

Mathematische Modelle für die kombinatorische Chemie und die molekulare Strukturaufklärung

Kurzfassung

Mathematische Modelle bilden eine unentbehrliche Grundlage in nahezu allen Bereichen von Wissenschaft und Technik. Immer häufiger werden auch Problemstellungen der organischen Chemie durch mathematische Modellierung simuliert und gelöst.

Motivation ist dabei oft die Suche nach neuen Wirkstoffen und Materialien mit angestrebten biologisch-pharmazeutischen oder physiko-chemischen Eigenschaften. Wurde eine entsprechende Verbindung synthetisiert und gefunden, besteht eine weitere wichtige Aufgabe darin, die *molekulare Struktur* der oft noch unbekannt Substanz zu bestimmen. Zu diesen Zweck verwendet man hauptsächlich spezielle physiko-chemische Eigenschaften, die aus *spektroskopischen* Methoden gewonnen werden.

Bei der Suche nach neuen Wirkstoffen und Materialien finden immer häufiger Techniken der *kombinatorischen Chemie* Verwendung. Dabei werden aus mehreren Sätzen chemischer Bausteine sämtliche Kombinationen synthetisiert und anschließend auf ihre biologisch-pharmazeutische Wirksamkeit getestet oder hinsichtlich angestrebter physiko-chemischer Eigenschaften untersucht. Der enorme Vorteil dieser Methode besteht darin, dass sowohl die Synthese als auch das Screening in hohem Maße automatisiert und parallelisiert werden kann. Obwohl somit erstaunliche Durchsatzraten erzielt werden, mahnen Kosten- und Zeitgründe zur gewissenhaften Planung und zur automatischen Auswertung derartiger Experimente.

Die Optimierung kombinatorisch-chemischer Experimente und die automatisierte molekulare Strukturbestimmung werfen vielfältige Probleme auf, die nach mathematischer Modellierung mit Hilfe von algebraisch-kombinatorischen Konstruktionsalgorithmen, graphentheoretischen Invarianten und statistischen Lernverfahren gelöst werden können. Die entscheidenden Problemstellungen betreffen die

- **Strukturgenerierung:**

In der kombinatorischen Chemie benötigt man Methoden, um virtuelle kombinatorische Bibliotheken zu konstruieren. Meist werden solche Strukturräume durch *Reaktanden* und *Reaktionen* definiert. In dieser Arbeit werden Algorithmen zur *reaktionsbasierten* Strukturgenerierung beschrieben.

Für die molekulare Strukturaufklärung werden Algorithmen verwendet, die ausgehend von der *Bruttoformel* eines Analyten unter Berücksichti-

gung struktureller *Restriktionen* mögliche Strukturformeln generieren können. Strategien zur *bruttoformelbasierten* Strukturgenerierung werden vorgestellt.

Wichtige Methoden für beide Problemkreise bilden *kanonische Nummerierung* und *ordnungstreue Erzeugung*.

- Suche nach Beziehungen zwischen Struktur und Eigenschaft:
Um Eigenschaften für die Strukturen virtueller kombinatorischer Bibliotheken vorhersagen zu können, verwendet man *quantitative Struktur-Eigenschafts-Beziehungen (QSPR)*. Diese werden zuvor anhand einer tatsächlich synthetisierten und gescreenten kleineren *realen* Bibliothek ermittelt.
Die *computer-unterstützte molekulare Strukturaufklärung (CASE)* verfolgt das Ziel, ausgehend von spektroskopisch gemessenen Eigenschaften einer unbekanntem chemischen Verbindung ihre molekulare Struktur zu bestimmen.
Die mathematischen Werkzeuge für diese Aufgaben sind *molekulare* bzw. *spektrale Deskriptoren* und *statistische Lernverfahren*.

Im ersten Teil der Arbeit werden die benötigten Modelle und Methoden beschrieben: Die Darstellung von chemischen Verbindungen, molekularen *Substrukturen* und chemischen *Reaktionen*, *ordnungstreue* und *zielgerichtete Erzeugung* zur bruttoformelbasierten Strukturgenerierung, *Kernstruktur-Ligand-Anlagerungen* und Konstruktion nach dem *Netzwerkprinzip* für die reaktionsbasierte Strukturgenerierung. Auf Seiten des *überwachten statistischen Lernens* werden *Klassifikation* und *Regression* durch *lineare Modelle*, *künstliche neuronale Netze*, *Support-Vektor-Maschinen*, *Entscheidungsbäume* und *k-nächste Nachbarn* vorgestellt.

Der zweite Teil enthält konkrete Anwendungen zu Problemstellungen aus der Chemie: Kombinatorische Bibliotheken werden generiert, z.B. nach *Ugis* Siebenkomponentenreaktion. Mit Hilfe von verschiedenen molekularen Deskriptoren (*arithmetische*, *topologische* und *geometrische Indizes*, *Substruktur-Vielfachheiten*) und statistischen Lernverfahren werden QSPR bestimmt, verglichen und zur Vorhersage herangezogen. Problemstellungen sind dabei *Siedepunkte* von *Decanen*, *physikalische Dichten* von *Propylacrylaten* und die *antibakterielle Aktivität* von *Quinolonen*. Die Untersuchungen zur CASE zielen auf die *Interpretation* und *Verifikation* von Massenspektren. Rankingfunktionen für Brutto- und Strukturformeln werden definiert und getestet. Verschiedene Klassifikationsverfahren zur Bestimmung von MS-Klassifikatoren werden verglichen. Weiterführende Studien untersuchen u.a. die Möglichkeiten *hochauflösender MS*.

Schlüsselworte: Strukturgenerierung, kombinatorische Chemie, molekulare Deskriptoren, quantitative Struktur-Eigenschafts-Beziehungen, molekulare Strukturaufklärung, Massenspektren-Klassifikation.