

**Ein Porträt des Arbeitskreises
Prof. Dr. Johann Gasteiger**

**25 Jahre
Forschung und Entwicklung
in Chemoinformatik**

**Computer-Chemie-Centrum
und
Institut für Organische Chemie
Universität Erlangen-Nürnberg
91052 Erlangen**

<http://www2.chemie.uni-erlangen.de>



Oktober 2001

Einleitung

Zunehmend setzt sich die Erkenntnis durch, dass die Chemoinformatik gleichberechtigt neben die Bioinformatik treten muss, um den langen Weg vom Gen zum Medikament beschreiten zu können. Dabei hat die Chemoinformatik - früher verkleidet in Namen wie Chemie-Information oder Computerchemie - gerade auch in Deutschland eine lange Tradition. Hier entstanden, nicht zuletzt durch die Förderung im Rahmen des Fachinformationsprogramms der Bundesregierung, einige der weltführenden Chemie-Datenbanken. Unsere Arbeitsgruppe konnte in den letzten 25 Jahren wichtige Impulse zur Entwicklung des Gebiets der Chemoinformatik geben und die Mitarbeiter entfalteten ihre Kenntnisse und ihr Wissen in den verschiedensten Anwendungsbereichen. Wir wollen hier einen Überblick über unsere wichtigsten Forschungsgebiete geben und zeigen, in welchen Anwendungen sie Früchte gebracht haben.



1. Grundlagen zu Chemie-Datenbanken

Computerrepräsentation chemischer Strukturen und Reaktionen

Vor über hundert Jahren entwickelten die Chemiker eine eigene, international verstandene graphische Sprache: Strukturformeln und Reaktionsgleichungen. Um die Möglichkeiten der Computertechnologie zur Verarbeitung chemischer Information nutzen zu können, musste diese Sprache in eine elektronische Form umgesetzt werden.

Dabei waren eine Reihe von Problemen zu lösen: eindeutige Strukturrepräsentation, Ringerkennung, Behandlung von π -Systemen, Aromatenerkennung, Tautomerie, Repräsentation der Stereochemie sowie Codierung der molekularen Chiralität.

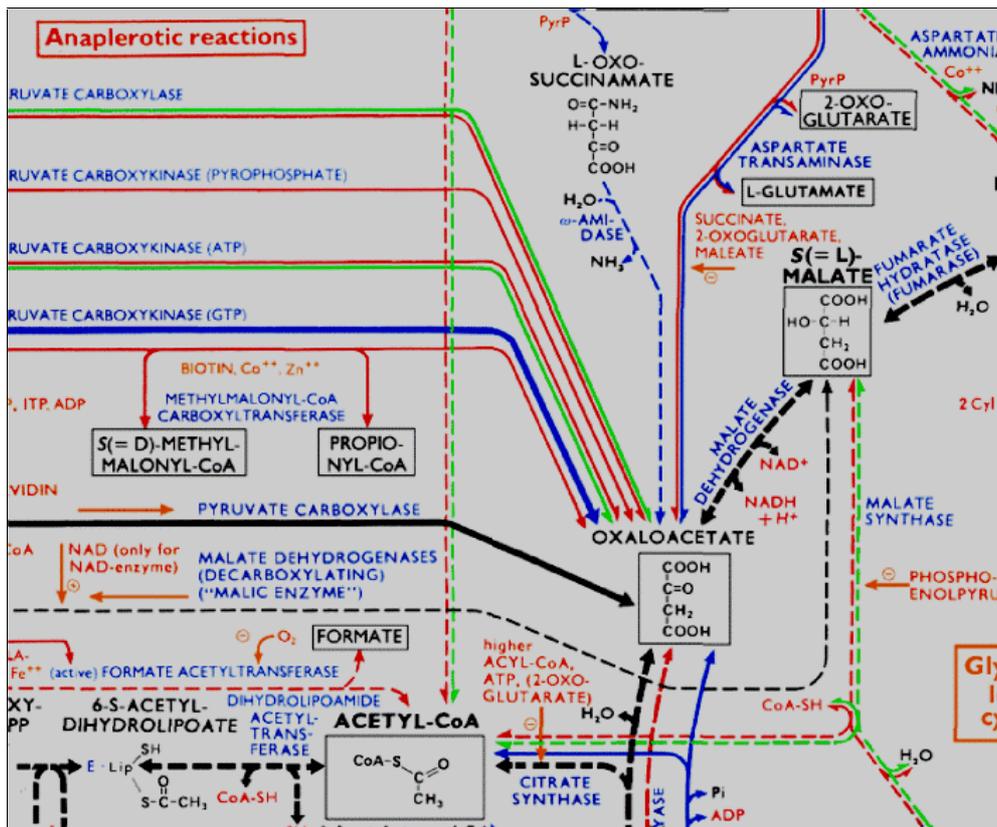
Anwendungen

Eines der stolzesten Produkte, zu dem unsere Arbeiten beitragen konnten, ist die Beilstein-Datenbank mit Faktendaten zu 7 Millionen organischen Verbindungen. Selbstverständlich haben viele Personen zu so einem monumentalen Werk beigetragen, jedoch ohne der tatkräftigen und weitsichtigen Leitung durch *Clemens Jochum*, dem ersten Doktoranden aus der Arbeitsgruppe, wäre es nicht zustande gekommen. Er hat das hier erworbene Rüstzeug sehr effizient eingesetzt und anfänglich auf die Beratung durch die Gruppe bei der Entwicklung der Datenstruktur zurückgreifen können.

Heinz Saller und *Peter Löw* haben schon vor 15 Jahren den Sprung gewagt, ihr in der Gruppe erworbenes Wissen in die Gründung einer Firma, *ChemoData*, umzusetzen. *InfoChem*, die Nachfolgefirma von ChemoData, vertreibt nun Struktur- und Reaktionsdatenbanken und entwickelt innovative Chemie-Informationssysteme, vor allem für Verlage wie Springer, Wiley-VCH und Thieme Verlag.

Die Expertise der Arbeitsgruppe in der Computerverarbeitung chemischer Strukturen und Reaktionen wurde zum Aufbau der ChemInform Reaktionsdatenbank eingesetzt. Im Auftrag des Fachinformationszentrums (FIZ) Chemie, Berlin und unter der Projektleitung von *Johann Gasteiger* führte ChemoData die wesentlichen Softwareentwicklungsarbeiten durch. Die ChemInform Reaktionsdatenbank ist zur Zeit die weltweit erfolgreichste inhouse-Reaktionsdatenbank; sie wird vom FIZ Chemie, Berlin, erstellt und von MDL Information Systems, Inc., San Leandro, CA, USA, vermarktet.

Wolf-Dietrich Ihlenfeldt hat mit dem **CACTVS** System ein umfassendes Programmpaket zur elektronischen Verarbeitung chemischer Strukturen, Reaktionen und Spektren erstellt. Auf dieser Grundlage wurde die Datenbank des National Cancer Institutes, die mit über 250.000 Strukturen derzeit größte kostenlos zugängliche Chemiedatenbank, erstellt.



In jüngster Zeit haben wir uns Biochemischen Reaktionswegen zugewendet. Das Poster „Biochemical Pathways“, ursprünglich von Boehringer Mannheim (nun Roche) erstellt, wurde in einem Kooperationsprojekt mit dem Spektrum Akademischer Verlag, mit LionBioscience und den Universitäten Passau und Mannheim in eine Reaktionsdatenbank konvertiert. Biochemische Reaktionen lassen sich nun mit allen modernen Suchmethoden über Strukturen und Reaktionen analysieren.

Vor einigen Jahren wurde die Firma *Molecular Networks GmbH* gegründet, um in einer Kooperation mit der Universität Erlangen-Nürnberg, teilweise aufbauend auf Software aus der Arbeitsgruppe, Verfahren zur Auswertung chemischer Information und zur Wirkstoffentwicklung zu erstellen.

Mitarbeiter

Clemens Jochum, Deutsche Bank AG, Eschborn
Heinz Saller und *Peter Löw*, InfoChem GmbH, München
Christian Hiller, SAP, Redmond, WA, USA
Bernd Christoph (†), FIZ Chemie, Berlin
Leo Gann, Siemens AG, München
Bruno Bienfait, ChemCodes, Research Triangle Park, NC, USA
João Aires de Sousa, Lissabon, Portugal
Dietrich Trümbach, Biomax, München
Wolf-Dietrich Ihlenfeldt, Universität Erlangen-Nürnberg

2. Wissensbasierte Systeme

Simulation chemischer Reaktionen und Syntheseplanung

Von Anfang an beschäftigte sich die Gruppe mit der Entwicklung von Programmsystemen zur Reaktionssimulation und zur Syntheseplanung. Ein wesentliches Merkmal all dieser Arbeiten ist, dass chemische Reaktionen streng mechanistisch gesehen werden. Dementsprechend mussten Modelle und Verfahren zur Berechnung physikochemischer Effekte entwickelt werden, um die Triebkräfte chemischer Reaktionen und Reaktionsmechanismen modellieren zu können. Die Arbeiten mündeten in drei Systeme: **EROS** (Elaboration of Reactions for Organic Synthesis) zur Simulation und Vorhersage organisch-chemischer Reaktionen und **WODCA** (Workbench for the Organization of Data for Chemical Applications) zur Planung organisch-chemischer Synthesen. Mit dem Anwachsen von Reaktionsdatenbanken wurde es zunehmend interessanter, aus diesen Informationsquellen Wissen über chemische Reaktionen abzuleiten. Dazu wurde das Programmsystem **CORA** (Classification of Organic Reactions for Applications) entwickelt.

Anwendungen

Das Syntheseplanungsprogramm WODCA wird von *Molecular Networks* vertrieben und ist bei einigen chemischen Firmen bereits im praktischen Einsatz.

Mitarbeiter

Kazumi Yuki, Sumitomo Chemical Co., Ltd., Tokio, Japan

Peter Rose, Pfizer, La Jolla, CA, USA

Roland Wanke, Siemens AG, München

Ulrich Hondelmann, ecomed, Buchloe

Wolfgang Witzenbichler, Frankfurt

John Rose, University of South Carolina, Columbia, SC, USA

Akio Tanaka, Sumitomo Chemical Co., Ltd., Osaka, Japan

Lingran Chen, MDL Information Systems, Inc., San Leandro, CA, USA

Ralf Fick, Roche, Penzberg

Susanne Bauerschmidt, Organon, Oss, NL

Robert Höllering, Creon-Lab-Control AG, Köln

Larissa Steinhauer, DANET, Darmstadt

Matthias Pförtner, Molecular Networks, Erlangen

Oliver Sacher, Molecular Networks, Erlangen

Thomas Kostka, ChemCodes, Erlangen

Achim Herwig, Universität Erlangen-Nürnberg

Jörg Marusczyk, Universität Erlangen-Nürnberg

Markus Sitzmann, Universität Erlangen-Nürnberg

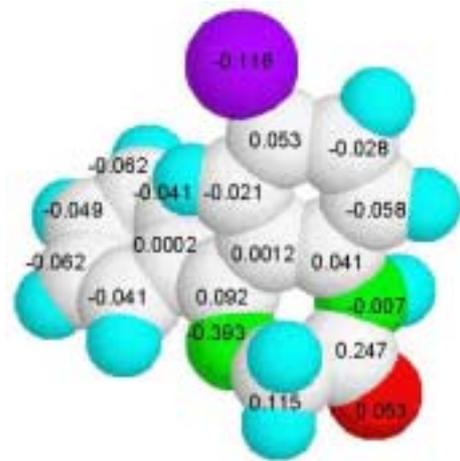
3. Chemisches Wissen

Physikochemische Effekte

Die Modellierung von Reaktionsmechanismen verlangte nach der Abschätzung der chemischen Reaktivität. Dazu wurden empirische Verfahren zur Berechnung fundamentaler physikochemischer Effekte, wie Ladungsverteilung, induktiver, sterischer, Resonanz- und Polarisierbarkeits- Effekt sowie von Bildungswärmen und Bindungsdissoziationsenergien entwickelt. Bei diesen Verfahren wurde großer Wert auf hohe Rechengeschwindigkeiten gelegt, um auch größere Datensätze von Millionen von Strukturen, wie sie in kombinatorischen Bibliotheken auftreten, behandeln zu können.

Anwendungen

Das **PEOE**-Verfahren zur Berechnung von Ladungsverteilungen ist praktisch in allen Molecular Modeling Programmpaketen eingebaut. Das Programmpaket **PETRA** (Parameter Estimation for the Treatment of Reactivity Applications) zur Berechnung einer Reihe physikochemischer Effekte lässt sich auf große Datensätze anwenden und wird von *Molecular Networks* vertrieben.



Mitarbeiter

Mario Marsili, Università di l' Aquila, Italien
Maria D. Guillen, Universidad Vitoria-Gasteiz, Spanien
Michael G. Hutchings, DyStar UK, Manchester, UK
Angela Fröhlich, Freising
Monica Fato, Italien
Klaus Rafeiner, empolis professional services, Gütersloh
Thomas Kleinöder, Universität Erlangen-Nürnberg

4. Wirkstoffforschung

Methoden zur Leitstrukturfindung und -optimierung

Die gleichen physikochemischen Effekte, die den Bruch und das Knüpfen von Bindungen in chemischen Reaktionen verursachen, sind auch für die Bindung eines biologisch aktiven Liganden an seinen Protein-Rezeptor verantwortlich. Es lag also auf der Hand, dass wir uns früher oder später dem Zusammenhang zwischen chemischer Struktur und biologischer Aktivität zuwenden würden. Um die dreidimensionale Welt der Moleküle zu erschließen, wurde der automatische 3D-Strukturgenerator **CORINA** (COoRdINAtes) erstellt sowie Methoden zur Behandlung der konformativen Flexibilität aufgebaut (**ROTATE** und **GAMMA**).

Die Entwicklung von Verfahren zum Auffinden von Leitstrukturen, zur Aufstellung quantitativer Struktur-Wirkungsbeziehungen und zur Optimierung von Wirkstoffen sowie Fragen zur Ähnlichkeit und Diversität kombinatorischer Bibliotheken standen in den letzten Jahren im Zentrum unserer Arbeiten. Eine entscheidende Rolle spielen dabei Strukturcodierungsverfahren (**AUTOCORR**, **SURFACE** und **ARC**).

Anwendungen

Von Anfang an wurden diese Arbeiten, gefördert vom BMBF, in enger Kooperation mit Pharmafirmen durchgeführt: *Merck KGaA*, *BASF*, *Byk Gulden* und *Solvay Pharmaceuticals*. Der Erfahrungsaustausch mit unseren Partnern in diesen Firmen sowie den Universitäten Tübingen und Marburg war entscheidend für unsere Fortschritte. Weitere Kooperationen bildeten sich mit *Pfizer*, USA, *Organon*, NL und *Janssen Pharmaceuticals*, Belgien, heraus.

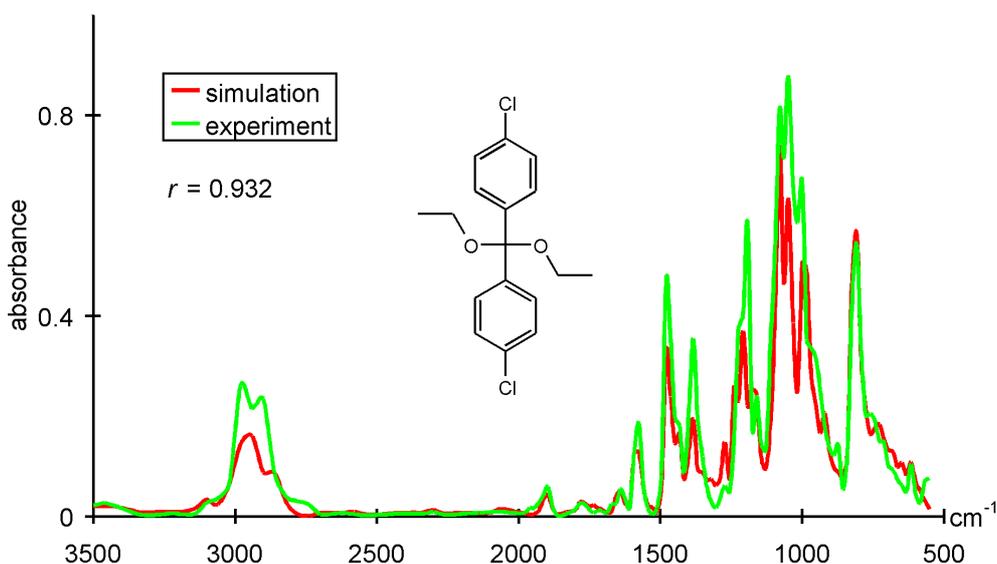
CORINA wird nun von *Molecular Networks* vertrieben. Weltweit existieren über 60 Installationen, insbesondere bei den großen Pharmafirmen. Darüberhinaus entwickelt und vertreibt Molecular Networks Software zur Leitstruktursuche und -optimierung.

Mitarbeiter

Christine Rudolph, Aventis, Frankfurt/Main
Jens Sadowski, AstraZeneca, Göteborg, S
Soheila Anzali, Merck KGaA, Darmstadt
Jarosław Polański, Katowice, PL
Markus Wagener, Organon, Oss, NL
Sandra Handschuh, Boehringer Ingelheim, Biberach
Andreas Teckentrup, Boehringer Ingelheim, Biberach
Christof H. Schwab, Molecular Networks, Erlangen
Alexander von Homeyer, Universität Erlangen-Nürnberg
Ai-Xia Yan, Universität Erlangen-Nürnberg

5. Chemische Analytik

Simulation und Auswertung von Spektren



Um Spektren zu interpretieren, greift der Chemiker auf seine Erfahrung zurück und dementsprechend ist der Zusammenhang zwischen der Struktur einer Verbindung und seinen Spektren ein klassisches Anwendungsgebiet für automatische Lernverfahren geworden.

Wir investierten viel Arbeit in die Simulation von Massenspektren, eine endgültige Lösung konnten wir aber noch nicht erzielen. Dafür haben wir schöne Erfolge bei der Simulation und automatischen Auswertung von Infrarotspektren erzielen können, und neuerdings gibt es vielversprechende Ergebnisse bei der Simulation von ¹H NMR Spektren.

Anwendungen

Auch diese Arbeiten wurden anfänglich vom BMBF, dann vom DFN-Verein in Kooperation mit *Chemical Concepts*, Weinheim, gefördert. Das Verfahren zur Simulation von Infrarotspektren wird über das Internet weltweit kostenlos zur Benutzung angeboten.

Mitarbeiter

Wolfgang Hanebeck, Merck KGaA, Darmstadt
Klaus-Peter Schulz, Arizona State University, Tempe, AZ, USA
Jan Schuur, TechniData, Walldorf
Valentin Steinhauer, DANET, Darmstadt
Paul Selzer, Novartis, Basel, CH
Markus Hemmer, Creon-Lab-Control AG, Köln

6. Lernen aus Daten

Neuronale Netze und Genetische Algorithmen

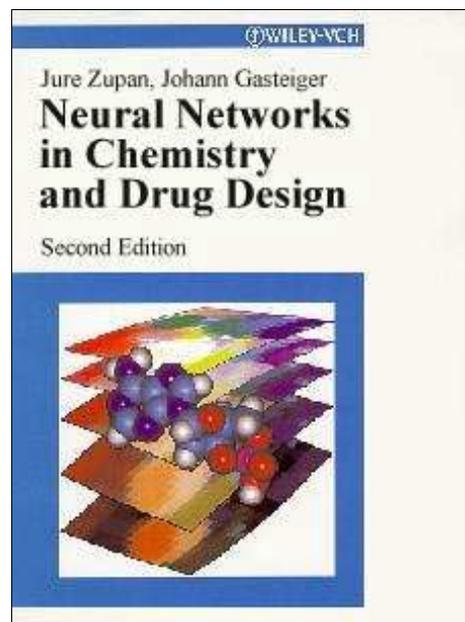
Aus Daten zu lernen ist eines der Grundprinzipien der Chemie. Von Anfang an setzten wir dafür statistische Verfahren und Mustererkennungsmethoden ein. Gleich nach dem Aufkommen leistungsfähiger künstlicher neuronaler Netze wandten wir uns diesen induktiven Lernverfahren zu. Gegenwärtig nehmen sie eine zentrale Rolle in unseren Arbeiten ein; praktisch jedes Gruppenmitglied verwendet sie früher oder später in seinen wissenschaftlichen Arbeiten. Die Ermittlung der Zusammenhänge zwischen Struktur und biologischer Aktivität, Reaktivitätsmodelle, Vergleich kombinatorischer Bibliotheken, Wissensextraktion aus Reaktionsdatenbanken, Bestimmung strategischer Bindungen in der Synthesepaltung, Modellierung der Zusammenhänge zwischen Struktur und Spektren sind Problemstellungen, die wir mit neuronalen Netzen behandeln. Desweiteren setzen wir genetische Algorithmen für schwierige Optimierungsprobleme ein, vor allem zur Ermittlung von Pharmakophoren, um die strukturellen Voraussetzungen für eine gewünschte biologische Aktivität, auch ohne Kenntnis der Rezeptorstruktur, festzustellen.

Anwendungen

Zusammen mit *Prof. Jure Zupan*, National Institute of Chemistry, Ljubljana, Slowenien, machten wir vor fast 10 Jahren die ersten Schritte in die aufregende Welt der künstlichen neuronalen Netze. Ein gemeinsam verfasstes Buch (*J. Zupan, J. Gasteiger: Neural Networks in Chemistry and Drug Design*) liegt nun schon in zweiter Auflage vor und unterstreicht somit die Bedeutung dieses Gebiets.

Mitarbeiter

Xinzhi Li, Kanada
Vera Simon, AKR Computer, Ismaning
Lothar Terfloth, Universität Erlangen-Nürnberg
Giorgi Lekishvili, Universität Erlangen-Nürnberg
Ulrike Burkard, Universität Erlangen-Nürnberg



7. Lernen mit neuen Medien

Vernetztes Studium-Chemie

Mit unseren Erfahrungen im Einsatz des Computers in der Chemie und der Verwendung des World Wide Webs zur Informationsvermittlung war es kein großer Schritt mehr, dieses Medium auch zur Unterstützung der Lehre einzusetzen.



Anwendung

Zusammen mit weiteren 15 Arbeitskreisen an 14 deutschen Hochschulen wird unter der Federführung des FIZ CHEMIE, Berlin, im Rahmen des vom BMBF geförderten Projekts „Vernetztes Studium - Chemie“ der gesamte Wissensstoff des Basisstudiums Chemie multimedial aufbereitet. In unserer Gruppe werden die Themenbereiche Chemie-Information, Chemoinformatik sowie Chemie für Mediziner bearbeitet.

Mitarbeiter

Universität Erlangen-Nürnberg:

Thomas Engel, Angelika Hofmann, Axel Schunk, Simon Spycher

Danksagungen

Es erforderte viel Mut und Glauben, in den letzten 25 Jahren das Gebiet der Chemoinformatik mit aufzubauen; ich möchte mich bei meinen Mitarbeitern dafür bedanken, dass sie mit mir diese Schritte gegangen sind und immer an den Erfolg geglaubt haben.

Die Finanzierung unserer Forschungsarbeiten erfolgte fast ausschließlich über Drittmittel: als erstes durch die Deutsche Forschungsgemeinschaft. BMBF-Mittel erlaubten uns dann unsere Arbeiten in bedeutend größerem Rahmen durchzuführen. Wichtig waren auch immer Mittel aus der Industrie: sehr früh setzte ICI, UK, Vertrauen in unsere Entwicklungsarbeiten, gefolgt von Sumitomo Chemical Industries, Japan, sowie von Shell Forschung, Schwabenheim und in jüngster Zeit Pfizer, USA. Im Rahmen von BMBF-Projekten konnten wir fruchtbare Kooperationen mit Merck KGaA, BASF, Byk Gulden und Solvay Pharmaceuticals aufbauen. Auf unbürokratische Weise erhielten wir Unterstützung vom Verband der Chemischen Industrie. Der Freistaat Bayern und die Universität Erlangen-Nürnberg haben uns durch die Anmietung von Räumen ausgezeichnete Arbeitsbedingungen zur Verfügung gestellt. Aber nochmals: mein tiefster Dank gilt meinen Mitarbeitern. Und nicht zuletzt nahmen mir meine Sekretärinnen viel Last von den Schultern: Sabine Sell, Jill Seidlitz, Angela Döbler und Ulrike Scholz.



Prof. Dr. Johann Gasteiger
Computer-Chemie-Centrum und
Institut für Organische Chemie
Universität Erlangen-Nürnberg
Nägelsbachstr. 25
D-91052 Erlangen
Tel.: 09131/85-26570
Fax: 09131/85-26566
E-Mail: gasteiger@chemie.uni-erlangen.de
<http://www2.chemie.uni-erlangen.de/>