Algorithmische Methoden für kombinatorische chemische Bibliotheken

Louis Bellmann[1]

Abstract: Computergestützte Methoden sind seit Jahrzehnten ein integraler Bestandteil des Wirkstoffentwurfs. Hierfür werden große Molekülmengen digital prozessiert und zusammengefasst. Diese chemischen Bibliotheken werden beispielsweise nach Molekülen durchsucht, die interessante Eigenschaften im Rahmen einer bestimmten Anwendung aufweisen und als Leitstruktur für ein neues Medikament innerhalb eines Forschungsprojekts verwendet werden könnten. Hierbei spielt sowohl die Qualität als auch die Quantität der in einer Bibliothek enthaltenen Moleküle eine entscheidende Rolle. Klassischerweise wird die Molekülmenge einer Bibliothek enumeriert repräsentiert und durchsucht, das heißt jedes Molekül wird einzeln betrachtet. Dadurch skaliert der benötigte Speicherplatz und die beanspruchte Rechenzeit für die Durchsuchung der Bibliothek linear mit der Anzahl der enthaltenen Moleküle. In dieser Dissertation werden neuartige algorithmische Verfahren und Datenstrukturen entwickelt, die einen kombinatorischen Ansatz verfolgen. Dabei werden Ideen aus der kombinatorischen Chemie aufgegriffen: Durch eine begrenzte Menge chemischer Bausteine und Reaktionen wird ein kombinatorischer Raum von Produkten implizit aufgespannt, der diese um mehrere Größenordnung übersteigen kann. Die so gebildeten kombinatorischen Bibliotheken sind in der Lage mit weniger Ressourcen eine weitaus größere Anzahl von Molekülen abzubilden als klassische enumerierte Bibliotheken. Die drei in der Dissertation erarbeiteten algorithmischen Verfahren bieten jeweils neue Funktionalitäten für kombinatorische Bibliotheken und sind mit diesem Ansatz in der Lage auf Milliarden von Molekülen effizient zu operieren.

1 Einleitung

Seit den 1950er Jahren werden für den Wirkstoffentwurf chemische Daten in digitaler Form verarbeitet, gesammelt und computergestützte Methoden zu ihrer Durchsuchung entwickelt und verwendet. [RK57] Das „chemische Universum“ aller synthetisch zugänglichen Moleküle und Naturstoffe wird auf über 10^{60} geschätzt [BMG96] und ist damit bei Weitem zu groß und unerforscht, um eine für ihn repräsentative Menge an Molekülen auf ihre Eignung als Wirkstoff-Kandidaten für neue Arzneimittel zu testen. Stattdessen werden Sammlungen von Molekülen, sogenannte chemische Bibliotheken verwendet, die bestimmte Gruppen von Molekülen, kommerziell verfügbare oder bereits identifizierte Wirkstoff-Kandidaten enthalten. Für die erfolgreiche Suche nach einem neuen Arzneimittel spielt dabei sowohl die Größe der chemische Bibliothek, als auch die Qualität der enthaltenen Moleküle eine entscheidende Rolle. Im Stand der Technik wird jedes Molekül einer Bibliothek enumeriert und zur Analyse sowie bei der Suche nach Wirkstoff-Kandidaten einzeln algorithmisch prozessiert. Damit skaliert der Ressourcenbedarf an Laufzeit und Speicherplatz linear mit der

[1] Universitätsklinikum Hamburg-Eppendorf, Institut für angewandte Medizininformatik, Christoph-Probst-Weg 1, 20251 Hamburg, Deutschland, l.bellmann@uke.de

Größe einer chemischen Bibliothek. Um dieses Problem zu lösen, wurden kombinatorische chemische Bibliotheken entwickelt. [RS01] Hierbei werden kleinere Moleküle, sogenannte chemische Bausteine, in der Bibliothek zusammengefasst. Zusätzlich wird eine Menge von chemischen Reaktionen definiert, mit denen die chemischen Bausteine synthetisch zu Produkten kombiniert werden können. Allerdings werden diese Produkte nicht explizit enumeriert, sondern nur implizit durch die Bausteine und Reaktionen beschrieben. Dadurch lässt sich mithilfe der kombinatorischen Explosion eine potenziell große Menge von Produkte durch eine begrenzte Anzahl chemischer Bausteine und Reaktionen beschreiben.
Da algorithmische Verfahren im Stand der Technik jedes Molekül einzeln prozessieren, wird eine kombinatorische Bibliothek durch diese Ansatz faktisch enumeriert und der positive Effekt der impliziten kombinatorischen Explosion negiert. Prominente kombinatorische Bibliotheken [En21; OT21; Wu21] enthalten Milliarden von Molekülen und sind damit nicht mehr praktikabel enumerierbar ohne einen großen Aufwand an Laufzeit und Speicherplatz aufzuwenden. [HG19] In dieser Dissertation [Be22a] wurden deshalb neue Methoden und Datenstrukturen entwickelt, die den kombinatorischen Charakter dieser Bibliotheken ausnutzen und damit einen deutlich geringeren Ressourcenbedarf erzielen oder die Prozessierung dieser Bibliotheken erst ermöglichen. Sie operieren auf chemischen Bausteinen und Reaktionen und explorieren dabei den kombinatorischen Raum der Produkte ohne diese explizit zu enumerieren. Hierbei werden drei Verfahren vorgestellt, die erstmals die Substruktur-basierte Ähnlichkeitssuche, die Schnittmengenberechnung und die Bestimmung physikochemischer Eigenschaftsverteilungen für kombinatorische Bibliotheken ermöglichen. Die resultierenden Software-Lösungen werden bereits in der Praxis angewandt und tragen damit durch neue Möglichkeiten für Medizinalchemiker aktiv zum computergestützten Wirkstoffentwurf bei.

2 Topologische Fragmenträume

Um algorithmischen Verfahren ein effizientes Prozessieren zu ermöglichen, erarbeiten wir zunächst den *topologischen Fragmentraum*, eine kompakte Datenstruktur für kombinatorische Bibliotheken. Hierbei werden Moleküle durch einen Graphen repräsentiert. Die Knoten eines Molekulargraphen stellen die Atome des Moleküls dar, seine Kanten die kovalenten Bindungen. Molekulargraphen chemischer Bausteine bezeichnen wir als *Fragmente*, sie enthalten zusätzlich sogenannte *Linker*. Linker markieren die Verknüpfungsstellen zwischen chemischen Bausteinen, die durch die verwendeten Reaktionen definiert sind. Durch die Hinzunahme der Linker und die daraus resultierende Präprozessierung der chemischen Reaktionen bildet ein Fragment einen Subgraph aller Produkte, die durch Kombination des Fragments mit anderen Fragmenten gebildet werden kann. Diese Datenstrukturen bauen auf einem früheren Konzept von Fragmenträumen [RS01] auf und erweitern dieses um die Kompatibilität mit Ringschluss-Reaktionen sowie den *Topologiegraphen*, eine übergeordnete Datenstruktur. Chemische Bausteine, die eine gleiche reaktive Gruppen besitzen, können innerhalb einer Reaktion an der gleichen Stelle verwendet werden. Folglich bilden sich natürlicherweise Gruppen von chemischen Bausteinen, deren Fragmente ebenfalls alle

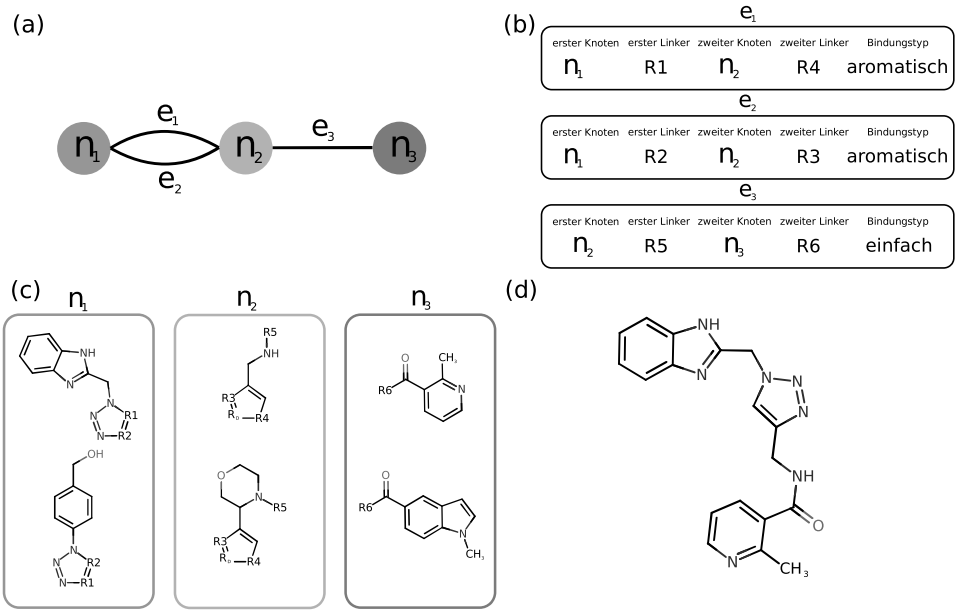

Abb. 1: (a) Ein Topologiegraph mit drei Knoten und drei Kanten. (b) Für jede Kante des Topologiegraph das Paar adjazenter Knoten, kompatibler Linker und der Bindungstyp der repräsentierten Bindung. (c) Die in den Knoten enthaltenen Fragmente. (d) Ein Beispielprodukt, enstanden aus der Kombination der drei obersten Fragmente.

die gleiche Linker-Konfiguration aufweisen. In Abbildung 1 (c) sind drei Gruppen mit jeweils zwei Fragmenten dargestellt. Diese Gruppen werden im Topologiegraphen durch Knoten repräsentiert. Die Kanten des Topologiegraphen stellen die kovalenten Bindungen dar, die während den Reaktionen zwischen den chemischen Bausteinen gebildet werden. Sie enthalten den Typ der geschlossenen Bindung, sowie ein paar von kompatiblen Linkern zwischen denen die Bindung geschlossen wird. Ein topologischer Fragmentraum besteht aus mindestens einem Topologiegraphen. Der Topologiegraph aus Abbildung 1 (a) enthält neben drei Knoten für die drei Fragment-Gruppen aus Abbildung 1 (c) noch drei Kanten, die in Abbildung 1 (c) definiert sind. Jedes Produkt entsteht somit aus einer Kombination von drei Fragmenten, bei der ein Ring zwischen den Fragmenten aus Gruppe n_1 und n_2 gebildet wird. Insgesamt lässt sich sowohl von einem Menschen, als auch von einem algorithmischen Verfahren anhand eines Topologiegraphen leicht die gemeinsame Struktur aller Produkte identifizieren, die mithilfe eines Syntheseprotokolls generiert werden können.

3 Kombinatorische topologische Ähnlichkeitssuche

Eine weitverbreitete Strategie zur Identifizierung von Wirkstoff-Kandidaten ist die Extraktion von Molekülen einer enumerierten chemischen Bibliothek, die zu einem Anfragemolekül

ähnlich sind. Eines der prominentesten Werkzeug der Ähnlichkeitssuche sind molekulare Fingerabdrücke, [MM16; RH10] die ein Molekül mithilfe einer Menge von Hashwerten oder als Bitfolge repräsentieren. Hierbei steht jeder Hashwert bzw. jedes gesetzte Bit für eine bestimmte chemische Struktur, die als zusammenhängender, induzierter Teilgraph im Molekulargraphen des repräsentierten Moleküls vorkommt. Damit sind molekulare Fingerabdrücke sehr kompakt und eignen sich neben der topologischen Ähnlichkeitssuche auch für das maschinelle Lernen. [Ya19]
Ein Ziel dieser Dissertation ist die Entwicklung der ersten Methode für topologische Ähnlichkeitssuchen in kombinatorischen chemischen Bibliotheken. Zunächst erarbeiten wir allerdings den Connected Subgraph Fingerprint (CSFP), [BPR19] einen molekularen Fingerabdruck speziell für dieses Anwendungsfeld. Für die Erzeugung eines CSFP werden zunächst mithilfe eines Branch-and-Bound-Ansatzes alle zusammenhängenden, induzierten Teilgraphen eines Molekulargraphen bestimmt. Nun wird für jeden Teilgraphen ein Hashwert errechnet, wobei wir ein bekanntes Verfahren zur Kanonisierung von Molekulargraphen [WWW89] adaptieren. Zusätzlich zur topologischen Struktur des Teilgraphen enthält der Hashwert auch Informationen über die chemischen Eigenschaften der repräsentierten Atome und Bindungen. Durch die Betrachtung aller zusammenhängenden, induzierten Teilgraphen liefert der CSFP eine dichte Beschreibung der topologischen Charakteristik eines Moleküls. Daneben erfüllt der CSFP eine Teilmengenrelation, die für die in Abschnitt 4 vorgestellten Methoden sehr hilfreich ist: Der CSFP eines Fragments ist in den CSFPs aller Produkte enthalten, die mithilfe des Fragments in einem topologischen Fragmentraum generiert werden können. Dadurch lassen sich die topologischen Eigenschaften von Produkten durch die CSFPs von Fragmenten approximieren, ohne diese Produkte explizit zu enumerieren. Wir evaluieren die Eignung des CSFP für den Medikamententwurf mithilfe eines Benchmarks für molekulare Fingerabdrücke. [RL13] Hier konnte der CSFP zeigen, dass er in der enumerierten topologischen Ähnlichkeitssuche vergleichbar und teilweise besser als der Stand der Technik bei der Prädiktion von Bioaktivität abschneidet. Zusätzlich zeigen wir, dass auf einem repräsentativen Datensatz keine Hashkollisionen des CSFP vorkommen und damit auch einzelne chemische Substrukturen eindeutig repräsentiert werden.
Die kombinatorische algorithmische Methode SpaceLight [BPR21] baut direkt auf dem CSFP auf und verwendet somit ebenfalls eine auf chemischen Substrukturen basierende Beschreibung molekularer Ähnlichkeit. Die zwei bereits existierenden Verfahren zur kombinatorischen Ähnlichkeitssuche [RS01; SKR21] verfolgen die Strategie des größten gemeinsamen Teilgraphens [SKR21] bzw. einen Pharmakophor-basierten Ansatz [RS01] und unterscheiden sich damit stark von der hier erarbeiteten Methode SpaceLight. Für die Ähnlichkeitssuche mit SpaceLight sind ein oder mehrere Anfragemoleküle, sowie ein topologischer Fragmentraum gegeben. Das Ziel ist es nun die ähnlichsten Moleküle des Fragmentraums zu identifizieren, ohne alle Produkte explizit zu enumerieren. Im ersten Schritt wird das Anfragemolekül mit einer Branch-and-Bound-Strategie in zusammenhängende Teilgraphen partitioniert. Hierbei werden nur Partitionen P generiert, die topologisch ähnlich zu mindestens einem Topologiegraphen T des Fragmentraums sind. Dafür muss ein Isomorphismus ϕ zwischen T und dem Graphen G_P, der aus Kontraktion der Partitionsklassen von P entsteht, existieren. Im zweiten Schritt wird nun für jede topolo-

gisch ähnliche Partition P und jeden gefunden Isomorphismus ϕ eine Ähnlichkeitssuche durchgeführt. Hierbei wird der CSFP einer Partitionsklasse $p \in P$ mit den CSFPs aller Fragmente aus dem Knoten $v = \phi(p)$ verglichen und ein Ähnlichkeitswert berechnet. Die höchsten Ähnlichkeitswerte der Fragmente werden zu gewichteten Summe kombiniert, die den Ähnlichkeitswert des entsprechende Produkts ergeben. Insgesamt werden die Produkte mit den höchsten Ähnlichkeitswerten ausgegeben. Durch diese Strategie wird nur eine kleine Anzahl der ähnlichsten Fragmentkombinationen wirklich erzeugt und eine Enumeration des kompletten Produktraums verhindert.
Um SpaceLight zu validieren, vergleichen wir die generierten Ergebnisse mit denen einer enumerierten Ähnlichkeitssuche. Wir können zeigen, dass die Beschreibung von Ähnlichkeit durch Spacelight korreliert zum klassischen enumerierten Ansatz durch molekulare Fingerabdrücke ist. Im Folgenden untersuchen wir das Laufzeitverhalten von SpaceLight auf den REAL Space [En21] (10^{10} implizite Produkte) und dem KnowledgeSpace [De10] (10^{14} implizite Produkte). Beide Räume sind in einer SQLite Datenbank abgespeichert, die jeweils unter 2 GB Speicherplatz benötigt. 500 Moleküle wurden für die Anfrage zufällig ausgewählt [IS05]. Wir verwenden für die Ähnlichkeitssuche mit SpaceLight openSUSE Leap 15 auf einer Intel Core i5-6500 64-Bit-Architektur mit 3,2 GHz und 16 GB Arbeitsspeicher. Die Analyse wird sowohl sequenziell als auch mit drei parallelen Prozessen durchgeführt. Für den Vergleich ziehen wir neben dem CSFP auch den ECFP [RH10], einen prominenten molekularen Fingerabdruck heran. Die Ergebnisse aus Abbildung 2 (a) zeigen, dass SpaceLight mit drei paralellen Prozessen in weniger als zehn Sekunden die 10^{10} impliziten Produkte des REAL Space durchsucht. Ein Großteil der Laufzeit wird dabei auf das Einladen der Daten verwandt. Bei mehreren Anfragemolekülen muss allerdings, anders als bei diesem Laufzeitexperiment, der Fragmentraum nur ein einziges Mal eingelesen werden, weshalb dieser Schritt gesondert betrachtet werden kann. Nach Abzug verbleiben nur grob drei Sekunden für die eigentliche Ähnlichkeitssuche. Die Ergebnisse in Abbildung 2(b) für die Suche im KnowledgeSpace sind vergleichbar und teilweise sogar leicht besser. Da der Produktraum des KnowledgeSpace den des REAL Space um grob vier Größenordnungen übersteigt, zeigt dies die Stärke des kombinatorischen Ansatzes von SpaceLight. Die größte momentan existierende enumerierte chemische Bibliothek [Ir20] umfasst grob 10^9 Produkte, also ungefähr um eine bzw. fünf Größenordnungen kleinere Produktmenge. Zur Speicherung und Suche dieser Bibliothek werden über 100TB Festplattenspeicher und über 100 Prozessorkerne verwendet. Die Autoren veranschlagen für eine Ähnlichkeitssuche im Mittel mehrere Minuten. Gerade in diesem Vergleich zeigt sich die Effizienz der Methode SpaceLight, die mit handelsüblichen PCs auf deutlich größeren Bibliotheken innerhalb von Sekunden operieren kann.

4 Analyse und Vergleich kombinatorischer Bibliotheken

Neben der Durchsuchbarkeit chemischer Bibliotheken spielt auch deren Analyse eine große Rolle für den Medikamententwurf. Beispielsweise kann eine chemische Bibliothek mit im Mittel hydrophoben Molekülen für die Suche nach Wirkstoffen geeignet sein, die in den

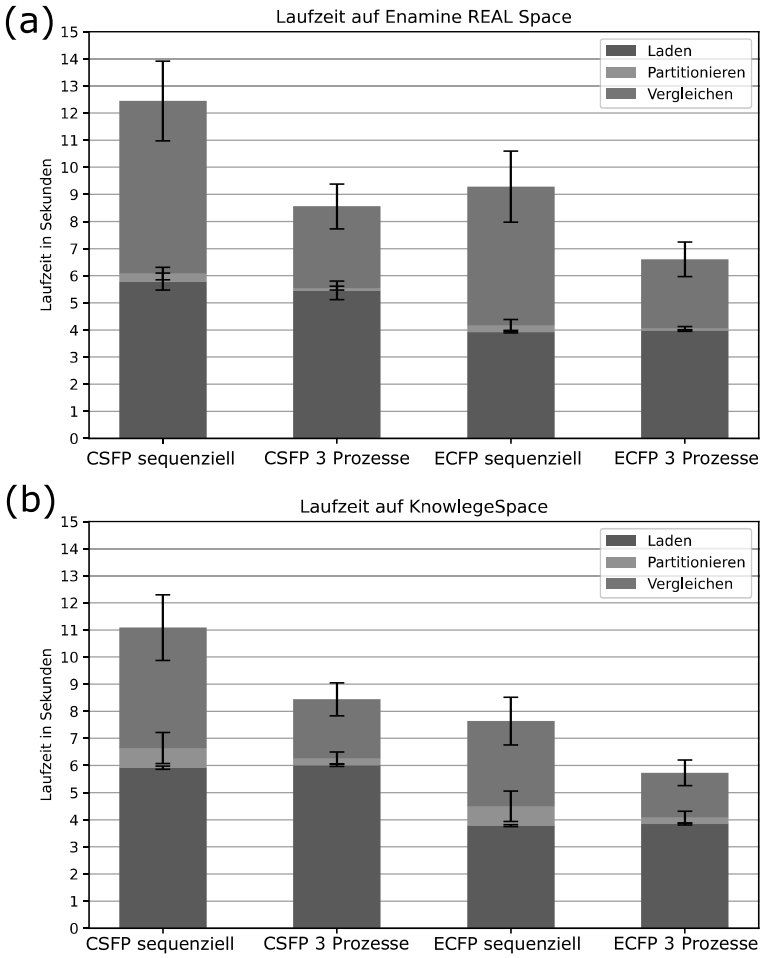

Abb. 2: Durchschnittliche Laufzeiten von SpaceLight auf dem REAL Space in (a) und KnowledgeSpace in (b). Die Laufzeit ist aufgeteilt in die Zeit zum Laden des Fragmentraums aus der Datenbank (blau), Partitionierung des Anfragemoleküls (orange) und Vergleich der molekularen Fingerabdrücke (grün).

Zellkern gelangen sollen. [Wa08] Die gleiche Bibliothek kann aber aufgrund der geringen Wasserlöslichkeit ihrer Moleküle für andere Projekte ungeeignet sein. Ebenfalls kann die Frage interessant sein, welche und wieviele Moleküle gemeinsam in zwei chemischen Bibliotheken existieren. Mit dieser Information kann beispielsweise Wissen über Synthese oder Bioaktivität verknüpft und Preise verglichen werden. Um diese Fragestellungen erstmals auch für kombinatorische Bibliotheken beantworten zu können, entwickeln wir zwei neuartige algorithmische Verfahren, die wie SpaceLight einen kombinatorischen Ansatz verfolgen.

Die innerhalb dieser Dissertation erarbeitete Methode SpaceCompare [Be22b] ermöglicht erstmals eine exakte Schnittmengenberechnung zweier kombinatorischer Bibliotheken ohne die Enumeration ihrer Produkte. Ähnlich wie bei SpaceLight werden hier chemische Substrukturen, repräsentiert durch den CSFP, verwendet. Hier wird die in Abschnitt 3 erläuterte Teilmengeneigenschaft des CSFP und dessen vollständige Beschreibung aller Substrukturen ausgenutzt. Zunächst werden alle sogenannten *kreuzende Subtrukturen* (induzierte, zusammenhängende Teilgraphen) identifziert, die bei der Kombination von Fragmenten entstehen und nicht schon innerhalb der Fragmente als Teilgraph vorkommen. Wir konnten eine Strategie zur Identifikation aller kreuzenden Substrukturen erarbeiten, die einen Partitionierungsansatz verfolgt und die Enumeration des Produktraums vermeidet. Nun erzeugen wir für jedes Fragment F einen *erweiterten CSFP*, der aus dem ursprünglichen CSFP hervorgeht und zusätzlich einen CSFP Hashwert für jede kreuzende Substruktur enthält, die bei der Kombination von F mit weiteren Fragmenten entsteht. Damit sind alle Substrukturen eines Produkts in mindestens einem erweiterten CSFP eines seiner Fragmente enthalten. Sei P ein Produkt, dass in zwei topologischen Fragmenträumen $\mathbb{F}$ und $\mathbb{F}'$ enthalten ist. Seien $F = \{F_1, \ldots, F_n\}$ die zugrundeliegende Fragmentkombination in $\mathbb{F}$ und $F' = \{F'_1, \ldots, F'_m\}$ in $\mathbb{F}'$. Sei weiterhin $c(M)$ der CSFP eines Moleküls M und $c_e(M)$ dessen erweiterter CSFP. Dann ergibt sich insgesamt

$$\begin{aligned} c(F_1) \cup \cdots \cup c(F_n) \subseteq c(P) \subseteq c_e(F'_1) \cup \cdots \cup c_e(F'_m) \\ \Rightarrow c(F_1) \cup \cdots \cup c(F_n) \subseteq c_e(F'_1) \cup \cdots \cup c_e(F'_m) \end{aligned}$$

In diesem Fall sprechen wir davon, dass F' F überdeckt. Dabei ist F' *minimal überdeckend*, wenn kein $F'' \subseteq F'$ ebenfalls F überdeckt. Damit haben wir eine notwendige Bedingung von Fragmenten und Fragmentkombinationen gefunden, damit sie zur Schnittmenge zweier Bibliotheken beitragen. Weiterhin ist diese Bedingung ohne explizite Betrachtung von Produkten überprüfbar.
Die Eingabe für SpaceCompare sind nun zwei topologische Fragmenträume $\mathbb{F}$ und $\mathbb{F}'$. In einem ersten Schritt werden Fragmente identifiziert und aussortiert, die durch keine Fragmentkombination des anderen Raums überdeckt werden. Dabei werden zu jedem verbleibenden Fragment alle minimal überdeckenden Kombinationen abgespeichert. In einem nächsten Schritt werden die überdeckten Fragmente in einer Branch-and-Bound-Strategie zu überdeckten Kombinationen erweitert. Im finalen Schritt werden die Produkte zu den überdeckten Kombinationen gebildet und deren Schnittmenge zwischen $\mathbb{F}$ und $\mathbb{F}'$ mithilfe einer kanonisierten Darstellung bestimmt. Auf diese Weise konnte erstmals die Schnittmenge dreier kommerzieller kombinatorischer Bibliotheken [En21; OT21; Wu21] bestimmt werden. Die drei Räume teilen sich weniger als 2% ihrer Produkte und tragen damit jeweils Milliarden von individuellen Produkten für den Medikamententwurf bei.
Das in dieser Dissertation vorgestellte algorithmische Verfahren SpaceProp [BKR22] ermöglicht erstmals eine exakte Berechnung physikochemischer Eigenschaftsverteilungen aller Produkte einer kombinatorischen Bibliothek. Hierbei werden die Wasserlöslichkeit und das Gewicht von Produkten untersucht, sowie ihre Fähigkeit Wasserstoffbrückenbindungen einzugehen. Der Eigenschaftswert eines Produktes ergibt sich als die Summe von Eigenschaftswerten seiner Atome. Dabei hängt der Wert eines Atoms von seinen

chemischen Eigenschaften und auch seiner Umgebung ab. Vergleichbar zu den kreuzenden Substrukturen der SpaceCompare Methode, kann sich diese Umgebung teilweise erst durch Kombination von Fragmenten bilden. Die Idee von SpaceProp ist es, Fragmente in eine *interne Komponente* und *externe Komponente* zu unterteilen. Der Eigenschaftswert von Atomen der internen Komponente hängt nicht von der Kombination mit anderen Fragmenten ab. Für Atome der externen Komponente ist das Gegenteil der Fall. Nun können Fragmente mit der gleichen externen Komponente gruppiert und gleichzeitig prozessiert werden. Damit kann eine Enumeration des Produktraums verhindert und trotzdem eine exakte Eigenschaftsverteilung bestimmt werden. Mithilfe von SpaceProp konnten wir erstmalig zeigen, dass ein Großteil der Produkte prominenter kombinatorischer Bibliotheken [En21; OT21; Wu21] weitverbreitete Kriterien für die orale Verfügbarkeit [Li97] erfüllen und damit für die Arzneimittelforschung geeignet sind.

5 Fazit und Ausblick

Die in dieser Dissertation beschriebenen algorithmischen Verfahren dienen der Durchsuchung und Analyse von kombinatorischen chemischen Bibliotheken für den Medikamentenentwurf. Jede Methode bietet für ihre jeweilige Problemstellung neuartige Funktionalitäten, die bisher nicht für dieses Feld existierten. Im Gegensatz zu klassischen, auf Enumeration basierenden Ansätzen, sind diese Verfahren in der Lage Billionen und mehr Moleküle effizient algorithmisch zu prozessieren. Zusammen mit der wachsenden Zahl öffentlich verfügbarer kombinatorischer Bibliothken, [HG19] werden mit steigendem Interesse für diese Technologien Milliarden neuer Moleküle für die Medizinalforschung zugänglich gemacht. Ein nächster Schritt könnte die Entwicklung kombinatorischer Methoden im Bereich des räumlichen Strukturvergleichs sein. [Wa22] Damit wäre eine großer Teil der Methodiken des computergestützten Medikamententwurfs auch für kombinatorische Bibliotheken verfügbar. Hier könnte beispielsweise der kompakte räumliche Molekulardeskriptor *Ray Volume Matrix* [Pe20] mit dem Partitionierungsansatz von SpaceLight kombiniert werden, um Partitionsklassen von Proteinbindetaschen mit Volumenbeschreibungen von Fragmenten zu vergleichen.

Literaturverzeichnis

[Be22a] Bellmann, L.: Algorithmische Methoden für kombinatorische chemische Bibliotheken. Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky, 2022.

[Be22b] Bellmann, L.; Penner, P.; Gastreich, M.; Rarey, M.: Comparison of Combinatorial Fragment Spaces and Its Application to Ultralarge Make-on-Demand Compound Catalogs. J. Chem. Inf. Model. 62/3, S. 553–566, 2022.

[BKR22] Bellmann, L.; Klein, R.; Rarey, M.: Calculating and Optimizing Physicochemical Property Distributions of Large Combinatorial Fragment Spaces. J. Chem. Inf. Model./, 2022.

[BMG96] Bohacek, R. S.; McMartin, C.; Guida, W. C.: The Art and Practice of Structure-Based Drug Design: a Molecular Modeling Perspective. Med. Res. Rev. 16/1, S. 3–50, 1996.

[BPR19] Bellmann, L.; Penner, P.; Rarey, M.: Connected Subgraph Fingerprints: Representing Molecules Using Exhaustive Subgraph Enumeration. J. Chem. Inf. Model. 59/11, S. 4625–4635, 2019.

[BPR21] Bellmann, L.; Penner, P.; Rarey, M.: Topological Similarity Search in Large Combinatorial Fragment Spaces. J. Chem. Inf. Model. 61/1, S. 238–251, 2021.

[De10] Detering, C.; Claussen, H.; Gastreich, M.; Lemmen, C.: KnowledgeSpace-a Publicly Available Virtual Chemistry Space. J. Cheminformatics 2/1, S. 1–1, 2010.

[En21] Enamine Ltd.: REAL Space, zuletzt zugegriffen 05/08/2021, 2021, URL: https://enamine.net/library-synthesis/real-compounds/real-space-navigator.

[HG19] Hoffmann, T.; Gastreich, M.: The Next Level in Chemical Space Navigation: Going Far Beyond Enumerable Compound Libraries. Drug Discov. Today/, 2019.

[Ir20] Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A.: ZINC20-A Free Ultralarge-scale Chemical Database for Ligand Discovery. J. Chem. Inf. Model. 60/12, S. 6065–6073, 2020.

[IS05] Irwin, J. J.; Shoichet, B. K.: ZINC- a Free Database of Commercially Available Compounds for Virtual Screening. J. Chem. Inf. Model. 45/1, S. 177–182, 2005.

[Li97] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J.: Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. Adv. Drug Deliv. Rev. 23/1-3, S. 3–25, 1997.

[MM16] Muegge, I.; Mukherjee, P.: An Overview of Molecular Fingerprint Similarity Search in Virtual Screening. Expert Opin. Drug Discov. 11/2, S. 137–148, 2016.

[OT21] OTAVAchemicals Ltd.: CHEMriya, zuletzt zugegriffen 23/08/2021, 2021, URL: https://www.otavachemicals.com/products/chemriya.

[Pe20] Penner, P.; Martiny, V.; Gohier, A.; Gastreich, M.; Ducrot, P.; Brown, D.; Rarey, M.: Shape-Based Descriptors for Efficient Structure-Based Fragment Growing. J. Chem. Inf. Model. 60/12, S. 6269–6281, 2020.

[RH10] Rogers, D.; Hahn, M.: Extended-Connectivity Fingerprints. J. Chem. Inf. Model. 50/5, S. 742–754, 2010.

[RK57] Ray, L. C.; Kirsch, R. A.: Finding Chemical Records by Digital Computers. Science 126/3278, S. 814–819, 1957.

[RL13] Riniker, S.; Landrum, G. A.: Open-Source Platform to Benchmark Fingerprints for Ligand-based Virtual Screening. J. Cheminf. 5/1, S. 26, 2013.

[RS01] Rarey, M.; Stahl, M.: Similarity Searching in Large Combinatorial Chemistry Spaces. J. Comput. Aided Mol. Des. 15/6, S. 497–520, 2001.

[SKR21] Schmidt, R.; Klein, R.; Rarey, M.: Maximum Common Substructure Searching in Combinatorial Make-on-Demand Compound Spaces. J. Chem. Inf. Model./, 2021.

[Wa08] Wasan, K. M.; Brocks, D. R.; Lee, S. D.; Sachs-Barrable, K.; Thornton, S. J.: Impact of Lipoproteins on the Biological Activity and Disposition of Hydrophobic Drugs: Implications for Drug Discovery. Nature Reviews Drug Discovery 7/1, S. 84–99, 2008.

[Wa22] Warr, W. A.; Nicklaus, M. C.; Nicolaou, C. A.; Rarey, M.: Exploration of Ultralarge Compound Collections for Drug Discovery. Journal of Chemical Information and Modeling 62/9, S. 2021–2034, 2022.

[Wu21] WuXi AppTec: GalaXi, zuletzt zugegriffen 05/08/2021, 2021, URL: https://www.labnetwork.com/frontend-app/p/#!/library/virtual.

[WWW89] Weininger, D.; Weininger, A.; Weininger, J. L.: SMILES. 2. Algorithm for Generation of Unique SMILES Notation. J. Chem. Inf. Comput. Sci. 29/2, S. 97–101, 1989.

[Ya19] Yang, M.; Tao, B.; Chen, C.; Jia, W.; Sun, S.; Zhang, T.; Wang, X.: Machine Learning Models Based on Molecular Fingerprints and an Extreme Gradient Boosting Method Lead to the Discovery of JAK2 Inhibitors. J. Chem. Inf. Model. 59/12, S. 5002–5012, 2019.

Louis Bellmann wurde am 12. September 1991 in Hamburg geboren. Er studierte an der Universität Hamburg Mathematik mit Schwerpunkt Graphentheorie und erhielt 2017 seinen Masterabschluss. Von 2018 bis 2022 promovierte er als wissenschaftlicher Mitarbeiter von Prof. Dr. Matthias Rarey in der Gruppe Algorithmisches Molekulardesign. In dieser Zeit entwickelte er Verfahren und Softwarelösungen, die bereits jetzt in der Arzneimittelforschung eingesetzt werden. Zurzeit arbeitet er als Postdoktorand am Universitätsklinikum Hamburg-Eppendorf an neuen Methoden zur Prävention und Analyse häufiger Krankheitsbilder. Zu seinen Forschungsschwerpunkten gehören Graphentheorie, Kombinatorik, Chemieinformatik und Medizininformatik.