

Ein konsistentes chemieinformatisches Framework für automatisiertes virtuelles Screening¹

Sascha Urbaczek²

Abstract: Virtuelles Screening ist mittlerweile zu einem integralen Bestandteil der industriellen und akademischen Arzneimittelforschung geworden. Es wird eingesetzt, um sehr große Substanzdatenbanken mit der Hilfe von computerbasierten Methoden auf eine überschaubare Zahl vielversprechender Wirkstoffkandidaten zu reduzieren. Um dies zu erreichen, werden komplexe mehrstufige Arbeitsabläufe benötigt, die aktuell nur von Spezialisten durchgeführt werden können. Das Ziel der vorliegenden Arbeit war die Entwicklung einer zuverlässigen Basis für die Entwicklung von Software, die auch Medizinalchemikern den Zugang zur computergestützten Wirkstoffsuche ermöglicht. Das Ergebnis ist ein konsistentes chemieinformatische Software-Framework (NAOMI), das die anspruchsvollen Anforderungen eines solchen Anwendungsszenarios erfüllt.

1 Einleitung

Die computergestützte Wirkstoffsuche ist ein wichtiger Bestandteil der Entwicklung von Arzneimittelwirkstoffen und spielt insbesondere bei der Erzeugung und Auswahl vielversprechender Leitstrukturen eine zentrale Rolle. In der Literatur finden sich etliche Beispiele für ihren erfolgreichen Einsatz bei der Suche nach neuen Medikamenten [MS11]. Mit der zunehmenden Verbreitung hat jedoch auch die Kritik an der Qualität, Zuverlässigkeit und allgemeinen Anwendbarkeit computergestützter Methoden zugenommen, insbesondere in Hinblick auf das virtuelle Screening [Sc10, Ri10, Sc12b]. Die Vorhersagegenauigkeit ist noch bei weitem nicht ausreichend, um ohne aufwändige Analyse und Interpretation durch Experten verlässliche Ergebnisse zu produzieren. Und eine vollständige Automatisierung scheint in Anbetracht der Komplexität der Wirkstoffsuche sogar mehr oder weniger illusorisch.

Aus diesem Grund wurde angeregt, dass der zukünftige Erfolg des Feldes nicht nur von der Entwicklung neuer Methoden und Konzepte, sondern auch von der Fähigkeit zur Integration der Erfahrungen und Strategien aus der klassischen Medizinalchemie abhängen wird [Ku03]. Ein entscheidender Schritt in diese Richtung könnte sein, Medizinalchemiker zu befähigen, selbst stärker auf computergestützte Methoden zurückzugreifen [RM12]. Auf diese Weise könnten experimentell arbeitende Forscher zum Beispiel von einem besseren Verständnis für dreidimensionale Aspekte der Interaktionen von Protein und gebundenen Molekülen durch Visualisierung in Screening-Applikationen profitieren. Zudem bieten computergestützte Methoden ihnen die Werkzeuge, um ihre Hypothesen effizient zu formulieren und nachvollziehbar zu validieren. Das Ziel, Medizinalchemiker stärker in die

¹ Englischer Titel der Dissertation: A consistent cheminformatics framework for automated virtual screening

² Fakultät für Mathematik, Informatik und Naturwissenschaften der Universität Hamburg

computergestützte Wirkstoffsuche zu integrieren, wird jedoch das zukünftige Design chemieinformatischer Software wesentlich beeinflussen. Nach Ritchie sollten Anwendungsprogramme für diesen Zweck “well-thought-out, suitable for their needs, [and] able to generate useful, timely and valid results” sein [RM12].

Das Ziel der vorliegenden Arbeit war der Aufbau einer soliden Basis für die Entwicklung von Software, die eine Integration von Medizinalchemikern in die computergestützte Wirkstoffsuche erleichtert. Das Ergebnis dieser Bemühungen ist ein neues chemieinformatisches Software-Framework (NAOMI) für virtuelles Screening, welches speziell auf die damit verbundenen Anforderungen angepasst wurde. Erstens erlaubt das NAOMI-Framework Medizinalchemikern, basierend auf innovativen und intuitiv verständlichen Konzepten, ihre Erfahrung und ihr Expertenwissen genau an den Stellen der Berechnungen einzubringen, die für den Erfolg ihrer Projekte maßgeblich sind. Zweitens umfasst NAOMI zahlreiche neue chemieinformatische Methoden, die auf einem konsistenten internen chemischen Modell aufbauen und eine effiziente Ausführung der einzelnen Schritte des virtuellen Screenings erlauben. Die so erreichte Effizienz ist jedoch nicht nur eine notwendige Voraussetzung für Hochdurchsatz-Screening, sondern spielt auch eine zentrale Rolle bei interaktiven Anwendungen, die in Kombination mit einer intuitiven Benutzerschnittstelle eine Schlüsselrolle in der Integration von Medizinalchemikern in die computergestützte Wirkstoffsuche spielen. Drittens wurde viel Wert darauf gelegt sicherzustellen, dass die Ergebnisse der verschiedenen Rechenschritte chemisch sinnvoll sind und dass ein hoher Grad an Konsistenz zwischen den verschiedenen Komponenten des Prozesses sichergestellt ist. Dies ist von entscheidender Bedeutung, da viele Schritte automatisiert werden müssen, um zu erreichen, dass Medizinalchemiker sich auf die für sie relevanten Teilspekte konzentrieren können. Wie in den Publikationen dieser kumulativen Dissertation gezeigt wird, sind zahlreiche Methoden in NAOMI selbst relevante Beiträge in ihren jeweiligen Anwendungsgebieten und ihre Kombination erlaubt den Aufbau effizienter, stark automatisierter und hoch-adaptiver Screening-Prozesse.

Im Folgenden werden die konzeptionellen und algorithmischen Beiträge des NAOMI-Frameworks zum Feld des computergestützten Wirkstoffentwurfs vorgestellt. Diese beinhalten die zuverlässige Interpretation von Molekülstrukturen aus chemischen Dateiformaten, Methoden zum Prozessieren, Speichern und Durchsuchen von großen Molekülmengen, die Vorhersage und Bewertung von intermolekularen Wechselwirkungen im Kontext von Protein-Ligand-Komplexen und die Identifikation vielversprechender chemischer Strukturen mit Hilfe eines index-basierten Docking-Ansatzes. In den folgenden drei Abschnitten, welche die Phasen des strukturbasierten Wirkstoffentwurfs widerspiegeln, werden die Ansätze des NAOMI-Frameworks zur Lösung spezifischer Probleme aus der jeweiligen Phase beschrieben. Der Text schließt mit einer Zusammenfassung und einem Ausblick auf mögliche Erweiterungen und zukünftige Anwendungen.

2 Screening-Bibliothek

Die Erstellung einer fokussierten Bibliothek ist der erste Schritt eines virtuellen Screenings. Den Ausgangspunkt bildet oftmals eine große Moleküldatenbank, aus der eine

Teilmenge mit für die jeweilige Fragestellung geeigneten Eigenschaften ausgewählt wird. Hierbei können zahlreiche Aspekte, wie beispielweise physiko-chemische und strukturelle Eigenschaften, kommerzielle Verfügbarkeit oder die aktuelle Patentlage, eine wesentliche Rolle spielen. Um die Laufzeiten der Berechnungen zu verringern und unnötigen Aufwand bei der Analyse der Resultate zu vermeiden, enthält die Screening-Bibliothek im optimalen Falle nur solche Moleküle, die auch tatsächlich als Wirkstoffe in Frage kommen. Neben diesen allgemeinen Bedingungen, die auch im experimentellen Kontext ihre Gültigkeit haben, existieren jedoch zusätzlich noch weitere, für das computergestützte Screening spezifische, Aspekte. Zum einen muss die Qualität der Molekül Daten sichergestellt werden, da chemisch invalide Strukturen letztendlich zu falschen und somit nutzlosen Ergebnissen führen. Zum anderen ist aus Gründen der Effizienz und Konsistenz oftmals eine Normalisierung der chemischen Strukturen mit anschließender Abspeicherung in chemischen Datenbanken sinnvoll. Abhängig vom gewählten Screening-Ansatz kann es zudem noch notwendig sein, zusätzliche Eigenschaften, z.B. dreidimensionale Atomkoordinaten oder Molekülkonformationen, zu berechnen.

In der NAOMI-Publikation³ wurde das hierarchische chemische Modell eingeführt, das den Kern des NAOMI-Frameworks bildet. Es wurde speziell für die effiziente Prozessierung von im Kontext der Wirkstoffsuche relevanten, d.h. organischen, Molekülen entwickelt. Eine der wichtigsten Aufgaben ist die Bereitstellung aller benötigten strukturellen Informationen und chemischen Deskriptoren für die Entwicklung und Implementierung unterschiedlicher chemieinformatischer Methoden und Algorithmen. Obwohl es keinen direkten Weg gibt, die chemischen Modelle verschiedener chemieinformatischer Software-Bibliotheken hinsichtlich Qualität untereinander zu vergleichen, zeigen die auf der Konvertierung von chemischen Dateiformaten basierenden Untersuchungen aus der Publikation zumindest indirekt, dass das chemische Modell des NAOMI-Frameworks robuster als das anderer Frameworks und insbesondere in Bezug auf die Interpretation und Konvertierung chemischer Dateiformate hochgradig konsistent ist. Die Integrität der strukturellen Daten kann selbst bei Verwendung unterschiedlicher Repräsentationen und multiplen Konvertierungsschritten zuverlässig gewährleistet werden. Zudem zeichnen sich die zugrundeliegenden Algorithmen durch eine höhere Effizienz gegenüber vergleichbaren Methoden aus. In Hinblick auf das virtuelle Screening ist die Prozessierung von Informationen aus chemischen Dateiformaten ein notwendiger erster Schritt, der gewöhnlich für große Datenmengen durchgeführt werden muss. Das bedeutet, dass die zugehörigen Prozesse neben einer hohen Effizienz auch einen hohen Automatisierungsgrad aufweisen müssen und zudem sichergestellt sein muss, dass die chemische Information korrekt interpretiert und zuverlässig an darauffolgende Verfahrensschritte propagiert wird. Das NAOMI-Framework lehnt deshalb alle gemäß des internen Modells invaliden Strukturen ab, um die akkurate Beschreibung der zugrundeliegenden Moleküle zu jedem Zeitpunkt sicherzustellen. Darüber hinaus wurden verschiedene Korrekturmechanismen entwickelt, welche die Arbeit mit inkonsistenten Daten bis zu einem gewissen Grade ermöglichen. Insbesondere die letzten beiden Aspekte sind zentral, wenn nicht mit stark kurierten Datensätzen gearbeitet wird.

³ Urbaczek, S.; Kolodzik, A.; Fischer, J.R.; Lippert, T.; Heuser, S.; Groth, I.; Schulz-Gasch, T.; Rarey, M.: NAO-MI: On the Almost Trivial Task of Reading Molecules from Different File Formats. *J. Chem. Inf. Model.*, 51(12):3199-3207, 2011.

In der URF-Publikation⁴ wurde das Konzept der Unique Ring Families (URF) als Grundlage für die Identifikation und Beschreibung von Ringsystemen und Ringen innerhalb des NAOMI-Frameworks beschrieben. Ringe sind ein allgegenwärtiges Strukturmotiv in Molekülen und beeinflussen deren physiko-chemischen Eigenschaften maßgeblich. Im chemieinformatischen Kontext sind insbesondere die Anzahl der Ringe in einem Molekül sowie die Ringzugehörigkeit einzelner Atome für zahlreiche Anwendungen von großer Bedeutung. Aus diesem Grund wurde eine Vielzahl unterschiedlicher Konzepte und Algorithmen zur Beschreibung und Bestimmung von Ringen in Molekülen entwickelt, deren Vielfalt sich aus der fehlenden Eindeutigkeit bei der Identifikation von Zyklen in Graphen und der daraus resultierenden Notwendigkeit, heuristische Kriterien für chemisch sinnvolle Ringe zu definieren, erklärt. Letztere wiederum sind, wie oftmals in chemieinformatischen Anwendungen, ein Kompromiss zwischen chemischer Intuition und technischer Realisierbarkeit und hängen stark vom Anwendungskontext ab. Die innerhalb des NAOMI-Frameworks verwendeten URF sind der erste veröffentlichte Ringdeskriptor, der die drei wichtigsten Kriterien für chemisch sinnvolle Ringe gleichzeitig erfüllt. Erstens verhalten sich die URF in Bezug auf Größe polynomiell, d.h. die Anzahl der Ringe wächst nicht exponentiell mit der Zahl der Atome. Zweitens sind die URF eindeutig, so dass der Deskriptor unabhängig von dem zur Berechnung verwendeten Algorithmus und der Struktur der Input-Daten ist. Drittens entsprechen die URF, im Gegensatz zu zahlreichen anderen Konzepten, insbesondere in Hinblick auf die Ringzugehörigkeit einzelner Atome sehr stark der chemischen Intuition. In der Publikation wurde zudem ein Algorithmus beschrieben, mit dem die URF in polynomieller Zeit berechnet werden können, so dass, wie in der Validierung der Methode gezeigt, selbst für sehr komplizierte Fälle keine Probleme hinsichtlich der Laufzeit auftreten.

Das chemische Datenbanksystem des NAOMI-Frameworks, die MolDB, wurde gemeinsam mit einem interaktiven Tool zur Bearbeitung von Moleküldatensätzen (MONA) in der MONA-Publikation⁵ vorgestellt. Das zentrale Konzept der MolDB ist die Molekülmenge (Molecule Set), die als Menge von in Hinblick auf ihre Topologie paarweise unterschiedlichen Molekülen definiert ist. Molekülmengen stellen gewöhnlich Untermengen der gesamten Moleküldatenbank dar und werden durch Anwendung spezieller Filterkombinationen zusammengestellt. Die MolDB implementiert verschiedene Filtertypen unterschiedlicher Komplexität sowie Operationen zur Erstellung und Manipulation von Molekülmengen. Die Zusammenstellung von Screening-Bibliotheken auf Basis von Molekülmengen ist ein sehr intuitives Konzept und stellt somit auch für Medizinalchemiker keine Hürde dar. Die Operationen der MolDB sind zudem sehr effizient, so dass interaktive Abläufe auch mit mehreren Millionen Molekülen problemlos realisiert werden können. Dies ist ein sehr wichtiger Faktor, da in vielen Situationen die optimale Kombination von Filtern nicht von Anfang an bekannt ist, sondern basierend auf den Ergebnissen des letzten Schrittes iterativ erarbeitet werden muss. Hierbei sind lange Wartezeiten zwischen den Operationen hinderlich. MONA als Applikation stellt genau diese Funktionalität bereit und kann so-

⁴ Kolodzik, A.; Urbaczek, S.; Rarey, M.: Unique Ring Families: A Chemically Meaningful Description of Molecular Ring Topologies. *J. Chem. Inf. Model.*, 52(8):2013-2021, 2012.

⁵ Hilbig, M.; Urbaczek, S.; Heuser, S.; Groth, I.; Rarey, M.: MONA - Interactive Manipulation of Molecule Collections. *J. Cheminform.*, 5(1):38, 2013.

mit als Beispiel dienen, wie das NAOMI-Framework zur Implementierung effizienter und interaktiver Software für Medizinalchemiker verwendet werden kann.

Die Konzepte und Algorithmen des NAOMI-Frameworks zur Behandlung von Tautomeren und Protonierungszuständen, zusammenfassend als Protomere bezeichnet, wurden im Rahmen der VSC-Publikation⁶ vorgestellt. Nach Sayle [Sa10] bestehen im chemieinformatischen Kontext in Bezug auf Protomere fünf wesentliche Aufgaben, nämlich Vergleich, Normalisierung, Enumeration, Auswahl und Vorhersage. Die ersten beiden Themenkomplexe sind eng miteinander verwoben, da die Erstellung einer normalisierten Darstellung oftmals auch den ersten Schritt für den Vergleich zweier Moleküle darstellt. Mit Enumeration ist die bloße Generierung aller (theoretisch) möglichen Protomere gemeint, wohingegen Auswahl und Vorhersage eine Bewertung der resultierenden Molekülstrukturen, z.B. in Bezug auf Stabilität, miteinschließen. Auswahl bedeutet die Einschränkung der (theoretisch) möglichen Gesamtmenge auf eine für den aktuellen Kontext relevante, aber ungeordnete, Untermenge. Vorhersage beinhaltet dann auch die Bestimmung der Stabilitätsordnung, und somit der Verhältnisse der Strukturen untereinander. Die Basis für die Behandlung von Protomeren innerhalb des NAOMI-Frameworks bildet das Valence State Combination (VSC)-Modell. Dieses adressiert, im Gegensatz zu vielen anderen Konzepten, durch Berücksichtigung von Mesomerie, Tautomerie und Ionisierung konsequent und konsistent alle relevanten Dimensionen des Problems und ermöglicht somit eine vollständige Systematisierung der möglichen Protomere von Molekülen. Auf Basis des VSC-Modells wurden für die weiter oben erwähnten fünf Aufgaben Lösungen innerhalb des NAOMI-Frameworks entwickelt. Der Nachweis der hohen Konsistenz dieser Lösungen, die sich insbesondere darin ausdrückt, dass die Ergebnisse unabhängig davon sind, welches Protomer als Input verwendet wird, wurde anhand der verschiedenen in der Publikation vorgestellten Validierungsprozeduren erbracht. Zudem konnte auf Basis kurrer Datensätze eine sehr gute Übereinstimmung zwischen den von Experten ausgewählten Protomerformen und den Vorhersagen auf Basis des internen Bewertungsschemas gezeigt werden.

3 Protein-Struktur

Die Vorhersage eines realistischen Bindungsmodus mit Hilfe eines Docking-Algorithmus setzt die detaillierte Kenntnis der Bindetasche des Proteins auf atomarer Ebene voraus. Die dreidimensionale Struktur des Zielproteins ist somit, neben der Screening-Bibliothek, die zweite notwendige Voraussetzung für die Durchführung eines virtuellen Screenings. Die bei weitem wichtigste Quelle für experimentell bestimmte Strukturinformationen von Proteinen und Protein-Ligand-Komplexen ist die Protein Data Bank (PDB) [Be00] mit aktuell mehr als 90000 Einträgen. Die dort bereitgestellten Daten lassen sich jedoch im Allgemeinen nicht direkt für virtuelle Screenings einsetzen, sondern müssen durch verschiedene Arbeitsschritte teilweise sehr aufwendig aufbereitet werden. Die Strukturdaten in der PDB werden in Form von dreidimensionalen Atomkoordinaten unter Verwendung eines speziellen chemischen Dateiformats zur Verfügung gestellt. Um diese Daten für den virtuellen Wirkstoffentwurf nutzbar zu machen, werden Methoden benötigt, die zuverlässig

⁶ Urbaczek, S.; Kolodzik, A.; Rarey, M.: The Valence State Combination Model: A Generic Framework for Handling Tautomers and Protonation States. *J. Chem. Inf. Model.*, 54(3):756-766, 2014.

Molekülstrukturen aus diesen dreidimensionalen Atomkoordinaten ableiten können. Im Gegensatz zu den weiter oben beschriebenen Methoden zur Prozessierung von Molekülen einer Screening-Bibliothek, müssen bei der Interpretation von Daten aus der PDB immer auch experimentelle Ungenauigkeiten berücksichtigt werden.

In der PDB-Publikation⁷ wurde der Algorithmus zur Ableitung von Molekülstrukturen aus dreidimensionalen Atomkoordinaten des NAOMI-Frameworks vorgestellt. Die Evaluierung der Ergebnisse erfolgte anhand des Vergleichs der von der Methode aus PDB-Daten abgeleiteten Moleküle mit bekannten Referenzstrukturen. Im Rahmen einer umfassenden Analyse konnte gezeigt werden, dass der Algorithmus in der großen Mehrheit der Fälle selbst bei stark verzerrten Geometrien das erwartete Ergebnis liefert. Die Ursachen von Abweichungen wurden detailliert untersucht und konnten in den meisten Fällen auf experimentellen Ungenauigkeiten in den Atomkoordinaten zurückgeführt werden, die sich nicht eindeutig auflösen ließen. Weiterhin wurde der Algorithmus mit anderen bereits publizierten Methoden verglichen. Hierbei konnte gezeigt werden, dass durch Rückgriff auf das robuste chemische Modell des NAOMI-Frameworks Fehler anderer Methoden vermieden werden konnten. Zudem zeichnet sich das NAOMI-Framework auch bei der Prozessierung von PDB-Daten wieder durch eine hohe Effizienz im Vergleich zu anderen Frameworks aus.

Ein weiterer zentraler Aspekt bei der Aufbereitung von Strukturinformationen aus der PDB ist die Berechnung von Koordinaten für Wasserstoffatome, da diese im Normalfall experimentell nicht bestimmt werden können. In der Protoss-Publikation⁸ wurde die Methode zur Bestimmung von Wasserstoffposition innerhalb von Protein-Ligand-Komplexen, die auf einer früheren Arbeit aus der Arbeitsgruppe [LR09] aufbaut, beschrieben. Die wesentliche Erweiterung besteht vor allem in der umfassenden Berücksichtigung von Protomeren auf Seite des Liganden sowie der Verwendung des NAOMI-Frameworks zur konsistenten Beschreibung von Molekülen. Das Ziel der Erweiterung war, die relativen Stabilitäten der beteiligten chemischen Gruppen bei der Vorhersage des optimalen Wasserstoffbrückennetzwerks zu berücksichtigen. Protoss ist somit ein Beispiel für die von Sayle als Vorhersage bezeichnete Aufgabe im Kontext von Protein-Ligand-Komplexen. Ziel ist eine Menge von relevanten Protomeren zu erzeugen, aus denen dann die wahrscheinlichste Form basierend auf Stabilitätsbetrachtungen ausgewählt wird. Die umfassende Berücksichtigung von Protomeren bei der Identifikation des optimalen Wasserstoffbrückennetzwerks in Protein-Ligand-Komplexen stellt ein Alleinstellungsmerkmal des Protoss-Ansatzes dar. Die prinzipielle Notwendigkeit der Berücksichtigung derartiger Effekte konnte anhand einer umfassenden Analyse der verschiedenen tautomerisierbaren und ionisierbaren Substrukturen von Molekülen innerhalb der PDB nachgewiesen werden. Die hohe Qualität der Vorhersage wurde zudem durch verschiedene Experimente belegt und mit den Ergebnissen anderer Methoden verglichen. Hierzu wurde die Anzahl der unerwünschten Interaktionen, d.h. die räumliche Überlappung von Wasserstoffatomen oder Interaktion zweier Akzeptor-Atome, sowie die Abweichung von der erwarteten Protomerform her-

⁷ Urbaczek, S.; Kolodzik, A.; Heuser, S.; Groth, I.; Rarey, M.: Reading PDB: Perception of Molecules from 3D Atomic Coordinates. *J. Chem. Inf. Model.*, 53(1):76-87, 2013.

⁸ Bietz, S.; Urbaczek, S.; Schulz, B.; Rarey, M.: Protoss: A Holistic Approach to Predict Tautomers and Protonation States in Protein-Ligand Complexes. *J. Cheminform.*, 6(1):12, 2014.

angezogen. In beiden Hinsichten übertrifft Protoss existierende Ansätze und zeichnet sich zudem, wie zahlreiche Methoden des NAOMI-Frameworks, durch eine hohe Effizienz aus.

4 Virtuelles Screening

Wie weiter oben bereits erwähnt wurde, stellen Screening-Bibliothek und aufbereitete Proteinstruktur die Voraussetzungen für das strukturbasierte virtuelle Screening dar. Den methodischen Kern des Ansatzes bildet das Docking, eine Methode die verwendet wird, um die Art wie Moleküle an Proteine binden, den sogenannten Bindungsmodus, vorherzusagen und energetisch zu bewerten. Das zugrundeliegende algorithmische Problem ist eine räumliche Ausrichtung des Moleküls zu finden, die geometrisch in die Bindetasche des Proteins passt und die zudem einer bioaktiven Konformation, d.h. der tatsächlichen Konformation des Liganden im Bindungszustand, entspricht. Dies setzt die gleichzeitige Berücksichtigung von Translation und Rotation innerhalb der Bindetasche, der inhärenten konformationellen Flexibilität von Molekülstrukturen sowie der Effekte, die Interaktionen zwischen Proteinen und den gebundenen Liganden bestimmen, voraus. Diese enorme Vielfalt an Einflussfaktoren macht Docking zu einem hochgradig komplexen Problem.

Die im Rahmen der cRAISE-Publikation⁹ beschriebene Docking-Engine des NAOMI-Frameworks ist eine Weiterentwicklung des TrixX-Ansatzes [SR07, SR09] aus der gleichen Arbeitsgruppe. Die zugrundeliegende Technologie baut auf einer deskriptorbasierten Bitmap-Suche auf und trägt den Namen RAISE (Rapid Index-based Screening Engine). Kern der Methode sind die sogenannten RAISE-Deskriptoren, die für Molekül und Bindetasche berechnet werden und deren Vergleich die Feststellung von räumlicher und chemischer Komplementarität erlaubt. Durch den Einsatz von Bitmap Indices beim Vergleich der Deskriptoren kann eine extrem hohe Effizienz erreicht werden. Eine wichtige Eigenschaft der RAISE-Technologie ist die Unterstützung sogenannter Pharmakophore. Diese beschreiben spezielle Anforderungen an den Bindungsmodus, z.B. bestimmte Atome des Proteins, die in Wechselwirkung mit dem Liganden treten müssen, oder bestimmte Regionen in der Bindetasche, in denen der Ligand nicht liegen darf. cRAISE ist in der Lage, derartige Randbedingungen direkt während der Platzierung des Liganden in der Bindetasche, und nicht nachgelagert, zu berücksichtigen und somit eine höhere Effizienz durch frühzeitiges Verwerfen ungeeigneter Lösungen zu erreichen. Neben der Steigerung der Geschwindigkeit spielen Pharmakophore jedoch auch bei der Integration von Medizinalchemikern eine wichtige Rolle, da sie eine Möglichkeit bieten, den Screening-Prozess auf Basis der eigenen Erfahrung aktiv zu steuern.

Die RAISE-Technologie und die darauf aufbauende Docking-Methodik (cRAISE) basieren vollständig auf dem NAOMI-Framework und profitieren von der konsistenten Beschreibung von Molekülen und Proteinen. Obwohl RAISE die grundlegenden Prinzipien des TrixX-Ansatzes adaptiert hat, bestehen doch beträchtliche Unterschiede vor allem in Hinblick auf die interne Modellierung von Interaktionen zwischen Protein und Ligand. Die resultierende RAISE-Screening-Pipeline kombiniert alle Methoden und Konzepte, die

⁹ Henzler, A.M.; Urbaczek, S.; Hilbig, M.; Rarey, M.: An Integrated Approach to Knowledge-Driven Structure-Based Virtual Screening. *J. Comput.-Aided Mol. Des.*, 28(9):927-939, 2014.

im Rahmen der Dissertation entwickelt wurden, und ermöglicht einen hochgradig effizienten Screening-Workflow, der zu sehr großen Teilen automatisch ablaufen kann, aber Intervention und Steuerung an verschiedenen zentralen Stellen zulässt.

Die vorgestellten Modelle, Konzepte und Methoden eignen sich jedoch nicht nur für den Einsatz im Kontext des strukturbasierten Screenings. In den iRAISE-¹⁰ und COMPASITE-Publikationen¹¹ werden alternative Anwendungen vorgestellt. Erstere beschreibt einen Ansatz für inverses virtuelles Screening, letztere eine Methode zum Vergleich von Proteinbindetaschen. Beide Beispiele belegen die große Anwendungsbreite von RAISE-Technologie und NAOMI-Framework.

5 Zusammenfassung und Ausblick

Im Rahmen der vorgestellten Dissertation [Ur] wurde ein konsistentes chemieinformatische Software-Framework (NAOMI) entwickelt. NAOMI basiert auf einem robusten chemischen Modell und umfasst eine Vielzahl innovativer Konzepte und Methoden, die zusammen die Basis der RAISE-Screening-Pipeline bilden. Jede einzelne Komponente wurde mit dem Ziel entwickelt, Screening-Berechnung auf eine intuitive und nachvollziehbare Art und Weise durchzuführen und anschließend auf ihre Eignung zur Umsetzung prinzipiell vollständig automatisierter Screening-Anwendungen hin untersucht. In mehreren Evaluierungsstudien wurde gezeigt, dass die zugrundeliegenden Modelle und Algorithmen gleichzeitig effizient und verlässlich sind und so eine ausgewogene Kombination von automatisierten und interaktiven Schritten möglich machen. Dies ist wiederum eine wichtige Voraussetzung für die Implementierung von flexiblen Screening-Workflows für unterschiedliche Anwendungsbereiche.

Die neu entwickelten Methoden stellen jeweils signifikante Verbesserungen gegenüber existierenden Methoden aus den jeweiligen Bereichen dar. Dies betrifft im Besonderen die Umwandlung chemischer Dateiformate, die Interpretation von Molekülen auf der Grundlage von dreidimensionalen Koordinaten, die Normalisierung von Molekülstrukturen und die Generierung von Protomeren. Zusätzlich sind viele der vorgestellten Konzepte wichtige Beiträge zu einer nachhaltigen Lösung von chemieinformatischen Problemen, wie zum Beispiel die Repräsentation von Ringen durch URF und die Behandlung von Protomeren auf der Grundlage des VSC-Modells. Sowohl die Datenbank für Moleküle (MolDB) als auch die Erweiterung für Proteine (ProteinDB) erlauben eine effiziente Speicherung und Prozessierung der jeweiligen Strukturen und stellen damit die Basis für interaktive Funktionalität im Rahmen von Screening-Workflows dar. Die Berücksichtigung von Protomeren bei der Bewertung und Vorhersage von Wasserstoffbrückennetzwerken in Protein-Ligand-Komplexen sowie die Integration der Methode in den Docking-Prozess¹² stellen ein Alleinstellungsmerkmal des NAOMI-Frameworks dar. Die RAISE-Technologie ist

¹⁰ Schomburg, K.T.; Bietz, S.; Briem, H.; Henzler, A.M.; Urbaczek, S.; Rarey, M.: Facing the Challenges of Structure-based Target Prediction by Inverse Virtual Screening. *J. Chem. Inf. Model.*, 54(6):1676-1686, 2014.

¹¹ v. Behren, M.; Volkamer, A.; Henzler, A.M.; Schomburg, K.T.; Urbaczek, S.; Rarey, M.: Fast protein binding site comparison via an index-based screening technology. *J. Chem. Inf. Model.*, 53(2):411-422, 2013.

¹² Bisher noch nicht veröffentlicht

die Grundlage für eine sehr effiziente Screening-Engine, die zudem eine Formulierung von intuitiven pharmakophoren Beschränkungen unterstützt. Die zusätzlichen RAISE-Anwendungen, inverses Screening und der Vergleich von Bindetaschen, verdeutlichen das Potential des Ansatzes zur Lösung verschiedener chemieinformatischer Probleme.

Obwohl der aktuelle Stand der Pipeline einen wichtigen Meilenstein für die Entwicklung interaktiver Screening-Workflows darstellt, gibt es noch immer genügend Raum für Verbesserungen. Zum einen könnte die Qualität der Docking-Ergebnisse durch eine nachgelagerte Optimierungsprozedur basierend auf der in der gleichen Arbeitsgruppen entwickelten HYDE-Methode [Sc12a] verbessert werden. Dies wurde zwar auf Software-Ebene schon umgesetzt, die notwendige Evaluierung steht jedoch noch aus. In Hinblick auf die hohe Effizienz des RAISE-Ansatzes wäre die Berücksichtigung von Proteinflexibilität, ein bisher immer noch ungelöstes Problem im Bereich des virtuellen Screenings, eine weitere mögliche Anwendung. Zudem ist die RAISE-Technologie nicht auf strukturbasierte Ansätze beschränkt und könnte auch als Basis für die Entwicklung von ligandbasierten Verfahren dienen. Natürlich bietet auch das NAOMI-Framework durch seine hohe Konsistenz und Effizienz einen guten Ausgangspunkt für die Entwicklung neuer und verbesserter chemieinformatischer Methoden. Als Beispiele seien hier neue Konzepte für die Visualisierung von Molekülmengen auf Basis von Ähnlichkeiten oder Methoden aus dem Bereich der Fragmenträume genannt.

Ein weiteres wichtiges Ziel der Entwicklung des NAOMI-Frameworks war die Schaffung einer Basis für die Realisierung von Software-Applikationen, die auch von Medizinalchemikern genutzt werden können. MONA sowie das von der Firma BioSolveIT entwickelte Tool SeeSAR¹³ sollen als Belege dienen, dass auch dieses Ziel erreicht werden konnte.

Literaturverzeichnis

- [Be00] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E.: The Protein Data Bank. *Nucleic Acids Res.*, 28(1):235–242, 2000.
- [Ku03] Kubinyi, H.: Drug Research: Myths, Hype and Reality. *Nat. Rev. Drug. Discov.*, 2(8):665–668, 2003.
- [LR09] Lippert, T.; Rarey, M.: Fast Automated Placement of Polar Hydrogen Atoms in Protein-Ligand Complexes. *J. Cheminform.*, 1(1):13, 2009.
- [MS11] Matter, H.; Sotriffer, C.: Applications and Success Stories in Virtual Screening. In (Sotriffer, C., Hrsg.): *Virtual Screening - Principles, Challenges, and Practical Guidelines*, Jgg. 48 in *Methods and Principles in Medicinal Chemistry*, S. 319–358. Wiley-VCH, Weinheim, 2011.
- [Ri10] Ripphausen, P.; Nisius, B.; Peltason, L.; Bajorath, J.: Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications. *J. Med. Chem.*, 53(24):8461–8467, 2010.
- [RM12] Ritchie, T.J.; McLay, I.M.: Should Medicinal Chemists do Molecular Modelling? *Drug Discov. Today*, 17(11-12):534–537, 2012.

¹³ <https://www.biosolveit.de/SeeSAR/>

- [Sa10] Sayle, R.A.: So you think you understand tautomerism? *J. Comput.-Aided Mol. Des.*, 24(6-7):485–496, 2010.
- [Sc10] Schneider, G.: Virtual Screening: An Endless Staircase? *Nat. Rev. Drug Discov.*, 9(4):273–276, 2010.
- [Sc12a] Schneider, N.; Hindle, S.; Lange, G.; Klein, R.; Albrecht, J.; Briem, H.; Beyer, K.; Clausen, H.; Gastreich, M.; Lemmen, C.; Rarey, M.: Substantial Improvements in Large-scale Redocking and Screening Using the Novel HYDE Scoring Function. *J. Comput.-Aided Mol. Des.*, 26(6):701–723, 2012.
- [Sc12b] Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J.L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K. and Agrafiotis, D. K.: Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.*, 52(4):867–881, 2012.
- [SR07] Schellhammer, I.; Rarey, M.: TrixX: Structure-Based Molecule Indexing for Large-Scale Virtual Screening in Sublinear Time. *J. Comput.-Aided Mol. Des.*, 21(5):223–238, 2007.
- [SR09] Schlosser, J.; Rarey, M.: Beyond the Virtual Screening Paradigm: Structure-Based Searching for New Lead Compounds. *J. Chem. Inf. Model.*, 49(4):800–809, 2009.
- [Ur] Urbaczek, S.: A Consistent Cheminformatics Framework for Automated Virtual Screening. Dissertation, Universität Hamburg, Juni 2015.



Sascha Urbaczek hat sein Studium der Chemie an der Universität Heidelberg mit dem Diplom abgeschlossen. Während des Grundstudiums erhielt er das Jubiläumsspendium des Verbandes der chemischen Industrie und war anschließend Stipendiat der Studienstiftung des deutschen Volkes. Nach einer Diplomarbeit im Bereich der theoretischen Chemie wechselte er im Sommer 2008 an die Universität Hamburg, um dort eine Dissertation im Bereich Chemieinformatik anzufertigen. Im Anschluß daran entwickelte er die im Rahmen der Promotionszeit entstandene Software für den industriellen Einsatz bei der Beiersdorf AG und der BioSolveIT GmbH weiter. Seit Mitte 2014 arbeitet er als IT Unternehmensberater bei der Platinion GmbH und beschäftigt sich schwerpunktmäßig mit Analyse, Entwurf und Umsetzung von anspruchsvollen IT-Lösungen für Unternehmen.