

Praxis der Naturwissenschaften
Chemie in der Schule

2/62

Heft Nr. 2/62. Jahrgang
März 2013

**Pillen und
Pflverchen –
Chemie und
Pharmazie**

Jenseits von Google

Chemie im Computer- und Informationszeitalter

M. Rarey und T. Schwabe

Der Computer, ein wichtiges Werkzeug für die Chemie: Mit dem Internet als Datenquelle stehen dem Chemiker heute vielfältige Informationen zur Verfügung. Computermodelle und -experimente komplementieren die Laborforschung und ermöglichen dabei ganz neue Wege.

Stichwörter: Chemieinformatik, Computerchemie, Computergestütztes Molekulares Design (CAMD)

1 Einleitung

Der Start des World Wide Webs vor ca. 20 Jahren markiert den Eintritt ins Informationszeitalter, welcher auch einen grundsätzlichen Wandel in den Naturwissenschaften eingeleitet hat. Gemeinsam mit Automatisierungstechnik im Laborumfeld ist der Systemwechsel hin zur datenintensiven Forschung in allen Bereichen der Lebenswissenschaften spürbar. Am prominentesten sind hier sicherlich die Genomprojekte, allen voran das Human-Genome-Project, welches 2001 erstmals die vollständige Sequenz des menschlichen Genoms veröffentlichen konnte [1]. Auch im Umfeld der Chemie werden durch Automatisierung und Systematisierung von Experimenten große Datenmengen erzeugt. Das Internet bietet die Plattform, um diese jedem Wissenschaftler auf der Welt verfügbar zu machen. Dabei ist für Chemiker der Umgang mit großen Informationsmengen schon immer ein wichtiger Bestandteil der Forschung gewesen. Chemische Reaktionen und Synthesen, physikalische Eigenschaften und die Verknüpfung zu experimentellen Daten wie Spektren und Strukturen für Millionen von Verbindungen haben schon frühzeitig zur computergestützten Datenverarbeitung geführt. Mit dem Chemical Abstract Service, der im Jahre 1966 erste chemische Datenbanken etablierte, kann die Chemie als Pionier des wissenschaftlichen Datenmanagements betrachtet werden (siehe www.cas.org). Eine historisch lange und starke Verknüpfung der Chemie mit der Informatik ist daher auch nicht verwunderlich. Jedoch stellen chemische Anwendungen für Informatiker bis heute eine große Herausforderung dar. Schon die Repräsentation chemischer Verbindungen mit all ihren isomeren Formen im Kontext von Datenbanken mit

entsprechender Suchfunktionalität ist heute noch nicht zufrieden stellend gelöst. Öffentliche Datenbanken, die Informationen über mehrere hundert Millionen Moleküle liefern, einfach zu nutzende Anfragesysteme, die ohne Verwendung systematischer Namen auskommen und auch Teilstruktur- und Musteranfragen erlauben, sind dabei heute von großer Bedeutung (siehe auch Tab. 1). Natürlich ist dies nur eine Facette der Chemieinformatik, die heute vielfältige Applikationen vom Datenmanagement bis zur Vorhersage physikochemischer und biochemischer Eigenschaften von Verbindungen verfügbar macht [2, 3]. Die moderne lebenswissenschaftliche Forschung ist heute ohne Einsatz chemieinformatischer Techniken undenkbar.

Neben der Chemieinformatik bietet der Einsatz von Computern aber auch Möglichkeiten, grundlegende chemische Fragestellungen anzugehen. Im Zentrum der Überlegungen steht dabei die computergestützte Modellierung und Simulation von chemischen Systemen. Dies geschieht aus unterschiedlichen Gründen, z. B., weil mit einem solchen Modell Fragestellungen untersucht werden können, die experimentell nur schwer oder gar nicht zugänglich sind. An Hand dieser Modelle lassen sich aber auch physikalische Eigenschaften von neuen

Molekülen vorhersagen, sodass aufwendige Experimente im Vorfeld vermieden werden können. Das Entwickeln, Verbsen und Anwenden solcher Modelle ist das Gebiet der Computerchemie. Die Verfahren lassen sich durch den Einsatz von Computer und Software automatisieren und erlauben so, dass in kürzerer Zeit wesentlich mehr Moleküle zu geringeren Kosten untersucht werden können. Durch die Erfolge der letzten Jahrzehnte und den stetigen Anstieg der Rechenleistung von Computern wird die Computerchemie daher auch immer interessanter für die chemische und pharmazeutische Industrie.

Der vorliegende Artikel soll einen Einblick in die Welt der Chemieinformatik und der Computerchemie gewähren. Der Fokus liegt vor allem auf der Tatsache, dass der Computer bereits an vielen Stellen Einzug in die Chemie gefunden hat. Dabei ist die Chemieinformatik und Computerchemie nicht nur als interessantes wissenschaftliches Feld zu betrachten. Zukünftige Chemiker werden vielmehr den Computer als tägliches Arbeitswerkzeug in den Laboren erleben.

2 Chemieinformatik

2.1 Chemie trifft Graphentheorie

Am Anfang nahezu jeder chemieinformatischen Entwicklung steht die Frage der

URL	Inhalt	# Moleküle
www.cas.org	Chemical Abstract Service, größte und umfangreichste chemische Datensammlung, nicht frei verfügbar	67 Mio.
www.chemspider.com	Suchmaschine zu Informationen über Moleküle im Netz	26 Mio.
pubchem.ncbi.nih.gov	Datenbank mit Molekülen und assoziierten biologischen Essays	
www.ebi.ac.uk/chembl	Datenbank mit bioaktiven Molekülen	1,4 Mio.
zinc.docking.org	Datenbank mit kommerziell erhältlichen, wirkstoffähnlichen Molekülen	21 Mio.
www.ccdc.cam.ac.uk/csd	Kristallstrukturen niedermolekularer Verbindungen, nicht frei verfügbar	0,6 Mio.
www.rcsb.org	Protein-Datenbank PDB; Kristallstrukturen von Biomolekülen wie Proteine, RNA und DNA	83 000

Tab. 1: Wichtige Internet-Ressourcen für Chemiker

Chemische Fragestellung	Graphalgorithmus
Ist eine Bindung Teil eines Rings?	Berechnung des Bizusammenhangs
Zerlegung von Ringsystemen in elementare Ringe	Berechnung von Zyklenbasen
Sind zwei Moleküle identisch?	Bestimmung der Graph-Isomorphie
Ist ein Molekül als Teilstruktur in einem anderen enthalten und wenn ja, wo?	Bestimmung der Subgraph-Isomorphie
Was ist der größte gemeinsame Teil zweier Moleküle?	Berechnung des maximalen gemeinsamen Teilgraphen (MCS)

Tab. 2: Zusammenhang zwischen einfachen Fragen auf Molekülstrukturen und Graphalgorithmen

Die SMILES-Notation

Die SMILES-Notation ist eine textuelle Beschreibung von Molekülen, die sich direkt aus der Molekülstruktur ableitet:

Atome werden durch Zeichenketten in eckigen Klammern zusammengefasst. Atome der Elemente C, O, N, S, P, Cl, Br, I können auch ohne Klammern geschrieben werden, vorausgesetzt es liegen keine weiteren Beschreibungsattribute vor. Wasserstoffe werden implizit angenommen bis die Standardzahl der Valenzen in der neutralen Form erreicht ist. C, O, N, S in aromatischen Systemen werden durch Kleinbuchstaben repräsentiert. Weitere Attribute erlauben die Charakterisierung von Masse, Ladung, Anzahl gebundener Wasserstoffe und Chiralität.

Kovalente Bindungen ergeben sich zunächst durch die Atommachbarschaften in der Zeichenkette. Als Standardfall wird die Einfachbindung angenommen, die durch kein Symbol repräsentiert wird, die Doppelbindung wird durch ein =-Symbol, die Dreifachbindung durch ein #-Symbol, die aromatische Bindung durch ein :-Symbol dargestellt.

Verzweigungen werden durch Klammerpaare repräsentiert, das der schließenden Klammer folgende Atom ist kovalent mit dem Atom vor der öffnenden Klammer verbunden. Zusätzliche kovalente Bindungen für Ringschlüsse werden durch einander zugeordnete Zahlenpaare dargestellt.

Eine vollständige Beschreibung der SMARTS-Notation findet man unter www.daylight.com/dayhtml/doc/theory/theory.smiles.html.

Verbindung	SMILES
Wasser	O
Hydroxidion	[OH-]
Ethan	CC
Ethanol	CCO
Formaldehyd	C=O
Demethylamin	CNC
Essigsäure	CC(=O)O
L-Alanin	C[C@H](N)C(=O)O
D-Alanin	C[C@@H](N)C(=O)O
Cyclohexan	C1CCCCC1
Benzol	c1ccccc1
Purin	c1ncnc(c1)nc[nH]2
Adamantan	C12CC3CC(C1)CC(C2)C3

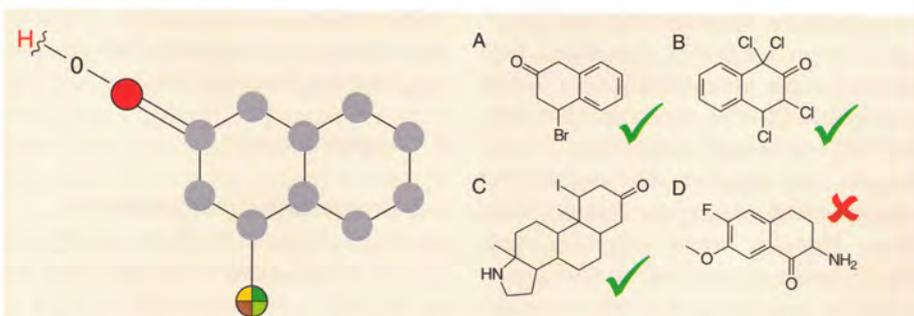
Molekülrepräsentation im Computer. Textuelle Beschreibungen eignen sich nur in den seltensten Fällen, nahezu alle Berechnungen erfolgen auf der Molekülstruktur. Aufgebaut aus Atomen und kovalenten Bindungen entspricht diese dem mathematischen Konstrukt eines sogenannten Graphen bestehend aus Knoten (den Atomen) und Kanten (den Bindungen), die wiederum mit Markierungen (dem Element, der Ladung, dem Bindungstyp, etc.) ausgestattet sind. Da Graphen in der Mathematik und Informatik intensiv bearbeitet wurden, steht eine Vielzahl von Verfahren zur Berechnung von Grapheneigenschaften auch für chemische Fragestellungen zur Verfügung. Wichtige Beispiele hierzu sind in Tabelle 2 zusammengefasst.

Wie so häufig steckt bei der Umsetzung der bekannten Graphalgorithmen für chemische Fragen der Teufel im Detail. Als ein Beispiel seien hier Tautomere genannt. Wenn ein Molekül in zwei unterschiedlichen tautomeren Formen vorliegt, unterscheiden sich die Molekülstrukturen, die Identität wird somit nicht mehr ohne weiteres erkannt. Natürlich gibt es nicht für alle chemischen Fragestellungen Algorithmen von der Stange, häufig werden diese von Chemieinformatikern maßgeschneidert entwickelt [4].

2.2 Repräsentation chemischer Strukturen in Ein- und Ausgaben

Aus Sicht des Benutzers chemieinformatischer Software stellt sich natürlich die Frage, wie Moleküle in Computersysteme ein- und wieder ausgegeben werden. Bezüglich der Eingabe scheint die Antwort zunächst offensichtlich, chemische Verbindungen haben Namen, die einer systematischen Nomenklatur folgen. Die Überführung eines systematischen Namens in eine Molekülstruktur ist jedoch eine sowohl für den Benutzer als auch für den Computer recht komplexe Aufgabe, insbesondere wenn Trivialnamen und sprachspezifische Aspekte berücksichtigt werden sollen. Da das Zeichnen von Strukturen sehr zeitintensiv ist, hat man schon früh in der Chemieinformatik nach alternativen Textdarstellungen gesucht. In den vergangenen Jahren hat sich schließlich weitgehend die SMILES-Notation als alternative Beschreibungsform durchgesetzt (siehe Abb. 1) [5]. Sie kann zwar nicht phonetisch umgesetzt werden, lässt sich jedoch sehr einfach und präzise in Molekülstrukturen überführen. Nahezu jede Chemie-Software und Internet-Ressource kann Eingaben in SMILES-Notation verarbeiten, sodass heute jeder Che-

Abb. 1: Beispiel eines SMARTS-Musters und seiner grafischen Darstellung



SMARTS-Sprache zur Beschreibung chemischer Muster

Die Abbildung links zeigt ein chemisches Muster bestehend aus zwei anellierten, sechselementigen Kohlenstoffringen. Die Atome sind spezifiziert als aromatisch oder aliphatisch, die Bindungen als einfach oder aromatisch. Das Ringgerüst trägt zwei Substituenten, einmal ein beliebiges Halogen (Fluor, Brom, Iod oder Chlor) und in *meta*-Stellung zu diesem ein Sauerstoffatom, welches über eine Einfach- oder Doppelbindung an das Ringsystem gebunden ist und keine weitere Bindung zu einem Schweratom hat. Mit diesem Muster werden beispielsweise die Moleküle A, B und C getroffen. Molekül D wird nicht getroffen, da hier das Sauerstoffatom eine weitere Bindung an ein Schweratom hat und an der spezifizierten Substitutionsposition kein Halogen ist.

SMARTS: [C,c]1[C,c]2[C,c]([C,c][C,c][C,c]1)[C,c][C,c](-,[OD1])[C,c][C,c]([Cl,Br,I,F])2

Abb. 2: SMARTS-Sprache zur Beschreibung chemischer Muster

miker damit vertraut sein sollte. Für viele Datenbank-Anwendungen in der Chemie ist die Beschreibung von Molekülstrukturen nicht ausreichend. In der Regel werden verallgemeinernde Formulierungen für Suchanfragen verwendet. Solche „chemischen Muster“ sind im einfachsten Fall Teilstrukturen wie ein Phenylring oder eine Hydroxylgruppe, sie können aber auch aus komplexen logischen Formulierungen aufgebaut sein. Auf der Basis von SMILES hat sich in der Chemieinformatik die SMARTS-Notation zur Beschreibung chemischer Muster etabliert [6]. Im Gegensatz zu SMILES ist diese Notation bereits recht komplex und nicht so intuitiv verständlich. Erste graphische Darstellungen chemischer Muster und Editoren, die diese erstellen, sollen hier die Brücke zur praktischen Anwendung durch jeden Chemiker bilden [7]. Abbildung 2 zeigt ein Beispiel eines SMARTS-Musters und seiner graphischen Darstellung.

Zur Ausgabe von chemischen Strukturen werden heute primär Strukturdiagramme verwendet. Sowohl Softwaresysteme als auch Web-Anwendungen verfügen heute faktisch immer über graphische Benutzerschnittstellen, so dass die Darstellung von Strukturdiagrammen in der für Chemiker vertrauten Form erfolgen kann. Oft wird übersehen, dass sich für den Chemieinformatiker hinter dieser Darstellungsform ein hochkomplexes Informatik-Problem mit dem Namen SDG (structure diagram generation) verbirgt [8]. Gemeint ist die automatische Erzeugung eines Strukturlayouts, welches nicht nur den IUPAC Empfehlungen folgt, sondern auch den häufig hohen ästhetischen Ansprüchen der Betrachter genügt. Auch wenn Softwaresysteme für mehr als 98% der Moleküle in typischen Datenbanken für pharmazeutische und biotechnologische Anwendungen Strukturdiagramme automatisch erzeugen können – dieses Problem ist noch nicht vollständig gelöst.

Dreidimensionale Darstellungen wie sie in Modelling-Softwarepaketen üblich sind, stellen hingegen aufgrund der enormen Graphikleistung heutiger Computer keine technische Herausforderung mehr dar. Sie sind heute in der lebenswissenschaftlichen Forschung nicht mehr wegzudenken, ermöglichen sie doch, eine räumliche Vorstellung beispielsweise von der Passung zwischen Enzymen und Substraten oder Inhibitoren zu erlangen. In den vergangenen zwei Jahrzehnten haben sie ein räumlich-mechanistisches Verständnis von Bindungs- und Reaktionsprozessen ermöglicht, das jedem Chemiker bei der

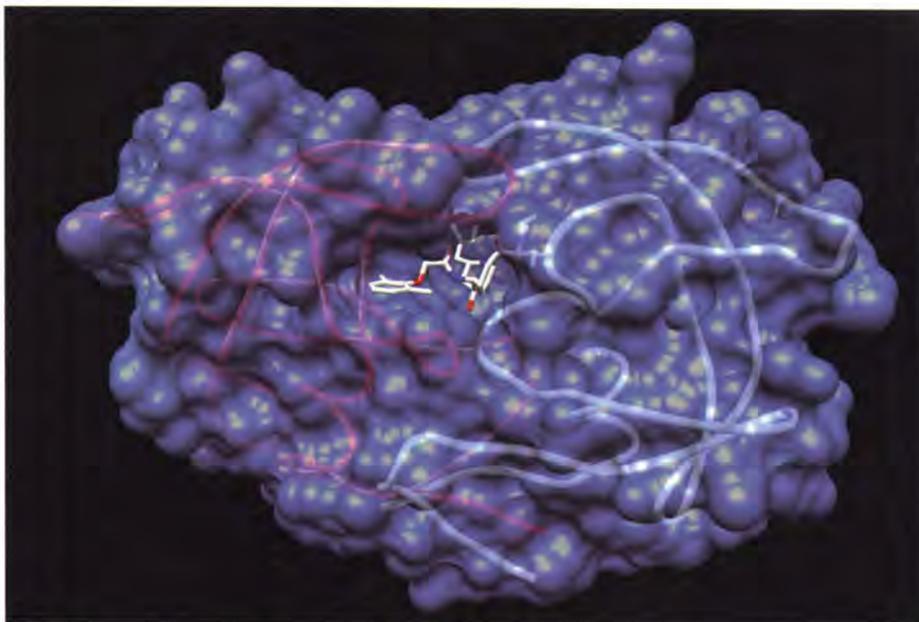


Abb. 3: Protein-Ligand-Komplex, hier der Komplex der HIV-Protease mit dem Inhibitor Lopinavir, einem Inhibitor der dritten Generation

Entwicklung von funktionalen Molekülen wie Pharmaka eine große Hilfe sein kann (siehe hierzu Abb. 3).

2.3 Vorhersage physikochemischer Eigenschaften und Bioaktivität

Die computergestützte Verwaltung chemischer Datenbestände und auch die Visualisierung molekularer Systeme stellt die erste Säule chemieinformatischer Anwendungen dar. Von gleicher Bedeutung sind die computergestützten Vorhersageverfahren, die molekulare Eigenschaften auf Basis der Molekülstruktur berechnen. Berechnungen mit dem Computer können so die Anzahl der Laborexperimente reduzieren, was in der Regel mit deutlichen Einsparungen von Kosten und Zeit verbunden ist. Führt man sich die große Anzahl an chemischen Verbindungen, die heute beispielsweise für die pharmazeutische Forschung verfügbar sind (siehe auch Tab. 1), vor Augen, sind die Vorteile computergestützter Analysen offensichtlich. Die Forschung konzentriert sich dabei natürlich nicht nur auf verfügbare, bereits synthetisierte Verbindungen. Computermodellerte chemische Räume übersteigen die Anzahl verfügbarer Verbindungen um viele Größenordnungen. Beispielsweise verfügt die weltgrößte Pharmafirma Pfizer (USA) über kombinatorische Syntheseprotokolle, die zu über 3 Billionen Verbindungen führen [9]. Computerverfahren ermöglichen, diese Verbindungen bereits zu untersuchen, bevor sie chemisch synthetisiert werden.

Die Arbeiten von Corwin Hansch Mitte der 60er Jahre werden heute als Pio-

nierleistungen des computergestützten Wirkstoffentwurfs betrachtet. Hansch entwickelte lineare Gleichungssysteme, die physikochemische Parameter von Molekülen mit einer speziellen biologischen Aktivität korrelieren [10]. Das Herstellen eines numerischen (quantitativen) Zusammenhangs zwischen Struktur und Aktivität (QSAR, *quantitative structure-activity relationships*) war bahnbrechend. Abbildung 4 zeigt beispielhaft eine QSAR-Analyse nach Free-Wilson. Heute verfügt die Informatik über eine Vielzahl von Methoden, sogenannte maschinelle Lernverfahren, um solche Zusammenhänge automatisch aus Daten abzuleiten. Diese Lernverfahren werden vielfältig zur Vorhersage sowohl von physikochemischen Eigenschaften als auch von Bioaktivität verwendet. Da die Vorhersage schon einfacher Eigenschaften aus der Molekülstruktur allein noch eine große Herausforderung darstellt, bieten Lernverfahren eine Brücke, indem experimentelle Ergebnisse von Molekülmengen zur Ableitung von Modellen eingesetzt werden.

Die immer größer werdende Zahl experimentell bestimmter räumlicher Strukturen legen nahe, diese auch in Vorhersageverfahren intensiv zu nutzen. In der Protein-Datenbank PDB (siehe auch Tab. 1) stehen heute bereits mehr als 80000 makromolekulare Strukturen zur Verfügung, darunter eine Vielzahl von Protein-Ligand-Komplexen. Da die Bindung kleiner Moleküle an Proteine nicht nur für die Funktionsaufklärung, sondern auch für den Entwurf neuer bioaktiver Verbindungen essentiell ist, werden Com-

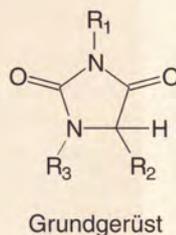
Die Free-Wilson-Analyse

Ziel einer Free-Wilson-Analyse ist es, die Relevanz von R-Gruppen für einen biologischen Effekt zu ermitteln. Als Eingabe dient eine Serie von Molekülen mit gleichem Grundgerüst, an dem unterschiedliche Substituenten variiert werden. Für die Serie wird der biologische Effekt, beispielsweise die Affinität zu einem Protein, experimentell bestimmt. Im folgenden wird ein Gleichungssystem erstellt: Jede Zeile repräsentiert ein vermessenes Molekül, jede Spalte der Matrix einen möglichen Substituenten, nach R-Gruppen sortiert. Die Matrix enthält eine ‚1‘, wenn der Substituent in dem Molekül enthalten ist, eine ‚0‘ in allen anderen Fällen. Jeder mögliche Substituent wird mit einem freien Parameter gewichtet, rechts im Gleichungssystem sind die biologischen Messergebnisse enthalten. Durch Regressionsanalyse können nun die freien Parameter bestimmt werden. Für jede Kombination von Substituenten ermöglichen sie nun die Vorhersage der Bindungsaffinität, in dem die freien Parameter der vorkommenden Substituenten miteinander kombiniert werden.

Modell:

$$\text{biol. Aktivität} = \mu + \sum_{\text{R-Gruppen } i} \sum_{\text{Substituenten } j} a_{ij} \delta_{ij}$$

$$\delta_{ij} = \begin{cases} 1 & \text{Substituent } s_{ij} \text{ an } R_i \\ 0 & \text{sonst} \end{cases}$$



R.-Grp.	Substituent
R ₁	s ₁₁ -H
	s ₁₂ -Cl
	s ₁₃ -methyl
R ₂	s ₂₁ -H
	s ₂₂ -methyl
	s ₂₃ -hydroxyl
R ₃	s ₃₁ -H
	:

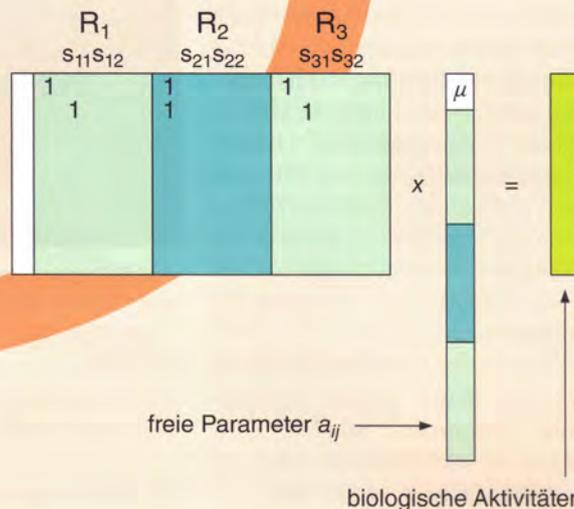


Abb. 4: Die Free-Wilson-Analyse

puterverfahren zur Vorhersage von Protein-Ligand-Komplexen entwickelt. Die zentralen Schritte dieser Verfahren sind die Vorhersage der Komplexgeometrie (das *Docking*) und im Anschluss der freien Bindungsenthalpie (das *Scoring*) [11]. Ändert das Protein beim Bindungsprozess seine Konformation nur unwesentlich, können Dockingverfahren heute in vielen Fällen die korrekte Komplexgeometrie vorhersagen. Die Prädiktion der freien Bindungsenthalpie gestaltet sich jedoch als eine sehr schwierige Aufgabe, sodass bisher nur grobe Abschätzungen erreicht werden. Docking und Scoring werden heute bereits routinemäßig im sogenannten virtuellen Screening für die Suche nach neuen Wirkstoffen eingesetzt.

3 Computerchemie

3.1 Die Hierarchie von Computermodellen in der Chemie

Bekanntlich laufen physikalische (und somit auch chemische) Prozesse auf verschiedenen Zeit- und Größenskalen ab. Je nach Fragestellung kann und muss man daher auf unterschiedliche Modelle zurückgreifen, möchte man diese Prozesse mit Hilfe des Computers simulieren. Die Wahl des Modells ist dabei eine reine Kos-

ten-Nutzen-Analyse. Kann das Modell die Fragestellung hinreichend genau beantworten, und lässt sich das Problem in akzeptabler Zeit berechnen? Prinzipiell ist mit der Entwicklung der Quantenmechanik die Modellentwicklung für die Chemie abgeschlossen worden, liefert sie doch alle wesentlichen physikalischen Gesetzmäßigkeiten für das Verhalten (sub-)atomarer Teilchen. Wendet man die geltenden quantenmechanischen Gleichungen jedoch ohne weitere Vereinfachungen an, so kann bereits die Berechnung eines einzigen Wassermoleküls im Vakuum zu Rechenzeiten von einigen Tagen (und sogar mehr) führen. Chemische Fragestellungen, gerade solche im Bereich der Medikamentenentwicklung, erfordern daher weitere Approximationen [12].

Dies ist in Abbildung 5 „Skalen vs. Modelle“ gezeigt. Dabei überlappen die Rechenmodelle am unteren linken Rand mit Modellen der Teilchenphysik, und am oberen rechten Rand gehen sie in jene der Materialwissenschaften über. Ein wesentlicher Forschungsschwerpunkt der Computerchemie liegt darin, den abgedeckten Skalenbereich zu erweitern, indem entweder effizientere Computerprogramme auf immer schnelleren Rechnern realisiert werden, oder indem neue Näherungsver-

fahren entwickelt werden, die eine schnellere Berechnung erlauben, ohne jedoch wesentliche Verluste in ihrer Genauigkeit aufzuweisen.

Ein besonderes Augenmerk liegt hierbei auf sogenannten Multiskalen-Ansätzen, die zwei oder mehr der aufgezeigten Methoden kombinieren, um so das Anwendungsspektrum zu erweitern. Dabei ist besonders die Verknüpfung von klassischen und quantenmechanischen Methoden (QM/MM-Verfahren) von Interesse, da sich hiermit auch große Biomoleküle gut beschreiben lassen. Die Idee hierbei ist, dass der relevante Teil des chemischen Systems (etwa das aktive Zentrum eines Proteins) auf quantenmechanischem Niveau erfasst wird, während der inerte Teil des Systems (das Proteingerüst) durch Kraftfelder beschrieben wird. Diese Ansätze ermöglichen eine große Rechenzeitersparnis.

3.2 Anwendbarkeit und Limitierungen der Modelle

Die bereits erwähnte Kosten-Nutzen-Analyse kann letztlich auch zu dem Ergebnis führen, dass eine Fragestellung (in dieser Form) nicht simuliert werden kann. So existiert derzeit weder ein Verfahren, um ein ganzes Protein quantenmechanisch

zu berechnen, noch könnte ein Kraftfeld jemals den Tunneleffekt des H-Atoms in einer Wasserstoffbrückenbindung beschreiben. Wo liegen heute also die Grenzen des Machbaren und welche Fragen können wir damit beantworten? Die hier gegebenen Antworten darauf verstehen sich selbstverständlich nur als grobe Richtwerte und hängen im Einzelnen von vielen Detailfragen ab; außerdem sollen sie sich auf die zwei häufigsten Verfahren der Computerchemie beschränken: die Molekülmechanik (Kraftfeldverfahren) [13] und die Dichtefunktionaltheorie (DFT) [14].

Die Molekülmechanik ist das Pendant zum Stäbchenmodell für Moleküle und entspricht am ehesten der chemischen Intuition. Es gibt definierte Bindungen zwischen den Atomkernen, die die relative Anordnung der Atome bestimmen. Dabei werden die Bindungsstärken als Kraftkonstante einer Feder modelliert, und die Dynamik eines solchen Systems gehorcht den Newtonschen Gesetzen der klassischen Mechanik. Die Atome befinden sich im Kraftfeld der Atom-Atom-Wechselwirkungen. Das genaue Potenzial dieser Kraftfelder muss empirisch festgelegt werden, wobei entweder experimentelle Daten oder theoretische Berechnungen auf Basis der Quantenmechanik Verwendung finden. Hiermit lassen sich Dynamiksimulationen von einigen 100 000 Atomen über einen Zeitraum von wenigen Nanosekunden realisieren, wobei die Umsetzung den Zugang zu sehr leistungsfähigen Rechenclustern und mehrere Wochen Rechenzeit erfordert.

Im Gegensatz dazu steht die DFT, die alle Moleküleigenschaften an Hand der Elektronendichte bestimmen kann. Hierzu ist im Prinzip allein die Kenntnis der relevanten Naturkonstanten notwendig, weshalb man von einer *ab-initio*-Methode (also einer Methode, die keine empirische Parametrisierung benötigt) sprechen kann. Dies hat den Vorteil, dass selbst Systeme, über die keine experimentellen Erkenntnisse vorliegen, verlässlich untersucht werden können. Da die Methode auf der Elektronenstruktur und nicht auf atomistischen Kraftfeldern beruht, lassen sich hiermit andere Phänomene beschreiben, wie etwa der Bruch und die Neubildung einer kovalenten

Bindung oder spektroskopische Eigenschaften. Das Limit liegt hier bei einigen hundert Atomen, wenn es etwa um die Aufklärung einer Struktur oder die relative Energie zweier chemischer Systeme geht. Gleichfalls können Dynamiksimulationen durchgeführt werden, doch verringert sich hier die Anzahl der Atome um eine Größenordnung und die Simulationszeit beschränkt sich auf etwa eine Nanosekunde (was für viele chemische

von Punktmutationen auf die Struktur und Dynamik von Proteinen durch diese Studien untersuchen. Die daraus gewonnenen Erkenntnisse über biochemische Prozesse sind dabei natürlich nicht nur genereller Natur, sondern können auch helfen, den Effekt von Medikamenten besser zu verstehen. Ebenso können damit die Auswirkungen von Gendefekten bestimmter Erkrankungen untersucht werden, indem die Folgen der Mutation auf die Aktivität von Proteinen simuliert wird.

Die noch aufwendigeren quantenmechanischen Verfahren finden oft eher einen indirekten Einsatz in der Medikamentenentwicklung. Eine wichtige Fragestellung ist z. B. die Kenntnis von pKs-Werten von noch nicht synthetisierten Verbindungen. Die verlässliche Vorhersage erfordert oft Rechnungen auf DFT-Niveau. Und auch zur Aufklärung der katalytischen Reaktion eines Proteins, welches als potenzielles Target für ein neues Medikament identifiziert worden ist, bedarf es unterstützender DFT-Rechnungen. In diesem Fall kommen vor allem die bereits

erwähnten multiskalaren QM/MM-Verfahren zum Zuge. Wesentlich ist dabei, dass die Rechnungen nicht unbedingt anstelle von, sondern in Kombination mit dem Experiment durchgeführt werden, um ein weiteres Instrument zur Analyse von chemischen Prozessen zur Hand zu haben.

Die steigende Rechenleistung und immer bessere Methoden erlauben inzwischen aber auch erste Studien, in denen bestimmte Moleküleigenschaften gezielt mit quantenchemischen Modellen optimiert werden. Dies wird in nächster Zeit eine immer stärkere Verbreitung finden und sicherlich einen der neuen Trends bei der Anordnung von Computern in der Chemie darstellen. Hierzu werden aber auch das Wissen und die Erfahrung aus der Chemieinformatik benötigt, so dass sich langfristig die Felder Computerchemie und Chemieinformatik stärker verzahnen werden.

4 Chemieinformatik und Computerchemie im Schulunterricht?

Während die wichtige Rolle der Informatik für die Biowissenschaften den meisten Studieninteressierten beispielsweise durch die Genomprojekte bekannt ist, kennen nur wenige die Möglichkeiten

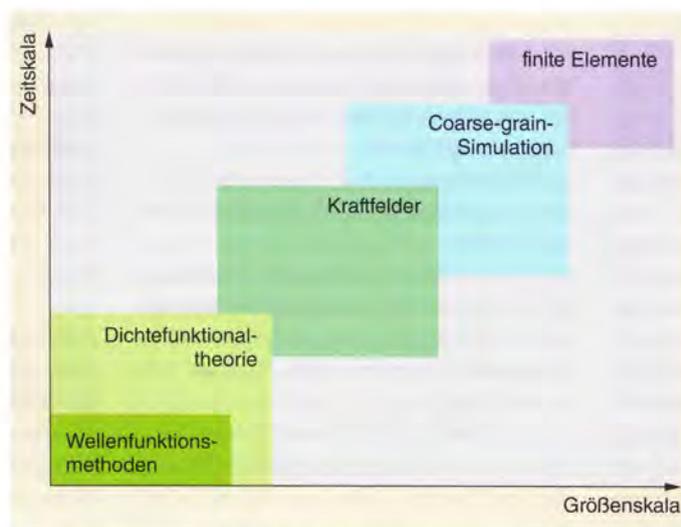


Abb. 5: „Skalen vs. Modelle“: Schematische Zuordnung verschiedener Modelle in der Computerchemie zu Zeit- und Größenskalen

Prozesse nicht ausreichend ist). Der benötigte Rechenaufwand hängt stark von der Systemgröße und der gewünschten Genauigkeit ab, und kann Minuten bis zu Wochen auf modernen Rechnern betragen. DFT liefert wesentlich exaktere Ergebnisse über eine größere Bandbreite von chemischen Systemen als Kraftfelder, dennoch beruht sie (in ihrer derzeitigen Formulierung) nur auf approximativen Modellen, und die Untersuchung und Verbesserung ihrer Anwendbarkeit ist ein aktives Forschungsfeld der Computerchemie.

3.3 Anwendungen von Computermodellen

Mit den unterschiedlichen Methoden (Molekülmechanik und DFT) wird sehr unterschiedlichen Problemen nachgegangen. Die Kraftfeldverfahren finden vor allem in der Simulation von dynamischen Prozessen Anwendung, etwa bei der zeitlichen Entwicklung eines Systems oder der Einstellung von (Lösungs-)Gleichgewichten. Dabei kann eine Simulation z. B. bei der Interpretation von NMR-Daten von Proteinen in Lösung helfen, indem die bevorzugten Konformationen, in denen das Protein in Lösung vorliegt, simuliert werden. Weiter lassen sich auch die Folgen

und den Stellenwert von Computerverfahren in der Chemie. Dabei ist der Computer für Chemiker nicht einfach nur ein Werkzeug, der die bekannten Arbeitsweisen unterstützt. Vielmehr ermöglicht er ganz neue Forschungsansätze und bietet Möglichkeiten, die sich anders nicht erschließen ließen. Diese Tatsache, welche sich in den kommenden Jahren immer weiter im Forschungs- und Entwicklungsumfeld manifestieren wird, sollte auch im Unterricht entsprechend widerspiegelt werden. Für Chemiker stehen Informationen in großen Datenbanken häufig frei im Internet zur Verfügung. Eine optimale Nutzung dieser Ressourcen ist nur mit computergestützten Methoden möglich und bereits heute für Chemiker von großer Bedeutung. Schülerinnen und Schüler sollte ein Bewusstsein für computerchemische Methoden vermittelt werden. Dies kann bei einfachen Internet-Recherchen beginnen und mit ersten eigenen Experimenten am Computer vertieft werden. Wichtig hierbei ist, dass der Computer nicht als Ersatz für das eigentliche Experiment verstanden wird, sondern als ergänzendes Werkzeug mit eigenen Untersuchungs-

möglichkeiten. Der Computer bietet - wie für so vieles - eben nicht nur eine (preiswerte) Alternative, sondern ein Mehr an Erfahrungsmöglichkeiten. ■

Literatur

- [1] International Human Genome Sequencing Consortium, *Initial sequencing and analysis of the human genome*. *Nature* 409 (6822), 860 (2001)
- [2] J. Gasteiger, T. Engel, *Cheminformatics*. WileyVCH, Weinheim 2003
- [3] A. R. Leach, V. J. Gillet, *An introduction to cheminformatics*. Springer, Dordrecht 2005
- [4] J. M. Amigó, J. Gálvez, V. M. Villar, *A review on molecular topology: applying graph theory to drug discovery and design*. *Naturwissenschaften* 96 (7), 749 (2009)
- [5] D. Weininger, *SMILES. 2. Algorithm for Generation of Unique SMILES Notation*. *JCICS* 29 (2), 97 (1989)
- [6] Daylight Chemical Information Systems, Inc. (Aliso Viejo, CA), *Daylight Theory Manual, version 4.9, 2008*, <http://www.daylight.com/dayhtml/doc/theory/index.html>. Accessed August 2012
- [7] K. Schomburg, H.-C. Ehrlich, K. Stierand, M. Rarey, *From Structure Diagrams to Visual Chemical Patterns*. *JCIM* 50 (9), 1529 (2010)

- [8] H. Helson, *Structure diagram generation*. *Reviews in Computational Chemistry*, 13 (11), 313 (1999)
- [9] M. Boehm, T.Y. Wu, H. Claussen, C. Lemmen, *Similarity searching and scaffold hopping in synthetically accessible combinatorial chemistry spaces*. *Journal of Medicinal Chemistry* 51 (8), 2468 (2008)
- [10] C. Hansch, *A Quantitative Approach to Biochemical Structure-Activity Relationships*. *Accounts of Chemical Research*, 2, 1969, 232–239
- [11] C. Sotriffer, *Virtual Screening*. WileyVCH, Weinheim 2011
- [12] F. Jensen, *Introduction to Computational Chemistry*. WileyVCH, Chichester 2007
- [13] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego, London 2002
- [14] W. Koch, M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*. Wiley-VCH, Weinheim 2001

Anschrift der Verfasser

Prof. Dr. Matthias Rarey und JProf. Dr. Tobias Schwabe, Universität Hamburg, Zentrum für Bioinformatik, Bundesstraße 43, 20146 Hamburg, E-Mail: [rarey|schwabe]@zbh.uni-hamburg.de

Paracelsus

Begründer der medizinischen Chemie

H. Remane

Herrn PD Dr. Hans-Georg Bartel (Physikochemiker und Chemiehistoriker in Berlin) zum 70. Geburtstag gewidmet

Die von Paracelsus (1493–1541) begründete medizinische Chemie brachte für die Chemie einen grundlegenden inhaltlichen Wandel und forcierte deren Abkehr vom alchemistischen Tun. Für die Pharmazie leitete sie revolutionäre Veränderungen ein.

Stichworte: Paracelsus, Geschichte der medizinischen Chemie

1 Einführung

Seit den Anfängen der Menschheit werden chemische und pharmazeutische Prozeduren verwendet, um das Leben zu verbessern und angenehmer zu machen bzw. um Kranke zu heilen. Doch lange Zeit geschah dies auf rein empirischer Grundlage. Chemie und Pharmazie als wissenschaftliche Disziplinen begannen ihre Entwicklung erst seit der Renaissance. Entscheidende Paradigmenwechsel erfolgten gegen Ende des 18. Jahrhunderts. In der Chemie legte der französische Naturforscher Antoine Laurent

Lavoisier (1743–1794) mit seiner Oxidationstheorie den Grundstein für die Entwicklung der quantitativen Chemie und die Herausbildung der klassischen Chemie des 19. Jahrhunderts. In Medizin und Pharmazie lösten die Ideen und das Wirken von Paracelsus (vgl. Abb. 1) einen revolutionären Umschwung aus [1]. Im vorliegenden Beitrag soll die Rolle von Paracelsus als Wegbereiter der modernen Chemie und der Pharmazie unter Einbeziehung bereits vorliegender Arbeiten [2, 3, 10, 14–16] diskutiert werden.

2 Der Lebensweg von Paracelsus

Der Lebenslauf von Paracelsus ist uns nur lückenhaft überliefert. Er wurde Ende 1493

als Sohn des Arztes Wilhelm von Hohenheim (um 1457–1534) im Schweizer Kanton Schwyz geboren. Um 1500 wurde der Vater Stadtarzt in Villach (Kärnten), wo auch der Sohn seine Schulbildung erhielt. Sein Medizinstudium führte ihn dann an verschiedene Universitäten in Deutschland und Italien. Nach eigenen Angaben ist er an der 1391 gegründeten Universität Ferrara im Jahre 1515 zum „Doctor beider arzneyen“ (Doctor utriusque medicinae) promoviert worden. Mit dem Wissen, das er beim Studium erworben hatte, war er aber nicht zufrieden. Das Medizinstudium der damaligen Zeit war geprägt durch die dogmatische Vermittlung des Wissens der