

**Entwicklung neuer Deskriptoren
für vergleichende molekulare Feldanalysen
und
Verfahren zur Beschreibung
der Selektivität von Serinproteasen**

**Dissertation
zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)**

dem
Fachbereich Pharmazie
der Philipps-Universität Marburg
vorgelegt von

Markus Böhm
aus Auerbach / Opf.

Marburg / Lahn 2002

Vom Fachbereich Pharmazie der Philipps-Universität Marburg
als Dissertation angenommen am:

13. Februar 2002

Erstgutachter:

Prof. Dr. G. KLEBE

Zweitgutachter:

Prof. Dr. M. SCHLITZER

Tag der mündlichen Prüfung:

13. Februar 2002

Die vorliegende Arbeit wurde auf Anregung von Herrn Prof. Dr. G. KLEBE am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Mai 1998 bis November 2001 durchgeführt.

Mein herzlicher Dank gilt:

- Herrn Prof. Dr. G. KLEBE für die interessante Aufgabenstellung, die sehr freundliche Unterstützung bei der Durchführung dieser Arbeit sowie für die Möglichkeit, das Gebiet des rationalen Wirkstoffdesigns unter verschiedenen Gesichtspunkten kennenzulernen. Die Kombination aus Anleitung und Freiheit zu eigenständiger Arbeit stellte eine angenehme Form der Betreuung dar.
- Herrn Prof. Dr. M. SCHLITZER für die freundliche Betreuung bei der strukturbasierten Entwicklung von Hemmstoffen der Farnesyltransferase. Die damit verbundene Gelegenheit einer interdisziplinären Zusammenarbeit mit seiner Gruppe war eine lohnende Bereicherung der wissenschaftlichen Arbeit.
- Herrn A. WEBER für die zahlreichen Diskussionen, die vielen erfolgreichen Stunden gemeinsam vor dem Rechner sowie für das Korrekturlesen.
- Herrn D. KUHN für seine Hilfestellung bei der Ähnlichkeitssuche von Bindetaschen der Farnesyltransferase sowie für viele anregende Gespräche.
- Herrn Dr. M. T. STUBBS für die hilfreichen Ratschläge zum kristallographischen und biologischen Hintergrund der Arbeit.
- Herrn Dr. H. GOHLKE und Herrn Dr. W. NISSINK für viele anregende Gespräche und für die Bereitstellung einiger hilfreicher Programme.
- Herrn H.-D. GERBER für das Korrekturlesen und für die freundliche, zeitweilige Bereitstellung einer häuslichen Unterkunft.
- Herrn T. MIETZNER für seine beständige Auskunftsbereitschaft bei mathematischen und programmtechnischen Fragen.
- Herrn Dr. J. STÜRZEBECKER für die Bereitstellung der Strukturen und biologischen Daten der in dieser Arbeit verwendeten Serinprotease-Inhibitoren.
- allen Mitgliedern der Arbeitsgruppe für eine angenehme und freundliche Arbeitsatmosphäre sowie für viele interessante und fachliche Diskussionen.

Teile dieser Arbeit wurden bereits veröffentlicht:

i) Tagungsbeiträge

Böhm, M. & Klebe, G. (2001), 3D QSAR and Chemometrical Analyses to Study Selectivity Features of Serine Proteases, Vortrag präsentiert beim 15. Darmstädter Molecular-Modelling-Workshop, Darmstadt, 22. – 23. Mai 2001.

Böhm, M. & Schlitzer, M. (2000), 3D QSAR Studies of benzophenone-based farnesyltransferase inhibitors, Poster präsentiert beim Schliersee-Symposium „Oncology: New Aspects“ der Gesellschaft Deutscher Chemiker, Schliersee, 18. – 20. Oktober 2000.

Böhm, M. & Schlitzer, M. (2000), 3D-QSAR Untersuchungen von Benzophenon Farnesyltransferase Inhibitoren, Poster präsentiert bei der Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Münster, 05. – 07. Oktober 2000, publiziert in: *Arch. Pharm. Pharm. Med. Chem.* **333**, Suppl. 2, 39.

Böhm, M. & Klebe, G. (2000), Development of new H-Bond descriptors and their application in 3D QSAR Analyses, Poster präsentiert beim 13th European Symposium on Quantitative Structure-Activity Relationships, Düsseldorf, 27. August – 01. September 2000.

Böhm, M. & Klebe, G. (2000), Development of new H-bond descriptors and their application in 3D QSAR analyses, Poster präsentiert beim 14. Darmstädter Molecular-Modelling-Workshop, Darmstadt, 30. – 31. Mai 2000.

Böhm, M. & Schlitzer, M. (1999), Nicht-Peptidische, Nicht-Prenylische Bisubstratanaloga Inhibitoren der Farnesyltransferase: Docking Studien, Poster präsentiert bei der Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Frankfurt, 06. – 09. Oktober 1999, publiziert in: *Arch. Pharm. Pharm. Med. Chem.* **332**, Suppl. 2, 5.

ii) Aufsätze

Böhm, M. & Klebe, G. (2002). 3D QSAR and chemometrical analyses to study selectivity features of serine proteases, in Vorbereitung.

Böhm, M. & Klebe, G. (2002). Development of new hydrogen-bond descriptors and their application to comparative molecular field analyses. *J. Med. Chem.*, im Druck.

Böhm, M., Mitsch, A., Sattler, I. & Schlitzer, M. (2001). Exploration of novel aryl binding sites of farnesyltransferase using molecular modeling and benzophenone-based farnesyltransferase inhibitors. *J. Med. Chem.* **44**, 3117-3124.

Sakowski, J., Böhm, M., Sattler, I., Dahse, H. M. & Schlitzer, M. (2001). Synthesis, molecular modeling, and structure-activity relationship of benzophenone-based CAAX-peptidomimetic farnesyltransferase inhibitors. *J. Med. Chem.* **44**, 2886-2899.

Schlitzer, M., Böhm, M. & Sattler, I. (2000). Non-peptidic, non-prenylic bisubstrate farnesyltransferase inhibitors. Part 3: structural requirements of the central moiety for farnesyltransferase inhibitory activity. *Bioorg. Med. Chem.* **8**, 2399-2406.

Schlitzer, M., Böhm, M., Sattler, I. & Dahse, H. M. (2000). Design, synthesis and early structure-activity relationship of farnesyltransferase inhibitors which mimic both the peptidic and the prenylic substrate. *Bioorg. Med. Chem.* **8**, 1991-2006.

Böhm, M., Stürzebecher, J. & Klebe, G. (1999). Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **42**, 458-477.

Klebe, G., Böhm, M., Dullweber, F., Grädler, U., Gohlke, H. & Hendlich, M. (1999). Structural and Energetic Aspects of Protein-Ligand Binding in Drug Design. In *Molecular Modelling and Prediction of Bioactivity* (Gundertofte, K. & Jorgensen, F., eds.), pp. 103-110, KLUWER Academic / Plenum Publ., New York.

INHALTSVERZEICHNIS

1	Entwicklung wissensbasierter Deskriptoren.....	1
1.1	Problemstellung.....	1
1.2	Ansätze aus der Literatur.....	5
1.3	Theorie und Methoden	7
1.3.1	Deskriptoren zur Charakterisierung kleiner Moleküle	7
1.3.2	Validierung der Deskriptoren durch vergleichende Feldanalysen nach dem CoMSIA-Verfahren.....	14
1.3.3	Deskriptoren zur Beschreibung von Wasserstoffbrücken-Eigenschaften	19
1.3.4	Approximation durch anisotrope Gaußfunktionen	20
1.3.5	Approximation durch isotrope Gaußfunktionen.....	22
1.4	Ergebnisse und Diskussion.....	30
1.4.1	Statistische Auswertung der QSAR-Modelle	30
1.4.2	Interkorrelation einzelner Feldbeiträge	31
1.4.3	Charakterisierung von Wasserstoffbrücken-Eigenschaften.....	34
1.4.4	Approximation der Felder aus SuperStar	36
1.4.5	Graphische Auswertung der Konturdiagramme	40
1.5	Zusammenfassung	46
2	Selektivitätsuntersuchungen am Beispiel von Serinproteasen	48
2.1	Problemstellung.....	48
2.2	Ansätze aus der Literatur.....	49
2.3	Untersuchung von Liganden mittels 3D-QSAR Verfahren.....	58
2.3.1	Aufbau des Ligand-Datensatzes	58
2.3.2	Strukturelles Alignment der Inhibitoren.....	64
2.3.3	Generierung separater Einzelmodelle.....	70
2.3.4	Selektivitätsmodelle durch Differenzbildung von Affinitäten	80
2.3.5	Selektivitätsmodelle aus Kombination von Einzelmodellen	95
2.4	Untersuchung von physikochemischen Eigenschaften in Proteinbindetaschen durch chemometrische Auswertungen	98
2.4.1	Aufbau des Protein-Datensatzes	98
2.4.2	Physikochemische Charakterisierung der Proteinbindetaschen	100

2.4.3	Chemometrische Analyse der molekularen Feldbeiträge	106
2.4.4	Selektivitätsmodell durch differenzielle Hauptkomponentenanalyse	114
2.4.5	Selektivitätsmodell durch Differenzen von Wechselwirkungsfeldern	117
2.4.6	Analyse diskriminierender Faktoren zwischen Trypsin und Trypsin-Mutanten	119
2.5	Zusammenfassung	120
3	Struktur-basiertes Design von Hemmstoffen der Farnesyltransferase	123
3.1	Einleitung	123
3.1.1	Der Ras-Signaltransduktionsweg	124
3.1.2	Die Familie der Ras-Proteine	126
3.1.3	Posttranslationale Modifikationen	129
3.2	Die Farnesyltransferase	131
3.2.1	Struktur und Funktion der Farnesyltransferase	132
3.2.2	Farnesyltransferase-Inhibitoren	137
3.2.3	Wirkmechanismus von Farnesyltransferase-Inhibitoren	140
3.3	Rationales Design von Farnesyltransferase-Inhibitoren	142
3.3.1	Benzophenon-basierte Farnesyltransferase-Inhibitoren	143
3.3.2	Thiolfreie Farnesyltransferase-Inhibitoren	145
3.3.3	Flexible Docking-Studien und GRID-Analysen	146
3.3.4	Analyse der postulierten Arylbindestellen	152
3.4	Zusammenfassung	158
4	Anhang	160
4.1	Verwendete Hardware	160
4.2	Verwendete Programme	160
4.3	Statistische Parameter der PLS-Analyse	161
4.3.1	Kreuzvalidierung	161
4.3.2	Nicht-Kreuzvalidierung (Endvalidierung)	163
4.3.3	Vorhersage unbekannter Verbindungen	164
	Literaturverzeichnis	165

1 ENTWICKLUNG WISSENSBASIERTER DESKRIPTOREN

1.1 Problemstellung

Eine Methode zur Berücksichtigung von Wasserstoffbrücken-Eigenschaften kleiner Moleküle bei einer vergleichenden Feldanalyse wurde mit dem CoMSIA-Verfahren verwirklicht (Klebe *et al.*, 1994a; Klebe, 1998; Klebe & Abraham, 1999). CoMSIA (*Comparative Molecular Similarity Indices Analysis*) ist eine Methode im Bereich der dreidimensionalen quantitativen Struktur-Wirkungsbeziehungen (*Quantitative Structure-Activity Relationship*, 3D-QSAR), in der molekulare Eigenschaften mit biologischen Aktivitäten von Molekülen eines Datensatzes korreliert werden. Ähnlich der weit verbreiteten CoMFA-Methode (*Comparative Molecular Field Analysis*), die ein Lennard-Jones-(6–12)- und ein Coulomb-Potenzial zur Berechnung von sterischen bzw. elektrostatischen Wechselwirkungsenergien verwendet (Cramer *et al.*, 1988; Cramer *et al.*, 1993), werden hierbei abstandsabhängige Ähnlichkeitsindizes berechnet. Dabei werden sterische, elektrostatische, hydrophobe und Wasserstoffbrücken-Eigenschaften zur Beschreibung der Moleküle herangezogen.

Die räumliche Festlegung der Wasserstoffbrücken-Eigenschaften wird in dem CoMSIA-Verfahren durch eine begrenzte Anzahl von Regeln festgelegt, die sich aus experimentellen Werten ableitet (Klebe, 1994). Zunächst wird das betreffende Molekül hinsichtlich vorhandener Wasserstoffbrücken-Donoren und -Akzeptoren untersucht. Daraufhin werden Pseudoatom-Positionen an den Stellen im Raum erzeugt, wo sich potenzielle Wasserstoffbrücken-Bindungspartner in einer virtuellen Proteinbindetasche befinden könnten. Die Platzierung dieser Pseudozentren erfolgt über vordefinierte Distanz-, Winkel- und Torsionswinkel-Werte einer Datenbank. Diese Größen werden mit Hilfe experimenteller Daten aus Kristallstrukturen parametrisiert, die das betreffende Wasserstoffbrücken-tragende Fragment besitzen. Um jede platzierte Pseudoatom-Position wird eine kugelförmige Gaußfunktion erzeugt, wobei ein stark vereinfachtes Schema zur Wichtung der einzelnen Funktionen ebenfalls aus den kristallographischen Daten abgeleitet wird. Abb. 1-1 zeigt die mit CoMSIA erzeugten Donor- und Akzeptorfelder eines Amidfragments zusammen mit den

entsprechenden Pseudoatom-Positionen. Die dreidimensionalen Gaußfunktionen sind als Konturen mit drei verschiedenen Farbniveaus angedeutet.

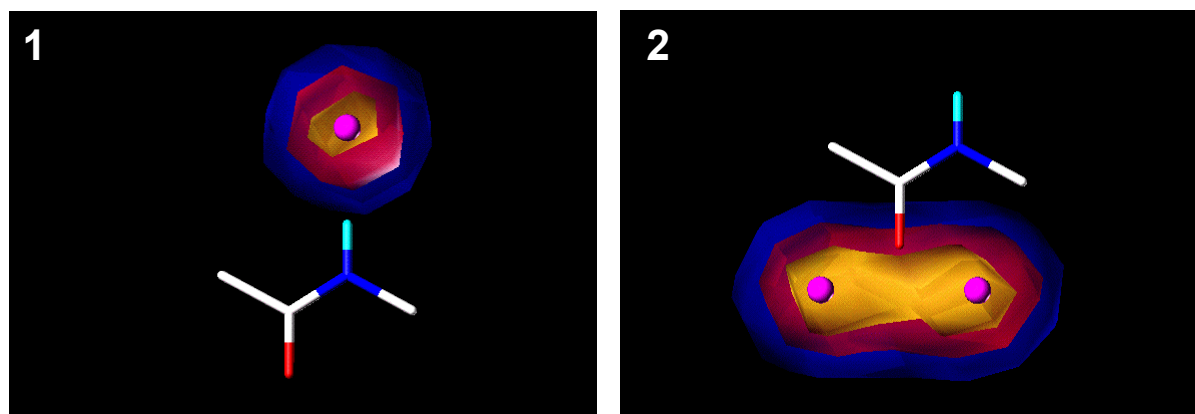


Abb. 1-1: CoMSIA Donor- und Akzeptorfeld einer aliphatischen Amidgruppe und ihre korrespondierenden Pseudoatom-Positionen. Das Donorfeld (1) wird durch ein Pseudoatom (violett) erzeugt, welches auf der verlängerten Achse der NH-Gruppe in einem Abstand von 1,9 Å zum Wasserstoffatom positioniert wird. Zwei Pseudozentren entlang der freien Elektronenpaare des Sauerstoffs in einem Abstand von 1,8 Å bilden das Akzeptorfeld (2). An den jeweiligen Pseudozentren werden kugelförmige Gaußfunktionen generiert, die günstige Regionen für Akzeptorgruppen (1) oder Donorgruppen (2) einer virtuellen Bindetasche darstellen. Diese Dichteverteilung ist auf drei Niveaus konturiert, wobei die Abweichung von einer Idealgeometrie der Wasserstoffbrücken-Wechselwirkung von gelb nach blau zunimmt.

Ein Nachteil der regelbasierten Implementierung der Wasserstoffbrücken-Eigenschaften in CoMSIA ist ihre stark vereinfachte und idealisierte Beschreibung. Es ermöglicht zwar die schnelle und einfache Berechnung von Deskriptoren, jedoch gehen wichtige und für eine spätere Verwendung nutzbare Informationen verloren. Ein neuer, wesentlich detaillierter Ansatz zur Definition von Wasserstoffbrücken-Eigenschaften geht daher von einer direkten Verwendung kristallographischer Daten aus der CSD (*Cambridge Structural Database*) aus, einer Datenbank mit mehr als 250.000 Kristallstrukturen¹ kleiner organischer und metallorganischer Verbindungen (Allen *et al.*, 1991; Allen & Kennard, 1993; Lommerse & Taylor, 1997). Basierend auf der CSD wurde die Datenbank IsoStar (Bruno *et al.*, 1997; Cole *et al.*, 1998) entwickelt, die Informationen über nichtbindende Wechselwirkungen in Form

¹ Am 5. Oktober 2001 wurde die 250.000ste Kristallstruktur in der CSD registriert.
[http:// www.ccdc.cam.ac.uk](http://www.ccdc.cam.ac.uk)

sogenannter *scatterplots* bereitstellt. Damit können die Verteilungen bestimmter funktioneller Gruppen (Kontaktgruppen) um andere Fragmente (Zentralgruppen) bestimmt werden. Die aktuelle IsoStar-Version 1.3 enthält ca. 18.000 Einträge von *scatterplots*, die aus mehr als 300 Zentralgruppen und 48 Kontaktgruppen zusammengesetzt sind. Aus den experimentellen Verteilungen lassen sich Wahrscheinlichkeitsdichten berechnen, deren Konturierung räumliche Bereiche markiert, in denen die Wahrscheinlichkeit für das Auftreten einer bestimmten Gruppe um eine andere groß ist (Abb. 1-2).

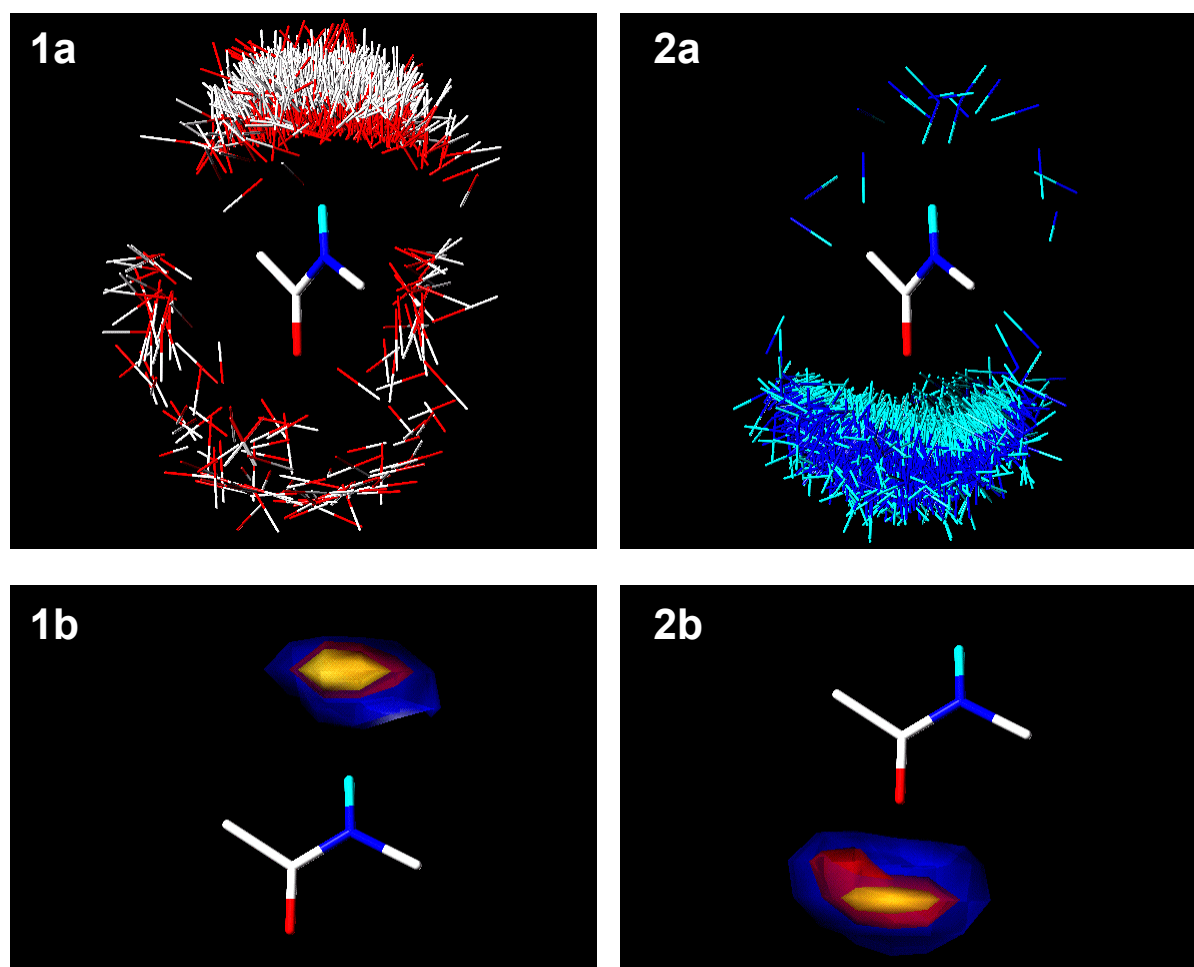


Abb. 1-2: Experimentell beobachtete Verteilungen von Carbonyl- (**1a**) und NH-Kontaktgruppen (**2b**) um ein aliphatisches Amidfragment. Aus diesen Verteilungen können entsprechende Wahrscheinlichkeitsverteilungen berechnet und angezeigt werden (**1b** und **2b**). Die hier dargestellten Dichten sind auf drei Niveaus (gelb: 8; rot: 4; blau: 2) konturiert. In dem gelb konturierten Bereich beispielsweise ist die Wahrscheinlichkeit, eine Carbonyl- (**1b**) bzw. NH-Gruppe (**2b**) zu finden, achtmal höher als die durchschnittliche Wahrscheinlichkeit.

Das Programm SuperStar (Verdonk *et al.*, 1999; Verdonk *et al.*, 2001; Boer *et al.*, 2001) verwendet die Informationen aus IsoStar, indem es eine zu untersuchende Proteinbindetasche in geeignete Zentralfragmente zerlegt und die jeweiligen *scatterplots* auf die dazugehörigen Fragmente überlagert. Nach einer Konvertierung und Normalisierung in entsprechende Wahrscheinlichkeitsdichten können diese als Wechselwirkungsfelder visualisiert und ausgewertet werden. Abb. 1-3 zeigt das Wechselwirkungsfeld einer Thermolysin-Bindetasche, das mit einer Carbonyl-Kontaktgruppe erzeugt wurde.

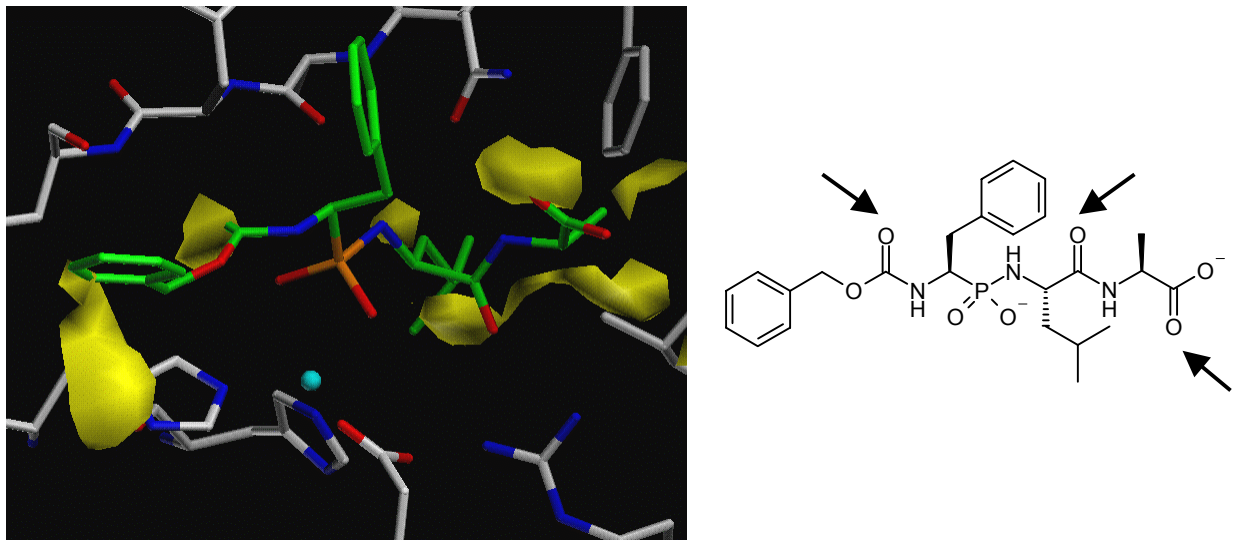


Abb. 1-3: Mit SuperStar berechnetes Wechselwirkungsfeld (links) einer Carbonyl-Kontaktgruppe in der Thermolysin-Bindetasche (PDB-Code 4TMN). Die gelben Konturen geben günstige Bereiche an, in denen ein potenzieller Ligand mögliche Carbonylgruppen positionieren kann. Der im Komplex enthaltene Thermolysin-Inhibitor weist drei Carbonylgruppen auf (rechts), die sich in den gelb gefärbten Bereichen befinden. Der Ligand wurde zur Berechnung des Wechselwirkungsfeldes vorher entfernt.

Bisher war die Anwendung und Validierung von Wechselwirkungsfeldern aus SuperStar auf die Analyse von Proteinbindetaschen begrenzt. Diese Felder können jedoch auch für kleine Moleküle generiert und somit als Deskriptoren zur Beschreibung der Wasserstoffbrücken-Eigenschaften verwendet werden. Die Entwicklung dieser Deskriptoren und ihre Anwendung im Bereich der vergleichenden molekularen Feldanalysen wird im folgenden beschrieben. Zunächst wird jedoch ein Überblick über literaturbekannte Methoden zur Erzeugung von Wechselwirkungsfeldern zur Charakterisierung von Proteinbindetaschen als auch von kleinen Molekülen gegeben.

1.2 Ansätze aus der Literatur

Ein Verfahren zur Bestimmung von Wechselwirkungsfeldern wurde mit dem weitverbreiteten Programm GRID (GRID, 2001) von Goodford (Goodford, 1985; Boobbyer *et al.*, 1989; Wade *et al.*, 1993; Wade & Goodford, 1993) realisiert. Es basiert auf einem parametrisierten Kraftfeldansatz und konturiert Bereiche innerhalb der Proteinbindetasche gemäß den Wechselwirkungsenergien einzelner Sonden auf festgelegten Gitterpunkten. Dabei stehen spezifische Sonden zur Verfügung, die einzelne Atome (*single-atom probe*; zum Beispiel phenolisches OH, amidisches NH, Wasser), chemisch-funktionelle Gruppen (*multi-atom probe*; zum Beispiel aliphatische Carboxylat-, aromatische Amidgruppe) oder hydrophobe (Sonde *DRY*) und amphiphile Eigenschaften (Sonde *BOTH*) repräsentieren können.

Bei dem MCSS-Ansatz (*multiple copy simultaneous search*) aus der Karplus-Gruppe (Miranker & Karplus, 1991; Caflisch *et al.*, 1993; Caflisch, 1996; English *et al.*, 2001) werden energetisch günstige Positionen und Orientierungen einer Vielzahl funktioneller Gruppen mittels des Kraftfelds CHARMM (Brooks *et al.*, 1983) berechnet. Dabei werden kleine Molekülfragmente in einer gegebenen Proteinbindetasche verteilt und anschließend mit dem Kraftfeld energieminiert. Um sowohl hydrophile als auch hydrophobe Merkmale der Bindetasche analysieren zu können, werden geladene (zum Beispiel Acetat, Methylammonium), polare (zum Beispiel Methanol, N-Methylacetamid), aromatische (zum Beispiel Toluol, Naphthalin), aliphatische (zum Beispiel Propan, Methylcyclohexan) als auch benutzerdefinierte Fragmente positioniert.

Die Natur der Direktionalität von Wasserstoffbrücken wird in HSITE (Danziger & Dean, 1989a; Danziger & Dean, 1989b) berücksichtigt, um geeignete Positionen dafür in einer gegebenen Proteinbindetasche zu sondieren. Dieser Ansatz basiert auf kristallographischen Informationen, die aus der CSD extrahiert wurden. Ausgehend von HSITE wurde das Programm HBMAP (Mills & Dean, 1996; Mills *et al.*, 1997) entwickelt, das ebenfalls ein überlagertes, gemittelttes Bild der Wechselwirkungsmuster im Kristallverband („*composite crystal-field information*“) von experimentellen Daten aus der CSD ableitet. Daraus konnten ebenfalls Wahrscheinlichkeiten potenzieller Wasserstoffbrücken-Positionen ermittelt werden.

Das Programm X-SITE (Singh *et al.*, 1991; Laskowski *et al.*, 1996; Wallace *et al.*, 1996) verwendet, ähnlich wie GRID, räumliche Kontaktverteilungen beruhend auf 163 dreiatomigen Fragmenten. Daraus konnten für insgesamt 26 verschiedene Atomtypen günstige Wechselwirkungsregionen in der Bindetasche identifiziert werden. Allerdings bestand die Datenbasis bei diesem Verfahren aus einem begrenzten Datensatz von 83 hochaufgelösten Proteinstrukturen der PDB (*Protein Data Bank*) (Bernstein *et al.*, 1977; Berman *et al.*, 2000).

Wie schon bei der Problemstellung erläutert, nutzt das in dieser Arbeit angewendete Programm SuperStar (Verdonk *et al.*, 1999; Verdonk *et al.*, 2001; Boer *et al.*, 2001) die in IsoStar (Bruno *et al.*, 1997; Cole *et al.*, 1998) enthaltenen räumlichen Informationen über nichtbindende Wechselwirkungen aus Kristallstrukturen niedermolekularer Verbindungen der CSD. Daraus können Wahrscheinlichkeitsdichten für Kontakte mit Atomen funktioneller Gruppen auf Gitterpositionen in einer Proteinbindetasche berechnet werden. Das Programm SuperStar und seine Anwendung werden in den nächsten Kapiteln ausführlich erläutert.

Weiterhin ermöglicht die Extraktion von Informationen aus experimentellen Daten eine Ableitung von sogenannten Bewertungsfunktionen (*scoring functions*), die zur Vorhersage von Bindungsaffinitäten eines Liganden an einem Rezeptor herangezogen werden können. Im Rahmen eines wissensbasierten Ansatzes werden im allgemeinen aus kristallographisch bestimmten Proteinstrukturen Häufigkeitsverteilungen für interatomare Wechselwirkungen berechnet; die erhaltenen radialen Verteilungsfunktionen werden anschließend unter Anwendung des „inversen Boltzmann’schen Gesetzes“ (Sippl, 1993; Sippl, 1995) in „Freie-Energie-Beiträge“ (*potentials of mean force*) bzw. „wissensbasierte Potenziale“ umgesetzt.

So ermittelten Muegge *et al.* (Muegge & Martin, 1999; Muegge *et al.*, 1999) aus 697 kristallographisch bestimmten Protein-Ligand-Komplexen „Helmholtz’sche Freie Wechselwirkungsenergien“ bzw. einen „PMF-Score“ (*potential of mean force*) unter Verwendung von 16 bzw. 34 Atomtypen für Proteine und Liganden. Die Verwendung eines Volumenkorrekturterms und maximale Abstände zwischen den betrachteten Atomen bis zu 12 Å bei der Herleitung der Paarverteilungsfunktionen ermöglichte dabei eine implizite Berücksichtigung der durch Wasser bedingten Beiträge.

Mit BLEEP (*Biomolecular Ligand Energy Evaluation Protocol*) entwickelten Mitchell *et al.* ebenfalls ein atombasiertes *potential of mean force* (PMF) zur Charakterisierung von

Protein-Ligand-Wechselwirkungen (Mitchell *et al.*, 1999b). Aus 351 strukturell diversen, hochaufgelösten Protein-Kristallstrukturen in der PDB wurden zu 40 unterschiedlichen Atomtypen – darunter auch polare Wasserstoffatome – insgesamt 820 mögliche, distanzabhängige Paarpotenziale abgeleitet. Zusätzlich zu diesem Protein-Ligand-Potenzial wurde eine weitere Version von BLEEP erzeugt, die unter Einbeziehen von Wassermolekülen auch Wechselwirkungen zwischen Protein und Wasser in der Bindetasche berücksichtigt. Die Berechnung von Bindungsenergien zu 90 Protein-Ligand-Komplexen resultierte in einem Korrelationskoeffizienten von 0,74 bezüglich experimentell bestimmter Affinitäten (Mitchell *et al.*, 1999a).

Von Gohlke *et al.* (Gohlke *et al.*, 2000a; Gohlke & Klebe, 2001) wurden Strukturinformationen aus Protein-Ligand-Komplexen ebenfalls in atombasierte, statistische Präferenzen umgewandelt. Die zugrundeliegenden Daten der PDB wurden mit Hilfe der Datenbank ReLiBase (*Receptor Ligand Database*) gewonnen. In einem GRID-ähnlichen Ansatz können mit Hilfe dieser distanzabhängigen Paarpräferenzen Potenzialwerte an Gitterpunkten berechnet werden. Durch die Visualisierung der Ergebnisse in Form von Konturdiagrammen lassen sich somit – spezifisch für einen Ligandatotyp – günstige Bereiche (*hot spots*) in Proteinbindetaschen identifizieren (Gohlke *et al.*, 2000b).

1.3 Theorie und Methoden

1.3.1 Deskriptoren zur Charakterisierung kleiner Moleküle

Der erste Schritt bei der Anwendung von SuperStar auf kleine Moleküle ist die Zerlegung des in Frage kommenden Moleküls in geeignete Fragmente. Die Fragmentierung muss dabei so erfolgen, dass alle Molekülteile als Zentralfragment in der Datenbank von IsoStar vorhanden sind. Bei der Analyse von Proteinbindetaschen ist diese Aufgabe überschaubar, da lediglich die 20 natürlich vorkommenden Aminosäuren untersucht werden müssen. Aufgrund der wesentlich höheren Vielfalt und Komplexität kleiner Moleküle ist die Auswahl geeigneter Fragmente für ihre korrekte und sinnvolle Charakterisierung entscheidend. Dabei können einerseits zu kleine und damit zu stark verallgemeinerte Bruchstücke die physikochemischen Eigenschaften der Liganden nicht ausreichend beschreiben. Andererseits können zu große und



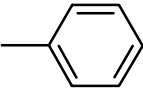





detailreiche Fragmente ebenfalls nicht herangezogen werden, da die dazugehörige Datenmenge oftmals keine statistisch signifikante Auswertung erlaubt. Ein Ligand, der beispielsweise eine Sulfonamid-Gruppe aufweist, würde in seinen physikochemischen Eigenschaften am besten durch ein einzelnes Sulfonamid-Fragment beschrieben werden. Der dazugehörige Eintrag in der IsoStar-Datenbank enthält jedoch nicht genügend Daten, die eine aussagekräftige statistische Auswertung zulassen würde. Daher ist es ratsam, die Sulfonamid-Gruppe in ein Sulfon-Fragment und ein planares oder pyramidales NH-Fragment² zu zerlegen. Beide Molekülteile weisen eine genügend große Datenlage in IsoStar auf und beschreiben ausreichend genau die physikochemischen Parameter des Sulfonamids (Akzeptor- und Donorfunktion).

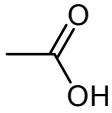
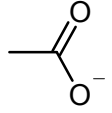
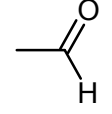
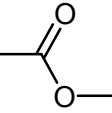
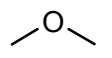
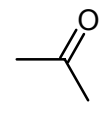
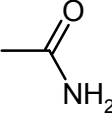
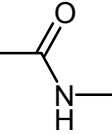
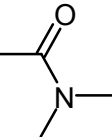
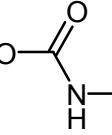
Diesen Überlegungen folgend wurde eine Auswahl geeigneter Zentralfragmente zur Beschreibung kleiner Moleküle getroffen. Aus der bestehenden IsoStar-Datenbank, die in der Version 1.3 über 300 Zentralgruppen enthält, wurden insgesamt 46 Fragmente selektiert und in einer separaten Fragment-Datenbank abgelegt (Tab. 1-1). Diese ausgewählten Gruppen weisen eine ausreichende Menge an Daten auf, die eine statistische Signifikanz der nachfolgenden Analysen gewährleisten. Weiterhin können mit dieser Sammlung an Fragmenten ein Großteil der in der medizinischen Chemie vorkommenden, kleinen Moleküle charakterisiert werden.

Einige Einträge in Tab. 1-1 sind als Fragmente mit nur schwachen Donor- oder Akzeptor-Eigenschaften markiert, insbesondere alle Einträge der CH-enthaltenden Gruppe. Diese Zentralfragmente weisen in IsoStar zwar eine statistisch signifikante Datenmenge auf, die Maximalwerte der dazugehörigen Wahrscheinlichkeitsdichten (*maximum propensity*, p_{max}) bei der Konvertierung heben sich aber kaum von einer Zufallswahrscheinlichkeit ab. Da sich die weitere Arbeit speziell auf die Berechnung von Wasserstoffbrücken-Eigenschaften konzentrierte, wurde deutlich zwischen Fragmenten mit starken und schwachen Donor- und Akzeptor-Eigenschaften unterschieden. Das ausschlaggebende Kriterium für die Einteilung eines Fragments in eine schwache Donor- oder Akzeptorgruppe war $p_{max} < 3$.

² Eine Analyse der in der CSD vorhandenen Kristallstrukturen mit einer Sulfonamidgruppe ergab, dass die Amideinheit sowohl in einer planaren als auch pyramidalen Form (ca. 50:50) vorliegt.

Tab. 1-1: Ausgewählte Zentralfragmente aus der IsoStar-Datenbank zur Beschreibung kleiner Moleküle. Die 46 Einträge sind in sechs Hauptgruppen unterteilt, die nur jeweils die angegebenen Atomtypen enthalten. Neben den Namen der Fragmente, wie sie in IsoStar bezeichnet werden (englische Ausdrücke), und den Strukturformeln sind die zugewiesenen physikochemischen Eigenschaften (D: Donor; A: Akzeptor) angegeben.

Nr.	IsoStar Name	Struktur ^(a)	D / A
C, H enthaltende Gruppe			
1	<i>methyl</i>	—CH_3	(D) ^(b)
2	<i>methylene</i>		(D)
3	<i>tertiary CH</i>		(D)
4	<i>phenyl</i>		(D)
5	<i>aromatic CH</i>		(D)
N, C, H enthaltende Gruppe			
6, 7	<i>aliphatic NH₂</i>	$\text{C}_{\text{al}}\text{—NH}_2$	D, A
8	<i>aromatic NH₂, planar</i>	$\text{C}_{\text{ar}}\text{—NH}_2$	D
9, 10	<i>aromatic NH₂, pyramidal</i>	$\text{C}_{\text{ar}}\text{—}\bar{\text{N}}\text{H}_2$	D, (A)
11	<i>amino, charged</i>	—NH_3^+	D
12	<i>aliphatic amine, charged</i>		D
13	<i>tert. ammonium, charged</i>		D
14	<i>planar ring N</i>		A
15	<i>aromatic NH, planar</i>		D
16	<i>aliphatic cyano</i>	$\text{C}_{\text{al}}\text{—C}\equiv\text{N}$	A
17	<i>aromatic cyano</i>	$\text{C}_{\text{ar}}\text{—C}\equiv\text{N}$	A

Nr.	IsoStar Name	Struktur ^(a)	D / A
O, C, H enthaltende Gruppe			
18, 19	<i>carboxylic acid, cis</i>		D, A
20	<i>carboxylate</i>		A
21, 22	<i>formyl</i>		(D), A
23, 24	<i>aliphatic OH</i>	$C_{al}-OH$	D, A
25, 26	<i>aromatic OH</i>	$C_{ar}-OH$	D, A
27	<i>aliphatic ester</i>		A
28	<i>aliphatic ether</i>		A
29	<i>aliphatic ketone</i>		A
N, O, C, H enthaltende Gruppe			
30, 31	<i>carbamoyl</i>		D, A
32, 33	<i>nitro</i>	$-NO_2$	D, A
34, 35	<i>aliphatic amide</i>		D, A
36	<i>N,N-disubstituted amide</i>		A
37, 38	<i>carbamate, trans</i>		D, A

Nr.	IsoStar Name	Struktur ^(a)	D / A
S, P enthaltende Gruppe			
39	<i>sulfonato, (phosphonato)</i> ^(c)	$\begin{array}{c} \text{O} \\ \parallel \\ \text{---S---O}^- \\ \parallel \\ \text{O} \end{array} \quad \begin{array}{c} \text{O} \\ \parallel \\ \text{---P---O}^- \\ \\ \text{O}^- \end{array}$	A
40	<i>sulfone, (phosphinato)</i> ^(c)	$\begin{array}{c} \text{O} \\ \parallel \\ \text{---S---} \\ \parallel \\ \text{O} \end{array} \quad \begin{array}{c} \text{O} \\ \parallel \\ \text{---P---} \\ \\ \text{O}^- \end{array}$	A
Halogen enthaltende Gruppe			
41	<i>aliphatic chloro</i>	$\text{C}_{\text{al}}\text{---Cl}$	D
42	<i>aromatic chloro</i>	$\text{C}_{\text{ar}}\text{---Cl}$	D
43	<i>aliphatic bromo</i>	$\text{C}_{\text{al}}\text{---Br}$	D
44	<i>aromatic bromo</i>	$\text{C}_{\text{ar}}\text{---Br}$	D
45	<i>aliphatic iodo</i>	$\text{C}_{\text{al}}\text{---I}$	D
46	<i>aromatic iodo</i>	$\text{C}_{\text{ar}}\text{---I}$	D

(a) C_{al} : aliphatisches (sp^3) C-Atom; C_{ar} : aromatisches oder sp^2 C-Atom

(b) Fragmente mit nur schwachen Donor- oder Akzeptor-Eigenschaften sind in Klammern gesetzt (siehe Erläuterung im Text)

(c) Aufgrund der geringen Datenlage bei den Phosphonato- und Phosphinato-Gruppen wurden die jeweiligen Einträge der Sulfonato- bzw. Sulfon-Gruppen verwendet

Ähnliche Überlegungen wie bei der Auswahl der Zentralfragmente wurden auch bei der Selektion geeigneter Kontaktgruppen (*probe atoms*) angestellt. Es fanden von den in IsoStar vorhandenen 48 Kontaktgruppen nur diejenigen Berücksichtigung, die zu allen Fragmenten aus Tab. 1-1 eine ausreichende Datenmenge aufweisen konnten. Daraus wurde wiederum eine nicht-redundante Gruppe von Kontaktatomen selektiert, die alle wichtigen physikochemischen Eigenschaften (zum Beispiel hydrophobe und Wasserstoffbrücken-Eigenschaften) widerspiegeln sollten. In Tab. 1-2 sind die daraus resultierten Kontaktgruppen zusammen mit ihren zugewiesenen physikochemischen Eigenschaften aufgeführt.

Tab. 1-2: Ausgewählte Kontaktgruppen aus der IsoStar-Datenbank zur Beschreibung der physikochemischen Eigenschaften kleiner Moleküle. Neben dem Namen der Kontaktgruppen, wie sie in IsoStar bezeichnet werden, sind ihre zugewiesenen physikochemischen Eigenschaften angegeben.

Nr.	IsoStar Name	Eigenschaft
1	<i>carbonyl oxygen</i>	Akzeptor
2	<i>amino hydrogen</i> ^(a)	Donor
3	<i>alcohol hydrogen</i>	Donor
4	<i>alcohol oxygen</i>	Akzeptor
5	<i>aliphatic carbon</i>	sterisch / hydrophob
6	<i>aromatic carbon</i>	sterisch / hydrophob
7	<i>oxygen atom</i> ^(b)	Akzeptor

(a) Wird in IsoStar unter „Any NH“ aufgeführt

(b) Generisches Sauerstoffatom; beinhaltet alle Kontaktgruppen mit terminalen Sauerstoffatomen, zum Beispiel Carbonyl-, Carboxylat-, Nitrogruppen

Abb. 1-4 zeigt ein Beispiel für die Vorgehensweise bei der Berechnung der Deskriptoren mit SuperStar. Ein Thermolysin-Inhibitor wird in seine strukturellen Fragmente (vgl. Tab. 1-1) zerlegt. Die in dem Liganden enthaltene Phosphonamid-Gruppe ist als solche zwar in der IsoStar-Datenbank vorhanden; sie wurde jedoch aufgrund einer unzureichenden Datenmenge nicht komplett miteinbezogen, sondern weiter in eine Phosphonato- und eine planare Amingruppe aufgeteilt. Die Carbonsäureamid-Gruppe hingegen wurde nicht in eine Carbonyl- und eine planare Amingruppe getrennt, da für dieses Fragment genügend Einträge in IsoStar vorhanden sind. Die als schwache Donorgruppen gekennzeichneten Methyl- und tertiären CH-Gruppen (vgl. Tab. 1-1) werden bei der Berechnung der Donor-Eigenschaften nicht berücksichtigt. Die zu den vier Fragmenten gehörenden *scatterplots* werden überlagert und in eine Wahrscheinlichkeitsdichte konvertiert. Diese geben Auskunft darüber, an welchen Stellen die Wahrscheinlichkeit groß ist, eine bestimmte Kontaktgruppe zu finden. In

Abb. 1-4 wurde die NH-Kontaktgruppe (*amino hydrogen*) verwendet; prinzipiell lassen sich jedoch alle in Tab. 1-2 aufgeführten Kontaktgruppen zur Charakterisierung des Liganden heranziehen.

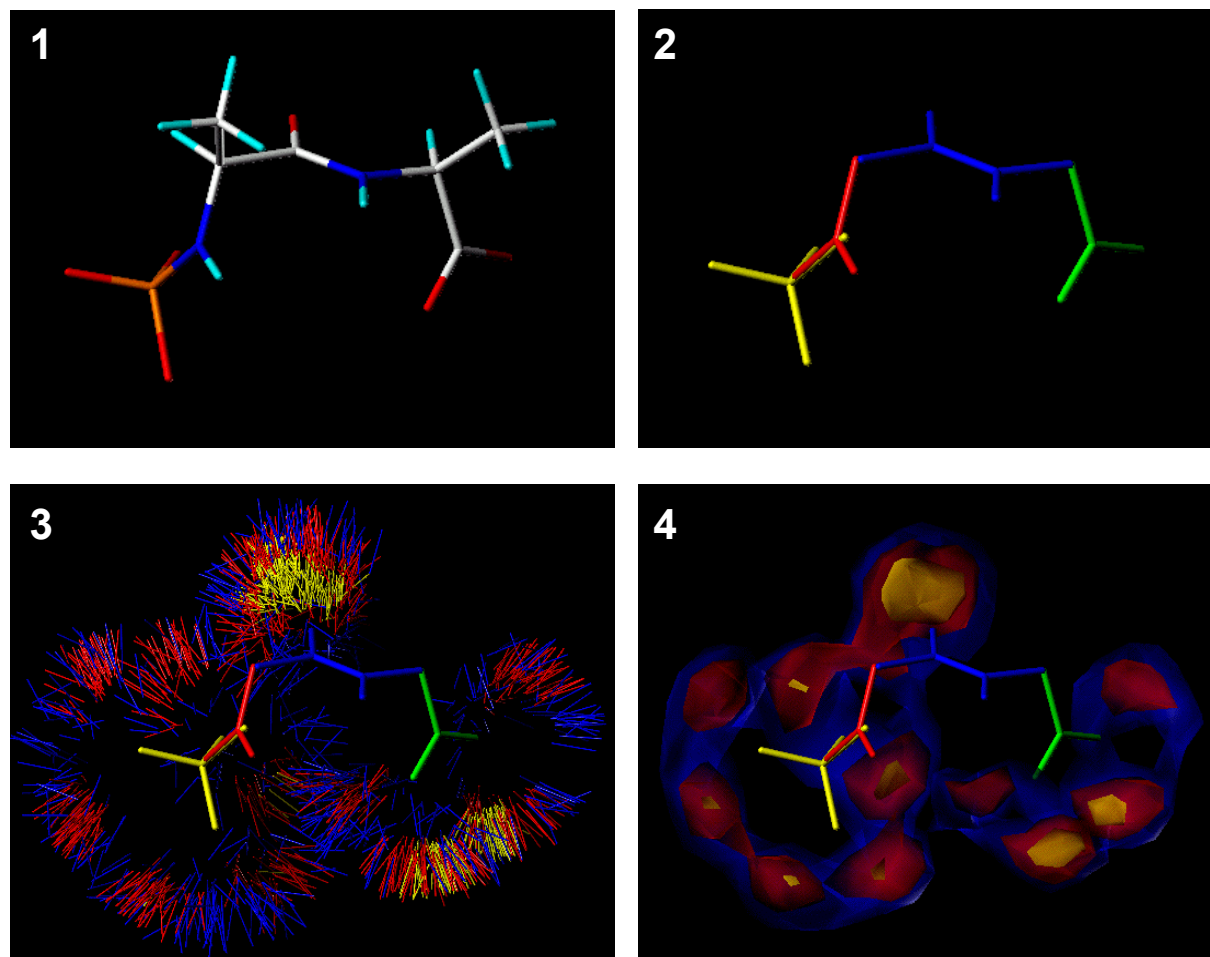


Abb. 1-4: Berechnung der Donor-Eigenschaften eines Thermolysin-Inhibitors. Der Ligand (1) wird in strukturelle Fragmente zerlegt; es werden eine Carboxylat- (grün), eine aliphatische Amid- (blau), eine planare Amin- (rot) und eine Phosphonatogruppe (gelb) gefunden (2). Die beiden Methyl- und tertiären CH-Gruppen besitzen nur schwache Donor-Eigenschaften (siehe Text) und sind daher nicht berücksichtigt. Die *scatterplots* werden auf die Fragmente überlagert (3) und in Wahrscheinlichkeitsdichten umgewandelt (4). Die Kontaktgruppen der *scatterplots* und die Dichtekontur sind in drei Abstufungen (gelb: 8; rot: 4; blau: 2) abgebildet. Zur Berechnung der Donor-Eigenschaften wurde die NH-Kontaktgruppe (*amino hydrogen*) verwendet. Beispielsweise ist in dem rot konturierten Bereich die Wahrscheinlichkeit, das H-Atom einer NH-Kontaktgruppe zu finden, vier mal höher als die durchschnittliche Wahrscheinlichkeit.

1.3.2 Validierung der Deskriptoren durch vergleichende Feldanalysen nach dem CoMSIA-Verfahren

Eine Validierung der mit SuperStar erzeugten Deskriptoren erfolgte durch ihren Einsatz in vergleichenden molekularen Feldanalysen (*comparative molecular field analyses*). Die Deskriptoren wurden zur Berechnung von molekularen Feldern – ähnlich zu einem gewöhnlichen CoMFA- oder CoMSIA-Ansatz – herangezogen und im Rahmen einer quantitativen Struktur-Wirkungsbeziehung analysiert. Die Durchführung einer vergleichenden molekularen Feldanalyse soll im folgenden kurz erläutert werden.

Die Grundlage von quantitativen Struktur-Wirkungsbeziehungen ist die Schaffung einer Beziehung zwischen der Wirkung (biologische Aktivität, Bindungsaffinitäten) einer Verbindung und ihrer zugrundeliegenden Struktur (Kubinyi, 1993; Kubinyi, 1994; Böhm *et al.*, 1996). Aufgrund der im allgemeinen fehlenden Informationen über die Rezeptorstruktur ist die Berechnung von Absolutwerten der Bindungsaffinitäten nicht möglich. Daher werden bei QSAR-Analysen die relativen Unterschiede der Verbindungen eines Datensatzes ermittelt und diese mit den biologischen Daten korreliert. Die Voraussetzung für einen relativen Vergleich der 3D-Strukturen untereinander ist ihre räumliche Überlagerung (*alignment*). Darunter versteht man die räumliche Orientierung der Wirkstoffmoleküle in der Art, wie sie in der Bindetasche vorliegen würden. Das Ziel dabei ist, eine bioaktive Konformation der meist flexiblen Moleküle anzugeben und ihre relative Orientierung zueinander zu bestimmen. Mittlerweile gibt es eine Vielzahl von Überlagerungsverfahren, auf die hier nicht weiter eingegangen werden kann. Ein kürzlich erschienener Übersichtsartikel (Lemmen & Lengauer, 2000) gibt jedoch einen umfassenden Überblick über derzeit existierende Überlagerungsmethoden. Liegt nun eine Serie von Molekülen in ihrer dreidimensionalen, überlagerten Form vor, so können ihre zu vergleichenden molekularen Felder berechnet werden. Dabei wird zunächst ein Gitterkasten (*grid box*) um die Moleküle des Datensatzes gelegt; die Größe der Gitterbox wird so gewählt, dass alle Moleküle weiträumig innerhalb eines Abstands von 4 - 5 Å zum Rand eingebettet sind; die Maschenweite des Gitters beträgt gewöhnlich 1 - 2 Å. An jedem Gitterpunkt werden anschließend mit Hilfe einer Sonde (*probe atom*) molekulare Wechselwirkungsenergien berechnet. Die Gesamtheit aller Wechselwirkungsenergien ergibt das „molekulare Feld“ eines Moleküls. In einer nachfolgenden Korrelationsanalyse werden diese Felder miteinander verglichen und mit den zugehörigen Bindungsaffinitäten in Bezug gesetzt (vergleichende molekulare Feldanalyse).

Als Referenz und Vergleichsbasis der Validierung dienen in dieser Arbeit die Ergebnisse aus den QSAR-Modellen, die bei der Verwendung der standardmäßig verwendeten CoMSIA-Felder aufgestellt werden (Klebe & Abraham, 1999). Wie zu Beginn der Problemstellung schon angedeutet, wird das molekulare Feld in CoMSIA mit Hilfe von Ähnlichkeitsindizes bezüglich eines Sondenatoms berechnet (Klebe *et al.*, 1994a). An einem Gitterpunkt q ergibt sich der Ähnlichkeitsindex A_F für ein Molekül j mit der Eigenschaft k wie folgt (Gl. 1-1):

$$A_{F,k}^q(j) = - \sum_i w_{probe,k} \cdot w_{ik} \cdot \exp(-\alpha \cdot x_{iq}^2) \quad \text{Gl. 1-1}$$

Die Ähnlichkeit an dem Gitterpunkt q wird durch die Eigenschaft des Sondenatoms ($w_{probe,k}$) und durch die physikochemischen Eigenschaften der Atome i (w_{ik}) bestimmt. Die Abstandsabhängigkeit zwischen Gitterpunkt q und Atom i – festgelegt durch den Abstand x_{iq} – wird über eine Gaußfunktion beschrieben. Die Eigenschaften des Sondenatoms $w_{probe,k}$ werden auf Standardwerte gesetzt: Sonde mit Radius 1 Å, Ladung +1, Hydrophobizität +1, Donor-Eigenschaft +1 und Akzeptor-Eigenschaft +1. Die Breite der Gaußkurve wird durch den Abschwächungsfaktor α (*attenuation factor*) bestimmt. Ein großer Wert für α bedeutet einen schnellen Abfall der Abstandsfunktion. Dies ermöglicht eine stärkere Fokussierung auf lokale Ähnlichkeiten. Bei einem kleinen Abschwächungsfaktor hingegen wird der globalen Beschreibung der Moleküle ein höheres Gewicht beigemessen (Abb. 1-5).

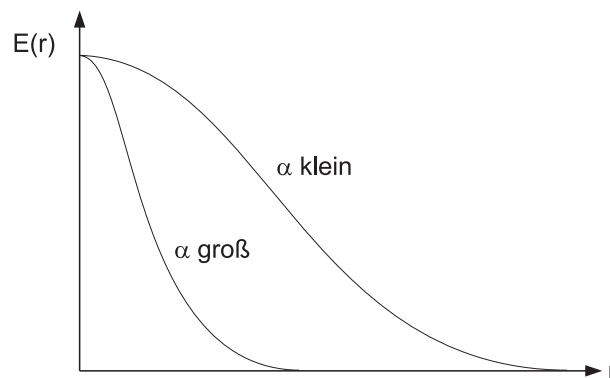


Abb. 1-5: Abhängigkeit der Breite einer Gaußfunktion vom Wert des Abschwächungsfaktors α .

Ähnlichkeitsindizes können für eine Vielzahl von Eigenschaften k bestimmt werden. Die Voraussetzung für die Berechnung ist eine Beschreibung der Moleküle über atombasierte

Eigenschaftswerte w_{ik} . Für die Charakterisierung von Rezeptor-Ligand-Wechselwirkungen sind vor allem sterische, elektrostatische, hydrophobe und Wasserstoffbrücken-Eigenschaften von Bedeutung. In CoMSIA werden daher standardmäßig fünf Felder (sterisch, elektrostatisch, hydrophob, Donor und Akzeptor) für ein zu untersuchendes Molekül berechnet, die diese physikochemischen Eigenschaften repräsentieren.

Im Falle der SuperStar-Deskriptoren ergibt sich das molekulare Feld eines Liganden aus der Berechnung der Wahrscheinlichkeitsverteilungen bezüglich einer bestimmten Kontaktgruppe, die aus der Konvertierung der *scatterplots* resultieren. An den Schnittpunkten des zugrundeliegenden Gitters werden für jedes Molekül die Werte der Wahrscheinlichkeitsdichten abgelegt und direkt in der nachfolgenden PLS-Analyse mit den Bindungsaffinitäten korreliert. Zur Beschreibung der physikochemischen Eigenschaften der untersuchten Liganden wurden die sieben Felder berechnet, die sich bei der Verwendung der Kontaktgruppen aus Tab. 1-2 (Seite 13) ergeben.

Weiterhin wurden zur Validierung der entwickelten SuperStar-Deskriptoren mit dem Programm GRID (Goodford, 1985; Boobbyer *et al.*, 1989; Wade *et al.*, 1993; Wade & Goodford, 1993) berechnete Felder als Vergleich herangezogen. Dazu kamen fünf Sondenatome zur Auswahl, deren physikochemischen Eigenschaften denen von SuperStar ähnlich sein sollten (Tab. 1-3).

Tab. 1-3: Ausgewählte Sondenatome aus GRID zur Beschreibung der physikochemischen Eigenschaften kleiner Moleküle. Neben dem Namen der Sondenatome, wie sie in GRID bezeichnet werden, und der entsprechenden chemischen Gruppe sind ihre zugewiesenen physikochemischen Eigenschaften angegeben.

Nr.	GRID Name	chemische Gruppe	Eigenschaft
1	O	<i>sp² carbonyl oxygen</i>	Akzeptor
2	N1:	<i>sp³ NH with lone pair</i>	Donor
3	O1	<i>alkyl hydroxy group</i>	Donor / Akzeptor
4	C3	<i>methyl CH₃ group</i>	sterisch / hydrophob
5	C1=	<i>sp² CH aromatic or vinyl</i>	sterisch / hydrophob

Die Durchführung aller Analysen mit CoMSIA, GRID und SuperStar³ erfolgte derart, dass die resultierenden statistischen und graphischen Ergebnisse direkt miteinander vergleichbar waren. So wurden für sämtliche Feldberechnungen die gleichen Koordinaten des Gitterkastens mit einem einheitlichen Gitterabstand von 1 Å verwendet. Für die fünf CoMSIA-Felder wurden in dieser Arbeit die in SYBYL vorgegebenen Standardparameter herangezogen ($\alpha = 0,3$). Die partiellen Atomladungen der Liganden wurden mit einem semiempirischen Verfahren nach der AM1-Methode (Dewar *et al.*, 1985) mit dem Programm MOPAC (Stewart, 1990) ermittelt. Ein SPL-Makro ermöglichte die Importierung der mit SuperStar und GRID zunächst extern berechneten Felder in eine SYBYL-Tabelle. Die Korrelation der Bindungsaffinitäten mit den Deskriptoren der molekularen Feldbeiträge wurde über eine PLS-Analyse⁴ (*Partial Least Square*) durchgeführt (Wold *et al.*, 1984). Eine Skalierung (*CoMFA standard scaling*) ermöglichte dabei die gleiche Gewichtung der einzelnen Felder untereinander.

Die Vorhersagekraft der erzeugten Modelle wurde mittels Kreuzvalidierung überprüft, wobei die Methode *leave-one-out* (LOO) zur Anwendung kam, bei der immer nur ein Molekül dem Trainingsdatensatz entnommen und seine biologische Aktivität aufgrund des mit den verbleibenden Verbindungen abgeleiteten Modells vorhergesagt wird. Die Kreuzvalidierung erfolgte mit der SAMPLS-Methode (*Sample-distance PLS*), eine schnellere Variante der PLS-Analyse. Das Verfahren beruht auf der Reduktion der erklärenden Variablen auf eine $n \times n$ Kovarianzmatrix bei n Verbindungen in einem Datensatz (Bush & Nachbar, 1993). Als Auswahlkriterium zur Bestimmung der optimalen Anzahl an Komponenten während der Kreuzvalidierung diente der q^2 -Wert⁵. Die Aufnahme jeder zusätzlichen Komponente zum Modell musste dabei mit einer Erhöhung des q^2 -Wertes von mindestens 5% einhergehen. Diese Vorgehensweise – auch als „*principle of parsimony*“ bekannt – entspricht dem Ziel einer PLS-Analyse, ein Modell mit möglichst wenigen, signifikanten Komponenten zu erstellen (Thibaut *et al.*, 1993; Kubinyi & Abraham, 1993). Die so erhaltene Zahl entsprach meist auch der Anzahl an Komponenten, für die s_{PRESS} (*standard deviation of predictions*) den kleinsten Wert annahm.

³ Die verwendeten Programme sind in Kap. 4.2 nochmals kurz aufgeführt.

⁴ Die statistischen Größen der PLS-Analyse werden in Kap. 4.3 erläutert.

⁵ andere Bezeichnung: kreuzvalidierter r^2 -Wert

Mit der gleichen Zahl an Komponenten wurde schließlich eine nicht-kreuzvalidierte Analyse durchgeführt. Um eine Reduktion der Rechenzeit zu ermöglichen, wurde die Option „*column filtering*“ in SYBYL eingeschaltet. In diesem Fall werden alle Variablen, deren Standardabweichung an einem Gitterpunkt bezüglich aller Moleküle kleiner als ein zuvor festgelegter Grenzwert (*minimum-sigma*) ist, auf Null gesetzt und bei der eigentlichen PLS-Analyse somit nicht berücksichtigt. Der Grenzwert wurde in dieser Arbeit für jede Analyse so gesetzt, dass mindestens 10% der Variablen in die PLS-Analyse eingingen. Diese Vorgehensweise war erforderlich, da die Feldwerte bei CoMSIA und SuperStar keine physikochemischen Einheiten besitzen, sondern in willkürlichen Einheiten bzw. Wahrscheinlichkeitswerten angegeben werden. Im Gegensatz dazu werden bei CoMFA Lennard-Jones- und Coulomb-Potenziale in Energieeinheiten von kcal/mol ausgedrückt. Standardmäßig wird der *minimum-sigma* Grenzwert bei CoMFA-Analysen auf 2 kcal/mol gesetzt, was im Durchschnitt eine Datenreduktion um 90% zur Folge hat; dies entspricht somit der gewählten Vorgehensweise der Berücksichtigung von 10% aller Variablen.

Die oben beschriebenen Verfahren zur Erzeugung von molekularen Deskriptoren – CoMSIA, GRID, SuperStar – wurden an einem bereits zuvor in der Literatur beschriebenen Datensatz von Thermolysin-Inhibitoren (DePriest *et al.*, 1993; Klebe *et al.*, 1994a; Klebe & Abraham, 1999) validiert. Die experimentell bestimmten Bindungsaffinitäten aller Verbindungen lagen in Form von pK_i -Werten ($-\log K_i$) vor. Ein Trainingsdatensatz von 61 Inhibitoren wurde zum Aufstellen der jeweiligen QSAR-Modelle verwendet. Weitere 15 Verbindungen, die nicht in dem Trainingsdatensatz enthalten waren, dienten in Form eines Testdatensatzes zur Bestimmung der externen Vorhersagefähigkeit der entwickelten Modelle. Dazu wurde eine weitere statistische Größe (*predictive r^2*) bestimmt, die eine quantitative Abschätzung der Vorhersagekraft zulässt (Cramer *et al.*, 1988).

Die physikochemischen Eigenschaften der Moleküle werden in CoMSIA durch sterische, elektrostatische, hydrophobe und Wasserstoffbrückenfelder beschrieben. Diese Felder verhalten sich jedoch nicht linear unabhängig voneinander, sondern sind in komplexer Weise miteinander korreliert (Böhm *et al.*, 1999). Bei der Erstellung der QSAR-Modelle mit SuperStar wurden insgesamt sieben Eigenschaftsfelder berücksichtigt. Zur Überprüfung, inwieweit die Feldbeiträge interkorrelierten, wurden einige mögliche Kombinationen der sieben Felder gebildet und damit jeweils eine PLS-Analyse durchgeführt.

1.3.3 Deskriptoren zur Beschreibung von Wasserstoffbrücken-Eigenschaften

Wie in Kapitel 1.3.1 erläutert, lässt sich zur Charakterisierung von Molekülen prinzipiell jede Kontaktgruppe, die in IsoStar vorhanden ist, verwenden. Eine nicht-redundante Gruppe von Kontaktatomen, die zu allen Zentralfragmenten aus Tab. 1-1 eine signifikante Anzahl an Daten aufweisen, ist in Tab. 1-2 aufgeführt. Im weiteren Verlauf dieser Arbeit wurde auf die detaillierte Darstellung von Wasserstoffbrücken-Eigenschaften näher eingegangen. Dabei kamen zwei geeignete Kontaktgruppen zur Auswahl, die diese Eigenschaften am besten repräsentieren konnten: (1.) zur Charakterisierung von Akzeptor-Eigenschaften wurde das Sauerstoffatom der Carbonyl-Kontaktgruppe (kurz: CO) herangezogen; (2.) die Donor-Eigenschaften wurden von dem H-Atom der NH-Kontaktgruppe (*amino hydrogen*) repräsentiert (siehe Tab. 1-2).

Um im weiteren eine einheitliche Nomenklatur beizubehalten, sei an dieser Stelle darauf hingewiesen, dass bei der Berechnung eines *Akzeptorfeldes* alle Akzeptorgruppen eines Liganden berücksichtigt werden und dabei die *Donor-Eigenschaften* in die Proteinumgebung des Moleküls projiziert werden. Dieses erfolgt unter Verwendung der NH-Kontaktgruppe. Umgekehrt werden bei der Benutzung der CO-Kontaktgruppe die funktionellen Donorgruppen zur Berechnung des *Donorfeldes* berücksichtigt, die wiederum die *Akzeptor-Eigenschaften* um den Liganden abbilden. Die Begriffe „erzeugendes Feld“ und „umgebende Eigenschaft“ verhalten sich bezogen auf die jeweiligen Deskriptoren also komplementär.

Die alleinige Verwendung dieser beiden Deskriptoren zu einer vollständigen Charakterisierung der Liganden erscheint jedoch nicht ausreichend. Um zunächst eine direkte Vergleichbarkeit zu den derzeit implementierten Wasserstoffbrücken-Deskriptoren in CoMSIA zu ermöglichen, wurden die ursprünglichen CoMSIA Donor- und Akzeptorfelder durch die entsprechenden SuperStar-Felder ausgetauscht. Dies bedeutet, dass bei einer vollständigen Beschreibung eines Moleküls das sterische, elektrostatische und hydrophobe Feld von CoMSIA berechnet wird, während die Donor- und Akzeptorfelder mit SuperStar ermittelt und anschließend mit den übrigen Feldern kombiniert werden.

Eine Alternative zur weiteren Verwendung der drei Felder aus CoMSIA besteht in einem ausschließlichen Einsatz von SuperStar-Feldern zur Beschreibung der Moleküleigenschaften. Dabei stellt sich die Frage, inwieweit die Zentralfragmente mit nur schwachen Donor- oder

Akzeptor-Eigenschaften – besonders die CH-enthaltenden Gruppen (siehe Tab. 1-1) – berücksichtigt werden müssen. Daher wurden in diesem Fall zwei vergleichende Analysen durchgeführt, wobei einmal alle Fragmente aus Tab. 1-1 zur Feldberechnung verwendet und anschließend als Alternative die Gruppe mit den CH-enthaltenden Fragmenten von der Analyse ausgeschlossen wurde.

1.3.4 Approximation durch anisotrope Gaußfunktionen

Ein großer Nachteil bei der direkten Verwendung von IsoStar *scatterplots* ist die relativ zeitaufwändige Berechnung der Eigenschaftsfelder: ein zu untersuchendes Molekül wird zunächst in Fragmente zerlegt; jedes dieser Fragmente wird in der IsoStar-Datenbank gesucht, wobei die entsprechenden *scatterplots* anschließend überlagert werden; die einzelnen *scatterplots* werden vereinigt und in eine Dichtekontur (*density map*) konvertiert; diese wird zu einer Wahrscheinlichkeitsdichte (*propensity map*) normiert und schließlich als Eigenschaftsfeld konturiert und graphisch angezeigt. Bei der Analyse einer Reihe von Liganden müssen oft Hunderte von Fragmenten durchsucht werden. Meist sollen auch mehrere Eigenschaften angezeigt werden, so dass sich die Rechenzeit bei der Verwendung mehrerer Kontaktgruppen entsprechend multipliziert. Typische Datensätze weisen ca. 50 – 100 Liganden auf; bei einer durchschnittlichen Zahl von 10 Fragmenten pro Molekül kann dies eine Bearbeitung von über 1.000 zuzuweisenden Fragmenten ergeben. Bei der Analyse mehrerer Eigenschaften erhöht sich diese Zahl entsprechend.

Ein Ansatz zu einer deutlichen Reduktion der Rechenzeit ist die Approximation der ursprünglichen Wahrscheinlichkeitsdichten in geeignete, schnell berechenbare Deskriptoren. Nissink *et al.* entwickelte eine Methode, die von SuperStar berechnete Deskriptoren durch nicht-sphärische, *anisotrope Gaußfunktionen* anzunähern versucht (Nissink *et al.*, 2000). Anisotropie bedeutet in diesem Zusammenhang, dass die Form dieser dreidimensionalen Deskriptoren in drei Raumrichtungen unterschiedlich ist. Durch diese Art der Approximation lassen sich zum Beispiel auch „bananenförmige“ Wahrscheinlichkeitsdichten ausdrücken. Die Notwendigkeit für eine solche Annäherung begründet sich darin, dass die allgemeine Gestalt der ursprünglichen SuperStar-Deskriptoren nicht kugelförmig⁶ ist, sondern einen mehr oder

⁶ oder: isotrop

weniger deformierten Verlauf im Raum annimmt, wie beispielsweise in Abb. 1-2 zu erkennen ist. Die mathematische Form der nicht-sphärischen, anisotropen Funktion ist in Gl. 1-2 dargestellt.

$$p^c(i) = h \cdot f_1 \cdot f_2 \cdot f_3 \quad \text{Gl. 1-2}$$

Der Wahrscheinlichkeitswert p an einem bestimmten Punkt i ist gegeben durch die Höhe h an einem vordefinierten Zentrum c multipliziert mit drei Gauß-artigen Abstandsfunktionen f_n ($n = 1, 2, 3$). Diese Abstandsfunktionen (*decay functions*) sorgen für eine Abnahme der Wahrscheinlichkeitswerte (δ_1 , δ_2 und δ_3) in drei Raumrichtungen (Abb. 1-6).

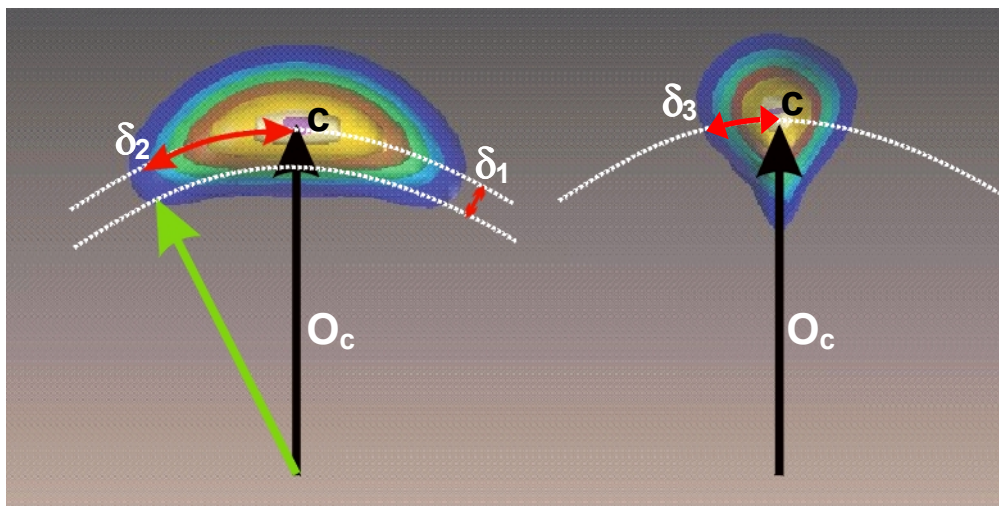


Abb. 1-6: Darstellung der nicht-sphärischen, anisotropen Gaußfunktion. Der rechte Teil der Abbildung wurde durch eine vertikale 90° Drehung erzeugt. Der Basisvektor O_c zeigt in Richtung des vordefinierten Zentrums c ; an diesem Punkt besitzt die Wahrscheinlichkeitsdichte ihren höchsten Wert. Drei Abstandsfunktionen sind für eine unterschiedliche Abnahme des Wahrscheinlichkeitswerts in drei Richtungen (δ_1 , δ_2 und δ_3) verantwortlich.

Die Abstandsfunktionen f_n sind als gewöhnliche Gaußfunktionen mit den Koeffizienten c_n gegeben (Gl. 1-3). Die Öffnung der Gaußfunktionen σ_n ($n = 1, 2, 3$) ist abhängig von der jeweiligen Verteilung der Wahrscheinlichkeitsdichte und muss in einem Anpassungsprozeß (*fitting*) für jede Dichte, d.h. für jede Kombination von Zentralfragment mit Kontaktgruppe, einzeln bestimmt werden.

$$f_n(\delta_n) = \exp(-c_n \cdot \delta_n^2); \quad c_n = \frac{1}{2\sigma_n^2} \quad \text{Gl. 1-3}$$

Das Ergebnis ist eine Liste mit nur wenigen Parametern pro Typ einer Wahrscheinlichkeitsdichte, die in einer einfachen Textdatei abgespeichert werden kann. Bei der Analyse einer Proteinbindetasche oder eines Datensatzes von Liganden muss nun nicht mehr auf die *scatterplots* selbst zurückgegriffen werden, sondern nur auf die Parameter der entsprechenden Kombination von Zentralfragment und Kontaktgruppe. Nissink *et al.* ermittelten zum Beispiel bei der Untersuchung der Bindetasche von Proteinase K (PDB-Code 2PRK) eine Reduktion der Rechenzeit von 85 auf 14 Sekunden; im allgemeinen konnte eine Steigerung der Rechengeschwindigkeit um den Faktor fünf und höher erzielt werden.

In dieser Arbeit wurden die anisotropen Gaußfunktionen für die Berechnung der Wasserstoffbrücken-Deskriptoren eingesetzt. Die dazugehörigen Parameter für alle Fragmente aus Tab. 1-1 in Kombination mit den CO- und NH-Kontaktgruppen (Tab. 1-2) waren verfügbar. Es wurde der direkte Vergleich mit den ursprünglichen SuperStar-Deskriptoren hinsichtlich einer Verbesserung der daraus resultierenden Modelle bei gleichzeitiger Erhöhung der Rechengeschwindigkeit durchgeführt. Ähnlich, wie im vorherigen Kapitel beschrieben, wurden wiederum zwei getrennte Analysen mit und ohne Berücksichtigung der CH-enthaltenden Fragmente vollzogen.

1.3.5 Approximation durch isotrope Gaußfunktionen

Eine vergleichbare Methode, die Eigenschaftsfelder aus SuperStar durch schneller berechenbare Deskriptoren zu ersetzen, ist die Anpassung der ursprünglichen Wahrscheinlichkeitsdichten durch eine begrenzte Anzahl von *isotropen Gaußfunktionen*. Im Gegensatz zu den im vorherigen Kapitel erläuterten anisotropen Gaußfunktionen zeichnen

sich diese Funktionen durch ihre Rotationssymmetrie in alle Raumrichtungen aus. Es handelt sich also um sphärische, kugelförmige Funktionen. Die mathematische Form einer sphärischen Gaußfunktion ist durch Gl. 1-4 gegeben.

$$p^c(i) = h \cdot \exp(-\alpha \cdot x_i^2) \quad \text{Gl. 1-4}$$

Der Wahrscheinlichkeitswert p an einem bestimmten Punkt i wird berechnet durch die Höhe h im Zentrum c des Koordinatensystems (Mittelpunkt der Sphäre), multipliziert mit einer exponentiellen Abstandsfunktion (vgl. Gl. 1-3). Der Parameter α bestimmt die Breite der Abstandsfunktion und x_i definiert den Abstand zwischen dem Zentrum c und dem Punkt i . Um die meist ungleichmäßigen Wahrscheinlichkeitsdichten möglichst gut approximieren zu können, wird nicht nur eine, sondern mehrere kugelförmige Gaußfunktionen nacheinander in eine vorgegebene, ursprüngliche Dichte eingepasst (Gl. 1-5).

$$P(i) = \sum_m p^c(i) \quad \text{Gl. 1-5}$$

Dabei wird der Wahrscheinlichkeitswert P an einem bestimmten Punkt i durch die Summe der Werte aller einzelnen sphärischen Gaußfunktionen m (Gl. 1-4) errechnet. Die Anpassung von mehreren sphärischen Gaußfunktionen an eine bestehende Wahrscheinlichkeitsdichte erfolgt in mehreren Schritten.

Zuerst wird die Wahrscheinlichkeitsverteilung aus einem gegebenen *scatterplot* aus der IsoStar-Datenbank generiert. Dazu wird eine dreidimensionale Gitterbox mit regelmäßigen Gitterabständen auf den *scatterplot* gelegt. Die Anzahl der Kontaktgruppen innerhalb jedes einzelnen Gitterwürfels wird gezählt und die daraus zunächst bestimmten Dichtewerte durch eine Normalisierung in eine Wahrscheinlichkeitsdichte konvertiert. Die Durchführung dieser Normalisierung ist von Verdonk *et al.* im Detail beschrieben worden (Verdonk *et al.*, 1999). In dieser Publikation wurde ein Gitterabstand von 0,7 Å für die Validierung von SuperStar verwendet. Im Hinblick auf den später geplanten Einsatz dieser Deskriptoren in einem vergleichenden Feldanalyse-Verfahren, wo der Einsatz von Gitterweiten kleiner als 1 Å extrem rechenintensiv wird (Standardanalysen verwenden normalerweise Gitterabstände von nur 2 Å), wurde in der hier verwendeten Approximation geprüft, ob neben der

voreingestellten Gitterweite von $0,7 \text{ \AA}$ nicht auch ein Abstand von 1 \AA für die Auswertungen ausreicht.

Der nächsten Schritt ist die Feststellung des ersten globalen Maximums mit dem höchsten Wert der Wahrscheinlichkeitsdichte. An die Stelle eines solchen Maximums wird das Zentrum einer sphärischen Gaußfunktion (im weiteren als „Gaußhut“ bezeichnet) gelegt und ihre relative Höhe an den Wert der Wahrscheinlichkeitsdichte an dieser Position angepasst. Dabei erfolgt in einem Voro Optimierungsschritt die Positionierung des Gaußhutes zunächst nur räumlich, so dass er möglichst gut in die Referenzdichte eingepasst wird. Bei einer anschließenden Nachoptimierung wird dann seine Höhe an die Dichte angepasst. Im Gegensatz zu der „*peak fitting*“ Routine in SuperStar (Verdonk *et al.*, 2001), bei der jeder Gaußhut individuell in seiner Breite kalibriert wird, sind hier alle positionierten Gaußhüte auf eine gleiche, vorher definierte (siehe unten) Breite gesetzt.

Anschließend erfolgt die Berechnung einer Differenzdichte D , indem die durch den platzierten Gaußhut erzeugte Dichte P (vgl. Gl. 1-5) von der ursprünglichen Wahrscheinlichkeitsdichte P_0 subtrahiert wird (Gl. 1-6).

$$D(i) = P_0(i) - P(i) \quad \text{Gl. 1-6}$$

In der verbleibenden Differenzdichte wird erneut das nächste globale Dichtemaximum gesucht, an dieser Stelle ein Gaußhut platziert und durch räumliche Verschiebung an die Differenzdichte angepasst. Die relative Höhe wird an den Wert der Wahrscheinlichkeitsdichte angepasst, die Breite des Gaußhutes entspricht wie oben beschrieben dem zuvor definierten Wert. Wiederum wird eine Differenzdichte zu der im vorherigen Schritt verbleibenden Dichte gebildet und ein neuer Gaußhut in das Dichtemaximum gelegt. Dieser rekursive Prozess wird nun solange wiederholt, bis das Dichtemaximum der verbleibenden Differenzdichte unter einem bestimmten, vordefinierten Grenzwert fällt. Der Grenzwert wurde in dieser Arbeit auf den Wahrscheinlichkeitswert $p = 1$ gesetzt; dies entspricht genau dem Niveau, an dem das Auftreten einer bestimmten Kontaktgruppe gleich der Zufallswahrscheinlichkeit wird. In einem letzten Nachoptimierungsschritt werden alle zugeordneten Gaußhüte in ihrer Gesamtheit nochmals an die ursprüngliche Wahrscheinlichkeitsdichte angepasst.

Die kartesischen Koordinaten und die Höhen aller gefundenen Gaußhüte werden zunächst in einer einfachen Textdatei gespeichert. Um diese Informationen relativ einfach in den bestehenden Programm-Code von CoMSIA zu integrieren und einen schnellen Zugriff für die nachfolgenden Berechnungen der Deskriptoren zu ermöglichen, werden die kartesischen Koordinaten in interne Koordinaten umgewandelt. Somit ist die Lokalisierung jedes einzelnen Gaußhutes zu drei definierten Atomen des dazugehörigen Zentralfragments durch Abstand, Winkel und Torsionswinkel möglich. Die unterschiedliche Wichtung der Gaußhüte untereinander erfolgt über ihre relativen Höhenwerte.

Bei der Berechnung der Deskriptoren eines gegebenen Moleküls wird dieses zunächst wieder in seine funktionellen Gruppen zerlegt und in der Fragmentdatenbank gesucht. Zu jedem gefundenen Fragment wird der entsprechende Eintrag mit den dazugehörigen Gaußhüten abgerufen, die Zentren der Gaußhüte mittels der zugewiesenen internen Koordinaten platziert und mit den gespeicherten Gewichtungswerten versehen. An jedem Gitterpunkt werden nun die Wahrscheinlichkeitswerte über Gl. 1-4 und Gl. 1-5 berechnet und können anschließend graphisch konturiert werden.

Als Beispiel zeigt Abb. 1-7 die ursprünglichen und angepassten Wahrscheinlichkeitsverteilungen von NH-Kontaktgruppen um eine Carboxylatgruppe als Zentralfragment.

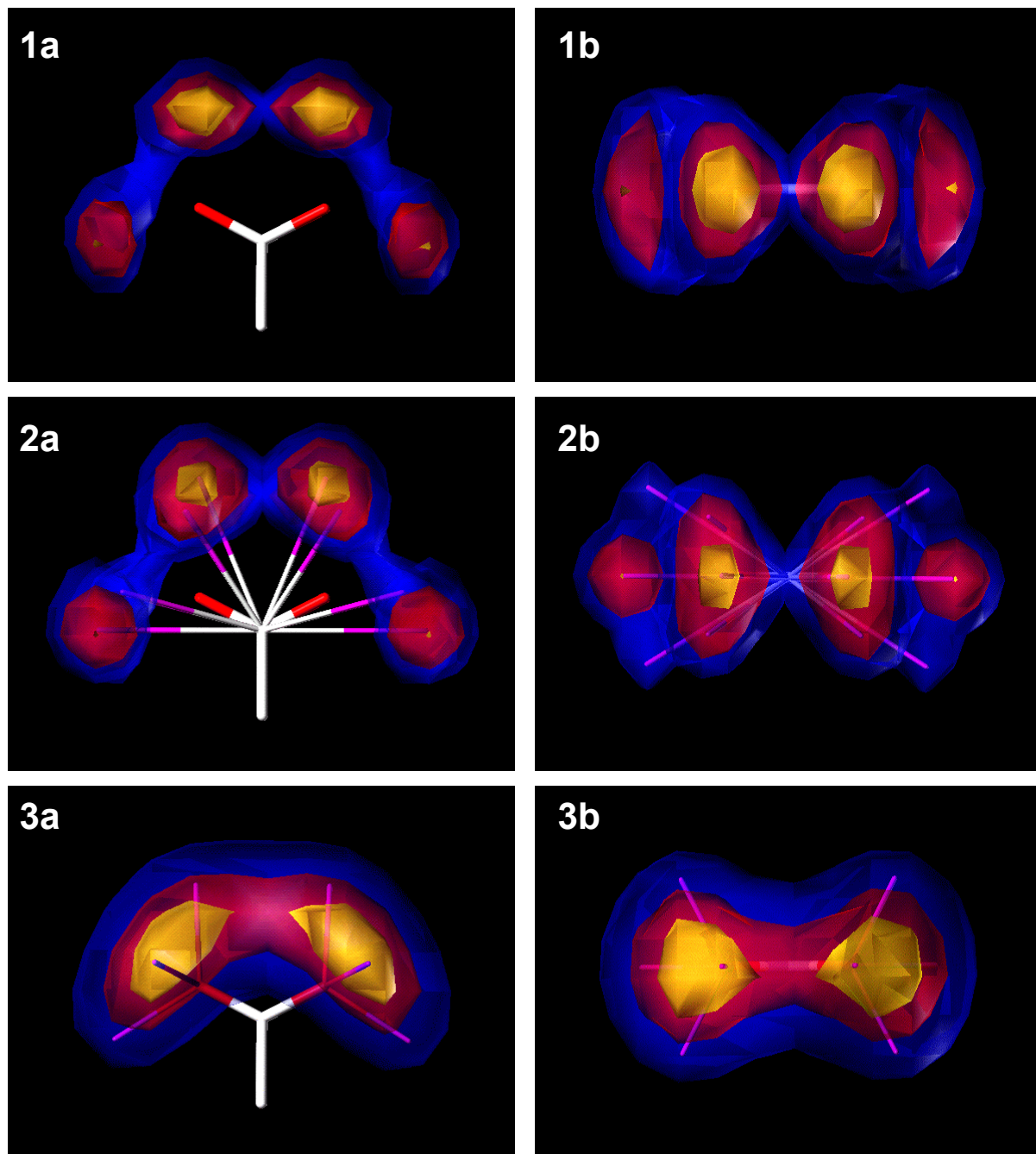


Abb. 1-7: Wahrscheinlichkeitsverteilung von NH-Kontaktgruppen um eine Carboxylatgruppe aus SuperStar (1) und die mit Hilfe von sphärischen Gaußfunktionen angepasste Wahrscheinlichkeitsdichte (2). Die Dichte ist auf drei Niveaus (gelb: 8; rot: 4; blau: 2) konturiert. Die 12 violetten Linien verbinden den Carboxylat-Kohlenstoff mit den Zentren der einzelnen Gaußhüte. Im Vergleich dazu ist die Dichteverteilung angegeben, wie sie das Akzeptorfeld von CoMSIA erzeugt (3). Hier positioniert jedes Sauerstoffatom jeweils vier Pseudoatome, die die Donor-Eigenschaften der Carboxylatgruppe repräsentieren sollen. Die dazugehörigen Dichten sind auch auf drei Niveaus konturiert, die Werte entsprechen jedoch keiner abgestuften Wahrscheinlichkeit. Abbildungen links (a): Frontalansicht; Abbildungen rechts (b): Ansicht von oben (horizontale 90° Drehung).

Die oben beschriebene Approximation der ursprünglichen Wahrscheinlichkeitsdichten aus SuperStar durch mehrere sphärische Gaußfunktionen wurde für alle Fragmente aus Tab. 1-1 durchgeführt. Um diese angepassten Deskriptoren zur Beschreibung der Akzeptor- und Donor-Eigenschaften kleiner Moleküle einsetzen zu können, wurden CO- (*carbonyl oxygen*) bzw. NH- (*amino hydrogen*) Kontaktgruppen zur Berechnung herangezogen.

Eine Abschätzung der Güte der durchgeführten Approximationen kann durch die Berechnung eines sogenannten Ähnlichkeitsindex (*similarity index*) erfolgen. Ein häufig verwendetes Ähnlichkeitsmaß stellt der *Carbo-Index*⁷ dar (Carbo *et al.*, 1980). Ursprünglich wurde er zum Vergleich der Elektronendichten zweier Moleküle herangezogen. Der Carbo-Index S_{AB_Carbo} zweier Verteilungen A und B mit den zugehörigen Wahrscheinlichkeitswerten P^A und P^B (vgl. Gl. 1-5) an den Gitterpunkten i berechnet sich nach Gl. 1-7 wie folgt:

$$S_{AB_Carbo} = \frac{\sum_i P_i^A \cdot P_i^B}{\sqrt{\sum_i (P_i^A)^2 \cdot \sum_i (P_i^B)^2}} \quad \text{Gl. 1-7}$$

Ein Nachteil dieses Index ist, dass unterschiedliche Skalierungen zweier zu vergleichender Verteilungen nicht berücksichtigt werden. Es wird also nur die Form der beiden Dichten in Betracht gezogen. Hodgkin *et al.* entwickelte daher ein Ähnlichkeitsmaß, das auch verschiedene Größen der Verteilungswerte beurteilt (Hodgkin & Richards, 1987). Die mathematische Form des *Hodgkin-Index*⁸ $S_{AB_Hodgkin}$ zweier Verteilungen A und B mit den zugehörigen Wahrscheinlichkeitswerten P^A und P^B an den Gitterpunkten i ist in Gl. 1-8 gegeben.

$$S_{AB_Hodgkin} = \frac{2 \sum_i P_i^A \cdot P_i^B}{\sum_i (P_i^A)^2 + \sum_i (P_i^B)^2} \quad \text{Gl. 1-8}$$

Der Wertebereich des Carbo- und Hodgkin-Index variiert von -1 bis $+1$. Ein Ähnlichkeitsindex von 0 entspricht „keiner Ähnlichkeit“; ein Wert von $+1$ würde eine „perfekte Ähnlichkeit“ bedeuten; eine „perfekte Unähnlichkeit“ (*dissimilarity*) ist bei einem Ähnlichkeitsindex von -1 gegeben.

⁷ andere Bezeichnung: Kosinus-Koeffizient

⁸ andere Bezeichnung: Dice-Koeffizient

Die Güte der Approximation einer gegebenen Wahrscheinlichkeitsdichte durch Gaußhüte ist unter anderem von dem Parameter α abhängig, der die Öffnung der einzelnen Gaußfunktionen festlegt (vgl. Gl. 1-4 und Abb. 1-5, Seite 15). Um eine optimale Anpassung zu erreichen, wurde dieser Wert daher in einem Zahlenbereich von $\alpha = 0,3$ bis 1 variiert und ein jeweiliger Ähnlichkeitsindex berechnet. Hierbei wurde der Hodgkin-Index (Gl. 1-8) verwendet, der (wie weiter oben erläutert) neben der Form auch die möglicherweise unterschiedliche Skalierung der ursprünglichen und angepassten Verteilungen berücksichtigt.

Kleine Zahlenwerte für α (zum Beispiel $\alpha = 0,3$) erzeugen Gaußfunktionen mit einer großen Öffnung; folglich werden während der Approximation nur wenige Gaußhüte in der ursprünglichen Wahrscheinlichkeitsdichte platziert. Daraus resultiert eine – mehr oder weniger stark – vereinfachte angepasste Dichte, die viele Details der ursprünglichen Dichte nicht mehr beinhaltet. Zusätzliche Gaußhüte lassen sich durch eine Erhöhung von α positionieren (zum Beispiel $\alpha = 1$), da die Öffnung der Gaußfunktionen dadurch verringert wird. Es bleibt mehr Information erhalten und die ursprüngliche Verteilung kann somit detaillierter beschrieben werden.

Um dies an einem Beispiel zu erläutern, sind in Abb. 1-8 die approximierten Wahrscheinlichkeitsverteilungen einer CO-Kontaktgruppe um ein geladenes Aminfragment ($R-NH_2^+-R'$) dargestellt, die mit verschiedenen Werten für α berechnet wurden. Höhere Werte von α (geringere Breite) erlauben die Platzierung einer größeren Zahl von Gaußhüten und führen zu einer exakteren Beschreibung der ursprünglichen Dichte.

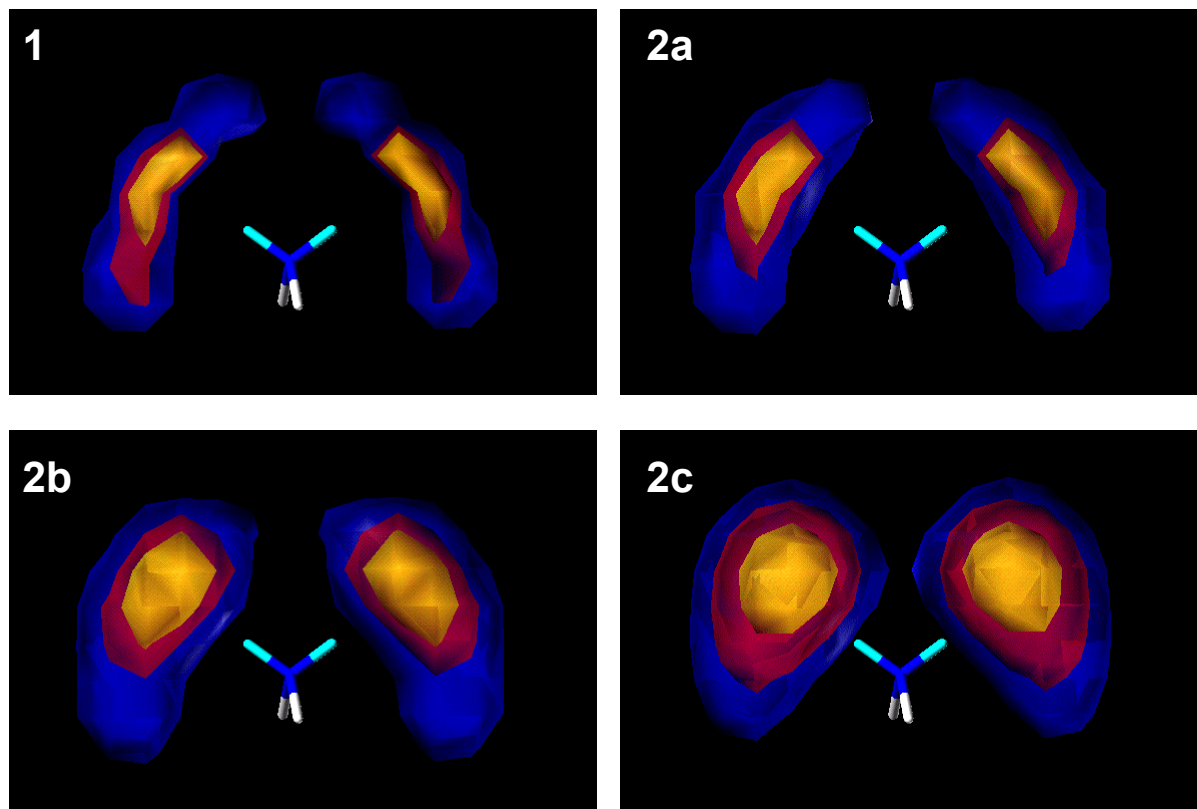


Abb. 1-8: Abhängigkeit der angepassten Wahrscheinlichkeitsverteilungen von der Breite der sphärischen Gaußfunktionen bestimmt durch den Parameter α . Neben der ursprünglichen Dichte aus SuperStar (**1**) wurden die angepassten Dichten einer CO-Kontaktgruppe um ein geladenes Aminfragment für $\alpha = 1,5$ (**2a**), $\alpha = 0,7$ (**2b**) und $\alpha = 0,3$ (**2c**) berechnet. Für die Approximation der jeweiligen Dichten wurden dabei 34 (**2a**), 14 (**2b**) bzw. 6 (**2c**) Gaußhüte verwendet. Die Zentren dieser Gaußhüte sind aus Gründen der Übersichtlichkeit ausgeblendet.

1.4 Ergebnisse und Diskussion

1.4.1 Statistische Auswertung der QSAR-Modelle

Die statistischen Ergebnisse der QSAR-Modelle, die unter Verwendung der Felder aus CoMSIA, GRID und SuperStar aufgestellt wurden, sind in Tab. 1-4 aufgelistet.

Tab. 1-4: Zusammenfassung der Ergebnisse aus den PLS-Analysen unter Verwendung der Felder, die mit CoMSIA, GRID und SuperStar erzeugt wurden.

	CoMSIA	GRID	SuperStar	
			alle Fragmente	keine CH Fragmente
q^2	0,536	0,562	0,573	0,481
S_{press}	1,471	1,416	1,425	1,556
r^2	0,860	0,843	0,951	0,903
S	0,807	0,849	0,482	0,673
F	67,8	75,0	175,0	102,3
Komponenten	5	4	6	5
min-sigma	1,0	1,0 ^(a)	0,5	0,1
predictive r^2	0,308	0,304	0,492	0,535
Gitterweite	1 Å			

(a) Angabe in kcal/mol

Alle berechneten QSAR-Modelle weisen ein hohes Maß an interner Vorhersagekraft auf. Unter Verwendung der fünf in SYBYL standardmäßig implementierten CoMSIA-Felder (sterisch, elektrostatisch, hydrophob, Donor und Akzeptor) ergab sich bei der Kreuzvalidierung unter Verwendung von fünf PLS-Komponenten ein q^2 -Wert von 0,536 und ein konventioneller r^2 -Wert von 0,860. Die Vorhersage der 15 unbekanntten Verbindungen des Testdatensatzes, die nicht in die Ableitung des Modells einfließen, resultierte in einem *predictive* r^2 -Wert von 0,308.

Ähnliche Ergebnisse konnten für das Modell erhalten werden, das mit den fünf Feldern aus GRID erstellt wurde (Tab. 1-3). Hierbei ergab sich ein etwas höherer q^2 -Wert von 0,562 ($r^2 = 0,843$) bei nur vier relevanten PLS-Komponenten. Die externe Vorhersagekraft des GRID-Modells (*predictive* $r^2 = 0,304$) entsprach annähernd der des CoMSIA-Modells.

Im Fall von SuperStar wurden, wie in Kap. 1.3.2 angedeutet, zwei getrennte PLS-Analysen unter Verwendung der sieben Felder durchgeführt, die aus den Kontaktgruppen der Tab. 1-2 (Seite 13) generiert wurden. Die erste Analyse berücksichtigte bei der Feldberechnung alle Zentralfragmente, die in Tab. 1-1 (Seite 9) aufgelistet sind, während eine weitere Analyse diejenigen Fragmente ignorierte, die nur CH-enthaltende Gruppen aufwies (vgl. Tab. 1-1). In beiden Fällen konnten statistisch signifikante Modelle erhalten werden, wobei ersteres Modell einen höheren q^2 -Wert von 0,573 ($r^2 = 0,951$) ergab. Ein deutlicher Abfall in den statistischen Ergebnissen wurde bei letzterem Modell mit einem q^2 -Wert von 0,481 ($r^2 = 0,903$) beobachtet. Offensichtlich ist die Berücksichtigung hydrophober CH-Fragmente (zum Beispiel Methyl oder Phenyl) entscheidend für die Feldberechnung. Dies erscheint im Falle der Verwendung von Kontaktgruppen wie *aliphatic carbon* oder *aromatic carbon* plausibel, denn eine damit verbundene Beschreibung sterischer und hydrophober Eigenschaften kann bei Vernachlässigung hydrophober Zentralfragmente nicht zufriedenstellend erfolgen. Ein gewisser Verlust an Information machte sich auch in der Reduktion der Zahl der PLS-Komponenten von sechs auf fünf bemerkbar. Der *minimum-sigma* Grenzwert, der jedes Mal so eingestellt wurde, dass mindestens 10% der Variablen in die PLS-Analyse einfließen, verringerte sich ebenfalls von 0,5 auf 0,1. Überraschenderweise resultierte aus dem zweiten SuperStar-Modell mit der niedrigeren internen Vorhersagefähigkeit ein höherer *predictive* r^2 -Wert von 0,535 im Vergleich zum ersteren Modell (*predictive* $r^2 = 0,492$). Beide Werte jedoch sind signifikant höher als die der CoMSIA- und GRID-Modelle.

1.4.2 Interkorrelation einzelner Feldbeiträge

Werden mehrere molekulare Felder zur Beschreibung der physikochemischen Eigenschaften von Molekülen herangezogen, so ist die Möglichkeit einer Interkorrelation gegeben. Verglichen mit einem gewöhnlichen CoMFA-Ansatz, bei dem nur zwei Felder – sterisch und elektrostatisch – in das QSAR-Modell eingehen, beschreiben einige der sieben Felder im Falle von SuperStar ähnliche Eigenschaften. Beispielsweise repräsentieren drei der

sieben Kontaktgruppen (*carbonyl oxygen, alcohol oxygen, oxygen atom*, vgl. Tab. 1-2) vorhandene Akzeptor-Eigenschaften eines Liganden und erzeugen daher Felder mit einem gewissen Grad an redundanter Information. Es ist jedoch schwierig, den Grad der Abhängigkeit zwischen den molekularen Feldern zu ermitteln.

In einer früheren Arbeit wurde das Maß der Interkorrelation zwischen den fünf CoMSIA-Feldern für einen Datensatz von Serinprotease-Inhibitoren analysiert (Böhm *et al.*, 1999). Dabei wurden für alle 31 möglichen Kombinationen von fünf Feldern die q^2 -Werte der generierten QSAR-Modelle berechnet. Analog dazu kamen in dieser Arbeit sieben Felder aus SuperStar zum Einsatz. Dazu wurden aus den insgesamt 127 Kombinationsmöglichkeiten 36 Feldkombinationen ausgewählt und damit jeweils ein QSAR-Modell erstellt. Die zugehörigen q^2 -Werte sind in Abb. 1-9 aufgetragen.

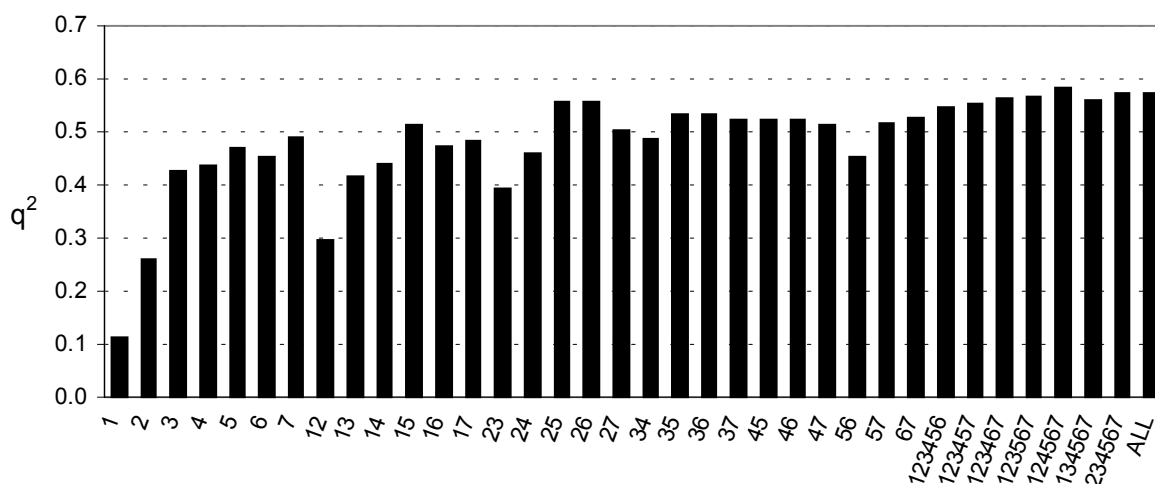


Abb. 1-9: Erhaltene q^2 -Werte für 36 ausgewählte Kombinationen an molekularen Feldern einer SuperStar-Analyse. Die Zahlen an der x-Achse entsprechen der Kombination an Kontaktgruppen, deren Nummern in Tab. 1-2 (Seite 13) aufgelistet sind.

Wie aus Abb. 1-9 ersichtlich wird, variierten die meisten q^2 -Werte im Bereich von 0,4 bis 0,6. Die gleichbleibend hohen Zahlen – auch bei Verwendung von mehr Feldern – deuten auf eine starke Korrelation der Felder untereinander hin. Es konnten sogar statistisch signifikante Modelle mit nur einem einzigen Feld (zum Beispiel durch Kontaktgruppe 5 oder 7; siehe Tab. 1-2) oder zwei Feldern (zum Beispiel durch Kontaktgruppen 2 und 5 oder 2 und 6; siehe Tab. 1-2) erzeugt werden. Die Anwendung einer größeren Zahl von Feldern hatte dabei keine

signifikante Erhöhung der statistischen Ergebnisse zur Folge. So stieg beispielsweise der q^2 -Wert von 0,557 (Kontaktgruppen 2 und 6) auf nur 0,573 (alle sieben Kontaktgruppen) an. Diese Zahl ließ sich sogar noch auf $q^2 = 0,584$ erhöhen, wenn die Kontaktgruppe 3 (*alcohol hydrogen*) bei der Generierung des Modells ignoriert wurde. Auf der anderen Seite zeigten Modelle stark unterschiedliche q^2 -Werte, obwohl die Kontaktgruppen, aus denen die einzelnen Felder erzeugt wurden, ähnliche physikochemische Eigenschaften aufwiesen. So ergab sich zum Beispiel ein statistisch nicht signifikanter q^2 -Wert von 0,26 beim Einsatz der Kontaktgruppe 2 (*amino hydrogen*) und ein wesentlich höherer Wert von 0,426 für Kontaktgruppe 3 (*alcohol hydrogen*). Obwohl beide Kontaktgruppen in ähnlicher Weise die Donor-Eigenschaften eines Liganden beschreiben sollten, resultierten aus der Alkoholgruppe signifikantere Ergebnisse; dies liegt möglicherweise daran, dass die Alkoholgruppe teilweise auch Akzeptor-Eigenschaften enthält. Ebenfalls unterschiedliche Resultate fanden sich für die Akzeptor-Eigenschaften beschreibenden Kontaktgruppen 1 (*carbonyl oxygen*) mit $q^2 = 0,114$ und 4 (*alcohol oxygen*) mit $q^2 = 0,438$. Auch hier kann es wieder ausschlaggebend sein, dass die Alkoholgruppe sowohl Akzeptor- als auch Donor-Eigenschaften repräsentiert, während die Carbonylgruppe nur als Akzeptorgruppe fungiert.

In einer kürzlich erschienenen Publikation selektierten Verdonk *et al.* aus den in der IsoStar-Datenbank enthaltenen Kontaktgruppen eine nicht-redundante Menge von 16 Kontaktgruppen mit Hilfe von *Clustering*-Methoden aus (Verdonk *et al.*, 2001). Die dabei zugrundeliegenden – ähnlich den hier angewandten – Kriterien basierten auf einem Ähnlichkeitsvergleich der einzelnen *scatterplots* in IsoStar und der Verfügbarkeit der Daten. Dabei wurden annähernd die gleichen Kontaktgruppen selektiert, die auch in der vorliegenden Arbeit (Tab. 1-2) Berücksichtigung fanden. Anstelle der Kontaktgruppe „*aliphatic carbon*“ (C-Atom einer aliphatischen CH-Gruppe) verwendeten die Autoren die von ihren physikochemischen Eigenschaften vergleichbare Methylgruppe. Die Übereinstimmung der Auswahl an Kontaktgruppen unterstützt die in dieser Arbeit getroffenen Selektionskriterien hinsichtlich einer relativ geringen Redundanz an Information und der statistischen Signifikanz der *scatterplots* durch die zugrundeliegende Datenlage. Darüber hinaus wählten Verdonk *et al.* weitere Kontaktgruppen aus, deren Verfügbarkeit an Daten jedoch nur eine Analyse von Proteinbindetaschen (bzw. deren 20 Aminosäuren) zulässt. Eine Verwendung dieser Gruppen zur Charakterisierung kleiner Moleküle erscheint angesichts einer wesentlich größeren Vielfalt an funktionellen Gruppen nicht sinnvoll.

1.4.3 Charakterisierung von Wasserstoffbrücken-Eigenschaften

Nach der Validierung der SuperStar-Felder in Kap. 1.4.1 mit allen aufgeführten Kontaktgruppen konzentrierte sich die weitere Arbeit auf die Charakterisierung der Wasserstoffbrücken-Eigenschaften. Es wurden dazu „gemischte“ QSAR-Modelle erstellt, die zum einen die sterischen, elektrostatischen und hydrophoben Eigenschaften der CoMSIA-Felder und zum anderen die Donor- und Akzeptorfelder aus SuperStar beinhalteten. Die beiden letzteren wurden mit Hilfe der NH- und CO-Kontaktgruppen aus Tab. 1-2 erzeugt. Die statistischen Ergebnisse der Modelle sind in Tab. 1-5 („SuperStar Felder“) zusammengefasst.

Tab. 1-5: Zusammenfassung der Ergebnisse aus den PLS-Analysen bei Verwendung der SuperStar-Felder zur Beschreibung der Donor- und Akzeptor-Eigenschaften und der sterischen, elektrostatischen und hydrophoben CoMSIA-Felder. Die Ergebnisse der anisotropen und isotropen Anpassung der SuperStar-Felder sind ebenfalls dargestellt.

	SuperStar Felder		Angepasste Felder (anisotrope Approx.)		Angepasste Felder (isotrope Approx.)	
	alle Fragmente	keine CH Fragmente	alle Fragmente	keine CH Fragmente	$\alpha = 0,3$	$\alpha = 0,7$
q^2	0,592	0,615	0,662	0,667	0,632	0,670
S_{press}	1,393	1,352	1,267	1,258	1,335	1,263
r^2	0,918	0,877	0,958	0,952	0,936	0,956
S	0,626	0,766	0,447	0,479	0,555	0,460
F	100,2	63,9	205,0	177,4	111,4	165,5
Komponenten	6	6	6	6	7	7
min-sigma	0,5	0,5	0,5	0,5	0,5	0,5
predictive r^2	0,548	0,448	0,385	0,431	0,449	0,420
Gitterweite	1 Å					

Die „gemischten“ Modelle zeigten statistisch signifikante Korrelationen mit q^2 -Werten $> 0,5$. Die Verwendung aller Zentralfragmente aus Tab. 1-1 (Seite 9) zur Beschreibung der Moleküleigenschaften resultierte in einem q^2 -Wert von 0,592. Ein leicht verbessertes Ergebnis ($q^2 = 0,615$) konnte erzielt werden, wenn die Fragmente der CH-enthaltenden Gruppen (Tab. 1-1) während der Feldberechnung nicht miteinbezogen wurden. Der konventionelle r^2 -Wert

fiel hierbei jedoch von 0,918 auf 0,877 zurück. Im Vergleich zu den Ergebnissen aus den PLS-Analysen in Tab. 1-4 (Seite 30) konnte mit den gemischten Modellen eine signifikante Verbesserung erreicht werden. So erhöhte sich der q^2 -Wert von 0,536 (CoMSIA) bzw. 0,573 (SuperStar) auf 0,592 (CoMSIA/SuperStar) unter Berücksichtigung aller Zentralfragmente. Die Kombination beider Verfahren scheint daher geeignet zur besseren Charakterisierung physikochemischen Eigenschaften kleiner Moleküle. Die sterischen, elektrostatischen und hydrophoben Felder sollten also idealerweise durch ursprüngliche CoMSIA-Felder repräsentiert werden, wohingegen die Beschreibung der Wasserstoffbrücken-Eigenschaften durch SuperStar vorteilhafter erscheint. Dieses Argument wird zusätzlich durch die Tatsache erhärtet, dass der q^2 -Wert sogar nochmals verbessert wurde, wenn anstelle der fünf CoMSIA-Felder ($q^2 = 0,536$) nur das sterische, elektrostatische und hydrophobe Feld zum Erstellen des Modells herangezogen wurde ($q^2 = 0,578$; in Tabelle nicht aufgeführt). Die beiden Wasserstoffbrücken-Felder erniedrigten also in diesem Fall die statistischen Ergebnisse. Dahingegen konnte der q^2 -Wert bei einer Erweiterung um das Donor- und Akzeptorfeld aus SuperStar von 0,578 auf 0,592 gesteigert werden.

Die Unterschiede in den Resultaten fielen nochmals deutlicher aus, wenn die Fragmente der CH-enhaltenden Gruppen während der Feldberechnung unberücksichtigt blieben. Während das Einbeziehen dieser Fragmente mit ihrem sterisch-hydrophoben Charakter entscheidend bei der Erstellung eines statistisch signifikanten Modells waren, wenn alleine die SuperStar-Felder zum Einsatz kamen (vgl. Tab. 1-4), schienen sie zur Beschreibung der Wasserstoffbrücken-Eigenschaften nicht essenziell zu sein. Im Gegenteil, der q^2 -Wert verbesserte sich bei ihrem Ausschluss sogar von 0,592 auf 0,615. Die optimale Anzahl der Komponenten (sechs) und der *minimum-sigma* Grenzwert (0,5) veränderte sich im Gegensatz zum reinen SuperStar-Modell in Tab. 1-4 nicht. Dies lässt darauf schließen, dass die CH-Fragmente im ersteren Fall (nur SuperStar-Felder) wichtig für die Charakterisierung der sterischen und hydrophoben Eigenschaften sind, wohingegen diese Aufgabe bei den gemischten Modellen explizit von dem sterischen und hydrophoben CoMSIA-Feld übernommen wird. Die Fragmente der CH-enhaltenden Gruppen besitzen zwar eine gewisse Fähigkeit zur Ausbildung von Wasserstoffbrücken, ihr Beitrag zu den Donor-Eigenschaften ist aber relativ gering. So liegen die maximalen Wahrscheinlichkeitswerte der dazugehörigen Dichteverteilungen in allen Fällen bei $p_{max} < 3$ (vgl. Tab. 1-1). Die externe Vorhersagekraft

wurde durch den Einsatz der „gemischten“ CoMSIA/SuperStar-Modelle mit *predictive* r^2 -Werten von 0,548 (alle Fragmente) bzw. 0,448 (keine CH-Fragmente) ebenfalls verbessert.

1.4.4 Approximation der Felder aus SuperStar

Die statistischen Ergebnisse der QSAR-Modelle konnten nochmals verbessert werden, wenn zur Beschreibung der Wasserstoffbrücken-Eigenschaften anstelle der ursprünglichen SuperStar-Felder die angepassten Wahrscheinlichkeitsdichten (siehe Kap. 1.3.4 und 1.3.5) herangezogen wurden (Tab. 1-5). Im Falle der anisotropen Approximation waren relativ ähnliche q^2 -Werte zu beobachten, wenn alle Zentralfragmente zur Berechnung der molekularen Felder verwendet wurden ($q^2 = 0,662$) oder wenn die CH-Fragmente unberücksichtigt blieben ($q^2 = 0,667$). Offensichtlich spielen diese hydrophoben Gruppen auch bei den approximierten Verteilungen keine signifikante Rolle (vgl. Kap. 1.4.3) und wurden daher im weiteren während der Berechnung der Wasserstoffbrücken-Felder ignoriert. Die Anpassung der SuperStar-Felder durch einen Satz kugelförmiger Gaußfunktionen (isotrope Approximation) resultierte in Modellen vergleichbarer Güte zu denen der anisotropen Annäherung. Beispielsweise ergaben sich q^2 -Werte von 0,632 und 0,670, wenn der Parameter α (Öffnung der Gaußhüte) auf 0,3 bzw. 0,7 festgelegt wurde (vgl. Gl. 1-4). Da die statistischen Ergebnisse signifikant von dem Wert der variablen Größe α abhängen, wurde der Einfluss dieses Parameters weiter im Detail untersucht.

Die Abhängigkeit der angepassten Wahrscheinlichkeitsdichten von der Öffnung der sphärischen Gaußfunktionen – bestimmt durch den Parameter α – ist in Abb. 1-8 dargestellt. Am Beispiel der Verteilung einer CO-Kontaktgruppe um ein geladenes Aminfragment wurden die approximierten Dichten für unterschiedliche Werte von α berechnet und konturiert. Wenn α größere Werte annahm, konnte eine höhere Zahl von Gaußhüten während der Optimierung in die ursprüngliche Dichte platziert werden. Dies führte zu einer exakteren Approximation der ursprünglichen SuperStar-Felder und konnte durch die Berechnung des Hodgkin-Ähnlichkeitsindizes (HI, Gl. 1-8) quantitativ bestimmt werden. So erhöhte sich der Wert von HI von 0,44 ($\alpha = 0,3$) und 0,76 ($\alpha = 0,7$) auf 0,95 ($\alpha = 1,5$), wobei während der Optimierung 6, 14 bzw. 34 Gaußhüte in der ursprünglichen Dichte positioniert wurden. Da die Verteilungen

in Abb. 1-8 zwei räumlich voneinander getrennte Wahrscheinlichkeitsbereiche aufweisen, wurde also jede isolierte Dichte durch 3, 7 bzw. 17 Gaußhüte beschrieben.

Um die Qualität der in dieser Arbeit berechneten, approximierten Dichten abzuschätzen, wurde der Hodgkin-Index für die Wahrscheinlichkeitsverteilungen aller Zentralfragmente aus Tab. 1-1 (keine CH-Fragmente) berechnet und ein Mittelwert bestimmt. Neben der Abhängigkeit des Parameters α wurde auch der Einfluss der Gitterweite bei der Berechnung der Wahrscheinlichkeitsdichten aus den *scatterplots* ermittelt (siehe Kap. 1.3.5). Abb. 1-10 zeigt die Abhängigkeit des Hodgkin-Indexes von der Variation des Parameters α zwischen 0,3 und 1 und des Gitterabstands zwischen 0,6 Å und 1 Å.

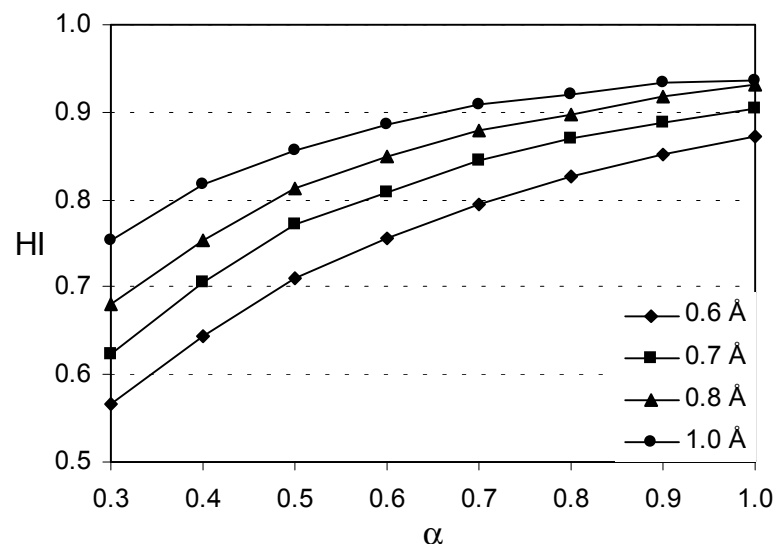


Abb. 1-10: Abhängigkeit des mittleren Hodgkin-Index (HI) von dem Parameter α (Breite der Gaußhüte) und der Gitterweite, die während der Berechnung der Wahrscheinlichkeitsdichten aus den *scatterplots* vorgegeben ist. Der Ähnlichkeitsindex steigt mit größeren Werten von α (von $\alpha = 0,3$ bis 1) und größeren Gitterabständen (von 0,6 bis 1 Å). Die angegebenen Zahlenwerte für HI entsprechen jeweils den Mittelwerten der Hodgkin-Indexes, für die angepasste Wahrscheinlichkeitsdichten berechnet wurden.

Aus Abb. 1-10 lässt sich erkennen, dass der Mittelwert aller berechneten Hodgkin-Indexes mit größerem Wert für α und dem Gitterabstand steigt. So verbessert sich zum Beispiel der Ähnlichkeitsindex HI von 0,62 ($\alpha = 0,3$) auf 0,91 ($\alpha = 1$) bei einer Gitterweite von 0,7 Å. Die Erhöhung des mittleren Hodgkin-Index bei größerem α ergibt sich – wie weiter oben schon

erläutert – aus der Platzierung einer größeren Zahl von Gaußhüten während der Optimierung. Der Abfall des Ähnlichkeitsindex bei einer Verkleinerung des Gitterabstands lässt sich dadurch erklären, dass bei einer kleineren Maschenweite mehr Datenpunkte zur Berechnung von HI zur Verfügung stehen und dadurch auf dem Gitter wesentlich mehr Details der Wahrscheinlichkeitsverteilung abgelegt werden können. Um diese zusätzliche Information bei der Approximation jedoch zu berücksichtigen, müssten wiederum mehr Gaußhüte in der Dichte positioniert werden. So genügt beispielsweise ein Wert von $\alpha = 0,7$ bei einem Gitterabstand von 1 \AA , um einen mittleren Hodgkin-Index $> 0,9$ zu erreichen. Bei einer kleineren Maschenweite von $0,7 \text{ \AA}$ – dies entspricht im Vergleich zu 1 \AA einer ca. 3-fach höheren Menge an Gitterpunkten⁹ – muss auf $\alpha = 1$ erhöht werden, um einen vergleichbar hohen mittleren Ähnlichkeitsindex zu erhalten.

In Anbetracht der obigen Ergebnisse erscheint es daher sinnvoll, einen relativ hohen Zahlenwert für α zu setzen, um die ursprünglichen Wahrscheinlichkeitsdichten aus SuperStar möglichst präzise zu approximieren. Daher wurden in einer weiteren Analyse mehrere QSAR-Modelle für verschiedene Werte von α generiert und die jeweiligen q^2 -Werte graphisch aufgetragen. Aus Abb. 1-11 lässt sich keine signifikante Abhängigkeit der q^2 -Werte aus dem Parameter α ableiten. Im Vergleich zum Verhalten des Hodgkin-Index (Abb. 1-10) erscheint eine Verbesserung der statistischen Werte bei einer Erhöhung von α nicht möglich. Im Gegenteil – QSAR-Modelle mit den höchsten q^2 -Werten konnten für $\alpha \approx 0,5$ erhalten werden, auch wenn die Unterschiede in den Zahlenwerten nicht signifikant waren. Eine Erklärung für den leichten Abfall der q^2 -Werte bei höherem α ist nicht trivial, spielen doch eine Vielzahl von Parametern in der PLS-Analyse eine Rolle. So wurde bei der Berechnung der molekularen Felder ein Gitter mit einer Maschenweite von 1 \AA verwendet. Das Ablegen sämtlicher Informationen der Felder ist auf diese Gitterpunkte beschränkt. Wird nun der α -Wert erhöht, so werden anfangs zwar mehr Gaußhüte positioniert; die dazugehörigen Werte der Wahrscheinlichkeitsdichte können aber nur an den Gitterpunkten in 1 \AA Abstand gespeichert werden.

⁹ Berechnung: $(1 \text{ \AA} / 0,7 \text{ \AA})^3 \approx 2,92$

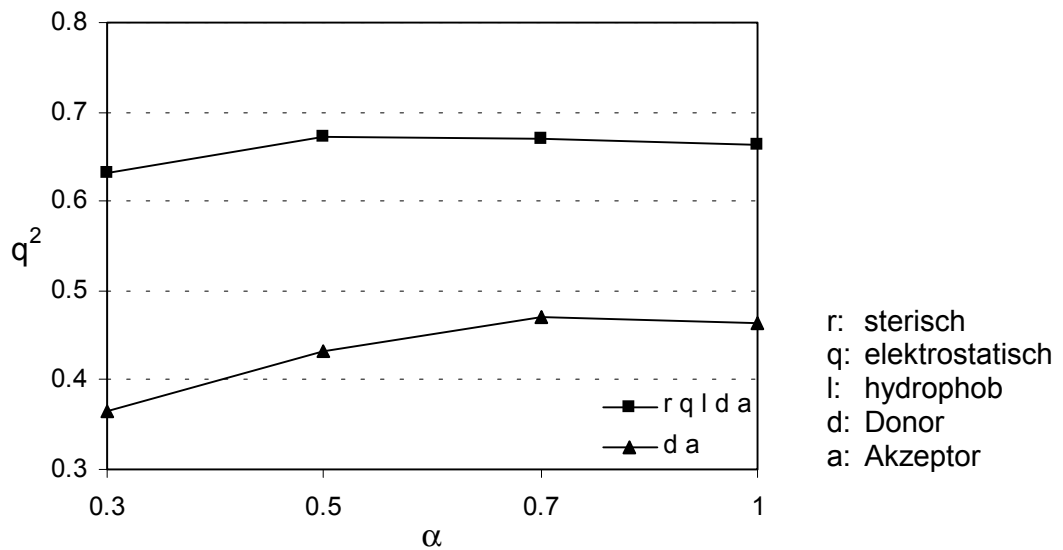


Abb. 1-11: Erhaltene q^2 -Werte für unterschiedliche Werte von α bei der Berechnung der isotrop approximierten Wasserstoffbrücken-Felder. Wenn das Modell nur durch Donor- und Akzeptorfelder (d, a) beschrieben wird, resultieren q^2 -Werte $< 0,5$ mit einem Maximum von 0,47 bei $\alpha = 0,7$. Fließen die Informationen aller fünf Felder (r, q, l, d, a) in die Analyse ein, ergeben sich q^2 -Werte zwischen 0,6 und 0,7. Die beste Wahl für α liegt in diesem Fall zwischen 0,5 und 0,7 mit einem maximalen q^2 -Wert von ca. 0,67. Während der Konvertierung der Wahrscheinlichkeitsdichten für die Wasserstoffbrücken-Felder wurde ein Gitterabstand von 0,7 Å gewählt.

Ein Ansatz, um mehr Information aus den vorgegebenen Dichten zu extrahieren, wäre daher eine Verkleinerung der Gitterweite bei der Berechnung der molekularen Felder. Die Änderung des Gitterabstands von 1 Å auf zum Beispiel 0,7 Å würde die Zahl an Gitterpunkten zwar um den Faktor 3 (siehe oben) erhöhen; diese Datenmenge ist aber mit einem erheblichen Rechenaufwand während der nachfolgenden PLS-Analyse verbunden (üblicherweise werden molekulare Felder in CoMFA auf einem Gitter mit 2 Å Maschenweite berechnet). In diesem Fall müssten die statistischen Auswertungen mit Methoden der Variablenfilterung und -selektion (zum Beispiel mit dem Programm GOLPE; siehe Anhang) verknüpft werden.

Die Beschränkung der Datengewinnung aufgrund des gewählten Gitterabstands zur Bestimmung der molekularen Felder erklärt auch die Verbesserung der Modelle, wenn der Gitterabstand während der Konvertierung der Dichten aus den *scatterplots* von 0,7 auf 1 Å vergrößert wurde. So erhöhte sich für $\alpha = 0,7$ der q^2 -Wert – jedoch nicht signifikant – von

0,46 auf 0,52 (nur Donor- und Akzeptorfeld) bzw. von 0,67 auf 0,71 (alle fünf Felder). Für die folgende graphische Auswertung der Konturdiagramme wurde auf das Modell zurückgegriffen, das mit $\alpha = 0,7$ und einer Gitterweite von 1 Å während der Konvertierung der Dichten generiert wurde ($q^2 = 0,71$).

Wie in Kap. 1.3.4 schon erwähnt, ist die schnelle Berechnung der molekularen Feldern aus den angenäherten SuperStar-Wahrscheinlichkeitsdichten von entscheidendem Vorteil. Nissink *et al.* konnten mit Hilfe der anisotrop approximierten SuperStar-Felder die durchschnittliche Rechenzeit zur Charakterisierung einer mittleren Proteinbindetasche um den Faktor fünf reduzieren (Nissink *et al.*, 2000). In dieser Arbeit beanspruchte beispielsweise die Bestimmung der Wasserstoffbrücken-Felder (CO- und NH-Kontaktgruppe) für 76 Thermolysin-Inhibitoren mit den ursprünglichen SuperStar-Deskriptoren ca. 30 min. Rechenzeit. Diese konnte sowohl bei Verwendung der anisotrop als auch der isotrop angepassten SuperStar-Felder auf sieben bis acht Minuten verringert werden, was einer Reduktion der Rechenzeit um ungefähr den Faktor vier entspricht.

1.4.5 Graphische Auswertung der Konturdiagramme

Die Interpretation der erhaltenen Modelle erfolgte über eine graphische Konturierung der Koeffizienten aus der PLS-Analyse. In dieser Arbeit wurden dazu die häufig verwendeten, standardisierten Koeffizienten $stdev*coeff$ berechnet (Cramer *et al.*, 1993). Die Feldbeiträge ergeben sich dabei aus dem Produkt der einzelnen Koeffizienten, die aus der PLS-Analyse erhalten wurden, mit der Standardabweichung der Funktionswerte aller Moleküle an einem Gitterpunkt. Der Vorteil dieser Beschreibung im Vergleich zu der Darstellung der reinen Koeffizienten ($coeff$) ist eine Abschwächung relativ großer Werte an den Schnittpunkten, die mit einer nur kleinen Standardabweichung verknüpft sind. Dadurch werden zahlenmäßig große Werte herunterskaliert, die bei vielen Molekülen des Datensatzes annähernd gleich und somit für die Varianz uninteressant sind, und die wesentlichen Effekte können deutlicher hervorgehoben werden. Die Bestimmung geeigneter Niveaus für die Konturierung erfolgte iterativ durch die Auswertung der Histogramme, die eine Verteilung der absoluten Zahlenwerte der standardisierten Koeffizienten angeben. Die einzelnen Felder wurden jeweils als Bereiche mit zwei unterschiedlichen Konturniveaus verschiedenfarbig dargestellt. Sie

zeigen Bereiche an, in denen Änderungen der Feldbeiträge parallel oder gegenläufig zu den Affinitätsänderungen laufen.

Die graphische Interpretation der Feldbeiträge erfolgte am Beispiel der Wasserstoffbrücken-Eigenschaften, die aus den in dieser Arbeit entwickelten SuperStar-Deskriptoren erhalten wurden. Dabei flossen in diesem Fall die Donor- und Akzeptorfelder der isotrop approximierten Wahrscheinlichkeitsdichten (vgl. Kap. 1.3.5 und 1.4.4) in die aufgestellten Modelle ein (Gitterweite 1 Å, $\alpha = 0,7$). Ein Vergleich der Konturdiagramme, die aus den ursprünglichen SuperStar-Feldern oder aus den Feldern der anisotrop angenäherten Dichten resultierten, zeigte qualitativ ähnliche Ergebnisse, so dass hier auf ihre weitere Auswertung verzichtet wurde. Die ursprünglichen Wasserstoffbrücken-Deskriptoren aus CoMSIA dienten bei der Interpretation als Referenz- und Vergleichsbasis.

In Abb. 1-12 sind die Konturdiagramme der Akzeptor-Eigenschaften aus der CoMSIA- und CoMSIA/SuperStar-Analyse dargestellt. Farblich gekennzeichnete Regionen innerhalb der Proteinbindetasche markieren Bereiche, in denen mögliche Wasserstoffbrücken-Bindungspartner des Enzyms mit funktionellen Gruppen des Liganden interagieren können. Wie schon in Kap. 1.3.3 (Feld \leftrightarrow Eigenschaft) darauf hingewiesen wurde, ergeben sich die aus dem Modell resultierenden Akzeptor-Eigenschaften des Enzyms aus der vergleichenden Korrelationsanalyse der (aus den Donorgruppen abgeleiteten) Donorfelder der Thermolysin-Liganden.

Auf den ersten Blick zeigen die Konturdiagramme, die aus den Feldern von CoMSIA (Abb. 1-12, oben) und SuperStar (Abb. 1-12, unten) resultieren, qualitativ ähnliche Ergebnisse. Beispielsweise werden die drei rot gefärbten Regionen sowohl bei der Korrelation der CoMSIA- als auch der SuperStar-Felder in den Konturdiagrammen angezeigt. Zwei dieser Konturen, die eine Steigerung der Bindungsaffinität bei Anwesenheit einer Akzeptorgruppe aus dem Protein erwarten lässt, umgeben jeweils ein Carbonyl-Sauerstoffatom der terminalen Carbamidgruppe von Asn 111 und Asn 112. Dies stimmt mit der Beobachtung überein, dass diese Aminosäuren oftmals als Akzeptorgruppen auftreten und Wechselwirkungen mit Wasserstoffbrücken-Donoren potenzieller Thermolysin-Inhibitoren eingehen. Der dritte, rot konturierte Bereich, der ebenfalls als günstig für Akzeptorgruppen des Enzyms vorhergesagt wird, befindet sich in der Nähe der OH-Gruppe von Tyr 157.

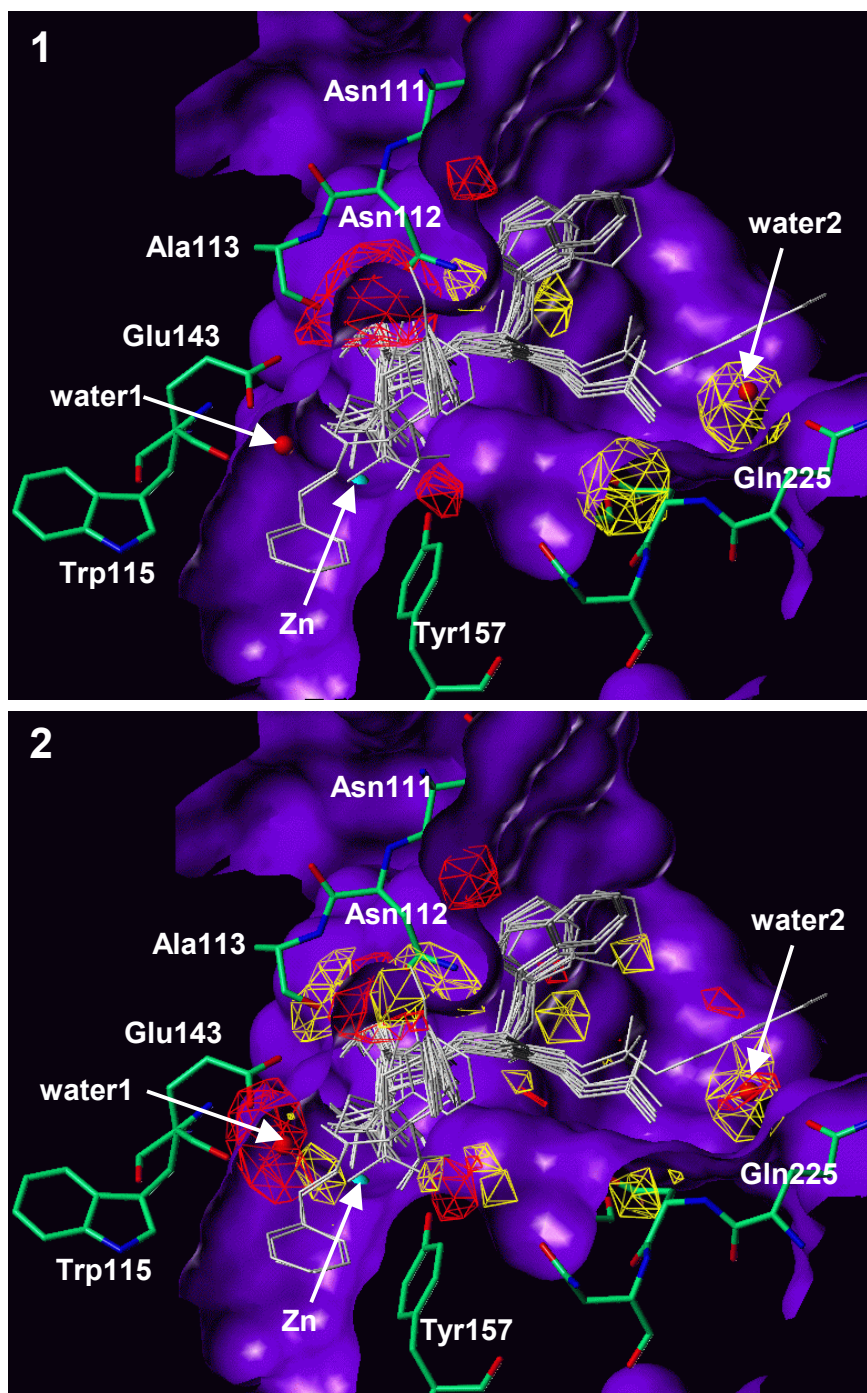


Abb. 1-12: Konturdiagramme der Akzeptor-Eigenschaften berechnet aus $stdev \cdot coeff$ Feldbeiträgen. Wichtige Seitenketten (grün) und ein Teil der überlagerten Liganden (weiß) sind zusammen mit der lösungsmittelzugänglichen Oberfläche der Thermolysin-Bindetasche abgebildet. Strukturell bedeutende Wassermoleküle (rot) und Zn (cyan) sind als Kugeln dargestellt. **1:** Modell mit ursprünglichem CoMSIA-Donorfeld; **2:** Modell mit isotrop angenähertem SuperStar-Feld (CO-Kontaktgruppe). Für Akzeptorgruppen günstige Bereiche des Proteins sind rot markiert (Konturniveau **1:** +0,001; **2:** +0,015). Keine Steigerung der Bindungsaffinität wird für funktionelle Gruppen mit Akzeptor-Eigenschaften in gelb gefärbten Bereichen (Konturniveau **1:** -0,005; **2:** -0,007) erwartet.

In dem Diagramm der SuperStar-Felder deutet eine zusätzliche rote Kontur in der Nähe eines strukturell bedeutenden Wassermoleküls (Abb. 1-12, *water1*) auf weitere Akzeptor-Eigenschaften hin. Dieses Wasser kann daher – als Akzeptorgruppe fungierend – eine Wasserstoffbrücke zwischen Thermolysin-Liganden und den funktionellen Gruppen der Hauptkette von Trp 115 vermitteln.

In gelb markierten Bereichen der Konturdiagramme kann sich die Anwesenheit von Akzeptorgruppen aus dem Protein negativ auf die zu erwartende Bindungsaffinität auswirken. In beiden Diagrammen wird die terminale Aminogruppe der Seitenkette im Asn 112 von einer der beiden gelben Konturen umgeben. Die Vorhersage des Modells, wonach eine Akzeptorgruppe des Enzyms in diesem Bereich zur Affinitätssteigerung fehlen sollte, stimmt mit der Tatsache überein, dass die NH₂-Gruppe der Aminosäure Asn 112 ausschließlich Donor-Eigenschaften aufweist. Es sei jedoch darauf hingewiesen, dass die Konturierung der Akzeptor-Eigenschaften keine Aussagen darüber zulässt, ob sich Donorgruppen aus dem Protein günstig oder ungünstig auf die Affinität auswirken. Aus dem Hinweis, dass ein Fehlen einer Akzeptorgruppe günstig sein soll, lässt sich also nicht zwingend schließen, dass sich umgekehrt die Anwesenheit einer Donorgruppe günstig auf die zu erwartende Bindungsaktivität auswirkt.

Im Unterschied zum CoMSIA-Konturdiagramm markiert das SuperStar-Modell zusätzlich einen Bereich in der Nähe der Hauptkettenatome von Ala 113, die eine Anwesenheit einer Protein-Akzeptorgruppe in dieser Region zur Steigerung der Bindungsaffinität verbietet. Dies steht im Einklang damit, dass in der Nähe dieser Markierung die NH-Funktion der Hauptketten-Amidgruppe von Ala 113 lokalisiert ist. Die Carbonylgruppe der gleichen Aminosäure befindet sich allerdings ebenfalls ganz in der Nähe der gelb gefärbten Kontur.

Eine weiterer, gelb gefärbter Bereich umgibt ein Wassermolekül (Abb. 1-12, *water2*) und kann ebenfalls mit einer vom Modell geforderten Präsenz funktioneller Gruppen korreliert werden. Dieses zweite, strukturell wichtige Wassermolekül vermittelt eine Wasserstoffbrücke zwischen Liganden und dem Carbamid-Sauerstoffatom von Gln 225. Das zugrundeliegende Modell sagt voraus, dass dieses Wassermolekül zur Affinitätssteigerung eines gebundenen Thermolysin-Liganden mit diesem nicht über seine Akzeptorfunktion in Wechselwirkung treten sollte.

Die Konturdiagramme der Donor-Eigenschaften aus der CoMSIA- und CoMSIA/SuperStar-Analyse zeigt Abb. 1-13. Wie schon bei den Akzeptor-Eigenschaften erläutert, ergeben sich die aus dem Modell resultierenden Donor-Eigenschaften des Proteins aus der Korrelation der in den Thermolysin-Inhibitoren vorhandenen Akzeptorgruppen (Akzeptorfeld).

Wiederum konnten qualitativ ähnliche, graphische Ergebnisse aus dem CoMSIA-Modell (Abb. 1-13, oben) auch mit dem SuperStar-Modell (Abb. 1-13, unten) erhalten werden. Drei wesentliche, in blau angedeutete Konturen zeigen an, dass in diesen Regionen Donorgruppen aus dem Enzym vorteilhaft für die Affinitätssteigerung der Inhibitoren sind. In Übereinstimmung mit der Proteinumgebung befinden sich diese markierten Bereiche in der Nähe funktioneller Proteingruppen mit Donor-Eigenschaften. Es sind dies NH-Gruppen der Hauptketten von Asn 112, Trp 115 und Arg 203 (letztere in Abb. 1-13 nicht sichtbar). Eine zusätzliche, blau markierte Kontur bei den SuperStar-Eigenschaften kann in diesem Fall den funktionellen Gruppen des Proteins nicht zugeordnet werden.

Die gelb gefärbten Konturen in der Nähe des katalytisch wirksamen Zink-Ions markieren eine Region, in der die Anwesenheit von Donorgruppen des Enzyms nicht von Vorteil wäre. Dieser relativ große Bereich entsteht durch die Korrelation unterschiedlicher zink-bindender Gruppen, die in den Inhibitoren des Thermolysin-Datensatzes vorhanden sind. Da diese Metall-koordinierenden Ligandgruppen normalerweise Sauerstoff- oder Stickstoffatome mit freien Elektronenpaaren aufweisen, erscheint es günstig, wenn in der Nähe des markierten Bereichs Akzeptorgruppen vorhanden wären. Dies muss aber nicht zwingend der Fall sein, da eine durch das Modell geforderte Abwesenheit von Donorgruppen nicht automatisch zur Folge hat, dass die Anwesenheit von Akzeptorgruppen eine Affinitätssteigerung bewirkt.

Eine zusätzliche gelbe Kontur im SuperStar-Diagramm befindet sich in der Nähe eines strukturell wichtigen Wassermoleküls (Abb. 1-13, *water1*). Die daraus folgende Forderung, dass dieses Wassermolekül keine Donor-Eigenschaften besitzen sollte, steht im Einklang mit der Interpretation der Akzeptor-Eigenschaften (Abb. 1-12). Dort wurde für dieses Strukturwasser zur Erhöhung der Ligandaffinitäten eine derartige Orientierung verlangt, dass es seine Akzeptor-Eigenschaften zu den Liganden hin präsentiert.

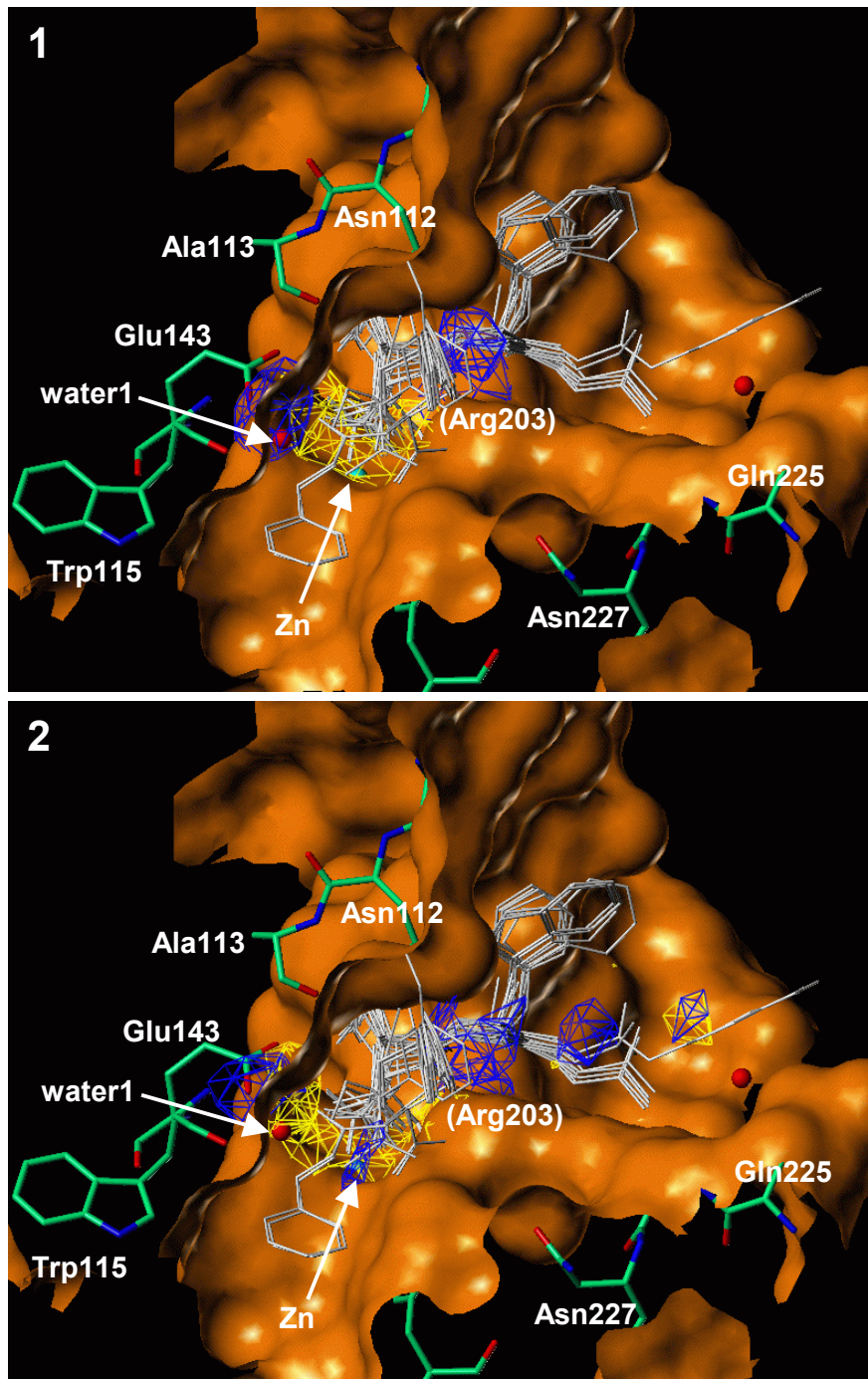


Abb. 1-13: Konturdiagramme der Donor-Eigenschaften berechnet aus $stdev \cdot coeff$ Feldbeiträgen. Wichtige Seitenketten (grün) und ein Teil der überlagerten Liganden (weiß) sind zusammen mit der lösungsmittelzugänglichen Oberfläche der Thermolysin-Bindetasche abgebildet. Strukturell bedeutende Wassermoleküle (rot) und Zn (cyan) sind kugelförmig dargestellt. **1:** Modell mit ursprünglichem CoMSIA-Akzeptorfeld; **2:** Modell mit isotrop angenähertem SuperStar-Feld (NH-Kontaktgruppe). Für Donorgruppen günstige Bereiche des Proteins sind blau markiert (Konturniveau **1:** +0,003; **2:** +0,015). Keine Steigerung der Bindungsaffinität wird für funktionelle Gruppen mit Donor-Eigenschaften in gelb gefärbten Bereichen (Konturniveau **1:** -0,008; **2:** -0,014) erwartet.

Zusammenfassend lässt sich über die graphische Interpretation der CoMSIA- und SuperStar-Modelle sagen, dass sowohl bei den Akzeptor- als auch den Donor-Eigenschaften qualitativ ähnliche Konturdiagramme erzeugt wurden. Aus der exakten Beschreibung der Wasserstoffbrücken-Eigenschaften durch die SuperStar-Deskriptoren ergeben sich dabei aber – verglichen mit den stark vereinfachten CoMSIA-Deskriptoren – detailliertere Feldbeiträge, deren zusätzlicher Informationsgehalt hilfreich für ein verbessertes Design neuer Liganden sein kann.

1.5 Zusammenfassung

In diesem Kapitel wurde die Entwicklung wissensbasierter Deskriptoren zur physikochemischen Charakterisierung kleiner Moleküle vorgestellt. Diese kamen erfolgreich für vergleichende molekulare Feldanalysen zum Einsatz, wobei signifikante Verbesserungen der statistischen Ergebnisse als auch Vorteile bei der graphischen Interpretation resultierten.

Grundsätzlich lassen sich durch eine Auswertung intermolekularer Kontakte von experimentellen Kristallstrukturen beliebige Deskriptoren zur Beschreibung intermolekularer Wechselwirkungen ableiten und für eine vergleichende molekulare Feldanalyse verwenden. In dieser Arbeit wurde eine Gruppe von sieben diversen Kontaktgruppen gewählt, die repräsentativ die wichtigsten physikochemischen Eigenschaften wie zum Beispiel hydrophobe oder Wasserstoffbrücken-Eigenschaften beinhalten sollten.

In einer vergleichenden Analyse konnte dann eine signifikante Verbesserung der Ergebnisse festgestellt werden, wenn zur Beschreibung von Wasserstoffbrücken-Eigenschaften die Wahrscheinlichkeitsverteilungen aus SuperStar herangezogen wurden. Daher wurde ein gemischtes Modell vorgeschlagen, bei dem die Charakterisierung von sterischen, elektrostatischen und hydrophoben Eigenschaften unter Verwendung der CoMSIA-Felder erfolgte, während die Donor- und Akzeptor-Eigenschaften am besten mit Hilfe der SuperStar-Felder repräsentiert werden konnten. Die experimentelle Verteilung von Carbonyl-Kontaktgruppen um entsprechende Zentralfragmente mit einer Donorfunktion erlaubte die Darstellung von Akzeptor-Eigenschaften; die Donor-Eigenschaften konnten mittels der NH-Kontaktgruppe beschrieben werden. Der Einsatz beider Sonden erlaubte einen direkten

Vergleich der ursprünglichen Wasserstoffbrücken-Deskriptoren aus CoMSIA mit denen der durch SuperStar abgeleiteten Deskriptoren anhand der statistischen Ergebnisse und der graphischen Interpretation der jeweils abgeleiteten Modelle.

Die signifikante Verbesserung der statistischen Ergebnisse lässt den Schluss zu, dass die neu entwickelten SuperStar-Deskriptoren zusätzliche Informationen enthalten, die in die Korrelationen der entwickelten QSAR-Modelle einfließen können. Diese detaillierteren Deskriptoren sind auch bei der graphischen Auswertung der Konturdiagramme erkennbar; die im Vergleich zu den CoMSIA-Diagrammen zusätzlich konturierten Bereiche erlauben weitreichendere Aussagen über geforderte Eigenschaften in einer potenziellen Proteinumgebung. Damit lassen sich in einer fortgeschrittenen Phase eines Inhibitor-Designprojektes präzise Schlussfolgerungen darüber ableiten, wie eine vorgegebene Leitstruktur bezüglich ihrer Wasserstoffbrücken-Eigenschaften optimiert werden sollte.

Die direkte Berechnung der SuperStar-Deskriptoren bei der Ermittlung der molekularen Felder großer Datensätze stellte sich als sehr rechenintensiv heraus. Eine deutliche Reduktion der Rechenzeit konnte durch eine Approximation der ursprünglichen SuperStar-Deskriptoren erzielt werden, wobei gleichzeitig auf die Erhaltung möglichst vieler Details geachtet wurde. Überraschenderweise wurden die statistischen Ergebnisse bei Verwendung der angenäherten Deskriptoren sogar verbessert. Ein Grund dafür könnte die Elimination des „Hintergrundrauschens“ sein, das bei direkter Verwendung der experimentellen Information in den SuperStar-Feldern ungefiltert in die Analyse eingeht (Nissink *et al.*, 2000).

2 SELEKTIVITÄTSUNTERSUCHUNGEN AM BEISPIEL VON SERINPROTEASEN

2.1 Problemstellung

Ein wichtiger Aspekt in der heutigen Wirkstoffforschung ist die Entwicklung von Arzneistoffen, die zum Beispiel ihre inhibitorische Aktivität nur selektiv an einem Enzym oder Rezeptor entfalten. Somit lassen sich stets auftretende Nebenwirkungen auf ein geringeres Maß reduzieren. Daher ist es erstrebenswert, im Zyklus der Arzneistoffentwicklung neben einer Optimierung der Bindungsaffinität eines Liganden an einem Rezeptor so früh wie möglich auch Aspekte der Selektivität einfließen zu lassen. In der Vergangenheit wurden selektiv wirksame Verbindungen meist mehr oder weniger zufällig gefunden, nachdem vorher eine Vielzahl von Molekülen synthetisiert, getestet und verworfen werden musste. Im Rahmen eines rationalen Wirkstoffdesigns stehen heutzutage Methoden zur Verfügung, die eine zielgerichtete Entwicklung möglichst selektiv bindender Inhibitoren ermöglichen. So lassen sich beispielsweise bei Kenntnis der dreidimensionalen Struktur konkurrierender Rezeptoren durch einen visuellen Vergleich der dazugehörigen Ligandbindetaschen Bereiche auffinden, die nur bei einem der Rezeptoren vorhanden oder besonders gut ausgeprägt sind. So konnte zum Beispiel die Selektivität von Thrombin-Inhibitoren gegenüber Trypsin (beide Enzyme gehören zur Klasse der Serinproteasen; siehe weiter unten) durch eine geeignete Auswahl an Substituenten erhöht werden, die möglichst komplementär die vorhandenen Bindetaschenbereiche von Thrombin ausfüllen (Böhm *et al.*, 1996).

Eine interessante Möglichkeit der Selektivitätsuntersuchung besteht darin, die Affinitäten von Inhibitoren zu mehreren, strukturell verwandten Proteinen untereinander zu vergleichen. Mit Hilfe von 3D-QSAR-Verfahren lassen sich Informationen aus einer größeren Anzahl von Liganden, die an verschiedene Rezeptoren binden und deren experimentelle Affinitäten bekannt sind, extrahieren. Unterschiedliche Verfahren der Kombination der gewonnenen QSAR-Modelle lassen Rückschlüsse zu, welche Kriterien entscheidend für die Selektivität der Liganden hinsichtlich eines Proteins sind. Bei diesem Verfahren fließen neben der experimentell bestimmten Bindungsaffinität keinerlei Informationen von Seiten des Rezeptors

ein. Liegen jedoch keine oder nur wenige Informationen von bindenden Liganden vor, lassen sich auch Aussagen über Selektivitätskriterien durch die Analyse mehrerer Proteinstrukturen treffen. Mit chemometrischen Methoden wie beispielsweise einer Hauptkomponentenanalyse können die exponierten physikochemischen Eigenschaften mehrerer Bindetaschen miteinander verglichen und somit für eine Selektivität diskriminierende Bereiche aufgefunden werden. Beide Ansätze zur Ermittlung von Kriterien für Selektivität – 3D-QSAR-Analysen und vergleichende Bindetaschenanalysen – wurden im Rahmen der vorliegenden Arbeit durchgeführt. Dabei lag der Schwerpunkt bei diesen Untersuchungen in der Suche nach praktikablen und adäquaten Strategien (*proof of principle*) zur Bestimmung von selektivitätsbestimmenden Faktoren. Diese Arbeiten erheben daher keinen Anspruch auf eine vollständige Analyse aller Selektivitätskriterien, die aus den zugrundeliegenden experimentellen Daten gewonnen werden können.

2.2 Ansätze aus der Literatur

Sowohl 3D-QSAR-Methoden zur Untersuchung einer Serie von Liganden als auch vergleichende Analysen strukturell ähnlicher Rezeptorstrukturen wurden in der Literatur zahlreich eingesetzt, um diejenigen Eigenschaften zu extrahieren, die für die Selektivität eines Liganden an eine Rezeptorstruktur entscheidend sind.

Eine Möglichkeit zur Untersuchung der Selektivität von Liganden zwischen zwei verschiedenen Rezeptorstrukturen ist die Entwicklung zweier getrennter QSAR-Modelle. Dazu müssen jedoch die Bindungsaffinitäten der Liganden zu beiden Proteinen bekannt sein. So untersuchten Agarwal *et al.* eine Serie von Tetrahydropyridinylindol-Agonisten, deren Affinitätswerte bezüglich der Serotonin-5-HT_{1A}- und 5-HT₂-Rezeptoren bekannt waren (Agarwal *et al.*, 1993). Zu jedem Rezeptorsubtyp wurde ein separates CoMFA-Modell erstellt und die resultierenden Konturdiagramme der Koeffizienten graphisch dargestellt. Durch den visuellen Vergleich dieser Konturen konnten Aussagen darüber getroffen werden, welche sterischen und elektrostatischen Eigenschaften entscheidend für die Bindung der Liganden zu der jeweiligen Rezeptorstruktur sind.

Einen Schritt weiter sind Wong *et al.* bei der Untersuchung einer Reihe von Imidazo-1,4-diazepin-Derivaten auf ihre selektive Bindung an Diazepam-empfindliche (DI) und Diazepam-unempfindliche (DS) Benzodiazepin-Rezeptoren gegangen (Wong *et al.*, 1993). Neben der Entwicklung zweier separater CoMFA-Modelle wurde zusätzlich ein „Selektivitätsmodell“ erstellt, das als abhängige Variablen die Differenzen der Bindungsaffinitäten (als pK_i -Werte) verwendete. Die Interpretation der damit gewonnenen Konturdiagramme erlaubte direkte Rückschlüsse auf selektivitätsbestimmende Faktoren zu beiden Benzodiazepin-Rezeptoren.

Auf ähnliche Weise wurde eine Serie von Serinprotease-Inhibitoren analysiert, deren Affinitäten zu Thrombin, Trypsin und Faktor Xa bekannt waren (Böhm *et al.*, 1999). Neben der Erstellung vorhersagefähiger CoMFA- und CoMSIA-Einzelmodelle zu allen drei Enzymen wurden Selektivitätsmodelle generiert, indem anstelle der Bindungsaffinitäten die Differenzen der pK_i -Werte zweier zu betrachtender Proteine in die PLS-Analyse einfließen. Daraus resultierten ebenfalls statistisch signifikante Modelle, aus deren graphischer Auswertung Bereiche in den Bindetaschen hervorgehoben wurden, die zur Diskriminierung der Selektivität zwischen den beiden Proteinen beitragen. Im Falle des Selektivitätsmodells von Thrombin und Trypsin wurden im wesentlichen zwei Bereiche in der Nähe der „60er-Schleife“ (siehe Beschreibung der Thrombin-Bindetasche, Seite 65) markiert, in denen die sterischen und elektrostatischen Eigenschaften der Liganden Einfluss auf ihre Selektivität zu beiden Proteinen nehmen. Diese Resultate stimmen damit überein, dass diese Schleife eine zusätzliche Bindetasche abgrenzt, die charakteristisch und einzigartig für Thrombin ist und somit wesentlich zur Selektivität dieses Enzyms beiträgt.

Die Resultate einer Selektivitätsstudie können auch in eine Datenbanksuche zur Auffindung selektiv bindender Strukturen einfließen. Wilcox *et al.* erstellten einen Datensatz von 16 strukturell diversen Dopamin-Agonisten, deren Affinitäten zu den Dopamin-D₁- und D₂-Rezeptoren bekannt waren (Wilcox *et al.*, 1998). Neben der Analyse der einzelnen CoMFA-Modelle¹⁰ zu den D₁- bzw. D₂-Aktivitätswerten wurden aus den am affinsten bindenden Liganden zu den jeweiligen Subrezeptoren zwei unterschiedliche Pharmakophormodelle erstellt. Erste Datenbanksuchen aus den distanzbasierten

¹⁰ Die Autoren verwenden hierbei die Bezeichnung „*differential QSAR*“, da der gleiche Datensatz für zwei Affinitätsreihen verwendet wird.

Pharmakophoren konnten unter Einbezug von Exklusionsvolumen verfeinert werden, die aus den sterischen Eigenschaften der entsprechenden Konturdiagramme extrahiert wurden. Dabei wurden für die D₁- und D₂-Pharmakophormodelle unterschiedliche Treffer in der Datenbank gefunden. Inwieweit diese auch wirklich selektiv an den jeweiligen Rezeptorstrukturen binden, wurde in der Arbeit nicht diskutiert.

In einer weiteren Studie untersuchten Wilcox *et al.* (Wilcox *et al.*, 2000) den gleichen Datensatz, nur diesmal mit Affinitätswerten zum Wildtyp (WT) und zu drei unterschiedlichen Mutanten (S193A, S194A, S197A) des D₂-Rezeptors. Die Vorgehensweise war analog zu der vorherigen: zu jeder Affinitätsreihe wurden CoMFA-Einzelmodelle erstellt und die daraus resultierenden Konturdiagramme graphisch interpretiert. Unterschiedliche Konturbereiche aus den WT- und Mutationsdaten gaben Hinweise auf eine mögliche Rezeptorumgebung. So wurden in einer weiteren Untersuchung vier der verwendeten D₂-Agonisten in ein Homologie-modelliertes D₂-Rezeptor-Modell eingedockt. Die aus den CoMFA-Modellen gewonnenen Einsichten bezüglich der drei mutierten Aminosäuren konnten zusammen mit schon bekannten Mutagenesedaten zur Erklärung eines verbesserten Bindungsmodells beitragen. In einem „*dual-template*“-Ansatz stand außerdem die Auswirkung eines anderen Alignments bei der Erstellung eines CoMFA-Modells mit den S193A-Mutationsdaten auf dem Prüfstand. Dazu wurde ein zweites Pharmakophormodell erstellt, wobei die Wahl auf einen D₂-Agonisten als Templat fiel, der die größte Affinitätssteigerung beim Wechsel vom WT zur S193A-Mutante aufweisen konnte. Während die beiden Modelle bei Verwendung der WT-Affinitätsdaten keine signifikanten Unterschiede zeigten, konnte die Vorhersagefähigkeit von D₂-Agonisten eines Testdatensatzes mit dem Mutationsmodells basierend auf dem neu erstellten Templat gesteigert werden.

Morón *et al.* demonstrierten anhand ihrer Untersuchungen von MAO-Inhibitoren¹¹, wie bei unbekannter Rezeptorstruktur Aussagen über eine mögliche Proteinumgebung getroffen werden können (Moron *et al.*, 2000). Dazu wurden aus einem Datensatz von Indolylmethylamin-Derivaten, deren Inhibitor konstanten zu MAO-A und MAO-B bekannt waren, zwei separate CoMFA-Modelle erstellt. Der visuelle Vergleich der Konturdiagramme beider Modelle ergab, dass die 5-Position des Indol-Rings bestimmend für die Selektivität zu den entsprechenden Rezeptorstrukturen ist. Aufgrund von Mutagenese-Experimenten war

¹¹ MAO = **M**ono**a**mino**x**idase

weiterhin bekannt, dass die Aminosäuren Phe208 und Ile199 entscheidend für die Substratspezifität von MAO-A bzw. MAO-B sind. Aus diesen Ergebnissen wurden mit Hilfe quantenmechanischer Rechnungen theoretische Enzym-Inhibitor-Modelle erstellt. Die daraus abgeleiteten Enthalpiedifferenzen zwischen freiem und gebundenem Liganden konnten mit den experimentell beobachteten Affinitäts- und Selektivitätswerten in Übereinstimmung gebracht werden.

Die bisherigen Beispiele zeigten Möglichkeiten zur Bestimmung von selektivitätsbestimmenden Faktoren anhand der Untersuchung einer Serie von Liganden. Wenn jedoch die dreidimensionale Struktur der Rezeptoren bekannt ist, so lassen sich daraus Informationen gewinnen, welche Bereiche in einer Bindetasche entscheidend für die Selektivität sind. Zur physikochemischen Charakterisierung einer Proteinbindetasche wird häufig das Programm GRID (GRID, 2001; Goodford, 1985) verwendet, das an den Schnittpunkten eines in der Bindetasche einbeschriebenen Gitters Wechselwirkungsenergien zu verschiedenen Sondenatomen berechnet. Pastor *et al.* analysierten damit die Bindetaschen einer bakteriellen (PDB-Code 4DFR) und einer rekombinant-hergestellten humanen (PDB-Code 1DHF) Dihydrofolat-Reduktase (DHFR) für insgesamt 41 GRID-Sonden (Pastor & Cruciani, 1995). Die daraus erhaltene Datenmatrix aus 82 Zeilen (entspricht 2x 41 Sonden-Protein-Paaren) und 6.480 Spalten (Anzahl Gitterpunkte) wurde einer Hauptkomponentenanalyse (*principal component analysis*, PCA) unterzogen¹². Die Auswertung des dabei resultierenden *score*-Plots, bei dem üblicherweise die ersten beiden Hauptkomponenten (*principal component*, PC) gegeneinander aufgetragen werden, deckte diejenigen Sondenatome auf, die eine Erhöhung der Affinität und Selektivität bezüglich der jeweiligen Proteine herbeiführen können. Die für die Selektivität diskriminierenden GRID-Sonden besitzen bezüglich der ersten Hauptkomponente *score*-Werte mit hohen Absolutwerten; die zweite Hauptkomponente hingegen separiert die Sonden nach der Stärke ihrer Affinität zu beiden Proteinen. Die mit diesen Sondenatomen verknüpften, chemischen Gruppen können daher bei der Entwicklung neuer, selektiv bindender Inhibitoren eingesetzt werden. Zusätzlich lassen sich durch die graphische Auswertung des *loading*-Plots auch die räumlichen Positionen bestimmen, an denen die verantwortlichen physikochemischen Eigenschaften zur Diskriminierung der Selektivität beitragen.

¹² Das PCA-Verfahren wird in Kap. 2.4 ausführlich beschrieben.

Die gleiche Vorgehensweise wurde von Filipponi *et al.* zur Charakterisierung selektivitätsbestimmender Faktoren zwischen den Rezeptorstrukturen von COX-1 und COX-2 herangezogen¹³ (Filipponi *et al.*, 2000). Da die entzündungshemmende Wirkung bei gleichzeitig nur geringen Nebenwirkungen eine selektive Inhibierung von COX-2 voraussetzt, könnte die Entwicklung selektiv wirkender COX-2-Inhibitoren einen Fortschritt für die Therapie darstellen. Bei dieser Untersuchung wurden die Kristallstrukturen von COX-1 (PDB-Code 1PRH) und COX-2 (PDB-Code 5COX) herangezogen und – nach der räumlichen Überlagerung ihrer Bindetaschen – die Wechselwirkungsfelder zu insgesamt 42 GRID-Sonden berechnet. Die resultierende Datenmatrix, bestehend aus 84 Zeilen (entspricht 2x 42 Sonden-Protein-Paaren) und 11.362 Spalten (Anzahl Gitterpunkte), wurde wie bei Pastor *et al.* einer PCA unterzogen. Die Auswertung des *score*-Plots lieferte diejenigen Sonden, die am stärksten zu einer Diskriminierung der Selektivität beitragen können. In diesem Fall ließen sich die Amidin- und Arylamidin-Sonden als selektivitätsbestimmende Fragmente identifizieren. Obwohl der Einsatz von Amidgruppen bei der Entwicklung selektiver COX-2-Inhibitoren nicht verbreitet ist, fanden die Autoren eine Verbindung, die eine abgewandelte Amidgruppe enthält und selektiv an COX-2 bindet ($IC_{50,COX-2} = 4,2 \mu M$; $IC_{50,COX-1} > 100 \mu M$). Die anschließende graphische Konturierung der Variablen aus dem *loading*-Plot ermöglichte zusätzlich die räumliche Bestimmung der Bereiche, die zu der Selektivität bezüglich beider Proteine beitragen. Aber auch die Auswertung der Beiträge von GRID-Sonden, für die keine Beiträge zur Selektivitätsdiskriminierung zu erwarten sind, kann beim Design neuer entzündungshemmender Arzneistoffe hilfreich sein. So zeigten sich beispielsweise Carboxylat- und Arylcarboxylat-Sonden als nicht-diskriminierend bezüglich ihrer Selektivität zu COX-1 und COX-2. Dieses Ergebnis stimmte mit der Beobachtung überein, dass viele literaturbekannte, nicht-selektive COX-Inhibitoren (zum Beispiel Flurbiprofen¹⁴, Indometacin¹⁵ oder auch Acetylsalicylsäure) eine Carboxylatgruppe tragen. Eine gewisse Flexibilität des Proteins konnte in dem Programm GRID durch die Benutzung der Direktive „*MOVE* = 1“ berücksichtigt werden; während der Berechnung der Wechselwirkungsfelder behandelte GRID dabei die Seitenketten bestimmter Aminosäuren (zum Beispiel Lysin, Arginin, etc.) als flexibel¹⁶. Somit konnte zum Beispiel der Einfluss der

¹³ COX = Cyclooxygenase

¹⁴ zur Gruppe der Arylpropionsäure-Derivate gehörend

¹⁵ zur Gruppe der Indoleessigsäure-Derivate gehörend

¹⁶ standardmäßig ist die Direktive „*MOVE* = 0“ gesetzt, das den Rezeptor als rigide behandelt

geschlossenen Konformation der Ligand-Eintrittsstelle auf die Ergebnisse der Selektivitätsuntersuchungen bestimmt werden.

Die Kombination Ligand- und Protein-basierter selektivitätsdeterminierender Verfahren wurde von Matter *et al.* am Beispiel der Matrixmetalloprotease (MMP) erprobt (Matter & Schwab, 1999). Hierbei waren sowohl die Kristallstrukturen zweier MMPs (MMP-3 und MMP-8) als auch eine Reihe von Inhibitoren mit Affinitätswerten zu beiden Proteinen bekannt. Analog zu den oben erläuterten Beispielen wurden auch hier zunächst getrennt für jedes Protein konsistente und voraussagekräftige QSAR-Einzelmodelle generiert und interpretiert. Neben dem CoMFA-Verfahren kamen auch CoMSIA- und GRID-Felder (phenolische OH-Gruppe) zur physikochemischen Charakterisierung der Liganden zum Einsatz. Die Kombination der Affinitätswerte von MMP-3 und MMP-8 durch Quotientenbildung der IC_{50} -Werte¹⁷ führte zu den entsprechenden „Selektivitätsmodellen“. Die Auswertung der dazugehörigen Konturdiagramme gab Aufschluss darüber, welche Bereiche mit welchen Eigenschaften bestimmend für die Selektivität zu einem bestimmten Protein sind. Die auf diesem Wege aufgefundenen Bereiche stimmten auch mit den physikochemischen Gegebenheiten selektiv bindender Inhibitoren des Datensatzes in diesem Bereich überein. Zusätzlich wurden die Bindetaschen der beiden MMPs mit insgesamt 38 GRID-Sonden analysiert und die daraus resultierende Datenmatrix, wie schon bei Pastor *et al.* (siehe oben) beschrieben, einer PCA unterworfen. Die Interpretation der aus den *loading*-Plots erhaltenen Konturdiagramme führte zu Selektivitätsbetrachtungen, die in Einklang mit den Ergebnissen der QSAR-Selektivitätsmodelle gebracht werden konnten. Somit lieferte die Auswertung sowohl der Ligand- als auch der Rezeptor-basierten Modelle konsistente Resultate.

Die Kombination von QSAR- und PCA-Verfahren ist auch dann möglich, wenn keine Informationen über eine Rezeptorstruktur vorliegen. Fichera *et al.* untersuchten einen Datensatz von 20 Cannabinoiden, deren Affinitäten als K_i -Werte gegenüber den Cannabinoid-Rezeptoren CB_1 und CB_2 bekannt waren, die beide zur Familie der G-Protein-gekoppelten Rezeptoren gehören (Fichera *et al.*, 2000). Zunächst wurden GRID-Wechselwirkungsfelder

¹⁷ Selektivitätsfaktor = $\log (IC_{50,MMP-3} / IC_{50,MMP-8})$

mit einem Carboxylat-Sondenatom berechnet und nach der Variablenselektion mit GOLPE¹⁸ (GOLPE, 2001) mit den CB₁- und CB₂-Affinitätswerten über eine PLS-Analyse korreliert. Die graphische Darstellung der PLS-Koeffizienten (PLS *partial weights*) resultierte beim CB₂-Rezeptormodell in einen zusätzlichen, beim CB₁-Modell nicht hervorgehobenen Bereich, der hinsichtlich seines Beitrags zur Selektivität diskutiert wurde. Um weitere selektivitätsbestimmende Faktoren zu bestimmen, wählten die Autoren zwei Inhibitoren aus, die jeweils selektiv an CB₁ und an CB₂ binden, und ermittelten die mit zehn GRID-Sonden berechneten Wechselwirkungsfelder. Die 20-reihige Datenmatrix (entspricht 2x 10 Sonden-Protein-Paaren) wurde einer PCA unterworfen (siehe weiter oben), jedoch mit dem Unterschied, dass anstelle zweier Rezeptorstrukturen nun die Felder um zwei selektiv bindende Liganden untersucht wurden. Im *score*-Plot ließen sich wiederum diejenigen GRID-Sonden selektieren, die entscheidend für die Selektivität der Liganden sind¹⁹.

Ein Nachteil der bisher vorgestellten Hauptkomponentenanalysen der GRID-Wechselwirkungsfelder von Proteinen hinsichtlich ihrer selektivitätsbestimmenden Eigenschaften war zum einen, dass immer nur zwei Proteine paarweise miteinander verglichen werden konnten. Prinzipiell erlaubt diese Methode auch die Verwendung von mehr als zwei Proteinen; die Auswertung der entsprechenden PCA-Plots würde sich dabei jedoch als äußerst komplex gestalten. Da andererseits immer mehr dreidimensionale Rezeptorstrukturen bekannt werden, wäre eine Methode wünschenswert, die einen chemometrischen Vergleich einer Vielzahl von Proteinen zulässt. Weiterhin konnten mit der beschriebenen Methode zwar die räumlichen Bereiche innerhalb der Proteinbindetasche bestimmt werden, die diskriminierend zur Selektivität beitragen, jedoch konnten die dafür entscheidenden GRID-Sondenatome diesen Bereichen nicht zugeordnet werden.

Eine neue Methode, die Consensus-Hauptkomponentenanalyse (*consensus principal component analysis*, CPCA) in Kombination mit dem Programm GRID, kann diese Beschränkungen aufheben (Kastenholz *et al.*, 2000). Als Datenbasis zur Validierung dieses Verfahrens analysierten Kastenholz *et al.* insgesamt neun Proteinstrukturen der Serinproteasen Thrombin (4x), Trypsin (3x) und Faktor Xa (2x). Nach der räumlichen Überlagerung der

¹⁸ Die Verfahren SRD (*smart region definition*) und FFD (*fast factorial design*) zur Variablenselektion sind in GOLPE implementiert.

¹⁹ Fraglich ist hierbei jedoch, inwieweit aus der Analyse von nur zwei Liganden umfassende Aussagen über die Selektivität auf die Cannabinoid-Rezeptoren gemacht werden können.

Strukturen wurden die Wechselwirkungsfelder zu zehn ausgewählten GRID-Sonden für jede Bindetasche berechnet. Die Datenmatrix (9 Proteine x 10 Sonden x 42.772 Gitterpunkte) wurde abweichend zur bisherigen Methode aufgebaut²⁰: jede Zeile der Matrix entspricht einer Proteinstruktur; in diesem Fall besteht die Datenmatrix aus neun Zeilen. Die Spalten nehmen nacheinander die Datenpunkte der Wechselwirkungsfelder für jedes Sondenatom auf; die Datenmatrix besteht also aus zehn Blöcken mit jeweils der gleichen Anzahl an Gitterpunkten pro Block²¹. Diese geänderte Anordnung der Datenmatrix hat bei der nachfolgenden PCA zur Folge, dass im zugehörigen *score*-Plot keine Sonden-Protein-Paare als Objekte (entsprechend den Zeilen der Matrix) aufgetragen werden, sondern die einzelnen Proteine selbst. In diesem Fall beinhaltet der *score*-Plot neun Datenpunkte, die den neun gewählten Serinproteasen entsprechen. Die Auftragung der ersten beiden Hauptkomponenten ermöglichte in diesem Fall eine eindeutige Separierung der drei verschiedenen Proteinstrukturen von Thrombin, Trypsin und Faktor Xa. Die Einteilung der Spalten in einzelne Blöcke entsprechend den verwendeten Sonden erlaubt zudem eine Skalierung, die den Einfluss eines jeden Blocks auf das Modell gleichsetzt. Die jeweiligen Wichtungskoeffizienten (*block unscaled weights*, BUW) werden durch eine Normalisierung der Varianzen jedes Blocks erhalten. Die CPCA-Methode kann nun die Datenmatrix insgesamt (*super level*) als auch jeden Block getrennt voneinander (*block level*) bewerten, wobei der Einfluss der einzelnen Blöcke auf die Gesamtheit mit Hilfe der *superweights* bewertet werden kann. So sind Aussagen darüber möglich, welche GRID-Sonden besonders diskriminierend auf die Selektivität der einzelnen Proteinklassen (Thrombin, Trypsin, Faktor Xa) wirken; zusätzlich können die Bereiche dieser Sonden in den Proteinbindetaschen bestimmt werden. Die Verwendung sogenannter *active plots* erleichtert die Handhabung der Selektivitätsstudien²². Für nähere Details und eine Beschreibung zusätzlicher Möglichkeiten dieses Verfahrens sei auf die entsprechende Literaturstelle (Kastenholz *et al.*, 2000) und auf das GOLPE-Manual (GOLPE, 2001) verwiesen.

Kürzlich wurde das oben beschriebene GRID/CPCA-Verfahren von Ridderström *et al.* für Selektivitätsuntersuchungen von Cytochrom-P450-Proteinen (CYP) eingesetzt (Ridderström *et al.*, 2001). Unter anderem spielen die Enzyme der CYP 2C-Subfamilie eine zentrale Rolle

²⁰ Mit der gewöhnlichen GRID/PCA-Methode würde die Datenmatrix aus 90 Zeilen (9 Proteine x 10 Sonden) und 42.772 Spalten (Anzahl Gitterpunkte) bestehen.

²¹ entspricht $10 \times 42.772 = 427.720$ Spalten

²² Die hier beschriebenen chemometrischen Verfahren (BUW, CPCA, *active plots*, etc.) sind in GOLPE implementiert.

bei der Metabolisierung einer Vielzahl von Arzneistoffen (Forth *et al.*, 2001). Basierend auf der Kristallstruktur von CYP 2C5 wurden zunächst strukturelle Homologiemodelle der Enzyme 2C8, 2C9, 2C18 und 2C19 (Sequenzhomologie $\approx 80\%$) erstellt. Innerhalb des aktiven Zentrums der 5 Proteinstrukturen wurden mit zehn ausgewählten GRID-Sonden Wechselwirkungsfelder berechnet und mittels der CPCA-Methode chemometrisch untersucht. Die Auswertung der *superweights* (siehe weiter oben) ermöglichte eine Auswahl vier repräsentativer Sonden (DRY, C3, N1 und O), die hydrophobe, sterische, Donor- und Akzeptor-Eigenschaften wiedergeben. Die Auftragung der ersten drei Hauptkomponenten (PC) im *score*-Plot ermöglichte eine eindeutige Separierung von CYP 2C8 (1. PC), 2C18 (2. PC) und 2C9 (3. PC). Weitere Untersuchungen zeigten, dass vor allem die hydrophoben Eigenschaften und die Geometrie der Bindetaschen entscheidend für die Selektivität der einzelnen Enzyme dieser Subfamilie sind. Zusätzlich konnte mit Hilfe des PCA-Modells ein inverses Pharmakophormodell²³ bezüglich CYP 2C9 erstellt werden, das in Übereinstimmung mit den physikochemischen Eigenschaften eines eingedockten Diclofenac-Moleküls (Substrat von CYP 2C9) war.

Weitere computergestützte Methoden zur Analyse selektivitätsbestimmender Faktoren werden im folgenden kurz aufgezählt: Montorsi *et al.* verwendeten 2D-Deskriptoren, die Größe und Form von Ligandmolekülen extrahieren (*size and shape descriptors*), zur Beschreibung einer Serie von α_{1a} -, α_{1b} - und α_{1d} -adrenergen Rezeptor-Antagonisten und konnten über eine PLS-Analyse ebenfalls Selektivitätsmodelle zu den jeweiligen Subrezeptoren erstellen (Montorsi *et al.*, 1998). Anstelle einer konventionellen PLS-Analyse zur Bestimmung der Selektivität zwischen dem Serotonin-5HT_{1A}- und dem α_1 -adrenergen Rezeptor wurde von López-Rodríguez *et al.* ein künstliches neuronales Netz (*artificial neural network*, ANN) benutzt (Lopez-Rodriguez *et al.*, 1999). Untersuchungen zur Selektivität von Celecoxib zu COX-2 im Vergleich zu COX-1 wurden von Price *et al.* mit Hilfe von Monte-Carlo-Simulationen und Freien-Energie-Rechnungen durchgeführt (Price & Jorgensen, 2001). Palomer *et al.* untersuchten die Positionierung einzelner Wassermoleküle in den Bindetaschen von COX-1 und COX-2 und den sich daraus ergebenden Einfluss auf ihre Selektivität (Palomer *et al.*, 2000).

²³ Pharmakophormodell basierend auf Rezeptorinformation (nicht Ligandeigenschaften)

2.3 Untersuchung von Liganden mittels 3D-QSAR Verfahren

Wie in der Einleitung schon erläutert wurde, besteht mit Hilfe von 3D-QSAR-Methoden die Möglichkeit, Aussagen über die Selektivität einer Reihe gegebener Inhibitoren zu treffen, die an mehreren Rezeptorstrukturen binden können. Durch die Verknüpfung der somit erzeugten Modelle lassen sich Rückschlüsse ziehen, welche Eigenschaften in bestimmten Bereichen der Liganden entscheidend für die Selektivität zu dem jeweiligen Rezeptor sind und wie sich diese gegebenenfalls steigern lässt.

2.3.1 Aufbau des Ligand-Datensatzes

2.3.1.1 Auswahl der Liganden

Entscheidend für den Erfolg der erzeugten QSAR-Modelle ist die Auswahl der Liganden des zu verwendenden Datensatzes und ihre „sinnvolle“ Überlagerung in einer bioaktiven Konformation. Die Erfassung der Selektivität von Liganden hinsichtlich eines Proteins wurde in dieser Arbeit anhand eines Datensatzes von 80 Serinprotease-Inhibitoren bestimmt, deren inhibitorische Aktivitäten in Form von K_i -Werten bzw. pK_i -Werten ($-\log K_i$) gegenüber den Enzymen Thrombin, Trypsin, Faktor Xa und Urokinase experimentell bekannt waren.

In einer früheren Studie (Böhm *et al.*, 1999) wurden erste Selektivitätsstudien anhand eines Datensatzes von 72 Inhibitoren durchgeführt, deren Affinitäten zu Thrombin, Trypsin und Faktor Xa bekannt waren. Die dort verwendeten Liganden bauten alle auf einer Benzamidin-Grundstruktur auf (Abb. 2-1).

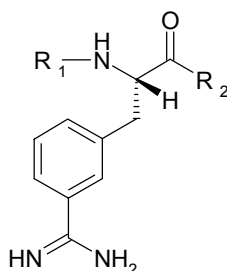


Abb. 2-1: Grundstruktur der Verbindungen mit den Substituenten R_1 und R_2 , die bei den früheren Selektivitätsuntersuchungen (Böhm *et al.*, 1999) verwendet wurde.

Die Variation des Datensatzes erfolgte durch verschiedene Substituenten R_1 und R_2 an dem Grundgerüst. Bei diesen Untersuchungen stellte sich die durch die gleiche Grundstruktur vorgegebene strukturelle Homogenität des Datensatzes bei den Auswertungen im Hinblick auf selektivitätsdiskriminierende Faktoren als nachteilig heraus. Weiterhin waren die experimentellen Affinitätswerte bezüglich der einzelnen Proteine teilweise stark miteinander korreliert. Der Korrelationskoeffizient der pK_i -Werte zwischen Thrombin und Trypsin lag mit einem Wert von $r = 0,85$ in einem Bereich, der die Aussagekraft anschließender Untersuchungen über die Selektivität beider Proteine prinzipiell beschränkte. Ebenfalls ein Nachteil des zunächst gewählten Datensatzes war der zu geringe Wertebereich der biologischen Aktivitäten bezüglich der Bindung an Faktor Xa. Zum Ableiten eines signifikanten QSAR-Modells lag die gegebene Streuung der Affinitätsdaten über drei logarithmischen Einheiten an der unteren Grenze. Folglich ergab sich ein Modell mit niedrigerem q^2 -Wert im Vergleich zu den Thrombin- und Trypsin-Modellen und die Resultate der sich anschließenden Selektivitätsuntersuchungen waren nur wenig aussagekräftig.

Da der Schwerpunkt der vorliegenden Arbeit jedoch auf einer Bestimmung selektivitätsbestimmender Faktoren lag, wurde ein erweiterter Datensatz erstellt, der obige Einschränkungen bezüglich strukturell zu homogener Strukturen und zu geringer Streuung der Affinitäten aufheben sollte. Dieser im stärkeren Maße heterogener Datensatz setzte sich aus insgesamt 80 Liganden zusammen, die in elf verschiedene Strukturklassen unterteilt wurden. In Abb. 2-2 ist jeweils eine repräsentative Strukturformel pro Klasse und ihre chemischen Bezeichnung dargestellt. Allen vorliegenden Inhibitoren gemein ist ihre benzamidinische Grundstruktur. Neben der TAPAP²⁴-Grundstruktur des vormals verwendeten Datensatzes (vgl. Abb. 2-1) finden sich in dem erweiterten Datensatz auch Strukturen wie einfach substituierte Benzamide oder Naphthamide, die große Bereiche der Serinprotease-Bindetaschen unbesetzt lassen. Sterisch anspruchsvolle Liganden hingegen wurden durch die Aufnahme von Bisbenzamidinen, TIPPS-²⁵ und NAPAP²⁶-Derivaten berücksichtigt. Eine Variation der Kettenlänge zwischen Benzamidin und der restlichen Struktur wurde mit den Glycinsäure-, Buttersäure- und Valeriansäure-Derivaten (zweiterer nicht in Abb. 2-2) realisiert.

²⁴ TAPAP = N^α-Tosyl-amidinophenylalanin-piperidid

²⁵ TIPPS = Tri-isopropyl-phenylsulfonyl

²⁶ NAPAP = N^α-(2-Naphthylsulfonylglycyl)-amidinophenylalanin-piperidid

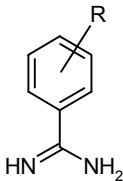
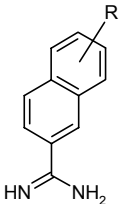
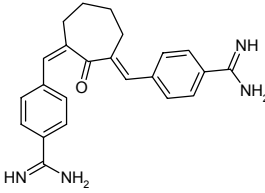
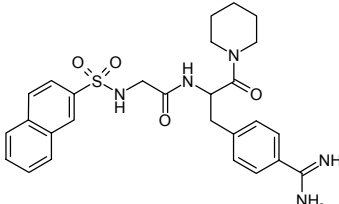
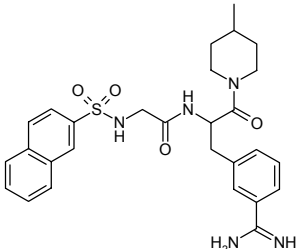
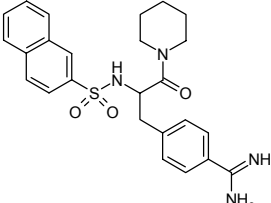
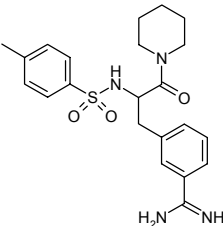
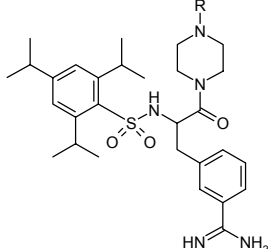
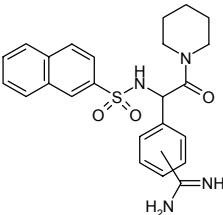
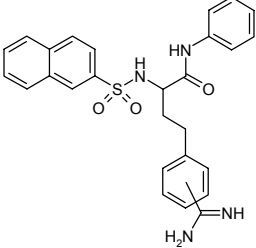
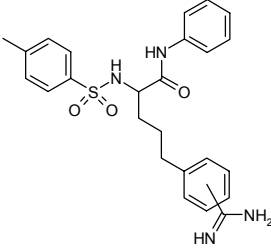
Benzamidine (BA)	Naphthamidine (NAPHTH)	Bisbenzamidine (BISBENZ)	4-NAPAP-Derivate (4NAPAP)
			
3-NAPAP-Derivate (3NAPAP)	4-TAPAP-Derivate (4TAPAP)	3-TAPAP-Derivate (3TAPAP)	TIPPS-Derivate (TIPPS)
			
Glycinsäure-Deriv. (GLYCIN)	Buttersäure-Derivate (BUTYRO)	Valeriansäure-Derivate (VALERI)	
			

Abb. 2-2: Übersicht der elf verschiedenen Strukturklassen des verwendeten Datensatzes und ihre chemische Bezeichnung. Die in Klammern gesetzten Bezeichnungen entsprechen den Abkürzungen der Strukturen, wie sie in der nachfolgenden Tab. 2-1 verwendet werden. Zu jeder Strukturklasse ist ein stellvertretender Inhibitor gezeigt.

Die experimentellen Bindungsaffinitäten der 80 Inhibitoren in Form ihrer pK_i -Werte ($-\log K_i$) bezüglich Thrombin, Trypsin, Faktor Xa und Urokinase sind in Tab. 2-1 aufgeführt. Die Genauigkeit der Affinitätsbestimmung ist mit einem durchschnittlichen Fehler von 10 – 15%, maximal 25% Abweichung angegeben (Stürzebecher *et al.*, 1995). Die Messbedingungen bei der Bestimmung der Bindungskonstanten sind ebenfalls in dieser Literaturstelle beschrieben und sollen hier nicht weiter erläutert werden.

Tab. 2-1: Bindungsaffinitäten der 80 Verbindungen, sortiert nach ihren elf verschiedenen Strukturklassen. Die Affinitätswerte sind als pK_i -Werte ($-\log K_i$) gegeben. Zusätzlich wurden zu jeder Affinitätsreihe Mittelwert, Standardabweichung, Maximal- und Minimalwert und der sich daraus ergebende Wertebereich berechnet.

Name (Nr_Strukturklasse_Substituent)	Thrombin pK_i	Trypsin pK_i	Faktor Xa pK_i	Urokinase pK_i
01_BA_H	3,658	4,456	3,387	3,745
02_BA_NH2	3,000	4,602	4,022	3,854
03_BA_CH2OPH	4,180	5,046	5,310	4,523
04_BA_CH2COPH	4,161	5,328	4,745	4,143
05_BA_CH2COCOOH	3,102	4,745	4,367	3,745
06_BA_CH2CH2COPH	3,538	4,398	3,602	3,854
07_NAPHTH_H	4,071	4,824	4,180	5,284
08_NAPHTH_NH2	4,180	5,284	4,699	5,237
09_NAPHTH_NHAC	3,921	5,027	4,268	4,959
10_NAPHTH_ONBU	4,921	5,444	4,824	5,268
11_NAPHTH_OCH2PH	5,469	5,678	5,076	5,337
12_NAPHTH_COOME	4,201	5,328	3,000	5,208
13_NAPHTH_CH2COPH	4,745	5,585	4,658	5,620
14_NAPHTH_DAIICHI	4,155	6,585	7,620	4,481
15_BISBENZ_CH7CH4AM	6,495	7,046	7,886	5,469
16_BISBENZ_CH5CH3AM	5,276	6,201	6,310	5,018
17_BISBENZ_CH26OHCH24AM	5,569	5,456	4,886	4,658
18_BISBENZ_PHPH3AM	5,538	6,745	5,638	5,553
19_BISBENZ_OCH2PH4AM	3,854	5,319	3,237	5,201
20_BISBENZ_OCH25O4AM	5,319	5,921	5,260	4,824
21_BISBENZ_OET3AM	5,495	5,824	6,523	5,229
22_BISBENZ_OET4AM	3,000	4,745	5,553	3,959
23_BISBENZ_OH4AM	4,222	5,222	6,102	4,509
24_BISBENZ_HPHNHPH4AM	3,886	5,824	4,092	5,469
25_4NAPAP_BNAS_PPD	8,222	6,161	5,102	3,638
26_4NAPAP_BA_PPD	5,699	4,959	4,051	3,000
27_4NAPAP_BNAS_PZD-NH	6,260	5,585	4,602	3,155
28_4NAPAP_BNAS_PRO-OH	6,292	5,678	4,796	3,097
29_4NAPAP_TOS_NHPH	4,824	4,796	5,347	4,032
30_4NAPAP_TOS_OME	6,131	4,770	4,699	3,301
31_3NAPAP_BNAS_PPD-4ME	5,444	4,337	4,602	3,000
32_3NAPAP_BNAS_PZD-NSME	5,051	7,066	4,538	4,131
33_3NAPAP_BNAS_NHME	5,018	5,036	5,523	3,509
34_3NAPAP_3AMPS_OH	4,432	5,252	5,658	4,959
35_3NAPAP_TOS_OME	5,409	4,796	6,076	4,180
36_3NAPAP_MTR_OME	6,377	5,432	5,796	3,959
37_3NAPAP_TBUPS_OME	5,854	5,367	6,420	3,886
38_3NAPAP_BNAS_OIPR	5,921	5,009	6,620	4,310
39_4TAPAP_BNAS_PPD	6,377	4,638	3,824	3,155

Name (Nr_Strukturklasse_Substituent)	Thrombin pK _i	Trypsin pK _i	Faktor Xa pK _i	Urokinase pK _i
40_4TAPAP_BNAS_PZD-NAC	4,886	4,456	4,284	3,222
41_4TAPAP_BNAS_OH	3,638	3,854	3,699	3,000
42_4TAPAP_BNAS_ONBU	6,432	4,658	4,886	4,180
43_3TAPAP_H_PPD-4ME	4,569	4,538	3,638	3,097
44_3TAPAP_BNAS_PPD-4PPO	4,824	6,000	4,658	4,149
45_3TAPAP_BNAS_PPD-4BZL	5,108	6,824	4,854	5,076
46_3TAPAP_BNAS_PIP-OME	7,155	5,721	4,337	3,097
47_3TAPAP_BNAS_PIP-3MOME	6,284	5,745	4,444	3,301
48_3TAPAP_BNAS_NIP-OET	6,444	7,699	4,377	5,125
49_3TAPAP_BNAS_NIP-OBZL	6,523	6,796	5,066	4,770
50_3TAPAP_BNAS_INIP-OME	7,770	7,444	4,367	5,032
51_3TAPAP_BNAS_INIP-OCHX	5,678	7,097	5,119	5,000
52_3TAPAP_BNAS_INIP-NBZL	5,509	6,569	4,456	4,027
53_3TAPAP_BNAS_INIP-OBZL	6,678	7,174	5,092	5,194
54_3TAPAP_BNAS_2PPN3COOM	6,921	5,620	5,824	4,027
55_3TAPAP_BNAS_PZD-NSME	8,481	6,721	4,658	4,367
56_3TAPAP_BNAS_PZD-NBZL	5,699	6,398	5,194	3,839
57_3TAPAP_TOS_NHME	4,745	4,337	4,398	3,301
58_3TAPAP_TMEPS_NHME	4,357	3,000	3,194	3,000
59_3TAPAP_BNAS_NHIPR	5,208	4,796	3,444	3,000
60_3TAPAP_BNAS_OME	6,553	5,602	5,602	3,523
61_3TAPAP_BNAS_ONBU	7,886	6,585	5,509	4,237
62_3TAPAP_BNAS_TCD	6,921	5,824	5,638	4,284
63_3TAPAP_BNAS_PRO-OME	4,553	3,553	3,745	3,000
64_TIPPS_NIP-OBZL	6,046	6,071	6,046	5,921
65_TIPPS_INIP-OH	5,056	5,194	5,770	4,292
66_TIPPS_INIP-OET	6,678	7,086	5,377	6,143
67_TIPPS_PZD-NBUO	5,921	7,268	5,585	4,924
68_TIPPS_PZD-NCOOET	6,174	6,921	5,824	6,215
69_TIPPS_NHME	5,420	4,585	4,959	4,000
70_TIPPS_NH-PIPYL	5,987	6,097	3,745	3,469
71_TIPPS_NHCH2-TFN	7,081	5,693	5,137	3,310
72_GLYCIN_3AM_TOS_PYR	5,387	6,215	5,523	5,244
73_GLYCIN_3AM_TOS_MOR	5,149	6,237	4,854	4,538
74_GLYCIN_4AM_BNAS_PPD	4,310	3,770	4,328	3,000
75_BUTYRO_4AM_TOS_NHPH	4,398	6,469	4,921	4,523
76_BUTYRO_4AM_BNAS_NHPH	4,469	6,244	5,284	4,056
77_BUTYRO_4AM_BNAS_MOR	3,886	4,569	4,745	3,000
78_VALERI_3AM_BNAS_MOR	7,022	5,959	5,444	3,097
79_VALERI_4AM_BNAS_PPD	6,481	5,301	4,921	4,658
80_VALERI_4AM_TOS_NHPH	4,796	5,387	5,398	5,092
Mittelwert	5,379	5,570	4,940	4,260
Standardabweichung	1,210	0,969	0,927	0,878
Maximalwert	8,481	7,699	7,886	6,215
Minimalwert	3,000	3,000	3,000	3,000
Wertebereich	5,481	4,699	4,886	3,215

Wie aus Tab. 2-1 zu erkennen ist, erstreckt sich der Wertebereich der experimentellen Bindungsaffinitäten zu allen vier Proteinen auf mehr als die für eine statistische Analyse notwendigen, drei logarithmischen Einheiten (Thibaut *et al.*, 1993). Für Thrombin, Trypsin und Faktor Xa wird die geforderte Bandbreite mit 5,5 (Thrombin), 4,7 (Trypsin) bzw. 4,9 (Faktor Xa) Zehnerpotenzen sogar deutlich überschritten. Nur für Urokinase existieren derzeit (noch) keine hochaffinen benzamidinischen Inhibitoren (Stürzebecher *et al.*, 1999). Der in diesem Datensatz berücksichtigte Inhibitor mit der höchsten Bindungsaffinität gegenüber Urokinase (68_TIPPS_PZD-NCOOET) weist einen pK_i -Wert von 6,2 ($K_i \approx 0,6 \mu\text{M}$) auf; bezüglich Thrombin, Trypsin und Faktor Xa konnten hingegen Inhibitoren im nanomolaren Bereich ($K_i \approx 3 - 20 \text{ nM}$) aufgenommen werden.

2.3.1.2 Streuung und Interkorrelation der Affinitätsdaten

Die untersuchten Liganden besitzen, wie schon erwähnt, unterschiedliche Bindungsaffinitäten bezüglich der vier Serinproteasen. Daher wurde in einem weiteren Schritt überprüft, wie stark die biologischen Aktivitäten von Protein zu Protein miteinander korrelieren. Eine hohe Datenkorrelation würde darauf hinweisen, dass die zu berechnenden Modelle nicht unabhängig voneinander sind und daher speziell bei den Selektivitätsmodellen Schwierigkeiten im Hinblick auf eine Auswertung und Interpretation zu erwarten sind.

Abb. 2-3 zeigt die wechselseitigen Auftragungen der Bindungsaffinitäten zu Thrombin, Trypsin, Faktor Xa und Urokinase. Gleichzeitig wurde jeweils ein Korrelationskoeffizient r zur quantitativen Abschätzung der Interkorrelation berechnet. Der Korrelationskoeffizient r liegt in einem theoretischen Wertebereich von 0 und 1, wobei Zahlen nahe 1 auf eine hohe Interkorrelation hindeuten. Die höchsten Korrelationen mit $r = 0,58$, $0,52$ und $0,41$ wurden zwischen den Daten für Thrombin und Trypsin, Urokinase und Trypsin bzw. Faktor Xa und Trypsin errechnet. Dagegen interkorrelieren die Werte für Thrombin und Faktor Xa oder Urokinase und Faktor Xa mit $r = 0,28$ bzw. $0,36$ deutlich geringer. Die Affinitätsdaten bezüglich Thrombin und Urokinase verhalten sich mit einem Korrelationskoeffizienten von $r = 0,02$ praktisch völlig unabhängig voneinander.

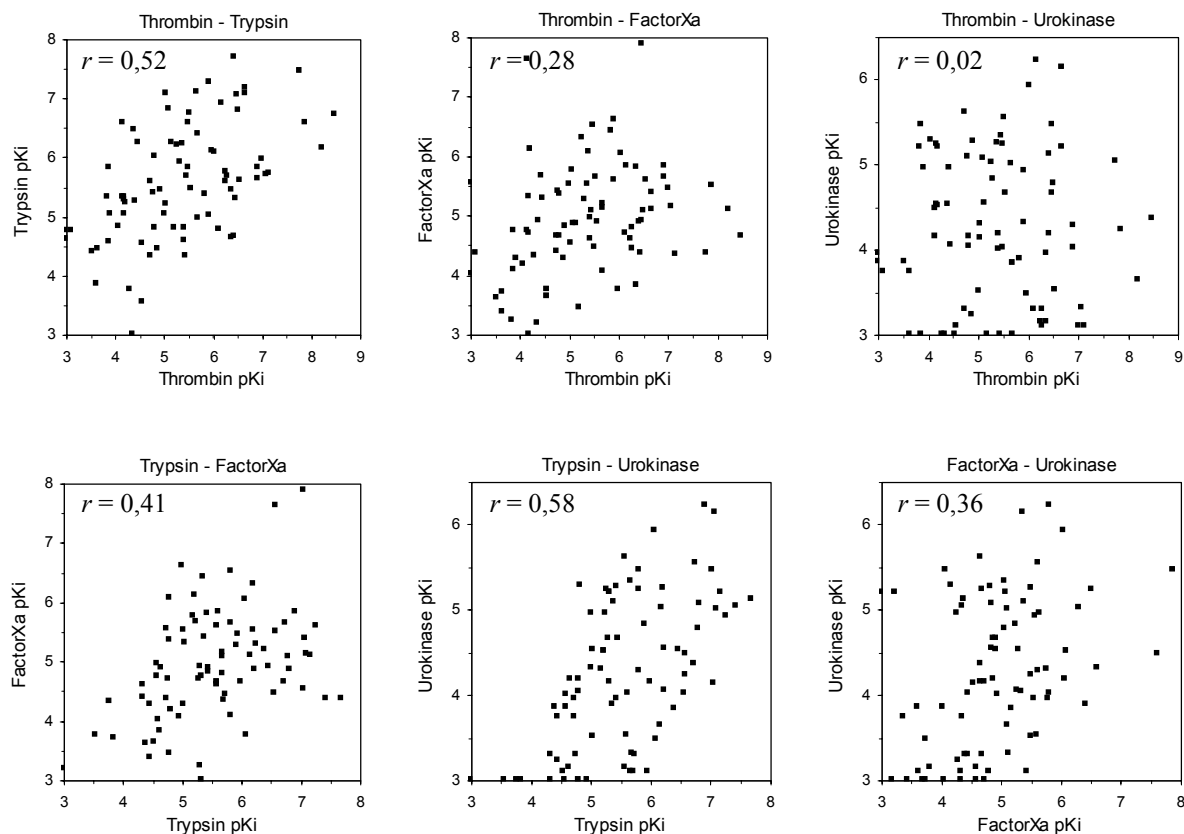


Abb. 2-3: Paarweise, wechselseitige Auftragung der experimentellen p_{K_i}-Werte der Inhibitoren bezüglich Thrombin, Trypsin, Faktor Xa und Urokinase. Der Korrelationskoeffizient r erlaubt eine quantitative Abschätzung der zugrundeliegenden Datenkorrelation.

Im Vergleich zu dem Korrelationskoeffizienten von $r = 0,85$ zwischen Thrombin und Trypsin des ursprünglichen Datensatzes (Böhm *et al.*, 1999) wäre daher bei dem neu erstellten, strukturell heterogen zusammengesetzten Datensatz zu erwarten, dass die erzeugten QSAR-Modelle und die daraus abgeleiteten Selektivitätsmodelle aussagekräftige Ergebnisse liefern sollten.

2.3.2 Strukturelles Alignment der Inhibitoren

Der Aufbau der Liganden des Datensatzes erfolgte zunächst manuell, wobei die Lage und Konformation von Referenzliganden aus experimentell bestimmten Kristallstrukturen als Template berücksichtigt wurden. Für jede Strukturklasse aus Abb. 2-2 lag zumindest eine, meist in Trypsin bestimmte Kristallstruktur vor. Diese Kristallstrukturen wurden zusammen

mit den Referenzliganden aufeinander überlagert, indem jeweils die C_{α} -Atome der Aminosäuren His 57, Asp 102 und Ser 195 der katalytischen Triade (Dodson & Wlodawer, 1998) und Asp189 der Spezifitätstasche für das rigide Alignment herangezogen wurden (RMS-Fit). Zu jeder der vier Serinproteasen (Thrombin, Trypsin, Faktor Xa und Urokinase) wurde eine stellvertretende Kristallstruktur ausgewählt (PDB-Codes 1ETS, 1PPH, 1HCG und 1LMW). Um diese vier Proteinstrukturen wurde im Bereich der Ligand-Bindestellen jeweils die lösungsmittelzugängliche Oberfläche (*Connolly surface*) berechnet (Connolly, 1983). Diese wird erzeugt, indem eine Kugel mit dem Radius 1,4 Å (Größe eines Wassermoleküls) auf der Oberfläche des Proteins abgerollt wird. Abb. 2-4 gibt einen Überblick über die somit erzeugten Proteinbindetaschen (*active site*) von Thrombin, Trypsin, Faktor Xa und Urokinase. Dabei zeigen sich unterschiedlich stark ausgeprägte Bereiche der Untereinheiten der Bindetaschen, die für die jeweilige Selektivität der Enzyme zu einem gegebenen Liganden verantwortlich sind.

2.3.2.1 Beschreibung der Proteinbindetaschen

Die S1-Tasche ist bei allen vier Serinproteasen vorhanden, wobei Breite und Tiefe der Taschen geringfügig variieren²⁷. Tief in dieser sogenannten Spezifitätstasche liegt Asp 189 vergraben, das bevorzugt mit basischen NH-Gruppen (zum Beispiel Amidine, Guanidine, Pyridine) Salzbrücken bildet (Mack *et al.*, 1995; Feng *et al.*, 1997; Stürzebecher *et al.*, 1997; Gabriel *et al.*, 1998; Phillips *et al.*, 1999). Mittlerweile sind aber auch Kristallstrukturen bekannt, in denen hydrophobe, nicht-basische Gruppen in die S1-Tasche binden (Lumma *et al.*, 1998; Tucker *et al.*, 1998). Die zu Asp 189 benachbarten Aminosäuren Gly 219 und Gly 216 können zusätzliche Wasserstoffbrückenbindungen zu einem Liganden ausbilden.

Thrombin besitzt im Vergleich zu den anderen Proteinen die engste und am stärksten strukturierte Bindetasche (Stubbs *et al.*, 1992; Bode & Stubbs, 1993; Stubbs & Bode, 1993a; Stubbs & Bode, 1993b; Stubbs & Bode, 1995; Bode & Stubbs, 1995; Wagner *et al.*, 1998). So wird die S2-Tasche – charakteristisch für Thrombin – durch eine sogenannte „60er-Schleife“ abgegrenzt, die durch die kurze Sequenz Tyr 60A, Pro 60B, Pro 60C und Trp 60D gebildet wird. Diese Schleife sitzt wie ein „Deckel“ auf der Spezifitätstasche (in Abb. 2-4 durch

²⁷ Somit wird auch durch die S1-Tasche die Selektivität zu Liganden bestimmt. Dieses Merkmal spielte aber in dieser Arbeit keine Rolle, da die benz- bzw. naphthamidinischen Inhibitoren des Datensatzes keine starke strukturelle Variation im Hinblick auf die Besetzung dieser Untertasche aufweisen.

gestrichelte Linie gekennzeichnet) und sorgt so für die hohe Substratspezifität von Thrombin. Im Gegensatz dazu ist die S2-Tasche, in der Literatur auch als proximale (in der Nähe zur katalytischen Triade) Tasche bezeichnet, bei Faktor Xa schon deutlich weniger ausgeprägt. Eine gewisse Eingrenzung dieser Tasche ist dort durch die Anordnung des sterisch anspruchsvollen Tyr 99 und Gln 61 in diesem Bereich gegeben. Im Falle von Trypsin und Urokinase wird die S2-Tasche durch die lösungsmittelzugängliche Oberflächen kaum noch als tief vergrabene Kavität ausgewiesen.

Ein weiterer, bei Thrombin stark ausgebildeter Bindetaschenbereich stellt die S3-Tasche dar. Dieser oftmals als distale (entfernt von der katalytischen Triade) Tasche bezeichnete Bereich wird durch die Seitenketten von Glu 97A, Asn 98, Leu 99, Ile 174, Glu 217 und Trp 215 gebildet. Zudem ist die Aminosäureseitenkette von Trp 215 zusammen mit denen von Tyr 60A, Trp 60D und Leu 99 auch verantwortlich für eine bevorzugte Bindung von aromatischen Resten. Daher wird die S3-Tasche bei Thrombin auch als Arylbindetasche bezeichnet. Die Anwesenheit einer durch Tyr 99, Phe 174 und Trp 215 gebildeten hydrophoben Tasche spielt eine entscheidende Rolle bei den Selektivitätseigenschaften von Faktor Xa (Stubbs & Bode, 1994; Stubbs *et al.*, 1995; Stubbs, 1996; Bode *et al.*, 1997; Gabriel *et al.*, 1998; Whitlow *et al.*, 1999; Zhao *et al.*, 2000). Diese in eine S3- und S4-Tasche unterteilte Region vermag bevorzugt große hydrophobe als auch aromatische Reste eines Liganden aufzunehmen. Bei Trypsin hingegen ist die zuletzt genannte Arylbindetasche wiederum nur geringfügig ausgeprägt; lediglich die Seitenketten von Asn 97, Thr 98 und Leu 99 flankieren die obere Grenze einer angedeuteten Bindetasche (vgl. Abb. 2-4). Im Falle der Urokinase-Struktur wird die Ausbildung einer S3-Tasche in dieser Region mit der sterischen Besetzung durch einen Insertionsloop (Asp 97, Thr 97A, Leu 97B, Ala 98 und His 99) blockiert. Dafür zeigt sich bei Urokinase eine deutliche Ausprägung der S1'-Bindetasche, die eine Bindung von hydrophoben und sterisch anspruchsvollen Substituenten ermöglicht (Stürzebecher *et al.*, 1999; Katz *et al.*, 2000; Nienaber *et al.*, 2000; Zeslawska *et al.*, 2000; Klinghofer *et al.*, 2001; Verner *et al.*, 2001).

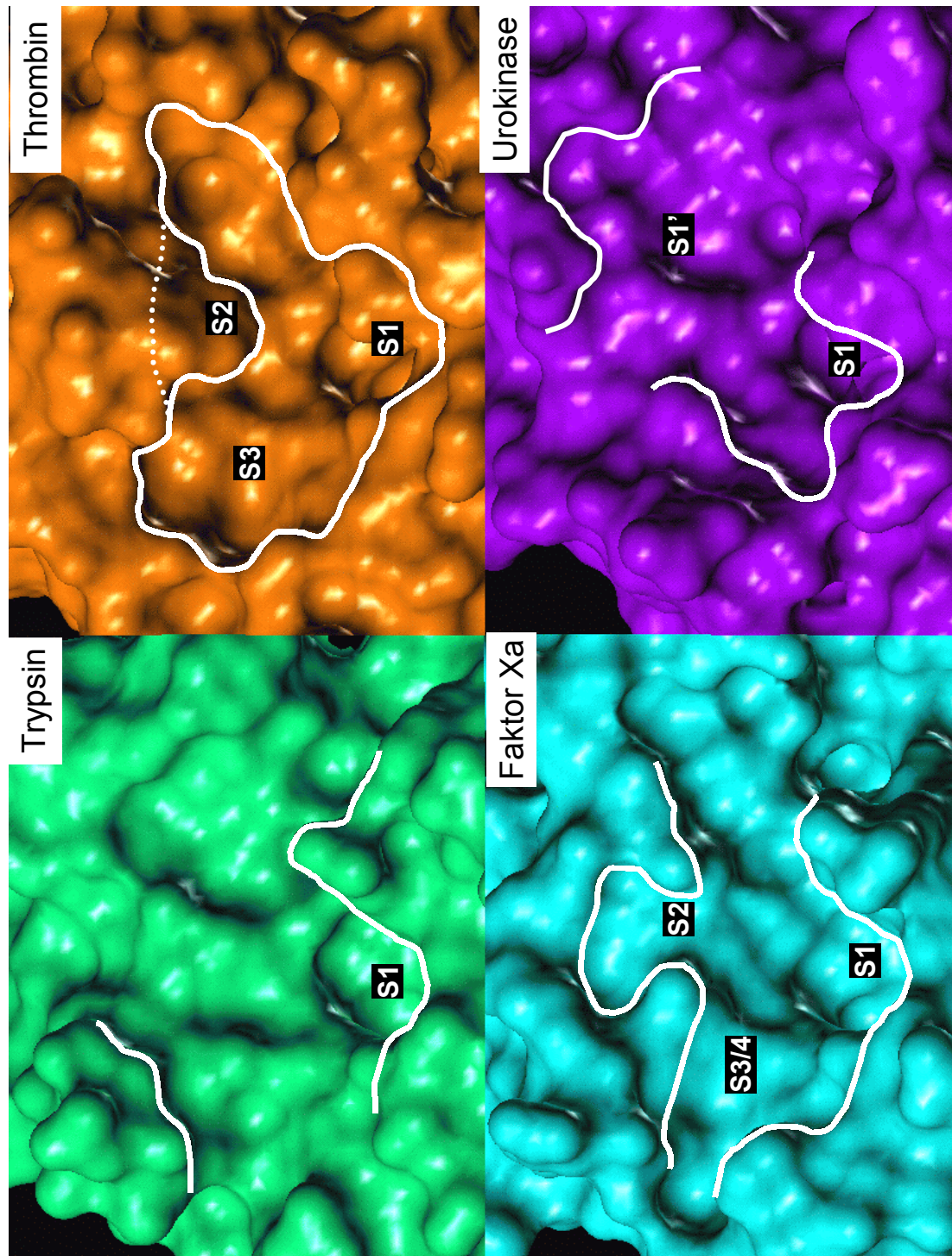


Abb. 2-4: Darstellung der Proteinbindetaschen von Thrombin (orange), Trypsin (grün), Faktor Xa (cyan) und Urokinase (violett) mit ihren lösungsmittelzugänglichen Oberflächen. Die für das jeweilige Protein entscheidenden Bindetaschenbereiche (S1, S1', S2, S3 bzw. S3/4) sind gekennzeichnet und farblich hervorgehoben.

2.3.2.2 Manuelle Platzierung der Liganden

Beim manuellen Aufbau der Inhibitoren des Datensatzes konnten unter Berücksichtigung der lösungsmittelzugänglichen Oberfläche von Thrombin sterische Überlappungen mit der Proteinstruktur weitgehend vermieden werden. Die Entscheidung für Thrombin als Referenz für den Aufbau der Liganden lag darin begründet, dass die dazugehörige Bindetasche den geringsten Raum für die Liganden zur Verfügung stellt und somit die Positionierung der einzelnen Substituenten auf wenige Freiheitsgrade beschränkt wird. Beim Aufbau der Liganden wurde beachtet, dass vorhandene Seitenketten möglichst nur energetisch bevorzugte Torsionswinkel aufweisen. Bestanden mehrere Möglichkeiten zur Wahl eines geeigneten Konformationswinkels, wurde eine systematische Konformationsanalyse durchgeführt. Dadurch konnten energetisch günstige Konformationen aufgefunden werden.

Zur Adaption von Bindungslängen und -winkel wurden die aufgebauten Strukturen mit dem in SYBYL implementierten Tripos-Kraftfeld (Clark *et al.*, 1989) energieminimiert. Dabei wurde jedoch darauf geachtet, dass sich die vorselektierten Konformationswinkel nicht mehr wesentlich änderten. Hierzu wurde die Option „steepest descent“ mit einer möglichst kleinen Anzahl an Minimierungsschritten (50 – 100) verwendet.

Die Protonierungszustände und damit die zugewiesenen Atomtypen der Moleküle wurden so eingestellt, wie die Verbindungen unter Annahme physiologischer Bedingungen (pH 7) vorliegen sollten. Dazu wurden alle basischen Amidgruppen protoniert (Atomtyp C.cat und N.pl3), für alle Carbonsäurereste wurde ein deprotonierter Zustand (Atomtyp O.co2) angenommen. Basische Amine, die ausschließlich an C.3-Atomen bzw. H-Atomen gebunden waren, wurden als protoniert (Atomtyp N.4) angesehen. Aromatische Amine und Amide hingegen wurden nicht protoniert. Zugleich wurde kontrolliert, dass die formale Gesamtladung eines Liganden den Wertebereich von -1 und +1 möglichst nicht übersteigt.

2.3.2.3 Kraftfeldminimierung in den Proteinbindetaschen

Um eine spezifische Adaption der Ligandkonformation an die Rezeptorstrukturen der einzelnen Proteine zu ermöglichen, wurden die zunächst manuell aufgebauten Ligandstrukturen mit Hilfe des Kraftfelds MAB (Gerber & Muller, 1995; Gerber, 1998), das

im dem Modellierungsprogramm MOLOC²⁸ (MOLOC, 2001) implementiert ist, innerhalb der Bindetaschen von Thrombin, Trypsin, Faktor Xa und Urokinase energieminiert. Mit diesem Schritt wurde eine gemeinsame, für alle Proteine als Referenz geltende Überlagerung aufgegeben. Dagegen kam für jede Serinprotease ein spezielles, auf die individuelle Bindetasche adaptiertes Alignment zum Tragen. Dabei werden Bindungslängen und -winkel auf die im MAB-Kraftfeld vorgegebenen Parameter angepasst, und mögliche, günstige Wechselwirkungen (sterisch, elektrostatisch, Wasserstoffbrücken) funktioneller Gruppen mit der jeweiligen Rezeptorstruktur können ausgebildet werden. Es wurde jedoch darauf geachtet, dass sich die vorgegebenen Torsionswinkel der per Hand aufgebauten Geometrien nicht mehr wesentlich änderten.

Zuerst wurden die Konformationen der Liganden bezüglich der Bindetasche von Thrombin (PDB-Code 1ETS) energieminiert. Da die Inhibitoren bereits manuell in der Thrombin-Tasche aufgebaut worden waren, zeigten sich hier nur geringe Abweichungen der minimierten Strukturen zu den ursprünglichen Geometrien. Durch die Optimierung wurden hauptsächlich noch vorhandene sterische Abstoßungen eliminiert und Wasserstoffbrückenbindungen zu geeigneten funktionellen Gruppen der Proteinbindetasche ausgebildet.

Zur Anpassung der Ligandkonformationen an die Bindetaschenstruktur von Faktor Xa wurden die in der Thrombin-Tasche minimierten Inhibitoren in die Bindetasche von Faktor Xa (PDB-Code 1HCG) platziert und wiederum mit MOLOC energieminiert. Dabei erfolgten die größten Konformationsänderungen im Bereich der S2 und S3/4-Taschen. Ausschlaggebend für diese Verschiebungen waren hauptsächlich Liganden, die in der S2-Bindetasche sterisch anspruchsvolle Substituenten platzierten. Dieser Bereich ist aber in der Struktur von Faktor Xa durch die Seitenkette von Tyr 99 besetzt und führte bei der Minimierung zu einer Verdrängung der betreffenden Ligandsubstituenten.

Analog verhielt es sich bei der anschließenden Adaption der Inhibitorstrukturen in der Bindetasche von Urokinase (PDB-Code 1LMW). Alle Substituenten der Liganden, die in der S3- (Thrombin) bzw. S3/4-Tasche (Faktor Xa) positioniert waren, kollidierten mit den Aminosäure-Seitenketten des 97er-Insertionsloops von Urokinase (siehe weiter oben). Folglich wurden diese Reste in Richtung der S2-Tasche wegorientiert oder auch aus der

²⁸ Die verwendeten Programme sind in Kap. 4.2 nochmals kurz aufgeführt.

Bindetasche verdrängt, so dass die Substituenten teilweise zum Lösungsmittel hin platziert wurden. Viele Substituenten, die im Falle von Thrombin und Faktor Xa die S2-Tasche besetzten, wurden während der Minimierung in der Urokinase-Struktur in Richtung der S1'-Bindetasche neu positioniert. Kristallstrukturen strukturell verwandter Urokinase-Inhibitoren, die diese Reste ebenfalls in die S1'-Tasche orientieren, untermauern die Richtigkeit dieser Konformationsänderungen in Urokinase verglichen mit Thrombin und Faktor Xa. Allerdings ist die Zahl experimentell zugänglicher Urokinase-Strukturen im Vergleich zu bekannten Thrombin- und Faktor Xa-Strukturen derzeit noch sehr gering; so befanden sich in der PDB (Stand: April 2001) nur drei Einträge von Urokinase-Komplexen, die auch einen geeigneten Liganden gebunden hatten (PDB-Codes 1EJN, 1FV9, 1LMW). Somit gelang eine Überprüfung der minimierten Strukturen in der Bindetasche von Urokinase auf sinnvolle Konformationen nur unzureichend. Der durch die Optimierung erhaltene Datensatz von in der Urokinase-Bindetasche energieminierten Inhibitoren ist daher im Hinblick auf seine Verwendung kritisch zu betrachten.

Wesentlich einfacher gestaltete sich hingegen die Minimierung der Inhibitoren in der Trypsin-Bindetasche. Da zu jeder Strukturklasse der Liganden aus Abb. 2-2 zumindest eine Komplexstruktur mit Trypsin als Röntgenstruktur vorlag, konnten diese Konformationen direkt übernommen und mit MOLOC an die Referenzstruktur von Trypsin (PDB-Code 1PPH) angepasst werden. Die Minimierung dieser Strukturen hatte keine signifikanten Änderungen der Konformationen zur Folge, lediglich die Ausrichtung funktioneller Gruppen einiger Liganden hinsichtlich ihrer intermolekularen Wechselwirkungen zum Protein wurde den Parametern des MAB-Kraftfelds angepasst.

2.3.3 Generierung separater Einzelmodelle

Bevor aussagekräftige Modelle erstellt werden können, die selektivitätsbestimmende Faktoren verschiedener Enzyme untereinander erschließen, sollten zunächst separate Untersuchungen der einzelnen beteiligten Proteine durchgeführt werden. Erst wenn dabei signifikante und konsistente Modelle aus den Ligandaffinitäten bezüglich jeweils eines Proteins resultieren, erscheint der nächste Schritt – die Erzeugung von Selektivitätsmodellen – sinnvoll.

Daher wurden zunächst unter Verwendung des Inhibitor-Datensatzes vier separate Korrelationsmodelle („Einzelmodelle“) zu den untersuchten Serinproteasen Thrombin, Trypsin, Faktor Xa und Urokinase generiert. Aufgrund eines solchen Modells lassen sich prinzipiell Aussagen treffen, welche physikochemischen Eigenschaften der Liganden in einem bestimmten Bereich der betreffenden Proteinbindetasche zu- oder abträglich für die Bindungsaffinität zu einem bestimmten Protein sind. Die Einzelmodelle wie auch die im Anschluss aufgestellten Selektivitätsmodelle wurden unter Einsatz der CoMSIA-Methode generiert, die schon in Kap. 1 beschrieben wurde. Kurz zusammengefasst ermöglicht dieses Verfahren die Berechnung physikochemischer Eigenschaften anhand sogenannter Ähnlichkeitsindizes (vgl. Gl. 1-1, Seite 15) an den Schnittstellen eines Gitterkastens, der über die zu untersuchenden Moleküle gelegt wird. Standardmäßig werden fünf Eigenschaften – sterisch, elektrostatisch, hydrophob, Wasserstoffbrücken-Donor und -Akzeptor – ermittelt und als molekulare Felder zur Charakterisierung der Liganden eines Datensatzes gespeichert. Ein spezielles chemometrisches Verfahren, die PLS-Analyse (Wold *et al.*, 1984), versucht die experimentell bestimmten Bindungsaffinitäten (abhängige Variablen) mit den molekularen Feldern der Liganden (unabhängige Variablen) zu korrelieren und in einem QSAR-Modell zusammenzufassen. Die entsprechenden Affinitätswerte der Inhibitoren zu den jeweiligen Proteinen wurden in Form von pK_i -Werten ($-\log K_i$) verwendet. In Kap. 4.3 (ab Seite 161) sind die statistischen Parameter aus der PLS-Analyse im Detail erläutert.

Sämtliche statistischen Analysen wurden auf Grundlage der molekularen Felder durchgeführt, die mit einem Gitter der Maschenweite 2 Å erzeugt wurden. Stichprobenartige PLS-Analysen mit einem 1 Å Gitter resultierten in relativ ähnlichen statistischen Werten. Wie schon in einer früheren Arbeit gezeigt wurde (Böhm *et al.*, 1999), ist die CoMSIA-Methode bezüglich der Wahl der Gitterweite robuster als dies bei CoMFA der Fall ist. Dies lässt sich dadurch erklären, dass bei CoMSIA Ähnlichkeitsindizes berechnet werden, deren Abstandsabhängigkeit durch eine Gaußfunktion definiert ist. Aufgrund ihrer Form lassen sich die Ähnlichkeitsindizes an jedem Punkt im Raum – also auch innerhalb des Molekülgerüsts – berechnen. Der Kurvenverlauf der Lennard-Jones- und Coulomb-Potenziale bei CoMFA erfordert hingegen die Einführung eines sogenannten *cutoff*-Parameters, um die extrem stark ansteigenden Funktionswerte bei kleiner werdenden Abständen zwischen Sondenatom und dem jeweiligen Atom des untersuchten Moleküls auf ein endliches Maß zu beschneiden. Die Unstetigkeit des sich hieraus ergebenden Potenzialverlaufs ist ein Grund für die auftretenden

Schwankungen der statistischen Werte bei CoMFA, wenn unterschiedliche Gitterweiten verwendet werden. Aufgrund der knapp achtfach größeren Datenmenge bei Verwendung des Gitters mit 1 Å Maschenweite und der damit verbundenen längeren Rechenzeit wurde daher dem 2 Å Gitter der Vorzug gegeben. Ein Nachteil bei seiner Verwendung ist jedoch eine gröbere Rasterung der Konturdiagramme, die eine graphische Interpretation schwieriger gestaltete. Tab. 2-2 zeigt die Parameter der eingesetzten Gitter mit 2 Å und 1 Å Gitterweite.

Tab. 2-2: Gitterparameter zur Berechnung der molekularen Felder.

(Werte in Å)	2 Å Gitter			1 Å Gitter		
	x	y	z	x	y	z
Untere Grenze	-16,0	-12,0	-12,0	-16,0	-12,0	-12,0
Obere Grenze	16,0	16,0	12,0	16,0	16,0	12,0
Gitterweite	2,0	2,0	2,0	1,0	1,0	1,0
Schrittzahl	17	15	13	33	29	25
Gitterpunkte	3.315			23.925		

Die räumlichen Abmessungen des Gitterkastens wurden zunächst mit Hilfe der CoMSIA-Routine in SYBYL automatisch so bestimmt, dass alle Moleküle des Datensatzes (minimiert in Thrombin) darin eingebettet sind und einen Mindestabstand von 4 Å zum Rand des Gitters aufweisen. Dann wurde im Vorgriff auf die Untersuchungen in Kap. 2.4, bei denen physikochemische Eigenschaften seitens des Proteins extrahiert wurden, das Gitter soweit expandiert, dass die direkt in die Bindetaschen hineinstehenden Aminosäurereste der vier verwendeten Serinproteasen (Thrombin, Trypsin, Faktor Xa und Urokinase) damit ebenfalls umschlossen wurden. Dieses Vorgehen hat den Vorteil, dass der Berechnung der molekularen Felder von Ligand- und Proteinseite jeweils die gleichen Gitterausmaße zugrunde liegen und einen späteren Vergleich beider Methoden vereinfacht.

Die Durchführung der Analysen mit CoMSIA erfolgte in der Weise, wie schon in Kap. 1 (vgl. Kap. 1.3.2, Seite 17) ausführlich beschrieben wurde. Kurz zusammengefasst, wurden für die fünf eingesetzten CoMSIA-Felder (sterisch, elektrostatisch, hydrophob, Donor und Akzeptor) die in SYBYL vorgegebenen Standardparameter herangezogen, die durch α

definierte Breite der Gauß'schen Abstandsfunktion wurde auf $\alpha = 0,3$ gesetzt. Die partiellen Atomladungen der Liganden wurden nach der AM1-Methode (Dewar *et al.*, 1985) mit MOPAC (Stewart, 1990) ermittelt. Die Option *CoMFA standard scaling* bei der PLS-Analyse setzt die Gewichtung der einzelnen Felder untereinander auf ein gleiches Niveau. Die *leave-one-out* Kreuzvalidierung überprüft die generierten Modelle auf ihre interne Vorhersagefähigkeit, wobei wiederum die schnellere SAMPLS-Variante eingesetzt wurde. Als Auswahlkriterium zur Bestimmung der optimalen Anzahl an Komponenten während der Kreuzvalidierung diente, wie in Kap. 1, der q^2 -Wert, der sich im Falle der Aufnahme einer zusätzlichen Komponente zum QSAR-Modell um mindestens 5% erhöhen musste („*principle of parsimony*“). Mit der gleichen Zahl an Komponenten wurde schließlich eine nicht-kreuzvalidierte Analyse mit der Option *column filtering* durchgeführt. Dabei werden alle Variablen, deren Standardabweichung an einem Gitterpunkt bezüglich aller Moleküle kleiner als ein zuvor festgelegter Grenzwert (*minimum-sigma*) ist, auf Null gesetzt und bei der eigentlichen PLS-Analyse somit nicht berücksichtigt. Somit hatte auch die oben beschriebene Erweiterung des Gitters keinen Einfluss auf das endgültige Modell, da diese äußeren Bereiche durch die Filterung eliminiert wurden. Der Grenzwert *minimum-sigma* wurde für jede Analyse so gesetzt, dass ca. 10% der Variablen in die PLS-Analyse einfließen (vgl. Kap. 1).

2.3.3.1 Verwendung eines einheitlichen Alignments für die QSAR-Modelle

In einer ersten Analyse wurden die molekularen Felder der Inhibitoren des Datensatzes berechnet, die in der Bindetasche von Thrombin minimiert wurden. Die statistischen Ergebnisse der vier QSAR-Modelle, die aus den Korrelationen mit den pK_i -Werten ($-\log K_i$) bezüglich der vier Serinproteasen resultierten (vgl. Tab. 2-1), sind in Tab. 2-3 dargestellt.

Die aus der Kreuzvalidierung erhaltenen q^2 -Werte lagen in einem Bereich von 0,326 für Faktor Xa und 0,511 für Trypsin und entsprechen somit statistisch signifikanten QSAR-Modellen ($q^2 > 0,3$). Die Vorhersagefähigkeit des Modells von Faktor Xa ist jedoch entsprechend niedriger zu bewerten als die des Modells von Trypsin. Erstaunlicherweise lag der kreuzvalidierte Wert, der aus den Thrombin-Aktivitätswerten resultierte, mit $q^2 = 0,391$ im unteren Bereich der Signifikanzskala.

Tab. 2-3: Zusammenfassung der Ergebnisse aus den PLS-Analysen unter Verwendung der in Thrombin minimierten Liganden und der Bindungsaffinitäten (pK_i -Werte) der vier Serinproteasen.

	Einzelmodelle (minimiert in Thrombin-Bindetasche)			
	Thrombin	Trypsin	Faktor Xa	Urokinase
q^2	0,391	0,511	0,326	0,444
S_{press}	0,963	0,669	0,781	0,685
r^2	0,805	0,770	0,830	0,861
S	0,544	0,459	0,392	0,343
F	70,3	77,0	65,6	82,8
Komponenten	4	3	5	5
min-sigma	0,5	0,5	0,5	0,5
Gitterweite	2 Å			

Aufgrund der Minimierung der Inhibitoren in der Thrombin-Bindetasche wäre eigentlich zu erwarten gewesen, dass insbesondere das Thrombin-Modell höhere statistische Werte liefern sollte. Die Anpassung der Strukturen an die strukturellen Gegebenheiten von Thrombin bedeutet scheinbar nicht gleichzeitig eine Optimierung der Liganden bezüglich ihres Verhaltens in der PLS-Analyse. Eine genauere Untersuchung der während der Kreuzvalidierung vorhergesagten Affinitätswerte mit den experimentellen Daten zeigte stärkere Abweichungen vor allem bei den bisbenzamidinischen Inhibitoren. Deren Aktivitäten erstrecken sich über einen relativ großen Bereich (pK_i von 3,0 bis 6,5), was umso mehr verwundert, weil sich diese Streuung zumindest auf den ersten Blick nicht in vergleichbar großen strukturellen Unterschieden widerspiegelt.

Die graphische Auswertung der QSAR-Modelle anhand der dazugehörigen Konturdiagramme ist in Abb. 2-5 am Beispiel der sterischen und Wasserstoffbrücken-Eigenschaften erläutert.

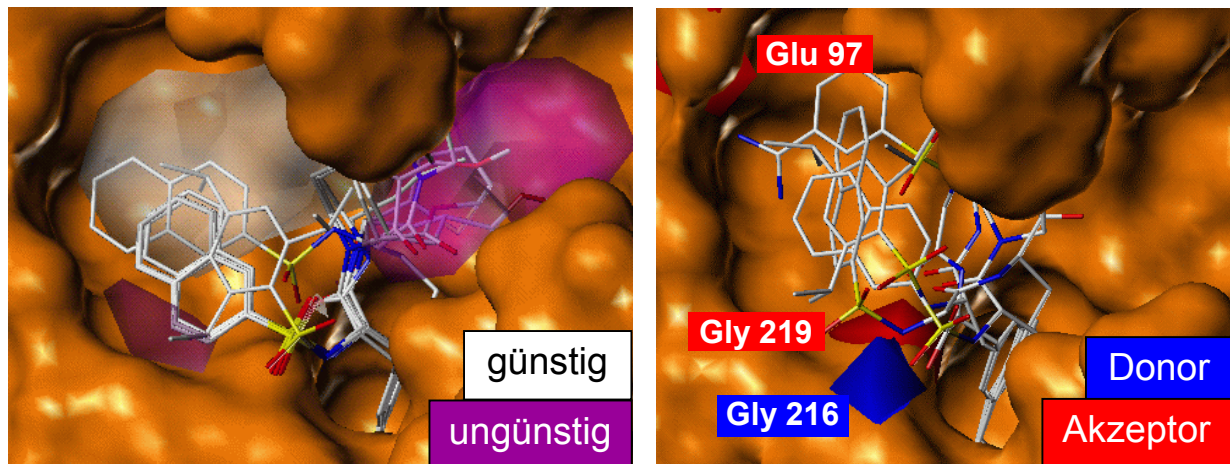


Abb. 2-5: Konturdiagramme der sterischen (1) und Wasserstoffbrücken-Eigenschaften (2) des Thrombin-Modells. Die im Modell berücksichtigten Inhibitoren wurden in der Bindetasche von Thrombin minimiert. 1: Weiß konturierte Bereiche zeigen Regionen, in denen die Anwesenheit sterisch anspruchsvoller Gruppen sich positiv auf die Bindungsaffinität auswirkt; violett gefärbte Bereiche sollten für eine Verbesserung der Affinität hingegen vermieden werden. 2: Rote Konturen weisen darauf hin, dass sich dort die Anwesenheit von Akzeptorgruppen auf Proteinseite günstig auf die Aktivitäten der Liganden auswirkt; daher sollten mögliche Donorgruppen eines Liganden in der Nähe dieser Kontur zu einer Steigerung der Bindungsaffinität beitragen. Umgekehrt werden in blau gefärbten Bereichen Donorgruppen auf Proteinseite favorisiert; folglich wirken sich Ligandgruppen mit Akzeptorfunktion in der Nähe der blauen Kontur positiv auf die Affinität aus.

Die Konturierung der sterischen Eigenschaften des Thrombin-Modells in Abb. 2-5 macht deutlich, dass eine Besetzung der S2- und S3-Tasche von Thrombin mit sterisch anspruchsvollen Gruppen zu einer Verbesserung der Affinitätswerte führt. Dies stimmt überein mit einer Vielzahl literaturbekannter, hochaffiner Thrombin-Inhibitoren, die diese Taschen sterisch beanspruchen (Obst *et al.*, 1997; Stürzebecher *et al.*, 1997; Tucker *et al.*, 1997a; Tucker *et al.*, 1997b; Tucker *et al.*, 1997c; Wagner *et al.*, 1998). Die violette Kontur im Bereich der S1'- und S2'-Taschen hingegen zeigt an, dass sich ihre sterische Besetzung ungünstig auf die Bindungsaffinität potenzieller Liganden auswirkt. Eine weitere Region, deren Besetzung mit sterisch anspruchsvollen Gruppen als ungünstig vorhergesagt wird, befindet sich im unteren Bereich der S3-Bindetasche; diese markierte Region ragt teilweise schon in die lösungsmittelzugängliche Oberfläche des Proteins. Zu große Substituenten (zum Beispiel TIPPS-Derivate; siehe Seite 60), die diese Kontur sterisch besetzen, sollten demnach nur unter Affinitätsverlust in der S3-Tasche binden. Die Konturierung der

Wasserstoffbrücken-Eigenschaften des Thrombin-Modells markiert Bereiche, deren Besetzung durch mögliche Donor- oder Akzeptorgruppen des Proteins als günstig vorhergesagt wird. Tatsächlich werden in diesen konturierten Bereichen funktionelle Gruppen der Aminosäuren Glu 97, Gly 216 und Gly 219 gefunden, die den geforderten Eigenschaften des Modells entsprechen. Komplementär zu diesen auf der Proteinseite vorhergesagten Eigenschaften wirken sich Akzeptor- bzw. Donorgruppen eines Liganden in der Nähe dazu vorteilhaft auf die Bindungsaffinität aus.

2.3.3.2 Auswirkungen des Alignments auf die Ergebnisse

Nachdem die anhand der Affinitätsdaten für die vier Serinproteasen abgeleiteten Modelle bisher auf Bindungsmodi der Liganden basierten, die in der Thrombin-Bindetasche minimiert wurden, fand im weiteren eine Optimierung der Bindungsgeometrien in der Bindetasche von Faktor Xa statt. Unterschiede in den statistischen Ergebnissen als auch bei der graphischen Auswertung können eine Abschätzung darüber geben, inwieweit die QSAR-Modelle abhängig von einer Adaption des Alignments der zugrundeliegenden Strukturen sind. Tab. 2-4 zeigt die statistischen Ergebnisse der Modelle bei Verwendung der in Faktor Xa minimierten Bindungsmodi.

Tab. 2-4: Zusammenfassung der Ergebnisse aus den PLS-Analysen unter Verwendung der in Faktor Xa minimierten Liganden und der Bindungsaffinitäten (pK_i -Werte) der vier Serinproteasen.

	Einzelmodelle (minimiert in Faktor Xa-Bindetasche)			
	Thrombin	Trypsin	Faktor Xa	Urokinase
q^2	0,515	0,534	0,306	0,405
S_{press}	0,889	0,655	0,735	0,696
r^2	0,915	0,844	0,797	0,810
S	0,373	0,380	0,397	0,394
F	116,4	71,2	51,9	56,1
Komponenten	6	5	5	5
min-sigma	0,5	0,5	0,5	0,5
Gitterweite	2 Å			

Es konnten in allen vier Fällen statistisch signifikante Werte mit q^2 -Werten $> 0,3$ erhalten werden. Ähnlich wie bei den in Thrombin minimierten Strukturen (vgl. Tab. 2-3) ergab sich das beste Modell bei Verwendung der Trypsin-Affinitäten mit einem $q^2 = 0,534$. Erstaunlicherweise konnte mit den in Faktor Xa angepassten Liganden ein besseres Thrombin-Modell erzeugt werden als dies mit den in der Thrombin-Bindetasche minimierten Verbindungen der Fall war. Im Gegensatz dazu weist das Modell mit den Affinitätsdaten von Faktor Xa einen q^2 -Wert von nur 0,306 auf, liegt also an der unteren statistischen Signifikanzgrenze. Eigentlich wäre – ähnlich zu den Thrombin-minimierten Strukturen (siehe oben) – ein besseres Faktor Xa-Modell zu erwarten gewesen, da die Inhibitoren auch an die Bindetasche von Faktor Xa angepasst wurden. Dies bestätigt oben getroffene Annahme, dass die Optimierung eines Datensatzes bezüglich einer Bindetasche nicht zwangsläufig zu Modellen mit höherer statistischer Signifikanz führen muss. Hierfür ist ebenfalls wichtig, dass die Liganden eines Datensatzes räumlich gleichverteilt vorliegen und die Unterschiede in den physikochemischen Eigenschaften der Verbindungen untereinander auch mit ihren Affinitätswerten korrelieren. Diese Gegebenheiten können jedoch bei verschiedenen Bindungsmodi der Liganden und einem dadurch veränderten Alignment variieren und so zu den beobachteten Unterschieden in den statistischen Ergebnissen führen.

Ein weiterer Grund für die vergleichsweise niedrigen q^2 -Werte einiger Modelle ist die relativ hohe strukturelle Heterogenität der Liganden des in dieser Arbeit verwendeten Datensatzes. Im Vergleich zu den in einer früheren Arbeit (Böhm *et al.*, 1999) untersuchten Serinprotease-Inhibitoren, die von einem einzigen Grundgerüst abgeleitet waren und an nur zwei Stellen Substitutionen aufwiesen, basiert der jetzt verwendete Datensatz auf mehr als zehn strukturell unterschiedlichen Klassen. Im allgemeinen gestaltet sich die Generierung konsistenter QSAR-Modelle schwieriger, wenn strukturell heterogenere Verbindungen in eine Struktur-Wirkungsbeziehung einfließen, die die physikochemischen Eigenschaften der Liganden mit den entsprechenden Bindungsaffinitäten korrelieren kann. Der strukturell relativ homogene Datensatz der früheren Arbeit konnte daher auch hohe statistische Ergebnisse mit q^2 -Werten $> 0,6$ erzielen. Außerdem gestaltete es sich im vorliegenden Fall als äußerst schwierig, mit einem Datensatz statistisch signifikante Modelle bezüglich der Affinitäten zu den vier Serinproteasen gleichzeitig zu erzeugen.

Am Beispiel des Modells von Faktor Xa sind in Abb. 2-6 die Konturdiagramme der sterischen und hydrophoben Eigenschaften gezeigt.

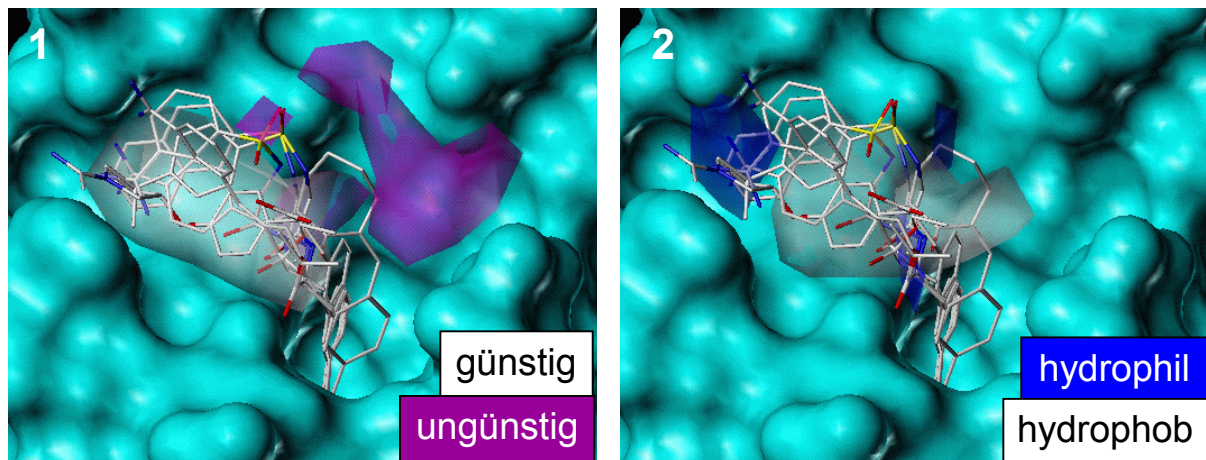


Abb. 2-6: Konturdiagramme der sterischen (1) und hydrophoben Eigenschaften (2) des Faktor Xa-Modells. Die im Modell berücksichtigten Inhibitoren wurden in der Bindetasche von Faktor Xa minimiert. 1: Weiß konturierte Bereiche zeigen Regionen, in denen die Anwesenheit sterisch anspruchsvoller Gruppen positiv mit der Bindungsaffinität korreliert; die Besetzung violetter Bereiche sollte für eine Verbesserung der Affinität hingegen vermieden werden. 2: Weiße Konturen markieren Bereiche, deren Besetzung mit hydrophoben Gruppen zu einer Erhöhung der Bindungsaffinität führen; entsprechend werden hydrophile Gruppen in blauen Bereichen zur Aktivitätserhöhung favorisiert.

Das Konturdiagramm der sterischen Eigenschaften des Faktor Xa-Modells markiert einen großen Teil der S3/4-Tasche, der als günstig für die Positionierung sterisch anspruchsvoller Gruppen vorhergesagt wird. Tatsächlich sind viele potenzielle, selektiv bindende Faktor Xa-Inhibitoren bekannt, die diese Region besetzen (Klein *et al.*, 1998; Galemme *et al.*, 2000). Vergleichbar zu den sterischen Eigenschaften des Thrombin-Modells (vgl. Abb. 2-5) ist auch hier die Besetzung der S1'/S2'-Tasche ungünstig für eine Erhöhung der Bindungsaffinität. Zusätzlich sagt das Modell von Faktor Xa auch einen Teil der S2-Tasche, dessen Besetzung beim Thrombin-Modell favorisiert wird, als ungünstig für eine sterische Besetzung vorher. Dies stimmt ebenfalls mit der Analyse literaturbekannter Faktor Xa-Inhibitoren überein, die nicht notwendigerweise sterisch anspruchsvolle P2-Substituenten (zur S2-Tasche zugehörig) besitzen und trotzdem Affinitäten im nanomolaren Bereich aufweisen (Gong *et al.*, 2000; Rai *et al.*, 2001; Shrader *et al.*, 2001).

Aus der Analyse der hydrophoben Eigenschaften des Faktor Xa-Modells resultiert ein weiß konturierter Bereich in der S3-Tasche, der die Besetzung mit hydrophoben Gruppen als günstig vorhersagt. In seiner Nachbarschaft befindet sich eine weitere, blau gefärbte Kontur, die angibt, dass hier Gruppen mit hydrophilen Eigenschaften positiv mit der Bindungsaffinität korrelieren. Beide Konturen spiegeln die Merkmale vieler hochpotenten Inhibitoren für Faktor Xa in der Literatur wider. Alle diese Verbindungen weisen neben spezifischen P1-Substituenten (zur S1-Tasche zugehörig) ein hydrophobes Zentrum und nachbarständig dazu terminale Substituenten mit hydrophilen Eigenschaften auf, die in der S3/4-Tasche zu liegen kommen.

Besonders erwähnenswert ist die Gruppe der bisamidinischen Faktor Xa-Inhibitoren, die diese Kriterien erfüllen und auch in dem in dieser Arbeit verwendeten Datensatz berücksichtigt sind (vgl. Bisbenzamidine in Abb. 2-2). Eine Amidin-Gruppe bildet in der S1-Spezifitätstasche eine Salzbrücke zu Asp 189 aus; der im wesentlichen hydrophobe Kernbereich der Bisamidine wechselwirkt mit den aromatischen Seitenketten von Tyr 99, Phe 174 und Trp 215, während die zweite basische Amidin-Gruppe mit den Carbonylgruppen von Glu 97, Thr 98 und Ile 175 am Rande der S3/4-Tasche interagiert (Stubbs, 1996; Renatus *et al.*, 1998). Mittlerweile ist eine Vielzahl von Bisamidin-Inhibitoren bekannt, die vor allem gegenüber Thrombin selektiv an Faktor Xa binden (Maduskuie *et al.*, 1998; Shaw *et al.*, 1998; Whitlow *et al.*, 1999; Phillips *et al.*, 1999; Guilford *et al.*, 1999). Ein bekannter Faktor Xa-Inhibitor, die Daiichi-Verbindung DX-9065a²⁹ (Hara *et al.*, 1994; Kumura *et al.*, 2000), die auch in der klinischen Phase erste Fortschritte zeigt (Herbert *et al.*, 1996; Murayama *et al.*, 1999), weist beispielsweise eine ca. 50.000-fach höhere Affinität zu Faktor Xa ($K_i = 0,04 \mu\text{M}$ im Vergleich zu Thrombin ($K_i > 2.000 \mu\text{m}$) auf³⁰ (Katakura *et al.*, 1993).

²⁹ IUPAC-Name: (+)-2S-2-[4-[[[(3S)-1-Acetimido-3-pyrrolidinyl]oxy]phenyl]-3-[7-amidino-2-naphthyl]propansäure Hydrochlorid Pentahydrat

³⁰ In dieser Arbeit wurden für die Daiichi-Verbindung (Nr. 14 in Tab. 2-1) die Werte $K_i = 0,024 \mu\text{M}$ (Faktor Xa) und $K_i = 70 \mu\text{M}$ (Thrombin) verwendet.

2.3.4 Selektivitätsmodelle durch Differenzbildung von Affinitäten

Wie bei der Beschreibung literaturbekannter Ansätze schon diskutiert (siehe Kap. 2.2, Seite 49), können Informationen über selektivitätsbestimmende Kriterien dadurch ermittelt werden, indem die Affinitätswerte bezüglich zweier Proteine verknüpft werden und so in ein „Selektivitätsmodell“ einfließen. Dabei ist es gleichwertig, ob die Quotienten von K_i - oder IC_{50} -Werten berechnet oder die Differenzen der zugehörigen pK_i - bzw. pIC_{50} -Werte gebildet werden³¹. Das Ergebnis dieser Affinitätsdifferenzen (im folgenden auch als „Selektivitätsfaktoren“ bezeichnet) ist ein Maß für die Selektivität eines Liganden zu einem Protein gegenüber dem anderen Protein.

2.3.4.1 Streuung der Affinitätsdifferenzen

Die berechneten Selektivitätsfaktoren sind in Tab. 2-5 für die 80 Verbindungen des Datensatzes zu allen sechs möglichen, paarweisen Kombinationen der vier Serinproteasen Thrombin (Thr), Trypsin (Try), Faktor Xa (Fxa) und Urokinase (Uro) aufgeführt. Dabei bedeuten beispielsweise die Werte in der Spalte „Thr-Fxa“, dass die Differenzen zwischen den pK_i -Werten von Thrombin und Faktor Xa ($\Delta pK_i = pK_{i,Thr} - pK_{i,Fxa}$) gebildet wurden. Hohe positive Zahlen weisen daher auf selektiv an Thrombin bindende Inhibitoren hin, wohingegen negative Zahlen mit hohen Absolutwerten die Selektivität zu Faktor Xa andeuten.

Der Wertebereich der in Tab. 2-5 berechneten Selektivitätsfaktoren erstreckt sich bei allen sechs paarweisen Kombinationen auf mehr als drei logarithmische Einheiten und sollte somit ausreichend für die Erzeugung statistisch signifikanter, aussagekräftiger QSAR-Modelle sein. Die Kombinationen Thrombin und Faktor Xa bzw. Thrombin und Urokinase weisen sogar eine Bandbreite von über sieben bzw. sechs Zehnerpotenzen auf.

³¹ Dabei gilt: $pK_{i,1} - pK_{i,2} = \log(K_{i,2} / K_{i,1})$; entsprechend bei IC_{50} - bzw. pIC_{50} -Werten

Tab. 2-5: Affinitätsdifferenzen (Selektivitätsfaktoren) der 80 Verbindungen des Datensatzes, sortiert nach ihren elf verschiedenen Strukturklassen. Die sechs möglichen, paarweisen Kombinationen zwischen Thrombin (Thr), Trypsin (Try), Faktor Xa (Fxa) und Urokinase (Uro) sind als Differenzen der jeweiligen pK_i -Werte gegeben (ΔpK_i).

Name (Nr_Strukturklasse_Substituent)	Thr-Try ΔpK_i	Thr-Fxa ΔpK_i	Thr-Uro ΔpK_i	Try-Fxa ΔpK_i	Try-Uro ΔpK_i	Fxa-Uro ΔpK_i
01_BA_H	-0,798	0,270	-0,087	1,069	0,711	-0,358
02_BA_NH2	-1,602	-1,022	-0,854	0,580	0,748	0,168
03_BA_CH2OPH	-0,865	-1,129	-0,342	-0,264	0,523	0,787
04_BA_CH2COPH	-1,167	-0,584	0,018	0,583	1,185	0,602
05_BA_CH2COCOOH	-1,642	-1,264	-0,642	0,378	1,000	0,622
06_BA_CH2CH2COPH	-0,860	-0,064	-0,316	0,796	0,544	-0,252
07_NAPHTH_H	-0,753	-0,110	-1,213	0,643	-0,460	-1,104
08_NAPHTH_NH2	-1,104	-0,519	-1,056	0,585	0,047	-0,538
09_NAPHTH_NHAC	-1,106	-0,347	-1,038	0,759	0,068	-0,691
10_NAPHTH_ONBU	-0,523	0,097	-0,347	0,620	0,176	-0,444
11_NAPHTH_OCH2PH	-0,209	0,393	0,131	0,602	0,341	-0,262
12_NAPHTH_COOME	-1,127	1,201	-1,007	2,328	0,120	-2,208
13_NAPHTH_CH2COPH	-0,840	0,087	-0,875	0,927	-0,035	-0,962
14_NAPHTH_DAIICHI	-2,430	-3,465	-0,327	-1,035	2,104	3,138
15_BISBENZ_CH7CH4AM	-0,551	-1,391	1,026	-0,840	1,577	2,418
16_BISBENZ_CH5CH3AM	-0,925	-1,034	0,258	-0,109	1,183	1,292
17_BISBENZ_CH26OHCH24AM	0,113	0,683	0,911	0,570	0,798	0,228
18_BISBENZ_PHPH3AM	-1,207	-0,101	-0,015	1,106	1,192	0,085
19_BISBENZ_OCH2PH4AM	-1,465	0,617	-1,347	2,082	0,118	-1,964
20_BISBENZ_OCH25O4AM	-0,602	0,059	0,495	0,661	1,097	0,436
21_BISBENZ_OET3AM	-0,329	-1,028	0,266	-0,699	0,595	1,294
22_BISBENZ_OET4AM	-1,745	-2,553	-0,959	-0,808	0,786	1,594
23_BISBENZ_OH4AM	-1,000	-1,881	-0,287	-0,881	0,713	1,594
24_BISBENZ_HPHNHPH4AM	-1,938	-0,205	-1,582	1,732	0,355	-1,377
25_4NAPAP_BNAS_PPD	2,061	3,119	4,584	1,059	2,523	1,464
26_4NAPAP_BA_PPD	0,740	1,648	2,699	0,908	1,959	1,051
27_4NAPAP_BNAS_PZD-NH	0,675	1,658	3,105	0,983	2,430	1,447
28_4NAPAP_BNAS_PRO-OH	0,615	1,497	3,196	0,882	2,581	1,699
29_4NAPAP_TOS_NHPH	0,028	-0,523	0,792	-0,551	0,764	1,315
30_4NAPAP_TOS_OME	1,361	1,432	2,830	0,071	1,469	1,398
31_3NAPAP_BNAS_PPD-4ME	1,106	0,842	2,444	-0,265	1,337	1,602
32_3NAPAP_BNAS_PZD-NSME	-2,015	0,513	0,920	2,528	2,935	0,407
33_3NAPAP_BNAS_NHME	-0,018	-0,505	1,509	-0,487	1,528	2,014
34_3NAPAP_3AMPS_OH	-0,820	-1,226	-0,527	-0,406	0,293	0,699
35_3NAPAP_TOS_OME	0,613	-0,667	1,228	-1,280	0,615	1,895
36_3NAPAP_MTR_OME	0,945	0,581	2,418	-0,364	1,473	1,837
37_3NAPAP_TBUPS_OME	0,487	-0,566	1,968	-1,054	1,480	2,534
38_3NAPAP_BNAS_OIPR	0,912	-0,699	1,611	-1,611	0,699	2,310
39_4TAPAP_BNAS_PPD	1,738	2,553	3,222	0,814	1,483	0,669

Name (Nr_Strukturklasse_Substituent)	Thr-Try ΔpK_i	Thr-Fxa ΔpK_i	Thr-Uro ΔpK_i	Try-Fxa ΔpK_i	Try-Uro ΔpK_i	Fxa-Uro ΔpK_i
40_4TAPAP_BNAS_PZD-NAC	0,430	0,602	1,664	0,172	1,234	1,062
41_4TAPAP_BNAS_OH	-0,216	-0,061	0,638	0,155	0,854	0,699
42_4TAPAP_BNAS_ONBU	1,774	1,546	2,251	-0,228	0,477	0,706
43_3TAPAP_H_PPD-4ME	0,031	0,930	1,472	0,899	1,441	0,541
44_3TAPAP_BNAS_PPD-4PPO	-1,176	0,166	0,675	1,342	1,851	0,509
45_3TAPAP_BNAS_PPD-4BZL	-1,716	0,254	0,032	1,970	1,748	-0,222
46_3TAPAP_BNAS_PIP-OME	1,434	2,818	4,058	1,384	2,624	1,240
47_3TAPAP_BNAS_PIP-3MOME	0,539	1,840	2,983	1,301	2,444	1,143
48_3TAPAP_BNAS_NIP-OET	-1,255	2,067	1,319	3,322	2,574	-0,748
49_3TAPAP_BNAS_NIP-OBZL	-0,273	1,457	1,753	1,730	2,026	0,296
50_3TAPAP_BNAS_INIP-OME	0,326	3,403	2,738	3,077	2,412	-0,665
51_3TAPAP_BNAS_INIP-OCHX	-1,419	0,559	0,678	1,978	2,097	0,119
52_3TAPAP_BNAS_INIP-NBZL	-1,060	1,053	1,482	2,113	2,542	0,429
53_3TAPAP_BNAS_INIP-OBZL	-0,496	1,586	1,484	2,082	1,980	-0,102
54_3TAPAP_BNAS_2PPN3COOM	1,301	1,097	2,894	-0,204	1,593	1,797
55_3TAPAP_BNAS_PZD-NSME	1,760	3,824	4,115	2,064	2,355	0,291
56_3TAPAP_BNAS_PZD-NBZL	-0,699	0,505	1,860	1,204	2,559	1,355
57_3TAPAP_TOS_NHME	0,407	0,347	1,444	-0,061	1,036	1,097
58_3TAPAP_TMEPS_NHME	1,357	1,163	1,357	-0,194	0,000	0,194
59_3TAPAP_BNAS_NHIPR	0,412	1,764	2,208	1,352	1,796	0,444
60_3TAPAP_BNAS_OME	0,951	0,951	3,030	0,000	2,079	2,079
61_3TAPAP_BNAS_ONBU	1,301	2,377	3,649	1,076	2,348	1,272
62_3TAPAP_BNAS_TCD	1,097	1,283	2,637	0,186	1,540	1,354
63_3TAPAP_BNAS_PRO-OME	1,000	0,808	1,553	-0,192	0,553	0,745
64_TIPPS_NIP-OBZL	-0,025	0,000	0,125	0,025	0,150	0,125
65_TIPPS_INIP-OH	-0,138	-0,714	0,763	-0,576	0,901	1,477
66_TIPPS_INIP-OET	-0,408	1,301	0,535	1,709	0,944	-0,766
67_TIPPS_PZD-NBUO	-1,347	0,336	0,996	1,683	2,343	0,661
68_TIPPS_PZD-NCOOET	-0,747	0,350	-0,041	1,097	0,706	-0,391
69_TIPPS_NHME	0,835	0,462	1,420	-0,374	0,585	0,959
70_TIPPS_NH-PIPYL	-0,110	2,242	2,519	2,352	2,628	0,276
71_TIPPS_NHCH2-TFN	1,388	1,944	3,771	0,556	2,383	1,827
72_GLYCIN_3AM_TOS_PYR	-0,827	-0,136	0,143	0,692	0,971	0,279
73_GLYCIN_3AM_TOS_MOR	-1,088	0,295	0,611	1,383	1,699	0,316
74_GLYCIN_4AM_BNAS_PPD	0,540	-0,018	1,310	-0,558	0,770	1,328
75_BUTYRO_4AM_TOS_NHPH	-2,071	-0,523	-0,125	1,548	1,946	0,398
76_BUTYRO_4AM_BNAS_NHPH	-1,776	-0,815	0,413	0,960	2,189	1,228
77_BUTYRO_4AM_BNAS_MOR	-0,683	-0,859	0,886	-0,176	1,569	1,745
78_VALERI_3AM_BNAS_MOR	1,064	1,579	3,925	0,515	2,862	2,347
79_VALERI_4AM_BNAS_PPD	1,180	1,561	1,824	0,380	0,643	0,263
80_VALERI_4AM_TOS_NHPH	-0,591	-0,602	-0,296	-0,011	0,296	0,306
Mittelwert	-0,191	0,439	1,120	0,629	1,310	0,681
Standardabweichung	1,094	1,300	1,484	1,033	0,853	1,024
Maximalwert	2,061	3,824	4,584	3,322	2,935	3,138
Minimalwert	-2,430	-3,465	-1,582	-1,611	-0,460	-2,208
Wertebereich	4,491	7,289	6,166	4,933	3,395	5,346

Die ebenfalls angegebenen Mittel-, Maximal- und Minimalwerte der Differenzen geben Aufschluss darüber, inwieweit eine ausgewogene Verteilung der Selektivitätsfaktoren gegenüber beiden Proteinen gegeben ist. So existieren im Falle der Selektivitäten zu Thrombin und Trypsin gleichermaßen Verbindungen, die hochselektiv an Thrombin als auch an Trypsin binden. Der Inhibitor 25_4NAPAP_BNAS_PPD bindet beispielsweise mit einem Selektivitätsfaktor von +2,061 mehr als 100-fach stärker an Thrombin als an Trypsin; die Daiichi-Verbindung 14_NAPHTH_DAIICHI hingegen bindet ca. 270-fach stärker (Selektivitätsfaktor -2,430) an Trypsin als an Thrombin³². Noch stärker fallen die selektivitätsdiskriminierenden Faktoren beim Vergleich von Thrombin und Faktor Xa aus: Inhibitor 55_3TAPAP_BNAS_PZD-NSME mit einem Wert von +3,824 zeigt eine 6.600-fach höhere Affinität zu Thrombin; dagegen bindet die Daiichi-Struktur 14_NAPHTH_DAIICHI knapp 3.000-fach stärker an Faktor Xa (Selektivitätsfaktor -3,465). Dies ist dadurch erklärbar, dass erstere Verbindung Substituenten besitzt, die optimal die S2- und S3-Bindetasche von Thrombin ausfüllen; die bekannte Faktor Xa-selektive Daiichi-Verbindung ist so aufgebaut, dass sie geeignete Wechselwirkungen mit der S3/4-Tasche von Faktor Xa eingehen kann (siehe Kap. 2.3.3, Seite 79). Ein Mittelwert der Selektivitätsfaktoren von +1,120 zwischen Thrombin und Urokinase (Tab. 2-5) deutet auf eine signifikante Verschiebung der Affinitätsdifferenzen zugunsten affinerer Bindung an Thrombin hin. So bindet die selektivste Urokinase-Verbindung³³ 24_BISBENZ_HPHNHPH4AM nur knapp 40-fach stärker (Selektivitätsfaktor -1,582) an Urokinase als an Thrombin, wohingegen der Inhibitor 25_4NAPAP_BNAS_PPD mit der höchsten Selektivität zu Thrombin fast um den Faktor von 40.000 diskriminiert (Selektivitätsfaktor +4,584). Noch extremer fällt das Ungleichgewicht bei Trypsin und Urokinase aus (Mittelwert +1,310): hier weist der Inhibitor mit der größten Urokinase-Selektivität 07_NAPHTH_H eine nur knapp 3-fach höhere Affinität zu Urokinase auf (Selektivitätsfaktor -0,460). Dies ist dadurch erklärbar, dass bis jetzt noch keine hochaffinen Inhibitoren zu Urokinase gefunden wurden, wie es bei Thrombin und Trypsin bereits gelungen ist (vgl. Tab. 2-1 und Bemerkung auf Seite 63). Bezogen auf die absoluten Zahlenwerte ist somit eine größere Zahl von Inhibitoren auch affiner zu Thrombin und Trypsin als zu Urokinase.

³² Berechnung: $10^{2,061} \approx 115$ bzw. $10^{2,430} \approx 270$

³³ die am besten zwischen Urokinase und Thrombin diskriminierende Verbindung bei gleichzeitig besserer Affinität gegenüber Urokinase

2.3.4.2 Verwendung eines einheitlichen Alignments für die Selektivitätsmodelle

Entsprechend den sechs Kombinationen von Selektivitätsfaktoren aus Tab. 2-5 können – wie bei den vorher berechneten Einzelmodellen – die gewünschten Selektivitätsmodelle generiert werden. Bei diesen Modellen kommen anstelle der Affinitäten zu den einzelnen Rezeptoren als abhängige Variablen die Affinitätsdifferenzen zwischen zwei Proteinen zum Einsatz. Wie schon bei den Einzelmodellen separat durchgeführt, ließen sich Modelle zu den unterschiedlich berechneten Bindungsmodi erzeugen, indem die Platzierungen verwendet wurden, die durch Minimierung der Strukturen in einem Referenzprotein (hier: Thrombin bzw. Faktor Xa) entstanden. Tab. 2-6 fasst die Ergebnisse der PLS-Analysen zusammen, die mit den in Thrombin minimierten Strukturen erhalten wurden.

Tab. 2-6: Zusammenfassung der Ergebnisse aus den PLS-Analysen der sechs paarweisen Kombinationsmöglichkeiten unter Verwendung der in Thrombin minimierten Bindungsmodi und der Selektivitätsfaktoren aus Tab. 2-5.

	Selektivitätsmodelle (minimiert in Thrombin-Bindetasche)					
	Thr-Try	Thr-Fxa	Thr-Uro	Try-Fxa	Try-Uro	Fxa-Uro
q^2	0,409	0,354	0,529	0,328	0,287	0,385
S_{press}	0,856	1,055	1,062	0,842	0,715	0,805
r^2	0,811	0,640	0,845	0,552	0,401	0,648
S	0,484	0,788	0,609	0,687	0,655	0,510
F	72,9	62,1	92,7	43,1	47,5	42,4
Komponenten	4	2	4	2	1	3
min-sigma	0,5	0,5	0,5	0,5	0,5	0,5
Gitterweite	2 Å					

Bei Annahme einer unteren Signifikanzgrenze von $q^2 = 0,3$ ergibt die Analyse der aus den Selektivitätsmodellen resultierenden Parametern statistisch signifikante Resultate mit q^2 -Werten von 0,328 bis 0,529. Das Modell aus der Kombination von Trypsin und Urokinase liegt bei Verwendung von nur einer Komponente mit $q^2 = 0,287$ knapp unter dieser Grenze. Dies liegt zum einen an dem relativ niedrigen Wertebereich (ca. 3,4 ΔpK_i -Einheiten) der dazugehörigen Selektivitätsfaktoren; zum anderen sind die Affinitäten zu Trypsin und

Urokinase deutlicher interkorreliert ($r = 0,58$; vgl. Abb. 2-3 auf Seite 64) als bei den anderen Kombinationen. Nur die Affinitäten von Thrombin und Trypsin korrelieren ähnlich stark miteinander ($r = 0,52$); dies kann jedoch wahrscheinlich durch den größeren Wertebereich der Affinitätsdifferenzen von knapp $4,5 \Delta pK_i$ -Einheiten kompensiert werden.

Analog zu den Selektivitätsmodellen, die auf den Thrombin-minimierten Strukturen basierten, wurden unter Verwendung der Affinitätsdifferenzen auch QSAR-Modelle generiert, bei denen die in Faktor Xa eingepassten Inhibitoren zugrunde lagen. Tab. 2-7 fasst die resultierenden statistischen Parameter zusammen.

Tab. 2-7: Zusammenfassung der Ergebnisse aus den PLS-Analysen der sechs paarweisen Kombinationsmöglichkeiten unter Verwendung der in Faktor Xa minimierten Bindungsmodi und der Selektivitätsfaktoren aus Tab. 2-5.

Selektivitätsmodelle (minimiert in Faktor Xa-Bindetasche)						
	Thr-Try	Thr-Fxa	Thr-Uro	Try-Fxa	Try-Uro	Fxa-Uro
q^2	0,561	0,363	0,643	0,331	0,300	0,354
S_{press}	0,741	1,110	0,929	0,831	0,700	0,800
r^2	0,893	0,909	0,905	0,626	0,400	0,784
S	0,365	0,419	0,481	0,622	0,649	0,462
F	110,7	108,7	125,1	37,9	46,3	61,0
Komponenten	5	6	5	3	1	4
min-sigma	0,5	0,5	0,5	0,5	0,5	0,5
Gitterweite	2 Å					

Auch hier ergaben sich für alle sechs Selektivitätsmodelle statistisch signifikante Modelle mit q^2 -Werten größer bzw. genau gleich 0,3 im Falle des Modells von Trypsin und Urokinase. In diesem Fall wurde (wie in Tab. 2-6) die Verwendung von nur einer Komponente bei der Kreuzvalidierung als optimal angegeben. Beim direkten Vergleich der q^2 -Werte mit den Ergebnissen der Thrombin-minimierten Strukturen fällt der gleichbleibende Trend der Resultate relativ zueinander auf, teilweise konnten jedoch signifikant höhere Werte bei Verwendung der Faktor Xa-eingepassten Liganden erzielt werden.

So resultierten in Tab. 2-7 die besten Modelle bei den Kombinationen von Thrombin und Trypsin mit $q^2 = 0,561$ und bei Thrombin und Urokinase mit $q^2 = 0,643$. Die entsprechenden Werte der Selektivitätsmodelle, die in Thrombin minimiert wurden, liegen mit $q^2 = 0,409$ bzw. $q^2 = 0,529$ jeweils niedriger. Die statistischen Ergebnisse der übrigen vier Modelle sind relativ ähnlich zu denen aus Tab. 2-6. Die Gründe für die auch hier schlechte Bewertung des Differenzmodells aus Trypsin und Urokinase entsprechen denen des Modells, das auf dem Thrombin-Alignment beruht, da sich die Datenlage zur Generierung der Modelle nicht verändert hat.

Insgesamt gestaltet es sich jedoch als äußerst komplex, die genauen Ursachen der abweichenden Ergebnisse zu bestimmen. Zu einer fundierten Begründung müsste theoretisch untersucht werden, welche Bindungsmodi bei den Minimierungen in den unterschiedlichen Proteinbindetaschen besonders stark verändert werden und ob diese Abweichungen mit den Bindungsaffinitäten bzw. deren Differenzen korrelieren.

2.3.4.3 *Graphische Interpretation der Selektivitätsmodelle*

Die graphische Interpretation der erhaltenen Selektivitätsmodelle erlaubt einen tieferen Einblick, welche Faktoren entscheidend für die Selektivität zu einem bestimmten Proteinpaaar sein können. Die Auswertung der Konturdiagramme erlaubt Aussagen darüber zu treffen, welche Bereiche der Proteinbindetasche diskriminierend für eine Selektivität sind und welche physikochemischen Eigenschaften dafür verantwortlich sind. Im folgenden werden einige Beispiele aufgeführt und gezeigt, wie die graphische Interpretation der Selektivitätsmodelle erfolgen kann. Aus Platz- und Übersichtlichkeitsgründen kann in diesem Zusammenhang keine vollständige Auswertung aller möglichen Konturdiagramme erfolgen.

In Abb. 2-7 sind die Konturdiagramme der sterischen Eigenschaften der beiden Thrombin- und Faktor Xa-Einzelmodelle als auch des Selektivitätsmodells von Thrombin und Faktor Xa dargestellt.

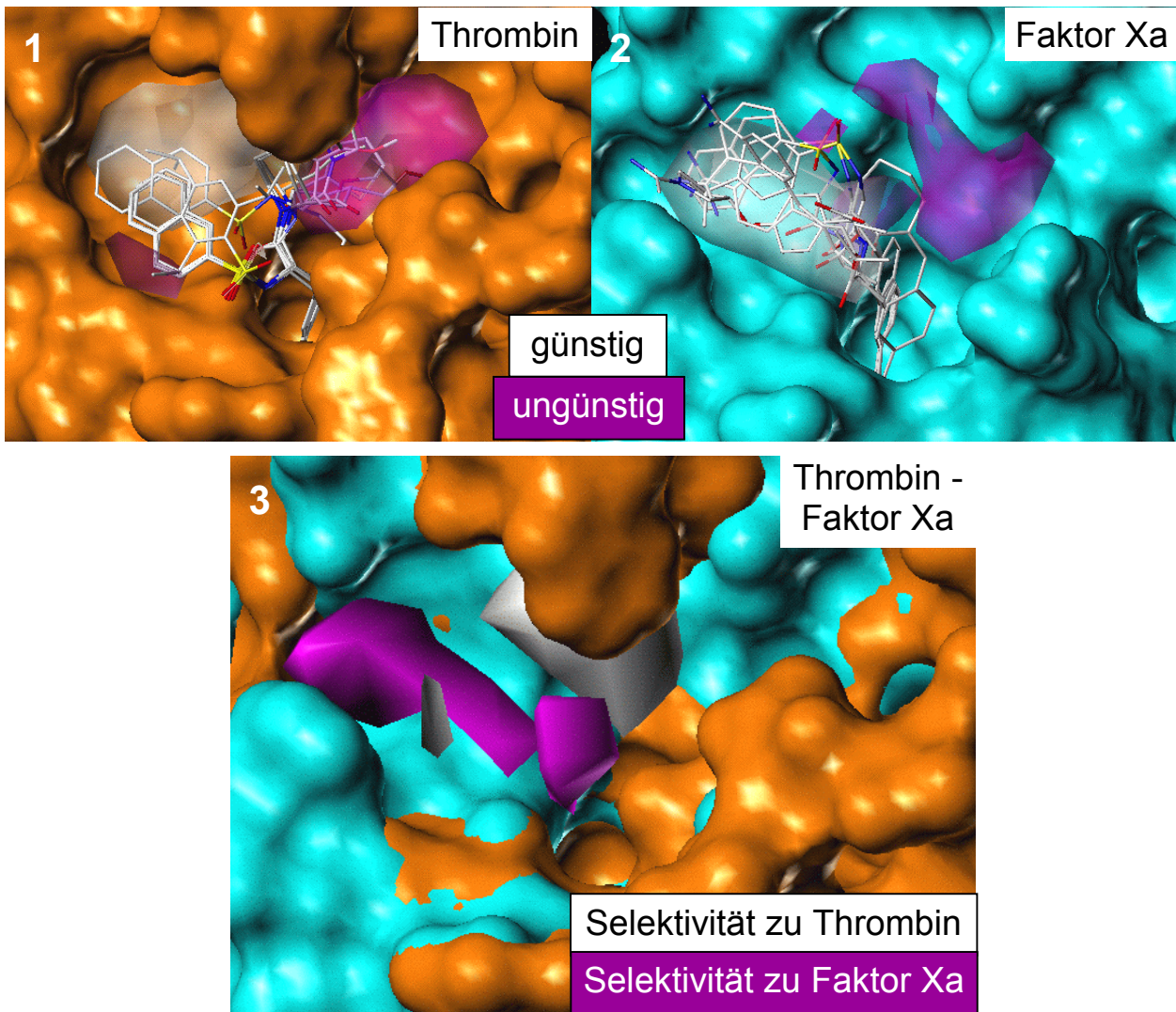


Abb. 2-7: Konturdiagramme der sterischen Eigenschaften der Einzelmodelle von Thrombin (1) und Faktor Xa (2) und des Selektivitätsmodells (3), das aus den Affinitätsdifferenzen zwischen Thrombin und Faktor Xa erzeugt wurde. Die bei allen Modellen verwendeten Inhibitoren wurden in der Bindetasche von Thrombin minimiert. Die Konturdiagramme 1 und 2 entsprechen denen aus Abb. 2-5(1) und Abb. 2-6(1). 1+2: Weiß konturierte Bereiche zeigen Regionen, in denen die Anwesenheit sterisch anspruchsvoller Gruppen positiv mit der Bindungsaffinität korreliert; violette Bereiche sollten für eine Verbesserung der Affinität hingegen sterisch unbesetzt bleiben. 3: Die Besetzung des weiß konturierten Bereichs mit sterisch anspruchsvollen Gruppen erhöht die Selektivität eines Liganden zu Thrombin, während ihre Anwesenheit in dem violetten Bereich die Selektivität zu Faktor Xa fördert. Zusätzlich sind die lösungsmittelzugänglichen Oberflächen beider Proteine übereinandergelegt.

Während bei den Einzelmodellen die konturierten Bereiche angeben, ob eine bestimmte physikochemische Eigenschaft dort platziert zu einer Verbesserung oder Reduktion der Bindungsaffinität führt, markieren die Konturen des Selektivitätsmodells die Bereiche, die diskriminierend für die Steigerung oder Reduktion der Selektivität eines möglichen Liganden zu dem jeweiligen Protein sind. Die erzeugten Konturen des Selektivitätsmodells lassen sich gedanklich durch den direkten Vergleich der Konturen der beiden Einzelmodelle nachvollziehen. Während im Falle des Thrombin-Modells die sterische Besetzung der S2-Tasche und des oberen Bereichs der S3-Tasche favorisiert wird, liegt dieser Bereich bei Faktor Xa eher im unteren Teil der S3/4-Tasche. Wenn nun eine „räumliche Differenz“ der beiden weißen Konturen gebildet wird, so bleibt in der Tat nur der weiß konturierte Bereich übrig, der bei dem Selektivitätsmodell markiert ist.

Die selektivitätsbestimmenden Bereiche stimmen auch gut mit den räumlichen Gegebenheiten in den Proteinbindetaschen von Thrombin und Faktor Xa überein. Der weiß konturierte Bereich, dessen sterische Besetzung die Selektivität zu Thrombin erhöhen soll, befindet sich unterhalb der 60er-Schleife von Thrombin im Bereich der S2-Tasche. Diese Bindetasche wird in der Literatur dafür beschrieben, dass ihre sterische Ausfüllung eine Verbesserung im Hinblick auf affinere und selektivere Thrombin-Bindung erzielt (Wagner *et al.*, 1998). Die violette Kontur hingegen, verantwortlich für die Erhöhung der Selektivität zu Faktor Xa, liegt genau in dem für Faktor Xa-Inhibitoren bekannten Bereich der S3/4-Tasche (Stubbs *et al.*, 1995).

Die graphische Auswertung des Selektivitätsmodells spiegelt auch die Eigenschaften der zugrundeliegenden Inhibitoren wider. In Abb. 2-8 ist das gleiche Konturdiagramm aus Abb. 2-7 zusammen mit einigen Verbindungen, die stark selektiv an Thrombin bzw. Faktor Xa binden, dargestellt.



Abb. 2-8: Konturdiagramm der sterischen Eigenschaften des Selektivitätsmodells zu Thrombin und Faktor Xa entsprechend Abb. 2-7(3). Zusätzlich sind einige selektiv bindende Inhibitoren zu Thrombin (orange) und Faktor Xa (cyan) dargestellt.

Aus Abb. 2-8 wird ersichtlich, dass die selektiv an Thrombin bindenden Inhibitoren den weiß konturierten Bereich gut ausfüllen. Insbesondere sind dies Verbindungen, die in die S2-Tasche von Thrombin Piperidin- und Piperazinreste positionieren, die zusätzlich in *para*-Stellung noch Methylsulfonyl- oder Methylcarboxylat-Substituenten aufweisen, wie bei 55_3TAPAP_BNAS_PZD-NSME (Selektivitätsfaktor +3,824) oder 50_3TAPAP_BNAS_INIP-OME (Faktor +3,403). Auf der anderen Seite finden sich selektiv an Faktor Xa bindende Inhibitoren, die die violett konturierte Region besetzen; dies wird besonders gut von den bisamidinischen Inhibitoren wie beispielsweise 14_NAPHTH_DAIICHI (Selektivitätsfaktor – 3,465) oder 22_BISBENZ_OET4AM (Faktor –2,553) erfüllt.

Ein weiteres Beispiel zur Interpretation der Selektivitätsmodelle zeigt die graphische Auswertung der selektivitätsbestimmenden Faktoren zwischen Thrombin und Urokinase. In Abb. 2-9 werden zum einen die sterischen Eigenschaften der Einzelmodelle von Thrombin und Urokinase dargestellt; im direkten Vergleich dazu ist das Konturdiagramm der sterischen Eigenschaften des Selektivitätsmodells gezeigt, das aus den Affinitätsdifferenzen zwischen Thrombin und Urokinase generiert wurde.

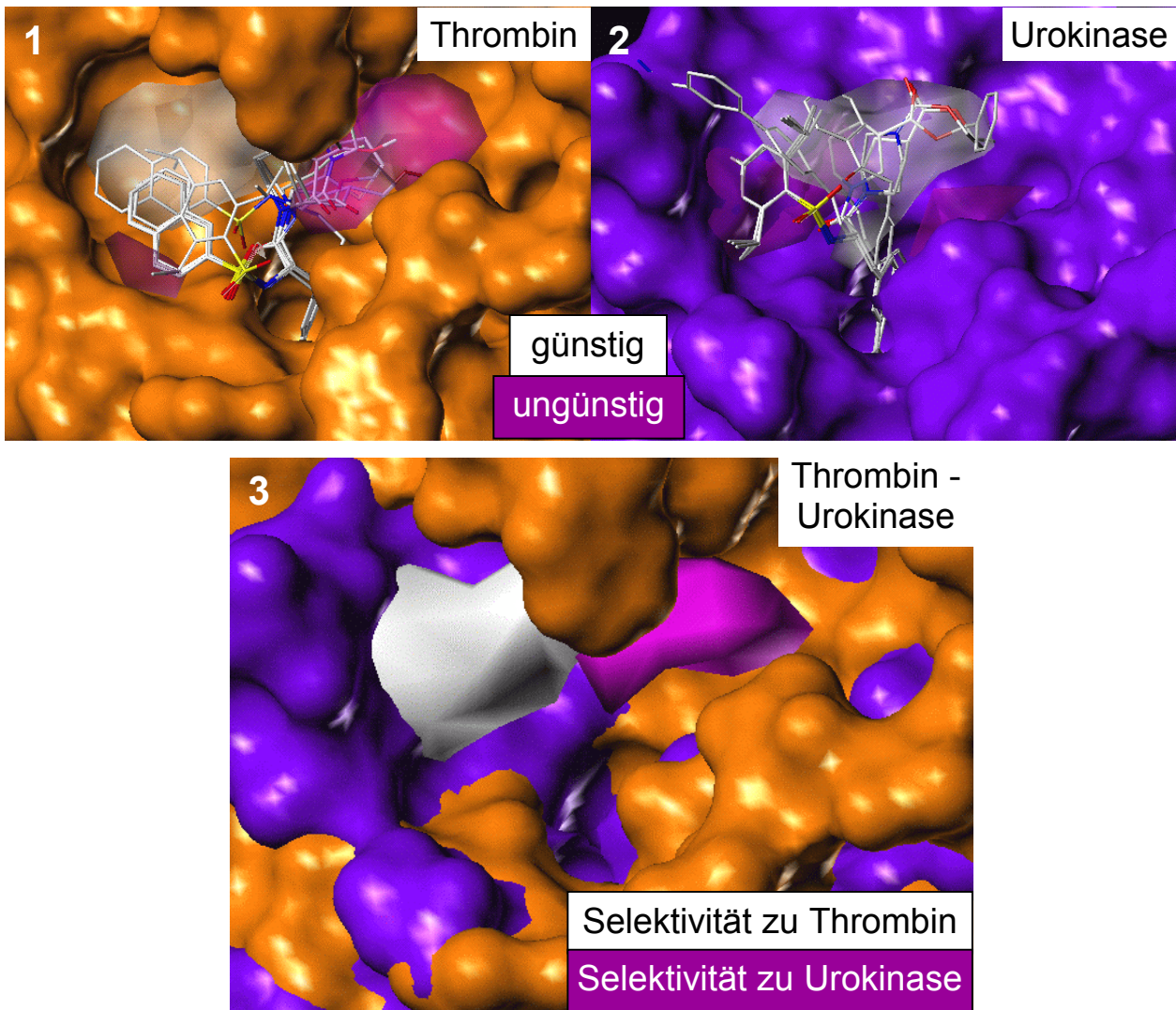


Abb. 2-9: Konturdiagramme der sterischen Eigenschaften der Einzelmodelle von Thrombin (1) und Urokinase (2) und des Selektivitätsmodells (3) aus den Affinitätsdifferenzen zwischen Thrombin und Urokinase. Die im Modell verwendeten Inhibitoren wurden in der Bindetasche von Thrombin minimiert. Das Konturdiagramm 1 entspricht dem aus Abb. 2-5(1). 1+2: Die weiß konturierten Bereiche zeigen Regionen, in denen die Anwesenheit sterisch anspruchsvoller Gruppen positiv mit der Bindungsaffinität korreliert; violette Bereiche sollten für eine Verbesserung der Affinität hingegen sterisch unbesetzt bleiben. 3: Die räumliche Ausfüllung der weißen Kontur erhöht die Selektivität eines Liganden zu Thrombin, während die Anwesenheit sterisch anspruchsvoller Gruppen in dem violetten Bereich die Selektivität zu Urokinase fördert. Die lösungsmittelzugänglichen Oberflächen beider Proteine (Thrombin: orange; Urokinase: blau) sind übereinandergelegt.

Die sterischen Eigenschaften des Thrombin-Modells wurden schon weiter oben diskutiert. Das Konturdiagramm des Urokinase-Modells zeigt eine ausgedehnte Kontur im Bereich der S2- und S1'-Tasche, die als günstig für die Besetzung mit sterisch anspruchsvollen Gruppen vorhergesagt wird. Dies wird durch die Platzierung einiger Liganden in Urokinase bestätigt, die röntgenstrukturanalytisch bestimmt werden konnten (PDB-Codes 1EJN, 1FV9, 1LMW). Dahingegen wird der Bereich, der bei Thrombin und Faktor Xa der S3- bzw. S3/4-Tasche entspricht, bei dem Urokinase-Modell als ungünstig zur Besetzung mit sterisch anspruchsvollen Gruppen vorhergesagt. Obwohl die in Thrombin minimierten Inhibitoren als Grundlage des Urokinase-Modells dienten, können die Ergebnisse der graphischen Auswertung der Konturen mit den tatsächlichen Gegebenheiten in der Bindetasche von Urokinase in Einklang gebracht werden. Das Konturdiagramm der sterischen Eigenschaften des Selektivitätsmodells von Thrombin und Urokinase markiert wiederum die Bereiche in der Bindetasche, die auf die Selektivität eines Liganden zu beiden Proteinen Einfluss nehmen. So wird, ähnlich zu dem Selektivitätsmodell von Thrombin und Faktor Xa, eine Region (weiße Kontur) unterhalb der 60er-Schleife und im oberen Bereich der S3-Tasche markiert, deren Besetzung die Affinität eines Liganden zugunsten von Thrombin, relativ zu Urokinase gesehen, verbessern sollte. Im Unterschied zum vorherigen Modell wird zusätzlich eine Kontur im Bereich der S1'-Bindetasche angezeigt, deren sterische Besetzung die Selektivität hinsichtlich Urokinase erhöhen kann. Beide Konturen spiegeln die räumliche Situation der Proteinbindetaschen beider Proteine wider; dies wird vor allem deutlich, wenn die überlagerten lösungsmittelzugänglichen Oberflächen betrachtet werden.

Neben den aus den Selektivitätsmodellen vorhergesagten, diskriminierenden Eigenschaften sterischer Natur lassen sich prinzipiell auch die vier anderen Eigenschaften, die mit CoMSIA berechnet wurden, untersuchen. Tatsächlich müssten zu einer vollständigen Interpretation der graphischen Resultate alle fünf Eigenschaften gleichzeitig berücksichtigt werden, da die entsprechenden Feldbeiträge auch in das Modell eingeflossen sind. Die Auswertung der hydrophoben Beiträge ist in Abb. 2-10 gezeigt. Neben den Konturdiagrammen der Einzelmodelle, die bei Verwendung der Affinitätsdaten von Thrombin und Faktor Xa erzeugt wurden, zeigt die graphische Auftragung der hydrophoben Eigenschaften des Selektivitätsmodells wiederum die Bereiche, die diskriminierend bezüglich der Selektivität zu einem der beiden Proteine sind.

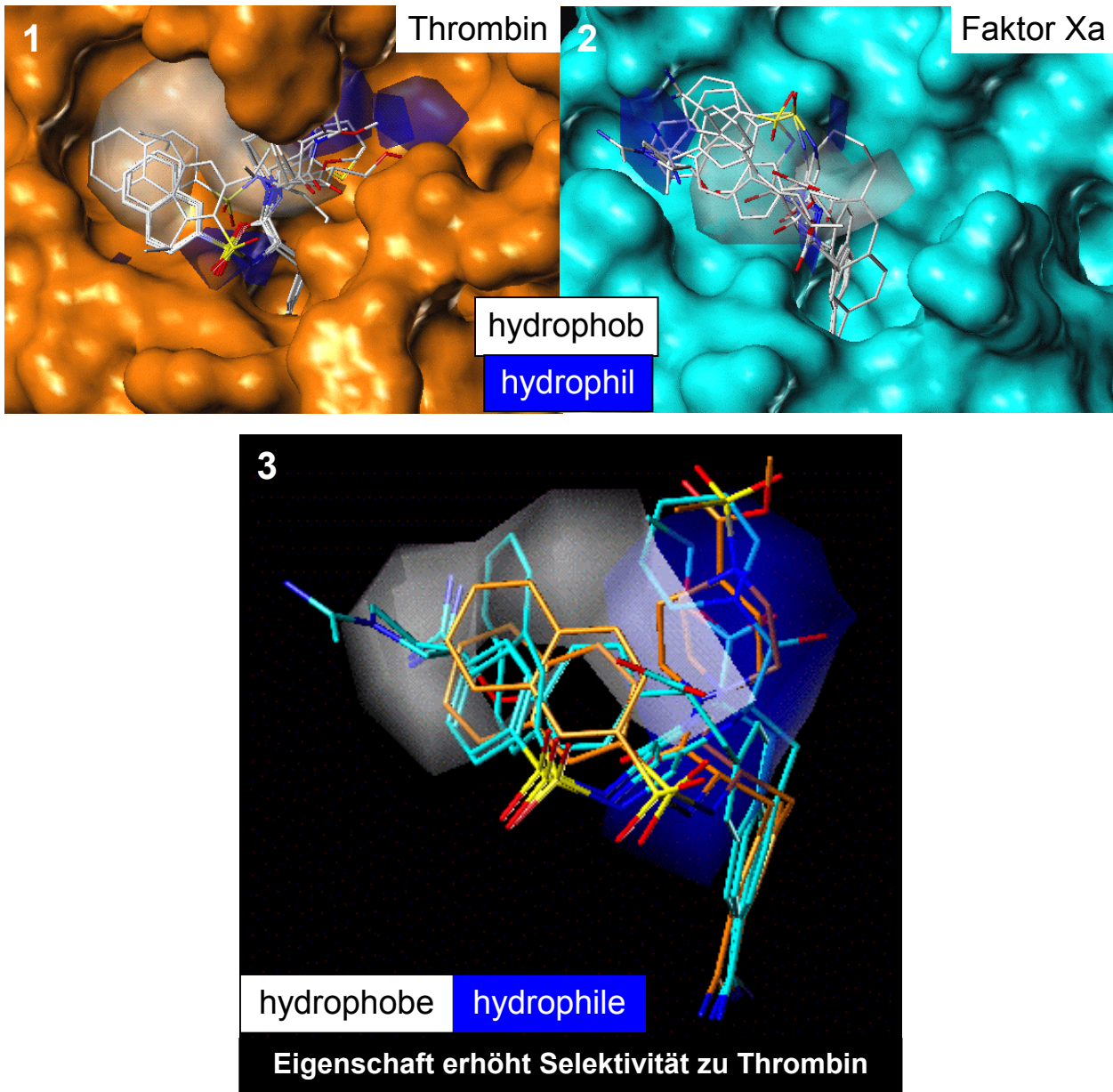


Abb. 2-10: Konturdiagramme der hydrophoben Eigenschaften der Einzelmodelle von Thrombin (1) und Faktor Xa (2) und des Selektivitätsmodells (3) zwischen Thrombin und Faktor Xa. Die im Modell verwendeten Inhibitoren wurden in der Bindetasche von Thrombin minimiert. Das Konturdiagramm 2 entspricht dem aus Abb. 2-6(2). 1+2: Weiße Konturen markieren Bereiche, deren Besetzung mit hydrophoben Gruppen mit einer Erhöhung der Bindungsaffinität einhergeht; entsprechend werden hydrophile Gruppen in blauen Bereichen zur Aktivitätserhöhung favorisiert. 3: Die Besetzung der weißen Kontur mit hydrophoben bzw. der blauen Kontur mit hydrophilen Gruppen steigert die Selektivität zu Thrombin; hingegen wird die Selektivität zu Faktor Xa erhöht, wenn die weiße Kontur mit hydrophilen bzw. die blaue Kontur mit hydrophoben Gruppen belegt werden können. Zusätzlich sind einige selektiv bindende Inhibitoren zu Thrombin (orange) und Faktor Xa (cyan) dargestellt.

Die Auswertung der Konturdiagramme erfolgt analog zu den vorherigen Beispielen. Die hydrophoben Eigenschaften der Einzelmodelle geben die Bereiche in der Bindetasche wieder, die verantwortlich für eine Erhöhung oder Verminderung der Affinitäten sind, wenn die markierten Konturen mit hydrophoben bzw. hydrophilen Substituenten besetzt werden. Bei Thrombin wird die Besetzung der S2- und S3-Taschen mit hydrophoben Gruppen und der S1'-Tasche mit eher hydrophilen Gruppen favorisiert. Demgegenüber wird im Falle von Faktor Xa nur der rechte Teil der S3/4-Tasche als günstig für die Ausfüllung mit hydrophoben Substituenten vorhergesagt; der andere Teil im Bereich der Aminosäuren Glu 97, Thr 98 und Ile 175 am äußeren Rand der S3/4-Tasche sollte mit hydrophilen Gruppen besetzt werden. Diese Konturierung geht auf die bisamidinischen, hochaffinen Faktor Xa-Inhibitoren des Datensatzes zurück, die diese Region mit Amidin-Substituenten besetzen.

Die Interpretation des Konturdiagramms des Selektivitätsmodells in Abb. 2-10 zeigt, dass die Selektivität zu Thrombin erhöht werden kann, wenn einerseits hydrophobe Gruppen den weiß markierten Bereich bzw. hydrophile Gruppen das blau hervorgehobene Areal besetzen. Umgekehrt lassen sich selektiv an Faktor Xa bindende Inhibitoren entwickeln, wenn deren hydrophobe Substituenten in der blauen Markierung bzw. hydrophile Gruppen in die weiß umrandete Region zu liegen kommen.

Die in Abb. 2-10 gezeigten, selektiv an Thrombin (orange) oder Faktor Xa (cyan) bindenden Inhibitoren spiegeln diese diskriminierende Faktoren auch wider: einige hochselektive Thrombin-Liganden weisen zum einen rein hydrophobe Gruppen (zum Beispiel Naphthyl-Substituenten) auf, die in der S3-Tasche geeignet binden; andererseits besitzen diese Verbindungen P2-Substituenten, die eher von hydrophiler Natur sind (Piperidin- und Piperazingruppen mit Sulfonyl- und Carboxyl-Substituenten). Die bisamidinischen Inhibitoren mit ihren terminalen Amidin-Substituenten erklären die Selektivitätssteigerung durch hydrophile Gruppen in der S3/4-Tasche, die Wechselwirkungen mit den Carbonylgruppen von Glu 97, Thr 98 und Ile 175 eingehen können (Stubbs *et al.*, 1995).

Ähnliche Ergebnisse wurden bei der Untersuchung der selektivitätsbestimmenden Faktoren zwischen Thrombin und Urokinase bezüglich ihrer hydrophoben Eigenschaften gefunden (Abb. 2-11).

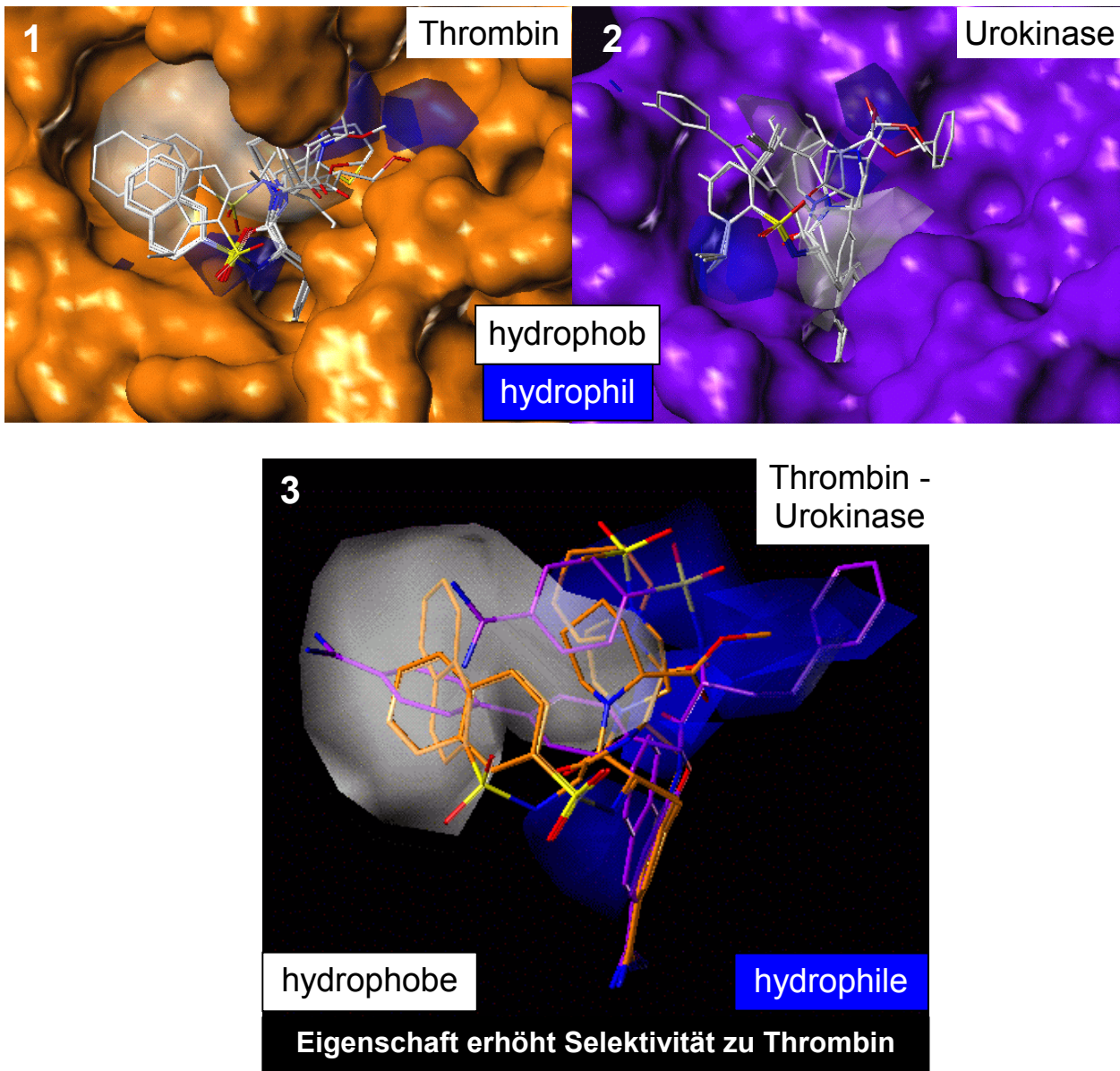


Abb. 2-11: Konturdiagramme der hydrophoben Eigenschaften der Einzelmodelle von Thrombin (1) und Urokinase (2) und des Selektivitätsmodells (3) zwischen Thrombin und Urokinase. Die im Modell verwendeten Inhibitoren wurden in der Bindetasche von Thrombin minimiert. 1+2: Weiße Konturen markieren Bereiche, deren Besetzung mit hydrophoben Gruppen mit einer Erhöhung der Bindungsaffinität einhergeht; entsprechend werden hydrophile Gruppen in blauen Bereichen zur Aktivitätserhöhung favorisiert. 3: Die Besetzung der weißen Kontur mit hydrophoben bzw. der blauen Kontur mit hydrophilen Gruppen steigert die Selektivität zu Thrombin; hingegen wird die Selektivität zu Urokinase erhöht, wenn die weiße Kontur mit hydrophilen bzw. die blaue Kontur mit hydrophoben Gruppen belegt werden können. Zusätzlich sind einige selektiv bindende Inhibitoren zu Thrombin (orange) und Urokinase (violett) dargestellt.

Im Falle des Urokinase-Einzelmodells sollten sich zur Steigerung der Bindungsaffinität hydrophobe Gruppen im zentralen Bereich oberhalb der S1-Tasche und in Richtung S1'-Tasche befinden, während Teile der S2- und S3-Tasche durch hydrophilere Substituenten besetzt werden sollten. Das Konturdiagramm des Selektivitätsmodells zeigt hingegen, dass stärker hydrophobe Gruppen in der S3-Tasche bzw. hydrophile Substituenten in der S2-Tasche die Selektivität zu Thrombin steigern. Soll die Affinität zu Urokinase verbessert werden und die zu Thrombin gleichzeitig vermindert werden, so müssten Inhibitoren entwickelt werden, die im zentralen Bereich und Teilen der S1'-Tasche hydrophobe Eigenschaften aufweisen. Weiterhin könnte die Selektivität zu Urokinase durch Positionierung hydrophiler Gruppen in die S3-Tasche verbessert werden. Wenn jedoch gleichzeitig – wie gefordert – die sterischen Eigenschaften des Selektivitätsmodells berücksichtigt werden, die eine sterische Besetzung dieser Region als ungünstig vorhersagt, erscheint es schwierig, beide Forderungen gleichzeitig zu erfüllen. Auch wenn sich hydrophobe Gruppen in der S3-Bindetasche als günstig für eine verbesserte Urokinase-Selektivität erweisen, so kann dem aus sterischen Gründen nur schwer nachgekommen werden. Es bleibt also immer abzuwägen, welche der insgesamt fünf³⁴ physikochemischen Eigenschaften auch als sinnvoll für die Entwicklung selektiver Inhibitoren erachtet werden können.

2.3.5 Selektivitätsmodelle aus Kombination von Einzelmodellen

Neben der in Kap. 2.3.4 beschriebenen und auch in der Literatur verwendeten Methode, durch Differenzbildung der Affinitätswerte Aussagen über selektivitätsbestimmende Faktoren ableiten zu können, hat dieser Ansatz auch gewisse Nachteile. So muss zum Beispiel die Annahme identischer Bindungsmodi der Liganden getroffen werden; dies bedeutet, dass die Konformationen und relativen Orientierungen in den beiden betrachteten Proteinbindetaschen jeweils gleich sind. Im Falle der Bindetaschen von Thrombin und Trypsin oder Faktor Xa und Trypsin erscheint diese Annahme weitgehend erfüllt zu sein. Die Bindungsmodi der Inhibitoren unterscheiden sich jedoch deutlich bei ihrer Bindung an Thrombin und Faktor Xa oder bei der Bindung an Thrombin und Urokinase. Daher wäre eine Berücksichtigung der jeweils optimierten Bindungsmodi wünschenswert, die durch Minimierung in den zugehörigen Bindetaschen erzielt werden konnten. Ein weiterer Nachteil ist die für diese

³⁴ im Falle der fünf verwendeten Felder in CoMSIA

Methode notwendige Berechnung der Affinitätsdifferenzen. Eine Differenzbildung von experimentell bestimmten Affinitätswerten hat aufgrund der Fehlerfortpflanzung meist eine Vergrößerung der Messfehler zur Folge. Wenn einem experimentellen K_i -Wert eine Fehlertoleranz von durchschnittlich 15% zugrunde liegt, so kann sich dieser Fehler bei einer Kombination zweier pK_i -Werte im ungünstigsten Falle aufaddieren³⁵.

Daher wurde eine vorteilhaftere Methode zur Gewinnung selektivitätsbestimmender Faktoren entwickelt, bei der zwei Einzelmodelle direkt miteinander kombiniert werden. Dies gelang dadurch, dass die Differenz der aus den Modellen stammenden QSAR-Koeffizienten zu jeweils zwei separat erzeugten Einzelmodellen gebildet wurde. Die Koeffizienten in der Form $stdev*coeff$ wurden vorher normalisiert, so dass beide Modelle gleich gewichtet waren. Die Differenzwerte wurden an jedem Gitterpunkt einer zu betrachteten Eigenschaft (sterisch, elektrostatisch, hydrophob, etc.) berechnet und anschließend in einem Konturdiagramm graphisch dargestellt. Diese mathematische Vorgehensweise entspricht der gedanklichen Bildung einer „räumlichen Differenz“ zweier Konturdiagramme (vgl. Seite 88). Somit konnte bei diesem Verfahren auch auf eine Verwendung von Differenzwerten der Bindungsaffinitäten verzichtet werden (vgl. Kap. 2.3.4).

Ein weiterer Vorteil dieser Methode ist, dass jedes Modell getrennt voneinander entwickelt werden konnte, also unter Berücksichtigung der für jede Bindetasche optimierten Bindungsmodi. Im Falle des Selektivitätsmodells zwischen Thrombin und Faktor Xa wurde also das Thrombin-Modell herangezogen, das unter Verwendung der in der Thrombin-Bindetasche minimierten Inhibitoren erstellt wurde, und dem Faktor Xa-Modell gegenübergestellt, das mit den in der Faktor Xa-Tasche optimierten Bindungsmodi der Liganden erzeugt wurde. Zu beiden Modellen wurden anschließend die Differenzen der QSAR-Koeffizienten an jedem Gitterpunkt bezüglich einer physikochemischen Eigenschaft gebildet und graphisch konturiert. In Abb. 2-12 sind zum Beispiel die sterischen Eigenschaften der Selektivitätsmodelle zwischen Thrombin und Faktor Xa gezeigt, wie sie mit der in Kap. 2.3.4 beschriebenen Methode (Differenzen der Bindungsaffinitäten) und mit dem in diesem Abschnitt verwendeten Alternativansatz (Differenzen der QSAR-Koeffizienten) gefunden wurden.

³⁵ Rechenbeispiel: angenommener Fehler $\cong 0,5$ log. Einheiten

$pK_{i,1} = 8,0 \pm 0,5$; $pK_{i,2} = 5,0 \pm 0,5 \Rightarrow pK_{i,1} - pK_{i,2} = 3,0 \pm 1,0$

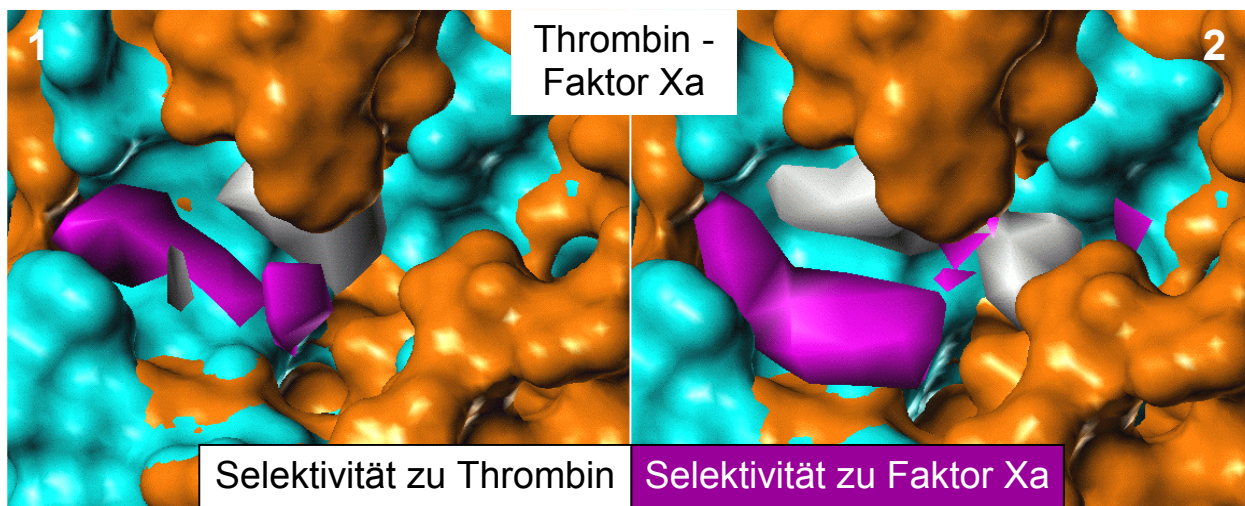


Abb. 2-12: Konturdiagramme der sterischen Eigenschaften, erzeugt mit zwei methodisch unterschiedlichen Selektivitätsanalysen. **1:** Generierung des QSAR-Modells aus berechneten Affinitätsdifferenzen; vgl. Abb. 2-7(3). **2:** Entwicklung zweier separater Einzelmodelle unter Verwendung individueller Überlagerungsmodelle; anschließend Differenzbildung der QSAR-Koeffizienten der sterischen Eigenschaften beider Modelle. **1+2:** Die sterische Besetzung der weiß konturierten Bereiche erhöht die Selektivität eines Liganden gegenüber Thrombin, während die Anwesenheit sterischer Gruppen in den violetten Bereichen die Selektivität zu Faktor Xa fördert.

Aus Abb. 2-12 sind qualitativ ähnliche Konturdiagramme zu ersehen. Der markierte Bereich in der S3/4-Tasche, der die Selektivität zu Faktor Xa erhöht, hebt sich bei der in diesem Abschnitt verwendeten Methode etwas stärker hervor. Zusätzlich zu der Kontur in der Nähe der S2-Tasche, die als selektivitätsbestimmend für Thrombin vorhergesagt wird, erscheint eine weitere Kontur im Bereich der S1'-Tasche, deren sterische Besetzung die Ligandaffinität ebenfalls zugunsten von Thrombin zu steigern vermag.

Diese hier vorgestellte, neuartige Methode der Selektivitätsanalyse bedarf jedoch noch weiterer Auswertungen, um gesicherte Aussagen bezüglich der Faktoren und Eigenschaften zu treffen, die entscheidend für eine Diskriminierung eines Liganden zu einem bestimmten Protein sein könnten. So ist noch unklar, inwieweit eine Differenzbildung von Koeffizienten zweier unterschiedlicher QSAR-Modelle chemometrisch aufeinander abgestimmte Modelle ergibt; weitere Einflüsse wie die Normalisierung der Koeffizienten oder die geeignete Wahl, in welcher Form diese Koeffizienten verwendet werden sollten (zum Beispiel *coeff* oder *stdev*coeff*), müssen ebenfalls genauer ausgewertet werden.

2.4 Untersuchung von physikochemischen Eigenschaften in Proteinbindetaschen durch chemometrische Auswertungen

Eine Methode zur Untersuchung selektivitätsbestimmender Faktoren von Protein-Ligand-Wechselwirkungen sollte – unabhängig vom Informationsgehalt dazu bindenden Liganden – die Auswertung exponierter physikochemischer Eigenschaften in Proteinbindetaschen sein. Der dazu notwendige chemometrische Ansatz einer Hauptkomponentenanalyse (PCA) solcher Eigenschaften und dessen Anwendung auf eine Reihe von Serinproteasen wird im folgenden diskutiert. Die in dieser Arbeit durchgeführten chemometrischen Untersuchungen lehnen sich an die von Kastenholz *et al.* beschriebenen Methodik (vgl. Seite 55) der Consensus-Hauptkomponentenanalyse (*consensus principle component analysis*, CPCA) an (Kastenholz *et al.*, 2000). Mit diesem Verfahren ist es möglich, eine größere Zahl von Proteinen gleichzeitig zu analysieren. Außerdem können Aussagen abgeleitet werden, welche Eigenschaften in welchen Bereichen einer Proteinbindetasche einen Einfluss auf die Selektivität ausüben können. In der vorliegenden Arbeit wurde mit einer erweiterten Datenbasis von 16 Proteinstrukturen gearbeitet; neben dem meist verwendeten GRID-Programm wurden auch andere Ansätze zur physikochemischen Charakterisierung der Proteinbindetaschen eingesetzt.

2.4.1 Aufbau des Protein-Datensatzes

Der erste Schritt zur Untersuchung von selektivitätsbestimmenden Faktoren von Seiten des Rezeptors ist der Aufbau einer entsprechender Datenbasis. Hierzu wurden eine Reihe von experimentell verfügbaren Kristallstrukturen der Serinproteasen Thrombin, Trypsin, Faktor Xa und Urokinase aus der PDB ausgewählt. Dabei wurden nur Einträge mit einer hohen kristallographischen Auflösung berücksichtigt. Weiterhin wurde darauf geachtet, dass diese Proteine auch einen in der Kristallstruktur gebundenen Liganden aufweisen. Zusätzlich wurden vier Trypsin-Mutanten aufgenommen, die zum Teil noch nicht in der PDB veröffentlicht sind (Reyda *et al.*, in preparation; Stubbs *et al.*, in press) und für die hier durchgeführten Studien zur Verfügung standen. Somit wurde ein Datensatz von insgesamt 16 experimentell verfügbaren Serinproteasen mit jeweils drei Thrombin-, Trypsin-, Faktor Xa- und Urokinase-Strukturen und weiteren vier Trypsin-Mutanten erstellt (Tab. 2-8).

Tab. 2-8: Aufstellung der 16 untersuchten Proteinstrukturen mit den zugehörigen PDB-Codes und ihren kristallographischen Auflösungen. Die nicht in der PDB veröffentlichten Trypsin-Mutanten sind in Klammern gesetzt.

Nr.	PDB-Code	Protein	Auflösung [Å]
1	1A2C	Thrombin	2,1
2	1D4P	Thrombin	2,1
3	1ETR	Thrombin	2,2
4	1AZ8	Trypsin	1,8
5	1F0T	Trypsin	1,8
6	1MTS	Trypsin	1,9
7	1F0R	Faktor Xa	2,1
8	1FJS	Faktor Xa	1,9
9	1HCG	Faktor Xa	2,2
10	1EJN	Urokinase	1,8
11	1FV9	Urokinase	3,0
12	1LMW	Urokinase	2,5
13	1QL9	Trypsin-Mutante	2,3
14	(<i>y99_03</i>)	Trypsin-Mutante	–
15	(<i>triple_03</i>)	Trypsin-Mutante	–
16	(<i>zen_tri_03</i>)	Trypsin-Mutante	–

Die in Tab. 2-8 enthaltenen Trypsin-Mutanten sind so gestaltet, dass die auftretenden Punktmutationen die Eigenschaft von Faktor Xa nachahmen sollen. So weisen die Strukturen von 1QL9 und *y99_03* die Punktmutationen N97E (Asn 97 → Glu 97) und L99Y (Leu 99 → Tyr 99) auf. Die Proteine *triple_03* und *zen_tri_03* enthalten eine Dreifachmutation mit den Punktmutationen N97E und L99Y (wie vorher) und zusätzlich S190A (Ser 190 → Ala 190) und Y172S (Tyr 172 → Ser 172). Alle Punktmutationen wurden so durchgeführt, dass die vormals für Trypsin kodierenden, ausgetauschten Aminosäuren mit denen von Faktor Xa identisch sind.

Die 16 Proteinstrukturen aus Tab. 2-8 wurden für weitere Analysen zunächst aufbereitet. So wurden sämtliche Liganden, Wassermoleküle und Ionen (Na^+ , Ca^{2+} , SO_4^{2-} , etc.) entfernt. Anschließend konnten die so erzeugten, von Liganden befreiten Proteine für einen räumlichen

Vergleich der Eigenschaften miteinander überlagert werden. Dazu wurden die bei allen Serinproteasen vorhandenen, konservierten Aminosäuren der katalytischen Triade (His 57, Asp 102 und Ser 195) und zusätzlich Asp 189 der S1-Spezifitätstasche herangezogen und deren C_α-Atome nach der RMS-Methode (*root mean square*) rigide superpositioniert.

2.4.2 Physikochemische Charakterisierung der Proteinbindetaschen

2.4.2.1 Berechnung der Wechselwirkungsfelder mit GRID

Die am häufigsten verwendete Methode zur Charakterisierung physikochemischer Eigenschaften von Proteinbindetaschen stellt die Analyse mit dem Programm GRID (Goodford, 1985; Boobbyer *et al.*, 1989; Wade *et al.*, 1993; Wade & Goodford, 1993) dar. Eine Vielzahl von Sondenatomen, die in GRID integriert sind, können die physikochemischen Eigenschaften einer Bindetasche ausleuchten und angeben, in welchen Bereichen die Positionierung eines bestimmten, dem Sondenatom entsprechenden Atomtyps günstig ist.

Aus den vorhandenen GRID-Sondenatomen wurden in der vorliegenden Arbeit vier diverse Sondenatome ausgewählt, die jeweils verschiedene physikochemische Eigenschaften (Donor, Akzeptor, sterisch, hydrophob) repräsentieren sollen (Tab. 2-9). In Abb. 2-13 sind die Konturen von drei Wechselwirkungsfeldern in den Bindetaschen von Thrombin und Faktor Xa gezeigt.

Tab. 2-9: Ausgewählte Sondenatome aus GRID zur Beschreibung der physikochemischen Eigenschaften der Proteinbindetaschen. Neben den Namen der Sondenatome, wie sie in GRID bezeichnet werden, und den entsprechenden chemischen Gruppen sind die jeweils zugewiesenen physikochemischen Eigenschaften angegeben.

Nr.	GRID Name	chemische Gruppe	Eigenschaft
1	N1	<i>neutral flat NH eg. amide</i>	Donor
2	O	<i>sp² carbonyl oxygen</i>	Akzeptor
3	C3	<i>methyl CH₃ group</i>	sterisch
4	DRY	<i>hydrophobic probe</i>	hydrophob

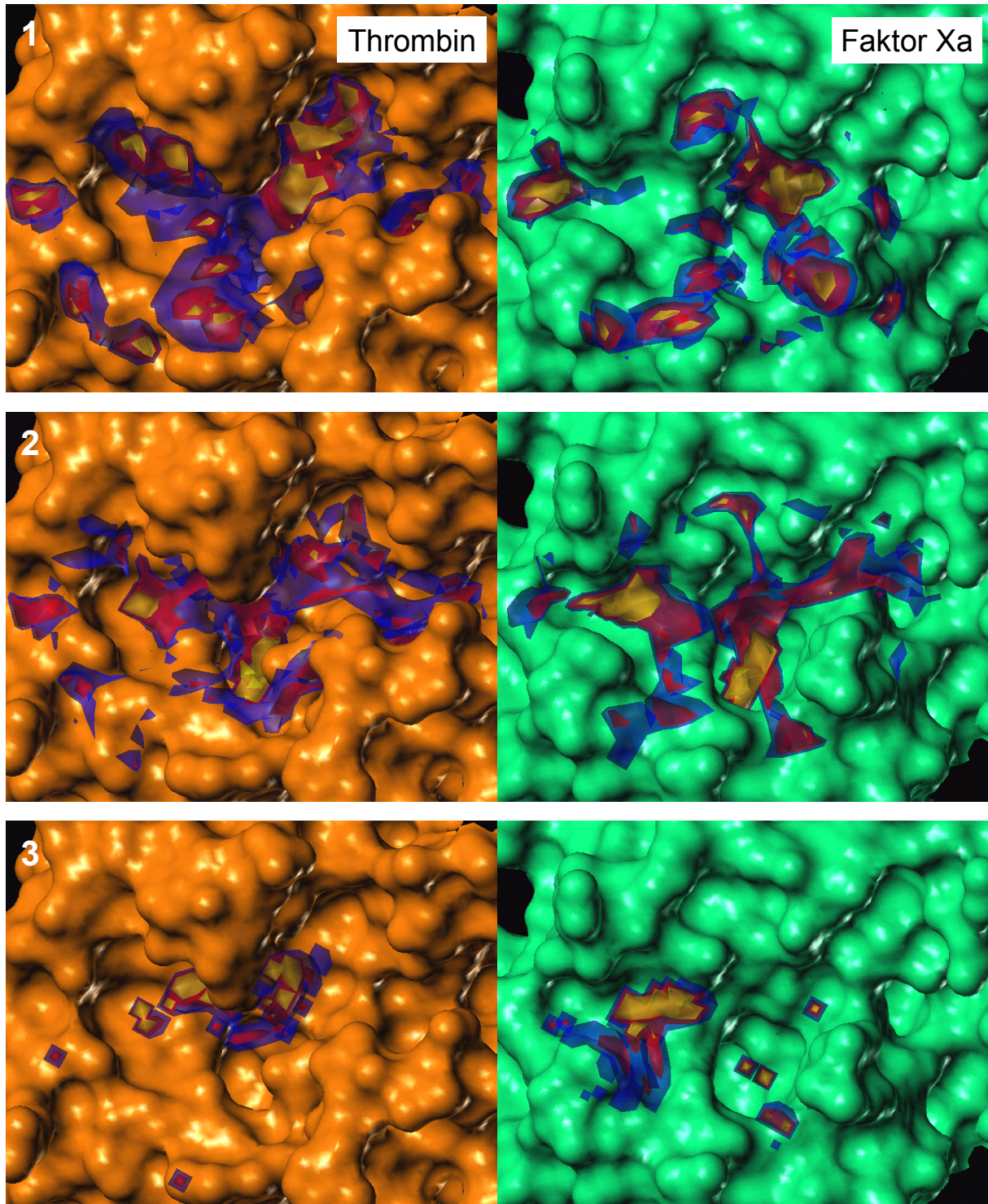


Abb. 2-13: Berechnete Wechselwirkungsfelder bezüglich der Bindetaschen von Thrombin und Faktor Xa, die mit GRID-Sondenatomen unterschiedlicher Eigenschaften berechnet wurden. Die Potenzialwerte der Dichtekontur sind in drei Abstufungen (von blau über rot nach gelb zunehmend günstiger) abgebildet **1**: Sonde „O“ (Akzeptor-Eigenschaften); **2**: Sonde „C3“ (sterische Eigenschaften); **3**: Sonde „DRY“ (hydrophobe Eigenschaften).

2.4.2.2 Berechnung der Wechselwirkungsfelder mit SuperStar

Eine ähnliche Beschreibung der Proteinbindetasche erfolgte mit dem Programm SuperStar (Verdonk *et al.*, 1999; Verdonk *et al.*, 2001; Boer *et al.*, 2001), dessen Anwendung auf kleine Moleküle in Kap. 1 bereits ausführlich dargestellt wurde. Während bei GRID ein parametrisiertes Kraftfeld zur Bestimmung der Wechselwirkungsfelder dient, beruhen die in SuperStar zugrundeliegenden Informationen ausschließlich auf experimentellen Daten der CSD. Daraus können Häufigkeitsverteilungen funktioneller Gruppen um Zentralfragmente in Form von sogenannten *scatterplots* (wie zum Beispiel Abb. 1-2, Seite 3) extrahiert und in Wahrscheinlichkeitsdichten konvertiert werden. Die somit erzeugten Konturfelder geben Bereiche in der Proteinbindetasche an, in denen die Wechselwirkung zu einer bestimmten funktionellen Gruppe günstig sein sollte.

Der Einsatz mehrerer Kontaktgruppen, die jeweils unterschiedliche physikochemische Eigenschaften repräsentieren sollen, führt zu einer Beschreibung der Bindetasche hinsichtlich sterischer, hydrophober, Donor- und Akzeptor-Eigenschaften. In dieser Arbeit wurden zur Charakterisierung der Proteinbindetaschen – wie zuvor bei GRID – insgesamt vier Kontaktgruppen aus SuperStar verwendet (Tab. 2-10). Als Beispiel sind in Abb. 2-14 die Konturen von drei unterschiedlichen Wechselwirkungsfeldern in den Bindetaschen von Thrombin und Faktor Xa gezeigt.

Tab. 2-10: Ausgewählte Kontaktgruppen aus SuperStar zur Beschreibung der physikochemischen Eigenschaften der Proteinbindetaschen. Neben dem Namen der Kontaktgruppen, wie sie in SuperStar bezeichnet werden, sind ihre zugewiesenen physikochemischen Eigenschaften angegeben.

Nr.	SuperStar Name	Eigenschaft
1	<i>amino hydrogen</i>	Donor
2	<i>carbonyl oxygen</i>	Akzeptor
3	<i>aliphatic carbon</i>	sterisch / hydrophob
4	<i>aromatic carbon</i>	aromatisch / hydrophob

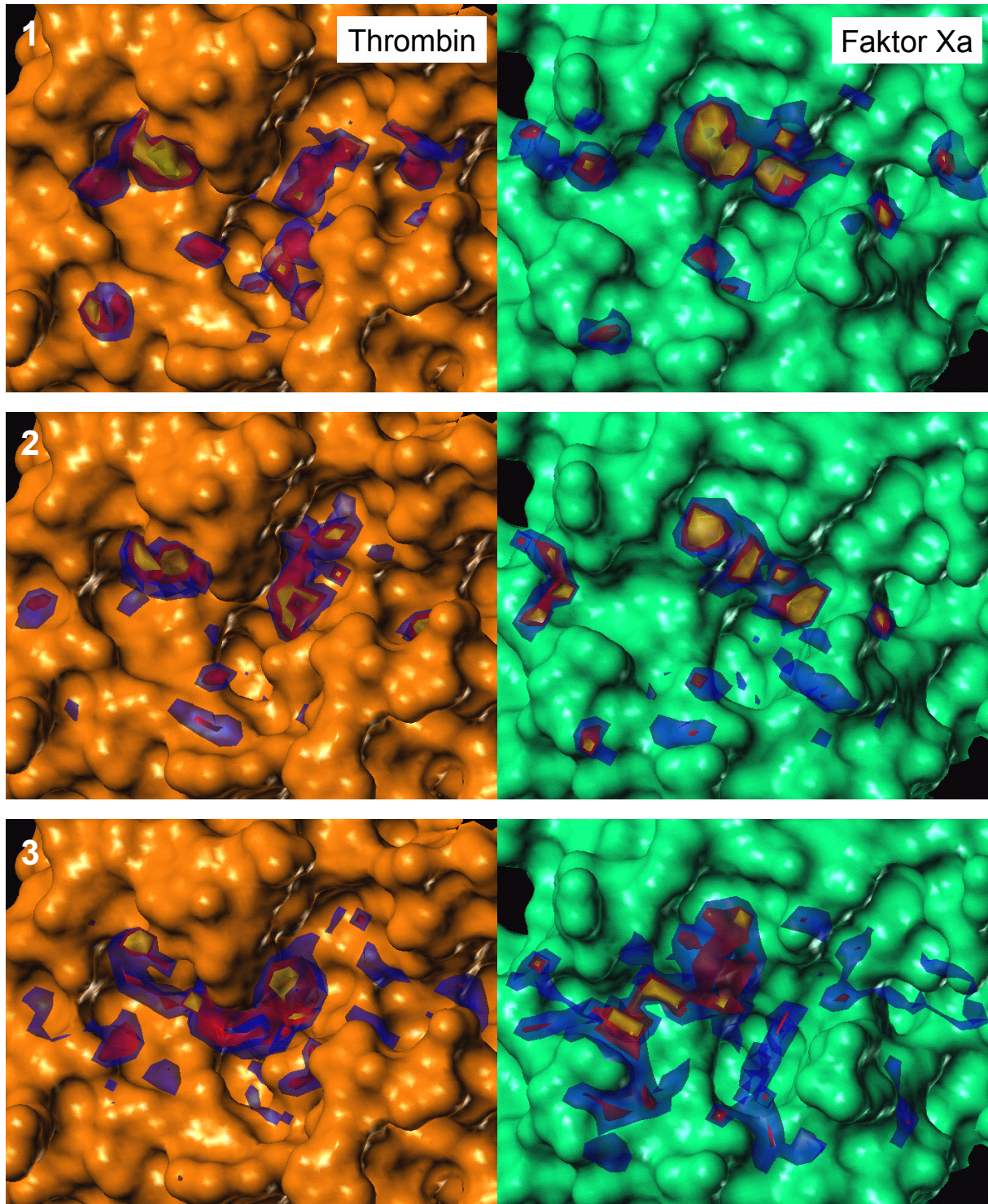


Abb. 2-14: Berechnete Wechselwirkungsfelder bezüglich der Bindetaschen von Thrombin und Faktor Xa, die mit SuperStar-Kontaktgruppen unterschiedlicher Eigenschaften berechnet wurden. Die Wahrscheinlichkeitswerte der Dichtekontur sind in drei Abstufungen (Werte von blau über rot nach gelb zunehmend günstiger) abgebildet. **1:** Kontaktgruppe *amino hydrogen* (Donor-Eigenschaften); **2:** Kontaktgruppe *carbonyl oxygen* (Akzeptor-Eigenschaften); **3:** Kontaktgruppe *aliphatic carbon* (sterische Eigenschaften).

2.4.2.3 Berechnung der Wechselwirkungsfelder mit DrugScore

Ein weiterer, wissensbasierter Ansatz zur Beschreibung struktureller Eigenschaften einer Proteinbindetasche wurde mit dem Programm DrugScore von Gohlke *et al.* realisiert (Gohlke *et al.*, 2000b; Gohlke & Klebe, 2001). Aus den in der Scoringfunktion enthaltenen distanzabhängigen Paarpräferenzen bezüglich einer Reihe unterschiedlicher Atomtypen können Kontaktpräferenzen (Potenziale) an Gitterpunkten innerhalb einer Bindetasche berechnet werden. Die Visualisierung dieser Potenziale in Form von Konturdiagrammen zeigt günstige Bereiche (*hot spots*) für eine Platzierung spezifischer Atomtypen.

Zur Charakterisierung der Proteinstrukturen wurden von den in DrugScore vorhandenen Atomtypen – wie bei GRID und SuperStar – vier Sondenatome ausgewählt, die jeweils verschiedene physikochemische Eigenschaften (Donor, Akzeptor, sterisch, hydrophob) repräsentieren (Tab. 2-11). Die aus den Potenzialwerten berechneten *hot spots* für drei unterschiedliche Atomtypen in den Bindetaschen von Thrombin und Faktor Xa sind in Abb. 2-15 graphisch dargestellt.

Tab. 2-11: Atomtypen aus DrugScore zur Beschreibung der physikochemischen Eigenschaften der Proteinbindetaschen. Neben den entsprechenden chemischen Gruppen sind die zugewiesenen physikochemischen Eigenschaften angegeben.

Nr.	Atomtyp	chemische Gruppe	Eigenschaft
1	N.3	<i>sp³ amine</i>	Donor
2	O.2	<i>sp² oxygen</i>	Akzeptor
3	C.3	<i>sp³ carbon</i>	sterisch / hydrophob
4	C.ar	<i>aromatic carbon</i>	aromatisch / hydrophob

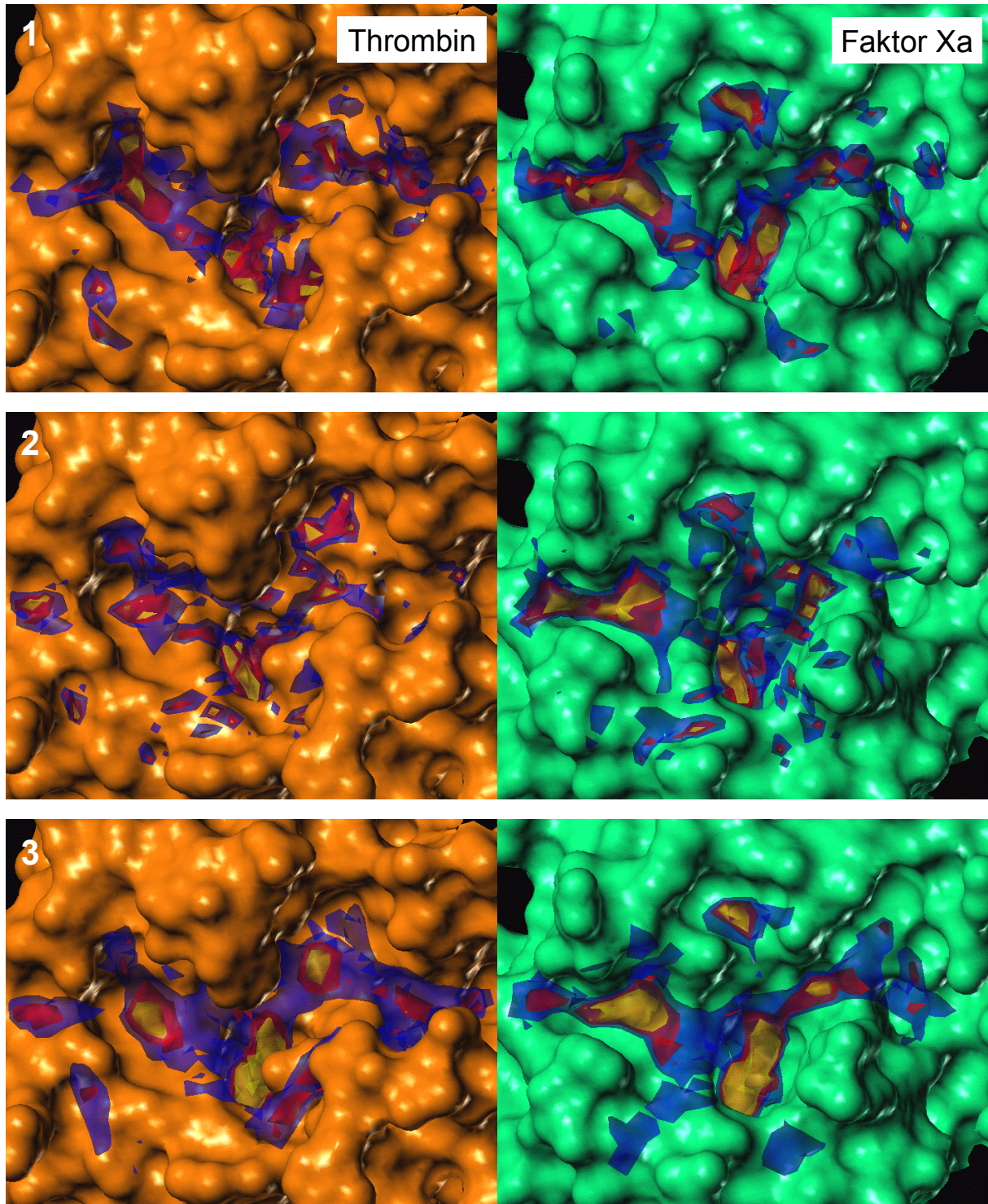


Abb. 2-15: Konturdiagramm der DrugScore-Potenzialwerte (*hot spots*) bezüglich der Bindetaschen von Thrombin und Faktor Xa. Die Potenzialwerte der Dichtekontur sind in drei Abstufungen (Werte von blau über rot nach gelb zunehmend günstiger) abgebildet. **1:** Atomtyp N.3 (Donor-Eigenschaft); **2:** Atomtyp O.2 (Akzeptor-Eigenschaft); **3:** Atomtyp C.3 (sterische Eigenschaft).

2.4.3 Chemometrische Analyse der molekularen Feldbeiträge

Die mit den verschiedenen, im letzten Abschnitt vorgestellten Verfahren (GRID, SuperStar und DrugScore) berechneten Wechselwirkungsfelder in den Bindetaschen der 16 Proteine des Datensatzes wurden mittels einer Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) chemometrisch analysiert (vgl. hierzu auch Kap. 2.2, Seite 55).

2.4.3.1 Generierung der X-Matrix

Zunächst wurden dafür die Zahlenwerte der berechneten Wechselwirkungsfelder für alle Proteine zu den jeweils verwendeten Sonden in einer Datenmatrix, der sogenannten X-Matrix abgelegt. In Abb. 2-16 ist der schematische Aufbau einer X-Matrix, wie sie in dieser Arbeit verwendet wurde, dargestellt.

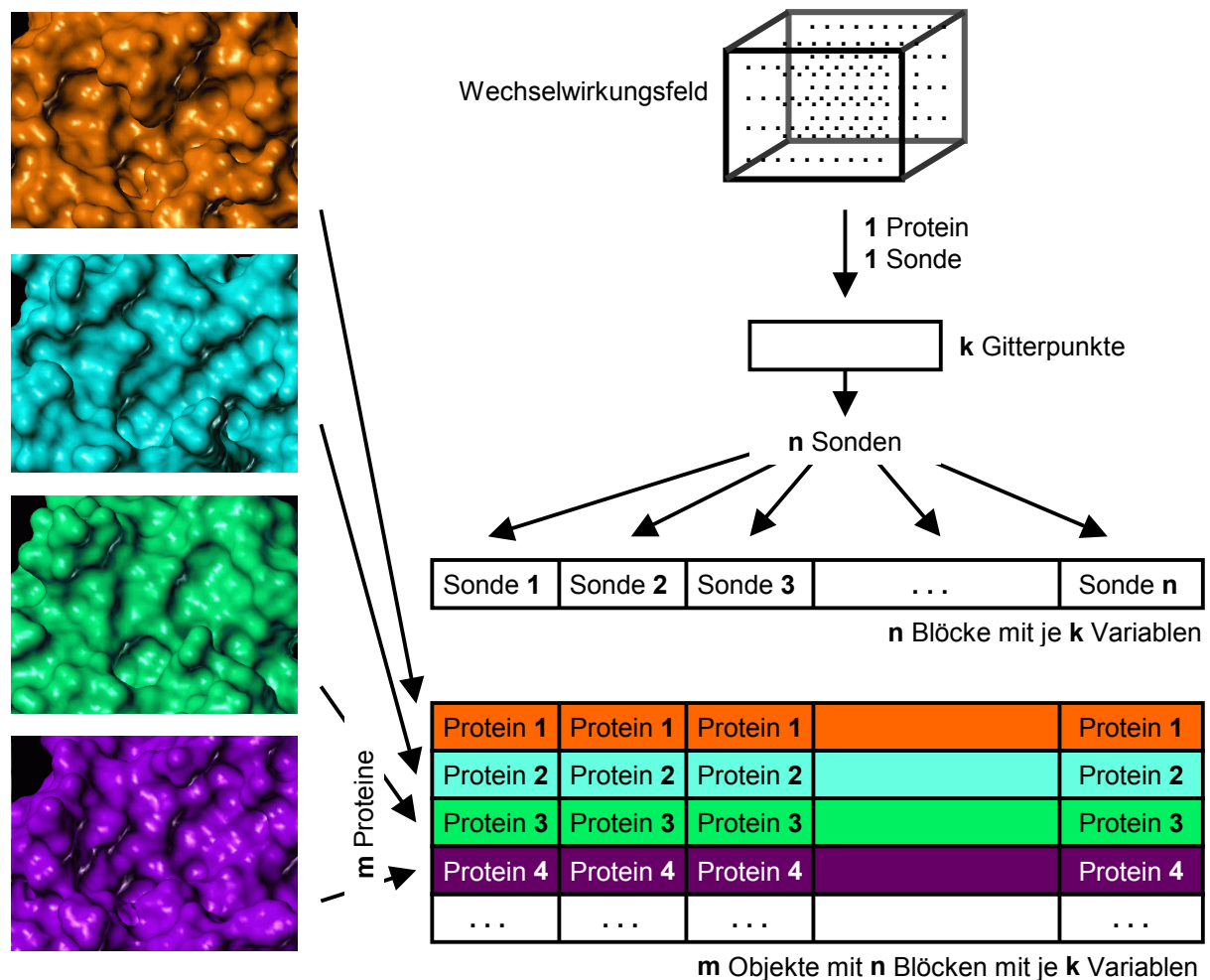


Abb. 2-16: Schematische Darstellung des Aufbaus einer X-Matrix, die die Zahlenwerte der Wechselwirkungsfelder einer Reihe von Proteinen zu mehreren Sonden enthält.

Aus Abb. 2-16 lässt sich ersehen, dass zur Erstellung einer X-Matrix zunächst das Wechselwirkungsfeld eines Proteins zu einer bestimmten Sonde an allen k Gitterpunkten berechnet wird. Die resultierenden Zahlenwerte werden in einer Zeile mit k Spalten abgelegt. Anschließend wird das Feld für die nächste Sonde berechnet und ebenfalls in einer Zeile mit k Spalten abgelegt, wobei die neu hinzugekommene Reihe an die schon bestehende Zeile angehängt wird. Somit entsteht eine Zeile mit n Blöcken, die jeweils k Spalten beinhaltet und insgesamt $n \cdot k$ Variablen aufweist. Dieser Vorgang wird für das nächste Protein mit den gleichen Sonden wiederholt. Die neu erzeugte Reihe wird als zweite Zeile in die entstehende Datenmatrix abgelegt. Für insgesamt m Proteine resultiert somit eine X-Matrix mit m Zeilen (oder Objekten) und n Blöcken mit jeweils k Variablen; sie enthält dabei insgesamt $m \cdot n \cdot k$ Zahlenwerte.

2.4.3.2 Variablenfilterung

Wie sich leicht vorstellen lässt, ist die Zahl der in der X-Matrix enthaltenen Feldwerte bei der Untersuchung mehrerer Proteine und Sonden in einem Gitter, das eine Proteinbindetasche einschließt, entsprechend groß. Ein Zahlenbeispiel soll dies verdeutlichen: Für die Untersuchung der Bindetaschen der einzelnen Serinproteasen wurde in dieser Arbeit ein Gitter mit 1 Å Gitterabstand und mit $33 \cdot 29 \cdot 25 = 23.925$ Gitterpunkten verwendet (vgl. Tab. 2-2, Seite 72). Die Berechnung der Felder zu den jeweils vier ausgewählten Sondenatomen (vgl. Tab. 2-9, Tab. 2-10 und Tab. 2-11) resultierte in $23.925 \cdot 4 = 95.700$ Variablen³⁶. Bei Berücksichtigung von 16 Proteinen entspricht dies insgesamt $95.700 \cdot 16 = 1.531.200$ Werten³⁶, die in eine chemometrischen Analyse einfließen würden. Aufgrund der großen Zahl an Variablen wurde vor der eigentlichen Hauptkomponentenanalyse eine Variablenfilterung³⁷ mit dem Programmpaket GOLPE durchgeführt³⁸. Damit ist eine mehrstufige Filterung der Variablen in die verbleibenden, sogenannten aktiven Variablen möglich. So sind von den oben genannten, insgesamt 95.700 Variablen der Datenmatrix nur ein Teil überhaupt aktiv, das heißt sie weisen bezüglich der 16 Proteine mindestens einen Funktionswert > 0 auf.

³⁶ In diesem Zusammenhang sei auf den Unterschied zwischen einer Variablen und einem Zahlenwert (engl.: *value*) hingewiesen: eine Variable bezieht sich immer auf eine Spalte der X-Matrix und wird daher aus mehreren Zahlenwerten (hier: für 16 Proteine) gebildet. Obiges Beispiel weist daher 95.700 Variablen mit jeweils 16 Werten auf.

³⁷ Eine Variablenfilterung ist unabhängig von einer nachgeschalteten Variablenselektion.

³⁸ Die verwendeten Programme sind in Kap. 4.2 nochmals kurz aufgeführt.

Ein erster Schritt der Filterung besteht darin, nur die Gitterpunkte zu berücksichtigen, die sich innerhalb der Proteinbindetaschen der untersuchten Serinproteasen befinden. Viele Punkte des rechtwinkligen Gitters befinden sich außerhalb der Bindetaschen oder im inneren Bereich des Proteins. In GOLPE besteht daher die Möglichkeit, nur den Bereich des Gitters zu berücksichtigen, der von einem vorgegebenen Templatmolekül besetzt wird („*cutout tool*“). Als Templat wurde hierzu ein „Superligand“ erstellt, der durch eine ODER-Verknüpfung aller Liganden der Inhibitor-Datensatzes (siehe Kap. 2.3.1) erzeugt wurde. Alle Gitterpunkte, die sich außerhalb des Bereichs von 4 Å um diese Templatstruktur befanden, wurden ignoriert und die dazugehörigen Variablen als inaktiv, das heißt ihre Werte auf Null gesetzt.

Günstige Wechselwirkungsbereiche (attraktive Wechselwirkungen) werden zum Beispiel in GRID in Form negativer Potenzialwerte berechnet, während positive Potenzialwerte auf repulsive Wechselwirkungen hinweisen. Standardmäßig wird in GRID der obere Grenzwert für repulsive Wechselwirkungen auf 5 kcal/mol gesetzt. Dieser Grenzwert wurde in GOLPE zur Reduzierung der aktiven Variablen auf 0 kcal/mol herabgesetzt („*maximum cutoff*“). Somit bleiben die repulsiven Wechselwirkungen im weiteren Verlauf unberücksichtigt. Zusätzlich lassen sich dadurch auch Schwierigkeiten bei der späteren graphischen Interpretation vermeiden, die sich aus der gleichzeitigen Behandlung positiver und negativer Zahlenwerte bei der Hauptkomponentenanalyse ergeben können.

In einem weiteren Schritt wurden auch alle Funktionswerte $< 0,01$ auf Null gesetzt („*zeroing values*“) und alle Variablen mit einer Standardabweichung $< 0,03$ inaktiviert („*SD cutoff*“). Eine letzte Filterung bestand in der Eliminierung unerwünschter „*n-level-Variablen*“, wobei n eine Zahl zwischen 2 und 4 darstellt. Dabei handelt es sich um Variablen, die nur n unterschiedliche Funktionswerte im Datensatz annehmen. Beispielsweise nimmt eine *2-level-Variable* nur zwei verschiedene Zahlenwerte an, wobei ein Zahlenwert nur bei maximal drei Objekten vorkommen kann. Diese Variablen mit ihren ungünstigen Werteverteilungen bereiten bei einer Hauptkomponentenanalyse oft Schwierigkeiten und sollten daher vor der Analyse ausgefiltert werden.

2.4.3.3 Durchführung der Hauptkomponentenanalyse

Somit verbleiben durch die ganzen Filterungsmaßnahmen nur mehr wenige aktive Variablen (meist 10 – 20% der ursprünglichen Variablen), die in einer nachfolgenden Hauptkomponentenanalyse (PCA) untersucht wurden. Das Prinzip einer solchen Analyse ist in Abb. 2-17 schematisch dargestellt. Eine empfehlenswerte Einführung in die Hauptkomponentenanalyse in Form eines Tutorials wird von Wold *et al.* gegeben (Wold *et al.*, 1987).

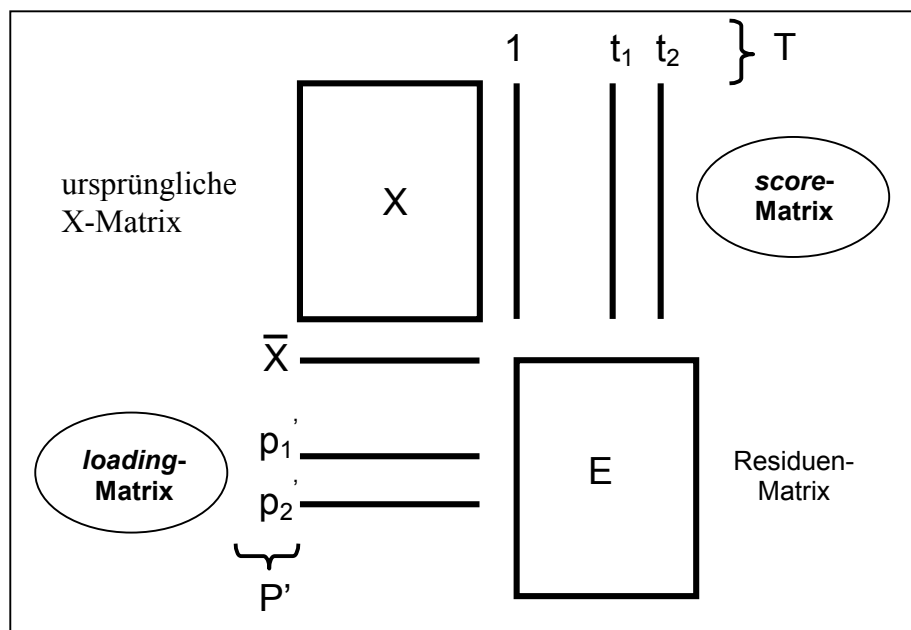


Abb. 2-17: Schematische Darstellung der Funktionsweise einer Hauptkomponentenanalyse (PCA). Die X-Matrix, die die ursprünglichen Daten enthalten, wird in eine *score*-Matrix *T* und eine *loading*-Matrix *P'* zerlegt. Die *score*-Matrix *T* enthält einige wenige Hauptkomponenten (hier: t_1 und t_2) und die *loading*-Matrix *P'* die dazugehörigen *loadings* (hier: p_1 und p_2). Nicht vom Modell erfasste Variablen verbleiben in der Residuenmatrix *E*.

Das eigentliche Ziel einer Hauptkomponentenanalyse ist die Vereinfachung einer vorgegebenen Datenstruktur in der zugrundeliegenden X-Matrix. Anschaulich gesehen erfolgt dies durch eine Koordinatendrehung im *n*-dimensionalen Raum, der durch die X-Matrix aufgespannt wird, mit dem Ziel, die Varianz in den Daten durch Auftragung in einem dimensionsreduzierten (zwei-, dreidimensionalen) Raum darzustellen. Mathematisch gesehen wird die X-Matrix zu einem Produkt zweier kleinerer Matrizen umgeformt, der *score*-Matrix

T und der *loading*-Matrix P' . Die *score*-Matrix enthält einige wenige Hauptkomponenten (*principal components*, PC), die die Objekte der X-Matrix – in diesem Fall die einzelnen Proteine – beschreiben. Die Extraktion relevanter Informationen aus der Datenstruktur durch einige Hauptkomponenten erfolgt auf der Weise, dass die Gesamtvarianz der X-Matrix möglichst gut repräsentiert wird. Je nach zugrundeliegender Datenlage werden meist 3 – 4 Hauptkomponenten benötigt, um einen großen Teil der Gesamtvarianz der Datenmatrix zu erklären. Der restliche, nicht-erklärbare Teil der X-Matrix verbleibt in der Residuen-Matrix E . Die zur *score*-Matrix entsprechende *loading*-Matrix enthält die dazugehörigen *loadings* und gibt an, wie die ursprünglichen Variablen zu den einzelnen Hauptkomponenten beitragen.

In Tab. 2-12 ist der prozentuale Anteil der Gesamtvarianz der jeweiligen Datenmatrix gegeben, der durch die ersten drei Hauptkomponenten der PCA-Modelle von GRID, SuperStar und DrugScore erklärt werden kann. Die ersten drei Hauptkomponenten des DrugScore- und GRID-Modells können dabei 48% bzw. 39% der Varianz erklären, während bei SuperStar mit gleicher Komponentenzahl nur 28% erklärbar sind.

Tab. 2-12: Erklärung der Gesamtvarianz von X bei Verwendung der ersten drei Hauptkomponenten.

Hauptkomponente	GRID	SuperStar	DrugScore
	[%]		
1	15,9	12,0	17,7
2	27,6	20,8	35,7
3	39,1	28,1	47,6

In einem *score*-Plot werden normalerweise die ersten zwei oder drei Hauptkomponenten aufgetragen, wobei jeder markierte Punkt einem einzelnen Objekt (Proteinbindetasche) entspricht. Dabei befinden sich Objekte mit ähnlichen Eigenschaften in den *score*-Plots an ähnlichen Stellen. Falls das PCA-Modell die Eigenschaften zweier Objekte in der zugrundeliegenden Datenstruktur als unterschiedlich bewertet, werden diese Objekte im zweidimensionalen PC-Raum voneinander separiert.

2.4.3.4 Auswertung des GRID-Modells

Die Auftragung der ersten drei Komponenten des PCA-Modells, das aus den mit GRID berechneten Wechselwirkungsfeldern erstellt wurde, ist in Abb. 2-18 dargestellt.

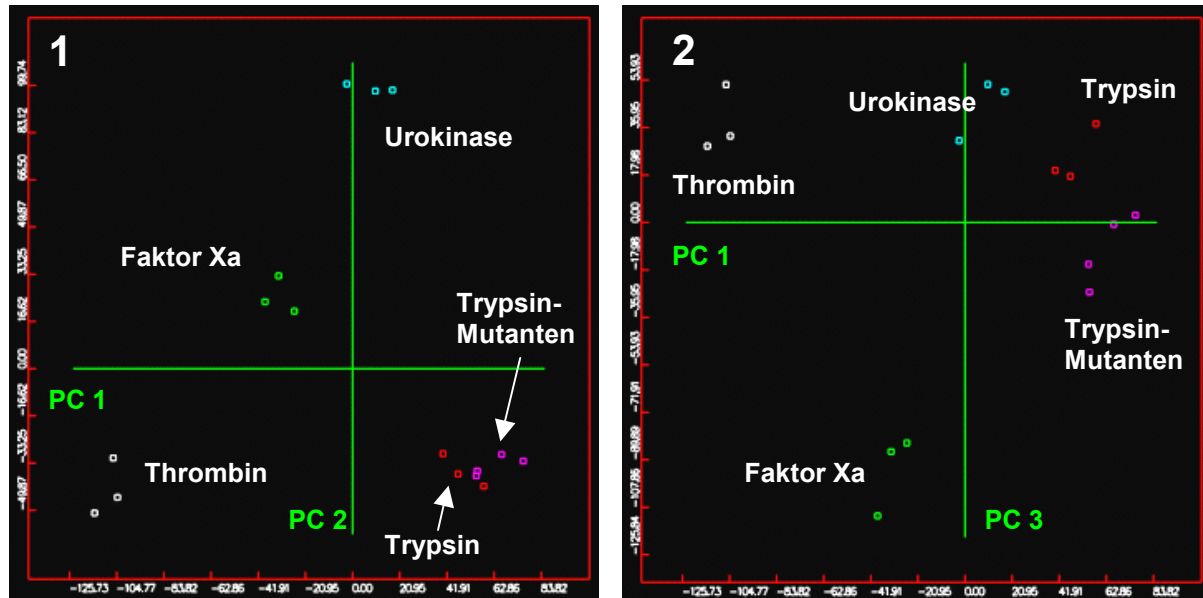


Abb. 2-18: Auftragung der ersten drei Hauptkomponenten des PCA-Modells, das aus den für 16 Proteine berechneten GRID-Feldern mit vier Sondenatomen generiert wurde. Objekte (Proteinbindetaschen) mit ähnlichen Eigenschaften finden sich im PC-Raum nahe beieinander und bilden einen Cluster. Räumlich voneinander getrennte Objekte werden von dem PCA-Modell als unterschiedlich in ihren Eigenschaften betrachtet. Die Strukturen von Thrombin (weiß), Trypsin (rot), Faktor Xa (grün), Urokinase (cyan) und der Trypsin-Mutanten (violett) finden sich jeweils in einem Cluster. **1:** Horizontale Achse: PC 1; vertikale Achse: PC 2. Die erste Hauptkomponente des Modells separiert dabei die Thrombin-Strukturen von denen des Trypsins und der Trypsin-Mutanten. Die Urokinase-Strukturen werden von der zweiten Hauptkomponente abgetrennt. **2:** Horizontale Achse: PC 1; vertikale Achse: PC 3. Die dritte Hauptkomponente diskriminiert hauptsächlich die Faktor Xa-Strukturen und trennt diese von den übrigen Proteinen ab. Zusätzlich vermag die dritte Hauptkomponente zu einem gewissen Grad zwischen den Trypsin-Strukturen und den Trypsin-Mutanten zu diskriminieren.

Aus den zweidimensionalen *score*-Plots des GRID-Modells in Abb. 2-18 wird ersichtlich, dass die verschiedenen Serinproteasen (Thrombin, Trypsin, Faktor Xa und Urokinase) jeweils einen separaten Cluster bilden. Die strukturelle Verwandtschaft der jeweiligen Serinproteasen wird also von dem PCA-Modell, das wohlgerne nur Informationen aus der Bindetasche

verarbeitet, als ähnlich erkannt. Die erste Hauptkomponente kann dabei die Strukturen von Thrombin von denen des Trypsins und der Trypsin-Mutanten separieren. Die Urokinase- und Faktor Xa-Strukturen hingegen werden erst durch die zweite bzw. dritte Hauptkomponente diskriminiert und von den anderen Proteinen abgetrennt. Daneben wird bei der Auftragung der ersten gegen die dritte Hauptkomponente ersichtlich, dass letztere zusätzlich zu einem gewissen Grad zwischen den Wildtyp-Strukturen des Trypsins und den Trypsin-Mutanten diskriminieren. Eine eindeutige und vollständige Separation kann bei Verwendung der GRID-Felder jedoch nicht erreicht werden.

2.4.3.5 Auswertung des SuperStar-Modells

Um zu überprüfen, ob die wesentlichen strukturellen Eigenschaften der Proteinbindetaschen auch durch die Berechnung der SuperStar-Deskriptoren herausgearbeitet wurden, sind die Felder der 16 Proteine, die mit SuperStar ermittelt wurden, ebenfalls einer Hauptkomponentenanalyse unterzogen worden. Abb. 2-19 zeigt die *score*-Plots des SuperStar-Modells wiederum unter Verwendung der ersten drei Hauptkomponenten.

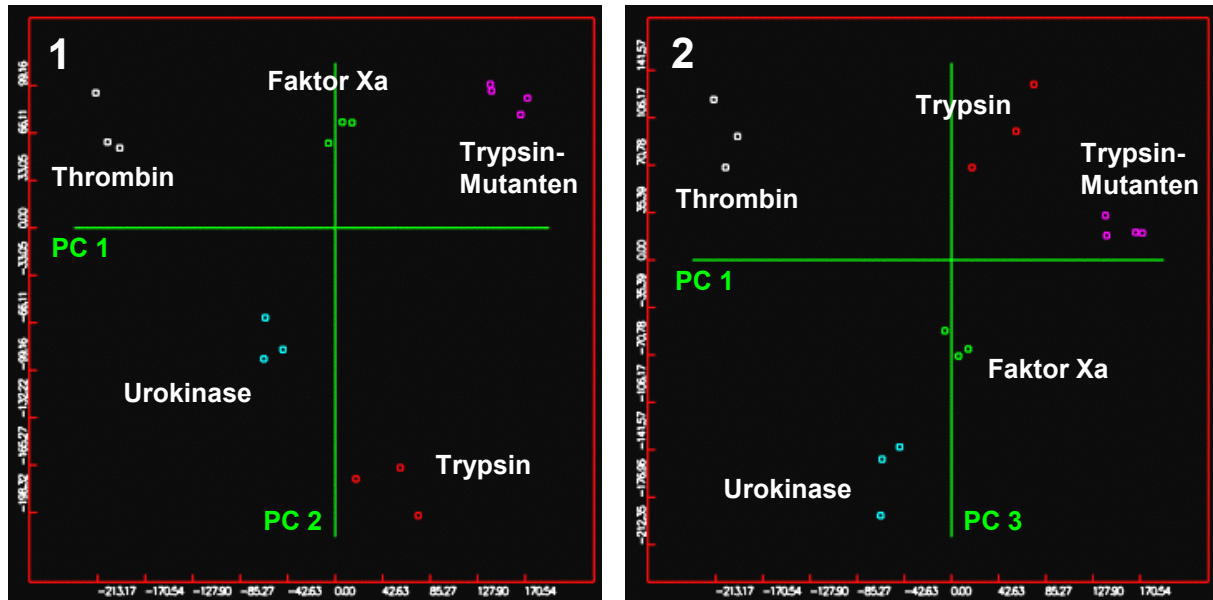


Abb. 2-19: Auftragung der ersten drei Hauptkomponenten des PCA-Modells, das aus den berechneten SuperStar-Feldern mit vier Kontaktatomen generiert wurde. 1: Die erste Hauptkomponente des Modells separiert die Strukturen von Thrombin von denen des Faktor Xa und der Trypsin-Mutanten. Die Trypsin-Strukturen werden von der zweiten Hauptkomponente abgetrennt. 2: Die dritte Hauptkomponente diskriminiert insbesondere die Urokinase-Strukturen und separiert diese von den übrigen Proteinen.

Auch im Falle des SuperStar-Modells (Abb. 2-19) bilden die verschiedenen Serinproteasen (Thrombin, Trypsin, Faktor Xa, Urokinase und Trypsin-Mutanten) jeweils separate Cluster. Dabei können die Strukturen von Thrombin, Faktor Xa und den Trypsin-Mutanten von der ersten Hauptkomponente separiert werden; die Strukturen von Trypsin und Urokinase hingegen werden erst durch die zweite bzw. dritte Hauptkomponente diskriminiert und von den anderen Proteinen abgetrennt. Interessanterweise ermöglicht dieses Modell eine deutliche Separation der Trypsin-Mutanten von den Wildtyp-Strukturen (vgl. hierzu Kap. 2.4.6, Seite 119). Bei einer Analyse nur der ersten drei Hauptkomponenten im Falle der SuperStar-Felder muss jedoch gleichzeitig berücksichtigt werden, dass damit lediglich 28% der Gesamtvarianz erklärt werden konnten (vgl. Tab. 2-12).

2.4.3.6 Auswertung des DrugScore-Modells

Abb. 2-20 zeigt die *score*-Plots der ersten drei Komponenten des PCA-Modells, das aus den mit DrugScore berechneten Wechselwirkungsfeldern der 16 Proteine erstellt wurde.

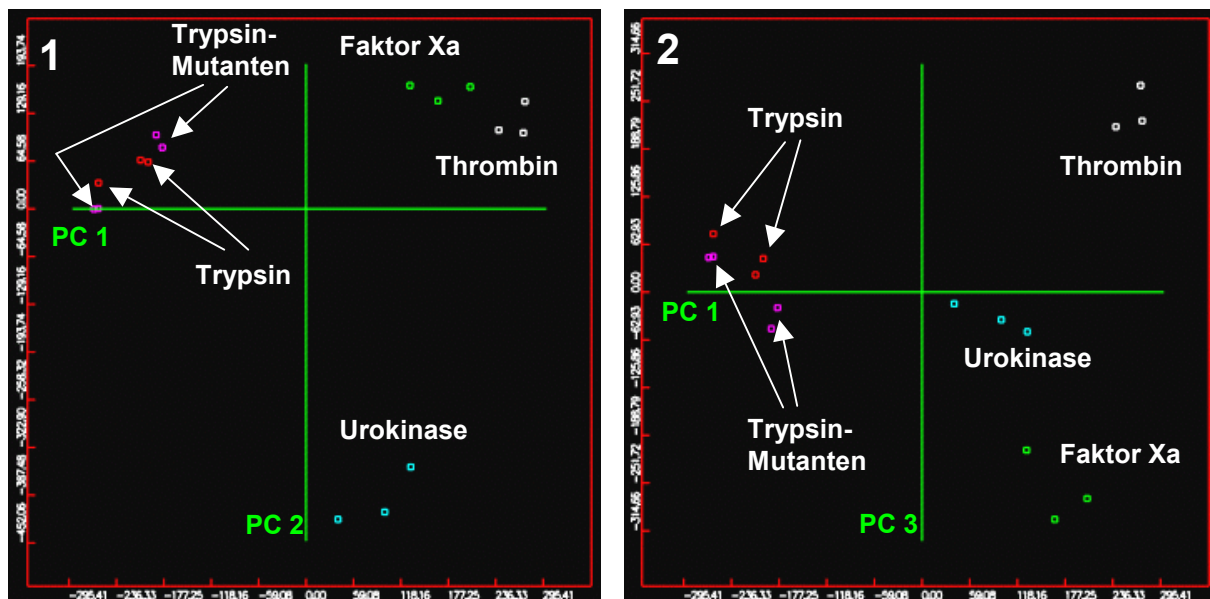


Abb. 2-20: Auftragung der ersten drei Hauptkomponenten des PCA-Modells, das aus den berechneten DrugScore-Feldern mit vier Sondenatomen generiert wurde. **1:** Die erste Hauptkomponente des Modells separiert die Strukturen von Thrombin/Faktor Xa von denen des Trypsins bzw. der Trypsin-Mutanten. Die Urokinase-Strukturen werden von der zweiten Hauptkomponente diskriminiert. **2:** Die dritte Hauptkomponente trennt die Faktor Xa- von den Thrombin-Strukturen ab. Die Trypsin-Mutanten (violett) können von den Wildtyp-Strukturen durch die ersten drei Hauptkomponenten nicht separiert werden.

Ähnlich zu den *score*-Plots des GRID- und SuperStar-Modells vermag das Modell mit den DrugScore-Feldern (Abb. 2-20) die einzelnen Proteinstrukturen zu klassifizieren. Dabei werden die Thrombin/Faktor Xa-Strukturen von denen des Trypsins bzw. der Trypsin-Mutanten durch die erste Hauptkomponente abgetrennt, wohingegen die zweite Hauptkomponente die Urokinase-Strukturen von den restlichen separieren kann. Bei Betrachtung nur der ersten beiden Komponenten fällt auf, dass im Gegensatz zu dem SuperStar-Modell sowohl bei den Thrombin- und Faktor Xa-Strukturen als auch bei den Trypsin/Trypsin-Mutanten keine Separation zu beobachten ist. Erst unter Hinzunahme der dritten Hauptkomponente werden die Strukturen von Faktor Xa von denen des Thrombins deutlich diskriminiert. Jedoch können die Trypsin-Wildtyp-Strukturen von den Trypsin-Mutanten auch bei Berücksichtigung der ersten drei Hauptkomponenten nicht separiert werden.

Aus den obigen Beobachtungen (unterschiedliches Diskriminierungsvermögen zwischen Trypsin-Strukturen und Trypsin-Mutanten) lässt sich daher schließen, dass die berechneten Felder von GRID, SuperStar und DrugScore bei der Hauptkomponentenanalyse zu unterschiedlich zusammengesetzten Hauptkomponenten kommen, in die auch die Sonden – trotz einer ähnlichen Wahl der dazugehörigen physikochemischen Eigenschaften – mit abweichenden Beiträgen eingehen.

2.4.4 Selektivitätsmodell durch differenzielle Hauptkomponentenanalyse

Die bisher gezeigten Darstellungen des *score*-Plots der PCA-Modelle aus einer Hauptkomponentenanalyse ermöglichte eine genauere Beschreibung der Objekte im Hauptkomponenten-Raum. Die Aufklärung der Beziehungen einzelner Objekte zueinander ergab Rückschlüsse darauf, welche Proteine Ähnlichkeiten zueinander aufweisen und daher im PC-Raum in Clustern erscheinen, und welche Proteine vom Modell als unterschiedlich in ihren Eigenschaften angesehen werden und von den restlichen Proteinen separiert werden. Die durch die Hauptkomponenten-Modelle geschaffenen Beziehungen der Objekte und somit auch der Eigenschaften der Proteine zueinander können auch dazu benutzt werden, Aussagen darüber zu treffen, welche Faktoren im Eigenschaftsraum für die Separation der Objekte verantwortlich sind. Wie weiter oben erläutert, werden Informationen über die Beziehung der Hauptkomponenten-Variablen zu den ursprünglichen Variablen im Eigenschaftsraum in der

loading-Matrix abgelegt. Jeder Punkt im Hauptkomponenten-Raum ist also über die *loadings* mit den realen Variablen verknüpft. Diese Variablen entsprechen den jeweiligen Eigenschaften der Sonden an den einzelnen Gitterpunkten zu den 16 Proteinstrukturen.

Um also Informationen über die Selektivität zwischen den verschiedenen Proteinklassen zu gewinnen, besteht in GOLPE im Rahmen einer differenziellen Hauptkomponentenanalyse die Möglichkeit der Verwendung sogenannter *active plots*. Dazu wird ein Vektor im Hauptkomponenten-Raum gezogen, der zwei Objekte bzw. Proteine im *score*-Plot miteinander verbindet und deren Selektivitätseigenschaften untersucht werden sollen. Anschließend werden die Differenzen für die erste und zweite Hauptkomponente zu diesen zwei Punkten gebildet. Diese Differenzen werden dann über die entsprechenden *loadings* auf die ursprünglichen Variablen im Eigenschaftsraum in Differenzen entsprechender Pseudofelder zurückprojiziert. Das Ergebnis kann in einem Konturdiagramm dargestellt werden, welches die Unterschiede der gewählten Objekte im PC-Raum im Hinblick auf eine bestimmte Sonde markiert. Es werden also diese Variablen (Gitterpunkte bezüglich einer Sonde) konturiert, die am meisten dazu beitragen, zwischen den selektierten Objekten zu differenzieren. Es sei insbesondere darauf hingewiesen, dass sich für jede gewählte Sonde unterschiedliche Konturdiagramme ergeben; es können also neben einer räumlichen Diskriminierung selektivitätsbestimmender Eigenschaften in den Bindetaschen auch Aussagen darüber getroffen werden, welche Sonden dafür verantwortlich sind (vgl. Beschreibung der CPCA-Methode, Seite 55).

Ein Beispiel zur Untersuchung selektivitätsbestimmender Faktoren mit Hilfe von *active plots* ist in Abb. 2-21 gezeigt. Ausgehend von den berechneten SuperStar-Feldern wurde eine Consensus-Hauptkomponentenanalyse (*consensus PCA*, CPCA) durchgeführt und die ersten beiden Hauptkomponenten in einem 2D-*score*-Plot aufgetragen. Bei einer CPCA können zusätzlich zu dem normalen *score*-Plot³⁹ einer PCA die *score*-Plots der einzelnen Blöcke⁴⁰ aus der X-Matrix ermittelt werden. Diese bei einer Normalisierung skalierten (*block unscaled weights*, BUW) Blöcke entsprechen den einzelnen Sonden, die in die Datenmatrix eingeflossen sind.

³⁹ wird bei einer CPCA als *superscore* bezeichnet

⁴⁰ werden bei einer CPCA als *block score* bezeichnet

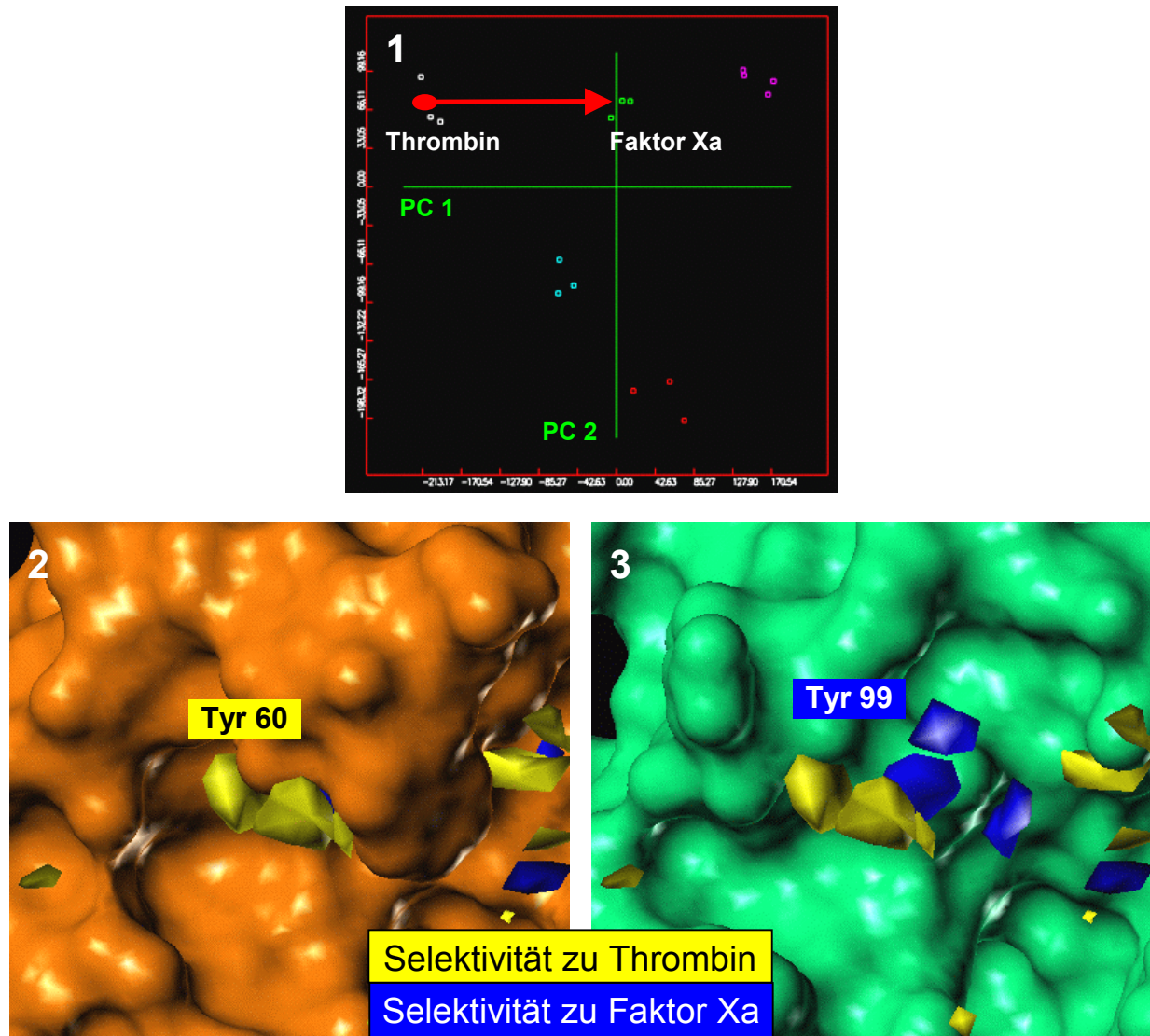


Abb. 2-21: In dem *score*-Plot des CPCA-Modells (*active plot*) wird ein Vektor (rot) von der Stelle, an dem die Thrombin-Strukturen lokalisiert sind, zu dem Punkt der Faktor Xa-Strukturen gezogen (1). Die zu diesen Objekten berechneten Differenzen der ersten beiden Hauptkomponenten werden in den ursprünglichen Raum projiziert. Die Konturierung der Differenzen im Eigenschaftsraum zeigt Regionen, die am meisten dazu beitragen, zwischen den zuvor gewählten Objekten, also zwischen Thrombin und Faktor Xa, zu diskriminieren (2+3). Der hier dargestellte *score*-Plot (1) wurde aus den *block scores* der SuperStar-Kontaktgruppe *carbonyl oxygen* erzeugt. Die Konturdiagramme bezüglich dieser Kontaktgruppe sind zusammen mit den lösungsmittelzugänglichen Oberflächen von Thrombin (2) und Faktor Xa (3) gezeigt. Gelb konturierte Bereiche weisen auf Regionen hin, die bezüglich der *carbonyl oxygen* Sonde diskriminierend für Thrombin sind; blaue Konturen differenzieren entsprechend für Faktor Xa.

Die in Abb. 2-21 konturierten, selektivitätsbestimmenden Bereiche zwischen Thrombin und Faktor Xa bezüglich der SuperStar-Kontaktgruppe *carbonyl oxygen* können direkt mit den strukturellen Gegebenheiten der jeweiligen Proteinbindetaschen in Beziehung gesetzt werden. Das Konturdiagramm in Abb. 2-21 zeigt Bereiche der Proteinbindetaschen, die als selektivitätsdiskriminierend für diese Sonde gefunden wurden. Dabei wurden überwiegend die Regionen um die aromatischen OH-Gruppen von Tyr 60 (Thrombin) und Tyr 99 (Faktor Xa) konturiert. Beide, in diesem Fall als Wasserstoffbrücken-Donor fungierende, funktionelle Gruppen tragen also zur Differenzierung der Sonde in diesem Bereich bei.

2.4.5 Selektivitätsmodell durch Differenzen von Wechselwirkungsfeldern

Die Möglichkeit einer Selektivitätsbestimmung mit Hilfe der *active plots* beschränkt sich in GOLPE auf eine Auftragung nur der ersten beiden Hauptkomponenten. Da im DrugScore-Modell die Thrombin-Strukturen von den Faktor Xa-Strukturen erst unter Verwendung der dritten Hauptkomponente separiert werden konnten (vgl. Abb. 2-20), ist die Anwendung einer differenziellen Hauptkomponentenanalyse, wie sie mit den SuperStar-Feldern durchgeführt wurde, hier nicht möglich.

Jedoch können in diesem Fall auch Aussagen über diskriminierende Bereiche in Proteinbindetaschen getroffen werden, indem die Differenz der Wechselwirkungsfelder zweier zu untersuchender Proteine (hier: Thrombin und Faktor Xa) gebildet wird. Dabei werden an jedem Gitterpunkt eines Wechselwirkungsfeldes (zum Beispiel Feld des C.3-Sondeatoms von DrugScore) die Differenzen der Funktionswerte zu beiden Proteinbindetaschen berechnet und in einem Konturdiagramm graphisch dargestellt (vgl. Differenzbildung der QSAR-Koeffizienten; Kap. 2.3.5, Seite 95). Abb. 2-22 zeigt das Konturdiagramm des Differenzfeldes zwischen Thrombin und Faktor Xa bei Verwendung des C.3-Sondeatoms von DrugScore (vgl. Einzelfelder Abb. 2-15).

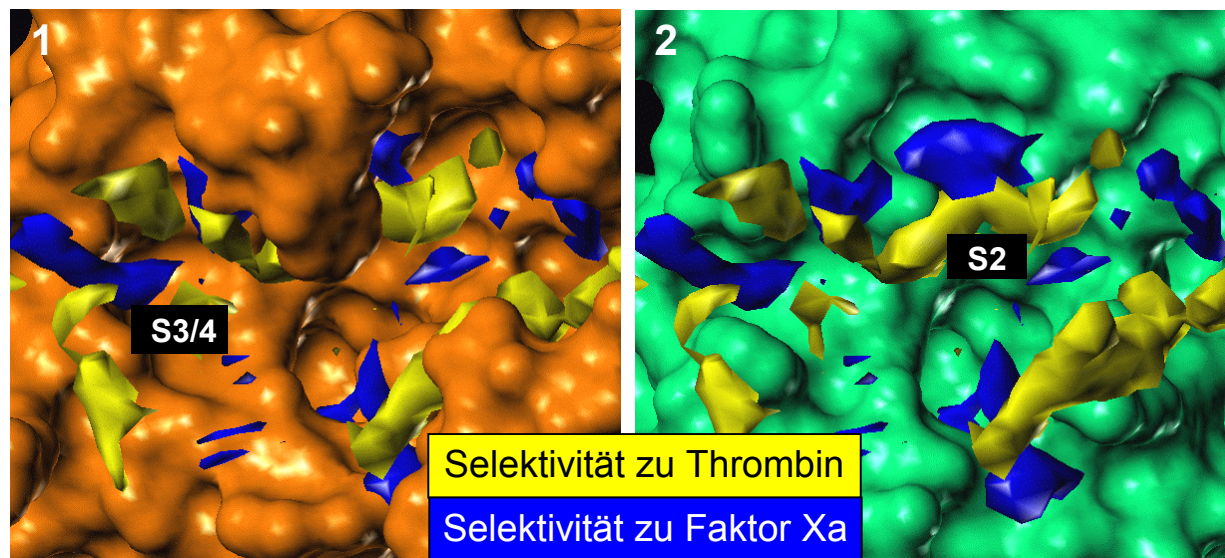


Abb. 2-22: Differenz der Wechselwirkungsfelder zwischen Thrombin und Faktor Xa. Die Konturdiagramme bezüglich des C.3-Atomtyps von DrugScore sind zusammen mit den lösungsmittelzugänglichen Oberflächen von Thrombin (1) und Faktor Xa (2) gezeigt. Gelb konturierte Bereiche weisen auf Regionen hin, die bezüglich der jeweiligen Sonde diskriminierend für Thrombin sind; blaue Konturen differenzieren entsprechend für Faktor Xa.

Die Konturierung der Differenzfelder bezüglich der C.3-Sonde von DrugScore in Abb. 2-22 zeigt, dass der Bereich der S2-Tasche hauptsächlich diskriminierend für Thrombin ist. Ein Teil der S3/4-Tasche hingegen ist für die gleiche Sonde selektivitätsbestimmend für Faktor Xa. Beide Beobachtungen stimmen mit den strukturellen Gegebenheiten der Bindetaschen von Thrombin und Faktor Xa überein, die eine sterische Besetzung der S2-Tasche (Thrombin) bzw. S3/4-Tasche (Faktor Xa) favorisieren.

Erstaunlicherweise stimmen die selektivitätsdiskriminierenden Faktoren, die bei einer chemometrischen Analyse der Proteinstrukturen erhalten wurden, gut mit den Ergebnissen der Selektivitätsmodelle überein, die aus dem Inhibitor-Datensatz – also aus reinen Ligandinformationen (vgl. Kap. 2.3) – mit 3D-QSAR-Methoden erhalten wurden. In Abb. 2-23 sind beispielsweise die graphischen Ergebnisse der sterischen Eigenschaften gegenübergestellt, die aus dem Selektivitätsmodell mit CoMSIA (Differenz der QSAR-Koeffizienten) und dem Differenzfeld der Proteinbindetaschen (C.3-Sonde von DrugScore) zwischen Thrombin und Faktor Xa erhalten wurden.

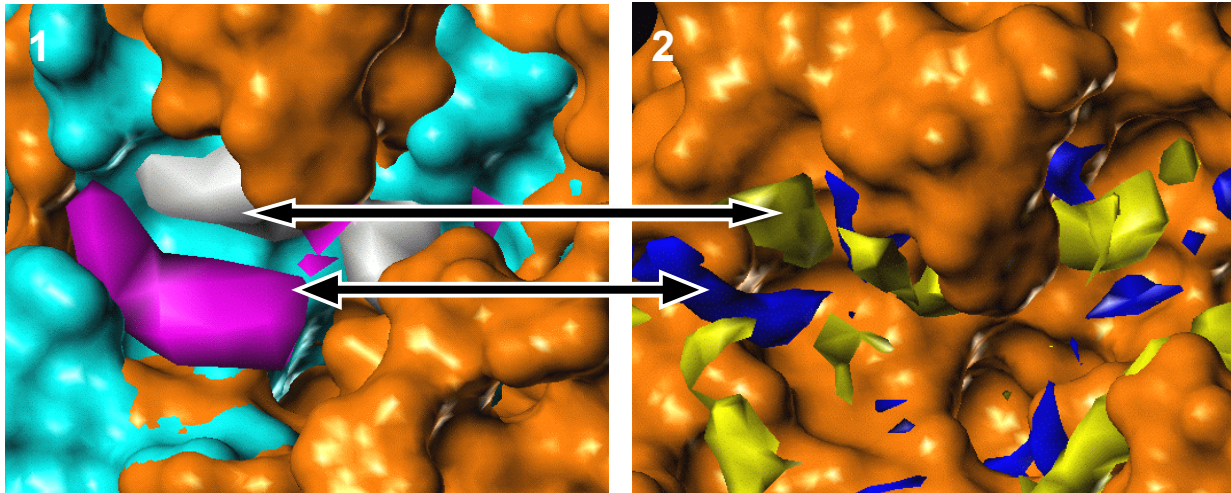


Abb. 2-23: Vergleich der Konturdiagramme der sterischen Eigenschaften, wie sie aus den Selektivitätsmodellen mit CoMSIA (1) und aus der Differenz der Wechselwirkungsfelder (2) zwischen Thrombin und Faktor Xa erhalten wurden. 1: entspricht dem Konturdiagramm in Abb. 2-12(2); 2: entspricht dem Konturdiagramm in Abb. 2-22(1).

2.4.6 Analyse diskriminierender Faktoren zwischen Trypsin und Trypsin-Mutanten

Wie an dem *score*-Plot des SuperStar-Modells in Abb. 2-19 (Seite 113) schon erläutert wurde, konnten die Strukturen der Trypsin-Mutanten von den Wildtyp-Strukturen des Trypsins bei Verwendung der SuperStar-Felder diskriminiert werden. Eine differenzielle Hauptkomponentenanalyse, wie sie schon zwischen Thrombin und Faktor Xa durchgeführt wurde (vgl. Abb. 2-21), ließ sich daher ebenfalls für eine genauere Analyse der Trypsin-Strukturen und den Mutanten einsetzen. Die Auswertung des Konturdiagramms in Abb. 2-24 zeigt Regionen in den Proteinbindetaschen, die bezüglich der Donor-Eigenschaften (Kontaktgruppe *amino hydrogen*) diskriminierend zwischen Trypsin und den Trypsin-Mutanten sind. Im wesentlichen werden drei Bereiche in der Bindetasche hervorgehoben, die sich in der Nähe der Aminosäure-Positionen 99, 151 und 217 befinden. Gerade an diesen Positionen wurden aber auch die entsprechenden Aminosäuren der Trypsin-Strukturen mutiert. So wurde an der Position 99 die Aminosäure Leucin (Wildtyp) gegen ein Tyrosin (Mutante) ausgetauscht. Während die Seitenkette von Leucin keine Donor-Eigenschaften exponiert (keine gelbe Kontur in diesem Bereich), zeigt an dieser Stelle die Tyrosin-Seitenkette bei der Trypsin-Mutante eine für Donor-Eigenschaften günstige Region.

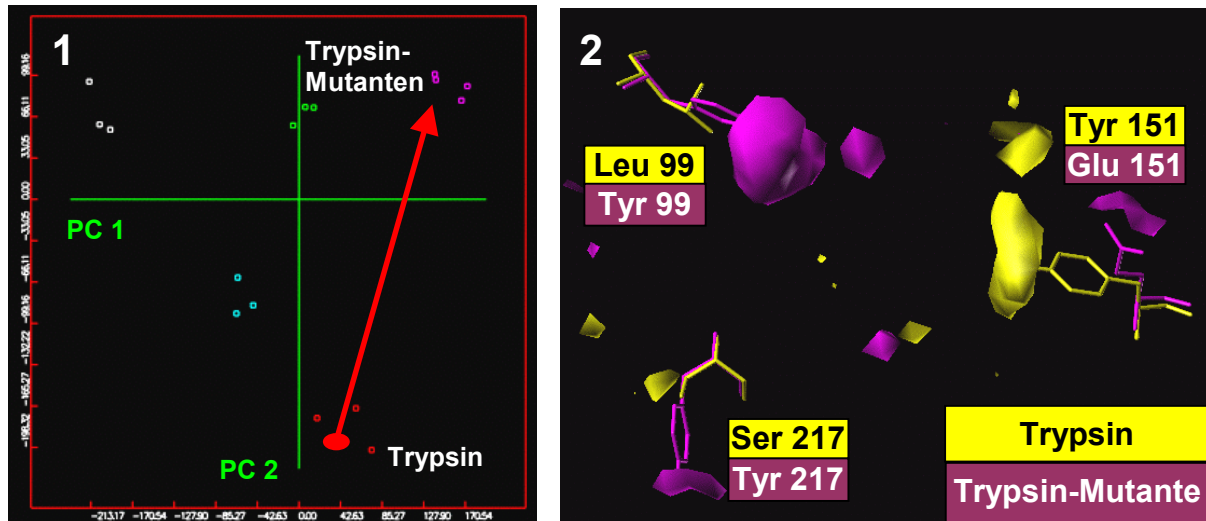


Abb. 2-24: In dem *score*-Plot des CPCA-Modells (*active plot*) wird ein Vektor (rot) von der Position der Trypsin-Strukturen zu dem Punkt der Trypsin-Mutanten gezogen (1). Die Konturierung der Differenzen zeigt Regionen, die am meisten dazu beitragen, zwischen Trypsin und den Trypsin-Mutanten zu diskriminieren (2). Der hier dargestellte *score*-Plot entspricht dem *block score* der SuperStar-Kontaktgruppe *amino hydrogen*. Gelb konturierte Bereiche markieren Regionen, die bezüglich dieser Sonde diskriminierend für Trypsin sind; violette Konturen diskriminieren entsprechend für die Trypsin-Mutanten. Die zu den diskriminierenden Konturen gehörenden Aminosäuren sind farblich hervorgehoben.

Ähnlich verhält es sich in der Umgebung der Position 151, an dem ein Tyrosin (Wildtyp) gegen ein Glutamat (Mutation) ausgetauscht wurde. In diesem Fall können beide Aminosäuren (Tyrosin: OH-Gruppe; Glutamat: Carboxylatgruppe) in unmittelbarer Nachbarschaft ihre Donor-Eigenschaften exponieren. Entsprechend zeigen an der Position 217 (Mutation Serin gegen Tyrosin) die für Donor-Eigenschaften günstige Bereiche ebenfalls in räumlich unterschiedliche Bereiche, die zu einer Diskriminierung bei der differentiellen Hauptkomponentenanalyse beitragen.

2.5 Zusammenfassung

Die in diesem Kapitel vorgestellten Studien zur Ermittlung von Eigenschaften, die selektivitätsbestimmend für Proteine der Familie von Serinproteasen sein können, lassen sich in zwei Bereiche einteilen. Zum einen wurde gezeigt, wie Informationen einer Reihe von

Liganden, die an mehrere Proteinstrukturen binden können, im Rahmen einer dreidimensionalen quantitativen Struktur-Wirkungsbeziehung zur Erstellung von Selektivitätsmodellen benutzt werden können. Dazu wurden zunächst zu jedem Protein – Thrombin, Trypsin, Faktor Xa und Urokinase – separate, vorhersagekräftige Einzelmodelle generiert. Sie erlauben Aussagen, welche Bereiche der jeweiligen Proteinbindetasche mit den dazugehörigen Eigenschaften sich günstig oder ungünstig auf die Affinität eines Liganden auswirken können.

Anschließend konnten durch eine direkte Gegenüberstellung der Bindungsaffinitäten analoge Modelle erstellt werden, die Aussagen über selektivitätsbestimmende Bereiche in einer Bindetasche ermöglichen. Dazu wurden die Differenzen der Affinitätswerte zu jeweils zwei Enzymen gebildet und die so erzeugten Selektivitätsfaktoren als abhängige Variablen zur Korrelation mit den physikochemischen Eigenschaften der Inhibitorstrukturen herangezogen. Die markierten Bereiche der damit erstellten Konturdiagramme verwiesen auf Bereiche, in denen eine bestimmte Eigenschaft die Selektivität in Richtung des einen oder anderen Proteins steigert. In einem weiteren Schritt konnten unterschiedliche Bindungsmodi der Liganden in den jeweiligen Rezeptorstrukturen dadurch berücksichtigt werden, indem nicht die Affinitäten zu zwei Proteinen, sondern die Koeffizienten zweier separater QSAR-Modelle kombiniert wurden. Dabei konnte bei der graphischen Auswertung beider Selektivitätsanalysen gezeigt werden, dass beide Modelle zueinander konsistente Ergebnisse lieferten.

Der zweite Ansatz bestand darin, selektivitätsbestimmende Faktoren direkt aus der Zusammensetzung der Proteinbindetaschen zu extrahieren. Verschiedene Methoden zur Berechnung von Wechselwirkungsfeldern in Proteinbindetaschen wurden verglichen. Neben dem oftmals eingesetzten GRID-Verfahren wurden zusätzlich Deskriptoren aus SuperStar und DrugScore zur Charakterisierung der Bindetaschen herangezogen. Mit einem erweiterten Verfahren der Hauptkomponentenanalyse – der Consensus-Hauptkomponentenanalyse (CPCA) – wurden die anhand verschiedener Sonden berechneten Wechselwirkungsfelder chemometrisch ausgewertet. Der Vorteil der angewendeten CPCA-Methode bestand vor allem darin, mit Hilfe einer differenziellen Hauptkomponentenanalyse Bereiche in den Proteinbindetaschen zu ermitteln, die im Hinblick auf eine bestimmte Probe diskriminierend wirken und somit selektivitätsbestimmend für ein bestimmtes Enzym sein können.

Ein abschließender Vergleich beider Ansätze zur Ermittlung selektivitätsdiskriminierender Faktoren – 3D-QSAR-Analysen bzw. chemometrische Auswertung physikochemischer Eigenschaften von Proteinbindetaschen mit Hilfe von Hauptkomponentenanalysen – zeigte eine gute Übereinstimmung der herausgefilterten Kriterien. Dies ist umso erfreulicher, da bei beiden Verfahren völlig unterschiedliche Informationen (3D-QSAR: ligandbasiert; PCA: proteinbasiert) zur Selektivitätsbestimmung ausgewertet wurden.

3 STRUKTUR-BASIERTES DESIGN VON HEMMSTOFFEN DER FARNESYLTRANSFERASE

3.1 Einleitung

Die Tumorthherapie von Krebserkrankungen, die für ca. 30% der Todesfälle in Deutschland verantwortlich sind, stellt eine der großen medizinischen Herausforderungen der heutigen Zeit dar. Während eine chirurgische Entfernung und Strahlentherapie bei lokalisierten Tumoren meist noch möglich ist, gewinnt die Chemotherapie bei fortgeschrittenem Tumorwachstum und bei Ausbildung von Metastasen zunehmend an Bedeutung. Die Wirkung heutiger Chemotherapeutika basiert vor allem auf einem Eingriff in den Zellzyklus und führt zum Tod unkontrolliert proliferierender Zellen. Dabei werden jedoch neben den malignen Tumorzellen auch normale, insbesondere schnellwachsende Zellen geschädigt. Diese unselektive Zytotoxizität vieler Chemotherapeutika stellt somit einen limitierenden Faktor bei der Behandlung von Tumoren dar. Neben der Suche nach Wirkstoffen mit geringeren toxischen Nebenwirkungen wird daher unter Berücksichtigung der Zell- und Pathophysiologie von Tumoren auch eine Reihe neuer Therapieansätze geprüft. Hierzu zählen die Anwendung monoklonaler Antikörper gegen spezifische Antigene der Tumorzelle bzw. gegen Wachstumsfaktor-Rezeptoren, die Therapie mit Hemmstoffen der Tumorangiogenese und Metastasierung sowie mit Immunmodulatoren, die eine spezifische Immunantwort gegen den Tumor fördern sollen. Die Applikation spezifischer Gene in Tumorzellen wie zum Beispiel Tumorsuppressorgene (Gentherapie) oder das Ausschalten bestimmter Gene durch Antisense-DNA oder -RNA (Antisense-Therapie) zur Behandlung von Tumoren kann derzeit noch nicht abschließend beurteilt werden (Forth *et al.*, 2001).

Ein weiterer Ansatzpunkt bei der Entwicklung neuartiger Therapiekonzepte ist der Einsatz von Substanzen zur Induktion der Differenzierung oder Apoptose (programmierter Zelltod) von Tumorzellen. Eine derzeit vielversprechende Möglichkeit stellt der Eingriff in den sogenannten Ras-Signaltransduktionsweg dar (Kap. 3.1.1); dabei spielt die Familie der Ras-Proteine eine wichtige Rolle bei der Übertragung eines biologischen Signals (Kap. 3.1.2). Ihre Aktivität erreichen diese membranständigen Enzyme erst durch weitere Veränderungen ihrer Struktur, den sogenannten posttranslationalen Modifikationen (Kap. 3.1.3).

3.1.1 Der Ras-Signaltransduktionsweg

Der Ras-Signaltransduktionsweg gilt heute als Paradigma für die Weiterleitung eines biologischen Signals aus dem extrazellulären Raum in den Zellkern (Travis, 1993; Gibbs *et al.*, 1994). Im Laufe einer mehrstufigen Reaktionsabfolge werden dabei Signale von membranständigen Rezeptor-Tyrosinkinasen auf zytosolische Serin- und Threoninkinasen übertragen, was letztendlich zu der Auslösung einer zellulären Antwort führt (Abb. 3-1).

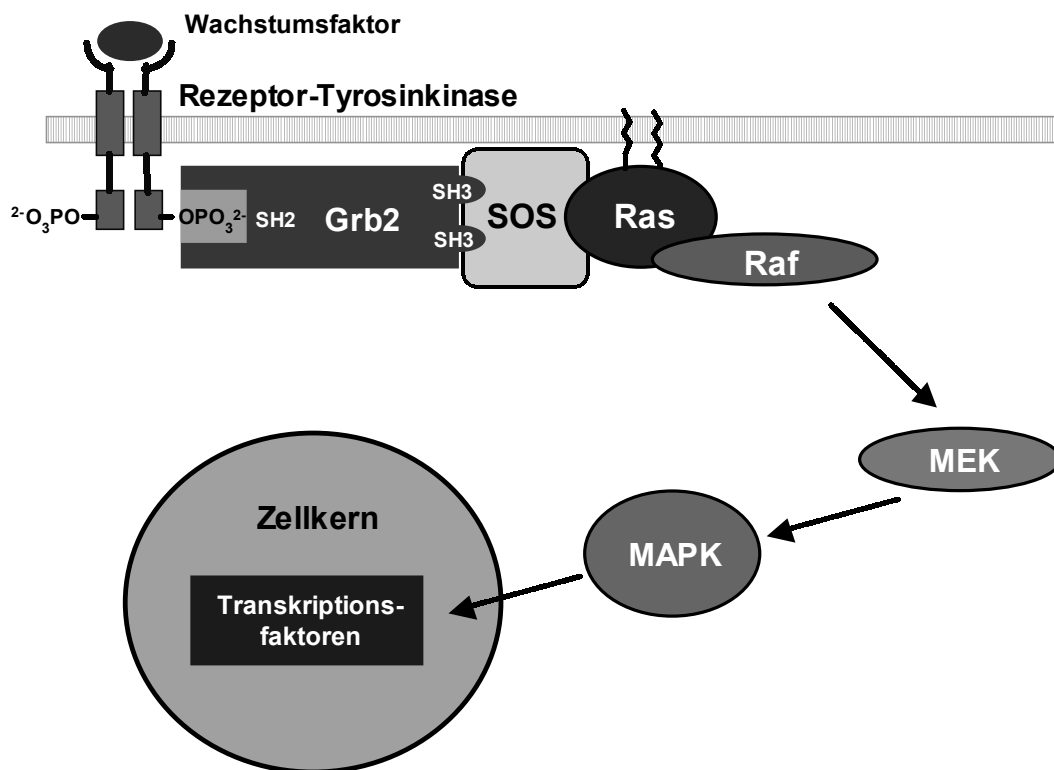


Abb. 3-1: Vereinfachte schematische Darstellung des Ras-Signaltransduktionsweges. Bindet ein extrazellulärer Ligand (Wachstumsfaktor) an seinen Rezeptor (Rezeptor-Tyrosinkinase), löst dies eine Reihe von nachgeschalteten Reaktionen aus (genauere Beschreibung siehe Text). Die membranständigen Ras-Proteine spielen dabei eine entscheidende Rolle bei der Signalweiterleitung. Das extrazelluläre Signal wird schließlich in den Zellkern weitergegeben und resultiert in einer entsprechenden Antwort der betroffenen Zellen wie Differenzierung und Proliferation.

Die Weiterleitung eines biologischen Signals wird durch die Bindung von Wachstumsfaktoren an die extrazelluläre Domäne ihrer dazugehörigen Rezeptoren eingeleitet. Wichtige Wachstumsfaktoren sind zum Beispiel der epidermale Wachstumsfaktor (*Epidermal*

Growth Factor, EGF), das menschliche Wachstumshormon (*human Growth Hormone*, hGH), der Insulin-verwandte Wachstumsfaktor (*Insuline-like Growth Factor*, IGF) und der Blutplättchenwachstumsfaktor (*Platelet-Derived Growth Factor*, PDGF), der bei der Wundheilung eine Rolle spielt. Diesen Liganden ist gemein, dass sie an einen transmembranständigen Rezeptor binden und eine Dimerisierung dieses monomeren Rezeptors verursachen. Hierdurch kommt es in einem autokatalytischen Prozess zur Phosphorylierung spezifischer Tyrosine, die im zytoplasmatischen Teil der Rezeptor-Tyrosinkinase (RTK) lokalisiert sind. Diese „Kreuzphosphorylierung“ steigert zum einen die Tyrosinkinase-Aktivität, andererseits wird dadurch eine hochaffine Bindetasche für weitere Signalproteine ausgebildet, die eine Erkennungsdömane für Phosphotyrosinreste besitzen. Somit können die phosphorylierten Tyrosinreste an die SH2-Domäne⁴¹ des Grb2-Proteins⁴² binden, das mit seiner SH3-Domäne⁴¹ wiederum ein weiteres Protein, das SOS-Protein⁴³, aktiviert. Das SOS-Protein katalysiert im weiteren Verlauf den wichtigsten Schritt in der Signaltransduktionskaskade, den Austausch des an Ras gebundene GDP gegen GTP. Dadurch wird das Ras-Protein in seinen aktiven Zustand versetzt, was eine Änderung seiner Konformation zur Folge hat. Eine detaillierte Beschreibung von Ras und seine zentrale Bedeutung als „molekularer Schalter“ bei der Signaltransduktion findet sich im nächsten Abschnitt. Die kurzfristig aktivierte Form von Ras ermöglicht eine Wechselwirkung mit dem in der Signalkaskade nachgeschalteten Protein, der Serin-/Threonin-spezifischen Proteinkinase Raf⁴⁴. Durch die Aktivierung von Raf werden im weiteren eine ganze Reihe intrazellulärer Phosphorylierungsreaktionen ausgelöst. Raf bindet und phosphoryliert MEK⁴⁵, die ihrerseits wiederum einen Phosphatrest auf weitere MAP-Kinasen⁴⁶ überträgt. Diese drei in Reihe geschalteten Kinasen bilden das sogenannte MAP-Kinase-Modul; sie werden alternativ auch als MAP-Kinase (MAPK), MAP-Kinase-Kinase (MAPKK; entspricht MEK) und MAP-Kinase-Kinase-Kinase (MAPKKK; entspricht Raf) bezeichnet. Die aktivierten MAP-Kinasen wandern schließlich in den Zellkern und phosphorylieren dort Transkriptionsfaktoren (*c-jun* und *c-fos*), die eine zelluläre Antwort auf das ursprüngliche

⁴¹ SH = *Sarcoma Homology*; ein Protein mit diesem Strukturelement wurde zuerst als das tumorverursachende Prinzip des Vogel-Sarkom-Virus entdeckt.

⁴² Grb = *Growth factor receptor binding protein*

⁴³ SOS = *Son Of Sevenless*; ein Ausdruck aus der *Drosophila*-Genetik

⁴⁴ Raf = *Rapid fibrosarcoma* oder *Ras activated factor*

⁴⁵ MEK setzt sich aus MAP und ERK (*Extracellular signal Regulated Kinase*) zusammen

⁴⁶ MAP = *Mitogen Activated Protein*

Signal des extrazellulären Wachstumsfaktors vermitteln. Das Resultat ist eine Differenzierung und Proliferation der betroffenen Zellen.

In letzter Zeit hat sich jedoch herausgestellt, dass die verschiedenen Funktionen von Ras nicht mit einer einzigen (linearen) Kaskade von MAP-Kinasen erklärt werden können. Weitere Signale, die Ras vor- und nachgeschaltet sind, müssen berücksichtigt werden, um die Vielfalt der aufgeklärten Ras-Funktionen zu verstehen. So ist mittlerweile bekannt, dass Ras-Proteine nicht nur von Rezeptor-Tyrosinkinase (RTK) gesteuert werden, sondern auch von zytoplasmatischen Tyrosinkinase (zum Beispiel Src) und von G-Protein-gekoppelten Rezeptoren (*G-Protein Coupled Receptor*, GPCR) aktiviert werden können. Andererseits interagiert Ras nicht nur mit Raf, sondern mit einer Vielzahl weiterer Effektoren (McCormick, 1996). Neben dem schon angesprochenen Signalpfad des MAP-Kinase-Moduls ist die Signalweitergabe des MEKK-Pfades ebenfalls für die Zelldifferenzierung und -proliferation verantwortlich (Minden *et al.*, 1994). Dahingegen wird der von Ras angestoßene Rac/Rho-Weg (*cell morphology pathway*) die Aufgabe der Organisation des Zytoskeletts zugeschrieben (Joneson *et al.*, 1996). Ein weiterer, wichtiger Effektor ist die Lipidkinase Phosphatidylinositol-3'-Kinase (PI3K), die die Produktion von Phosphoinositiden, den unmittelbaren Vorläufern von *second messenger* Botenstoffen, reguliert (Yan *et al.*, 1998). Leider gestaltet sich die Aufklärung des vollständigen Netzwerks der Signaltransduktionskaskade als außerordentlich schwierig; anzunehmen ist jedoch, dass Ras immer mehrere Signalwege gleichzeitig steuert. Somit hängt die letztendliche Signalwirkung von einer Vielzahl Ras-abhängiger und -unabhängiger Signalpfade ab, die sowohl parallel verlaufen als auch miteinander verknüpft sein können (Rowinsky & Patnaik, 2000).

3.1.2 Die Familie der Ras-Proteine

Eine zentrale Bedeutung innerhalb des Ras-Signaltransduktionsweges kommt den membranständigen Ras-Proteinen zu, die die Aktivität der Signalkaskade beeinflussen können. Das Protein Ras⁴⁷ steht dabei stellvertretend für eine Klasse von Proteinen, den sogenannten kleinen Guanylnukleotid-bindenden Proteinen. Ihre gemeinsamen Eigenschaften sind ein relativ niedriges Molekulargewicht von 20 – 25 kD (188 oder 189 Aminosäuren) und

⁴⁷ Ras = *Rat sarcoma*; die Gene der Ras-Proteine wurden zuerst bei Retroviren gefunden, die bei Ratten Bindegewebstumore (Sarkome) auslösen.

die Fähigkeit, Guaninnukleotide zu binden. Die Ras-Proteine fungieren als „molekulare Schalter“ und liegen in zwei verschiedenen Zuständen vor. Der inaktive Zustand („Schalter aus“) liegt bei GDP-gebundenem Ras vor und geht bei einer Bindung von GTP in den aktiven Zustand („Schalter ein“) über (Abb. 3-2).

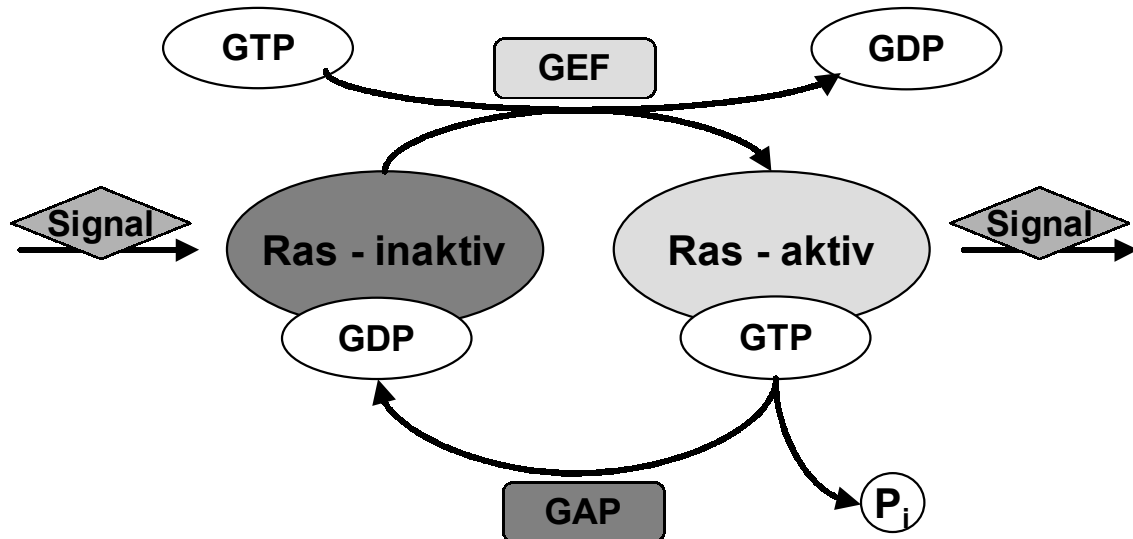


Abb. 3-2: Schematische Darstellung des Ras-Zyklus. Das in der GDP-gebundenen Form inaktive Ras wird bei einem eintreffenden Signal durch die Wirkung von Austauschfaktoren (GEF) aktiviert. Das in der aktivierten Form GTP-gebundene Ras-Protein kann in seiner veränderten Konformation Wechselwirkungen mit Effektoren eingehen. Damit werden weitere, nachgeschaltete Proteine aktiviert und das eingehende Signal weitergegeben. Durch Abspaltung eines Phosphatrestes P_i aus dem gebundenen GTP kehrt das Ras-Protein wieder in seinen ursprünglichen, inaktiven Zustand zurück. Die dazu notwendige GTPase-Aktivität wird durch Interaktion mit weiteren Austauschfaktoren (GAP) aktiviert.

Im zunächst inaktiven Zustand ist Ras an GDP (Guanosindiphosphat) gebunden. Bei einem ankommenden Signal wird es durch die Wirkung von Guaninnukleotid-Austauschfaktoren (*Guanine nucleotide Exchange Factor*, GEF) aktiviert, die eine Dissoziation von GDP und somit die Bindung von GTP (Guanosintriphosphat) ermöglichen. Innerhalb des Ras-Signaltransduktionsweges übernimmt das SOS-Protein (vgl. Kap. 3.1.1) die Aufgabe des Austauschfaktors. Das so aktivierte Ras-Protein verändert seine Konformation derart, dass eine Bindung mit Effektoren⁴⁸ und damit eine Aktivierung weiterer, in der Signaltransduktionskaskade nachgeschalteter Proteine, wie zum Beispiel der MAP-Kinasen

⁴⁸ Moleküle, die spezifisch mit der GTP-Form wechselwirken und ein Signal weiterleiten

(vgl. Kap. 3.1.1), ermöglicht wird. Die Signalweiterleitung wird durch die Abspaltung eines Phosphatrestes (P_i) von GTP unterbrochen. Die dazu erforderliche, relativ langsame GTPase-Aktivität ist auf den Ras-Proteinen selbst vorhanden („intrinsische Aktivität“), wird jedoch durch Ras-spezifische, GTPase-aktivierende Proteine (*GTPase Activating Protein*, GAP) beschleunigt. Das dadurch entstandene, GDP-gebundene Ras liegt wieder in seiner ursprünglichen, inaktiven Form vor (Wittinghofer & Waldmann, 2000).

Auftretende Punktmutationen von Gly 12, Gly 13 oder Gln 61 im aktiven Zentrum von Ras haben jedoch zur Folge, dass die Ras-Proteine ihre intrinsische GTPase-Aktivität verlieren. Folglich kann Ras nicht mehr inaktiviert werden und verbleibt in seiner aktiven Form. Dadurch werden kontinuierlich Wachstumssignale an die nachgeschalteten Proteine der Phosphorylierungskaskade in den Zellkern weitergegeben. Mutierte Ras-Proteine verlieren also ihre Fähigkeit zur Kontrolle der Zellteilung und wirken somit als Onkogene⁴⁹. Dies führt letztendlich zu einer malignen Entartung der Zelle und der Entstehung eines Tumors (Gibbs *et al.*, 1996; Cox & Der, 1997).

Aus den drei bekannten humanen *Ras*-Genen *H-*, *K-* und *N-Ras* sind vier verschiedene, strukturell und funktionell homologe Isoformen von Ras-Proteinen identifiziert worden: H-Ras, K-RasA, K-RasB und N-Ras. Mutierte Ras-Proteine treten bei durchschnittlich 25 – 30% aller menschlichen Tumoren auf, wobei die Häufigkeit bei einzelnen Tumoren (zum Beispiel Pankreas, Kolon, Lunge) bei 60% und höher liegen kann (Tab. 3-1).

Tab. 3-1: Vorkommen mutierter Ras-Proteine in verschiedenen Tumoren.

Tumortyp	Ras-Protein	Häufigkeit (%)
Pankreas	K-Ras	80 – 90
Kolon	K-Ras	30 – 60
Dünndarm	H-Ras	31
Lunge	K-Ras	27 – 60
Leber	K- und N-Ras	12 – 26
Schilddrüse	H-, K- und N-Ras	0 – 60
Leukämie	K- und N-Ras	6 – 40

⁴⁹ Gene, die in Tieren oder Zellkulturen Tumoren induzieren können

Die Mutation der Isoform K-RasB tritt in humanen Tumoren am häufigsten in Erscheinung (Leonard, 1997).

3.1.3 Posttranslationale Modifikationen

Ein Ansatzpunkt bei der Entwicklung neuer Antitumorwirkstoffe liegt in der Beeinflussung des Wachstumssignals onkogener Ras-Proteine („Anti-Ras-Therapie“). Damit Ras seine Funktion als „molekularer Schalter“ ausüben kann, müssen sowohl normale als auch mutierte Ras-Proteine an der Innenseite der Zellmembran fest verankert sein. Dazu durchläuft Ras verschiedene Stufen einer sogenannten posttranslationalen Modifikation. Die biologisch zunächst inaktive, zytosolische Vorstufe von Ras wird also in mehreren Reaktionen so modifiziert, dass es schließlich an der Zellmembran fixiert werden kann (Abb. 3-3).

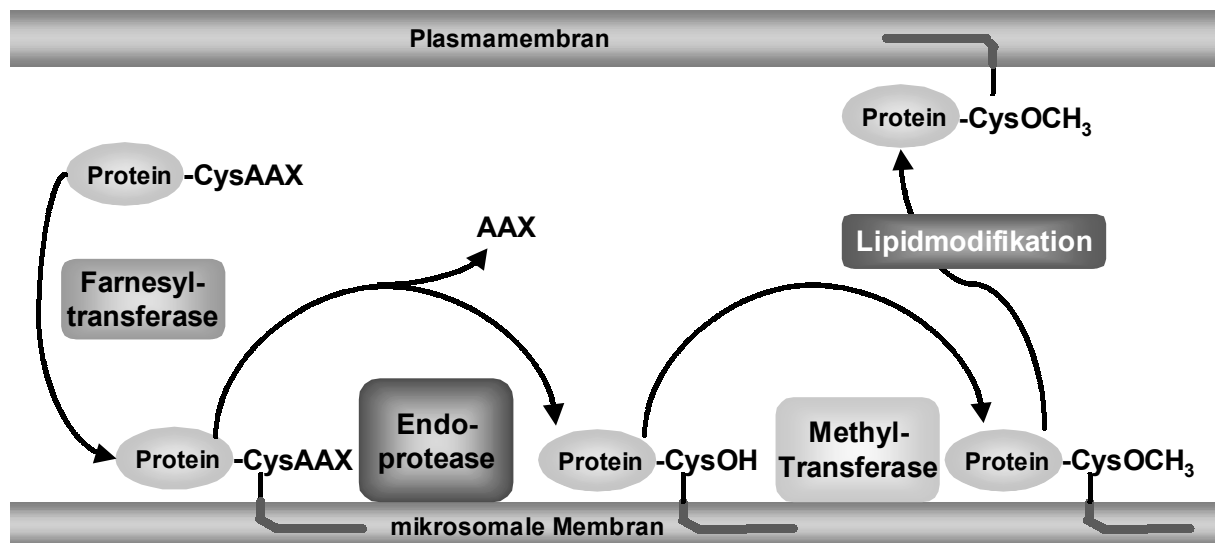


Abb. 3-3: Schematische Darstellung der verschiedenen Reaktionen bei der posttranslationalen Modifikation von Ras-Proteinen. In einem ersten Schritt überträgt die Farnesyltransferase einen Farnesylrest auf die SH-Gruppe der Cysteinkette der C-terminalen CAAX-Sequenz. Anschließend erfolgt Abspaltung des AAX-Teils durch eine spezifische Endoprotease. Das freie C-terminale Cystein wird durch eine Methyltransferase in einen Methylester umgewandelt. Die letztendliche Verankerung von Ras in der Plasmamembran erfolgt durch zusätzliche Lipidmodifikationen, wobei Palmitoylreste auf die SH-Gruppen weiterer Cysteinreste in der unmittelbaren Nachbarschaft der CAAX-Sequenz übertragen werden. Die dabei mögliche Beteiligung einer mutmaßlichen Palmitoyltransferase konnte im Experiment bis jetzt noch nicht eindeutig geklärt werden.

Der erste und für die Ras-Aktivität entscheidende Schritt bei der posttranslationalen Modifizierung ist die Übertragung eines Farnesylrestes von Farnesylpyrophosphat auf die SH-Funktion des Cysteinrestes der C-terminalen CAAX-Sequenz von Ras (Zhang & Casey, 1996). Diese für die Farnesyltransferase spezifische Erkennungssequenz („CAAX-Box“) setzt sich aus vier Aminosäuren zusammen; dabei steht „C“ für Cystein, „A“ für eine aliphatische Aminosäure (meist Valin, Leucin und Isoleucin) und „X“ für einen Methionin-, Serin- oder Glutaminrest. Durch eine spezifische Endoprotease (*Isoprenylated Protein Protease*, IPPase) wird das C-terminale AAX-Tripeptid der CAAX-Sequenz abgespalten. Eine Methyltransferase sorgt für eine anschließende reversible Methylierung des endständigen S-Farnesylcysteins. Bei H- und N-Ras erfahren zusätzliche Cysteinreste, die sich unmittelbar in der Nähe des farnesylierten Cysteins befinden, eine weitere Lipidmodifikation durch die Übertragung von Palmitoylresten auf die Thiolgruppen dieser Aminosäuren. Während die Farnesylierung essenziell für die Funktion von Ras im normalen und transformierten Zustand der Zelle ist (Kato *et al.*, 1992; Khosravi-Far & Der, 1994; Cox & Der, 1997), scheint die Palmitoylierung für die zelltransformierende Aktivität von Ras keine zwingende Voraussetzung zu sein; sie erhöht jedoch die Hydrophobizität und erleichtert die Verankerung des Ras-Proteins in der Plasmamembran. Somit ist diese zusätzliche posttranslationale Modifizierung wichtig für eine vollständige Entfaltung der biologischen Aktivität (Duder & Gelb, 1996). K-RasB weist anstelle einer Palmitoylierungsstelle eine Lysin-reiche Region in der Nähe der CAAX-Sequenz auf (Long *et al.*, 2000). Diese basischen, positiv geladenen Aminosäuren können elektrostatische Wechselwirkungen mit negativ geladenen Phospholipiden der Membran eingehen und so zu einer stärkeren Fixierung von Ras in der Plasmamembran beitragen (Ghomashchi *et al.*, 1995).

Prinzipiell bieten sich alle vier Reaktionsschritte der posttranslationalen Modifikation (Farnesyltransferase, Endoprotease, Methyltransferase, Palmitoyltransferase) als Angriffspunkt für die Entwicklung neuer Wirkstoffe an, die den Reifungsprozess der inaktiven, zytosolischen Vorstufenform von Ras unterbinden können. Da jedoch nur der erste Schritt, die Übertragung eines Farnesylrestes, obligatorisch für die Funktion der Ras-Proteine ist, erscheint die Suche nach Hemmstoffen der Farnesyltransferase am vielversprechendsten zu sein (Leonard, 1997; Qian *et al.*, 1997; Sebti & Hamilton, 1998). So wurde in Zellkulturen und an verschiedenen Tiermodellen bereits gezeigt, dass Inhibitoren der Farnesyltransferase

die transformierende Aktivität mutierter Ras-Proteine unterdrücken können (Cox & Der, 1997).

3.2 Die Farnesyltransferase

Die Farnesyltransferase (FTase), ein Zink-Metalloenzym, gehört zur Familie der Prenyltransferasen (Casey & Seabra, 1996; Tamanoi & Sigman, 2001). Zu dieser Gruppe zählen neben der FTase zwei Varianten der Geranylgeranyltransferase (GGTase), die als GGTase I und GGTase II bezeichnet werden. Sowohl bei der FTase als auch der GGTase I handelt es sich um ein heterodimeres Protein mit einer α - und einer β -Untereinheit. Die 48 kD große α -Untereinheit ist bei beiden Proteinen identisch, während die β -Untereinheit mit 46 kD bei der FTase bzw. 43 kD bei der GGTase I unterschiedlich ist (Seabra *et al.*, 1991). Beide Proteine erkennen ihre Substrate anhand der CAAX-Sequenz; dabei steht in beiden Fällen „C“ für Cystein und A für aliphatische Aminosäuren (meist Valin, Leucin und Isoleucin). Die endständige Aminosäure hingegen ist entscheidend für die Spezifität des Substrats zu FTase und GGTase I. Wenn sich an der Stelle von „X“ die Aminosäuren Methionin, Serin oder Glutamin befinden, bindet das Substrat spezifisch an FTase; dahingegen ist das Substrat spezifisch für GGTase I, wenn die „X“-Position durch Leucin und Phenylalanin besetzt ist. Beide Enzyme transferieren Prenylreste aus den entsprechenden Pyrophosphaten auf die Thiolfunktion des C-terminalen Cysteins der CAAX-Erkennungssequenz. FTase überträgt dabei einen Farnesylrest (C-15 Isopreneinheit) und GGTase I einen Geranylgeranylrest (C-20 Isopreneinheit) auf das zu prenylierende Substrat (Zhang & Casey, 1996; Roskoski & Ritchie, 1998).

Im Unterschied zu FTase und GGTase I ist die GGTase II als heterotrimeres Protein aufgebaut. Es setzt dabei solche Substrate um, die an ihrem C-Terminus eine CC-, CXC- oder CCX-Erkennungssequenz aufweisen. „C“ steht in diesem Fall wieder für die Aminosäure Cystein, während „X“ eine beliebige Aminosäure sein kann (Seabra *et al.*, 1992a; Seabra *et al.*, 1992b).

3.2.1 Struktur und Funktion der Farnesyltransferase

Mittlerweile sind eine Reihe von Kristallstrukturen der FTase bekannt, darunter die zuerst gelöste Struktur des freien Enzyms mit einer Auflösung von 2,3 Å (Park *et al.*, 1997) oder die Struktur des Rattenenzym im Komplex mit dem Co-Substrat Farnesylpyrophosphat (FPP) bei einer Auflösung von 3,4 Å (Long *et al.*, 1998). Kürzlich wurde auch die Kristallstruktur des ternären Komplexes (PDB-Code 1QBQ) aus FTase, einem FPP-Analogen und eines CAAX-Peptidomimetikums (*N*-Acetyl-Cys-Val-Ile-selenoMetOH) veröffentlicht (Strickland *et al.*, 1998).

Wie schon weiter oben erwähnt, ist die FTase ein heterodimeres Metalloprotein bestehend aus einer α - und einer β -Untereinheit mit einem Molekulargewicht von 48 bzw. 46 kD. Die Sekundärstruktur beider Untereinheiten basiert überwiegend aus α -Helizes, wobei die α -Untereinheit 15 α -Helizes und die β -Untereinheit insgesamt 14 α -Helizes aufweisen (Abb. 3-4).

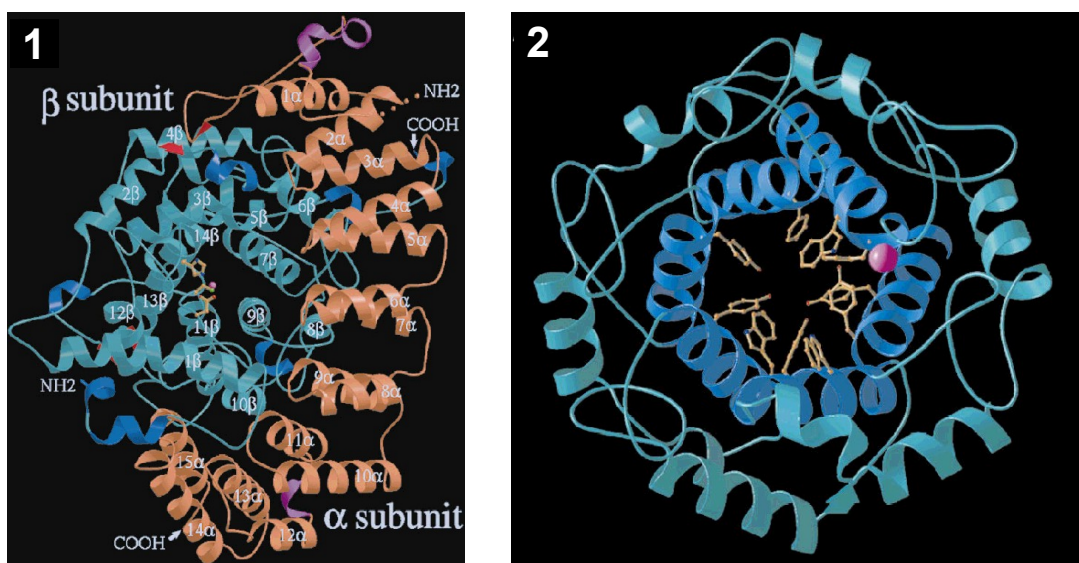


Abb. 3-4: Darstellung der Sekundärstruktur von FTase. 1: (Seitenansicht) Das Heterodimer ist aus einer α - (rot) und einer β -Untereinheit (blau) aufgebaut; beide Untereinheiten bestehen überwiegend aus α -Helizes. 2: (Frontalansicht, horizontale 90° Drehung) Konservierte aromatische Seitenketten (rot) bilden eine hydrophobe Tasche, die das Substrat bindet; das Zink-Ion (magenta) ist als Kugel dargestellt. Nähere Beschreibung der Sekundärstruktur siehe Text.

In der α -Untereinheit formen sich Helix 2 bis 15 zu sieben paarweisen Haarnadelstrukturen (*helical hairpins*) und bilden dadurch eine rechtsgewundene Superhelix aus. Zwölf α -Helizes der β -Untereinheit sind zu einer Faszstruktur (α - α -barrel) gefaltet. Sechs dieser zwölf α -Helizes sind paarweise parallel zueinander ausgerichtet und stellen den Boden der Faszstruktur dar. Antiparallel dazu bilden die sechs verbleibenden α -Helizes der β -Untereinheit – jeweils paarweise parallel zueinander – die Form des Fasses. Eine Seite der α - α -Faszstruktur ist nach außen hin geöffnet und formt einen tunnelförmigen Hohlraum mit einer Tiefe von ca. 14 Å und einem Durchmesser von ca. 15 – 16 Å. Diese hydrophobe Tasche wird von konservierten, aromatischen Aminosäureseitenketten (hauptsächlich Tryptophan und Tyrosin) ausgekleidet und bildet somit das aktive Zentrum der FTase.

Das aktive Zentrum besteht aus zwei Bindetaschenbereichen, die sich an der Stelle eines gebundenen Zink-Ions treffen. Das katalytisch aktive Zink-Ion koordiniert dabei die Thiolgruppe des Cysteinrestes der CAAX-Sequenz eines Substrats oder thiolhaltigen Inhibitors. Dieses Zink-Ion ist mit den Aminosäuren Cys 299 β , Asp 297 β und His 362 β sowie einem Wassermolekül fünffach koordiniert (Abb. 3-5).

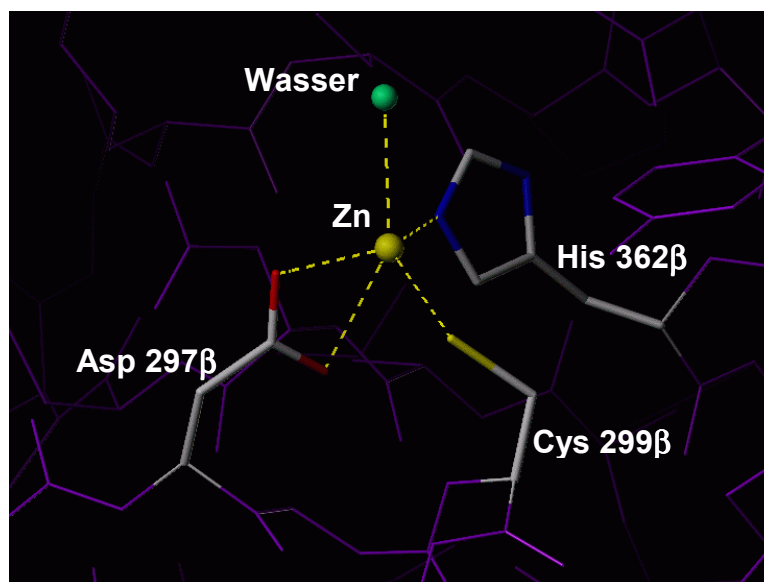


Abb. 3-5: Darstellung der Zink-Bindestelle der FTase. Bei ungebundenem Substrat wird das katalytisch aktive Zink-Ion von den Aminosäuren Asp 297 β , Cys 299 β , His 362 β und einem Wassermolekül fünffach koordiniert. Das strukturell konservierte Wasser wird bei Substratbindung von der Thiofunktion des Cysteins verdrängt.

Das konservierte Wassermolekül wird bei Bindung eines Substrats von der Thiofunktion des Cysteins der CAAX-Sequenz verdrängt. Eine der beiden Bindetaschen exponiert relativ hydrophile Aminosäuren an die Oberfläche des Enzyms und interagiert – teilweise mit Wassermolekülen überbrückt – mit den Aminosäuren des CAAX-Tetrapeptids. Die Bindestelle des zweiten Substrats Farnesylpyrophosphat (FPP) stellt eine Region (*isoprenoid binding cleft*) mit aromatischen Aminosäureseitenketten (Trp 102 β , Tyr 205 β , Tyr 251 β , Trp 303 β) dar. Ihre Größe reicht aus, um den Farnesylrest, nicht jedoch das um fünf Kohlenstoffeinheiten längere Geranylgeranylpyrophosphat (GGPP) aufzunehmen („*molecular ruler hypothesis*“). Neben dem katalytisch wirksamen Zink-Ion besitzt die FTase in ihrem aktiven Zentrum ein Magnesium-Ion, welches mit dem Pyrophosphatrest von FPP wechselwirkt (Abb. 3-6).

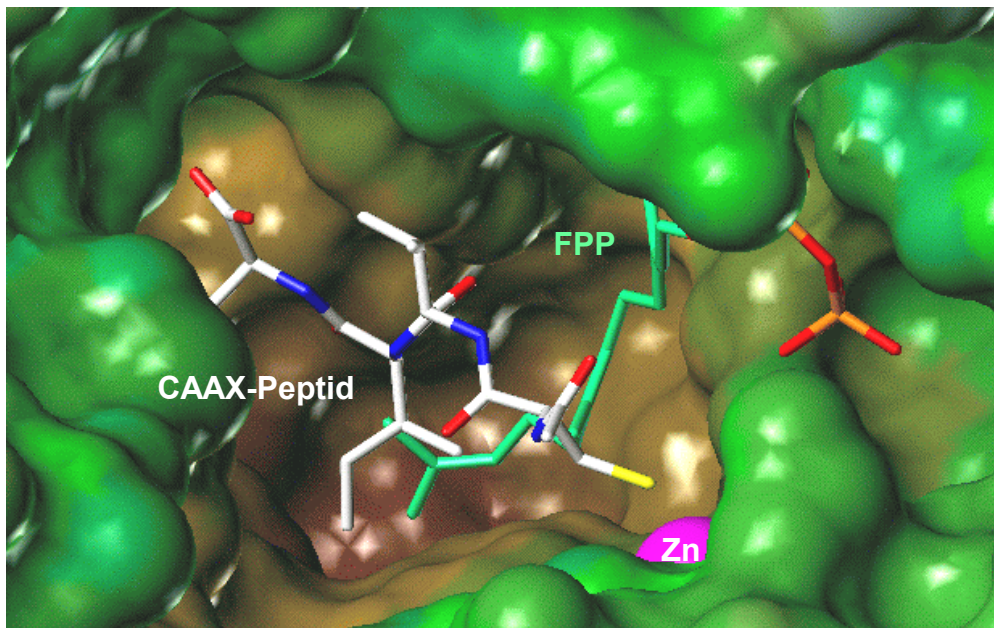


Abb. 3-6: Darstellung des aktiven Zentrums der Farnesyltransferase aus der Kristallstruktur des ternären Komplexes von FTase, einem FPP-Analogen und eines CAAX-Peptidomimetikums (PDB-Code 1QBQ). Das CAAX-Tetrapeptid (C-Atome weiß) bindet in eine hydrophile Tasche und ist durch die Thiolgruppe seines Cysteins mit dem Zink-Ion (magenta) koordiniert (das Magnesium-Ion ist nicht gezeigt). Der Farnesylrest von Farnesylpyrophosphat (C-Atome grün) bindet in einer zweiten, hydrophoben Tasche. Das aktive Zentrum der FTase ist mit seiner lösungsmittelzugänglichen Connolly-Oberfläche gezeigt; zusätzlich sind die hydrophoben Eigenschaften der Aminosäuren auf der Enzymoberfläche dargestellt (braun: hydrophob; grün: hydrophil).

Der katalytische Mechanismus des Transfers eines Farnesylrestes von FPP auf die SH-Funktion des Cysteins ist noch nicht vollständig aufgeklärt. Zunächst bindet FPP in die hydrophobe Isoprenoid-Bindetasche. Dabei werden die Konformationen der Seitenketten von Arg 202 β und Met 193 β derart verändert (Abb. 3-7), dass hydrophobe Wechselwirkungen zwischen der FTase und FPP verstärkt werden (Long *et al.*, 1998).

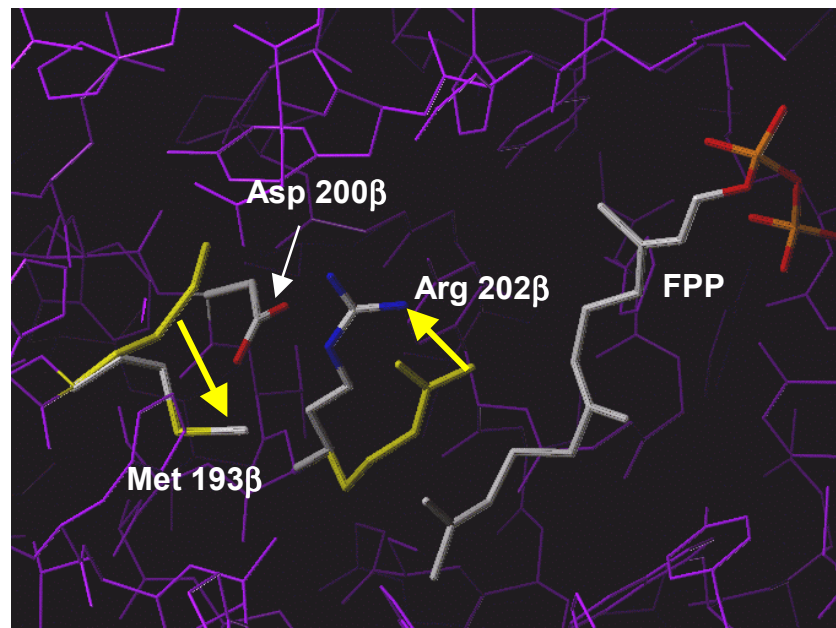


Abb. 3-7: Anpassung der Isoprenoid-Bindetasche bei Bindung von FPP. Die Konformationen der Seitenketten von Arg 202 β und Met 193 β verändern sich so (gelbe Pfeile), dass hydrophobe Wechselwirkungen mit FPP verstärkt werden (Konformationen der Apo-Struktur gelb dargestellt). Gleichzeitig wird die Konformation von Arg 202 β durch die Seitenketten von Met 193 β und Asp 200 β stabilisiert.

Anschließend bindet das CAAX-Substrat in einer gestreckten Konformation, wobei die Thiolgruppe des Cysteins an das Zink-Ion koordiniert; somit liegt temporär ein ternärer Komplex – gebildet aus FTase, FPP und dem eigentlichen Substrat – vor. Für die Übertragung des Farnesylrestes von FPP auf die Thiolfunktion des Cysteins wird sowohl ein elektrophiler als auch nukleophiler Mechanismus diskutiert (Abb. 3-8).

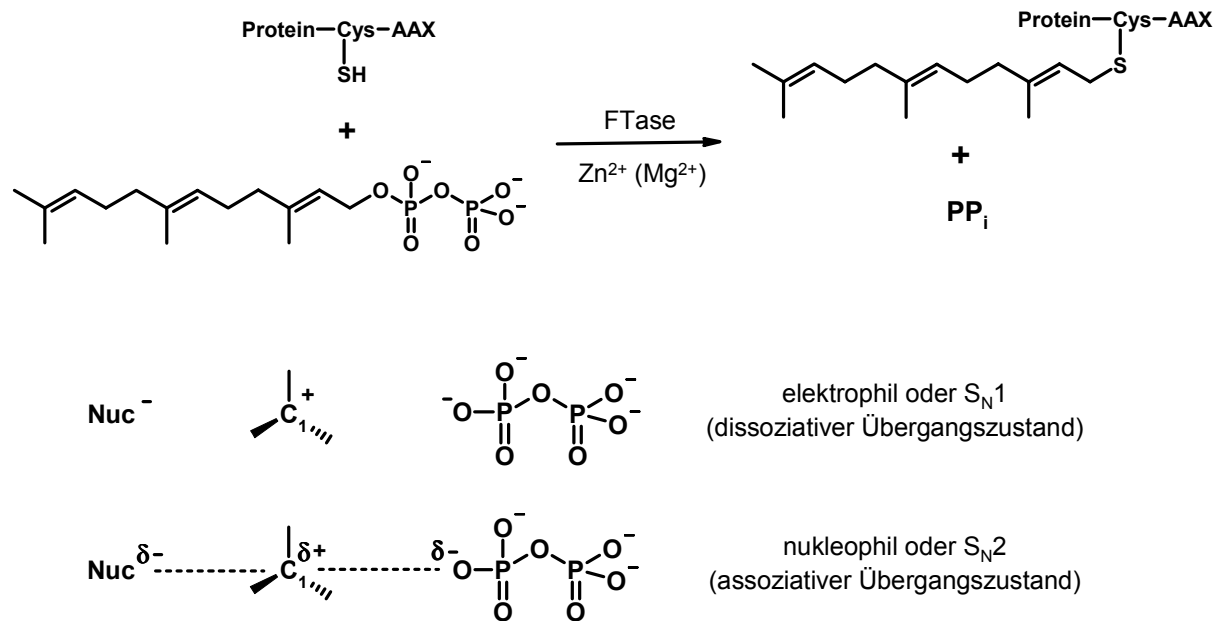


Abb. 3-8: Schema der Farnesylierungsreaktion der FTase und der zwei möglichen (dissoziativen und assoziativen) Übergangszustände.

Im Falle eines dissoziativen Übergangszustands (elektrophile oder $\text{S}_{\text{N}}1$ -Reaktion) könnte das mit dem Pyrophosphat koordinierte Magnesium-Ion das kurzzeitige Auftreten eines Carbokations an dem C1-Atom von FPP stabilisieren (Long *et al.*, 2000); die Rolle des Magnesium-Ions ist aber derzeit noch nicht abschließend geklärt (Saderholm *et al.*, 2000). In der Literatur wird auch ein assoziativer Übergangszustand (nukleophile oder $\text{S}_{\text{N}}2$ -Reaktion) postuliert, bei dem eine partiell positive Ladung am C1 von FPP und partiell negative Ladungen am Schwefelatom des Thiols und an dem zu C1 benachbarten Sauerstoffatom des Pyrophosphats vorliegen (Wu *et al.*, 1999; Huang *et al.*, 2000). In der Kristallstruktur des ternären Komplexes aus FTase, FPP und CAAX-Substrat (PDB-Code 1QBQ) wurde zwischen dem C1-Atom von FPP und dem Schwefelatom des Cysteins ein Abstand von mehr als 8 Å festgestellt. Daher ist davon auszugehen, dass diese für eine Reaktion zu große Distanz zunächst verringert werden muss. Aus der Kristallstruktur von FTase mit einem bereits farnesylierten Produkt konnte abgeleitet werden, dass sich im Laufe des Reaktionspfades der Pyrophosphatteil inklusive dem C1-Atom von FPP auf das Tetrapeptid zubewegt und somit die Reaktion in Gang setzt (Tamanoi & Sigman, 2001). Das neu gebildete, farnesylierte Produkt wird schließlich in einem geschwindigkeitsbestimmenden Schritt von einem neuen Molekül FPP aus dem Enzym verdrängt (Tschantz *et al.*, 1997).

3.2.2 Farnesyltransferase-Inhibitoren

Die ersten in der Literatur beschriebenen Farnesyltransferase-Inhibitoren (FTI) basierten auf den Substraten der FTase-Reaktion, also auf dem CAAX-Tetrapeptid-Bindungsmotiv und der FPP-Struktur. Auch die Kombination beider Substrate als sogenannte Bisubstrat-Analoga zur Entwicklung von Hemmstoffen ist möglich. Alle drei Ansatzpunkte – Entwicklung von CAAX-Peptidomimetika, FPP-Analoga und Bisubstrat-Analoga – wurden verfolgt und führten zu hochaffinen FTIs, die sich teilweise schon in der klinischen Phase II und III befinden (Abb. 3-9).

Völlig neue Leitstrukturen von Inhibitoren der FTase wurden durch ein *Screening* von Naturstoffen (Leonard, 1997) und chemischen Substanzbibliotheken (Perola *et al.*, 2000) gefunden, die nach ihrer Optimierung und Weiterentwicklung ebenfalls vielversprechende Resultate zeigen.

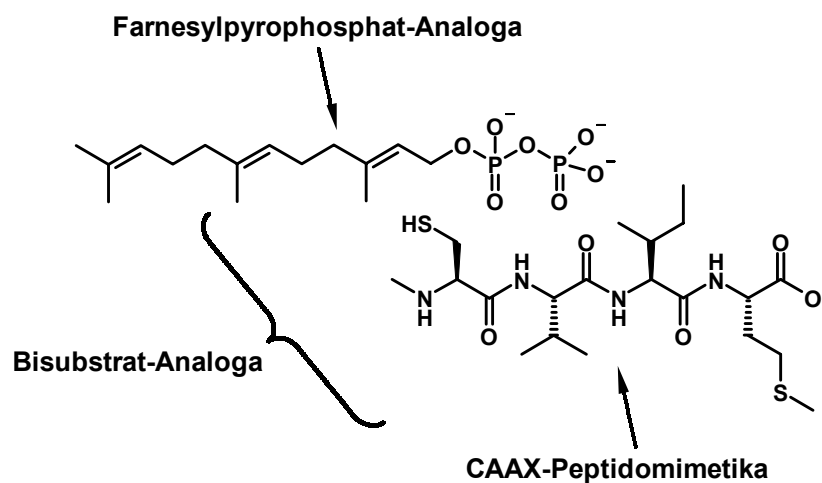


Abb. 3-9: Mögliche Ansatzpunkte zur Entwicklung von Hemmstoffen der Farnesyltransferase.

Im folgenden soll an ausgewählten Beispielen die Entwicklung von FTIs aufgezeigt werden. Einen umfassenden Überblick über entwickelte Hemmstoffe und ihren Stand in der klinischen Phase geben folgende Übersichtsartikel (Williams, 1998; Williams, 1999; End, 1999; Bell, 2000; Rowinsky & Patnaik, 2000; Sebti & Hamilton, 2000).

Relativ frühzeitig wurde erkannt, dass schon leicht abgewandelte CA₁A₂X-Tetrapeptide von der FTase zwar gebunden, aber nicht mehr umgesetzt werden können. So wurde zum Beispiel aus dem Substrat-Tetrapeptid CVIM (Cystein, Valin, Isoleucin, Methionin) durch Austausch der aliphatischen Aminosäure Isoleucin gegen Phenylalanin (aromatischer Rest) der Hemmstoff CVFM mit einem IC₅₀-Wert von 27 nM entwickelt. Die Nachteile dieser noch peptidischen Inhibitoren – Empfindlichkeit gegenüber enzymatischem Abbau und fehlende Membrangängigkeit – konnten mittels Ersatz der amidischen Struktureinheiten durch ihre entsprechenden Amine („reduzierte Amide“) oder ähnliche bioisostere Gruppen teilweise behoben werden. Einige Beispiele von FTIs, die durch Substitution der aliphatischen Aminosäuren A₁ und A₂ erhalten wurden, sind in Abb. 3-10 gezeigt (Schlitzer, 1998).

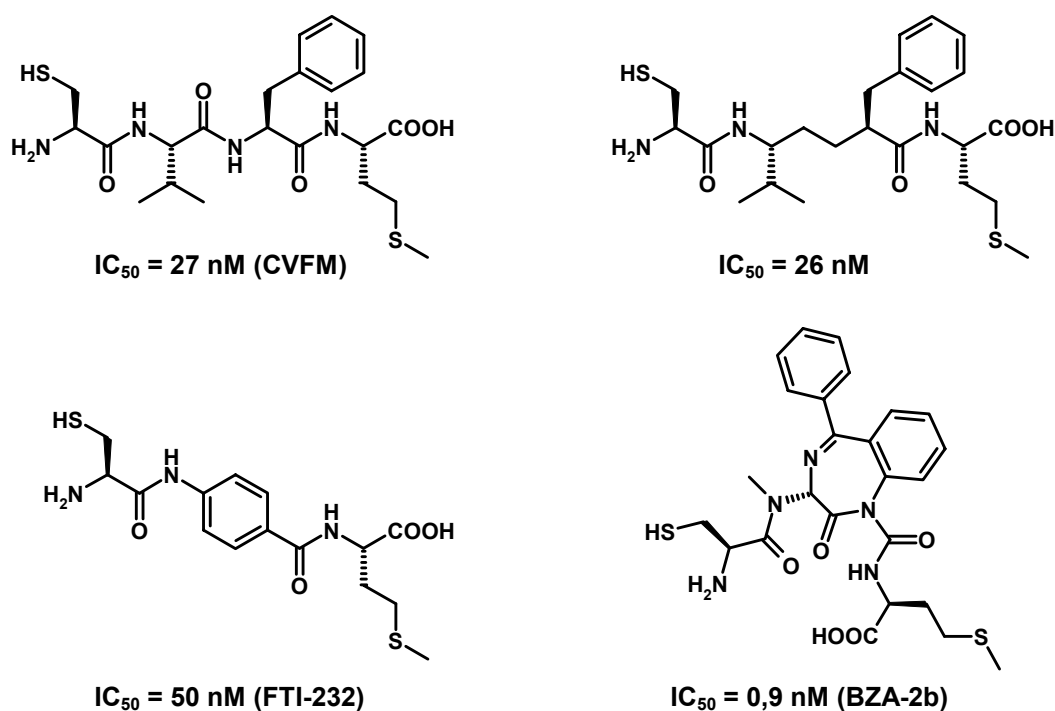


Abb. 3-10: CAAX-Peptidomimetika, die durch den Ersatz des zentralen A₁A₂-Peptids erhalten wurden.

Weitere Modifizierungen an der peptidischen CAAX-Struktur wurden durch den Ersatz der C-terminalen Aminosäure „X“ mit isosteren Struktureinheiten erreicht (Abb. 3-11).

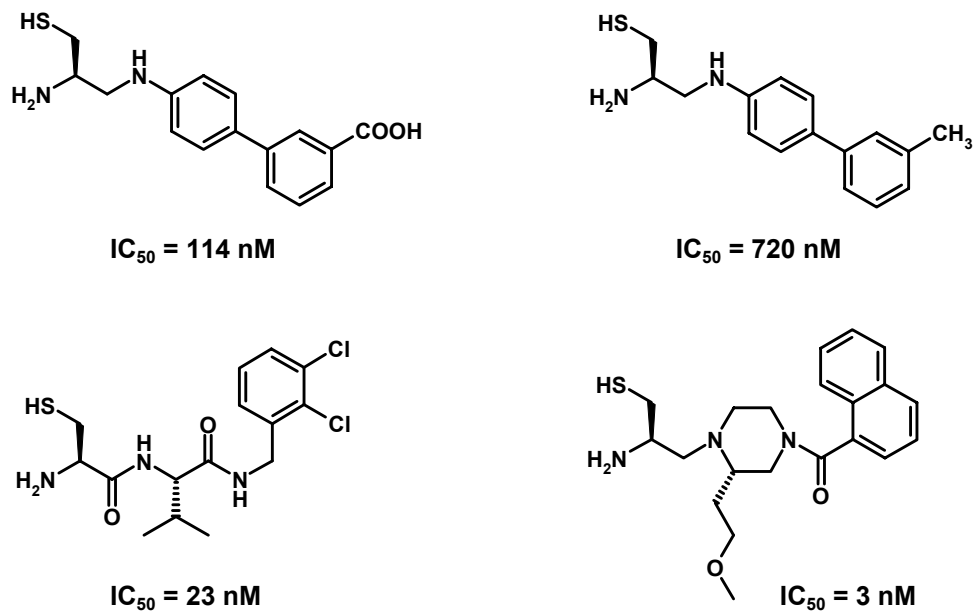


Abb. 3-11: CAAX-Peptidomimetika, deren C-terminale Aminosäure an der „X“-Position durch andere Gruppen ersetzt (erste Reihe) oder vollständig entfernt wurde (zweite Reihe).

So wurde zum Beispiel die Aminosäure Methionin durch eine *meta*-ständige Benzoesäure ersetzt ($IC_{50} = 114 \text{ nM}$). Auch die Carboxylatgruppe, die als funktionelle Gruppe von der Aminosäure zunächst noch übernommen wurde, konnte durch eine Methylgruppe ersetzt werden, was allerdings einen gewissen Wirkungsverlust ($IC_{50} = 720 \text{ nM}$) zur Folge hatte. Sogar die ersatzlose Entfernung des strukturellen Äquivalents der endständigen Aminosäure „X“ führte zu teilweise hochpotenten, peptidomimetischen FTIs mit IC_{50} -Werten unter 10 nM (Schlitzer, 1998).

Ein entscheidender Schritt in der Entwicklung neuer FTIs lag in dem erfolgreichen Ersatz der SH-Funktion des Cysteins, welche bis jetzt bei allen CAAX-Peptidomimetika als gemeinsames Strukturmerkmal vertreten war. Freie Thiolgruppen sind in Wirkstoffen meist unerwünscht, da zum einen die Mercaptogruppe oxidationsempfindlich ist und zum anderen von thiolhaltigen ACE-Hemmern bekannt ist, dass bei dauerhafter Anwendung stark beeinträchtigende Nebenwirkungen⁵⁰ auftreten können (Forth *et al.*, 2001). So wurden andere Zink-koordinierende Gruppen wie Imidazole, Lactame und Pyridine erfolgreich als alternative

⁵⁰ z.B. bei Captopril: Die vereinzelt auftretenden Hauterscheinungen (Exantheme; allergische Reaktion) und Geschmacksveränderungen konnten eindeutig auf die freie Thiolgruppe zurückgeführt werden. Ein genauer Mechanismus dieser Schädigungen ist jedoch nicht bekannt.

Bindungselemente für die Thiolgruppe des Cysteins eingesetzt. Auch der Ersatz dieser Fragmente durch Phenyl- oder Cyclohexylgruppen, die keine Zink-bindenden Gruppen aufweisen, führte zu affinen Inhibitoren. Abb. 3-12 zeigt einige ausgesuchte Beispiele von thiofreien, hochpotenten FTIs (Bell, 2000).

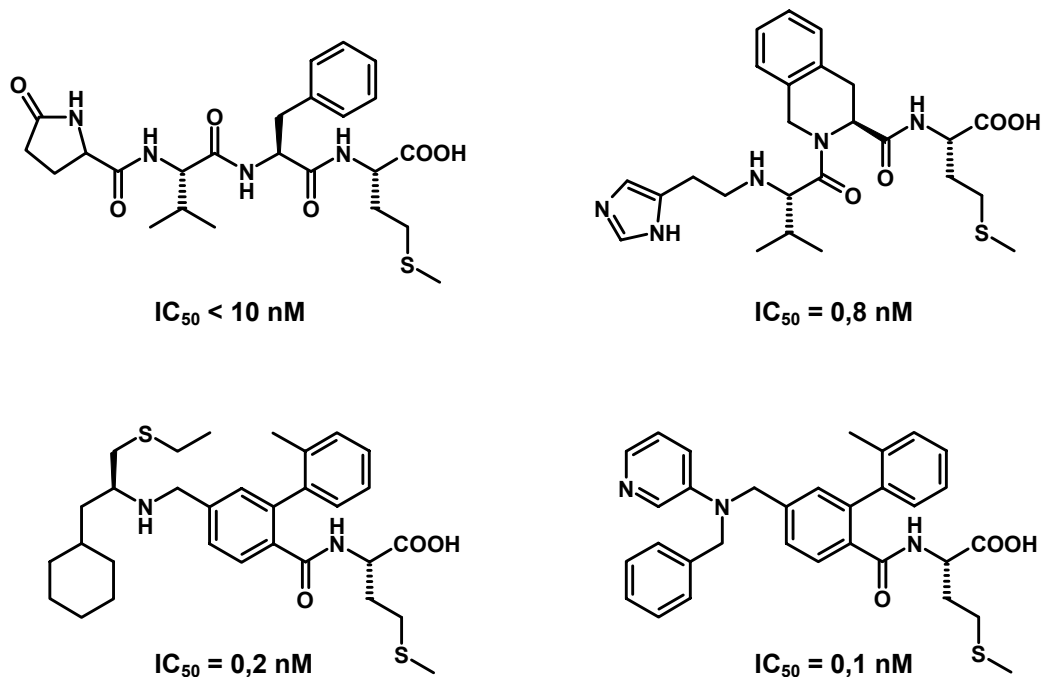


Abb. 3-12: Thiofreie CAAX-Peptidomimetika, bei denen die Aminosäure Cystein durch bioisostere (Lactam, Imidazol, Pyridin) oder andere funktionelle Gruppen ersetzt wurde.

Im weiteren wird auf eine Beschreibung von FPP-Analoga und Bisubstrat-Analoga (Schlitzer *et al.*, 2000a; Schlitzer *et al.*, 2000b) als Hemmstoffe der FTase verzichtet. Zum einen ist die Zahl an Verbindungen dieser Strukturklasse, verglichen mit CAAX-Peptidomimetika, weitaus geringer; außerdem leiten sich die im folgenden Abschnitt (Kap. 3.3) vorgestellten Inhibitoren ebenfalls von der CAAX-Grundstruktur ab. Für einen umfassenden Einblick über derzeit entwickelte FPP- und Bisubstrat-Analoga sei jedoch nochmals auf die Übersichtsartikel hingewiesen, die auf Seite 137 aufgeführt sind.

3.2.3 Wirkmechanismus von Farnesyltransferase-Inhibitoren

Obwohl die FTIs ursprünglich mit dem Ziel entwickelt wurden, die Farnesylierung von Ras zu unterdrücken und somit die transformierende Aktivität zu unterbinden, blockieren

Hemmstoffe der FTase auch die posttranslationale Modifikation anderer Proteine mit einer CAAX-Erkennungssequenz. Mittlerweile sind mehr als 20 solcher Proteine bekannt, die zunächst farnesyliert werden müssen, um ihre Funktionalität zu entfalten (Rowinsky & Patnaik, 2000); eine Vielzahl weiterer Proteine mit der CAAX-Sequenz sind bis jetzt noch nicht charakterisiert. Daneben ist die Farnesylierung normaler, nicht-mutierter Ras-Proteine genauso wichtig für die Lebensfähigkeit einer Zelle. Aus diesem Grund wurde von einer allgemeinen Zytotoxizität von FTIs ausgegangen, was jedoch im Experiment nicht bestätigt werden konnte. In Zellkulturen hemmten die Wirkstoffe das Wachstum von transformierten Zellen weit stärker als das normaler Zellen; auch die Behandlung von Tumoren in Tieren konnte mit relativ geringen Nebenwirkungen über mehrere Wochen durchgeführt werden. Mehrere Erklärungen für diese geringe Toxizität sind möglich: zum einen ist denkbar, dass einige geranylgeranylierte Proteine den Verlust an Ras-Aktivität kompensieren können. Gleichfalls könnte ein alternativer Signalweg aktiviert werden, wenn Ras nicht an der Zellmembran verankert ist. Die Inhibierung der FTase könnte die Ras-Aktivität auch nur unter den für die Transformation benötigten Grad herabsetzen, aber dennoch ausreichende Aktivität für eine normale Lebensfähigkeit der Zellen ermöglichen (Wittinghofer & Waldmann, 2000).

Die intensive Erforschung vieler FTIs zeigte, dass diese Hemmstoffe in Zellkulturen eine antiproliferative Aktivität gegenüber einer Vielzahl von transformierten Zelllinien und Tumoren in Tieren haben. Darunter befinden sich auch solche, die keine mutierten Ras-Proteine aufweisen. Es ist daher nicht anzunehmen, dass die mit einer Blockierung der Farnesylierung einhergehenden Effekte nur mit einer ausbleibenden Ras-Aktivität in Zusammenhang gebracht werden können; vielmehr erscheint es wahrscheinlicher, dass ein oder mehrere mit Ras nicht identische Proteine für die vielfältigen Wirkungen von FTIs verantwortlich sind. Daher wurde in der Literatur verstärkt darüber diskutiert, welche anderen Proteine neben Ras als Ziele von Hemmstoffen der FTase in Frage kommen können; eine intensive Suche nach dem „Target X“ wurde betrieben (Cox & Der, 1997). Eine interessante Klasse von Proteinen bildet dabei die Mitglieder der Rho-Proteinfamilie, die für die Transformation durch Ras benötigt werden. Insbesondere RhoB, ein durch Wachstumsfaktorsignale aktiviertes Protein, erscheint ein attraktives Ziel für FTIs zu sein (Lebowitz *et al.*, 1997a). Aber auch andere Mitglieder der Rho-Familie, beispielsweise Rnd3, Rho8, RhoD und RhoE werden farnesyliert und sind somit potenzielle Zielstrukturen. Mittlerweile sind weitere, Ras-ähnliche Strukturen wie zum Beispiel Rheb gefunden worden,

die für die Modulation der Ras-Aktivität verantwortlich sind. Ein weiteres Protein – Rap2 – wird ebenfalls farnesyliert, seine Funktion ist allerdings noch unbekannt (Farrell *et al.*, 1993).

Neben der Suche nach weiteren, möglichen Zielstrukturen wurde die Wirkung von FTIs auf zellulärer Ebene untersucht. Hier zeigen sich im wesentlichen vier Effekte: (1.) Zum einen inhibieren Wirkstoffe der FTase das verankerungsunabhängige Wachstum einer Reihe verschiedener Tumorzellen. Die besondere Fähigkeit dieser Zellen, in weichem Agar zu wachsen, wird von FTIs gehemmt; dies wurde zuerst bei *H-Ras* transformierten Zellen beobachtet. (2.) Weiterhin verursachen Hemmstoffe der FTase Veränderungen des Zellzyklus, je nach untersuchter Zelllinie vor allem zum G1/S- oder G2/M-Übergang. Bei einigen Zelllinien wurde eine Akkumulation von Zellen der G0/G1-Phase induziert, während parallel dazu die Zahl an Zellen der S-Phase abnahm. (3.) Ein ebenfalls interessanter Befund bei Verwendung von FTIs ist eine beobachtete Tumorregression von Brust- und Speicheldrüsenkarzinomen. Der Mechanismus der Tumorverkleinerung war zunächst unklar; aktuelle Studien zeigen, dass Inhibitoren der FTase die Apoptose in Zellkulturen auslösen können (Lebowitz *et al.*, 1997b). Dieser „programmierte Selbstmord“ der Zelle verläuft dabei unabhängig von dem Tumorsuppressor-Protein p53, einem wichtigen Vermittler der Apoptose. Da Resistenzen gegenüber üblichen Chemotherapeutika häufig durch Mutationen an p53 entstehen, kann diese Unabhängigkeit auch relevant für die klinische Anwendung sein, wenn FTIs trotz mutiertem p53 wirksam gegen Tumoren bleiben. (4.) Schließlich ist der Einfluss von Hemmstoffen der FTase auf das Zytoskelett der Zelle und ihre Zellmorphologie von Bedeutung. Diese Wirkungen auf Mikrotubuli, Mittelfilamente und Aktinfilamente sind allerdings noch nicht vollständig aufgeklärt (Prendergast *et al.*, 1994).

3.3 Rationales Design von Farnesyltransferase-Inhibitoren

Auch wenn der exakte Wirkmechanismus von FTIs noch viele Fragen aufwirft, stellen die zahlreichen positiven Befunde an Zellkulturen und Tiermodellen – wie niedrige Allgemeintoxizität und beobachtete Tumorregressionen – die Hemmung der FTase als derzeit vielversprechenden Ansatzpunkt für die Entwicklung neuartiger Chemotherapeutika dar. Wie in Kap. 3.2.2 angesprochen, können Hemmstoffe der FTase prinzipiell auf drei verschiedenen Wegen entwickelt werden. Zum einen lassen sich durch Abwandlung der CAAX-Sequenz

peptidomimetische Substratanaloga darstellen; diese Möglichkeit wird in der Literatur am weitaus häufigsten angewendet. Daneben wurden FTIs auch als Mimetika des Farnesylpyrophosphats (FPP) und Analoga beider Substrate, des CAAX-Peptids und FPP's, als sogenannte Bisubstrat-Analoga entwickelt. Der im folgenden beschriebene Ansatz geht von einem rationalen Design von CAAX-mimetischen Inhibitoren der FTase aus.

3.3.1 Benzophenon-basierte Farnesyltransferase-Inhibitoren

Ausgehend von der Struktur des CAAX-Tetrapeptids (CVIM) konnten im Laufe der Entwicklung peptidomimetische FTIs dargestellt werden, die auf einer Benzophenon-Grundstruktur beruhen. Die sukzessive Abwandlung des Substrats zu dieser neuen Klasse von FTase-Hemmstoffen ist in Abb. 3-13 gezeigt.

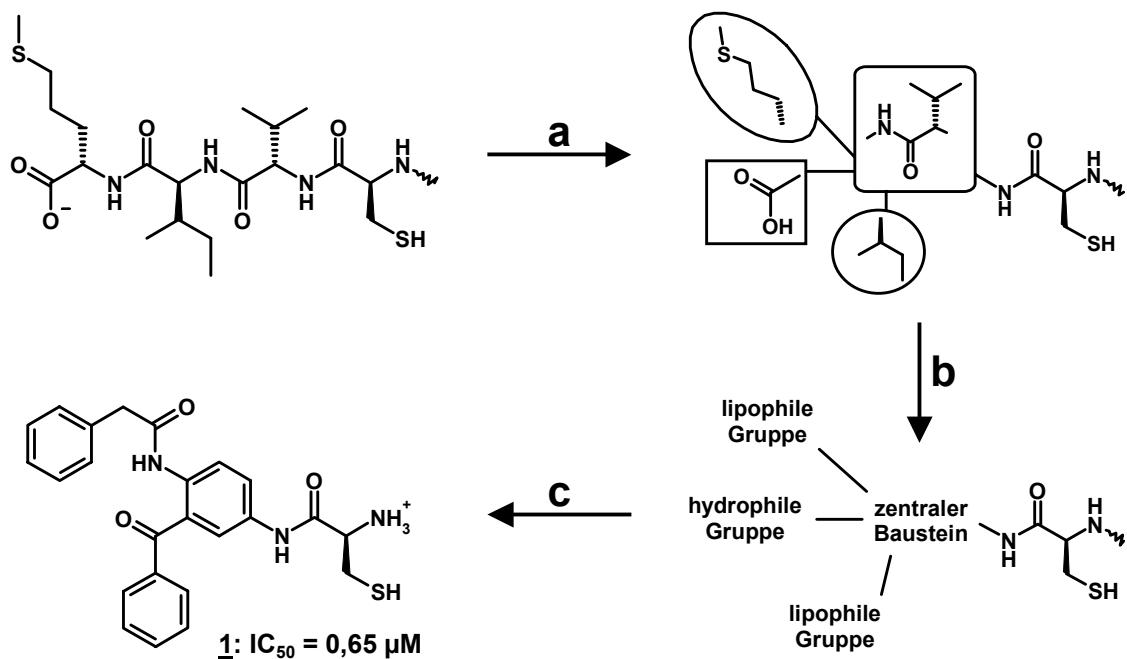


Abb. 3-13: Abwandlung des CAAX-Tetrapeptids zu der neuen Klasse von FTIs mit einer Benzophenon-Grundstruktur. Das CVIM-Substrat wurde zunächst in seine funktionellen Einheiten zerlegt (a). Aus der Analyse dieser Strukturfragmente resultierte ein Pharmakophormodell mit einem zentralen Baustein, zwei lipophilen Gruppen und einem hydrophilen Bereich (b). Durch die Besetzung der jeweiligen Bereiche mit geeigneten Fragmenten, die diese Eigenschaften widerspiegeln, resultierte eine neue Klasse von Hemmstoffen (c).

Die neu abgeleitete Hemmstoff-Klasse weist eine Benzophenon-Grundstruktur als zentralen Baustein auf; gleichzeitig wird durch den zweiten Phenylring die Isoleucin-Seitenkette des ursprünglichen Substrats ersetzt. Der zweite lipophile Bereich wird durch eine Benzylgruppe repräsentiert. Die zwischen dem Benzophenon und der Benzylgruppe befindliche Carbonylgruppe der Amideinheit soll hingegen den hydrophilen Bereich repräsentieren. Der in Abb. 3-13 dargestellte Benzophenon-Inhibitor **1** weist einen IC_{50} -Wert von $0,65 \mu\text{M}$ auf. Weitere Modifizierungen der Benzophenon-Grundstruktur führten zu verbesserten Hemmstoffen mit inhibitorischen Konstanten im nanomolaren Bereich (Sakowski *et al.*, 2001). Abb. 3-14 zeigt einige repräsentative Inhibitoren (**2–5**), die im Bereich der Benzylamidgruppe der Grundstruktur abgeändert wurden.

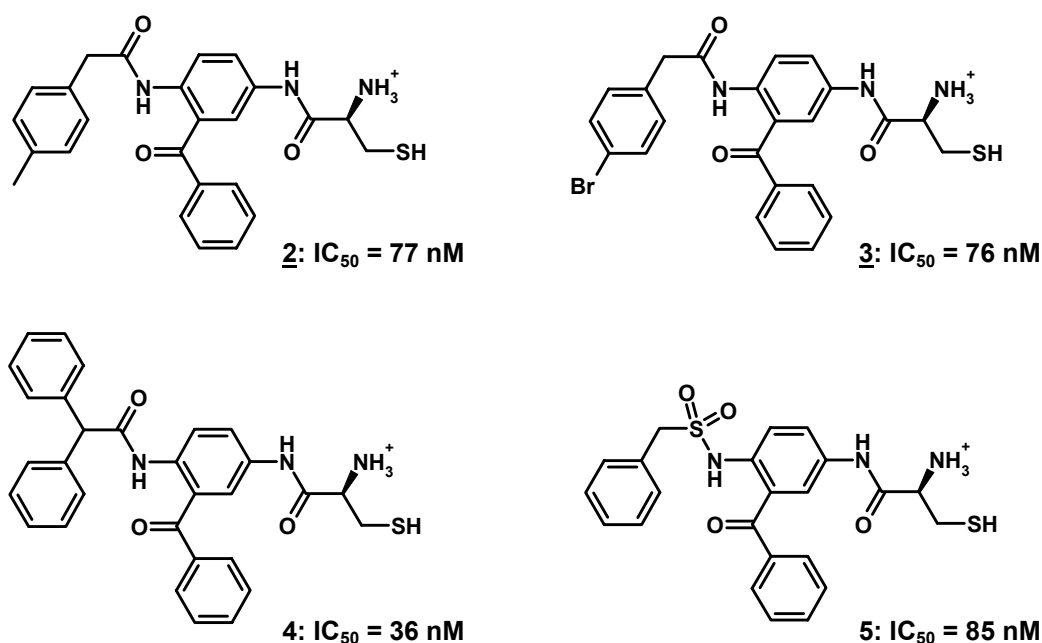


Abb. 3-14: Modifizierte Strukturen der Benzophenon-Grundstruktur mit verbesserten IC_{50} -Werten.

Eine flexible Überlagerung von **2** auf das CAAX-Tetrapeptid mit dem Programm SEAL⁵¹ (Klebe *et al.*, 1994b) zeigte, dass die Strukturfragmente des Pharmakophormodells mit gleichen physikochemischen Eigenschaften gut miteinander zur Deckung gebracht werden konnten (Abb. 3-15).

⁵¹ Die verwendeten Programme sind in Kap. 4.2 nochmals kurz aufgeführt.

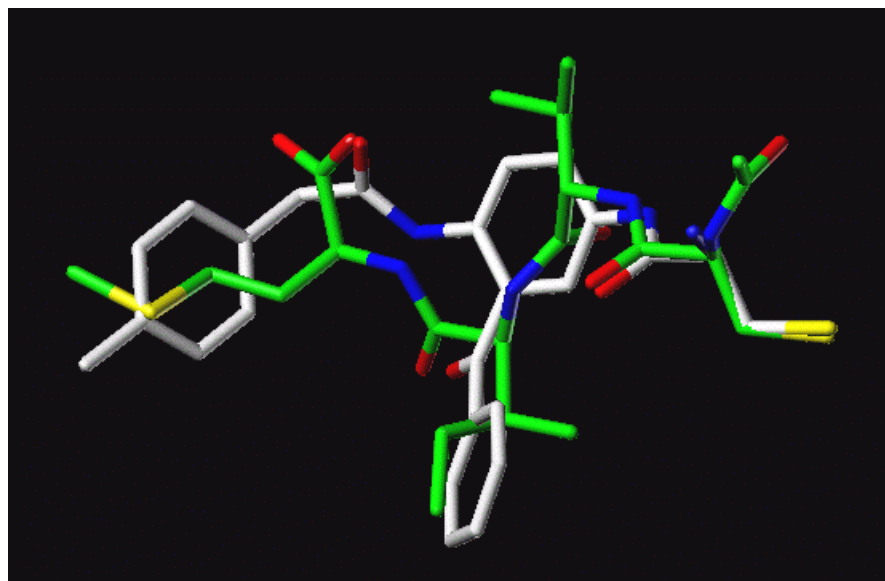


Abb. 3-15: Flexible Überlagerung des Inhibitors **2** (C-Atome weiß) auf die Struktur des CAAX-Peptids CVIM (C-Atome grün). Die Konformation der CVIM-Referenzstruktur wurde aus der Kristallstruktur des ternären Komplexes aus FTase, CVIM und einem FPP-Analogon übernommen (PDB-Code 1QBQ). Die flexible Überlagerung wurde mit dem Programm SEAL durchgeführt.

Wie in Abb. 3-15 zu erkennen, wurde der terminale Phenylring des Benzophenons auf die Seitenkette des Isoleucins überlagert (lipophiler Bereich). Die *para*-Tolylgruppe wurde in dem Bereich positioniert, der von der Methionin-Seitenkette besetzt ist (lipophiler Bereich). Die Carbonylgruppe der amidischen Verbindung der Benzyl- und Benzophenoneinheit konnte räumlich mit der terminalen Carboxylatgruppe des Tetrapeptids überlagert werden (hydrophiler Bereich).

3.3.2 Thiofreie Farnesyltransferase-Inhibitoren

Wie in Kap. 3.2.2 schon erläutert, sind freie Thiolgruppen in Wirkstoffen meist nicht erwünscht, da sie zum einen stark oxidationsempfindlich sind und andererseits ernsthafte Nebenwirkungen verursachen können. Der Ersatz der Mercaptogruppe durch ebenfalls Zinkkoordinierende Gruppen wie Imidazole, Lactame und Pyridine führte dabei in vielen Fällen zu weiterhin potenten Hemmstoffen im nanomolaren Bereich. Daher wurde bei den folgenden

Modifikationen der Benzophenon-Grundstruktur ein Ersatz der Cysteinfunktion angestrebt. (Abb. 3-16).

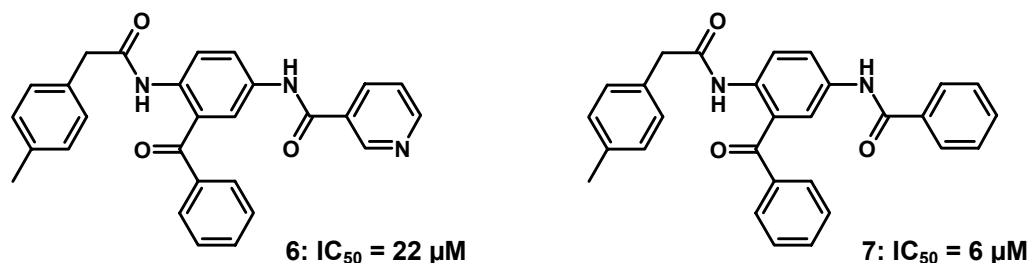


Abb. 3-16: Ein Ersatz der Cysteingruppe der Benzophenon-Struktur **2** durch thioolfreie Gruppen (Pyridin-, Phenylgruppe) resultiert in einem Wirkungsverlust der Inhibitoren (**6**, **7**).

Wie mittlerweile in der Literatur allgemein bekannt ist, muss die Thiolgruppe des Cysteins nicht zwingend durch Heterozyklen ersetzt werden. Auch arylische Substituenten, die keine Zink-koordinierenden funktionellen Gruppen aufweisen, lassen sich scheinbar als Ersatz für die Thiolfunktion verwenden; zumindest ist kein merklicher Wirkungsverlust festzustellen (Augeri *et al.*, 1998; Augeri *et al.*, 1999; O'Connor *et al.*, 1999). Wie aus Abb. 3-16 zu ersehen ist, verbessert sich sogar der IC₅₀-Wert bei den Benzophenon-Inhibitoren, wenn anstelle der Pyridin- eine Phenylgruppe eingesetzt wird. Daher wurde in der Literatur die Existenz von mindestens einer bis dahin unbekanntes Arylbindestelle postuliert (Breslin *et al.*, 1998; Ciccarone *et al.*, 1999). Bis jetzt gibt es jedoch noch keine schlüssigen Hinweise darauf, wo sich diese Taschen im aktiven Zentrum der FTase befinden könnten.

3.3.3 Flexible Docking-Studien und GRID-Analysen

Um diese Bindestellen näher eingrenzen zu können, wurden sowohl thioolfreie Inhibitoren aus der Literatur (Breslin *et al.*, 1998; Anthony *et al.*, 1999; Augeri *et al.*, 1999; O'Connor *et al.*, 1999) als auch ausgewählte Benzophenon-Strukturen (Schlitzer *et al.*, 1999; Sakowski *et al.*, 2001) flexibel in die Bindetasche der FTase gedockt (Böhm *et al.*, 2001). Abb. 3-17 und Abb. 3-18 zeigen die Strukturformeln ausgewählter thioolfreier FTIs, die mit dem Docking-Programm FlexX (Rarey *et al.*, 1996) untersucht wurden. Dabei kamen Inhibitoren zum Einsatz (Abb. 3-17), deren Arylgruppen kurzketzig zum zentralen Bereich des Liganden verknüpft sind („kurzkettige Inhibitoren“). Zusätzlich wurden auch Verbindungen

herangezogen (Abb. 3-18), deren endständige Arylgruppen längerkettig mit dem Rest des Liganden verbunden sind („langkettige Inhibitoren“).

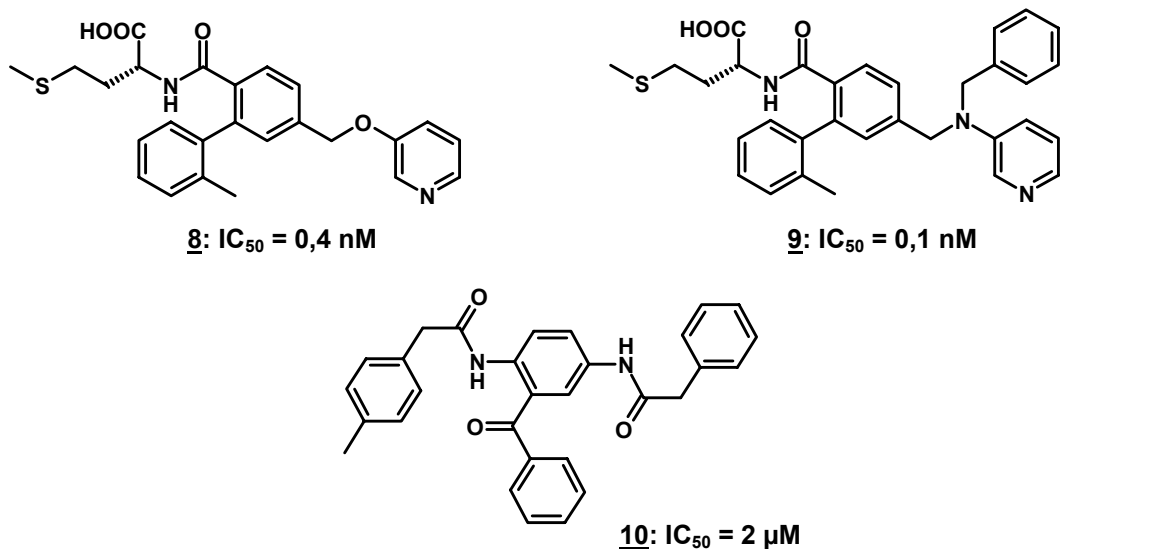


Abb. 3-17: Literaturbekannte, kurz-kettige Inhibitoren (8, 9) und eine Benzophenon-Struktur (10), die zum flexiblen Docking in die Bindetasche der FTase verwendet wurden.

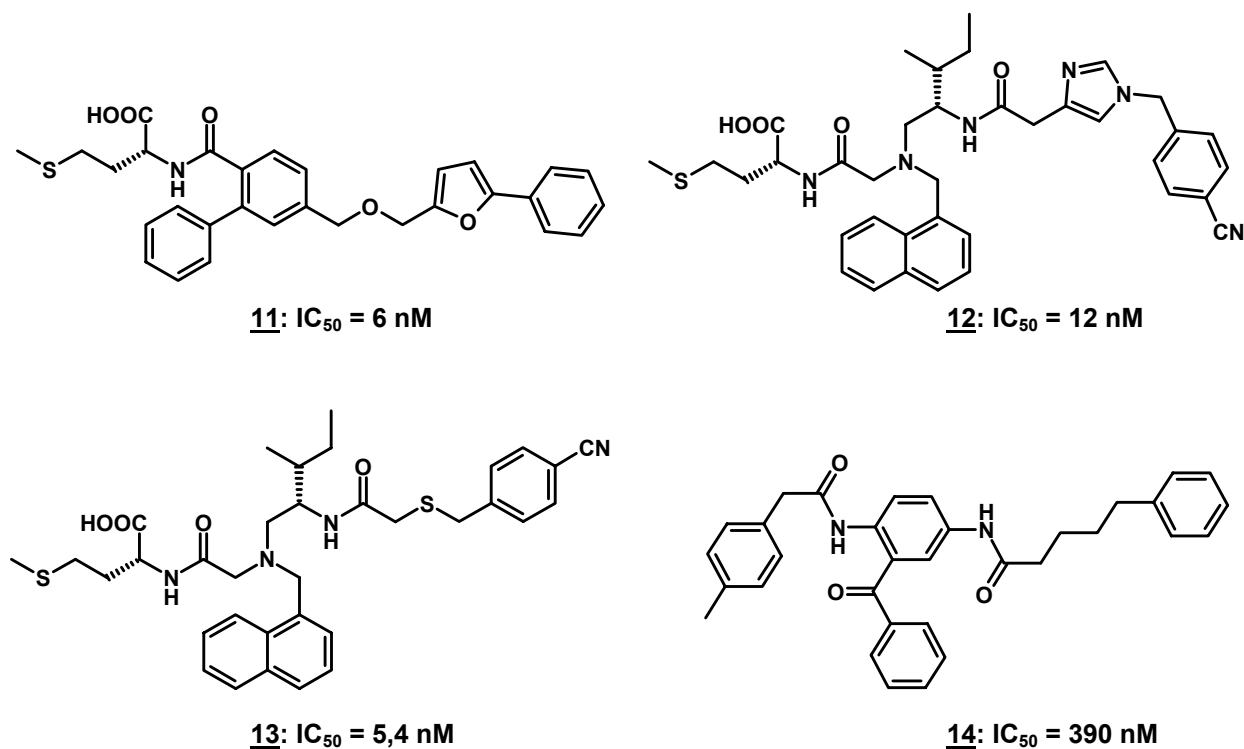


Abb. 3-18: Literaturbekannte, langkettige Inhibitoren (11-13) und ein Benzophenon-Inhibitor (14), die zum flexiblen Docking in die Bindetasche der FTase verwendet wurden.

Für die Docking-Studien der FTIs wurde die Kristallstruktur des ternären Komplexes – bestehend aus FTase, einem FPP-Analogen und *N*-acetyl-Cys-Val-Ile-selenoMetOH (PDB-Code 1QBQ) – verwendet. Die Koordinaten des CAAX-Peptids sowie aller Wassermoleküle wurden aus der Kristallstruktur entfernt. Da die literaturbekannten Inhibitoren und die Benzophenon-Strukturen jeweils als CAAX-Peptidomimetika entwickelt worden sind, wurden neben dem Zink-Ion auch die Atompositionen des FPP-Analogons in die Proteinstruktur der FTase miteinbezogen. Für das flexible Docking der ausgewählten Literaturverbindungen **8**, **9**, **11**, **12** und **13** fand als Basisfragment die räumliche Position des Methionin-Fragments HOOC-CH_α-NH der CAAX-Kristallstruktur im aktiven Zentrum der FTase Verwendung. Im Falle der Benzophenon-Strukturen **10** und **14** wurde das zum Methionin als isoster angenommene Toly-OC-NH-Fragment als Kopfgruppe herangezogen; die Koordinaten dieses Fragments konnten aus der flexiblen Überlagerung mit dem CVIM-Tetrapeptid (vgl. Abb. 3-15) erhalten werden.

Bei der anschließenden sukzessiven Platzierung der verbleibenden Ligandfragmente in das aktive Zentrum der FTase sucht das Docking-Programm günstige Wechselwirkungspartner in der Bindetasche, während gleichzeitig sterische Überlappungen mit der Proteinoberfläche vermieden werden. Die von FlexX gefundenen 100 besten Lösungen wurden – aufsteigend sortiert nach ihren mit der regressionsbasierten Bewertungsfunktion in FlexX berechneten Bindungsenergien – zunächst graphisch angezeigt. Offensichtlich ungünstige Lösungen wie zum Beispiel Strukturen, die zu einem großen Teil in das Lösungsmittel ragten und somit keine spezifischen Wechselwirkungen mit dem aktiven Zentrum der FTase ausbilden konnten, fanden keine Berücksichtigung. Aus dem verbleibenden Cluster an gedockten Lösungen wurde anschließend jeweils eine repräsentative Struktur ausgewählt.

In Abb. 3-19 sind die repräsentativen Lösungen aus den Docking-Läufen mit den kurzkettigen Inhibitoren **8-10** dargestellt.

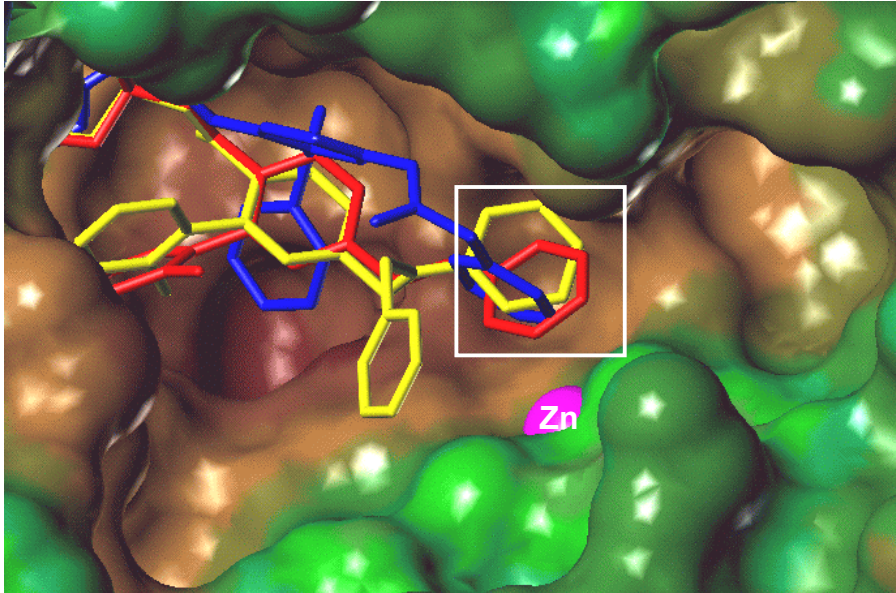


Abb. 3-19: Resultate der durchgeführten Docking-Läufe mit den kurzkettigen Inhibitoren **8** (rot), **9** (gelb) und **10** (blau) in das aktive Zentrum der FTase. Die terminalen Arylgruppen binden in eine lipophile Tasche (weiß umrandet) in der Nähe des Zink-Ions (magenta). Die lösungsmittelzugängliche Oberfläche ist zusammen mit den hydrophoben Eigenschaften des aktiven Zentrums dargestellt (braun: hydrophob; grün: hydrophil). Die Struktur von FPP wurde in die Connolly-Oberfläche integriert.

Aus den Docking-Läufen mit den langkettigen Inhibitoren **11-14** ergaben sich zwei separate Cluster, die zwei alternativen Vorschlägen für denkbare Bindungsmodi entsprechen und als repräsentative Lösungen in Abb. 3-20 (A und B) dargestellt sind.

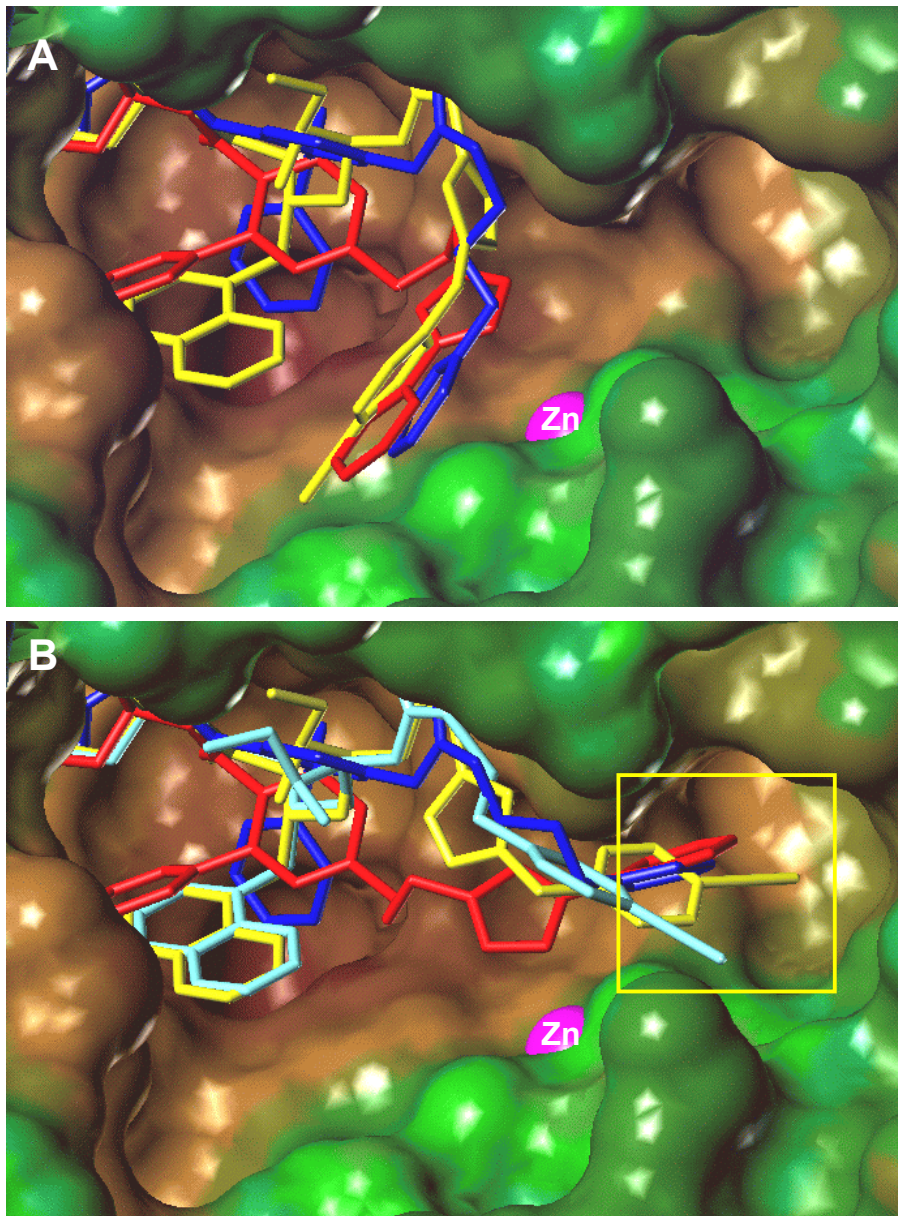


Abb. 3-20: Als denkbare Lösungen werden zwei alternative Bindungsmodi vorgeschlagen: Docking-Läufe mit den langkettigen Inhibitoren **11** (rot), **12** (gelb), **13** (cyan) und **14** (blau) in das aktive Zentrum der FTase. **A**: Die terminalen Arylgruppen binden in einem Bereich unterhalb des Zink-Ions (magenta). **B**: Die terminalen Arylgruppen binden in einer alternativen Tasche (gelb umrandet) weiter entfernt von dem Zink-Ion (magenta). **A+B**: Die lösungsmittelzugängliche Oberfläche ist zusammen mit den hydrophoben Eigenschaften des aktiven Zentrums dargestellt (braun: hydrophob; grün: hydrophil). Die Struktur von FPP wurde in die Connolly-Oberfläche integriert.

Die Docking-Lösungen der kurzkettigen Inhibitoren **8**, **9** und **10** exponieren in allen drei Fällen ihre terminalen Arylgruppen in einer lipophilen Bindetasche in der Nähe des Enzymgebundenen Zink-Ions (Abb. 3-19). Diese im weiteren als „nahe Bindetasche“ bezeichnete Region wird durch die Seitenketten von Tyr 300 β , Trp 303 β , Tyr 361 β und Tyr 365 β gebildet. Somit wurden von FlexX vor allem hydrophobe Wechselwirkungen mit den Arylsubstituenten (Pyridin-, Phenylgruppe) der Inhibitoren **8-10** vorgeschlagen.

Die Strukturen der langkettigen Verbindungen **11-14** hingegen sind offensichtlich sterisch zu anspruchsvoll, um in diesen Bereich des aktiven Zentrums zu binden (Abb. 3-20). Für diese Inhibitoren wurden von FlexX zwei räumlich separate Bereiche gefunden, die prinzipiell die Arylsubstituenten aufnehmen können. Eine dieser Regionen befindet sich im unteren Bereich des aktiven Zentrums der FTase (Abb. 3-20 A). Hier wurde eine hydrophobe Wechselwirkung mit dem Phenylring von Tyr 361 β angezeigt. In diese Region platzierte FlexX auch den N-Benzylsubstituenten von **9**. Es wurde jedoch keine Lösung für die Verbindung **13** gefunden, bei der deren endständige Arylgruppe in diesen Bereich positioniert wird.

Ein alternativer Bindungsbereich für die Arylsubstituenten der langkettigen Inhibitoren **11-14** ist in Abb. 3-20 B dargestellt. Diese als „ferne Bindetasche“ bezeichnete Region befindet sich entfernt vom aktiven Zentrum der FTase „östlich“ des Zink-Ions. Sie wird im wesentlichen von den Seitenketten der Aminosäuren Tyr 300 β , Leu 295 β und den drei Lysinseitenketten Lys 294 β , Lys 353 β und Lys 356 β definiert. Obwohl keine experimentelle Strukturbestimmung vorliegt, erscheint es plausibel, dass die zweite, ferne Bindetasche für eine Aufnahme der Arylsubstituenten der langkettigen Inhibitoren besser geeignet ist. So scheint diese Tasche ein besseres Komplement für aromatische Ringe darzustellen als der Arylbindebereich im unteren Bereich des aktiven Zentrums. Die relativ flache – wenn auch hydrophobe – molekulare Oberfläche ist stärker zum Lösungsmittel exponiert; eine Besetzung durch hydrophobe Gruppen sollte eine ungünstigere Bilanz für die Desolvatation bedeuten.

3.3.4 Analyse der postulierten Arylbindestellen

Um die im letzten Abschnitt postulierten Arylbindestellen näher identifizieren zu können, wurden neben den flexiblen Docking-Studien zusätzlich Wechselwirkungsfelder im aktiven Zentrum der FTase und seiner Umgebung berechnet und analysiert. Das dabei verwendete Programm GRID⁵² (GRID, 2001) basiert auf einem Kraftfeldansatz und konturiert Bereiche innerhalb einer Proteinbindetasche gemäß den Wechselwirkungsenergien einzelner Sonden (Goodford, 1985). Je nach Wahl einer eigenschaftsspezifischen Sonde können unterschiedliche Wechselwirkungsmuster in der Bindetasche hervorgehoben werden.

Um Bereiche des aktiven Zentrums ermitteln zu können, die günstig für eine Platzierung von Arylsubstituenten sind, wurde die GRID-Sonde „C1=“ (sp²-hybridisiertes, aromatisches CH-Fragment) verwendet. Von jedem Schnittpunkt eines dreidimensionalen Gitters aus, das weiträumig das aktive Zentrum der FTase umschließt⁵³, wurden die Wechselwirkungsenergien des Proteins zu der ausgewählten Sonde berechnet und die erzielten Energiewerte auf verschiedenen Niveaus konturiert. Abb. 3-21 zeigt die Regionen in der Bindetasche, die die Platzierung einer Arylgruppe favorisieren, zusammen mit den Seitenketten der Aminosäuren, die diese Taschen im wesentlichen definieren.

In Abb. 3-21 zeigt das Konturdiagramm des Wechselwirkungsfeldes der „C1=“-Sonde deutlich zwei separate Bereiche, die besonders günstig für die Platzierung von aromatischen CH-Gruppen sind. Bei einem Vergleich von Abb. 3-21 mit Abb. 3-19 bzw. Abb. 3-20 B wird ersichtlich, dass diese von GRID als günstig bestimmten Bereiche exakt mit den Positionen der Arylgruppen übereinstimmen, die mit FlexX in das aktive Zentrum der FTase gedockt wurden. Die sich gegenseitig bestätigenden Resultate zweier verschiedener Ansätze – flexibles Docking und Analyse der Bindetasche – unterstützen umso mehr die Existenz der hier postulierten (nahen und fernen) Arylbindestellen (Böhm *et al.*, 2001).

⁵² Die verwendeten Programme sind in Kap. 4.2 nochmals kurz aufgeführt.

⁵³ Dimensionen des Gitters: 22 Å x 30 Å x 20 Å; Gitterweite: 0,5 Å

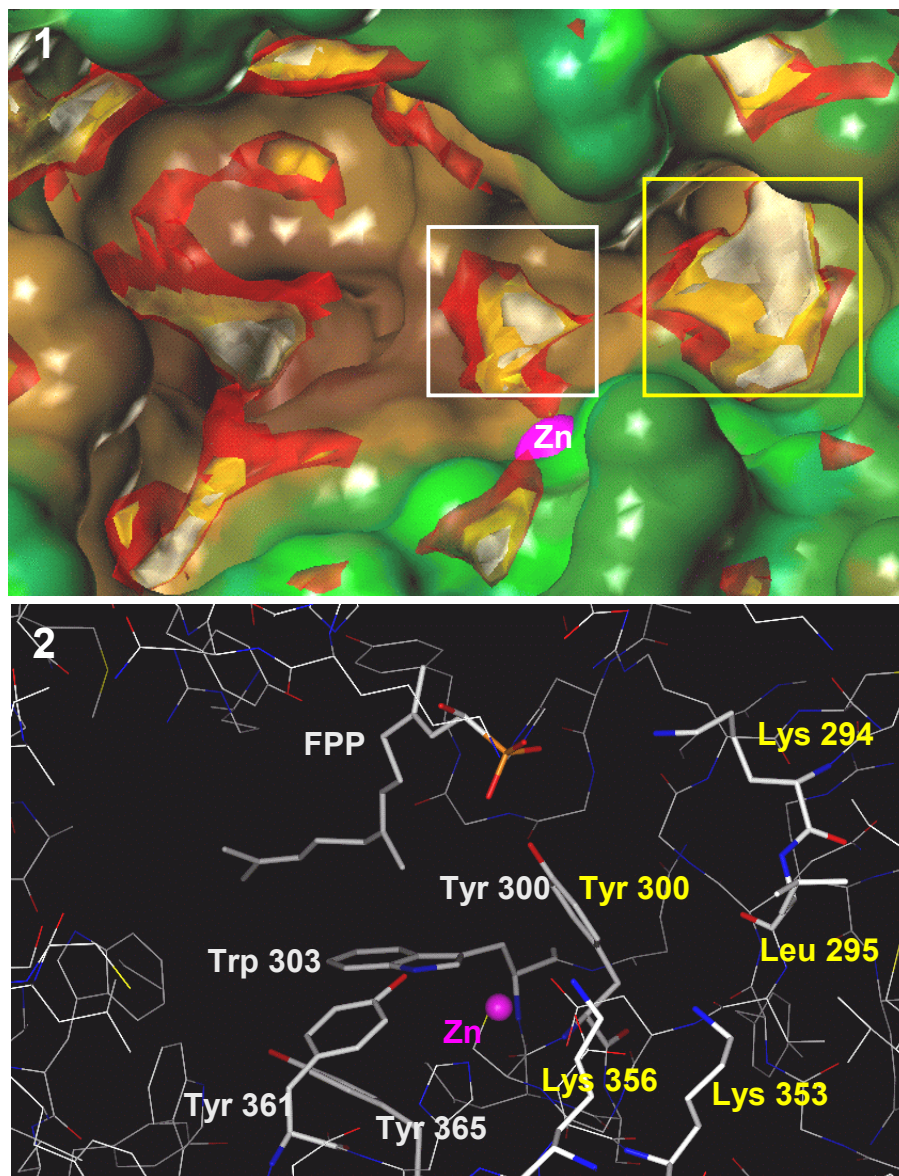


Abb. 3-21: Postulierung zweier neuer Arylbindestellen. **1:** Konturdiagramm der mit GRID berechneten Wechselwirkungsfelder mit dem Sondenatom „C1=“ (sp^2 CH aromatic). Die Felder sind bei drei Energieniveaus konturiert (rot: $-2,25$ kcal/mol; gelb: $-2,75$ kcal/mol; weiß: $-3,25$ kcal/mol). Die Regionen der postulierten nahen (weiß) und fernen (gelb) Arylbindestellen sind hervorgehoben. **2:** Darstellung der Aminosäuren, die die beiden postulierten Arylbindestellen definieren. Die Seitenketten zu der nahen (weiß) und fernen (gelb) Tasche sind angezeigt. Die aromatische Seitenkette von Tyr 300 β ist beiden Bindetaschen zugeordnet.

Einige langkettige Benzophenon-Inhibitoren, die Nitrosubstituenten an den terminalen Arylgruppen besitzen, zeigten eine auffallend gute Hemmaktivität. Im Gegensatz dazu fielen die IC_{50} -Werte vergleichbarer Inhibitoren mit einem terminalen Chlorsubstituenten deutlich schlechter aus (Abb. 3-22).

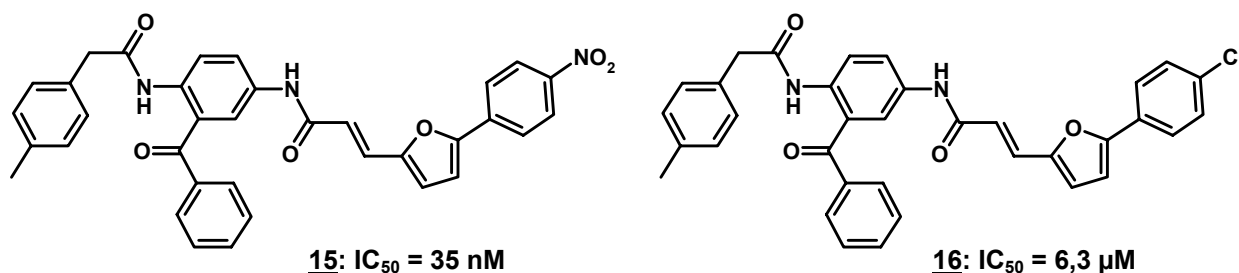


Abb. 3-22: Ausgewählte langkettige Benzophenon-Inhibitoren. Während ein Nitrosubstituent an der terminalen Arylgruppe die Affinität verbessern kann ($IC_{50} = 100 \text{ nM}$ ohne NO_2 -Gruppe), verschlechtert sich deutlich der Wert bei einer Chlorsubstitution.

Deshalb wurden weitere GRID-Wechselwirkungsfelder berechnet, die günstige Bereiche für Nitro- und Chlorgruppen im aktiven Zentrum der FTase bestimmen sollten. Dazu wurden die GRID-Sonden „ON“ (Nitro) und „CL“ (Chloro) herangezogen (Abb. 3-23).

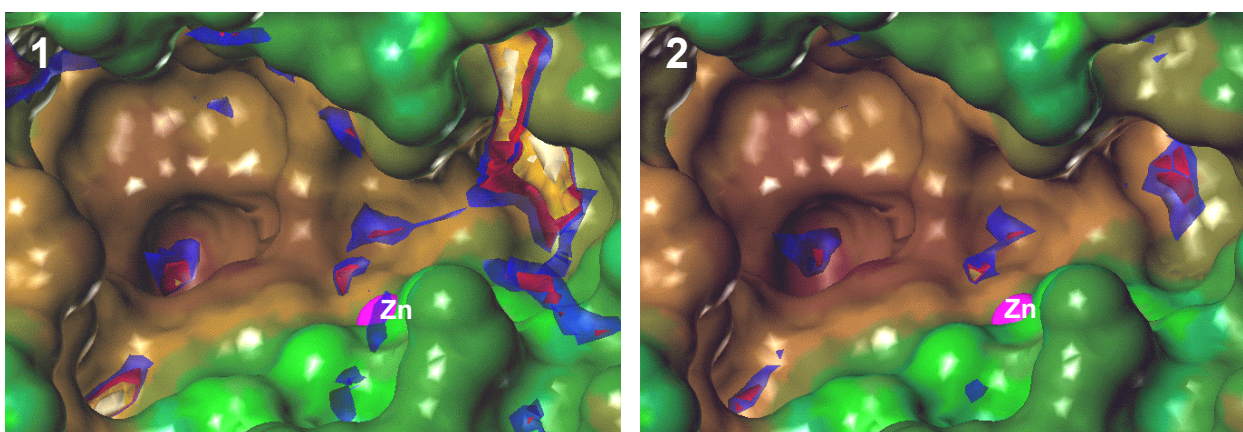


Abb. 3-23: Konturdiagramm der mit GRID berechneten Wechselwirkungsfelder mit den Sondenatomen „NO“ (1) und „CL“ (2). Die Felder sind jeweils bei drei gleichen Energieniveaus konturiert (blau: -5 kcal/mol ; rot: -6 kcal/mol ; gelb: -7 kcal/mol ; weiß: -8 kcal/mol). 1: Der Bereich der fernen Bindetasche wird als besonders günstig für die Positionierung von Nitrogruppen bestimmt. 2: Chlorsubstituenten werden in der fernen Bindetasche von GRID wenig favorisiert.

Wie aus Abb. 3-23 ersichtlich wird, können die unterschiedlichen IC_{50} -Werte der Verbindungen aus Abb. 3-22 mit den Befunden aus den GRID-Feldern in Einklang gebracht werden. Die ferne Arylbindestelle scheint für eine Aufnahme von Nitrogruppen besser geeignet zu sein als bei einer Besetzung mit Chlorsubstituenten. Die Analyse der Aminosäuren im Bereich der fernen Arylbindetasche zeigt mehrere Lysinseitenketten (Lys 164 α , Lys 294 β , Lys 353 β und Lys 356 β). Ein Grund für die vergleichsweise hohe Hemmaktivität von nitrosubstituierten, langkettigen Benzophenon-Inhibitoren könnte darin begründet liegen, dass diese partiell negativ geladenen Nitrogruppen günstige Wechselwirkungen mit den protonierten Aminofunktionen der (flexiblen) Lysinseitenketten eingehen (Abb. 3-24).

Um die Hypothese einer günstigen Wechselwirkung zwischen Nitrogruppen und Lysinseitenketten weiter zu untermauern, wurde mit dem Programm CavBase (Schmitt *et al.*, 2001) eine Suche nach Protein-Ligand-Komplexen mit einem vergleichbaren Wechselwirkungsmuster durchgeführt. Dabei werden zu einer vorgegebenen Referenztasche solche Proteine in der PDB aufgefunden, deren Bindetaschenbereiche ähnliche physikochemische Eigenschaften (aliphatisch, aromatisch, Donor, Akzeptor) mit strukturell vergleichbarer Geometrie aufweisen. Eine Suche nach ähnlichen Eigenschaftsmustern der fernen Arylbindetasche führte neben weiteren Lösungen zu zwei Proteinstrukturen (PDB-Codes 1QMH, 1B1C) mit Bindetaschenbereichen, die ähnliche physikochemische Eigenschaften zu der fernen Arylbindestelle aufweisen konnten (Abb. 3-25).

Die Bindetaschenbereiche der gefundenen Proteinstrukturen zeigen in beiden Fällen ausgedehnte Oberflächenregionen mit exponierten Donor-Eigenschaften, die von Aminosäuren mit basischen Seitenketten (Lysin, Arginin) erzeugt werden. Die in der Nähe gebundenen Liganden (Citrat, Flavinmononukleotid) weisen funktionelle Gruppen (Carboxylat-, Phosphatgruppen) mit dazu komplementären Akzeptor-Eigenschaften auf. Eine Übertragung dieser Konstellation der gefundenen Bindetaschen – Wechselwirkung von Akzeptorgruppen der Liganden mit basischen Aminosäuren – auf die strukturellen Verhältnisse der fernen Arylbindetasche unterstützt daher die Hypothese, dass günstige Wechselwirkungen der Nitrogruppe von **15** mit basischen Seitenketten möglich sind.

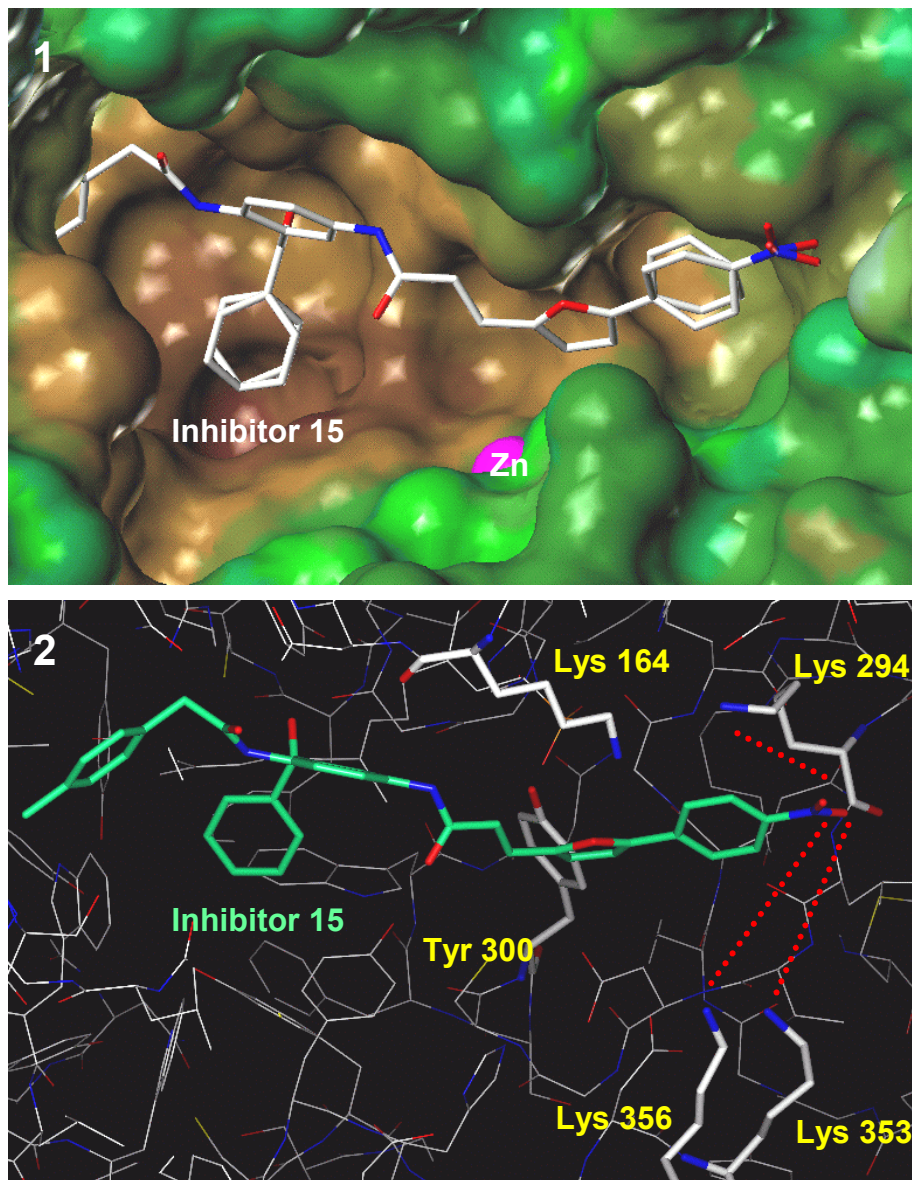


Abb. 3-24: Ergebnis der flexiblen Docking-Studie des Inhibitors **15** in das aktive Zentrum der FTase. **1**: Die beiden besten Lösungen mit annähernd gleichem Bindungsmodus sind zusammen mit der lösungsmittelzugänglichen Oberfläche gezeigt. Der Nitrophenylsubstituent wurde in die ferne Arylbindetasche platziert. **2**: Darstellung der flexiblen Lysinseitenketten (gelb) in der Nähe der fernen Arylbindestelle, die potenzielle Wechselwirkungen (rot) mit der Nitrogruppe des langkettigen Benzophenon-Inhibitors **15** (C-Atome grün) ausüben können. Zusätzlich ist die aromatische Seitenkette von Tyr 300 β markiert.

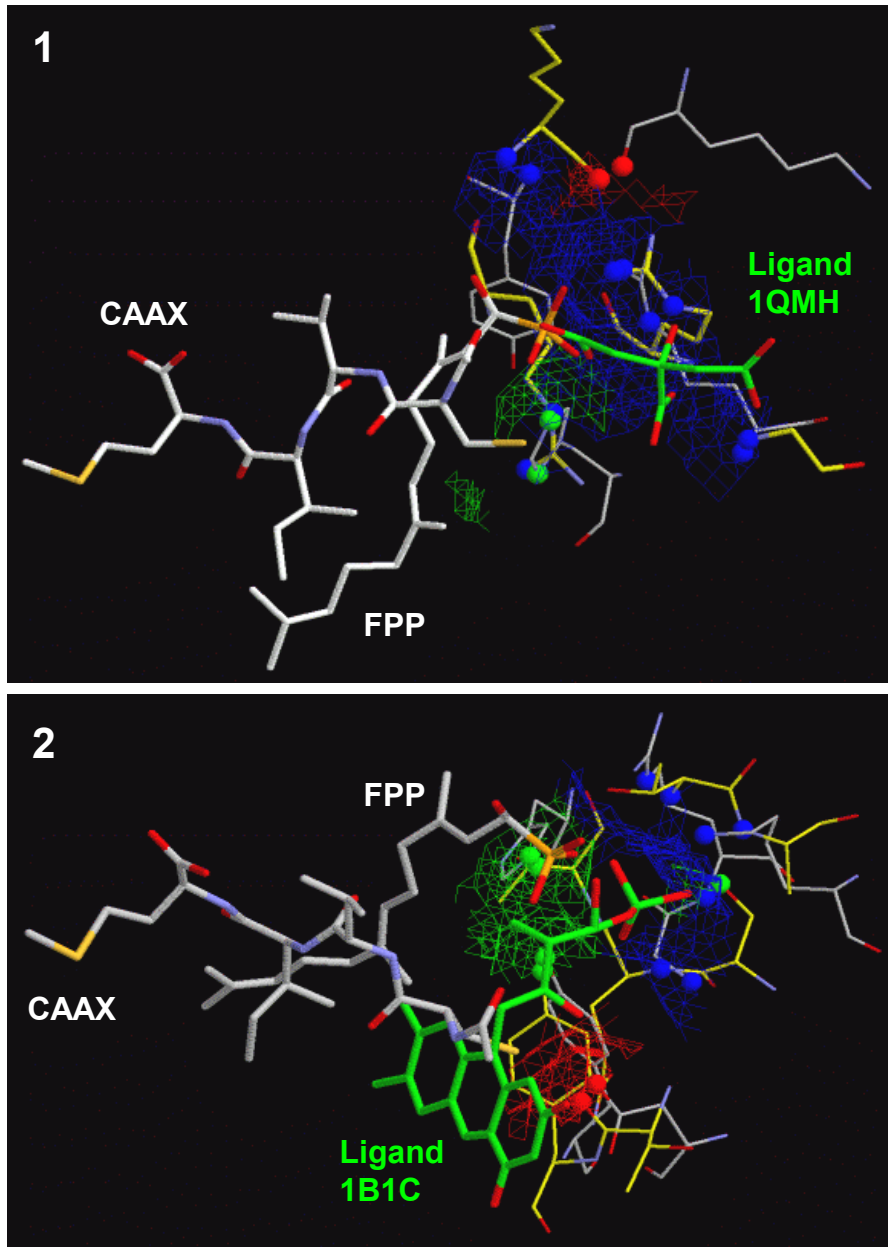


Abb. 3-25: Darstellung der Bindetaschenbereiche von zwei gefundenen Proteinstrukturen (PDB-Codes: 1QMH, 1B1C), die zu der fernen Arylbindetasche ähnliche physikochemischen Eigenschaften aufweisen. Oberflächenbereiche mit übereinstimmenden Eigenschaften sind farblich hervorgehoben (blau: Donor-Eigenschaften; rot: Akzeptor-Eigenschaften; grün: Donor- und Akzeptor-Eigenschaften). Die in den ermittelten Bindetaschen gebundenen Liganden sind grün dargestellt (**1**: Citrat; **2**: Flavinmononukleotid). Zur räumlichen Orientierung sind die Bindungsmodi von CAAX und FPP in der FTase (PDB-Code 1QBQ) abgebildet.

3.4 Zusammenfassung

In diesem Kapitel wurde die Entwicklung neuer Hemmstoffe der Farnesyltransferase (FTase) mit Hilfe verschiedener Ansätze des struktur-basierten Designs aufgezeigt.

Ausgehend von der Struktur des CAAX-Tetrapeptids als Substrat der FTase wurde im Laufe der Entwicklung eine neue Klasse von peptidomimetischen Inhibitoren dargestellt, die auf einer Benzophenon-Grundstruktur beruhen. Die flexible Überlagerung eines Inhibitors dieser Klasse auf das peptidische Substrat mit dem Programm SEAL zeigte dabei, dass Strukturfragmente mit gleichen physikochemischen Eigenschaften gut miteinander zur Deckung gebracht werden können.

Ein Ersatz der Cysteinfunktion der Benzophenon-Grundstruktur durch hydrophobe Arylsubstituenten führte zu einer Weiterentwicklung von thiofreien Farnesyltransferase-Inhibitoren. Die Tatsache, dass bei dieser Modifikation kein merklicher Wirkungsverlust festzustellen war, legte eine Postulierung von mindestens einer bis dahin unbekanntem Arylbindestelle im aktiven Zentrum der FTase nahe. Um diese Bindestelle(n) näher eingrenzen zu können, wurden mit dem Docking-Programm FlexX thiofreie Inhibitoren aus der Literatur als auch Benzophenon-Strukturen in die Bindetasche der FTase flexibel eingedockt. Dabei wurde die als „nahe Arylbindestelle“ bezeichnete Tasche meist von kurzkettigen arylischen Substituenten besetzt. Bei den Inhibitoren mit langkettigen Arylfunktionen ergaben sich zwei alternative Vorschläge für denkbare Bindungsmodi, wobei eine dieser beiden Bindestellen („ferne Arylbindestelle“) für eine Aufnahme von Arylsubstituenten besser geeignet erschien.

Die so gefundenen, postulierten Arylbindestellen wurden im weiteren näher untersucht. Eine Bindetaschenanalyse der FTase mit dem Programm GRID zeigte, dass die mit Hilfe der Docking-Studien identifizierten Bindestellen ebenfalls günstige Wechselwirkungsbereiche für aromatische Sondenatome darstellen. Diese somit übereinstimmenden Resultate zweier methodisch unterschiedlicher Ansätze (flexibles Docking und Analyse der Bindetasche) konnten daher die Existenz der vermuteten Arylbindestellen weiter untermauern.

Zusätzlich wurden einige langkettige Benzophenon-Inhibitoren mit terminalen Nitrosubstituenten, die eine auffallend hohen Hemmaktivität besitzen, genauer untersucht.

Eine GRID-Analyse der Bindetasche mit einem Sondenatom, das günstige Bereiche für Nitrogruppen anzeigt, bestätigte die Vermutung, dass die ferne Arylbindestelle für eine Aufnahme von Nitrogruppen besonders geeignet ist. Mehrere Lysinseitenketten im Bereich dieser Tasche könnten günstige Wechselwirkungen mit einer partiell negativ geladenen Nitrogruppe hervorrufen.

Zur weiteren Bekräftigung diese Hypothese wurde mit dem Programm CavBase eine Suche nach Protein-Ligand-Komplexen durchgeführt, die ein ähnliches Wechselwirkungsmuster zur fernen Arylbindetasche aufweisen. Daraus resultierten insbesondere zwei Proteinstrukturen, deren Bindetaschenbereiche vergleichbare physikochemische Eigenschaften aufweisen konnten. In beiden Fällen zeigten die Bindetaschen Oberflächenregionen mit exponierten Donor-Eigenschaften, die von Aminosäuren mit basischen Seitenketten erzeugt werden. Die Anwesenheit der in diesen Taschen gebundenen Liganden mit funktionellen Akzeptorgruppen untermauert somit die Annahme, dass auch die Nitrogruppe des hochaffinen Inhibitors günstige Wechselwirkungen mit den basischen Lysinseitenketten der fernen Arylbindestelle eingehen kann.

4 ANHANG

4.1 Verwendete Hardware

Alle Berechnungen und Visualisierungen in dieser Arbeit wurden auf einer Silicon Graphics O2 mit einem 225 MHz MIPS R10.000 Prozessor mit 256 MByte RAM durchgeführt.

4.2 Verwendete Programme

Die Berechnungen zur vergleichenden molekularen Feldanalyse und die Visualisierung aller Ergebnisse dieser Arbeit wurden mit dem Programmpaket SYBYL (SYBYL, 2000) durchgeführt. Die partiellen Atomladungen der Liganden wurden semiempirisch nach dem AM1-Verfahren (Dewar *et al.*, 1985) mit dem Programm MOPAC (Stewart, 1990; MOPAC, 2000) berechnet.

Zur Berechnung der wissensbasierten Deskriptoren wurden die Datenbank IsoStar (Bruno *et al.*, 1997) und eine angepasste Version des Programms SuperStar (Verdonk *et al.*, 1999) verwendet. Die Importierung der externer Felder aus SuperStar wurde mit Hilfe eines SPL-Makros implementiert. SPL (*SYBYL Programming Language*) ist eine in SYBYL implementierte Skriptsprache.

Das Docking der Farnesyltransferase-Inhibitoren erfolgte mit dem flexiblen Docking-Programm FlexX, Version 1.7.6 (Rarey *et al.*, 1996). Zur räumlichen Positionierung der Basisfragmente wurden die Optionen MAPREF und PLACEBAS (Modus *perturbate*) verwendet. Alle Standardparameter wurden übernommen; einzig der Wert von MAX_ENERGY wurde auf 10 kJ/mol angehoben (Standard: 0 kJ/mol).

Die flexible Überlagerung der Benzophenon-Leitstruktur auf das CAAX-Peptidomimetikums CVIM wurde mit dem Programm SEAL durchgeführt (Klebe *et al.*, 1994b).

Das Programm GRID wurde in Kap. 1 zur Berechnung von Wechselwirkungsfeldern kleiner Moleküle verwendet und diente dort als Vergleichsbasis zu den in dieser Arbeit entwickelten SuperStar-Deskriptoren. Weiterhin wurde GRID zur Bindetaschenanalyse der 16 Serinproteasen von Kap. 2.4 und der Farnesyltransferase von Kap. 3 eingesetzt.

Für die Minimierung der Serinprotease-Inhibitoren zur spezifischen Adaption in der Proteinbindetasche wurde das im Programm MOLOC (MOLOC, 2001) implementierte Kraftfeld MAB (Gerber & Muller, 1995; Gerber, 1998) verwendet.

Die chemometrischen Analysen zur Selektivitätsbestimmung der Serinproteasen erfolgte mit dem Programmpaket GOLPE (GOLPE, 2001).

4.3 Statistische Parameter der PLS-Analyse

4.3.1 Kreuzvalidierung

Ein oder mehrere Moleküle werden dem Datensatz nach dem Zufallsprinzip entnommen und mit den verbleibenden Molekülen wird ein Modell aufgestellt. Die Affinitäten der entnommenen Verbindungen werden anhand dieses Modells vorhergesagt und die Differenz zwischen vorhergesagten ($pK_{i,pred}$) und experimentell beobachteten ($pK_{i,act}$) Affinitäten gebildet. Das Herausnehmen der Verbindungen wird solange wiederholt, bis jedes Molekül wenigstens einmal vorhergesagt wurde. Die am häufigsten angewendete Form der Kreuzvalidierung ist die „*leave-one-out*“-Methode, bei der immer nur ein Molekül entnommen und seine biologische Aktivität vorhergesagt wird. Diese Vorgehensweise wird bei einem Datensatz aus n Verbindungen genau n -mal wiederholt. Die Summe aller quadratischen Differenzen ergibt den Wert von PRESS (*Predictive Residual Sum of Squares*) (Gl. 4-1).

$$PRESS = \sum_n (pK_{i,pred} - pK_{i,act})^2 \quad \text{Gl. 4-1}$$

Da der PRESS-Wert abhängig von der gewählten Skala der Variablen ist, wurde der statistische Parameter q^2 (*squared correlation coefficient of cross-validation*) eingeführt (Gl. 4-2). Der Wert SD ist die Summe der quadratischen Differenzen zwischen den experimentellen Werten und dem Mittelwert $pK_{i,mean}$ des Datensatzes (Gl. 4-3).

$$q^2 = 1 - \frac{PRESS}{SD} \quad \text{Gl. 4-2}$$

$$SD = \sum_n (pK_{i,mean} - pK_{i,act})^2 \quad \text{Gl. 4-3}$$

Ein q^2 -Wert von 1 (PRESS = 0) würde auf ein perfektes Modell hinweisen, denn die vorhergesagten Affinitäten entsprechen in diesem Fall genau den experimentellen Werten. Im Falle von $q^2 = 0$ lässt sich PRESS = SD aus Gl. 4-2 ableiten. Bei einem Vergleich von Gl. 4-1 und Gl. 4-3 folgt daraus $pK_{i,pred} = pK_{i,mean}$. Für die Vorhersage der Affinitäten könnte also jeweils der Mittelwert des Datensatzes verwendet werden; folglich liegt kein vorhersagefähiges Modell vor. Modelle mit q^2 -Werten kleiner als Null weisen eine noch geringere Vorhersagefähigkeit auf. Ein Modell wird im allgemeinen für $q^2 > 0,3$ als statistisch signifikant akzeptiert (Agarwal *et al.*, 1993; Cramer *et al.*, 1993).

Die statistische Größe S_{PRESS} (*standard deviation of predictions*) ermöglicht eine quantitative Fehlerabschätzung bei der Kreuzvalidierung (Gl. 4-4). Dieser Wert ist wie PRESS von der Skalierung der eingesetzten Variablen abhängig und stellt ein Maß für die erwartete Unsicherheit einer vorhergesagten Affinität dar (n: Anzahl der Verbindungen im Datensatz; a: Anzahl der verwendeten Komponenten).

$$S_{PRESS} = \sqrt{\frac{PRESS}{n-a-1}} \quad \text{Gl. 4-4}$$

Alternativ dazu wird auch der Parameter SDEP (*Standard Deviation of Error of Predictions*) zur Fehlerabschätzung eingesetzt (Gl. 4-5).

$$SDEP = \sqrt{\frac{PRESS}{n}} \quad \text{Gl. 4-5}$$

Diese Größe lässt jedoch die Zahl der verwendeten Komponenten unberücksichtigt. Ein Modell mit gleichem PRESS und weniger Komponenten erhält einen niedrigeren Wert für $SPRESS$. Dies entspricht dem „*principle of parsimony*“ – ein Modell mit möglichst wenigen Komponenten bei bester Vorhersagekraft zu erstellen.

4.3.2 Nicht-Kreuzvalidierung (Endvalidierung)

Unter Verwendung der in der Kreuzvalidierung gewählten, optimalen Zahl von Komponenten wird eine abschließende PLS-Analyse durchgeführt. Analog zu den Gleichungen der Kreuzvalidierung ergeben sich die statistischen Parameter RSS, r^2 und S. Dabei entspricht RSS (*Residual Sum of Squares*) dem PRESS-Wert der kreuzvalidierten Analyse (Gl. 4-6); anstelle der vorhergesagten ($pK_{i,pred}$) erscheinen die angepassten (*fitted*) Bindungsaffinitäten ($pK_{i,fit}$). Der Parameter r^2 (*squared correlation coefficient*) ist dem q^2 -Wert der Kreuzvalidierung äquivalent (Gl. 4-7) und S (*standard deviation*) entspricht der Standardabweichung $SPRESS$ der kreuzvalidierten Analyse (Gl. 4-8).

$$RSS = \sum_n (pK_{i,fit} - pK_{i,act})^2 \quad \text{Gl. 4-6}$$

$$r^2 = 1 - \frac{RSS}{SD} \quad \text{Gl. 4-7}$$

$$S = \sqrt{\frac{RSS}{n - a - 1}} \quad \text{Gl. 4-8}$$

Zur Überprüfung der statistischen Signifikanz wird häufig auch der Fischer-Wert F angegeben, der als Verhältnis von erklärten zu unerklärten Ergebnissen (über r^2 ausgedrückt) definiert ist und möglichst große Werte annehmen sollte (Harnett & Murphy, 1980).

$$F = \frac{r^2}{1 - r^2} \quad \text{Gl. 4-9}$$

4.3.3 Vorhersage unbekannter Verbindungen

Die Güte eines aufgestellten Modells lässt sich mit Hilfe eines Testdatensatzes abschätzen. Dabei werden Verbindungen, die nicht im Trainingsdatensatz vorhanden sind, vorhergesagt und mit ihren bereits bekannten, experimentell bestimmten Affinitäten verglichen. Analog zu dem q^2 -Wert der Kreuzvalidierung (*crossvalidated r^2*) wird hierfür der *predictive r^2* berechnet (Gl. 4-10).

$$\text{predictive } r^2 = 1 - \frac{PRESS}{SD} \quad \text{Gl. 4-10}$$

In diesem Fall entspricht PRESS der Summe der quadratischer Differenzen zwischen vorhergesagten und experimentellen Aktivitäten der Moleküle des Testdatensatzes (vgl. Gl. 4-1). Der Wert SD wird durch die Summe aller quadratischen Differenzen zwischen experimentellen Werten und dem Mittelwert gebildet. Zur Bestimmung des Mittelwerts wurden in dieser Arbeit die Verbindungen des Trainingsdatensatzes verwendet. Diese Vorgehensweise wurde auch von Cramer *et al.* (Cramer *et al.*, 1988) und anderen (Kim *et al.*, 1998) vorgeschlagen, da die Aktivitätsdaten innerhalb des Trainingsatzes gewöhnlich über einen größeren Aktivitätsbereich streuen. Die Unterschiede des *predictive r^2* bei der Verwendung der Moleküle des Trainings- und Testdatensatzes oder nur des Testdatensatzes zur Bestimmung des Mittelwerts waren jedoch in allen Fällen nur minimal.

LITERATURVERZEICHNIS

- Agarwal, A., Pearson, P. P., Taylor, E. W., Li, H. B., Dahlgren, T., Herslof, M., Yang, Y., Lambert, G., Nelson, D. L., Regan, J. W. & Martin, A. R. (1993). Three-dimensional quantitative structure-activity relationships of 5-HT receptor binding data for tetrahydropyridinylindole derivatives: a comparison of the Hansch and CoMFA methods. *J. Med. Chem.* **36**, 4006-14.
- Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). The development of version-3 and version-4 of the Cambridge Structural Database system. *J. Chem. Inf. Comput. Sci.* **31**, 187-204.
- Allen, F. H. & Kennard, O. (1993). 3D Search and Research Using the Cambridge Structural Database. *Chem. Des. Automation News* **8**, 31-37.
- Anthony, N. J., Gomez, R. P., Schaber, M. D., Mosser, S. D., Hamilton, K. A., O'Neil, T. J., Koblan, K. S., Graham, S. L., Hartman, G. D., Shah, D., Rands, E., Kohl, N. E., Gibbs, J. B. & Oliff, A. I. (1999). Design and in vivo analysis of potent non-thiol inhibitors of farnesyl protein transferase. *J. Med. Chem.* **42**, 3356-68.
- Augeri, D. J., Janowick, D., Kalvin, D., Sullivan, G., Larsen, J., Dickman, D., Ding, H., Cohen, J., Lee, J., Warner, R., Kovar, P., Cherian, S., Saeed, B., Zhang, H., Tahir, S., Ng, S. C., Sham, H. & Rosenberg, S. H. (1999). Potent and orally bioavailable noncysteine-containing inhibitors of protein farnesyltransferase. *Bioorg. Med. Chem. Lett.* **9**, 1069-74.
- Augeri, D. J., O'Connor, S. J., Janowick, D., Szczepankiewicz, B., Sullivan, G., Larsen, J., Kalvin, D., Cohen, J., Devine, E., Zhang, H., Cherian, S., Saeed, B., Ng, S. C. & Rosenberg, S. (1998). Potent and selective non-cysteine-containing inhibitors of protein farnesyltransferase. *J. Med. Chem.* **41**, 4288-300.
- Bell, I. M. (2000). Inhibitors of protein prenylation 2000. *Exp. Opin. Ther. Patents* **10**, 1813-31.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-42.

- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-42.
- Bode, W., Brandstetter, H., Mather, T. & Stubbs, M. T. (1997). Comparative analysis of haemostatic proteinases: structural aspects of thrombin, factor Xa, factor IXa and protein C. *Thromb. Haemost.* **78**, 501-11.
- Bode, W. & Stubbs, M. T. (1993). Spatial structure of thrombin as a guide to its multiple sites of interaction. *Semin. Thromb. Hemost.* **19**, 321-33.
- Bode, W. & Stubbs, M. T. (1995). Die Raumstruktur des Thrombins: ein Wegweiser zu seinen vielfältigen Funktionen. *Hämostaseologie* **15**, 1-13.
- Boer, D. R., Kroon, J., Cole, J. C., Smith, B. & Verdonk, M. L. (2001). Superstar: comparison of csd and pdb-based interaction fields as a basis for the prediction of protein-ligand interactions. *J. Mol. Biol.* **312**, 275-87.
- Böhm, H. J., Klebe, G. & Kubinyi, H. (1996). *Wirkstoffdesign - Der Weg zum Arzneimittel*, Spektrum Akademischer Verlag, Heidelberg Berlin Oxford.
- Böhm, M., Mitsch, A., Wissner, P., Sattler, I. & Schlitzer, M. (2001). Exploration of novel aryl binding sites of farnesyltransferase using molecular modeling and benzophenone-based farnesyltransferase inhibitors. *J. Med. Chem.* **44**, 3117-24.
- Böhm, M., Stürzebecher, J. & Klebe, G. (1999). Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **42**, 458-77.
- Boobbyer, D. N., Goodford, P. J., McWhinnie, P. M. & Wade, R. C. (1989). New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **32**, 1083-94.
- Breslin, M. J., deSolms, J., Giuliani, E. A., Stokker, G. E., Graham, S. L., Pompliano, D. L., Mosser, S. D., Hamilton, K. A. & Hutchinson, J. H. (1998). Potent, non-thiol inhibitors of farnesyltransferase. *Bioorg. Med. Chem. Lett.* **8**, 3311-6.
- Brooks, R. R., Bruccoleri, B. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187-217.

- Bruno, I. J., Cole, J. C., Lommerse, J. P., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). IsoStar: a library of information about nonbonded interactions. *J. Comput.-Aided Mol. Des.* **11**, 525-37.
- Bush, B. L. & Nachbar, R. B., Jr. (1993). Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **7**, 587-619.
- Caflisch, A. (1996). Computational combinatorial ligand design: application to human alpha-thrombin. *J. Comput.-Aided Mol. Des.* **10**, 372-96.
- Caflisch, A., Miranker, A. & Karplus, M. (1993). Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* **36**, 2142-67.
- Carbo, R., Leyda, L. & Arnau, M. (1980). An electron density measure of the similarity between two compounds. *Int. J. Quantum Chem.* **17**, 1185-9.
- Casey, P. J. & Seabra, M. C. (1996). Protein prenyltransferases. *J. Biol. Chem.* **271**, 5289-92.
- Ciccarone, T. M., MacTough, S. C., Williams, T. M., Dinsmore, C. J., O'Neill, T. J., Shah, D., Culberson, J. C., Koblan, K. S., Kohl, N. E., Gibbs, J. B., Oliff, A. I., Graham, S. L. & Hartman, G. D. (1999). Non-thiol 3-aminomethylbenzamide inhibitors of farnesyl-protein transferase. *Bioorg. Med. Chem. Lett.* **9**, 1991-6.
- Clark, M., Cramer, R. D., III & Van Opdenbosch, N. (1989). Validation of the General Purpose Tripos 5.2 Force Field. *J. Comp. Chem.* **10**, 982-1012.
- Cole, J. C., Taylor, R. & Verdonk, M. L. (1998). Directional preferences of intermolecular contacts to hydrophobic groups. *Acta Crystallogr. D. Biol. Crystallogr.* **54**, 1183-93.
- Connolly, M. L. (1983). Solvent accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-13.
- Cox, A. D. & Der, C. J. (1997). Farnesyltransferase inhibitors and cancer treatment: targeting simply Ras? *Biochim. Biophys. Acta* **1333**, F51-71.
- Cramer, R. D., III, DePriest, S. A., Patterson, D. E. & Hecht, P., Eds. (1993). The developing practice of Comparative Molecular Field Analysis. 3D QSAR in Drug Design: Theory, Methods and Applications. Edited by Kubinyi, H. Leiden: ESCOM.

- Cramer, R. D., III, Patterson, D. E. & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **110**, 5959-67.
- Danziger, D. J. & Dean, P. M. (1989a). Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces. *Proc. R. Soc. Lond. B. Biol. Sci.* **236**, 101-13.
- Danziger, D. J. & Dean, P. M. (1989b). Automated site-directed drug design: the prediction and observation of ligand point positions at hydrogen-bonding regions on protein surfaces. *Proc. R. Soc. Lond. B. Biol. Sci.* **236**, 115-24.
- DePriest, S. A., Mayer, D., Naylor, C. B. & Marshall, G. R. (1993). 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors: a comparison of CoMFA models based on deduced and experimentally determined active site geometries. *J. Am. Chem. Soc.* **115**, 5372-84.
- Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. (1985). AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902-9.
- Dodson, G. & Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347-52.
- Duder, T. & Gelb, M. H. (1996). Palmitoylation of H-Ras facilitates membrane binding, activation of downstream effectors and meiotic maturation in *Xenopus* oocytes. *J. Biol. Chem.* **271**, 11541-7.
- End, D. W. (1999). Farnesyl protein transferase inhibitors and other therapies targeting the Ras signal transduction pathway. *Invest. New Drugs* **17**, 241-58.
- English, A. C., Groom, C. R. & Hubbard, R. E. (2001). Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* **14**, 47-59.
- Farrell, F. X., Yamamoto, K. & Lapetina, E. G. (1993). Prenyl group identification of rap2 proteins: a ras superfamily member other than ras that is farnesylated. *Biochem. J.* **289**, 349-55.
- Feng, D. M., Gardell, S. J., Lewis, S. D., Bock, M. G., Chen, Z., Freidinger, R. M., Naylor-Olsen, A. M., Ramjit, H. G., Woltmann, R., Baskin, E. P., Lynch, J. J., Lucas, R., Shafer, J. A., Dancheck, K. B., Chen, I. W., Mao, S. S., Krueger, J. A., Hare, T. R., Mulichak, A. M. & Vacca, J. P. (1997). Discovery of a novel, selective, and orally bioavailable class of

- thrombin inhibitors incorporating aminopyridyl moieties at the P1 position. *J. Med. Chem.* **40**, 3726-33.
- Fichera, M., Cruciani, G., Bianchi, A. & Musumarra, G. (2000). A 3D-QSAR study on the structural requirements for binding to CB(1) and CB(2) cannabinoid receptors. *J. Med. Chem.* **43**, 2300-9.
- Filipponi, E., Cecchetti, V., Tabarrini, O., Bonelli, D. & Fravolini, A. (2000). Chemometric rationalization of the structural and physicochemical basis for selective cyclooxygenase-2 inhibition: toward more specific ligands. *J. Comput.-Aided Mol. Des.* **14**, 277-91.
- Forth, W., Henschler, D., Rummel, W., Förstermann, U. & Starke, K. (2001). *Allgemeine und spezielle Pharmakologie und Toxikologie*, 8. Auflage, Urban & Fischer, München.
- Gabriel, B., Stubbs, M. T., Bergner, A., Hauptmann, J., Bode, W., Stürzebecher, J. & Moroder, L. (1998). Design of benzamidine-type inhibitors of factor Xa. *J. Med. Chem.* **41**, 4240-50.
- Galemmo, R. A., Jr., Wells, B. L., Rossi, K. A., Alexander, R. S., Dominguez, C., Maduskuie, T. P., Stouten, P. F., Wright, M. R., Aungst, B. J., Wong, P. C., Knabb, R. M. & Wexler, R. R. (2000). The de novo design and synthesis of cyclic urea inhibitors of factor Xa: optimization of the S4 ligand. *Bioorg. Med. Chem. Lett.* **10**, 301-4.
- Gerber, P. R. (1998). Charge distribution from a simple molecular orbital type calculation and non-bonding interaction terms in the force field MAB. *J. Comput.-Aided Mol. Des.* **12**, 37-51.
- Gerber, P. R. & Muller, K. (1995). MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput.-Aided Mol. Des.* **9**, 251-68.
- Ghomashchi, F., Zhang, X. & Liu, L. (1995). Binding of prenylated and polybasic peptides to membranes: affinities and intervesicle exchange. *Biochemistry* **34**, 11910-8.
- Gibbs, J. B., Kohl, N. E., Koblan, K. S., Omer, C. A., Sepp-Lorenzino, L., Rosen, N., Anthony, N. J., Conner, M. W., deSolms, S. J., Williams, T. M., Graham, S. L., Hartman, G. D. & Oliff, A. (1996). Farnesyltransferase inhibitors and anti-Ras therapy. *Breast Cancer Res. Treat.* **38**, 75-83.
- Gibbs, J. B., Oliff, A. & Kohl, N. E. (1994). Farnesyltransferase inhibitors: Ras research yields a potential cancer therapeutic. *Cell* **77**, 175-8.

- Gohlke, H., Hendlich, M. & Klebe, G. (2000a). Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **295**, 337-56.
- Gohlke, H., Hendlich, M. & Klebe, G. (2000b). Predicting binding modes, binding affinities and "Hot Spots" for protein-ligand complexes using a knowledge-based scoring function. *Persp. Drug Discov. Design* **20**, 115-44.
- Gohlke, H. & Klebe, G. (2001). Statistical potentials and scoring functions applied to protein-ligand binding. *Curr. Opin. Struct. Biol.* **11**, 231-5.
- GOLPE, version 4.5; Multivariate Infometric Analysis (MIA) Srl.: Viale dei Castagni 16, Perugia, Italy. <http://www.miasrl.com>
- Gong, Y., Pauls, H. W., Spada, A. P., Czekaj, M., Liang, G., Chu, V., Colussi, D. J., Brown, K. D. & Gao, J. (2000). Amido-(propyl and allyl)-hydroxybenzamidines: development of achiral inhibitors of factor Xa. *Bioorg. Med. Chem. Lett.* **10**, 217-21.
- Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849-57.
- GRID, version 19; Molecular Discovery Ltd.: West Way House, Elms Parade, Oxford, U.K. <http://www.moldiscovery.com>
- Guilford, W. J., Shaw, K. J., Dallas, J. L., Koovakkat, S., Lee, W., Liang, A., Light, D. R., McCarrick, M. A., Whitlow, M., Ye, B. & Morrissey, M. M. (1999). Synthesis, characterization, and structure-activity relationships of amidine-substituted (bis)benzylidene-cycloketone olefin isomers as potent and selective factor Xa inhibitors. *J. Med. Chem.* **42**, 5415-25.
- Hara, T., Yokoyama, A., Ishihara, H., Yokoyama, Y., Nagahara, T. & Iwamoto, M. (1994). DX-9065a, a new synthetic, potent anticoagulant and selective inhibitor for factor Xa. *Thromb. Haemost.* **71**, 314-9.
- Harnett, D. L. & Murphy, J. L. (1980). The F-Test. In *Introductory Statistical Analysis 2* edit., pp. 430. Addison-Wesley, Reading, Mass.
- Herbert, J. M., Bernat, A., Dol, F., Herault, J. P., Crepon, B. & Lormeau, J. C. (1996). DX 9065A a novel, synthetic, selective and orally active inhibitor of factor Xa: in vitro and in vivo studies. *J. Pharmacol. Exp. Ther.* **276**, 1030-8.
- Hodgkin, E. E. & Richards, W. G. (1987). Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem. Quantum Biol. Symp.* **14**, 105-10.

- Huang, C., Hightower, K. E. & Fierke, C. A. (2000). Mechanistic studies of rat protein farnesyltransferase indicate an associative transition state. *Biochemistry* **39**, 2593-602.
- Joneson, T., White, M. A. & Wigler, M. H. (1996). Stimulation of membrane ruffling and MAP kinase activation by distinct effectors of Ras. *Science* **271**, 810-2.
- Kastenholz, M. A., Pastor, M., Cruciani, G., Haaksma, E. E. & Fox, T. (2000). GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* **43**, 3033-44.
- Katakura, S., Nagahara, T., Hara, T. & Iwamoto, M. (1993). A novel factor Xa inhibitor: structure-activity relationships and selectivity between factor Xa and thrombin. *Biochem. Biophys. Res. Commun.* **197**, 965-72.
- Kato, K., Cox, A. D. & Hisaka, M. M. (1992). Isoprenoid addition to Ras protein is the critical modification for its membrane association and transforming activity. *Proc. Natl. Acad. Sci. USA* **89**, 6403-7.
- Katz, B. A., Mackman, R., Luong, C., Radika, K., Martelli, A., Sprengeler, P. A., Wang, J., Chan, H. & Wong, L. (2000). Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chem. Biol.* **7**, 299-312.
- Khosravi-Far, R. & Der, C. J. (1994). The Ras signal transduction pathway. *Cancer Metastasis Rev.* **13**.
- Kim, K. H., Greco, G. & Novellino, E. (1998). A critical review of recent CoMFA applications. *Persp. Drug Discov. Design* **12**, 257-315.
- Klebe, G. (1994). The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **237**, 212-35.
- Klebe, G. (1998). Comparative molecular similarity indices analysis: CoMSIA. *Persp. Drug Discov. Design* **12-14**, 87-104.
- Klebe, G. & Abraham, U. (1999). Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.* **13**, 1-10.
- Klebe, G., Abraham, U. & Mietzner, T. (1994a). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **37**, 4130-46.

- Klebe, G., Mietzner, T. & Weber, F. (1994b). Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. *J. Comput.-Aided Mol. Des.* **8**, 751-78.
- Klein, S. I., Czekaj, M., Gardner, C. J., Guertin, K. R., Cheney, D. L., Spada, A. P., Bolton, S. A., Brown, K., Colussi, D., Heran, C. L., Morgan, S. R., Leadley, R. J., Dunwiddie, C. T., Perrone, M. H. & Chu, V. (1998). Identification and initial structure-activity relationships of a novel class of nonpeptide inhibitors of blood coagulation factor Xa. *J. Med. Chem.* **41**, 437-50.
- Klinghofer, V., Stewart, K., McGonigal, T., Smith, R., Sarthy, A., Nienaber, V., Butler, C., Dorwin, S., Richardson, P., Weitzberg, M., Wendt, M., Rockway, T., Zhao, X., Hulkower, K. I. & Giranda, V. L. (2001). Species specificity of amidine-based urokinase inhibitors. *Biochemistry* **40**, 9125-31.
- Kubinyi, H. (1993). *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM, Leiden.
- Kubinyi, H. (1994). The key to the castle. II. Hansch analysis, 3D QSAR and de novo design. *Pharm. Unserer Zeit* **23**, 281-90.
- Kubinyi, H. & Abraham, U., Eds. (1993). Practical problems in PLS analyses. 3D QSAR in Drug Design: Theory, Methods and Applications. Edited by Kubinyi, H. Leiden: ESCOM.
- Kumura, T., Hino, M., Yamane, T., Tominaga, K. & Tatsumi, N. (2000). DX-9065a, a specific factor Xa inhibitor, as a universal anticoagulant for blood collection tubes. *Clin. Chim. Acta* **294**, 27-35.
- Laskowski, R. A., Thornton, J. M., Humblet, C. & Singh, J. (1996). X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins. *J. Mol. Biol.* **259**, 175-201.
- Lebowitz, P. F., Casey, P. J., Prendergast, G. C. & Thissen, J. A. (1997a). Farnesyltransferase inhibitors alter the prenylation and growth-stimulating function of RhoB. *J. Biol. Chem.* **272**, 15591-4.
- Lebowitz, P. F., Sakamuro, D. & Prendergast, G. C. (1997b). Farnesyl transferase inhibitors induce apoptosis of Ras-transformed cells denied substratum attachment. *Cancer Res.* **57**, 708-13.

- Lemmen, C. & Lengauer, T. (2000). Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **14**, 215-32.
- Leonard, D. M. (1997). Ras farnesyltransferase: a new therapeutic target. *J. Med. Chem.* **40**, 2971-90.
- Lommerse, J. P. & Taylor, R. (1997). Characterising non-covalent interactions with the Cambridge Structural Database. *J. Enzyme Inhib.* **11**, 223-43.
- Long, S. B., Casey, P. J. & Beese, L. S. (1998). Cocystal structure of protein farnesyltransferase complexed with a farnesyl diphosphate substrate. *Biochemistry* **37**, 9612-8.
- Long, S. B., Casey, P. J. & Beese, L. S. (2000). The basis for K-Ras4B binding specificity to protein farnesyltransferase revealed by 2 Å resolution ternary complex structures. *Structure Fold. Des.* **8**, 209-22.
- Lopez-Rodriguez, M. L., Morcillo, M. J., Fernandez, E., Rosado, M. L., Orensanz, L., Beneytez, M. E., Manzanares, J., Fuentes, J. A. & Schaper, K. J. (1999). Design and synthesis of 2-[4-[4-(m-(ethylsulfonamido)-phenyl) piperazin-1-yl]butyl]-1,3-dioxoperhydropyrrolo[1,2-c]imidazole (EF-7412) using neural networks. A selective derivative with mixed 5-HT_{1A}/D₂ antagonist properties. *Bioorg. Med. Chem. Lett.* **9**, 1679-82.
- Lumma, W. C., Jr., Witherup, K. M., Tucker, T. J., Brady, S. F., Sisko, J. T., Naylor-Olsen, A. M., Lewis, S. D., Lucas, B. J. & Vacca, J. P. (1998). Design of novel, potent, noncovalent inhibitors of thrombin with nonbasic P-1 substructures: rapid structure-activity studies by solid-phase synthesis. *J. Med. Chem.* **41**, 1011-3.
- Mack, H., Pfeiffer, T., Hornberger, W., Böhm, H. J. & Hoffken, H. W. (1995). Design, synthesis and biological activity of novel rigid amidino-phenylalanine derivatives as inhibitors of thrombin. *J. Enzyme Inhib.* **9**, 73-86.
- Maduskuie, T. P., Jr., McNamara, K. J., Ru, Y., Knabb, R. M. & Stouten, P. F. (1998). Rational design and synthesis of novel, potent bis-phenylamidine carboxylate factor Xa inhibitors. *J. Med. Chem.* **41**, 53-62.
- Matter, H. & Schwab, W. (1999). Affinity and selectivity of matrix metalloproteinase inhibitors: a chemometrical study from the perspective of ligands and proteins. *J. Med. Chem.* **42**, 4506-23.
- McCormick, F. (1996). Ras biology in atomic detail. *Nat. Struct. Biol.* **3**, 653-5.

- Mills, J. E. & Dean, P. M. (1996). Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput.-Aided Mol. Des.* **10**, 607-22.
- Mills, J. E., Perkins, T. D. & Dean, P. M. (1997). An automated method for predicting the positions of hydrogen-bonding atoms in binding sites. *J. Comput.-Aided Mol. Des.* **11**, 229-42.
- Minden, A., Lin, A. & McMahon, M. (1994). Differential activation of ERK and JNK mitogen-activated protein kinases by Raf-1 and MEKK. *Science* **266**, 1719-23.
- Miranker, A. & Karplus, M. (1991). Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **11**, 29-34.
- Mitchell, J. B. O., Laskowski, R. A., Alex, A., Forster, M. J. & Thornton, J. M. (1999a). BLEEP - potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comp. Chem.* **20**, 1177-1185.
- Mitchell, J. B. O., Laskowski, R. A., Alex, A. & Thornton, J. M. (1999b). BLEEP - potential of mean force describing protein-ligand interactions: I. Generating potential. *J. Comp. Chem.* **20**, 1165-1176.
- MOLOC molecular modeling package. <http://www.moloc.ch>
- Montorsi, M., Menziani, M. C., Cocchi, M., Fanelli, F. & De Benedetti, P. G. (1998). Computer modeling of size and shape descriptors of alpha 1-adrenergic receptor antagonists and quantitative structure-affinity/selectivity relationships. *Methods* **14**, 239-54.
- MOPAC: a general molecular orbital package, version 6.0; QCPE #455: J. J. P. Stewart, Stewart Computational Chemistry, 15210 Paddington Circle, Colorado Springs CO 80921
- Moron, J. A., Campillo, M., Perez, V., Unzeta, M. & Pardo, L. (2000). Molecular determinants of MAO selectivity in a series of indolylmethylamine derivatives: biological activities, 3D-QSAR/CoMFA analysis, and computational simulation of ligand recognition. *J. Med. Chem.* **43**, 1684-91.
- Muegge, I. & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **42**, 791-804.

- Muegge, I., Martin, Y. C., Hajduk, P. J. & Fesik, S. W. (1999). Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *J. Med. Chem.* **42**, 2498-503.
- Murayama, N., Tanaka, M., Kunitada, S., Yamada, H., Inoue, T., Terada, Y., Fujita, M. & Ikeda, Y. (1999). Tolerability, pharmacokinetics, and pharmacodynamics of DX-9065a, a new synthetic potent anticoagulant and specific factor Xa inhibitor, in healthy male volunteers. *Clin. Pharmacol. Ther.* **66**, 258-64.
- Nienaber, V. L., Davidson, D., Edalji, R., Giranda, V. L., Klinghofer, V., Henkin, J., Magdalinos, P., Mantei, R., Merrick, S., Severin, J. M., Smith, R. A., Stewart, K., Walter, K., Wang, J., Wendt, M., Weitzberg, M., Zhao, X. & Rockway, T. (2000). Structure-directed discovery of potent non-peptidic inhibitors of human urokinase that access a novel binding subsite. *Structure Fold. Des.* **8**, 553-63.
- Nissink, J. W. M., Verdonk, M. L. & Klebe, G. (2000). Simple knowledge-based descriptors to predict protein-ligand interactions. methodology and validation. *J. Comput.-Aided Mol. Des.* **14**, 787-803.
- Obst, U., Banner, D. W., Weber, L. & Diederich, F. (1997). Molecular recognition at the thrombin active site: structure-based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structures of two thrombin-inhibitor complexes. *Chem. Biol.* **4**, 287-95.
- O'Connor, S. J., Barr, K. J., Wang, L., Sorensen, B. K., Tasker, A. S., Sham, H., Ng, A.-C., Cohen, J., Devine, E., Cherian, S., Saeed, B., Zhang, H., Lee, J. Y., Warner, R., Tahir, S., Kovar, P., Ewing, P., Alder, J., Mitten, M., Leal, J., Marsh, K., Bauch, J., Hoffman, D. J., Sebti, S. M. & Rosenberg, S. H. (1999). Second-generation peptidomimetic inhibitors of protein farnesyltransferase demonstrating improved cellular potency and significant in vivo efficacy. *J. Med. Chem.* **42**, 3701-10.
- Palomer, A., Perez, J. J., Navea, S., Llorens, O., Pascual, J., Garcia, L. & Mauleon, D. (2000). Modeling cyclooxygenase inhibition. Implication of active site hydration on the selectivity of ketoprofen analogues. *J. Med. Chem.* **43**, 2280-4.
- Park, H. W., Boduluri, S. R., Moomaw, J. F., Casey, P. J. & Beese, L. S. (1997). Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution. *Science* **275**, 1800-4.
- Pastor, M. & Cruciani, G. (1995). A novel strategy for improving ligand selectivity in receptor-based drug design. *J. Med. Chem.* **38**, 4637-47.

- Perola, E., Xu, K., Kollmeyer, T. M., Kaufmann, S. H., Prendergast, F. G. & Pang, Y. P. (2000). Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* **43**, 401-8.
- Phillips, G., Davey, D. D., Eagen, K. A., Koovakkat, S. K., Liang, A., Ng, H. P., Pinkerton, M., Trinh, L., Whitlow, M., Beatty, A. M. & Morrissey, M. M. (1999). Design, synthesis, and activity of 2,6-diphenoxypyridine-derived factor Xa inhibitors. *J. Med. Chem.* **42**, 1749-56.
- Prendergast, G. C., Davide, J. P., deSolms, S. J., Giuliani, E. A., Graham, S. L., Gibbs, J. B., Oliff, A. & Kohl, N. E. (1994). Farnesyltransferase inhibition causes morphological reversion of ras- transformed cells by a complex mechanism that involves regulation of the actin cytoskeleton. *Mol. Cell. Biol.* **14**, 4193-202.
- Price, M. L. & Jorgensen, W. L. (2001). Rationale for the observed COX-2/COX-1 selectivity of celecoxib from Monte Carlo simulations. *Bioorg. Med. Chem. Lett.* **11**, 1541-4.
- Qian, Y., Sebti, S. M. & Hamilton, A. D. (1997). Farnesyltransferase as a target for anticancer drug design. *Biopolymers* **23**, 25-41.
- Rai, R., Kolesnikov, A., Li, Y., Young, W. B., Leahy, E., Sprengeler, P. A., Verner, E., Shrader, W. D., Burgess-Henry, J., Sangalang, J. C., Allen, D., Chen, X., Katz, B. A., Luong, C., Elrod, K. & Cregar, L. (2001). Development of potent and selective factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **11**, 1797-800.
- Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470-89.
- Renatus, M., Bode, W., Huber, R., Stürzebecher, J. & Stubbs, M. T. (1998). Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: implications for the inhibition of factor Xa, tPA, and urokinase. *J. Med. Chem.* **41**, 5445-56.
- Reyda, S., Klebe, G., Sohn, C., Stabe, K., Ullmann, D., Jakubke, H.-D. & Stubbs, M. T. (in preparation). Introduction of Factor Xa ligand binding site into trypsin reveals structural plasticity in serine proteinases.
- Ridderström, M., Zamora, I., Fjellström, O. & Andersson, T. B. (2001). Analysis of Selective Regions in the Active Sites of Human Cytochromes P450, 2C8, 2C9, 2C18, and 2C19 Homology Models Using GRID/CPCA. *J. Med. Chem.* **44**, 4072-81.

- Roskoski, R., Jr. & Ritchie, P. (1998). Role of the carboxyterminal residue in peptide binding to protein farnesyltransferase and protein geranylgeranyltransferase. *Arch. Biochem. Biophys.* **356**, 167-76.
- Rowinsky, E. K. & Patnaik, A. (2000). The development of protein farnesyltransferase and other *ras*-directed therapeutics for malignant diseases. *Exp. Opin. Emerg. Drugs* **5**, 161-99.
- Saderholm, M. J., Hightower, K. E. & Fierke, C. A. (2000). Role of metals in the reaction catalyzed by protein farnesyltransferase. *Biochemistry* **39**, 12398-405.
- Sakowski, J., Böhm, M., Sattler, I., Dahse, H. M. & Schlitzer, M. (2001). Synthesis, molecular modeling, and structure-activity relationship of benzophenone-based CAAX-peptidomimetic farnesyltransferase inhibitors. *J. Med. Chem.* **44**, 2886-99.
- Schlitzer, M. (1998). Hemmstoffe der Farnesyltransferase: Ein neuer Ansatz zur Entwicklung potentieller Krebstherapeutika. *Pharm. Unserer Zeit* **27**, 278-88.
- Schlitzer, M., Böhm, M. & Sattler, I. (2000a). Non-peptidic, non-prenylic bisubstrate farnesyltransferase inhibitors. Part 3: structural requirements of the central moiety for farnesyltransferase inhibitory activity. *Bioorg. Med. Chem.* **8**, 2399-406.
- Schlitzer, M., Böhm, M., Sattler, I. & Dahse, H. M. (2000b). Design, synthesis and early structure-activity relationship of farnesyltransferase inhibitors which mimic both the peptidic and the prenylic substrate. *Bioorg. Med. Chem.* **8**, 1991-2006.
- Schlitzer, M., Sattler, I. & Dahse, H. M. (1999). Different amino acid replacements in CAAX-tetrapeptide based peptidomimetic farnesyltransferase inhibitors. *Arch. Pharm.* **332**, 124-32.
- Schmitt, S., Hendlich, M. & Klebe, G. (2001). Von der Struktur zur Funktion: ein neuer Ansatz zum Erkennen von funktionellen Ähnlichkeiten zwischen Proteinen unabhängig von Sequenz- und Faltungshomologie. *Angew. Chem.* **113**, 3237-41.
- Seabra, M. C., Brown, M. S., Slaughter, C. A., Sudhof, T. C. & Goldstein, J. L. (1992a). Purification of component A of Rab geranylgeranyl transferase: possible identity with the choroideremia gene product. *Cell* **70**, 1049-57.
- Seabra, M. C., Goldstein, J. L., Sudhof, T. C. & Brown, M. S. (1992b). Rab geranylgeranyl transferase. A multisubunit enzyme that prenylates GTP-binding proteins terminating in Cys-X-Cys or Cys-Cys. *J. Biol. Chem.* **267**, 14497-503.

- Seabra, M. C., Reiss, Y., Casey, P. J., Brown, M. S. & Goldstein, J. L. (1991). Protein farnesyltransferase and geranylgeranyltransferase share a common alpha subunit. *Cell* **65**, 429-34.
- Sebti, S. M. & Hamilton, A. D. (1998). New approaches to anticancer drug design based on the inhibition of farnesyltransferase. *Drug Discov. Today* **3**, 26-32.
- Sebti, S. M. & Hamilton, A. D. (2000). Farnesyltransferase and geranylgeranyltransferase I inhibitors in cancer therapy: important mechanistic and bench to bedside issues. *Exp. Opin. Invest. Drugs* **9**, 2767-82.
- Shaw, K. J., Guilford, W. J., Dallas, J. L., Koovakkaat, S. K., McCarrick, M. A., Liang, A., Light, D. R. & Morrissey, M. M. (1998). (Z,Z)-2,7-Bis(4-amidinobenzylidene)cycloheptan-1-one: identification of a highly active inhibitor of blood coagulation factor Xa. *J. Med. Chem.* **41**, 3551-6.
- Shrader, W. D., Young, W. B., Sprengeler, P. A., Sangalang, J. C., Elrod, K. & Carr, G. (2001). Neutral inhibitors of the serine protease factor Xa. *Bioorg. Med. Chem. Lett.* **11**, 1801-4.
- Singh, J., Saldanha, J. & Thornton, J. M. (1991). A novel method for the modelling of peptide ligands to their receptors. *Protein Eng.* **4**, 251-61.
- Sippl, M. J. (1993). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput.-Aided Mol. Des.* **7**, 473-501.
- Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **5**, 229-35.
- Stewart, J. J. (1990). MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **4**, 1-105.
- Strickland, C. L., Windsor, W. T., Syto, R., Wang, L., Bond, R., Wu, Z., Schwartz, J., Le, H. V., Beese, L. S. & Weber, P. C. (1998). Crystal structure of farnesyl protein transferase complexed with a CaaX peptide and farnesyl diphosphate analogue. *Biochemistry* **37**, 16601-11.
- Stubbs, M. T. (1996). Structural aspects of factor Xa inhibition. *Curr. Pharm. Des.* **2**, 543-52.

- Stubbs, M. T. & Bode, W. (1993a). A model for the specificity of fibrinogen cleavage by thrombin. *Semin. Thromb. Hemost.* **19**, 344-51.
- Stubbs, M. T. & Bode, W. (1993b). A player of many parts: the spotlight falls on thrombin's structure. *Thromb. Res.* **69**, 1-58.
- Stubbs, M. T. & Bode, W. (1994). Coagulation factors and their inhibitors. *Curr. Opin. Struct. Biol.* **4**, 823-32.
- Stubbs, M. T. & Bode, W. (1995). The clot thickens: clues provided by thrombin structure. *Trends Biochem. Sci.* **20**, 23-8.
- Stubbs, M. T., Huber, R. & Bode, W. (1995). Crystal structures of factor Xa specific inhibitors in complex with trypsin: structural grounds for inhibition of factor Xa and selectivity against thrombin. *FEBS Lett.* **375**, 103-7.
- Stubbs, M. T., Oschkinat, H., Mayr, I., Huber, R., Angliker, H., Stone, S. R. & Bode, W. (1992). The interaction of thrombin with fibrinogen. A structural basis for its specificity. *Eur. J. Biochem.* **206**, 187-95.
- Stubbs, M. T., Reyda, S., Dullweber, F., Möller, M., Klebe, G., Dorsch, D., Mederski, W. W. K. R. & Wurziger, H. (in press). pH-dependent binding modes observed in trypsin crystals - Lessons for structure-based drug design. *ChemBioChem*.
- Stürzebecher, J., Prasa, D., Hauptmann, J., Vieweg, H. & Wikström, P. (1997). Synthesis and structure-activity relationships of potent thrombin inhibitors: piperazides of 3-amidinophenylalanine. *J. Med. Chem.* **40**, 3091-9.
- Stürzebecher, J., Prasa, D., Wikström, P. & Vieweg, H. (1995). Structure-activity relationships of inhibitors derived from 3-amidinophenylalanine. *J. Enzyme Inhib.* **9**, 87-99.
- Stürzebecher, J., Vieweg, H., Steinmetzer, T., Schweinitz, A., Stubbs, M. T., Renatus, M. & Wikstrom, P. (1999). 3-Amidinophenylalanine-based inhibitors of urokinase. *Bioorg. Med. Chem. Lett.* **9**, 3147-52.
- SYBYL molecular modeling package, version 6.7; Tripos Inc.: 1699 South Hanley Road, Suite 303, St. Louis, MO 63144. <http://www.tripos.com>
- Tamanoi, F. & Sigman, D. S., Eds. (2001). Protein Lipidation. Third edit. Vol. XXI. The Enzymes. San Diego / London: Academic Press.

- Thibaut, U., Folkers, G., Klebe, G., Kubinyi, H., Merz, A. & Rognan, D., Eds. (1993). Recommendations for CoMFA studies and 3D QSAR publications. 3D QSAR in Drug Design: Theory, Methods and Applications. Edited by Kubinyi, H. Leiden: ESCOM.
- Travis, J. (1993). Novel anticancer agents move closer to reality. *Science* **260**, 1877-8.
- Tschantz, W. R., Furfine, E. S. & Casey, P. J. (1997). Substrate binding is required for release of product from mammalian protein farnesyltransferase. *J. Biol. Chem.* **272**, 9989-93.
- Tucker, T. J., Brady, S. F., Lumma, W. C., Lewis, S. D., Gardell, S. J., Naylor-Olsen, A. M., Yan, Y., Sisko, J. T., Stauffer, K. J., Lucas, B. J., Lynch, J. J., Cook, J. J., Stranieri, M. T., Holahan, M. A., Lyle, E. A., Baskin, E. P., Chen, I. W., Dancheck, K. B., Krueger, J. A., Cooper, C. M. & Vacca, J. P. (1998). Design and synthesis of a series of potent and orally bioavailable noncovalent thrombin inhibitors that utilize nonbasic groups in the P1 position. *J. Med. Chem.* **41**, 3210-9.
- Tucker, T. J., Lumma, W. C., Lewis, S. D., Gardell, S. J., Lucas, B. J., Baskin, E. P., Woltmann, R., Lynch, J. J., Lyle, E. A., Appleby, S. D., Chen, I. W., Dancheck, K. B. & Vacca, J. P. (1997a). Potent noncovalent thrombin inhibitors that utilize the unique amino acid D-dicyclohexylalanine in the P3 position. Implications on oral bioavailability and antithrombotic efficacy. *J. Med. Chem.* **40**, 1565-9.
- Tucker, T. J., Lumma, W. C., Lewis, S. D., Gardell, S. J., Lucas, B. J., Sisko, J. T., Lynch, J. J., Lyle, E. A., Baskin, E. P., Woltmann, R. F., Appleby, S. D., Chen, I. W., Dancheck, K. B., Naylor-Olsen, A. M., Krueger, J. A., Cooper, C. M. & Vacca, J. P. (1997b). Synthesis of a series of potent and orally bioavailable thrombin inhibitors that utilize 3,3-disubstituted propionic acid derivatives in the P3 position. *J. Med. Chem.* **40**, 3687-93.
- Tucker, T. J., Lumma, W. C., Mulichak, A. M., Chen, Z., Naylor-Olsen, A. M., Lewis, S. D., Lucas, R., Freidinger, R. M. & Kuo, L. C. (1997c). Design of highly potent noncovalent thrombin inhibitors that utilize a novel lipophilic binding pocket in the thrombin active site. *J. Med. Chem.* **40**, 830-2.
- Verdonk, M. L., Cole, J. C. & Taylor, R. (1999). SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **289**, 1093-108.
- Verdonk, M. L., Cole, J. C., Watson, P., Gillet, V. & Willett, P. (2001). SuperStar: improved knowledge-based interaction fields for protein binding sites. *J. Mol. Biol.* **307**, 841-59.

- Verner, E., Katz, B. A., Spencer, J. R., Allen, D., Hataye, J., Hruzewicz, W., Hui, H. C., Kolesnikov, A., Li, Y., Luong, C., Martelli, A., Radika, K., Rai, R., She, M., Shrader, W., Sprengeler, P. A., Trapp, S., Wang, J., Young, W. B. & Mackman, R. L. (2001). Development of serine protease inhibitors displaying a multicentered short (2.3 Å) hydrogen bond binding mode: inhibitors of urokinase-type plasminogen activator and factor Xa. *J. Med. Chem.* **44**, 2753-71.
- Wade, R. C., Clark, K. J. & Goodford, P. J. (1993). Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **36**, 140-7.
- Wade, R. C. & Goodford, P. J. (1993). Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **36**, 148-56.
- Wagner, J., Kallen, J., Ehrhardt, C., Evenou, J. P. & Wagner, D. (1998). Rational design, synthesis, and X-ray structure of selective noncovalent thrombin inhibitors. *J. Med. Chem.* **41**, 3664-74.
- Wallace, A. C., Laskowski, R. A., Singh, J. & Thornton, J. M. (1996). Molecular recognition by proteins: protein-ligand interactions from a structural perspective. *Biochem. Soc. Trans.* **24**, 280-4.
- Whitlow, M., Arnaiz, D. O., Buckman, B. O., Davey, D. D., Griedel, B., Guilford, W. J., Koovakkat, S. K., Liang, A., Mohan, R., Phillips, G. B., Seto, M., Shaw, K. J., Xu, W., Zhao, Z., Light, D. R. & Morrissey, M. M. (1999). Crystallographic analysis of potent and selective factor Xa inhibitors complexed to bovine trypsin. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 1395-404.
- Wilcox, R. E., Huang, W. H., Brusniak, M. Y., Wilcox, D. M., Pearlman, R. S., Teeter, M. M., DuRand, C. J., Wiens, B. L. & Neve, K. A. (2000). CoMFA-based prediction of agonist affinities at recombinant wild type versus serine to alanine point mutated D2 dopamine receptors. *J. Med. Chem.* **43**, 3005-19.
- Wilcox, R. E., Tseng, T., Brusniak, M. Y., Ginsburg, B., Pearlman, R. S., Teeter, M., DuRand, C., Starr, S. & Neve, K. A. (1998). CoMFA-based prediction of agonist affinities at recombinant D1 vs D2 dopamine receptors. *J. Med. Chem.* **41**, 4385-99.

- Williams, T. M. (1998). Inhibitors of protein farnesylation 1998. *Exp. Opin. Ther. Patents* **8**, 553-69.
- Williams, T. M. (1999). Inhibitors of protein prenylation 1999. *Exp. Opin. Ther. Patents* **9**, 1263-80.
- Wittinghofer, A. & Waldmann, H. (2000). Ras - ein molekularer Schalter bei der Tumorentstehung. *Angew. Chem.* **112**, 4360-83.
- Wold, S., Esbensen, K. & Geladi, P. (1987). Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **2**, 37-52.
- Wold, S., Ruhe, A., Wold, H. & Dunn III, W. J. (1984). The collinearity problem in linear regression. The partial least squares approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **5**, 735-43.
- Wong, G., Koehler, K. F., Skolnick, P., Gu, Z. Q., Ananthan, S., Schonholzer, P., Hunkeler, W., Zhang, W. & Cook, J. M. (1993). Synthetic and computer-assisted analysis of the structural requirements for selective, high-affinity ligand binding to diazepam-insensitive benzodiazepine receptors. *J. Med. Chem.* **36**, 1820-30.
- Wu, Z., Demma, M., Strickland, C. L., Radisky, E. S., Poulter, C. D., Le, H. V. & Windsor, W. T. (1999). Farnesyl protein transferase: identification of K164 alpha and Y300 beta as catalytic residues by mutagenesis and kinetic studies. *Biochemistry* **38**, 11239-49.
- Yan, J., Roy, S. & Apolloni, A. (1998). Ras isoforms vary in their ability to activate Raf-1 and phosphoinositide 3-kinase. *J. Biol. Chem.* **273**, 24052-6.
- Zeslawska, E., Schweinitz, A., Karcher, A., Sondermann, P., Sperl, S., Stürzebecher, J. & Jacob, U. (2000). Crystals of the urokinase type plasminogen activator variant beta(c)-uPA in complex with small molecule inhibitors open the way towards structure-based drug design. *J. Mol. Biol.* **301**, 465-75.
- Zhang, F. L. & Casey, P. J. (1996). Protein Prenylation: Molecular Mechanism and Functional Consequences. *Annu. Rev. Biochem.* **65**, 241-69.
- Zhao, Z., Arnaiz, D. O., Griedel, B., Sakata, S., Dallas, J. L., Whitlow, M., Trinh, L., Post, J., Liang, A., Morrissey, M. M. & Shaw, K. J. (2000). Design, synthesis, and in vitro biological activity of benzimidazole based factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **10**, 963-6.

Lebenslauf

Markus Böhm

Geburtstag: 11. Mai 1973
Geburtsort: Auerbach / Opf.

1983 – 1992 Gymnasium in Pegnitz
Mai 1992 Allgemeine Hochschulreife

1992 Immatrikulation im Studiengang Diplom-Chemie
an der Universität Regensburg

Oktober 1994 Diplom-Vorexamen
Sept. – Nov. 1996 Forschungspraktikum bei Firma Henkel KGaA, Düsseldorf
März 1997 Diplom-Hauptexamen

1997 – 1998 Anfertigung der Diplomarbeit „*Vergleichende Feldanalysen zur
Abschätzung der Bindungsaffinität von Serinprotease-Inhibitoren*“
unter Anleitung von Herrn Prof. Dr. Gerhard Klebe,
Institut für Pharmazeutische Chemie, Philipps-Universität Marburg

Januar 1998 Verleihung des akademischen Grades „Diplom-Chemiker“
der Universität Regensburg

1998 – 2002 Anfertigung der vorliegenden Dissertation
unter Anleitung von Herrn Prof. Dr. Gerhard Klebe,
Institut für Pharmazeutische Chemie, Philipps-Universität Marburg

1998 – 2001 Wissenschaftlicher Mitarbeiter am selbigen Institut;
Betreuung des Studentenpraktikums „Arzneistoff-Analytik“
(5. Semester Pharmazie)

Erklärung

Ich versichere, dass ich meine Dissertation

*„Entwicklung neuer Deskriptoren für vergleichende molekulare Feldanalysen
und Verfahren zur Beschreibung der Selektivität von Serinproteasen“*

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, im Januar 2002

Markus Böhm