

**Neuartige, empirische Scoring-Modelle für
Protein-Ligand-Komplexe
und
computergestützte Entwicklung von
Hsp70-Inhibitoren**

**Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der
Julius-Maximilians-Universität Würzburg**

vorgelegt von
David Thorsten Zilian
aus Pforzheim

Würzburg, 2014



Eingereicht bei der Fakultät für Chemie und Pharmazie am: _____

Gutachter der schriftlichen Arbeit:

1. Gutachter: _____

2. Gutachter: _____

Prüfer des öffentlichen Promotionskolloquiums:

1. Prüfer: _____

2. Prüfer: _____

3. Prüfer: _____

Datum des öffentlichen Promotionskolloquiums: _____

Doktorurkunde ausgehändigt am: _____

Für Karla und meine Eltern

Publikationen

Die vorliegende Arbeit ist im Zeitraum zwischen Dezember 2009 und Oktober 2013 am Institut für Pharmazie und Lebensmittelchemie an der Julius-Maximilians-Universität Würzburg entstanden. Die Arbeit beschäftigt sich sowohl mit der methodischen Entwicklung als auch mit der praktischen Anwendung von Techniken der computergestützten Wirkstoffentwicklung.

Im Rahmen der Arbeit sind eine Reihe von Publikationen entstanden:

1. M. Hein, D. Zilian und C. A. Sotriffer. „Docking compared to 3D-pharmacophores: The scoring function challenge“. In: *Drug Discovery Today: Technologies*. 3D Pharmacophore Elucidation and Virtual Screening 7.4 (2010), e229–e236
2. B. Merget, D. Zilian, T. Müller und C. A. Sotriffer. „MycPermCheck: the *Mycobacterium tuberculosis* permeability prediction tool for small molecules“. In: *Bioinformatics* 29.1 (2013), S. 62–68
3. D. Zilian und C. A. Sotriffer. „SFCscore^{RF}: A Random Forest-based scoring function for improved affinity prediction of protein–ligand complexes“. In: *Journal of Chemical Information and Modeling* 53.8 (2013), S. 1923–1933

Die Arbeit wurde durch folgende Posterbeiträge auf internationalen Konferenzen präsentiert:

1. Zilian, D. und Sotriffer, C.A.: *Combining SFCscore with Random Forests leads to improved affinity prediction for protein-ligand complexes*. Gemeinsame ÖPhG - DPhG-Jahrestagung, Innsbruck, 2011.
2. Zilian, D. und Sotriffer, C.A.: *Combining SFCscore with Random Forests leads to improved affinity prediction for protein-ligand complexes*. 8th German Conference on Chemoinformatics. Goslar, 2012

Teile des in Referenz [3] veröffentlichten Materials wurden in dieser Dissertation verwendet. Die Verwendung beziehungsweise die Adaption des Materials erfolgt mit Genehmigung der American Chemical Society (©2013). Die folgende Tabelle gibt an, an welchen Stellen der Arbeit und in welchem Ausmaß dieses Material verwendet wurde.

Stelle in der Dissertation	Ausmaß der Verwendung	Ursprünglich publiziert in Ref. [3] auf Seite
Kapitel 5.1, Seite 69 ff.	Text reproduziert, modifiziert und erweitert;	1924
Kapitel 5.2, Seite 71 ff.	Text reproduziert, modifiziert und erweitert; verwendete Abbildungen: 5.2	1925-26
Kapitel 5.3, Seite 74 ff.	Text reproduziert, modifiziert und erweitert; verwendete Abbildungen: 5.3, 5.4, 5.5, 5.6; verwendeten Tabelle: 5.1, 5.2, 5.3, 5.4.	1925-29
Kapitel 6.1, 6.2 und 6.3, Seite 83	Text reproduziert, modifiziert und erweitert	1924, 28-30
Kapitel 6.4 Seite 90 ff.	Text reproduziert, modifiziert und erweitert;	Seite 1925
Kapitel 6.5 und 6.6, Seite 91 ff.	Text reproduziert, modifiziert und erweitert; verwendete Abbildungen: 6.7 a); verwendete Tabellen: 6.2	1930-31
Kapitel 7 Seite 105 ff.	Text reproduziert, modifiziert und erweitert; verwendete Abbildungen: 7.3 g)-i)	1931-32

Danksagung

Die vorliegende Arbeit wäre ohne die Unterstützung, die ich von verschiedensten Seiten erhalten habe, nicht in dieser Form möglich gewesen.

Der größte Dank gebührt in diesem Zusammenhang Prof. Dr. Christoph A. Sotriffer, der mir die Möglichkeit gegeben hat, mich mit den interessanten Fragestellungen des computergestützten Wirkstoffdesigns zu befassen und diese Arbeit in seinem Arbeitskreis anzufertigen. In meiner gesamten Promotionszeit stand er mir konstruktiven Ratschlägen und guten Ideen zur Seite und hat einen großen Anteil am Gelingen dieser Arbeit. Ich bedanke mich deswegen für die fachlich wie menschlich hervorragende Betreuung.

Frau Prof. Dr. Ulrike Holzgrabe und Herrn Prof. Dr. Thomas Dandekar möchte ich für die Übernahme der Aufgaben des Zweit- bzw. Drittprüfer dieser Arbeit danken.

Der Klinischen Forschergruppe 216 (KFO 216) danke ich für die finanzielle Unterstützung und die hervorragende interdisziplinäre Zusammenarbeit. Besonders bedanken möchte ich mich in diesem Zusammenhang bei Frau Prof. Dr. Ulrike Holzgrabe und Dr. Andreas Hartung für die organische Synthese der Tetrahydroisochinolinon-Derivate, bei Prof. Dr. Ralf Bargou und Dr. Manik Chatterjee sowie ihren Mitarbeiterinnen Elisabeth Müller und Heike Schraut für die biologischen Untersuchungen sowie bei Dr. Clemens Grimm für die Lösung der Kristallstrukturen von bHsp70.

Am Institut für Pharmazie und Lebensmittelchemie danke ich Frau Möhler und Frau Ebner (für die Unterstützung in den bürokratischen Untiefen der Universitätsverwaltung), Christine Schneider (für die Hilfe bei der Organisation des 1. Semesters), Dr. Bernd Reyer (für die Unterstützung bei vielen computerbezogenen Problemen und Projekten) sowie der Werkstatt (für die Unterstützung bei verschiedensten Dingen).

Allen, die mit mir gemeinsam das 1. Semester betreut haben, möchte ich für die gute Zusammenarbeit und die gemeinsame Zeit danken. Ganz besonders möchte ich mich in diesem Zusammenhang bei Annette Albrecht bedanken.

Ein besonderer Dank gilt auch meinen Kollegen aus dem Arbeitskreis Sotriffer: Dr. Daniel Cappel, Dr. Martin Sippel, Dr. Monika Nocker, Dr. Benjamin Schaefer, Dr. Armin Welker, Dr. Christine Topf, Dr. Constanze Waltenberger, Ulrich Peinz, Michael Hein, Manuel Krug, Steffen Wagner, Benjamin Merget, Thomas Willmes, Raphael Dives, Yogesh Narkhede, Daniel Schuster, Sarah Wehle und Max Kuhn. Es war mir immer eine Freude mit euch zu arbeiten. Die vielen gemeinsamen Momente auf und außerhalb der Arbeit (von denen hoffentlich auch noch einige folgen werden), machen meine Promotionszeit für mich unvergesslich.

Meiner Familie, besonders meinen Eltern Meike und Norbert Zilian und meiner Freundin Karla, möchte ich für die immerwährende und bedingungslose Unterstützung in allen Belangen danken.

Zu guter Letzt möchte ich mich bei allen bedanken, die diese Arbeit mit Interesse lesen und hoffe, dass sich das Interesse nicht auf diese Danksagung beschränkt.

David Zilian, im August 2014

Inhaltsverzeichnis

1. Einleitung	15
1.1. Scoring-Funktionen	16
1.2. Hsp70	20
I. Entwicklung von neuartigen empirischen Scoring-Modellen zur Vorhersage von Bindungsaffinität von Protein-Ligand-Komplexen	25
2. Einleitung	27
2.1. Bindungsaffinität und Protein-Ligand-Wechselwirkungen	27
2.2. SFCscore - empirische Scoring-Funktionen zur Vorhersage von Bindungsaffinitäten	29
2.3. Evaluierung von SFCscore	31
3. Erstellung eines Trainingsdatensatzes mit ausgewogener Affinitätsverteilung	39
3.1. Zusammenstellung des Datensatzes	39
3.2. Variablen-Selektion und Ableitung der Funktionen	41
3.3. Ergebnisse	44
3.4. Ausblick	45
4. Klassifizierungsmodelle	47
4.1. Trainingsdaten	47
4.2. Generierung der Bindeposen	48
4.3. Generierung von Trainings- und Testdatensätzen	56
4.4. Ableitung der Modelle	59
4.5. Ergebnisse und Auswertung	61
4.6. Zusammenfassung und Diskussion	66
5. SFCscore^{RF} -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten	69
5.1. Trainingsdatensatz	69
5.2. Ableitung des RandomForest-Modells	71

5.3. Scoring-Performance	74
6. CSAR 2012 Exercise	83
6.1. Datensätze	83
6.2. Docking-Validierung	85
6.3. Scoring-Validierung	89
6.4. CSAR-Docking	90
6.5. Scoring-Performance	91
6.6. Erkennung von inaktiven Substanzen (Ranking-Performance)	93
6.7. CSAR 2012 Exercise: Ergebnisse aller Teilnehmer	94
6.8. Einfluss der Docking-Posen	95
6.9. Statistische Validität	99
6.10. Zusammenfassung	103
7. Applicability Domain	105
7.1. PCA-Dichtekarten	106
7.2. Numerische Ansätze	110
7.3. RandomForest-spezifische Parameter	112
7.4. Diskussion und Ausblick	113
II. Entwicklung von Inhibitoren und strukturdynamische Untersuchungen der Hsp70 Chaperon-Familie	115
8. Grundlagen	117
8.1. Hsp70-Struktur	117
8.2. Funktioneller Zyklus	119
8.3. Allosterische Zustände	119
8.4. Hsp70 als Target	121
9. Vorarbeiten, Kooperationen und Zielsetzung	123
9.1. Initiales virtuelles Screening	123
9.2. Biologische Testung	125
9.3. Synthese	125
9.4. Kristallstruktur	125
9.5. Zielsetzung dieser Arbeit	127
10. Weiterentwicklung der Leitstruktur C066-0757	129
10.1. Docking des Liganden C066-0757	131
10.2. Strukturbasiertes Design von Derivaten	140
10.3. Zusammenfassung und Diskussion	149

11. Molekulardynamische Simulationen	153
11.1. Einleitung	153
11.2. Methode und Durchführung	154
12. Ergebnisse: Molekulardynamische Simulationen	159
12.1. Dynamik des Gesamtsystems	159
12.2. Domänenbewegung	164
12.3. Dynamik der Interface-Bindetasche	167
12.4. Mutations-Region	180
12.5. Linker	184
12.6. Nukleotid-Bindedomäne	184
12.7. Substrat-Bindedomäne	191
12.8. Zusammenfassende Diskussion	193
III. Zusammenfassung/Summary	197
13. Zusammenfassung	199
14. Summary	203
IV. Anhang	207

1. Einleitung

Das computergestützte Wirkstoffdesign hat sich in den vergangenen Jahren zu einem wichtigen Bestandteil in der pharmazeutischen Forschung entwickelt. Computermodelle finden sowohl bei der initialen Suche nach biologisch aktiven Substanzen (virtuelles Screening) als auch bei der Optimierung von Leitstrukturen Verwendung. Die verwendeten Methoden lassen sich in strukturbasierte und in ligandbasierte Ansätze einteilen.

Dem ligandbasierten Ansatz liegt eine Bibliothek bereits bekannter aktiver Verbindungen zugrunde. Häufig verwendete Techniken aus dem ligandbasierten Bereich sind QSAR-Methoden oder auch ligandbasierte Pharmakophore-Modelle. Eine genaue Kenntnis der biologischen Zielstruktur ist bei ligandbasierten Methoden nicht nötig.

Bei einem strukturbasierten Ansatz steht das Zielprotein hingegen im Mittelpunkt, da die zentralen Techniken des strukturbasierten Wirkstoffdesigns, das Docking und das Scoring, ein genaues Modell der zu adressierenden Bindetasche voraussetzen.

Der erste Teil dieser Arbeit beschäftigt sich mit der Entwicklung von Scoring-Funktionen. Um deren Relevanz beim virtuellen Screening (VS) besser einordnen zu können, wird im Folgenden der generelle Workflow eines strukturbasierten VS skizziert, wobei sich die Darstellung an den Lehrbüchern von Klebe [4] und Sottriffer [5] orientiert.

Virtuelles Screening Ein VS kann man sich als hierarchischen Prozess vorstellen, bei dem große Moleküldatenbanken mit zunehmend genaueren Methoden durchsucht werden, um eine letztendlich geringe Anzahl von potentiell aktiven Substanzen zu erhalten. Am Anfang werden häufig eher grobe Kriterien, wie beispielsweise physikochemische Parameter, verwendet, um die Datenbank vorzufiltern. Die Filtermethoden werden dann sukzessive spezifischer. Welche Methoden verwendet werden, hängt stark vom konkreten Szenario, also beispielsweise den verfügbaren Informationen zum Target und zu aktiven Verbindungen und dem gewünschten Filtereffekt, ab.

Oftmals werden Pharmakophormodelle verwendet, mit denen große Datenbanken schnell nach spezifischen Merkmalen durchsucht werden können. Ein Beispiel für ein proteinspezifisches Filterwerkzeug ist MycPermCheck [2], welches die Wahrscheinlichkeit berechnet, dass eine Verbindung die Zellwand von *Mycobacterium tuberculosis* permeieren kann.

Das Ziel dieser Filterschritte ist, die Verbindungsdatenbank soweit zu reduzieren, dass der Einsatz von Docking und Scoring-Methoden möglich wird.

Beim Docking geht es darum, den Bindemodus eines Liganden in der Bindetasche vorherzusagen. Eine Docking-Methode besteht aus einem Algorithmus zur Generierung

von möglichen Posen des Liganden in der Bindetasche und einer Scoring-Funktion zur Bewertung und Auswahl der günstigsten Posen. Der Begriff "Pose" bezeichnet dabei die Konfiguration, die Konformation sowie die relative Orientierung des Liganden in der Bindetasche. Zur Generierung möglicher Posen können unterschiedliche Methoden wie das Multikonformer-Docking (z.B. im Programm FRED [6, 7]), der inkrementelle Aufbau in der Bindetasche (FlexX [8]) oder stochastische Suchmethoden (GOLD [9], AutoDock [10]) eingesetzt werden. Alle drei Klassen von Suchalgorithmen sind in der Lage, gute Ergebnisse bei der Vorhersage von Bindeposen zu liefern [11]. Eine detailliertere Einführung zu Docking-Algorithmen findet sich beispielsweise bei Rognan [12].

Auch auf Seiten der Scoring-Funktionen stehen drei Klassen zur Verfügung. Bei Kraftfeld-basierten Scoring-Funktionen wird die Affinität mit Hilfe der nicht-kovalenten Terme von molekularmechanischen Kraftfeldern abgeschätzt. Wissensbasierte Funktionen basieren hingegen auf Atom-Atom-Kontakten und werden anhand von großen Kristallstrukturdatenbanken parametrisiert. Die dritte Gruppe bilden die empirischen Scoring-Funktionen. Hier basiert die Affinitätsvorhersage auf der (meist linearen) Kombination von Deskriptoren. Die Deskriptoren beschreiben einzelne Charakteristika der Ligandbindung, die auf Basis eines gegebenen Protein-Ligand-Komplexes berechnet werden. Die Parametrisierung von empirischen Scoring-Funktionen erfolgt ebenfalls mittels Kristallstrukturdatenbanken, wobei zusätzlich ein experimenteller Affinitätswert benötigt wird. Ein guter Überblick zum Thema Scoring-Funktionen findet sich bei Sotriffer und Matter [13].

Die vorliegende Arbeit beschäftigt sich mit der Entwicklung empirischer Scoring-Funktionen. Konkrete Grundlage der Arbeit ist das Programm SFCscore, das in Kapitel 2.2 vorgestellt wird.

1.1. Scoring-Funktionen

Scoring-Funktionen werden nicht ausschließlich zum Vergleich von Docking-Posen benötigt, sondern auch zur Vorhersage der potentiellen Aktivität einer Verbindung eingesetzt. Die so erstellte Ranglisten (*Rankings*) sind eine wichtige Entscheidungsgrundlage für die Auswahl der Verbindungen, die experimentellen Testungen unterzogen werden sollen. Scoring-Funktionen stellen also eine zentrale Technik im strukturbasierten virtuellen Screening dar und sind für dessen Erfolg von entscheidender Bedeutung.

Leistungsstand von Scoring-Funktionen Bei der Beurteilung des Leistungsstandes von Scoring-Funktionen sollten die verschiedenen Einsatzgebiete getrennt voneinander betrachtet werden. Konkret bewertet werden die Docking-Performance, also die Fähigkeit korrekte Bindeposen zu erkennen; die Ranking-Performance, die angibt, wie gut eine Funktion Verbindungen anhand ihrer biologischen Aktivität ordnen kann; sowie die Scoring-Performance, die aussagt, wie präzise eine Funktion die biologische Aktivität vorhersagen kann. Im folgenden werden vor allem die Ergebnisse der Studien

von Cheng et al. [14] und von Warren et al. [11] zitiert, die den Leistungsstand von Scoring-Funktionen zu Beginn der hier vorliegenden Arbeit sehr gut widerspiegeln.

Die Docking-Performance von Scoring-Funktionen wird meist durch die Erfolgsrate beim Auffinden von kristallnahen Posen in einem Datensatz, der mit falschen Posen (*geometric decoys*) angereichert ist, bestimmt. Als "kristallnah" werden in der Regel Posen mit einer RMS-Abweichung von $<2 \text{ \AA}$ zur Kristallstruktur angesehen. Warren et al. erhielten für sieben target-spezifische Datensätze sehr stark schwankenden Ergebnisse, die sowohl von der Zielstruktur als auch von der verwendeten Docking/Scoring-Kombination abhängig waren. Die Erfolgsquoten bewegten sich zwischen 0 und 90% [11]. Cheng et al. haben einen großen, heterogen zusammengesetzten Datensatz von 195 Kristallstrukturen verwendet, der je Kristallstruktur mit bis zu 100 *geometric decoys* angereichert wurde. Für die 14 getesteten Scoring-Funktionen wurden Erfolgsquoten von 50 bis 90 % ermittelt [14]. Zusammengefasst ist die Docking-Performance von Scoring-Funktionen also keineswegs immer zufriedenstellend. Durch die Verwendung einer passenden Docking/Scoring-Kombination lässt sich jedoch in den meisten Fällen ein erfolgreiches Docking durchführen.

Die Ranking Performance ist eine sehr wichtige Eigenschaft, wenn es um die Auswahl von Verbindungen für experimentelle Testungen geht. Für den praktischen Einsatz im virtuellen Screening kommt es dabei nicht darauf an, ob die ersten 10 Plätze in genau der richtigen Reihenfolge vorhergesagt werden, als vielmehr darauf, im oberen Bereich der Rangliste möglichst viele aktive Verbindungen anzureichern und inaktive Verbindungen auszusortieren. Dies wird auch als "Enrichment" bezeichnet. Eine sinnvolle Einschätzung des *Enrichments* kann man nur aus Datensätzen erhalten, die inaktive Verbindungen enthalten, wie beispielsweise der von Warren verwendete. Hier konnte für jede Targetklasse mindestens eine Methode gefunden werden, die ein *Enrichment* von > 5 innerhalb der ersten 10% der Rangliste generiert (d.h. in den ersten 10% der Rangliste finden sich anteilmäßig 5-mal mehr aktive Verbindungen als im ganzen Datensatz). Die Performance war auch hier stark vom Target bzw. von der verwendeten Methode abhängig. Keine Methode konnte für alle Targetklassen gute Ergebnisse erzielen. Überraschenderweise konnte kein Zusammenhang zwischen guten Enrichment-Faktoren und guten Ergebnissen beim Docking beobachtet werden [11].

Die Vorhersage von Bindungsaffinitäten (Scoring-Performance) kann als Königsdisziplin für Scoring-Funktionen angesehen werden. Da durch die Verbesserung der Affinitätsvorhersage implizit die beiden anderen Bereiche mit verbessert werden, wird sie oftmals als Hauptparameter bei der Entwicklung neuer Funktionen herangezogen. Klassischerweise wird der Pearson-Korrelationskoeffizient (R_P) zwischen vorhergesagter und experimentell ermittelter Affinität als Kriterium für die Scoring-Performance verwendet. Zusätzlich zum Pearson-Koeffizienten werden häufig noch die Korrelationskoeffizienten nach Spearman (ρ oder R_s) und Kendall (τ) angegeben. Im Unterschied zur Pearson-Korrelation werden dabei nicht die numerischen Werte korreliert, sondern deren relative Rangfolge. Somit sind ρ und τ nicht von der Art der Korrelation abhängig, während der Pearson-Koeffizient einen linearen Zusammenhang voraussetzt. Der Unter-

schied zwischen ρ und τ liegt in der Bestrafung von falsch eingeordneten Datenpunkten. Während dies bei τ immer mit dem Wert 1 bestraft werden, wird bei der Berechnung von ρ das Ausmaß der Falscheinordnung berücksichtigt. Ein weiterer Parameter zur Abschätzung der Scoring-Performance ist der *Root Mean Square Error* (RMSE), also der durchschnittliche Fehler der Vorhersagen.

Cheng et al. erhielten für die beste der getesteten Funktionen eine Pearson-Korrelation von 0.64. Die Vorhersagefehler bewegten sich zwischen 1.8 und 2.3 pK_i -Einheiten [14]. Warren et al. ermittelten als besten Wert in ihren target-spezifischen Datensätzen eine Korrelation von 0.57. Allerdings ergaben sich für die meisten der untersuchten Targets überhaupt keine statistisch signifikanten Korrelationen [11].

Die Performance von SFCscore, das im Mittelpunkt dieser Arbeit steht, wird in Kapitel 2.3 detailliert vorgestellt. Dort wird man sehen, dass SFCscore bezüglich der Vorhersage von Bindungsaffinitäten, bereits zu den besten Funktionen gehört. Die Ergebnisse machen aber auch deutlich, dass eine dringende Notwendigkeit für Neu- und Weiterentwicklungen besteht, gerade wenn man die Relevanz eines akkuraten Scorings für den Erfolg des kompletten virtuellen Screenings berücksichtigt. Ziel dieser Arbeit ist die Verbesserung der Scoring- und der Ranking-Performance durch die Entwicklung neuer SFCscore-Funktionen.

Entwicklung von empirischen Scoring-Funktionen Bei empirischen Scoring-Funktionen gibt es drei Teilbereiche, die Ansatzpunkte für Neuentwicklungen bieten:

- die Datengrundlage (Trainingsdatensätze)
- die Deskriptoren
- die statistische Methode zur Ableitung einer Funktion.

Trainingsdatensätze Die Trainingsdaten sind die Grundlage jeder empirischen Scoring-Funktion und haben einen wesentlichen Einfluss auf die Qualität der Funktion. Dabei spielt sowohl die Qualität, also die Genauigkeit der experimentellen Daten, als auch die Größe der verwendeten Datensätze eine Rolle. Durch die immer größer werdende Zahl an Kristallstrukturen und Affinitätsdaten ergibt sich beinahe automatisch die Möglichkeit bessere Funktionen abzuleiten. Gerade vor diesem Hintergrund spielt aber auch die Ausgewogenheit der Datensätze, im Bezug auf die repräsentierten Targetklassen den abgedeckten Affinitätsbereich, eine wichtige Rolle.

Der erste praktische Teil dieser Arbeit (Kapitel 3) beschäftigt sich mit der Zusammenstellung eines neuen Datensatzes, der vor allem im Hinblick auf den abgedeckten Affinitätsbereich, eine größere Ausgewogenheit aufweist. Der Ansatz zielt in erster Linie auf die Schwäche von Scoring-Funktionen in der Anwendbarkeit für besonders hoch- oder niedrigaffine Verbindungen ab.

Die Arbeiten zur "Applicability Domain", die in Kapitel 7 beschrieben werden, befassen sich noch grundsätzlicher mit diesem Themenkomplex. Mittels der "Applicability Domain" soll der chemische Raum, der von einem bestimmten Datensatz beziehungsweise einer Scoring-Funktion abgedeckt wird, beschrieben werden. Daraus könnte man einerseits eine Abschätzung über die Vorhersagegenauigkeit einer Scoring-Funktion ableiten. Andererseits kann eine solche Methode beim gezielten Design von Trainingsdatensätzen helfen.

Deskriptoren Die Herausforderung bei der Entwicklung neuer Deskriptoren für das empirische Scoring besteht darin, die für die Protein-Ligand-Bindung wichtigen Beiträge mit möglichst geringem Rechenaufwand möglichst genau zu beschreiben und zu quantifizieren. Sowohl die Entwicklung neuer Deskriptoren als auch die Verbesserung bestehender Deskriptoren ist ein Feld für potentielle Neuentwicklung. Diese Arbeit befasst sich jedoch nicht mit der Entwicklung von Deskriptoren.

Ableitung der Funktionen Zur Ableitung von empirischen Scoring-Funktionen werden klassischerweise Methoden wie *Partial Least Square* (PLS) oder die *Multiple Linear Regression* (MLR) verwendet, die gut interpretierbare, lineare Modelle liefern. Dem stehen jedoch einige Nachteile gegenüber, beispielsweise die fehlende Robustheit bei der Verwendung von korrelierten oder unwichtigen Deskriptoren. Hinzu kommt, dass die Parametrisierung solcher Funktionen durch die zusätzlich meist notwendige Variablenselektion aufwendig ist und deswegen mitunter nicht zu einem optimalen Ergebnis führt.

Mit den Methoden aus dem Bereich des Machine Learning stehen eine ganze Reihe alternativer Verfahren zur Verfügung, die auch schon erfolgreich im Bereich des virtuellen Screenings und des Wirkstoffdesigns eingesetzt wurden. Einen guten Überblick über Methoden und Einsatzmöglichkeiten, mit direktem Bezug zum virtuellen Screening, findet sich bei Melville et al. [15]. Als Beispiel sei hier der *Support Vector Machine*-Algorithmus genannt, der von Jorissen und Gilson [16] oder von Li et al. [17] zur Entwicklung neuartiger Scoring-Funktionen verwendet wurde. Auch andere Methoden wie *Bayes'sche Klassifikatoren*, *Neuronale Netzwerke* und der *RandomForest*-Algorithmus wurden bereits erfolgreich in VS-Szenarien eingesetzt [15]. In dieser Arbeit wurde der *RandomForest*-Algorithmus, der auf Seite 59 ausführlich vorgestellt wird, verwendet, um verschiedene neue Scoring-Modelle abzuleiten.

Im ersten Schritt wurde der Algorithmus zur Ableitung eines target-spezifischen Klassifizierungsmodells eingesetzt. Dadurch war es möglich, sowohl aktive als auch inaktive Verbindungen zum Training zu verwenden. Das Ziel des Ansatzes war die verbesserte Erkennung von inaktiven Verbindungen. Die Analyse der Modelle und der Vergleich mit klassischen Scoring-Funktionen zeigt, dass dies erreicht wurde und dass ein solches Verfahren einen interessanten Ausgangspunkt zur Entwicklung proteinspezifischer Funktionen darstellt. Die Arbeiten zu den Klassifizierungsmodellen werden in Kapitel 4 beschrieben.

Der RandomForest-Algorithmus kann auch zur Erstellung von regressionsbasierten Modellen verwendet werden. In Kombination mit einem deutlich vergrößerten Trainingsdatensatz wurde in dieser Arbeit eine neue Scoring-Funktion zur Vorhersage von Bindungsaffinitäten entwickelt, die eine deutlich verbesserte Scoring-Performance zeigt und auch im internationalen Vergleich zu den besten Funktionen zählt. Die Ableitung und ausführliche Evaluierung dieser Funktion, die den größten Teil der hier durchgeführten Arbeiten ausmachen, sind in Kapitel 5 und 6 beschrieben.

1.2. Hsp70

Neben der methodischen Forschung umfasst die vorliegende Arbeit auch die praktische Anwendung von strukturbasierten Methoden zur Entwicklung neuer Inhibitoren für Chaperone der Hsp70-Familie.

“Heat-Shock”-Proteine 70 (Hsp70) sind eine Klasse molekularer Chaperone, die ihren Namen aufgrund ihres Molekulargewichts von etwa 70 kDa und ihrer Stress-Induzierbarkeit (beispielsweise durch einen Hitzeschock) tragen. Neben den stress-induzierten Formen existieren auch konstitutiv exprimierte Isoformen. Es sind mindestens 13 Isoformen bekannt, die in unterschiedlichen Zellkompartimenten und durch die Interaktion mit verschiedenen Co-Chaperonen spezifische Funktionen in der Zelle ausüben [18]. Die folgenden allgemeinen Grundlagen gelten für alle Isoformen. Der Begriff Hsp70 wird als Überbegriff für die Familie der Hsp70-Chaperone verwendet.

Zelluläre Funktion und Pathobiologie Hsp70-Chaperone sind an einer Vielzahl von zellulären Prozessen, vor allem der Faltung von neu synthetisierten Proteinen und der Verhinderung von Proteinaggregation, beteiligt. Die Funktionalität wird über ATP-kontrollierte Bindung von hydrophoben Peptiden ermöglicht [19, 20]. Es wird dabei eine sehr große Anzahl unterschiedlicher Substrate akzeptiert. Schätzungen besagen, dass Hsp70-Chaperone an der *de novo* Faltung von über 20% aller humanen Proteine beteiligt sind [20].

Die spezifische Funktionalität wird einerseits durch spezialisierte Isoformen, andererseits durch die Rekrutierung verschiedener Co-Chaperone und die Interaktion mit anderen Chaperonen und Klientenproteinen ermöglicht. Die genauen molekularen Mechanismen dieses komplexen Zusammenspiels sind nur bruchstückhaft bekannt.

Neben der reinen Faltung von Proteinsubstraten finden sich auch immer mehr Hinweise, dass viele für die Zellzykluskontrolle und die intrazelluläre Signaltransduktion benötigte Proteine durch die Bildung eines Komplexes mit Hsp70 in ihrer biologischen Aktivität beeinflusst werden. Beispiele hierfür sind nukleäre Rezeptoren (z.B. Steroidhormon-Rezeptoren), Kinasen (z.B. Ras/Raf) und Transkriptionsfaktoren (z.B. HSF-1) [20]. Vor diesem Hintergrund ist es nicht überraschend, dass Hsp70 mit einer Vielzahl von zellpathologischen Vorgängen wie Onkogenese [21], Neurodegeneration

[22, 23], Autoimmunerkrankungen [24] sowie viralen Infektionen [25] in Verbindung gebracht wird.

Hohe Expressionsraten von Hsp70 werden mit verschiedenen Krebsarten wie Brustkrebs, Prostatakrebs oder kolorektalen Karzinomen sowie verschiedenen Arten von Leukämie in Verbindung gebracht [26]. In dieser Arbeit steht die Rolle von Hsp70 bei der Entstehung und Therapie des Multiplen Myeloms im Mittelpunkt.

Multiplres Myelom Das multiple Myelom (MM) ist eine maligne Erkrankung, die zur Gruppe der B-Zell-Lymphome zählt. Die malignen B-Zellen sind vornehmlich im Knochenmark lokalisiert und produzieren nicht-funktionelle Antikörper oder Antikörperbruchstücke. Durch die Ablagerung der überschüssigen Antikörper und das entartete Wachstum der Plasmazellen kommen Symptome zustande, die von Knochenschmerzen über die teilweise Auflösung von Knochenstrukturen und Organfunktionsstörungen (oftmals der Niere) bis hin zum kompletten Organversagen führen können [27]. Weitere häufig auftretende Symptome sind Anämie-bedingte Müdigkeit, Infektneigung und Gewichtsverlust [28]. Das MM ist die dritthäufigste hämatologische Neoplasie und macht damit etwa 1% der Krebserkrankungen in Deutschland aus [27].

Die Pathogenese des MM beginnt mit der sogenannten MGUS (*Monoclonal Gammopathy of Uncertain Significance*), einer Vorstufe, die noch symptomfrei ist und in der Altersgruppe zwischen dem 45. und 75. Lebensjahr eine Prävalenz von 3.5% aufweist [29]. Durch unterschiedliche chromosomale Mutationen kann sich aus MGUS das klinische Bild des multiplen Myeloms entwickeln. Das Risiko liegt bei etwa 1% pro Jahr [30]. Das entartete Wachstum und die übermäßige Antikörperproduktion kommen durch die Aktivierung verschiedener zellulärer Signalkaskaden (z.B. PI3K/Akt, JAK/Stat, Raf/MEK/MAPK, NF κ B und Wnt) zustande, so dass die Pathogenese des MM als sehr heterogen anzusehen ist [31]. Die Aufklärung und Charakterisierung der onkogenen Signalwege und, darauf aufbauend, die Entwicklung von personalisierten Therapieschemata ist ein zentrales Thema der KFO 216, in deren Rahmen die vorliegende Arbeit entstanden ist.

Die allogene Stammzelltransplantation in Kombination mit einer Chemotherapie ist die Methode der Wahl bei der Therapie des MM. Auf Seiten der Chemotherapeutika konnten durch die Anwendung von Wirkstoffen wie Thalidomid, Lenalidomid und Bortezomib in den vergangenen Jahren einige Fortschritte erzielt werden [27]. Lenalidomid und Thalidomid gehören zur Gruppe der immunmodulierenden Substanzen und entfalten ihre Wirkung über verschiedene Wirkmechanismen, wie der Antiangiogenese, der Induktion von Caspase-8-vermittelter Apoptose und der Induktion einer antitumoralen Immunantwort [32, 33].

Bortezomib, ein Peptid-Analogon, ist ein Inhibitor des 26S-Proteasoms. Das 26S-Proteasom ist ein großer Proteinkomplex, der am Abbau von Proteinen beteiligt ist. Die Inhibition führt zu einer Beeinflussung verschiedener regulatorischer Proteine und Transkriptionsfaktoren. Dadurch werden Zellwachstum, Angiogenese und Metastasierung ver-

hindert [34].

Diese Wirkstoffe werden sowohl in der Induktions- als auch in der Erhaltungstherapie eingesetzt, meist in Kombination mit Glucocorticoiden (Prednison oder Dexamethason) und/oder Alkylantien (Melphalan, Cyclophosphamid). Bortezomib und Lenalidomid werden außerdem bei der Therapie eines Rezidivs erfolgreich eingesetzt. Obwohl initial ein gutes Ansprechen auf die MM-Therapie beobachtet wird, ist die Rezidivrate sehr hoch und geht oftmals mit einer Resistenzbildung einher. Es besteht daher ein dringender Bedarf nach neuen Therapieoptionen. Die heterogene Pathogenese sowie die Involvierung verschiedener zellulärer Regulationskaskaden legen eine Multitarget-Therapie nahe, bei der mehrere onkogene Mechanismen gleichzeitig attackiert werden.

Hsp70 als Zielstruktur für die Therapie des MM Hsp70 ist aufgrund der oben genannten Beteiligung an verschiedenen zellzyklus-regulierenden Signalwegen, die zu den onkogenen Kaskaden des MM zählen, ein sehr interessantes Zielprotein für die Entwicklung neuer Wirkstoffe.

Dieser Theorie folgend konnte bereits gezeigt werden, dass Hsp70 in MM-Zellen beziehungsweise allgemein in Tumorzellen stark überexprimiert ist und sowohl mit einer erhöhten Tumorüberlebensrate als auch mit einer schlechteren Prognose für den Patienten assoziiert ist [35, 36, 37, 38]. Ähnliche Beobachtungen konnten auch für die eng verwandte Gruppe der Hsp90-Chaperone gemacht werden [35]. Weiterhin konnte im Rahmen der KFO gezeigt werden, dass Hsp70 als Ausweichsystem Teile der Aufgaben von Hsp90 übernimmt, wenn Hsp90 inhibiert wird. Das lässt auf einen synergistischen Effekt beim parallelen Einsatz von Hsp70 und Hsp90-Inhibitoren schließen. Zudem gibt es experimentelle Hinweise, die Hsp70 mit der Entwicklung von Therapie-Resistenzen in Verbindung bringen [39, 40, 41, 42].

Strukturbiologie Hsp70 ist ein Monomer, welches aus zwei großen Domänen aufgebaut ist, die über ein Linker-Segment miteinander verbunden sind. Die beiden Domänen sind für die Bindung von Peptidsubstraten bzw. von ATP verantwortlich. Das Protein folgt dabei einem komplexen funktionellen Zyklus, in dessen Verlauf einerseits ATP gebunden, hydrolysiert und als ADP freigesetzt wird und andererseits hydrophobe Peptide gebunden beziehungsweise freigesetzt werden [43]. Die genauen Mechanismen dieses Zyklus sind nicht abschließend geklärt. Als gesichert gilt jedoch, dass das Protein in verschiedenen allosteren Zuständen vorliegt, die sich auch wesentlich durch die relative Stellung der beiden großen Domänen zueinander unterscheiden. Die Kontrolle dieses Zyklus verläuft zum einen durch die Bindung von Co-Chaperonen, zum anderen aber auch über eine direkte Kommunikation zwischen den beiden Domänen [43, 19]. Die Struktur und der funktionelle Zyklus werden in Kapitel 8 detailliert vorgestellt.

Aus Sicht der Wirkstoffentwicklung bietet sich auf den ersten Blick eine Adressierung der ATP- oder Substratbindetasche an. In beiden Fällen ist jedoch mit Selektivitätsproblemen zu rechnen, die eine gezielte Inhibition von Hsp70 erschweren würden.

Stattdessen wird in dieser Arbeit ein Ansatz verfolgt, bei dem gezielt die Kommunikation zwischen den Hsp70-Domänen durch die Adressierung der Interface-Region inhibiert werden soll.

1.2.1. Strukturbasiertes Design von Inhibitoren

Die hier durchgeführten Arbeiten beruhen zu großen Teilen auf den Vorarbeiten von Ulrich Peinz (AK Sotriffer), der eine potentielle Bindetasche in der Interface-Region identifiziert hat und in einem virtuellen Screening bereits eine Leitstruktur gefunden hat, die in dieser Region binden kann (vgl. Kapitel 9.1). Die in dieser Arbeit durchgeführten Untersuchungen lassen sich in zwei Themenkomplexe aufteilen.

Optimierung der Leitstruktur C066-0757 Der erste Teil beschäftigt sich mit der Weiterentwicklung der von Ulrich Peinz gefundenen Leitstruktur. Dazu wurde deren potentieller Bindemodus in der Interface-Bindetasche mit Hilfe von Docking- und Scoring-Methoden detailliert analysiert. Vor allem die Stereochemie der Verbindung stellte dabei eine Herausforderung dar. Mit Hilfe dieser Informationen wurde von Kooperationspartnern eine Substanzbibliothek erstellt und biologisch untersucht. Im Kontext der so erhaltenen Aktivitätsdaten wurden mit Hilfe von Dockinganalysen die potentiellen Bindemoden verifiziert bzw. falsifiziert sowie erste Struktur-Wirkungsbeziehungen abgeleitet.

Molekulardynamische Simulationen Eine große Herausforderung bei der Entwicklung neuer Wirkstoffe stellt die Flexibilität von Hsp70 dar. Wie weiter oben beschrieben, ist der funktionelle Zyklus des Proteins vor allem durch die relative Dynamik der beiden großen Domänen gekennzeichnet ist. Diese Dynamik wirkt sich auch auf die Topologie der Interface-Region und die Konformation der Bindetasche aus. Mit Hilfe von molekulardynamische Simulationen (MD-Simulationen) sollte dies genauer untersucht werden. Bei der Analyse der Trajektorien stand dementsprechend die Dynamik der Interface-Bindetasche im Mittelpunkt. Vor allem sollte untersucht werden, inwieweit die für das Docking verwendete Bindetaschenkonformation stabil ist und ob es alternative Konformationen gibt, die dann als Ausgangspunkt für weitere virtuelle Screenings dienen können.

Mögliche Kommunikationswege, die über die Interface-Region verlaufen und somit einen Ansatzpunkt für die Erklärung des Inhibitionsmechanismus darstellen könnten, wurden ebenfalls analysiert. Aufgrund der komplexen zellulären Aktivität von Hsp70 und einigen Einschränkungen, die aus der verwendeten Kristallstruktur resultieren, müssen diese Ergebnisse jedoch vorsichtig interpretiert werden. Eine dieser Einschränkungen ergibt sich aus zwei Punktmutationen im Interface-Bereich. Die MD-Trajektorien wurden deswegen auch hinsichtlich eines kritischen Einflusses dieser Mutationen auf die Dynamik des Gesamtsystems untersucht.

Teil I.

**Entwicklung von neuartigen
empirischen Scoring-Modellen zur
Vorhersage von Bindungsaffinität
von Protein-Ligand-Komplexen**

2. Einleitung

2.1. Bindungsaffinität und Protein-Ligand-Wechselwirkungen

Im Folgenden werden die wesentlichen Grundlagen zur Beschreibung der Bindungsaffinität von Protein-Ligand-Komplexen kurz dargestellt. Ausführliche Informationen finden sich in den Lehrbüchern von Klebe [4] sowie von Böhm und Schneider [44], an denen sich auch die folgende Darstellung orientiert.

Die zentrale Aufgabe von Scoring-Funktionen ist die Vorhersage der Bindungsaffinität eines Liganden (L) an ein Zielprotein (P). Für die nicht-kovalente, reversible Bindung gilt das Gleichgewicht:



Die Bindungsaffinität wird durch die Dissoziationskonstante K_D beschrieben. Diese lässt sich aus dem Massenwirkungsgesetz ableiten (Gl. 2.2).

$$K_D = \frac{[P_{aq}][L_{aq}]}{[PL_{aq}]} \quad (2.2)$$

Durch die thermodynamische Beziehung

$$\Delta G^0 = RT \cdot \ln K_D \quad (2.3)$$

lässt sich aus der Gleichgewichtskonstante die freie Bindungsenthalpie (ΔG^0) berechnen. Dabei ist R die allgemeine Gaskonstante und T die absolute Temperatur in Kelvin.

Die Dissoziationskonstante K_D ist allgemeingültig für die Bindungsaffinität zwischen einem Protein und einem Liganden jedweder Art. In dieser Arbeit steht die Affinität zwischen Enzymen und Inhibitoren im Mittelpunkt, die durch den K_i -Wert beschrieben wird. Experimentell wird der K_i -Wert oftmals in Gegenwart von kompetitiven Substraten gemessen und kann daher nicht immer mit dem K_D -Wert gleichgesetzt werden. Experimentell wird anstelle des K_i -Wertes oftmals der einfacher zu messende IC_{50} -Wert bestimmt. Dieser gibt die Inhibitorkonzentration an, bei der noch 50 % Enzymaktivität vorhanden sind. Im Unterschied zum K_i -Wert ist er aber keine Konstante und beispielsweise von der Enzymkonzentration abhängig. IC_{50} -Messungen aus verschiedenen Assays oder Laboren sind deswegen nur eingeschränkt vergleichbar.

2. Einleitung

Energetisch betrachtet, setzt sich ΔG aus einem enthalpischen und einem entropischen Term zusammen (Gleichung 2.4).

$$\Delta G = \Delta H - T\Delta S \quad (2.4)$$

Hier steht H für die Enthalpie, und S für die Entropie.

Der enthalpische Beitrag beschreibt, die beim Bindungsprozess umgewandelte Energie (ΔU), korrigiert um den Anteil der Volumenarbeit. Je mehr Energie durch die Bildung des Protein-Ligand-Komplexes frei wird, umso günstiger wirkt sich das auf die Bindungsaffinität aus. Die Haupttriebfeder für den enthalpischen Teil der Bindungsaffinität sind direkte Interaktionen zwischen Protein und Ligand. Die wichtigsten Beispiele für solche Interaktionen sind Wasserstoffbrücken, ionische Wechselwirkungen ("Salzbrücken"), polare Interaktionen sowie Metall-Komplexierungen. Entscheidend für den enthalpischen Beitrag sind jedoch nicht die absoluten Energien der einzelnen Interaktionen, sondern die Energieänderung im Vergleich zum Ausgangszustand, in dem Protein und Ligand jeweils solvatisiert in wässriger Lösung vorliegen. Gerade polare Gruppen bilden häufig auch mit Wasser sehr günstige Interaktionen, so dass die bei der die Bildung einer Protein-Ligand-Interaktion frei werdende Energie ganz oder teilweise kompensiert wird ("Desolvatationseffekte").

Ein weiterer wichtiger Interaktionstyp sind hydrophobe Wechselwirkungen, die durch die Ausbildung von hydrophoben Kontaktflächen zwischen Protein und Ligand zustande kommen. Der enthalpische Energiegewinn dieses Vorgangs erklärt sich vor allem daraus, dass Wassermoleküle aus den hydrophoben Bereichen freigesetzt werden und Wasserstoffbrücken zu anderen Wassermolekülen ausbilden können. Durch die Freisetzung der geordneten Wassermoleküle kommt es außerdem zu einem Entropiegewinn, der ebenfalls einen positiven Einfluss auf die Bindungsaffinität hat. Dies bezeichnet man auch als "hydrophoben Effekt".

Entropische Faktoren haben jedoch auch einen negativen Einfluss auf die Bindungsaffinität, da durch die Bildung eines Protein-Ligand-Komplexes immer die Translations- und Rotationsfreiheitsgrade der Bindungspartner eingeschränkt werden. Die entropischen Verluste, die dabei entstehen, müssen durch die enthalpischen Beiträge und den hydrophoben Effekt ausgeglichen beziehungsweise überkompensiert werden, um eine gute Bindungsaffinität zu erreichen. Neben der Einschränkung der Translation und globalen Rotation, die bei jeder Komplexbildung auftritt, haben auch die Freiheitsgrade von einzelnen Aminosäuren beziehungsweise die Einschränkung der Rotation von Einfachbindungen des Liganden einen Einfluss auf den entropischen Term. Der entropische Beitrag zur Bindung ist deswegen mitunter schwer zu quantifizieren.

Die Bildung eines Protein-Ligand-Komplexes ist aus thermodynamischer Sicht ein sehr komplexer Vorgang. Ausschlaggebend für die Bildung des Komplexes kann sowohl der enthalpische als auch der entropische Term sein. Man unterscheidet daher auch zwischen enthalpie-getriebener und entropie-getriebener Bindung.

2.2. SFCscore - empirische Scoring-Funktionen zur Vorhersage von Bindungsaffinitäten

Der Grundgedanke von empirischen Scoring-Funktionen ist es, die Bindungsaffinität in chemisch quantifizierbare Einzelbeiträge zu zerlegen. Dabei geht man zunächst von sogenannten Deskriptoren aus, die einzelne strukturelle Aspekte eines Komplexes numerisch beschreiben. Die Scoring-Funktion ist eine Kombination von mehreren, gewichteten Deskriptoren. Die Parametrisierung einer empirischen Funktion erfolgt anhand eines Trainingsdatensatzes, mit dessen Hilfe eine möglichst optimale Deskriptorenkombination und die Gewichtungsfaktoren ermittelt werden.

Der Ausgangspunkt der in dieser Arbeit durchgeführten Entwicklungen ist das Programm SFCscore, das (zu Beginn der Arbeit) acht empirische Scoring-Funktionen sowie die Algorithmen zur Berechnung von 66 Deskriptoren umfasste. Die 66 Deskriptoren bilden die Grundlage für die bereits existierenden Funktionen, können aber auch zur Entwicklung neuer Scoring-Modelle verwendet werden.

In den folgenden Abschnitten werden die wichtigsten Punkte bei der Entwicklung der SFCscore-Funktionen erläutert. Dies umfasst die verwendeten Trainingsdatensätze, die SFCscore-Deskriptoren sowie die Ableitungsmethoden. Daran schließt sich ein Überblick über den Leistungsstand von SFCscore an.

SFC-Datensätze Die für SFCscore verwendeten Daten stammen aus zwei Quellen. Zum einen wurden öffentlich verfügbare Daten aus der AffinDB [45] verwendet, zum anderen wurden über das *Scoring Function Consortium* (SFC) kristallographische Daten von verschiedenen pharmazeutischen Unternehmen zur Entwicklung von Scoring-Funktionen zur Verfügung gestellt. Die Daten aus der Industrie waren teilweise verblendet, d.h. es standen nur die berechneten Deskriptoren, nicht aber die zugrunde liegenden Koordinaten zur Verfügung. Insgesamt wurden 565 Komplexe aus industriellen Quellen sowie 290 Komplexe aus der PDB verwendet. Der SFCscore-Datensatz umfasste somit 855 Komplexe. Genauere Ausführungen zu Datensammlung, den zugrunde liegenden Auswahlkriterien sowie technische Details finden sich bei Sotriffer et al. [46].

Aus dieser Datengrundlage wurden insgesamt sechs Datensätze zum Training von SFCscore-Funktionen abgeleitet. Der größte Trainingsdatensatz umfasst dabei alle 855 zur Verfügung stehenden Komplexe. Ein weiterer Datensatz enthält die 290 Komplexe aus öffentlichen Quellen.

Ein zweiter öffentlicher Datensatz wurde mit Hilfe strengerer Filterkriterien aus dem 290-Datensatz rekrutiert. Dieser reduzierte Datensatz enthielt 229 Komplexe. Die öffentlich verfügbaren Daten haben den Vorteil, dass die Affinitätsdaten ausschließlich aus pK_i -Werten bestehen. Außerdem sind hier bei allen Komplexen die entsprechenden Kristallstrukturen verfügbar, was bei der Analyse und Weiterentwicklung der Modelle sehr hilfreich ist. Wiederum ausgehend vom 290-Datensatz wurden durch gezieltes Hinzufügen von Komplexen aus den industriellen Daten weitere spezialisierte

Datensätze erzeugt: Zum einen der *met*-Datensatz, der mit Metalloproteasen angereichert ist, und zum anderen der *ser*-Datensatz, bei dem Serinproteasen hinzugefügt wurden. Der letzte Datensatz wurde durch Filtern des *855*-Datensatzes nach Molekülgröße generiert. Die 130 Komplexe mit einer Größe < 300 Da bilden den *frag*-Datensatz. Weitere Details zur Auswahl und Zusammenstellung der Daten finden sich wiederum bei Sottriffer et al. [46].

Deskriptoren Die Deskriptoren bilden das Herzstück von SFCscore. Insgesamt können bis zu 66 Deskriptoren berechnet werden, die auch separat ausgegeben und beispielsweise zur Entwicklung eigener Funktionen verwendet werden können. Die von SFCscore berechneten Deskriptoren lassen sich grob in drei Gruppen einteilen.

Zum einen werden rein ligandspezifische Deskriptoren wie das Molekulargewicht des Liganden (*MW*), die Anzahl der Schweratome (*NAtoms*) oder die Anzahl der rotierbaren Bindungen (*n_rot*) berechnet.

Die zweite Gruppe besteht aus Deskriptoren für spezifische Wechselwirkungen. Dazu gehören Wasserstoffbrücken (*NHBonds*, *c_hb*, *n_hb* und *HBScore*), Metall-Komplexierungen (*met*, *RMScore*) und Interaktionen mit aromatischen Ringsystemen (*RRScore*, *RMScore*, *RKScore*).

Die dritte und weitaus größte Gruppe kann man unter dem Schlagwort "Oberflächendeskriptoren" zusammenfassen. Zu deren Berechnung werden die Oberflächen von Ligand und Protein in "POLAR", "HYDROPHOB" und "AROMATIC" eingeteilt. Die Oberfläche des Liganden wird außerdem noch in "buried" (in der Bindetasche vergraben) und "exposed" (lösungsmittel-exponiert) unterteilt. Basierend auf dieser Einteilung lassen sich dann Kontaktoberflächen berechnen, die einzeln oder kombiniert als Deskriptoren verwendet werden. Ein Beispiel dafür ist der Deskriptor *SURFC_PP*, der die polare Kontaktfläche zwischen Protein und Ligand angibt. Auf diese Weise kann man auch Strafterme generieren, beispielsweise gibt *PH_HP_surfc* den Kontakt zwischen polaren und hydrophoben Oberflächen an. Zusätzlich werden aus den Oberflächentermen auch Quotienten berechnet, welche die Oberflächen zueinander in Verhältnis setzen. Beispielsweise gibt der Deskriptor *PH_HP_surfc2TotLigSurf* den Anteil der Ligandoberfläche an, der unvorteilhaft in der Bindetasche vergraben ist, wohingegen *HHAAPP_surfc2TotLigSurf* den optimal vergrabenen Oberflächenanteil widerspiegelt.

Mit Hilfe dieser Oberflächendeskriptoren ist SFCscore in der Lage, die Bindungscharakteristiken eines Protein-Ligand-Komplexes sehr genau zu beschreiben. Der Schwerpunkt liegt dabei auf der Beschreibung der enthalpischen Beiträge, die durch die Oberflächenkomplementarität und gerichtete Wechselwirkungen entstehen. Entropisch getriebene Bindungsbeiträge wie der hydrophobe Effekt lassen sich über die Oberflächendeskriptoren ebenfalls abschätzen.

Weitere Details zur Parametrisierung und Berechnung der einzelnen Deskriptoren können in der SFCscore-Publikation nachgelesen werden [46].

Tabelle 2.1.: Überblick über die ursprünglichen SFCscore-Funktionen [46].

Funktion	Datensatzgröße	Ableitungsmethode	Deskriptoren (Komponenten)
sfc_290m	290	MLR	7
sfc_229m	229	MLR	7
sfc_frag	130	MLR	4
sfc_855	855	PLS	27 (6)
sfc_ser	466	PLS	23 (4)
sfc_met	341	PLS	22 (4)
sfc_290p	290	PLS	19 (5)
sfc_229p	229	PLS	24 (6)

Ableitung der Funktionen Zur Ableitung der Funktionen wurden die multivariaten Verfahren MLR und PLS verwendet, die auf Seite 41 genauer beschrieben sind.

Die Funktionen sind jeweils nach dem Datensatz benannt, auf dem sie basieren. Für den 229- und 290-Datensatz existiert sowohl eine PLS als auch eine MLR-Funktion. Dies wird durch den Buchstaben *m* bzw. *p* im Namen der Funktion angegeben. In Tabelle 2.1 sind die acht Funktionen mit Namen, Datensatzgröße, Ableitungsmethode und der Anzahl der verwendeten Deskriptoren aufgeführt. Im Fall der PLS-Funktionen ist zusätzlich angegeben, wie viele PCA-Komponenten für die Regression verwendet wurden.

Auch für die Details zur praktischen Durchführung bei der Ableitung der Funktionen und der Variablenselektion wird auf die originale Publikation verwiesen [46].

2.3. Evaluierung von SFCscore

Die erste Untersuchung, die im Rahmen der vorliegenden Arbeit durchgeführt wurde, war die Evaluierung der ursprünglichen SFCscore-Funktionen.

Bis zu diesem Zeitpunkt stammten die einzigen Anhaltspunkte zu deren Leistungsstand aus der SFCscore-Publikation selbst [46]. Hier wurde die Performance für den PDBbind-Datensatz von Wang et al. [47, 48] bestimmt, der seinerzeit einer der umfassendsten Benchmark-Datensätze war. Basierend auf der PDBbind-Version 2002 war von Wang und Mitarbeitern ein Test-Datensatz von 800 Komplexen erstellt worden, gegen den dann 14 weit verbreitete Scoring-Funktionen getestet wurden. Die beste Funktion war *X-Score::HMScore* mit einer Korrelation von 0.57 [49].

Die SFCscore-Funktionen wurden gegen einen Datensatz getestet, der dem von Wang et al. sehr ähnlich ist. Die Ergebnisse sollen hier nicht im Einzelnen wiedergegeben werden, da sie in der SFCscore-Publikation detailliert beschrieben sind. Qualitativ kann man aber zusammenfassen, dass die Funktionen Korrelationswerte von 0.523 (*sfc_frag*) bis 0.585 (*sfc_met*) ergaben. SFCscore ist den von Wang getesteten Funktionen also leicht überlegen [46].

Um das Leistungsniveau von SFCscore weiter zu charakterisieren, wurden alle verfügbaren Funktionen gegen den Benchmark-Datensatz von Cheng et al. [14] sowie gegen den *CSAR 2010*-Datensatz [50] getestet.

Benchmark-Datensätze: Cheng und CSAR-2010 Der *Cheng*-Datensatz besteht insgesamt aus 195 Komplexen, die aus der PDBbind Datenbank ausgewählt wurden [14]. Bei der Zusammenstellung des Datensatzes wurde die PDBbind (Version 2009) zunächst anhand der Proteinklassen geclustert. Aus allen Clustern, die mehr als 3 Mitglieder hatten, wurden dann genau 3 Komplexe basierend auf ihrer experimentellen Affinität ausgewählt. Es wurden jeweils der Vertreter mit der höchsten und der niedrigsten Affinität ausgewählt sowie derjenige, dessen Aktivität am nächsten am Mittelwert des Clusters lag. Der 195 Komplexe umfassende Datensatz rekrutiert sich aus 65 Dreier-Gruppen. Der Datensatz weist aufgrund der Auswahlmethode eine große Diversität bezüglich der Proteinstrukturen und eine große Variabilität bezüglich der Affinitäten auf. Er wird an viele Stellen der vorliegenden Arbeit zur Evaluierung verwendet und im Folgenden als *Cheng*-Benchmark bezeichnet.

Auch der *CSAR 2010*-Datensatz stellt eine Zusammenstellung von hoch-qualitativen Kristallstrukturen mit verfügbaren Affinitätsdaten dar. Der Datensatz umfasst insgesamt 343 Komplexe. Aus den in [50] genannten Gründen werden bei der Auswertungen jedoch nur 332 Komplexe berücksichtigt. Die Daten wurden aus akademischen sowie aus industriellen Quellen rekrutiert. Bei der Auswahl der Komplexe wurde Wert auf Diversität und Qualität der Strukturen sowie auf die Abdeckung eines möglichst großen Affinitätsspektrums gelegt [50].

Scoring-Performance In der Publikationen von Cheng et al. wurden insgesamt 14 Scoring-Funktionen getestet. Die beste Funktion, *X-Score::HMScore*, erreichte eine Pearson-Korrelation 0.64. Insgesamt wurde Werte zwischen 0.22 und 0.64 gemessen [14].

Im Vergleich dazu zeigen die SFCscore-Funktionen sehr gute Ergebnisse (Tabelle 2.2). Alle Funktionen würden im Vergleich mit den Cheng-Funktionen unter den Top 3 rangieren. Die beste Funktion, *sfc_met*, liegt mit einer Pearson-Korrelation von 0.64 gleichauf mit der besten von Cheng getesteten Funktion. Auch die Korrelationen der anderen SFCscore-Funktionen sind mit Werten zwischen 0.58 und 0.62 zufriedenstellend und liegen im selben Bereich, der schon in der originalen SFCscore-Publikation für den Wang-Datensatz gemessen wurde.

Tabelle 2.2.: Korrelationen und Vorhersagefehler von SFCscore für den *Cheng*-Datensatz (195 Komplexe). R_P : Pearson Korrelationskoeffizient; ρ : Spearman Korrelationskoeffizient; τ : Kendall Korrelationskoeffizient; RMSE: Root Mean Square Error; MedError: Median der Residuen.

Function	R_P	ρ	τ	RMSE	MedError
sfc_229m	0.59	0.64	0.46	1.93	1.17
sfc_229p	0.59	0.63	0.45	1.96	1.16
sfc_290m	0.59	0.65	0.46	1.95	1.16
sfc_290p	0.62	0.66	0.48	1.88	1.03
sfc_855	0.60	0.64	0.45	1.92	1.15
sfc_frag	0.58	0.64	0.46	2.69	1.57
sfc_met	0.64	0.68	0.50	1.84	1.09
sfc_ser	0.59	0.63	0.45	1.97	1.14

Die Scoring-Ergebnisse für den *CSAR 2010*-Datensatz (Tabelle 2.3) bestätigen die Ergebnisse des *Cheng*-Benchmarks. Auch hier liegen die Pearson-Korrelationen zwischen 0.58 und 0.62. Auch wenn in diesem Fall die interne Rangfolge der Funktionen etwas anders ist (sfc_ser ist hier die beste Funktion), werden die Werte der *Cheng*-Benchmarks hier insgesamt bestätigt.

Im Vergleich zu den anderen Funktionen, die mit dem *CSAR 2010*-Datensatz getestet wurden, liegt SFCscore allerdings eher im Mittelfeld. Die beiden besten Funktionen in der *CSAR*-Studie erreichen Pearson-Korrelationen von über 0.7 [51]. Allerdings werden die Ergebnisse in der Studie nur verblindet aufgeführt, so dass nicht klar ist, welche Ansätze dort erfolgreich waren.

Als generelle Einschränkung muss an dieser Stelle noch auf die Problematik der Überlappung zwischen Trainings- und Testdaten hingewiesen werden, die zu überoptimistischen Einschätzungen führen kann. Alle SFCscore-Datensätze enthalten potentiell Überlappungen mit den Testdaten. Würde man dies korrigieren, würde darunter aber wiederum die Vergleichbarkeit leiden, weil jede Funktion mit einem anderen Datensatz getestet werden würde.

Diese Einschränkung gilt auch für die in der *Cheng*- und der *CSAR*-Studie getesteten Funktionen. Dies unterstreicht die Notwendigkeit von allgemein anerkannten Benchmark-Datensätzen, die eine objektive Evaluierung und Vergleichbarkeit von Scoring-Funktionen ermöglichen.

Alle im Rahmen dieser Arbeit entwickelten Funktionen weisen keine Überlappungen zwischen Test- und Trainingsdaten auf, da die geplanten Testdaten von vornherein aus den Trainingsdatensätzen entfernt wurden.

Tabelle 2.3.: Korrelationen und Vorhersagefehler von SFCscore für den CSAR 2010-Datensatz (332 Komplexe). R : Pearson Korrelationskoeffizient; ρ : Spearman Korrelationskoeffizient; τ : Kendall Korrelationskoeffizient; RMSE: Root Mean Square Error; MedError: Median der Residuen.

Function	R	ρ	τ	RMSE	MedError
sfc_229m	0.58	0.58	0.41	1.86	1.21
sfc_229p	0.58	0.56	0.40	1.86	1.28
sfc_290m	0.58	0.57	0.41	1.84	1.21
sfc_290p	0.58	0.58	0.41	1.84	1.20
sfc_855	0.59	0.59	0.42	1.78	1.10
sfc_frag	0.62	0.60	0.43	2.61	1.57
sfc_met	0.60	0.59	0.42	1.79	1.12
sfc_ser	0.62	0.62	0.44	1.76	1.17

Docking-Performance Im Unterschied zur Scoring-Performance kann die Docking-Performance von SFCscore nicht mit den besten Funktionen aus anderen Studien mithalten. In Abbildung 2.1 sind die Erfolgsraten von SFCscore bei der Identifizierung von kristallnahen Posen im *Cheng Docking*-Datensatz dargestellt. In diesem Datensatz wurde jede der 195 Kristallstrukturen mit bis zu 100 Docking-Posen angereichert. Eine "kristallnahe" Pose ist mit einem RMSD-Wert von $< 2 \text{ \AA}$ zur Kristallstruktur definiert.

Während die besten Funktionen aus der Cheng-Studie Erfolgsraten von bis zu 90% erreichen [14], liegen die SFCscore-Funktionen im Bereich von 50 bis 60%. Betrachtet man nur die beste Pose, so erreicht keine SFCscore-Funktion die 50%-Marke. Im internen SFCscore-Vergleich sind interessanterweise die drei MLR-basierten Funktionen (sfc_frag, sfc_229m und sfc_290m) am erfolgreichsten. Die größere Anzahl an Deskriptoren, die für die PLS-Funktionen verwendet werden, bringt also offensichtlich keinen Vorteil bei der Vorhersage von Bindeposen.

Insgesamt ist die schlechtere Performance von SFCscore beim Docking jedoch nicht überraschend. Im Deskriptorensatz von SFCscore fehlen einige Terme, die gerade beim Vergleich von Bindeposen eine entscheidende Rolle spielen, beispielsweise ein Strafterm für interne Spannungsenergien der Liganden. Bei empirischen Funktionen beobachtet man häufig, dass sie für den Einsatz im Docking weniger geeignet sind. Die besten Funktionen aus der Cheng-Studie (*Gold:ASP*, *DrugScore^{PDB}*, *DS:PLP1*) sind beispielsweise alle wissensbasiert [14]. Da empirische Funktionen ausschließlich mit Kristallstrukturen, also mit nahezu optimalen Bindungsgeometrien, trainiert werden, kennen sie den Fall einer nicht optimierten Bindepose nicht. Michael Hein (AK Sotriffer) hat in seiner Doktorarbeit untersucht wie sich die Docking-Performance von SFCscore verbessern lässt [52].

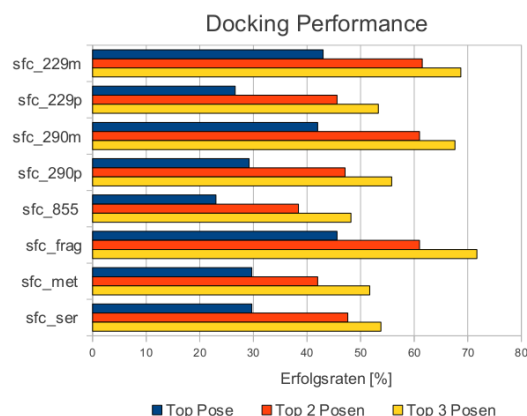


Abbildung 2.1.: Docking-Performance der SFCscore-Funktionen für den Docking-Benchmark Datensatz von Cheng et al. [14]

Ranking-Performance Die Ranking-Performance lässt sich anhand des *Cheng*- oder des *CSAR 2010*-Datensatzes quantitativ schwer evaluieren, da beide Datensätze keine inaktiven Verbindungen enthalten. Bei der praktischen Anwendung von Scoring-Funktionen im virtuellen Screening geht es aber vor allen Dingen darum, aktive und inaktive Verbindungen zu unterscheiden. Der in der *Cheng*-Studie verwendete Parameter für die Ranking-Performance (richtiges Sortieren der 65 Dreiergruppen, die jeweils aus aktiven Verbindungen bestehen) trifft deswegen nicht den Grundgedanken der Ranking-Performance.

Deutlich bessere Verfahren zur Bewertung der Ranking-Performance sind beispielsweise *Receiver-Operating-Characteristic*-Kurven (ROC-Kurven) [53] oder *Enrichment-Faktoren* [54]. Das Problem der Bewertung der Ranking-Performance ist jedoch nicht trivial, und es stehen noch eine Reihe weiterer Metriken zur Verfügung. Bei Truchon und Bayly findet sich ein Überblick sowie eine kritische Betrachtung der verfügbaren Methoden [55].

Für SFCscore können an dieser Stelle nur qualitative Aussagen zur Ranking-Performance gemacht werden, die auf einer Analyse der vorhergesagten Scores beruhen. Am wichtigsten ist dabei die Erkenntnis, dass alle SFCscore-Funktionen dazu tendieren, Werte im mittleren Affinitätsbereich (zwischen pK_i 5 und 8) vorherzusagen. In Abbildung 2.2 ist dies beispielhaft für die Funktion *sfc_229m* dargestellt.

Dementsprechend werden beim Scoring vor allem Komplexe mit sehr hohen oder sehr niedrigen Affinitäten nicht korrekt erfasst. Auch die praktische Erfahrung zeigt, dass die Separierung von aktiven und inaktiven Verbindungen zu den Schwächen von SFCscore, aber auch von Scoring-Funktionen im Allgemeinen, gehört.

Zielsetzung dieser Arbeit Die Zielsetzung dieser Arbeit ist die Verbesserung der Ranking- sowie der Scoring-Performance durch die Entwicklung neuer Scoring-Modelle.

Der Ansatzpunkt für die Verbesserung der Ranking-Performance war dabei die im

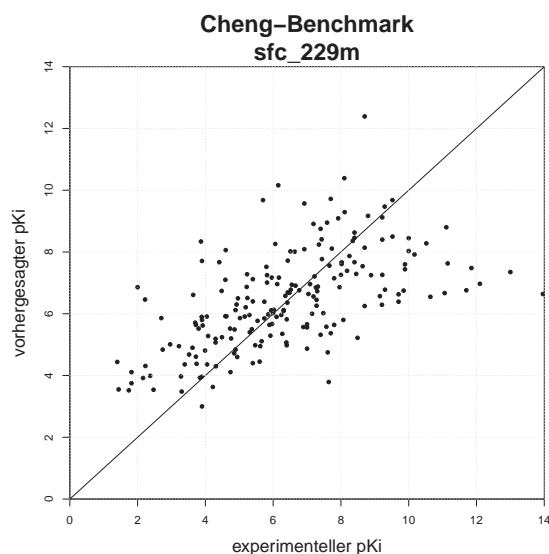


Abbildung 2.2.: Vorhergesagte gegen experimentelle Affinitätswerte der Funktion `sfc_229m` für den *Cheng*-Benchmark Datensatz

vorigen Kapitel beschriebene Tendenz von SFCscore, Affinitätswerte im mittleren pK_i -Bereich vorherzusagen, was sich vor allem auf die Unterscheidung von aktiven und inaktiven Verbindungen auswirkt.

Obwohl die Gründe für diese Schwächen sicherlich vielfältig sind, spielt die Zusammensetzung der Trainingsdatensätze eine wichtige Rolle. Hier spiegelt sich zum einen die Affinitätsverteilung der SFCscore-Vorhersagen, mit dem Schwerpunkt im mittleren Affinitätsbereich wider. Zudem enthielten die bisherigen Trainingsdatensätze keine inaktiven Verbindungen, weil die verwendeten Methoden und die verfügbaren Daten dies nicht zuließen. Die Verbesserung der Trainingsdatensätze unter diesem Gesichtspunkten stellt eine vielversprechende Strategie für die Entwicklung von Scoring-Funktionen mit besserer Ranking-Performance dar.

In einem ersten Ansatz wurde deswegen ein neuer Trainingsdatensatz zusammengestellt, der einen weiten Affinitätsbereich gleichmäßig abdeckt. Das Vorgehen und die Ergebnisse dieses Versuchs sind in Kapitel 3 beschrieben.

Um zusätzlich in der Lage zu sein auch inaktive Verbindungen für das Training zu verwenden, mussten zusätzlich neue Methoden zur Ableitung der Funktionen gefunden werden. Im Bereich des Machine Learning gibt es eine Reihe von Algorithmen, die dazu in der Lage sind. In dieser Arbeit wurde der RandomForest-Algorithmus verwendet, um die ersten SFCscore-basierten Klassifizierungsmodelle für die Unterscheidung von aktiven und inaktiven Verbindungen abzuleiten. Diese Arbeiten werden in Kapitel 4 vorgestellt.

Während diese Ansätze hauptsächlich auf eine Verbesserung der Ranking-Performance abzielen, beschäftigt sich der Hauptteil der Arbeit mit der Entwicklung einer neuen Funktion mit verbesserter Scoring-Performance. Der Einsatz des RandomForest-Algorithmus

in Kombination mit einem deutlich vergrößerten Trainingsdatensatz führte zur Ableitung einer neuartigen SFCscore-Funktion. Diese wurde sehr ausführlich mit verschiedenen Benchmark-Datensätzen getestet. Die Arbeiten dazu sind in den Kapiteln 5 und 6 beschrieben.

Ein wiederkehrendes Thema dieser Arbeit ist das Verhältnis zwischen den Trainingsdaten und Vorhersagen einer Funktion. In Kapitel 7 wird dieses Thema im Rahmen der "Applicability Domain" (AD) nochmals aufgegriffen. Die AD einer Funktion soll den chemischen Raum ihrer Trainingsdaten beschreiben und somit beispielsweise eine Abschätzung über die Verlässlichkeit von Vorhersagen ermöglichen. Die in Kapitel 7 vorgestellten Arbeiten sind dabei als Grundlage für weitere Entwicklungen zu sehen.

3. Erstellung eines Trainingsdatensatzes mit ausgewogener Affinitätsverteilung

3.1. Zusammenstellung des Datensatzes

Als Ausgangspunkt für die Erstellung des Datensatzes wurde der 229-Trainingsdatensatz gewählt. Dieser Datensatz bildete die Grundlage für die Ableitung der beiden SFCscore-Funktionen *sfc_229m* und *sfc_229p* und besteht ausschließlich aus öffentlich zugänglichen Komplexkristallstrukturen. Die Analyse der Affinitätsverteilung zeigt in diesem Datensatz eine Verteilung mit Schwerpunkt im pK_i -Bereich zwischen 5-7 (Abbildung 3.1 (a)). Um das Ziel eines möglichst gleichverteilten Datensatzes zu erreichen, mussten also in den sehr hohen und sehr niedrigen pK_i -Regionen Komplexe hinzugefügt werden. Als Datenquelle hierfür diente der *PDBbind refined*-Datensatz (Version 2007) [48, 47].

Da der *Cheng*-Datensatz zur Testung der neuen Funktionen verwendet werden sollte, wurden die entsprechenden Komplexe bei der Zusammenstellung nicht berücksichtigt. Auch alle Komplexe aus dem originalen 229-Trainingsdatensatz, die auch im *Cheng*-Datensatz vorkommen, wurden entfernt, so dass es keinerlei Überlappung zwischen den Test- und den Trainingsdaten gab.

Der verbliebene *PDBbind refined*-Datensatz wurde basierend auf den experimentellen Affinitätswerten in Gruppen eingeteilt. Jede Gruppe umfasste dabei eine pK_i -Einheit zwischen 0 und 11. Ausnahme ist die Gruppe zwischen 0 und 2, die sich über 2 pK_i -Einheiten erstreckt. In diesem Affinitätsbereich konnten insgesamt 20 Komplexe gefunden werden, die alle in den neuen Trainingsdatensatz übernommen wurden. Um nun eine Gleichverteilung zu erreichen, wurden für jede der neun weiteren Gruppen ebenfalls genau 20 Komplexe ausgewählt.

Für die Auswahl dieser Komplexe wurden folgenden Kriterien verwendet: Zunächst sollte durch die Auswahl der Komplexe versucht werden, die Korrelation zwischen Molekulargewicht und Bindungsaffinität zu minimieren. Diese Korrelation ist ein weit verbreitetes Problem v.a. von empirischen Scoring-Funktionen und führt zu trivialen Funktionen, deren Vorhersage hauptsächlich auf der Größe des Liganden beruht. Das zweite Kriterium war die Diversität der Protein-Targets. Um eine Überrepräsentation

3. Erstellung eines Trainingsdatensatzes mit ausgewogener Affinitätsverteilung

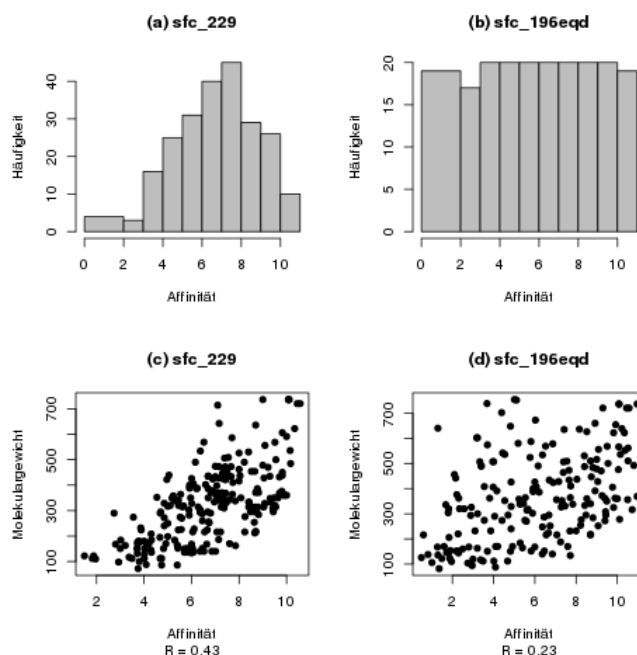


Abbildung 3.1.: Zusammenstellung des *196_eqd* Datensatzes. Die Abbildungen (a) und (b) zeigen die Verteilungen der Affinitätswerte im pK_i Bereich von 0-11. Während für den *sfc.229* Datensatz (a) der Schwerpunkt der Daten im Bereich zwischen 5-8 liegt, ist beim *196_eqd* Datensatz eine annähernde Gleichverteilung zu sehen. Die Abbildungen (c) und (d) zeigen die Beziehung zwischen dem Molekulargewicht und der Affinität in den beiden Datensätzen.

einzelner Targetklassen zu vermeiden, von denen es eine große Anzahl an Kristallstrukturen gibt und die häufig auch mit einer hohen Ähnlichkeit zwischen den Liganden einhergeht, wurde darauf geachtet, möglichst unterschiedliche Targets auszuwählen.

Basierend auf diesen Prinzipien wurde der Datensatz manuell zusammengestellt. Es sei darauf hingewiesen, dass durch die manuelle Auswahl sicherlich keines der Kriterien optimal erfüllt wurde, wie es vielleicht durch den Einsatz eines maschinellen Algorithmus möglich wäre. Trotzdem konnte ein Datensatz mit den genannten Schwerpunkten generiert werden.

Abbildung 3.1 stellt die Situation bezüglich Affinitätsverteilung und Affinitäts-Größen-Korrelation im Ausgangsdatsatz und im neuen Datensatz dar. Wie man sieht, konnte neben der Gleichverteilung über ein großes Affinitätsspektrum auch eine Verminderung der Pearson-Korrelation zwischen Molekulargewicht und Affinität von 0.43 auf 0.23 erreicht werden.

Der neue Datensatz umfasste ursprünglich genau 200 Komplexe. Im weiteren Verlauf der Arbeit mussten jedoch nochmals vier Komplexe gestrichen werden, da sie entweder

einen Fehler in SFCscore verursachten oder die Kristallstruktur sich als fehlerhaft erwies, so dass der finale Datensatz 196 Komplexe umfasste. Der Datensatz wird im Folgenden als "eqd_196" bezeichnet, wobei "196" für die Anzahl der Komplexe und "eqd" für "equal distribution" steht.

3.2. Variablen-Selektion und Ableitung der Funktionen

Zur Ableitung der neuen Scoring-Funktion wurden die Methoden PLS und MLR benutzt, die auch schon bei den originalen SFCscore-Funktionen zum Einsatz kamen. Es handelt sich um multivariate Methoden zur Ableitung von linearen Regressionsmodellen. Beiden Methoden werden im Folgenden kurz beschrieben. Weitergehende Informationen finden sich beispielsweise bei Bortz [56] oder Hartung [57], die auch zur Ausarbeitung der folgenden Zusammenfassung verwendet wurden.

Methode: MLR Im Unterschied zur klassischen linearen Regression hängt die Zielvariable bei der multiplen Regression von mehreren, unabhängigen Prediktorvariablen ab. Das resultierende Modell hat die Form:

$$y = b_1 * x_1 + b_2 * x_2 + \dots + b_z \quad (3.1)$$

Dabei ist y die Zielvariable, $x_{1,2,\dots}$ sind die unabhängigen Variablen, $b_{1,2,\dots}$ sind die Regressionskoeffizienten und b_z ist ein konstanter Term, der oftmals als y-Achsenabschnitt bezeichnet wird.

Die Aufgabe der Regressionsanalyse ist die Abschätzung der b -Werte. Dies geschieht sowohl bei der einfachen als auch bei der multiplen linearen Regression durch Minimierung der Residuen, also der Differenz zwischen der Regressionsgerade und dem tatsächlichen Wert eines Messpunktes bezogen auf die Zielvariable. Die Methode wird auch als *Methode der kleinsten Fehlerquadrate* bezeichnet.

Ein wichtiges Qualitätskriterium einer linearen Regression ist das Bestimmtheitsmaß (R^2). Dieses gibt den Anteil der durch die unabhängigen Variablen (also vom Regressionsmodell) erklärten Varianz der Zielvariable an und ist somit definiert als

$$R^2 = \frac{\text{Varianz der Regressionswerte}}{\text{Varianz der Zielvariablen}} \quad (3.2)$$

Das Bestimmtheitsmaß ist das Quadrat des Pearson-Korrelationskoeffizienten R_P .

R_P beziehungsweise R^2 nehmen durch die Verwendung von zusätzlichen unabhängigen Variablen bei der multiplen Regression immer zu. Um Modelle mit einer unterschiedlich großen Zahl von Prediktorvariablen zu vergleichen, müssen deswegen zusätzlich gewichtete Parameter wie das adjustierte Bestimmtheitsmaß oder der F-Wert

3. Erstellung eines Trainingsdatensatzes mit ausgewogener Affinitätsverteilung

berücksichtigt werden. Der F-Wert ist ein Maß für die Signifikanz von R^2 und ist mit der Zahl der Freiheitsgrade gewichtet. Er wird nach Gleichung (3.3) berechnet.

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k} \quad (3.3)$$

In Gleichung 3.3 steht k für die Zahl der Variablen und n für die Zahl der Datenpunkte. Der F-Wert wird in dieser Arbeit neben den Korrelationskoeffizienten als wichtigstes Kriterium für die Auswahl von linearen Modellen verwendet.

Methode: PLS Der Unterschied zwischen PLS und MLR besteht darin, dass bei der MLR direkt die Prediktorvariablen (im Fall von SFCscore also die Deskriptoren) für die Regressionsanalyse verwendet werden, während bei der PLS eine Hauptkomponentenanalyse (PCA) vorgeschaltet wird.

Bei der PCA geht es darum, hoch-dimensionale Datensätze in eine geringere Zahl von Variablen zusammenzufassen und dabei möglichst viel der ursprünglichen Information beizubehalten. Der Datensatz wird dazu in sogenannte Hauptkomponenten zerlegt. Eine Hauptkomponente ist eine Linearkombination (Variable + Gewichtungskoeffizient) aller Variablen des Datensatzes. Die erste Hauptkomponente ist dabei diejenige Kombination, die den geringsten Fehler zwischen der Geraden und den Datenpunkten im Variablenraum liefert, d.h. die den größten Teil der Varianz der Daten erklärt. Die zweite Hauptkomponente steht orthogonal zur ersten Hauptkomponente und erklärt das Maximum der verbleibenden Varianz. Dies lässt sich nun bis zur maximalen Anzahl an Variablen fortsetzen, wobei letztendlich wieder 100 % der Varianz erklärt werden würde.

Das Regressionsmodell wird bei der PLS basierend auf diesen Hauptkomponenten berechnet. Das Modell hat dann dieselbe generelle Form wie im vorigen Abschnitt beschrieben und es gelten dieselben Qualitätskriterien.

Variablen-Selektion Sowohl bei der Ableitung von PLS als auch von MLR-Funktionen ist eine Variablenselektion notwendig.

Die MLR ist zum einen anfällig gegenüber kokorrelierten Prediktorvariablen, die innerhalb des SFCscore-Deskriptorensatzes relativ häufig auftreten. Zum anderen führt die verwendete Datensatzgröße von 196 Komplexen dazu, dass maximal 5 bis 7 Deskriptorvariablen verwendet werden sollten, um ein "Overfitting" zu vermeiden.

Durch die Vorschaltung der PCA werden diese beiden Probleme bei der PLS umgangen. Zwar können auch hier nur wenige Hauptkomponenten für die eigentlich Regressionsanalyse verwendet werden. Diese sind jedoch nicht korreliert und enthalten zudem die wesentlichen Informationen des ursprünglichen Deskriptorensatzes.

Trotzdem empfiehlt sich auch bei der PLS die Durchführung einer milden Variablenselektion, weil der Deskriptorensatz einige Variablen enthält, die einen sehr geringe

Informationsgehalt für die Erklärung der Zielvariable liefern und somit das Vorhersagemodell hauptsächlich "verwässern".

Zur Durchführung der Variablenselektion existieren verschiedene Methoden, die sich grob in (heuristische) Suchen nach guten Deskriptor-Kombinationen und in Filter-Ansätze, welche iterativ Variablen eliminieren, einteilen lassen. Eine Übersicht über die zur Verfügung stehenden Methoden findet sich bei Blum und Langley [58]. Ob beim Schritt der Variablenselektion durch den Einsatz anderer Verfahren ein Optimierungspotential besteht, wurde an dieser Stelle nicht weiter untersucht. Stattdessen wurde die Methode verwendet, die sich bei Ableitung der ursprünglichen Funktionen bewährt hat (vgl. Sottriffer et al. [46]).

Diese basiert auf der relativen Wichtigkeit der Variablen in einem PLS-Modell. Diese wird durch Vergleich der standardisierten Regressionskoeffizienten ermittelt. Alle Variablen, deren standardisierter Regressionskoeffizient unter einem bestimmten Grenzwert liegt, werden dabei aus dem Modell eliminiert. Mit den verbleibenden Variablen wird dann ein neues Modell abgeleitet. Wenn keine Variable mehr den Grenzwert unterschreitet, ist das Stopp-Kriterium erreicht und die Variablenselektion beendet. Wichtige Parameter dieser Methode sind der Eliminationsgrenzwert und die Komponentenzahl für die lineare Regression. Das Modell, das im letzten Schritt berechnet wurde, kann dann direkt als prediktives Modell verwendet werden.

Die Berechnungen wurden mit der Software SYBYL-X¹ durchgeführt.

PLS-Modelle Es wurden verschiedene PLS-Modelle basierend auf unterschiedlichen Einstellungen für den Eliminationsgrenzwert und die Komponentenzahl erstellt. Als Qualitätskriterium für die Auswahl einer Funktion wurden der Korrelationskoeffizient (der Regression) sowie der F-Wert verwendet.

Das beste Modell beruhte auf insgesamt 19 Deskriptoren und 5 Hauptkomponenten. Der Eliminationsgrenzwert in der Variablenselektion war hier 4%. Eine genauere Analyse des Modells ergab jedoch einige physikalisch nicht plausible Deskriptorengewichtungen. So hat hier der n_{rot} -Deskriptor ein positives Vorzeichen, also einen positiven Einfluss auf die Bindungsaffinität. Des Weiteren haben einige Deskriptoren für die polaren Kontakte einen relativ großen negativen Einfluss auf das Modell, während sich polar-hydrophobe Kontakte positiv auswirken. Hinzu kommt ein relativ großer y-Achsenabschnitt von 5.87. Der y-Achsenabschnitt sollte bei einer solchen Funktion eigentlich möglichst nahe bei 0 liegen.

Insgesamt muss das Modell also als künstlich angesehen werden und kann nicht als sinnvolles Vorhersagemodell verwendet werden. Bei der Analyse der weiteren PLS-Modelle ergaben sich ähnliche Probleme, meist im Bezug auf die polaren Oberflächenkontakte.

¹SYBYL-X 1.0, Tripos International, St. Louis, Missouri, U.S.A.

MLR-Modelle Als Ausgangspunkt für die Entwicklung von MLR-Modellen wurde die Variablenselektion der PLS-Modelle (s.o.) verwendet. Diese wurden systematisch nach Variablenkombinationen (mit bis 7 Variablen) durchsucht, für die MLR-Modelle mit guten Regressionskoeffizienten und F-Werten berechnet werden konnten. Für die systematische Suche wurden die vorselektierten Deskriptorensätze verwendet, da die Kombinationsmöglichkeiten bei Verwendung aller Deskriptoren einen sehr hohen Rechenaufwand bedeutet hätten.

Basierend auf F-Wert und Korrelationskoeffizienten sowie der physikalischen Plausibilität wurde die beste Funktion für die Testung gegen den externen *Cheng*-Datensatz ausgewählt. Die Funktion besteht aus 5 Deskriptoren und hat einen Regressionskoeffizienten von 0.658, einen RMSE von 2.16 und einen F-Wert von 29.7. Diese Statistiken sind etwas schlechter als die Regressionsstatistiken der ursprünglichen SFCscore-Funktionen [46]. Da der Trainingsdatensatz aber durch die breite Affinitätsverteilung auch anspruchsvoller ist, ist dies nicht überraschend und muss nicht notwendigerweise eine schlechtere prediktive Qualität des Modells bedeuten.

In Tabelle 3.1 sind die Deskriptoren und Koeffizienten der neuen Funktion im Vergleich zu *sfc_229m* gezeigt. Wie man sieht, teilen sich beide Funktionen grundlegende Deskriptoren wie *n_rot*, *n_hb*, *AHPDI* und *TotBurSurf*. Auch der y-Achsenabschnitt ist ähnlich. Quantitativ haben die genannten Deskriptoren bei der neuen Funktion einen geringeren Einfluss auf die Gesamtvorhersage als bei *sfc_229m*. Einen relativ großen Einfluss hat beim neuen Modell der Deskriptor *HH_HA_AH_AA_surfc* (Summe der aromatischen und hydrophoben Kontaktflächen), der in der ursprünglichen Funktion nicht vorkommt. Die Deskriptoren für spezifische Interaktionen wie *RRScore*, *RMScore* und *met* treten dagegen nur bei *sfc_229m* auf.

3.3. Ergebnisse

Die Testung der neuen Funktion wurde mit dem *Cheng*-Datensatz durchgeführt. Die Ergebnisse sind in Tabelle 3.2 im Vergleich zu den ursprünglichen SFCscore-Funktionen zusammengefasst. Um den Einfluss der Variablenselektion etwas besser einschätzen zu können, wurde zusätzlich zu der neuen Funktion (*sfc_196_eqd*) auch eine Funktion abgeleitet, die auf der Deskriptorenauswahl der *sfc_229m*-Funktion beruht (*sfc_196_eqd_229varsel*).

Die Korrelationen für beide Funktionen liegen bei etwa 0.6. Der Vorhersagefehler beträgt knapp 2 pK_i-Einheiten. Damit wird das gleiche Niveau erreicht wie bei den ursprünglichen SFCscore-Funktionen. Die Annahme, dass der neue Trainingsdatensatz anspruchsvoller ist, hat sich bestätigt, da die Performance im Vergleich zu den Regressionsstatistiken nur leicht nachlässt.

Der Einfluss der Datensatz-spezifischen Variablenselektion ist ebenfalls erkennbar, da *sfc_196* etwas bessere Statistiken aufweist als *sfc_196_229varsel*. Der Effekt ist dabei zwar relativ gering, dies hängt jedoch auch damit zusammen, dass der neue Datensatz

Tabelle 3.1.: Deskriptoren und Gewichtungskoeffizienten der beiden Funktionen *sfc_196eqd* und *sfc_229m*.

Deskriptor	<i>sfc_196_eqd</i>	<i>sfc_229m</i>
n_rot	-0.147	-0.0663
n_hb	+0.174	+0.416
AHPDI	+0.0195	+0.0325
HH_HA_AH_AA_surfc	+0.00696	-
met	-	+0.636
RRScore	-	+0.145
RMScore	-	+1.407
TotBurSurf	+0.00501	+0.00541
Intercept	+1.696	+1.726

aus dem 229-Datensatz abgeleitet wurde und die Variablenselektion somit recht ähnlich ist. Der Einfluss eine adäquaten Variablenselektion wird dabei aber trotzdem deutlich.

Die erhoffte deutliche Verbesserung der Vorhersagegenauigkeit ist hingegen nicht erkennbar. Analysiert man die Ergebnisse in den verschiedenen Affinitätsbereichen genauer, so lässt sich feststellen, dass *sfc_196_eqd* im niedrigen Affinitätsbereich (unter 5) etwas bessere Vorhersagen liefert (Tabelle 3.3), was als Effekt der Trainingsdatenzusammensetzung angesehen werden kann. Eine bessere Erkennung von schwach oder nicht aktiven Verbindungen ist eine sehr wichtige Eigenschaft für Scoring-Funktionen. Das Ausmaß dieses Effekts ist jedoch relativ gering, so dass insgesamt das Ziel einer deutlich verbesserten Scoring-Funktion nicht erreicht wurde.

3.4. Ausblick

Obwohl der Ansatz (zugunsten der in Kapitel 4 beschriebenen Klassifizierungsmodelle) hier nicht weiter verfolgt wurde, ist das gezielte Design von Trainingsdatensätzen eine erfolgversprechende Strategie.

Zur Weiterentwicklung der hier verwendeten Methode wird es unumgänglich sein die manuelle Auswahl von Komplexen durch maschinellen Algorithmen zu ergänzen beziehungsweise zu ersetzen. Einerseits können dadurch deutlich größere Datenmengen verarbeitet werden. Andererseits kann auch gezielter darauf geachtet werden, artifizielle Deskriptorenverteilungen, wie sie hier beispielsweise für die polaren Kontaktoberflächen beobachtet wurden, zu vermeiden.

Die am Ende dieser Arbeit durchgeführten Untersuchungen zur "Applicability Do-

3. Erstellung eines Trainingsdatensatzes mit ausgewogener Affinitätsverteilung

Tabelle 3.2.: Vergleich der *196_eqd*-Funktionen (fett hervorgehoben) mit den ursprünglichen SFCscore-Funktionen bei der Vorhersage des *Cheng*-Datensatzes. R_P : Pearson Korrelations-Koeffizient; ρ : Spearman Rangkorrelation; RMSE: Root Mean Square Error.

Scoring Funktion	R_P	ρ	RMSE
sfc_229m	0.59	0.64	1.93
sfc_229p	0.59	0.63	1.96
sfc_290m	0.59	0.65	1.95
sfc_290p	0.62	0.66	1.88
sfc_855	0.60	0.64	1.92
sfc_frag	0.58	0.64	2.69
sfc_met	0.64	0.68	1.84
sfc_ser	0.59	0.63	1.97
sfc_196_eqd	0.60	0.63	1.93
sfc_196_eqd_229varsel	0.58	0.60	1.95

Tabelle 3.3.: Vorhersagefehler (RMSE) der Funktionen sfc_met und sfc_196_eqd für verschiedene Affinitätsbereiche.

experimenteller Affinitätsbereich	sfc_met	sfc_196_eqd
< 5	1.45	1.21
5 - 8	0.79	1.07
> 8	1.93	2.10

main" (vgl. Kapitel 7) stellen eine Grundlage für das gezielte Design von Trainingsdatensätzen dar.

4. Klassifizierungsmodelle

Durch die Ableitung von Klassifizierungsmodellen soll das Augenmerk noch stärker auf die Unterscheidung von aktiven und inaktiven Verbindungen gelegt werden. Im Unterschied zu klassischen Scoring-Funktionen erhält man als Ergebnis eines solchen Modells nicht mehr einen genauen Score, sondern nur noch eine Einteilung in "aktiv" oder "inaktiv". Auf den ersten Blick generiert ein solches Modell also weniger Information als eine klassische Scoring-Funktion. In vielen Fällen ist jedoch diese grundlegende Aussage schon sehr wertvoll.

4.1. Trainingsdaten

Klassifizierungsalgorithmen stellen grundlegend andere Anforderungen an die Trainingsdaten als Regressionsmethoden. Ein Vorteil besteht darin, dass keine genauen Affinitätsdaten notwendig sind, sondern eine - experimentell einfacher und verlässlicher zu erhaltende - qualitative Aussage über Aktivität oder Inaktivität ausreichend ist. Untersuchungen zeigen, dass die Aktivitätsdaten in öffentlich zugänglichen Datenbanken nicht zu unterschätzende Fehler aufweisen [59, 60]. Diese Fehler, die die Ableitung und die Testung von regressionbasierten Modellen beeinträchtigen, spielen bei Klassifikatoren keine oder nur eine untergeordnete Rolle.

Der eigentliche Vorteil der Methode besteht jedoch darin, dass auch negative Daten (d.h. inaktive Verbindungen) zum Training verwendet werden können. Die klassischen Funktionen sind hingegen ausschließlich mit aktiven Verbindungen trainiert, was ein wichtiger Grund für die Schwächen bei der Erkennung von inaktiven Verbindungen ist.

Gleichzeitig stellen die negativen Daten aber auch die größte Einschränkung der Methode dar. Zum einen ist es schwierig, die entsprechenden Daten in ausreichender Menge zu akquirieren, zum anderen stellt die Generierung von plausiblen Bindeposen eine große Herausforderung dar, die allerdings entscheidend für die Qualität eines solchen Modells ist.

Material: Directory of Useful Decoys (DUD) Das "Directory of Useful Decoys" (DUD) wurde von J. Irwin als Benchmark-Datensatz für Docking-Anwendungen und Scoring-Funktionen entwickelt [61]. Insgesamt enthält die Datenbank 2950 aktive Verbindungen, die sich auf 40 verschiedene Proteintargets verteilen. Zusätzlich zu den etwa 3000 aktiven Verbindungen, enthält die Datenbank knapp 100,000 *Decoys*, also inaktive Verbindungen.

Für jeden der aktiven Liganden wurden genau 36 *Decoys* generiert. Dazu wurde die ZINC Datenbank [62, 63] mit Hilfe von *CACTVS type 2 Fingerprints* gefiltert. Alle Einträge, die einen Tanimoto-Koeffizienten von über 0.9 zu einem der Liganden haben und diesem somit topologisch sehr ähnlich sind, wurden ausgeschlossen. Für die verbleibenden Einträge wurden physikochemische Deskriptoren mit Hilfe des Programms QuikProp (Schrödinger, LLC, New York, NY) berechnet. Basierend auf fünf Deskriptoren wurden dann für jeden Liganden 36 möglichst ähnliche *Decoys* ausgewählt. Weitere Details zur Zusammenstellung und der Auswahl der *Decoys* können bei Huang et al. [61] nachgelesen werden.

Die große Zahl an *Decoys*, sowie das gute Verhältnis zwischen aktiven und inaktiven Verbindungen machen den Datensatz zu einer ausgezeichneten Datenquelle für die Generierung von Klassifizierungsmodellen.

Eine große Einschränkung des DuD-Datensatzes ist, dass die Negativ-Daten nicht experimentell bestätigt sind. Es ist also möglich, dass einige der hier als "inaktiv" indizierten Verbindungen in Wirklichkeit aktiv sind.

Zur Ableitung der Modelle wurde nicht der komplette DuD-Datensatz verwendet, da eine methodisch valide Durchführung des Dockings für 40 verschiedene Targets nicht möglich gewesen wäre. Stattdessen wurden protein-spezifische Modelle für das mykobakterielle Enzym InhA erstellt, welches ein Target in anderen Projekten des Arbeitskreises ist.

Der DUD InhA-Datensatz besteht aus 85 aktiven Verbindungen sowie 3043 *Decoys*. In der DUD-Publikation ist als Quelle für die aktiven Verbindungen "personal communication" angegeben [61], weswegen neben den Strukturen keinen weiteren Informationen zu den Verbindungen direkt zugänglich waren.

4.2. Generierung der Bindeposen

Die Liganden und Proteine liegen in der DuD-Datenbank separat vor. Es müssen also sowohl aktive als auch inaktive Verbindungen zunächst in die Bindetasche gedockt werden. Bevor der eigentliche DuD-Datensatz gedockt werden kann, muss zunächst ein valides Docking-Schema gefunden werden. Im Fall von InhA wird dies durch eine Besonderheit der Bindetasche erschwert. Diese verfügt über einen flexiblen Loop, welcher die Größe der Bindetasche variiert und somit die Bindung verschieden großer Liganden erlaubt (*Induced Fit*). In Abbildung 4.1 ist dies anhand verschiedener Kristallstrukturen dargestellt.

Um das Docking vor diesem Hintergrund valide zu gestalten, wurde ein Cross-Docking von bekannten Kristallstrukturen durchgeführt.

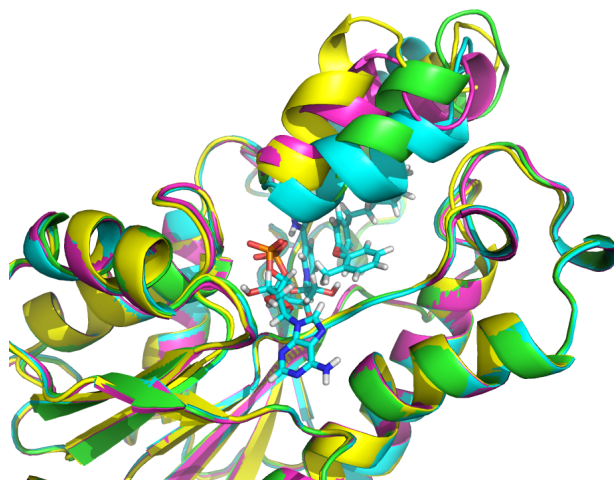


Abbildung 4.1.: InhA-Bindetaschenregion der PDB-Strukturen: 1P44(gelb), 2H7I(pink), 3FNE(grün), 2X23(türkis). Alle Strukturen sind auf 2X23 überlagert. In türkis sind der 2X23-Ligand und NAD⁺ dargestellt. Im oberen Teil der Abbildung erkennt man gut den flexiblen Teil der Bindetasche.

4.2.1. Etablierung und Validierung eines Docking-Schemas - Cross-Docking

Datensatz Für das Cross-Docking muss ein Datensatz von bekannten Kristallstrukturen mit möglichst unterschiedlichen Bindetaschenkonformationen erstellt werden. Dazu wurde die Protein Data Bank (PDB) [64] nach InhA-Kristallstrukturen durchsucht, bei denen die Bindetasche komplett aufgelöst ist und möglichst unterschiedliche Konformationen aufweist. Es wurden 11 Strukturen gefunden.

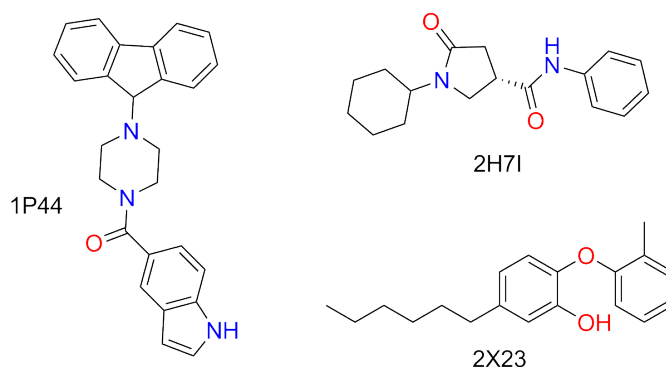


Abbildung 4.2.: Repräsentative Strukturklassen der Liganden des Cross-Dockings. 1P44: Piperazin-substituierter Trizyklus, 2H7I: Pyrrolidincarboxamid, 2X23: Diphenylether

4. Klassifizierungsmodelle

Auf Ligandenseite treten vier verschiedene Strukturklassen auf. Abbildung 4.2 zeigt drei der Liganden als Vertreter der hier relevanten Strukturklassen (Pyrrolidincarboxamide, Diphenylether, Trizyklen). Bei der vierten Strukturklasse (vertreten durch 1BVR) handelt es sich um einen langkettigen Fettsäureester, der in dieser Auswertung keine Rolle spielt. Die Pyrrolidincarboxamid und die Diphenylether sind dabei mehrfach vertreten, während die Liganden von 1P44 und 1BVR als einzelne Vertreter vorkommen. Im Falle der PDB-Struktur 3FNE wurde noch eine zweite sehr ähnliche Struktur (3FNH) gefunden, welche eine praktisch identische Proteinstruktur aufweist. Der ko-kristallisierte Ligand dieser Struktur wurde als zusätzlicher Ligand im Cross-Docking verwendet. Die Proteinstruktur wurde nicht berücksichtigt, da die Bindetasche identisch zu der von 3FNE ist, jedoch eine schlechtere Auflösung aufweist. Insgesamt wurden also 11 Proteinstrukturen und 12 Liganden verwendet. Tabelle 4.1 gibt einen Überblick über die verwendeten Liganden und Proteine.

Tabelle 4.1.: Liste der für das Cross-Docking verwendeten Proteine und Liganden aus der PDB.

Proteine	Auflösung (in Å)	Referenz	Liganden	Strukturklasse Ligand
1BVR	2.80	[65]	1BVR	Fettsäurethioester
1P44	2.70	[66]	1P44	tricyclisches Grundgerüst
2B36	2.80	[67]	2B36	Diphenylether
2H7I	1.62	[68]	2H7I	Pyrrolidincarboxamid
2H7L	1.73	[68]	2H7L	Pyrrolidincarboxamid
2H7M	1.62	[68]	2H7M	Pyrrolidincarboxamid
2H7N	1.90	[68]	2H7N	Pyrrolidincarboxamid
2H7P	1.86	[68]	2H7P	Pyrrolidincarboxamid
2NSD	1.90	[69]	2NSD	Diphenylether
3FNE	1.98	[70]	3FNE, 3FNH	Diphenylether
2X23	1.81	[71]	2X23	Diphenylether

Im Cross-Docking wird nun jeder der Liganden in jede Kristallstruktur gedockt. Die gedockte Pose wird dann mit der Referenzpose aus der Kristallstruktur verglichen. Auf diese Weise kann einerseits evaluiert werden, wie gut das Docking-Schema für die gegebene Bindetasche funktioniert. Andererseits kann auch ermittelt werden, welche Bindetaschenkonformationen benötigt werden, um für jeden der Liganden eine korrekte Pose zu erhalten. Für das Docking des DUD-Datensatzes ist diese Information besonders wichtig.

Methode: Das Docking-Program Glide Zum Docking wurde das Programm Glide [72, 73, 54] verwendet. Bei der Generierung der Posen wird hierbei zunächst mit vorgenerierten Ligandkonformationen gearbeitet. Mit Hilfe einer Reihe von hierarchischen Filtern, die zunehmend präziser werden, wird nach vielversprechenden Posen innerhalb der Bindetasche gesucht. Die Filter sind Gitter-basiert und müssen einmalig vor der Durchführung des eigentlichen Dockings berechnet werden. Der Konformationsraum des Liganden wird dabei systematisch durchsucht, so dass auf keine stochastischen Methoden zurückgegriffen werden muss. Obwohl einige Näherungen nötig sind, um vertretbare Rechenzeiten zu erreichen, lassen sich gute und reproduzierbare Ergebnisse erhalten.

Die Posen, die bei dieser initialen Suche gefunden werden, werden dann in der Bindetasche energieminiert. Dazu wird das Kraftfeld OPLS [74] in Kombination mit einem distanzabhängigen Ladungsmodell verwendet. Im letzten Optimierungsschritt werden für die besten Lösungen die Torsionswinkel mit Hilfe eines Monte-Carlo-Algorithmus optimiert [73]. Für die Auswahl der besten Posen wird die Scoring-Funktion *Emodel* verwendet, eine Kombination aus den Kraftfeldenergien des OPLS-Kraftfeldes und der Scoring-Funktion *GlideScore*. *GlideScore* ist eine Erweiterung der empirischen Scoring-Funktion *ChemScore* [75] und wird bei Glide zur Vorhersage von Bindungsaffinitäten und beim *Ranking* eingesetzt.

Die Methode *XP Glide* stellt eine Weiterentwicklung von Glide beim *Sampling* und vor allen Dingen beim *Scoring* dar. Durch die zusätzliche Einbeziehung von Desolvationseffekten und der verbesserten Erkennung von Bindungsmotiven, die einen großen Einfluss auf die Affinität haben, können hier nochmals bessere Ergebnisse, vor allem im Bereich der Erkennung von inaktiven Verbindungen, erzielt werden [72].

Cross-Docking: Setup und Durchführung Zur Durchführung des Cross-Dockings wurden im ersten Schritt zunächst alle Komplexstrukturen anhand ihrer Bindetasche überlagert. Dadurch wird später der Vergleich von verschiedenen Bindeposen möglich.

Danach wurden die Liganden aus der Bindetasche ausgeschnitten und mit Hilfe des Schrödinger Programms LigPrep¹ für das Docking aufbereitet. Zur Energieminimierung wurde dabei das Kraftfeld OPLS_2005 verwendet. Die Protonierungszustände wurden mit der Software Epik². [76, 77] bei einem pH Wert von 7 +/- 2 berechnet und dann nochmals visuell überprüft. Tautomere Strukturen wurden nicht gefunden. In einigen Fällen existieren stereoisomere Formen; da die aktive Konfiguration jedoch bekannt ist, wurden diese nicht verwendet.

Die 11 Proteinstrukturen wurden mit Hilfe des *Protein Preparation Wizard*³ für das Docking aufbereitet. Dabei wurde die Proteinstruktur durch das Zuweisen bzw. Prüfen

¹LigPrep, Version 2.5, Schrödinger, LLC, New York, NY, 2011

²Epik, Version 2.2, Schrödinger, LLC, New York, NY, 2011

³Schrödinger Suite 2011 Protein Preparation Wizard; Epik Version 2.2, Schrödinger, LLC, New York, NY, 2011; Impact version 5.7, Schrödinger, LLC, New York, NY, 2011; Prime Version 3.0, Schrödinger, LLC, New York, NY, 2011 [78]

der korrekten Atom- und Bindungstypen zunächst in das richtige Format gebracht. Vereinzelt traten dabei Probleme auf, die manuell bereinigt wurden. Sämtliche strukturellen Probleme (z.B. unvollständige Seitenketten), die außerhalb der Bindetaschenregion lagen, wurden ignoriert. Es wurden alle Wassermoleküle und Heteroatome (z.B. Moleküle aus dem Kristallisationspuffer) entfernt. Wasserstoffatome wurden zunächst anhand der Standardprotonierungszustände der Aminosäuren hinzugefügt. Auch die Protonierung des Co-Faktors (NAD^+) wurde angepasst.

Im Anschluss daran wurde das Wasserstoffbrücken-Netzwerk des gesamten Proteins optimiert. Dabei wird die Orientierung von polaren Wasserstoffatomen angepasst. Terminale Amine und Histidine können um 180° gedreht werden. Auch die Protonierungszustände werden hier nochmals geprüft. Durch eine kurze Energieminimierung der Wasserstoffatome wird das Wasserstoffbrücken-Netzwerk abschließend optimiert.

Die so generierten Strukturen wurden als Basis für das Docking mit Glide verwendet. Der erste Schritt für das Docking ist die Generierung des Gitters (*Grid*), welches Glide zur Posengenerierung benötigt. Hierbei wurden alle OH-Gruppen im Bereich der Bindetasche als frei drehbar definiert. Als Zentrum des Gitters wurde jeweils der kokristallisierte Ligand gewählt. Bei allen weiteren Parametern wurden die Standardeinstellungen verwendet.

Das eigentliche Docking wurde dann im XP (*Extra Precision*) Modus durchgeführt. Die Liganden wurden dabei flexibel behandelt. Die Inversion von pyramidalen Stickstoffen und die Generierung von zusätzlichen Ringkonformationen (vor dem eigentlichen Docking) war erlaubt. Das Auftreten von nicht-planaren Amididen wurde bestraft, aber nicht von vornherein ausgeschlossen. Die weiteren Parameter wurden auf ihren Standardwerten belassen. Das Docking wurde frei, ohne die Vorgabe von "constraints" oder ähnlichem, durchgeführt. Für jeden Liganden wurden bis zu 5 Posen generiert. Für den letzten Docking-Schritt, die *Post-Docking Minimierung*, sollten je Ligand 10 Posen berücksichtigt werden.

Auf diese Weise wurde jeder der 12 Liganden in jede der 11 Proteinstrukturen gedockt. Aus den 5 Posen wurde mit Hilfe des *GlideScores* die beste Pose ausgewählt und mit den Referenzposen aus der Kristallstruktur verglichen. Alternativ zu *GlideScore* wurde auch DSX [79] zur Auswahl der Posen verwendet. Da die Ergebnisse von *GlideScore* insgesamt aber besser waren, werden nur diese im folgenden Kapitel vorgestellt und analysiert.

Cross-Docking: Auswertung Die Ergebnisse des Cross-Dockings lassen sich am besten in einem sog. Levelplot darstellen (Abb. 4.3). Dabei wird die Abweichung (RMSD-Wert in Å) der Dockingposen zur jeweiligen Referenzstruktur anhand einer Farbskala dargestellt. Dunkelblau/violett bedeutet hierbei, dass der Bindemodus gut reproduziert werden konnte, während grün/gelb für eine größere Abweichung steht. Bei den Liganden von 1BVR und 1P44 gibt es außerdem weiße Felder. Diese bedeuten, dass hier das Docking nicht erfolgreich war, weil die Liganden nicht in die Bindetasche

passen.

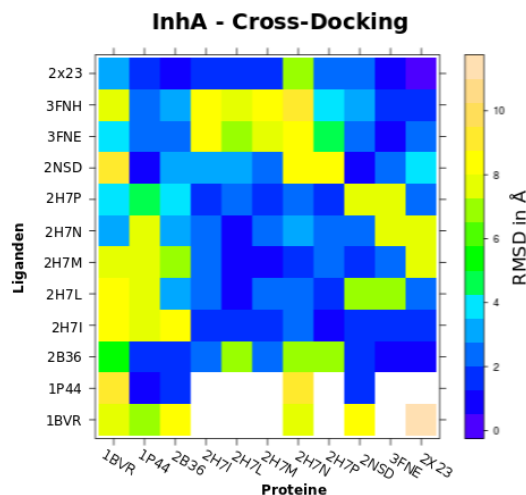


Abbildung 4.3.: Ergebnis des Cross-Docking. Die Farbskala zeigt die RMSD Abweichung (in Ångström) jedes Liganden (y-Achse) von seiner Kristallpose nach dem Docking in die jeweilige Proteinstruktur (x-Achse). Weiße Felder bedeuten, dass das Docking nicht erfolgreich war.

Der Levelplot zeigt, dass das Redocking in die ursprüngliche Bindetasche für die meisten Liganden gut funktioniert. Lässt man die Struktur 1BVR außen vor (s.u.) werden die Liganden mit einem durchschnittlich RMSD von 1.18 Å gedockt. Das bedeutet, dass Glide generell in der Lage ist, kristallnahe Posen zu finden. Die Ausnahme hierbei bildet 1BVR. Aufgrund der langen und flexiblen Kohlenwasserstoffkette dieses Liganden (Fettsäurethioester) war hier jedoch zu erwarten, dass der Bindemodus nicht exakt reproduziert werden kann. Auch das Docking des Liganden in die anderen Bindetaschen war dementsprechend nicht erfolgreich. Da die Methode aber ohnehin nicht auf das Docking von langkettigen Fettsäuren abzielt und diese auch nicht im DuD-Datensatz vorkommen, kann man dies bei der weiteren Analyse außer Acht lassen.

Die Ergebnisse des Cross-Dockings, also des Dockings der Liganden in die nicht-nativen Bindetaschen, kann man gut auf Basis der Ligandklassen interpretieren. So erhält man innerhalb der Gruppe der Diphenylether (2X23, 2FNH, 3FNE, 2NSD und 2B36) geringe RMSD-Werte. Die RMS-Abweichungen sind hier bei fast allen Kombinationen kleiner als 2 Å. Beim Docking der Diphenylether in die Pyrrolidincarboxamid-Bindetaschen erhält man hingegen deutlich größere Abweichungen. Die Liganden von 3FNH und 3FNE werden beispielsweise in keiner der 2H7*-Bindetaschen korrekt gedockt. In die etwas größere 1P44 Bindetasche passen die Diphenylether hingegen gut.

Für die Pyrrolidincarboxamide (2H7*) gilt ähnliches. Während die Ergebnisse inner-

halb der Gruppe zufriedenstellend sind, ist das Docking in andere Bindetaschen nur teilweise erfolgreich. Vor allem für die größeren Bindetaschen von 1P44 und 1BVR werden große RMS-Abweichungen erhalten. Zum Teil gilt das aber auch für die kleineren Diphenylether-Taschen.

Der 1P44-Ligand, der sich mit seinem trizyklischen Grundgerüst deutlich von den anderen Stoffklassen unterscheidet, konnte neben seiner Referenzbindetasche nur in vier weitere Zielstrukturen gedockt werden. Gute Posen findet man nur in 2NSD und 3FNE. In den anderen Fällen konnte kein Bindemodus erhalten werden, da die Taschen zu klein für den Liganden sind.

Das Hauptziel der Analyse ist es, eine Proteinstruktur oder eine Kombination von Proteinstrukturen zu finden, in die möglichst viele Strukturklassen korrekt gedockt werden können. Müsste man sich für eine einzelne Targetstruktur entscheiden, würde man auf Basis des Cross-Dockings wohl 2NSD auswählen. Hier werden, mit Ausnahme von zwei Pyrrolidincarboxamiden (2H7P und 2H7L), alle Liganden relativ gut gedockt. Allerdings sind auch die Ergebnisse für die Liganden von 3FNH, 3FNE und 2X23 mit Abweichungen im Bereich von 2,5 - 3,5 Å nicht optimal. Der Mittelwert aller RMSD-Werte (ohne 1BVR) läge bei 3,03 Å.

Um dieses Ergebnis noch etwas zu verbessern, wurde nach möglichen Kombinationen gesucht. Neben den RMSD-Werten müssen bei der Suche nach solchen Kombinationen auch die Scores berücksichtigt werden. In Abbildung 4.4 ist neben dem bereits bekannten RMSD-Levelplot (Abb. 4.4; oben links) zusätzlich der Levelplot für die *GlideScores* dargestellt (Abb. 4.4; oben rechts). Unter den beiden Levelplots sind die Tabellen mit den entsprechenden numerischen Werten angegeben.

In der Abbildung sind bereits zwei Kombinationsmöglichkeiten eingezeichnet. Würde man beispielsweise in 1P44 und 2H7N docken und jeweils die von GlideScore höher bewertete Pose auswählen (rote Kästen in der Tabelle), käme man auf einen durchschnittlichen RMSD-Wert von 2,97 Å. Verglichen mit den 3,03 Å von 2NSD ist das nur eine geringe Steigerung.

Eine andere Kombinationsmöglichkeit ist mit grünen Kästen dargestellt. Hier werden 2NSD und 2X23 kombiniert, was zu einem durchschnittlichen RMSD von 2,6 führt. Für die gegebenen Werte ist das die bestmögliche Zweier-Kombination. Noch höhere Genauigkeiten ließen sich durch eine Kombination mit einer der 2H7M-Strukturen erhalten. Betrachtet man aber die *GlideScores*, so sieht man, dass 2NSD und 2X23 generell höhere Vorhersagen erhalten. Diese Alternative macht deswegen wenig Sinn.

Mit der Kombination 2NSD/2X23 erhält man allerdings immer noch in zwei Fällen schlechte Dockingposen (bei 2H7N und 2H7P). Abbildung 4.5 zeigt den Kristall-Bindemodus des 2H7P-Liganden in der 2NSD Bindetasche (violett) sowie die korrespondierende Dockingpose (türkis). Der RMSD-Wert zwischen beiden Posen beträgt 8,71 Å. Man sieht jedoch, dass in beiden Fällen die essentielle Wasserstoffbrücke zu Tyrosin 158 gebildet wird. Der Unterschied ist, dass dies in der Kristallstruktur über die Carbonylgruppe am Pyrrolidin-Ring geschieht, bei der Docking-Pose jedoch über den Amid-Sauerstoff. Dies führt dazu, dass die beiden endständigen hydrophoben Reste

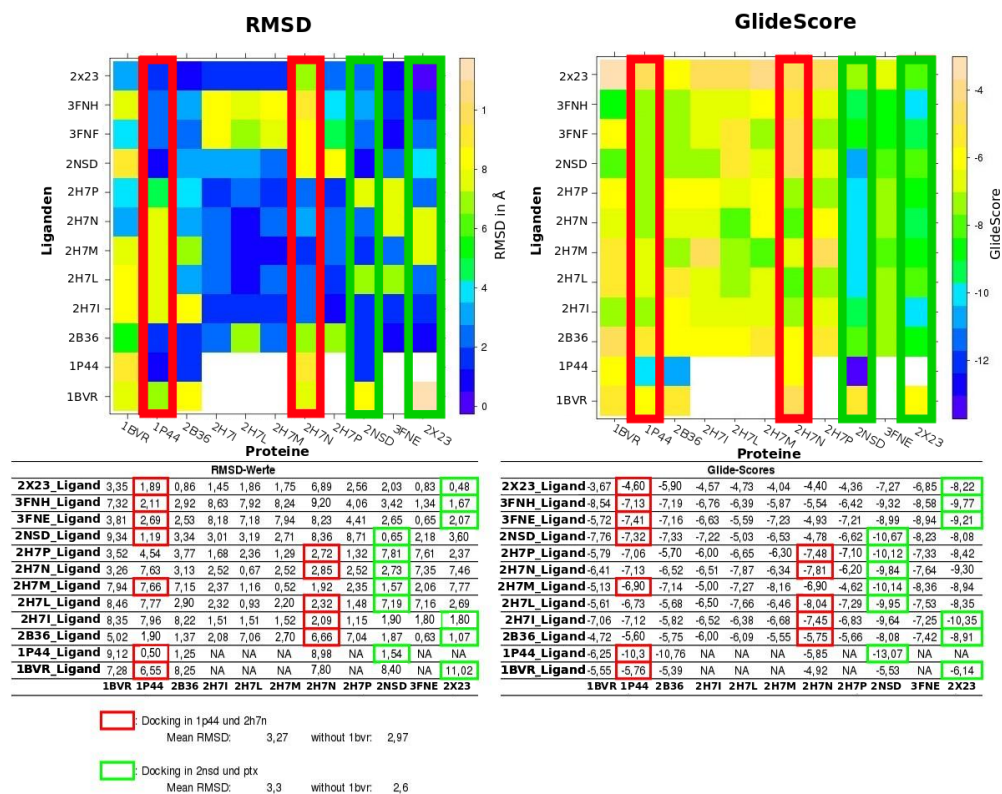


Abbildung 4.4.: Links: RMSD-Werte für das Cross-Docking als Heatmap und als Tabelle. Rechts: Glide-Scores der Top-Posen aus dem Cross-Docking als Heatmap und als Tabelle. Rot eingrahmt: Kombination von 1P44 und 2H7N, grün eingrahmt: Kombination von 2NSD und 2X23

vertauscht werden und der gesamte Ligand beim Docking um 180° gedreht vorliegt. Der etwas größere Chlor-substituierte Phenylrest passt nicht so gut in die engere 2NSD Bindetasche, weswegen dieser beim Docking auf der Lösungsmittel-exponierten Seite platziert wird und stattdessen der Cyclohexylrest im tief vergrabenen Teil der Bindetasche liegt. Innerhalb der fünf besten Posen für die Pyrrolidin-Carboxamide kommen meist beide Posen vor. Die Scores liegen sehr eng beieinander, so dass eine Unterscheidung mit Scoring-Methoden sehr schwierig ist. Da die Platzierung der entscheidenden funktionellen Gruppe aber weitgehend korrekt ist, ist der Bindemodus trotz der großen Abweichung zur Kristallstruktur akzeptabel.

Die Bindetaschen von 2NSD und 2X23 sind beide relativ klein. Es besteht deswegen die Gefahr, dass das Docking nicht für alle Liganden des DUD-Datensatzes erfolgreich ist. Aus diesem Grund wurde entschieden, zusätzlich ein Docking in die Struktur 1P44 durchzuführen. Trotz der relative geringen Auflösung (2.7 Å) liefert diese Struktur im Cross-Docking gute Ergebnisse. Die Bindetasche ist ausreichend groß, so dass für alle

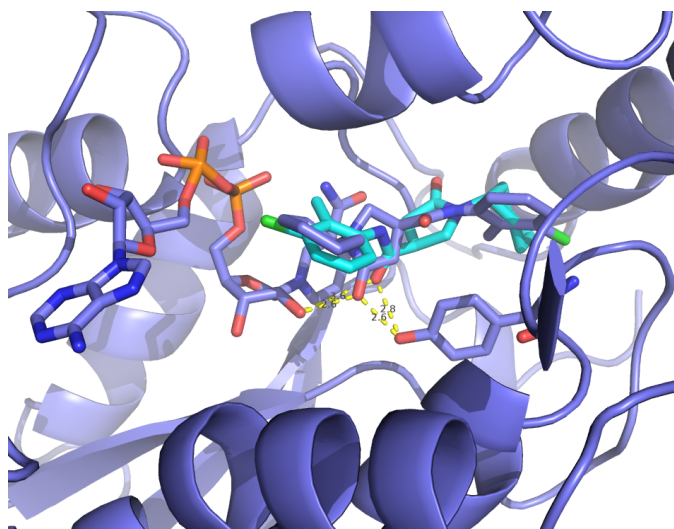


Abbildung 4.5.: Violett: Proteinstruktur 2NSD überlagert mit der Kristallstruktur des 2H7P-Liganden. Türkis: Docking-Pose des 2H7P-Liganden in der 2NSD-Struktur.

Liganden eine Docking-Lösung gefunden werden sollte. Die *GlideScores* von 1P44 liegen im Cross-Docking deutlich unter denen von 2NSD und 2X23. Es ist deswegen davon auszugehen, dass nur dann die 1P44-Pose verwendet wird, wenn das Docking in die anderen beiden Bindetaschen nicht erfolgreich ist.

4.3. Generierung von Trainings- und Testdatensätzen

DUD-Docking und Auswahl der Posen Im nächsten Schritt wurde der DuD InhA-Datensatz in die drei Proteinstrukturen gedockt. Dazu wurden die DUD-Liganden analog zu den Cross-Docking-Liganden mit Hilfe von LigPrep⁴ aufbereitet. Da in diesem Fall auch keine Informationen über aktive Konfiguration oder die Protonierung in der gebundenen Form vorlagen, wurden alle in Frage kommenden Stereoisomere, Tautomere und Protomere generiert. Für jeden Liganden waren maximal 32 Konfigurationen erlaubt. Der ursprüngliche DUD-Datensatz enthielt 3128 Verbindungen. Durch die zusätzlich generierten Strukturen ergab sich an dieser Stelle ein Datensatz mit insgesamt 3694 Einträgen (88 *Aktive*, 3606 *Decoys*).

Der komplette Datensatz wurde dann separat in jede Proteinstruktur gedockt. Auch das Docking erfolgte mit denselben Einstellungen wie beim Cross-Docking. Alle 88 *Aktiven* konnten erfolgreich in jede Targetstruktur gedockt werden. Von den 3606 *Decoys* konnten in 2X23 2816, in 2NSD 3561 und in 1P44 3604 Strukturen gedockt werden. Für jeden Liganden aus dem ursprünglichen DUD-Datensatz wurde dann, basierend auf dem

⁴LigPrep, Version 2.5, Schrödinger, LLC, New York, NY, 2011

GlideScore, genau eine Docking-Pose ausgewählt. Alle anderen Posen in den anderen Proteinstrukturen sowie die Posen von eventuell vorhandenen Isomeren oder Protomeren wurden im weiteren Verlauf nicht berücksichtigt. Der so entstandene gedockte Datensatz hatte dementsprechend wieder genau die Größe des ursprünglichen DUD-Datensatzes.

Auswahl der Testdaten Für die Evaluierung der Modelle verwendet man idealerweise einen externen Datensatz. Für das InhA-Modell waren solche Daten jedoch nicht in ausreichender Zahl verfügbar. Es konnte zwar eine kleine Anzahl an *Aktiven* und *Decoys* gesammelt werden, die nicht im DUD-Datensatz vorhanden sind; mit insgesamt 39 Verbindungen (6 *Aktive*, 33 *Decoys*) war der Datensatz aber zu klein und vor allem nicht heterogen genug zusammengesetzt, um eine verlässliche Evaluierung zu ermöglichen.

Aus diesem Grund wurde der gedockte DuD-Datensatz in Test- und Trainingsdaten aufgeteilt. Üblich ist hierbei etwa 20% des Datensatzes für die Testung zurückzuhalten und das Modell mit den restlichen 80% zu trainieren. Diese Werte wurden auch in dieser Arbeit verwendet. So wurde ein Random-Selection-Testdatensatz mit insgesamt 623 Komplexen generiert.

Was die Auswahl der Testobjekte angeht, gibt es wiederum unterschiedlich Ansätze. Die einfachste Methode ist, die Testdaten zufällig auszuwählen. Das hat den Nachteil, dass man, je nachdem wie die Datensätze dann zusammengesetzt sind, einen Bias in den Daten haben kann, der die statistische Auswertung verzerrt und zu einer Über- oder Unterschätzung der Performance führen kann. Bei einer zufälligen Auswahl empfiehlt es sich deshalb, den Prozess von Aufspaltung und Modellableitung mehrfach zu wiederholen.

Eine andere Möglichkeit besteht darin, die Testdaten gezielt auszuwählen, um mit den Testdaten den trainierten Bereich möglichst gut abzudecken. Man könnte in diesem Zusammenhang argumentieren, dass die wahre Performance so eher überschätzt wird und man nicht dieselben Ergebnisse für unbekannte Objekte (außerhalb des Trainingsbereichs) erwarten kann. Dies gilt jedoch unabhängig von der Zusammensetzung der Testdatensätze. Durch das gezielte Auswählen der Testobjekte lässt sich dagegen der zur Verfügung stehende Trainingsbereich optimal ausnutzen. Für die systematische Auswahl des Testdaten bietet sich der Einsatz von Clustering-Methoden an.

Methode: Clustering und PAM Cluster-Methoden sind multivariate Analyseverfahren, die zur Gruppe der unüberwachten Lernverfahren zählen. Sie spielen an unterschiedlichen Stellen in dieser Arbeit eine Rolle, beispielsweise zur Analyse von Liganddatensätzen oder bei der Auswertung der MD-Trajektorien (vgl. Kapitel 12).

Die grundlegende Idee der Clusteranalyse ist es, die Objekte eines Datensatz so zu gruppieren, dass sie Cluster bilden, in denen sich die Objekte untereinander möglichst ähnlich sind, während sie sich deutlich von Objekten aus anderen Clustern unterscheiden.

Für die Durchführung einer Clusteranalyse werden zwei Dinge benötigt. Zum einen

muss ein Maß für die Ähnlichkeit zwischen den Objekten definiert werden. Dies kann beispielsweise der (euklidische) Abstand zweier Objekte in einem hoch-dimensionalen Variablenraum sein (dann spielt auch noch die Methode der Distanzbestimmung eine Rolle), genauso gut können aber auch RMSD-Wert, Korrelationswerte oder andere Ähnlichkeitsmetriken verwendet werden. Zum anderen muss festgelegt werden, in welcher Weise die Objekte und Cluster untereinander verknüpft werden sollen (*Linkage*-Methode). Hier steht zum Beispiel die *single*-Methode, bei der für die Verknüpfung von verschiedenen Clustern immer die sich am nächsten liegenden Einzelobjekte verwendet werden, oder die *average*-Methode, bei der jeweils der Mittelwert des Clusters zur Verknüpfung verwendet wird, zur Verfügung.

Clustering-Verfahren lassen sich in eine Reihe von Untergruppen einteilen. In der vorliegenden Arbeit wird häufig das hierarchische Clustering verwendet. Hierbei wird, ausgehend vom kleinsten Ähnlichkeitswert, der Schwellenwert für die Verknüpfung immer weiter angehoben, bis letztendlich beim größtmöglichen Schwellenwert die Gesamtmenge der Daten ein Cluster bildet. Die Struktur der Daten lässt sich dadurch sehr schön in Form eines Dendrogramms darstellen.

Ein weitere Methode ist das sog. *PAM-Clustering*. PAM steht für "Partitioning Around Medoids" und gehört zur Gruppe der *Partitionierenden Verfahren*. Hier wird die Anzahl der gewünschten Cluster vor Beginn der Analyse festgelegt. Die Cluster werden dann durch Minimierung des durchschnittlichen Abstands innerhalb des Clusters berechnet. Beim PAM-Clustering wird für jeden Cluster ein sogenannter *Medoid* ermittelt. Dies ist dasjenige Objekt, welches den geringsten aufsummierten Abstand zu allen anderen Objekten des Clusters hat und somit als repräsentativ für den Cluster angesehen werden kann. Diese Methoden eignet sich hervorragend, um repräsentative Strukturen für den Testdatensatz auszuwählen.

Ein weiterführende Überblick zu Cluster-Methoden findet sich beispielsweise bei Kaufman und Rousseeuw [80].

Zur systematischen Auswahl der Testdaten wurde der im R-Paket "cluster" implementierte PAM-Algorithmus verwendet [81]. In der praktischen Umsetzung wurde die Aufteilung für *Aktive* und *Decoys* getrennt durchgeführt, um deren Anteile im Test- und Trainingsdatensatz ausgewogen zu halten. Für die *Aktiven* wurden dabei im ersten Schritt aus den 88 Objekten 15 Medoids (ca. 18%) ausgewählt bzw. der Datensatz in 15 Cluster unterteilt. In einem zweiten Schritt wurden die 15 Cluster dann nochmals einzeln einer PAM-Analyse unterzogen, wobei für alle Cluster, die mehr als 9 Objekte enthielten, 2 Repräsentanten ausgewählt wurden, für alle mit 3 bis 9 Mitgliedern 1 Medoid; für die sehr kleinen Cluster, die nur 1 oder 2 Objekte enthielten, wurde kein Medoid berechnet. Auf diese Weise wird sichergestellt, dass Ausreißer kein zu großes Gewicht in den Testergebnissen erhalten. Für den *Decoy*-Datensatz wurden 545 initiale Cluster (ca. 18% aus 3043) berechnet und dann analog vorgegangen. Der resultierende PAM-Testdatensatz enthielt 516 Komplexe (14 *Aktive* und 502 *Decoys*) und ist damit etwas kleiner als der zufällig ausgewählte Testdatensatz (623 Komplexe).

4.4. Ableitung der Modelle

Auf Basis der verbleibenden Trainingsdaten wurden dann zwei *RandomForest*-basierte Klassifizierungsmodelle abgeleitet.

Methode: RandomForest Der von Leo Breiman entwickelte *RandomForest*-Algorithmus (RF) zählt zur Gruppe der maschinellen Lern-Methoden (*Machine Learning*), genauer zu den *Ensemble*-Methoden [82].

Der Algorithmus leitet sich aus der weithin bekannten Methode der Entscheidungsbäume (*Decision Trees*) ab. In ihrer klassischen Form sind Entscheidungsbäume eine Klassifizierungsmethode. Ein Trainingsdatensatz wird dabei an jeder Verzweigung des Baums anhand einer Regel, die aus den Prediktorvariablen abgeleitet wird, in Untergruppen aufgeteilt. Das Ziel ist es, den Datensatz so aufzuteilen, dass eine möglichst gute Separierung der Zielklassen erreicht wird. Der Algorithmus kann allerdings auch als Regressionsverfahren zur Vorhersage von kontinuierlichen Werten verwendet werden. In diesem Fall ist das Ziel der Separierung, Objekte mit möglichst ähnlichen Zielvariablen zu gruppieren. Die beste Separierung wird durch Minimierung des *Mean Square Error* (MSE) der Zielvariable (Regression) bzw. durch Maximierung der Anreicherung (*Enrichment*) der Klassen (Klassifizierung) bestimmt. Für die Vorhersage wird einem unbekanntem Objekt der Durchschnittswert der jeweiligen Gruppe zugewiesen. Neben dem überwachten Lernen (Regression und Klassifizierung) ist auch unüberwachtes Lernen möglich. Hier werden die Objekte, völlig ohne Zielvariable, nach ihrer Ähnlichkeit im Variablenraum gruppiert.

Im Unterschied zu dieser klassischen Methode besteht ein *RandomForest* nun aus eine Vielzahl solcher Entscheidungsbäume. Die grundlegenden Eigenschaften und Einsatzmöglichkeiten (Klassifizierung, Regression, unüberwachtes Lernen) bleiben dabei erhalten. Durch die Verwendung einer großen Anzahl von Bäumen (*Ensemble*) ist ein RF-Modell jedoch sehr viel robuster und liefert bessere Vorhersagen. Diese Robustheit wird dadurch ermöglicht, dass sich die Bäume untereinander leicht unterscheiden. Dies wird auf zwei Wegen erreicht. Zum einen wird jeder Baum mit nur etwa zwei Dritteln der gesamten Trainingsdaten trainiert. Zum anderen steht an jedem einzelnen Entscheidungsknoten nur eine begrenzte, zufällige Auswahl an Prediktorvariablen zur Verfügung. Bei der Vorhersage eines unbekanntem Objekts werden die Einzelergebnisse aller Bäume gemittelt (Regression) bzw. es wird diejenige Klasse zugewiesen, welche von der Mehrheit der Bäume ermittelt wurde (Klassifizierung).

Ein großer Vorteil des RF-Ansatzes im Vergleich zu MLR und PLS ist, dass der Algorithmus nicht anfällig gegenüber unwichtigen Variablen ist. Da an jeder Verzweigung aus einer bestimmten Anzahl an Variablen ausgewählt werden kann, werden unwichtige Deskriptoren hier einfach ignoriert und beeinflussen das Modell nicht negativ. Aus dem selben Grund stellen auch korrelierte Deskriptoren kein Problem dar.

Ein weiterer Vorteil ist, dass für einen *RandomForest* nur zwei optimierbare Parameter existieren. Der erste ist die Anzahl der generierten Bäume (N_{Tree}). Svetnik und

Mitarbeiter haben nachgewiesen, dass eine erhöhte Anzahl an Bäumen ab einem bestimmten Grenzwert nur noch geringe Effekte auf die Verbesserung der Vorhersagen, jedoch wesentliche Effekte auf die Verlängerung der Rechenzeiten hat [83]. Der zweite optimierbare Parameter ist die Anzahl der unabhängigen Variablen, die dem Modell an jeder Verzweigung zur Verfügung stehen (m_{try}). Dieser Wert hat größeren Einfluss auf das Modell und muss je nach Einsatzzweck (Regression, Klassifizierung) unterschiedlich gewählt werden [83].

Bei der Validierung von Vorhersagemodellen ist die Evaluierung gegen möglichst große, unbekannte Datensätze der Gold-Standard (*externe Validierung*). Die *interne Validierung* mit Hilfe der Trainingsdaten ist aber gerade in der Entwicklungsphase eines Modells sehr wichtig, auch um die externen Daten möglichst lange zurückzuhalten. Als Methode zur internen Validierung bietet der RF-Algorithmus den *Out-of-Bag*-Fehler (OOB-Fehler). Dieser basiert auf der Tatsache, dass für das Training eines Baums nur etwa 66 % der Daten verwendet werden, während ein Drittel zurückbehalten wird. Deswegen existiert für jedes Trainingsobjekt eine bestimmte Anzahl an Bäumen, denen dieses Objekt nicht bekannt ist. Lässt man nun jedes Objekt nur von diesen Bäumen vorhersagen, erhält man den *Out-of-Bag*-Fehler. Dieser ist vergleichbar mit der Kreuzvalidierung einer PLS oder MLR-Funktion, hat aber den Vorteil, dass er direkt während der Generierung des Modells ohne zusätzlichen Aufwand berechnet werden kann. Der OOB-Fehler ist nützlich bei der Parametrisierung des Modells und bietet, wie man später sehen wird, eine durchaus realistische Einschätzung der Vorhersagequalität des Modells.

Die größte Schwäche der RF-Methode ist, dass das resultierende Modell sehr komplex ist und die Ergebnisfindung nicht auf direktem Weg nachvollziehbar ist. Das ist vor allem dann problematisch, wenn die Ergebnisse nicht den Erwartungen entsprechen und die Vorhersagen auf Basis der Deskriptoren interpretiert werden sollen. Es gibt jedoch indirekte Methoden, die einen Einblick in die Funktionsweise des Modells gewähren. So kann beispielsweise die sogenannte *Variable Importance*, also die Wichtigkeit der Variablen, berechnet werden. Dazu werden die Variablen während des Trainings einzeln, zufällig permutiert und die Abnahme der Vorhersagegenauigkeit gemessen. Je größer der Einfluss einer Variable auf die Vorhersage, desto größer ist diese Abnahme. Kombiniert man diese Variablen-Wichtigkeit mit einer Analyse der Deskriptorenverteilungen in den Trainings- und den Testdaten, lassen sich die Vorhersagen des Modells durchaus interpretieren.

Eine andere interessante Eigenschaft des Algorithmus' ist die Möglichkeit, die sogenannte *Proximity* zu berechnen. Dabei handelt es sich um einen Parameter, der angibt, wie ähnlich sich zwei Objekte sind. Zu diesem Zweck wird der Anteil der Entscheidungsbäume berechnet, in denen zwei Objekte in den selben Endknoten klassifiziert werden. Je größer der Anteil dieser gemeinsamen Bäume ist, desto ähnlicher sind sich zwei Objekte bezüglich der Verteilung der unabhängigen Variablen. So lässt sich beispielsweise die Ähnlichkeit eines Testobjekts zu den Trainingsdaten berechnen.

Ableitung der RF-Modelle Die Modelle wurden mit der Statistiksoftware R [84] und dem Paket “randomForest” [85] erstellt. Ein Ensemble von 500 Bäumen hat sich als ausreichend groß erwiesen. Der verwendete Variablenraum bestand aus allen SFCscore-Deskriptoren, mit Ausnahme der beiden stark größenabhängigen Deskriptoren *MW* (Molekulargewicht) und *NAtoms* (Anzahl schwerer Atome). Diese wurden nicht verwendet, um triviale Korrelationen zwischen Molekülgröße und Aktivität auszuschließen.

Die Modelle wurden mit einem für Klassifizierungsaufgaben relativ hohen m_{try} -Wert von 33 abgeleitet. Bei kleineren Werten nahm die Performance - sowohl für die interne als die externe Validierung - deutlich ab. Ein möglicher Grund dafür ist, dass relativ viele der SFCscore-Deskriptoren für diese spezifische Klassifizierungsaufgabe nicht wichtig sind, also keine relevante Information enthalten. Um eine gute Auftrennung zu erzielen, müssen deswegen an den jeweiligen Knoten mehr Variablen angeboten werden.

Es wurden zwei Modelle abgeleitet. Das erste beruht auf der systematischen Testdatenauswahl mit der PAM-Methode. Das Modell wird dementsprechend im Folgenden *PAM-Modell* genannt. Das zweite Modell basiert auf einer zufälligen Einteilung in Test- und Trainingsdaten. Dabei wurden 20% des Gesamtdatensatzes für Testzwecke zurückgehalten. Der Vorgang von Aufsplittung und Modellableitung wurde zehnfach durchgeführt, um ein statistisch valides Modell auswählen zu können. Die Ergebnisse des Modells, welches am nächsten an den durchschnittlichen *Out-of-Bag*-Statistiken der 10 Iterationen lag, werden in der folgenden Auswertung vorgestellt (*Random-Selection-Modell*).

4.5. Ergebnisse und Auswertung

Das Ergebnis einer Klassifizierung entspricht einem von vier Fällen: (1) Ein Objekt wird als aktiv vorhergesagt und ist tatsächlich aktiv; (2) das Objekt wird als aktiv vorhergesagt ist aber in Wirklichkeit inaktiv; (3) das Objekt wird als inaktiv vorhergesagt und ist tatsächlich inaktiv oder (4) es ist in Wirklichkeit aktiv.

Diese vier Fälle können sehr schön in einer Confusion-Matrix (Abb. 4.6) dargestellt werden, wobei oben links und unten rechts die korrekten Vorhersagen stehen. Im unteren linken Feld findet man die Zahl der als aktiv vorhergesagten Liganden, die aber in Wirklichkeit inaktiv sind, also die Falsch-Positiven (FP). Rechts oben findet sich dementsprechend die Zahl der Falsch-Negativen (FN) Verbindungen. Aus der Confusion-Matrix lassen sich direkt die wichtigsten Parameter für die Bewertung eines Klassifikators, wie die Genauigkeit und die Sensitivität, berechnen (siehe dazu Abb. 4.6).

Die Tabellen 4.2 a) bis d) zeigen die Confusion-Matrices für die interne (*Out-of-Bag*-Daten) und die externe Validierung (Testdaten, vgl. voriger Abschnitt) der beiden Modelle. In Tabelle 4.3 sind die Genauigkeit, die Präzision, die Sensitivität und die Falsch-Positiv-Rate (FP-Rate) für die vier Testszenarien aufgeführt. Die Genauigkeit gibt an, welcher Anteil der Testfälle tatsächlich korrekt vorhergesagt wird. Sie ist bei allen Modellen sehr hoch, was in diesem Fall aber nicht viel bedeutet, da durch den

4. Klassifizierungsmodelle

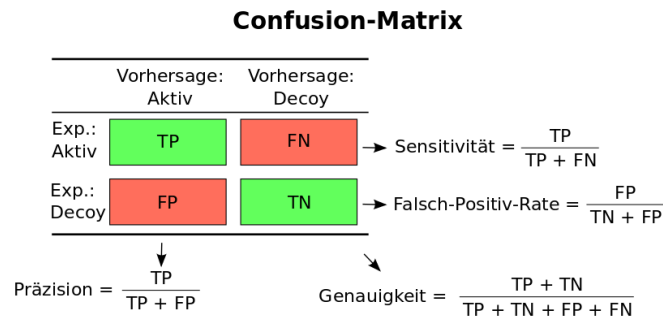


Abbildung 4.6.: Schematische Darstellung des Aufbaus und der Interpretation einer Confusion-Matrix. TP: True Positives, TN: True Negatives, FP: False Positives, FN: False Negatives

Tabelle 4.2.: Confusion-Matrices für die interne und externe Validierung der beiden Klassifizierungsmodelle. Horizontal sind die vorhergesagten Klassen, vertikal die tatsächlichen Klassen aufgetragen.

- a) Random-Selection-Modell, *OOB*-Datensatz (2493 Komplexe) b) Random-Selection-Modell, Testdatensatz (623 Komplexe)

	Aktiv _{calc}	Decoy _{calc}
Aktiv _{exp}	35	32
Decoy _{exp}	5	2421

	Aktiv _{calc}	Decoy _{calc}
Aktiv _{exp}	8	9
Decoy _{exp}	0	606

- c) PAM-Modell, *OOB*-Datensatz (2600 Komplexe)

	Aktiv _{calc}	Decoy _{calc}
Aktiv _{exp}	37	33
Decoy _{exp}	5	2525

- d) PAM-Modell, Testdatensatz (516 Komplexe)

	Aktiv _{calc}	Decoy _{calc}
Aktiv _{exp}	8	6
Decoy _{exp}	0	502

Tabelle 4.3.: Statistische Parameter zur Beurteilung der Klassifikatoren für die internen und externen Validierungsdaten.

Modell	Datensatz	Genauigkeit	Sensitivität	Präzision	FP-Rate
<i>Random-Selection</i>	OOB	0.99	0.52	0.86	0.0021
<i>Random-Selection</i>	Testdaten	0.99	0.47	1.0	0
<i>PAM</i>	OOB	0.99	0.53	0.88	0.0020
<i>PAM</i>	Testdaten	0.99	0.57	1.0	0

großen Überschuss an negativen Fällen auch ein Klassifikator, der nur negative Werte vorhersagt, eine hohe Genauigkeit hätte. Deswegen ist vor allem wichtig, wie gut die aktiven Verbindungen erkannt werden.

Die Sensitivität (oder Wiederfindungsrate) gibt an, welcher Anteil der positiven Fälle korrekt klassifiziert wird. Sie liegt hier zwischen 0.47 und 0.57. Es wird also jeweils etwa die Hälfte der aktiven Verbindungen korrekt erkannt. Der beste Wert von 0.57 wird für die PAM-Testdaten erreicht. Den schlechtesten Wert (0.47) erhält der Random-Selection-Testdatensatz. Die beiden *OOB*-Sensitivitäten liegen mit 0.52 und 0.53 sehr eng zusammen.

Die Präzision gibt an, welcher Anteil der als positiv vorhergesagten Fälle tatsächlich aktiv ist. Für die beiden externen Testdatensätze ist die Präzision 1, da alle aktiv vorhergesagten Fälle auch tatsächlich aktiv sind. Aus demselben Grund ist in beiden Fällen auch die Falsch-Positiv-Rate 0. Auch für die beiden *OOB*-Validierungen erhält man hohe Präzisionswerte (86% und 88%) sowie sehr kleine Falsch-Positiv-Raten.

Insgesamt erhält man in allen vier Testszenarien ähnliche Ergebnisse, was für die Robustheit der Methode spricht. Die gute Übereinstimmung zwischen *OOB*-Ergebnissen und externen Testergebnissen zeigt, dass der *OOB*-Fehler ein guter interner Qualitätsparameter für die Entwicklung von RF-Modellen ist.

Qualitativ betrachtet sind beide Modelle sehr gut in der Lage, *Decoys* zu erkennen. Im Gegenzug wird dafür aber nur etwa die Hälfte der tatsächlich aktiven Komplexe erkannt. Je nachdem, in welcher Phase des virtuellen Screenings das Modell eingesetzt wird, wäre es zu bevorzugen, wenn das Modell eine etwas höhere Sensitivität hätte (also mehr *Aktive* erkennen würde) auch wenn das zu Lasten der Präzision und der Falsch-Positiv-Rate ginge. Gerade wenn das Modell in einer frühen Phase des Screenings eingesetzt wird, um sicher inaktive Verbindungen herauszufiltern, sollten dabei möglichst wenige potentiell aktive Verbindungen verloren gehen.

Der Grund für die Tendenz, eher Falsch-Negative vorherzusagen als Falsch-Positive,

ist in der Zusammensetzung der Trainingsdaten zu sehen. Wollte man das Modell zu Gunsten einer höheren Sensitivität anpassen, müsste man entweder den Anteil an positiven Fällen im Trainingsdatensatz erhöhen oder die korrekte Vorhersage von positiven Fällen höher gewichten.

4.5.1. Vergleich mit anderen Scoring-Funktionen

Durch den Vergleich mit SFCscore und der zum Docking verwendeten *GlideScore*-Funktion kann man die Leistungsfähigkeit der Klassifikatoren noch besser einordnen.

Im Unterschied zu den binären Klassifikatoren bestimmen die Scoring-Funktionen für jeden Komplex einen genauen Score, der zur Aufstellung einer Rangliste genutzt werden kann. Geht man diese Rangliste nun vom ersten bis zum letzten Platz durch, kann man an jeder Position die oben gezeigten Parameter, wie die Genauigkeit oder Sensitivität berechnen, indem man an dieser Stelle die Grenze zwischen aktiv und inaktiv setzt. Trägt man dann Sensitivität gegen die Falsch-Positiv-Rate auf, erhält man eine sog. ROC-Kurve (Receiver-Operating-Characteristic) [53], die sich zu einem Standardverfahren für die Darstellung und Analyse von Scoring-Ergebnissen gegen Datensätze mit binärer Zielvariable entwickelt hat.

In Abbildung 4.7 sind die Ergebnisse der SFCscore-Funktionen (originale Funktionen und die neu entwickelte 196eqd-Funktion) sowie von *GlideScore* für die beiden externen Testdatensätze dargestellt.

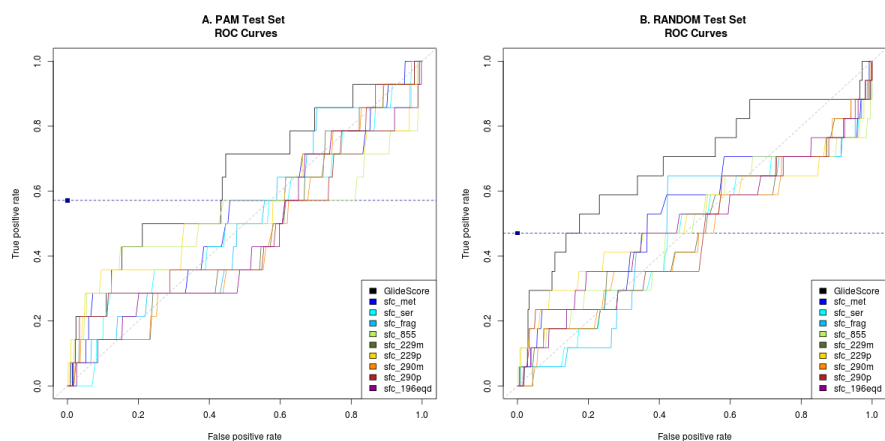


Abbildung 4.7.: ROC-Diagramm für den PAM Testdatensatz (A) und den Random-Selection-Testdatensatz (B). Der Wert für das entsprechende Klassifizierungsmodell ist durch ein blaues Quadrat markiert. Die blaue gestrichelte Linie (horizontal) markiert die Sensitivität des Klassifizierungsmodells.

In einem ROC-Diagramm stellt die Ursprungsgerade die Performance eines rein zufälligen Modells dar. Das Ergebnis ist umso besser, je höher die *Area-Under-the-*

Curve (AUC) ist. Optimal ist eine AUC von 1. Die AUC eines zufälligen Modells ist 0.5.

Die Ergebnisse in Abbildung 4.7 sind dementsprechend nicht zufriedenstellend, da alle Kurven nahe der Diagonalen verlaufen. Das Ergebnis ist jedoch nicht unerwartet, weil der InhA-Datensatz auch in der DuD-Publikation als sehr anspruchsvoll beschrieben ist und auch dort keine der getesteten Funktionen gute Ergebnisse liefern konnte [61]. Die *GlideScore*-Kurve (in schwarz dargestellt) zeigt für beiden Testdatensätze die besten Werte und ist den *SFCscore*-Funktionen überlegen. Beim *Random-Selection*-Testdatensatz ist der Unterschied deutlicher als beim PAM-Datensatz.

Für Klassifizierungsmodelle ist die Klassengrenze durch das Modell festgelegt und kann nicht variiert werden. Somit können die *RandomForest*-Modelle in einem ROC-Plot nur durch einen einzelnen Punkt dargestellt werden. In beiden Abbildungen ist die Performance des jeweiligen Klassifizierungsmodells als blauer Punkt eingezeichnet. Die gestrichelte horizontale Linie markiert die Sensitivität des Klassifizierungsmodells.

Für den PAM-Testdatensatz wird die Sensitivität des *RandomForest*-Modells von den besten Scoring-Funktionen (*GlideScore* und *sfc_met*) bei einer FP-Rate von ca. 0.5 erreicht. Das Klassifizierungsmodell erreicht diesen Wert gänzlich ohne falsch-positive Vorhersagen. Im Fall des *Random-Selection*-Testdatensatzes wird die Sensitivität des Klassifizierungsmodells von *GlideScore* mit einer FP-Rate von etwa 0.2 erreicht, von den *SFCscore*-Funktionen erst deutlich später. In beiden Fällen ist das *RandomForest*-Modell den klassischen Scoring-Funktionen also deutlich überlegen.

4.5.2. Interpretation der RandomForest-Modelle

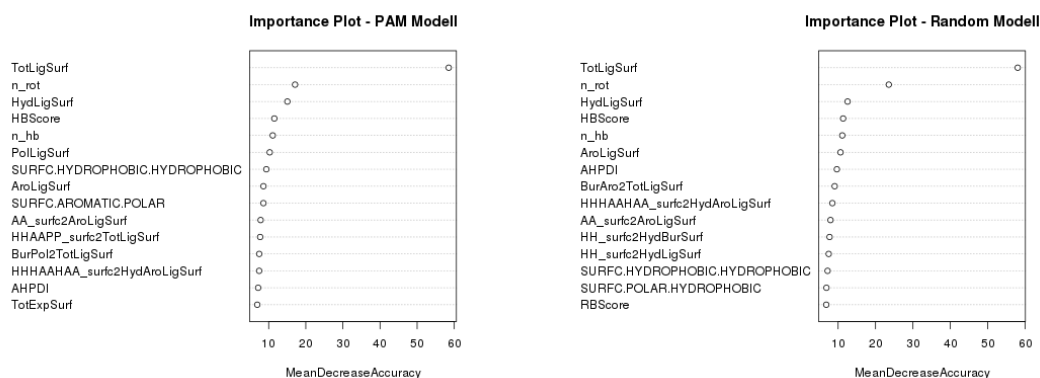
Ein Nachteil der *Random Forest*-Modelle ist, dass die zugrunde liegenden Mechanismen der Klassifizierung nicht direkt zugänglich sind und nicht im Einzelnen nachvollzogen werden können. Dadurch wird die Interpretation der Ergebnisse erschwert.

Eine Möglichkeit, trotzdem einen Eindruck von den entscheidenden Faktoren zu bekommen, sind die sogenannten "Importance Plots". Die Wichtigkeit der Deskriptoren wird dabei durch die Berechnung der "Mean Decrease Accuracy" (MDA) bestimmt (vgl. Seite 59f). In Abbildung 4.8 sind die 15 wichtigsten Deskriptoren der beiden Modelle aufgelistet. Da beide Modelle zu großen Teilen auf derselben Datengrundlage beruhen, ist es nicht überraschend, dass sich die beiden Plots ähnlich sind.

Sehr auffällig ist, dass der *TotLigSurf*-Deskriptor (Gesamtoberfläche des Liganden) mit weitem Abstand die wichtigste Rolle spielt. Das ist problematisch, weil es sich bei dem Deskriptor nicht um einen bindungsspezifischen Parameter handelt. Zudem ist die Ligandoberfläche sehr stark mit der Ligandgröße korreliert. Da die *DUD-Decoys* basierend auf ihrer physikochemischen Ähnlichkeit (u.a. dem Molekulargewicht) zu den aktiven ausgewählt wurden, ist es überraschend, dass der Deskriptor hier eine so prominente Rolle spielt.

Ähnlich verhält es sich mit der Anzahl der rotierbaren Bindungen, die ebenfalls bei der Generierung der inaktiven Verbindungen berücksichtigt wurde und trotzdem eine

4. Klassifizierungsmodelle



a) PAM-Modell

b) Random-Selection-Modell

Abbildung 4.8.: Random-Selection-Modell Importance-Plot zeigt die 15 wichtigsten Deskriptoren des Modells basierend auf der Abnahme der Genauigkeit bei zufälliger Permutation der Werte eines Deskriptors.

entscheidende Rolle bei der Separierung spielt. Danach fällt die relative Wichtigkeit der Deskriptoren schnell unter 10%. Allerdings gibt es eine große Anzahl von Deskriptoren mit einer Wichtigkeit im Bereich von 7 bis 10%. Auch wenn dies als Einzelbeitrag relativ gering ist, basiert die Vorhersage des Modells also insgesamt auf der Summe vieler unterschiedlicher Deskriptoren.

4.6. Zusammenfassung und Diskussion

Im diesem Kapitel wurde gezeigt, dass durch die Entwicklung eines SFCscore-basierten Klassifizierungsmodells im Vergleich zu den ursprünglichen SFCscore-Funktionen sowie der *GlideScore*-Funktion wesentlich höhere Genauigkeiten bei der Unterscheidung von aktiven und inaktiven Verbindungen erreicht werden können. Für die verbesserte Performance gibt es mehrere Gründe.

Zunächst führt sicherlich die proteinspezifische Ableitung der Funktion zu einer verbesserten Performance. Es ist anzunehmen, dass auch eine auf klassischem Wege abgeleitete proteinspezifische Funktion besser abschneiden würde als die allgemeinen SFCscore-Funktionen.

Darüber hinaus bestand hier jedoch, durch den Einsatz des RF-Algorithmus, die Möglichkeit alle SFCscore-Deskriptoren ohne vorhergehende Variablenselektion für die Ableitung eines Modells zu verwenden. Die Analyse der Wichtigkeit der Deskriptoren zeigt zwar, dass einzelne Deskriptoren (vor allem die *Total Ligand Surface*) einen vergleichsweise großen Einfluss auf die Vorhersage haben. Letztlich setzt sich die Gesamtvorhersage aber trotzdem aus sehr vielen unterschiedlichen Deskriptoren zusammen, die zwar einzeln betrachtet einen relativ geringen Beitrag leisten, in der Summe

jedoch letzten Endes die deutlich bessere Performance ermöglichen. Diese Annahme wird dadurch gestützt, dass auch einige der klassischen Funktionen die im Modell wichtigsten Deskriptoren (*TotLigSurf*, rotierbare Bindungen und Wasserstoffbrücken) enthalten (z.B. *sfc_229p* oder *sfc_met*), aber trotzdem nicht an die Vorhersagegenauigkeit des RF-Modells heranreichen.

Der dritte Grund für die bessere Performance ist, dass es durch den Einsatz des Klassifizierungsalgorithmus möglich wird, auch negative Daten bei der Modellableitung mit einzubeziehen. Inaktive Verbindungen sind den anderen SFCscore-Funktionen "unbekannt". Es ist also naheliegend, dass das Klassifizierungsmodell besser in der Lage ist, inaktive Verbindungen zu identifizieren.

Das Einbeziehen der negativen Daten stellt aber auch gleichzeitig die potentiell größte Fehlerquelle dar und schränkt die Anwendbarkeit der Methode ein. Problematisch ist zum einen die Rekrutierung einer ausreichenden Zahl von Negativdaten. In dieser Arbeit wurde aus diesem Grund der DUD-Datensatz verwendet, der allerdings nur experimentell nicht bestätigte Negativdaten enthält. Durch die zunehmende Menge an öffentlich verfügbaren Affinitätsdaten, beispielsweise aus *High-Throughput-Screening* (HTS) Experimenten, wird sich diese Situation in Zukunft sicherlich verbessern.

Die zweite Problematik betrifft die Generierung von Docking-Posen für inaktive Verbindungen. Dies erscheint zunächst paradox, da inaktive Verbindungen definitionsgemäß nicht an der Zielstruktur binden. Zum einen ist es aber so, dass inaktiv nicht notwendigerweise bedeutet, dass die Verbindungen absolut inaktiv sind, sondern nur, dass sie bei einer bestimmten Maximalkonzentration keine Aktivität zeigten. Deswegen kann man beispielsweise bei generischen Ligandenserien, die nur teilweise aktiv sind, auch oftmals den wahrscheinlichsten Bindemodus für die inaktiven Verbindungen mit gewisser Sicherheit abschätzen. Wichtig ist vor allen Dingen, dass die Docking-Methode sorgfältig validiert wird. Die gewissenhafte Zusammenstellung der Negativdaten, mit dem Hauptaugenmerk auf möglichst gesicherten Bindeposen der inaktiven Verbindungen, ist sicherlich ein Schlüsselement bei der Ableitung eines Klassifizierungsmodells.

Aus diesen Betrachtungen lassen sich folgende Rückschlüsse für den Einsatz von Klassifizierungsmethoden ziehen:

Zunächst eignen sich Klassifizierungsmodelle aufgrund der hohen Anforderungen an die Präzision des Dockings nur für den Einsatz als spezialisierte Modelle. Verfügt man jedoch über einen großen Datensatz an negativen Daten, von denen man zusätzlich mit großer Sicherheit die wahrscheinlichsten Bindemoden vorhersagen kann (s.o.), so ist die Entwicklung eines solchen Klassifizierungsmodells als Grundlage für ein virtuelles Screening sehr interessant.

Die Einsatzgebiete von Klassifizierungsmethoden liegen dabei hauptsächlich in den frühen und mittleren Phasen eines virtuellen Screenings, wo ein genaues Ranking der Verbindungen nicht notwendig ist. An dieser Stelle sei jedoch erwähnt, dass es auch Algorithmen gibt, die basierend auf binären Trainingsdaten zu numerischen Vorhersagen

kommen, mit denen sich auch ein Ranking aufstellen lässt. Ein Beispiel einer solchen Methoden ist der *Support Vector Machine*-Algorithmus, bei dem die Wahrscheinlichkeit der Klassenzugehörigkeit über den Abstand zu einer hochdimensionalen Trennebene abgeschätzt wird [86].

Hat man dagegen nur wenige inaktive Verbindungen oder lässt sich deren Bindemodus nur schwer definieren, macht die Ableitung eines solchen strukturbasierten Modells wenig Sinn.

Bei der Entwicklung von target-spezifischen Modellen ist in allen Fällen zu beachten, dass die Modelle auch auf Ligandseite in der Regel in einem eingeschränkten chemischen Raum trainiert wurden. Die Modelle können daher potentiell ungeeignet für die Vorhersagen von neuen Verbindungsklassen sein, vor allem wenn diese einen stark abweichenden Bindemodus aufweisen.

Der RF-Algorithmus an sich ist aufgrund seiner sehr geringen Parameter-Abhängigkeit und seiner Robustheit gegenüber "Rauschen" in den Daten sehr gut zur Entwicklung spezialisierter Modelle geeignet. Dabei ist sowohl die Entwicklung von regressionsbasierten Funktionen als auch von Klassifizierungsmodellen möglich. Mit der Ableitung eines regressionsbasierten RF-Modells beschäftigt sich das folgende Kapitel.

5. SFCscore^{RF} -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten

Der RandomForest-Algorithmus ist auch zur Ableitung von regressionbasierten Scoring-Funktionen geeignet. Sehr schön demonstriert wird dies durch die Funktion *RF-Score* von Ballester und Mitchell [87]. Hier wurden Atom-Atom-Kontakthäufigkeiten als Deskriptoren zur Ableitung einer RF-basierten Scoring-Funktion eingesetzt. Trotz der Einfachheit der zugrunde liegenden Deskriptoren konnte eine überraschend gute Korrelation von $R_P = 0.78$ für den *Cheng*-Datensatz erreicht werden [87].

Die Kombination eines regressionsbasierten RandomForest-Modells mit den SFCscore-Deskriptoren ist deswegen ein sehr interessanter Ansatz.

5.1. Trainingsdatensatz

Zusammenstellung Der verwendete Trainingsdatensatz ist aus dem *PDBbind refined*-Datensatz (Version 2007) [48, 47] abgeleitet. Es wurden zunächst alle Komplexe entfernt, die Bestandteil der geplanten Testdatensätze sind. Dies waren alle 195 Komplexe des *Cheng*-Datensatzes sowie weitere 99 Objekte, die im *CSAR 2010*-Datensatz enthalten sind. Darüber hinaus musste ein Komplex entfernt werden, weil er einen Fehler beim Scoring mit SFCscore verursachte, der nicht unmittelbar behoben werden konnte. Letztendlich enthielt der Trainingsdatensatz 1005 Komplexe. Somit ist es der größte Datensatz, der bisher zur Ableitung einer SFCscore-Funktion verwendet wurde. Er gehört auch generell zu den größten Datensätzen für die Ableitung von empirischen Scoring-Funktionen.

Aufbereitung Zunächst wurden alle Cofaktoren mit Hilfe des Programms SYBYL-X (SYBYL-X 1.0, Tripos International, St. Louis, Missouri, U.S.A.) aus den jeweiligen Proteinen extrahiert und als mol2-Files gespeichert. Wassermoleküle sowie andere Heteroatome (z.B. Komponenten aus dem Kristallisationspuffer) wurden gelöscht. Die Liganden werden in der PDBbind-Datenbank bereits als separate mol2-Files zur Verfügung gestellt, so dass hier keine weitere Bearbeitung notwendig war. Protein, Ligand und

5. SFCscore^{RF} -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten

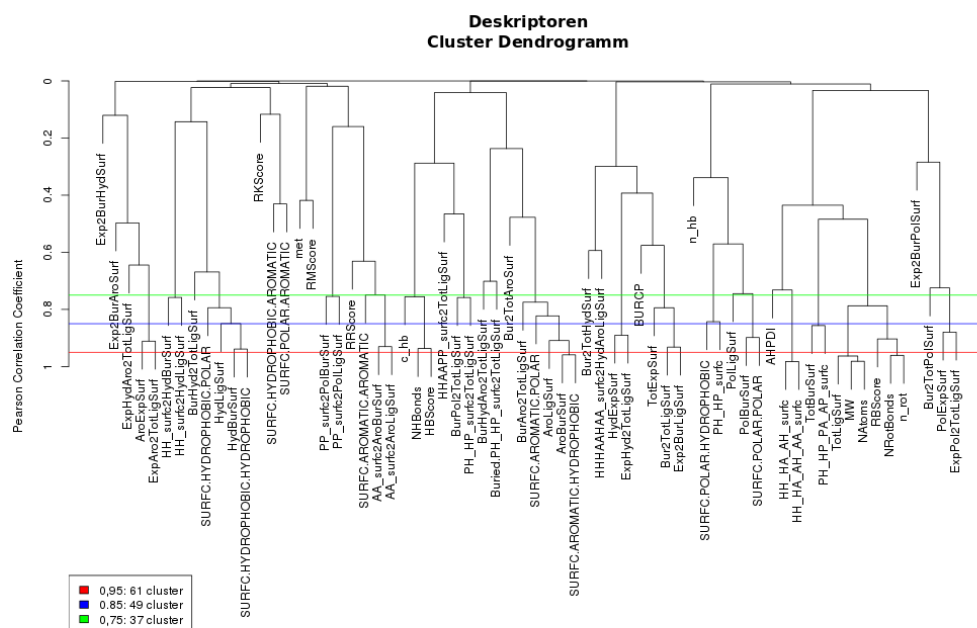


Abbildung 5.1.: Hierarchisches Clustering der 66 SFCscore-Deskriptoren basierend auf dem Pearson Korrelationskoeffizienten im Trainingsdatensatz. Farbige eingezeichnet sind die Korrelationswerte 0,95 (rot), 0,85 (blau) und 0,75 (grün). Würde man diese als Grenzwerte für die Clusterbildung verwenden, würde man 61, 49 oder 37 Cluster erhalten.

ggf. Cofaktoren wurden dann direkt, ohne weitere Optimierungen, zur Berechnung der SFCscore-Deskriptoren verwendet.

5.1.1. Deskriptoren

Der RF-Algorithmus ist grundsätzlich nicht anfällig gegenüber korrelierten Deskriptoren. Die Kolinearität der Deskriptorenmatrix soll an dieser Stelle trotzdem genauer untersucht werden. In Abbildung 5.1 sind alle 66 Deskriptoren basierend auf der Pearson-Korrelation geclustert (vgl. Cluster-Methoden auf Seite 57). Es wurde der Absolutwert des Pearson-Korrelationskoeffizienten verwendet, so dass sowohl negative als auch positive Korrelationen erfasst werden.

Erwartungsgemäß erhält man eine ganze Reihe korrelierter Deskriptoren. Würde man einen Korrelationsgrenzwert von 0,95 anlegen, so würde man 61 Cluster erhalten. Die Deskriptoren, die bei diesem Grenzwert bereits ein Cluster bilden, sind sehr hoch korreliert und enthalten im Wesentlichen dieselbe Information. In der Abbildung sind mit 0,85 und 0,75 noch zwei weitere Grenzwerte eingezeichnet. Bei einem Grenzwert von 0,75

erhält man nur noch 37 Cluster. Insgesamt gibt es also einen nicht geringen Grad an Kolinearität in der Deskriptorenmatrix. Wichtig ist dabei jedoch, dass es sich um 37 sehr kleine Cluster handelt (meist mit 2 oder 3 Mitgliedern) und es keine großen Agglomerate von sehr ähnlichen Deskriptoren gibt. Dies wäre auch für den RF-Algorithmus problematisch. Mit den vielen kleinen Clustern hat der RF-Algorithmus jedoch keine Probleme. Theoretisch können also alle Deskriptoren zur Ableitung des Modells verwendet werden.

Letztlich wurden 63 Deskriptoren zur Ableitung des Modells verwendet. Neben *MW* und *NAtoms*, die aufgrund der trivialen Korrelation zur Affinität schon in den vorigen Kapiteln nicht berücksichtigt wurden, wurde hier zusätzlich der Deskriptor *TotLigSurf* weggelassen. Dieser hatte im vorangegangenen Kapitel einen sehr großen Einfluss auf die Klassifizierungsmodelle. Die Clusteranalyse zeigt, dass er ebenfalls sehr stark mit dem Molekulargewicht korreliert ist.

5.2. Ableitung des RandomForest-Modells

Das *RandomForest*-Modell wurde mit der Statistiksoftware R [84] und dem Paket "randomForest" [85] erstellt. Die theoretischen Grundlagen des Algorithmus sind auf Seite 59 ff. beschrieben. Hier soll deswegen nur auf die praktischen Details eingegangen werden. Neben der Möglichkeit, den kompletten Deskriptorensatz zu verwenden, ist der geringe Parametrisierungsaufwand einer der Hauptvorteile der Methode. Die einzigen Parameter, die festgelegt werden müssen, sind die Zahl der Variablen, aus der die Separierungsvariable an jedem Knoten ausgewählt wird (m_{try}), sowie die Anzahl der Bäume, die generiert werden sollen (N_{Tree}).

Der m_{try} -Wert wurde durch ein systematisches Screening aller möglichen Werte (2 bis 63) bestimmt. Als Auswahlkriterium wurde der *OOB*-Fehler verwendet. Diese Methode ist stark an den von Ballester und Mitchell im Rahmen der RF-Score Publikation bereitgestellten R-Code angelehnt [87]. Diese Analyse ergab einen optimalen m_{try} -Wert von 21, was in gutem Einklang mit der Faustregel steht, dass der m_{try} -Wert für ein Regressionsmodell etwa einem Drittel der Gesamtdeskriptorenzahl entsprechen sollte.

Zudem muss die Anzahl der generierten Entscheidungsbäume (N_{Tree}) festgelegt werden. Eine systematische Untersuchung von Svetnik et al. zeigt, dass eine zu große Anzahl an Bäumen die Qualität des Modells nur noch mäßig erhöht, während die Rechenzeit deutlich ansteigt. Eine Größe von 500 Bäumen wurde dabei als ausreichend für ein robustes Modell befunden [83]. Um das zu bestätigen, wurden für den SFCscore-Datensatz Modelle mit 500, 1000 und 5000 Bäumen erstellt. Erwartungsgemäß lieferten alle drei Modelle in etwa dieselben *OOB*-Ergebnisse, so dass für die weitere Entwicklung ein N_{Tree} -Wert von 500 verwendet wurde.

Da die Ableitung eines RF-Modells eine Zufallskomponente enthält und somit nicht deterministisch ist, wurden zunächst zehn Modelle mit den oben gefundenen Einstellungen ($m_{try} = 21$, $N_{tree} = 500$) abgeleitet. Dadurch sollte überprüft werden, ob es zu erheblichen Schwankungen in der Performance der unterschiedlichen Modelle kommt,

5. SFCscore^{RF} -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten

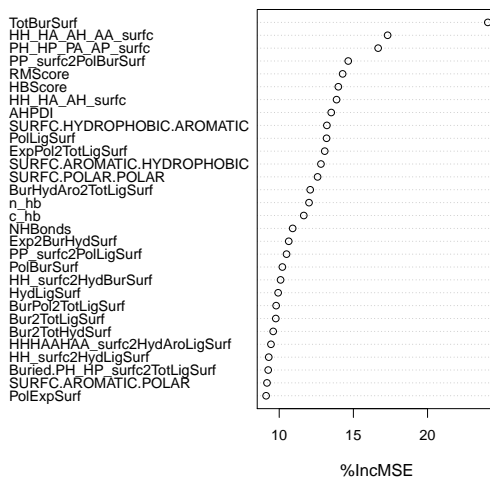


Abbildung 5.2.: Importance Plot der 30 wichtigsten Deskriptoren des SFCscore^{RF}-Modells. Die Wichtigkeit der Deskriptoren anhand der Zunahme des durchschnittlichen Vorhersagefehlers (Mean Standard Error) bei zufälliger Permutation der Deskriptorwerte gemessen. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

was ein Hinweis auf ein instabiles Modell wäre. Die *OOB*- und die Regressionstatistiken waren aber für alle zehn Modelle sehr ähnlich.

Um eine möglichst robuste Funktion zu erhalten, wurde das Modell ausgewählt, das am ehesten den durchschnittlichen Regressionstatistiken entsprach. Alle folgenden Analysen beziehen sich auf dieses Modell, das schlussendlich auch in den Programmcode von SFCscore implementiert wurde und dort als neue Scoring-Funktion zu Verfügung steht (SFCscore^{RF}).

5.2.1. Deskriptor-Importance

Die Wichtigkeit der einzelnen Deskriptoren in SFCscore^{RF} ist im Importance Plot (Abb. 5.2) dargestellt. Im Falle eines Regressionsmodells wird dazu die prozentuale Zunahme des Vorhersagefehlers (*IncMSE*) bei zufälliger Permutation der Deskriptorwerte bestimmt.

Abbildung 5.2 zeigt die 30 wichtigsten Deskriptoren von SFCscore^{RF}. Man sieht, dass der Deskriptor *TotBurSurf* (Total Buried Surface) den größten Einfluss hat. Im Unterschied zu *TotLigSurf*, der bei den Klassifizierungsmodellen sehr wichtig war und hier nicht mehr berücksichtigt wurde, ist die *Total Buried Surface* jedoch nicht rein ligandspezifisch. Sie gibt an, wie gut ein Ligand in der Bindetasche vergraben ist. Der Einfluss im Modell ist deswegen plausibel.

An den Stellen 2, 3 und 4 folgen dann mit *HH_HA_AH_AA_surfc*, *PH_HP_PA_AP_surfc* und *PP_surfc2PolBurSurf* drei Deskriptoren, die die Kontaktflächen zwischen Protein und Ligand genauer charakterisieren. An fünfter und sechster Stelle folgen mit *RMScore* und *HBScore* die beiden ersten Deskriptoren, die gerichtete Interaktionen beschreiben.

Auf diesen ersten Positionen werden also bereits sehr wichtige Beiträge zur Bindungsaffinität, wie Wasserstoffbrücken oder der hydrophobe Effekt, abgedeckt. Am wichtigsten sind aber definitiv die Oberflächen- und Kontaktflächen-Deskriptoren. Außer weiteren Wasserstoffbrücken-Deskriptoren finden sich keine anderen Deskriptoren für gerichtete Wechselwirkungen in den Top 30. Mit dem Deskriptor *PolLigSurf* findet sich erfreulicherweise auch nur ein einziger rein ligandabhängiger Wert in dieser Aufzählung.

Insgesamt haben auch hier, ähnlich wie bei den Klassifizierungsmodellen, viele Deskriptoren einen relativen Einfluss von unter 10%. Die Gesamtvorhersage kommt aber wieder durch die Summe der vielen kleineren Einzeleinflüsse zustande.

Überraschend ist der geringe Einfluss der rotierbaren Bindungen, die nicht unter den wichtigsten Deskriptoren auftauchen. Tatsächlich beträgt die *IncMSE* für die drei "rotatable-bond"-Deskriptoren jeweils nur etwa 8%. Eine mögliche Begründung für den relativen geringen Einfluss ist, dass die Vorhersagen von *SFCscore^{RF}* weniger größenabhängig sind und somit die rotierbaren Bindungen nicht mehr so häufig als Korrekturterm benötigt werden. Der tatsächliche Einfluss von rotierbaren Bindungen wird beim empirischen Scoring also möglicherweise oftmals überschätzt.

Wichtig ist in diesem Zusammenhang, dass die *IncMSE* nicht durch die Verwendung von mehreren korrelierten Deskriptoren beeinflusst wird. Daher wurde testweise ein Modell abgeleitet, bei dem nur ein "rotatable-bond"-Deskriptor (*n_rot*) berücksichtigt wurde. Die relative Wichtigkeit von *n_rot* stieg dabei nur unwesentlich von 8.4% auf 9.1%. Das belegt nochmals, dass der RF-Algorithmus sehr gut mit korrelierten Deskriptoren umgehen kann.

Ebenfalls überraschend ist, dass der *RMScore*-Deskriptor sehr wichtig zu sein scheint, während der *met*-Deskriptor, der eigentlich die Komplexierung eines Metallions besser abbilden sollte, deutlich weniger Relevanz besitzt. Dies könnte ein Hinweis darauf sein, dass die Parametrisierung und Balance der beiden Deskriptoren noch nicht optimal sind. Zum einen sind die Abstandskriterien des *met*-Deskriptors sehr strikt, so dass die Interaktion evtl. zu selten erkannt wird. Die Definition des *RMScore*-Deskriptors ist hingegen weniger strikt. Beispielsweise ist kein Winkelkriterium definiert. Dies führt möglicherweise dazu, dass *RMScore* oftmals anstelle von *met* ausgewählt wird. Eine andere Interpretation ist, dass die Interaktion von Ringsystemen mit Metallkationen wichtiger ist als oftmals angenommen.

Der Importance-Plot bietet insgesamt keinen Anlass, die physikalische Plausibilität des Modells anzuzweifeln.

5.3. Scoring-Performance

5.3.1. Regression und interne Validierung

Tabelle 5.1 zeigt die Regressions- und OOB-Statistiken des Modells. Für die Regression erhält man sehr beeindruckende Korrelationswerte von über 0.9. Der RF-Algorithmus ist also sehr gut in der Lage, die experimentellen Daten mit Hilfe der SFCscore-Deskriptoren zu reproduzieren. Es ist allerdings anzunehmen, dass diese Werte für die Vorhersage von unbekanntem Komplexen nicht gehalten werden können.

Tabelle 5.1.: Regression und interne Validierung des RF-Modells; R : Pearson Korrelationskoeffizient; ρ : Spearman Rangkorrelation; τ : Kendall Rangkorrelation; RMSE: Root Mean Square Error; MedError: Median der Residuen; OOB: *Out-of-Bag*. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

	R	ρ	τ	RMSE	MedError
Trainingsdaten (Regression)	0.97	0.98	0.87	0.60	0.33
OOB-Statistiken	0.71	0.72	0.53	1.49	0.85

Dafür spricht auch der *Out-of-Bag*-Fehler. Dieser kann als eine Art intrinsische Validierung betrachtet werden, da jeder Komplex des Testdatensatzes nur von denjenigen Bäumen vorhergesagt wird, bei denen er nicht Teil der Trainingsdatensatzes war. Der Pearson Koeffizient liegt bei 0.71, der durchschnittliche Vorhersage-Fehler beträgt 1.49 pK_i-Einheiten. Das ist (erwartungsgemäß) eine deutliche Verschlechterung im Vergleich zu den Regressionsstatistiken. In Abbildung 5.3 ist dieser Unterschied in Form von Streudiagrammen veranschaulicht.

Trotzdem stellen auch Korrelationen im Bereich von 0.7 noch eine deutliche Steigerung im Vergleich zu den ursprünglichen Funktionen dar.

5.3.2. Benchmark-Datensätze

Cheng-Benchmark Tabelle 5.2 zeigt die Ergebnisse der ursprünglichen SFCscore-Funktionen und des neuen RF-Modells für die 195 Komplexe des *Cheng*-Datensatzes. Sowohl die drei Korrelationskoeffizienten als auch die Vorhersagefehler sind bei SFCscore^{RF} deutlich besser. Mit einer Pearson-Korrelation von annähernd 0.8 erreicht SFCscore^{RF} dasselbe Niveau wie die RF-Score Funktion von Ballester und Mitchell, für die bis dato die besten Korrelationswerte für diesen Datensatz veröffentlicht wurden [87]. Somit liegt SFCscore^{RF} auch im internationalen Vergleich in der Spitzengruppe.

Eine mögliche Einschränkung dieser Ergebnisse ist, dass der *Cheng*-Datensatz eine systematische Auswahl von Komplexen aus der PDBbind-Datenbank darstellt, die

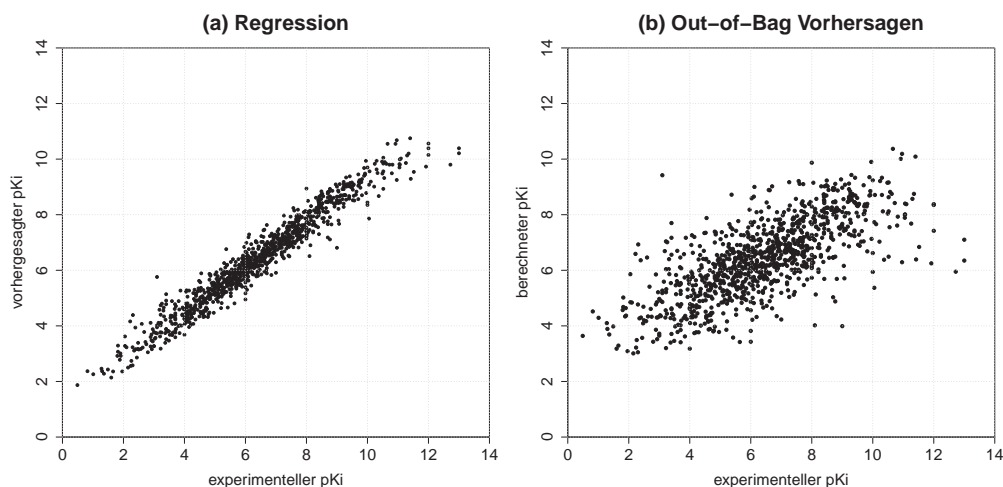


Abbildung 5.3.: Korrelationen zwischen den berechneten und den experimentell bestimmten Affinitätswerten für den Trainingsdatensatz; (a) Vorhersagen des gesamten Modells (Regression) (b) *Out-Of-Bag*-Vorhersagen. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

Tabelle 5.2.: Performance für den *Cheng*-Datensatz für SFCscore^{RF} (sfc_rf) und die ursprünglichen SFCscore-Funktionen. Beschriftung siehe Tabelle 5.1. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

Funktion	R	ρ	τ	RMSE	MedError
sfc_rf	0.78	0.79	0.59	1.56	1.06
sfc_229m	0.59	0.64	0.46	1.93	1.17
sfc_229p	0.59	0.63	0.45	1.96	1.16
sfc_290m	0.59	0.65	0.46	1.95	1.16
sfc_290p	0.62	0.66	0.48	1.88	1.03
sfc_855	0.60	0.64	0.45	1.92	1.15
sfc_frag	0.58	0.65	0.46	2.69	1.56
sfc_met	0.64	0.68	0.50	1.84	1.09
sfc_ser	0.59	0.63	0.45	1.97	1.14

auch zum Training von SFCscore^{RF} verwendet wurde. Der Trainingsdatensatz von SFCscore^{RF} deckt den chemischen Raum des *Cheng*-Datensatzes also sehr gut ab. Dies könnte in der Tat ein Vorteil gegenüber anderen Funktionen sein. Wie groß dieser Ein-

5. SFCscore^{RF} -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten

Tabelle 5.3.: Performance für den CSAR 2010-Datensatz für das RandomForest-Modell (sfc_rf) und die ursprünglichen SFCscore-Funktionen. Beschriftung siehe Tabelle 5.1. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

Funktion	R	ρ	τ	RMSE	MedError
sfc_rf	0.73	0.72	0.53	1.53	0.96
sfc_229m	0.58	0.58	0.41	1.86	1.21
sfc_229p	0.58	0.56	0.40	1.86	1.28
sfc_290m	0.58	0.57	0.40	1.84	1.21
sfc_290p	0.58	0.58	0.41	1.84	1.20
sfc_855	0.59	0.59	0.42	1.78	1.10
sfc_frag	0.62	0.60	0.43	2.61	1.57
sfc_met	0.60	0.59	0.42	1.79	1.12
sfc_ser	0.62	0.62	0.44	1.76	1.17

fluss im konkreten Fall aber tatsächlich ist, ist sehr schwierig abzuschätzen. Im Rahmen einer *Leave-Cluster-Out*-Validierung wird dies später nochmals genauer untersucht.

Dass man eine gute Übereinstimmung zwischen Trainings- und Testdaten auch von einer positiven Seite sehen kann, wird im Kapitel zur "Applicability Domain" (Kapitel 7) thematisiert. Es ist außerdem anzumerken, dass SFCscore^{RF} als einzige der SFCscore-Funktionen keine direkte Überlappung zwischen den Trainingsdaten und dem *Cheng*-Datensatz aufweist.

CSAR 2010-Benchmark Auch für den *CSAR 2010*-Datensatz lässt sich eine deutliche Überlegenheit des neuen Modells feststellen. Die Korrelationswerte von SFCscore^{RF} sind im Vergleich zu den ursprünglichen Funktionen um etwa 0.15 Einheiten besser, so dass die Tendenzen des *Cheng*-Datensatzes bestätigt werden.

Während sich die Vorhersagefehler bei beiden Datensätzen auf demselben Niveau bewegen, sind die Korrelationswerte hier etwas schlechter als beim *Cheng*-Datensatz. Dafür gibt es zwei mögliche Gründe. Zum einen könnte es sich um eine Auswirkung des oben beschriebenen Datensatz-Bias handeln, der zu einer Überschätzung der Performance beim *Cheng*-Datensatz führt. Zum anderen scheint der *CSAR 2010*-Datensatz generell etwas anspruchsvoller zu sein, da alle SFCscore-Funktionen etwas schwächere Werte liefern.

Auch im Vergleich zu den Funktionen, die in der CSAR-Studie selbst getestet wurden, schneidet SFCscore^{RF} sehr gut ab. Die beste Funktion der *CSAR 2010*-Publikation erreicht einen Pearson-Korrelation von 0.76 [51]. SFCscore^{RF} würde sich mit 0.73 an

zweiter Stelle einreihen und gehört somit auch für diesen Datensatz zu den besten Funktionen [3].

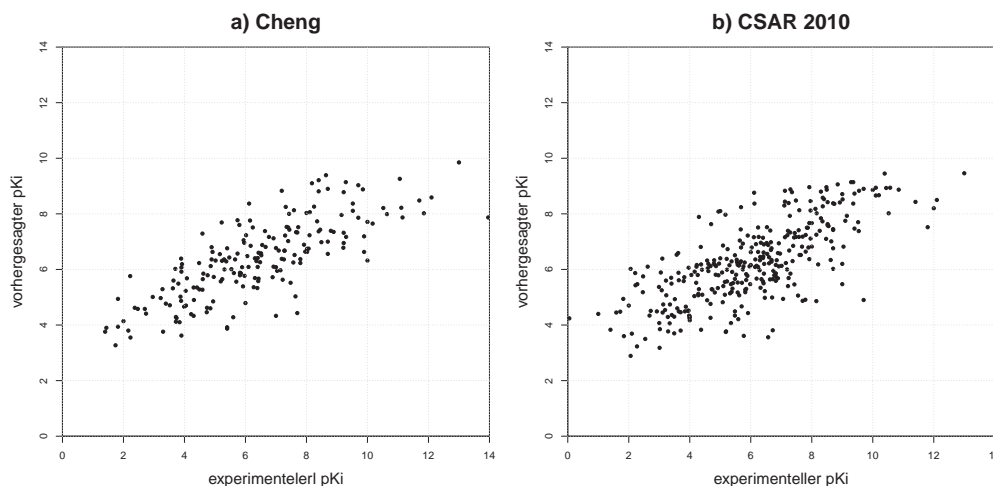


Abbildung 5.4.: Streudiagramme: SFCscore^{RF}-Vorhersage vs. experimentelle Werte für die beiden Testdatensätze *Cheng* (a) und *CSAR 2010* (b). Teilabbildung a) mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

Abbildung 5.4 zeigt die Streudiagramme für die beiden Datensätze. Die Korrelation lässt sich in beiden Fällen erkennen. Ähnlich wie schon in Abbildung 5.3 b) ist die Regressionsgerade jedoch deutlich flacher als die Ursprungsgerade, was bedeutet, dass vor allem niedrig- und hoch-affine Komplexe nicht korrekt vorhergesagt werden - ein Phänomen, das man bei allen SFCscore-Funktionen beobachten kann.

Insgesamt zeigen die besseren Korrelationskoeffizienten und die geringeren Vorhersagefehler, dass SFCscore^{RF}, im Bezug auf die Affinitätsvorhersage für große, heterogen zusammengesetzte Datensätze, eine deutliche Verbesserung der ursprünglichen SFCscore-Funktionen darstellt.

5.3.3. Leave-Cluster-Out Validierung

Mit Hilfe der Leave-Cluster-Out Validierung (LCO-Validierung) soll untersucht werden, inwiefern sich die gemeinsame Datenquelle von SFCscore^{RF}-Trainingsdatensatz und Cheng-Datensatz auf die Ergebnisse der Scoring-Performance auswirkt. Die Methodik wurde von Kramer und Gedeck vorgeschlagen [88].

Die LCO-Validierung ist eine Variante der klassischen Kreuzvalidierung. Dabei wird der Datensatz basierend auf den Targetklassen geclustert. Jeder dieser Cluster, der eine Mindestgröße von 10 Komplexen hat, wird dann von einem Modell vorhergesagt, welches mit allen anderen Komplexen des Datensatzes trainiert wurde. Kramer und Gedeck haben für die Analyse den *PDBbind refined*-Datensatz in der Version von 2010

5. $SFCscore^{RF}$ -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten

verwendet. Um eine bessere Vergleichbarkeit zu gewährleisten, wurde dieser Datensatz auch für die hier gezeigten Analysen verwendet.

Die Ergebnisse für die 23 Cluster sind in Tabelle 5.4 zusammengefasst und in Abbildung 5.5 visualisiert. Man erhält dabei für manche Cluster sehr gute Ergebnisse (z.B. Trypsin), während für andere Klassen praktisch keine Korrelation zu beobachten ist. Es ist allgemein bekannt, dass Scoring-Funktionen in Abhängigkeit von der Zielstruktur schwankende Ergebnisse liefern. Die Studien von Warren [11], Cheng [14] und Wang [49] bestätigen dies. Auch für die ursprünglichen $SFCscore$ -Funktionen konnten Fluktuationen für targetspezifische Datensätze beobachtet werden [46]. Vor diesem Hintergrund sind die Ergebnisse nicht überraschend.

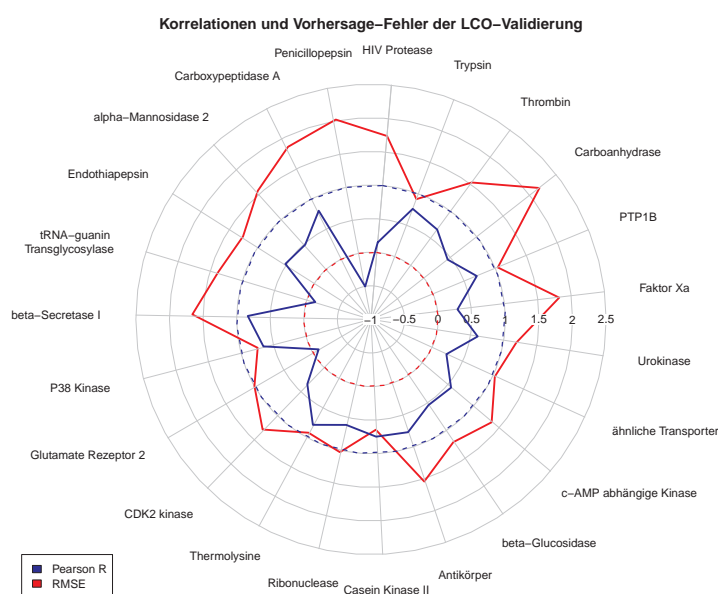


Abbildung 5.5.: Graphische Darstellung der LCO-Validierung für den RandomForest Ansatz. Die blaue Linie stellt den Korrelationskoeffizienten dar, der optimalerweise 1 beträgt (blaue gestrichelte Linie). Die rote Linie steht für den RMSE der Vorhersagen, der möglichst nahe bei Null sein sollte (rote gestrichelte Linie). Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

Die Methodik der LCO-Validierung weist zudem einige kritische Punkte auf, welche die Aussagekraft der Ergebnisse einschränken. Zunächst wird bei dieser Vorgehensweise nicht die eigentliche Funktion $SFCscore^{RF}$ bewertet. Stattdessen wird für jeden Cluster ein eigenes Modell abgeleitet. Durch das Heraushalten von kompletten Targetklassen aus den Trainingsdaten werden teilweise zu pessimistische Einschätzungen bezüglich der Scoring-Performance des eigentlichen Modells getroffen. Am Beispiel des *HIV Protease*-Clusters lässt sich dies gut zeigen.

Bei den 187 *HIV-Protease*-Komplexen handelt es sich ausnahmslos um sehr große Liganden, die größtenteils sehr gut in der Bindetasche vergraben sind. Verwendet man alle *HIV-Protease*-Komplexe für den Testdatensatz, so fehlen solche Beispiele im Trainingsdatensatz. Abbildung 5.6 b) zeigt den großen Unterschied, der dadurch in der Verteilung des *TotBurSurf*-Deskriptors entsteht. Abbildung 5.6 a) zeigt, dass dieser Deskriptor einen sehr großen Einfluss auf die Vorhersagen des LCO-Modells hat. Die Affinitätsvorhersagen für die *HIV-Protease*-Komplexe liegen sehr einheitlich zwischen pK_i 7 und 8.5, während der Datensatz in Wirklichkeit ein Affinitätspektrum von 5 bis 12 abdeckt (Abbildung 5.6 c)). Eine Unterscheidung zwischen hoch- und niedrig-affinen Komplexen gelingt nicht, weil alle *HIV-Protease*-Komplexe aufgrund ihrer großen vergrabenen Oberfläche als eher hochaffin angesehen werden.

Verwendet man hingegen die *OOB*-Vorhersagen von SFCscore^{RF} zur Vorhersage des *HIV Protease*-Clusters, so erhält man immerhin eine Pearson-Korrelation von 0.60. Durch das Einbeziehen von ähnlichen Komplexen in den Trainingsdatensatz lässt sich die Scoring-Performance also deutlich verbessern. In diesem Sinne provoziert die Methodik der *LCO*-Validierung schlechte Ergebnisse und führt nicht unbedingt zu einer realistischen Einschätzung.

Ein zweiter kritischer Punkt betrifft die Größe der Cluster sowie den Affinitätsbereich, den die einzelnen Cluster abdecken. Aus Tabelle 5.4 geht hervor, dass nur die ersten vier Cluster eine Größe von mehr als 50 Komplexen aufweisen. Die restlichen 19 Cluster haben maximal 30, in den meisten Fällen sogar weniger als 20 Mitglieder. Hier besteht die Gefahr, dass die statistische Auswertung keine signifikanten Ergebnisse mehr liefert.

Ein gutes Beispiel hierfür ist Cluster P (*Glutamat-Rezeptor 2*). Hier wird eine negative Korrelation von -0.10 gemessen. Der RMSE ist mit 1.00 pK_i -Einheiten aber eigentlich gut. Der Grund für die sehr schlechte Korrelation ist der enge Affinitätsbereich, in dem viele der Komplexe liegen. 9 der 13 Komplexe liegen innerhalb von 1.5 pK_i Einheiten. Die Aufstellung einer korrekten Reihenfolge in einem so engen Rahmen kann nicht gelingen. Je kleiner der Datensatz, desto höher ist auch das Gewicht von einzelnen Ausreißern.

Die *LCO*-Ergebnisse sollten also sehr vorsichtig interpretiert werden. Die übergreifende Aussage, dass es wesentliche Unterschiede in der Performance gegen verschiedene Targets gibt, gilt jedoch auch für SFCscore^{RF} .

5. SFCscore^{RF} -Ein RandomForest-Modell zur Vorhersage von Bindungsaffinitäten

Tabelle 5.4.: Ergebnisse der Leave-Cluster-Out Validierung wie von Kramer und Gedeck vorgeschlagen [88]. Jeder Cluster wird mit einem Modell vorhergesagt, welches mit allen anderen Clustern trainiert wurde. R_P : Pearson Korrelationskoeffizient; RMSE: Root Mean Square Error. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

Zielstruktur	Cluster	Komplexe	R_P	RMSE
HIV Protease	A	187	0.15	1.74
Trypsin	B	73	0.76	0.91
Thrombin	C	53	0.66	1.53
Carboanhydrase	D	57	0.45	2.18
PTP1B	E	32	0.70	1.04
Faktor Xa	F	32	0.30	1.82
Urokinase	G	29	0.60	1.19
verschiedene Transportproteine	H	29	0.24	1.03
c-AMP abhängige Kinase	I	17	0.57	1.36
β -Glucosidase	J	17	0.54	1.20
Antikörper	K	16	0.77	1.54
Casein-Kinase II	L	15	0.75	0.64
Ribonuclease	M	15	0.61	1.03
Thermolysin	N	14	0.79	0.93
CDK2 Kinase	O	13	0.35	1.30
Glutamat-Rezeptor 2	P	13	-0.10	1.00
P38 Kinase	Q	13	0.65	0.74
β -Secretase I	R	12	0.84	1.66
tRNA-Guanine-Transglycosylase	S	12	-0.13	1.38
Endothiapepsin	T	11	0.52	1.27
α -Mannosidase 2	U	10	0.48	1.55
Carboxypeptidase A	V	10	0.80	1.85
Penicillopepsin	W	10	-0.50	2.02
Cluster mit 4-9 Komplexen	X	379	0.66	1.51
Cluster mit 2-3 Komplexen	Y	334	0.58	1.53
einzelne Komplexe	Z	316	0.47	1.69

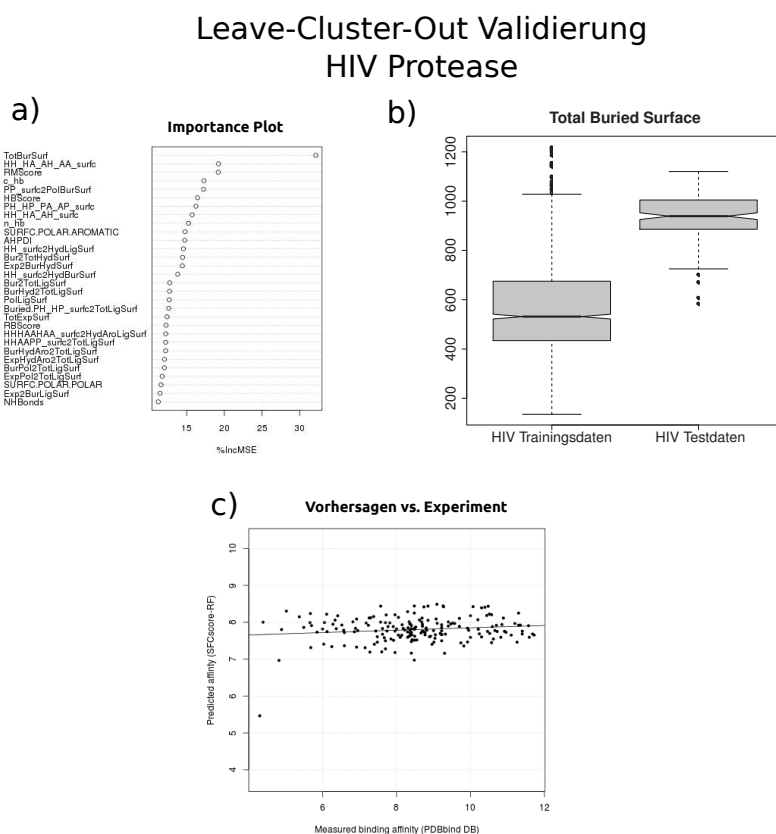


Abbildung 5.6.: (a) Importance Plot für das Modell zur Vorhersage des HIV-Protease-Clusters. (b) Der Boxplot vergleicht die Verteilung des TotBurSurf-Deskriptors für den HIV-Testdatensatz der LCO Validierung mit dem entsprechenden Trainingsdatensatz. Der Deskriptor TotBurSurf gibt die vergrabene Oberfläche des Liganden an. (c) Das Streudiagramm zeigt den Zusammenhang zwischen vorhergesagten und experimentell bestimmten Affinitäten des HIV-Protease-Clusters. Teilabbildung b) wurde mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

6. CSAR 2012 Exercise

Bei den bisher verwendeten Benchmarks (*Cheng* und *CSAR 2010*) handelt es sich um Kristallstrukturdatensätze, die im Bezug auf die Proteinstrukturen sowie auf die Liganden sehr heterogen zusammengesetzt sind.

Beim realen Einsatz von Scoring-Funktionen gibt es jedoch einige entscheidende Unterschiede zu diesen Testszenarien. Zum einen werden Scoring-Funktionen meist nicht zur Bewertung von Kristallstrukturen, sondern in der Regel zur Bewertung von Docking-Posen eingesetzt. Zum anderen werden so gut wie immer proteinspezifische Datensätze bewertet, so dass die target-spezifische Performance einer Funktion im realen Einsatz von erheblicher Relevanz ist.

Im *CSAR 2012*-Benchmark [89] wurde versucht, diesen speziellen Anforderungen besser Rechnung zu tragen. Dazu wurden von den Organisatoren vier Benchmark-Datensätze zusammengestellt. Jeder Datensatz bestand aus einer oder mehreren Ligandserien, die in der Industrie oder in der akademischen Forschung als Inhibitoren für das jeweilige Target entwickelt wurden.

Da keine Kristallstrukturen zur Verfügung gestellt wurden, musste der komplette Prozess, von der Aufbereitung der Ligand- und Proteinstrukturen über das Docking bis hin zum Scoring und Rescoring, von den Teilnehmern selbst durchgeführt werden. Die Anforderungen dieses Benchmarks gehen also über das reine Scoring hinaus. Der Grundgedanke der Initiatoren war, dass dadurch sowohl die Docking-Performance als auch die Scoring- und die Ranking-Performance getestet werden können.

In Kapitel 2.3 ist beschrieben, dass SFCscore bei der Erkennung von Bindemoden (Docking-Performance) nicht zu den besten Funktionen gehört. Auch bei SFCscore^{RF} liegt der Fokus eindeutig auf der Vorhersage von Affinitäten und dem Einsatz als Rescoring-Funktion. Da die Generierung korrekter Bindeposen jedoch unerlässlich für das Scoring ist, wurde der Docking-Schritt hier mit dem Programm Glide und der korrespondierenden *GlideScore*-Funktion durchgeführt. Zum Scoring wurden dann alle SFCscore-Funktionen einschließlich SFCscore^{RF} eingesetzt.

6.1. Datensätze

Insgesamt wurden von den CSAR-Organisatoren vier Datensätze zur Verfügung gestellt [89]. Aufgrund des engen Zeitrahmens bis zu den jeweiligen Abgabeterminen wurden nur drei dieser Datensätze bearbeitet.

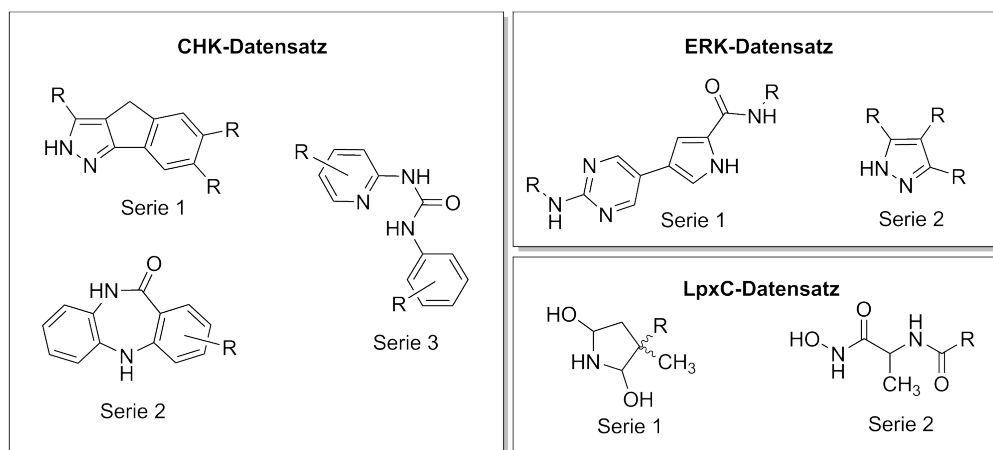


Abbildung 6.1.: Ligandklassen der drei CSAR-Datensätze

CHK 1 Checkpoint-Kinase 1 (CHK1) ist eine Kinase, die an der Zellzykluskontrolle beteiligt ist. Die zu adressierende Bindetasche ist die ATP-Bindetasche. Der Datensatz besteht aus 47 Verbindungen, die sich in drei Strukturklassen unterteilen lassen. Die erste Serie besteht aus einer Gruppe aromatischer Trizyklen mit einer Pyrazol-Substruktur. Das gemeinsame Strukturmerkmal von Serie 2 ist ein Trizyklisches Grundgerüst das aus zwei Phenyl- und einer zentralen 1,4-Diazepan-5-on-Einheit besteht. Die dritte Gruppe bilden unterschiedlich substituierte Harnstoffderivate (Serie 3). Die Grundstrukturen sind in Abbildung 6.1 dargestellt.

ERK 2 Extracellular-signal-regulated Kinase 2 (ERK2) ist ebenfalls in verschiedenen Phasen des Zellzyklus aktiv. Auch dieser Liganddatensatz besteht aus Inhibitoren, die die ATP-Bindetasche adressieren sollen. Der Datensatz besteht aus 39 Verbindungen, die sich in zwei generische Serien aufteilen lassen (Abbildung 6.1). Serie 1 hat als gemeinsame Strukturmerkmale eine 2-Aminopyrimidin- und eine Pyrrol-Einheit, sowie eine sekundäres Amid. Bei Serie 2 handelt es sich um heterogen substituierte Pyrazole. Außerdem gibt es noch 4 Liganden, die keiner der beiden Gruppen angehören und auch keine weiteren gemeinsamen strukturellen Merkmale haben.

LpxC UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamin Deacetylase (LpxC) ist eine Amidase, die an der bakteriellen Lipopolysaccharidsynthese beteiligt ist. Die Bindetasche enthält ein Zink-Ion als Cofaktor. Der Datensatz besteht aus 16 strukturell sehr ähnlichen Liganden, die entweder eine Hydroxamsäure-Gruppe oder eine Hydantoin-Gruppe zur Komplexierung des Zink-Ions enthalten und ansonsten sehr ähnlich substituiert sind (Abbildung 6.1).

6.2. Docking-Validierung

Für die erfolgreiche Affinitätsvorhersage ist eine korrekte Vorhersage des Bindemodus unerlässlich. Um das Docking für die drei Targetklassen zu validieren, wurde ein Cross-Docking mit Hilfe bekannter Kristallstrukturen durchgeführt. Hierzu wurde zunächst die *Protein Data Bank* [64] nach geeigneten Kristallstrukturen durchsucht. Es wurden nur Komplexe mit einer guten Auflösung von weniger als 2.5 Å verwendet. Alle PDB-Strukturen wurden visuell auf Vollständigkeit der Bindetasche und die korrekte Positionierung des Liganden überprüft.

Setup: Cross-Docking Für das Cross-Docking wurden die Liganden aus der Bindetasche extrahiert und mit LigPrep ¹ aufbereitet. Zur Vorhersage der Protonierungszustände wurde der von den CSAR-Autoren angegebene pH-Wert des Assays, mit einem Spielraum von +/- 0.5 pH-Einheiten, verwendet. Für diese Berechnungen wurde das Programm Epik ² eingesetzt. Protomere und Tautomere wurden beim Docking berücksichtigt. Zusätzliche Stereoisomere wurden hingegen nicht generiert. Alle weiteren Parameter wurden auf ihren Standardeinstellungen belassen.

Die Proteinstrukturen wurden mit dem "Protein Preparation Wizard" ³ aufbereitet. Die dabei verwendeten Einstellung und das Vorgehen sind analog zu den auf Seite 51 beschriebenen.

Für das Glide-Docking wurde der XP-Modus verwendet. Abweichend von den Standardeinstellungen wurden alle OH-Gruppen in der Bindetasche als "frei rotierbar" definiert. Jeder Ligand wurde auf diese Weise in jede Proteinstruktur gedockt. Je Protein-Ligand-Kombination wurden 10 Docking-Lösungen ausgegeben.

6.2.1. Chk1

Für Chk1 wurden elf passende PDB-Strukturen gefunden. Die Liganden dieser Komplexe sind größtenteils Adenin-Analoga. Als einzige Substanzklasse aus dem CSAR-Datensatz ist die Serie 1 (Pyrazol-Derivate) vertreten, für die mit 2E9N ein strukturell eng verwandter Vertreter existiert. Die anderen Verbindungen enthalten meist mehrere stickstoffhaltige Heterozyklen wie Piperazin (1ZYS), Pyrimidin (2BR1) oder Indazol (2C3L). 1ZLT ist ein Azepion und ähnelt in der Grundstruktur somit der CSAR Serie 2. Harnstoffderivate (Serie 3) finden sich keine.

Abbildung 6.2 zeigt den RMSD-Levelplot als Ergebnis des Cross-Dockings. Zur Auswahl der besten Docking-Lösung wurde die Scoring-Funktion *GlideScore* verwendet, da sie bessere Ergebnisse lieferte als die Funktion DSX [79].

¹LigPrep, Version 2.5, Schrödinger, LLC, New York, NY, 2012 [76, 77]

²Epik, Version 2.3, Schrödinger, LLC, New York, NY, 2012

³Schrödinger Suite 2012 Protein Preparation Wizard; Epik version 2.3, Schrödinger, LLC, New York, NY, 2012; Impact version 5.8, Schrödinger, LLC, New York, NY, 2012; Prime version 3.1, Schrödinger, LLC, New York, NY, 2012 [78]

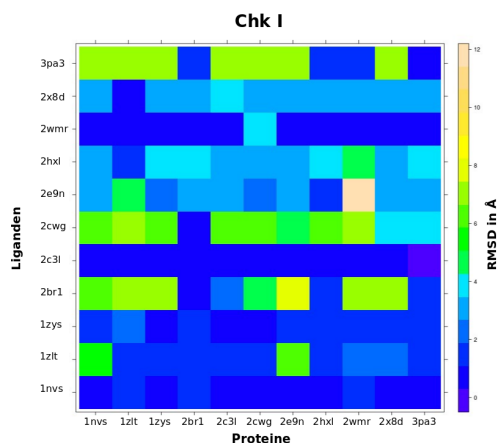


Abbildung 6.2.: Cross-Docking Ergebnisse für den Chk-Datensatz für die von *GlideScore* am besten bewerteten Lösungen

Sowohl Redocking als auch das Cross-Docking funktionieren größtenteils gut. Das Redocking liefert in den meisten Fällen RMSD-Werte < 2 Å. Nur in einem Fall (2CWG) beobachtet man eine Abweichung von mehr als 2 Å. Beim Cross-Docking fällt auf, dass die Liganden 2BR1, 2CWG und 3PA3 in vielen Proteinstrukturen einen relativ hohen RMSD von 3-4 Å haben.

Der Ligand 2E9N, der Ligand sehr eng mit der CSAR Serie 1 verwandt ist, erhält in den meisten Fällen RMSD-Werte zwischen 2 und 4 Å. Eine Analyse dieser Posen zeigt jedoch, dass vorhergesagte Bindemodus des 2E9N-Liganden in den meisten Fällen sehr gut ist. Die erhöhte RMS-Abweichung kommt durch die unterschiedliche Positionierung des Cyclohexyl-Restes zustande, der außerhalb der Bindetasche liegt. Abbildung 6.3 zeigt die Docking-Pose des Liganden im Vergleich zur Kristallstruktur.

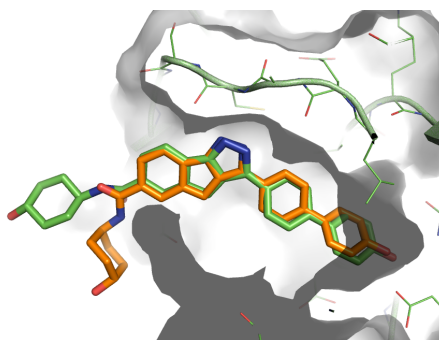


Abbildung 6.3.: Grün: Kristallstruktur von 2E9N; Orange: Glide Docking-Pose

Als Proteinstruktur für das Docking wird von den CSAR-Organisatoren ebenfalls 2E9N empfohlen. Im Cross-Docking erhält man für diese Proteinstruktur nur zwei Liganden mit deutlich erhöhten RMSD-Werten. Diese Liganden unterscheiden sich

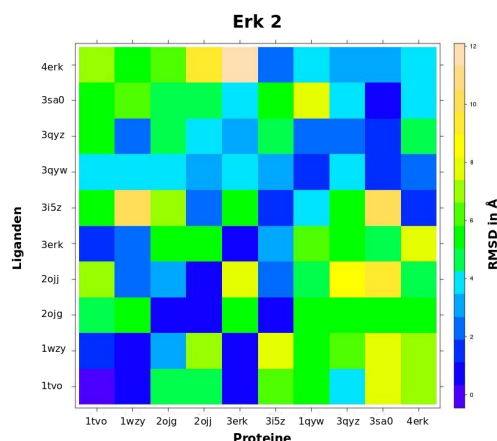


Abbildung 6.4.: Cross-Docking Ergebnisse für den Erk-Datensatz für die von *GlideScore* am besten bewerteten Lösungen

strukturell jedoch deutlich von den CSAR-Liganden, so dass für das Docking der CSAR-Liganden die vorgeschlagene Struktur 2E9N mit den hier ermittelten Docking-Einstellungen verwendet wird.

6.2.2. Erk2

Für Erk2 wurden zehn ausreichend gut aufgelöste und vollständige PDB-Strukturen gefunden. Der Datensatz enthält auch Vertreter aus den beiden großen Ligandenserien des CSAR-Datensatzes. Vertreter der Serie 1 finden sich in 3I5Z, 3ERK, 4ERK und 3QYW. Die Pyrazol-Substruktur der Serie 2 ist in den Liganden 2OJG und 2OJJ zu finden. Bei den anderen Verbindungen handelt es sich um zwei Pyridazin-Derivate (1TVO, 1WZY), ein Deoxyadenosin-Derivat (3QYZ) sowie ein hydroxyliertes Xanthen-Derivat (3SA0).

Die Aufbereitung und das Docking der Strukturen erfolgten analog zum Vorgehen für den Chk-Datensatz. Der Levelplot für die besten Posen (lt. *GlideScore*) ist in Abbildung 6.4 dargestellt.

Die Ergebnisse fallen im Vergleich zum Chk-Datensatz deutlich schlechter aus. Vor allem beim Cross-Docking werden viele Fälle mit großen RMS-Abweichungen beobachtet. Ein Grund dafür kann sein, dass die Bindetasche flexibel ist und sich bei der Bindung dem Liganden anpasst. Für diese These spricht, dass das Redocking in den meisten Fällen sehr gut funktioniert und die falschen Vorhersagen beim Docking in "fremde" Bindetaschen auftreten.

Die Einführung eines van-der-Waals Skalierungsfaktors sowie die Anwendung des "Induced Fit"-Protokolls für Glide⁴ führten nicht zu einer Verbesserung der Ergebnisse.

⁴Schrödinger Suite 2012 Induced Fit Docking protocol; Glide version 5.8, Schrödinger, LLC, New York, NY, 2012; Prime version 3.1, Schrödinger, LLC, New York, NY, 2012, [90, 91, 92]

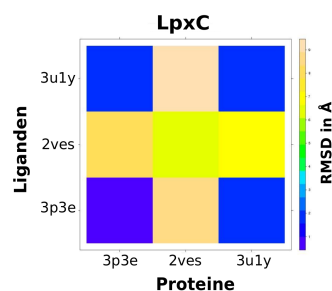


Abbildung 6.5.: Cross-Docking Ergebnisse für den LpxC-Datensatz

Es wurden stattdessen durch die Vergrößerung des Konformationsraums und der Freiheitsgrade beim Docking eher noch mehr falsche Posen generiert.

Ein weiteres Problem beim Docking dieser Verbindungen ist, dass die meisten eine große Anzahl potentieller Wasserstoffbrücken-Donoren und -Akzeptoren haben. Dadurch werden sehr viele verschiedene Interaktionsmuster in der Bindetasche möglich. Das Docking ist also aus verschiedenen Gründen als problematisch anzusehen.

Die sechs Strukturen, die den CSAR-Verbindungen am ähnlichsten sind, erhalten für das Docking in die Bindetasche von 3I5Z einen durchschnittlichen RMSD-Wert von 2.01 Å. 3I5Z wird auch von den CSAR-Organisatoren zum Docking empfohlen. Man kann deswegen davon ausgehen, dass für das Docking der CSAR-Liganden in die Struktur 3I5Z keine essentiellen Induced-Fit-Mechanismen berücksichtigt werden müssen. Aus diesem Grund wurde entschieden, das Docking des CSAR-Datensatzes mit den ursprünglichen Glide-Einstellungen in die PDB-Struktur 3I5Z durchzuführen.

6.2.3. LpxC

Für LpxC konnten nur drei Kristallstrukturen gefunden werden. Bei allen drei Liganden handelt es sich um Hydroxamsäure-Derivate. Die Substitutionsmuster der Liganden 3P3E und 3U1Y sind denen der CSAR-Liganden sehr ähnlich. Der Ligand aus 2VES ist hingegen mit zwei sterisch anspruchsvollen Gruppen substituiert und unterscheidet sich deswegen topologisch sehr deutlich.

Die Datenaufbereitung und das Docking wurden wieder analog zum Vorgehen bei den anderen beiden Datensätzen durchgeführt. Bei der Ligandenaufbereitung gab es jedoch einen wesentlichen Unterschied: Das Programm Epik kann bei der Generierung von Protonierungszuständen "Metal-Binding-States" berechnen. Dabei handelt es sich um Protonierungszustände, die beim gegebenen pH-Wert eigentlich nicht vorkommen, die aber durch die Komplexbildung eines Metall-Ions bei der Bindung des Liganden begünstigt werden können. Die "Metal-Binding-States" spielen vor allem beim Docking der Hydantion-Derivate (CSAR Serie 1) eine Rolle.

Für die beiden Strukturen 3P3E und 3U1Y funktioniert sowohl das Cross-Docking als auch das Redocking problemlos (Abbildung 6.5). Das Docking von 2VES ist hingegen in

keinem Fall erfolgreich. Aufgrund der großen strukturellen Unterschiede zu den CSAR-Liganden kann dies jedoch in diesem Fall vernachlässigt werden.

Die Analyse der Bindeposen von 3P3E und 3U1Y zeigt, dass der Bindemodus in beiden Fällen durch die Topologie der Liganden und der Bindetasche sowie durch die Schlüsselinteraktion der Metallkomplexierung durch die Hydroxamsäure-Gruppe sehr genau definiert ist. Man kann deswegen davon ausgehen, dass auch das Docking der CSAR-Verbindungen in die Struktur 3P3E keine Probleme bereiten wird.

6.3. Scoring-Validierung

Im zweiten Schritt sollte die Scoring-Performance von SFCscore für die drei Targets überprüft werden. In einem realen Szenario würde man ebenso vorgehen, um herauszufinden welche Funktion die verlässlichsten Ergebnisse liefert.

Zur Validierung des Scorings wurde die ChEMBL-Datenbank [93] nach Verbindungen mit experimentellen Affinitätswerten für die drei Targets durchsucht. So konnten die folgenden Datensätze generiert werden.

- CHK: 13 Verbindungen (pK_i-Werte)
- ERK: 41 Verbindungen (pK_i-Werte)
- LpxC: 42 Verbindungen (IC₅₀-Werte)

Im Fall von LpxC konnten nicht ausreichend viele Verbindungen mit pK_i-Werten gefunden werden. Es wurden deswegen 42 Verbindungen mit IC₅₀-Werten ausgewählt, die aber, da sie alle aus demselben Assay stammen, ebenfalls eine solide Grundlage für die Scoring-Validierung darstellen.

Die Liganden wurden mittels der validierten Methode in die entsprechende Proteinstruktur gedockt. Für jeden Liganden wurden wieder 10 Lösungen generiert. Danach wurde ein Rescoring mit DSX durchgeführt. Die am besten bewertete Posen von Glide und DSX wurden jeweils mit SFCscore bewertet.

Wie schon bei der Docking Validierung lieferten die *GlideScore*-Posen auch hier die etwas besseren Ergebnisse. In Tabelle 6.1 sind die resultierenden Korrelationskoeffizienten für alle drei Datensätze zusammengefasst.

Für den Chk-Datensatz werden sehr gute Korrelationen erhalten. Alle SFCscore-Funktionen liefern hier bessere Korrelationen als die beiden weit verbreiteten Funktionen *DSX* und *GlideScore*. Da der Datensatz jedoch nur 13 Komplexe enthält, müssen diese Ergebnisse vorsichtig interpretiert werden.

Für den Erk-Datensatz beobachtet man große Schwankungen (zwischen 0.39 und 0.74) und insgesamt etwas schlechtere Korrelationen als beim Chk-Datensatz. Die beiden besten Funktionen sind *sfc_rf* und *sfc_frag*. Das sind auch die einzigen beiden Funktionen, die besser abschneiden als die Vergleichsfunktionen *GlideScore* und *DSX*.

Tabelle 6.1.: Pearson Korrelationskoeffizienten von SFCscore, *GlideScore* und *DSX* für die drei ChEMBL-Datensätze. Die Auswahl der Docking-Posen basiert auf der Bewertung von *GlideScore*.

Funktion	Chk1	Erk2	LpxC	Funktion	Chk1	Erk2	LpxC
sfc_229m	0.90	0.50	0.64	sfc_met	0.82	0.48	0.57
sfc_290m	0.90	0.48	0.63	sfc_855	0.78	0.61	0.52
sfc_229p	0.88	0.45	0.57	sfc_rf	0.89	0.73	0.52
sfc_290p	0.87	0.39	0.61	-DSX	0.73	0.61	0.27
sfc_frag	0.82	0.74	0.55	- <i>GlideScore</i>	0.72	0.64	0.55
sfc_ser	0.84	0.47	0.56				

Im Fall von LpxC ist die Performance innerhalb von SFCscore wieder sehr konstant. Die Werte liegen zwischen 0.52 und 0.64, also etwa in dem Bereich, der auch in den Evaluierungen mit den diversen Testdatensätzen (*CSAR 2010* und *Cheng*) ermittelt wurde. Die SFCscore-Funktionen schneiden größtenteils etwas besser als *GlideScore* und deutlich besser als *DSX* ab. Die Performance von SFCscore^{RF} ist mit $R_p = 0.52$ dagegen eher enttäuschend, da die neue Funktion damit im SFCscore-internen Vergleich zu den schwächeren Funktionen zählt.

6.4. CSAR-Docking

Zur Aufbereitung der CSAR-Datensätze wurden zunächst die als Smiles-Strings zur Verfügung stehenden Liganden mit Hilfe von LigPrep⁵ aufbereitet. Die Aufbereitung umfasste die Konvertierung in dreidimensionale Strukturen, sowie die Generierung von Protomeren und Tautomeren. Bei der Generierung von 3D-Strukturen wurden, wo erforderlich, mehrere energetisch günstige Ringkonformationen erstellt. Die Ringkonformationen wurden dann nochmals visuell überprüft. In einigen Fällen (v.a. bei den Chk1-Liganden) mussten manuell zusätzliche Ringkonformationen generiert werden, um passenden Docking-Posen zu erhalten. Bei der Generierung von Protomeren und Tautomeren wurden die von den CSAR-Organisatoren vorgegebenen pH-Werte (± 0.5 pH-Einheiten) der jeweiligen Assays zur Berechnung der Protonierungszustände verwendet.

Für das Docking wurden die Proteinstrukturen, die während der Docking-Validierung (Kapitel 6.2) aufbereitet wurden, sowie die dort validierten Einstellungen verwendet. Existierten für einen Liganden mehrere Protomere oder Tautomere oder wurden unterschiedliche Isomere generiert, so wurden diese zunächst separat gedockt. Bei der Posen-

⁵Version 2.5, Schrödinger, LLC, New York, 2011

Auswahl wurde dann je Ligand nur eine Pose berücksichtigt, egal wieviele verschiedene Varianten ursprünglich gedockt wurden. In jedem Docking-Run wurden zehn Lösungen generiert und mit *GlideScore* und *DSX* bewertet.

Eine grundlegende Annahme beim Docking war, dass die generischen Ligandserien in einem konservierten Bindemodus vorliegen müssen. Eine visuelle Überprüfung der am besten bewerteten Posen ergab jedoch, dass dies beim Chk- sowie beim Erk-Datensatz nicht immer der Fall war. Deswegen wurde letzten Endes für diese beiden Datensätze eine manuelle Auswahl der Posen vorgenommen. Dazu wurden pharmakophorartige Kriterien (auf Basis der bekannten Kristallstrukturen) definiert. Im Falle von Chk mussten die Liganden zwei Wasserstoffbrücken zum Proteinrückgrat (Cys85 und Glu87) bilden. Auch für den Erk-Datensatz wurden zwei Wasserstoffbrücken zu den Aminosäuren Met106 und Asp104 als Kriterien definiert.

Für den größten Teil der Liganden ließ sich unter den zehn Glide-Posen mindestens eine Pose finden, die den Kriterien entsprach. Wurden mehrere Posen gefunden, wurde diejenige mit dem besseren *GlideScore* ausgewählt. In seltenen Fällen ließ sich keine passende Pose finden. Da eine Affinitätsvorhersage auf Basis eines höchstwahrscheinlich nicht korrekten Bindemodus nicht sinnvoll ist, wurde diese Liganden nicht mit SFCscore bewertet und stattdessen als *inaktiv* deklariert.

Beim Chk-Datensatz war das bei sechs Verbindungen der Fall. Beim Erk-Datensatz wurden vier Liganden nicht berücksichtigt. Beim Docking der LpxC-Liganden gab es hingegen keine Probleme. Hier war auch keine manuelle Posen-Auswahl notwendig.

Die Posen wurden dann mit SFCscore bewertet. In der nun folgenden Auswertung liegt der Schwerpunkt auf der Bewertung der Scoring-Performance. Allerdings ist dies oftmals sehr eng mit den Erfolgsraten beim Docking verknüpft, so dass auch die Docking-Ergebnisse genauer analysiert werden.

6.5. Scoring-Performance

Tabelle 6.2 zeigt die Scoring-Statistiken von SFCscore für die drei gedockten CSAR-Datensätze. Es sind jeweils die Korrelationskoeffizienten nach Pearson und Spearman sowie die RMSE-Werte angegeben.

Chk1 Basierend auf der Scoring-Validierung (Kapitel 6.3) waren für den Chk-Datensatz gute Ergebnisse zu erwarten. Gerade vor diesem Hintergrund sind die tatsächlichen Werte überraschend schlecht. Die besten Funktionen sind *sfc_229m* und *sfc_290m*, die eine Pearson Korrelation von 0.38 erreichen. Die anderen Funktionen liegen zwischen 0.18 und 0.30. Die schlechtesten Ergebnisse liefert *sfc_rf* ($R = -0.02$). Eine genauere Analyse zeigt, dass vor allem für die Ligandserie 3 (Harnstoffderivate) teilweise hohe negative Korrelationen erzielt werden (Abbildung 6.6 (a)). Für Serie 1 erhält man hingegen relativ gute Ergebnisse (mit Werten zwischen 0.4 und 0.6). Für

Tabelle 6.2.: Scoring-Ergebnisse für die CSAR-Datensätze. R : Pearson Korrelationskoeffizient; ρ : Spearman Korrelationskoeffizient; RMSE: Root Mean Square Error. Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

	CHK 1			ERK 2			LpxC		
	R	ρ	RMSE	R	ρ	RMSE	R	ρ	RMSE
sfc_rf	-0.02	-0.02	1.28	0.49	0.45	1.13	0.90	0.76	0.55
sfc_229m	0.38	0.31	1.25	0.45	0.47	1.11	0.56	0.02	1.06
sfc_229p	0.30	0.11	1.59	0.35	0.37	1.96	0.70	0.48	0.73
sfc_290m	0.38	0.30	1.25	0.44	0.45	1.13	0.52	0.02	1.05
sfc_290p	0.23	0.15	1.31	0.44	0.46	1.16	0.58	0.10	0.91
sfc_855	0.19	0.17	1.24	0.47	0.44	1.30	0.55	0.43	0.81
sfc_frag	0.24	0.20	1.27	0.54	0.53	1.50	0.80	0.57	1.10
sfc_met	0.18	0.12	1.36	0.44	0.48	1.29	0.63	0.43	0.89
sfc_ser	0.22	0.17	1.29	0.46	0.48	1.13	0.51	0.35	0.96

Serie 2 sind die Korrelationen sehr inkonsistent. Einige Funktionen, wie sfc_229m/p oder sfc_frag, erreichen hier immerhin Werte von über 0.4.

Die schwachen Korrelationen werden durch die Vorhersagefehler etwas relativiert. Diese sind mit Werten zwischen 1.24 (sfc_855) und 1.59 (sfc_229p) pK_i-Einheiten eigentlich zufriedenstellend und liegen unter den Werten die in den großen Benchmark-Datensätzen erreicht wurden. Es stellt sich deswegen, ähnlich wie schon bei der LCO-Validierung, die Frage, wie aussagekräftig die Korrelationswerte an dieser Stelle sind. In Abschnitt 6.9 wird dies ausführlicher diskutiert.

Eine Erklärung für die großen Unterschiede zwischen den Ligandklassen liefert die Analyse der Dockingposen (vgl. Abschnitt 6.8).

Erk2 Die Korrelationen für den Erk-Datensatz sind deutlich besser als die des Chk-Datensatzes. Mit Werten zwischen 0.35 und 0.54 erreichen sie allerdings ebenfalls nicht das erhoffte Niveau. In Anbetracht der Docking- und der Scoring-Validierung ist dieses Ergebnis jedoch weniger überraschend. Die beiden Funktionen, die in der Scoring-Validierung noch sehr gute Vorhersagen gezeigt haben (sfc_rf und sfc_frag), können diese hier leider nicht bestätigen. Die Vorhersagefehler sind mit Werten zwischen 1.11 und 1.50 jedoch gut.

Im Unterschied zu den Chk-Daten lassen sich keine eindeutigen Unterschiede zwischen den Ligandserien beobachten (Abb. 6.6 (b)). Serie 1 wird etwas besser vorhergesagt als

Serie 2. Ein großer Unterschied besteht nur bei sfc_855. Serie 3 besteht, wie oben beschrieben, nur aus 4 Liganden, weswegen die Korrelationen nicht aussagekräftig sind. Auch bei diesem Datensatz hat die Qualität der Docking-Posen einen wesentlichen Einfluss auf das Scoring (siehe 6.8).

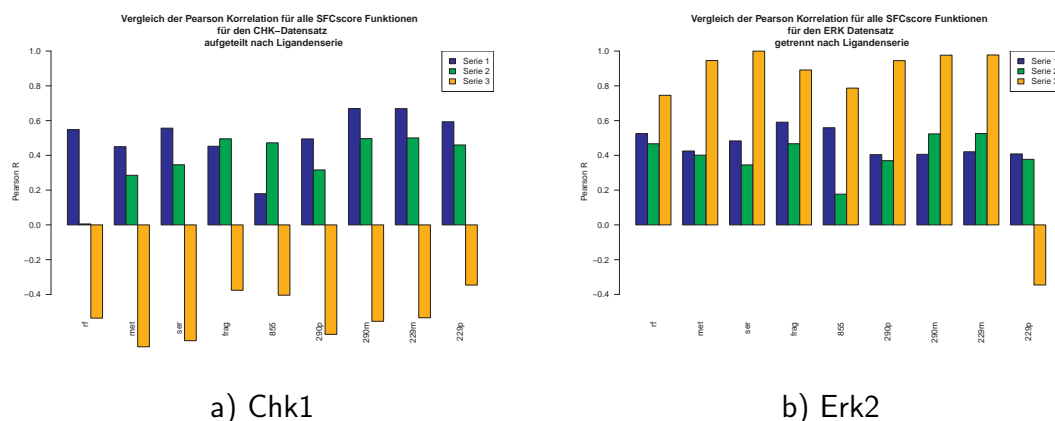


Abbildung 6.6.: Ligandserien-spezifische Pearson-Korrelationen für den Chk-Datensatz (a) und den Erk-Datensatz (b).

LpxC Für den LpxC-Datensatz werden mit Korrelationen zwischen 0,52 und 0,90 insgesamt die besten Ergebnisse ermittelt. Für die meisten Funktionen liegt der durchschnittliche Vorhersagefehler unter einer pKi-Einheit, was ebenfalls sehr gut ist. Allerdings waren in diesem Datensatz nur 8 (von insgesamt 16) Verbindungen überhaupt aktiv. Eine Auswertung für die einzelnen Ligandenserien ist in diesem Fall nicht möglich, da alle aktiven Verbindungen zur Gruppe der Hydroxamsäure-Derivate gehören.

6.6. Erkennung von inaktiven Substanzen (Ranking-Performance)

Unter den 47 Liganden des Chk-Datensatzes waren 9 inaktive Verbindungen. Die auf Basis des Dockings als "inaktiv" deklarierten Liganden waren jedoch alle aktiv. Im Fall von Erk waren alle 39 Verbindungen tatsächlich aktiv. Folglich waren auch hier die als "inaktiv" deklarierten Substanzen aktiv. Im LpxC-Datensatz war die komplette Gruppe der Hydantoine inaktiv, während alle Hydroxamsäure-Derivate aktiv waren.

Auf Basis der SFCscore-Vorhersagen ist ebenfalls nicht erkennbar, dass die inaktiven Verbindungen eher schlechtere Scores erhalten, so dass man insgesamt sagen muss, dass es hier weder beim Docking noch beim Rescoring gelungen ist, die inaktiven Verbindungen zu identifizieren.

Die Gründe dafür sind vielfältig. Ein Punkt ist, dass SFCscore nicht auf das Erkennen von falschen und ungünstigen Bindeposen oder inaktiven Substanzen trainiert ist.

In diesem speziellen Fall kommt hinzu, dass sehr ähnliche Liganden bewertet werden sollten, bei denen kleine Veränderungen in den Substitutionsmustern den Unterschied zwischen "aktiv" und "inaktiv" ausmachen. Sofern aus dieser unterschiedlichen Substitution nicht deutlich andere Bindemoden resultieren, ist es für eine Scoring-Funktion generell sehr schwierig, diese Unterscheidung zu treffen. Dementsprechend ist die Wahrscheinlichkeit größer, die Inaktiven schon beim Docking zu identifizieren, was hier ja auch - leider ohne Erfolg - versucht wurde.

Beim LpxC-Datensatz gibt es noch eine weitere Erklärung, warum die Hydantoine nicht als inaktiv erkannt wurden. Wie oben beschrieben, wurde die Hydantoin-Serie sowohl der protonierte als auch der deprotonierte Zustand beim Docking berücksichtigt, obwohl dieser beim pH-Wert des Assays (7.0) eigentlich nicht auftritt. Mit dem Programm *MoKa* [94, 95, 96] wurde ein pK_S -Wert für die Hydantion-Gruppe von etwa 10 berechnet. Die deprotonierten Spezies wurden trotzdem beim Docking mit Glide berücksichtigt, weil die Möglichkeit besteht, dass durch die verbesserte Zinkkomplexierung die deprotonierter Form in der Bindetasche begünstigt ist. Tatsächlich wurden die Bindeposen der deprotonierten Form beim Docking von Glide besser bewertet, obwohl ein Strafterm für den Protonierungszustand berücksichtigt wurde. Dementsprechend wurden diese Posen für das Rescoring mit SFCscore ausgewählt. Hätte man stattdessen die protonierten Liganden verwendet, so wären schon anhand der Docking-Scores die Hydroxamsäuren als die aktivere Verbindungsklasse erkannt worden. Auch diese Vorhersage ist aber letzten Endes sehr schwierig zu treffen, was man auch daran sieht, dass die Hydantoine gezielt als Inhibitoren für die LpxC-Bindetasche entwickelt wurden [89].

6.7. CSAR 2012 Exercise: Ergebnisse aller Teilnehmer

Am CSAR Benchmark Exercise 2011-2012 haben insgesamt 20 Forschungsgruppen teilgenommen. In der Veröffentlichung von Damm-Ganamet et al. [97] wurden die Ergebnisse aller Teilnehmer zusammengefasst. Von den CSAR-Organisatoren wurden außerdem noch zusätzliche Daten zur Verfügung gestellt. Die folgenden Betrachtungen basieren auf diesen beiden Quellen.

Die Analyse zeigt, dass insgesamt gesehen die besten Korrelationen für den LpxC-Datensatz erhalten werden. Mehr als die Hälfte der teilnehmenden Funktionen erreichen hier zufriedenstellende Person-Korrelationen von über 0.7. SFCscore^{RF} gehört mit 0.9 aber auch in diesem Vergleich zu den besten Funktionen. Für den Chk1-Datensatz wurden hingegen nicht nur für SFCscore, sondern auch insgesamt sehr schlechte Korrelationen ermittelt. Die beste Methode erreicht eine Korrelation von 0.38. Es ist also in keinem der über 40 von den Teilnehmern verwendeten Ansätze gelungen, den Chk-

Datensatz korrekt vorherzusagen. Auch der Erk-Datensatz konnte nur von zwei Methoden mit Korrelationen von über 0.6 vorhergesagt werden. SFCscore liegt dabei im relativen Vergleich mit Werten zwischen 0.35 und 0.54 im Mittelfeld.

Die Ranking-Performance und die Erkennung von inaktiven Verbindungen ist ebenfalls bei keiner der teilnehmenden Funktionen für alle Datensätze zufriedenstellend. Die besten Ergebnisse wurde für den Urokinase-Datensatz erhalten, der hier nicht bearbeitet wurde. Der LpxC-Datensatz war in dieser Beziehung die größte Herausforderung.

Die für SFCscore beobachteten Tendenzen stimmen also mit den allgemeinen Ergebnissen von Damm-Ganamet überein. Für detailliertere Auswertungen bezüglich des Vergleichs der CSAR 2012-Teilnehmer wird auf [97] verwiesen.

6.8. Einfluss der Docking-Posen

Ein wichtiger Einflussfaktor für die Scoring- und Ranking-Performance ist die Korrektheit der verwendeten Dockingposen. Nach der Einreichung aller Ergebnisse wurden von den CSAR-Organisatoren für jeden der Datensätze einige Komplex-Kristallstrukturen zur Verfügung gestellt. Für den Chk-Datensatz waren es 14, für den ERK-Datensatz 12 und für den LpxC-Datensatz 4 Strukturen. Mit Hilfe dieser Strukturen kann die Qualität der verwendeten Docking-Posen beurteilt werden.

Tabelle 6.3 zeigt die RMSD-Werte der hier verwendeten Glide-Dockingposen im Vergleich mit den durchschnittlichen Ergebnissen der anderen Teilnehmer [97]. Man sieht, dass das Docking sowohl für den Chk- als auch für den Erk-Datensatz sehr anspruchsvoll war und dass insgesamt relativ schlechte Ergebnisse erzielt wurden. Für den LpxC-Datensatz sind die Docking-Ergebnisse dagegen sehr gut (sowohl für das hier geführte Glide-Docking als auch im CSAR-Durchschnitt). Die hier verwendeten Glide-Posen sind für alle drei Datensätze etwas besser als der jeweilige Gesamtdurchschnitt.

Chk1 Für den Chk-Datensatz wird für die hier verwendete Glide-Methode ein durchschnittlicher Docking-Fehler von 2.56 Å erhalten. Es bestehen jedoch sehr große Unterschiede zwischen den drei Ligandserien.

Vor allem für die Harnstoffderivate (Serie 3) werden viele falsche Posens vorhergesagt. Dabei werden zwar die geforderten Wasserstoffbrücken-Kriterien erfüllt, der Ligand liegt jedoch um 180° gedreht in der Bindetasche, weil das "falsche" Stickstoffatom der Harnstoffgruppe als Wasserstoffbrücken-Donor agiert. Abbildung 6.7 a) stellt die Situation exemplarisch für den Liganden chk_35 dar (RMSD = 8.00 Å). Insgesamt sind Kristallstrukturen für fünf Harnstoffderivate verfügbar, davon erhalten drei einen RMSD von > 5 Å.

Auch in der Serie 2 gibt es Unsicherheiten bezüglich der Dockingposen. Die Lactam-Substruktur dieser Serie kann die geforderten Interaktionen mit dem Proteinerückgrat in zwei verschiedenen Posen erfüllen. Die Konformation des tricyclischen Ringsystems ist dabei ausschlaggebend dafür, welche Variante bevorzugt wird. Abbildung 6.7 b) zeigt

Tabelle 6.3.: Zusammenfassung der RMSD-Werte (in Å) für den Vergleich zwischen gedockten und kristallisierten Strukturen. Die Daten wurden von den CSAR-Organisatoren zur Verfügung gestellt. [97]

Datensatz	Docking-Posen	# Liganden	Min	25%tile	Median	75%tile	Max
Chk	diese Arbeit (Glide)	14	0.48	1.15	2.56	5.52	8.00
	CSAR-Durchschnitt	14	0.23	1.44	3.50	5.90	20.56
Erk	diese Arbeit (Glide)	12	0.77	1.82	4.61	6.64	10.60
	CSAR-Durchschnitt	12	0.39	2.26	5.03	7.00	13.67
LpxC	diese Arbeit (Glide)	4	0.83	0.83	0.86	0.99	1.30
	CSAR-Durchschnitt	4	0.36	0.89	1.09	1.47	13.96

dies beispielhaft für den Liganden chk_20. Der Effekt tritt bei zwei der vier Kristallstrukturen der Serie 2 auf.

Für die Serie 1 werden hingegen sehr gute Dockingposen erhalten. Für alle fünf Kristallstrukturen konnten passende Posen generiert werden. Es ist sicher kein Zufall, dass für diese Gruppe auch die besten Affinitätsvorhersagen erhalten wurden.

Erk2 Für den Erk-Datensatz sind die Docking-Ergebnisse insgesamt noch etwas schlechter als für den Chk-Datensatz. Das Hauptproblem besteht hier darin, dass alle Liganden, unabhängig aus welcher Serie, viele stickstoffhaltige funktionelle Gruppen (meist Heterozyklen) enthalten. Die beiden geforderten Wasserstoffbrücken können deswegen meist von unterschiedlichen Substrukturen erfüllt werden, so dass kein konsistenter Bindemodus vorliegen muss. Die Analyse der Kristallstrukturen bestätigt dies.

In Abbildung 6.8 a) ist exemplarisch der Ligand erk_37 dargestellt. Die geforderten Wasserstoffbrücken werden in der Docking-Pose von einer Pyrazolon-Gruppe gebildet, während in der Kristallstruktur eine Aminopyrazol-Substruktur für die Wechselwirkungen verantwortlich ist. Der RMSD-Wert beträgt dementsprechend 10.60 Å.

Insgesamt konnten von den zwölf kristallisierten Liganden nur fünf mit einem RMSD-Wert von unter 2.5 Å gedockt werden. Der Median der RMSD-Werte aller CSAR-Teilnehmer beträgt 5.03 Å.

LpxC Für den LpxC-Datensatz werden sehr gute Docking-Posen erhalten. Abbildung 6.8 b) zeigt den Vergleich von kristallisierter Pose und Docking-Pose des Liganden lpxc_2. Der Bindemodus ist durch die Schlüsselinteraktion mit dem Cofaktor Zink und die enge Topologie der Bindetasche vorgegeben, so dass es hier zu keinerlei Problemen

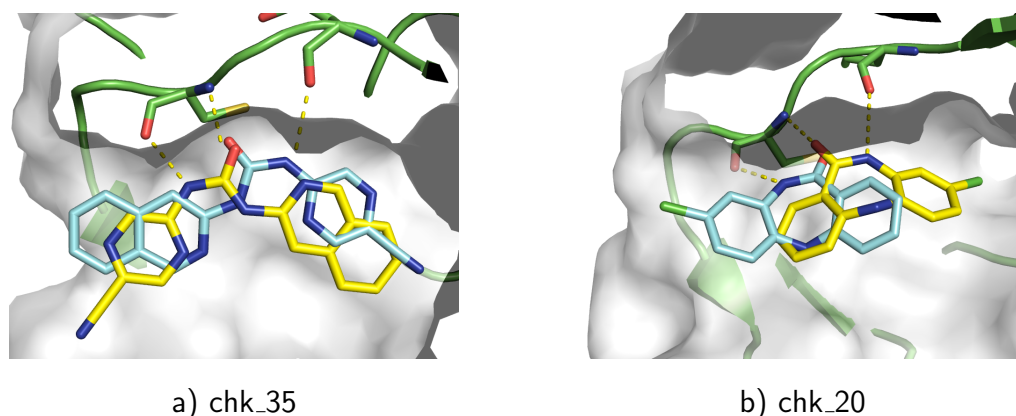


Abbildung 6.7.: a) Vergleich der gedockten (gelb) und der kristallisierten (cyan) Pose des Liganden chk_35. Der sehr hohe RMSD-Wert von 8.00 Å kommt zustande, weil die Backbone-Wasserstoffbrücke von unterschiedlichen Stickstoffatomen der Harnstoff-Substruktur gebildet wird. b) Der Ligand chk_20 liegt ebenfalls in einem um 180° gedrehten Bindemodus vor, weil der Lactamring die geforderten Wasserstoffbrücken in zwei unterschiedlichen Konformationen erfüllen kann. Der RMSD-Wert beträgt hier 6.10 Å. Teilabbildung a) mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society.

beim Docking kommt. Das stellt sicherlich eine wichtige Grundlage für die guten Affinitätsvorhersagen von SFCscore dar.

Im Fall von Erk und Chk wird also die Scoring-Performance erheblich durch die Verwendung nicht-korrekturer Docking-Posen überlagert. Weil für die meisten Verbindungen keine Kristalldaten zur Verfügung stehen, ist es schwierig, diesen Einfluss genau zu quantifizieren.

Die beste Abschätzung erhält man aus dem Vergleich der Scores der kristallisierten Posen mit denen der entsprechenden gedockten Lösungen. Abbildung 6.9 zeigt dies für den Erk2- und den Chk1-Datensatz. Die Kristallstrukturen liefern in fast allen Fällen die besseren Vorhersagen (durchschnittlich um etwa 0.25 pK_i-Einheiten). Einzige Ausnahme sind die Vorhersagen von sfc_frag für den Erk-Datensatz.

Erwähnenswert ist dabei, dass nicht nur falsch vorhergesagte Posen einen negativen Einfluss auf die Affinitätsvorhersage haben, sondern dass teilweise auch relativ kleine Unterschiede zwischen gedockter und kristallisierter Pose eine Auswirkung haben.

Ein Beispiel hierfür ist der Ligand chk_6, ein Imidazol-Derivat bei dem der RMSD-Wert zwischen Docking- und Kristall-Pose nur etwa 0.5 Å beträgt (Abb. 6.10). Der vorhergesagte Score von SFCscore^{RF} beträgt für die gedockte Pose 7.0 und für die Kristallstruktur 7.5. Der experimentelle Affinitätswert liegt mit pK_i = 8.8 nochmals etwas darüber. Kleine Unterschiede in der Positionierung des Liganden beeinflussen die Deskriptorenverteilung also schon so, dass deutlich Unterschiede in den Vorhersagen

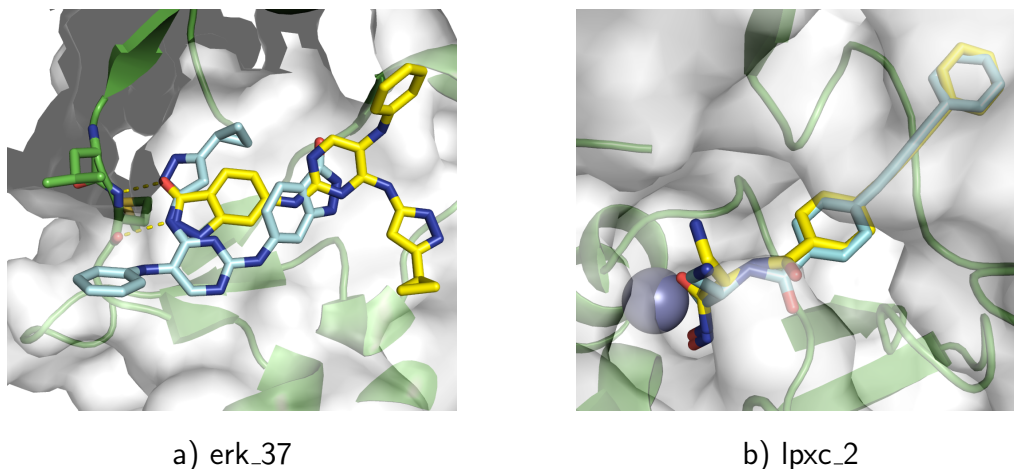


Abbildung 6.8.: Vergleich von kristallisierter Pose (cyan) und Docking-Pose (gelb) von a) erk_37 (RMSD = 10.60 Å) und b) lpxc_2 (RMSD = 0.32 Å).

zustande kommen können.

Im Fall von chk_6 sind die Werte der Wasserstoffbrücken-Deskriptoren in der gedockten Struktur etwas schlechter, weil die Winkel und Abstände für die Interaktionen nicht so optimal sind, wie in der Kristallstruktur. Beispielsweise hat der *HBScore*-Deskriptor in der Kristallstruktur einen Wert von 2.78, während er bei der gedockten Pose nur 2.28 beträgt. Diese Diskrepanz reicht selbstverständlich noch nicht aus um den Unterschied zwischen den beiden Vorhersagen vollständig zu erklären; die Effekte summieren sich jedoch. Da *SFCscore^{RF}* deutlich mehr Deskriptoren verwendet als die anderen *SFCscore*-Funktionen, ist davon auszugehen, dass sich dies hier stärker auswirkt. Eine Erklärung für dieses Verhalten ist, dass die *SFCscore*-Funktionen anhand von Kristallstrukturen trainiert wurden und somit auf die dort auftretenden Abstände und Winkel optimiert sind.

Vor diesem Hintergrund wäre es interessant, eine Funktion mit gedockten Komplexen zu trainieren. Dabei muss allerdings sichergestellt werden, dass die Docking-Lösungen sehr nahe an der Wirklichkeit liegen. Beispielsweise könnte man, ausgehend von Kristallstrukturen eine Art Minimierung in der Bindetasche vornehmen. Dadurch bliebe der Bindemodus erhalten, würde aber mit den Parametern des Docking-Algorithmus optimiert. Trainiert man eine Funktion auf einem solchen Datensatz, sollte diese bezüglich der Dockinglösungen robuster sein, wäre aber andererseits auf den verwendeten Docking-Algorithmus spezialisiert.

Es wäre auch interessant zu untersuchen, ob solche Phänomene speziell bei der Kombination von Glide und *SFCscore* auftreten oder ob es sich um ein generelles Problem bei der Vorhersage von Docking-Posen mit *SFCscore*-Funktionen handelt.

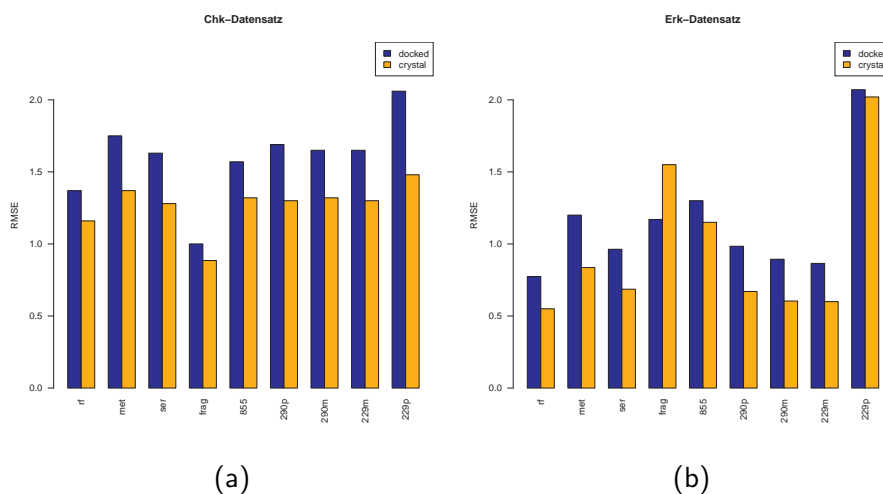


Abbildung 6.9.: Vergleich der RMSE-Werte für die Vorhersagen von SFCscore für die kristallisierten Posen und die entsprechenden Docking-Posen; (a) Chk-Datensatz (b) Erk-Datensatz.

6.9. Statistische Validität

Neben den Korrelationen sind in Tabelle 6.2 auch die RMSE-Werte der Vorhersagen eingetragen. Diese liegen größtenteils zwischen 1.0 und 1.5 pK_i-Einheiten. Damit ist die durchschnittliche Vorhersagegenauigkeit beim CSAR 2012-Datensatz besser als in den beiden großen Benchmark-Datensätzen (*Cheng* und *CSAR 2010*). Die Korrelationswerte des *CSAR 2012*-Datensatzes sind hingegen größtenteils unbefriedigend. Ähnlich wie schon bei der Auswertung der LCO-Validierung stellt sich deswegen die Frage, inwieweit die Korrelationen statistisch belastbar sind.

Angeregt durch die Gesamtergebnisse der *CSAR 2012 Exercise* hat sich auch Heather Carlson, die Initiatorin des *CSAR 2012-Benchmarks*, mit dieser Fragestellung beschäftigt und beleuchtet in einer begleitenden Publikation einige interessante Aspekte, die bei der Auswertung und dem Vergleich von Scoring-Ergebnissen oftmals nicht ausreichend berücksichtigt werden [98].

Unter anderem schlägt sie eine Formel vor, mit der die theoretisch notwendige Datensatzgröße berechnet werden kann, die benötigt wird, damit der Korrelationsunterschied zwischen zwei Funktionen als statistisch signifikant angesehen werden kann:

$$N \approx 4 * (1 - R^2)^2 * (z_{\alpha/2} / \Delta R)^2 + 3 \quad (6.1)$$

Dabei steht N für die Datensatzgröße, R für den kleineren der zu vergleichenden Pearson-Korrelationswerte und ΔR für den Korrelationsunterschied. Der Term $z_{\alpha/2}$ gibt das gewünschte Konfidenzniveau an. Ein Wert von 1.0 entspricht dem Konfidenzniveau von 67%, 1.64 wäre ein Konfidenzniveau von 90% und 1.96 entsprechen 95% Konfidenz

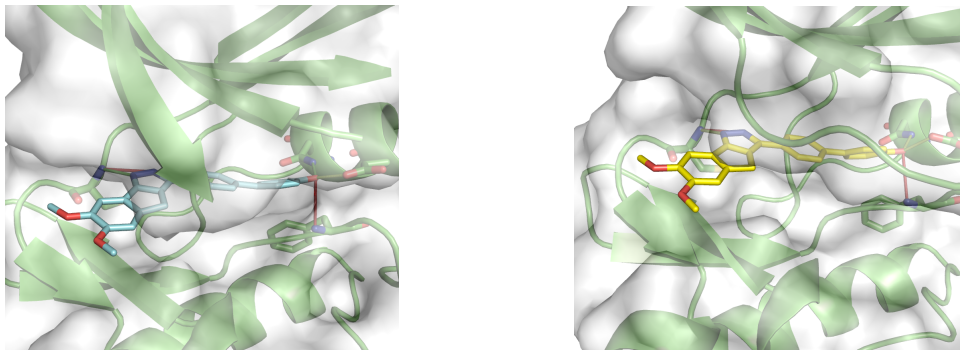


Abbildung 6.10.: Vergleich der kristallisierten (cyan) und gedockten (gelb) Pose von chk_6. Die von SFCscore bewerteten Wasserstoffbrücken (HBScore-Deskriptor) sind als Linien eingezeichnet, die Farbe zeigt die Qualität der Interaktion an (grün: gut, rot: schlecht).

[98, 99]. Die theoretische Ableitung von Formel 6.1 findet sich bei Bonett und Wright [99]. Wichtig ist, dass hier nicht die Signifikanz eines einzelnen Korrelationswertes, sondern eines Korrelationsunterschiedes bestimmt wird. Aus der Formel geht hervor, dass man umso größere Datensätze braucht, je kleiner der Korrelationsunterschied (ΔR) und die absoluten Korrelationen (R) sind.

CSAR 2012-Datensätze Tabelle 6.4 a) zeigt die resultierenden theoretischen Datensatzgrößen für den CSAR Chk-Datensatz. Als Konfidenzniveau wurden 90% verwendet. Wie man sieht, ist bei der tatsächlichen Datensatzgröße von 38 Komplexen keiner der Korrelationsunterschiede zwischen den SFCscore-Funktionen signifikant. Selbst für die Aussage, dass *sfc_rf* ($R_P = -0.02$) für diesen Datensatz schlechter geeignet ist als *sfc_229m* ($R_P = 0.38$) würde man beinahe die doppelte Datenmenge (70 Komplexe) benötigen. Für den Erk- und den LpxC-Datensatz lassen sich ebenfalls keine signifikanten Korrelationsunterschiede berechnen (Daten nicht gezeigt).

Löst man Gleichung (6.1) nach ΔR auf und setzt die tatsächliche Datensatzgröße (Chk-Datensatz 38 Komplexe) als N ein, so kommt man bei einer theoretischen Korrelation von 0.70 auf ein ΔR von 0.28. Das bedeutet, man könnte bei dieser Datensatzgröße selbst bei einer sehr guten Korrelation (0.70) erst bei einer Vergleichskorrelation von 0.42 von einem signifikanten Unterschied sprechen.

Mit sinkender Korrelation nimmt die Unsicherheit weiter zu. Bei der tatsächlich besten SFCscore-Korrelation von 0.38 hat man bereits ein Unsicherheitsintervall von 0.47.

Scoring-Validierung Interessant ist auch ein Blick auf die Scoring-Validierungsdaten (vgl. Kapitel 6.3). Tabelle 6.4 b) zeigt beispielhaft für Erk, dass auch diese Datensätze

viel zu klein waren, um als Basis für die Auswahl der besten Scoring-Funktion dienen zu können. Selbst für die größten hier gemessenen Unterschiede (z.B. zwischen *sfc_rf* und *sfc_290p*) bräuchte man 64 Komplexe, um von einem signifikanten Unterschied sprechen zu können. Für die anderen beiden Validierungsdatensätze (*Chk* und *LpxC*), bei denen die Korrelationen noch enger zusammen liegen, lässt sich ebenfalls in keinem Fall ein signifikanter Unterschied belegen.

Die Performance der SFCscore-Funktionen muss auf Basis der zur Verfügung stehenden Daten sowohl für die Validierungsdatensätze als auch für die *CSAR 2012*-Datensätze als gleichwertig angesehen werden. Vor diesem Hintergrund ist es nicht überraschend, dass sich die Ergebnisse der Scoring-Validierung, was die relative Performance der Funktionen angeht, nicht auf die *CSAR*-Daten übertragen ließen.

Cheng- und CSAR 2010-Benchmark Dieselbe Berechnung kann man auch für die beiden großen Benchmarkdatensätze *Cheng* und *CSAR 2010* durchführen (Tabelle 6.4 c) und d)). Erfreulicherweise ist hier der Korrelationsunterschied zwischen $SFCscore^{RF}$ und den ursprünglichen SFCscore-Funktionen für ein Konfidenzniveau von 90% signifikant (grüne Felder in den Tabellen). In zwei Fällen (orange eingefärbte Felder) würden für einen signifikanten Unterschied geringfügig größere Datensätze benötigt. Die Korrelationsunterschiede zwischen den ursprünglichen SFCscore-Funktionen sind dagegen nicht signifikant, so dass diese in den beiden großen Benchmark-Datensätzen ebenfalls als gleichwertig anzusehen sind. Da die Werte aber sehr eng beieinander liegen, war dies ohnehin anzunehmen.

6. CSAR 2012 Exercise

a) CSAR Chk-Datensatz (38 Komplexe)

	rf (-0.02)	229m (0.38)	229p (0.30)	290m (0.38)	290p (0.23)	855 (0.19)	frag (0.24)	met (0.18)	ser (0.22)
rf (-0.02)	N.A.	70	108	70	175	247	162	272	190
229m (0.38)	70	N.A.	1395	N.A.	432	280	490	255	384
229p (0.30)	108	1395	N.A.	1395	1972	829	2657	702	1525
290m (0.38)	70	N.A.	1395	N.A.	432	280	490	255	384
290p (0.23)	175	432	1972	432	N.A.	6250	96506	4032	97425
855 (0.19)	247	280	829	280	6250	N.A.	4001	>100000	11109
frag (0.24)	162	490	2657	490	96506	4001	N.A.	2801	24358
met (0.18)	272	255	702	255	4032	100728	2801	N.A.	6298
ser (0.22)	190	384	1525	384	97425	11109	24358	6298	N.A.

b) Erk Validierungs-Datensatz (41 Komplexe)

	rf (0.73)	229m (0.50)	229p (0.45)	290m (0.48)	290p (0.39)	855 (0.61)	frag (0.74)	met (0.48)	ser (0.47)
rf (0.73)	N.A.	121	93	106	70	327	9402	108	103
229m (0.50)	121	N.A.	2740	11176	593	485	106	15933	7749
229p (0.45)	93	2740	N.A.	10125	1850	264	83	7606	15519
290m (0.48)	106	11176	10125	N.A.	946	351	94	N.A.	260585
290p (0.38)	70	593	1850	946	N.A.	154	64	868	1058
855 (0.61)	327	485	264	351	154	N.A.	252	369	331
frag (0.74)	9402	106	83	94	64	252	N.A.	96	92
met (0.48)	108	15933	7606	N.A.	868	369	96	N.A.	80430
ser (0.47)	103	7749	15519	260585	1058	331	92	80430	N.A.

c) Cheng-Datensatz (195 Komplexe)

	rf (0.78)	229m (0.59)	229p (0.59)	290m (0.59)	290p (0.62)	855 (0.60)	frag (0.58)	met (0.64)	ser (0.59)
rf (0.78)	NA	128	130	128	164	143	123	205	129
229m (0.59)	128	N.A.	N.A.	N.A.	4247	20544	131144	1425	N.A.
229p (0.59)	130	N.A.	N.A.	N.A.	4778	27154	73770	1520	N.A.
290m (0.59)	128	N.A.	N.A.	N.A.	4247	20544	131144	1425	N.A.
290p (0.62)	164	4247	4778	4247	N.A.	13502	3107	7081	4500
855 (0.60)	143	20544	27154	20544	13502	N.A.	10708	2482	23499
frag (0.58)	123	131144	73770	131144	3107	10708	N.A.	1192	96351
met (0.64)	205	1425	1520	1425	7081	2482	1192	N.A.	1471
ser (0.59)	129	N.A.	N.A.	N.A.	4500	23499	96351	1471	N.A.

d) CSAR 2010 (332 Komplexe)

	rf (0.73)	229m (0.58)	229p (0.58)	290m (0.58)	290p (0.58)	855 (0.59)	frag (0.62)	met (0.60)	ser (0.62)
rf (0.73)	N.A.	209	204	208	209	244	322	252	340
229m (0.58)	209	N.A.	535607	N.A.	N.A.	21206	3307	14728	2708
229p (0.58)	204	535607	N.A.	N.A.	535607	14881	2871	10934	2383
290m (0.58)	208	N.A.	N.A.	N.A.	N.A.	18703	3150	13264	2592
290p (0.58)	209	N.A.	535607	N.A.	N.A.	21206	3307	14728	2708
855 (0.59)	244	21206	14881	18703	21206	N.A.	8552	502491	6207
frag (0.62)	322	3307	2871	3150	3307	8552	N.A.	11185	258927
met (0.60)	252	14728	10934	13264	14728	502491	11185	N.A.	7768
ser (0.62)	340	2708	2383	2592	2708	6207	258927	7768	N.A.

Tabelle 6.4.: Theoretisch benötigte Datensatzgrößen um eine statistisch signifikante Aussage (Konfidenzniveau 90%) zu den Korrelationsunterschieden für verschiedene in dieser Arbeit verwendete Datensätze treffen zu können (berechnet nach Gleichung (6.1) [98]). Grün: Unterschied signifikant; orange: knapp nicht signifikant; rot: deutlich nicht signifikant, N.A.: Wert nicht berechenbar (gleiche Korrelation).

6.10. Zusammenfassung

Die CSAR 2012-Datensätze stellen im Vergleich zu den bisher verwendeten Benchmark-Datensätzen eine deutlich größere Herausforderung dar. Zum einen werden target-spezifische Ligandserien bewertet. Zum anderen wurden keine Kristallstrukturen zur Verfügung gestellt, so dass die Protein und Ligandaufbereitung sowie das Docking selbstständig durchgeführt werden mussten.

Für die drei bearbeiteten Datensätze konnten für alle Funktionen gute Vorhersagefehler beobachtet werden. Die Spannweite reicht dabei von 0.55 bis 1.96. Größtenteils bewegen sich die RMSD-Werte zwischen 1.2 und 1.5 pK_i-Einheiten und liegen somit unter den Werten die in den großen Benchmark-Datensätzen (*Cheng* und *CSAR 2010*) erreicht wurden.

Für die Korrelationen zwischen Vorhersage und experimentellen Affinitätswerten konnten jedoch nur bei einem Datensatz (LpxC) gute Werte (R_p zwischen 0.52 und 0.90) ermittelt werden. Für den Chk- und den Erk-Datensatz liegen die Pearson-Korrelationen hingegen größtenteils bei unter 0.5. Der Vergleich der Korrelationen ist allerdings bei allen drei Targets aufgrund der geringen Datensatzgrößen (8 bis 41 Komplexe) nur bedingt aussagekräftig. Dies zeigt sich in einer Analyse der statistischen Validität dieser Werte, die ergibt, dass die Korrelationsunterschiede zwischen den Funktionen in den meisten Fällen als nicht signifikant angesehen werden müssen. Die dabei gewonnen Erkenntnisse zu den mindestens benötigten Datensatzgrößen für die Validierung von Scoring-Funktionen sind sicherlich für die zukünftige Entwicklung und Anwendung von Scoring-Funktionen von großem Nutzen.

Die Erkennung von inaktiven Verbindungen war bei keinem der *CSAR 2012*-Datensätze erfolgreich. Dieses Gebiet bleibt also, trotz der deutlichen Fortschritte bei der Vorhersage von Bindungsaffinitäten der beispielsweise durch SFCscore^{RF} erreicht wurde, eine Schwachstelle von SFCscore und von Scoring-Funktionen im allgemeinen. Dies zeigt sich vor allen Dingen in anwendungsorientierten Testszenarien, wie dem *CSAR 2012* Benchmark.

Die detaillierte Analyse der Scoring-Ergebnisse zeigt auch, dass der Einfluss von inkorrekten Docking-Posen (vor allem beim Chk- und beim Erk-Datensatz) eine sehr große Rolle spielte. Darüber hinaus wurde festgestellt, dass SFCscore auch bei der Bewertung von korrekten Docking-Posen eine schwächere Performance liefert als bei der Bewertung von Kristallstrukturen. Das ist ein wichtiger Ansatzpunkt für weitere Entwicklungen.

7. Applicability Domain

Trotz der genannten Einschränkungen zeigt der CSAR-Benchmark eine Sache sehr deutlich: Die Scoring-Performance, die für große, heterogen zusammengesetzte Datensätze gemessen wird, lässt sich nicht ohne Weiteres auf die Vorhersage von target-spezifischen Datensätzen übertragen. Genau diese spezifische Performance ist jedoch der entscheidende Faktor für die praktische Anwendung.

Mit Hilfe der bereits vorhandenen neun SFCscore-Funktionen lässt sich dabei ein deutlich breiteres Spektrum an Komplexen abdecken als mit jeder Funktion alleine. Tabelle 7.1 zeigt die Scoring-Performance, wenn man jeden Komplex des *Cheng*-Datensatzes jeweils mit der SFCscore-Funktion bewertet, die die beste Vorhersage liefert. Man erhält dadurch eine Korrelation von über 0.9 und deutlich geringere Vorhersagefehler im Vergleich zu den einzelnen Funktionen.

Dieses Beispiel zeigt, dass praktisch jeder der *Cheng*-Komplexe bereits mit einer sehr hohen Genauigkeit vorhergesagt werden könnte, wenn man die jeweils passende Funktion verwenden würde. Dadurch wird auch der große Vorteil einer Methode deutlich, die in der Lage wäre *a priori* vorauszusagen, welche Funktion für einen bestimmten Komplex oder Datensatz zur Anwendung kommen sollte.

Die in diesem Kapitel vorgestellten Ansätze basieren auf der Hypothese, dass empirische Scoring-Funktionen nur dann verlässliche Vorhersagen liefern, wenn sich der vorhergesagte Komplex in einem chemischen Raum befindet, welcher von den Trainingsdaten der Funktion abgedeckt wird. Dieser chemische Raum wird im Folgenden auch als "Applicability Domain" (AD) bezeichnet. Eine passende deutsche Übersetzung wäre "Kalibrierbereich", was sehr gut den Grundgedanken widerspiegelt, dass Scoring-Funktionen, ähnlich wie experimentelle analytische Methoden, nur in einem bestimmten, kalibrierten Bereich präzise Vorhersagen liefern können. Im computer-gestützten Wirkstoffdesign wurde der Begriff "Applicability Domain" bisher vor allem durch die Anwendung im Bereich von ligandbasierten QSAR-Methoden geprägt [100, 101, 102].

Eine weitere Grundhypothese ist, dass die AD von SFCscore-Funktionen mit Hilfe der SFCscore-Deskriptoren beschrieben werden kann. Im Folgenden werden einige Experimente zur Definition der AD vorgestellt, die als Grundlage für die Entwicklung einer Methode zur Abschätzung der Vorhersageverlässlichkeit zu sehen sind.

7. Applicability Domain

Tabelle 7.1.: Performance der SFCscore-Funktionen für den *Cheng*-Datensatz (Tabelle 5.2). In der letzte Zeile sind die Ergebnisse eingetragen, die man erhalten würde, wenn man jeden Komplexe mit der jeweils besten der neun verfügbaren SFCscore-Vorhersagen bewertet.

Funktion	Pearson R	Spearman ρ	Kendall τ	RMSE	MedError
229m	0.59	0.64	0.46	1.93	1.17
229p	0.59	0.63	0.45	1.95	1.15
290m	0.59	0.64	0.46	1.93	1.16
290p	0.62	0.66	0.47	1.87	1.02
met	0.64	0.68	0.49	1.83	1.09
ser	0.59	0.63	0.47	1.86	1.12
frag	0.57	0.64	0.46	2.70	1.56
855	0.60	0.63	0.45	1.90	1.12
rf	0.78	0.79	0.59	1.56	1.06
Beste Vorhersage	0.93	0.94	0.81	1.03	0.38

7.1. PCA-Dichtekarten

Methode SFCscore berechnet insgesamt 66 Deskriptoren. Für die Ableitung von SFCscore^{RF} wurden 63 Deskriptoren verwendet. Die AD von SFCscore^{RF} wird nach obiger Definition also durch einen 63-dimensionalen Raum beschrieben. Ein solch hochdimensionaler Raum ist sehr abstrakt und nicht unmittelbar visualisierbar. Ziel dieses Abschnitts ist die Entwicklung einer Methode, mit der der von den Trainingsdaten abgedeckte chemische Raum und die relative Lage einzelner Objekte in diesem Raum visualisiert werden können.

Dazu wird die Hauptkomponentenanalyse (PCA) verwendet. Wie auf Seite 41 ff. beschrieben, stellt die PCA eine Möglichkeit dar, einen hoch-dimensionalen Variablenraum in beliebig viele Hauptkomponenten aufzuteilen, wobei der Informationsverlust minimal gehalten wird. Um das "Rauschen" in den Daten schon von vornherein zu reduzieren, wurden im Fall von SFCscore^{RF} nur die 19 bedeutendsten Deskriptoren, die alle eine relative Wichtigkeit von mehr als 10% für das Modell haben (siehe Kapitel 5.2.1), berücksichtigt. Abbildung 7.1 (a) zeigt das Streudiagramm für die ersten beiden Hauptkomponenten, die zusammengenommen noch über 60% der Information des ursprünglichen Datenraums enthalten. Jeder Punkt des Diagramms stellt einen Komplex des Trainingsdatensatzes dar.

Basierend auf dieser Verteilung lässt sich eine zwei-dimensionale Dichtekarte berechnen. Dabei wird jeder Punkt des Diagramms durch eine bivariate Gauss'sche Verteilung

dargestellt. Die in Abbildung 7.1 (b) eingezeichnete Dichteverteilung kommt durch die Summierung der normalisierten Werte dieser Funktion zustande. Die roten Bereiche stellen sehr stark populierte Regionen dar. Die gelben und grünen Regionen sind dagegen weniger gut abgedeckt.

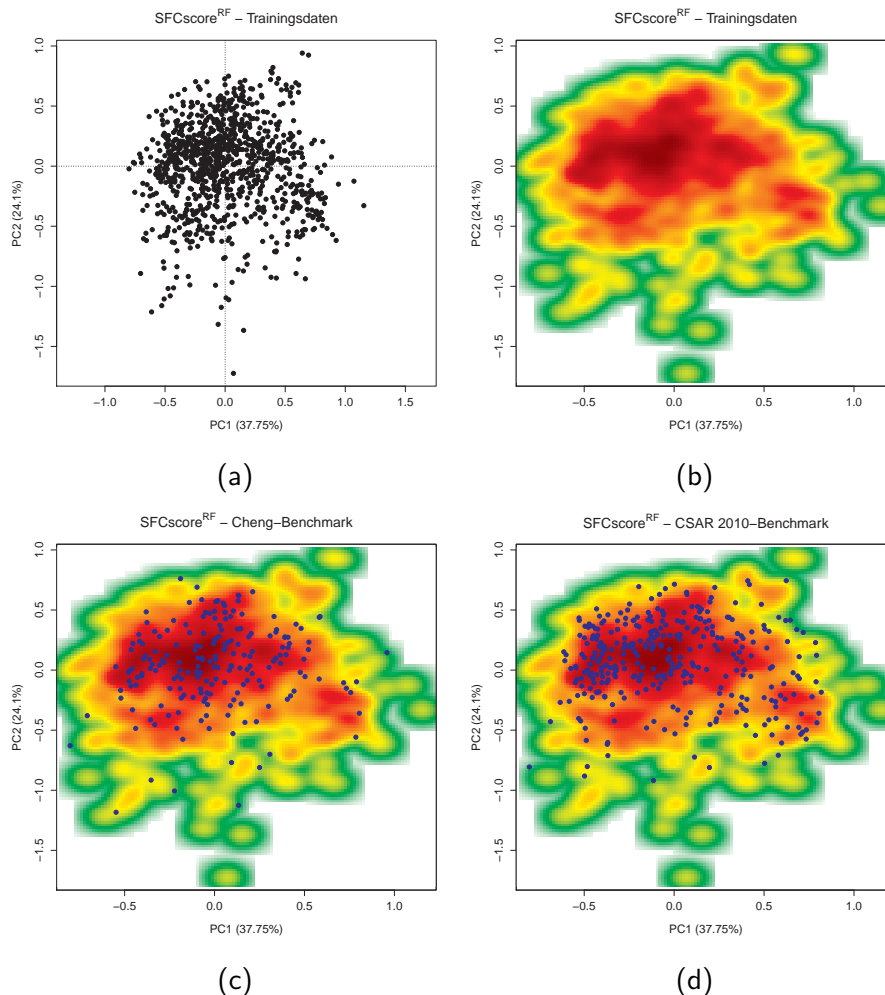


Abbildung 7.1.: (a) Streudiagramm der ersten beiden Hauptkomponenten der PCA-Analyse (b) Dichteverteilung der Trainingsobjekte in diesem Datenraum (c) Lage der *Cheng* Testkomplexe (d) Lage der *CSAR 2010* Testkomplexe

Cheng-Benchmark In diese Dichtekarte lassen sich nun die Koordinaten von Testobjekten einzeichnen. In Abbildung 7.1 (c) wurde das für den *Cheng*-Datensatz gemacht. Man kann gut erkennen, dass die große Mehrzahl der Komplexe im roten (also gut abgedeckten) Bereich liegt. Das passt zu den sehr guten Vorhersagen von SFCscore^{RF}

7. Applicability Domain

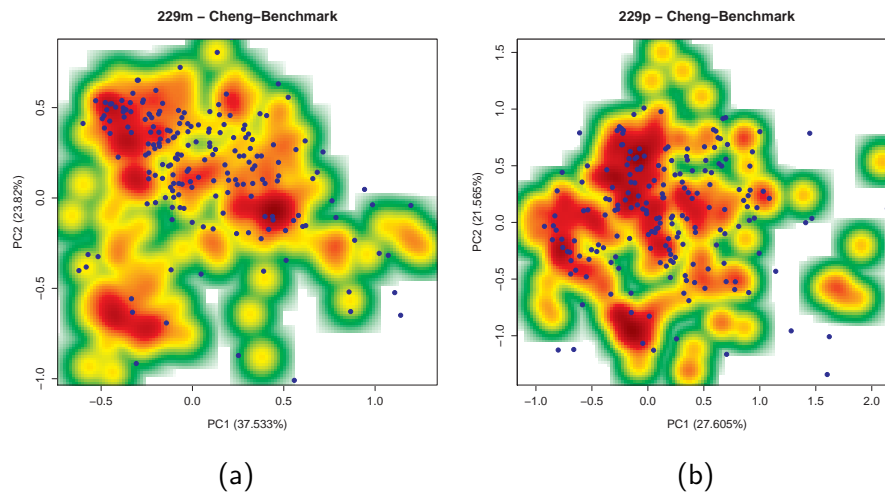


Abbildung 7.2.: Dichtekarte (AD) von (a) sfc_229m und (b) sfc_229p. In blau sind die Komplexe des *Cheng*-Datensatzes eingezeichnet.

für diesen Datensatz. In Teilabbildung (d) sind die Testobjekte des *CSAR 2010*-Datensatzes eingezeichnet. Tendenziell liegen hier etwas mehr Komplexe in den äußeren Bereichen der AD. Insgesamt ist aber auch hier die Übereinstimmung groß, was ebenfalls zur Scoring-Performance passt.

Für die MLR- und PLS-basierten SFCscore-Funktionen lassen sich ebenfalls Dichtekarten generieren. Da bei diesen Funktionen schon während der Ableitung eine Variablenselektion durchgeführt wurde, kann der Schritt hier entfallen und der gesamte Deskriptorensatz der jeweiligen Funktion kann direkt für die PCA verwendet werden. Das weitere Vorgehen ist analog zu jenem für SFCscore^{RF}.

In Abbildung 7.2 sind die entsprechenden Dichtekarten für die beiden Funktionen sfc_229m und sfc_229p eingezeichnet. Beide Funktionen basieren auf demselben Trainingsdatensatz, wurden aber mit unterschiedlichen Deskriptorensätzen und Methoden abgeleitet. Aus diesem Grund unterscheiden sich auch die Dichtekarten. Man sieht, dass beide Funktionen "Dichte-Lücken" aufweisen, die dazu führen, dass einige der Trainingskomplexe des *Cheng*-Datensatzes (blaue Punkte) in schlecht abgedeckten Bereichen und somit am Rand bzw. außerhalb des Kalibrierbereichs liegen.

Dadurch wird sehr anschaulich dargestellt, dass SFCscore^{RF} aufgrund des größeren Trainings- und Deskriptorendatensatzes den ursprünglichen Funktionen überlegen ist.

CSAR 2012 Trägt man anstatt der *Cheng*-Daten die *CSAR 2012*-Komplexe in die Dichtekarten ein, so erhält man die in Abbildung 7.3 gezeigten Ergebnisse.

Die ersten drei Teilabbildungen zeigen die Dichtekarten für den *CHK*-Datensatz, für den sehr schlechte Korrelationswerte erzielt wurden (siehe Tabelle 6.2). Basierend auf den PCA-Dichtekarten müsste man eigentlich davon ausgehen, dass SFCscore^{RF} auch

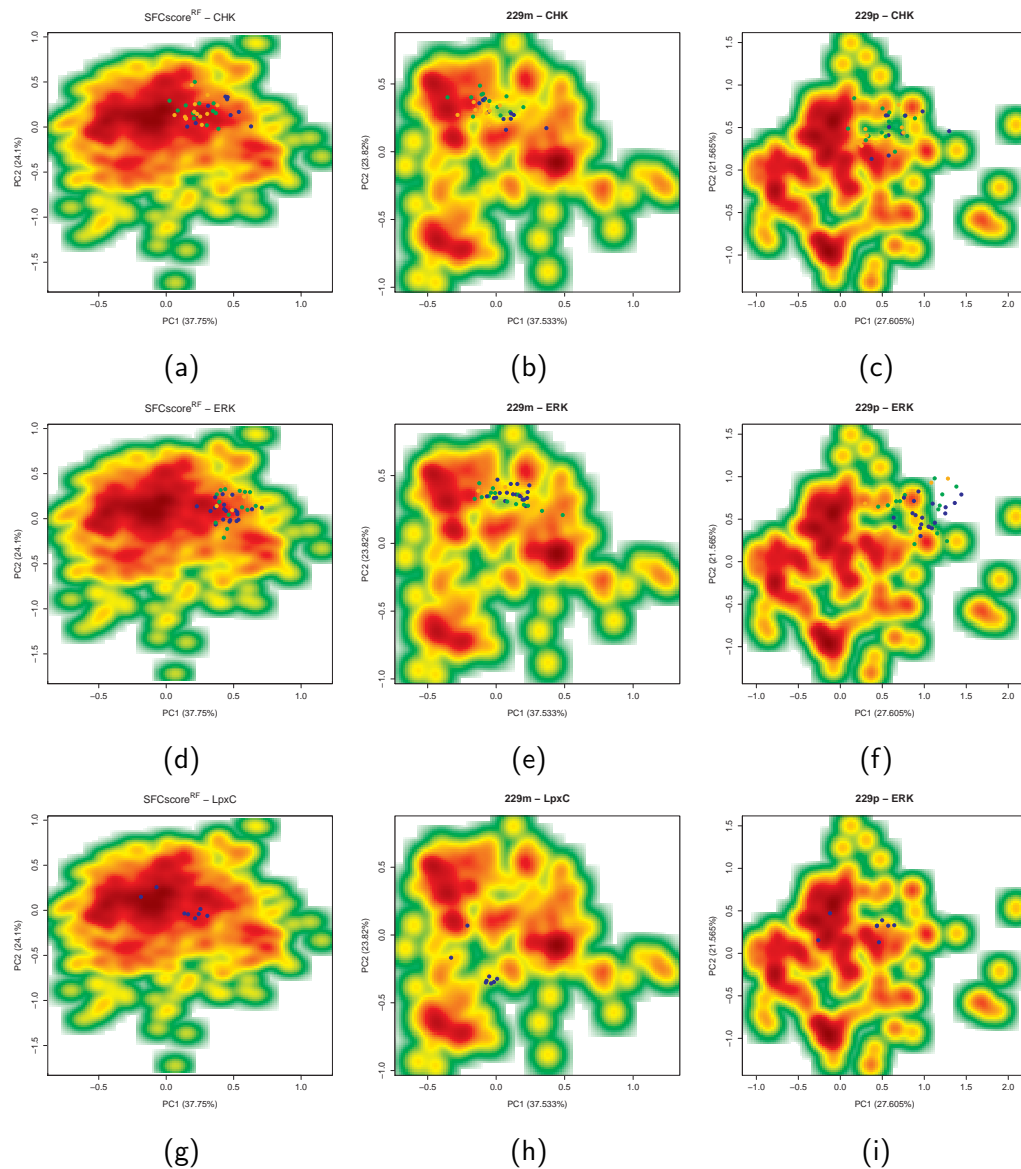


Abbildung 7.3.: PCA-Maps der Funktionen *sfc_rf*, *sfc_229m* und *sfc_229p* für die CSAR 2012 Datensätze (a)-(c): CHK-Datensatz, (d)-(f) ERK-Datensatz, (g)-(i) LpxC-Datensatz. Die Testobjekte sind außerdem nach Ligandserien koloriert. Blau: Serie 1, grün: Serie 2, orange: Serie 3. Teilabbildungen g) - i): Mit Erlaubnis adaptiert aus [3]. © (2013) American Chemical Society

in diesem Fall den anderen beiden Funktionen überlegen ist, da sich alle Chk-Komplexe innerhalb des gut abgedeckten Bereichs der Karte befinden. Im Fall von `sfc_229m` liegen ebenfalls die meisten Komplexe in gut populierten Regionen, während sich bei `sfc_229p` viele Punkte in den äußeren Regionen befinden. Die Chk-Komplexe sind außerdem gemäß dem Schema aus Abb. 6.6 (Seite 93) nach Ligandserien koloriert. Es lässt sich jedoch nicht erkennen, dass es diesbezüglich zu Unterschieden in der Verteilung der Komplexe kommt. Die unterschiedliche Performance der verschiedenen Ligandserien kann somit nicht erklärt werden.

Gleiches gilt für den Erk-Datensatz, für den man auf Basis der Dichtekarten eine Überlegenheit von `sfc_rf` und `sfc_229m` gegenüber `sfc_229p` vorhersagen würde. Tatsächlich ist es so, dass `sfc_229p` mit 1.96 einen deutlich höheren Vorhersagefehler (RMSE) liefert als `sfc_rf` oder `sfc_229m` (1.13 bzw. 1.11).

Für den LpxC-Datensatz lässt sich eine gute Beziehung zwischen den Dichtekarten und der tatsächlichen Performance ableiten. Hier findet man, vor allem im Fall von `sfc_229m`, eine ungünstige Position der Testobjekte, während diese bei `sfc_rf` und `sfc_229p` größtenteils im gut abgedeckten Bereich liegen. Tatsächlich ist `sfc_229m`, was den Vorhersagefehler angeht, die schwächste der drei Funktionen.

Zusammenfassung Die Dichtekarten sind eine interessante visuelle Hilfe für die Analyse der AD und die Auswahl von Scoring-Funktionen. Zur weiteren Evaluierung der Methode wäre die Testung mit anderen Datensätzen, die sich noch deutlicher von den Trainingsdatensätzen unterscheiden, interessant.

Genau wie für das Scoring der Komplexe ist es auch für die Berechnung der Dichtekarten ein korrekter Bindemodus als Grundlage der Deskriptorenberechnung unerlässlich. Der Bindemodus bestimmt wesentlich die Lage der einzelnen Punkte innerhalb der Dichtekarte. Aus diesem Grund ist die Position der Erk- und Chk-Komplexe nur unter Vorbehalt auswertbar bzw. für alle nicht korrekt gedockten Komplexe müssen die Positionen als falsch angesehen werden.

Der Nachteil der Methodik ist, dass die Karten nur visuell interpretiert werden können und man keine numerischen Werte erhält, die einen objektiveren Vergleich ermöglichen würden. Im Folgenden werden deswegen Ansätze für die Entwicklung von numerischen Modellen untersucht.

7.2. Numerische Ansätze

Auch für die hier vorgestellten Analysen ist die grundlegende Hypothese, dass Testobjekte nur dann verlässlich vorhergesagt werden können, wenn es möglichst viele Trainingsobjekte mit einer ähnlichen Deskriptorenverteilung gibt. Um das Ausmaß dieser Ähnlichkeit zu charakterisieren, wurden folgenden drei Parameter entwickelt, die auf den euklidischen Distanzen im multivariaten Datenraum basieren:

- **Durchschnittliche Distanz**

Hier wird die euklidische Distanz eines Testobjekts zu allen Trainingsobjekten in einem bestimmten Datenraum berechnet. Die durchschnittliche Distanz kann entweder für den gesamten Datensatz oder auch nur für eine Teilmenge (beispielsweise die 20 nächsten Objekte) berechnet werden. Ein kleinerer durchschnittlicher Abstand sollte dabei zu einer höheren Verlässlichkeit der Vorhersage führen.

- **Anzahl naher Nachbarn**

Für diesen Parameter wird die Anzahl der Trainingsobjekte gezählt, die innerhalb einer bestimmten Distanz liegen ("nahe Nachbarn"). Eine größere Anzahl naher Nachbarn sollte dabei wiederum für bessere Vorhersageverlässlichkeit sorgen.

- **Lokale Varianz der Zielvariable**

Dieser Parameter unterscheidet sich von den anderen beiden dadurch, dass nicht nur die Deskriptoren betrachtet werden, sondern auch die Zielvariable. Hierbei wird die Standardabweichung in der Affinität der Trainingsobjekte gemessen, die sich in der Nähe des Testobjekts befinden. Eine hohe Standardabweichung der Affinität der "nahen Nachbarn" macht die Vorhersagen in dieser Region weniger verlässlich. Im Bereich der Ligand-basierten QSAR-Methoden wird ein ähnliches Prinzip unter dem Stichwort "Activity Cliffs" angewendet [103, 104].

Alle drei Werte sind stark von ihrer Parametrisierung abhängig. Für die "Durchschnittliche Distanz" und die "Lokale Varianz" muss die Anzahl der nahen Trainingsobjekte definiert werden. Die Festlegung des optimalen Distanzgrenzwerts für die Berechnung der nahen Nachbarn ist ebenfalls problematisch. Je nach verwendetem Datenraum liefert der Parameter sehr unterschiedliche Ergebnisse. Die in den folgenden Analysen verwendeten Werte wurden empirisch festgelegt, mit dem Ziel, einen möglichst guten Zusammenhang zwischen dem Parameter und den Residuen der Vorhersagen zu erzielen. Als Datenraum kommt für alle drei Parameter sowohl der ursprüngliche Deskriptorenraum als auch der mittels einer PCA generierte Datenraum in Frage. Verwendet man den ursprünglichen Deskriptorenraum, ist eine Normalisierung der Daten notwendig, um eine gleichmäßige Gewichtung der Variablen zu gewährleisten.

In Abbildung 7.4 sind die drei Parameter gegen die Residuen der SFCscore^{RF}-Vorhersagen für die Komplexe des *Cheng*-Datensatzes aufgetragen. Die Plots basieren dabei auf denselben Daten, wie die PCA-Karten aus dem vorigen Abschnitt; d.h. als Datenraum wurden die ersten beiden Hauptkomponenten der PCA-Analyse der 19 bedeutendsten Deskriptoren des RF-Modells verwendet. Für sich genommen, zeigt keiner der Parameter einen direkten Zusammenhang mit den Residuen der Vorhersagen. Idealerweise würde man auf den Plots erkennen, dass sich in den Bereichen mit höherer Verlässlichkeit nur Objekte mit geringeren Residuen befinden. In den Bereichen mit geringerer Verlässlichkeit sind hingegen sowohl große als auch kleine Residuen zu erwarten (die Vorhersage in diesen Bereichen ist ja nur unsicher und nicht per se falsch). Die Tendenz ist am ehesten in Teilabbildung (c) ("Nahe Nachbarn") zu erkennen.

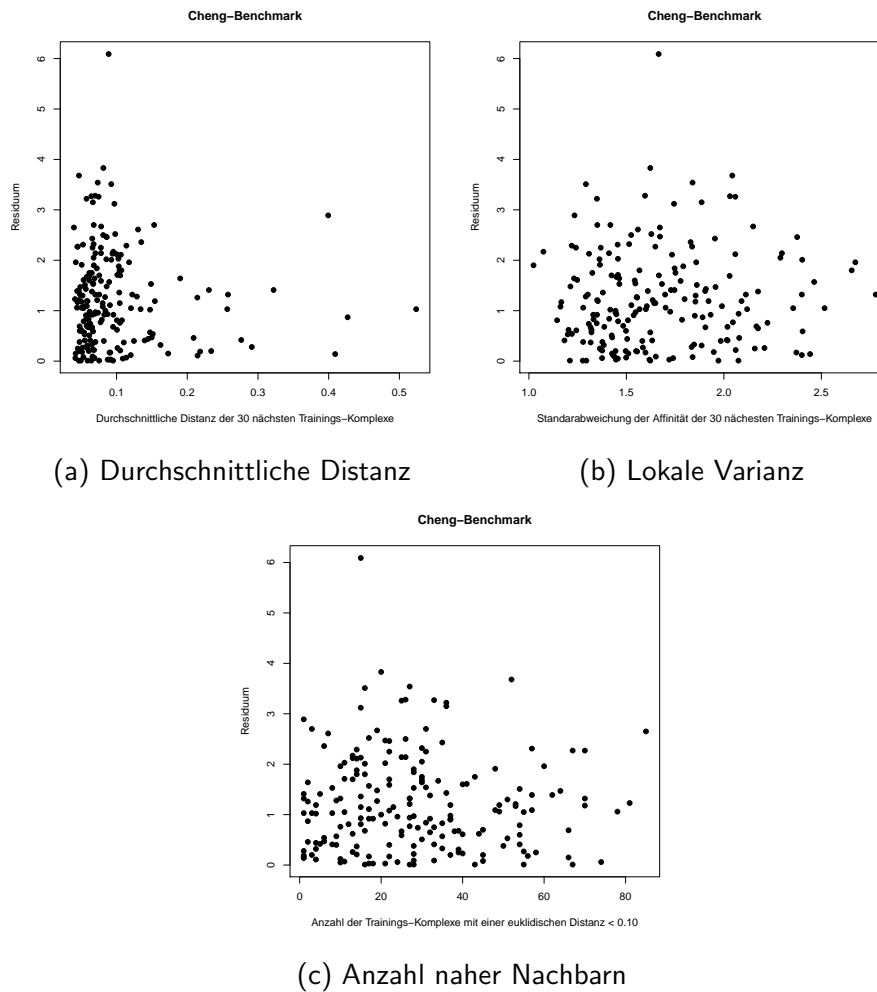


Abbildung 7.4.: Distanz-basierte Parameter zur Bestimmung der Vorhersageverlässlichkeit von $SFCscore^{RF}$. Als Datenraum wurden die ersten beiden Hauptkomponenten einer PCA der 19 wichtigsten Deskriptoren des RF-Modells verwendet. a) Durchschnittliche Distanz zu den nächsten 30 Komplexen des Trainingsdatensatzes. b) Standardabweichung der Affinität der nächsten 30 Komplexe des Trainingsdatensatzes. c) Anzahl der Komplexe des Trainingsdatensatzes, die einen euklidischen Abstand von weniger als 0.1 haben.

7.3. RandomForest-spezifische Parameter

Während die im vorigen Abschnitt vorgestellten Methoden theoretisch für alle $SFCscore$ -Funktionen anwendbar sind, sind die in diesem Abschnitt untersuchten Parameter nur für die RF-Funktion verfügbar. Es wird dabei einer der intrinsischen Parameter des *RandomForest*-Algorithmus, die sogenannte "RF-Proximity", verwendet. Diese stellt

ein Maß für die Ähnlichkeit zwischen zwei Objekten dar. Sie wird über den Anteil an Entscheidungsbäumen berechnet, bei denen zwei Komplexe demselben Endknoten zugeordnet werden. Je mehr Endknoten sich zwei Objekte teilen, desto ähnlicher sind sie sich. Der große Vorteil dieser Methode ist, dass die Ähnlichkeit direkt vom Modell ermittelt wird und somit die Deskriptoren genau so gewichtet werden wie im Vorhersagemodell.

Die Proximities des *Cheng*- und des *CSAR 2010*-Datensatzes zum SFCscore^{RF}-Trainingsdatensatz wurden auf diese Weise berechnet. In Abbildung 7.5 sind die durchschnittlichen Proximities zu den 20 nächsten Trainingskomplexen gegen die Residuen der SFCscore^{RF}-Vorhersagen aufgetragen. Man kann deutlich erkennen, dass vor allem beim *Cheng*-Datensatz ein guter Zusammenhang besteht. So kommen ab einem Proximity-Wert von 10% nur noch Vorhersagen mit Residuen unter 2 pK_i-Einheiten vor. Vorhersagen mit größeren Abweichungen haben geringere Proximities. Auch für den *CSAR 2010* ist dieser Trend erkennbar, allerdings gibt es hier auch eine größere Anzahl an Ausreißern.

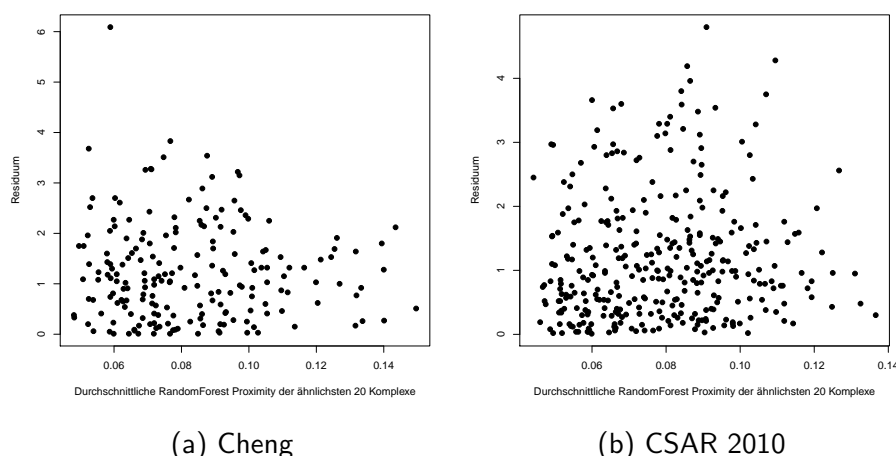


Abbildung 7.5.: Streudiagramm der RandomForest Proximities gegen die Residuen der Vorhersagen der SFCscore^{RF}-Funktion.

In beiden Fällen ist jedoch die Zahl von Objekten mit hohen Verlässlichkeitswerten relativ gering. Da die PCA-Karten gezeigt haben, dass beide Datensätze gut zur "Applicability Domain" von SFCscore^{RF} passen, sollte man hier mehr Objekte mit höherer Vorhersageverlässlichkeit erwarten.

7.4. Diskussion und Ausblick

Insgesamt ist keiner der hier vorgestellten Parameter direkt in der Lage, die Vorhersageverlässlichkeit von SFCscore ausreichend gut abzuschätzen.

Die RF-Proximity stellt jedoch in jedem Fall einen interessanten Ausgangspunkt für weitere Untersuchungen dar. Ein gutes Modell wird letztlich eine Kombination mehrerer Parameter sein. Neben den SFCscore-Deskriptoren ist auch die Anwendung weiterer Parameter denkbar, die nicht direkt für das Scoring benötigt werden, aber beispielsweise für die Beschreibung bestimmter Charakteristiken der Bindetasche hilfreich sein können. Zur Vorhersage eines konkreten Konfidenzwertes kommen sowohl klassische Regressionsmodelle als auch Klassifizierungsmodelle, die nur eine grobe Einteilung (z.B. in "sehr verlässlich", "weniger verlässlich" und "gar nicht verlässlich") vornehmen, in Frage.

Für eine bessere Evaluierung ist die Verwendung von anderen Testdatensätze anzuraten, die sich deutlicher von den SFCscore-Trainingsdaten unterscheiden. Eine interessante Alternative wäre die Untersuchung der Leave-Cluster-Out Datensätze (Kapitel 5.3.3). Hierbei handelt es sich um proteinspezifische Daten, und es wurde im Rahmen dieser Arbeit bereits diskutiert, dass zumindest in einigen Fällen eine große Diskrepanz zwischen Trainings- und Testdaten besteht.

Obwohl also noch viele Fragen offen bleiben, ist das Ziel einer solchen Methode - die Abschätzung wie gut eine Funktion zur Vorhersage eines Datensatzes geeignet ist - ein sehr interessanter Ansatzpunkt zur Verbesserung der Scoring- und Ranking-Performance von Scoring-Funktionen.

Teil II.

Entwicklung von Inhibitoren und strukturdynamische Untersuchungen der Hsp70 Chaperon-Familie

8. Grundlagen

8.1. Hsp70-Struktur

Homologe Proteine Das eigentliche Zielprotein der im folgenden vorgestellten Arbeiten ist humanes Hsp70 (hHsp70), dessen biologischer Hintergrund in der Einleitung (Seite 20f) vorgestellt wurde. Aus strukturebiologischer Sicht stehen für das humane Protein vergleichsweise wenige Daten zur Verfügung. Es existieren jedoch eine ganze Reihe homologer Proteine aus anderen Organismen, die teilweise experimentell sehr viel besser untersucht sind.

Ein Großteil der strukturellen und funktionellen Erkenntnisse, die im Folgenden vorgestellt werden, basiert auf Untersuchungen des *E.coli*-Homologs DnaK.

Die computerchemischen Untersuchungen in dieser Arbeit wurden mit einer Kristallstruktur von bovinem Hsp70 (bHsp70) (vgl. Kapitel 9.4) durchgeführt. Dieses weist eine wesentlich höhere Sequenzidentität zum humanen Hsp70 auf und stellt somit ein besseres Modellsystem dar als die bakteriellen Proteine. Eine Kristallstruktur von humanem Hsp70 existiert nicht.

Aufbau Allen Hsp70-Homologen ist gemein, dass es sich um Monomere handelt, die aus zwei großen Domänen aufgebaut sind. Abbildung 8.1 zeigt die PDB-Struktur 1YUW, der zu Beginn dieser Arbeit einzigen kristallisierten Zwei-Domänen-Struktur von bHsp70 [105].

Die N-terminale Domäne ist dabei für die ATPase-Aktivität verantwortlich und wird daher auch als *Nucleotide Binding Domain* (NBD) bezeichnet. Sie besteht aus zwei Subdomänen (I und II), welche ihrerseits in zwei Untereinheiten (a und b) aufgeteilt sind [26]. Die Bindetasche für ATP liegt vergraben zwischen den beiden Domänen I und II (Abb. 8.1).

Die zweite große Domäne ist für die Substratbindung verantwortlich und wird deswegen als SBD (*Substrate Binding Domain*) bezeichnet. Sie gliedert sich ebenfalls in zwei Subdomänen, von denen eine eine α -helikale Tertiärstruktur aufweist und die andere größtenteils in einer β -Faltblattstruktur vorliegt (Abb 8.1). Die beiden Domänen werden deswegen auch als α - bzw. β -Subdomäne bezeichnet. Die Substratbindetasche befindet sich zwischen den Schleifen der β -Subdomäne. Sie kann durch die α -Domäne geöffnet bzw. verschlossen werden, wodurch die Substrataffinität reguliert wird. Aus diesem Grund wird die α -Subdomäne auch als LID-Region bezeichnet. Die Bindetasche weist eine geringe Sequenzspezifität auf, so dass eine große Anzahl verschiedener Sub-

strate gebunden werden kann. Die helikale α -Domäne ist in der abgebildeten Struktur (1YUW) nicht vollständig. Das C-terminale Ende ist um etwa 100 Aminosäuren trunziert, weswegen dieser Bereich auch nicht komplett in der helikalen Struktur vorliegt und teilweise in der Substratbindetasche vergraben ist. NBD und SBD sind über ein kurzes Linker-Segment verknüpft.

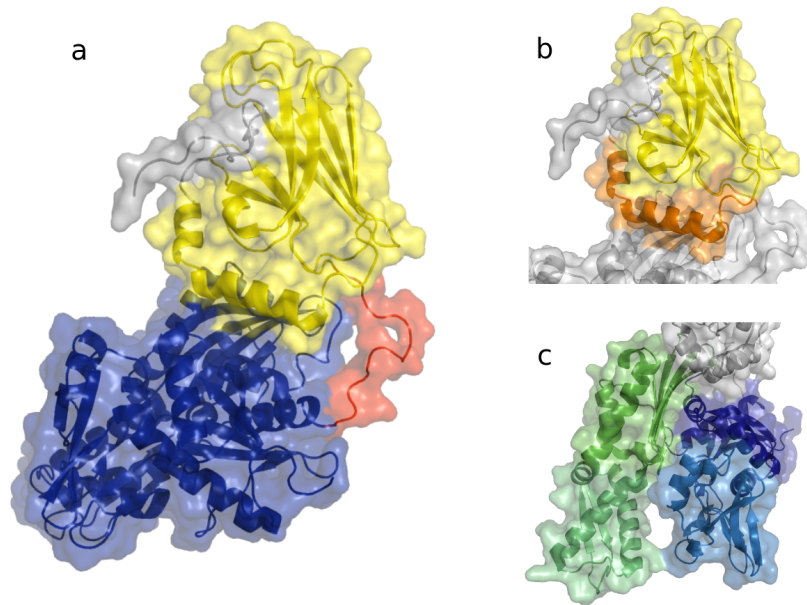


Abbildung 8.1.: a: Die Zwei-Domänen-Struktur 1YUW (bHsp70) zeigt eine ausgeprägte Kontaktfläche zwischen der NBD (blau) und der SBD (gelb). Beide Domänen sind über ein Linkersegment (rot) verbunden. b: Die SBD teilt sich in eine β -Domäne (gelb) und eine α - oder LID-Domäne (orange). Die LID-Domäne ist trunziert und liegt nicht in der nativen Sekundärstruktur vor (grau). Dieser Teil ist zudem artifiziell in die Substratbindetasche, die sich zwischen den oberen Schleifen der β -Domäne befindet, gebunden. c: Die NBD teilt sich strukturell in zwei Subdomänen, I und II, die ihrerseits wieder in Teile, a und b, unterteilt werden. Ia: dunkelblau, Ib: hellblau, IIa: dunkelgrün, IIb: hellgrün. Die NBD Bindetasche liegt in einer tiefen Furche zwischen den Subdomänen I und II.

8.2. Funktioneller Zyklus

Funktionell gesehen laufen in Hsp70-Proteinen zwei verschiedene Prozesse ab. Zum einen wird in der NBD ATP hydrolysiert (ATPase-Aktivität), zum anderen werden in der SBD Peptidsubstrate gebunden. Diese beiden Vorgänge laufen nicht unabhängig voneinander ab, sondern beeinflussen sich gegenseitig.

Die Mechanismen dieser Kommunikation sind nicht vollständig bekannt. Es gilt als sicher, dass sowohl die direkte Kommunikation zwischen den Hsp70-Domänen als auch der Einfluss von bestimmten Co-Chaperonen eine Rolle spielen. Die meisten Daten zu dieser Fragestellung gibt es zum homologen Protein DnaK aus *E.coli*. Der funktionelle Zyklus von DnaK wird meist folgendermaßen beschrieben [43, 20]:

Ausgehend von einem DnaK-ATP-Komplex wird, vermittelt durch Co-Chaperone aus der Gruppe der J-Proteine (DnaJ), ein Substrat an der SBD gebunden. J-Proteine sind eine Gruppe von Proteinen, die eine konservierte, etwa 70 AS große J-Domäne als konserviertes Merkmal aufweisen, welche für die Interaktion mit Hsp70 verantwortlich ist. Es wurden bislang mindestens 40 J-Protein-Gene im menschlichen Genom identifiziert [26].

Im so entstehenden DnaK-DnaJ-Substrat-Komplex wird die Aktivierungsenergie für die ATP-Hydrolyse herabgesetzt und ATP wird in ADP und P_i gespalten. In diesem Zustand ist das Substrat sehr stark an die SBD gebunden.

Die Dissoziation von ADP wird durch den Einfluss von *Nucleotid-Exchange-Faktoren* (NEF) wie GrpE oder Bag-1 ermöglicht. Der so entstehende NBD(apo)-Komplex hat eine sehr hohe ATP-Affinität und erlaubt gleichzeitig die Dissoziation des Peptidsubstrats, so dass hier schnell der Ausgangszustand des Zyklus (NBD(atp)-SBD(apo)) wieder hergestellt wird.

Diese Grundfunktionalität wird durch die Assoziation von weiteren Co-Chaperonen und Klientenproteinen reguliert. Als wichtige Gruppen von Co-Chaperonen sind an diesen Stellen noch die *TPR (Tetratricopeptid) containing proteins* zu nennen. Diese binden mit ihrer TPR-Substruktur an eine EEVD-Sequenz am C-Terminus von Hsp70 und steuern die Kopplung mit weiteren aktiven Domänen, wodurch die Bildung eines Chaperon-Komplexes mit spezifischer Funktionalität ermöglicht wird [26].

8.3. Allosterische Zustände

Aus obigen Erläuterungen folgt, dass Hsp70 verschiedene allosterische Zustände einnehmen kann. Selbst ohne den Einfluss von Co-Chaperonen, ergeben sich bei drei möglichen Zuständen der NBD (ATP, ADP und Apo) und zwei Zuständen der SBD (Substrat, Apo) theoretisch bis zu sechs allosterischen Zuständen. Die Struktur der isolierten Domänen ist - wiederum hauptsächlich für DnaK - gut untersucht.

Bei der NBD geht man davon aus, dass im ATP-Zustand die Nucleotidbindetasche in einer geschlossenen Konformation vorliegt, während sie im ADP- und im Apo-Zustand

weiter geöffnet ist. Der Apo-Zustand wird oft als Übergangszustand zwischen ATP und ADP-Form angesehen. Die physiologische Relevanz des Apo-Zustandes ist nicht geklärt und wird kritisch diskutiert [43].

Die strukturellen Änderungen der SBD finden dagegen hauptsächlich im LID-Bereich statt, der die Substratbindetasche öffnet bzw. schließt und somit die Substratspezifität reguliert. Die β -Faltblatt-Struktur, welche die eigentliche Bindetasche bildet, bleibt stabil. Man geht davon aus, dass die Substratbindetasche im ATP-Zustand geöffnet ist, weswegen auch oft von einer *offenen* Konformation gesprochen wird, während sie im ADP-Zustand durch die LID-Domäne geschlossen wird.

Die strukturellen Änderungen in NBD und SBD gehen mit einer Änderungen der relativen Position von NBD und SBD einher. Die genauen Mechanismen und die resultierenden Konformationen des Gesamtproteins sind unklar, vor allem weil bisher nur sehr wenige Zwei-Domänen-Strukturen kristallisiert werden konnten [43].

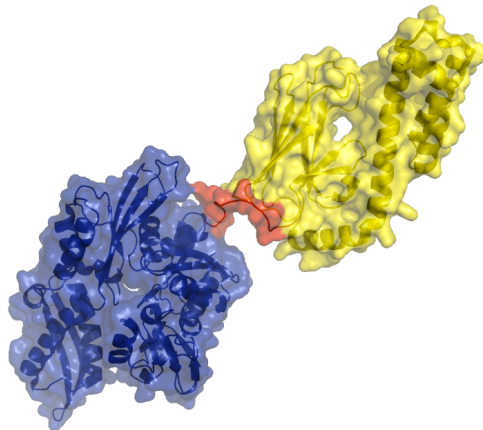


Abbildung 8.2.: Zwei-Domänen Struktur von E.coli DnaK (PDB: 2KHO [106]) im allosteren Zustand NBD(ADP)-SBD(sub). In dieser Struktur ist kein direkter Kontakt zwischen NBD und SBD zu erkennen.

Von Bertelsen und Mitarbeitern wurde eine Struktur von Wildtyp-DnaK (wt-Dnak) im allosteren Zustand NBD(ADP)-SBD(sub) (PDB 2KHO) gelöst [106]. Abbildung 8.2 zeigt, dass in dieser Struktur kein direkter Kontakt zwischen NBD und SBD besteht. Die Autoren folgern daraus, dass die allosterische Kommunikation ausschließlich im ATP- und nicht im ADP-Zustand stattfindet. Aufgrund der schnellen ATP-Hydrolyse lässt sich der ATP-Zustand nicht kristallisieren. Laut den Autoren dieser Studie stellt deswegen eine Kristallstruktur von Sse1 (PDB 2QXL), ein Hsp70-ähnliches Protein aus *S. cerevisiae* [107], welches ATP zwar bindet, aber nicht hydrolysiert, ein mögliches Modell des NBD-SBD-Interfaces von Hsp70 dar [106, 43]. Allerdings weist dieses Protein sehr große Sequenzunterschiede zu Hsp70 auf. Zum Beispiel weist der unter den Hsp70-Homologen (ansonsten) sehr stark konservierte Linker eine andere Sequenz auf.

Insgesamt unterscheidet sich die Tertiärstruktur deutlich von DnaK und von bHsp70.

Die These, dass Hsp70 im ATP-Zustand in einer Form vorliegt, in der ein direktes Interface zwischen NBD und SBD besteht, während der ADP-Zustand eine offene Form ohne direkten Kontakt der Domänen aufweist, wird auch von Mapa et al. untermauert, die das mitochondriale Hsp70 (SSc1) mittels Fluoreszenz-Spektroskopie untersucht haben [19]. In dieser Studie wurde jedoch auch festgestellt, dass der ADP-Zustand sehr heterogen ist und viele verschiedene Konformationen aufweisen kann.

Die in Abbildung 8.1 gezeigte 1YUW-Struktur befindet sich im allosteren Zustand NBD(apo)-SBD(sub) und sollte somit dem allosteren Zustand von 2KHO (Abb. 8.2) eigentlich ähnlich sein. 1YUW weist jedoch eine ausgeprägte Kontaktfläche zwischen NBD und SBD auf und widerspricht somit der These vom offenen ADP-Zustand. Es gibt deswegen Bedenken, dass es sich bei 1YUW um eine biologisch nicht relevante Konformation handeln könnte [43].

1YUW weist zum nativen Protein zwei wesentliche Unterschiede auf. Zum einen ist, wie oben beschrieben, der LID-Bereich der SBD trunkiert und die Substratbindetasche durch den nicht-nativen Rest der LID-Domäne besetzt. Zum anderen handelt es sich um eine Mutante, bei der Glu213 und Asp214 jeweils durch ein Alanin ersetzt sind. Vor allem aufgrund dieser Mutationen wird die Struktur von manchen Experten als artifiziell angesehen. Aus experimenteller Sicht kann dem entgegen gehalten werden, dass es sich um ein biologisch funktionelles Protein handelt [105].

In der vorliegenden Arbeit wurde mit Hilfe von MD-Simulationen untersucht, inwieweit die Mutation einen Einfluss auf die Stabilität des NBD-SBD-Interfaces und die Gesamtkonformation hat. Grundlage dieser Arbeit ist die von Dr. Clemens Grimm gelöste Kristallstruktur, die große Ähnlichkeiten mit der 1YUW-Struktur aufweist und auf dem gleichen Protein basiert. Die Auflösung ist mit 1.9 Å jedoch deutlich besser als bei 1YUW. Weitere Details zur Kristallstruktur finden sich in Kapitel 9.4.

8.4. Hsp70 als Target

Aus den bisherigen Betrachtungen zur Struktur und zur biochemischen Funktion von Hsp70 kann man auf verschiedene Angriffspunkte für die Entwicklung neuer Wirkstoffe schließen.

Offensichtlich ist zunächst die Adressierung der beiden Bindetaschen für ATP und Substrate. Tatsächlich wurden für beide Bindetaschen bereits Inhibitoren entwickelt [108, 109]. In beiden Fällen ist jedoch mit Spezifitätsproblemen zu rechnen.

Der komplexe funktionelle Zyklus und die vielfältigen Interaktionen von Hsp70 mit anderen Proteinen eröffnen jedoch weitere Möglichkeiten für die Inhibition (oder auch die Stimulation) der Hsp70-Funktion. Dementsprechend sind aktive Substanzen aus einem breiten chemischen Spektrum bekannt. Als Beispiele sind zu nennen: Spergualin-ähnliche Verbindungen [110], Dihydropyrimidine [111], Fettsäuren [112], Thiophen-2-carboxamide [113], 2-Phenylethynsulfonamid [114] und ein Rhodocyanin (MKT-077)

[115]. Die Verbindung MKT-077 hat es dabei bis in die klinische Phase I geschafft, wo sie allerdings aufgrund von Nierentoxizität scheiterte [116].

Bei der Mehrzahl dieser Verbindungen ist der genaue Interaktionsmechanismus mit Hsp70 nicht bekannt. In einigen Fällen gibt es zwar experimentelle Hinweise, in welchen Bereichen Interaktionen stattfinden. Ein experimenteller Beweis in Form einer Kristallstruktur existiert jedoch in keinem Fall. Im fehlenden Verständnis dieser Zusammenhänge ist auch der Grund zu sehen, warum die Weiterentwicklung dieser Substanzen stagniert. Bei Evans et al. findet sich ein guter Überblick zu diesem Themenkomplex [26].

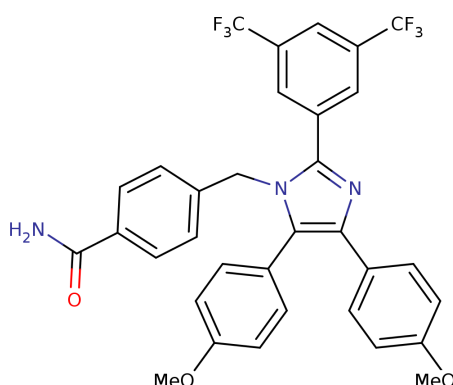


Abbildung 8.3.: Struktur des Hsp70-Inhibitors Apoptozol

Eine weitere Verbindung, die eine Aktivität gegen Hsp70 besitzt, ist das in Abbildung 8.3 gezeigte Apoptozol (Az) [117]. Es handelt sich um ein Imidazol-Derivat, welches ursprünglich in einem Ganzzell-Screening bei der Suche nach Apoptose-induzierenden Substanzen gefunden wurde. Mittels Affinitätschromatographie wurden Hsp70 bzw. Hsc70 als Zielstrukturen von Az identifiziert. Die mit Hilfe von *Surface Plasmon Resonance*-Spektroskopie (SPR) bestimmten Dissoziationskonstanten ergaben Werte von $0.21 \mu M$ (Hsc-Az) bzw. $0.14 \mu M$ (Hsp-Az) [117].

Auch bei Apoptozol war der genaue Angriffspunkt für die Interaktion mit Hsp70 zunächst nicht bekannt. Später konnte durch experimentelle Studien gezeigt werden, dass der wahrscheinliche Angriffspunkt von Apoptozol im Bereich der ATP-Bindetasche der NBD liegt [118]. Apoptozol stellt den Ausgangspunkt für die Entdeckung der Leitstruktur C066-0757 dar (vgl. Kapitel 9.1).

9. Vorarbeiten, Kooperationen und Zielsetzung

Die vorliegende Arbeit beschäftigt sich mit der Entwicklung neuer Inhibitoren für Hsp70 sowie mit der Untersuchung der Dynamik des Proteins. In beiden Teilbereichen konnte von der engen Zusammenarbeit mit Kooperationspartnern innerhalb der KFO 216 und der Universität Würzburg profitiert werden. Zu nennen sind dabei: Dr. Andreas Hartung (AK Holzgrabe), der die organischen Synthesen durchgeführt hat, Elisabeth Müller und Dr. Manik Chatterjee, die für die biologischen Testungen verantwortlich waren, sowie Dr. Clemens Grimm, der die Kristallstruktur, die in beiden Teilen der Arbeit verwendet wurde, gelöst hat.

In den folgenden Abschnitten werden diese Arbeiten kurz vorgestellt, da sie eine wichtige Grundlage für die Durchführung und Auswertung der vorliegenden Arbeit sind. Davor wird kurz die Arbeit von Ulrich Peinz vorgestellt, auf dessen ligand- und strukturbasierten Designansätzen die vorliegende Arbeit aufbaut.

9.1. Initiales virtuelles Screening

Ausgangspunkt für die Arbeiten von Ulrich Peinz waren der Hsp70-Inhibitor Apoptozol (Az) (vgl. Abschnitt 8.4) sowie die Zwei-Domänen-Struktur von bHsp70 (1YUW)[105].

Der Ansatz bestand zunächst darin, mögliche Bindetaschen für Apoptozol in der Struktur 1YUW zu identifizieren. Dazu wurden Docking-Studien durchgeführt, die zu dem Ergebnis führten, dass nicht die Nucleotid- oder die Substratbindetasche die wahrscheinlichsten Angriffspunkte von Apoptozol sind, sondern die Interfaceregion zwischen NBD und SBD. Eine Analyse mit dem Tool *SiteFinder* aus dem MOE-Softwarepaket [119] bestätigte, dass es sich um eine vielversprechende Bindetasche handelt (Abb. 9.1 a).

In einem strukturbasierten Ansatz diente die PDB-Struktur 1YUW als Ausgangspunkt für die Entwicklung eines Pharmakophor-Modells. Dazu wurde eine Hotspot-Analyse mit den Programmen GRID [120, 121, 122] und DrugScore^{CSD} [123, 124, 125] durchgeführt, um günstige Interaktionsstellen innerhalb der Bindetasche zu identifizieren. Mit Hilfe der so gefundenen Hotspots wurde ein Pharmakophor-Modell erstellt, das neben zwei aromatischen Features auch zwei Akzeptor/Donor-Funktionen und eine reine Donor-Funktion, mit entsprechenden projizierten Features, enthielt (Abbildung 9.1

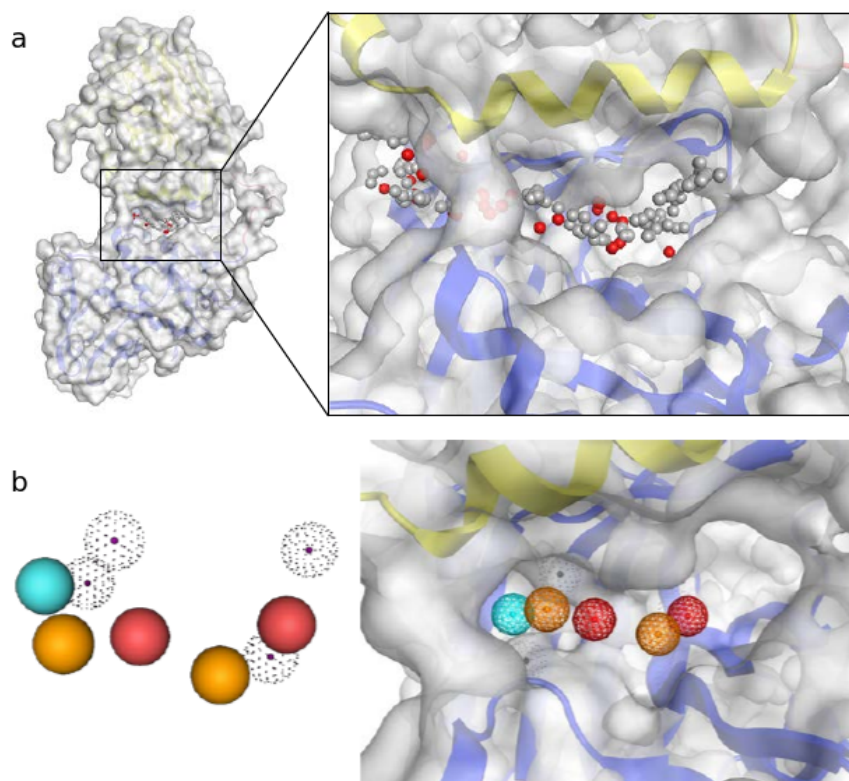


Abbildung 9.1.: a) Ergebnis der *SiteFinder*-Analyse. Die Kugeln markieren Kontaktstellen, die gut in der Proteinoberfläche vergraben sind (grau: hydrophobe Kontakte, rot: hydrophile Kontakte) und somit eine potentielle Bindetasche für Inhibitoren darstellen. b) Auf Basis der Hotspot-Analysen wurde ein Pharmakophor-Modell entwickelt. Orange: aromatisches Feature; cyan: H-Brücken-Donor; rot: H-Brücken-Donor/Akzeptor; gepunktet: projizierte Donor/Akzeptor-Funktionen auf Proteinseite.

b). Mit diesem Modell wurden 256 Verbindungen in zwei großen Moleküldatenbanken (mit insgesamt über 3 Mio. Verbindungen) gefunden.

Zusätzlich wurde ein ligandbasierter Ansatz verfolgt, bei dem mit dem Programm FTreesXL [126, 127] eine Ähnlichkeitssuche für Apoptozol durchgeführt wurde. In der gruppeninternen Moleküldatenbank (Inhouse-DB) [128], die für das virtuelle Screening aufbereitete Verbindungen aus der ZINC-Datenbank [63, 62] enthält, wurden so 189 weitere Verbindungen gefunden.

Die insgesamt 454 Verbindungen wurden dann mit Hilfe des Docking-Programms AutoDock3.0 in die Bindetasche gedockt. Alle Verbindungen, die bessere Docking-Scores zeigten als Apoptozol, wurden dann nochmals einer visuellen Inspektion unterzogen, an deren Ende neun Verbindungen für die biologischen Testungen ausgewählt wurden. Erfreulicherweise zeigten 5 dieser 9 Verbindungen tatsächlich eine biologische

Aktivität. Die Verbindung C066-0757, ein Tetrahydroisochinolinon, zeigte mit einem EC_{50} -Wert von $0.7 \mu\text{M}$ die höchste Aktivität. Kapitel 10 dieser Arbeit beschäftigt sich mit der Weiterentwicklung dieser Leitstruktur.

9.2. Biologische Testung

Die biologische Testung der initialen Verbindungen von Ulrich Peinz und der im weiteren Verlauf der Arbeiten synthetisierten Derivate wurde im Arbeitskreis von Dr. Manik Chatterjee von Elisabeth Müller durchgeführt.

Die pharmakodynamischen Effekte wurden mit Hilfe der INA-6-Zelllinie bestimmt, die ein etabliertes MM-Modell darstellt [129]. Nicht-maligne mononukleäre Zellen aus peripherem Blut (PBMC) wurden zur Toxizitätsbestimmung verwendet.

Beide Zelllinien wurden 72 Stunden mit den getesteten Verbindungen inkubiert. Die Viabilität der Zellen - also der Anteil lebender und apoptotischer Zellen - wurde mittels Annexin V-FITC- und Propidiumiodidfärbung bestimmt. Durch die Testung unterschiedlicher Inhibitorkonzentrationen wurden EC_{50} -Werte berechnet.

Für einige Substanzen wurden außerdem Untersuchungen zum Wirkmechanismus und zu den zellulären Effekten durchgeführt. Die Expressionslevel von Hsp70, Hsp90 und einiger wichtiger Klientenproteine wurden mittels SDS-PAGE und Westernblot-Analysen untersucht. In einem Apoptose-Assay wurden die anti-apoptotischen Effekte von HSP70, die durch Caspase 3 und 9 sowie den Apoptosis-Inducing Factor (AIF) vermittelt werden, untersucht.

9.3. Synthese

Andreas Hartung aus dem Arbeitskreis von Frau Prof. Holzgrabe hat die Synthesen des ursprünglichen Screening-Hits (C066-0757) sowie einer ganzen Reihe von Derivaten durchgeführt.

Im ersten Reaktionsschritt wurde durch die Kondensation von Aminen und Aldehyden eine Reihe unterschiedlich substituierte Imine aufgebaut. Diese können mit Homophthalsäureanhydrid zu Isochinolinon-Carbonsäuren umgesetzt werden. Mit Hilfe von zwei unterschiedlichen Syntheserouten konnten diastereoselektiv *cis*- bzw. *trans*-Produkte erhalten werden. Im letzten Reaktionsschritt erfolgt die Amidierung der Carbonsäuren mittels Aktivierung als gemischte Anhydride und somit die Umsetzung in die gewünschten Tetrahydroisochinolinon-Carboxamide (Abb. 9.2) [130].

9.4. Kristallstruktur

Dr. Clemens Grimm (Lehrstuhl für Biochemie, Universität Würzburg) hat Kristallisationsexperimente mit bovinem Hsp70 (1-554, E213A, D214A) durchgeführt. Das Protein

9. Vorarbeiten, Kooperationen und Zielsetzung

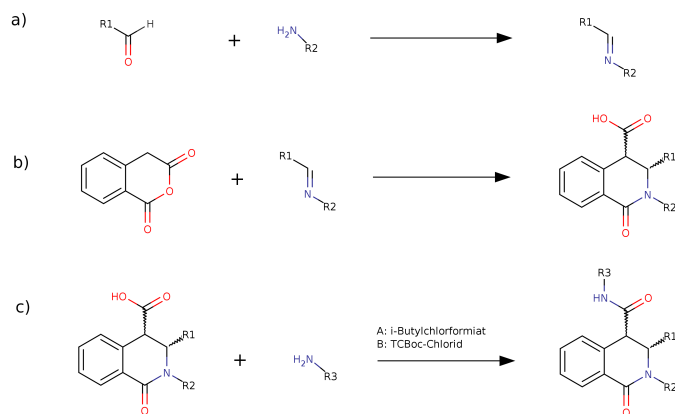


Abbildung 9.2.: Schematische Darstellung der drei essentiellen Schritte bei der Synthese der Tetrahydroisochinolinon-Carboxamide. a) Kondensation von Aminen und Aldehyden zu Iminen. b) Umsetzung der Imine mit Homophthal säureanhydrid zu Isochinolinon-Carbonsäuren. c) Amidierung der Carbonsäuren durch Aktivierung mit A: *i*-Butylchlorformiat oder B: 2,2,2-Trichlor-1,1-dimethylethylchloroformiat [130].

wurde sowohl mit Apoptozol als auch mit dem Screening Hit C066-0757 ko-kristallisiert. Leider konnte in beiden Fällen keine ausreichende Elektronendichte in der Interface-Bindetasche gemessen werden, um die Bindung der Liganden dort nachzuweisen. Mögliche Gründe dafür sind die schlechte Löslichkeit sowie die für Kristallisationszwecke unter Umständen zu schwachen Inhibitionskonstanten. Stattdessen wurde in der Bindetaschenregion ein Molekül Trimethylamin-N-oxid (TMAO) gefunden. Dieses ist in hohen Konzentrationen im Kristallisationspuffer enthalten. Möglicherweise verhinderte dies die Bindung der Liganden.

In Abbildung 9.3 ist die Kristallstruktur dargestellt, die aus dem Ko-Kristallationsexperiment mit Apoptozol resultierte und große Ähnlichkeit mit 1YUW aufweist. Die Struktur wird im Folgenden mit dem internen Titel *hsc_az* bezeichnet und ist die Grundlage für die hier durchgeführten Arbeiten.

Die deutlich bessere Auflösung von 1.9 Å - im Vergleich zu 2.6 Å bei 1YUW - macht diese Struktur zu einem ausgezeichneten Ausgangspunkt für computerchemische Untersuchungen. Allerdings konnte das Linker-Segment, das NBD und SBD verbindet, hier nicht ausreichend gut aufgelöst werden und fehlt deswegen in der Struktur *hsc_az*. Für die Docking-Analysen in der Bindetaschen-Region hat dies keine Auswirkungen, weil die Linker-Region hier nicht unmittelbar involviert ist. Für die MD-Simulationen (Kapitel 11) musste der Linker hingegen durch Homologiemodelling nachträglich in die Struktur eingefügt werden (siehe dazu Kapitel 11.2.2).

In der Zwischenzeit wurde eine durch ein weiteres Refinement bearbeitete Version von *hsc_az* in der PDB mit dem Code 4FL9 veröffentlicht.

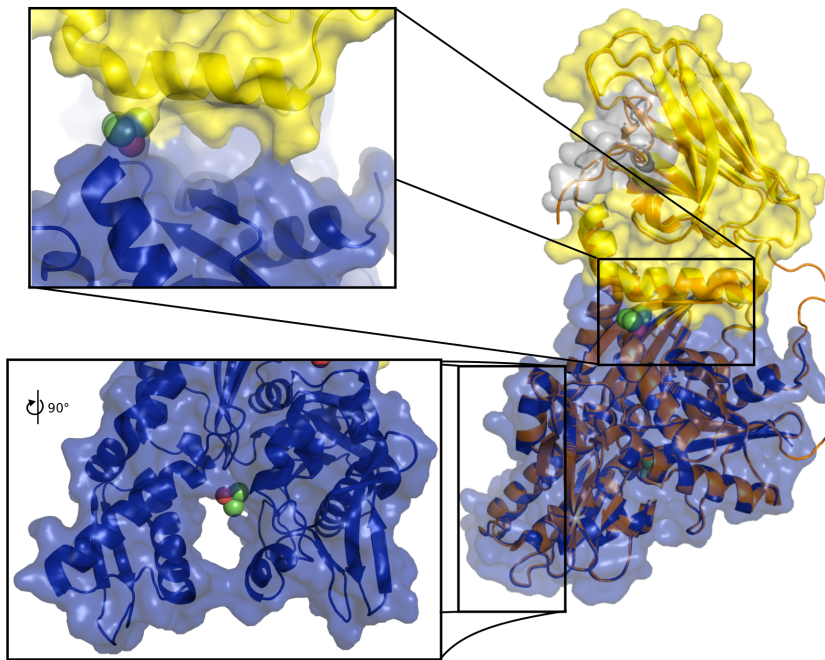


Abbildung 9.3.: Die von Dr. Clemens Grimm gelöste Kristallstruktur *hsc_az* (blau/gelb) weist große Ähnlichkeiten zur PDB-Struktur 1YUW (rechts, orange) auf. Anstatt der ko-kristallisierten Liganden ist im Interface zwischen NBD (blau) und SBD (gelb) ein TMAO-Molekül gebunden, welches Bestandteil des Kristallisationspuffers war. Ähnlich wie bei 1YUW liegt das trunkeerte Ende der LID-Domäne (grau) nicht in helikaler Sekundärstruktur vor und ist in die Substratbindetasche gebunden.

9.5. Zielsetzung dieser Arbeit

Die Zielsetzung dieser Arbeit war in erster Linie die Weiterentwicklung der Leitstruktur C066-0757, wobei vor allem Aufklärung und Verifizierung des Bindemodus in der Interface-Region von Hsp70 im Mittelpunkt stand. Es sollten dabei Docking- und Scoring-Methoden für die Entwicklung von neuen Derivaten eingesetzt werden, die nach Synthese und Testung durch die oben genannten Kooperationspartner, neue Erkenntnisse zum Bindemodus der Substanzklasse liefern. Die entsprechenden Arbeiten sind in Kapitel 10 beschrieben.

In einem zweiten Ansatz wurden molekulardynamische Simulationen für Hsp70 durchgeführt. Zielsetzung war hier zum einen allgemeine Informationen über die Stabilität der Kristallstruktur *hsc_az* zu erhalten, zum anderen sollte die Dynamik der Interface-Region und daraus folgend der Einfluss auf die Bindung von Liganden in diesem Bereich untersucht werden. Das methodische Vorgehen bei der Durchführung der Simulationen ist in Kapitel 11, die Auswertung der Ergebnisse findet sich in Kapitel 12.

10. Weiterentwicklung der Leitstruktur C066-0757

Wie bereits erwähnt, wurde die Verbindung C066-0757 nicht als Inhibitor der beiden offensichtlichen Bindetaschen in der NBD oder der SBD von Hsp70 entwickelt. In beiden Fällen wäre mit Selektivitätsproblemen im weiteren Entwicklungsverlauf zu rechnen. Stattdessen wird die Interface-Region zwischen NBD und SBD adressiert, mit dem Hintergedanken, die allosterische Kommunikation zwischen den beiden Hauptdomänen zu unterbinden und so eine sehr spezifische Hemmung zu erreichen.

Dieser Teil der Arbeit beschäftigt sich mit der Weiterentwicklung von C066-0757. Der erste Schritt ist dabei die Aufklärung des Bindemodus von C066-0757. Als Grundlage dessen werden im Folgenden die von Ulrich Peinz gefundene Interface-Bindetasche sowie die Leitstruktur C066-0757 vorgestellt.

Im Anschluss daran (Kapitel 10.1) werden die Docking-Ergebnisse für C066-0757 vorgestellt und die möglichen Bindemoden diskutiert. Basierend auf diesen Docking-Ergebnissen wurde eine Substanzbibliothek mit Derivaten von C066-0757 erstellt und umfassenden biologischen Untersuchungen unterzogen. Kapitel 10.2 befasst sich mit der Interpretation der biologischen Daten vor dem Hintergrund der Docking-Ergebnisse dieser Verbindungen.

10.0.1. Interface-Bindetasche

Abbildung 10.1 zeigt die von Ulrich Peinz identifizierte Bindetasche in der Interface-Region zwischen SBD (gelb) und NBD (blau). Auf Seiten der SBD ist vor allem die α -helikale Subdomäne und auf Seiten der NBD die Subdomäne IA an der Ausbildung der Bindetasche beteiligt. Die Vergrößerung zeigt, dass die Topologie der Bindetasche hauptsächlich durch die beiden Salzbrücken zwischen den Seitenketten von Asp152 und Lys524 (links in der Abbildung) sowie zwischen Arg171 und Glu513 (rechts) geprägt ist.

Dadurch entstehen zwei tiefe Bindetaschen, die in der Abbildung und im Rest dieser Arbeit mit *S1* und *S2* bezeichnet werden. In beiden Fällen handelt es sich jedoch nicht um geschlossene Bindetaschen, da sie jeweils von zwei Seiten lösungsmittelzugänglich sind.

Die *S1*-Bindetasche stellt tatsächlich eher einen überbrückten Tunnel dar, der direkt auf der anderen Seite der Salzbrücke wieder lösungsmittelzugänglich ist. In der Kristallstruktur *hsc_az* bindet an dieser Stelle das TMAO-Molekül, welches jedoch in

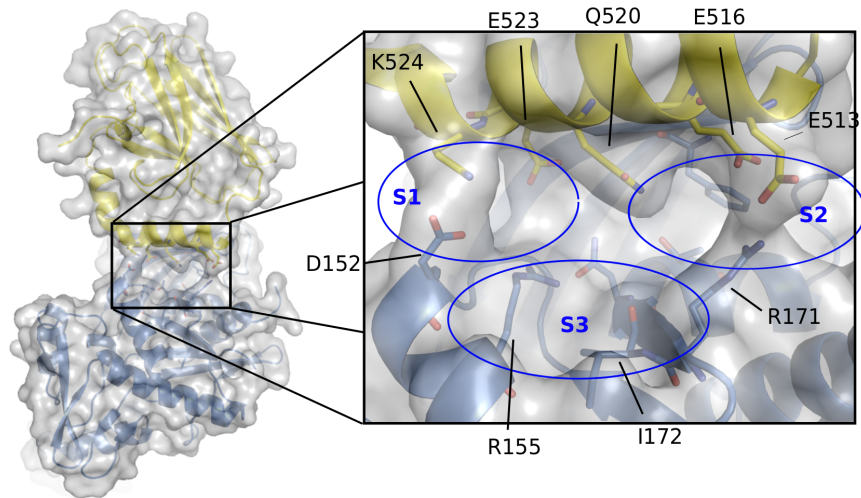


Abbildung 10.1.: Bindetasche in der Interface-Region der *hsc_az*-Struktur. Die Unteretaschen S1 und S2 werden durch Salzbrücken gebildet, die die Topologie der Bindetasche wesentlich prägen. Die S3-Bindetasche ist eher oberflächlich und wird nur von Aminosäuren der NBD gebildet.

Abbildung 10.1 entfernt wurde. Die von Ulrich Peinz durchgeführte Hotspotanalyse der 1YUW-Struktur ergab, dass in der S1-Bindetasche günstige Potentiale für aromatische Gruppen bestehen. Eine eher hydrophile para-Substitution ist aufgrund der Lösungsmittelzugänglichkeit auf der anderen Seite der Bindetasche naheliegend. Die S2-Bindetasche ist deutlich größer als die S1-Tasche. Wie schon erwähnt ist auch sie von der anderen Seite lösungsmittlexponiert, jedoch findet man hier eine deutlich ausgeprägte hydrophobe Region, welche für das Design von Inhibitoren interessant sein könnte.

Die dritte Bindetasche, die in Abbildung 10.1 eingezeichnet ist, wird ausschließlich von Aminosäuren der NBD gebildet. Die S3-Bindetasche ist deutlich flacher als die beiden anderen Taschen, durch den hohen Anteil hydrophober Aminosäuren (z.B. Ile172) ist sie jedoch ebenfalls interessant für das Design neuer Inhibitoren.

Insgesamt weisen vor allem die tiefer im Protein vergrabenen Teile der Bindetasche eher hydrophobe Charakteristika auf, wie man auch anhand der von Ulrich Peinz durchgeführten Bindetaschenanalyse (Abb. 9.1 a) erkennen kann. In den oberflächlichen Proteinbereichen finden sich, vor allem auf Seiten der SBD, viele polare Seitenketten (Glu523, Gln520, Glu516). Diese sind größtenteils flexibel, und eine Beteiligung an der Bindung von Liganden im Sinne eines *Induced-Fit*-Mechanismus ist durchaus denkbar. Die Dynamik der Bindetasche wird später anhand von MD-Simulationen untersucht.

10.0.2. Screening Hit: C066-0757

Abbildung 10.2 zeigt die Struktur des Screening-Hits C066-0757. Es handelt es sich um ein dreifach substituiertes Tetrahydroisochinolin-1-on. An den Positionen 2 und 3 befindet sich jeweils eine *para*-Methoxyphenyl-Einheit, Position 4 ist mit einer *N*-Methylpyridincarboxamid-Gruppe substituiert.

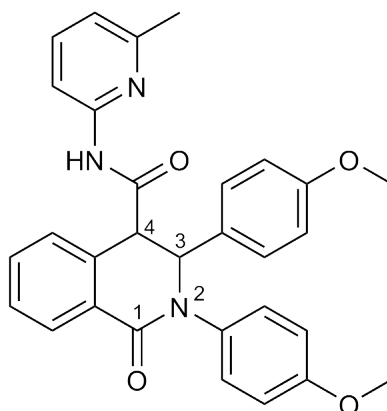


Abbildung 10.2.: Struktur des initialen Screening Hits C066-0757

Stereoisomerie Da das Molekül an den Positionen 3 und 4 zwei Chiralitätszentren aufweist, sind insgesamt vier Konfigurationsisomere möglich. Für die kommerziell bestellte Verbindung lagen keine Informationen über die stereochemische Zusammensetzung vor, so dass zunächst unklar war, welches Stereoisomer für die biologische Aktivität primär verantwortlich ist.

Durch die diastereomer-spezifische Syntheseroute von Andreas Hartung konnte gezeigt werden, dass nur die beiden *trans*-Enantiomere eine biologische Aktivität besitzen. Als aktives Enantiomer kommen somit das *R,R*- oder das *S,S*-Isomer (Abbildung 10.3) in Frage. Mit Hilfe von Docking-Studien wird im folgenden untersucht, welche Bindepasen für die beiden Formen möglich sind, um so Rückschlüsse über das wahrscheinlich aktive Enantiomer ziehen zu können.

10.1. Docking des Liganden C066-0757

Beide Enantiomere wurden getrennt in die Interface-Bindetasche von *hsc_az* gedockt. Zum Docking wurde das Programm AutoDock3.0 [10] verwendet, da hier bereits eine etablierte Methode aus dem initialen Screening von Ulrich Peinz existierte.

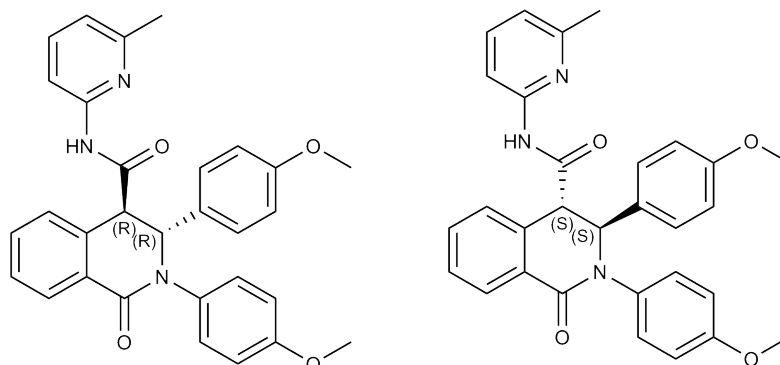


Abbildung 10.3.: Mögliche aktive Isomere von C066-0757

Methode: AutoDock Ähnlich zum bisher in dieser Arbeit verwendeten Glide-Algorithmus, basiert auch die Energiefunktion von AutoDock 3.0 auf einem Potentialgitter, welches vor der Durchführung des Dockings für jeden Atomtyp des Liganden separat berechnet wird.

Dazu wurde die Proteinstruktur *hsc_az* folgendermaßen aufbereitet: Das Programm PROTONATE (aus dem AMBER-Softwarepaket [131, 132]) wurde verwendet, um dem Protein polare Wasserstoffatome hinzuzufügen. Mit dem Skript *add_chrgsol* wurden dann Kollman Partiaalladungen [133] zugewiesen, sowie die Solvatationsparameter [10] berechnet.

Basierend auf diesen Parametern wurden die Interaktionspotentiale mit dem Programm AutoGrid berechnet. Als Mittelpunkt des Gitternetzes wurde das terminale C-Atom von Gln520 definiert, welches sich im Zentrum der Interfaceregion befindet (siehe Abbildung 10.1). In jede Richtung wurden 73 Gitterpunkte mit einem Abstand von 0.375 Å generiert.

Die Liganden wurden mit SYBYL-X (SYBYL-X Molecular Modeling Software, Version 1.0, Tripos Inc.: St. Louis, MO, USA) aufbereitet. Dabei wurden Wasserstoffatome hinzugefügt und Gasteiger-Marsili-Ladungen berechnet. Anschließend wurden die Liganden im Tripos-Kraftfeld energieminiert. Rotierbare Bindungen wurden mit dem Modul *AutoTors* identifiziert und definiert.

Autodock verwendet zum Sampling möglicher Bindeposen einen stochastischen Suchalgorithmus. Es stehen verschiedene Methoden wie das *Simulated Annealing* (SA) (ein Algorithmus aus der Gruppe der Monte-Carlo Methoden) oder *genetische Algorithmen* (GA) zur Verfügung. Methode der Wahl ist in der Regel eine Hybridmethode, bei der ein klassischer GA mit einem lokalen Optimierungsalgorithmus (nach Solis und Wets) kombiniert wird (*Lamarckian Genetic Algorithm*). Diese Methode wurde auch in der vorliegenden Arbeit mit den folgenden Parametern eingesetzt:

- maximale Zahl an Energiebewertungen: $3.0 \cdot 10^6$ bei maximal 27000 Generationen;

- Mutationsrate: 0.2, Crossoverrate: 0.8, Elitism-Wert: 1;
- Wahrscheinlichkeit für eine lokale Optimierung (Solis und Wets): 6 %;
- maximal 300 Iterationen bei der lokalen Suche.

Die Posen wurden mit einem RMSD-Grenzwert von 1 Å nach Konformationsfamilien gruppiert.

Als Scoring-Funktion wird eine empirische Methode verwendet, die anhand von 30 Protein-Ligand-Komplexen kalibriert wurde und die neben van der Waals-Kräften und elektrostatischen Wechselwirkungen auch Torsions- und Desolvatationsenergien berücksichtigt [10].

Ringkonformationen Bei der praktischen Durchführung des Dockings kommt ein weiterer Aspekt hinzu, der sich aus der Struktur von C066-0757 ergibt: Durch das hydrierte Ringsystem sind bei beiden Enantiomeren zwei verschiedene Ringkonformationen möglich. Dadurch kommt sowohl eine axiale als auch eine äquatoriale Stellung der Substituten zum Ringsystem in Frage. Abbildung 10.4 zeigt, dass sich die beiden äquatorialen Konformationen trotz ihrer Enantiomerie sehr ähnlich sind. Die beiden Konformationen mit axialen Substituenten unterscheiden sich dagegen deutlich voneinander.

Aus energetischer Sicht ist dabei jeweils die Konformation mit axialer Substituentenstellung bevorzugt, da sie eine um etwa $27 \frac{\text{kJ}}{\text{mol}}$ geringere potentielle Energie aufweist (axiale Substituenten: $212 \frac{\text{kJ}}{\text{mol}}$, äquatoriale Substituentenstellung: $239 \frac{\text{kJ}}{\text{mol}}$; berechnet mit Schrödinger LigPrep¹ nach Energieminimierung (500 Zyklen) im OPLS_2005-Kraftfeld [74]). Für Bindung an das Protein sind jedoch beide Konformationen zugänglich und müssen beim Docking berücksichtigt werden. Da beim Docking mit Autodock die Einfachbindungen innerhalb des Ringsystems als nicht rotierbar angesehen werden, kann der Übergang zwischen den Rotameren hier nicht simuliert werden. Die vier Varianten (zwei Enantiomere mit je zwei Ringkonformationen) müssen deswegen separat behandelt werden.

10.1.1. Docking-Ergebnisse

Im Folgenden werden die Ergebnisse der vier Docking-Runs einzeln vorgestellt. Darauf aufbauend werden dann Hypothesen zu möglichen Bindemoden sowie dem wahrscheinlich aktiven Enantiomer und dessen Konformation abgeleitet.

S,S-Enantiomer mit äquatorialen Substituenten ((S,S)_{aq})

Das Clustering der 50 Docking-Posen ergibt für das (S,S)_{aq}-Enantiomer fünf Gruppen von Bindeposen, die je zwischen 6 und 12 Mitglieder haben.

¹LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2012

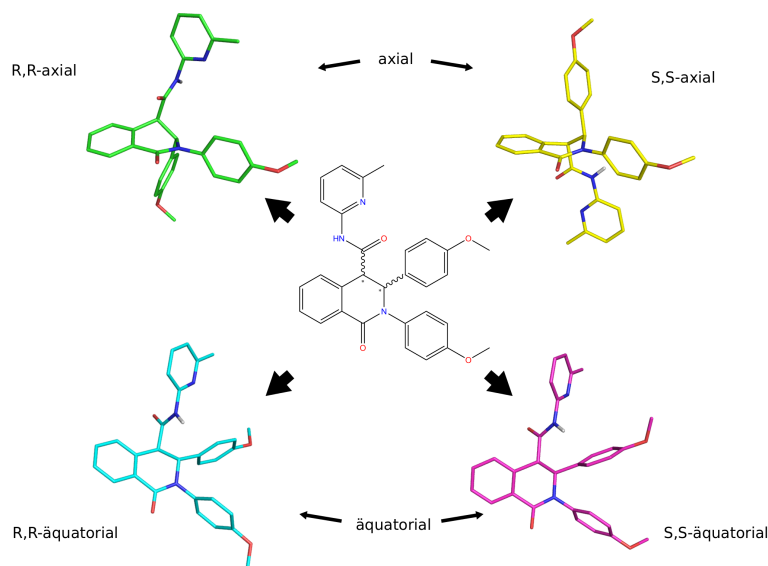


Abbildung 10.4.: Durch das hydrierte Ringsystem des Tetrahydroisochinolinon-Grundgerüsts von C066-0757 sind für die beiden trans-Enantiomere je zwei Konformationen (jeweils mit axialer und äquatorialer Substituentenstellung) möglich.

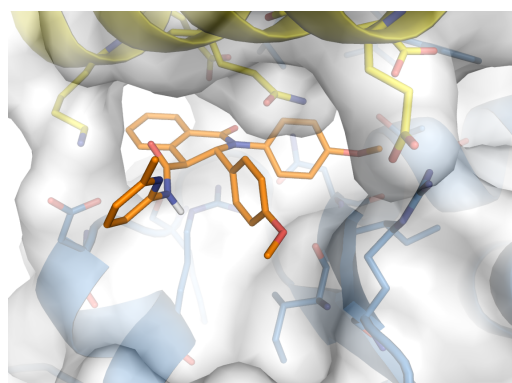
Bei einer Gruppe ist der Ligand jedoch nicht in der Interface-Bindetasche plaziert, sondern in einer Cavität, die sich auf der anderen Seite der S1-Bindetasche befindet. Da die in Kapitel 10.0.1 vorgestellte Bindetasche Ausgangspunkt und Zielstruktur dieser Arbeit ist, werden diese Posen im Folgenden nicht weiter analysiert.

Es kommen also vier Bindeposen in Frage. Dabei lässt sich weder anhand der relativen Häufigkeit noch mittels des Docking-Scores eine einzelne Gruppe favorisieren. In Abbildung 10.5 ist jeweils der Vertreter mit dem besten Docking-Score aus jeder Gruppe dargestellt.

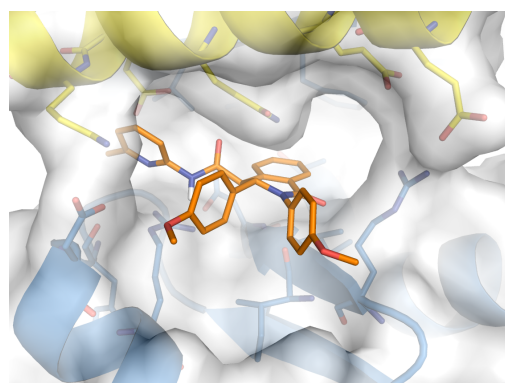
Der Ligand ist in allen vier Fällen deutlich unterschiedlich in der Bindetasche positioniert. Dies hat mit dem planaren und teilweise symmetrischen Aufbau des Moleküls zu tun, der unterschiedliche Orientierungen in der Bindetasche zulässt. Begünstigt wird dies durch das Fehlen von gerichteten Schlüsselinteraktionen (z.B. einer Wasserstoffbrücke), welche den Liganden in einer bestimmten Orientierung fixieren würden.

Die beiden ersten gezeigten Posen (Abbildung 10.5 a und b) adressieren vornehmlich die beiden Bindetaschen S1 und S2 und sind deswegen beide sehr tief in der Bindetasche vergraben. Sie erhalten in Vergleich zu den anderen beiden Posen auch die besseren Scores. Die beiden anderen Posen (c und d) sind nicht so tief vergraben und adressieren hauptsächlich die S3-Bindetasche.

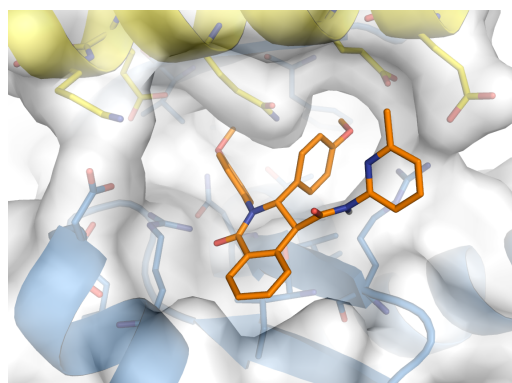
Trotz der teilweise guten Docking-Scores erscheint es aufgrund der inkonsistenten Docking-Ergebnisse und des Fehlens von gerichteten Interaktionen, eher unwahrscheinlich, dass eine dieser Posen für die biologische Aktivität verantwortlich ist.



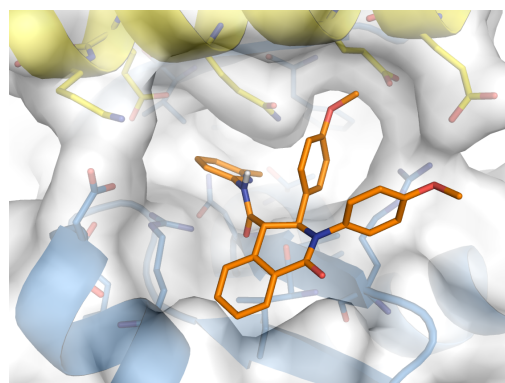
a) Pose 26
Docked Energy: $-6.59 \frac{\text{kcal}}{\text{mol}}$



b) Pose 36
Docked Energy: $-6.53 \frac{\text{kcal}}{\text{mol}}$



c) Pose 4
Docked Energy: $-5.82 \frac{\text{kcal}}{\text{mol}}$



d) Pose 15
Docked Energy: $-5.67 \frac{\text{kcal}}{\text{mol}}$

Abbildung 10.5.: Die Abbildungen zeigen jeweils die am besten bewertete AutoDock-Pose aus den vier Konformationsfamilien des $(S,S)_{\text{aq}}$ -Enantiomers.

R,R -Enantiomer mit äquatorialen Substituenten ($(R,R)_{\text{aq}}$)

Die Ergebnisse für das R,R -Enantiomer mit äquatorialen Substituenten sind sehr ähnlich zu jenen des $(S,S)_{\text{aq}}$ -Enantiomers. Da sich die beiden Enantiomere in dieser Konformation sterisch nur geringfügig unterscheiden (vgl. 10.4), ist das nicht überraschend.

Auch bei diesem Docking erhält man eine große Anzahl von meist recht kleinen Clustern. Insgesamt finden sich neun Cluster, die drei oder mehr Mitglieder haben. Jedoch treten auch hier eine Reihe von Lösungen auf, bei denen der Ligand nicht direkt in der gewünschten Bindetasche plaziert ist und die deswegen für diese Auswertung nicht berücksichtigt wurden. Im Wesentlichen lassen sich daher auch hier vier Cluster herausfiltern, die in Abbildung 10.6 gezeigt sind.

Die Bindemoden ähneln qualitativ jenen des $(S,S)_{\text{aq}}$ -Enantiomers. Die beiden Posen,

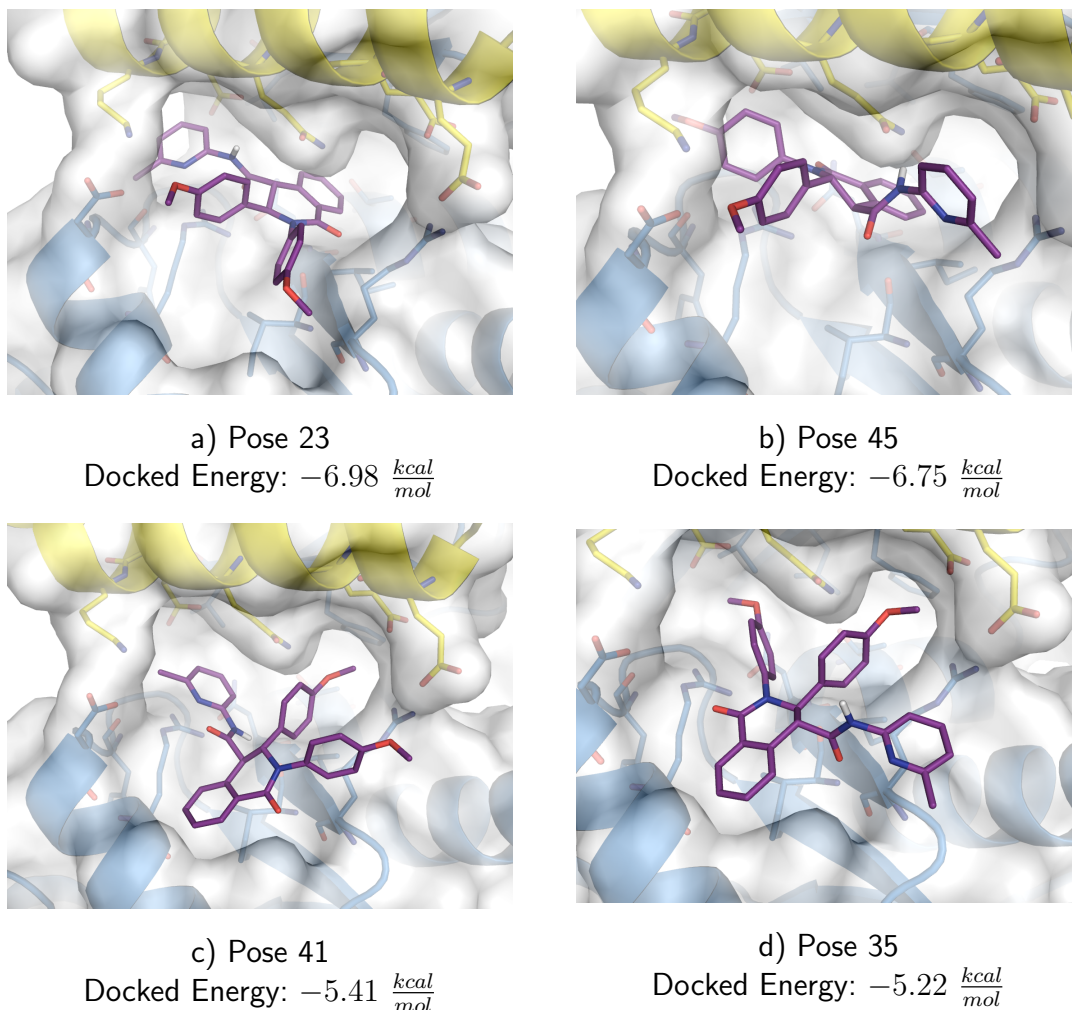


Abbildung 10.6.: Die Abbildungen zeigen jeweils die am besten bewertete AutoDock-Pose aus den vier Clustern des (R,R)_{aq} Enantiomers

die tiefer vergraben sind (Abbildung 10.6 a und b) und die S1- und S2-Bindetasche adressieren, erhalten auch hier die besseren Scores. Das Fehlen einer essentiellen gerichteten Wechselwirkung führt dazu, dass der Ligand auf unterschiedliche Weise in der Tasche plaziert wird.

Vor allem aufgrund der sehr inkonsistenten Docking-Ergebnisse, erscheint es unwahrscheinlich, dass es sich bei diesem Enantiomer um die aktive Form handelt. Die Ergebnisse legen außerdem nahe, dass, sollte sich die äquatoriale Substituentenstellung als aktiv erweisen, beide Enantiomere eine biologische Aktivität aufweisen sollten, da sie die selben Bindeposen einnehmen können.

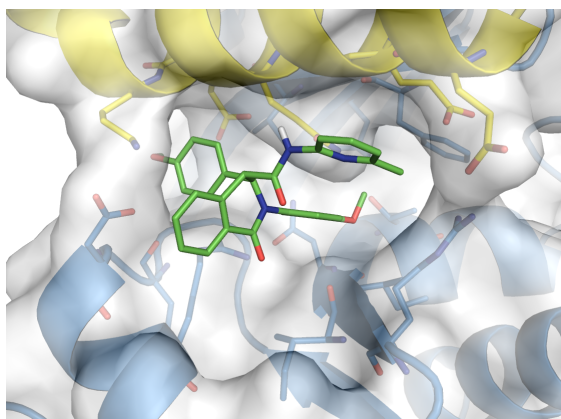


Abbildung 10.7.: Top-Pose (Run 14) des $(R,R)_{ax}$ Enantiomers. Der Autodock-Score beträgt $-7.07 \frac{kcal}{mol}$.

(R,R) -Enantiomer mit axialen Substituenten ($(R,R)_{ax}$)

Ganz im Gegensatz zu den äquatorialen Konformationen ergeben 47 der 50 Docking-Runs für das $(R,R)_{ax}$ -Enantiomer die gleiche Bindepose. In der Regel ist das ein positives Qualitätskriterium in einem Docking-Experiment. Oftmals finden sich dann starke Interaktionen, welche den Liganden in einer bestimmten Position fixieren. Allerdings ist es nicht zwingend so, dass der am häufigsten gefundene Bindemodus auch der biologisch korrekte ist.

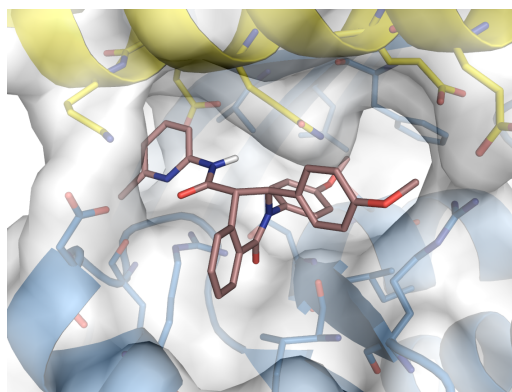
In Abbildung 10.7 ist die beste Pose basierend auf dem AutoDock-Score dargestellt. Die beiden tiefen Bindetaschen werden von den *para*-Methoxyphenyl-Substituenten adressiert. Der Amid-Substituent wird dagegen nicht in der Tasche vergraben, sondern ist lösungsmittel exponiert und steht in Kontakt mit der α -Helix der SBD.

Die Haupttriebkraft für diesen Bindemodus ist die gute sterische Komplementarität zwischen Ligand und Bindetasche. So ist beispielsweise dem Methoxyphenylrest in der S1-Bindetasche beinahe optimal zur Ausbildung einer Kation- π -Interaktion (mit Lys524) platziert. Der Docking-Score ist allerdings nicht wesentlich besser als für die beiden äquatorialen Liganden. Mit Ausnahme der angesprochenen Kation- π -Interaktion finden sich auch hier keine starken gerichteten Wechselwirkungen. Solche Interaktionen sind zwar nicht unbedingt erforderlich, bei einer spezifischen Bindung aber doch zumindest wahrscheinlich. Aus Sicht der Wirkstoffentwicklung wäre es wünschenswert, solche Interaktionen zu identifizieren, nicht zuletzt weil dadurch die Verifizierung des korrekten Bindemodus vereinfacht wird.

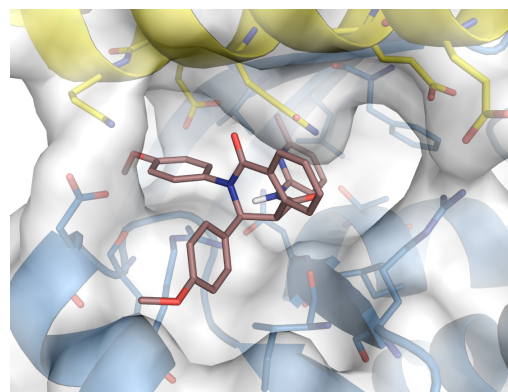
Löst man sich etwas von der Vorstellung eines starren Proteins, so wäre es denkbar, dass in diesem Bindemodus eine Wasserstoffbrücke zwischen Gln520 und dem Amid-Stickstoff des Liganden zustande kommt.

S,S-Enantiomer mit axialen Substituenten ((S,S)_{ax})

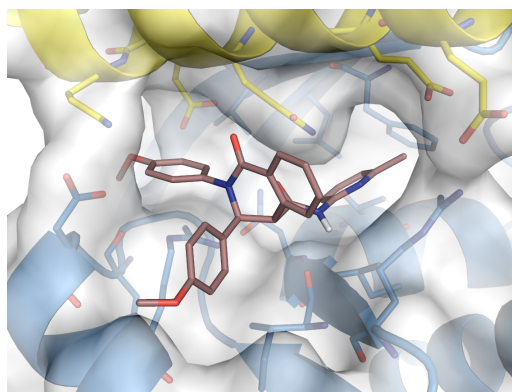
Weniger eindeutig sind die Ergebnisse beim Docking des S,S-Enantiomers mit axialer Substituentenstellung. Man erhält hier drei Cluster mit möglichen Bindeposen. Der größte Cluster hat 26 Mitglieder, die beiden kleineren 9 bzw. 7. Jeweils ein Vertreter ist in Abbildung 10.8 dargestellt.



a) Pose 13
Docked Energy: $-8.80 \frac{\text{kcal}}{\text{mol}}$



b) Pose 2
Docked Energy: $-8.43 \frac{\text{kcal}}{\text{mol}}$



c) Pose 32
Docked Energy: $-8.37 \frac{\text{kcal}}{\text{mol}}$

Abbildung 10.8.: Beste Docking-Lösung der drei Bindeposengruppen für das (S,S)_{ax}-Enantiomer.

Die insgesamt beste Pose (Abb. 10.8 a) entspricht hinsichtlich der Positionierung des Isochinolingrundgerüsts und des 2-N-Substituenten sehr gut der Top-Pose des (R,R)_{ax}-Enantiomers. Die beiden axialen Substituenten liegen dabei vertauscht vor, so dass der Methylpyridin-Rest nun in der S1-Tasche plaziert und der zweite Methoxyphenyl-Rest solvensexponiert ist.

Die anderen beiden Posen (Abb. 10.8 b und c) sind sich sehr ähnlich und unterscheiden sich nur in der Stellung der Amidgruppe und dementsprechend des Methylpyridin-Substituenten, der in beiden Fällen die S2-Tasche adressiert. Wie in Abbildung 10.8

b ersichtlich, ist so vor allem der Methylrest sehr gut im hydrophoben Teil der Binde-tasche vergraben. Der dritte Bindemodus (Abbildung 10.8 c) zeichnet sich hingegen vor allem durch eine Wasserstoffbrücke zwischen dem Amid-Stickstoff des Liganden und der Carbonylgruppe von Ile172 aus.

Obwohl die Posen ansonsten auch hauptsächlich von der guten sterischen Komple-mentarität mit der Bindetasche profitieren, sind die Docking-Scores mit Werten zwischen -8.37 und $-8.80 \frac{\text{kcal}}{\text{mol}}$ sehr viel besser als die der anderen drei Enantiomere.

10.1.2. Rescoring

Tabelle 10.1.: Docking-Ergebnisse für die vier möglichen Startkonformationen von C066-0757. Aufgeführt sind die besten AutoDock-Posen, die in den vorangegan-genen Kapiteln besprochen wurden. Zusätzlich sind die DSX-Rescoring Ergeb-nisse für diese Posen aufgeführt. In Klammern ist zusätzlich der relative Rang des Pose im DSX-Ranking sowie der Rang der besten Pose aus dem jeweiligen Cluster aufgeführt.

Enantiomer	AutoDock-Run	AutoDock-Score	DSX-Score (Rang/ Bester Rang aus Cluster)	Abbildung
$(S,S)_{\text{aq}}$	26	-6.59	-85.4 (29/29)	10.5 a
	36	-6.53	-100.3 (11/7)	10.5 b
	4	-5.82	-73.5 (39/39)	10.5 c
	15	-5.67	-87.6 (26/23)	10.5 d
$(R,R)_{\text{aq}}$	23	-6.98	-87.6 (25/25)	10.6 a
	45	-6.75	-88.5 (24/11)	10.6 b
	41	-5.41	-82.0 (37/22)	10.6 c
	35	-5.22	-93.4 (14/13)	10.6 d
$(R,R)_{\text{ax}}$	14	-7.07	-105.9 (30/1)	10.7
$(S,S)_{\text{ax}}$	13	-8.80	-116.6 (9/1)	10.8 a
	2	-8.43	-113.1 (24/12)	10.8 b
	32	-8.37	-116.3 (13/3)	10.8 c

In Tabelle 10.1 sind die AutoDock-Ergebnisse numerisch zusammengefasst. Zusätzlich zeigt die Tabelle die Ergebnisse des Rescorings mit DSX [79]. Dabei ist der Score für die entsprechende AutoDock-Pose angegeben. In Klammern ist der Rang dieser Pose

in der DSX-Rangliste sowie der Rang der besten Pose aus dem entsprechenden Cluster angegeben.

Das Rescoring mit DSX ergibt für die Posen mit axialer Substituentenstellung deutlich bessere Scores als für jene mit äquatorialen Substituenten. Die Posen des S,S-Enantiomers mit axialen Substituenten erhalten insgesamt die besten Scores. Das Rescoring mit DSX bestätigt somit qualitativ des Ranking des AutoDock-Scores.

Bei den Konformationen mit äquatorialen Substituenten fällt außerdem auf, dass die zuvor angesprochenen "Artefakte", die nicht in der gewünschten Bindetasche liegen, von DSX besser bewertet werden und die hier gezeigt Posen deswegen im DSX-Ranking eher weiter hinten auftauchen. Bei den Autodock-Scores konnte man diese Präferenz nicht feststellen.

10.1.3. Schlussfolgerung

Das Docking der vier Enantiomere legt nahe, dass die Verbindung mit axialer Substituentenstellung am ehesten im Interface von Hsp70 binden kann. Diese These wird zusätzlich durch experimentelle Studien gestützt, die zeigen, dass der Ligand in freier Lösung ebenfalls in axialer Form auskristallisiert [130].

Offen bleibt aber weiterhin die Frage, welches Enantiomer für die Aktivität verantwortlich ist. Sowohl der AutoDock-Score als auch der DSX-Score bevorzugen das S,S-Enantiomer. Die Möglichkeit, dass das R,R-Enantiomer, trotz der schwächeren Scoring-Ergebnisse, für die Aktivität verantwortlich ist oder zu dieser beiträgt, sollte - vor allem im Hinblick auf dessen sehr einheitliches Docking - weiterhin in Betracht gezogen werden.

10.2. Strukturbasiertes Design von Derivaten

Das in Kapitel 9.3 vorgestellte Syntheschema zeigt, dass die Synthese neuer Liganden durch Derivatisierung an den Positionen 2, 3 und 4 gut realisierbar ist (vgl. Abb. 9.2). Durch Variation der Edukte wurde so eine Substanzbibliothek mit unterschiedlichen Substituenten erstellt. Als wichtige Voraussetzung für die Weiterentwicklung der Substanzen standen dabei die Verifizierung bzw. Falsifizierung der in Frage kommenden Bindeposen sowie die Aufstellung von Struktur-Wirkungsbeziehungen im Vordergrund. Die Optimierung der Aktivität war zunächst zweitrangig. Die genaue Zusammenstellung war von der synthetischen Zugänglichkeit bzw. der kommerziellen Verfügbarkeit der Edukte geleitet.

Das Docking dieser Substanzbibliothek wurde mit der in Kapitel 10.1 beschriebenen Methode durchgeführt. Es wurden für jedes Enantiomer zunächst beide möglichen Ringkonformationen berücksichtigt.

Auch die biologischen Messwerte beziehen sich auf das *trans*-Enantiomergemisch. Wie in Kapitel 9.2 beschrieben, werden dabei die EC₅₀-Werte der Substanzen für INA-6

Zellen und PBMCs ermittelt. Die Aktivität der Verbindungen wird mittels der Zellviabilität und nicht durch einen protein-spezifischen Assay gemessen. Für die Aufstellung von Struktur-Wirkungsbeziehungen hat dies einen großen Einfluss, da die Modulation der Aktivität nicht ausschließlich von der Bindungsaffinität der Verbindung am Protein abhängt, sondern beispielsweise durch zelluläre Regulationseffekte überlagert wird. Es ist deswegen nicht zu erwarten, dass man anhand der Docking-Posen und Scoring-Vorhersagen direkte quantitative Beziehungen zu den biologischen Daten herstellen kann. Die folgende Auswertung konzentriert sich deswegen vorrangig auf qualitative Aussagen darüber, an welchen Stellen eine Substitution sinnvoll erscheint und inwiefern sich das mit den vorhandenen Bindungsmodellen erklären lässt.

Zum Zeitpunkt der Erstellung und Testung der Substanzbibliothek war noch nicht klar, welches Enantiomer für die biologische Aktivität verantwortlich ist. Durch eine chirale HPLC-Trennung ist es später gelungen nachzuweisen, dass nur das R,R-Enantiomer biologisch aktiv ist [130]. Hinsichtlich der Bindungskonformation erscheint eine axiale Substituentenstellung wahrscheinlicher. Zum einen zeigen die Docking-Ergebnisse von C066-0757 für diese Variante sowohl konsistentere Posen als auch bessere Scores. Zum anderen würde man bei einer äquatorialen Substituentenstellung erwarten, dass beide Enantiomere eine Aktivität zeigen, da diese sich sterisch sehr ähnlich sind und man im Wesentlichen dieselben Dockingposen erhält (vgl. Kapitel 10.1.1).

Obwohl ein eindeutiger Beweis zur aktiven Konformation also noch fehlt, sprechen diese Indizien dafür, dass das R,R-Enantiomer mit axialer Substituentenstellung das biologisch aktive Isomer darstellt. Dementsprechend werden im folgenden die Posen der (R,R)_{ax}-Enantiomere zur Interpretation der biologischen Daten verwendet. Die Quelle für die biologischen Daten ist die Dissertation von Dr. Andreas Hartung [130]. Die Bezeichnung der synthetisierten Verbindungen wurde ebenfalls aus diesem Dokument entnommen, um eine einfachere Vergleichbarkeit zu ermöglichen.

10.2.1. Variationen an den Positionen 2 und 3

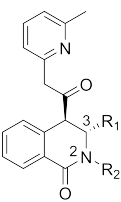






















Die erste Gruppe von Derivaten wurde durch die Variation der beiden 4-Methoxyphenyl-Substituenten erhalten. Mit Ausnahme der halogenierten Derivate sind alle Substanzen dieser Gruppen noch ohne Kenntnis der in Kapitel 10.1.1 vorgestellten Posen entstanden. Da der initiale Hit (Syntheseprodukt: trans-**24a**) mit einem EC₅₀-Wert von 0.7 μ M schon eine relativ gute Aktivität zeigte, wurden die Variationen hier zunächst relativ gering gehalten (Tabelle 10.2).

Zunächst wurden die Methoxy-Substituenten alternierend sowie parallel durch Wasserstoffatome ersetzt. Dies führte im Falle des R1-Substituenten (trans-**24h**) zu einem Aktivitätsverlust und im Falle des R2-Substituenten zu einem leichten Aktivitätsgewinn (trans-**24i**). Bei trans-**24h** ist der Aktivitätswert aufgrund von Präzipitation während des Assays nur eingeschränkt interpretierbar. Ersetzt man beide Methoxy-Gruppen durch Wasserstoffatome (trans-**24m**), verringert sich die Aktivität.

Das Docking dieser vier Liganden ergibt zwei mögliche Posen (Abbildung 10.9).

10. Weiterentwicklung der Leitstruktur C066-0757

Tabelle 10.2.: Biologische Aktivitäten der Verbindungen mit Derivatisierungen an den Positionen 2 und 3.

Verbindung	R1	R2	INA-6 EC ₅₀ (±sdv) [μM]	PBMC EC ₅₀ [μM]
 trans- 24a			0.68 (± 0.04)	> 100
trans- 24h			> 50 ^a	-
trans- 24i			0.41 (±0.03)	> 100
trans- 24m			2.3 (±0.6)	>100
trans- 36a			10.1 (±0.1)	> 100
trans- 36b			57 (±15)	> 100
trans- 36c			1.3 (±0.2)	> 100
trans- 36d			6.9 (± 0.3)	> 100
trans- 24r			30 (± 2)	80
trans- 24n			6.4 (±0.7)	> 100
trans- 24o			2.9 (±0.3)	> 100

^a Präzipitation unter Assay-Bedingungen.

Die erste Pose entspricht jener, die für den initialen Screening Hit in Kapitel 10.1.1 beschrieben wurde. Der Methylpyridin-Rest ist hier Richtung SBD α -Helix und Solvens orientiert. Die Phenylsubstituenten adressieren die S1- und die S2-Bindetasche.

In der zweiten möglichen Pose liegt die Methylpyridin-Substruktur hingegen in der S3-Bindetasche. Die Phenylsubstituenten adressieren weiterhin die S1- und S2-Tasche, liegen allerdings vertauscht vor. Die Scores für beide Posen liegen jeweils sehr nahe beieinander, wobei die Pose 2 zwar tendenziell den etwas höheren Score erhält, allerdings deutlich seltener auftritt als Pose 1. Den besten Score dieser Serie erhält die Ausgangsverbindung trans-**24a** in Pose 1. Beide Posen liefern jedoch keine schlüssige Erklärung für die Aktivitätsmodulation (wie beispielsweise eine Wasserstoffbrücke). Es ist deswegen anzunehmen, dass vor allem Solvatationseffekte für die unterschiedlichen Aktivitäten verantwortlich sind.

Bei den Verbindungen trans-**36a**, trans-**36b**, trans-**36c** und trans-**36d** handelt es

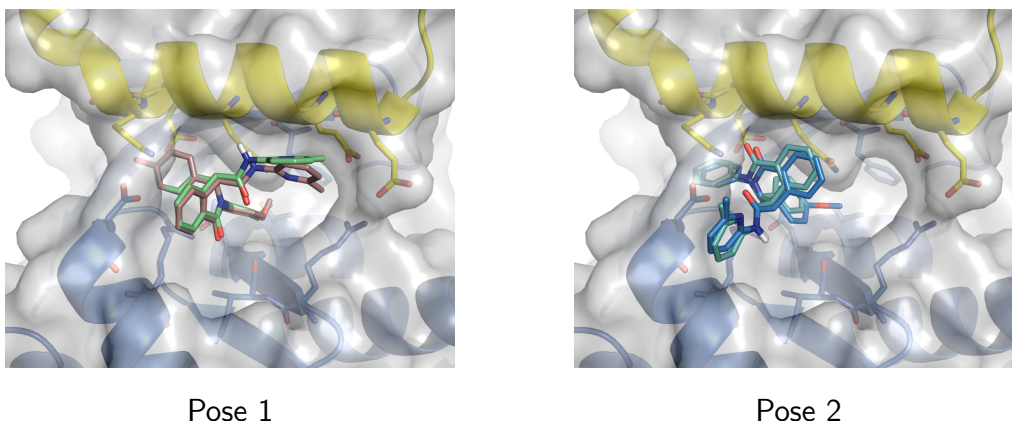


Abbildung 10.9.: Top-Posen der Liganden **trans-24a**, **trans-24h**, **trans-24i** und **trans-24m** mit axialer Substituentenstellung. **Trans-24a** und **trans-24h** werden in einer Pose gedockt, die der in Kapitel 10.1.1 für den initialen Screening-Hit beschriebenen entspricht (Pose 1). Bei den Liganden **trans-24i** und **trans-24m** ist hingegen der Methylpyridyl-Rest in Richtung der S3-Bindetasche orientiert. Die beiden Methoxyphenylreste adressieren weiterhin die S1- und S2-Tasche, liegen allerdings vertauscht vor (Pose 2).

sich um Carbonsäure-Derivate, bei denen die Methoxysubstituenten in gleicher Weise variiert wurden wie in der zuvor besprochenen Gruppe. Qualitativ erhält man innerhalb der Gruppe dieselben Aktivitätstrends wie in der Carboxamid-Gruppe. Absolut gesehen bewegen sich die Aktivitäten jedoch auf einem deutlich schwächeren Niveau. Beim Docking nehmen die Carbonsäure-Derivate bevorzugt die Pose 2 ein (siehe Tabelle 10.3). Die Gruppe erhält vergleichsweise geringe Scores, was in gutem Einklang mit den biologischen Aktivitäten steht.

In der dritten Gruppe wurden die beiden Methoxygruppen jeweils parallel durch Hydroxyl- (**trans-24r**), Chlor- (**trans-24o**) und Fluor-Substituenten (**trans-24n**) ersetzt. Keine der Verbindungen zeigt eine verbesserte Aktivität im Vergleich zu **trans-24a**. Man beobachtet stattdessen einen deutlichen Aktivitätseinbruch durch die Einführung von Hydroxy-Gruppen. Dies lässt sich durch einen erhöhten Desolvatisierungsterm und eine ungünstigere Platzierung dieser in der Bindetasche begründen. Dies zeigt zudem, dass die beiden Methoxygruppen der ursprünglichen Verbindungen nicht vorrangig aufgrund ihrer hydrophilen Eigenschaften (als Wasserstoffbrückenakzeptoren) an der Bindung beteiligt sind.

Qualitativ wird dieser Zusammenhang von der AutoDock Scoring-Funktion nur für **trans-24r** korrekt vorhergesagt. Die ungünstige Platzierung der polaren Hydroxygruppen in den Bindetaschen S1 und S2 führt zu einem verringerten Score. Quantitativ kann das Ausmaß des Aktivitätseinbruchs hingegen erwartungsgemäß nicht vorhergesagt werden. Ähnliches gilt für die halogenierten Derivate, die ebenfalls einen schwächeren Score als die Ausgangsverbindung erhalten, ohne dass die Scoring-Funktion in der Lage ist die

Tabelle 10.3.: AutoDock-Ergebnisse der Verbindungen mit Variationen an den Positionen 2 und 3 ((R,R)_{ax}-Enantiomere). Aufgeführt ist der Docking-Run der jeweiligen Pose, der Rang der Pose, die Größe des korrespondierenden Clusters (in Klammern) sowie "Estimated Free Energy of Binding" der AutoDock Scoring-Funktion (in $\frac{kcal}{mol}$). Pose 1 und Pose 2 bezieht sich auf die in Abbildung 10.9 gezeigten Bindemoden.

Verbindung	Pose 1			Pose 2		
	Run	Rang	Score	Run	Rang	Score
trans- 24a	14	1 (47)	-12.69	-	-	-
trans- 24h	32	1 (32)	-10.79	5	2 (1)	-11.08
trans- 24i	25	2 (41)	-10.41	28	1 (1)	-10.89
trans- 24m	44	5 (25)	-19.89	3	1 (3)	-10.83
trans- 36a	49	3 (12)	-8.68	48	1 (13)	-9.25
trans- 36b	-	-	-	11	2 (4)	-9.03
trans- 36c	43	5 (4)	-7.69	42	1 (9)	-8.96
trans- 36d	-	-	-	42	1 (12)	-8.61
trans- 24r	29	1 (44)	-10.50	-	-	-
trans- 24o	4	2 (5)	-11.10	32	1 (4)	-11.74
trans- 24n	26	5 (20)	-10.08	39	1 (5)	-10.83

experimentellen Aktivitätseinbußen genau zu quantifizieren.

Zusammengenommen lassen diese Derivate folgende Rückschlüsse zu. Der Methylpyridyl-Substituent hat eine deutlich positive Auswirkung auf die Aktivität. Anhand der Docking-Posen kann dies durch die Ausbildung von zusätzlichen hydrophoben Interaktionsflächen, die auch einen verbesserten Score zur Folge haben, erklärt werden. Das Docking legt zudem nahe, dass die beiden Methoxyphenyl-Substituenten tief in der Bindetasche vergraben vorliegen und dort ebenfalls günstige Kontaktflächen ausbilden. Eine direkte Wechselwirkung im Sinne einer Wasserstoffbrücke ist nicht zu erkennen.

Die AutoDock Scoring-Funktion ist in der Lage die grundlegenden Tendenz dieser Ligandenserie zu reproduzieren, nämlich, dass die Variation der Substituenten an Position 2 und 3 mit leichten Aktivitätseinbußen einhergeht. Eine genaue Quantifizierung dieser Modulationen gelingt hingegen nicht. Dies war allerdings auch nicht zu erwarten, da die Aktivitätsänderungen teilweise sehr gering sind und somit ausserhalb der Genauigkeitsgrenze von Scoring-Funktionen liegen. Hinzu kommt die zuvor beschriebene Problematik, dass es sich bei den Aktivitätsdaten um einen Ganz-Zell-Readout han-

delt und somit die Aufstellung von Struktur-Wirkungsbeziehungen in diesem Detailgrad sicherlich nicht möglich ist.






Die Inkubation mit PBMCs zeigt, dass keines dieser Derivate eine unspezifische Zelltoxizität aufweist, was für einen spezifischen Wirkmechanismus spricht.

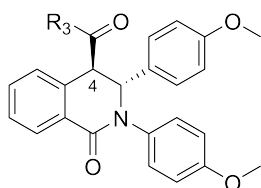
10.2.2. Variationen der Amid-Substituenten an Position 4

Weitere Derivate wurden durch Variation des Amid-Substituenten an Position 4 des Isochinolinon-Grundgerüsts erhalten (Tabelle 10.4). Bei den Derivaten **trans-24b** und **trans-24c** ist die Methylpyridyl-Gruppe durch einen Toly- bzw. eine Pyridyl-Rest ersetzt. Die Auswirkungen auf die Aktivität sind eher gering. Durch Elimination des Methylrestes wird die Aktivität leicht verbessert, während ein Toly-Rest eine leicht schwächere Aktivität bewirkt.

Interessanter sind die Derivate **trans-24e** und **trans-24f**, bei denen der Substituent nur aus einem sekundären bzw. einem tertiären Amin besteht. Während die Eliminierung der Methylpyridin-Gruppe bei **trans-24e** die Aktivität schon signifikant einschränkt und etwa auf das Niveau der Carbonsäure-Derivate absenkt, bewirkt die Einführung eines tertiärenamins an dieser Stelle einen vollständigen Aktivitätsverlust. Die Wasserstoffbrücken-Donor Funktionalität an dieser Stelle ist also von essentieller Bedeutung für die Aktivität.

Tabelle 10.4.: Biologische Aktivität der Verbindungen mit Variation des Amid-Substituenten in Position 4.

Verbindung	R3	INA-6 EC ₅₀ (±sdv) [μM]	PBMC EC ₅₀ [μM]
trans-24a		0.68 (±0.04)	> 100
trans-24b		0.88 (±0.10)	> 100
trans-24c		0.53 (±0.03)	> 100
trans-24e		7.1 (±0.3)	> 100
trans-24f		> 100	-



Das Docking ergibt auch hier die beiden aus Abbildung 10.9 bekannten Posen. Pose 1 wird bei den Liganden **trans-24b** und **trans-24f** etwas besser bewertet. Pose 2 erhält bei den Liganden **trans-24c** und **trans-24e** die etwas besseren Score, tritt aber bei allen vier Liganden sehr selten auf. Innerhalb der Serie zeigt kein Derivat eine besseren Score

Tabelle 10.5.: AutoDock-Ergebnisse der Derivate mit variierten Amid-Substituenten ((R,R)_{ax}-Enantiomere). Aufgeführt ist der Docking-Run der jeweiligen Pose, der Rang der Pose, die Größe des korrespondierenden Clusters (in Klammern) sowie "Estimated Free Energy of Binding" der AutoDock Scoring-Funktion (in $\frac{kcal}{mol}$). Die Bezeichnung Pose 1 und Pose 2 bezieht sich auf die in Abbildung 10.9 gezeigten Bindemoden.

Verbindung	Pose 1			Pose 2		
	Run	Rang	Score	Run	Rang	Score
trans- 24a	14	1 (47)	-12.69	-	-	-
trans- 24b	40	1 (26)	-11.04	45	3 (1)	-10.55
trans- 24c	12	1 (35)	-10.76	34	2 (1)	-11.08
trans- 24e	41	1 (23)	-9.42	27	2 (3)	-9.64
trans- 24f	39	1 (21)	-10.02	34	2 (2)	-9.92

als die Ausgangsverbindung, was gut mit den experimentellen Werten übereinstimmt. Da in keiner der Posen der (R,R)_{ax}-Enantiomere eine Wasserstoffbrücke identifiziert werden kann, ist der Aktivitätsverlust von trans-**24f** hingegen nicht erklärbar. Auf Seite 138f wurde gezeigt, dass das (S,S)_{ax}-Enantiomer von C066-0757 eine Pose zeigt, zu der diese Verhalten besser passen würde. Dabei interagiert die Amidgruppe über eine Wasserstoffbrücke mit dem Sauerstoff von Ile172.

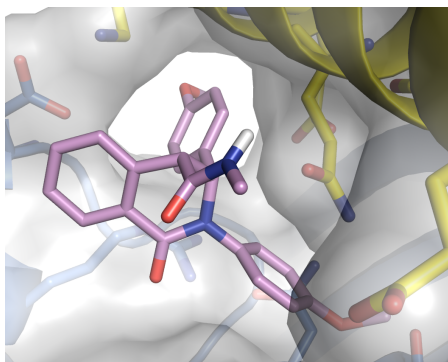


Abbildung 10.10.: Eine konformative Anpassung der Seitenkette von Gln520 würde die Bildung einer Wasserstoffbrücke ermöglichen. So ließe sich der Aktivitätsverlust von Substanzen mit tertiären Aminen erklären. Beispielhaft ist hier die Verbindung trans-**24e** ((R,R)_{ax}-Enantiomer) dargestellt.














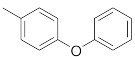

Das Bindungsmodell muss in diesem Bereich also noch weiter verfeinert werden. Eine Möglichkeit besteht darin, dass die Bindetaschenkonformation durch einen *Induced-Fit*-Mechanismus verändert wird. Es wurde schon erwähnt, dass es vor allem im Bereich der

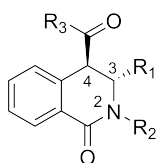
SBD α -Helix einige flexible, polare Seitenketten gibt, die hier involviert sein könnten. Ein möglicher Kandidat ist die Aminosäure Gln520. Diese befindet sich sehr zentral in der Bindetasche und ist - zumindest in der Kristallstruktur - in keine Interaktionen mit anderen Seitenketten involviert. Abbildung 10.10 zeigt die Pose von trans-**24e**, in der eindeutig keine Wasserstoffbrücke zum Protein besteht. Durch relativ geringe Rotationen von Gln520 und Ligand wäre in diesem Bereich jedoch die Ausbildung einer Wasserstoffbrücke denkbar.

10.2.3. Einführung sterisch anspruchsvoller Reste

Durch die Einführung von sterisch anspruchsvolleren Resten an allen drei Variationspositionen sollte die Charakteristik der Bindung weiter untersucht werden. Die Reste wurden dabei so gewählt, dass die Liganden nicht mehr in der Lage sein sollten, im postulierten Bindemodus zu binden. Dabei ist zu beachten, dass zu kleine Reste durch die Proteinflexibilität ausgeglichen werden können. Wählt man hingegen zu große Substituenten, würde dies bei jeglicher Art von spezifischer Bindung einen Aktivitätsverlust bewirken und somit keine spezifische Aussage mehr erlauben. Die Variationsmuster dieser Derivate bewegen sich im Bereich einzelner Methylgruppen bis hin zu größeren Gruppen wie zum Beispiel Phenylresten.

Tabelle 10.6.: Derivate mit sterisch anspruchsvollen Substituenten an verschiedenen Positionen.

Verbindung	R1	R2	R3	INA-6 EC ₅₀ (\pm sdv) [μ M]	PBMC EC ₅₀ [μ M]
trans- 24a				0.68 (\pm 0.04)	> 100
trans- 24d				0.95 (\pm 0.09)	> 100
trans- 24g				> 100	-
trans- 24p				43 (\pm 12) ^b	-
trans- 24q				> 100 ^b	-



^aVerbindung präzipitiert unter Assay-Bedingungen.

Die Einführung einer zusätzlichen Methylgruppe in meta-Position des R3-Substituenten (trans-**24d**) führt zu einer leichten Abschwächung der Aktivität. Der AutoDock-Score

Tabelle 10.7.: AutoDock-Ergebnisse für die Derivate mit sterisch anspruchsvollen Substituenten ((R,R)_{ax}-Enantiomere). Aufgeführt ist der Docking-Run der jeweiligen Pose, der Rang der Pose, die Größe des korrespondierenden Clusters (in Klammern) sowie "Estimated Free Energy of Binding" der AutoDock Scoring-Funktion (in $\frac{kcal}{mol}$). Die Bezeichnung Pose 1 und Pose 2 bezieht sich auf die in Abbildung 10.9 gezeigten Bindemoden.

Verbindung	Pose 1			Pose 2		
	Run	Rang	Score	Run	Rang	Score
trans- 24a	14	1 (47)	-12.69	-	-	-
trans- 24d	24	1 (20)	-11.00	31	11 (1)	-9.98
trans- 24g	17	1 (22)	-12.17	-	-	-
trans- 24p	-	-	-	47	1 (3)	-11.75
trans- 24q	25	1 (34)	-12.80	-	-	-

ist für diese Verbindung um relativ deutliche $1.69 \frac{kcal}{mol}$ schwächer (Tabelle 10.7).

Die Einführung einer Tetrahydronaphthylgruppe an Position R3 (trans-**24g**) führt hingegen zu einem kompletten Aktivitätsverlust. Dieser Ligand wurde eigentlich entwickelt, um eine der Posen des (S,S)_{ax}-Enantiomers, bei denen dieser Substituent in der S2-Bindetasche positioniert ist, zu verifizieren. In den Posen des (R,R)_{ax}-Enantiomers ist der Substituent hingegen solvensexponiert, so dass Bindepose 1 weiterhin möglich ist. Der Aktivitätsverlust ist anhand der Docking-Ergebnisse des (R,R)_{ax}-Enantiomers daher nicht erklärbar. Beim Docking wird für die Verbindung eine ähnlich hohe Aktivität wie für die Ausgangsverbindung vorhergesagt (vgl. Abbildung 10.9). Das Docking spricht hier also eher für (S,S)_{ax} als das aktive Enantiomer. Eine mögliche Erklärung für den Affinitätsverlust besteht darin, dass der Tetrahydronaphthylrest mit der essentiellen Wasserstoffbrücke, die vom Amidstickstoff ausgeht, interferiert, indem er die Ausbildung der nötigen Konformation auf Ligand- oder Proteinseite behindert. Ein andere Möglichkeit ist, dass die Verbindungen doch nicht im bisher favorisierten Bindemodus 1 binden. Wie man in Tabelle 10.7 sieht, kann trans-**24g** beispielsweise nicht in Pose 2 binden.

Beim Liganden trans-**24p** wurden die beiden Methoxyphenyl-Substituenten jeweils durch einen *meta*-Xylolrest ersetzt. Auch dies führt zu einem deutlichen Aktivitätsverlust, allerdings präzipitierte die Verbindung unter Assay-Bedingungen. Die Ergebnisse sollten deswegen nur vorsichtig interpretiert werden. Das Docking zeigt deutlich, dass vor allem die sehr enge S1-Bindetasche von den Xylol-Resten nicht adressiert werden kann. Beim Docking ist Bindemodus 1 für trans-**24p** dementsprechend nicht möglich. In seltenen Fällen wird eine dem Bindemodus 2 ähnliche Pose vorhergesagt. Dabei ist der Ligand etwas tiefer in Richtung S1-Tasche vergraben. Dadurch können die Methyl-

gruppen des Xylol-Rests am gegenüberliegenden Ausgang der Tasche platziert werden. Die Pose kommt unter den 50 Posen nur drei Mal vor, weist aber einen durchaus guten Score von $11.75 \frac{\text{kcal}}{\text{mol}}$ auf. Durch die parallele Substitution von **trans-24p** kann man nicht ableiten, welcher der beiden Substituenten die enge S1-Bindetasche adressiert. Eine alternierende Substitution wäre deswegen sehr interessant.

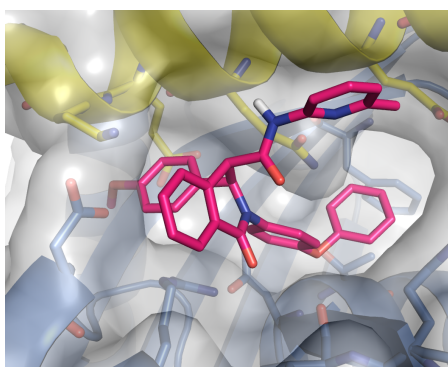


Abbildung 10.11.: Top-Pose von **trans-24q**. Der vergrößerte Phenoxyphenyl-Substituent passt in die S2-Bindetasche, bildet aber keine optimalen Kontaktflächen zum Protein aus. Die Rotationsfreiheit der Seitenkette von Gln520 ist stark eingeschränkt.

Bei der Verbindung **trans-24q** wurde die Methoxyphenyl-Gruppe an Position R2 durch ein Phenoxyphenyleinheit ersetzt. Die Verbindung zeigt ebenfalls keine Aktivität, wobei auch hier Präzipitate während des Assays beobachtet wurden. Das Docking ergibt, dass der größere Rest eigentlich noch in die Bindetasche S2 passen sollte (Abbildung 10.11). Diese Pose erhält sogar einen etwas höheren Score als die Ausgangsverbindung. Die Positionierung der Phenoxyphenyl-Gruppe in einer der anderen Subtaschen ist nicht möglich. Eine andere mögliche Erklärung für die eingeschränkte experimentelle Aktivität ist, dass der größere Rest mit der Ausbildung von essentiellen Interaktionen interferiert. Konkret wäre vorstellbar, dass die Rotationsfreiheit von Gln520 eingeschränkt wird und somit die nötige Geometrie zur Ausbildung einer Wasserstoffbrücke mit dem Liganden nicht mehr möglich ist.

10.3. Zusammenfassung und Diskussion

Die Bindetasche im Interface zwischen NBD und SBD wurde mit Hilfe von Docking-Methoden untersucht. Ausgangspunkt war der Screening-Hit C066-0757. Aus experimentellen Studien war bekannt, dass eines der *trans*-Enantiomere für die Aktivität verantwortlich ist. Das Docking des Diastereomenpaars mit unterschiedlichen Startkonformationen in die Bindetasche der von Clemens Grimm gelösten Proteinstruktur ergab, dass der Ligand höchstwahrscheinlich mit axialer Substituentenstellung bindet.

Trotzdem konnte zunächst keine eindeutig bevorzugt Bindepose identifiziert werden, so dass beim Design von weiteren Derivaten mehrere mögliche Posen in Betracht gezogen wurden.

Im Mittelpunkt der Zusammenstellung der Substanzbibliothek stand dementsprechend die Verifizierung und Falsifizierung der in Frage kommenden Bindeposen sowie die Aufstellung von Struktur-Wirkungsbeziehungen. Die daraus resultierenden biologischen Daten lassen sich folgendermaßen zusammenfassen [130]:

1. Nur die R,R-Enantiomere sind biologisch aktiv.
2. Carboxamide sind aktiver als die analogen Carbonsäuren. Eine aromatische Substitution des Carboxamids hat einen positiven Einfluss auf die Aktivität.
3. Die Methoxyfunktionen an den Phenylsubstituenten modulieren die Aktivität. Trägt der Substituent an Position 2 eine Methoxyfunktion hat dies einen negativen Einfluss auf die Aktivität. Im Fall des Phenylsubstituenten an Position 3 wirkt sich die Methoxyfunktion hingegen positiv auf die Aktivität aus. Insgesamt sprechen die Ergebnisse dafür, dass beide Methoxyfunktionen im Protein vergraben sind und hauptsächlich aufgrund günstiger Kontaktflächen zur Aktivität beitragen. An einer Wasserstoffbrücke ist wohl hingegen keine der beiden Gruppen beteiligt.
4. Die Substitution des sekundären Amid-Stickstoffs durch ein tertiäres Amid führt zum kompletten Aktivitätsverlust. Das spricht dafür, dass die Amidgruppe in eine Wasserstoffbrückenbindung involviert ist, die essentiell für die Bindung der Liganden ist.
5. Die Einführung sterisch anspruchsvoller Reste an Position 2 sowie am Amid-Substituenten führen zu einem signifikanten Aktivitätsverlust.
6. Toxische Effekte auf PBMCs werden nur in Ausnahmefällen beobachtet.

Diese sehr detaillierten und spezifischen Struktur-Wirkungsbeziehungen sowie das Fehlen von toxischen Effekten sind insgesamt sehr positiv zu bewerten und weisen auf einen spezifischen Wirkmechanismus der Substanzklasse hin.

Die ersten drei Punkte können mit Hilfe von Docking-Simulationen zu großen Teilen erklärt werden. Für die Punkte 4 und 5 ist das Bindungsmodell hingegen nicht präzise genug. Zur Verfeinerung des Modells ist vor allem eine Untersuchung der Flexibilität der Bindetasche erforderlich. Einige der experimentellen Ergebnisse (z.B. der Aktivitätsverlust bei Einführung eines tertiären Amins) können schon durch relativ kleine konformationale Veränderungen in der Bindetasche begründet werden.

Insgesamt muss die Bindetasche ohnehin als flexibel angesehen werden, da sie aus zwei Domänen der Proteins aufgebaut ist, die sich nachweislich unabhängig voneinander

bewegen und die im allosteren Zyklus des Proteins verschiedene Konformationen einnehmen. Bei der Kristallisation wurden die Liganden vermutlich durch einen Überschuss TMAO aus der Bindetasche verdrängt, was die Annahme stützt, dass es durch die Bindung der Liganden zu Induced-Fit-Effekten kommen kann. Die molekulardynamischen Untersuchungen, die im nächsten Kapitel (ab Seite 153) beschrieben werden, sind in diesem Zusammenhang sehr wichtig.

Bei einer kritischen Betrachtung muss jedoch gesagt werden, dass die Docking-Simulationen auf einer Reihe noch nicht eindeutig belegter Hypothesen basieren. So fehlt bisher ein schlüssiger experimenteller Nachweis, dass die Interface-Region tatsächlich den Angriffspunkt der Tetrahydroisochinoline darstellt. Neben der Blockade von Nukleotid- oder Substratbindetasche, was den Docking-Analysen von Ulrich Peinz zufolge eher unwahrscheinlich ist, kommt beispielsweise auch die Inhibition von Protein-Protein-Interaktionen mit Co-Chaperonen und Klientenproteinen und somit die Verhinderung der Bildung des funktionellen Chaperonkomplexes in Frage. Balaburski und Mitarbeiter haben 2013 einen Inhibitor präsentiert, der potentiell in einer bisher unbekanntem Bindetasche in der LID-Region der SBD bindet [134]. Es bestehen also theoretisch sehr vielfältige Möglichkeiten zur Inhibition von Hsp70.

Darüber hinaus ist auch noch nicht einwandfrei nachgewiesen, ob Hsp70 intrazellulär überhaupt die (alleinige) Zielstruktur der Tetrahydroisochinoline darstellt. So haben Rothweiler und Mitarbeiter die Protein-Protein-Interaktion zwischen MDM2 und p53 als Zielstruktur einer Reihe von Tetrahydroisochinolinonen identifiziert, die ein sehr ähnliches Substitutionsmuster aufweisen, wie die hier vorgestellten [135]. Auch MDM2 und p53 sind in die zelluläre Stressantwort involviert und die Inhibition hat einen proapoptischen Effekt. Für die hier vorgestellten Liganden weisen die zellulären Effekte, beispielsweise die Regulation von Klientenproteinen, jedoch deutlich auf eine Blockade der Hsp70-Funktion hin.

Trotzdem ist für die Weiterführung des strukturbasierten Designs dringend weitere experimentelle Evidenz notwendig, die belegt, dass - und wenn möglich wie - die Verbindungen an Hsp70 binden. Die Lösung einer Komplex-Kristallstruktur würde dabei nach wie vor die Ideallösung darstellen. Hierzu wurden inzwischen spezielle Liganden entwickelt und synthetisiert, die zusätzlich mit einer Trimethylamin-Substruktur substituiert sind. Diese Substruktur soll das TMAO-Molekül aus der Bindetasche verdrängen und so die Lösung eines Co-Kristalls von Hsp70 und Ligand ermöglichen. Zusätzlich wird an der Etablierung enzymspezifischer Assays für Hsp70, wie z.B. einem ATPase-Assay, gearbeitet. Die Bindung an Hsp70 könnte auch mittels *Isothermal Titration Calorimetry* (ITC) oder *Surface Plasmon Resonance* (SPR) nachgewiesen werden. Entsprechenden Experimente werden derzeit in der Gruppe von Dr. Clemens Grimm durchgeführt.

Trotz der vielen noch offenen Punkte überwiegen die positiven Aspekte, wie die gute biologische Aktivität in Kombination mit deutlichen Struktur-Wirkungsbeziehungen und das Fehlen von unspezifischer Toxizität. Die Weiterentwicklung dieser Substanzklasse als Inhibitoren von Hsp70 ist deswegen in jedem Fall als vielversprechend anzusehen.

11. Molekulardynamische Simulationen

11.1. Einleitung

Die Bildung eines Protein-Ligand-Komplexes ist ein dynamischer Vorgang, wobei die Bindungspartner auch im komplexierten Zustand Teile ihrer Flexibilität beibehalten. Die Untersuchung dieser Dynamik kann wichtige Hinweise für das Verständnis der Protein-Ligand-Bindung und somit für das strukturbasierte Wirkstoffdesign liefern.

Die Dynamik von Hsp70 wurde daher mit Hilfe molekulardynamischer Simulationen untersucht. Dabei sollten verschiedene Fragen beantwortet werden. Im Mittelpunkt der Analysen steht jedoch die Bindetasche der Tetrahydroisochinolinone in der Interfaceregion zwischen NBD und SBD, die im Hinblick auf ihre Stabilität, ihren Konformationsraum und den Einfluss verschiedener Faktoren (z.B. die Gegenwart eines Liganden) untersucht wird.

Bei allen Simulation spielt auch die Frage nach der generellen Stabilität der Proteinkonformation *hsc_az* eine Rolle. In Kapitel 8.3 wurde beschrieben, dass die *hsc_az*-Konformation von einigen Experten als Artefakt angesehen wird, welches nur aufgrund der Mutationen, der trunkeierten LID-Domäne und Kristallpackungseffekten entsteht [43], *in vivo* allerdings keine Rolle spielt. Diese These würde entkräftet werden, wenn das Protein in den hier durchgeführten Simulationen diesen Zustand beibehält.

In Kapitel 8.2 und 8.3 ist außerdem beschrieben, dass Hsp70 (bzw. DnaK) in seinem funktionellen Zyklus verschiedene allosterische Zustände einnimmt, die sich wesentlich durch die Stellung von NBD und SBD zueinander unterscheiden. Es ist allerdings nicht zu erwarten, dass im Rahmen der hier durchgeführten Simulationen der Übergang zu einem dieser allosterischen Zustände simuliert werden kann. Zum einen liegt die hier simulierte Zeitskala (ca. 100 ns) deutlich unter den Zeiträumen, die experimentell für den Übergang zwischen den allosterischen Zuständen ermittelt wurden (Millisekunden bis Sekunden [19]). Zum anderen sind an den Übergängen *in vivo* Co-Chaperone beteiligt, die in den MD-Simulationen nicht berücksichtigt werden.

11.1.1. Durchgeführte Simulationen

Apo-Simulation In der Apo-Simulation wurde die Dynamik der Proteinstruktur *hsc_az*, so wie sie von Dr. C. Grimm gelöst wurde, über einen Zeitraum von 100 ns

simuliert. Da in dieser Struktur die Linker-Region (zwischen den Aminosäuren 380 und 395) nicht vorhanden ist, wurde dieser Proteinbereich aus der PDB-Struktur 1YUW eingefügt (vgl. Kapitel 11.2.2). Darüber hinaus wurden keine Änderungen vorgenommen.

Docked-Simulation In einer zweiten Simulation wurde *hsc_az* im Komplex mit Apoptozol simuliert, welches durch Docking mit Glide in der Interface-Bindetasche plaziert wurde. Durch diese Simulation soll der Effekt eines im Interface gebundenen Liganden auf die Stabilität der Bindetasche und des Proteins insgesamt untersucht werden. Als Ligand wurde Apoptozol gewählt, weil es zu diesem Zeitpunkt der einzige durch einen proteinspezifischen Assay bestätigte Inhibitor von Hsp70 war und die Vorlage zur Entdeckung der Tetrahydroisochinolinone lieferte. Zur Aufklärung der spezifischen Struktur-Wirkungsbeziehungen der Tetrahydroisochinolinone sind in Zukunft weitere Simulationen mit Vertretern dieser Substanzklasse notwendig.

Wildtyp-Simulation In der dritten Simulation wurden die beiden mutierten Aminosäuren Ala213 und Ala214 in ihre Wildtyp-Formen (Glu213 und Asp214) zurückmutiert. Dadurch soll untersucht werden, ob die Mutationen zu Alanin, die zum Zweck der besseren Kristallisierbarkeit durchgeführt wurden, einen kritischen Einfluss auf die Bindetaschen- oder Proteinkonformation haben.

ATP-Simulation In der vierten Simulation wurde ein ATP-Molekül in der Nukleotid-Bindetasche der NBD plaziert. Bei der Analyse dieser Trajektorie stehen der Einfluss von ATP auf die Konformation der NBD sowie allosterische Effekte von ATP auf die Interface-Region und die Interface-Bindetasche im Mittelpunkt.

11.2. Methode und Durchführung

11.2.1. Grundlagen

Bei molekulardynamischen Simulationen werden die Kräfte, die auf die Atome eines Systems wirken, nach den Gesetzen der klassischen Mechanik berechnet.

Die Atome werden dabei als Massepunkte angesehen, die über Federn miteinander verknüpft sind. Die kovalenten Kräfte, die in diesem System auf jedes einzelne Atom wirken, werden in einem molekularmechanischen Kraftfeld als Funktion der Bindungslängen, Bindungswinkel und Torsionswinkel modelliert.

Zusätzlich wird jedes Atom durch van-der-Waals-Kräfte und durch elektrostatische Wechselwirkungen beeinflusst. Zur Abschätzung der van-der-Waals-Kräfte wird das Lennard-Jones-Potential verwendet, während die elektrostatische Kraft nach dem Coulomb'schen Gesetz berechnet wird. Gemeinsam bilden diese beiden Terme den nicht-kovalenten Anteil eines molekularmechanischen Kraftfeldes.

Es stehen verschiedene Kraftfelder wie AMBER [136, 137, 131, 132], CHARMM [138] oder GROMOS [139] zur Verfügung. Die Unterschiede liegen jeweils in der Parametrisierung der atomspezifischen Kraftfeldparameter, wie den Federkonstanten oder den *van-der-Waals*-Parametern. Die Parametrisierung eines Kraftfelds erfolgt auf der Basis von quantenchemischen Berechnungen und experimentellen Untersuchungen. Trotz der unterschiedlichen Parametrisierung liefern die verschiedenen Kraftfelder grundsätzlich recht ähnliche Ergebnisse [140].

Auf Basis dieser Kraftfelder lassen sich durch Integration der Newton'schen Bewegungsgleichung die Vektoren (Geschwindigkeit und Position) aller Atome in Abhängigkeit von der Zeit berechnen.

Weiterführende Details zur Berechnung der Kraftfeldterme, der Parametrisierung der Kraftfelder und den Methoden zur Integration der Bewegungsgleichung finden sich im Lehrbuch "Molecular Modelling: Principles and Applications" von Andrew Leach [141] in den Kapiteln 4 bis 7, die auch zur Ausarbeitung dieses Abschnitts verwendet wurden.

11.2.2. Protein-Aufbereitung

Als Grundlage für alle vier Simulationen diente die Proteinstruktur *hsc_az*. Wie in Kapitel 9.4 beschrieben, fehlt hier die Linker-Region zwischen den Aminosäuren 385 und 395. Aus diesem Grund wurde die Struktur mit dem Programm Maestro ¹ mit der 1YUW-Struktur überlagert, der fehlende Bereich aus 1YUW ausgeschnitten und in *hsc_az* eingefügt. Mit Hilfe der Software Prime ² [142, 143] wurde dann eine Energieminimierung im OPLS_2005 Kraftfeld durchgeführt, um durch die neu eingefügten Reste verursachte Spannungen zu beheben. Daher wurden nur der Linker und die direkt angrenzenden Proteinbereiche für die Minimierung berücksichtigt.

Für die Startstruktur der Wildtyp-MD wurden - ebenfalls mit Prime - die beiden mutierten Aminosäuren (Ala213 und Ala214) in ihre nativen Aminosäuren rückmutiert (A213E/A214D) und energieminiert.

Mit *tLeap* aus AMBER10 [131] wurden dem Protein Wasserstoffatome hinzugefügt. Dabei wurden die Standardprotonierungszustände der Aminosäuren verwendet; Histidine wurden am ϵ -Stickstoff protoniert. Im Anschluss wurden den Proteinatomen die Kraftfeldparameter des ff99SB-Kraftfelds aus AMBER10 zugewiesen.

11.2.3. Ligand-Aufbereitung

Apoptozol Für die Docked-Simulation wurde Apoptozol zunächst entsprechend der in Kapitel 10.1 beschriebenen Methode mit AutoDock3.0 in die Interface-Bindetasche von *hsc_az* gedockt. Im Unterschied zum Docking von Ulrich Peinz konnte hier jedoch

¹Maestro, Version 9.3, Schrödinger, LLC, New York, NY, 2012

²Suite 2012: Prime, version 3.1, Schrödinger, LLC, New York, NY, 2012

kein günstiger Bindemodus gefunden werden. Die 50 Docking-Runs ergaben 34 unterschiedliche Cluster. Der beste Docking-Score von $-3.47 \frac{\text{kcal}}{\text{mol}}$ ist ebenfalls vergleichsweise schwach. Der von Ulrich Peinz für Apoptozol in 1YUW ermittelte Bindemodus, ist aufgrund der unterschiedlichen Stellung von Gln520 in *hsc_az* nicht möglich.

Zum Vergleich wurde daher ein Docking mit Glide durchgeführt. Die Methode entspricht der in Kapitel 4.2.1 beschriebenen. Auch hier konnte der 1YUW-Bindemodus nicht reproduziert werden. In den zehn resultierenden Docking-Lösungen traten im Wesentlichen zwei mögliche Bindemoden auf. Die beste Pose (GlideScore = 5.842) zeigt eine gute Übereinstimmung mit dem Bindemodus, der für C066-0757 ((R,R)_{ax}-Enantiomer) als wahrscheinlich gilt (vgl. Kapitel 10.1.1). Dabei liegen die beiden Methoxyphenyl-Reste von Az in den tiefen Subtaschen S1 und S2. Der hydrophobe, Trifluormethyl-substituierte Phenylrest füllt die S3-Tasche aus und die - im Fall von Apoptozol endständige - Carboxamidgruppe ist in Richtung α -Helix der SBD orientiert (Abbildung 11.1). Diese Pose wurde als Startstruktur für die MD-Simulation ausgewählt.

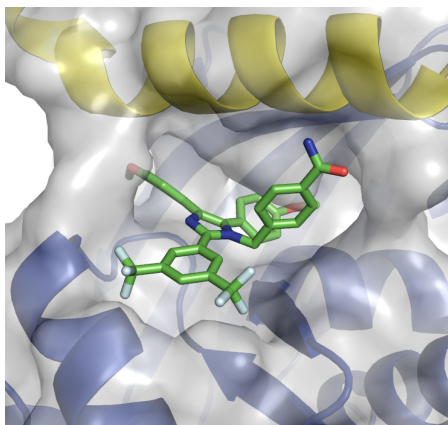


Abbildung 11.1.: Die beste Glide-Pose von Az in *hsc_az* zeigt grundlegend dieselben Bindungscharakteristika wie C066-0757 und dessen Derivate und diente daher als Startstruktur der Docked-MD.

ATP Für die ATP-Simulation wurde *hsc_az* mit der PDB-Struktur 1KAX überlagert. 1KAX ist eine Kristallstruktur der isolierten NBD-Domäne von bovinem Hsp70 mit gebundenem ATP. Nach der Überlagerung wurde ATP sowie das als Cofaktor benötigte Magnesiumion aus 1KAX ausgeschnitten und in die *hsc_az*-Struktur eingefügt. Mit Hilfe des Glide-Algorithmus wurde die ATP-Pose in der Bindetasche optimiert. Dabei wurde jedoch kein komplettes Docking durchgeführt, sondern nur die gegebene Pose auf Basis des GlideScore optimiert. Der grundsätzliche Bindemodus bleibt dabei erhalten. Die Konformation einiger Seitenketten, die enge Kontakte zum ATP-Molekül aufweisen, wurden zuvor mit Hilfe von Prime im OPLS_2005-Kraftfeld angepasst.

Für beide Liganden wurden die atomaren Partialladungen mittels der RESP-Methode [144] festgelegt. Das elektrostatische Potential wurde dazu auf HF/6-31G*-Niveau mit Gaussian 03 [145] berechnet. Die Kraftfeldparameter wurden mit Hilfe der Module *antechamber* und *parmchk* [146] basierend auf dem GAFF-Kraftfeld [147] zugewiesen.

11.2.4. Equilibrierung

Die Startstrukturen der vier Simulationen wurden anhand des folgenden Schemas equilibriert. Zunächst wurde eine kurze Energieminimierung über 200 Zyklen mit Hilfe eines impliziten Solvensmodells nach der Generalized-Born-Methode [148, 149] durchgeführt.

Nachdem mit *tLeap* Natriumionen bis zur elektrischen Neutralität hinzugefügt wurden, wurde das System in einer TIP3P-Wasserbox solvatisiert [150]. Die Größe dieses virtuellen Wasserbads wurde dabei so gewählt, dass in jeder Raumrichtung 7 bis 10 Å Abstand zwischen dem Protein und den Rändern des Wasserbads bestanden. Um Randeffect zu vermeiden wurden außerdem periodische Randbedingungen gewählt. Das bedeutet, dass Atome, die das Wasserbad auf einer Seite verlassen im selben Moment auf der gegenüberliegenden Seite wieder in das System eintreten.

Unter Fixierung aller Proteinatome wurde das System dann zunächst über 20 ps von 100 K auf 300 K aufgeheizt und innerhalb von 5 ps wieder auf 100 K abgekühlt. Im Anschluss wurde das komplette System über weitere 25 ps schrittweise wieder auf 300 K aufgeheizt. Diese Berechnungen wurden mit *sander* aus Amber10 [151] durchgeführt.

11.2.5. Production Run

Die Simulationen wurden dann bei konstantem Druck (1 bar) und konstanter Temperatur (300 K) über einen Zeitraum von mindestens 100 ns fortgeführt. Für die Apo- und die ATP-Simulation wurden genau 100 ns berechnet. Die Wildtyp- und die Docked-Simulation wurden über 113 bzw. 118 ns simuliert. Die Berechnungen wurden mit NAMD 2.9 unter Verwendung der zugewiesenen AMBER-Kraftfeldparameter durchgeführt [152].

Für die Berechnung der nicht-kovalenten Kraftfeldterme wurde im Fall der van-der-Waals Wechselwirkungen ein Abbruchradius von 12 Å verwendet. Zur Berechnung von elektrostatischen Wechselwirkungen wurde die Particle-Mesh-Ewald-Methode [153] eingesetzt.

Für die Integration der Newton'schen Bewegungsgleichung wurde ein Zeitschritt von 2 fs verwendet. Mit diesem Zeitintervall werden alle wesentlichen Bewegungen des Systems erfasst. Die Valenzschwingungen der Wasserstoffbindungen, die deutlich schneller sind, wurden mittels des SHAKE-Algorithmus [154] fixiert.

Die Atomkoordinaten wurden in Intervallen von 1 ps gespeichert. Bei der Analyse der Ergebnisse wurde meist jeder hunderste dieser Snapshots verwendet. Bei einer Laufzeit von 100 ns wurden also insgesamt 1000 Snapshots analysiert. Dadurch ist anhand der

11. Molekulardynamische Simulationen

Snapshot-Nummer auch sehr einfach der jeweilige Zeitpunkt ersichtlich, an dem Konformation auftritt. Der Snapshot ATP-275 entspricht beispielsweise der Konformation der ATP-Simulation bei 27.5 ns. Bei der Auswertung wird diese Nomenklatur zur Bezeichnung der Snapshots verwendet.

Für die RMSD- und Abstandsberechnungen wurde das Programm VMD [155] verwendet. Die Visualisierungen wurden mit VMD und PyMol [156] erstellt. Bei der statistische Analyse der Daten und der Erstellung der Diagramme kam die Statistiksoftware R [84] zum Einsatz.

12. Ergebnisse: Molekulardynamische Simulationen

12.1. Dynamik des Gesamtsystems

Für die Analyse der MD-Trajektorien ist es wichtig, dass das System equilibriert ist, d.h. das die Gesamtenergie des Systems einen konstanten Wert beibehält bzw. sich innerhalb eines konstanten Bereichs bewegt. In den vier durchgeführten Simulationen beobachtet man jeweils zu Beginn einen Drift der Gesamtenergie. Im Docked- und im Apo-System erkennt man nach etwa 1 ns eine Stabilisierung der Gesamtenergie. Im Wildtyp-System dauert dies 1.5 ns. Das Apo-System hat mit knapp 2 ns die längste Equilibrierungszeit. Verantwortlich für den Energie-Drift ist in allen Fällen der elektrostatische Kraftfeldterm, während die anderen Terme bereits nach deutlich kürzerer Zeit stabile Werte annehmen (Daten nicht gezeigt). In den folgenden Abbildungen ist jeweils der gesamte Verlauf der Simulationen gezeigt. Eine Auswertung ist jedoch nur innerhalb des equilibrierten Zeitraums sinnvoll.

In Abbildung 12.1 sind die RMSD-Verläufe der vier Simulationen mit der jeweiligen Startstruktur als Referenz gezeigt. Man erhält in allen Fällen zu Beginn deutliche Anstiege der RMSD-Werte, die auch deutlich über die Equilibrierungsphasen (s.o.) hinweg verlaufen. Einzig das Apo-System bleibt zunächst bei einem RMSD-Wert von unter 2 Å. Wie bei den anderen Simulationen stellt sich aber auch hier ein stabiler Verlauf erst bei einem RMSD von 3 bis 4 Å zur Startstruktur ein.

Über die gesamte Simulationszeit betrachtet bewegen sich die Werte aller Simulationen meist zwischen 3 und 5 Å. Die - absolut gesehen - größten Abweichungen zur Startstruktur erhält man bei der Docked-Simulation, die mit 4.48 Å auch den größten durchschnittlichen RMSD-Wert zeigt. Die Standardabweichungen zeigen, dass die Apo- und die Docked-Simulation stabiler verlaufen als die ATP- und die Wildtyp-Simulation. Absolut gesehen sind diese Werte für die Simulation eines Proteins vergleichsweise hoch, weswegen das System in allen vier Fällen als verhältnismäßig flexibel anzusehen ist.

Mit Hilfe eines Per-Residue-RMSF-Plots kann analysiert werden, in welchen Proteinbereichen diese Flexibilität hauptsächlich auftritt (vgl. Abbildung 12.2).

Für die NBD erkennt man, dass die außen liegenden Subdomänen IB und IIB tendenziell beweglicher sind als die weiter innen befindlichen Subdomänen IA und IIA. Im

12. Ergebnisse: Molekulardynamische Simulationen

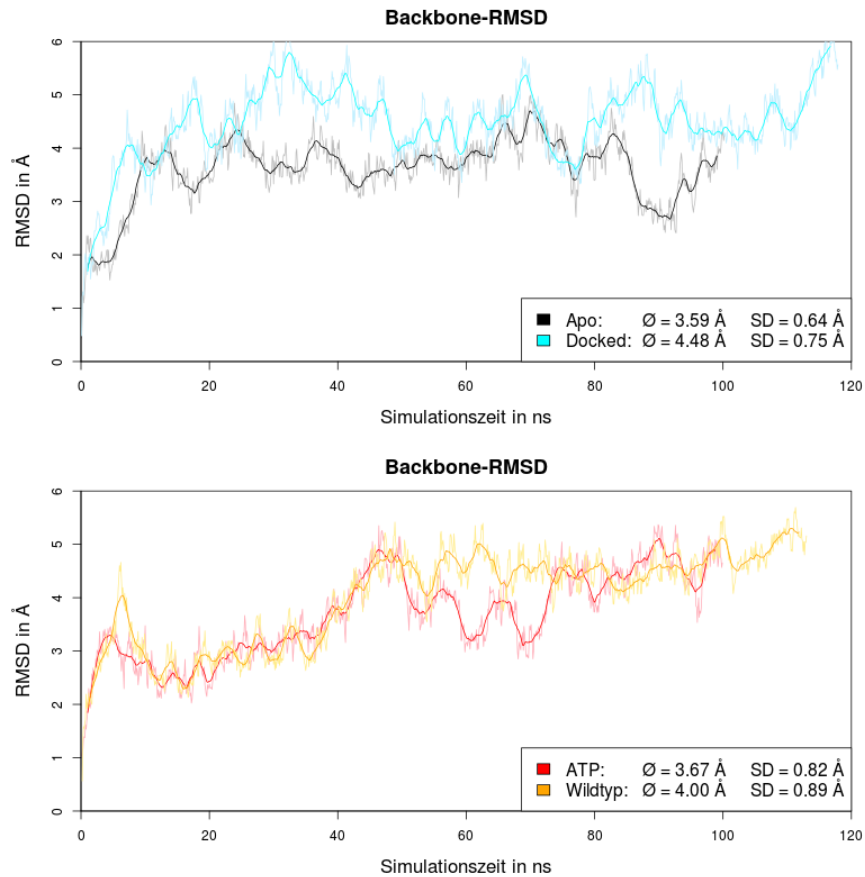


Abbildung 12.1.: RMSD-Verläufe der vier Simulationen. Die Referenzstruktur für die Überlagerung und die RMSD-Berechnung war die Startstruktur der jeweiligen Simulation.

Vergleich der vier Simulationen zeigt vor allem die ATP-Simulation (rote Linie) deutlich erhöhte RMSF-Werte in der NBD. Dieses eigentlich unerwartete Verhalten, welches durch das gebundene ATP-Molekül verursacht wird, wird in Kapitel 12.6 analysiert. Bei der Betrachtung des Linkerbereichs fällt auf, dass dieser in der Wildtyp-MD eine deutlich erhöhte und in der Docked-MD eine deutlich verringerte Beweglichkeit im Vergleich zu Apo-Trajektorie zeigt. Für die SBD lassen sich keine signifikanten Unterschiede zwischen den Simulationen erkennen.

Die berechneten RMSF-Werte stimmen größtenteils gut mit den experimentellen B-Faktoren (graue Linie) der *hsc_az*-Struktur überein.

Methode: 2D-RMSD-Berechnungen und Snapshot-Clustering Mit Hilfe von 2D-RMSD-Berechnungen kann man den kompletten Konformationsraum von MD-Trajektorien analysieren. Auch der Vergleich von Konformationen aus verschiedenen

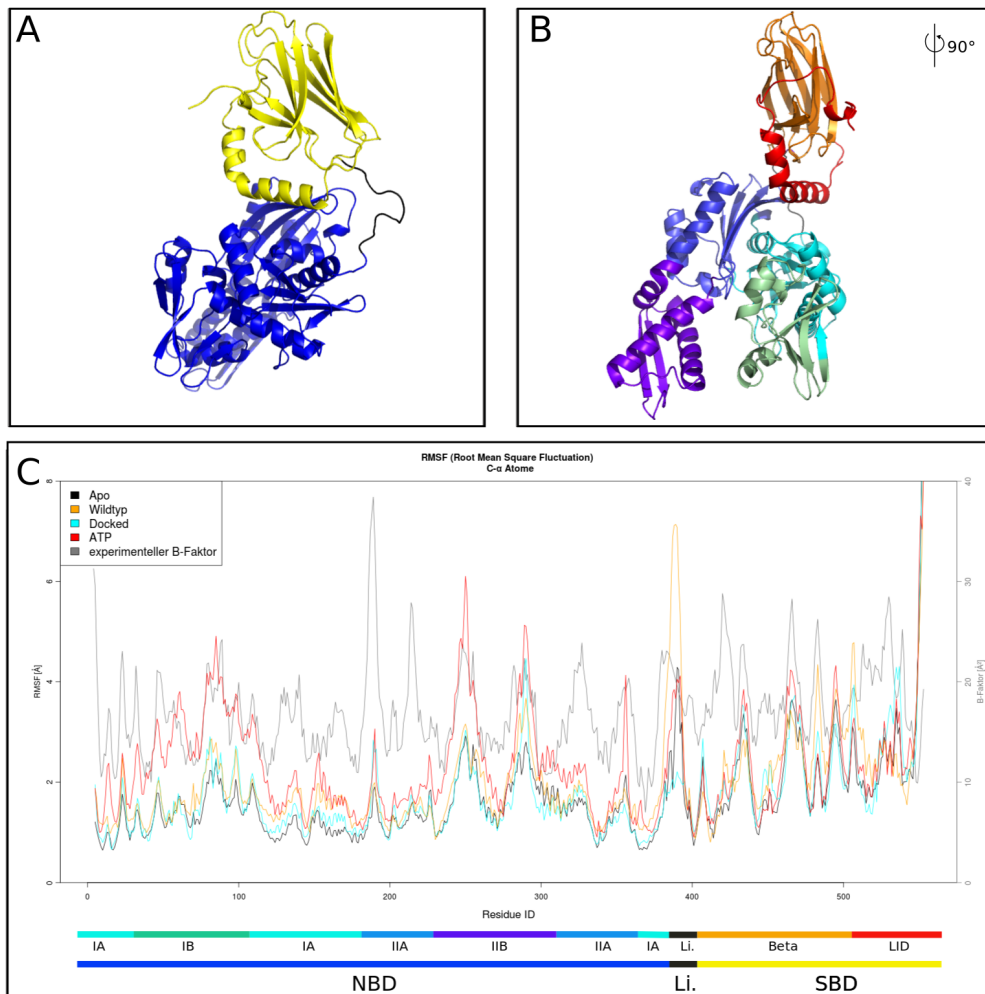


Abbildung 12.2.: (A) Übersicht über die MD-Startstruktur: NBD (blau), Linker (schwarz) und SBD (gelb); (B) Proteinstruktur in einer zu (A) um 90° gedrehten Ansicht, gefärbt nach Subdomänen: IA: cyan, AS 1–37, 112–184 und 363–380; IB: grün, AS 38–111; IIA: lila, AS 185–227 und 310–362; IIB: violett, AS 228–309; Linker: schwarz, AS 380–395; SBD β -Domäne: orange, AS 396–505; SBD LID-Domäne: rot, 506–554; (C) RMSF-Werte (Root Mean Square Fluctuation in Å) der C- α Atome der Simulationen nach Überlagerung auf die jeweilige Startstruktur. Die Balken im unteren Teil markieren den jeweiligen Protein-Bereich nach demselben Farbschema wie in (a) und (b).

Trajektorien ist möglich. Dazu werden paarweise RMSD-Werte zwischen allen Snapshots berechnet. Für jede Simulation werden dabei vier verschiedene Matrices berechnet, wobei die Snapshots jeweils auf die Startstruktur jeder Simulation überlagert werden. Es resultieren Matrices, die graphisch in Form eines 2D-RMSD-Plots dargestellt wer-

den können, wie er in Abbildung 12.3 gezeigt ist. Zusätzlich kann die Matrix auch als Grundlage für ein Clustering verwendet werden, mit dessen Hilfe die Snapshots in Gruppen mit ähnlichen Konformationen eingeteilt werden. Mit Hilfe des PAM-Algorithmus (vgl. Kapitel 4.3) lassen sich aus diesen Gruppen dann repräsentative Strukturen der jeweiligen Cluster ermitteln.

Zur Untersuchung des Konformationsraums des Gesamtsystems wurde das komplette Proteinrückgrat aller vier Trajektorien jeweils mit der Startstruktur jeder Simulation überlagert. Auf diese Weise wurden Translations- und Rotationsbewegungen, die für diese Analyse keine Rolle spielen, eliminiert. Die RMSD-Berechnungen wurden dann ebenfalls basierend auf den Atomen des Proteinrückgrats durchgeführt.

Die Aminosäuren 545 bis 554 wurden weder bei der Überlagerung noch bei der RMSD-Berechnung berücksichtigt. Es handelt sich hierbei um den Teil der Proteinkette, der aufgrund der trunkierten LID-Domäne nicht in der helikalen Tertiärstruktur vorliegt. Die Aminosäuren ragen dabei als ungefaltete Peptidkette aus der Substratbindetasche heraus, sind in allen vier Simulationen sehr flexibel und leisten keinen Beitrag zur Proteinkonformation. Sie wurden daher in keiner der nachfolgenden Analysen berücksichtigt.

Wie schon durch die Backbone-RMSD-Verläufe angedeutet, weisen alle vier Simulationen vor allem zu Anfang deutliche Konformationsänderungen auf. Im weiteren Verlauf beobachtet man bei der Apo- und bei der Docked-MD, dass die Konformation über einen Großteil der Simulationszeit erhalten bleibt. Bei der Wildtyp- und bei der ATP-Simulation werden hingegen im Verlauf der Simulation weitere Konformationen eingenommen.

Beim Vergleich der vier Simulationen fällt auf, dass das ATP-System die deutlichsten Unterschiede zu den anderen Trajektorien aufweist. Dies liegt am Einfluss, den das ATP-Molekül auf die NBD ausübt. Auch die Konformationen der Docked-MD sind über weite Teile der Simulation unterschiedlich zu denen der Apo- und der Wildtyp-Simulation.

Insgesamt sind die Unterschiede zwischen den Trajektorien mit RMSD-Werten von über 7 Å sehr groß. Die gefundenen Konformationsfamilien sind vergleichsweise inhomogen, was man daran erkennt, dass sie im 2D-RMSD-Plot nicht deutlich voneinander abgegrenzt sind. Dies ist damit zu begründen, dass durch die Betrachtung des Gesamtproteins viele unterschiedliche Effekte gleichzeitig erfasst werden und einander überlagern. Ein Clustering und eine visuelle Analyse der repräsentativen Strukturen (Daten nicht gezeigt) ergibt, dass ein großer Teil der Flexibilität auf die Bewegung der beiden großen Domänen in Relation zueinander zurückzuführen ist (vgl. Kapitel 12.2). Die Effekte, die beispielsweise durch Apoptozol in der Interface-Bindetasche oder die Punktmutationen bei der Wildtyp-MD ausgelöst werden, lassen sich deswegen nur bei isolierter Betrachtung der einzelnen Proteinbereiche analysieren.

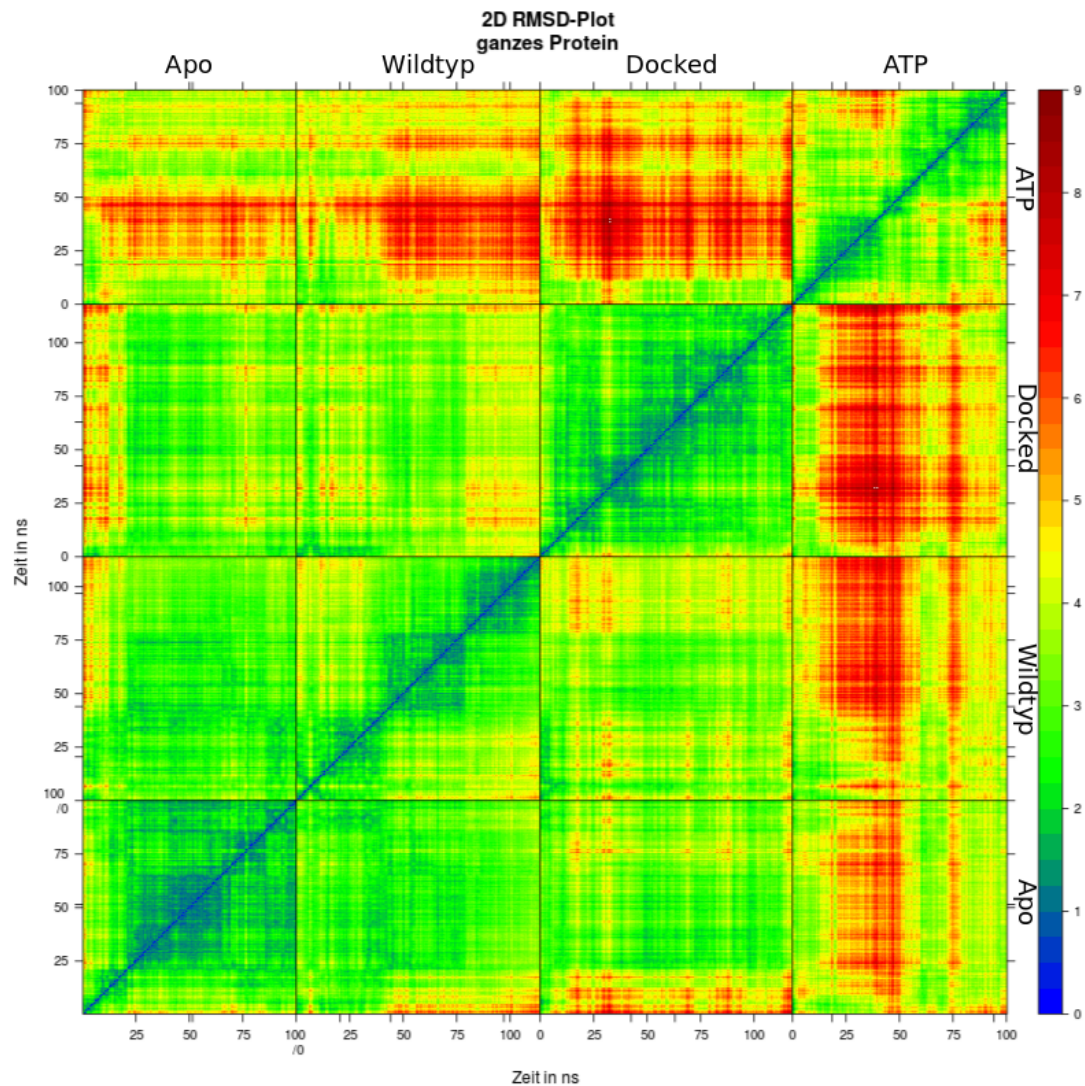


Abbildung 12.3.: 2D-RMSD-Plot für das Proteinrückgrat. Die Überlagerung und RMSD-Berechnung erfolgte für die Rückgrat-Atome der Aminosäuren 5 bis 544. Die verschiedenen Matrices sind in der horizontalen auf die Startstruktur der jeweiligen Referenz-Simulation überlagert (d.h. die vier unteren Matrices sind auf die Startstruktur der Apo-Simulation überlagert usw.). Obwohl der Plot qualitativ entlang der diagonalen gespiegelt ist, kommt es, aufgrund der unterschiedlichen Referenzstrukturen, zu leichten numerischen Unterschieden beim Vergleich von zwei korrespondierenden Matrices in der unteren und oberen Hälfte. Dies gilt ebenso für die im späteren Verlauf der Arbeit gezeigten 2D-RMSD-Plots.

12.2. Domänenbewegung

Die Dynamik des Gesamtsystems wird von der Bewegung von NBD und SBD relativ zueinander dominiert. Zur Untersuchung des Konformationsraumes dieser Domänenbewegung wurden 2D-RMSD-Berechnungen durchgeführt, wobei die Snapshots jeweils anhand einer Domäne überlagert und dann die RMSD-Werte der jeweils anderen Domäne berechnet wurden.

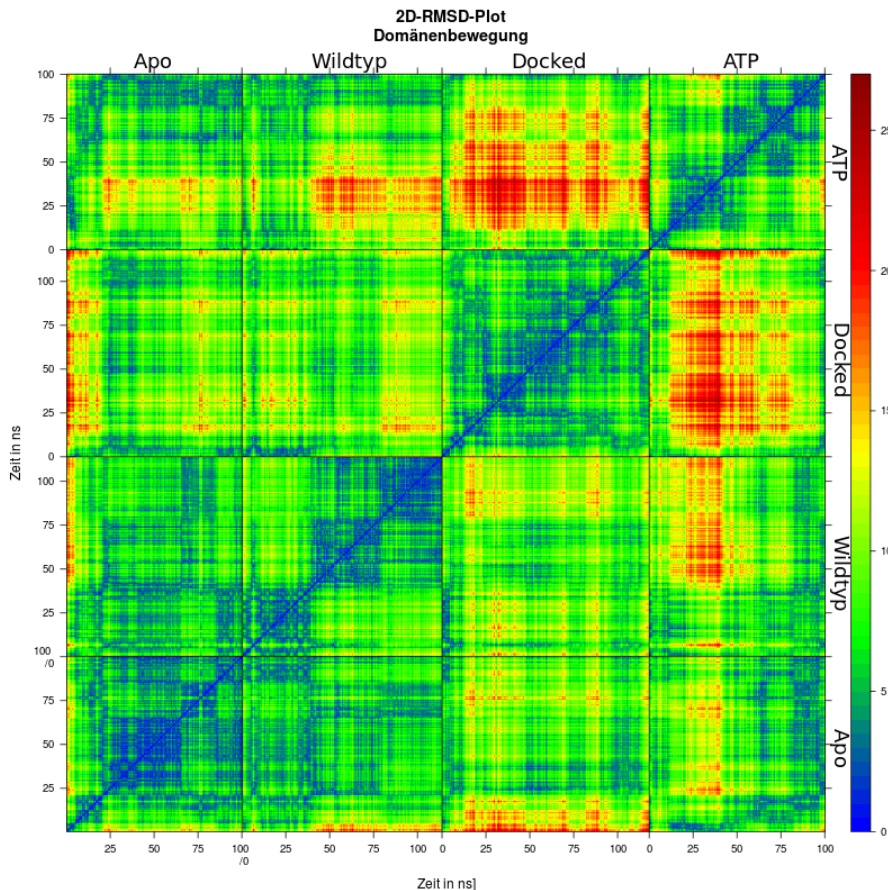


Abbildung 12.4.: 2D-RMSD-Plot für die Domänenbewegung von SBD und NBD. Alle Strukturen wurden auf die NBD überlagert. Die RMSD-Berechnung erfolgte für die SBD.

Durch dieses Vorgehen erhält man sehr große RMSD-Werte (von bis zu 20 Å), was die Relevanz dieser Domänenbewegung für das Gesamtsystem belegt. In Abbildung 12.4 ist der Plot für die Überlagerung auf die NBD und die RMSD-Berechnung der SBD gezeigt. Die Werte der SBD eignen sich besser für diese Analyse, weil die SBD eine geringere interne Flexibilität zeigt und die RMSD-Werte somit weniger stark durch andere Effekte überlagert werden (vgl. Kapitel 12.7). Der Vergleich der vier Simulationen zeigt für das

Apo- und das Wildtyp-System ähnlich Verläufe, wobei man bei beiden Systemen zwei bis drei Konformationsfamilien erkennen kann. Die Docked-Simulation zeigt hingegen insgesamt nur eine sehr moderate Flexibilität bezüglich der Domänenbewegung. Das im Interface gebundene Apoptozol hat hier also einen stabilisierenden Effekt. Im Gegensatz dazu wird die Flexibilität durch die Bindung von ATP an der NBD deutlich erhöht. Die ATP-Simulation zeigt, sowohl was die Flexibilität innerhalb der Trajektorie angeht, als auch im Vergleich zu den anderen Simulationen, die größten RMSD-Werte.

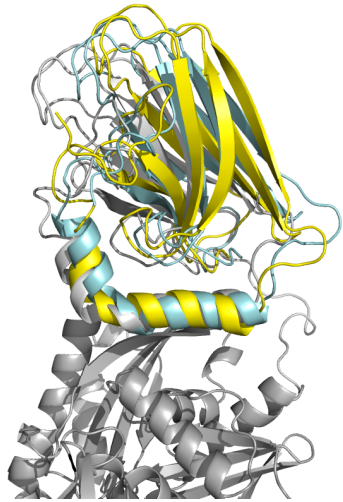
Um die Snapshots im Hinblick auf die SBD-Konformation zu gruppieren wurde ein hierarchisches Clustering der 2D-RMSD-Matrix durchgeführt. Bei einem Grenzwert von 14 Å erhält man daraus sechs Cluster. Aus jedem dieser Cluster wurde mit Hilfe des PAM-Algorithmus (vgl. Kapitel 4.3, Seite 57) eine repräsentative Struktur ausgewählt. Mit Hilfe dieser Strukturen lässt sich der Konformationsraum der SBD in Relation zur NBD gut visualisieren (Abbildung 12.5).

Abbildung 12.5 A zeigt zwei Snapshots aus Apo- (gelb) und Wildtyp-Simulation (türkis) im Vergleich zur Kristallstruktur *hsc.az* (grau). Die α -Domäne ist hier nach rechts verschoben, was sich vor allem in den peripheren Bereichen der β -Domäne auswirkt. Diese Konformationsfamilien bilden die beiden größten Cluster und machen anteilmäßig mehr als die Hälfte aller Snapshots aus. Das Clustering ergab außerdem eine weitere Konformationsfamilie mit einer ähnlichen Struktur, die aus Gründen der Übersichtlichkeit in der Abbildung nicht dargestellt ist. Die Konformationsfamilie kommt in allen vier Trajektorien vor.

Abbildung 12.5 B zeigt eine Verschiebung der SBD in dieselbe Richtung. Zusätzlich ist hier die Helix-Region aus der Ebene herausgedreht, was zu einer noch deutlicheren Abweichung im Vergleich zur Kristallstruktur führt. Diese Konformation kommt hauptsächlich in der Docked-Simulation vor und wird wesentlich durch die Anwesenheit von Apoptozol induziert und stabilisiert. Vereinzelt finden sich auch Vertreter in der Wildtyp- und der Apo-Trajektorie.

In den Abbildungen 12.5 C und D ist der α -Helix-Bereich hingegen in die andere Richtung verschoben (in der Abbildung nach links). In der Konformation des Snapshots Wildtyp-868 (Abbildung 12.5 C) ist außerdem der Winkel zwischen Helix-A und Helix-B deutlich flacher (vgl. dazu Kapitel 12.7). Die Auswirkungen auf die β -Domäne im Vergleich zur Kristallstruktur sind insgesamt aber gering. Die Konformation kommt in der Apo, Wildtyp und ATP-Simulation vor. Abbildung 12.5 D zeigt den Snapshot ATP-275, bei dem Helix-A aus der Ebene gedreht ist und so eine Verschiebung nach links und - in der Abbildung schlecht erkennbar - nach hinten bewirkt. Diese Konformation kommt ausschließlich im ersten Teil der ATP-Trajektorie vor. Ein Zusammenhang mit der geöffneten Nukleotid-Bindetasche ist deswegen naheliegend.

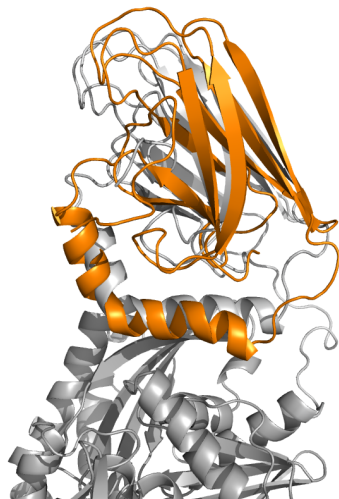
Die repräsentativen Strukturen zeigen, dass die Interface-Region und die relative Stellung der Helix-A zur NBD charakteristisch für die jeweiligen Konformationen sind. Dies wird auf Seiten der SBD sicherlicher dadurch unterstützt, dass sich die Bewegung der Helix-A aufgrund des in die Substratbindetasche gedockten Abschnitts der LID-Domäne direkt auf die SBD überträgt (vgl. Kapitel 12.7). Der Einfluss der Konformation in der



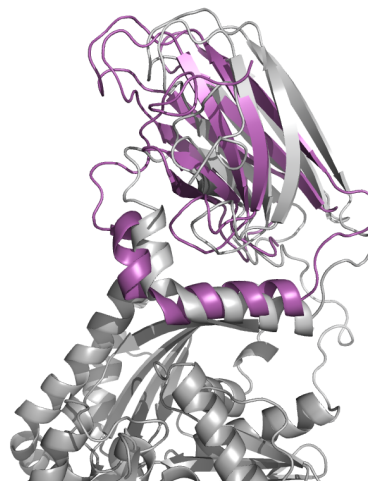
A: Apo-383 (gelb) und Wildtyp-308 (türkis)



B: Docked-267



C: Wildtyp-868



D: ATP-275

Abbildung 12.5.: Die Domänenbewegung der SBD relativ zur NBD kann in verschiedenen Richtungen erfolgen. Die Snapshots (farbig) sind jeweils anhand der NBD auf die Kristallstruktur (grau) überlagert. A: Apo-383 (gelb) und Wildtyp-308 (türkis). Ein dritter Cluster mit einer ähnlichen Struktur ist aus Gründen der Übersichtlichkeit nicht gezeigt. B: Docked-267 (blau) C: Wildtyp-868 (orange) D: ATP-275 (violett)

Interface-Region auf die Dynamik des Gesamtsystems stellt in jedem Fall ein Argument für die Adressierung der Interface-Bindetaste zur Inhibition von Hsp70 dar.

12.3. Dynamik der Interface-Bindetasche

Der Konformationsraum der Interface-Bindetasche wurde zunächst mittels eines 2D-RMSD-Plots visualisiert (Abbildung 12.6). Folgende Aminosäuren wurden dabei als Teil der Interface-Bindetasche behandelt: 146-155, 170-181, 215-222, 376-379 und 512-525. Die Überlagerung und RMSD-Berechnung erfolgten jeweils für die Atome des Proteinrückgrats.

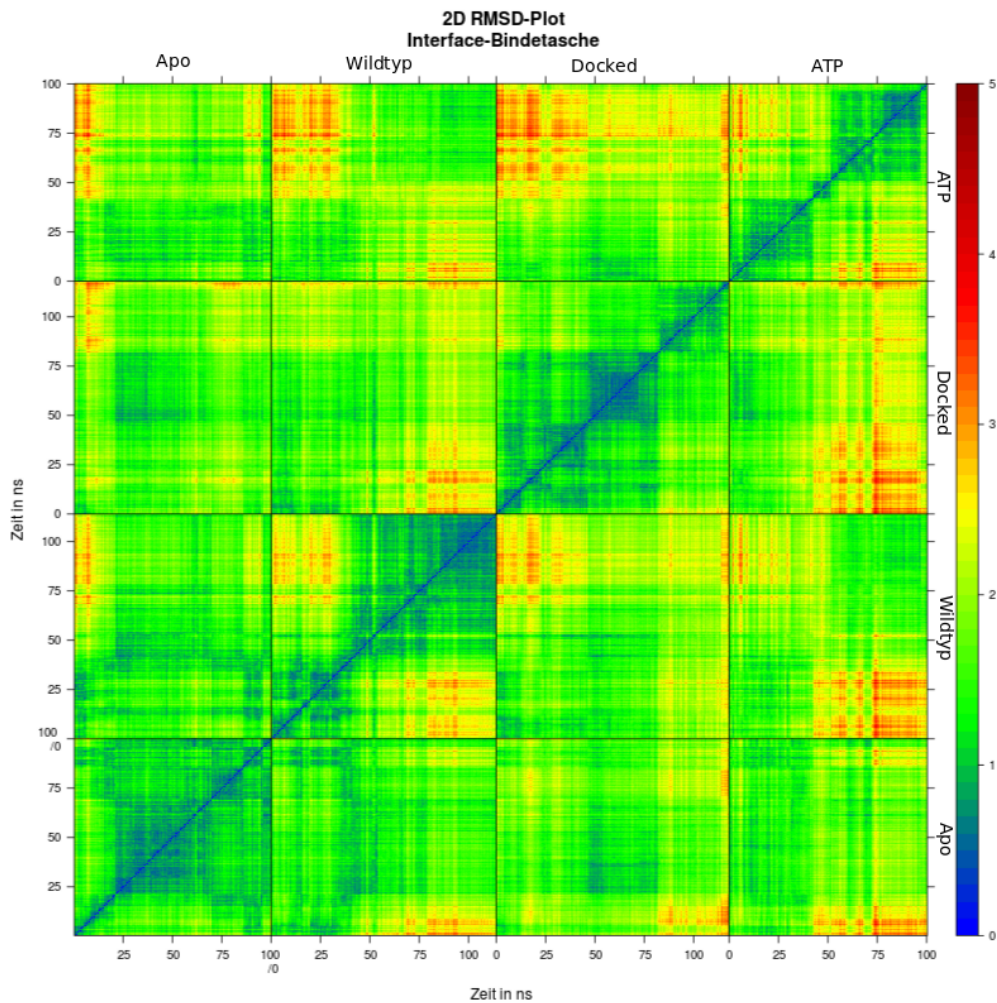


Abbildung 12.6.: 2D-RMSD Plot für die Interface-Region; Alignment und RMSD-Berechnung für die Backbone-Atome der Aminosäuren 146 bis 155, 170 bis 181, 215 bis 222, 376 bis 379, 512 bis 525.

Es zeigt sich in allen vier Systemen eine Flexibilität der Bindetasche. Die insgesamt stabilsten Konformationen beobachtet man in der Docked- und der Apo-Trajektorie (maximale RMSD-Werte innerhalb der Trajektorien: ca. 3 Å). Das Wildtyp- und das

ATP-System zeigen, mit RMSD-Werten bis zu 5 Å (innerhalb der Trajektorie), eine etwas größere Flexibilität.

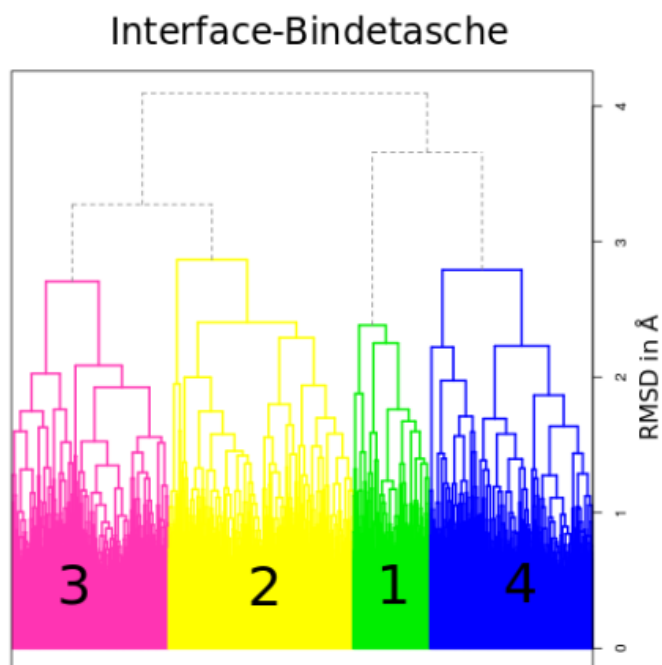


Abbildung 12.7.: Hierarchisches Clustering der Konformationen der Interface-Bindetasche aus allen vier Trajektorien. Bei einem Grenzwert von 3 Å erhält man 4 Konformationsfamilien, die farblich hervorgehoben sind.

Eine hierarchische Clusteranalyse aller Snapshots ergibt bei einem Grenzwert von 3 Å über alle Trajektorien vier Konformationsfamilien (Abbildung 12.7). Die mit dem PAM-Algorithmus berechneten repräsentativen Snapshots sind in Abbildung 12.8 gezeigt.

Der Hauptteil der Flexibilität entfällt auf die Helix-A (SBD) und deren relative Bewegung zu Asp152 bzw. der Helix um Asp152 (im Folgenden als 152-Helix bezeichnet) auf Seiten der NBD.

Der Snapshot Apo-52 (Cluster 1, grün) repräsentiert die kleinste Konformationsfamilie, die, über alle vier Trajektorien gesehen, in 570 Snapshots auftritt. Vertreter dieser Familie sind am Beginn aller Trajektorien zu finden. In der Docked- und der ATP-Simulation bleibt diese Startstruktur nur kurz erhalten (im 2D-RMSD-Plot nicht erkennbar). Die Konformation tritt vor allem in der Wildtyp-Simulation (während der ersten ca. 35 ns) und der Apo-Simulation (während der ersten 10-15 ns sowie ab etwa 90 ns) auf. Die Konformationsfamilie ist gekennzeichnet durch die Salzbrücke von Asp152 (NBD) mit Lys524, welche bei allen Vertretern zu beobachten ist. Dadurch wird

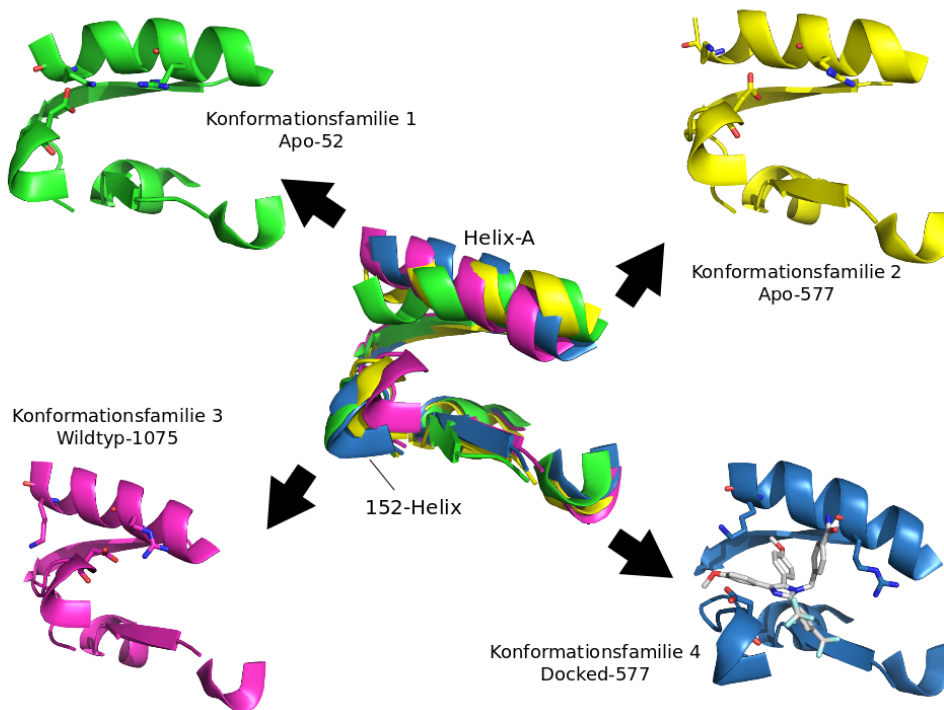


Abbildung 12.8.: Repräsentative Snapshots für die Konformationen der Interface-Bindetasche aus allen vier Trajektorien. Die Dynamik wird hauptsächlich durch die Stellung der 152-Helix(NBD) relativ zur Helix-A (SBD) dominiert.

die 152-Helix relativ zur Helix-A auf deren C-terminaler Seite fixiert. Strukturell kommt das der Kristallstruktur *hsc_az* sehr nahe.

Der Snapshot Apo-577 (Cluster 2, gelb) repräsentiert die größte Konformationsfamilie. Sie tritt über weite Strecken der Apo-Simulation (zwischen 15 ns und 90 ns Simulationszeit) sowie in den ersten 40 ns der ATP-Simulation auf. In der Wildtyp- und der Docked-Simulation kommt die Konformation nur unregelmäßig vor. Die Struktur zeichnet sich dadurch aus, dass die Salzbrücke zwischen Asp152 und Lys524 nicht mehr besteht, weil sich die 152-Helix relativ zur Helix-A verschoben hat. Asp152 ist dadurch nun auch in Wasserstoffbrücken-Reichweite zu Arg517. Vor allem in den Apo Snapshots zwischen 20 und 60 ns interagiert Asp152 stabil über Wasserstoffbrücken mit Arg517 (siehe dazu 12.10). Allerdings ist diese Interaktion nicht charakteristisch für alle Vertreter. Im letzten Drittel der Apo-MD sowie im ersten Teil der ATP-MD tritt beispielsweise auch die Salzbrücke Asp152-Lys524 wieder auf. Die Konformationsfamilie kann als Zwischenzustand gesehen werden, bei dem sowohl die Asp152-Lys524 als auch die Asp152-Arg517 Interaktion möglich ist.

Auch in Cluster 3 (repräsentiert durch Snapshot Wildtyp-1075) ist eine deutliche re-

lative Verschiebung der 152-Helix in Richtung der Mitte von Helix-A zu sehen. Diese ist noch stärker ausgeprägt als in Cluster 2 (vgl. Abbildung 12.8). Der Cluster rekrutiert sich aus Snapshots aus der zweiten Hälfte der Wildtyp- und der ATP-Trajektorien. Ein wesentliches Merkmal dieser Konformationsfamilie ist die Ausbildung einer Wasserstoffbrückeninteraktion zwischen Asp152 und Arg517, was durch die Verschiebung von Asp152 ermöglicht wird. Eine Salzbrücke zwischen Asp152 und Lys524 ist hingegen nicht möglich, da die Entfernung zwischen den Seitenketten zu groß ist.

Die Konformation aus Snapshot Docked-577 (blau) entspricht wieder eher der Kristallstruktur bzw. Konformationsfamilie 1 (grün). Der Cluster rekrutiert sich hauptsächlich aus der Docked-MD (vom Zeitpunkt 3 ns bis Ende). Im Unterschied zu Cluster 1 sind hier durch den Einfluss von Apoptozol die Backbone-Atome von Helix A und Asp152 deutlich weiter voneinander entfernt (vgl. Abbildung 12.8). Details zum Einfluss von Apoptozol auf die Konformation der Bindetasche werden in Kapitel 12.3.3 besprochen. Interessanterweise gibt es auch am Beginn der ATP-Simulation über einen Zeitraum von 7-8 ns einige Vertreter dieser Konformationsfamilie. Da diese aber im restlichen Verlauf der ATP-Simulation nicht mehr auftritt, scheint zur Stabilisierung über einen längeren Zeitraum ein Ligand wie Apoptozol nötig zu sein.

12.3.1. Interaktionsmuster von Asp152

Die repräsentativen Strukturen zeigen, dass die Dynamik der Interface-Bindetasche wesentlich durch das Interaktionsmuster von Asp152 beeinflusst wird. Die wichtigsten Interaktionspartner von Asp152 sind Arg155, Arg517, Gln520 und Lys524.

In Abbildung 12.9 sind die Positionen der fünf Seitenketten in der Kristallstruktur dargestellt. Charakteristisch für die Kristallstruktur und die Konformationsfamilie 1 ist die Salzbrücke zwischen Asp152 und Lys524. Die anderen Seitenketten sind zu Simulationsbeginn hingegen nicht in Richtung Asp152 orientiert.

Abbildung 12.10 zeigt die Abstandsverläufe zu den beiden SBD-Aminosäuren Lys524 (schwarz) und Arg517 (blau). Im Apo-System treten diese beiden Aminosäuren abwechselnd als Interaktionspartner von Asp152 in Erscheinung. Nur in zwei kurzen Zeiträumen bei 60 und bei 80 ns besteht keine der beiden Interaktionen über längere Zeit. Dies entspricht dem Übergang von Konformationsfamilie 1 zu Konformationsfamilie 2. Am Ende der Apo-Simulation vergrößert sich der Abstand zu Arg517 wieder sehr deutlich. Hier nimmt die Bindetasche wieder eine Kristallstruktur-ähnlich Konformation ein.

In der Wildtyp-Simulation bleibt die Salzbrücke zu Lys524 innerhalb der ersten 30 bis 40 ns zunächst erhalten. In diesem Zeitraum befindet sich die Interface-Region noch in einem Startstruktur-ähnlichen Zustand (Konformationsfamilie 1). Danach beobachtet man über einen Zeitraum von 40 ns sehr instabile Abstandsverläufe. In diesem Zeitraum fluktuiert die Bindetaschenkonformation zwischen den Konformationsfamilien 1, 2 und 3. Erst ab etwa 80 ns führt die Ausbildung der Interaktion zwischen Asp152-Arg517 zur stabilen Ausbildung der Bindetaschenkonformation 3 (Abbildung 12.8, magenta), die bis zum Ende der Simulation erhalten bleibt.

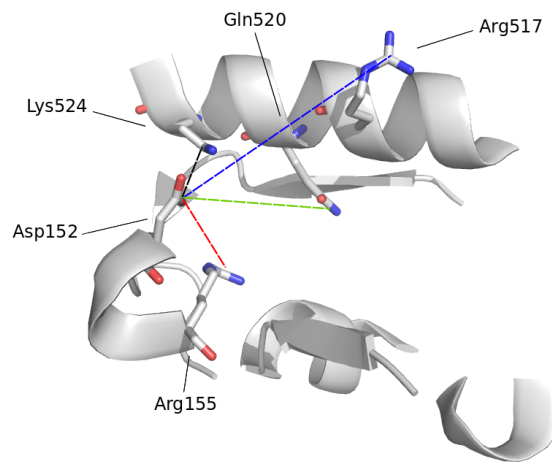


Abbildung 12.9.: Interface-Region der Kristallstruktur hsc_az mit den Aminosäuren Asp152, Arg155, Arg517, Gln520 und Lys524.

Einen qualitativ ähnlichen Verlauf zeigt das ATP-System. Auch hier bleibt zu Beginn der Simulation die Lys524-Salzbrücke bestehen, wobei der Abstandsverlauf jedoch nicht so stabil ist, wie beim Apo- oder beim Wildtyp-System. Möglicherweise spielt der Einfluss von ATP auf die Dynamik der NBD hier eine Rolle (vgl. Kapitel 12.6). Nach etwa 50 ns steigt der Abstand zu Lys524 dann stark an. Gegen Ende der Simulation bildet sich eine Interaktion mit Arg517 - und damit die Konformationsfamilie 3 - aus. Auch der 2D-RMSD-Plot (Abbildung 12.6) zeigt, dass das Wildtyp- und das ATP-System zum Ende der Simulation ähnliche Bindetaschen-Konformationen aufweisen.

In allen drei Trajektorien gibt es Zeiträume, in denen Asp152 weder mit Lys524 noch mit Arg517 interagiert. In Abbildung 12.11 sind daher zusätzlich die Abstandsverläufe zu Arg155 und Gln520 dargestellt.

In der ATP-Simulation werden die "Interaktionslücken" von Asp152 sehr gut durch diese beiden Aminosäuren ausgefüllt. Die Interaktion mit Arg155 (rot) erfolgt im Zeitraum zwischen 20 und 50 ns und ist dabei sogar stabiler als die Lys524 Salzbrücke. Zwischen 40 und 80 ns dominiert hingegen die Interaktion mit Gln520.

Die Wasserstoffbrücke mit Arg155 lässt sich auch über weite Teile der Apo-Simulation beobachten. Sie kann sowohl parallel zur Lys524-Salzbrücke als auch zur Arg517 Interaktion bestehen. Eine Hypothese ist, dass Arg155 die Asp152-Seitenkette in einer mittleren Position fixiert, aus der noch Interaktionen in beiden Richtungen möglich sind (Abbildung 12.8 Snapshot Apo-577; gelb). Findet eine stärkere Verschiebung in Richtung Arg517 statt (Cluster 3), spielt die Arg155-Interaktion hingegen keine Rolle mehr.

Während die Interaktionsmuster von Asp152 charakteristisch für die Konformation der Bindetasche sind, zeigt sich vor allem in der Wildtyp-Simulation, dass die Stabilität der Interfaceregion insgesamt davon nicht beeinflusst wird. Im Zeitraum zwischen 50

und 80 ns ist die Seitenkette von Asp152 in Richtung Solvens orientiert und wechselwirkt längere Zeit nicht mit der SBD. Die Konformation der Interfaceregion bleibt aufgrund der großen Kontaktfläche zwischen NBD und SBD auch in diesem Zeitraum dennoch erhalten.

Bei den bisherigen Analysen ausgenommen war die Docked-Simulation. Für dieses System zeigen die Abstandsverläufe in Abbildung 12.10 und Abbildung 12.11, dass ausschließlich die Lys524-Salzbrücke auftritt. Die anderen Interaktionen sind nicht möglich, weil sie durch Apoptozol verhindert werden (vgl. Abschnitt 12.3.3).

Die Relevanz der Asp152-Interaktionen wird auch durch Mutationsexperimente von Jiang et al. untermauert, die zeigen dass Asp152-Mutationen einen Einfluss auf funktionelle Parameter von HSP70 haben. Beispielsweise wird die Abbaurate von Clathrin-Oligomeren, der Einfluss der ATP-Bindung auf die Fluoreszenz von Trp90 und die Unterdrückung der Linker-Proteolyse durch ATP durch die Mutation D152K vermindert [105]. Bei verschiedenen Asp152-Mutationen wurde außerdem eine erhöhte ATPase-Aktivität und Auxillin-Bindung festgestellt. Diese Ergebnisse weisen darauf hin, dass die allosterische Kommunikation zwischen NBD und SBD, welche zur Regulierung dieser Prozesse notwendig ist, durch Asp152 beeinflusst wird.

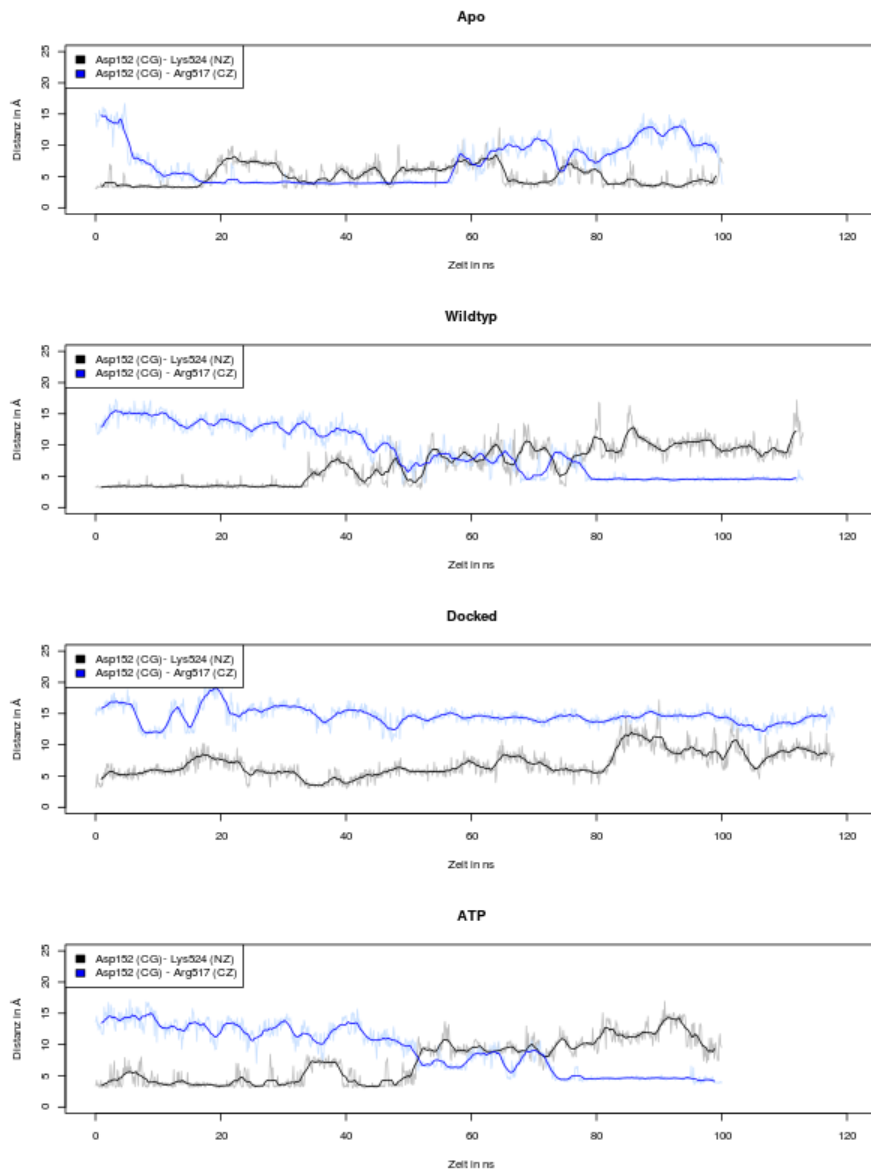


Abbildung 12.10.: Abstandsverläufe für die Interaktionen von Asp152 mit Lys524 (schwarz) und Arg517 (blau). Der Abstand wurde zwischen dem CG-Atom von Asp152 und den terminalen N-Atomen von Lys524 bzw. Arg517 gemessen.

12. Ergebnisse: Molekulardynamische Simulationen

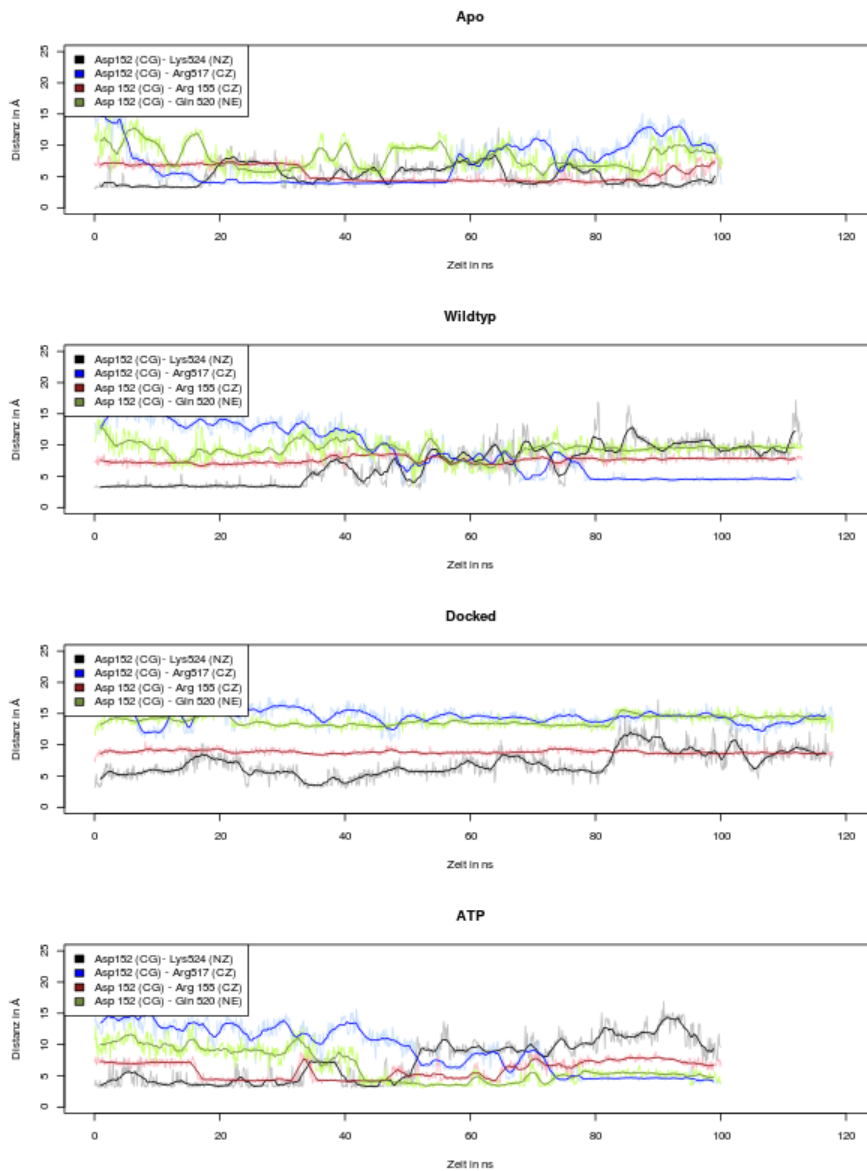


Abbildung 12.11.: Abstandsverläufe für die Interaktionen von Asp152 mit Arg155 (rot) und Gln520 (grün). Zum besseren Vergleich sind außerdem die aus Abbildung 12.10 bekannten Abstände zu Arg517 (blau) und Lys524 (schwarz) eingezeichnet.

12.3.2. Kontakte zwischen NBD-Subdomäne II und Helix-B

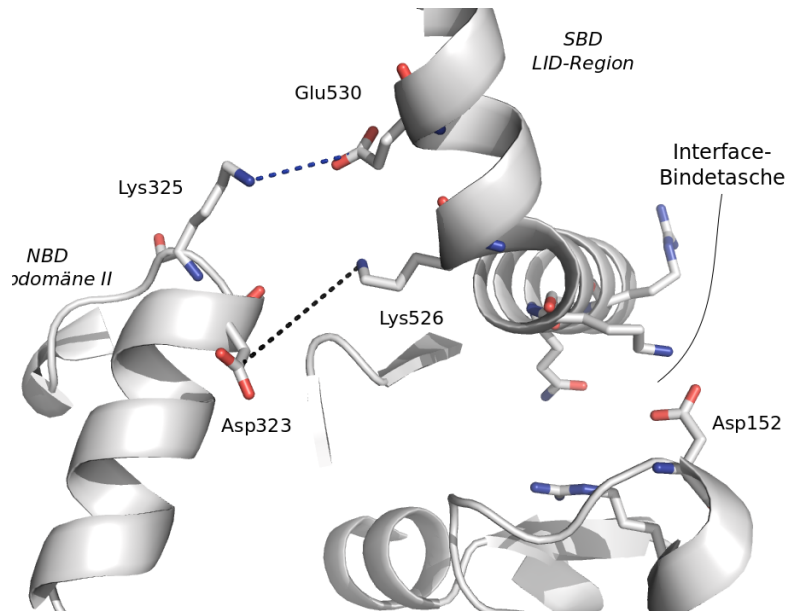


Abbildung 12.12.: Der nach oben “abgeknickte” Abschnitt der LID-Region steht durch zwei Salzbrücken zwischen Asp323 und Lys526 sowie zwischen Lys325 und Glu530 in Kontakt mit der NBD-Subdomäne II.

Ein weiteres Interaktionspaar, welches laut Jiang et al. einen Einfluss auf die allosterische Kommunikation hat, ist Lys325-Glu530 [105]. Lys325 befindet sich in der IIA-Subdomäne der NBD und Glu530 im “aufsteigenden” Helix B Bereich der SBD (Abbildung 12.12), so dass hier eine direkte Interaktion zwischen NBD und LID-Bereich besteht, dem im funktionellen (nicht trunkierten) Protein ein Einfluss auf die Substrat-Affinität zugeschrieben wird [43].

In Abbildung 12.13 sind die Abstände zwischen Lys325 und Glu530 sowie zwischen Asp323 und Lys526, einem weiteren Interaktionspaar zwischen NBD und dem “aufsteigenden” Teil der LID-Domäne, dargestellt. Für die Interaktion Lys325-Glu530 (blaue Linie) fällt zunächst auf, dass der Abstand meist stark fluktuiert. Im Apo-System kann man jedoch für den Großteil der Trajektorie von einer ionischen Wechselwirkung ausgehen, da der Abstand durchschnittlich unter 5 Å liegt und hier insgesamt den stabilsten Verlauf zeigt. Eine Zusammenhang mit den Konformationsfamilien der Bindetasche der Apo-MD lässt sich nicht ableiten.

In der Wildtyp-Trajektorie ist der durchschnittliche Abstand hingegen deutlich größer. Vor allem im zweiten Teil der Simulation (ab 40 ns) tritt die Wechselwirkung praktisch nicht mehr über einen längeren Zeitraum auf. Der Abstand liegt im Mittel bei etwa 10 Å. Die Interaktion wird aufgrund der stark in Richtung N-Terminus von Helix-A verschobenen NBD (Abbildung 12.8: Konformationsfamilie 3) behindert.

12. Ergebnisse: Molekulardynamische Simulationen

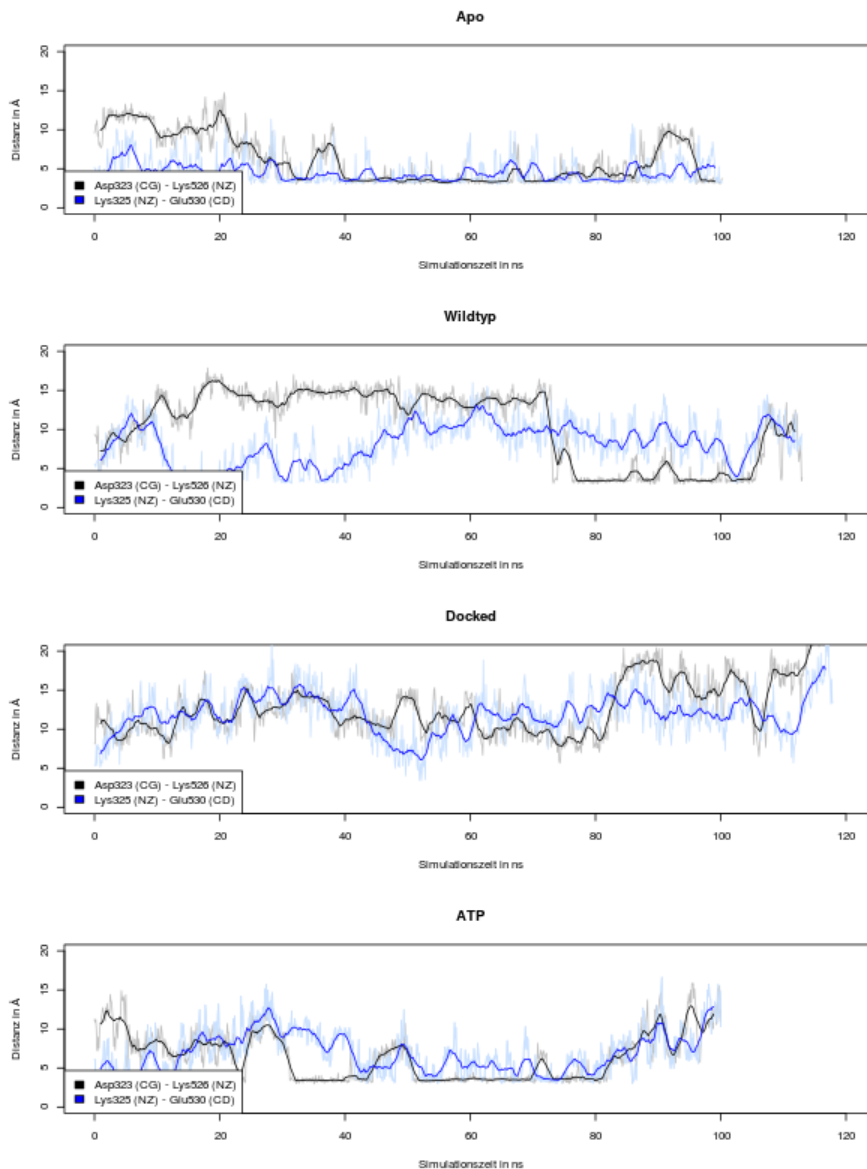


Abbildung 12.13.: Abstände für die Salzbrücken zwischen Asp323 und Lys526 sowie zwischen Lys325 und Glu530, die das Interface zwischen NBD Subdomäne IIA und SBD Helix-B stabilisieren.

Auch bei der ATP-Simulation zeigt sich eine große Fluktuation des Abstands Lys325-Glu530. Durchschnittlich ist er zwar geringer als in der Wildtyp-Trajektore. Eine stabile Interaktion wie im Apo-System kommt jedoch nicht zustande.

Die zweite in Abbildung 12.13 eingezeichnete Interaktion ist die Salzbrücke zwischen Asp323 und Lys526, die sich ebenfalls in Subdomäne IIA bzw. Helix-B befinden. Im

Apo-, Wildtyp- und ATP-System sieht man jeweils über längere Zeiträume eine starke ionische Wechselwirkung mit einem Abstand von etwa 4 Å. Diese treten bevorzugt dann auf, wenn die 152-Helix zur Mitte von Helix A verschoben ist, also in den Konformationsfamilien 2 und 3. Allerdings ist auch diese Wechselwirkung nicht obligatorisch, wie man beispielsweise bei der Wildtyp-Simulation zwischen 40 und 60 ns sieht. Die Interaktion ist somit ebenfalls eher an der Stabilisierung als an der Induktion der Konformation beteiligt.

An den Abstandsverläufen der Docked-Trajektorie erkennt man, dass durch die von Apoptozol induzierte Interface-Konformationen der NBD-SBD-Kontakt in diesem Bereich ebenfalls eingeschränkt wird.

Es ist möglich, dass der Einfluss dieser beiden Salzbrücken *in vivo* eine noch größere Auswirkung auf die Dynamik des Proteins hat, weil sie die Bewegung der LID-Domäne weg von der SBD- β -Domäne unterstützen können. Aufgrund der trunkeierten LID-Domäne ist das in diesen Simulationen nicht zu beobachten (vgl. Kapitel 12.7).

12.3.3. Dynamik von Apoptozol und dessen Einfluss auf die Konformation der Bindetasche

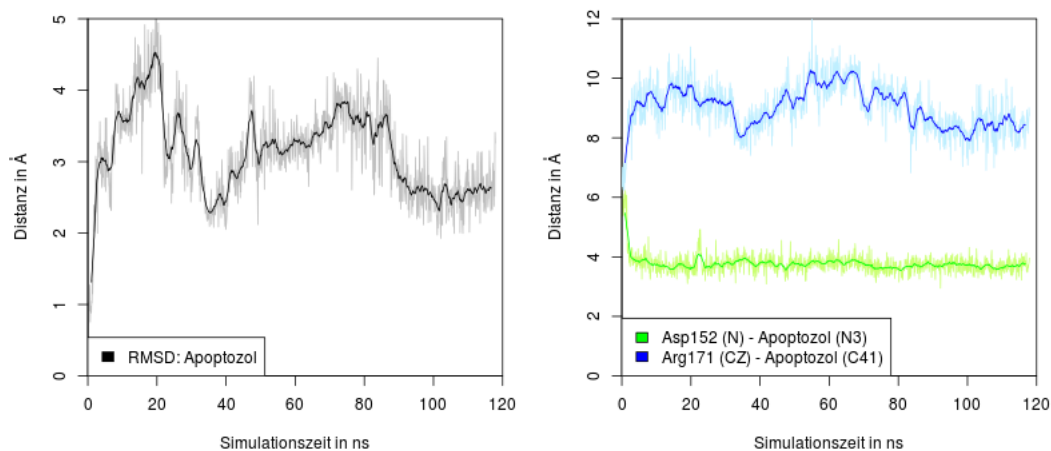
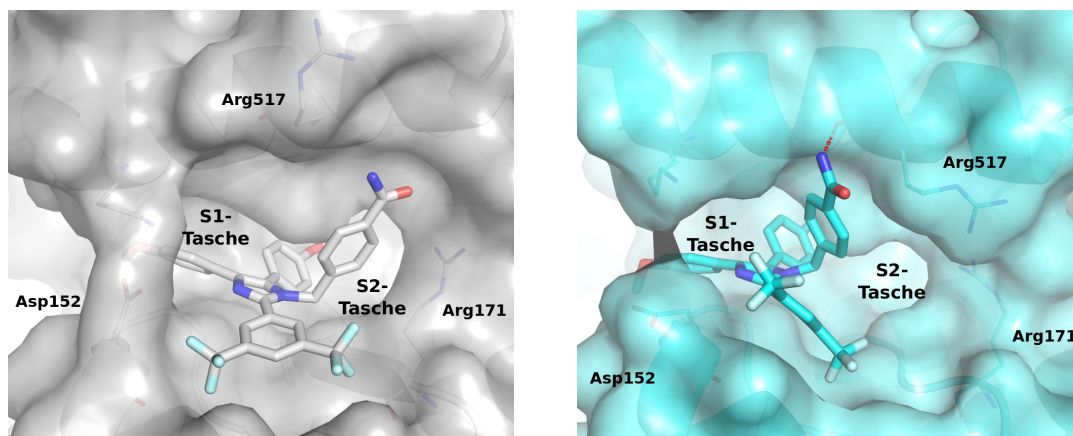


Abbildung 12.14.: Links: RMSD-Entwicklung von Apoptozol (berechnet für alle Schweratome) in Relation zur Startstruktur nach Überlagerung auf die Bindetasche. Rechts: Abstandsentwicklung von Apoptozol zu den Aminosäuren Asp152 und Arg171.

Das Apoptozol-Molekül zeigt gleich zu Beginn der Simulationen eine deutliche Anpassung des Bindemodus, die mit einer Bewegung innerhalb der Bindetasche einhergeht (Abbildung 12.14). Dabei schiebt sich Apoptozol in Richtung der C-terminalen Endes

von Helix-A. Die Abstandsverläufe von Apoptozol zu den Aminosäuren Arg171 und Asp152 verdeutlichen dies. Schon nach wenigen ns nimmt der Abstand zu Asp152 deutlich ab, während der zu Arg171 größer wird. Die Abstände bleiben dann im Verlauf der Simulation auf ihrem jeweiligen Niveau.



A. MD-Startstruktur

B. Docked-383

Abbildung 12.15.: Apoptozol verschiebt sich gleich zu Beginn der Docked-Simulation in Richtung des C-terminalen Endes der Helix-A. Die Abbildung zeigt, wie sich dabei der S1-Bereich der Bindetasche weitet und die Salzbrücke zwischen Asp152 und Lys524 deutlich abgeschwächt wird. Durch das "Herunterklappen" der Seitenkette von Arg517 wird die Bindetasche im Bereich der S2-Tasche hingegen verengt und nicht mehr von Apoptozol adressiert.

Um die Bewegung von Apoptozol visualisieren zu können, wurden alle Snapshots anhand der Interface-Bindetasche überlagert. Danach wurden die RMSD-Werte des Apoptozol-Moleküls berechnet, die bei diesem Vorgehen die relative Bewegung von Apoptozol in der Bindetasche widerspiegeln. Abbildung 12.15 zeigt die repräsentative Struktur des größten Clusters (ausgewählt mit Hilfe des PAM-Algorithmus) im Vergleich zur Startstruktur der Simulation.

Die beiden Strukturen zeigen deutlich, wie sich Apoptozol zwischen Asp152 und Helix A bewegt und somit SBD und NBD auseinander schiebt. Dies führt dazu, dass sich die Bindetasche im Bereich der S1-Subtasche im Vergleich zur Kristallstruktur deutlich vergrößert (Abbildung 12.15).

Auf der N-terminalen Seite von Helix-A klappt - ebenfalls zu einem sehr frühen Zeitpunkt der Simulation - die Seitenkette von Arg517 nach unten und verengt somit die Subtasche S2. Diese Subtasche wird im Verlauf der Simulation daher nicht von Apoptozol adressiert.

Die Arg517-Seitenkette besetzt vorrangig der Bereich den Bindetasche, der in der Docking-Pose vom *para*-Carboxamid-Substituenten von Apoptozol adressiert wird. Dieser bewegt sich in der Simulation ebenfalls in Richtung C-Terminus von Helix-A.

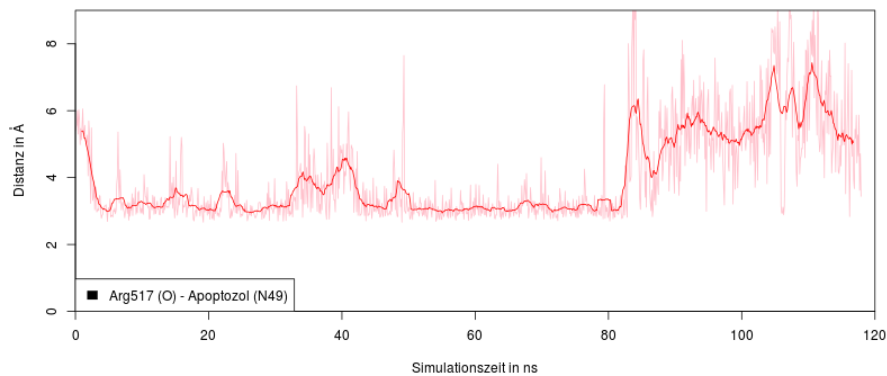


Abbildung 12.16.: Die Carboxamid-Gruppe von Apoptozol interagiert in der Simulation über einen langen Zeitraum mit dem Sauerstoffatom von Arg517. Die Abbildung zeigt den Abstand zwischen dem Carboxamidstickstoff und dem Sauerstoffatom des Proteinerückgrats im Verlauf der Simulation. Innerhalb der ersten 80 ns beträgt der Abstand durchschnittlich 3.3 Å.

Während der ersten 80 ns der Simulation befindet sich der Amidstickstoff in einem Abstand von 2.7 bis 3.2 Å zum Carbonyl-Sauerstoffatom von Arg517 (Abbildung 12.16). Diese Wasserstoffbrückeninteraktion stellt die einzige gerichtete Wechselwirkung dar, die in der Simulation über einen längeren Zeitraum auftritt.

Das Fehlen von gerichteten Wechselwirkungen hat auch zur Folge, dass Apoptozol im gesamten Simulationsverlauf eine eingeschränkte Flexibilität beibehält (vgl. RMSD-Verlauf Abbildung 12.14). Der Bindemodus bleibt aber im Verlauf der Simulation insgesamt erhalten, wie man an der Überlagerung der repräsentativen Strukturen der beiden größten Cluster des PAM-Clusterings gut erkennen kann (Abbildung 12.17).

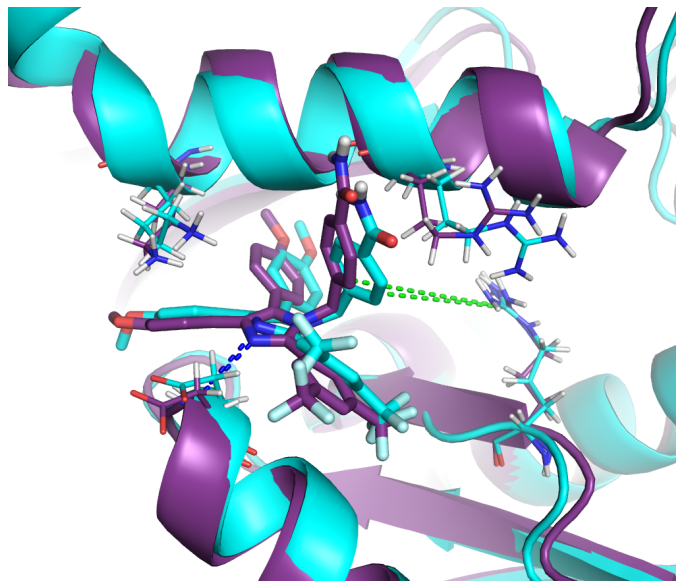


Abbildung 12.17.: Die Überlagerung der beiden Snapshots Docked-383 und Docked-513 zeigt, dass sich der Bindemodus von Apoptozol nach der initialen Konformationsanpassung kaum verändert. Abweichungen (Abbildung 12.14) kommen durch Rotation der Substituenten zustande. Die beiden gezeigten Strukturen sind die repräsentativen Strukturen der beiden größten Konformationsfamilien von Apoptozol, die mit Hilfe eine PAM-Clusterings nach Überlagerung auf die Bindetasche ausgewählt wurden.

12.4. Mutations-Region

Die Kristallstruktur *hsc_az* weist im Vergleich zum Wildtyp von bHsp70 zwei Punktmutationen der Aminosäuren Glu213 und Asp214 auf. Durch die Mutation zu Alanin wird die Kristallisation der Proteins ermöglicht. Da sich beide Aminosäuren im Interface zwischen NBD und SBD befinden, gibt es Bedenken, dass die zusätzlichen hydrophoben Kontakte die Tertiärstruktur des Proteins kritisch beeinflussen und es sich bei der Zwei-Domänen-Struktur mit geschlossenem Interface somit um ein "Artefakt" handelt, welches *in vivo* nicht vorkommt [43].

Abbildung 12.18 zeigt, dass sich die beiden mutierten Reste auf der Rückseite der Interface-Bindetasche befinden und tatsächlich direkt in die Interaktion zwischen NBD und SBD involviert sind. Die Hauptinteraktionspartner auf Seiten der SBD sind Arg416 und Asn417. Auf Seiten der NBD ist neben den mutierten Aminosäuren die Loopregion um Ala191 im weiteren Simulationsverlauf am Domänenkontakt beteiligt (Abbildung 12.18).

Das Interaktionsmuster in der Kristallstruktur ist hauptsächlich durch hydrophobe Kontakte charakterisiert, in die die beiden mutierten Aminosäuren direkt involviert sind

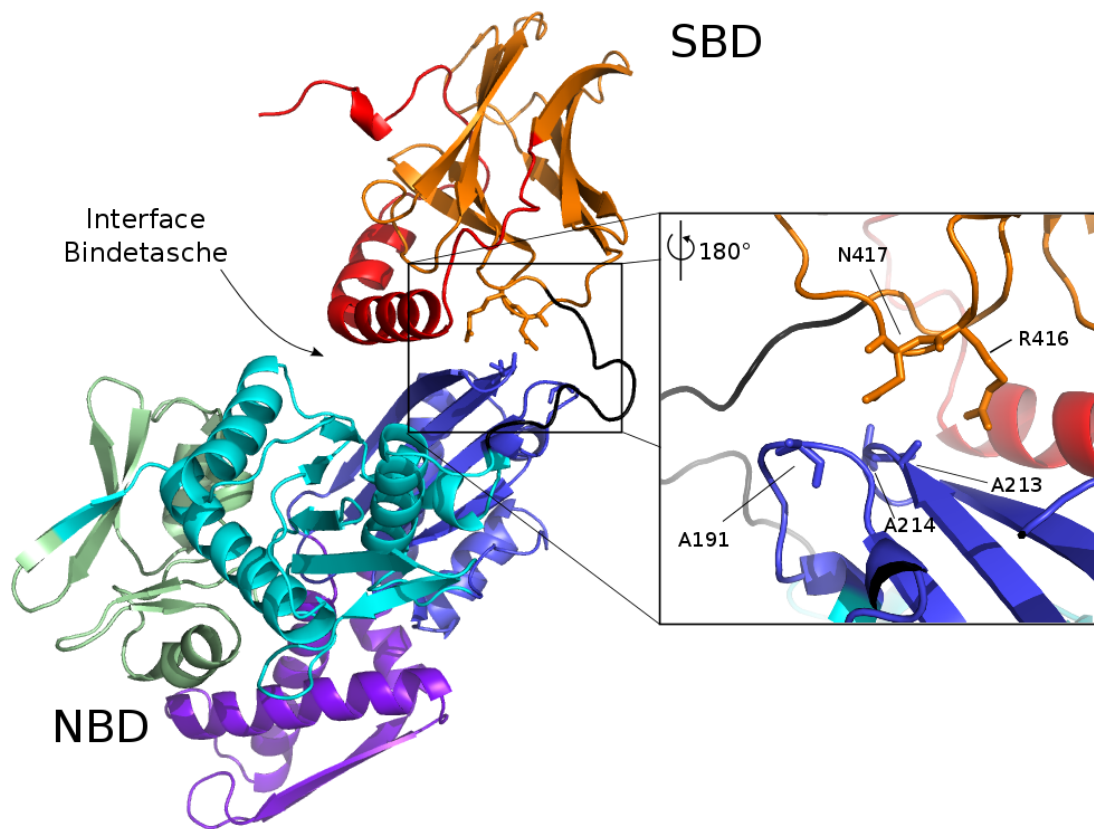


Abbildung 12.18.: Die mutierten Aminosäuren 213 und 214 sind direkt am Kontakt zwischen NBD und SBD beteiligt. Die Region um die mutierten Aminosäuren befindet sich von der Interface-Bindetasche aus gesehen auf der Rückseite des Proteins.

(Abbildung 12.19). Während der Equilibrierung und in den ersten Nanosekunden der Apo-Simulation bildet sich eine Wasserstoffbrücke zwischen dem Stickstoffatom von Ala214 und dem Seitenketten-Sauerstoff von Asn417 (Abbildung 12.19 B). Bei der Simulation des rückmutierten Wildtyps (Wildtyp-MD) ist dies aufgrund der größeren Seitenketten von Glu213 und Asp214 nicht der Fall.

In Abbildung 12.20 sind die Abstandsverläufe der wichtigsten Interaktionen der Apo- und der Wildtyp-Simulation dargestellt. Im Apo-System bleibt die Wasserstoffbrücke zwischen Ala214 und Asn417 (blaue Linie) nur in den ersten 10 ns der Simulation bestehen. Danach wechselt Asn417 seinen Interaktionspartner und bildet eine Wasserstoffbrücke mit dem Backbone-Sauerstoff von Ala191 aus. Diese bleibt über einen sehr langen Zeitraum stabil. Der repräsentative Snapshot dieser Konformation (Abbildung 12.20, gelb) zeigt, wie die beiden mutierten Reste hier überbrückt werden. Die Konformation

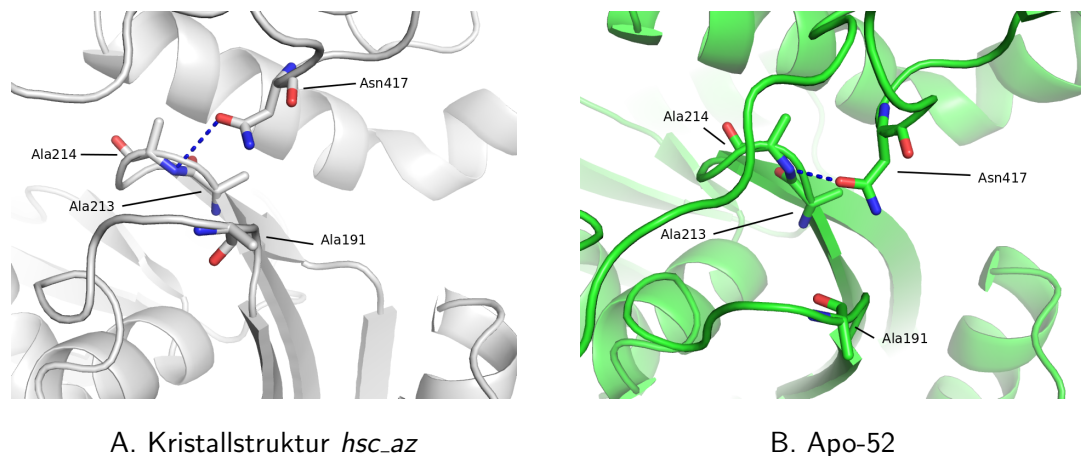


Abbildung 12.19.: A.: In der Kristallstruktur *hsc_az* stehen die mutierten Aminosäuren in direktem (hydrophobe Kontakt) mit der SBD. Eine Wasserstoffbrücken-Interaktion ist aus geometrischen Gründen nicht möglich. B.: Nach 5.2 ns Simulationszeit hat sich eine Wasserstoffbrücke (blaue Linie) zwischen Ala214 und Asn417 gebildet.

des Proteinrückgrats und somit die Gesamtkonformation in der Region wird davon aber nur geringfügig beeinflusst (vgl. Abbildung 12.19). Zum Ende der Apo-Simulation bei etwa 85 ns verliert Asn417 seinen Kontakt mit der NBD und orientiert sich in Richtung Arg416. Allerdings stabilisiert sich der Abstand bei knapp unter 4 Å und ist somit geringfügig größer als der Wasserstoffbrückenabstand. Das Proteinrückgrat behält aber auch in dieser Phase seine Konformation bei, und es kommt zu keiner Destabilisierung der Kontaktfläche. Der hydrophobe Kontakt der beiden mutierten Aminosäuren bleibt im gesamten Simulationsverlauf bestehen.

In der Wildtyp-Simulation bildet sich hingegen ein anderes Interaktionsmuster aus. Zu Beginn der Simulation werden NBD und SBD durch die größeren Seitenketten der rückmutierten Aminosäuren zunächst auseinander geschoben, der Kontakt zwischen NBD und SBD bleibt jedoch zu jedem Zeitpunkt bestehen.

Erst nach 40 ns bildet sich eine Wasserstoffbrücke zwischen dem Backbone-Sauerstoff von Glu213 und dem Seitenketten-Stickstoff von Asn417 (Abbildung 12.20 rechts: blaue Linie), die bis zum Ende der Simulation erhalten bleibt. Der Abstand des Säureamid-Stickstoffs zum zentralen C-Atom der Carbonsäuregruppe von Asp214 bleibt ebenfalls über weite Strecken der Simulation bei unter 4 Å, wodurch die Konformation weiter stabilisiert wird. Eine Interaktion mit Ala191 lässt sich in der Wildtyp-MD hingegen zu keinem Zeitpunkt beobachten.

Insgesamt bleibt die Mutationsregion also sowohl in der Apo- als auch in der Wildtyp-Simulation sehr stabil. Das NBD-SBD-Interface wird hauptsächlich durch hydrophobe Kontakte stabilisiert. In den Simulationen bilden sich im Bereich der mutierten Aminosäuren zusätzlich Wasserstoffbrücken, welche die Konformation in der Mu-

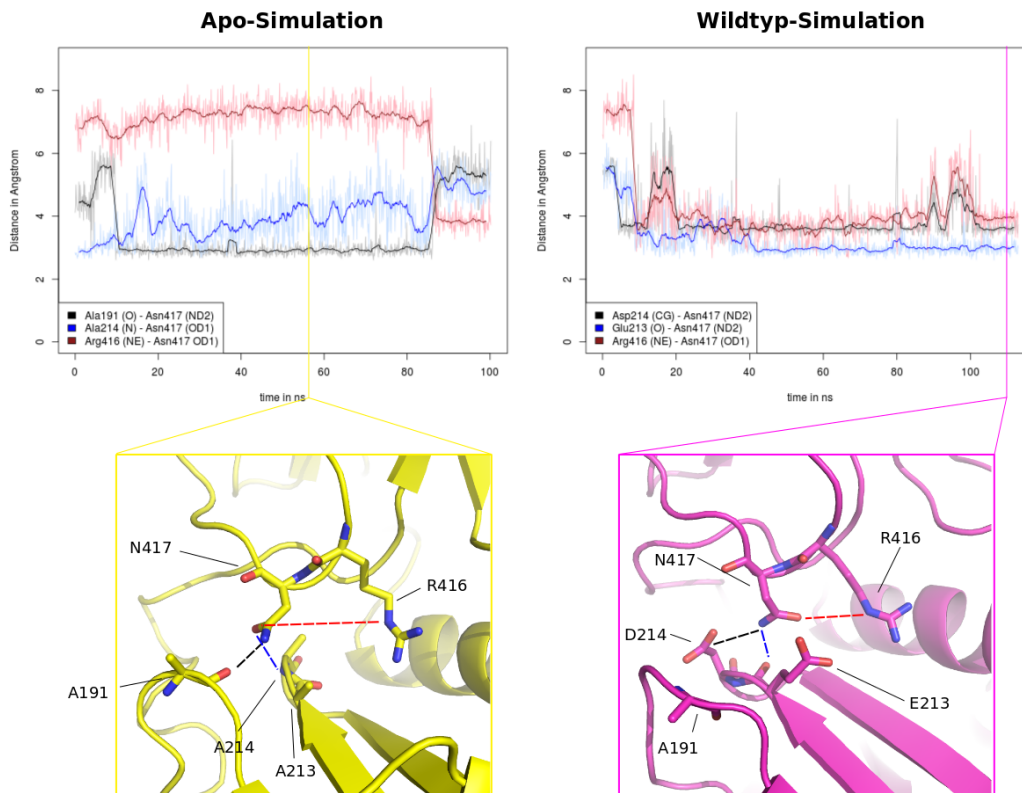


Abbildung 12.20.: Die Interaktionsmuster in der Mutationsregion unterschieden sich zwischen Apo- und Wildtyp-MD deutlich. Dargestellt sind für beide Simulationen je drei wichtige Interaktionen. In der Legende sind in Klammern die Atome, die für die Abstandmessung verwendet wurden, angegeben. Im unteren Teil ist eine repräsentative Struktur für beide Simulationen dargestellt (gelb: Apo-577, magenta: Wildtyp-1075).

tationsregion jedoch nicht kritisch beeinflussen und somit eher einen unterstützenden Charakter haben.

Die Einführung der beiden polaren Seitenketten im Wildtyp-System beeinflusst zwar das Muster der Wasserstoffbrückeninteraktionen, hat aber keinen Einfluss auf die Gesamtkonformation der NBD-SBD-Kontaktfläche. Ein Zusammenhang zwischen den Interaktionen in der Mutationsregion und der Domänenbewegung (vgl. Kapitel 12.2) kann ebenfalls nicht abgeleitet werden.

12.5. Linker

Zu Beginn des Ergebnisteils wurden die Per-Residue-RMSF-Werte aller vier Simulationen gezeigt (Abbildung 12.2 auf Seite 161). Dort war zu sehen, dass der Linker in der Wildtyp-Simulation eine deutlich höhere Flexibilität aufweist.

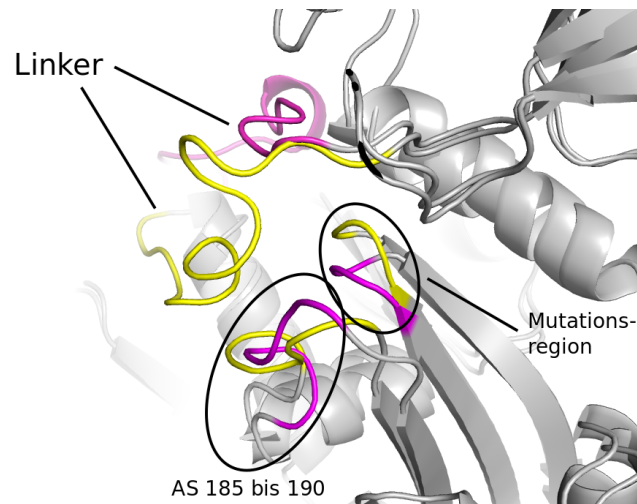


Abbildung 12.21.: In der Apo-Simulation (Apo-577, gelb) steht der Linker häufig in hydrophobem Kontakt mit der NBD. In der Wildtyp-MD (Wildtyp-1075, magenta) ist dies deutlich seltener der Fall, was zu einer erhöhten Flexibilität des Linkers führt.

Eine Analyse der Linker-Dynamik zeigt, dass dieser in der Wildtyp-Simulation deutlich seltener mit der NBD (hauptsächlich mit den Aminosäuren 185-190) interagiert (Abbildung 12.21) und deswegen diese erhöhte Flexibilität aufweist.

Die Region um die mutierten Aminosäuren steht räumlich in sehr engem Kontakt mit dem Bereich, in dem der Linker in die SBD übergeht. Es ist deswegen denkbar, dass die Mutation der beiden polaren Aminosäuren (Glu213 und Asp214) zu Alanin zu verbesserten hydrophoben Kontakten führt und somit der ebenfalls hydrophobe Linker in diesem Bereich stabilisiert wird (Abbildung 12.21).

Die Stabilisierung des Linkers durch die Mutationen könnte auch eine Erklärung für die bessere Kristallisierbarkeit der mutierten Form sein.

12.6. Nukleotid-Bindedomäne

Die NBD besteht aus zwei großen Subdomänen I und II, die ihrerseits nochmals in zwei Subdomänen (a und b) aufgeteilt werden. Die Dynamik der NBD wird durch die relative Bewegung der Domänen I und II zueinander dominiert, die ein Öffnen bzw. ein Schließen der Nukleotid-Bindetasche bewirkt.

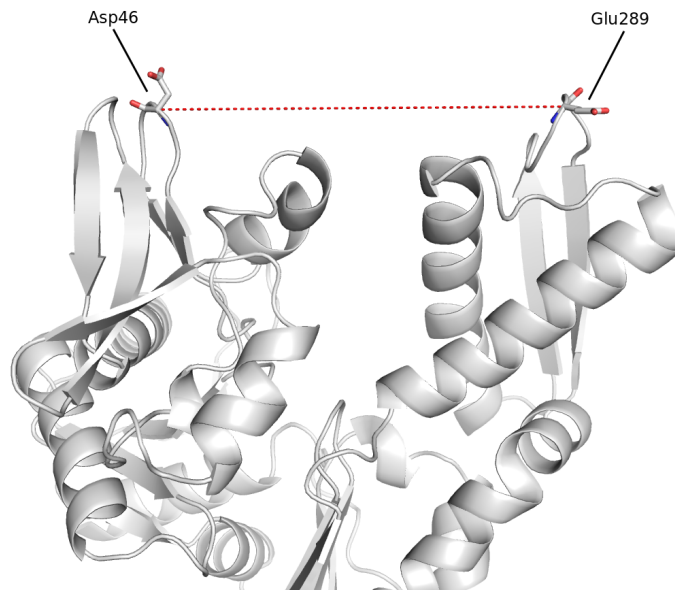


Abbildung 12.22.: Die Distanz zwischen den C-alpha Atomen von Asp46 und Glu289 ist ein Indikator für den Öffnungszustand der NBD. Hier ist der Zustand in der Kristallstruktur *hsc_az* gezeigt.

Ein Indikator für den Öffnungszustand ist der Abstand zwischen den Aminosäuren Asp46 und Glu289, die sich jeweils an den äußeren Enden der beiden NBD-Subdomänen befinden (12.22). Zu Beginn der Simulationen beträgt der Abstand etwa 32 Å, was auf den geöffneten Zustand hindeutet. Bei Werten unter 27 Å ist die Bindetasche hingegen geschlossen. Abbildung 12.23 zeigt den Abstand zwischen den $C\alpha$ -Atomen der beiden Aminosäuren im Verlauf der Simulationen.

Sowohl die Apo- als auch die Wildtyp-Trajektorie entwickeln sich in Richtung des geschlossenen Zustandes. Nach etwa 20 ns werden in beiden Trajektorien Abstände von unter 27 Å erreicht. Im weiteren Verlauf der Simulationen bewegt sich der Abstand zwischen 25 und 27 Å. Das Docked-System verhält sich zu Beginn sehr ähnlich und erreicht ebenfalls nach etwa 20 ns den geschlossenen Zustand. Im Unterschied zu den anderen beiden Simulationen beobachtet man jedoch im weiteren Verlauf wieder eine Öffnung der NBD. Nach etwa 60 ns Simulationszeit steigen die Abstandswerte wieder auf über 30 Å an, so dass ab diesem Zeitpunkt bis zum Ende der Simulation wieder der geöffnete NBD-Zustand vorliegt. Dieses Verhalten ist vor allem interessant, weil es ein Hinweis auf einen allosteren Einfluss der Interface-Region auf die Dynamik der NBD darstellt. Obwohl die zugrunde liegenden molekularen Mechanismen noch weiter untersucht werden müssen, kann ein solcher allosterer Effekt möglicherweise als Grundlage zur Erklärung eines Inhibitionsmechanismus dienen.

In Abbildung 12.23 ist auch der Abstandsverlauf der ATP-Simulation gezeigt, der

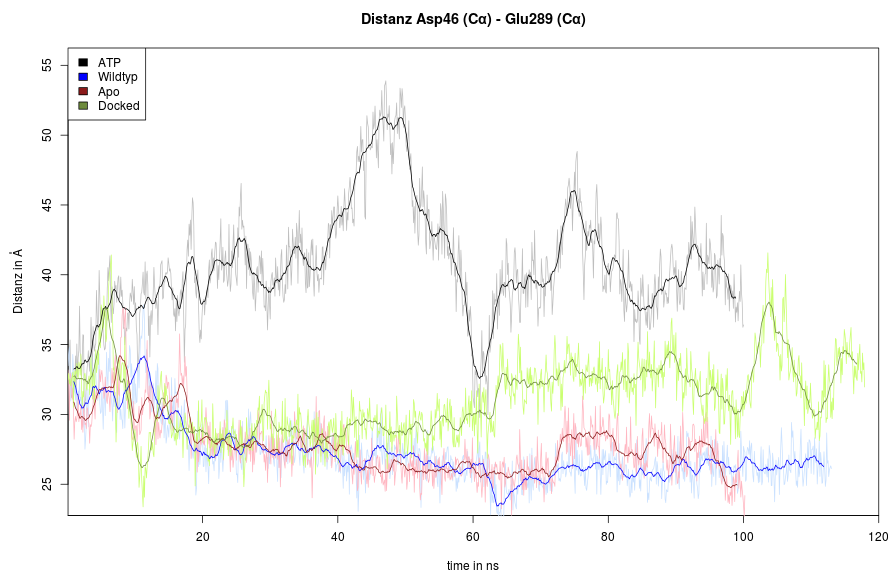


Abbildung 12.23.: Die vier Simulation zeigen deutlich unterschiedliche Verläufe des Asp46-Glu289-Abstandes. Während sich im Apo- und Wildtyp-System jeweils der geschlossene NBD-Zustand entwickelt, erkennt man bei der Docked-Simulation, dass der geöffnete Zustand der Kristallstruktur über weite Strecken der Simulation erhalten bleibt. Im ATP-System ergibt sich - entgegen der Erwartungen - eine deutliche Öffnung der Nukleotid-Bindetasche.

sich sehr deutlich von den andern Simulationen unterscheidet. Man beobachtet hier eine Öffnung der ATP-Bindetasche, die umgehend nach Simulationbeginn eintritt. Durchschnittlich bewegt sich der Asp46-Glu289-Abstand im Bereich von 40 Å. Im Zeitraum zwischen 40 und 60 ns fluktuiert der Abstand sehr stark und erreicht einem Maximalwert von über 50 Å, was mit einer sehr weiten Öffnung der Nukleotidbindetasche einhergeht. Ebenfalls in diesem Zeitraum fällt der Abstand auch zum einzigen Mal seit Beginn der Simulation unter 35 Å. Beide Zustände können jedoch nicht über einen längeren Zeitraum stabilisiert werden. Dieses Verhalten widerspricht den eigentlichen Erwartungen, dass die Bindung von ATP eine geschlossene NBD-Konformation induziert. Eine Analyse der Dynamik des ATP-Moleküls und dessen Einfluss auf die NBD-Konformation folgt in Kapitel 12.6.1.

2D-RMSD-Berechnungen und repräsentative Konformationen Auch im 2D-RMSD-Plot ist der unterschiedliche Konformationsraum der NBD in der ATP-MD im Vergleich zu den anderen Simulationen deutlich zu erkennen (Abbildung 12.24). Die Anpassungsphase bei der Wildtyp- und der Apo-Simulation zu Simulationsbeginn und die Stabilisierung der Konformation zwischen 20 und 100 ns wird ebenfalls deutlich.

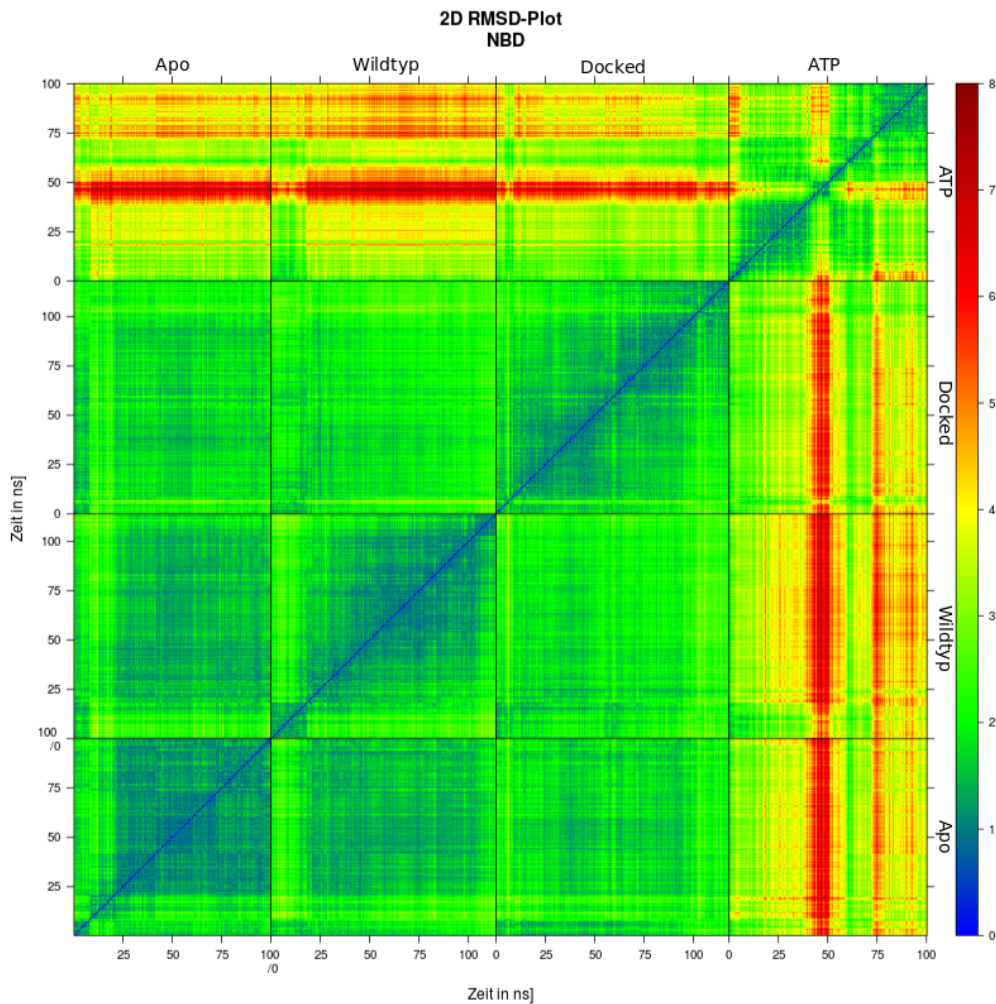


Abbildung 12.24.: Der 2D-RMSD-Plot der NBD zeigt vor allem bei der ATP-MD einen signifikant unterschiedlichen Verlauf. Überlagerung und RMSD-Berechnung für die Aminosäuren 5 bis 380.

Für das Docked-System zeigt der 2D-RMSD-Plot vor allem zu Beginn der Simulation Flexibilität. Das Schließen und Öffnen der Nukleotidbindetasche, welches der Asp46-Glu289-Abstandsverlauf gezeigt hat, lässt sich hingegen hier nicht erkennen. In Abbildung 12.25 sind die repräsentativen Strukturen der Konformationsfamilien gezeigt, die man aus einem hierarchischen Clustering mit einem Grenzwert von 4 Å erhält. Abbildung 12.25a zeigt den Vergleich zwischen der Kristallstruktur *hsc_az* und dem Snapshot Wildtyp-133 (violett). Der Cluster setzt sich aus den Anfangsnapshots der vier Trajektorien zusammen und ist deswegen der Kristallstruktur sehr ähnlich (Backbone-RMSD: 0.93 Å). In der Wildtyp-MD bleibt diese Konformation etwa 20 ns lang stabil. Abbildung 12.25 b zeigt zwei Snapshots aus der Apo-Simulation im Vergleich zu dieser Struktur.

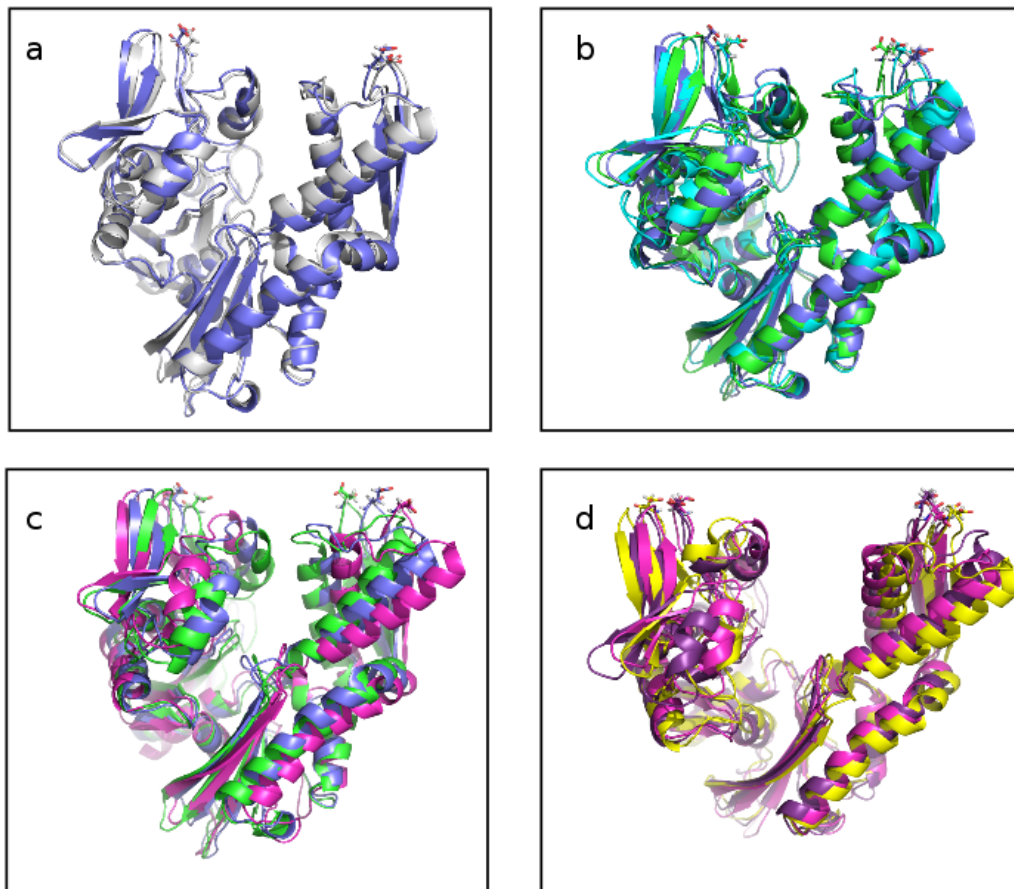


Abbildung 12.25.: Repräsentative Strukturen für Konformationsfamilien der NBD: (a) Grau: Kristallstruktur *hsc_az*, Lila: Wildtyp-133; (b) Lila: Wildtyp-133, Grün: Apo-441, Türkis: Apo-136; (c) Lila: Wildtyp-133, Grün: Apo-441, Magenta: ATP-316; (d) Magenta: ATP-316, Gelb: ATP-450, Violett: ATP-802. In allen Screenshots sind außerdem die Aminosäuren Asp46 und Glu289 eingezeichnet, deren Abstand als Indikator für den Öffnungszustand der Domänen dient.

Beide weisen eine geschlossener Nukleotid-Bindetasche auf. Der in türkis dargestellte Snapshot Apo-136 repräsentiert einen sehr kleinen Cluster, der nur über 10 ns stabil ist und ausschließlich in der Apo-MD vorkommt. Die grüne Struktur repräsentiert dagegen den mit 2936 Snapshots größten Cluster und stellt die Konformation dar, die über weite Teil der Apo-, Wildtyp- und Docked-MD auftritt.

Die Öffnung der NBD in der ATP-Simulation ist in Teilabbildung 12.25 c durch den Vergleich der Snapshots ATP-316 (magenta) mit der geschlossenen Struktur (grün) und

der MD-Startstruktur (violett) dargestellt. Im Verlauf der ATP-Simulation werden noch zwei weitere Konformationsfamilien gefunden, die im Vergleich zu ATP-316 noch weiter geöffnete NBD-Bindetaschen aufweisen. Der Snapshot ATP-450 repräsentiert dabei die am weitesten geöffnete Konformation (Abbildung 12.25 d, gelb).

12.6.1. Dynamisches Verhalten und Einfluss des ATP-Moleküls

Abbildung 12.26 A zeigt den Bindemodus von ATP nach 0.9 ns Simulationszeit. Die Adenosin-Substruktur interagiert dabei ausschließlich mit der Subdomäne II (im linken Teil der Abbildung), während der Triphosphat-Rest über Wasserstoffbrücken mit Thr14 und Lys71 der Subdomäne I wechselwirkt. Außerdem erfolgt durch die Phosphatgruppen eine dreifache Koordinierung des Co-Faktors Mg^{2+} , welcher seinerseits auch von Asp199 komplexiert wird und so eine weitere Verbindung zur Subdomäne II herstellt. Die Konformation bleibt in dieser Form jedoch nur in den ersten 3 ns der Simulation erhalten.

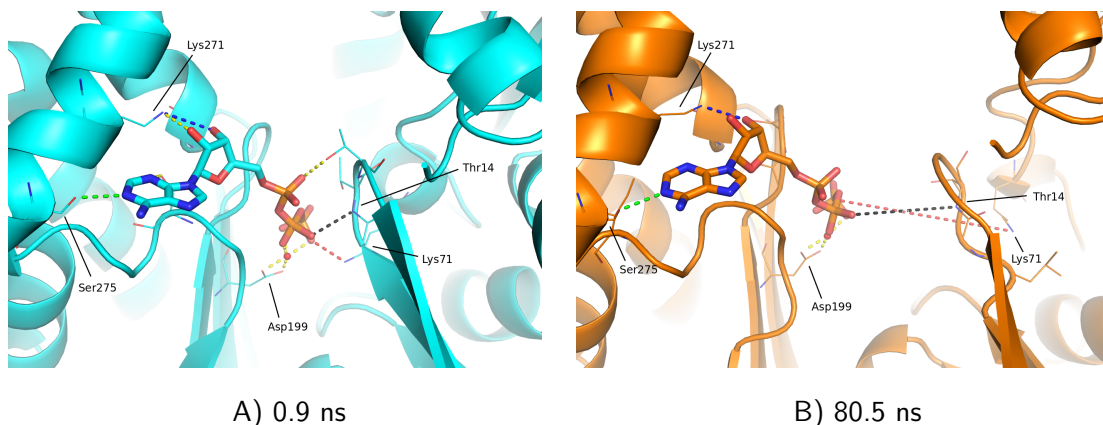


Abbildung 12.26.: Nucleotid-Bindetasche mit gebundenem ATP zum Zeitpunkt A) 0.9 ns und B) 80.5 ns. Potentielle Wasserstoffbrücken sind als gestrichelte Linien eingezeichnet. Der zeitliche Verlauf einiger Abstände (grün, blau, rot, schwarz) ist in Abbildung 12.27 dargestellt.

Die beiden NBD-Subdomänen bewegen sich - wie oben beschrieben - im Verlauf der Simulation voneinander weg. Das ATP-Molekül verliert dabei den Kontakt zur Subdomäne I vollständig, bleibt aber im kompletten Simulationsverlauf fest an Subdomäne II gebunden (Abbildung 12.26 B). Die Abstandsverläufe zwischen ATP und den Aminosäuren Lys71 und Thr14 (beide Subdomäne I) sowie Lys271 und Ser275 (Subdomäne II) (Abbildung 12.27) zeigen dies deutlich.

Vor allem die Interaktionen zwischen ATP und Subdomäne I sind in der Simulation also nicht stark genug, um die halb geöffnete NBD-Konformation zu stabilisieren bzw. den geschlossenen Zustand zu induzieren. Aufgrund von experimentellen Daten

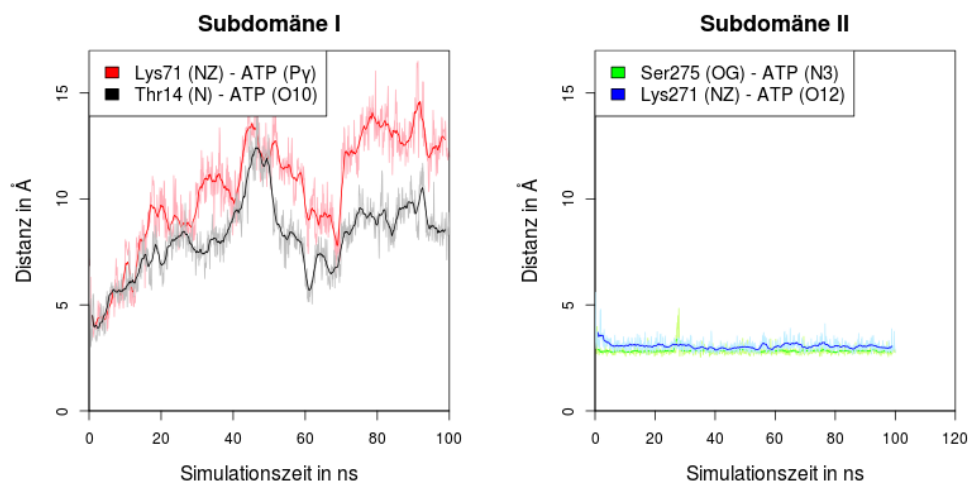


Abbildung 12.27.: Distanz zwischen ATP und ausgewählten Wechselwirkungspartnern in Subdomäne I und II.

wäre dies eigentlich zu erwarten gewesen. Für DnaK ist bekannt, dass die Nukleotid-Bindetasche im ATP-Zustand geschlossen und im ADP Zustand geöffnet ist [43]. In der Kristallstruktur 1KAX, die die Vorlage für die hier verwendete ATP-Pose lieferte, liegt das Protein ebenfalls in der geschlossenen NBD-Konformation vor (der Asp46-Glu289-Abstand beträgt dort 27 Å).

Dass der Übergang zur geschlossenen Form in der hier durchgeführten Simulation nicht beobachtet werden kann, hängt vermutlich in erster Linie damit zusammen, dass der Bindemodus von ATP in 1KAX nicht fehlerfrei auf *hsc_az* übertragen werden kann. Zum einen liegt die NBD in 1KAX - wie oben beschrieben - in einer stärker geschlossenen Konformation vor, was sich natürlich auch auf die Nukleotidbindetasche auswirkt. Zudem wird die Bindetaschenkonformation in *hsc_az* von einem TMAO-Molekül aus den Kristallisationspuffer beeinflusst, der in diesem Bereich bindet.

Da vor allem die Interaktionen zwischen ATP und der NBD-Subdomäne I offensichtlich nicht ausreichend stark waren, um die geschlossene NBD-Konformation zu induzieren, sollte für weitere Untersuchungen der NBD-Dynamik ein besonderes Augenmerk auf diesem Bereich liegen. Experimentelle Daten weisen zudem darauf hin, dass drei Kaliumionen in die Interaktion zwischen ATP und dem Protein involviert sind [157], die hier nicht berücksichtigt wurden. Für die *in vivo*-Dynamik der NBD und der Nukleotid-Bindetasche spielen zudem Co-Chaperone eine wichtige Rolle, deren Einfluss in diesen Simulationen nicht berücksichtigt werden kann.

Für weitere Untersuchungen in diesem Bereich bietet sich sicherlich eine Simulation der 1KAX-Struktur an, um zu untersuchen ob und wie die geschlossene NBD-Konformation in diesem System stabilisiert wird.

12.7. Substrat-Bindedomäne

Die SBD besteht aus einer β -Domäne, die vorwiegend in β -Faltblattstruktur vorliegt, und einer α -Domäne. Die α -Domäne wird auch als LID-Domäne bezeichnet und weist eine helikale Tertiärstruktur auf. Sie wird strukturell in eine Helix-A und eine Helix-B unterteilt. Helix-A ist in der Kristallstruktur *hsc_az* direkt an der Interface-Interaktion und Ausbildung der Interface-Bindetasche beteiligt. Helix-B schließt sich direkt an Helix-A an, steht aber in einem Winkel von 90° zu dieser (vgl. Abbildung 8.1 und 12.29). In der Kristallstruktur ist Helix-B nur sehr kurz, da durch das trunkierte Konstrukt die helikale Tertiärstruktur in diesem Bereich nicht stabilisiert werden kann und die Peptikette stattdessen in ungefalteter Form in der Substratbindetasche gebunden ist (in Abbildung 12.29 schwarz dargestellt).

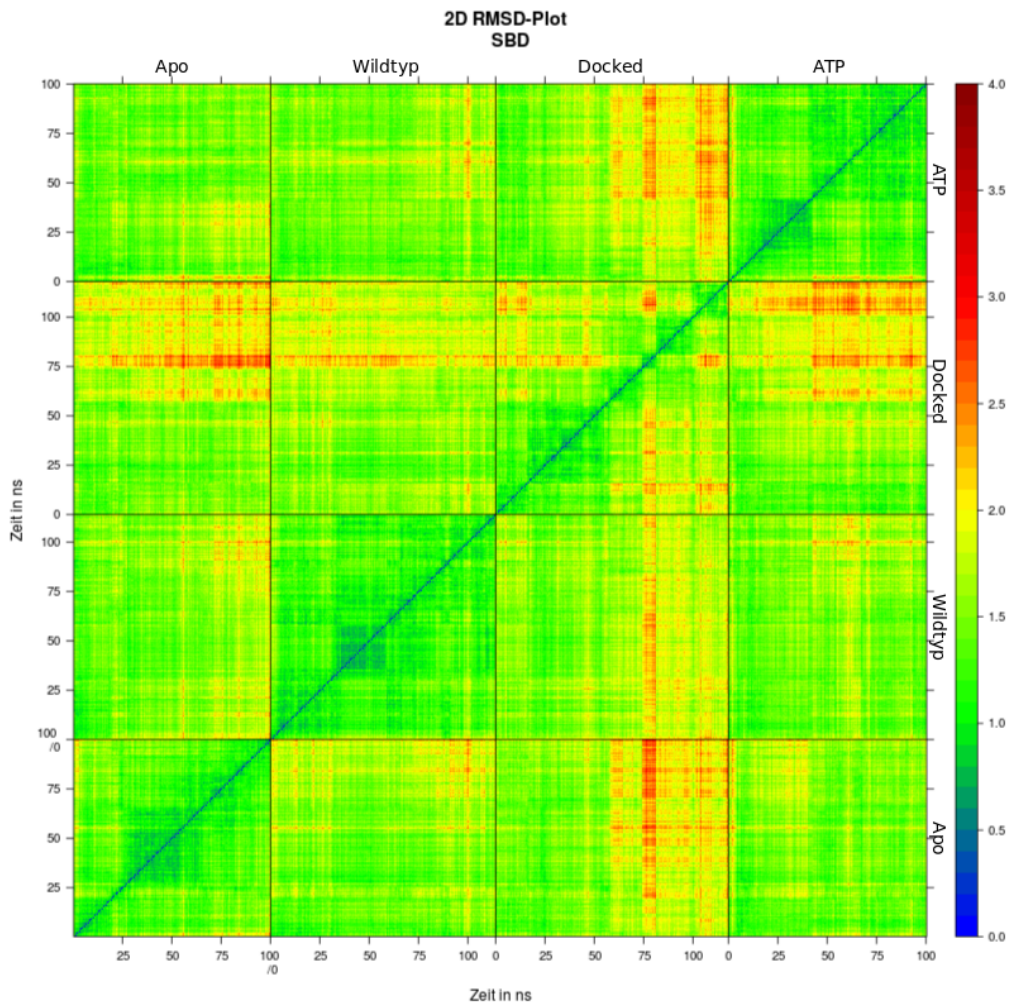


Abbildung 12.28.: 2D-RMSD Plot für die SBD; Überlagerung und RMSD-Berechnung für das Proteinrückgrat der Aminosäuren 395 bis 544.

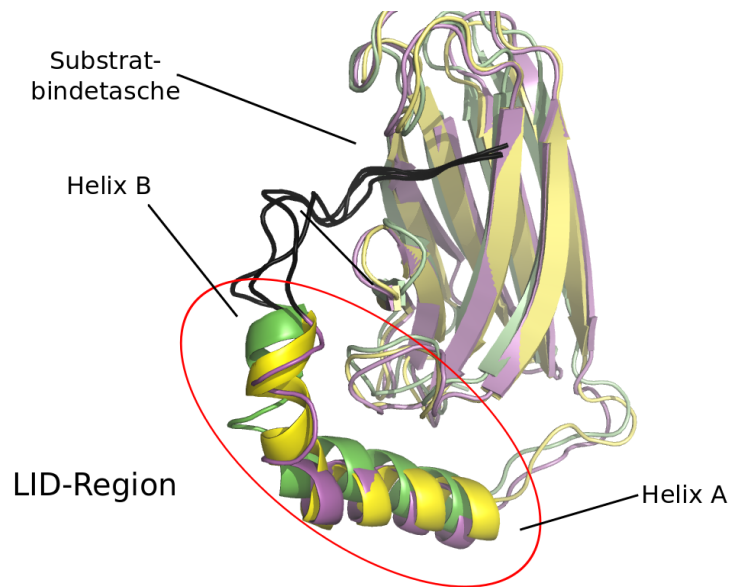


Abbildung 12.29.: Der Snapshot Wildtyp-138 (grün) ist der Repräsentant für die SBD-Hauptkonformation in allen vier Simulationen. Die Snapshots Docked-747 (violett) und Docked-646 (gelbe) treten im letzten Drittel der Docked-MD auf und sind durch eine leichte Verschiebung der LID-Region (Helix A und B) gekennzeichnet. Die Strukturen wurden auf das SBD-Proteinrückgrat (Aminosäuren 395 bis 544) der Startstruktur der jeweiligen Trajektorie überlagert. Der Bereich der von Helix-B, der aufgrund des truncierten Konstrukts nicht in der nativen Form vorliegt, ist schwarz dargestellt.

Die Konformation der Kristallstruktur bleibt in allen vier Simulationen über weite Strecken erhalten. Die RMSD-Werte im 2D-RMSD-Plot liegen größtenteils unter 2.5 Å. Die größte Konformationsänderung findet im zweiten Teil der Docked-Simulation statt. Hier werden RMSD-Werte von bis zu 4 Å beobachtet.

Ein hierarchisches Clustering mit einem Grenzwert von 3 Å ergibt einen sehr großen (3569 Snapshots) und zwei kleinere Cluster (651 bzw. 95 Snapshots). Bei den beiden kleinen Clustern handelt es sich um Konformationsfamilien, die im zweiten Teil der Docked-Simulation auftreten. Abbildung 12.29 zeigt die Überlagerung der repräsentativen Konformationen. Der wesentliche Konformationsunterschied findet sich in der LID-Region und ist dadurch gekennzeichnet, dass Helix-A relativ zur β -Domäne verschoben wird (in der Abbildung nach rechts und leicht nach unten). Dies überträgt sich auch auf Helix-B, die ebenfalls ihre relative Position zur β -Domäne ändert.

Experimentelle Daten weisen darauf hin, dass die LID-Region (v.a. Helix-B) an der allosteren Kommunikation beteiligt ist und durch einen "Deckel"-Mechanismus die Substratspezifität und Affinität der Substratbindetasche steuert [43, 19]. Die Dynamik der LID-Region im Vergleich zur β -Domäne ist deswegen eigentlich sehr interessant. Anhand der hier durchgeführten Simulationen kann die Dynamik der LID-Region jedoch nur sehr

eingeschränkt analysiert werden, weil die Domäne teilweise trunkiert und der ungefaltete Bereich der Helix-B zudem in der Substratbindetasche fixiert ist (vgl. Abbildung 12.29). Es muss davon ausgegangen werden, dass es sich hierbei um Kristallpackungseffekte handelt, die nicht die native Konformation widerspiegeln. Der "Deckel"-Mechanismus kann daher nicht beobachtet werden und die Flexibilität von Helix-A und einiger Bereiche von Helix-B kann allenfalls als Anhaltspunkt für einen solchen Mechanismus angesehen werden. Eine Analyse des allosteren Einflusses von ATP oder von Apoptozol auf die Konformation der SBD ist ebenfalls nicht möglich.

12.8. Zusammenfassende Diskussion

Die zentralen Fragestellungen, die mit Hilfe der MD-Simulationen beantwortet werden sollten, betreffen die Stabilität des Proteins insgesamt sowie das dynamische Verhalten der Interface-Bindetasche.

Zur Stabilität des Gesamtsystems lässt sich sagen, dass man in allen Simulationen deutliche Abweichungen von der Ausgangsstruktur *hsc_az* beobachten kann. Die $C\alpha$ -RMSD-Verläufe (Abbildung 12.1) verdeutlichen dies. Trotzdem bleibt die Grundstruktur mit einer ausgedehnten Kontaktfläche zwischen NBD und SBD, wie sie in den Kristallstrukturen *hsc_az* und 1YUW auftritt, in allen vier durchgeführten Simulationen erhalten. Obwohl für große Domänenbewegungen - wie es der Übergang zwischen den allosteren Zuständen von Hsp70 sicherlich ist - oftmals größere Zeitfenster benötigt werden, spricht die stabile Grundkonformation in den hier simulierten 100 ns doch dafür, dass es sich bei *hsc_az* um eine biologisch relevante Proteinkonformation handelt und nicht um ein Artefakt, welches aufgrund von Kristallpackungseffekten oder der mutierten Aminosäuren entstanden ist. Die Analysen zeigen, dass das Wildtyp-System (in dem die Aminosäuren rückmutiert wurden) durch die beiden zusätzlichen polaren Seitenketten zwar ein anderes Interaktionsmuster in der Mutationsregion bewirkt, eine Destabilisierung der NBD-SBD-Kontaktfläche oder eine signifikante Beeinflussung der Dynamik des Gesamtsystems tritt jedoch nicht auf.

Interessant ist die Beobachtung, dass das Wildtyp-System im Bereich des Linkers eine deutlich erhöhte Flexibilität aufweist. Die Alanin-Mutationen erlauben eine verbesserte hydrophobe Interaktionen zwischen der Mutationsregion und dem Linker, was dessen Flexibilität einschränkt. Dies kann ein Einflussfaktor für die bessere Kristallisierbarkeit der mutierten Proteine sein.

Die Interface-Bindetasche zeigt in den vier Simulationen - trotz der insgesamt stabilen Proteinstruktur - verschiedene Konformationen, die jeweils eng mit der relativen Stellung von NBD und SBD zueinander zusammenhängen. Vor allem im Apo-, Wildtyp- und ATP-System bleibt diese Flexibilität über den kompletten Simulationszeitraum erhalten. Die Gegenwart von Apoptozol in der Bindetasche beeinflusst die Konformation deutlich. Dabei ändert Apoptozol gleich zu Beginn der Simulation seinen Bindemodus, was zu einer veränderten Bindetaschentopologie führt. Apoptozol stabilisiert im weiteren

Verlauf der Simulation die Konformation der Bindetasche, so dass auch die NBD-SBD-Domänenbewegung eingeschränkt wird.

Zusammengenommen lässt dies zwei Rückschlüsse zu. Erstens, besteht ein deutlicher Zusammenhang zwischen der Dynamik der Interface-Bindetasche und der Dynamik des Gesamtsystems. Das ist ein wichtiger Hinweis darauf, dass die Bindung von Liganden im Interface einen Einfluss auf die allosterische Kommunikation zwischen NBD und SBD hat. Zweitens, zeigt die Apoptozol-Simulation, dass die Bindetaschenkonformation im Sinne eines "Induced-Fit"-Mechanismus durch die Bindung eines Liganden angepasst werden kann. Im Hinblick auf die Entwicklung der Tetrahydroisochinolinone werden MD-Simulationen in zukünftigen Arbeiten deswegen sehr wichtig sein, um potentielle Bindetaschenkonformationen für diese Substanzklasse zu untersuchen und die biologischen Aktivitätswerte besser verstehen und interpretieren zu können.

Neben der Interface- und Mutationsregion wurde auch die Dynamik der beiden großen Domänen (NBD und SBD) analysiert. Die SBD konnte jedoch aufgrund der trunkeierten LID-Region nur sehr eingeschränkt bewertet werden. Für die NBD konnte hingegen eine deutlich veränderte Dynamik durch die Bindung von Apoptozol beobachtet werden. Auch das ist ein Hinweis auf den allosterischen Einfluss der Interface-Bindetasche. Die Simulation mit gebundenem ATP ergab hingegen überraschenderweise eine Öffnung der NBD, was im Widerspruch zu experimentellen Daten, beispielsweise der Kristallstruktur 1KAX, steht. Ob dieses Verhalten primär durch eine unzureichende Startstruktur der ATP-Simulation oder die Nichtberücksichtigung von Cofaktoren zu begründen ist, kann an dieser Stelle nicht abschließend geklärt werden.

Im Rahmen einer kritischen Diskussion soll auch nicht unerwähnt bleiben, dass nicht ganz klar ist, in welchem Stadium des funktionellen Hsp-Zyklus die NBD-Apo-Struktur, wie sie in *hsc_az* und 1YUW beobachtet wird, einzuordnen ist. Die gängigste Hypothese zu den allosterischen Zuständen von Hsp70 geht davon aus, dass nur im allosterischen Zustand mit gebundenem ATP ein direkter Kontakt zwischen NBD und SBD stattfindet, während der ADP-Zustand - dem der NBD-Apo-Zustand in der Regel zugeordnet wird - mit einer offenen Proteinkonformation ohne Domänenkontakt assoziiert wird. Diese These wird beispielsweise in den Studien von Bertelsen [106], mit der PDB-Struktur 2KHO, und von Mapa [19] vertreten. Auch in anderen experimentellen Studien finden sich Hinweise, dass die Kommunikation zwischen den Domänen hauptsächlich im ATP-Zustand stattfindet [158, 159]. Oftmals wird dabei eine Beteiligung des Linker-Segments an der Interdomain-Kommunikation festgestellt, der vermutlich in einem hydrophoben Bereich zwischen den Subdomänen IA und IIA mit der NBD interagiert [159, 160].

In einer sehr umfassenden MD-Studie haben Chiappori und Mitarbeiter den Übergang zwischen der *offenen* und *geschlossenen* SBD-Konformation untersucht [161]. Als Modell für die offene Konformation wurde auch die Kristallstruktur 2KHO verwendet. Als Modell der geschlossenen Konformation diente ein nicht-allosterisches Hsp110-Konstrukt aus *Saccharomyces cerevisiae* (PDB 3C7N [162]). Durch verschiedene Simulationen jeweils mit ADP und ATP, konnte der Übergang zwischen den beiden Konformationen simuliert und mögliche Signalwege der allosterischen Kommunikation identifiziert werden.

Hinsichtlich der direkten Kommunikation zwischen NBD und SBD spielt in diesen Simulationen vor allem der Linker eine wichtige Rolle. Auch der "Deckel"-Mechanismus der LID-Region kann in diesen Simulationen beobachtet werden.

Ein Vergleich zwischen den hier durchgeführten Simulationen mit jenen von Chiapori et al. fällt jedoch schwer, da die Arbeiten von sehr unterschiedlichen Startpunkten ausgehen, die unter den Bedingungen einer MD-Simulation nicht ineinander überführbar sind. Eine zentrale Rolle des Linkers an der Interdomain-Kommunikation kann in den in der vorliegenden Arbeit durchgeführten Simulationen deswegen nicht beobachtet werden. Die Dynamik der LID-Region ist aufgrund der teilweise artifiziellen Konformation der SBD ebenfalls in der vorliegenden Arbeit ebenfalls nicht auswertbar.

Gerade vor diesem Hintergrund ist es sehr wichtig, dass komplementär zu den molekulardynamische Untersuchungen weitere experimentelle Untersuchungen durchgeführt werden. Dabei steht vor allem die weitere Charakterisierung der Bindung der Tetrahydroisochinolone an Hsp70 (beispielsweise durch SPR oder ITC-Messungen) sowie der biologischen Effekte der Inhibitoren (beispielsweise durch ATPase- oder Chaperon-Aktivitätsassays) im Vordergrund. Eng damit verknüpft ist auch der Nachweis der biologischen Relevanz der Proteinkonformation von *hsc_az*. Schlecht und Mitarbeiter haben für zwei andere Hsp70-Inhibitoren (PES und Ver-155008) entsprechende Experimente durchgeführt, die auch für die Tetrahydroisochinolinone notwendig sind [163].

Ein bessere biologische Datengrundlage würde einerseits die Weiterentwicklung der Tetrahydroisochinolinone mit Docking- und Scoring-Methoden vereinfachen, ermöglicht andererseits aber auch eine fokussiertere Analyse der MD-Trajektorien. Als komplementäre Technik können auch Mutationstudien - wie sie beispielsweise von Jiang et al. [105] durchgeführt wurden - dabei helfen, die Ergebnisse der molekulardynamischen Simulationen experimentell zu überprüfen bzw. neue Ansatzpunkte für die Auswertung der Trajektorien liefern.

Insgesamt sollte dabei das Augenmerk weiterhin stärker auf der Dynamik der Interface-Bindetasche als auf der mechanistischen Untersuchung des Gesamtsystems liegen. Die Simulation eines oder mehrerer Hsp70-Tetrahydroisochinolinon-Komplexe ist in diesem Zusammenhang sicherlich einer der nächsten Schritte. Das bis hierher erfolgreiche Design dieser Inhibitorenserie stellt in jedem Fall ein wichtiges Argument für die biologische Relevanz der *hsc_az*-Struktur und die Adressierung der Interfaceregion zum Zweck der Hsp70-Inhibition dar.

Teil III.

Zusammenfassung/Summary

13. Zusammenfassung

Techniken des computergestützten Wirkstoffdesigns spielen eine wichtige Rolle bei der Entwicklung neuer Wirkstoffe. Die vorliegende Arbeit befasst sich sowohl mit der Entwicklung als auch mit der praktischen Anwendung von Methoden des strukturbasierten Wirkstoffdesigns. Die Arbeit gliedert sich daher in zwei Teile.

Entwicklung empirischer Scoring-Funktionen Der erste Teil beschäftigt sich mit der Entwicklung von empirischen Scoring-Funktionen, die eine Schlüsselrolle im strukturbasierten computergestützten Wirkstoffdesign einnehmen. Grundlage dieser Arbeiten sind die empirischen Deskriptoren und Scoring-Funktionen aus dem SFCscore-Programmpaket.

Dabei wurde zunächst untersucht, wie sich die Zusammensetzung der Trainingsdaten auf die Vorhersagen von empirischen Scoring-Funktionen auswirkt. Durch die gezielte Zusammenstellung eines neuen Trainingsdatensatzes wurde versucht, die Spannweite der Vorhersagen zu vergrößern, um so vor allem eine bessere Erkennung von hoch- und niedrig-affinen Komplexen zu erreichen. Die resultierende Funktion erzielte vor allem im niedrig-affinen Bereich verbesserte Vorhersagen.

Der zweite Themenkomplex beschäftigt sich ebenfalls mit der verbesserten Separierung von aktiven und inaktiven Verbindungen. Durch den Einsatz der *Machine Learning*-Methode *RandomForest* wurden dazu Klassifizierungsmodelle abgeleitet, die im Unterschied zu den klassischen Scoring-Funktionen keinen genauen Score liefern, sondern die Verbindungen nach ihrer potentiellen Aktivität klassifizieren. Am Beispiel des mykobakteriellen Enzyms InhA konnte gezeigt werden, dass derartige Modelle den klassischen Scoring-Funktionen im Bezug auf die Erkennung von aktiven Verbindungen deutlich überlegen sind.

Der *RandomForest*-Algorithmus wurde im nächsten Schritt auch verwendet, um eine neue Scoring-Funktion zur Vorhersage von Bindungsaffinitäten abzuleiten. Diese Funktion wurde unter dem Namen SFCscore^{RF} in das SFCscore-Programmpaket implementiert. Die Funktion unterscheidet sich in einigen wesentlichen Punkten von den ursprünglichen SFCscore-Funktionen. Zum einen handelt es sich beim RF-Algorithmus um eine nicht-lineare Methode, die im Unterschied zu den klassischen Methoden, die zur Ableitung von Scoring-Funktionen eingesetzt werden, nicht von der Additivität der einzelnen Deskriptoren ausgeht. Der Algorithmus erlaubt außerdem die Verwendung aller verfügbaren SFCscore-Deskriptoren, was eine deutlich umfassendere Repräsentation von Protein-Ligand-Komplexen als Grundlage des Scorings ermöglicht. Für die Ableitung von

SFCscore^{RF} wurden insgesamt 1005 Komplexe im Trainingsdatensatz verwendet. Dieser Datensatz ist somit einer der größten, die bisher für die Ableitung einer empirischen Scoring-Funktion verwendet wurden.

Die Evaluierung gegen zwei Benchmark-Datensätze ergab deutlich bessere Vorhersagen von SFCscore^{RF} im Vergleich zu den ursprünglichen SFCscore-Funktionen. Auch im internationalen Vergleich mit anderen Scoring-Funktion konnten für beide Datensätze Spitzenwerte erreicht werden. Weitere ausgiebige Testungen im Rahmen einer Leave-Cluster-Out-Validierung und die Teilnahme am *CSAR 2012 Benchmark Exercise* ergaben, dass auch SFCscore^{RF} Performanceschwankungen bei der Anwendung an proteinspezifischen Datensätzen zeigt - ein Phänomen, dass bei Scoring-Funktionen immer beobachtet wird. Die Analyse der *CSAR 2012*-Datensätze ergab darüber hinaus wichtige Erkenntnisse im Bezug auf Vorhersage von gedockten Posen sowie über die statistische Signifikanz bei der Evaluierung von Scoring-Funktionen.

Die Tatsache, dass empirische Scoring-Funktionen innerhalb eines bestimmten chemischen Raums trainiert wurden, ist ein wichtiger Faktor für die protein-abhängigen Leistungsschwankungen, die in dieser Arbeit beobachtet wurden. Verlässliche Vorhersagen sind nur innerhalb des kalibrierten chemischen Raums möglich. In dieser Arbeit wurden verschiedene Ansätze untersucht, mit denen sich diese "Applicability Domain" für die SFCscore-Funktionen definieren lässt. Mit Hilfe von PCA-Analysen ist es gelungen die "Applicability Domain" einzelner Funktionen zu visualisieren. Zusätzlich wurden eine Reihe numerischer Deskriptoren getestet, mit den die Vorhersageverlässlichkeit basierend auf der "Applicability Domain" abgeschätzt werden könnte. Die RF-Proximity hat sich hier als vielversprechender Ausgangspunkt für weitere Entwicklungen erwiesen.

Hsp70 Der zweite Teil der Arbeit beschäftigt sich mit der Entwicklung neuer Inhibitoren für das Chaperon Hsp70, welches eine vielversprechende Zielstruktur für die Therapie des multiplen Myeloms darstellt. Grundlage dieser Arbeiten war eine Leitstruktur, die in einer vorhergehenden Arbeit entdeckt wurde und die vermutlich an einer neuartigen Bindestelle in der Interface-Region zwischen den beiden großen Domänen von Hsp70 angreift.

Die Weiterentwicklung und Optimierung dieser Leitstruktur, eines Tetrahydroisochinolinon-Derivats, stand zunächst im Vordergrund. Anhand detaillierter Docking-Analysen wurde der potentielle Bindemodus der Leitstruktur in der Interfaceregion von Hsp70 untersucht. Basierend auf diesen Ergebnissen wurde eine Substanzbibliothek erstellt, die von Kooperationspartnern innerhalb der KFO 216 synthetisiert und biologisch getestet wurde. Die Struktur-Wirkungsbeziehungen, die sich aus diesen experimentellen Daten ableiten lassen, konnten teilweise gut mit den erstellten Docking-Modellen korreliert werden. Andere Effekte konnten anhand der Docking-Posen jedoch nicht erklärt werden. Für die Entwicklung neuer Derivate ist deswegen eine umfassendere experimentelle Charakterisierung und darauf aufbauend eine Verfeinerung der Bindungsmodelle notwendig.

Strukturell handelt es sich bei Hsp70 um ein Zwei-Domänen-System, das verschiedene allosterische Zustände einnehmen kann. Um die Auswirkungen der daraus folgenden Flexibilität auf die Stabilität der Struktur und die Bindung von Inhibitoren zu untersuchen, wurden molekulardynamische Simulationen für das Protein durchgeführt. Diese zeigen, dass das Protein tatsächlich eine überdurchschnittlich hohe Flexibilität aufweist, die vor allem durch die relative Bewegung der beiden großen Domänen zueinander dominiert wird. Die Proteinkonformation, die in der Kristallstruktur *hsc_az* beobachtet wird, bleibt jedoch in ihrer Grundstruktur in allen vier durchgeführten Simulationen erhalten. Es konnten hingegen keine Hinweise dafür gefunden werden, dass die Mutationen, welche die für die strukturbasierten Arbeiten verwendete Kristallstruktur im Vergleich zum Wildtyp aufweist, einen kritischen Einfluss auf die Gesamtstabilität des Systems haben.

Obwohl die Interface-Region zwischen NBD und SBD also in allen Simulationen erhalten bleibt, wird die Konformation in diesem Bereich doch wesentlich durch die Domänenbewegung beeinflusst und variiert. Da dieser Proteinbereich den wahrscheinlichsten Angriffspunkt der Tetrahydroisochinolinone darstellt, wurde der Konformationsraum detailliert untersucht. Wie erwartet weist die Region eine nicht unerhebliche Flexibilität auf, welche zudem, im Sinne eines "Induced-Fit"-Mechanismus, durch die Gegenwart eines Liganden (Apoptozol) stark beeinflusst wird. Es ist daher als sehr wahrscheinlich anzusehen, dass die Dynamik der Interface-Region auch einen wesentlichen Einfluss auf die Bindung der Tetrahydroisochinolinone hat. Molekulardynamische Berechnungen werden deswegen auch in zukünftige Arbeiten auf diesem Gebiet eine wichtige Rolle spielen.

Die Analysen zeigen zudem, dass die Konformation der Interface-Region eng mit der Konformation des gesamten Proteins - vor allem im Bezug auf die relative Stellung von SBD und NBD zueinander - verknüpft ist. Das untermauert die Hypothese, dass die Interface-Bindetasche einen Angriffspunkt für die Inhibition des Proteins darstellt.

14. Summary

Methods of computational drug design play a crucial role in the development of new pharmaceutical drugs. The work presented here comprises the methodological development and the practical application of structure-based techniques in computational drug design.

SFCscore The first part of this dissertation focuses on the development of empirical scoring functions, which play an essential part in structure-based computer-aided drug design. The basis for this work are the empirical descriptors and scoring functions of the SFCscore software package.

First, the influence of the training data composition on the prediction of empirical scoring functions was analyzed. A new training data set was created to spread the prediction range of the function and thus achieve a better separation of high and low affinity binders. The resulting function indeed yielded better predictions in the low affinity area compared to the original functions.

In another approach, which also addresses the issue of discriminating *active* and *inactive* compounds, the *Machine Learning* method *RandomForest* was used to derive a classification model. Different to classical empirical scoring functions, this model no longer predicts a precise value but classifies the compounds according to their potential affinity as *active* or *inactive*. The example of the mycobacterial enzyme InhA showed that such models are clearly superior to different classical scoring function in terms of separating active and inactive compounds.

The *RandomForest* algorithm was also used to derive a new scoring function for the prediction of binding affinities. This new function was implemented into the SFCscore software package under the name SFCscore^{RF}. This new function differs from the original SFCscore functions in several essential points. On the one hand, the RF-algorithm is a non-linear method, which - in contrast to classical methods used for the derivation of empirical scoring functions - does not assume the additivity of the single descriptors. On the other hand, the algorithm allows for using the whole set of available SFCscore descriptors and is therefore able to utilize a more comprehensive representation of a protein ligand complex as the basis for the prediction. Additionally, the training data set used to derive SFCscore^{RF} comprised 1005 complexes. This training set is one of the largest data sets used to train an empirical scoring function.

The evaluation against two widely-used benchmark sets confirmed that SFCscore^{RF} yields superior predictions as compared to the original functions. The comparison with

other functions tested for these benchmarks shows that SFCscore^{RF} also achieves top results on an international level. Further analyses using a leave-cluster-out validation scheme and the participation in the *CSAR 2012 Benchmark Exercise* revealed that - similar to other scoring functions - SFCscore^{RF} shows varying performances when applied to protein-specific data sets. Additionally, by analysing the results of the CSAR 2012 data sets, valuable insight were gained regarding the prediction of docking poses and the statistical significance for the evaluation and comparison of scoring functions.

The fact that empirical scoring functions are trained within a certain chemical space, is an important reason for the target-dependent performance observed in this work. Reliable predictions can only be expected within the calibrated area. Different approaches for the definition of this “applicability domain” are presented in this work. PCA analyses have been used to create a two dimensional representation of the “applicability domain”. Additionally, different numerical descriptors have been tested to estimate the reliability of SFCscore predictions. The RF-proximity has been found to be a promising starting point for future research.

Hsp70 The development of new inhibitors for the molecular chaperone Hsp70 - a promising target in the therapy of multiple myeloma - comprises the second part of this dissertation. The basis for this work was a lead structure that was found in a previous work and attacks a novel binding pocket in the interface between the two domains of the Hsp70 protein.

The optimization and development of that lead structure - a tetrahydroisochinolinone - was the primary focus of the present work. Potential binding poses in the interface were elucidated by detailed docking analyses. Based on that information, a compound library was compiled, which was synthesized and biologically analyzed by cooperation partners within the CRU 216. The resulting structure activity relationships can partially be explained on the basis of the corresponding docking poses. However, some of the effects remain unexplained. For the further development of new derivatives a comprehensive experimental characterization of the current compounds is needed. This information can be used as a basis for the refinement of the existing binding models.

Hsp70 is a two-domain system, which can visit different allosteric states. To further investigate the effects of the resulting flexibility on the stability of the structure and on inhibitor binding, molecular dynamics simulations were conducted. These simulations show an above-average flexibility of the protein, which is primarily dominated by the movement of the two domains NBD and SBD relatively to each other. However, the basic conformation that is observed in the crystal structure *hsc_az*, which was used in this work, remains stable in all simulations. Furthermore, the trajectories showed no evidence that the mutations, in which *hsc_az* differs from the wild type protein, have a significant effect on the overall protein conformation.

Although, the overall conformation of the interface between NDB and SBD remains stable, the exact conformation in this area is significantly influenced by the domain

movement. As this region includes the binding pocket of the tetrahydroisochinolinones, the conformational space of this area was analyzed in detail. The analyses expectedly reveal a high flexibility in the interface area that is dominated by the SBD-NBD movement. Furthermore, it could be shown that the conformation and dynamics can be influenced by a bound ligand (apoptozole), in terms of an induced fit mechanism. It is highly probable that the binding of the tetrahydroisochinolinones trigger similar effects, influencing the binding mechanism of this compound class. Thus, molecular dynamics simulations should play a crucial role in the future development of new compounds.

The analyses also show that the dynamics of the interface region have large effects on the overall structure of the protein and vice versa. Especially, the relative orientation of NBD and SBD has a large impact on the binding pocket. This underlines the hypothesis that the interface region constitutes a promising target area for the inhibition of Hsp70.

Teil IV.

Anhang

Abkürzungsverzeichnis

AD Applicability Domain

Az Apoptozol

CHK1 Checkpoint-Kinase 1

ERK2 Extracellular-signal-regulated Kinase 2

FITC Fluoresceinisothiocyanat

GA genetischer Algorithmus

HSP70 Heat Shock Protein 70

HTS High Throughput Screening

LCO Leave-Cluster-Out

LpxC UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamin Deacetylase

MD Molekulardynamik

MLR Multiple Lineare Regression

MM Multiples Myelom

NMR Nuclear Magnetic Resonance

OOB Out-Of-Bag

PCA Principal Component Analysis (Hauptkomponentenanalyse)

PLS Partial Least Squares

RF Random Forest

RMSE Root Mean Square Error

ROC Receiver-Operating-Characteristic

TMAO Trimethylaminoxid

VS Virtuelles Screening

Literatur

- [1] M. Hein, D. Zilian und C. A. Sotriffer. „Docking compared to 3D-pharmacophores: The scoring function challenge“. In: *Drug Discovery Today: Technologies*. 3D Pharmacophore Elucidation and Virtual Screening 7.4 (2010), e229–e236.
- [2] B. Merget u. a. „MycPermCheck: the *Mycobacterium tuberculosis* permeability prediction tool for small molecules“. In: *Bioinformatics* 29.1 (2013), S. 62–68.
- [3] D. Zilian und C. A. Sotriffer. „SFCscore^{RF}: A Random Forest-based scoring function for improved affinity prediction of protein–ligand complexes“. In: *Journal of Chemical Information and Modeling* 53.8 (2013), S. 1923–1933.
- [4] G. Klebe. *Wirkstoffdesign Entwurf und Wirkung von Arzneistoffen*. Heidelberg: Spektrum Akademischer Verlag, 2009.
- [5] C. Sotriffer. *Virtual screening principles, challenges, and practical guidelines*. Weinheim, Germany: Wiley-VCH, 2011.
- [6] M. McGann. „FRED and HYBRID docking performance on standardized datasets“. In: *Journal of Computer-Aided Molecular Design* 26.8 (2012), S. 897–906.
- [7] M. McGann. „FRED pose prediction and virtual screening accuracy“. In: *Journal of Chemical Information and Modeling* 51.3 (2011), S. 578–596.
- [8] M. Rarey u. a. „A fast flexible docking method using an incremental construction algorithm“. In: *Journal of Molecular Biology* 261.3 (1996), S. 470–489.
- [9] M. L. Verdonk u. a. „Improved protein–ligand docking using GOLD“. In: *Proteins: Structure, Function, and Bioinformatics* 52.4 (2003), 609–623.
- [10] G. M. Morris u. a. „Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function“. In: *Journal of Computational Chemistry* 19.14 (1998), 1639–1662.
- [11] G. L. Warren u. a. „A critical assessment of docking programs and scoring functions“. In: *Journal of Medicinal Chemistry* 49.20 (2006), S. 5912–5931.
- [12] D. Rognan. „Ch. 6: Docking methods for virtual screening. Principles and recent advances“. In: *Virtual screening principles, challenges, and practical guidelines*. Hrsg. von C. Sotriffer. Weinheim, Germany: Wiley-VCH, 2011.

- [13] C. Sotriffer und H. Matter. „Ch. 7: The challenge of affinity prediction: Scoring functions for structure-based virtual screening“. In: *Virtual screening principles, challenges, and practical guidelines*. Hrsg. von C. Sotriffer. Weinheim, Germany: Wiley-VCH, 2011.
- [14] T. Cheng u. a. „Comparative assessment of scoring functions on a diverse test set“. In: *Journal of Chemical Information and Modeling* 49.4 (2009), S. 1079–1093.
- [15] J. L. Melville, E. K. Burke und J. D. Hirst. „Machine learning in virtual screening“. In: *Combinatorial Chemistry & High Throughput Screening* 12.4 (2009), S. 332–343.
- [16] R. N. Jorissen und M. K. Gilson. „Virtual screening of molecular databases using a Support Vector Machine“. In: *Journal of Chemical Information and Modeling* 45.3 (2005), S. 549–561.
- [17] L. Li, B. Wang und S. O. Meroueh. „Support vector regression scoring of receptor-ligand complexes for rank-ordering and virtual screening of chemical libraries“. In: *Journal of Chemical Information and Modeling* 51.9 (2011), S. 2132–2138.
- [18] M. Daugaard, M. Rohde und M. Jäättelä. „The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions“. In: *FEBS Letters* 581.19 (2007), S. 3702–3710.
- [19] K. Mapa u. a. „The conformational dynamics of the mitochondrial Hsp70 chaperone“. In: *Molecular Cell* 38.1 (2010), S. 89–100.
- [20] M. P. Mayer und B. Bukau. „Hsp70 chaperones: Cellular functions and molecular mechanism“. In: *Cellular and Molecular Life Sciences* 62.6 (2005), S. 670–684.
- [21] C. Jolly und R. I. Morimoto. „Role of the heat shock response and molecular chaperones in oncogenesis and cell death“. In: *Journal of the National Cancer Institute* 92.19 (2000), S. 1564–1572.
- [22] S. N. Witt. „Hsp70 molecular chaperones and Parkinson's disease“. In: *Biopolymers* 93.3 (2010), S. 218–228.
- [23] T. Kirkegaard u. a. „Hsp70 stabilizes lysosomes and reverts Niemann–Pick disease-associated lysosomal pathology“. In: *Nature* 463.7280 (2010), S. 549–553.
- [24] D. G. Millar u. a. „Hsp70 promotes antigen-presenting cell function and converts T-cell tolerance to autoimmunity in vivo“. In: *Nature Medicine* 9.12 (2003), S. 1469–1476.
- [25] M. P. Mayer. „Recruitment of Hsp70 chaperones: A crucial part of viral survival strategies“. In: *Reviews of Physiology, Biochemistry and Pharmacology* 153 (2005), S. 1–46.

- [26] C. G. Evans, L. Chang und J. E. Gestwicki. „Heat shock protein 70 (Hsp70) as an emerging drug target“. In: *Journal of Medicinal Chemistry* 53.12 (2010), S. 4585–4602.
- [27] Martin Kortüm u. a. *Leitlinie: multiples Myelom, Onkopedia*, <http://www.dgho-onkopedia.de>. 2013.
- [28] M. A. Dimopoulos u. a. „Pathogenesis and treatment of renal failure in multiple myeloma“. In: *Leukemia* 22.8 (2008), S. 1485–1493.
- [29] L. Eisele u. a. „Prevalence and progression of monoclonal gammopathy of undetermined significance and light-chain MGUS in Germany“. In: *Annals of Hematology* 91.2 (2012), S. 243–248.
- [30] R. A. Kyle und S. V. Rajkumar. „Multiple myeloma“. In: *New England Journal of Medicine* 351.18 (2004), S. 1860–1873.
- [31] K. Podar, D. Chauhan und K. C. Anderson. „Bone marrow microenvironment and the identification of new targets for myeloma therapy“. In: *Leukemia* 23.1 (2008), S. 10–24.
- [32] A. Mahindra u. a. „Latest advances and current challenges in the treatment of multiple myeloma“. In: *Nature Reviews Clinical Oncology* 9.3 (2012), S. 135–143.
- [33] European Medicines Agency. *Revlimid - Zusammenfassung des EPAR für die Öffentlichkeit, in Deutsch von der Website der europäischen Gesundheitsbehörde (EMA)*. 2013.
- [34] E. Mutschler u. a. *Arzneimittelwirkungen: Lehrbuch der Pharmakologie und Toxikologie*. Wissenschaftliche Verlagsgesellschaft, Stuttgart, 9. Auflage, 2008.
- [35] M. Chatterjee u. a. „The PI3K/Akt signaling pathway regulates the expression of Hsp70, which critically contributes to Hsp90-chaperone function and tumor cell survival in multiple myeloma“. In: *Haematologica* 98.7 (2013), S. 1132–1141.
- [36] D. R. Ciocca und S. K. Calderwood. „Heat shock proteins in cancer: Diagnostic, prognostic, predictive, and treatment implications“. In: *Cell Stress & Chaperones* 10.2 (2005), S. 86–103.
- [37] J. Nylandsted, K. Brand und M. Jäätelä. „Heat shock protein 70 is required for the survival of cancer cells“. In: *Annals of the New York Academy of Sciences* 926.1 (2000), 122–125.
- [38] D. D. Mosser und R. I. Morimoto. „Molecular chaperones and the stress of oncogenesis“. In: *Oncogene* 23.16 (2004), S. 2907–2918.
- [39] R. Nimmanapalli u. a. „HSP70 inhibition reverses cell adhesion mediated and acquired drug resistance in multiple myeloma“. In: *British Journal of Haematology* 142.4 (2008), 551–561.

- [40] V. L. Gabai, K. R. Budagova und M. Y. Sherman. „Increased expression of the major heat shock protein Hsp72 in human prostate carcinoma cells is dispensable for their viability but confers resistance to a variety of anticancer agents“. In: *Oncogene* 24.20 (2005), S. 3328–3338.
- [41] M. Pocaly u. a. „Overexpression of the heat-shock protein 70 is associated to imatinib resistance in chronic myeloid“. In: *Leukemia* 21.1 (2006), S. 93–101.
- [42] M. Jäättelä. „Escaping cell death: Survival proteins in cancer“. In: *Experimental Cell Research* 248.1 (1999), S. 30–43.
- [43] E. R. P. Zuiderweg u. a. „Allostery in the Hsp70 chaperone proteins“. In: *Molecular Chaperones*. Hrsg. von S. Jackson. Topics in Current Chemistry 328. Springer Berlin Heidelberg, 2013, S. 99–153.
- [44] H.-J. Böhm. *Protein-ligand interactions*. Methods and principles in medicinal chemistry, Band 19. Weinheim: Wiley-VCH, 2003.
- [45] P. Block u. a. „AffinDB: A freely accessible database of affinities for protein-ligand complexes from the PDB“. In: *Nucleic Acids Research* 34.suppl 1 (2006), S. D522–526.
- [46] C. A. Sotriffer u. a. „SFCscore: scoring functions for affinity prediction of protein–ligand complexes“. In: *Proteins: Structure, Function, and Bioinformatics* 73.2 (2008), 395–419.
- [47] R. Wang u. a. „The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures“. In: *Journal of Medicinal Chemistry* 47.12 (2004), S. 2977–2980.
- [48] R. Wang u. a. „The PDBbind database: Methodologies and updates“. In: *Journal of Medicinal Chemistry* 48.12 (2005), S. 4111–4119.
- [49] R. Wang u. a. „An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes“. In: *Journal of Chemical Information and Computer Sciences* 44.6 (2004), S. 2114–2125.
- [50] J. B. Dunbar u. a. „CSAR benchmark exercise of 2010: Selection of the protein–ligand complexes“. In: *Journal of Chemical Information and Modeling* 51.9 (2011), S. 2036–2046.
- [51] R. D. Smith u. a. „CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions“. In: *Journal of Chemical Information and Modeling* 51.9 (2011), S. 2115–2131.
- [52] M. Hein. „Entwicklung computergestützter Methoden zur Bewertung von Docking-Lösungen und Entwurf niedermolekularer MIP-Inhibitoren“. Diss. Würzburg: Universität Würzburg, 2014.

- [53] N. Triballeau u. a. „Virtual screening workflow development guided by the “Receiver Operating Characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4“. In: *Journal of Medicinal Chemistry* 48.7 (2005), S. 2534–2547.
- [54] T. A. Halgren u. a. „Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening“. In: *Journal of Medicinal Chemistry* 47.7 (2004), S. 1750–1759.
- [55] J.-F. Truchon und C. I. Bayly. „Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem“. In: *Journal of Chemical Information and Modeling* 47.2 (2007), S. 488–508.
- [56] J. Bortz. *Statistik für Human- und Sozialwissenschaftler*. 6., vollständig überarbeitete und aktualisierte Auflage. Springer-Lehrbuch. Heidelberg: Springer, 2005.
- [57] J. Hartung. *Multivariate Statistik*. 7., unveränderte Auflage. München: Oldenbourg, 2007.
- [58] A. L. Blum und P. Langley. „Selection of relevant features and examples in machine learning“. In: *Artificial Intelligence* 97.1–2 (1997), S. 245–271.
- [59] T. Kalliokoski u. a. „Comparability of mixed IC₅₀ data – A statistical analysis“. In: *PLoS ONE* 8.4 (2013), e61007.
- [60] C. Kramer u. a. „The experimental uncertainty of heterogeneous public K_i data“. In: *Journal of Medicinal Chemistry* 55.11 (2012), S. 5165–5173.
- [61] N. Huang, B. K. Shoichet und J. J. Irwin. „Benchmarking sets for molecular docking“. In: *Journal of Medicinal Chemistry* 49.23 (2006), S. 6789–6801.
- [62] J. J. Irwin und B. K. Shoichet. „ZINC—A free database of commercially available compounds for virtual screening“. In: *Journal of Chemical Information and Modeling* 45.1 (2005), S. 177–182.
- [63] J. J. Irwin u. a. „ZINC: A free tool to discover chemistry for biology“. In: *Journal of Chemical Information and Modeling* 52.7 (2012), S. 1757–1768.
- [64] H. M. Berman u. a. „The Protein Data Bank“. In: *Nucleic Acids Research* 28.1 (2000), S. 235–242.
- [65] D. A. Rozwarski u. a. „Crystal structure of the *Mycobacterium tuberculosis* enoyl-ACP reductase, InhA, in complex with NAD⁺ and a C16 fatty acyl substrate“. In: *The Journal of Biological Chemistry* 274.22 (1999), S. 15582–15589.
- [66] M. R. Kuo u. a. „Targeting tuberculosis and malaria through inhibition of enoyl reductase compound activity and structural data“. In: *Journal of Biological Chemistry* 278.23 (2003), S. 20851–20859.
- [67] T. J. Sullivan u. a. „High affinity InhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*“. In: *ACS Chemical Biology* 1.1 (2006), S. 43–53.

- [68] X. He u. a. „Pyrrolidine carboxamides as a novel class of inhibitors of enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*“. In: *Journal of Medicinal Chemistry* 49.21 (2006), S. 6308–6323.
- [69] X. He, A. Alian und P. R. Ortiz de Montellano. „Inhibition of the *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase InhA by arylamides“. In: *Bioorganic & Medicinal Chemistry* 15.21 (2007), S. 6649–6658.
- [70] J. S. Freundlich u. a. „Triclosan derivatives: Towards potent inhibitors of drug-sensitive and drug-resistant *Mycobacterium tuberculosis*“. In: *ChemMedChem* 4.2 (2009), S. 241–248.
- [71] S. R. Luckner u. a. „A slow, tight binding inhibitor of InhA, the enoyl-acyl carrier protein reductase from *Mycobacterium tuberculosis*“. In: *Journal of Biological Chemistry* 285.19 (2010), S. 14330–14337.
- [72] R. A. Friesner u. a. „Extra Precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes“. In: *Journal of Medicinal Chemistry* 49.21 (2006), S. 6177–6196.
- [73] R. A. Friesner u. a. „Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking Accuracy“. In: *Journal of Medicinal Chemistry* 47.7 (2004), S. 1739–1749.
- [74] W. Jorgensen, D. Maxwell und J. Tirado-Rives. „Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids“. In: *Journal of the American Chemical Society* 118.45 (1996), S. 11225–11236.
- [75] M. D. Eldridge u. a. „Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes“. In: *Journal of Computer-Aided Molecular Design* 11.5 (1997), S. 425–445.
- [76] J. R. Greenwood u. a. „Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution“. In: *Journal of Computer-Aided Molecular Design* 24.6-7 (2010), S. 591–604.
- [77] J. C. Shelley u. a. „Epik: A software program for pK_a prediction and protonation state generation for drug-like molecules“. In: *Journal of Computer-Aided Molecular Design* 21.12 (2007), S. 681–691.
- [78] G. M. Sastry u. a. „Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments“. In: *Journal of Computer-Aided Molecular Design* 27.3 (2013), S. 221–234.

- [79] G. Neudert und G. Klebe. „DSX: A knowledge-based scoring function for the assessment of protein–ligand complexes“. In: *Journal of Chemical Information and Modeling* 51.10 (2011), S. 2731–2745.
- [80] L. Kaufman und P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. Wiley, 2005.
- [81] M. Maechler u. a. *cluster: Cluster analysis basics and extensions*. R package version 1.14.4. 2013.
- [82] L. Breiman. „Random Forests“. In: *Machine Learning* 45.1 (2001), S. 5–32.
- [83] V. Svetnik u. a. „Random Forest: A classification and regression tool for compound classification and QSAR modeling“. In: *Journal of Chemical Information and Computer Sciences* 43.6 (2003), S. 1947–1958.
- [84] Team R Core. *R: A language and environment for statistical computing*. Wien, Österreich, 2013.
- [85] A. Liaw und M. Wiener. „Classification and regression by RandomForest“. In: *R News* 2.3 (2002), S. 18–22.
- [86] R. Burbidge u. a. „Drug design by machine learning: Support Vector Machines for pharmaceutical data analysis“. In: *Computers & Chemistry* 26.1 (2001), S. 5–14.
- [87] P. J. Ballester und J. B. O. Mitchell. „A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking“. In: *Bioinformatics (Oxford, England)* 26.9 (2010), S. 1169–1175.
- [88] C. Kramer und P. Gedeck. „Leave-Cluster-Out Cross-Validation is appropriate for scoring functions derived from diverse protein data sets“. In: *Journal of Chemical Information and Modeling* 50.11 (2010), S. 1961–1969.
- [89] J. B. Dunbar u. a. „CSAR data set release 2012: Ligands, affinities, complexes, and docking decoys“. In: *Journal of Chemical Information and Modeling* (2013).
- [90] W. Sherman u. a. „Novel procedure for modeling ligand/receptor induced fit effects“. In: *Journal of Medicinal Chemistry* 49.2 (2006), S. 534–553.
- [91] W. Sherman, H. S. Beard und R. Farid. „Use of an induced fit receptor structure in virtual screening“. In: *Chemical Biology & Drug Design* 67.1 (2006), 83–84.
- [92] R. Farid u. a. „New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies“. In: *Bioorganic & Medicinal Chemistry* 14.9 (2006), S. 3160–3173.
- [93] A. Gaulton u. a. „ChEMBL: A large-scale bioactivity database for drug discovery“. In: *Nucleic Acids Research* 40.D1 (2011), S. D1100–D1107.

- [94] F. Milletti und A. Vulpetti. „Tautomer preference in PDB complexes and its impact on structure-based drug discovery“. In: *Journal of Chemical Information and Modeling* 50.6 (2010), S. 1062–1074.
- [95] F. Milletti u. a. „Tautomer enumeration and stability prediction for virtual screening on large chemical databases“. In: *Journal of Chemical Information and Modeling* 49.1 (2009), S. 68–75.
- [96] F. Milletti u. a. „New and original pK_a prediction method using grid molecular interaction fields“. In: *Journal of Chemical Information and Modeling* 47.6 (2007), S. 2172–2181.
- [97] K. L. Damm-Ganamet u. a. „CSAR benchmark exercise 2011–2012: Evaluation of results from docking and relative ranking of blinded congeneric series“. In: *Journal of Chemical Information and Modeling* (2013).
- [98] H. A. Carlson. „Check your confidence: size really *does* matter“. In: *Journal of Chemical Information and Modeling* 53.8 (2013), S. 1837–1841.
- [99] D. G. Bonett und T. A. Wright. „Sample size requirements for estimating Pearson, Kendall and Spearman correlations“. In: *Psychometrika* 65.1 (2000), S. 23–28.
- [100] N. Fechner u. a. „Estimation of the applicability domain of kernel-based machine learning models for virtual screening“. In: *Journal of Cheminformatics* 2 (2010), S. 2.
- [101] T. I. Netzeva u. a. „Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52“. In: *Alternatives to Laboratory Animals: ATLA* 33.2 (2005), S. 155–173.
- [102] S. Dimitrov u. a. „A stepwise approach for defining the applicability domain of SAR and QSAR models“. In: *Journal of Chemical Information and Modeling* 45.4 (2005), S. 839–849.
- [103] D. Stumpfe u. a. „Recent progress in understanding activity cliffs and their utility in medicinal chemistry“. In: *Journal of Medicinal Chemistry* (2013).
- [104] P. Iyer u. a. „Activity landscapes, information theory, and structure-activity relationships“. In: *Molecular Informatics* (2013), n/a–n/a.
- [105] J. Jiang u. a. „Structural basis of interdomain communication in the Hsc70 chaperone“. In: *Molecular Cell* 20.4 (2005), S. 513–524.
- [106] E. B. Bertelsen u. a. „Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate“. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.21 (2009), S. 8471–8476.

- [107] Q. Liu und W. A. Hendrickson. „Insights into Hsp70 chaperone activity from a crystal structure of the yeast Hsp110 Sse1“. In: *Cell* 131.1 (2007), S. 106–120.
- [108] D. S. Williamson u. a. „Novel adenosine-derived inhibitors of 70 kDa heat shock protein, discovered through structure-based design“. In: *Journal of Medicinal Chemistry* 52.6 (2009), S. 1510–1513.
- [109] G. Kragol u. a. „The antibacterial peptide pyrrocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding“. In: *Biochemistry* 40.10 (2001), S. 3016–3026.
- [110] Y. Umeda u. a. „Synthesis and antitumor activity of spergualin analogues. I. Chemical modification of 7-guanidino-3-hydroxyacyl moiety“. In: *The Journal of Antibiotics* 38.7 (1985), S. 886–898.
- [111] S. W. Fewell u. a. „Small molecule modulators of endogenous and co-chaperone-stimulated Hsp70 ATPase activity“. In: *Journal of Biological Chemistry* 279.49 (2004), S. 51131–51140.
- [112] D. Mamelak und C. Lingwood. „The ATPase domain of Hsp70 possesses a unique binding specificity for 3'-sulfogalactolipids“. In: *The Journal of Biological Chemistry* 276.1 (2001), S. 449–456.
- [113] J. Cellitti u. a. „Small molecule DnaK modulators targeting the β -domain“. In: *Chemical Biology & Drug Design* 74.4 (2009), S. 349–357.
- [114] J. I.-J. Leu u. a. „A small molecule inhibitor of inducible heat shock protein 70“. In: *Molecular Cell* 36.1 (2009), S. 15–27.
- [115] R. Wadhwa u. a. „Selective toxicity of MKT-077 to cancer cells is mediated by its binding to the Hsp70 family protein mot-2 and reactivation of p53 function“. In: *Cancer Research* 60.24 (2000), S. 6818–6821.
- [116] D. J. Propper u. a. „Phase I trial of the selective mitochondrial toxin MKT077 in chemo-resistant solid tumours“. In: *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 10.8 (1999), S. 923–927.
- [117] D. R. Williams u. a. „An apoptosis-inducing small molecule that binds to heat shock protein 70“. In: *Angewandte Chemie International Edition* 47.39 (2008), 7466–7469.
- [118] H. J. Cho u. a. „A small molecule that binds to an ATPase domain of Hsc70 promotes membrane trafficking of mutant cystic fibrosis transmembrane conductance regulator“. In: *Journal of the American Chemical Society* 133.50 (2011), S. 20267–20276.
- [119] *Molecular Operating Environment (MOE), 2009.10 ; Chemical Computing Group Inc.* 1010 Sherbooke St. West Suite 910 Montreal, Canada, H3A 2R7, 2013.

- [120] R. C. Wade und P. J. Goodford. „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“. In: *Journal of Medicinal Chemistry* 36.1 (1993), S. 148–156.
- [121] R. C. Wade, K. J. Clark und P. J. Goodford. „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“. In: *Journal of Medicinal Chemistry* 36.1 (1993), S. 140–147.
- [122] C. A. Reynolds, R. C. Wade und P. J. Goodford. „Identifying targets for bio-reductive agents: Using GRID to predict selective binding regions of proteins“. In: *Journal of Molecular Graphics* 7.2 (1989), S. 103–108, 100.
- [123] H. F. G. Velec, H. Gohlke und G. Klebe. „DrugScore^{CSD} - knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction“. In: *Journal of Medicinal Chemistry* 48.20 (2005), S. 6296–6303.
- [124] H. Gohlke, M. Hendlich und G. Klebe. „Predicting binding modes, binding affinities and ‘hot spots’ for protein-ligand complexes using a knowledge-based scoring function“. In: *Perspectives in Drug Discovery and Design* 20.1 (2000), S. 115–144.
- [125] H. Gohlke, M. Hendlich und G. Klebe. „Knowledge-based scoring function to predict protein-ligand interactions“. In: *Journal of Molecular Biology* 295.2 (2000), S. 337–356.
- [126] M. Rarey und J. S. Dixon. „Feature Trees: A new molecular similarity measure based on tree matching“. In: *Journal of Computer-Aided Molecular Design* 12.5 (1998), S. 471–490.
- [127] *FTreesXL, BioSolveIT, Germany*. <http://www.biosolveit.de/FTreesXL>. 2013.
- [128] B. Schaefer. „Computergestützte Untersuchungen zur Inhibition und Dynamik der β -Ketoacyl-ACP-Synthase I (KasA) aus *Mycobacterium tuberculosis*“. Diss. Universität Würzburg, 2012.
- [129] R. Burger u. a. „Gp130 and ras mediated signaling in human plasma cell line INA-6: a cytokine-regulated tumor model for plasmacytoma“. In: *The Hematology Journal: The official Journal of the European Haematology Association (EHA)* 2.1 (2001), S. 42–53.
- [130] A. Hartung. „Entwicklung neuartiger pharmakologischer Wirkstoffe als Inhibitoren des HSF-1/HSP70- Systems zur Behandlung des Multiplen Myeloms“. Diss. Universität Würzburg, 2013.
- [131] D. Case u. a. *Amber 10*. University of California, San Francisco, 2008.

- [132] D. A. Case u. a. „The amber biomolecular simulation programs“. In: *Journal of Computational Chemistry* 26.16 (2005), S. 1668–1688.
- [133] S. J. Weiner u. a. „A new force field for molecular mechanical simulation of nucleic acids and proteins“. In: *Journal of the American Chemical Society* 106.3 (1984), S. 765–784.
- [134] G. M. Balaburski u. a. „A modified Hsp70 inhibitor shows broad activity as an anticancer agent“. In: *Molecular Cancer Research* 11.3 (2013), S. 219–229.
- [135] U. Rothweiler u. a. „Isoquinolin-1-one inhibitors of the MDM2–p53 interaction“. In: *ChemMedChem* 3.7 (2008), S. 1118–1128.
- [136] R. Salomon-Ferrer u. a. „Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald“. In: *Journal of Chemical Theory and Computation* 9.9 (2013), S. 3878–3888.
- [137] A. W. Götz u. a. „Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born“. In: *Journal of Chemical Theory and Computation* 8.5 (2012), S. 1542–1555.
- [138] B. R. Brooks u. a. „CHARMM: a program for macromolecular energy, minimization, and dynamics calculations“. In: *Journal of Computational Chemistry* 4.2 (1983), 187–217.
- [139] M. Christen u. a. „The GROMOS software for biomolecular simulation: GROMOS05“. In: *Journal of Computational Chemistry* 26.16 (2005), S. 1719–1751.
- [140] J. D. Durrant und J. A. McCammon. „Molecular dynamics simulations and drug discovery“. In: *BMC Biology* 9.1 (2011), S. 71.
- [141] A. R. Leach. *Molecular modelling: Principles and applications*. Auflage: 2. A. Harlow, England ; New York: Prentice Hall, 2001.
- [142] M. P. Jacobson u. a. „A hierarchical approach to all-atom protein loop prediction“. In: *Proteins: Structure, Function, and Bioinformatics* 55.2 (2004), 351–367.
- [143] M. P. Jacobson u. a. „On the role of the crystal environment in determining protein side-chain conformations“. In: *Journal of Molecular Biology* 320.3 (2002), S. 597–608.
- [144] C. Bayly u. a. „A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model“. In: *The Journal of Physical Chemistry* 97.40 (1993), S. 10269–10280.
- [145] M. J. Frisch u. a. „Gaussian 03, Revision C.02“. In: (2008). 26251 Gaussian, Inc., Wallingford, CT, 2004.
- [146] J. Wang u. a. „Automatic atom type and bond type perception in molecular mechanical calculations“. In: *Journal of Molecular Graphics & Modelling* 25.2 (2006), S. 247–260.

- [147] J. Wang u. a. „Development and testing of a general amber force field“. In: *Journal of Computational Chemistry* 25.9 (2004), S. 1157–1174.
- [148] J. Srinivasan u. a. „Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects“. In: *Theoretical Chemistry Accounts* 101.6 (1999), S. 426–434.
- [149] C. Still u. a. „Semianalytical treatment of solvation for molecular mechanics and dynamics“. In: *Journal of the American Chemical Society* 112.16 (1990), S. 6127–6129.
- [150] W. L. Jorgensen u. a. „Comparison of simple potential functions for simulating liquid water“. In: *The Journal of Chemical Physics* 79.2 (1983), S. 926–935.
- [151] D. Case u. a. *Amber 11*. University of California, San Francisco, 2010.
- [152] J. C. Phillips u. a. „Scalable molecular dynamics with NAMD“. In: *Journal of Computational Chemistry* 26.16 (2005), 1781–1802.
- [153] T. Darden, D. York und L. Pedersen. „Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems“. In: *The Journal of Chemical Physics* 98.12 (1993), S. 10089–10092.
- [154] J.-P. Ryckaert, G. Ciccotti und H. J. C. Berendsen. „Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes“. In: *Journal of Computational Physics* 23.3 (1977), S. 327–341.
- [155] W. Humphrey, A. Dalke und K. Schulten. „VMD: visual molecular dynamics“. In: *Journal of Molecular Graphics* 14.1 (1996), S. 33–38, 27–28.
- [156] Schrödinger LLC. „The PyMOL Molecular Graphics System, Version 1.3r1“. 2010.
- [157] M. C. O'Brien, K. M. Flaherty und D. B. McKay. „Lysine 71 of the chaperone protein Hsc70 is essential for ATP hydrolysis“. In: *The Journal of Biological Chemistry* 271.27 (1996), S. 15874–15878.
- [158] M. Vogel, M. P. Mayer und B. Bukau. „Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker“. In: *Journal of Biological Chemistry* 281.50 (2006), S. 38705–38711.
- [159] J. F. Swain u. a. „Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker“. In: *Molecular Cell* 26.1 (2007), S. 27–39.
- [160] A. Zhuravleva und L. M. Gierasch. „Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones“. In: *Proceedings of the National Academy of Sciences* 108.17 (2011), S. 6987–6992.

- [161] F. Chiappori u. a. „Molecular mechanism of allosteric communication in Hsp70 revealed by molecular dynamics simulations“. In: *PLoS Computational Biology* 8.12 (2012), e1002844.
- [162] J. P. Schuermann u. a. „Structure of the Hsp110:Hsc70 nucleotide exchange machine“. In: *Molecular Cell* 31.2 (2008), S. 232–243.
- [163] R. Schlecht u. a. „Functional analysis of Hsp70 inhibitors“. In: *PLoS ONE* 8.11 (2013), e78443.