

# **TrixX – Strukturbasierte virtuelle Suche in Wirkstoffdatenbanken in sublinearer Zeit**

Ingo Schellhammer

Abteilung für algorithmisches molekulares Design  
Zentrum für Bioinformatik Hamburg  
Bundesstraße 43  
20146 Hamburg  
Ingo\_Schellhammer@web.de

**Abstract:** Virtuelles Screening ist zu einem integralen und bedeutsamen Bestandteil der modernen Wirkstoffforschung geworden. Strukturbasierte virtuelle Screening-Verfahren analysieren die dreidimensionale Struktur des Zielmoleküls, dessen Funktion beeinflusst werden soll, um in großen virtuellen Screening-Bibliotheken nach Molekülen zu suchen, die hinsichtlich des Zielmoleküls biologisch aktiv sind. In den letzten Jahren haben die Anzahl virtueller Screening-Experimente sowie die durchschnittliche Größe der Screening-Bibliotheken stetig zugenommen. Pharmazeutische Forschungseinrichtungen sind dem hohen Rechenbedarf durch Einsatz moderner Computer-Hardware und massiver Parallelisierung begegnet. Softwareseitig wurden jedoch kaum Fortschritte bei der Effizienz der eingesetzten Algorithmen gemacht. Heutige strukturbasierte Screening-Werkzeuge basieren in der Regel auf sequenziellen Verarbeitungsstrategien, bei denen alle Moleküle einer Screening-Bibliothek nacheinander bewertet werden, so dass sich insgesamt eine lineare Laufzeit in den Anzahl der zu untersuchenden Moleküle ergibt. Im Rahmen dieser Dissertation wurde ein neuartiges Paradigma für die strukturbasierte Wirkstoffsuche entwickelt, welches erstmals eine sublineare Laufzeit in der Anzahl der zu untersuchenden Moleküle ermöglicht. Dieses Paradigma wurde prototypisch in einem Softwarewerkzeug (TrixX) umgesetzt und erfolgreich in Screening-Experimenten für elf Zielproteine mit hoher therapeutischer Relevanz angewendet. TrixX ist bis zu 60-mal schneller als das sehr verbreitete molekulare Docking-Werkzeug FlexX, welches in der Arbeit als Vergleichsmaßstab verwendet wurde. Im Gegensatz zur sequentiellen Verarbeitung aller Moleküle bei bestehenden Verfahren besteht das in TrixX umgesetzte Paradigma darin, zunächst die Struktur und die physikochemischen Eigenschaften der Bindetasche zu analysieren und dann gezielt Moleküle mit dazu komplementären Eigenschaften aus einem Katalog zu identifizieren. Der Katalog ordnet alle Moleküle eines Screening-Experiments nach ihren geometrischen und physikochemischen Eigenschaften. Der Katalog ist in einem relationalen Datenbanksystem mit indizierten Zugriffstabellen realisiert, um effiziente Suchanfragen zu ermöglichen. Zentrales Element des Katalogs ist ein hochselektiver Moleküldeskriptor, der die funktionellen Gruppen des Moleküls, ihren paarweisen Abstand, ihre bevorzugten Wechselwirkungsrichtungen sowie die Lage sterischer Masse beschreibt. Bis auf die Lage sterischer Masse werden die Moleküle nach all diesen Eigenschaften in einem  $B^+$ -Baum sortiert.

## 1 Einleitung

Ein wichtiges Feld der Chemieinformatik ist die Entwicklung computergestützter Verfahren für die Beschreibung von Molekülen und die Vorhersage ihrer biologischen Aktivität. Diese Verfahren kommen zum Beispiel in der pharmazeutischen Forschung zum Einsatz. Ausgangspunkt ist das so genannte pharmazeutische Target, ein Protein oder ein Rezeptor, der verantwortlich für einen Krankheitsverlauf verantwortlich ist. Mit Hilfe von Röntgenkristallographie oder NMR-Spektroskopie kann für viele Targets ihre dreidimensionale Struktur, d.h. Art und Lage ihrer Atome im Raum, aufgeklärt werden. Die biologische Funktion eines Targets wird über eine Bindetasche kontrolliert, an die molekulare Verbindungen andocken und damit z.B. eine Signalübertragung oder eine andere molekulare Reaktion auslösen. Ziel der Wirkstoffforschung ist es, solche Verbindungen zu identifizieren, die spezifisch und mit hoher Affinität an die Bindetasche des relevanten Targets binden und die gewünschte molekulare Reaktion auslösen bzw. hemmen. Virtuelle Screening-Verfahren ermöglichen eine gezielte Suche nach solchen Wirkstoffkandidaten *in silico*<sup>1</sup>. Die wichtigste Klasse von virtuellen Screening Verfahren sind Molekulare Docking-Verfahren: Sie bewerten die Bindungsaffinität einer Verbindung mit dem Target, in dem sie durch Rotation und Translation unterschiedlicher Molekülkonformere mögliche Platzierungen in der Bindetasche des Targets suchen und die resultierende Bindungsenergie approximieren [BK03]. Dabei sind die hohe Komplexität des Suchraums und die aufwändige Berechnung der Bindungsenergie die wesentlichen Herausforderungen aus Sicht der Informatik.

Heutige Docking-Verfahren benötigen für die Bewertung eines Moleküls auf einer leistungsfähigen Workstation etwa eine halbe bis ganze Minute Rechenzeit [KR04]. Da virtuelle Screening-Experimente häufig hunderttausende oder bis zu einer Million Verbindungen umfassen, ist die Entwicklung effizienter Docking-Algorithmen ein bedeutsames Forschungsgebiet geworden. Die meisten Verfahren für das Molekulare Docking folgen jedoch weiterhin dem Paradigma der sequentiellen Verarbeitung aller Verbindungen in einem Screening-Experiment und haben damit intrinsisch ein lineares Laufzeitverhalten in der Anzahl der zu untersuchenden Verbindungen. Einzige Ausnahmen hierzu sind Verfahren für das Screening kombinatorischer Bibliotheken und Cluster-basierte Screening-Verfahren:

- Kombinatorische Bibliotheken entstehen durch ein gemeinsames Basis- oder Ankerfragment, das an definierten Stellen mit einer vorgegebenen Menge von Teilfragmenten substituiert werden kann und somit unterschiedliche Verbindungen ergibt. Dadurch weisen die Verbindungen einer kombinatorischen Bibliothek strukturell bedingt einen hohen Grad an Gemeinsamkeit auf. Beim Molekularen Docking kann dies ausgenutzt werden, indem das gemeinsame Basisfragment einmalig für alle Verbindungen in der Bindetasche platziert wird. Die dabei entstehenden Startplatzierungen werden

---

<sup>1</sup> *In silico* bedeutet die Durchführung der Wirkstoffsuche im Computer. Klassische Verfahren der Wirkstoffsuche z.B. im Nasslabor bezeichnet man hingegen als *in vitro*-Verfahren.

dann für alle Verbindungen in der Bibliothek wieder verwendet und jeweils um die unterschiedlichen Substituenten der Verbindungen ergänzt [RL00].

- Cluster-basierte Verfahren nutzen Redundanzen in normalen, nicht-kombinatorischen Screening-Bibliotheken aus, indem sie die zu untersuchenden Verbindungen anhand physikochemischer und geometrischer Eigenschaften vorab in Cluster gruppieren. Dies kann anhand gemeinsamer molekularer Teilfragmente [SL01] oder basierend auf gleichen funktionellen Gruppen in ähnlicher räumlicher Anordnung [JT03] geschehen. Für das Screening wird in einem ersten Schritt ein repräsentativer Vertreter jedes Clusters in der Bindetasche platziert und energetisch bewertet. Für die Cluster der am besten bewerteten Vertreter werden dann in einem zweiten Schritt alle Verbindungen detailliert untersucht und bewertet. Damit ist das Laufzeitverhalten dieser Verfahren nicht mehr linear in der Anzahl der zu untersuchenden Verbindungen sondern in der Anzahl der unterschiedenen Cluster.

Im Rahmen dieser Arbeit wurde ein neues Paradigma für das strukturbasierte virtuelle Screening entwickelt, welches sich von der sequentiellen Verarbeitung von Verbindungen oder Gruppen von Verbindungen einer Screening-Bibliothek löst und stattdessen ausgehend von den Eigenschaften der Bindetasche gezielt nach solchen Verbindungen sucht, die die Eigenschaften und Nebenbedingungen der Bindetasche erfüllen (Target-getriebenes Screening<sup>2</sup>). Dieses Paradigma wurde prototypisch im Screening-Werkzeug TriXX umgesetzt. Im Folgenden werden die wesentlichen Methoden von TriXX kurz vorgestellt und ausgewählte Ergebnisse der experimentellen Validierung diskutiert.

## 2 Methoden

Das in TriXX umgesetzte Target-getriebene Screening-Paradigma wird im folgenden Abschnitt überblicksartig vorgestellt. Der zweite Abschnitt geht auf den in TriXX verwendeten molekularen Deskriptor näher ein. Für einen umfassenden Überblick sei an dieser Stelle auf die Dissertation des Autors verwiesen.

### 2.1 Grundlegende Vorgehensweise in TriXX

Das Software-Werkzeug TriXX besteht aus einer relationalen Datenbank<sup>3</sup>, in der die zu untersuchenden Verbindungen katalogisiert sind, und aus der Docking-Engine<sup>4</sup>, in der die rechenintensive Platzierung und Bewertung von Verbindungen durchgeführt wird. Ein virtuelles Screening-Experiment mit TriXX besteht insgesamt aus fünf Phasen (vgl. Abbildung 1):

---

<sup>2</sup> Patent angemeldet unter Aktenzeichen 10 2005 061 270.9 am 20.12.2005.

<sup>3</sup> Als Datenbankmanagementsystem wurde PostgreSQL 7.4 für Linux verwendet.

<sup>4</sup> Die Docking-Engine wurde in ANSI-C als eigenständiges Programm unter Linux entwickelt. Die Kommunikation zwischen Docking-Engine und Datenbank geschieht über das C-API von PostgreSQL.

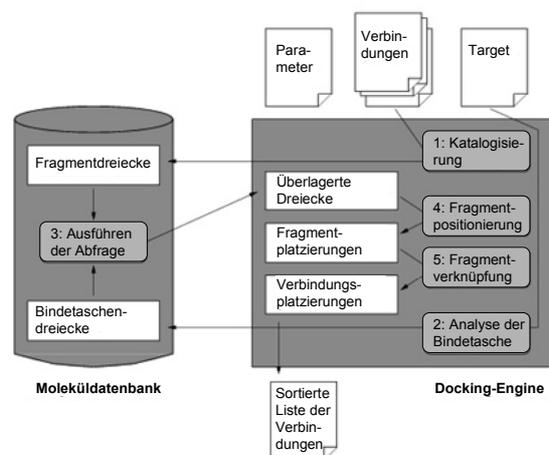


Abbildung 1: Ablauf eines Screening-Experiments in TriX

In der ersten Phase wird der Katalog von Verbindungen aufgebaut. Diese Phase ist ein Vorverarbeitungsschritt, der unabhängig vom Screening-Experiment einmalig für alle Verbindungen der Screening-Bibliothek durchzuführen ist. In dieser Phase werden die Verbindungen aus Textdateien eingelesen, funktionelle Gruppen identifiziert (im Folgenden CIACs<sup>5</sup> genannt) und die Verbindungen in wenige, möglichst starre Fragmente zerlegt. Fragmente werden mit Hilfe der unique SMILES-Notation [WW89] eindeutig gekennzeichnet, so dass das wiederholte Auftreten in anderen Verbindungen erkannt wird und sie nur einmalig im Katalog abgelegt werden. Für jedes Fragment werden gültige Konformere erzeugt. Ausgewählte Tripel funktioneller Gruppen eines Konformers bilden Dreiecke, mit denen die geometrischen und physikochemischen eines Fragmentkonformers beschrieben werden. Der Dreiecksdeskriptor ist das Schlüsselement für die effiziente Katalogisierung der Fragmente und wird in Abschnitt 2.2 näher beschrieben. Wenn ein Fragment noch nicht im Katalog beschrieben ist, werden alle Dreiecksdeskriptoren sowie unterstützende Informationen<sup>6</sup> in der Datenbank gespeichert.

Phasen 2 bis 5 sind Target-spezifisch und müssen in jedem Screening-Experiment wiederholt werden. Phase 2 übersetzt die atomare Darstellung der Target-Bindetasche in eine vereinfachte, diskrete Bindetaschendarstellung mit wenigen energetisch bevorzugten Punkten (SIACs<sup>7</sup>), die mit funktionellen Gruppen der Verbindungen bzw. der Fragmente überlagert werden. Die bevorzugten Punkte werden zu Tripeln zusammengefasst und die optimalen Eigenschaften eines komplementären Fragment-Dreiecks mit Hilfe des Dreiecksdeskriptors beschrieben. Die so entstandenen Bindetaschendreiecke werden ebenfalls in die Datenbank geschrieben.

<sup>5</sup> CIAC: compound interaction center (wir nur im Zusammenhang mit einer Verbindung und unabhängig von einem speziellen Fragment der Verbindung benutzt)

<sup>6</sup> Zu den unterstützenden Informationen gehören unter anderem die funktionellen Gruppen des Fragments sowie die Zuordnung des Fragments zu der Verbindung, aus der es durch Zerlegung identifiziert wurde.

<sup>7</sup> SIAC: site interaction center

In Phase 3 wird aus jedem Bindetaschendreieck eine Suchanfrage auf die Tabelle der Fragmentdreiecke gebildet. Diese Phase läuft innerhalb der Datenbank ab. Auf die Abstände der funktionellen Gruppen eines Bindetaschendreiecks sowie deren bevorzugte Wechselwirkungsrichtungen werden Toleranzen addiert bzw. subtrahiert. Für jedes Bindetaschendreieck entsteht so eine Bereichsanfrage, die die Art der funktionellen Gruppen, deren minimalen und maximalen Euklidischen Abstand sowie deren ungefähre bevorzugte Wechselwirkungsrichtung<sup>8</sup> vorgibt. Die Fragmentdreiecke sind nach genau diesen Eigenschaften über einen B<sup>+</sup>-Baum in der Datenbank indiziert abgelegt, so dass diese Bereichsanfragen sehr effizient durchgeführt werden können.

Die zu einem Bindetaschendreieck gefundenen Fragmentdreiecke werden in Phase 4 in der Docking-Engine weiter verarbeitet. Das zu einem Fragmentdreieck gehörige Fragmentkonformer wird dabei mit dem Bindetaschendreieck überlagert (Punktüberlagerung nach Kabsch [Ka76]). Die resultierende Fragmentplatzierung wird energetisch bewertet und optimiert.

Für die erfolgreich platzierten Fragmente wird in Phase 5 im Katalog nachgeschlagen, in welchen Verbindungen diese Fragmente auftreten. Die unabhängig voneinander platzierten Fragmente einer Verbindung werden dann wieder zusammengesetzt, sofern die relative Lage der Fragmente zueinander dies zulässt. Dabei ist es möglich und zulässig, dass nicht alle Fragmente der Verbindung platziert werden konnten oder ausreichen nah beieinander liegen, um zusammengesetzt werden zu können. Aus den zusammengesetzten Fragmenten einer Verbindung wird schließlich die Bindungsaffinität der Verbindung approximiert. Auf Basis dieses Werts können nachfolgend dann jene Verbindungen selektiert werden, deren Bindungsmodus und -affinität in weiteren Screening-Experimenten detaillierter untersucht werden sollen.

## 2.2 Dreiecksdeskriptor

Der Dreiecksdeskriptor ist die wesentliche Voraussetzung für den vorgestellten Screening-Ansatz von TrixX. Er muss ausreichend selektiv sein, um aus einer sehr hohen Zahl von Fragmentdreiecken diejenigen zu identifizieren, welche die Anforderungen der Bindetasche innerhalb vorgegebener Toleranzwerte erfüllen. Gleichzeitig müssen die Eigenschaften des Deskriptors sortierbar sein, damit sie in indizierten Zugriffsstrukturen wie B<sup>+</sup>-Bäumen verwaltet werden können. Die Deskriptoren in bestehenden Molekularen Docking-Verfahren kodieren in der Regeln Art und Abstand funktioneller Gruppen. Die Selektion passender Deskriptorinstanzen geschieht dann mit Hilfe von geometrischen Hashing-Verfahren (line/triangle matching). Der in TrixX verwendete Deskriptor erreicht eine deutlich höhere Selektivität, indem zusätzlich Richtungsinformationen und Informationen über die Lage sterischer Masse kodiert werden. Abbildung 2 verdeutlicht die Elemente des Dreiecksdeskriptors.

---

<sup>8</sup> Wechselwirkungsrichtungen werden mit Hilfe von Euler-Winkeln dargestellt, so dass die minimale und maximale Abweichung von der optimalen Wechselwirkungsrichtung durch Winkelbereiche der drei Euler-Winkel beschrieben werden können.

Drei funktionelle Gruppen oder FIACs<sup>9</sup> eines Fragmentkonformers definieren den funktionellen Typ (Wasserstoffbrücken-Akzeptoren FIAC1 und FIAC2 sowie die hydrophobe Gruppe FIAC3) und die Geometrie des Dreiecks (Seitenlängen  $d_{0,1}$ ,  $d_{1,2}$ ,  $d_{2,0}$ ). Für die eindeutige Beschreibung des Dreiecks und den direkten Vergleich mit anderen Dreiecken werden die funktionellen Gruppen anhand ihres physikochemischen Typs und ihrer Abstände voneinander in eine kanonische Reihenfolge gebracht ( $fiac_0$ ,  $fiac_1$ ,  $fiac_2$ ).

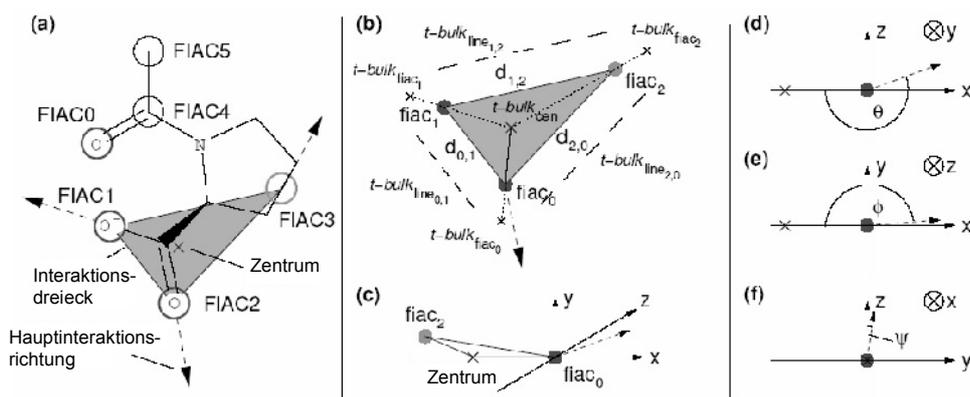


Abbildung 2: (a) Bildung eines Dreiecks aus drei funktionellen Gruppen (FIACs) des Fragments, (b) Darstellung aller Beschreibungselemente des Dreiecksdeskriptors, (c) Platzierung eines lokalen Koordinatensystems zur Beschreibung der bevorzugten Wechselwirkungsrichtung einer funktionellen Gruppe, und (d-f) Kodierung der Wechselwirkungsrichtung durch drei Eulerwinkel

Entlang der Dreiecksseiten kodieren Liniensegmente (z.B.  $t\text{-bulk}_{\text{line},0,1}$ ) das Vorhandensein sterischer Masse des Fragments bzw. der Bindetasche. Ein solches Segment entspricht einem Bit eines Bitvektors, so dass durch einfachen Bitvektorvergleich Überlappungen von Verbindung und Bindetasche erkannt werden können.

Jede funktionelle Gruppe einer Verbindung weist eine bevorzugte Wechselwirkungsrichtung auf (dargestellt durch gestrichelte Pfeile). Diese Richtung wird relativ zur Dreiecksebene und zum Mittelpunkt des Dreiecks beschrieben. Zu diesem Zweck wird ein lokales Koordinatensystem so in der funktionellen Gruppe zentriert, dass der Dreiecksmittelpunkt auf der negativen x-Achse und die in der kanonischen Ordnung vorangehende funktionelle Gruppe in der x-z-Ebene mit positivem z-Wert liegt (vgl. Abbildung 2(c)). Der Vektor der bevorzugten Wechselwirkungsrichtung wird dann in Euler-Winkel relativ zu den Achsen des lokalen Koordinatensystems übersetzt (vgl. Abbildung 2(d-f)).

Bis auf den Bitvektor zur Beschreibung der sterischen Masse sind alle Elemente des Deskriptors geeignet, Dreiecke nach ihren Eigenschaften zu sortieren und mit Hilfe von Bereichsanfragen zu selektieren.

<sup>9</sup> FIAC: fragment interaction center (funktionelle Gruppe eines Fragments)

### 3 Ergebnisse und Diskussion

Für die Validierung von TrixX wurde eine erste Screening-Bibliothek aus ca. 7.500 Wirkstoff-ähnlichen Verbindungen [SR01] und etwa 700 bekannten aktiven Verbindungen von elf Targets<sup>10</sup> mit hoher therapeutischer Relevanz aufgebaut. Für die Analyse des Laufzeitverhaltens wurden die 700 bekannten aktiven Verbindungen in einer weiteren Bibliothek mit 1,3 Millionen zufälligen Verbindungen aus öffentlich verfügbaren Anbieterkatalogen<sup>11</sup> vermischt.

#### 3.1 Strukturelle Eigenschaften des Molekülkatalogs

Die etwa 8.200 ausgewählten Verbindungen wurden mit TrixX katalogisiert, um die grundlegenden Annahmen bezüglich Redundanz zwischen den Verbindungen und ausreichender Starrheit der Fragmente zu validieren. Eine Verbindung wird von TrixX in durchschnittlich 3,2 Fragmente zerlegt. Insgesamt ergeben sich circa 17.000 unterschiedliche Fragmente, etwa 9.000 weitere Fragmente sind Wiederholungen von Fragmenten, die auch in anderen Verbindungen auftreten. Statt 3,2 müssen in der Datenbank entsprechend nur etwa 2 Fragmente physisch gespeichert werden. Experimente mit größeren Verbindungsbibliotheken zeigen, dass dieser Wert bei 700.000 Verbindungen auf unter 0,4 fällt.

Ein Fragment hat durchschnittlich 83 Konformationen und wird insgesamt durch circa 600 Fragmentdreiecke im Katalog beschrieben. Dies erlaubt eine kompakte Darstellung der Molekülflexibilität.

Die Trefferquote einer Suchanfrage auf die Tabelle der Fragmentdreiecke mit üblichen Anfrageswerten liegt bei etwa 0,01 bis 0,1 Prozent aller gespeicherten Fragmentdreiecke, wenn lediglich die Typen der funktionellen Gruppen und die tolerierten Dreiecksseitenlängen als Suchkriterien verwendet werden. Bei Hinzunahme der tolerierten Wechselwirkungsrichtungen als Suchkriterien sinkt die Trefferquote auf etwa 0,001 bis 0,01 Prozent aller gespeicherten Fragmentdreiecke. Das entspricht einer Steigerung der Selektivität des Dreiecksdeskriptors um den Faktor zehn.

#### 3.2 Anreicherung in virtuellen Screening-Experimenten

In elf virtuellen Screening-Experimenten wurde für jedes Target einzeln getestet, wie gut TrixX in der Lage ist, die in der Bibliothek bekannten aktiven Wirkstoffe des Target von den restlichen Verbindungen in der Bibliothek zu trennen. Als Vergleichswert wurde

---

<sup>10</sup> Verwendete Targets (mit PDB Code): Gelatinase A (1ck7); Cyclooxygenase 2 (1cx2); Sekretorische Phospholipase A<sub>2</sub> (1db4); Thrombin (1dwd); Estrogen (1err); Cyclin-abhängige Kinase 2 (1jvp); Thymidinkinase (1kim); Angiotensin-konvertierendes Enzym (1o86); Dihydrofolatreduktase (4dfr); HIV-Protease (4phv); Thermolysin (4tmn).

<sup>11</sup> Diese Bibliothek wurde durch Marcus Gastreich (BioSolveIT GmbH, St. Augustin) und Andreas Kämper (Max-Planck-Institut für Informatik, Saarbrücken) zusammengestellt und freundlicherweise für die Zwecke dieser Arbeit zur Verfügung gestellt.

hier die der Anteil der bekannten aktiven Wirkstoffe unter den 10 Prozent der von TrixX am höchsten bewerteten Wirkstoffe verwendet. Dieser Anteil ist je nach Target drei- bis achtmal höher als in der ursprünglichen Screening-Bibliothek; beim Target Dihydrofolat-Reduktase befinden sich unter den 10 Prozent am höchsten bewerteten Wirkstoffe sogar bereits alle bekannten Aktiven. Diese Ergebnisse belegen, dass TrixX in der Lage ist, aktive Wirkstoffe von nicht aktiven Verbindungen zu trennen.

Dieses Anreicherungsverhalten ist vergleichbar zum etablierten Molekularen-Docking-Verfahren FlexX [RK96]. FlexX erreicht bei den gleichen Experimenten in drei Fällen eine bessere Anreicherung als TrixX (1db4, 1kim, 1o86); in zwei Fällen zeigt TrixX eine bessere Anreicherung als FlexX (1err, 4dfr) und in den restlichen Fällen ergeben sich für FlexX und TrixX ähnliche Anreicherungen. In detaillierten Untersuchungen konnten nachgewiesen werden, dass die Unterschiede im Anreicherungsverhalten systematisch sind. TrixX zeigt ein besseres Anreicherungsverhalten als FlexX, wenn die aktiven Wirkstoffe eines Targets eher klein (weniger als 24 Heteroatome) und hydrophob (vier hydrophobe funktionelle Gruppen oder mehr) sind. Insgesamt erreicht TrixX eine leicht geringere aber noch vergleichbare Genauigkeit in der Anreicherung wie etablierte Molekulare Docking-Verfahren.

### 3.3 Laufzeitverhalten

Die 1,3 Millionen zufälligen Verbindungen wurden repräsentativ in fünf Bibliotheken zunehmender Größe (6.500, 13.000, 26.000, 52.000 und 130.000 Verbindungen) aufgeteilt und um die 700 bekannten aktiven Wirkstoffe der elf Targets ergänzt. Die fünf Bibliotheken wurden einmalig mit TrixX katalogisiert und danach in Screening-Experimenten für jedes Target verwendet. Dabei wurde Laufzeit und Speicherbedarf von TrixX gemessen<sup>12</sup>. Abbildung 3 zeigt die Entwicklung der durchschnittlichen Laufzeit bei zunehmender Größe der Screening-Bibliothek.

Deutlich erkennbar bei allen Targets ist der Rückgang der durchschnittlichen Laufzeit pro Verbindung mit wachsender Größe der Screening-Bibliothek. Die absolute Höhe der Laufzeit variiert zum Teil deutlich zwischen den Targets. Wesentlicher Einflussfaktor ist hier die Anzahl an Bindetaschendreiecken eines Targets, die für Abfragen auf der Datenbank genutzt werden<sup>13</sup>.

Sequentielle Molekulare Docking-Verfahren wie FlexX hingegen weisen eine konstante durchschnittliche Laufzeit. In Schellhammer et al. wurde die durchschnittliche Laufzeit von FlexX für die vier Target 1dwd, 1jvp, 1o86 und 4dfr auf etwa 25,20 Sekunden pro Verbindung bestimmt<sup>14</sup>. Für diese vier Targets liegt die durchschnittliche Laufzeit von

---

<sup>12</sup> Die Laufzeit- und Speicherbedarfsmessungen wurden mit einer 64-Bit-Version von TrixX auf einem SUN Fire Server mit 4 CPUs und 32 GB Hauptspeicher durchgeführt.

<sup>13</sup> Die Korrelation zwischen Anzahl Bindetaschendreiecke und Laufzeit pro Verbindung liegt über alle Targets und über alle Bibliotheksgrößen bei 0,934.

<sup>14</sup> Diese Experimente wurden auf einer Dual Xeon 2.4 GHz Linux Workstation durchgeführt. In separaten Experimenten wurde nachgewiesen, dass TrixX auf dieser Hardware ähnliche Laufzeiten erreicht wie auf der SUN-Umgebung.

TrixX im Falle der größten Bibliothek bei nur 0,42 Sekunden pro Verbindung. Das entspricht einer Beschleunigung um den Faktor 60.

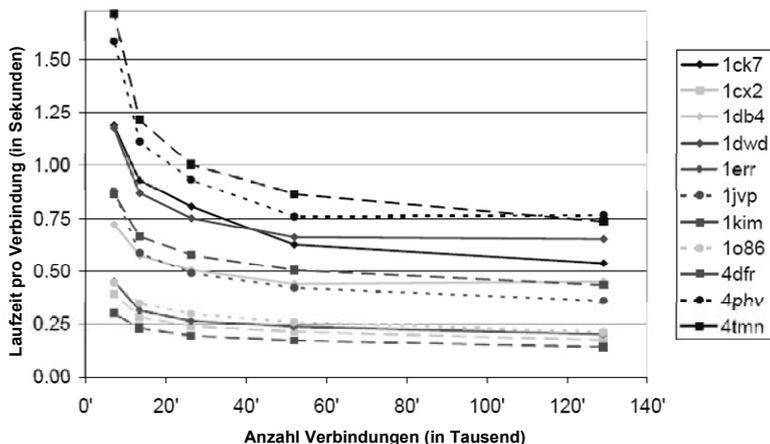


Abbildung 3: Durchschnittliche Laufzeit von TrixX pro Verbindung für unterschiedliche Größen der Screening-Bibliothek bei elf Targets

Ein zusätzlicher Laufzeitvergleich mit FlexX-Scan [SR04], einer Weiterentwicklung von FlexX um eine vereinfachte Bindetaschenbeschreibung, zeigt, dass TrixX auch etwa 20 mal schneller als FlexX-Scan ist. Nur dieser Anteil des Geschwindigkeitsgewinns ist demnach dem Target-getriebenen Paradigma von TrixX zuzuschreiben, während ein Faktor 3 der vereinfachten Bindetaschenbeschreibung zuzuschreiben ist.

## 4 Zusammenfassung

Mit TrixX wurde erstmalig ein Target-getriebenes Paradigma für das virtuelle Screening entwickelt und seine praktische Anwendbarkeit in Experimenten nachgewiesen. TrixX erreicht ein vergleichbares Anreicherungsverhalten wie das etablierte Molekulare Docking-Verfahren FlexX. Die Laufzeitanalysen bestätigen, dass die lineare Abhängigkeit der Laufzeit von der Anzahl der zu untersuchenden Verbindungen gebrochen werden konnte. Stattdessen wurde im Bereich der getesteten Bibliotheksgrößen ein sublinearer Anstieg der Laufzeit erreicht. In weiteren Untersuchungen muss dieses Laufzeitverhalten auch für größere Bibliotheken überprüft werden.

Für die praktische Anwendbarkeit des TrixX-Prototyps ist die Software zu parallelisieren, um auf großen Compute-Clustern eingesetzt werden zu können. Die klare Struktur des Datenbankkatalogs und die Darstellung der Bindetasche durch unabhängige Dreiecke sind bereits die wesentlichen Voraussetzungen für eine solche Parallelisierung.

## Literatur

- [BK03] Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 2003, 32, 335-373.
- [KR04] Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* 2004, 57, 225-242.
- [RL00] Rarey, M.; Lengauer, T. A Recursive Algorithm for Efficient Combinatorial Library Docking. *Persp Drug Discov Des* 2000, 20, 63-81.
- [SL01] Su, A. I.; Lorber, D. M.; Weston, G. S.; Baase, W. A.; Matthews, B. W. et al. Docking molecules by families to increase the diversity of hits in database screens: computational strategy and experimental evaluation. *Proteins* 2001, 42, 279-293.
- [JT03] Joseph-McCarthy, D.; Thomas IV, B. E.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-Based Molecular Docking to Account for Ligand Flexibility. *Proteins* 2003, 51, 172-188.
- [WW89] Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J Chem Inf Comput Sci* 1989, 29, 97-101.
- [Ka76] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* 1976, 32, 922-923.
- [SR01] Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J Med Chem* 2001, 44, 1035-1042.
- [RK96] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996, 261, 470-489.
- [SR04] Schellhammer, I.; Rarey, M. FlexX-Scan: fast, structure-based virtual screening. *Proteins* 2004, 57, 504-517.

**Ingo Schellhammer** wurde 1974 in Freckenhorst in Westfalen geboren und nahm bereits während seiner Abiturzeit mehrmals erfolgreich am Bundeswettbewerb Informatik teil. Er studierte von 1995 bis 2000 Wirtschaftsinformatik an der Westfälischen Wilhelms-Universität in Münster. 1998 ging er im Rahmen eines Austauschprogramms für ein Jahr an die Ecole Supérieure de Commerce de Montpellier in Frankreich und erwarb dort das Diplôme. Nach seinem Studium begann Ingo Schellhammer im Jahr 2000 als Berater beim Business Technology Office von McKinsey & Company in Hamburg. Dort arbeitete er zunächst zwei Jahre in der strategischen Beratung von Unternehmen bei IT-Management-Fragen. Mitte 2002 begann er bei Prof. Dr. Matthias Rarey am neu gegründeten Zentrum für Bioinformatik in Hamburg seine Promotion im Bereich der computergestützten Wirkstoffforschung. Nach Abschluss der Promotion im August 2005 setzte Ingo Schellhammer seine Arbeit bei McKinsey & Company fort.