zählen neben den oben genannten Bibliotheken auch das Framework `log4j` [175], welches für die Ausgabe von Meldungen zur Information des Benutzers oder bei auftretenden Fehlern eingesetzt wurde. Anstatt solche Meldungen auf die Standardausgabe auszugeben, werden die Meldungen dem Loggingsystem von `log4j` übergeben. Beim Erstellen des Programms legt der Programmierer lediglich den Typ der übergebenen Meldung fest, wobei zwischen Informations-, Warn- und Fehlermeldungen unterschieden wird. Zur Laufzeit des Programms können dann Ziel und Umfang der Ausgabe konfiguriert werden. Auf diese Weise kann etwa für die Benutzer eine Programmversion eingestellt werden, die nur wichtige Informationen ausgibt, während die Entwickler beispielsweise zur Suche nach Programmfehlern eine wortreichere Konfiguration wählen können.

C Eine Methode zur vorurteilsfreien Klassifikation von SIFT-Daten

Bei den Screening-Kampagnen der DIVERSet-Bibliotheken im SIFT-Assay (vgl. Kapitel 5) wurden für die Auswahl von Treffersubstanzen anhand der gemessenen 2D-SIFT-Spektren das auf den Summenobservablen S_1^5 basierende kontrollenbezogene Wirksamkeitsmaß (vgl. Abschnitt 3.2) sowie ein heuristisch festgelegter Grenzwert eingesetzt, der die wirksamen von den unwirksamen Testsubstanzen trennte (vgl. Abbildung 5.1). Obwohl diese Vorgehensweise erfolgreich zur Treffererkennung eingesetzt wurde, hat sie den Nachteil, dass ihr einige (sicherlich gut begründete) Vorurteile zugrunde liegen, wie beispielsweise die Annahme, dass die Summenobservable S_1^5 die anti-aggregatorische Wirksamkeit kodiert.

Um diesen Nachteil zu vermeiden, habe ich ein alternatives Verfahren zur Trefferauswahl entwickelt, welches die Testsubstanzen anhand der 2D-SIFT-Spektren vorurteilsfrei in Klassen von ähnlicher spektraler Charakteristik unterteilt. Das Verfahren basiert dazu auf einem selbstorganisierenden Klassifikationsalgorithmus, den ich im Rahmen meiner Diplomarbeit [176] in der Arbeitsgruppe von Prof. Paul Tavan entwickelt und in das dort existierende Softwarepaket *multivar* [177, 178] integriert hatte. Das Verfahren wurde zunächst in das Shell-Skript-basierte Paket SIFT-Viewer I und später in das Java-Werkzeug SIFT-Viewer II integriert. Dazu war es notwendig, das Paket *multivar* in die Sprache Java zu portieren. Die resultierende Java-Bibliothek *jGRBF* (generalisierte radiale Basisfunktionen) steht als gleichnamiges Modul im CVS zur Verfügung. Bei den *jGRBF*- und SIFT-Viewer II-bezogenen Programmieraufgaben unterstützte mich wiederum Simon Youssef als studentische Hilfskraft. Das Verfahren umfasst mehrere Stufen, die im Folgenden beschrieben werden.

C.1 Das Klassifikationsverfahren

Gemäß dem hier vorgestellten Verfahren werden die SIFT-Messdaten plattenweise prozessiert. Dazu werden die für eine Messplatte mit N Töpfchen w gewonnenen 2D-SIFT-Spektren \mathbf{x}_w , die gemäß Gleichung (3.5) $D = 18$ -dimensionale Vektoren sind, in einem Datensatz

$$\mathcal{X} = \{\mathbf{x}_w \mid w = 1, \dots, N\} \subset \mathbb{R}^D$$

zusammengefasst. Zunächst wird der Datensatz \mathcal{X} einer Ausreißerbereinigung mithilfe der nach dem Töpfchentyp trennenden Erkennungsmethode (vgl. Abschnitt 3.1.2) unterzogen, wobei als Prüfgrößen neben den Gesamtintensitäten $I_{\text{tot},r}$ und $I_{\text{tot},r}$ zusätzlich die gemäß Gleichung (3.6) gebildeten Summenobservablen S_1^{18} eingesetzt werden.

C.1.1 Vorverarbeitung

Extraktion von Merkmalen

Die verbleibenden $N_v \leq N$ validen Spektren \mathbf{x}_w werden einem mehrstufigen Transformationsprozess, der sogenannten „Vorverarbeitung“, unterzogen. Dieser Prozess der Vorverarbeitung wurde ursprünglich in der Gruppe von Paul Tavan zur Aufbereitung von Sprachsignalen in einem Spracherkennungssystem entwickelt [179, 180]; dort verwendete Bezeichnungen werden hier übernommen. Neben der Transformation der Spektren dient die Vorverarbeitung der Extraktion von Merkmalen. Zuerst wird hierbei für jedes Spektrum \mathbf{x}_w der sogenannte *Grauwert*^a

$$g_w = \sum_{i=1}^D x_{wi} \geq 0 \quad (\text{C.1})$$

als erstes Merkmal extrahiert. Zudem werden die Spektren \mathbf{x}_w in die grauwertbereinigten Spektren

$$\mathbf{x}'_w = \mathbf{x}_w - g_w \mathbf{1}$$

überführt, indem jeweils komponentenweise der Grauwert g_w subtrahiert wird. Anschließend wird töpfchenweise der sogenannte *Kontrast*

$$c_w = \sqrt{\sum_{i=1}^D (x'_{wi})^2} \quad (\text{C.2})$$

der grauwertbereinigten Spektren \mathbf{x}'_w als weiteres Merkmal gespeichert. Die Vektoren \mathbf{x}'_w werden schließlich in die kontrastnormierten Spektren

$$\mathbf{x}''_w = (1/c_w) \mathbf{x}'_w$$

überführt, indem sie jeweils mit dem Kehrwert des Kontrastes c_w skaliert werden. Die resultierenden grauwertbereinigten und kontrastnormierten Spektren \mathbf{x}''_w werden in dem Datensatz

$$\mathcal{X}'' = \{\mathbf{x}''_w \mid w = 1, \dots, N_v\}$$

zusammengefasst.

^aDer Grauwert g_w eines Spektrums \mathbf{x}_w stimmt mit der Summenobservable $S_1^{18}(w)$ überein.

Dimensionsreduktion durch Hauptkomponentenanalyse

Der Datensatz \mathcal{X}'' wird einer sogenannten Hauptkomponentenanalyse (principal component analysis, PCA) unterzogen [181]. Dabei wird zum einen der Mittelvektor

$$\mu = \frac{1}{N_v} \sum_{w=1}^{N_v} \mathbf{x}_w''$$

des Datensatzes \mathcal{X}'' gebildet, sodass die Spektren \mathbf{x}_w'' gemäß

$$\mathbf{x}_w''' = \mathbf{x}_w'' - \mu$$

in das Schwerpunktsystem transformiert werden können. Zum anderen wird die Kovarianzmatrix

$$\mathbf{K} \in \mathbb{R}^{D \times D} \quad \text{mit} \quad K_{mn} = \frac{1}{N_v} \sum_{w=1}^{N_v} x_{wm}''' x_{wn}''' \quad \text{für} \quad m, n \in \{1, \dots, D\}$$

des Datensatzes \mathcal{X}'' berechnet. Die Kovarianzmatrix \mathbf{K} wird durch Diagonalisierung in die Transformationsmatrix

$$\mathbf{T} \in \mathbb{R}^{D \times D}$$

überführt, deren Zeilen durch die Eigenvektoren von \mathbf{K} gebildet werden, die normiert und nach der Größe der zugehörigen Eigenwerte λ_i ($i = 1, \dots, D$) sortiert sind. Unter Verwendung der Transformationsmatrix \mathbf{T} können die D -dimensionalen Spektren \mathbf{x}_w''' in einen Unterraum der reduzierten Dimension $D_r < D$ projiziert werden. Dazu wird die Teilmatrix

$$M_{mn} = T_{mn} \quad m \in \{1, \dots, D_r\}, \quad n \in \{1, \dots, D\}$$

der Transformationsmatrix \mathbf{T} definiert, mit welcher die Vektoren \mathbf{x}_w''' multipliziert und dadurch in dimensionsreduzierte Vektoren

$$\mathbf{x}_w'''' = \mathbf{M}^t \mathbf{x}_w''' \in \mathbb{R}^{D_r} \tag{C.3}$$

überführt werden. Diese Transformation dient der Rauschunterdrückung.

Konstruktion von Merkmalsvektoren

Um die Varianz der Merkmale (C.1) und (C.2) auf eine ähnliche Größe zu bringen, wie jene der Komponenten von (C.3), wird der Mittelwert $\bar{\lambda} = (1/D_r) \sum_{i=1}^{D_r} \lambda_i$ der D_r größten

Eigenwerte λ_i der Kovarianzmatrix \mathbf{K} sowie die Varianzen $\text{var}_w(g_w)$ und $\text{var}_w(c_w)$ der Grauwerte g_w bzw. der Kontraste c_w herangezogen, um geeignete Faktoren

$$\kappa_g = \sqrt{\bar{\lambda}/\text{var}_w(g_w)} \quad \text{und} \quad \kappa_c = \sqrt{\bar{\lambda}/\text{var}_w(c_w)}$$

zu definieren, mit welchen die Grauwerte g_w und Kontraste c_w gemäß

$$g'_w = \kappa_g g_w \quad \text{bzw.} \quad c'_w = \kappa_c c_w$$

skaliert werden. Schließlich werden die dimensionsreduzierten Vektoren $\mathbf{x}_w'''' \in \mathbb{R}^D$ mit den skalierten Grauwerten g'_w und Kontrasten c'_w zu Merkmalsvektoren

$$\hat{\mathbf{x}}_w = (x_{w1}''', \dots, x_{wD_r}''', g'_w, c'_w)^t \in \mathbb{R}^{D_m} \quad \text{mit} \quad D_m = D_r + 2 \quad (\text{C.4})$$

in dem Merkmalsraum \mathbb{R}^{D_m} der Dimension $D_m = D_r + 2$ kombiniert. Die Merkmalsvektoren $\hat{\mathbf{x}}_w$ aller N_v validen Töpfchen w der gegebenen Platte werden in dem *Merkmalsdatensatz*

$$\hat{\mathcal{X}} = \{\hat{\mathbf{x}}_w \mid w = 1, \dots, N_v\} \subset \mathbb{R}^{D_m} \quad (\text{C.5})$$

zusammengefasst, der den Ausgangspunkt für die Bildung von Klassen darstellt.

C.1.2 Bildung von Klassen

Bildung eines Codebuchs und eines Dichtemodells

Zunächst wird der *univar*-Algorithmus von Kloppenburg und Tavan [177] dazu eingesetzt, eine reduzierte Repräsentation des Merkmalsdatensatzes $\hat{\mathcal{X}}$ in Form eines sogenannten *Codebuchs*

$$\mathcal{C} = \{\mathbf{c}_r \mid r = 1, \dots, M\} \subset \mathbb{R}^{D_m}$$

zu bilden, welches aus $M \ll N_v$ Codebuchvektoren \mathbf{c}_r besteht. Mit dem Codebuch ist ein Modell für die Dichte des Datensatzes $\hat{\mathcal{X}}$ in Gestalt einer Mischung

$$\hat{p}(\hat{\mathbf{x}} \mid \mathcal{C}, \sigma) = \frac{1}{M} \sum_{r=1}^M \hat{p}(\hat{\mathbf{x}} \mid \mathbf{c}_r, \sigma) \quad (\text{C.6})$$

aus M gleich gewichteten univariaten Gaußfunktionen

$$\hat{p}(\hat{\mathbf{x}} \mid \mathbf{c}_r, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^D} \exp\left[-\frac{(\hat{\mathbf{x}} - \mathbf{c}_r)^2}{2\sigma^2}\right], \quad (\text{C.7})$$

verbunden, dessen Parameter, die Zentren \mathbf{c}_r und die Breite σ , vom *univar*-Algorithmus optimiert werden, indem die sogenannte Likelihood dieser Parameter maximiert wird. Als Ergebnis liegen die Zentren \mathbf{c}_r an lokalen Mittelwerten des Merkmalsdatensatzes $\hat{\mathcal{X}}$, und die Breite σ ist im Sinne der Maximierung der Likelihood optimal [177].

Prototypische Merkmalsvektoren und scharfe Klassen

Gemäß dem Klassifikationsverfahren aus meiner Diplomarbeit [176] wird nun die Gaußbreite σ des Dichtemodells $\hat{p}(\hat{\mathbf{x}} | \mathcal{C}, \sigma)$ variiert, um eine Schar von Dichtemodellen auf verschiedenen räumlichen Skalen σ zu erzeugen. Mithilfe eines Gradientenaufstiegsverfahrens werden die Maxima der Dichtemodelle $\hat{p}(\hat{\mathbf{x}} | \mathcal{C}, \sigma)$ gefunden und als prototypische Merkmalsvektoren

$$\mathbf{y}_\sigma^j \quad \text{für } j = 1, \dots, N_\sigma \quad (\text{C.8})$$

aufgefasst. Der Index σ zeigt dabei an, dass die Lage der Prototypen \mathbf{y}_σ^j und ihre Anzahl N_σ von der gewählten Gaußbreite σ des Dichtemodells abhängen. Für genügend große Werte von σ besitzt das Dichtemodell ein einziges Maximum, und es wird ein Prototyp gebildet. Für abnehmende Werte von σ steigt die Zahl N_σ der Prototypen monoton an, bis schließlich für $\sigma \rightarrow 0$ jeder der M Codebuchvektoren \mathbf{c}_r einen Prototypen \mathbf{y}_σ^j definiert.

Das Gradientenaufstiegsverfahren gestattet die Zuordnung

$$\hat{\mathbf{x}} \rightarrow \mathbf{y}_\sigma^j \quad (\text{C.9})$$

von beliebigen Merkmalsvektoren $\hat{\mathbf{x}} \in \mathbb{R}^{D_m}$ zu den Prototypen \mathbf{y}_σ^j . Es kann also dazu eingesetzt werden, die Merkmalsvektoren $\hat{\mathbf{x}}_w$ der Töpfchen w auf verschiedenen Skalen σ den Prototypen \mathbf{y}_σ^j zuzuordnen. Diese Zuordnungen definieren *scharfe* Klassen

$$\mathcal{K}_\sigma^j(\hat{\mathcal{X}}) = \left\{ \hat{\mathbf{x}}_w \in \hat{\mathcal{X}} \mid \hat{\mathbf{x}}_w \rightarrow \mathbf{y}_\sigma^j \right\} \subset \hat{\mathcal{X}} \quad \text{für } j = 1, \dots, N_\sigma \quad (\text{C.10})$$

und zerlegen den Datensatz $\hat{\mathcal{X}}$ in N_σ disjunkte Teilmengen. Wie die Prototypen \mathbf{y}_σ^j hängen auch die Klassen $\mathcal{K}_\sigma^j(\hat{\mathcal{X}})$ von der gewählten Skala σ ab.

Die Implementierung dieses Klassifikationsalgorithmus in SIFT-Viewer II gestattet es, dass (im Quellcode) ein gewünschter Bereich $[N_{\min}, N_{\max}]$ vorgegeben werden kann, innerhalb dessen die Zahl N_σ der Prototypen liegen soll. Es wird dann ein dazu passendes Intervall $[\sigma_{\min}, \sigma_{\max}]$ für den Skalenparameter σ gefunden.

Identifikation einer Trefferklasse

Falls auf einer gewählten Auflösungsstufe σ eine Klasse $\mathcal{K}_\sigma^j(\hat{\mathcal{X}})$ existiert, welcher die Merkmalsvektoren $\hat{\mathbf{x}}_w$ aller Töpfchen $w \in \mathcal{P}$ mit Positiv-Kontrollen zugeordnet werden, so wird diese Klasse als *Trefferklasse*

$$\mathcal{T}_\sigma := \mathcal{K}_\sigma^j(\hat{\mathcal{X}}) \quad (\text{C.11})$$

definiert. Enthält die Klasse \mathcal{T}_σ auf der gewählten Skala σ neben den Merkmalsvektoren der Positiv-Kontrollen auch solche von Testsubstanzen, so werden diese als Treffer angesehen. Über diese scharfe Klassifikation hinaus soll eine unscharfe Zuordnung der Merkmalsvektoren $\hat{\mathbf{x}}_w$ von Testsubstanzen zur Trefferklasse \mathcal{T}_σ auf der gewählten Auflösungsstufe σ getroffen und anhand dieser ein neues Wirksamkeitsmaß konstruiert werden.

C.1.3 Konstruktion eines neuen Wirksamkeitsmaßes

Nach dem Bayesschen Satz definieren die Komponenten (C.7) des Dichtemodells (C.6) M *unscharfe* Klassen r , die durch die Zuordnungswahrscheinlichkeiten

$$\hat{P}(r | \hat{\mathbf{x}}, \mathcal{C}, \sigma) = \frac{\frac{1}{M} \hat{p}(\hat{\mathbf{x}} | \mathbf{c}_r, \sigma)}{\hat{p}(\hat{\mathbf{x}} | \mathcal{C}, \sigma)} \quad \text{für } r = 1, \dots, M \quad (\text{C.12})$$

charakterisiert sind. Dabei wird ein gegebener Merkmalsvektor $\hat{\mathbf{x}}$ gemäß den Wahrscheinlichkeiten $\hat{P}(r | \hat{\mathbf{x}}, \mathcal{C}, \sigma) \in [0, 1]$ den M Klassen r zugeordnet.

Die zu den *unscharfen* Klassen r gehörigen Codebuchvektoren \mathbf{c}_r werden gemäß (C.9) den Prototypen \mathbf{y}_σ^j zugeordnet ($\mathbf{c}_r \rightarrow \mathbf{y}_\sigma^j$), welche die *scharfen* Klassen $\mathcal{K}_\sigma^j(\hat{\mathcal{X}})$ definieren. Dadurch werden die M unscharfen Klassen r zu $N_\sigma \leq M$ unscharfen Klassen j vereinigt, die durch die Zuordnungswahrscheinlichkeiten

$$\hat{P}(j | \hat{\mathbf{x}}, \mathcal{C}, \sigma) = \sum_{r \in \{r | \mathbf{c}_r \rightarrow \mathbf{y}_\sigma^j\}} \hat{P}(r | \hat{\mathbf{x}}, \mathcal{C}, \sigma) \quad \text{für } j = 1, \dots, N_\sigma$$

charakterisiert sind. Hier werden die Zuordnungswahrscheinlichkeiten zu denjenigen Klassen r aufsummiert, deren Codebuchvektoren \mathbf{c}_r dem Prototypen \mathbf{y}_σ^j zugeordnet werden.

Eine der unscharfen Klassen j gehört zu der in (C.11) definierten Trefferklasse \mathcal{T}_σ ; ihr Index sei $j = t$. Die Zuordnungswahrscheinlichkeit des gegebenen Merkmalsvektors $\hat{\mathbf{x}}_w$ zu dieser Klasse t wird als neues *Wirksamkeitsmaß*

$$\hat{a} : \mathcal{W} \rightarrow [0, 1], \quad w \mapsto \hat{a}(w | \mathcal{C}, \sigma) := \hat{P}(t | \hat{\mathbf{x}}_w, \mathcal{C}, \sigma) \quad (\text{C.13})$$

definiert, welches Töpfchen w der Menge \mathcal{W} Werte aus dem Intervall $[0, 1]$ zuordnet. Je wahrscheinlicher der Merkmalsvektor $\hat{\mathbf{x}}_w$ eines gegebenen Töpfchens w zu der unscharfen Trefferklasse t gehört, desto näher liegt seine Wirksamkeit $\hat{a}(w | \mathcal{C}, \sigma)$ bei 1.

C.2 Eine exemplarische Anwendung

Um die Fähigkeiten des skizzierten Klassifikationsverfahrens zu demonstrieren, habe ich es exemplarisch auf die Screening-Messung einer DIVERSet-Platte angewendet. Abbildung C.1 zeigt die von SIFT-Viewer II generierte Repräsentation der verwendeten Messplatte, in welcher die Knöpfe nach den Typen der repräsentierten Töpfchen eingefärbt sind. Wie es beim primären Screening im SIFT-Assay üblich war, wurden in den Töpfchen der Zeilen A bis H und der Spalten 2 bis 11 die 80 Substanzen der ausgewählten DIVERSet-Platte, in der ersten Spalte in den Zeilen A, C und E Negativ-, in den Zeilen B, D und F Positiv- und in den Zeilen G und H Puffer-Kontrollen gemessen (vgl. das Plattenlayout in Abbildung 2.6).

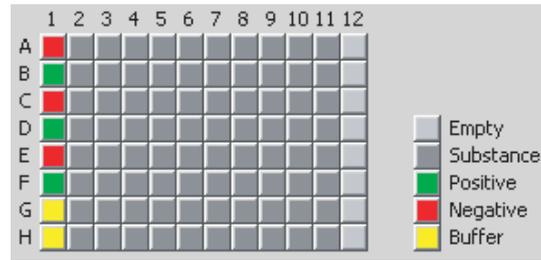


Abbildung C.1: Layout der für die Klassifikation verwendeten Plattenmessung. Gezeigt ist die von SIFT-Viewer II generierte Repräsentation der verwendeten Messplatte des DIVERSet-Screenings. Wie der Legende entnommen werden kann, sind die Knöpfe nach den Typen der repräsentierten Töpfchen gefärbt (vgl. Abbildung 3.8A). Beim primären DIVERSet-Screening wurden in Spalte 1 in den Zeilen A, C und E Negativ-, in den Zeilen B, D und F Positiv- und in den Zeilen G und H Puffer-Kontrollen gemessen. In den Töpfchen der Spalten 2 bis 11 befanden sich Testsubstanzen; Spalte 12 blieb leer (vgl. das Plattenlayout in Abbildung 2.6).

Die Daten der 88 befüllten Töpfchen wurden der Ausreißerererkennung unterzogen, wobei fünf Töpfchen aussortiert wurden. Die rohen 18-dimensionalen 2D-SIFT-Spektren \mathbf{x}_w der 83 verbliebenen Töpfchen wurden dem Vorverarbeitungsprozess aus Abschnitt C.1.1 unterzogen und dabei in Vektoren (C.3) der reduzierten Dimension $D_r = 5$ transformiert, die gemäß (C.4) mit den Grauwerten (C.1) und Kontrasten (C.2) zu Merkmalsvektoren $\hat{\mathbf{x}}_w$ der Dimension $D_m = 5 + 2$ zusammengesetzt wurden. Für den resultierenden Merkmalsdatensatz (C.5) wurde unter Einsatz des *univar*-Algorithmus ein Dichtemodell (C.6) aus $M = 9$ univariaten Gaußfunktionen gebildet. Mithilfe des Gradientenaufstiegsverfahrens wurden auf verschiedenen Skalen σ gemäß (C.8) Prototypen und gemäß (C.10) zugehörige scharfe Klassen \mathcal{K}_σ^j gebildet, deren Anzahlen zwischen $N_{\min} = 1$ und $N_{\max} = 9$ lagen.

Abbildung C.2 zeigt die Ergebnisse der Klassifikation auf einer Skala σ mit $N_\sigma = 2$ Klassen \mathcal{K}_σ^1 und \mathcal{K}_σ^2 . In **A** ist eine von SIFT-Viewer II generierte Repräsentation der Messplatte gezeigt, in welcher die Knöpfe nach der Zugehörigkeit der Merkmalsvektoren $\hat{\mathbf{x}}_w$ zu den beiden Klassen eingefärbt sind (\mathcal{K}_σ^1 rot, \mathcal{K}_σ^2 grün). Weiß durchgestrichene und schwarz hinterlegte Knöpfe markieren die Ausreißertöpfchen, denen auch die beiden Puffer-Kontrollen (Töpfchen G1 und H1) angehören. Die meisten Töpfchen wurden der roten Klasse \mathcal{K}_σ^1 zugeordnet. Die grüne Klasse \mathcal{K}_σ^2 umfasst alle drei Positiv-Kontrollen (Töpfchen B1, D1 und F1) und stellt daher gemäß (C.11) die Trefferklasse \mathcal{T}_σ dar. Dieser Klasse gehört auch das Töpfchen C3 an; die darin gemessene Testsubstanz wird als Treffer angesehen.

In **B** sind die Merkmalsvektoren $\hat{\mathbf{x}}_w$ für die validen Töpfchen dargestellt und gemäß der selben Farbkodierung eingefärbt. Die grün gefärbten Merkmalsvektoren, die der Trefferklasse $\mathcal{T}_\sigma = \mathcal{K}_\sigma^2$ zugeordnet wurden unterscheiden sich von jenen der Klasse \mathcal{K}_σ^1 besonders deutlich durch die ersten beiden Komponenten der dimensionreduzierenden Transformation (PC1 und PC2) sowie durch die skalierten Grauwerte und Kontraste.

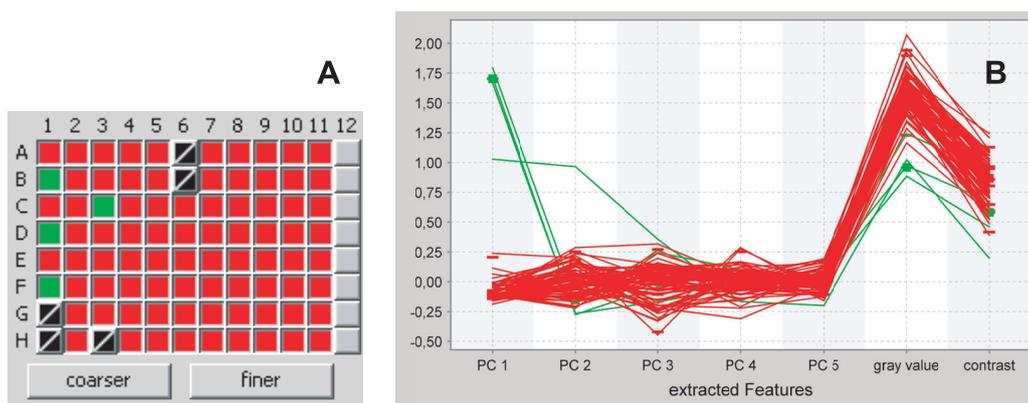


Abbildung C.2: Klassen und Merkmalsvektoren. Gezeigt sind Bildschrimschnappschüsse von SIFT-Viewer II, welche für die exemplarische Messplatte die Ergebnisse der Klassifikation auf einer Aulösungsstufe σ mit zwei Klassen \mathcal{K}_σ^1 und \mathcal{K}_σ^2 darstellen. **A** In der Plattenrepräsentation sind die Knöpfe gemäß der Zuordnung der Töpfchen zu den beiden Klassen \mathcal{K}_σ^1 (rot) und \mathcal{K}_σ^2 (grün, Trefferklasse) eingefärbt. **B** Die gemäß (C.4) gebildeten Merkmalsvektoren sind nach der selben Farbkodierung gefärbt. Die Knöpfe „coarser“ und „finer“ in **A** ermöglichen das Umschalten auf die nächstgrößere bzw. nächstfeinere Aulösungsstufe σ .

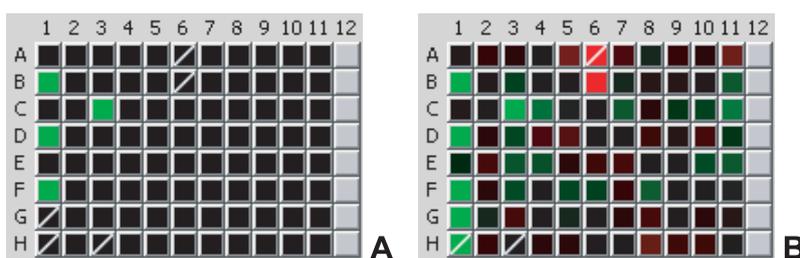


Abbildung C.3: Wirksamkeitsmaße im Vergleich. Die Knöpfe der gezeigten Plattenrepräsentationen sind in **A** gemäß dem Bayesschen Wirksamkeitsmaß (C.13) und zum Vergleich in **B** gemäß dem kontrollenbezogenen Maß (3.9) eingefärbt.

Abbildung C.3 zeigt zwei von SIFT-Viewer II generierte Repräsentationen, in welchen die Knöpfe nach den Wirksamkeiten eingefärbt sind, die in **A** gemäß dem Bayesschen Maß (C.13) und in **B** gemäß dem kontrollenbezogenen Maß (3.9) berechnet wurden. Die Wirksamkeiten aus **A**, welche die Zuordnungswahrscheinlichkeiten der Töpfchen zur Trefferklasse \mathcal{T}_σ angeben, liegen für die Töpfchen B1, D1 und F1 nahe bei 1,0, für das Töpfchen C3 bei 0,88 und für die übrigen Töpfchen bei Werten nahe 0,0. Töpfchen C3 stellt daher einen Primärtreffer im Sinne dieses Maßes dar. Die kontrollenbezogene Wirksamkeit (**B**) für Töpfchen C3 liegt bei 0,72, sodass die Substanz auch von diesem Maß als Treffer eingestuft wird. Auch den Testsubstanzen aus weiteren Töpfchen (beispielsweise C4: 0,31 und C11: 0,34) werden moderate Wirksamkeiten zugeordnet.

Das Verfahren stellt also eine alternative Methode zur Treffererkennung dar und wurde in geringem Umfang beim Screening der DIVERSet-Bibliotheken eingesetzt.

D Molekulardynamik-Simulationen von PrP

Von Konstanze Winklhofer und weiteren Mitarbeitern aus der Gruppe von Jörg Tatzelt (TP4) war das N2a-Zellkulturmodell (vgl. Abschnitt 2.3) dazu eingesetzt worden, den Einfluss von Mutationen des Prion-Protein-Gens auf die Faltung und Maturierung von PrP zu untersuchen [112]. Die Untersuchungen hatten gezeigt, dass die Punktmutationen M205S und M205R, d.h. die Ersetzung des hydrophoben Residuums M205 durch hydrophile Residuen, die Faltung und Maturierung von PrP stören [112]. Die Störungen wurden als Konsequenzen von mutationsinduzierten strukturellen Änderungen von PrP interpretiert, von denen angenommen wurde, dass sie Helix 1 und ihre Anbindung an Helix 3 betreffen, weil das mutierte Residuum M205 sich an der Bindestelle zwischen diesen beiden Helizes befindet (vgl. Abbildung 1.6). Darüber hinaus implizieren aktuelle Modelle für die Struktur von PrP^{Sc}, wie das in Abbildung 1.7 gezeigte, dass bei der Umfaltung von PrP^C nach PrP^{Sc} Helix 1 ihre Struktur verliert. Im Rahmen eines daraufhin von mir initiierten Projekts wurden wildtypisches PrP sowie die beiden Mutanten PrP-M205S und PrP-M205R mithilfe von Molekulardynamik (MD)-Simulationen untersucht. Mein Kollege Bernhard Schropp führte damals als studentische Hilfskraft diese Simulationen durch. Die Ergebnisse dieses Beitrags zur Grundlagenforschung der Prionkrankheiten habe ich gemeinsam mit M. Stork, B. Schropp und P. Tavan im *Biophysical Journal* publiziert [113].

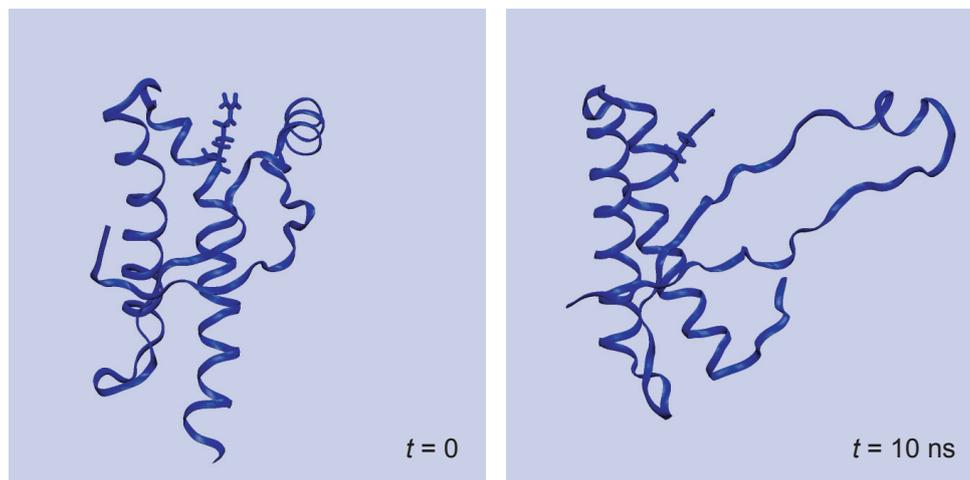


Abbildung D.1: Umgebungsabhängige Faltung von Helix 1 in PrP-M205R. Gezeigt sind zwei strukturelle Schnappschüsse der Mutante PrP-M205R zu Beginn ($t = 0$) der MD-Trajektorie und nach $t = 10$ ns. Das Proteinrückgrat ist jeweils als Band, das mutierte Residuum R205 in einer sogenannten Licorice (Lakritze)-Repräsentation dargestellt.

Die MD-Simulationen der PrP-Moleküle in wässriger Lösung zeigten, dass die native Struktur von PrP^C innerhalb weniger Nanosekunden stark deformiert wird, sobald die Punktmutationen M205S und M205R durchgeführt wurden. Im Fall von M205R ist die Deformation durch eine Bewegung von Helix 1 weg vom hydrophoben Kern des Proteins in die wässrige Umgebung und durch einen anschließenden strukturellen Zerfall charakterisiert. Abbildung D.1 veranschaulicht diesen Zerfall von Helix 1 anhand von zwei Schnappschüssen von PrP-M205R zum Beginn ($t = 0$) und zum Ende der Simulation nach $t = 10$ ns.

Der Zerfall legt nahe, dass die hydrophobe Anbindung von Helix 1 an Helix 3 bei dem Residuum M205 für die korrekte Faltung von Helix 1 in ihre stabile native Struktur notwendig ist. Das Residuum M205 erfüllt dadurch die Funktion einer intramolekularen Chaperone für die Faltung von Helix 1. Zusammen mit CD- und NMR-spektroskopischen Untersuchungen von Helix 1-Modellpeptiden in verschiedenen Lösungsmitteln [182, 183] legt der in der Simulation von PrP-M205R beobachtete Zerfall von Helix 1 nahe, dass die Stabilität von Helix 1 durch die Dielektrizitätskonstante der Umgebung gesteuert wird. Solange Helix 1 an den hydrophoben Kern des Proteins angelagert ist, kann sie wegen der dort erniedrigten Dielektrizitätskonstante in eine stabile helikale Struktur falten. In wässriger Umgebung ist die Helix jedoch instabil. Der Zerfall von Helix 1 in hydrophiler Umgebung unterstützt jene Strukturmodelle für PrP^{Sc}, die eine Umfaltung von Helix 1 postulieren.

Literaturverzeichnis

- [1] R. G. Will, J. W. Ironside, M. Zeidler, S. N. Cousens, K. Estibeiro, A. Alperovitch, S. Poser, M. Pocchiari, A. Hofman und P. G. Smith, *A new variant of Creutzfeldt-Jakob disease in the UK*, *Lancet*, 347, 921–925 (1996). 1
- [2] M. E. Bruce, R. G. Will, J. W. Ironside, I. McConnell, D. Drummond, A. Suttie, L. McCardsle, A. Chree, J. Hope, C. Birkett et al., *Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent*, *Nature*, 389, 498–501 (1997). 1
- [3] *The National Creutzfeldt-Jakob Disease Surveillance Unit: CJD Statistics (2006)*.
<http://www.cjd.ed.ac.uk/figures.htm> 1
- [4] *Bundesministerium für Bildung und Forschung: Förderschwerpunkt zur Erforschung von Therapiemöglichkeiten für menschliche TSE-Erkrankungen*.
<http://www.gesundheitsforschung-bmbf.de/de/396.php> 1, 17
- [5] S. B. Prusiner, *Prions*, *Proc. Natl. Acad. Sci. USA*, 95, 13363–13383 (1998). 1, 2, 3, 4
- [6] W. Hadlow, *Scrapie and kuru*, *Lancet*, ii, 289–290 (1959). 3
- [7] I. Klatzo, D. C. Gajdusek und V. Zigas, *Pathology of Kuru*, *Lab. Invest.*, 8, 799–847 (1959). 3
- [8] D. C. Gajdusek, C. J. Gibbs und M. Alpers, *Experimental transmission of a Kuru-like syndrome to chimpanzees*, *Nature*, 209, 794–796 (1966). 3
- [9] C. J. Gibbs, D. C. Gajdusek, D. M. Asher, M. P. Alpers, E. Beck, P. M. Daniel und W. B. Matthews, *Creutzfeldt-Jakob disease (spongiform encephalopathy): transmission to the chimpanzee*, *Science*, 161, 388–389 (1968). 3
- [10] I. Pattison und K. Jones, *The possible nature of the transmissible agent of scrapie*, *Vet. Rec.*, 80, 2–9 (1967). 3
- [11] J. S. Griffith, *Self-replication and scrapie*, *Nature*, 215, 1043–1044 (1967). 3
- [12] D. C. Bolton, M. P. McKinley und S. B. Prusiner, *Identification of a protein that purifies with the scrapie prion*, *Science*, 218, 1309–1311 (1982). 3
- [13] S. B. Prusiner, *Novel proteinaceous infectious particles cause scrapie*, *Science*, 216, 136–144 (1982). 3
- [14] S. B. Prusiner, D. C. Bolton, D. F. Groth, K. A. Bowman, S. P. Cochran und M. P. McKinley, *Further Purification and Characterization of Scrapie Prions*, *Biochemistry*, 21, 6942–6950 (1982). 3

- [15] A. Giese und H. A. Kretzschmar, *Prion-induced neuronal damage – the mechanisms of neuronal destruction in the subacute spongiform encephalopathies*, *Curr. Top. Microbiol. Immunol.*, 253, 203–217 (2001). 3
- [16] H. Büeler, A. Aguzzi, A. Sailer, R. A. Greiner, P. Autenried, M. Aguet und C. Weissmann, *Mice Devoid of PrP Are Resistant to Scrapie*, *Cell*, 73, 1339–1347 (1993). 3
- [17] S. B. Prusiner, D. Groth, A. Serban, R. Koehler, D. Foster, M. Torchia, D. Burton, S. L. Yang und S. J. DeArmond, *Ablation of the prion protein (PrP) gene in mice prevents scrapie and facilitates production of anti-PrP antibodies*, *Proc. Natl. Acad. Sci. USA*, 90, 10608–10612 (1993). 3
- [18] A. Sailer, H. Büeler, M. Fischer, A. Aguzzi und C. Weissmann, *No Propagation of Prions in Mice Devoid of PrP*, *Cell*, 77, 967–968 (1994). 3
- [19] N. Meyer, V. Rosenbaum, B. Schmidt, K. Gilles, C. Mirenda, D. Groth, S. B. Prusiner und D. Riesner, *Search for a putative scrapie genome in purified prion fractions reveals a paucity of nucleic acids*, *J. Gen. Virol.*, 72, 37–49 (1991). 4
- [20] J. G. Safar, K. Kellings, A. Serban, D. Groth, J. E. Cleaver, S. B. Prusiner und D. Riesner, *Search for a Prion-Specific Nucleic Acid*, *J. Virol.*, 79, 10796–10806 (2005). 4
- [21] F. E. Cohen, K. M. Pan, Z. Huang, M. Baldwin, R. J. Fletterick und S. B. Prusiner, *Structural clues to prion replication*, *Science*, 264, 530–531 (1994). 5
- [22] J. T. Jarrett und P. T. Lansbury, *Seeding 'one-dimensional crystallization' of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie?*, *Cell*, 73, 1055–1058 (1993). 5
- [23] P. T. Lansbury, *Mechanism of scrapie replication*, *Science*, 265, 1510 (1994). 5
- [24] M. Eigen, *Prionics or The kinetic basis of prion diseases*, *Biophys. Chem.*, 63, A1–A18 (1996). 5
- [25] M. Eigen, *BSE und das Prionen-Problem*, *Spektrum der Wissenschaft*, 4 / 2001, 40–49 (2001). 5, 6, 7
- [26] P. Tavan, *Vorlesung: Theoretische Grundlagen der molekularen Biophysik (WS 2005/2006)*, Ludwig-Maximilians-Universität München. 5, 6
- [27] M. Stork, A. Giese, H. A. Kretzschmar und P. Tavan, *Molecular Dynamics Simulations Indicate a Possible Role of Parallel beta-Helices in Seeded Aggregation of Poly-Gln*, *Biophys. J.*, 88, 2442–2451 (2005). 5, 11, 12, 93, 117, 118, 119
- [28] J. Masel, V. A. Jansen und M. A. Nowak, *Quantifying the kinetic parameters of prion replication*, *Biophys. Chem.*, 77, 139–152 (1999). 6
- [29] J. Masel und V. A. Jansen, *The measured level of prion infectivity varies in a predictable way according to the aggregation state of the infectious agent*, *Biochim. Biophys. Acta*, 1535, 164–173 (2001). 6
- [30] L. Manuelidis, *Transmissible Encephalopathies: Speculations and Realities*, *Viral Immunol.*, 16, 123–139 (2003). 7

- [31] M. Eiden, A. Buschmann, L. Kupfer und M. H. Groschup, *Synthetic Prions*, J. Vet. Med. B Infect. Dis. Vet. Public Health, 53, 251–256 (2006). 8
- [32] D. A. Kocisko, J. H. Come, S. A. Priola, B. Chesebro, G. J. Raymond, P. T. Lansbury und B. Caughey, *Cell-free formation of protease-resistant prion protein*, Nature, 370, 471–474 (1994). 8, 21
- [33] G. Legname, I. V. Baskakov, H.-O. B. Nguyen, D. Riesner, F. E. Cohen, S. J. DeArmond und S. B. Prusiner, *Synthetic Mammalian Prions*, Science, 305, 673–676 (2004). 8
- [34] G. P. Saborio, B. Permanne und C. Soto, *Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding*, Nature, 411, 810–813 (2001). 8
- [35] J. Bieschke, P. Weber, N. Sarafoff, M. Beekes, A. Giese und H. Kretzschmar, *Autocatalytic self-propagation of misfolded prion protein*, Proc. Natl. Acad. Sci. USA, 101, 12207–12211 (2004). 8
- [36] P. Weber, A. Giese, N. Piening, G. Mitteregger, A. Thomzig, M. Beekes und H. A. Kretzschmar, *Cell-free formation of misfolded prion protein with authentic prion infectivity*, Proc. Natl. Acad. Sci. USA, 103, 15818–15823 (2006). 8
- [37] G. C. Telling, M. Scott, J. Mastrianni, R. Gabizon, M. Torchia, F. E. Cohen, S. J. DeArmond und S. B. Prusiner, *Prion Propagation in Mice Expressing Human and Chimeric PrP Transgenes Implicates the Interaction of Cellular PrP with Another Protein*, Cell, 83, 79–90 (1995). 8
- [38] F. Wopfner, G. Weidenhofer, R. Schneider, A. von Brunn, S. Gilch, T. F. Schwarz, T. Werner und M. Schatzl, *Analysis of 27 Mammalian and 9 Avian PrPs Reveals High Conservation of Flexible Regions of the Prion Protein*, J. Mol. Biol., 289, 1163–1178 (1999). 9
- [39] J. Herms, T. Tings, S. Gall, A. Madlung, A. Giese, H. Siebert, P. Schurmann, O. Windl, N. Brose und H. Kretzschmar, *Evidence of Presynaptic Location and Function of the Prion Protein*, J. Neurosci., 19, 8866–8875 (1999). 9
- [40] N. Vassallo und J. Herms, *Cellular prion protein function in copper homeostasis and redox signalling at the synapse*, J. Neurochem., 86, 538–544 (2003). 9, 13
- [41] D. R. Brown, K. F. Qin, J. W. Herms, A. Madlung, J. Manson, R. Strome, P. E. Fraser, T. Kruck, A. vonBohlen, W. SchulzSchaeffer et al., *The cellular prion protein binds copper in vivo*, Nature, 390, 684–687 (1997). 9, 10
- [42] M. Mentler, A. Weiss, K. Grantner, P. del Pino, D. Deluca, S. Fiori, C. Renner, W. M. Klaucke, L. Moroder, U. Bertsch et al., *A new method to determine the structure of the metal environment in metalloproteins: investigation of the prion protein octapeptide repeat Cu(2+) complex*, Eur. Biophys. J., 34, 97–112 (2005). 9, 10
- [43] R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber und K. Wuthrich, *NMR structure of the mouse prion protein domain PrP(121–231)*, Nature, 382, 180–182 (1996). 9
- [44] R. Zahn, A. Z. Liu, T. Luhrs, R. Riek, C. von Schroetter, F. L. Garcia, M. Billeter, L. Calzolari, G. Wider und K. Wuthrich, *NMR solution structure of the human prion protein*, Proc. Natl. Acad. Sci. USA, 97, 145–150 (2000). 10

- [45] T. L. James, H. Liu, N. B. Ulyanov, S. Farr-Jones, H. Zhang, D. G. Donne, K. Kaneko, D. Groth, I. Mehlhorn, S. B. Prusiner et al., *Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform*, Proc. Natl. Acad. Sci. USA, 94, 10086–10091 (1997). 10
- [46] R. Glockshuber, S. Hornemann, R. Riek, G. Wider, M. Billeter und K. Wuthrich, *Three-dimensional NMR structure of a self-folding domain of the prion protein PrP(121-231)*, Trends Biochem. Sci., 22, 241–242 (1997). 10
- [47] R. Riek, S. Hornemann, G. Wider, R. Glockshuber und K. Wuthrich, *NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231)*, FEBS Lett., 413, 282–288 (1997). 10
- [48] D. G. Donne, J. H. Viles, D. Groth, I. Mehlhorn, T. L. James, F. E. Cohen, S. B. Prusiner, P. E. Wright und H. J. Dyson, *Structure of the recombinant full-length hamster prion protein PrP(29-231): The N terminus is highly flexible*, Proc. Natl. Acad. Sci. USA, 94, 13452–13457 (1997). 10
- [49] R. Riek, G. Wider, M. Billeter, S. Hornemann, R. Glockshuber und K. Wuthrich, *Prion protein NMR structure and familial human spongiform encephalopathies*, Proc. Natl. Acad. Sci. USA, 95, 11667–11672 (1998). 10
- [50] H. Liu, S. Farr-Jones, N. B. Ulyanov, M. Llinas, S. Marqusee, D. Groth, F. E. Cohen, S. B. Prusiner und T. L. James, *Solution structure of Syrian hamster prion protein rPrP(90-231)*, Biochemistry, 38, 5362–5377 (1999). 10
- [51] F. L. Garcia, R. Zahn, R. Riek und K. Wuthrich, *NMR structure of the bovine prion protein*, Proc. Natl. Acad. Sci. USA, 97, 8334–8339 (2000). 10
- [52] D. R. Perez und K. Wuthrich, *NMR assignment of the Xenopus laevis prion protein fragment xlPrP(98-226)*, J. Biomol. NMR, 31, 260–260 (2005). 10
- [53] D. A. Lysek, C. Schorn, L. G. Nivon, V. Esteve-Moya, B. Christen, L. Calzolari, C. von Schroetter, F. Fiorito, T. Herrmann, P. Guntert et al., *Prion protein NMR structures of cats, dogs, pigs, and sheep*, Proc. Natl. Acad. Sci. USA, 102, 640–645 (2005). 10
- [54] L. Calzolari, D. A. Lysek, D. R. Pérez, P. Güntert und K. Wüthrich, *Prion protein NMR structures of chickens, turtles, and frogs*, Proc. Natl. Acad. Sci. USA, 102, 651–655 (2005). 10
- [55] A. D. Gossert, S. Bonjour, D. A. Lysek, F. Fiorito und K. Wüthrich, *Prion protein NMR structures of elk and of mouse/elk hybrids*, Proc. Natl. Acad. Sci. USA, 102, 646–650 (2005). 10
- [56] D. C. Bolton, R. K. Meyer und S. B. Prusiner, *Scrapie PrP 27-30 Is a Sialoglycoprotein*, J. Virol., 53, 596–606 (1985). 10
- [57] E. Stimson, J. Hope, A. Chong und A. L. Burlingame, *Site-Specific Characterization of the N-Linked Glycans of Murine Prion Protein by High-Performance Liquid Chromatography/Electrospray Mass Spectrometry and Exoglycosidase Digestions*, Biochemistry, 38, 4885–4895 (1999). 10

- [58] N. Stahl, D. R. Borchelt, K. Hsiao und S. B. Prusiner, *Scrapie Prion Protein Contains a Phosphatidylinositol Glycolipid*, *Cell*, 51, 229–240 (1987). 10
- [59] S. Hornemann, C. Schorn und K. Wüthrich, *NMR structure of the bovine prion protein isolated from healthy calf brains*, *EMBO Rep.*, 5, 1159–1164 (2004). 10
- [60] B. W. Caughey, A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes und W. S. Caughey, *Secondary Structure-Analysis of the Scrapie-Associated Protein PrP 27-30 in Water by Infrared Spectroscopy*, *Biochemistry*, 30, 7672–7680 (1991). 11
- [61] K. M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. W. Huang, R. J. Fletterick, F. E. Cohen et al., *Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins*, *Proc. Natl. Acad. Sci. USA*, 90, 10962–10966 (1993). 11
- [62] J. T. Nguyen, H. Inouye, M. A. Baldwin, R. J. Fletterick, F. E. Cohen, S. B. Prusiner und D. A. Kirschner, *X-ray Diffraction of Scrapie Prion Rods and PrP Peptides*, *J. Mol. Biol.*, 252, 412–422 (1995). 11
- [63] H. Wille, M. D. Michelitsch, V. Guenebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard und S. B. Prusiner, *Structural studies of the scrapie prion protein by electron crystallography*, *Proc. Natl. Acad. Sci. USA*, 99, 3563–3568 (2002). 11, 13
- [64] C. Govaerts, H. Wille, S. B. Prusiner und F. E. Cohen, *Evidence for assembly of prions with left-handed beta-helices into trimers*, *Proc. Natl. Acad. Sci. USA*, 101, 8342–8347 (2004). 11, 13, 119
- [65] C. R. Raetz und S. L. Roderick, *A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase*, *Science*, 270, 997–1000 (1995). 11
- [66] S. B. Prusiner (Herausgeber), *Prion Biology and Diseases*, Cold Spring Harbor Laboratory Press (2004). 13
- [67] I. Jones, *Prions show their metal*.
http://www.chemsoc.org/chembytes/ezone/2002/jones_apr02.htm
14
- [68] X. Roucou und A. C. LeBlanc, *Cellular prion protein neuroprotective function: implications in prion diseases*, *J. Mol. Med.*, 83, 3–11 (2005). 13
- [69] B. Caughey und R. E. Race, *Potent inhibition of scrapie associated PrP accumulation by Congo red*, *J. Neurochem.*, 59, 768–771 (1992). 15
- [70] L. Ingrosso, A. Ladogana und M. Pocchiari, *Congo red prolongs the incubation period in scrapie-infected hamsters*, *J. Virol.*, 69, 506–508 (1995). 15
- [71] W. S. Caughey, L. D. Raymond, M. Horiuchi und B. Caughey, *Inhibition of protease-resistant prion protein formation by porphyrins and phthalocyanines*, *Proc. Natl. Acad. Sci. USA*, 95, 12117–12122 (1998). 15
- [72] S. Priola, B. Caughey und W. Caughey, *Novel therapeutic uses for porphyrins and phthalocyanines in the transmissible spongiform encephalopathies*, *Curr. Opin. Microbiol.*, 2, 563–566 (1999). 15

- [73] S. A. Priola, A. Raines und W. S. Caughey, *Porphyrin and phthalocyanine antiscrapie compounds*, *Science*, 287, 1503–6. (2000). 15
- [74] V. Perrier, A. C. Wallace, K. Kaneko, J. Safar, S. B. Prusiner und F. E. Cohen, *Mimicking dominant negative inhibition of prion replication through structure-based drug design*, *Proc. Natl. Acad. Sci. USA*, 97, 6073–6078 (2000). 15
- [75] K. Winklhofer und J. Tatzelt, *Cationic Lipopolyamines Induce Degradation of PrPSc in Scrapie-Infected Mouse Neuroblastoma Cells*, *Biol. Chem.*, 381, 463–469 (2000). 15, 30, 31
- [76] J. Tatzelt, S. Prusiner und W. Welch, *Chemical chaperones interfere with the formation of scrapie prion protein*, *EMBO J.*, 15, 6363–6373 (1996). 15
- [77] S. Gilch, K. F. Winklhofer, M. H. Groschup, M. Nunziante, R. Lucassen, C. Spielhaup-ter, W. Muranyi, D. Riesner, J. Tatzelt und H. M. Schätzl, *Intracellular re-routing of prion protein prevents propagation of PrP(Sc) and delays onset of prion disease*, *EMBO J.*, 20, 3957–3966 (2001). 15, 34
- [78] K. Doh-Ura, T. Iwaki und B. Caughey, *Lysosomotropic Agents and Cysteine Protease Inhibitors Inhibit Scrapie-Associated Prion Protein Accumulation*, *J. Virol.*, 74, 4894–4897 (2000). 15
- [79] C. Korth, B. C. May, F. E. Cohen und S. B. Prusiner, *Acridine and phenothiazine derivatives as pharmacotherapeutics for prion disease*, *Proc. Natl. Acad. Sci. USA*, 98, 9836–9841 (2001). 15, 75
- [80] B. C. H. May, A. T. Fafarman, S. B. Hong, M. Rogers, L. W. Deady, S. B. Prusiner und F. E. Cohen, *Potent inhibition of scrapie prion replication in cultured cells by bis-acridines*, *Proc. Natl. Acad. Sci. USA*, 100, 3416–3421 (2003). 15
- [81] S. J. Collins, V. Lewis, M. Brazier, A. F. Hill, A. Fletcher und C. L. Masters, *Quinacrine Does Not Prolong Survival in a Murine Creutzfeldt-Jakob Disease Model*, *Ann. Neurol.*, 52, 503–506 (2002). 15
- [82] K. Doh-ura, K. Ishikawa, I. Murakami-Kubo, K. Sasaki, S. Mohri, R. Race und T. Iwaki, *Treatment of Transmissible Spongiform Encephalopathy by Intraventricular Drug Infusion in Animal Models*, *J. Virol.*, 78, 4999–5006 (2004). 15
- [83] Y. Kobayashi, K. Hirata, H. Tanaka und T. Yamada, *Quinacrine administration to a patient with Creutzfeldt-Jakob disease who received a cadaveric dura mater graft – an EEG evaluation*, *Rincho Shinkeigaku*, 43, 403–408 (2003). 15
- [84] M. Nakajima, T. Yamada, T. Kusuhara, H. Furukawa, M. Takahashi, A. Yamauchi und Y. Kataoka, *Results of quinacrine administration to patients with Creutzfeldt-Jakob disease*, *Dement. Geriatr. Cogn. Disord.*, 17, 158–163 (2004). 15
- [85] R. Love, *Old drugs to treat new variant Creutzfeldt-Jakob disease*, *Lancet*, 358, 563–563 (2001). 15
- [86] *Medical Research Council, Clinical Trials Unit: PRION-1: Randomised trial of quinacrine in human prion disease.*
<http://www.ctu.mrc.ac.uk/studies/cjd.asp> 15

- [87] National Institutes of Health, *ClinicalTrials.gov: CJD (Creutzfeldt-Jakob Disease) Quinacrine Study*.
<http://www.clinicaltrials.gov/show/NCT00183092> 15
- [88] V. Gayraud, N. Picard-Hagen, C. Vigié, V. Laroute, O. Andréoletti und P.-L. Toutain, *A possible pharmacological explanation for quinacrine failure to treat prion diseases: pharmacokinetic investigations in a ovine model of scrapie*, *Br. J. Pharmacol.*, 144, 386–393 (2005). 15
- [89] Y. Huang, H. Okochi, B. C. H. May, G. Legname, S. B. Prusiner, L. Z. Benet, B. J. Guglielmo und E. T. Lin, *Quinacrine is mainly metabolized to mono-desethyl quinacrine by CYP3A4/5 and its brain accumulation is limited by P-glycoprotein*, *Drug Metab. Dispos.*, 34, 1136–1144 (2006). 15
- [90] B. C. H. May, J. Witkop, J. Sherrill, M. O. Anderson, P. B. Madrid, J. A. Zorn, S. B. Prusiner, F. E. Cohen und R. K. Guy, *Structure-activity relationship study of 9-aminoacridine compounds in scrapie-infected neuroblastoma cells*, *Bioorg. Med. Chem. Lett.*, 16, 4913–4916 (2006). 15
- [91] U. Bertsch, K. F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber, F. U. Hartl, P. Tavan, J. Tatzelt, H. A. Kretzschmar und A. Giese, *Systematic Identification of Antiprion Drugs by High-Throughput Screening Based on Scanning for Intensely Fluorescent Targets*, *J. Virol.*, 79, 7785–7791 (2005). 18, 25, 26, 29, 30, 31, 32, 36, 95, 96, 97, 99, 100, 102
- [92] U. Bertsch, K. F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber, F. U. Hartl, P. Tavan, J. Tatzelt, H. A. Kretzschmar und A. Giese, *Single Molecule Analysis of Protein Aggregation and Prions by SIFT*, *FEBS J.*, 272, 338–339 (2005). 18, 95
- [93] H. A. Kretzschmar, U. Bertsch, K. F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber, F. U. Hartl, P. Tavan, J. Tatzelt und A. Giese, *Single Molecule Analysis of Protein Aggregation and Prions by SIFT*, *J. Neuropath. Exp. Neur.*, 64, 441–441 (2005). 18, 95
- [94] U. Bertsch, A. Giese, H. Kretzschmar, P. Tavan, T. Hirschberger, J. Bieschke, P. Weber, K. F. Winklhofer, J. Tatzelt, F. U. Hartl et al., *Systematic Identification of New Anti-Prion Drugs by High-Throughput Screening Based on Scanning for Intensely Fluorescent Targets (SIFT)*, PCT/EP2005/005614, WO 2005/116640 A2. 18, 95, 103, 110, 112, 114
- [95] *ChemBridge Corporation, San Diego, CA, USA*.
<http://chembridge.com> 19, 44
- [96] O. F. Güner (Herausgeber), *Pharmacophore Perception, Development, and Use in Drug Design*, International University Line, La Jolla, California (2000). 19
- [97] P. Ehrlich, *Über den jetzigen Stand der Chemotherapie*, *Berichte der Deutschen Chemischen Gesellschaft*, 42, 17–47 (1909). 19
- [98] C. A. Lipinski, F. Lombardo, B. W. Dominy und P. J. Feeney, *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*, *Adv. Drug Deliv. Rev.*, 46, 3–26 (2001). 19, 20, 46
- [99] H. J. Verheij, *Leadlikeness and structural diversity of synthetic screening libraries*, *Mol. Div.*, 10, 377–388 (2006). 20

- [100] G. M. Rishton, *Nonleadlikeness and leadlikeness in biochemical screening*, Drug Discov. Today, 8, 86–96 (2003). 20
- [101] I. Vorberg, A. Buschmann, S. Harmeyer, A. Saalmuller, E. Pfaff und M. H. Groschup, *A novel epitope for the specific detection of exogenous prion proteins in transgenic mice and transfected murine cell lines*, Virology, 255, 26–31 (1999). 21
- [102] P. Schwille, J. Bieschke und F. Oehlenschläger, *Kinetic Investigations by Fluorescence Correlation Spectroscopy: The Analytical and Diagnostic Potential of Diffusion Studies*, Biophys. Chem., 66, 211–228 (1997). 22
- [103] J. Bieschke, A. Giese, W. Schulz-Schaeffer, I. Zerr, S. Poser, M. Eigen und H. Kretzschmar, *Ultrasensitive detection of pathological prion protein aggregates by dual-color scanning for intensely fluorescent targets*, Proc. Natl. Acad. Sci. USA, 97, 5468–5473 (2000). 22
- [104] A. Giese, J. Bieschke, M. Eigen und H. A. Kretzschmar, *Putting prions into focus: application of single molecule detection to the diagnosis of prion diseases*, Arch. Virol., Suppl. 16, 161–171 (2000). 22
- [105] *Evotec Technologies GmbH, Hamburg, Germany.*
<http://www.evotec-technologies.com> 23, 26, 29, 30, 76
- [106] M. Goedert, *Alpha-synuclein and neurodegenerative diseases*, Nat. Rev. Neurosci., 2, 492–501 (2001). 30
- [107] T. S. Ulmer, A. Bax, N. B. Cole und R. L. Nussbaum, *Structure and Dynamics of Micelle-bound Human alpha-Synuclein*, J. Biol. Chem., 280, 9595–9603 (2005). 30
- [108] L. A. Munishkina, C. Phelan, V. N. Uversky und A. L. Fink, *Conformational Behavior and Aggregation of alpha-Synuclein in Organic Solvents: Modeling the Effects of Membranes*, Biochemistry, 42, 2720–2730 (2003). 30
- [109] D. Butler, M. Scott, J. Bockman, D. Borchelt, A. Taraboulos, K. Hsiao, D. Kingsbury und S. Prusiner, *Scrapie-infected murine neuroblastoma cells produce protease-resistant prion proteins*, J. Virol., 62, 1558–1564 (1988). 31
- [110] D. R. Borchelt, A. Taraboulos und S. B. Prusiner, *Evidence for synthesis of scrapie prion proteins in the endocytic pathway*, J. Biol. Chem., 267, 16188–16199 (1992). 31
- [111] B. Caughey und G. J. Raymond, *The scrapie-associated form of PrP is made from a cell surface precursor that is both protease and phospholipase-sensitive*, J. Biol. Chem., 266, 18217–18223 (1991). 31
- [112] K. F. Winklhofer, J. Heske, U. Heller, A. Reintjes, W. Muranyi, I. Moarefi und J. Tatzelt, *Determinants of the in vivo folding of the prion protein - A bipartite function of helix 1 in folding and aggregation*, J. Biol. Chem., 278, 14961–14970 (2003). 32, 141
- [113] T. Hirschberger, M. Stork, B. Schropp, K. F. Winklhofer, J. Tatzelt und P. Tavan, *Structural instability of the prion protein upon M205S/R mutations revealed by molecular dynamics simulations*, Biophys. J., 90, 3908–3918 (2006). 33, 141

- [114] M. Eiden, G. J. Palm, W. Hinrichs, U. Matthey, R. Zahn und M. H. Groschup, *Synergistic and strain-specific effects of bovine spongiform encephalopathy and scrapie prions in the cell-free conversion of recombinant prion protein*, J. Gen. Virol., 87, 3753–3761 (2006). 34
- [115] I. N. Bronstein, K. A. Semendjajev, G. Musiol und H. Mühlig, *Taschenbuch der Mathematik*, Deutsch, Thun und Frankfurt am Main (1995). 39
- [116] F. E. Grubbs, *Procedures for Detecting Outlying Observations in Samples*, Technometrics, 11, 1–21 (1969). 40, 41
- [117] *Elsevier MDL, San Ramon, CA, USA: CTfile Formats.*
<http://www.mdli.com/downloads/public/ctfile/ctfile.jsp> 44
- [118] A. J. Leo, *Calculating log Poct from structures*, Chem. Rev., 93, 1281–1306 (1993). 46, 47
- [119] P. Ertl, B. Rohde und P. Selzer, *Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties*, J. Med. Chem., 43, 3714–3717 (2000). 46
- [120] *Daylight, Chemical Information Systems, Inc., Aliso Viejo, CA, USA.*
<http://www.daylight.com> 46, 89
- [121] A. Leo, C. Hansch und D. Elkins, *Partition coefficients and their uses*, Chem. Rev., 71, 525–616 (1971). 46
- [122] *Free Software Foundation, Inc.: CVS - Concurrent Versions System.*
<http://www.nongnu.org/cvs> 47
- [123] *Free Software Foundation, Inc.: Gawk.*
<http://www.gnu.org/software/gawk/gawk.html> 49, 84
- [124] *CREASO GmbH, Gilching, Germany.*
<http://www.creaso.de> 50
- [125] W.-D. Ihlenfeldt, *The CACTVS System Home Page.*
<http://www2.chemie.uni-erlangen.de/software/cactvs> 51, 80
- [126] *World Wide Web Consortium (W3C): HyperText Markup Language (HTML) Home Page.*
<http://www.w3.org/MarkUp> 51
- [127] *Ecma International: Standard ECMA-262 - ECMAScript Language Specification.*
<http://www.ecma-international.org/publications/standards/Ecma-262.htm> 53
- [128] *Leibniz-Rechenzentrum München: Oracle9i Datenbankmanagementsystem.*
<http://www.lrz-muenchen.de/services/datenhaltung/datenbanken/oracle9> 55
- [129] *Quest Software: Toad Data Modeler.*
http://www.quest.com/Toad_Data_Modeler 56, 57
- [130] P. P.-S. Chen, *The Entity-Relationship Model – Toward a Unified View of Data*, ACM Trans. Database Syst., 1, 9–36 (1976). 56

- [131] *Daylight Chemical Information Systems, Inc.: SMILES.*
<http://www.daylight.com/smiles> 59
- [132] *Open Babel: The Open Source Chemistry Toolbox.*
<http://openbabel.sourceforge.net> 59
- [133] *CambridgeSoft Corp., Cambridge, MA, USA.*
<http://www.cambridgesoft.com> 59, 71, 73
- [134] *Borland Software Corporation, Cupertino, CA, USA.*
<http://www.borland.com> 60, 129
- [135] *Spotfire, Inc., Somerville, MA, USA.*
<http://www.spotfire.com> 62
- [136] *JFree.org: JFreeChart.*
<http://www.jfree.org/jfreechart> 66, 132
- [137] *The Chemistry Development Kit.*
<http://almost.cubic.uni-koeln.de/cdk> 68, 132
- [138] *Adobe Systems Incorporated: PDF Reference.*
http://partners.adobe.com/public/developer/pdf/index_reference.html 69
- [139] *iText, a Free Java-PDF Library.*
<http://www.lowagie.com/iText> 69, 131
- [140] G. Thomas, *Medicinal Chemistry: An Introduction*, John Wiley & Sons, Ltd (2000). 75
- [141] H. Böhm, G. Klebe und H. Kubinyi, *Wirkstoffdesign - Der Weg zum Arzneimittel*, Spektrum - Akademischer Verlag (2002). 75, 104
- [142] *Accelrys, Inc., San Diego, CA, USA.*
<http://www.accelrys.com> 76
- [143] *Tripos Inc., St. Louis, MO, USA.*
<http://www.tripos.com> 76, 77, 88
- [144] C. Borgelt, *Neuro-Fuzzy Systeme, Institut für Wissens- und Sprachverarbeitung, Otto-von-Guericke-Universität Magdeburg: MoSS - Molecular Substructure Miner.*
<http://fuzzy.cs.uni-magdeburg.de/~borgelt/moss.html> 77
- [145] C. Borgelt und M. Berthold, *Mining Molecular Fragments: Finding Relevant Substructures of Molecules*, in *IEEE International Conference on Data Mining (ICDM)*, 51–58, IEEE Press, Piscataway, NJ, USA 2002, Maebashi, Japan (2002). 77, 78
- [146] H. Hofer, C. Borgelt und M. R. Berthold, *Large Scale Mining of Molecular Fragments with Wildcards*, in *Proc. 5th Int. Symposium on Intelligent Data Analysis (IDA2003, Berlin, Germany)*, 376–385, Springer-Verlag, Heidelberg, Germany (2003). 77
- [147] H. Hofer, C. Borgelt und M. R. Berthold, *Large Scale Mining of Molecular Fragments with Wildcards*, *Intell. Data Anal.*, 8, 495–504 (2004). 77

- [148] C. Borgelt, T. Meinl und M. R. Berthold, *MoSS: A Program for Molecular Substructure Mining*, in *Workshop Open Software for Data Mining (OSDM'05, Chicago, IL)*, 6–15, ACM Press, New York, NY, USA (2005). 77
- [149] C. Borgelt, M. R. Berthold und D. E. Patterson, *Molecular Fragment Mining for Drug Discovery*, in *Proc. 8th European Conf. on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'05, Barcelona, Spain)*, 1002–1013, Springer-Verlag, Heidelberg, Germany (2005). 77
- [150] V. Heiser, S. Engernann, W. Brocker, I. Dunkel, A. Boeddrich, S. Waelter, E. Nordhoff, R. Lurz, N. Schugardt, S. Rautenberg et al., *Identification of benzothiazoles as potential polyglutamine aggregation inhibitors of Huntington's disease by using an automated filter retardation assay*, *Proc. Natl. Acad. Sci. USA*, 99, 16400–16406 (2002). 81
- [151] *Morten's JavaScript Tree Menu*.
<http://www.treemenu.org> 84
- [152] *Leadscope, Inc., Columbus, OH, USA*.
<http://www.leadscope.com> 88
- [153] R. Clark, *OptiSim: An Extended Dissimilarity Selection Method for Finding Diverse Representative Subsets*, *J. Chem. Inf. Comput. Sci.*, 37, 1181–1188 (1997). 88, 90, 91
- [154] R. D. Brown und Y. C. Martin, *Use of structure Activity data to compare structure-based clustering methods and descriptors for use in compound selection*, *J. Chem. Inf. Comput. Sci.*, 36, 572–584 (1996). 88, 89
- [155] R. D. Brown und Y. C. Martin, *The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding*, *J. Chem. Inf. Comput. Sci.*, 37, 1–9 (1997). 88
- [156] *Daylight Theory Manual (2006)*.
<http://www.daylight.com/dayhtml/doc/theory/index.pdf> 89, 90
- [157] T. T. Tanimoto, *IBM Internal Report*, Technischer Bericht, IBM Technical Report. Series, November 17. (1957). 91
- [158] *Triplos: SARNavigator User's Guide Version 1.5 (2004)*. 91
- [159] N. W. Schiffer, S. A. Broadley, T. Hirschberger, P. Tavan, H. A. Kretzschmar, A. Giese, C. Haass, F. U. Hartl und B. Schmid, *Identification of Anti-prion Compounds as Efficient Inhibitors of Polyglutamine Protein Aggregation in a Zebrafish Model*, *J. Biol. Chem.*, 282, 9195–9203 (2007). 93, 115, 116, 118
- [160] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew und A. J. Olson, *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*, *J. Comp. Chem.*, 19, 1639–1662 (1998). 93, 117
- [161] M. J. S. Dewar und W. Thiel, *Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters*, *J. Am. Chem. Soc.*, 99, 4899–4907 (1977). 93
- [162] W. Humphrey, A. Dalke und K. Schulten, *VMD: Visual Molecular Dynamics*, *J. Mol. Graph.*, 14, 33–38 (1996). 94

- [163] *Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim am Rhein, Germany.*
<http://www.boehringer-ingelheim.de> 113
- [164] S. W. Davies, M. Turmaine, B. A. Cozens, M. DiFiglia, A. H. Sharp, C. A. Ross, E. Scherzinger, E. E. Wanker, L. Mangiarini und G. P. Bates, *Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation*, *Cell*, 90, 537–548 (1997). 115
- [165] E. Scherzinger, R. Lurz, M. Turmaine, L. Mangiarini, B. Hollenbach, R. Hasenbank, G. P. Bates, S. W. Davies, H. Lehrach und E. E. Wanker, *Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo*, *Cell*, 90, 549–558 (1997). 115
- [166] M. DiFiglia, E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel und N. Aronin, *Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain*, *Science*, 277, 1990–1993 (1997). 115
- [167] W. E. Klunk, J. W. Pettegrew und D. J. Abraham, *Quantitative-Evaluation of Congo Red Binding to Amyloid-like Proteins with a Beta-pleated Sheet Conformation*, *J. Histochem. Cytochem.*, 37, 1273–1281 (1989). 116
- [168] R. Nelson und D. Eisenberg, *Recent atomic models of amyloid fibril structure*, *Curr. Opin. Struct. Biol.*, 16, 260–265 (2006). 117
- [169] *Deutsches Patent- und Markenamt: DEPATISnet.*
<http://depatisnet.dpma.de> 128
- [170] X. Cai, Sui, S. Kasibhatla, J. Drewe, S. Reddy, P. und H.-Z. Zhang, *Substituted N'-(Arylcarbonyl)-Benzhydrazides, N'-(Arylcarbonyl)-Benzylidene-Hydrazides and Analogs as Activators of Caspases and Inducers of Apoptosis and the Use thereof*, WO/2002/089745. 128
- [171] *Sun Microsystems, Inc.: Java Technology.*
<http://java.sun.com> 129
- [172] *Eclipse - an open development platform.*
<http://www.eclipse.org> 129
- [173] *Oracle Technology Network.*
<http://www.oracle.com/technology> 131
- [174] E. Gamma, R. Helm, R. Johnson und V. John, *Design Patterns. Elements of Reusable Object-Oriented Software*, Addison Wesley (1995). 131
- [175] *Apache Software Foundation: Log4j project.*
<http://logging.apache.org/log4j> 132
- [176] T. Hirschberger, *Hierarchische Klassifikation durch selbstorganisierende neuronale Algorithmen*, Diplomarbeit, Lehrstuhl für BioMolekulare Optik, Fakultät für Physik, LMU München (2002). 133, 137

- [177] M. Kloppenburg und P. Tavan, *Deterministic Annealing for Density Estimation by Multivariate Normal Mixtures*, Phys. Rev. E, 55, 2089–2092 (1997). 133, 136
- [178] S. Albrecht, J. Busch, M. Kloppenburg, F. Metze und P. Tavan, *Generalized Radial Basis Function Networks for Classification and Novelty Detection: Self-Organization of Optimal Bayesian Decision*, Neural Networks, 13, 1075–1093 (2000). 133
- [179] F. Metze, *Integration neuronaler Methoden zur Phonemerkennung aus fließend gesprochener Sprache*, Diplomarbeit, Institut für Medizinische Optik, Fakultät für Physik, LMU München (1998). 134
- [180] A. Wörl, *Gehörorientierte Vorverarbeitung von Sprachsignalen für spracherkennende Systeme*, Diplomarbeit, Institut für Medizinische Optik, Fakultät für Physik, LMU München (2000). 134
- [181] K. Pearson, *On Lines and Planes of Closest Fit to Systems of Points in Space*, Phil. Mag., 2, 559–572 (1901). 135
- [182] S. A. Kozin, G. Bertho, A. K. Mazur, H. Rabesona, J. P. Girault, T. Haertle, M. Takahashi, P. Debey und G. H. B. Hoa, *Sheep prion protein synthetic peptide spanning helix 1 and beta-strand 2 (residues 142-166) shows beta-hairpin structure in solution*, J. Biol. Chem., 276, 46364–46370 (2001). 142
- [183] S. Megy, G. Bertho, S. A. Kozin, P. Debey, G. H. B. Hoa und J. P. Girault, *Possible role of region 152-156 in the structural duality of a peptide fragment from sheep prion protein*, Protein Sci., 13, 3151–3160 (2004). 142

Danksagung

Allen, die zum Gelingen meiner Doktorarbeit beigetragen haben, möchte ich an dieser Stelle herzlich danken.

Ganz besonders danke ich meinem Doktorvater Prof. Paul Tavan für die engagierte Betreuung meiner Arbeit. Ich danke ihm für das Vertrauen, das er mir schenkte, als er mir dieses reiz- und anspruchsvolle Projekt anbot, für die Freiheiten, die er mir bei der Gestaltung meiner Arbeit ließ, und für die große Unterstützung, die er mir, wo immer nötig, gab. Sein hoher Anspruch an die Qualität wissenschaftlichen Arbeitens und Schreibens ist mir zum Leitbild geworden. Für seinen persönlichen Einsatz und für die großzügige Finanzierung meiner Doktorarbeit im Rahmen des BMBF-Projektes (01 KO 0501) danke ich ihm sehr.

Meinen Partnern aus den Teilprojekten des Verbundes, Dr. Uwe Bertsch, Dr. Armin Giese und Prof. Hans Kretzschmar (TP1), Dr. Martin Eiden, Dr. Markus Geissen und Prof. Martin Groschup (TP3), Dr. Konstanze Winklhofer und Prof. Jörg Tatzelt (TP4) und Prof. Boris Schmidt (TP6) danke ich herzlich für die Aufnahme in die „Prion-Community“, für die intensive Zusammenarbeit und für viele interessante Gespräche und Diskussionen, die wir alltäglich, bei den regelmäßigen Treffen des Verbundes oder auf gemeinsam besuchten Konferenzen führten. Frau Dr. Rosi Lederer (ZNP) gilt mein Dank für die Organisation der Treffen und vieler weiterer Angelegenheiten des Verbundes.

Meinen Kollegen in der Arbeitsgruppe Tavan danke ich für die kollegiale Arbeitsatmosphäre, die von großem Interesse an der Arbeit des Anderen, von Hilfsbereitschaft und von gegenseitiger Unterstützung geprägt ist. Dieser Dank gebührt auch allen weiteren Angehörigen des BMO. Im einzelnen gilt mein Dank

- Verena Schultheis und Dr. Matthias Schmitz, mit denen ich das Zimmer teil(t)e, für die angenehme Atmosphäre, für viele hilfreiche Gespräche und Diskussionen sowie für die netten Ablenkungen zwischendurch,
- Martina Stork und Andreas Weiss für ihre Hilfe bei meinem Einstieg in das Gebiet der Prionkrankheiten und für die fruchtbare Zusammenarbeit bei verschiedenen Projekten auf diesem Gebiet,
- Simon Youssef für seine Unterstützung als studentische Hilfskraft bei diversen umfangreichen Programmieraufgaben,
- Bernhard Schropp, Martina Stork und Paul Tavan für ihre Beiträge zu unserem gemeinsamen Paper,

- Robert Denschlag und Martin Lingenheil dafür, dass sie mich an ihren teils hitzigen, stets interessanten Diskussionen teilnehmen ließen,
- Galina Babizki für die Wartung und den Betrieb der Drucker,
- Benjamin Rieff und Sebastian Bauer für ihren geistreichen Humor und für die Leichtigkeit des Seins, die sie neben ihren wissenschaftlichen Leistungen unserer Arbeitsgruppe zu Teil werden lassen,
- Rudolf Reichold, Dr. Arne Sieg und Dr. Karl-Heinz Mantel für die gute Zusammenarbeit bei der Systemverwaltung,
- den Sekretärinnen Frau Michaelis, Frau Widmann-Diermeier und Frau Podolski für ihre unkomplizierte Hilfsbereitschaft bei den vielen organisatorischen Kleinigkeiten des Alltags,
- der Werkstatt um Rudi Schwarz sowie Martina Stork für ihre Hilfe beim Anfertigen von Modellen des Prion-Proteins, die bei öffentlichen Veranstaltungen ausgestellt wurden,
- sowie Sebastian Bauer und Verena Schultheis für das sorgfältige Korrekturlesen meiner Dissertation.

Ausdrücklich gilt mein Dank für das angenehme Arbeitsumfeld auch den ehemaligen Mitgliedern unserer Arbeitsgruppe Drs. Heiko Carstens, Bernhard Egwolf, Gerald Mathias, Paul Strodel und Prof. Marco Nonella.

Meinen Freunden außerhalb des BMO bin ich dankbar für die Erinnerung an das Leben da draußen. Meinen Eltern, meinem Bruder und meiner näheren Familie danke ich von Herzen für ihren Glauben an mich sowie für die Fürsorge und Unterstützung, die sie mir geben seit ich sie kenne. Die stärkste persönliche Unterstützung während meiner Promotion gab mir meine Freundin Natalie Oliphant mit ihrer Liebe und Geduld. Sie nahm an meiner Arbeit Anteil, gab mir in schwierigen Momenten Kraft und schenkte mir erholsame Wochenenden. Dafür, dass sie immer für mich da ist, danke ich ihr von Herzen.