



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) **DE 10 2005 029 437 A1 2007.01.04**

(12)

## Offenlegungsschrift

(21) Aktenzeichen: **10 2005 029 437.5**

(22) Anmeldetag: **24.06.2005**

(43) Offenlegungstag: **04.01.2007**

(51) Int Cl.<sup>8</sup>: **C12Q 1/00 (2006.01)**

(71) Anmelder:

**Humboldt-Universität zu Berlin, 10099 Berlin, DE**

(74) Vertreter:

**BOEHMERT & BOEHMERT, 28209 Bremen**

(72) Erfinder:

**Thimm, Martin, 10243 Berlin, DE; Hougardy, Stefan, Dr., 12527 Berlin, DE; Ziegler, Valentin, 10315 Berlin, DE**

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:

**THIMM, M. u.a.: Comparison of 2D similarity and 3D superposition, Application to searching a conformational drug database, J. Chem. Inf. Comput. Sci. (2004) 44(5)1816-22;**  
**MICHALSKY, E. u.a.: SuperLigands - a database of**

**ligand structures derived from the Protein Data Bank, BMC Bioinformatics (19.05.2005)6(1)122; Datenbank PubMed bei NCBI, Adresse [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov).**

**nih.gov, Zusammenfassung zu: GOEDE, A. u.a.: SuperDrug: a conformational drug database, Bioinformatics (01.05.2005), Epub 02.02.2005) 21(9) 1751-3 [recherchiert am 15.11.2005]; Datenbank PubMed bei NCBI, Adresse [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov).**

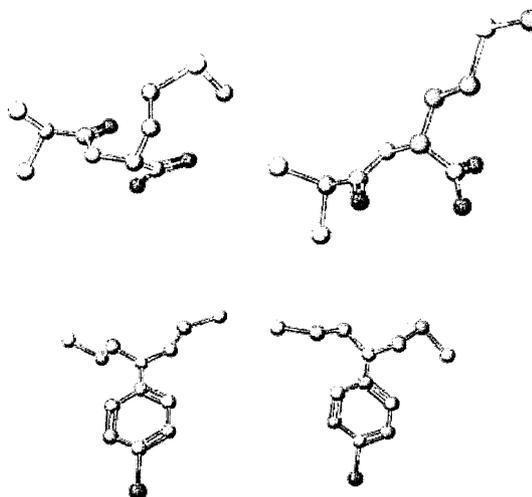
**nih.gov, Zusammenfassung zu: TERVO, A.J. u.a.: BRUTUS: optimization of a grid-based similarity function for rigid-body molecular superposition, 1. Alignment and virtual screening applications, J. Med. Chem. (16.06.2005)48(12)4076-86 [recherchiert am 15.11.2005];**

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gemäß § 44 PatG ist gestellt.

(54) Bezeichnung: **Verfahren und Vorrichtung zum computergestützten Auffinden von ähnlichen Molekülen**

(57) Zusammenfassung: Die Erfindung betrifft ein iteratives Verfahren zum computergestützten Auffinden von Molekülen mit Ähnlichkeiten zu einem oder mehreren Referenzmolekülen, bei welchen in einer Iteration mehrere Paare aus einer Referenzstruktur und einer Vergleichsstruktur in einander analoger Weise überlagert werden und ein Qualitätsmaß für jedes dieser Paare bestimmt oder abgeschätzt wird, wobei nach vorgegebenen Kriterien der Wert des Qualitätsmaßes für ein bestimmtes Paar zusammen mit der entsprechenden Überlagerungszuordnung als geltender optimaler Wert des Qualitätsmaßes oder einer der geltenden optimalen Werte des Qualitätsmaßes gespeichert wird. Die Erfindung betrifft weiterhin eine zugehörige Vorrichtung.



## Beschreibung

**[0001]** Die Erfindung betrifft ein computergestütztes iteratives Verfahren zum Auffinden von chemischen Molekülen mit Ähnlichkeiten zu einem oder mehreren Referenzmolekülen, das auf der Grundlage von strukturellen Daten, die in einer Datenbank vorgegeben sind oder von einem Benutzer eingegeben werden können, strukturelle Vergleiche anstellt und das oder diejenigen Vergleichsmoleküle bestimmt, welche chemisch, insbesondere stereochemisch gesehen die größten Übereinstimmungen mit dem Referenzmolekül aufweisen. Dies kann ein Molekül sein, welches insgesamt oder auch nur auf einem Teilbereich dem Referenzmolekül ähnlich ist. Letzteres kann insbesondere der Fall sein, wenn das Vergleichsmolekül größer als das Referenzmolekül ist und das Referenzmolekül in seiner Gesamtheit einem Teil des Vergleichsmoleküls ähnlich ist oder sogar identisch als Teilstruktur in dem Vergleichsmolekül enthalten ist. Das Ergebnis einer solchen Suche kann aber auch ein Vergleichsmolekül sein, welches auf einem Teilbereich eine Ähnlichkeit zu einem Teilbereich des Referenzmoleküls hat, ebenso wie das Vergleichsmolekül auch kleiner als das Referenzmolekül und in seiner Gesamtheit einem Teil des Referenzmoleküls ähnlich oder identisch mit diesem Teil sein kann. Derartige Vergleiche sind insbesondere in der pharmazeutischen Chemie und der Biochemie von großer Bedeutung.

### Stand der Technik

**[0002]** In der Regel ist vorab nicht bekannt, wie diejenigen Abschnitte der Moleküle beschaffen sind, in denen eine Ähnlichkeit gegeben ist. In den meisten Fällen ist dies auch für den Chemiker a priori nicht übersehbar. Computergestützte Verfahren gehen, wie auch die vorliegende Erfindung, davon aus, daß man alle möglichen Teilkonfigurationen des Referenzmoleküls mit allen möglichen entsprechenden Teilkonfigurationen aller Vergleichsmoleküle vergleichen muß, um sicher zu sein, daß man tatsächlich das optimale Ergebnis findet. Diese Aufgabe wird dadurch erschwert, daß zu einem Molekül mit einer bestimmten Konstitution, d.h. einer unverwechselbaren, für die Verbindung charakteristischen Anordnung der Atome (ohne Berücksichtigung von räumlichen Richtungen), gegebenenfalls mehrere Konfigurationen existieren und auch zu einer bestimmten Konfiguration immer noch mehrere mögliche Strukturen bestehen, die sich durch Rotation um einfache Bindungen ergeben. Die verschiedenen möglichen Strukturen eines Moleküls mit definierter Konstitution und Konfiguration werden als Konformere bezeichnet.

**[0003]** Für den computergestützten Vergleich gibt man in der Regel eine endliche Anzahl von Konformeren zu dem Referenzmolekül bzw. den Vergleichsmolekülen an. Insgesamt ist die Anzahl der Vergleiche, die man durchführen muß, so groß, daß eine vollständige Evaluierung aller möglichen Fälle praktisch nicht möglich ist. Vergleicht man beispielsweise organische Moleküle mit dreißig Nicht-Wasserstoffatomen und betrachtet man nur, was bereits eine Näherung ist, Teilstrukturen innerhalb eines Moleküls, die über Bindungen zusammenhängen, liegt die Anzahl der Vergleiche, die man für eine vollständige Evaluierung aller Möglichkeiten durchführen müßte, in der Größenordnung von  $10^{15}$ . Dies würde mit heutigen Geräten eine Rechenzeit von mehreren tausend Jahren erfordern. Für größere Moleküle wäre eine vollständige Evaluierung aller Möglichkeiten schlichtweg nicht möglich.

**[0004]** In der Mathematik bzw. Informatik sind Verfahren zum Lösen von kombinatorischen Optimierungsproblemen bekannt, bei denen die Anzahl der möglichen Lösungen so groß ist, daß die Berechnung aller möglichen Lösungen und die Auswahl der optimalen Lösungen aus den berechneten Lösungen nicht möglich ist. Hierzu zählt unter anderem auch das sogenannte Branch and Bound-Verfahren. Hierbei strukturiert man die möglichen Lösungen hierarchisch, wobei Lösungen der tieferen Hierarchieebenen die ihnen übergeordneten Lösungen der höheren Hierarchieebenen als Teillösung enthalten, wobei jede dieser Teillösungen auch bereits eine Gesamtlösung des gestellten Optimierungsproblems sein kann. Diese Hierarchie wird dabei zunächst entlang einem Zweig von oben nach unten, d.h. in jedem Schritt fortschreitend zu der jeweils tieferen Hierarchiestufe, abgearbeitet, wobei in jeder Iteration ein Qualitätsmaß, auch Score genannt, bestimmt wird, welches ein Maß für die Güte oder Qualität der bislang aufgefundenen (Teil)lösung ist. Auf jeder Hierarchiestufe wird überprüft, ob die bereits aufgefundene Teillösung durch Fortschreiten zu einer (beliebigen) tieferen Hierarchiestufe so verbessert werden kann, daß sie ein besseres Qualitätsmaß als die bislang aufgefundene (Teil)lösung mit dem besten Wert des Qualitätsmaßes liefert. Ist dies nicht der Fall, brauchen die tieferen Hierarchieebenen nicht mehr untersucht werden, da die beste bereits aufgefundene Lösung bereits eine bessere Lösung ist als alle Lösungen, die durch Einbeziehen der tieferen Hierarchieebenen gewonnen werden können. Auf diese Weise kann in einem einzigen Schritt eine sehr große Anzahl von möglichen Lösungen ausgeschlossen werden, ohne diese Lösungen explizit bestimmen zu müssen.

**[0005]** Stellt man nun fest, daß ein Fortschreiten zu tieferen Hierarchieebenen keine Lösung mit einem besseren Wert des Qualitätsmaßes liefern kann, evaluiert man typischerweise die Alternativen auf der gleichen

Hierarchieebene, die derselben übergeordneten Teillösung zugeordnet sind. Hat man diese Alternativen abgearbeitet, wobei wiederum idealerweise ein großer Teil der Lösung mittels einer Abschätzung, ob das Qualitätsmaß über die bereits bekannte beste Lösung hinaus verbessert werden kann, ausgeschlossen wird, geht man eine Hierarchiestufe höher und untersucht die alternativen Lösungen auf dieser Hierarchieebene, die einer gemeinsamen übergeordneten Lösung zugeordnet sind, und so weiter, bis der ganze Hierarchiebaum abgearbeitet ist.

**[0006]** Dieses Branch and Bound-Verfahren eignet sich sehr gut für die Übertragung auf chemische Sachverhalte. Will man zwei chemische Strukturen vergleichen, geht man zweckmäßigerweise so vor, daß man versucht, zunächst nur wenige Atome oder, allgemeiner gesprochen, Komponenten des Referenzmoleküls und des Vergleichsmoleküls zur Deckung zu bringen und, wenn dies hinreichend gut gelungen ist, zu versuchen, weitere Atome zur Deckung zu bringen. Dieses an der chemischen Struktur orientierte Vorgehen kann man dadurch auf die Methodik des Branch and Bound-Verfahrens übertragen, daß man die Hierarchieebenen so definiert, daß jede Hierarchieebene einer bestimmten Anzahl von Atomen oder Komponenten entspricht, die möglichst gut miteinander zur Deckung gebracht werden.

**[0007]** Das Branch and Bound-Verfahren ist teilweise auch bereits auf biologische und chemische Problemstellungen angewandt worden. Beispielsweise wird es in R. H. Latrop u.a., „A multiqueue branch-and-bound algorithm for anytime optimal search with biological applications“, Genome Informatiks 12 (2001), 73-82 für die Voraussage der Struktur eines Proteins und von Konformationen kleiner Moleküle verwendet. Ein abstrakter Branch and Bound-Algorithmus ohne konkreten Bezug zu einer chemischen oder biologischen Problemstellung ist beispielsweise in Ari Frank u.a., „A distance-based branch and bound feature selection algorithm“, in Uncertainty in Artificial Intelligence: Proceedings of the Nineteenth Conference (UAI-2003), Morgan Kaufmann Publishers, San Francisco, CA, 2003 beschrieben.

**[0008]** Bislang gibt es allerdings noch kein wirklich zufriedenstellendes Verfahren zum Vergleich von Molekülen unter Einbeziehung von Konformerstrukturen, bei dem sich die Zahl der potentiell möglichen Lösungen in einer Größenordnung von  $10^{30}$  bis  $10^{100}$  bewegen kann.

**[0009]** Es ist die Aufgabe der vorliegenden Erfindung, ein solches Verfahren und ein zugehöriges Computersystem bzw. ein entsprechendes Computerprogramm zur Verfügung zu stellen, welche eine optimale Lösung liefern können.

#### Aufgabenstellung

**[0010]** Erfindungsgemäß wird diese Aufgabe gelöst durch ein iteratives Verfahren zum computergestützten Auffinden eines oder mehrerer Moleküle mit Ähnlichkeiten, insbesondere strukturellen Ähnlichkeiten, zu einem oder mehreren Referenzmolekülen, die mehrere Komponenten aufweisen, auf der Grundlage von Informationen zu einer oder mehreren Referenzstrukturen, welche jeweils eine Struktur eines Referenzmoleküls darstellen, wobei die besagten Informationen die Lage von Komponenten in dem Referenzmolekül gemäß dieser Referenzstruktur, insbesondere in einem dreidimensionalen Raum, angeben, und von Informationen zu mehreren Vergleichsstrukturen, welche jeweils eine Struktur eines Vergleichsmoleküls darstellen, wobei die besagten Informationen jeweils die Lage von Komponenten in dem Vergleichsmolekül in dieser Vergleichsstruktur, insbesondere in einem dreidimensionalen Raum, angeben, bei dem in einer Iteration ein Teil eines Referenzmoleküls und ein Teil eines Vergleichsmoleküls verglichen werden, die jeweils einen Teil des Referenzmoleküls bzw. des Vergleichsmoleküls enthalten, die in einer früheren Iteration verglichen worden sind, wobei mindestens eine Iteration des Verfahrens umfaßt:  
 Auswahl einer ersten Referenzmenge von Komponenten eines Referenzmoleküls, welche einen Teil der Komponenten des Referenzmoleküls enthält, wobei die Anzahl der Komponenten in der ersten Referenzmenge kleiner als die Gesamtzahl der zu vergleichenden Komponenten in dem Referenzmolekül ist,  
 Auswahl einer ersten Vergleichsmenge von Komponenten eines Vergleichsmoleküls, welche die gleiche Anzahl von Komponenten wie die erste Referenzmenge besitzt und im Regelfall ebenfalls nur einen Teil der Komponenten des Vergleichsmoleküls enthält,  
 Bestimmen einer ersten Überlagerungszuordnung für eine erste Referenzstruktur, die dem besagten Referenzmolekül zugeordnet ist, und eine erste Vergleichsstruktur, welche dem besagten Vergleichsmolekül zugeordnet ist, wobei die erste Überlagerungszuordnung jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau eine Komponente der ersten Vergleichsmenge zuweist, wobei für mindestens eine weitere, sekundäre Referenzstruktur, welche zumindest eine Teilstruktur aufweist, die zu der ersten Referenzstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Kom-

ponente der ersten Referenzmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht,  
 und/oder für mindestens eine weitere, sekundäre Vergleichsstruktur, welche eine Teilstruktur aufweist, die zu der ersten Vergleichsstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Vergleichsmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht, bestimmt wird:  
 eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die primäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmenge die entsprechende Komponente der sekundären Referenzstruktur gemäß der Ähnlichkeitsbeziehung für die sekundäre Referenzstruktur derjenigen Komponente der ersten Vergleichsstruktur zuweist, welches die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist, und/oder  
 eine Überlagerungszuordnung für die erste Referenzstruktur und die sekundäre Vergleichsstruktur, welche jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau die Komponente zuweist, welche aufgrund der Kompatibilitätsbeziehung der sekundären Vergleichsstruktur derjenigen Komponente der ersten Vergleichsmenge entspricht, welche die erste Überlagerungszuordnung der betreffenden Komponente der ersten Referenzmenge zuweist,  
 und/oder  
 eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die sekundäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmenge die entsprechende Komponente der sekundären Referenzstruktur derjenigen Komponente der sekundären Vergleichsstruktur zuweist, welche der Komponente der ersten Vergleichsstruktur entspricht, welche die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,  
 wobei für jede dieser Überlagerungszuordnungen für eine Referenzstruktur mit einer Vergleichsstruktur ein Qualitätsmaß oder eine Abschätzung eines Qualitätsmaßes für die Ähnlichkeit der beiden Strukturen bestimmt wird,  
 und für ein oder mehrere Referenzmoleküle der Wert des Qualitätsmaßes einer Überlagerungszuordnung einer Struktur dieses Referenzmoleküls mit einer Struktur eines Vergleichsmoleküls, zusammen mit der zugehörigen Überlagerungszuordnung, als der geltende optimale Wert des Qualitätsmaßes oder als einer der geltenden optimalen Werte des Qualitätsmaßes gespeichert wird, falls ein Qualitätskriterium für den oder die geltenden optimalen Werte des Qualitätsmaßes erfüllt ist.

**[0011]** Vorzugsweise ist die Zahl der geltenden optimalen Werte des Qualitätsmaßes, die gespeichert werden, auf  $N$  begrenzt, wobei  $N$  eine natürliche Zahl größer oder gleich 1 ist. Der Wert  $N$  kann insbesondere eins sein. In diesem Fall kann zum Beispiel für ein Referenzmolekül jeweils immer nur der jeweils beste Wert des Qualitätsmaßes gespeichert werden.

**[0012]** Gemäß einer Ausführungsform der Erfindung kann das Qualitätskriterium eine oder mehrere der folgenden Bedingungen beinhalten:

- (a) es ist noch kein geltender optimaler Wert des Qualitätsmaßes gespeichert worden,
- (b) das Qualitätsmaß der besagten Überlagerungszuordnung ist besser als der schlechteste der geltenden optimalen Werte des Qualitätsmaßes,
- (c) der Wert des Qualitätsmaßes der besagten Überlagerungszuordnung ist besser als ein vorgegebener Schwellenwert des Qualitätsmaßes.

**[0013]** Das Qualitätskriterium kann dabei außer den Bedingungen (a), (b) und/oder (c) noch weitere Bedingungen beinhalten. Beispielsweise kann als Bedingung für eine Speicherung vorgesehen sein, daß die Zahl der geltenden optimalen Werte des Qualitätsmaßes, die bislang für dieses Referenzmolekül gespeichert wurden, kleiner als  $N$  ist, wobei  $N$  eine vorgegebene natürliche Zahl größer oder gleich 1 ist und die maximale Anzahl der geltenden optimalen Werte des Qualitätsmaßes angibt, die gespeichert werden. Dies stellt eine Erweiterung von (a) dar.

**[0014]** Wird nur ein optimaler Wert gespeichert ( $N = 1$ ), bedeutet die Bedingung (b), daß der Wert des Qualitätsmaßes nur dann gespeichert wird, wenn er besser ist als der geltende optimale Wert des Qualitätsmaßes. Ist  $N > 1$ , können maximal  $N - 1$  der geltenden optimalen Werte des Qualitätsmaßes, die für dieses Referenzmolekül gespeichert wurden, besser als das Qualitätsmaß der besagten Überlagerungszuordnung sein, wenn die Zahl der geltenden optimalen Werte auf  $N$  begrenzt ist. Die Bedingung (a) kann insbesondere dann erfüllt sein, wenn für dieses Referenzmolekül noch kein Qualitätsmaß für eine andere Überlagerungszuordnung berechnet wurde.

**[0015]** Mit der Bedingung (c) kann eine Mindestqualität der Überlagerungszuordnung vorgegeben werden.

Dies ist insbesondere sinnvoll, wenn eine Verbesserung gegenüber bereits bekannten ähnlichen Molekülen oder Strukturen erreicht werden soll oder wenn, z.B. aufgrund von stereochemischen Kriterien oder aufgrund von empirischen Daten, feststeht oder vorgegeben wird, daß Vergleichsmoleküle oder Vergleichsstrukturen, die ein bestimmtes Mindestmaß an Ähnlichkeit nicht aufweisen, kein sinnvolles Ergebnis bilden.

**[0016]** Werden die Bedingungen (b) und (c) kombiniert, wird der Wert nur dann als einer der geltenden optimalen Werte gespeichert, wenn er besser als der Schwellenwert ist. Werden (a) und (c) kombiniert und wird der Wert nur gespeichert, wenn er besser als der Schwellenwert ist. Es kann daher kein geltender optimaler Wert des Qualitätsmaßes gespeichert sein, obwohl in dem Verfahren schon mehrere Werte des Qualitätsmaßes bestimmt wurden. Wird dagegen die Bedingung (b) ohne die Bedingung (c) verwendet, werden die besten geltenden Werte, vorzugsweise die N besten geltenden Werte ( $N \geq 1$ ) gespeichert, ohne daß es darauf ankommt, wo dieser Wert bzw. diese N Werte auf einer absoluten Skala liegen.

**[0017]** Ist  $N = 1$ , läßt sich die Kombination der Kriterien (a), (b) und (c) dadurch implementieren, daß der Wert der Variablen, welche den geltenden optimalen Wert des Qualitätsmaßes angibt, zu Beginn des Verfahrens auf den besagten Schwellenwert gesetzt wird.

**[0018]** Die Verwendung der Bedingung (c) allein kann z.B. sinnvoll sein, wenn als Schwellenwert ein sehr guter Wert des Qualitätsmaßes vorgegeben wird und man grundsätzlich an allen Lösungen interessiert ist, die diesen Wert verbessern. In der Regel wird man allerdings auch in diesem Fall meistens die Zahl der Lösungen mit einem hinreichend großen N begrenzen, um den für die gespeicherten Lösungen verwendeten Speicherplatz zu begrenzen.

**[0019]** Das Qualitätskriterium kann, muß aber nicht in allen Iterationen des Verfahrens dasselbe sein. Beispielsweise kann vorgesehen sein, daß der vorangehend unter (c) genannte Schwellenwert im Laufe des Verfahrens geändert wird, wobei die Art und Weise, wie sich dieser Schwellenwert ändert, vorab festgelegt sein kann.

**[0020]** Gemäß einer Ausführungsform der Erfindung werden ein Referenzmolekül und mehrere Vergleichsmoleküle vorgegeben, wobei zu dem Referenzmolekül und den Vergleichsmolekülen jeweils mehrere Strukturen als Referenzstrukturen bzw. Vergleichsstrukturen vorgegeben sein können.

**[0021]** Das erfindungsgemäße Verfahren kann insbesondere beinhalten, daß für alle Paare aus einer Referenzstruktur (erste Referenzstruktur, sekundäre Referenzstrukturen) und einer Vergleichsstruktur (erste Vergleichsstruktur, sekundäre Vergleichsstrukturen) eine Überlagerungszuordnung bestimmt wird, ggf. mit der Einschränkung, daß nur solche Paare bzw. Überlagerungszuordnungen berücksichtigt werden, welche nicht bereits vorab ausgeschlossen wurden, z.B. weil festgestellt wurde, daß für diese Überlagerungszuordnung und zugehörige erweiterte Überlagerungszuordnungen, welche für das gleiche Paar von Referenzstruktur und Vergleichsstruktur denselben Komponenten der Referenzstruktur dieselben Komponenten der Vergleichsstruktur zuordnen, das Qualitätskriterium nicht erfüllt werden kann. Gemäß einer Ausführungsform der Erfindung werden in der Iteration alle sekundären Vergleichsstrukturen, welche in der vorgegebenen Kompatibilitätsbeziehung zu der ersten Vergleichsstruktur stehen, und/oder alle sekundären Referenzstrukturen in der Iteration herangezogen, welche in der besagten Kompatibilitätsbeziehung zu der ersten Referenzstruktur stehen, mit der Maßgabe, daß die entsprechenden Überlagerungszuordnungen nicht bereits vorab ausgeschlossen wurden.

**[0022]** Eine Komponente, aus der sich ein Referenzmolekül zusammensetzt, kann ein Atom oder eine zusammenhängende Gruppe von Atomen sein, die durch chemische Bindungen miteinander verbunden sind, beispielsweise eine lineare Teilkette oder eine Teilstruktur in einem organischen Molekül. Es kann sinnvoll sein, statt einzelnen Atomen Atomgruppen zu betrachten, zum Beispiel wenn sich alle betrachteten Moleküle in den Teilen, in denen sie verglichen werden, in solche Atomgruppen zerlegen lassen und es auf strukturelle Einzelheiten solcher Atomgruppen für den Vergleich nicht ankommt.

**[0023]** Die besagten Komponenten müssen nicht das vollständige Molekül bilden, sondern geben im Regelfall nur die für die Ähnlichkeit relevanten Teile des Moleküls wieder. Beispielsweise kann vorgesehen sein, daß man bei dem Vergleich von organischen Verbindungen Wasserstoffatome nicht berücksichtigt. Gemäß den bevorzugten Ausführungsformen bestehen die für den Vergleich signifikanten Teile des Moleküls aus den besagten Komponenten. Es kann auch vorgesehen sein, daß die zu vergleichende Moleküle vollständig oder im wesentlichen vollständig aus den besagten Komponenten bestehen.

**[0024]** Die Erfindung kann vorsehen, daß zu einem Referenzmolekül und/oder einem Vergleichsmolekül In-

formationen zu mehreren möglichen Strukturen des Referenzmoleküls bzw. des Vergleichsmoleküls, insbesondere mehreren vollständigen Strukturen, gespeichert sind. Die Erfindung kann allerdings auch vorsehen, daß zu einem, mehreren oder allen Referenzmolekülen Informationen zu nur einer Struktur bereitgestellt werden und/oder zu einem, mehreren oder allen Vergleichsmolekülen Informationen zu nur einer Struktur bereitgestellt werden.

**[0025]** In den besagten Informationen zu den Referenzstrukturen bzw. Vergleichsstrukturen kann die Lage aller Komponenten des Referenzmoleküls bzw. des Vergleichsmoleküls, aber auch gegebenenfalls nur die Lage bestimmter Komponenten enthalten sein. Dies richtet sich im wesentlichen danach, welche Information für den Vergleich zweier Moleküle wichtig ist.

**[0026]** Die sekundäre Referenzstruktur kann demselben, aber auch einem anderen Referenzmolekül wie die erste Referenzstruktur zugeordnet sein, ebenso, wie die sekundäre Vergleichsstruktur einem anderen oder demselben Vergleichsmolekül wie die erste Vergleichsstruktur zugeordnet sein kann. Gemäß bevorzugten Ausführungsformen ist die sekundäre Referenzstruktur demselben Referenzmolekül wie die erste Referenzstruktur und/oder die sekundäre Vergleichsstruktur demselben Vergleichsmolekül wie die erste Vergleichsstruktur zugeordnet.

**[0027]** Eine Iteration in dem vorangehend genannten Sinne muß sich nicht notwendigerweise identisch wiederholen, sondern ist als Teilprozeß zu verstehen, der in ähnlicher, aber nicht notwendigerweise identischer Weise wiederkehrt. Typischerweise wird in einer Iteration zumindest eine Überlagerungszuordnung einer Referenzstruktur mit einer Vergleichsstruktur gebildet und hierzu entweder ein Qualitätsmaß bestimmt oder abgeschätzt wird, um feststellen zu können, ob das Qualitätsmaß dieser Überlagerungszuordnung überhaupt das Qualitätskriterium für die geltenden optimalen Werte des Qualitätsmaßes erfüllen kann, z.B. ob das Qualitätsmaß besser als die N besten bisherigen Werte des Qualitätsmaßes bzw. der beste bisherige Wert des Qualitätsmaßes sein kann.

**[0028]** Eine Überlagerungszuordnung in dem vorangehend genannten Sinne ist so zu verstehen, daß man jeder Komponente aus einer ausgewählten Gruppe von Komponenten des Referenzmoleküls (erste Referenzmenge) jeweils genau eine Komponente aus einer gleichgroßen Gruppe von Komponenten des Vergleichsmoleküls (erste Vergleichsmenge) zuordnet. Stereochemisch gesehen bedeutet dies, wenn man Moleküle ähnlicher dreidimensionaler Strukturen zu Referenzmolekülen sucht, daß man festlegt, welche Komponenten, beispielsweise Atome, des Referenzmoleküls möglichst gut mit Komponenten des Vergleichsmoleküls zur Deckung gebracht werden sollen. Mathematisch würde man diesen chemischen Sachverhalt als umkehrbar eindeutige (bijektive) Abbildung der ersten Referenzmenge auf die erste Vergleichsmenge beschreiben.

**[0029]** Erfindungsgemäß wird nun in einer Iteration nicht nur, wie nach dem Stand der Technik, jeweils eine Referenzstruktur mit einer Vergleichsstruktur verglichen. Vielmehr wurde erfindungsgemäß erkannt, daß man eine erhebliche Zeitersparnis erreichen kann, wenn man ähnliche oder teilweise ähnliche Strukturen in der gleichen Iteration vergleicht. Was in diesem Sinne ähnlich ist, bestimmt sich im wesentlichen durch chemische bzw. stereochemische Kriterien, die in dem computergestützten Prozeß durch die Kompatibilitätsbeziehung implementiert werden.

**[0030]** Die Erfindung kann vorsehen, daß die Kompatibilitätsbeziehung zwischen der primären Referenzstruktur und der sekundären Referenzstruktur darin besteht, daß die besagte Teilstruktur der sekundären Referenzstruktur die gleiche Konstitution oder Konfiguration aufweist wie die durch die erste Referenzmenge definierte Teilstruktur der ersten Referenzstruktur und/oder daß die Kompatibilitätsbeziehung zwischen der ersten Vergleichsstruktur und der sekundären Vergleichsstruktur darin besteht, daß die Teilstruktur der sekundären Vergleichsstruktur die gleiche Konstitution oder Konfiguration aufweist wie die durch die erste Vergleichsmenge definierte Teilstruktur der ersten Vergleichsstruktur.

**[0031]** Eine ergänzende oder alternative Kompatibilitätsbeziehung könnte beispielsweise auch sein, daß die Komponenten, welche durch die Kompatibilitätsbeziehung den Komponenten der ersten Referenzmenge bzw. der ersten Vergleichsmenge zugeordnet werden, innerhalb bestimmter, vorgegebener Toleranzen relativ zueinander die gleiche Lage haben wie die entsprechenden Komponenten in dem Referenzmolekül bzw. in dem Vergleichsmolekül.

**[0032]** Es kann insbesondere vorgesehen sein, daß die gesamte sekundäre Referenzstruktur bzw. sekundäre Vergleichsstruktur eine Kompatibilitätsbeziehung zu der ersten Referenzstruktur bzw. der ersten Vergleichsstruktur aufweist, also daß z.B. die erste Referenzstruktur und die sekundäre Referenzstruktur jeweils die glei-

che Konstitution oder Konfiguration aufweisen.

**[0033]** Die Kompatibilitätsbeziehung, die für die sekundäre Referenzstruktur bzw. die sekundäre Vergleichsstruktur zugrunde gelegt wird, kann in beiden Fällen dieselbe sein. Dies ist jedoch nicht zwingend.

**[0034]** Gemäß einer besonders vorteilhaften Anwendung des erfindungsgemäßen Verfahrens kann vorgesehen sein, daß die Kompatibilitätsbeziehung darin besteht, daß die sekundäre Referenzstruktur ein Konformer zu der ersten Referenzstruktur und die sekundäre Vergleichsstruktur ein Konformer zu der ersten Vergleichsstruktur ist.

**[0035]** Die Erfindung kann vorsehen, daß in der besagten Iteration für alle Überlagerungszuordnungen, die in der Iteration bestimmt werden, entschieden wird, ob zu der jeweiligen Überlagerungszuordnung für alle erweiterten Überlagerungszuordnungen der betreffenden Referenzstruktur mit der betreffenden Vergleichsstruktur, welche zusätzlich zu den Zuordnungen der besagten Überlagerungszuordnung hinaus weitere Komponenten der Referenzstruktur umkehrbar eindeutig jeweils einer Komponente der Vergleichsstruktur zuordnen, das Qualitätskriterium für den oder die geltenden optimalen Werte des Qualitätsmaßes (z.B. ob der Wert des Qualitätsmaßes besser ist als einer der N geltenden optimalen Qualitätswerte des Qualitätsmaßes) nicht erfüllt werden kann, und daß in nachfolgenden Iterationen nur solche erweiterten Überlagerungszuordnungen berücksichtigt werden, für welche dies nicht der Fall ist.

**[0036]** Erfindungsgemäß kann vorgesehen sein, daß aus denjenigen Überlagerungszuordnungen, für die nicht festgestellt wurde, daß der Wert des Qualitätsmaßes für eine hierzu gehörige erweiterte Überlagerungszuordnung das Qualitätskriterium nicht erfüllen kann, ein Paar einer Überlagerungszuordnung, d.h. eine Referenzstruktur und eine Vergleichsstruktur, ausgewählt werden, wobei hierzu eine neue Referenzmenge gebildet wird, welche alle Komponenten der Referenzstruktur enthält, die in der besagten Überlagerungszuordnung einer Komponente der Vergleichsstruktur zugeordnet wurden, aber auch mindestens eine weitere zusätzliche Komponente des betreffenden Referenzmoleküls, und eine neue Vergleichsmenge gebildet wird, welche alle Komponenten der Vergleichsstruktur, die in der besagten Überlagerungszuordnung einer Komponente in der Referenzstruktur zugeordnet wurden, aber zumindest eine weitere Komponente der Vergleichsstruktur enthält, wobei diese neue Vergleichsmenge und diese neue Referenzmenge jeweils die gleiche Anzahl von Komponenten aufweisen. Die Referenzstruktur und die Vergleichsstruktur, die man hierfür auswählt, können, müssen aber nicht die erste Referenzstruktur und die erste Vergleichsstruktur der vorangegangenen Iteration sein.

**[0037]** Auf der Grundlage dieser neuen Referenzmenge und dieser neuen Vergleichsmenge bildet man nun eine erste erweiterte Überlagerungszuordnung, welche die Überlagerungszuordnung der Referenzstruktur zu der Vergleichsstruktur in der vorangehenden Iteration in dem Sinne umfaßt, daß diejenigen Zuordnungen von Komponenten der Referenzstruktur zu Komponenten der Vergleichsstruktur, die in der vorangehenden Iteration definiert wurden, erhalten bleiben und lediglich zu denjenigen Komponenten der neuen Referenzmenge und der neuen Vergleichsmenge, die bei der Überlagerungszuordnung in der vorangehenden Iteration noch nicht berücksichtigt waren, neue umkehrbar eindeutige Zuordnungen zwischen jeweils einer Komponente der Referenzstruktur und einer Komponente der Vergleichsstruktur definiert werden.

**[0038]** Gemäß einer bevorzugten Ausführungsform der Erfindung wird bei der Bildung der neuen Referenzmenge und der neuen Vergleichsmenge der ersten Referenzmenge und der ersten Vergleichsmenge jeweils genau eine weitere Komponente hinzugefügt. Die Anzahl der Komponenten in der neuen Referenzmenge und der neuen Vergleichsmenge ist also um 1 größer als die Anzahl der Komponenten in der ersten Referenzmenge und der ersten Vergleichsmenge.

**[0039]** Diese neue Referenzmenge und diese neue Vergleichsmenge entsprechen der ersten Referenzmenge bzw. der ersten Vergleichsmenge der vorangehenden Iteration und die entsprechend erweiterte Überlagerungszuordnung entspricht der ersten Überlagerungszuordnung der vorangehenden Iteration.

**[0040]** Weiterhin werden für alle Paare von Vergleichsstrukturen und Referenzstrukturen, die in der vorangehenden Iteration betrachtet wurden und für die nicht festgestellt wurde, daß das Qualitätsmaß für alle erweiterten Überlagerungszuordnungen zu der Überlagerungszuordnung zwischen den Strukturen des Paares, die in der vorangehenden Iteration betrachtet wurde, das Qualitätskriterium nicht erfüllen kann, erweiterte Überlagerungszuordnungen in der gleichen Weise wie in der vorangehenden Iteration auf der Grundlage der Kompatibilitätsbeziehung ausgehend von der ersten erweiterten Überlagerung gebildet, sofern sich auch für alle Komponenten der neuen Referenzmenge bzw. der neuen Vergleichsmenge eine Kompatibilitätsbeziehung in

dem vorangehend genannten Sinne herstellen läßt und mit der Maßgabe, daß die Zuordnung von Komponenten der zugehörigen Überlagerungszuordnung in der vorangehenden Iteration erhalten bleibt. Läßt sich für eine Überlagerungszuordnung einer Referenzstruktur und einer Vergleichsstruktur zu den Komponenten der neuen Referenzmenge und/oder der neuen Vergleichsmenge keine Kompatibilitätsbeziehung mehr herstellen oder führt diese Kompatibilitätsbeziehung, die vorzugsweise dieselbe ist wie in der vorangehenden Iteration, dazu, daß sich die Zuordnung von Komponenten gegenüber der entsprechenden Überlagerungszuordnung der vorangehenden Iteration ändern würde, wird in dieser Iteration zu dieser Überlagerungszuordnung keine erweiterte Überlagerungszuordnung bestimmt.

**[0041]** Für die in dieser nachfolgenden Iteration festgelegte Überlagerungszuordnungen wird nun wiederum in der vorangehend beschriebenen Weise das Qualitätsmaß bestimmt bzw. abgeschätzt und es wird der entsprechende Wert des Qualitätsmaßes, zusammen mit der entsprechenden Überlagerungszuordnung, als einer der geltenden optimalen Werte, insbesondere einer der N geltenden optimalen Werte, des Qualitätsmaßes abgespeichert, wenn das Qualitätskriterium erfüllt ist.

**[0042]** Stellt man in einer Iteration fest, daß die Überlagerungszuordnung nicht mehr erweiterbar ist, z.B. weil bereits alle Komponenten der Referenzstruktur einer Komponente der Vergleichsstruktur bzw. alle Komponenten einer Vergleichsstruktur einer Komponente der Referenzstruktur zugeordnet worden sind, oder daß alle zugehörigen erweiterten Überlagerungszuordnungen das Qualitätskriterium nicht erfüllen können, kehrt das Verfahren zu der Überlagerungszuordnung zurück, für die festgestellt wurde, daß es mehr als eine mögliche erweiterte Überlagerungszuordnung hierzu gibt und zu der noch nicht für alle möglichen erweiterten Überlagerungszuordnungen entweder das Qualitätsmaß bestimmt oder abgeschätzt wurde oder festgestellt wurde, daß das Qualitätskriterium nicht erfüllt werden kann, z.B. weil das Qualitätsmaß nicht besser sein kann als die N besten ermittelten Werte des Qualitätsmaßes. Hiervon ausgehend wird dann eine bzw. die verbleibende erweiterte Überlagerungszuordnung gebildet und dann entsprechend verfahren.

**[0043]** Gemäß einer Ausführungsform der Erfindung ist die sekundäre Referenzstruktur ein Konformer zu der ersten Referenzstruktur und die sekundäre Vergleichsstruktur ein Konformer zu der ersten Vergleichsstruktur.

**[0044]** Dabei kann vorgesehen sein, daß die Struktur jedes Konformers als starre Struktur behandelt wird, bei der jede Komponente, zum Beispiel jedes Atom, eine feste Position bezüglich den anderen Komponenten in dem Konformer besitzt.

**[0045]** Gemäß einer bevorzugten Ausführungsform der Erfindung werden genau ein Referenzmolekül, aber mehrere Vergleichsmoleküle betrachtet und es werden Moleküle gesucht, die diesem Referenzmolekül ähnlich sind.

**[0046]** Die Erfindung kann vorsehen, daß das Verfahren zu einem vorgegeben Referenzmolekül ähnliche Moleküle ermittelt, wobei dem Verfahren Informationen zur dreidimensionalen Struktur eines oder mehrere Konformere des Referenzmoleküls, welche die Lage von Atomen, allerdings nicht notwendigerweise allen Atomen, in dem Konformer angeben und Informationen zu der dreidimensionalen Struktur eines oder mehrere Konformere mindestens eines Vergleichsmoleküls, vorzugsweise mehrerer Vergleichsmoleküle, welche die Lage von Atomen, allerdings nicht notwendigerweise aller Atome, in dem Konformer angeben, zugrunde liegen und mindestens eine Iteration des Verfahrens umfaßt:

Auswahl einer ersten Referenzmenge von Atomen des Referenzmoleküls, wobei die Anzahl der Atome in der ersten Referenzmenge kleiner als die Anzahl der Atome in dem Referenzmolekül ist,

Auswahl einer ersten Vergleichsmenge von Atomen eines Vergleichsmoleküls, welche die gleiche Anzahl von Atomen wie die erste Referenzmenge besitzt,

Bestimmung einer ersten Überlagerungszuordnung für ein erstes Konformer des Referenzmoleküls und ein erstes Konformer des Vergleichsmoleküls, welche jedem Atom der ersten Referenzmenge umkehrbar eindeutig genau ein Atom der ersten Vergleichsmenge zuweist,

Bestimmung einer Überlagerungszuordnung für ein zweites Konformer des Referenzmoleküls zu einem Konformer des Vergleichsmoleküls und/oder für ein Konformer des Referenzmoleküls zu einem zweiten Konformer des Vergleichsmoleküls, welche jedem Atom der ersten Referenzmenge umkehrbar eindeutig genau dasselbe Atom der ersten Vergleichsmenge wie die erste Überlagerungszuordnung zuweist.

**[0047]** Die Erfindung kann vorsehen, daß die Iteration folgendes umfaßt:

Bestimmen eines Qualitätsmaßes oder einer Abschätzung für das Qualitätsmaß für jede Überlagerungszuordnung,

Speichern des Werts des Qualitätsmaßes als geltender optimaler Wert zusammen mit der zugehörigen Über-

lagerungszuordnung, falls ein Qualitätskriterium für den geltenden optimalen Wert erfüllt ist.

**[0048]** Das Qualitätskriterium kann dabei insbesondere die vorangehend genannte Bedingung (b) und/oder die vorangehend genannte Bedingung (c) enthalten, wobei die Zahl der geltenden optimalen Werte (N) eins ist. Es kann insbesondere vorgesehen sein, daß der Wert des Qualitätsmaßes als bislang geltender optimaler Wert gespeichert wird, wenn noch kein Qualitätsmaß für eine andere Überlagerungszuordnung berechnet wurde oder das Qualitätsmaß aller bisherigen Überlagerungszuordnungen schlechter war als das besagte Qualitätsmaß.

**[0049]** Gemäß einer bevorzugten Ausführungsform der Erfindung werden in der besagten Iteration alle Paare von Konformeren des Referenzmoleküls und Konformeren des Vergleichsmoleküls betrachtet, für welche nicht vorangehend festgestellt wurde, daß das Qualitätskriterium nicht erfüllt wird.

**[0050]** Die Erfindung kann vorsehen, daß die Iteration, welche der besagten Iteration folgt, umfaßt: Bestimmen einer erweiterten Überlagerungszuordnung für alle Paare von Konformeren des Referenzmoleküls und des Vergleichsmoleküls, für welche in einer früheren Iteration nicht festgestellt wurde, daß das Qualitätskriterium nicht erfüllt werden kann, wobei diese erweiterte Überlagerungszuordnung jeweils jedem Atom einer zweiten Referenzmenge von Atomen des Referenzmoleküls umkehrbar eindeutig jeweils genau ein Atom einer zweiten Vergleichsmenge von Atomen des Vergleichsmoleküls zuordnet, wobei die erste und die zweite Referenzmenge voneinander verschieden sind und alle Atome der ersten Referenzmenge in der zweiten Referenzmenge enthalten sind und die erste und die zweite Vergleichsmenge voneinander verschieden sind und alle Atome der ersten Vergleichsmenge in der zweiten Vergleichsmenge enthalten sind und wobei gemäß der erweiterten Überlagerungszuordnung jedes Atom aus der ersten Referenzmenge einem Atom aus der ersten Vergleichsmenge entsprechend der ersten Überlagerungszuordnung zugeordnet ist.

**[0051]** Gemäß einer Ausführungsform der Erfindung kann vorgesehen sein, daß verschiedene Paare aus einem Referenzmolekül und einem Vergleichsmolekül nacheinander abgearbeitet werden. In jeder Iteration wird also nur ein Paar aus einem Referenzmolekül und Vergleichsmolekül berücksichtigt und es wird erst dann zu einem neuen Paar aus Referenzmolekül und Vergleichsmolekül übergegangen, wenn alle möglichen Überlagerungszuordnungen zwischen allen Strukturen des Referenzmoleküls und allen Strukturen des Vergleichsmoleküls abgearbeitet worden sind, wobei das Abarbeiten auch und insbesondere beinhalten kann, daß eine Überlagerungszuordnung vorab ausgeschieden wird, weil festgestellt wurde, daß das Qualitätskriterium nicht erfüllt werden kann. Der beste Wert des Qualitätsmaßes (oder eine endliche Anzahl der besten Werte des Qualitätsmaßes) kann als Ergebnis des Vergleiches unabhängig von weiteren Vergleichen festgehalten werden.

**[0052]** Für den Vergleich des nächsten Pairs aus einem Referenzmolekül und einem Vergleichsmolekül kann vorgesehen sein, daß die N geltenden optimalen Werte zurückgesetzt werden. Es kann jedoch auch vorgesehen sein, z.B. wenn man nur an einem optimalen Paar von Referenzmolekül und Vergleichsmolekül interessiert ist, daß die N optimalen Werte des Qualitätsmaßes aus dem Vergleich eines anderen Pairs von Referenzmolekül und Vergleichsmolekül, als Anfangswerte für die N optimalen Werte für das aktuelle Paar verwendet wird oder auf der Grundlage der jeweils optimalen Werte des Qualitätsmaßes aus einem oder mehreren Vergleichen eines Referenzmoleküls und eines Vergleichsmoleküls ein Schwellenwert bestimmt wird, der in das Qualitätskriterium für den Vergleich des aktuellen Pairs aus Referenzmolekül und Vergleichsmolekül eingeht. Auf diese Weise ist es möglich, den Suchraum auf der Basis früherer Vergleiche eines Referenzmoleküls und eines Vergleichsmoleküls zu reduzieren.

**[0053]** Dieser Schwellenwert kann beispielsweise der beste Wert des Qualitätsmaßes sein, der bislang bei einem Vergleich eines Referenzmoleküls mit einem Vergleichsmolekül aufgefunden wurde, der kleinste der M besten bislang aufgefundenen Werte des Qualitätsmaßes, wobei M eine ganze Zahl größer oder gleich 1 ist, oder in einem vorgegebenen Verhältnis zu diesen Werten stehen (z.B. um einen bestimmten Prozentsatz über oder unter dem entsprechenden Wert liegen). Es kann vorgesehen sein, daß ein solcher Schwellenwert erst dann vorgegeben wird, wenn bereits eine bestimmte Anzahl von Paaren aus einem Vergleichsmolekül und einem Referenzmolekül abgearbeitet worden sind.

**[0054]** Es kann insbesondere vorgesehen sein, daß insgesamt nur ein Referenzmolekül betrachtet wird. Insbesondere kann vorgesehen sein, daß zu jedem Paar aus diesem Referenzmolekül und einem Vergleichsmolekül jeweils nur der optimale Wert des Qualitätsmaßes gespeichert wird.

**[0055]** Die Erfindung kann insbesondere vorsehen, daß zu jedem Paar eines Referenzmoleküls und eines Vergleichsmoleküls, ggf. geordnet nach dem jeweiligen Referenzmolekül, wenn mehrere Referenzmoleküle

betrachtet werden, der optimale Wert des Qualitätsmaßes gespeichert wird und eine Liste oder Tabelle der Paare (bzw., falls nur ein Referenzmolekül betrachtet wird, der Vergleichsmoleküle) und der zugehörigen Werte des Qualitätsmaßes erstellt wird. In einer Ausführungsform kann diese Liste auf diejenigen Vergleichsmoleküle beschränkt sein, welche einen Wert aus einer vorgegebenen Anzahl von besten Werten des Qualitätsmaßes ergeben haben.

**[0056]** Es kann vorgesehen sein, daß in dem Referenzmolekül die Komponenten der ersten Referenzmenge zusammen mit den Bindungen zwischen ihnen einen zusammenhängenden Teilabschnitt des Referenzmoleküls bilden und die Komponenten der ersten Vergleichsmenge zusammen mit den Bindungen zwischen ihnen in dem Vergleichsmolekül einen zusammenhängenden Teilabschnitt des Vergleichsmoleküls bilden.

**[0057]** Bei dieser Ausführungsform würden also die Komponenten der ersten Referenzmenge, wenn man sie aus dem Referenzmolekül zusammen mit den zwischen ihnen bestehenden Bindungen herauslösen würde, ein einziges zusammenhängendes (Teil)molekül bilden. Anders gesprochen kann man entlang den Bindungen zwischen den Komponenten der ersten Referenzmenge von jeder Komponente der ersten Referenzmenge zu einer beliebigen anderen Komponente der ersten Referenzmenge gelangen. Dies wäre von dem Fall zu unterscheiden, daß es eine oder mehrere Gruppen von Komponenten in der ersten Referenzmenge gibt, die nicht durch mindestens eine Bindung miteinander verbunden sind, so daß, wenn man diese Komponenten aus den Molekülen herauslösen würde, mehrere Teilmoleküle entstehen würden. Entsprechendes gilt gemäß dieser Ausführungsform auch für die Vergleichsmenge.

**[0058]** Entsprechend kann bei dieser Ausführungsform auch vorgesehen sein, daß die Kompatibilitätsbeziehung zwischen den Komponenten der ersten Referenzmenge bzw. den Komponenten der ersten Vergleichsmenge zu den Komponenten der sekundären Referenzstruktur bzw. der sekundären Vergleichsstruktur beinhaltet, daß den Komponenten der ersten Referenzmenge bzw. der ersten Vergleichsmenge jeweils ein zusammenhängender Teil der sekundären Referenzstruktur bzw. der sekundären Vergleichsstruktur durch die Kompatibilitätsbeziehung zugewiesen wird, d.h. daß die entsprechenden Komponenten der sekundären Referenzstruktur bzw. der sekundären Vergleichsstruktur, die aufgrund der Kompatibilitätsbeziehung jeweils einer Komponente der ersten Referenzmenge bzw. der ersten Vergleichsmenge entsprechen, in der gleichen Weise zusammenhängen müssen wie die Komponenten der ersten Vergleichsmenge bzw. der ersten Referenzmenge.

**[0059]** Die Erfindung kann vorsehen, daß das Qualitätsmaß den Abstand einander durch die Überlagerungszuordnung zugeordneter Komponenten, insbesondere den Abstand der Atome, berücksichtigt, der sich ergibt, wenn die Referenzstruktur mit der Vergleichsstruktur optimal zur Deckung gebracht wird.

**[0060]** Hierbei können für das Kriterium, daß die Strukturen optimal zur Deckung gebracht worden sind, übliche Kriterien angewendet werden, beispielsweise daß der mittlere quadratische Abstand der Komponenten minimal ist oder dergleichen.

**[0061]** Es kann auch vorgesehen sein, daß in das Qualitätsmaß die Anzahl der Komponenten der Referenzstruktur eingeht, welche durch die Überlagerungszuordnung umkehrbar eindeutig jeweils einer Komponente der Vergleichsstruktur zugeordnet werden.

**[0062]** Die Erfindung kann auch vorsehen, daß weitere Kriterien in das Qualitätsmaß eingehen, beispielsweise die chemische Natur der einander überlagerten Komponenten. Betrachtet man größere Einheiten von Atomen als Komponenten des Referenzmoleküls, so kann a priori bereits ein Qualitätsmaß festgelegt werden, welches berücksichtigt, ob die Komponenten, die einander zugeordnet werden, identisch sind oder nicht oder, in einer Weiterentwicklung dieser Ausführungsform, inwieweit diese Komponenten, für sich genommen, d.h. unabhängig von ihrer Lage in dem Referenzmolekül bzw. in dem Vergleichsmolekül, räumlich zur Deckung gebracht werden können. Beispielsweise kann ein Beitrag zu dem Qualitätsmaß aufgrund des letzteren Kriteriums dadurch festgelegt werden, daß die Schwerpunkte der isolierten Komponenten aufeinander gelegt werden und betrachtet wird, inwieweit sich die relevanten Atome der beiden Komponenten räumlich zur Deckung bringen lassen. Hierfür kann ein pauschalierter Beitrag zu dem Qualitätsmaß bestimmt werden, der in den Iterationen des Verfahrens bei der Bestimmung des Qualitätsmaßes berücksichtigt wird, ohne daß in den jeweiligen Iterationen die Überlagerung der Atome der einander zugeordneten Komponenten betrachtet wird. Man geht dann vielmehr davon aus, daß die Lage jeder Komponente durch einen Satz von dreidimensionalen Koordinaten festgelegt ist.

**[0063]** Nach einer Ausführungsform der Erfindung ist das Qualitätsmaß durch die folgende Formel gegeben:

$$SC = (N_U/N_{Mmin})C^{-rmsd}, \quad (1)$$

wobei

- $N_U$  die Anzahl der Komponenten der Referenzstruktur ist, die durch die Überlagerungszuordnung jeweils einer Komponente der Vergleichsstruktur zugewiesen werden,  
 $N_{Mmin}$  den kleineren Wert aus dem Paar angibt, welches durch die Anzahl der zu vergleichenden Komponenten in dem Referenzmolekül und die Anzahl der zu vergleichenden Komponenten in dem Vergleichsmolekül gebildet wird, mit anderen Worten die Anzahl der zu vergleichenden Komponenten in dem kleineren der beiden Moleküle,  
 $rmsd$  der mittlere quadratische Abstand der einander durch die Überlagerungszuordnung zugeordneten Komponenten ist und  
 $C$  eine Konstante ist.

**[0064]** Die Erfindung kann insbesondere vorsehen, daß  $rmsd$  der minimale mittlere quadratische Abstand ist, der erreicht werden kann, wenn die einander durch die Überlagerungszuordnung zugeordneten Komponenten der Referenzstruktur und der Vergleichsstruktur einander überlagert werden, d.h. wenn man versucht, die Lage dieser Komponenten in den beiden Strukturen möglichst gut zur Deckung zu bringen.

**[0065]** Der mittlere quadratische Abstand kann dabei durch die folgende Gleichung bestimmt werden:

$$rmsd = \sqrt{\frac{1}{N_U} \sum_{i=1}^{N_U} \|X_{Ri} - X_{Vi}\|^2}$$

wobei  $X_{Ri}$  den Ort der i-ten Komponente der Referenzstruktur und  $X_{Vi}$  den Ort der ihr zugehörigen i-ten Komponente der Vergleichsstruktur bezeichnet, die sich ergeben, wenn man versucht, unter Berücksichtigung der vorgegebenen Referenzstruktur und der vorgegebenen Vergleichsstruktur, für alle Komponenten, die einander in der Überlagerungsstruktur zugeordnet sind, möglichst nahe beieinander liegende Positionen, idealerweise identische Positionen zu finden.  $\| \|$  bezeichnet die euklidische Norm.

**[0066]** Das Qualitätsmaß nach (1) ist so definiert, daß eine bessere Übereinstimmung der verglichenen Strukturen zu einem höheren Wert des Qualitätsmaßes führt, wobei der maximale Wert 1 ist. Ein Wert des Qualitätsmaßes ist also besser als ein anderer Wert, wenn er größer ist. Ein Qualitätsmaß kann jedoch auch so definiert sein, daß das Qualitätsmaß um so kleiner ist, je besser die Übereinstimmung der verglichenen Strukturen ist. In diesem Fall bedeutet ein besserer Wert des Qualitätsmaßes einen kleineren Wert.

**[0067]** Nach einer bevorzugten Ausführungsform folgt das Verfahren einem Branch and Bound-Algorithmus.

**[0068]** Die Erfindung kann vorsehen, daß für eine Überlagerungszuordnung einer Referenzstruktur zu einer Vergleichsstruktur, bei welcher die Komponenten der Referenzstruktur, die Komponenten der Vergleichsstruktur zugeordnet werden, einen zusammenhängenden Teilabschnitt des Referenzmoleküls bilden, überprüft wird, welche erweiterten Überlagerungszuordnungen, bei denen die Zuordnung der Komponenten entsprechend der besagten Überlagerungszuordnung erhalten bleibt, existieren, bei denen die Komponenten eines zusammenhängenden Teils der Referenzstruktur einem zusammenhängenden Teil der Vergleichsstruktur zugeordnet werden, das Qualitätsmaß für diese möglichen erweiterten Überlagerungszuordnungen abgeschätzt wird und diese erweiterten Überlagerungszuordnungen in späteren Iterationen nicht berücksichtigt werden, wenn die Abschätzung ergibt, daß das Qualitätskriterium nicht erfüllt werden kann, zum Beispiel weil das Qualitätsmaß schlechter ist als die N besten Werte des Qualitätsmaßes, die bislang ermittelt wurden.

**[0069]** Dieses Abschneidekriterium ist zum Beispiel dann sinnvoll, wenn die Erweiterung der Überlagerungszuordnung entlang einer linearen Kette von Atomen, oder, allgemeiner gesprochen, Komponenten vorgenommen wird. In diesem Fall kann man absehen, mit wie vielen Erweiterungen um jeweils eine weitere Komponente bzw. ein weiteres Atom man zu dem Ende der Kette gelangt, an dem die Überlagerungszuordnung, jedenfalls in diese Richtung, nicht erweitert werden kann. In diesem Fall kann man die möglichen Erweiterungen der Überlagerungszuordnungen bestimmen und hierfür einfache Abschätzungen vornehmen. Wenn beispielsweise die vorangehend genannten Formel (1) verwendet wird, gilt in jedem Fall

$$SC \leq N_U/N_{Mmin}. \quad (2)$$

**[0070]** Wenn  $rmsd$  den quadratischen mittleren Abstand für diejenige gegenseitige Orientierung der Referenzstruktur und der Vergleichsstruktur bezeichnet, die die Überlagerungszuordnung ergibt, die den minimalen quadratischen mittleren Abstand ergibt, dann kann die Erfindung vorsehen, daß die Überlagerungszuordnung, die den minimalen quadratischen mittleren Abstand ergibt, die Überlagerungszuordnung ist, die die Überlagerungszuordnung ergibt, die den minimalen quadratischen mittleren Abstand ergibt.

renzstruktur und der Vergleichsstruktur angibt, bei welcher alle Komponenten in der Überlagerungszuordnung mit  $N_0$  überlagerten Komponenten bestmöglichst zur Deckung gebracht werden, so kann der Beitrag des Quadrats der Differenzen dieser Abstände zu dem quadratischen mittleren Abstand für eine erweiterte Überlagerungszuordnung, bei der  $m$  weitere Atome hinzugekommen sind, nur größer sein. Andererseits kann der Beitrag, den die hinzugekommenen Komponenten leisten, minimal 0 sein.

**[0071]** Dementsprechend läßt sich der mittlere Abstand für eine Überlagerungszuordnung mit  $N_0 + m$  Komponenten,  $rmsd(N_0 + m)$ , durch den quadratischen mittleren Abstand für eine Überlagerungszuordnung mit  $N_0$  Komponenten,  $rmsd(N_0)$ , wie folgt abschätzen:

$$rmsd(N_0 + m) \leq \sqrt{\frac{N_0}{N_0 + m}} rmsd(N_0) \quad (3)$$

**[0072]** Mit (2) und/oder (3) kann man zum Beispiel das Qualitätsmaß SC nach oben abschätzen.

**[0073]** Eine weitere Möglichkeit, den Beitrag von  $rmsd$  in (1) abzuschätzen, besteht darin, daß man den quadratischen mittleren Abstand durch die Differenz inneratomarer Abstände abschätzt. Es läßt sich zeigen, daß der Beitrag zum mittleren Abstand einer Überlagerungszuordnung zweier beliebiger Paare von Komponenten durch die Differenz der Abstände der jeweiligen Komponenten in dem jeweiligen Molekül nach unten abgeschätzt werden kann.

**[0074]** Die Erfindung kann vorsehen, daß für alle Konformere des Referenzmoleküls bzw. der Vergleichsmoleküle für die Abstände der Atome in dem jeweiligen Molekül Abstandsintervalle gespeichert sind, welche die Obergrenze und die Untergrenze dieses Abstandes für alle Konformere angibt, und daß bei der Abschätzung des Qualitätsmaßes für eine Überlagerungszuordnung eines Konformers eines Referenzmoleküls mit einem Konformer eines Vergleichsmoleküls eine Abschätzung des Qualitätsmaßes dadurch berechnet wird, daß auf der Grundlage der besagten Bereichsgrenzen das beste Qualitätsmaß berechnet wird, das sich für beliebige Werte innerhalb der Grenzen der Abstandsintervalle für die inneratomaren Abstände ergibt und überprüft wird, ob dieses bestmögliche Qualitätsmaß das Qualitätskriterium erfüllt, zum Beispiel indem dieses bestmögliche Qualitätsmaß mit dem geltenden optimalen Wert des Qualitätsmaßes bzw. den  $N$  geltenden optimalen Werten des Qualitätsmaßes verglichen wird.

**[0075]** Ist dieser Wert schlechter, im Falle der Formel (1) beispielsweise kleiner als der beste gespeicherte Wert des Qualitätsmaßes SC, muß man das Qualitätsmaß für die einzelnen Paare von Konformeren nicht mehr berechnen. Ist dieser bestmögliche Wert besser als der beste gespeicherte Wert oder einer der  $N$  besten gespeicherten Werte, kann man unter den möglichen Paaren von Konformeren dasjenige ermitteln, welches den besten Wert des Qualitätsmaßes ergibt. Es kann dabei vorgesehen sein, innerhalb der besagten Abstandsintervalle Unterintervalle vorab zu definieren, welche nur den Wertebereich der Atomabstände für eine bestimmte Gruppe von Konformeren in dem Molekül abdecken. Stellt sich heraus, daß für ein solches Unterintervall der bestmögliche Wert schlechter als der beste Wert ist, muß das Paar von Konformeren mit dem besten Wert außerhalb dieser Gruppe liegen.

**[0076]** Die Erfindung stellt auch ein Computersystem zum Durchführen eines iterativen Verfahrens zum Auffinden von Molekülen mit strukturellen Ähnlichkeiten zu einem oder mehreren Referenzmolekülen, die mehrere Komponenten aufweist, zur Verfügung, welches umfaßt: eine Einrichtung zum Speichern von Informationen zu einer oder mehreren Referenzstrukturen, welche jeweils eine Struktur eines Referenzmoleküls darstellen, wobei die besagten Informationen die Lage von Komponenten in dem Referenzmolekül gemäß dieser Referenzstruktur, insbesondere in einem dreidimensionalen Raum, angeben, und von Informationen zu mehreren Vergleichsstrukturen, welche jeweils eine Struktur eines Vergleichsmoleküls darstellen, wobei die zugehörigen Informationen die Lage von Komponenten in dem Vergleichsmolekül in dieser Vergleichsstruktur, insbesondere in einem dreidimensionalen Raum, angeben, mit einer Einrichtung zum Vergleichen eines Teils eines Referenzmoleküls und eines Teils eines Vergleichsmoleküls in einer Iteration des Verfahrens, die jeweils einen Teil des Referenzmoleküls bzw. des Vergleichsmoleküls enthalten, die in einer früheren Iteration verglichen worden sind, wobei das Computersystem eine Einrichtung zum Ausführen der folgenden Schritte in mindestens einer Iteration umfaßt:

Auswahl einer ersten Referenzmenge von Komponenten eines Referenzmoleküls, welche einen Teil der Komponenten des Referenzmoleküls enthält, wobei die Anzahl der Komponenten in der ersten Referenzmenge kleiner als die Gesamtzahl der zu vergleichenden Komponenten in dem Referenzmolekül ist,

Auswahl einer ersten Vergleichsmenge von Komponenten eines Vergleichsmoleküls, welche die gleiche An-

zahl von Komponenten wie die erste Referenzmenge besitzt,  
 Bestimmen einer ersten Überlagerungszuordnung für eine erste Referenzstruktur, die dem besagten Referenzmolekül zugeordnet ist, und eine erste Vergleichsstruktur, welche dem besagten Vergleichsmolekül zugeordnet ist, wobei die erste Überlagerungszuordnung jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau eine Komponente der ersten Vergleichsmenge zuweist,  
 wobei für mindestens eine weitere, sekundäre Referenzstruktur, welche zumindest eine Teilstruktur aufweist, die zu der primären ersten Referenzstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Referenzmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht,  
 und/oder für mindestens eine weitere, sekundäre Vergleichsstruktur, welche eine Teilstruktur aufweist, die zu der ausgewählten primären Vergleichsstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Vergleichsmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht, bestimmt wird:  
 eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die primäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmenge die entsprechende Komponente der sekundären Referenzstruktur derjenigen Komponente der ersten Vergleichsstruktur zuweist, welches die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,  
 und/oder  
 eine Überlagerungszuordnung für die erste Referenzstruktur und die sekundäre Vergleichsstruktur, welche jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau die Komponente zuweist, welche aufgrund der Kompatibilitätsbeziehung der sekundären Vergleichsstruktur derjenigen Komponente der ersten Vergleichsmenge entspricht, welche die erste Überlagerungszuordnung der betreffenden Komponente der ersten Referenzmenge zuweist,  
 und/oder  
 eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die sekundäre Vergleichsstruktur, welche für jede Komponente die ersten Referenzmengen die entsprechende Komponente der sekundären Referenzstruktur derjenigen Komponente der sekundären Vergleichsstruktur zuweist, welche der Komponente der ersten Vergleichsstruktur entspricht, welche die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,  
 und das Computersystem weiterhin eine Einrichtung zum Bestimmen oder Abschätzen eines Qualitätsmaßes aufweist, wobei das Computersystem weiterhin dafür eingerichtet ist, in dieser Iteration für eine oder mehrere Referenzmoleküle den Wert des Qualitätsmaßes einer Überlagerungszuordnung einer Struktur dieses Referenzmoleküls mit einer Struktur eines Vergleichsmoleküls zusammen mit der zugehörigen Überlagerungszuordnung als geltenden optimalen Wert des Qualitätsmaßes oder als einen der geltenden optimalen Werte des Qualitätsmaßes zu speichern, falls ein Qualitätskriterium für den oder die geltenden optimalen Werte des Qualitätsmaßes erfüllt ist.

**[0077]** Ein Computersystem ist dabei als geeignete Datenverarbeitungseinrichtung anzusehen. Es kann insbesondere mehrere Computer umfassen, die miteinander zusammenwirken, aber auch nur aus einem Computer bestehen.

**[0078]** Ein erfindungsgemäßes Computersystem kann eine oder mehrere Einrichtungen zum Durchführen eines Verfahrens wie vorangehend umschrieben aufweisen.

**[0079]** Die Erfindung stellt auch ein Computerprogramm zur Verfügung, welches Anweisungen für ein Computersystem enthält, die, wenn sie auf einem Computersystem ausgeführt werden, das Computersystem veranlassen, ein Verfahren wie vorangehend beschrieben auszuführen.

**[0080]** Die Erfindung stellt auch ein computerlesbares Speichermedium zur Verfügung, bei dem, insbesondere in maschinenlesbarer Form, ein Programm wie vorangehend beschrieben gespeichert ist.

**[0081]** Dieses computerlesbare Medium kann insbesondere computerlesbaren Programmcode enthalten, der, wenn er von einem Computer ausgeführt, den Computer veranlaßt, ein Verfahren wie vorangehend beschrieben auszuführen.

**[0082]** Das erfindungsgemäße Verfahren beruht auf der Erkenntnis, daß chemische Ähnlichkeiten der verglichenen Strukturen, etwa die strukturelle Ähnlichkeit von Konformeren, in vorteilhafter Weise benutzt werden können, um einerseits Rechenschritte, die sich für solche chemisch ähnlichen Strukturen wiederholen, einzusparen, indem die Verarbeitung von Paaren entsprechender Strukturen in dem gleichen Verfahrensabschnitt erfolgt, und andererseits aber auch, das Branch and Bound-Verfahren wesentlich effizienter zu gestalten, da

zu erwarten ist, daß jedenfalls bis zu einem gewissen Grade derartige ähnliche Strukturen zu einem ähnlichen Qualitätsmaß führen und dementsprechend Abschätzungen für eine Gruppe von ähnlichen Paaren von Strukturen kollektiv für diese Gruppe durchgeführt werden können. Weiterhin kann in dem sogenannten Bound-Schritt des Branch and Bound-Verfahrens, in dem entschieden wird, ob eine Erweiterung der Teillösung zu einem besseren Ergebnis führt, häufig eine Gruppe von Paaren solcher ähnlicher Strukturen kollektiv ausgeschieden werden kann. Die Kompatibilitätsbeziehung läßt sich in geeigneter Weise definieren, beispielsweise daß die einander ähnlichen Strukturen Konformere sind, daß die relativen Abstände bestimmter Atome nur innerhalb einer bestimmten Bandbreite voneinander abweichen oder dergleichen. Bei dem erfindungsgemäßen Verfahren wird vorzugsweise nach einer geometrisch ähnlichen Struktur durch eine Überlagerung gesucht. Dabei versucht man, vereinfacht gesprochen, die jeweiligen Referenzstruktur und die jeweilige Vergleichsstruktur so gut wie möglich zur Deckung zu bringen. Bei einer Überlagerung wird zunächst eine Überlagerungszuordnung festgelegt, in der die Komponenten der einen Struktur den Komponenten der anderen Struktur zugeordnet werden. In einem zweiten Schritt müssen dann die ausgewählten und durch die Überlagerungszuordnung einander zugeordneten Paare von Komponenten bestmöglich zur Deckung gebracht werden, indem eine der Strukturen im dreidimensionalen Raum einer geeigneten Transformation (Translation und Rotation im dreidimensionalen Raum) unterzogen wird. Qualitätsmaß für diesen Schritt ist der mittlere quadratische Abstand der überlagerten Komponenten (rmsd).

**[0083]** Die Berechnung derjenigen Transformationen, die zu einer gegebenen Überlagerungszuordnung der Komponenten der einen Struktur zu den Komponenten der anderen Struktur den kleinstmöglichen mittleren quadratischen Abstand rmsd liefert, kann mit Verfahren nach dem Stand der Technik effizient gelöst werden, vgl. zum Beispiel S. Umeyama, Least-squares estimation of transformation parameters between two-point patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (1991), S. 376-380.

**[0084]** Gemäß einer bevorzugten Ausführungsform der Erfindung werden zwei Strukturen als einander ähnlich im Sinne des Vergleichs betrachtet, wenn zum einen die Überlagerung groß ist, d.h. wenn möglichst viele Komponenten der einen Struktur solchen der anderen zugeordnet werden können und andererseits diese Komponenten, wenn sie der vorangehend erwähnten Translation und Rotation unterzogen worden sind, um sie bestmöglich zur Deckung zu bringen, möglichst nahe beieinander liegen. Im allgemeinen sind diese Ziele nicht immer gleichzeitig zu erreichen, da rmsd tendenziell um so größer wird, je mehr Komponenten einander in einer Überlagerung zugeordnet werden.

**[0085]** Das vorangehend genannte Qualitätsmaß SC gemäß Formel (1) trägt dieser Gegenläufigkeit Rechnung. Es berücksichtigt einerseits, daß möglichst viele Komponenten überlagert werden sollen und andererseits die geometrischen Abweichungen zwischen einander entsprechenden Komponenten nicht allzu groß sein dürfen, wenn die beiden Strukturen (Referenzstruktur und Vergleichsstruktur) in optimaler Weise zur Deckung gebracht worden sind.

**[0086]** Erfindungsgemäß kann vorgesehen sein, daß außer der geometrischen Struktur in dem Qualitätsmaß und/oder bei der Kompatibilitätsbeziehung weitere Größen, beispielsweise chemische oder biochemische Größen, berücksichtigt werden, die es gestatten, chemisch nicht sinnvolle Lösungen von vornherein auszuschließen. Dies kann die Rechenzeit noch einmal erheblich verkürzen.

**[0087]** Gemäß einem vorteilhaften Aspekt der Erfindung werden nicht alle möglichen Zuordnungen von Komponenten der einen Struktur zu Komponenten der anderen Struktur zum Vergleich zugelassen. Gemäß einer bevorzugten Ausführungsform wird als Eingabe zu jedem Molekül auch dessen Bindungsstruktur gespeichert. Dies wird ausgenutzt, um die Menge aller potentiellen Überlagerungszuordnungen dahingehend einzuschränken, daß bei der Überlagerung zusammenhängende Teilstrukturen, d.h. Teilstrukturen, bei denen zwischen den Komponenten Bindungen derart bestehen, daß jeder Teil der Teilstruktur mit dem Rest der Teilstruktur über wenigstens eine Bindung verbunden ist, auf zusammenhängende Teilstrukturen der anderen Struktur abgebildet werden.

**[0088]** Mathematisch bedeutet dies eine gewisse Approximation. Aus chemischer oder biologischer Sicht sind jedoch in aller Regel die optimalen Paare von Strukturen solche, bei denen solche zusammenhängenden Teilstrukturen einander zugeordnet sind. Testrechnungen der Erfinder haben weiterhin gezeigt, daß selbst dann, wenn das Verfahren ohne die Einschränkung, daß nur Komponenten aus zusammenhängenden Teilstrukturen der jeweiligen Moleküle einander zugeordnet werden, nur in äußerst wenigen Fällen solche Überlagerungszuordnungen zu optimalen Lösungen führen, die diese Einschränkungen nicht erfüllen.

**[0089]** Gemäß einer bevorzugten Ausführungsform der Erfindung, bei der das Kompatibilitätskriterium bein-

haltet, daß die vergleichbare oder kompatible Struktur ein Konformer ist, kann der Suchraum wesentlich dadurch reduziert werden, daß die Überlagerungszuordnung der Atome nur anhand der Bindungsstruktur der Moleküle, simultan für alle Konformationspaare, betrachtet wird. Gewissermaßen werden hierbei verschiedene Enumerationsbäume, die jeweils einem Konformerpaar entsprechen, übereinandergelegt.

**[0090]** Vorteilhafterweise wird die Enumeration der Atomzuordnungen so angelegt, daß jede mögliche Zuordnung nur genau einmal erzeugt wird und kein aufwendiges Zwischenspeichern von Teilzuordnungen während der Enumeration notwendig ist.

**[0091]** Beispielsweise kann wie folgt vorgegangen werden. Die Menge der zusammenhängenden Teilstrukturen, die sich sowohl in einer Referenzstruktur als auch in einer Vergleichsstruktur wiederfinden und also einander zugeordnet werden können, und die gleichzeitig dadurch gekennzeichnet sind, daß sie je eine eindeutig spezifizierte Komponente der Referenzstruktur bzw. der Vergleichsstruktur enthalten, können in eindeutiger Weise identifiziert werden mit einer Kollektion von Mengen, die ihrerseits dadurch gekennzeichnet sind, daß sie jeweils aus den Komponenten der Referenzstruktur bzw. der Vergleichsstruktur bestehen, die zu der eindeutig spezifizierten Komponente einen eindeutigen Abstand gemäß der Bindungsstruktur haben.

**[0092]** Diese Kollektion von Mengen läßt sich nun durch eine einfache Rekursion eindeutig erzeugen, und somit auch die entsprechenden zusammenhängenden Teilstrukturen, die diese eindeutig spezifizierten Komponenten enthalten.

**[0093]** Ist zusätzlich eine strikte Ordnung auf allen Paaren von Komponenten, eine aus der Referenzstruktur, eine aus der Vergleichsstruktur, definiert, so können alle diese Paare in einer eindeutigen Reihenfolge als die oben angesprochenen eindeutig spezifizierten Komponenten der Referenzstruktur bzw. der Vergleichsstruktur benutzt werden. Auf diese Weise kann dann in der Enumeration vermieden werden, daß die gleichen zusammenhängenden Teilstrukturen mehrfach erzeugt werden.

**[0094]** Für das Ausscheiden von erweiterten Überlagerungszuordnungen können unter anderem die folgenden Abschätzungen verwendet werden:

1. Keine Lösung kann besser sein als der Beitrag des ersten Terms in (1), d.h. die Anzahl der zugeordneten Atome muß groß genug sein (unabhängig von den Konformeren).
2. Es wird angenommen, daß die noch nicht zugeordneten Atome alle verlustfrei, d.h. mit zusätzlichem  $rmsd = 0$ , zugeordnet werden können (unabhängig von den Konformeren).
3. Der zu erwartende  $rmsd$ -Wert von noch nicht zugeordneten Atomen kann durch Betrachtung der interatomaren Abstände von je zwei Atomen in jedem Molekül (und für jedes noch aktuelle Konformer) nach unten abgeschätzt werden, d.h. man kann auf diese Weise den mindestens auftretenden zusätzlichen  $rmsd$ -Wert bestimmen. Allgemein läßt sich sagen, daß hohe Unterschiede der interatomaren Abstände in den Molekülen zu hohen  $rmsd$ -Werten führen. Dies macht man sich bei der erfindungsgemäßen Methode zunutze.
4. Anstatt interatomare Abstände für alle Konformerpaare explizit zu betrachten, speichert man vor Beginn des Enumerationsprozesses für jedes Paar von Atomen in einem Molekül dasjenige Intervall, das alle auftretenden Abstände dieser beiden Moleküle über alle Konformere enthält (in dem später erläuterten Fallbeispiel (vgl. [Fig. 3](#)) sind zum Beispiel Atom 12 und Atom 4 immer mindestens 3,70 Å und höchstens 5,55 Å voneinander entfernt). Eine solche Information macht es möglich, ohne aufwendige Berechnungen zu entscheiden, ob eine Erweiterung der Zuordnung der Komponenten noch sinnvoll ist, so die betrachteten Intervalle in den beiden Molekülen z.B. disjunkt sind. In vielen Fällen kann der entstehende  $rmsd$ -Wert schon aufgrund einer solchen Abschätzung gut genug von unten abgeschätzt werden. Auch können die zugehörigen Distanzmatrizen im Laufe der Enumeration mit Informationen des vorherigen Schrittes (d.h. einzelne Konformerpaare sind als nicht mehr relevant ausgeschieden) aktualisiert werden, was zu präziseren Aussagen führt.
5. Die Lage der noch nicht überlagerten Komponenten der Moleküle relativ zum bereits überlagerten Anteil (sowohl topologisch als auch geometrisch betrachtet) liefert zusätzliche Information über die Größe einer potentiell noch gewinnbringenden Erweiterung.

**[0095]** Um effizient zu enumerieren und den Suchraum stark einzuschränken, wird vorzugsweise die Reihenfolge der durchsuchten Möglichkeiten stark an der Zusammenhangsstruktur der Moleküle orientiert.

**[0096]** Gelangt die Enumeration an Stellen, an denen alle Atome fest zugeordnet sind (entweder einem bestimmten Atom des anderen Moleküls oder eben keinem anderen Atom), kann eine Situation eintreten, in der für beide Moleküle noch mehrere Überlagerungszuordnungen, zum Beispiel mehrere Paare von Konformeren,

zulässig sind. Eine exakte (aber aufwendige) Berechnung des Qualitätsmaßes ist in diesem Fall nicht unbedingt erforderlich. Man kann sich in dieser Situation Techniken bedienen, mit denen das Qualitätsmaß abgeschätzt wird und die denen zu Punkt 3 und 4 (s. o.) ähneln können. Auf diese Weise können einerseits schnell die aussichtsreichen exakten Berechnungen von (1) zuerst durchgeführt werden, andererseits kann so der Wert der bisher besten bekannten Lösung früher erhöht werden, was wiederum die Strategien für den Bound-Schritt effizienter macht.

**[0097]** Gemäß einer weiteren Ausführungsform der Erfindung werden nur solche Überlagerungen gesucht, deren Qualitätsmaß besser als ein vorgegebener Wert des Qualitätsmaßes ist. Beispielsweise kann, wenn das Qualitätsmaß gemäß (1) verwendet wird, vorgegeben sein, daß die aufgefundene optimale Zuordnung einen Wert von SC aufweisen muß, der größer oder gleich als ein Minimalwert,  $SC_{\min}$  ist. Bei dieser Ausführungsform kann man bereits zu Beginn den Wert des Qualitätsmaßes auf den Wert für die gewünschte Mindestqualität setzen, also bei dem vorangehenden Beispiel auf den Wert  $SC_{\min}$ . Dies beschleunigt die Enumeration erheblich.

**[0098]** Das erfindungsgemäße Verfahren hat den Vorteil, daß auch sehr unterschiedlich große Moleküle miteinander verglichen werden können. Obwohl das benutzte Qualitätsmaß weder monoton noch additiv ist, was bedeutet, daß gute Teillösungen nicht notwendigerweise zu guten Lösungen erweiterbar sind oder umgekehrt und die Qualität einer Lösung sich nicht durch die Qualität der Teillösungen bestimmt, ist es möglich, in effizienter Weise global optimale Lösungen, die ein chemisch sinnvolles Ergebnis liefern, zu berechnen. Dadurch, daß simultan alle Paare von Konformeren zweier betrachteter Moleküle behandelt werden können, wird ein enormer Geschwindigkeitsgewinn bei der Verarbeitung erzielt. Beispielsweise kann bei etwa 100 Konformeren pro Molekül mit dem erfindungsgemäßen Verfahren die Rechenzeit um einen Faktor 10.000 gegenüber einem Verfahren verkürzt werden, bei dem die einzelnen Konformerpaare separat verglichen werden.

#### Ausführungsbeispiel

**[0099]** Die Erfindung wird nachfolgend anhand eines modellhaften Ausführungsbeispiels mit weiteren Einzelheiten und anhand der beigefügten Zeichnungen näher erläutert.

**[0100]** [Fig. 1](#) zeigt die als Eingabe bei dem Ausführungsbeispiel verwendeten Konformere für NCI ID 52050 (oben) und für NCI ID 95173 (unten).

**[0101]** [Fig. 2](#) zeigt die optimale Überlagerung der Substanzen für das erörterte Ausführungsbeispiel.

**[0102]** [Fig. 3](#) zeigt die Strukturgraphen der Eingabemoleküle,

**[0103]** [Fig. 4](#) zeigt die ersten fünf Rekursionsschritte, wobei Punkte (...) noch nicht wahrgenommene Branch-Alternativen darstellen,

**[0104]** [Fig. 5](#) zeigt den Abstieg auf dem ersten Ast bis zum Ende, wobei die gestrichelten Kästen Branch-Alternativen darstellen, die abgeschnitten werden können,

**[0105]** [Fig. 6](#) zeigt einen zweiten Unterbaum der Überlagerung in Schritt 5,

**[0106]** [Fig. 7](#) zeigt die verbleibenden Schritte für das Überlagerungspaar 13A,

**[0107]** [Fig. 8](#) zeigt eine Situation, in der eine Überlagerung nicht sinnvoll erweiterbar ist, ohne die bereits überlagerten Atome komplett aus der momentanen Raumlage zu drehen.

**[0108]** Nachfolgend wird anhand eines einfachen Beispiels ein typischer Programmablauf eines erfindungsgemäßen Verfahrens erläutert.

**[0109]** Als Eingabe dienen die Substanzen NCI ID 52050 und NCI ID 95173 (Quelle: National Cancer Institute), wobei aus Gründen der Übersichtlichkeit nur zwei Konformere pro Substanz ausgewählt wurden. Als Zielfunktionskonstante wurde die Eulersche Zahl ( $e = 2,71828$ ) gewählt. Beide Moleküle enthalten jeweils 14 Nichtwasserstoffatome, somit gibt es ca.  $1,5 \cdot 10^{12}$  mögliche Überlagerungen. Da insgesamt 4 Möglichkeiten bestehen, ein Paar von Konformeren auszuwählen, besitzt der Lösungsraum für diese Eingabe sogar die vierfache Größe.

**[0110]** Mit einer einfachen Implementierung des erfindungsgemäßen Verfahrens, welche auf aufwendige Programmierung verzichtet und lediglich erfindungsgemäße Branch-and-Bound-Kriterien und Heuristiken verwendet, wurde ein Ablaufprotokoll für die Eingabe erstellt. Die Implementation findet in lediglich 2753 Rekursionsschritten die optimale Lösung des Überlagerungsproblems mit einem Wert des Qualitätsmaßes von 0.4387. Mit Absicht wurde hierbei ein Beispiel gewählt, welches keine gute (und somit eventuell „offensichtliche“) Lösung besitzt. Die optimale Überlagerung der beiden Substanzen ist in [Fig. 2](#) dargestellt. Man erkennt, daß die räumliche Übereinstimmung nur mäßig gut ist.

**[0111]** Die ersten 28 Rekursionsschritte, die das Verfahren auf dem Branch and Bound-Baum entlangläuft, umfassen bereits einen abgeschlossenen Teilbaum, in welchem konkrete Beispiele zu den verschiedenen Klassen von beschriebenen Bound-Kriterien zur Anwendung kommen. Eine vollständige Darstellung des Ablaufes würde den Rahmen dieser Beschreibung sprengen. Daher beschränkt sich die folgende Darstellung auf diese ersten 28 Rekursionsschritte, um verschiedene Einzelheiten des erfindungsgemäßen Verfahrens zu illustrieren.

**[0112]** Für die Computerimplementierung wird die Molekularstruktur auf einen Graphen abgebildet. Jedes Atom der beiden Moleküle (Wasserstoffatome ausgenommen) wird mit einem Knoten des Graphen identifiziert, die Atombindungen stellen die Kanten dar. Die entsprechenden Graphen für NCI ID 52050 und NCI ID 95173 sind in [Abb. 3](#) dargestellt.

**[0113]** Des weiteren werden für die einzelnen Konformere die interatomaren Distanzen aller Atompaare (ohne Berücksichtigung von Wasserstoffatomen) berechnet. Diese werden zu einer Matrix mit Intervallen möglicher Abstandswerte zusammengefaßt. In der nachfolgenden Tabelle ist die Distanzmatrix der beiden Konformere der Substanz NCI ID 52050 gezeigt, die bei diesem Ausführungsbeispiel betrachtet werden. Ein Eintrag (i, j) enthält den jeweils kleinsten und größten Wert des euklidischen Abstandes (in Å) der Atome i und j über alle Konformere.

Tabelle 1

	0	1	2	3	4	5
0	-	[1:50; 1:50]	[2:46; 2:47]	[3:82; 3:83]	[4:26; 4:55]	:::
1	[1:50; 1:50]	-	[1:35; 1:36]	[2:44; 2:45]	[2:81; 3:21]	:::
2	[2:46; 2:47]	[1:35; 1:36]	-	[1:47; 1:48]	[2:44; 2:47]	:::
3	[3:82; 3:83]	[2:44; 2:45]	[1:47; 1:48]	-	[1:49; 1:50]	:::
4	[4:26; 4:55]	[2:81; 3:21]	[2:44; 2:47]	[1:49; 1:50]	-	:::
5	[5:10; 5:29]	[3:81; 3:95]	[3:00; 3:53]	[2:28; 2:36]	[1:20; 1:22]	:::
6	[1:77; 1:78]	[2:70; 2:71]	[3:26; 3:83]	[4:59; 5:01]	[4:89; 5:30]	:::
7	[1:78; 1:79]	[2:70; 2:70]	[3:06; 3:61]	[4:44; 4:84]	[5:33; 5:87]	:::
8	[2:36; 2:36]	[1:20; 1:21]	[2:21; 2:24]	[2:70; 2:71]	[2:50; 3:26]	:::
9	[4:58; 4:88]	[3:27; 3:68]	[2:47; 2:47]	[1:52; 1:53]	[2:51; 2:52]	:::
10	[4:41; 6:23]	[3:19; 4:90]	[3:01; 3:83]	[2:53; 2:53]	[3:03; 3:03]	:::
11	[6:16; 7:57]	[4:94; 6:39]	[4:78; 5:22]	[4:00; 4:15]	[4:20; 4:72]	:::
12	[6:62; 8:94]	[5:24; 7:63]	[5:17; 6:57]	[4:14; 5:28]	[3:70; 5:55]	:::
13	[3:94; 5:01]	[2:73; 3:65]	[2:70; 3:36]	[2:38; 2:47]	[1:32; 1:33]	:::

**[0114]** Bemerkenswert ist, daß viele der Intervalle trotz der Beweglichkeit des gesamten Moleküls klein sind, weil lokal betrachtet starre Teilstrukturen, wie Kohlenstoffringe, existieren.

**[0115]** Das Branching startet mit einer undefinierten Überlagerung und legt schrittweise Zuordnungen von

Atomen der einen Substanz zu Atomen der zweiten Substanz fest. Noch nicht überlagerte Atome werden als offene Fälle betrachtet, die gegebenenfalls noch überlagert werden können. Allerdings sind dabei nur Zuordnungen erlaubt, die keine bereits zuvor im Baum betrachtete Überlagerung ergeben können.

**[0116]** Diese Einschränkung wird explizit bei der Auswertung aller Bound-Kriterien und Abschätzungen berücksichtigt. Wird an einer Stelle der Ausführung festgestellt, daß die aktuelle Überlagerung nicht mehr hinreichend erweitert oder verbessert werden kann, erfolgt ein Backtracking zur letzten offenen Alternative im Rekursionsbaum.

**[0117]** Zu Beginn wird willkürlich eine Zuordnung eines Atoms festgelegt. Dann werden alle Erweiterungen dieser ersten partiellen Überlagerung gebildet, sofern sie nicht nach einem der erfindungsgemäß vorgesehenen Kriterien verworfen werden können. Dabei wird zuerst entlang gemeinsamer Kanten im Strukturgraphen erweitert (siehe [Fig. 4](#)). Als Qualitätsmaß (Score) wird die Größe SC, die durch (1) definiert ist, verwendet.

**[0118]** In diesem Beispiel wird mit der Auswahl des Paares 13A (d.h. Atom 13 von NCI ID 52050 wird Atom A von NCI ID 95173 zugeordnet) begonnen. Für einelementige Überlagerungen gilt trivialerweise  $rmsd = 0$ , somit ist der Score =  $1/14 = 0,0714$ . In den nächsten vier Schritten wird die Überlagerung zu 2D 3C 4B 13A erweitert. Der Wert des Qualitätsmaßes (Score-wert) der partiellen Überlagerungen verbessert sich kontinuierlich auf 0,2397.

**[0119]** Im fünften Schritt wird für das Konformerpaar (1,2) eine Differenz der Abstände der Atome 2 und 13 einerseits und D und A andererseits von 1,33409 Å festgestellt. Selbst bei optimaler Überlagerung des Paares 13A 2D ergibt sich ein Beitrag von  $1/2 (1,334)^2 = 0,8899$  zu den Abstandsquadraten. Ohne die Überlagerung dieses Paares ergibt sich ein rmsd-Wert vom Betrag 0,208. Man kann leicht überprüfen, daß für alle erreichbaren Überlagerungsgrößen  $n$  gilt, daß

$$(n + 2) \exp\left(\sqrt{\frac{1}{n + 2} (0,8899 + 3(0,208)^2)}\right) \leq n \exp\left(\sqrt{\frac{1}{n} 3(0,208)^2}\right)$$

**[0120]** Somit muß das Konformerpaar (1, 2) in Erweiterungen dieser Überlagerung nicht mehr betrachtet werden, da der Wert von rmsd eine durch die innermolekularen Atomabstände definierte untere Grenze aufweist, die eine Abschätzung des Qualitätsmaßes für alle zugehörigen erweiterten Überlagerungen gestattet.

**[0121]** In den folgenden Schritten ([Fig. 5](#)) wird die Überlagerung auf 0F 1E 2D 3C 4B 7G 13A erweitert. Man beachte dabei, daß sich der Score in Schritt 6 gegenüber Schritt 5 verschlechtert. Dies ist jedoch kein hinreichendes Bound-Kriterium, da sich der Score in größeren Überlagerungen wieder verbessern kann. Im Unterbaum von 6 ist nur noch das zweite Konformer der ersten Substanz relevant. Somit kann die Distanzmatrix verschärft werden, was im allgemeinen eine Verkleinerung der Intervalle bedeutet. Da hier nur noch ein weiteres Konformer dieser Substanz eingegeben wurde, werden in diesem Fall aus den Intervallen sogar Punkte.

**[0122]** Durch Betrachtung der neuen Intervalle können alternative Unterbäume von Schritt 6 und 7 aufgrund der paarweisen Unterschiede in den interatomaren Distanzen komplett abgeschnitten werden.

**[0123]** Bis Schritt 8 kann keine Verbesserung gegenüber dem Score aus Schritt 5 mehr erreicht werden. Deshalb erfolgt ein Backtracking zur letzten offenen Alternative dieses Schrittes ([Abb. 7](#)). In diesem Unterbaum kann der Score auf 0,4273 verbessert werden.

**[0124]** Das Branching selbst kann beschleunigt werden, indem Abschätzungen des scores einer partiellen Überlagerung nach oben gemacht werden (Schritte 16 und 19). Beispielsweise kann der Wert von rmsd durch Betrachtung der interatomaren Distanzen nach unten beschränkt werden. Ist diese Schranke zu hoch, um ein neues Optimum zu bilden, kann auf die exakte Berechnung des rmsd-Wertes verzichtet werden.

**[0125]** Der rekursive Abstieg auf einem Ast des Baums kann beendet werden, wenn selbst beim Erweitern der Überlagerung mit allen noch potentiell freien Atomtupeln bei einem angenommenen rmsd-Wert von 0 kein neues Optimum mehr zu erreichen ist. Selbst wenn also alle noch verbleibenden Atome jeweils optimal, d.h. ohne quadratische Abweichung, überlagert werden könnten, also den Beitrag 0 zu dem Wert von rmsd liefern, würde dies immer noch nicht ausreichen, um eine bessere Überlagerung zu erzeugen.

**[0126]** Als Beispiel betrachte man die Überlagerung 0I 1H 2D 3C 4B 7J 13A im Schritt 10, welche einen Score von 0,1865 ergibt ( $rmsd = 0,9862$ ). Durch Betrachtung der Distanzintervalle wird festgestellt, daß diese Über-

lagerung maximal auf neun Atome erweiterbar ist. Dadurch verbessert sich der Score im besten Fall auf 0,2694. Der beste bereits gefundene Score (0,304) liegt über diesem Wert, die Rekursion wird hier also abgebrochen.

**[0127]** Nach Schritt 19 erfolgt ein Backtracking zur nächsten Alternative von Schritt 4. Der Score ist durch die Anzahl der überlagerten Atome nach oben beschränkt. Konkret ist der Score  $SC$  gemäß Formel (1) nach oben durch das Verhältnis  $N_U/N_{Mmin}$  beschränkt. Soll das derzeitige Optimum von 0,4272 überschritten werden, müssen demnach 5 Atome überlagert werden. Deshalb erfolgt im Schritt 20 noch keine Berechnung eines Score-Wertes (vgl. [Fig. 8](#)). In den Schritten 21-22 muß das Konformerpaar (2,2) nicht zur Score-Berechnung hinzugenommen werden. Eine Abschätzung des erreichbaren Scores aufgrund der innermolekularen Atomabstände für die nächsten zwei Schritte ergibt, daß das Optimum nicht verbessert werden kann. Jedoch kann die Distanzmatrix nicht wie im Schritt 5 verschärft werden, da in einer hinreichend großen Überlagerung immer noch ein Optimum erreicht werden könnte.

**[0128]** Es verbleibt noch ein alternativer Unterbaum von Schritt 3, und zwar eine Erweiterung der Überlagerung 4B 5C 13A. Jedoch ist erkennbar, daß die nicht überlagerten Komponenten der Substanzen eine unterschiedliche Lage besitzen. Aufgrund der Topologie und der Geometrie der beiden Substanzen ist klar, daß eine weitergehende Überlagerung bereits aufgrund der Topologie zu einem derart geringen Wert des Qualitätsmaßes führen wird, daß die entsprechenden Lösungen ausscheiden. [Fig. 8](#) zeigt, daß die Überlagerung von 4B 5C 13A nicht sinnvoll erweiterbar ist, ohne die bereits überlagerten Atome komplett aus der momentanen Raumlage zu drehen.

**[0129]** Bei NCI ID 95173 bilden die verbleibenden Atome eine zusammenhängende Struktur, die mit Atom C verbunden ist, wohingegen die verbleibende Komponente von NCI ID 52050 mit Atom 4 verbunden ist. Bei guter Überlagerung von 4B 5C 13A stehen die beiden Komponenten in unterschiedliche Bereiche des Raums ab (vgl. [Fig. 8](#)).

**[0130]** Somit können die momentan überlagerten Atome in größeren Erweiterungen der betrachteten Teilüberlagerung nicht hinreichend gut aufeinander bewegt werden, ohne an anderer Stelle den rmsd-Wert zu verschlechtern. Da kleine unpassende Teile von Überlagerungen generell verworfen werden können, braucht dieser Teilbaum nicht mehr weiterverfolgt zu werden. Es müssen keine Erweiterungen zu dem Paar 13A mehr erzeugt werden; das Verfahren wird mit einem nächstem Paar fortgesetzt.

**[0131]** Bei Betrachtung tiefer Unterbäume ist zu erkennen, daß sich der Wert des Qualitätsmaßes in den „Blättern“ gegenüber vorher betrachteten Unterbäumen verbessert (vgl. Unterbaum Schritt 9) oder dem bisherigen Optimum sehr nahe kommt (vgl. Unterbaum Schritt 23). Dies zeigt, daß die aufgestellten Bound-Kriterien gut sind. Des weiteren wird in den ersten 18 Schritten bereits ein Score von 0,4273 erreicht, was die Effektivität der Suchheuristik bestätigt.

**[0132]** Nachdem alle möglichen Paare von Konformeren des Referenzmoleküls und des Vergleichsmoleküls abgearbeitet worden sind, wird der beste aufgefundene Wert des Qualitätsmaßes zusammen mit den zugehörigen Konformeren und der zugehörigen Überlagerungszuordnung ausgegeben, abgespeichert oder in anderer Weise festgehalten.

**[0133]** Danach wird, sofern die Liste der Vergleichsmoleküle noch nicht abgearbeitet ist, zu dem nächsten Vergleichsmolekül übergegangen. Für den Vergleich dieses Vergleichsmoleküls mit dem Referenzmolekül kann man, unabhängig von den vorangegangenen Vergleichen, in der gleiche Weise wie vorangehend beschrieben vorgehen, d.h. man beginnt mit der Zuordnung eines Atoms des Referenzmoleküls zu einem Atom des Vergleichsmoleküls, setzt den zugehörigen Wert des Qualitätsmaßes als den besten geltenden Wert und arbeitet dann die zugehörigen Erweiterungen dieser Zuordnung ab. Insbesondere wenn man schon eine größere Anzahl von Vergleichen zwischen demselben Referenzmolekül und verschiedenen Vergleichsmolekülen durchgeführt hat, bietet es sich an, die Bedingung zu stellen, daß der Wert des Qualitätsmaßes größer sein muß als der kleinste von  $M$  Werten ( $M \geq 1$ ), die bislang als jeweils bester Wert bei einem Vergleich des Referenzmoleküls mit einem Vergleichsmolekül ermittelt wurden. Der Rechenaufwand wird dadurch erheblich verkleinert, da eine größere Anzahl von Lösungen in dem Bound-Schritt des Branch-and Bound Verfahrens ausgeschieden werden kann. Gleichzeitig kann man solche Vergleichsmoleküle mit vergleichsweise geringem Aufwand ausscheiden, die gegenüber anderen Vergleichsmolekülen eine schlechtere Lösung darstellen.

**[0134]** Die in den Ansprüchen, der Beschreibung und den Zeichnungen offenbarten Merkmale der Erfindung können sowohl einzeln als auch in beliebiger Kombination für die Verwirklichung der Erfindung in ihren ver-

schiedenen Ausführungsformen wesentlich sein.

### Patentansprüche

1. Iteratives Verfahren zum computergestützten Auffinden eines oder mehrerer Moleküle mit Ähnlichkeiten zu einem oder mehreren Referenzmolekülen, die mehrere Komponenten aufweisen, auf der Grundlage von Informationen zu einer oder mehreren Referenzstrukturen, welche jeweils eine Struktur eines Referenzmoleküls darstellen, wobei die besagten Informationen die Lage von Komponenten in dem Referenzmolekül gemäß dieser Referenzstruktur angeben, und von Informationen zu mehreren Vergleichsstrukturen, welche jeweils eine Struktur eines Vergleichsmoleküls darstellen, wobei die besagten Informationen jeweils die Lage von Komponenten in dem Vergleichsmolekül in dieser Vergleichsstruktur angeben, bei dem in einer Iteration ein Teil eines Referenzmoleküls und ein Teil eines Vergleichsmoleküls verglichen werden, die jeweils einen Teil des Referenzmoleküls bzw. des Vergleichsmoleküls enthalten, die in einer früheren Iteration verglichen worden sind,

wobei mindestens eine Iteration des Verfahrens umfaßt:

Auswahl einer ersten Referenzmenge von Komponenten eines Referenzmoleküls, welche einen Teil der Komponenten des Referenzmoleküls enthält, wobei die Anzahl der Komponenten in der ersten Referenzmenge kleiner als die Gesamtzahl der zu vergleichenden Komponenten in dem Referenzmolekül ist,

Auswahl einer ersten Vergleichsmenge von Komponenten eines Vergleichsmoleküls, welche die gleiche Anzahl von Komponenten wie die erste Referenzmenge besitzt,

Bestimmen einer ersten Überlagerungszuordnung für eine erste Referenzstruktur, die dem besagten Referenzmolekül zugeordnet ist, und eine erste Vergleichsstruktur, welche dem besagten Vergleichsmolekül zugeordnet ist, wobei die erste Überlagerungszuordnung jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau eine Komponente der ersten Vergleichsmenge zuweist,

wobei für mindestens eine weitere, sekundäre Referenzstruktur, welche zumindest eine Teilstruktur aufweist, die zu der ersten Referenzstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Referenzmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht,

und/oder für mindestens eine weitere, sekundäre Vergleichsstruktur, welche eine Teilstruktur aufweist, die zu der ersten Vergleichsstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Vergleichsmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht, bestimmt wird:

eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die primäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmenge die entsprechende Komponente der sekundären Referenzstruktur gemäß der Ähnlichkeitsbeziehung für die sekundäre Referenzstruktur derjenigen Komponente der ersten Vergleichsstruktur zuweist, welches die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,

und/oder

eine Überlagerungszuordnung für die erste Referenzstruktur und die sekundäre Vergleichsstruktur, welche jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau die Komponente zuweist, welche aufgrund der Kompatibilitätsbeziehung der sekundären Vergleichsstruktur derjenigen Komponente der ersten Vergleichsmenge entspricht, welche die erste Überlagerungszuordnung der betreffenden Komponente der ersten Referenzmenge zuweist,

und/oder

eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die sekundäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmenge die entsprechende Komponente der sekundären Referenzstruktur derjenigen Komponente der sekundären Vergleichsstruktur zuweist, welche der Komponente der ersten Vergleichsstruktur entspricht, welche die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,

wobei für jede dieser Überlagerungszuordnungen für eine Referenzstruktur mit einer Vergleichsstruktur ein Qualitätsmaß oder eine Abschätzung eines Qualitätsmaßes für die Ähnlichkeit der beiden Strukturen bestimmt wird,

und für eines oder mehrere Referenzmoleküle der Wert des Qualitätsmaßes einer Überlagerungszuordnung einer Struktur dieses Referenzmoleküls mit einer Struktur eines Vergleichsmoleküls zusammen mit der zugehörigen Überlagerungszuordnung als der geltende optimale Wert des Qualitätsmaßes oder als einer der geltenden optimalen Werte des Qualitätsmaßes gespeichert wird, falls ein Qualitätskriterium erfüllt ist.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Qualitätskriterium eine oder mehrere der folgenden Bedingungen beinhaltet:

– der Wert des Qualitätsmaßes der besagten Überlagerungszuordnung ist besser als der geltende optimale

Wert des Qualitätsmaßes bzw. besser als der schlechteste der geltenden optimalen Werte des Qualitätsmaßes,

– der Wert des Qualitätsmaßes der besagten Überlagerungszuordnung ist besser als ein vorgegebener Schwellenwert des Qualitätsmaßes.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die Kompatibilitätsbeziehung zwischen der primären Referenzstruktur und der sekundären Referenzstruktur darin besteht, daß die besagte Teilstruktur der sekundären Referenzstruktur die gleiche Konstitution oder Konfiguration aufweist wie die durch die erste Referenzmenge definierte Teilstruktur der ersten Referenzstruktur und/oder

daß die Kompatibilitätsbeziehung zwischen der ersten Vergleichsstruktur und der sekundären Vergleichsstruktur darin besteht, daß die Teilstruktur der sekundären Vergleichsstruktur die gleiche Konstitution oder Konfiguration aufweist wie die durch die erste Vergleichsmenge definierte Teilstruktur der ersten Vergleichsstruktur.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß in der besagten Iteration für alle Überlagerungszuordnungen, die in der Iteration bestimmt werden, entschieden wird, ob zu der jeweiligen Überlagerungszuordnung für alle erweiterten Überlagerungszuordnungen der betreffenden Referenzstruktur mit der betreffenden Vergleichsstruktur, welche zusätzlich zu den Zuordnungen der besagten Überlagerungszuordnung hinaus weitere Komponenten der Referenzstruktur umkehrbar eindeutig jeweils einer Komponente der Vergleichsstruktur zuordnen, das Qualitätskriterium nicht erfüllt werden kann, und daß in nachfolgenden Iterationen nur solche erweiterten Überlagerungszuordnungen berücksichtigt werden, für welche dies nicht der Fall ist.

5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die sekundäre Referenzstruktur ein Konformer zu der ersten Referenzstruktur und die sekundäre Vergleichsstruktur ein Konformer zu der ersten Vergleichsstruktur ist.

6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß das Verfahren zu einem vorgegeben Referenzmolekül ähnliche Moleküle ermittelt, wobei dem Verfahren Informationen zur dreidimensionalen Struktur eines oder mehrere Konformere des Referenzmoleküls, welche die Lage von Atomen in dem Konformer angeben und Informationen zu der dreidimensionalen Struktur eines oder mehrere Konformere mindestens eines Vergleichsmoleküls, vorzugsweise mehrerer Vergleichsmoleküle, welche die Lage von Atomen in dem Konformer angeben, zugrunde liegen und mindestens eine Iteration des Verfahrens umfaßt:

Auswahl einer ersten Referenzmenge von Atomen des Referenzmoleküls, wobei die Anzahl der Atome in der ersten Referenzmenge kleiner als die Anzahl der Atome in dem Referenzmolekül ist,

Auswahl einer ersten Vergleichsmenge von Atomen eines Vergleichsmoleküls, welche die gleiche Anzahl von Atomen wie die erste Referenzmenge besitzt,

Bestimmung einer ersten Überlagerungszuordnung für ein erstes Konformer des Referenzmoleküls und ein erstes Konformer des Vergleichsmoleküls, welche jedem Atom der ersten Referenzmenge umkehrbar eindeutig genau ein Atom der ersten Vergleichsmenge zuweist,

Bestimmung einer Überlagerungszuordnung für ein zweites Konformer des Referenzmoleküls zu einem Konformer des Vergleichsmoleküls und/oder für ein Konformer des Referenzmoleküls zu einem zweiten Konformer des Vergleichsmoleküls, welche jedem Atom der ersten Referenzmenge umkehrbar eindeutig genau dasselbe Atom der ersten Vergleichsmenge wie die erste Überlagerungszuordnung zuweist.

7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß die Iteration folgendes umfaßt:

Bestimmen eines Qualitätsmaßes oder einer Abschätzung für das Qualitätsmaß für jede Überlagerungszuordnung,

Speichern des Werts des Qualitätsmaßes als geltender optimaler Wert zusammen mit der zugehörigen Überlagerungszuordnung, falls ein Qualitätskriterium für den geltenden optimalen Wert erfüllt ist.

8. Verfahren nach Anspruch 6 oder 7, dadurch gekennzeichnet, daß die Iteration, welche der besagten Iteration folgt, umfaßt:

Bestimmen einer erweiterten Überlagerungszuordnung für alle Paare von Konformeren des Referenzmoleküls und des Vergleichsmoleküls, für welche in einer früheren Iteration nicht festgestellt wurde, daß das Qualitätskriterium nicht erfüllt werden kann, wobei diese erweiterte Überlagerungszuordnung jeweils jedem Atom einer zweiten Referenzmenge von Atomen des Referenzmoleküls umkehrbar eindeutig jeweils genau ein Atom einer zweiten Vergleichsmenge von Atomen des Vergleichsmoleküls zuordnet, wobei die erste und die zweite Referenzmenge voneinander verschieden sind und alle Atome der ersten Referenzmenge in der zweiten Referenzmenge enthalten sind und die erste und die zweite Vergleichsmenge voneinander verschieden sind und alle

Atome der ersten Vergleichsmenge in der zweiten Vergleichsmenge gehalten sind und wobei gemäß der weiteren Überlagerungszuordnung jedes Atom aus der ersten Referenzmenge einem Atom aus der ersten Vergleichsmenge entsprechend der ersten Überlagerungszuordnung zugeordnet ist.

9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß in dem Referenzmolekül die Komponenten der ersten Referenzmenge zusammen mit den Bindungen zwischen ihnen einen zusammenhängenden Teilabschnitt des Referenzmoleküls bilden und die Komponenten der ersten Vergleichsmenge zusammen mit den Bindungen zwischen ihnen in dem Vergleichsmolekül einen zusammenhängenden Teilabschnitt des Vergleichsmoleküls bilden.

10. Verfahren nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß das Qualitätsmaß den Abstand einander durch die Überlagerungszuordnung zugeordneter Komponenten berücksichtigt, der sich ergibt, wenn die Referenzstruktur mit der Vergleichsstruktur optimal zur Deckung gebracht wird.

11. Verfahren nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß in das Qualitätsmaß die Anzahl der Komponenten der Referenzstruktur eingeht, welche durch die Überlagerungszuordnung umkehrbar eindeutig jeweils einer Komponente der Vergleichsstruktur zugeordnet werden.

12. Verfahren nach einem der Ansprüche 1 bis 11, dadurch gekennzeichnet, daß das Qualitätsmaß durch die folgende Formel gegeben ist:

$$SC = (N_U/N_{Mmin})C^{-rmsd}, \quad (1)$$

wobei

$N_U$  die Anzahl der Komponenten der Referenzstruktur ist, die durch die Überlagerungszuordnung jeweils einer Komponente der Vergleichsstruktur zugewiesen werden,

$N_{min}$  den kleineren Wert aus dem Paar angibt, welches durch die Anzahl der Komponenten in dem Referenzmolekül und die Anzahl der Komponenten in dem Vergleichsmolekül gebildet wird,

$rmsd$  der mittlere quadratische Abstand der einander durch die Überlagerungszuordnung zugeordneten Komponenten ist und

$C$  eine Konstante ist.

13. Verfahren nach einem der Ansprüche 1 bis 12, dadurch gekennzeichnet, daß das Verfahren einem Branch and Bound-Algorithmus folgt.

14. Verfahren nach Anspruch 13, dadurch gekennzeichnet, daß für eine Überlagerungszuordnung einer Referenzstruktur zu einer Vergleichsstruktur, bei welcher die Komponenten der Referenzstruktur, die Komponenten der Vergleichsstruktur zugeordnet werden, einen zusammenhängenden Teilabschnitt des Referenzmoleküls bilden, überprüft wird, welche erweiterten Überlagerungszuordnungen, bei denen die Zuordnung der Komponenten entsprechend der besagten Überlagerungszuordnung erhalten bleibt, existieren, bei denen die Komponenten eines zusammenhängenden Teils der Referenzstruktur einem zusammenhängenden Teil der Vergleichsstruktur zugeordnet werden, daß das Qualitätsmaß für diese möglichen erweiterten Überlagerungszuordnungen abgeschätzt wird und daß diese erweiterten Überlagerungszuordnungen in späteren Iterationen nicht berücksichtigt werden, wenn die Abschätzung ergibt, daß das Qualitätskriterium nicht erfüllt werden kann.

15. Verfahren nach einem der Ansprüche 5 bis 14, dadurch gekennzeichnet, daß für alle Konformere des Referenzmoleküls bzw. der Vergleichsmoleküle für die Abstände der Atome in dem jeweiligen Molekül Abstandsintervalle gespeichert sind, welche die Obergrenze und die Untergrenze dieses Abstandes für alle Konformere angibt, und daß bei der Abschätzung des Qualitätsmaßes für eine Überlagerungszuordnung eines Konformers eines Referenzmoleküls mit einem Konformer eines Vergleichsmoleküls eine Abschätzung des Qualitätsmaßes dadurch berechnet wird, daß auf der Grundlage der besagten Bereichsgrenzen das bestmögliche Qualitätsmaß berechnet wird, das sich für beliebige Werte innerhalb der Grenzen der Abstandsintervalle für die inneratomaren Abstände ergibt und überprüft wird, ob dieses bestmögliche Qualitätsmaß das Qualitätskriterium erfüllt.

16. Computersystem zum Durchführen eines iterativen Verfahrens zum Auffinden von Molekülen mit strukturellen Ähnlichkeiten zu einem oder mehreren Referenzmolekülen, die mehrere Komponenten aufweisen, welches umfaßt:

eine Einrichtung zum Speichern von Informationen zu einer oder mehreren Referenzstrukturen, welche jeweils eine Struktur eines Referenzmoleküls darstellen, wobei die besagten Informationen die Lage von Komponen-

ten in dem Referenzmolekül gemäß dieser Referenzstruktur angeben, und von Informationen zu mehreren Vergleichsstrukturen, welche jeweils eine Struktur eines Vergleichsmoleküls darstellen, wobei die zugehörigen Informationen die Lage von Komponenten in dem Vergleichsmolekül in dieser Vergleichsstruktur angeben, mit einer Einrichtung zum Vergleichen eines Teils eines Referenzmoleküls und eines Teils eines Vergleichsmoleküls in einer Iteration des Verfahrens, die jeweils einen Teil des Referenzmoleküls bzw. des Vergleichsmoleküls enthalten, die in einer früheren Iteration verglichen worden sind, wobei das Computersystem eine Einrichtung zum Ausführen der folgenden Schritte in mindestens einer Iteration umfaßt:

Auswahl einer ersten Referenzmenge von Komponenten eines Referenzmoleküls, welche einen Teil der Komponenten des Referenzmoleküls enthält, wobei die Anzahl der Komponenten in der ersten Referenzmenge kleiner als die Gesamtzahl der zu vergleichenden Komponenten in dem Referenzmolekül ist,

Auswahl einer ersten Vergleichsmenge von Komponenten eines Vergleichsmoleküls, welche die gleiche Anzahl von Komponenten wie die erste Referenzmenge besitzt,

Bestimmen einer ersten Überlagerungszuordnung für eine erste Referenzstruktur, die dem besagten Referenzmolekül zugeordnet ist, und eine erste Vergleichsstruktur, welche dem besagten Vergleichsmolekül zugeordnet ist, wobei die erste Überlagerungszuordnung jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau eine Komponente der ersten Vergleichsmenge zuweist,

wobei für mindestens eine weitere, sekundäre Referenzstruktur, welche zumindest eine Teilstruktur aufweist, die zu der primären ersten Referenzstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Referenzmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht,

und/oder für mindestens eine weitere, sekundäre Vergleichsstruktur, welche eine Teilstruktur aufweist, die zu der ausgewählten primären Vergleichsstruktur in einer vorgegebenen Kompatibilitätsbeziehung steht, in welcher jede Komponente der ersten Vergleichsmenge umkehrbar eindeutig genau einer Komponente der besagten Teilstruktur entspricht, bestimmt wird:

eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die primäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmenge die entsprechende Komponente der sekundären Referenzstruktur derjenigen Komponente der ersten Vergleichsstruktur zuweist, welches die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,

und/oder

eine Überlagerungszuordnung für die erste Referenzstruktur und die sekundäre Vergleichsstruktur, welche jeder Komponente der ersten Referenzmenge umkehrbar eindeutig genau die Komponente zuweist, welche aufgrund der Kompatibilitätsbeziehung der sekundären Vergleichsstruktur derjenigen Komponente der ersten Vergleichsmenge entspricht, welche die erste Überlagerungszuordnung der betreffenden Komponente der ersten Referenzmenge zuweist,

und/oder

eine Überlagerungszuordnung für die sekundäre Referenzstruktur und die sekundäre Vergleichsstruktur, welche für jede Komponente der ersten Referenzmengen die entsprechende Komponente der sekundären Referenzstruktur derjenigen Komponente der sekundären Vergleichsstruktur zuweist, welche der Komponente der ersten Vergleichsstruktur entspricht, welche die erste Überlagerungszuordnung der besagten Komponente der ersten Referenzmenge zuweist,

und das Computersystem weiterhin eine Einrichtung zum Bestimmen oder Abschätzen eines Qualitätsmaßes aufweist, wobei das Computersystem weiterhin dafür eingerichtet ist, in dieser Iteration für eine oder mehrere Referenzmoleküle den Wert des Qualitätsmaßes einer Überlagerungszuordnung einer Struktur dieses Referenzmoleküls mit einer Struktur eines Vergleichsmoleküls zusammen mit der zugehörigen Überlagerungszuordnung als geltenden optimalen Wert des Qualitätsmaßes oder als einen der geltenden optimalen Werte des Qualitätsmaßes zu speichern, falls ein Qualitätskriterium erfüllt ist.

17. Computersystem nach Anspruch 16, gekennzeichnet durch eine oder mehrere Einrichtungen zum Durchführen eines Verfahrens nach einem der Ansprüche 1 bis 15.

18. Computerprogramm, welches Anweisungen für ein Computersystem enthält, die, wenn sie auf einem Computersystem ausgeführt werden, das Computersystem veranlassen, ein Verfahren nach einem der Ansprüche 1 bis 15 auszuführen.

19. Computerlesbares Speichermedium, bei dem, insbesondere in maschinenlesbarer Form, ein Programm gemäß Anspruch 18 gespeichert ist.

Es folgen 8 Blatt Zeichnungen

Anhängende Zeichnungen

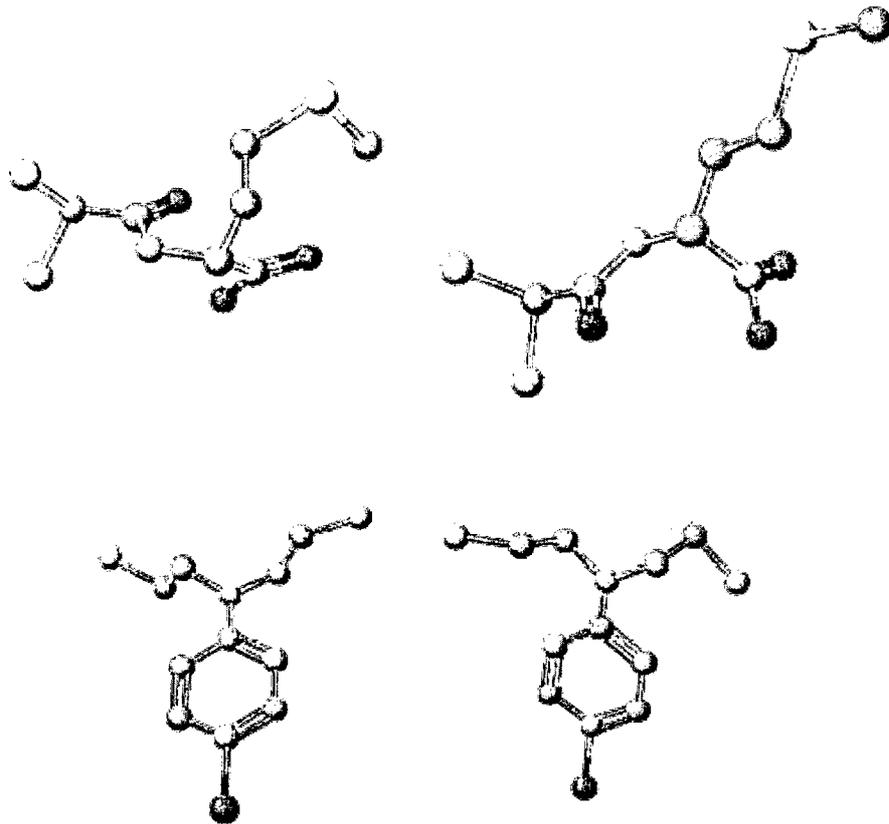


Fig. 1

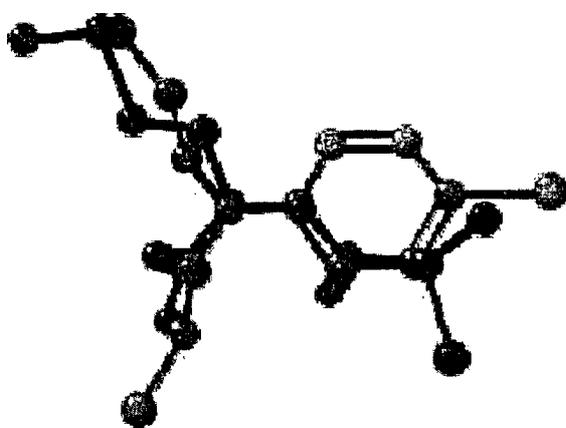


Fig. 2

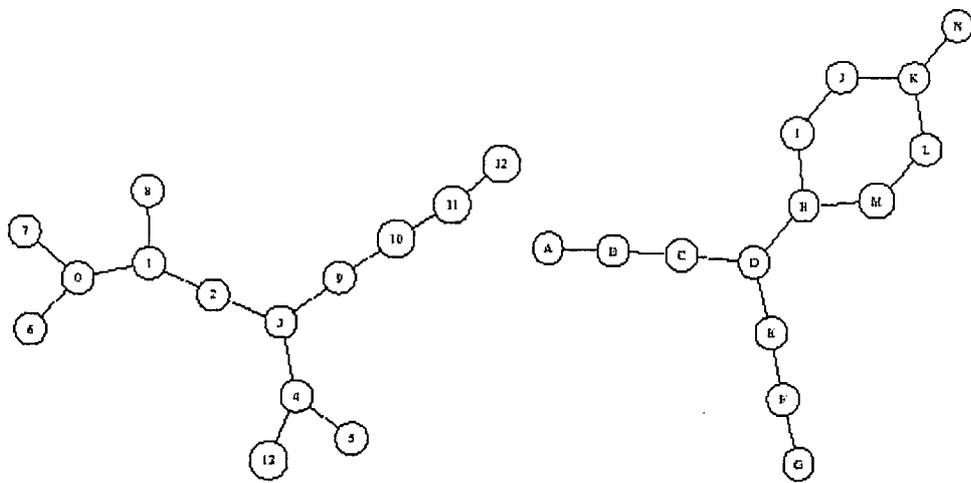


Fig. 3

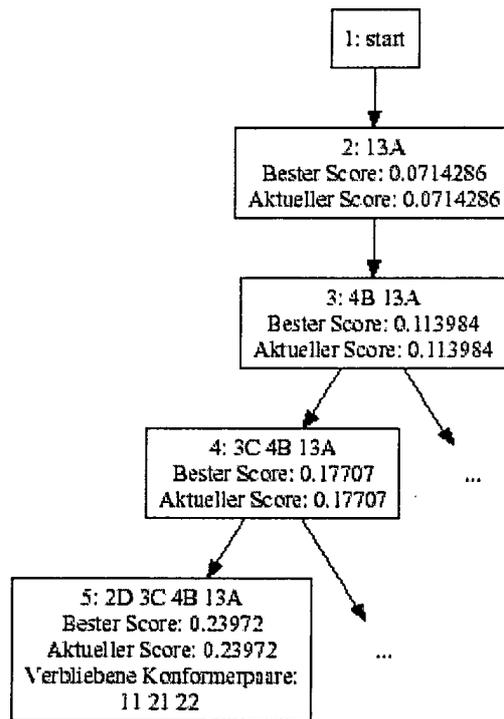


Fig. 4

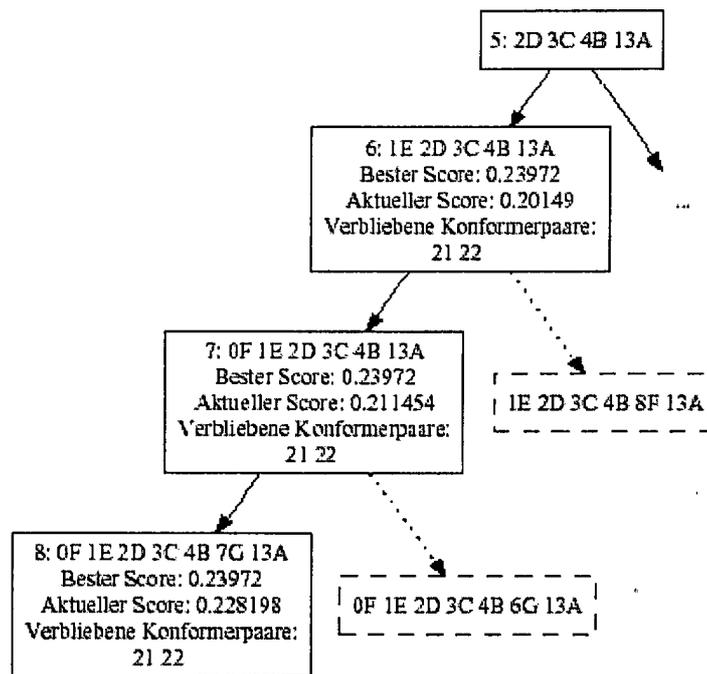


Fig. 5

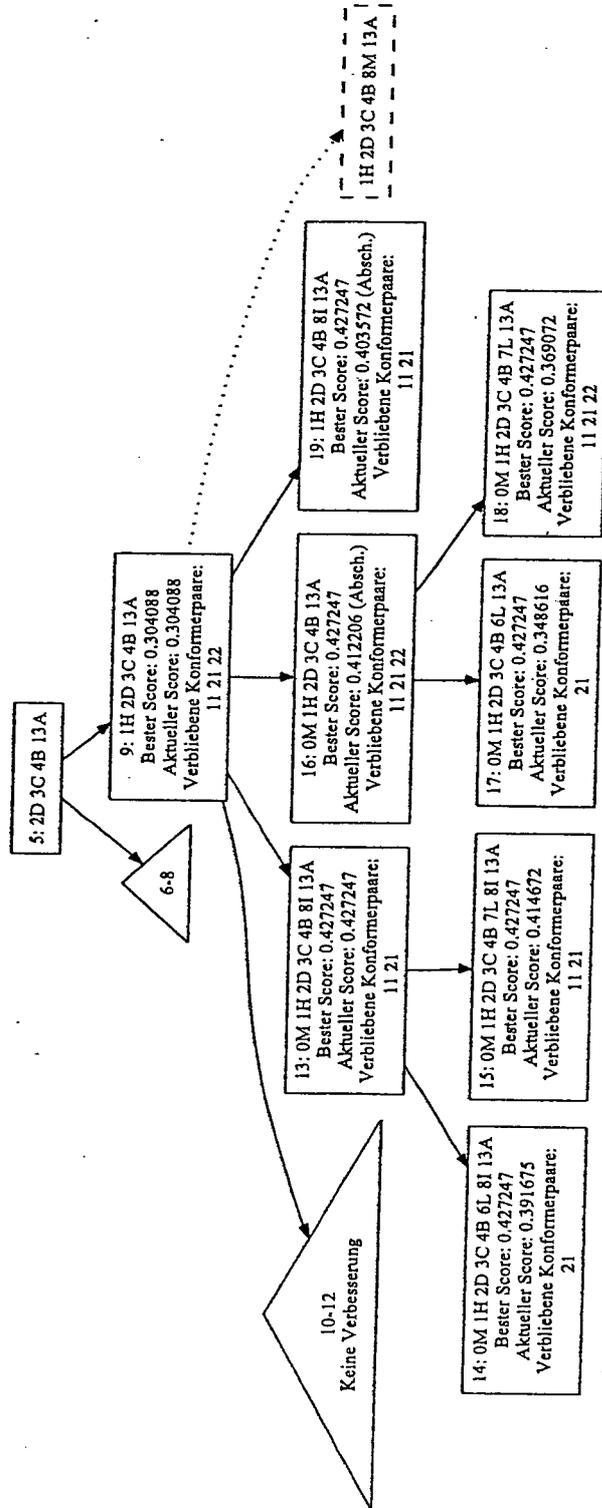


Fig. 6

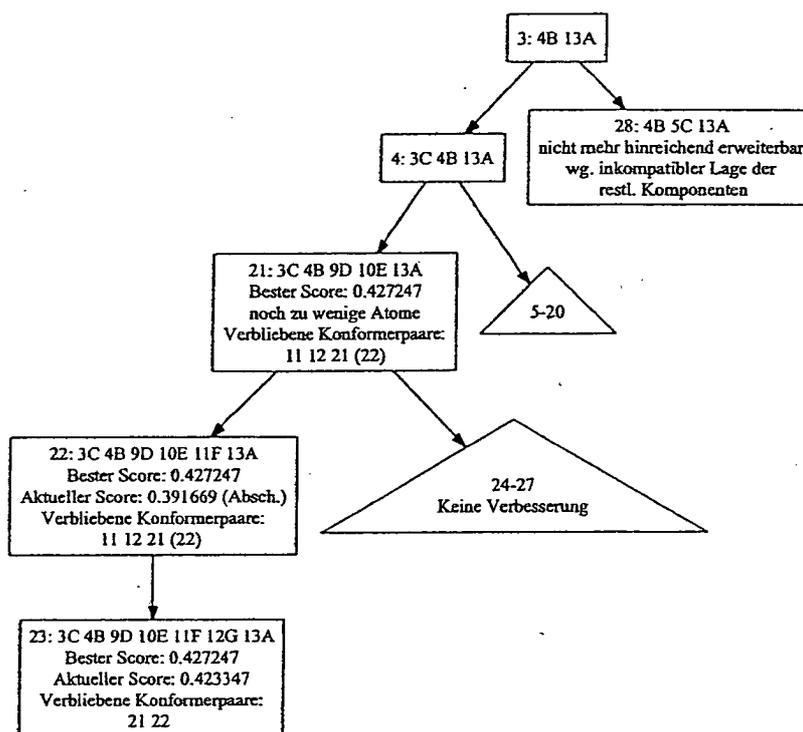


Fig. 7

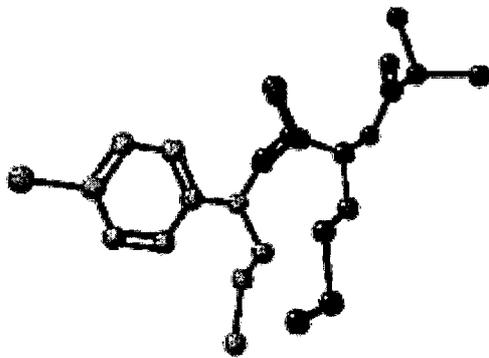


Fig. 8