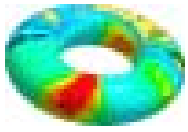
A Portrait of the Research Group of
Prof. Dr. Johann Gasteiger

25 years of Research and Development in Chemoinformatics

Computer-Chemie-Centrum
and
Institute of Organic Chemistry
University of Erlangen-Nürnberg
91052 Erlangen, Germany

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



October 2001

Introduction

It is increasingly realized that chemoinformatics has to work in concert with bioinformatics to find the long way from a gene to a drug. Chemoinformatics - previously called Chemical Information or Computer Chemistry - has a long history in Germany. Some of the world-leading databases in chemistry have been built under funding by the Fachinformationsprogramm of the Federal Government. In the last 25 years our research group could give some decisive impulses to the development of the field of chemoinformatics. The members of the group used their knowledge in a variety of applications. Here, a brief overview of our major research areas is given and it is shown where these endeavors have brought fruits.

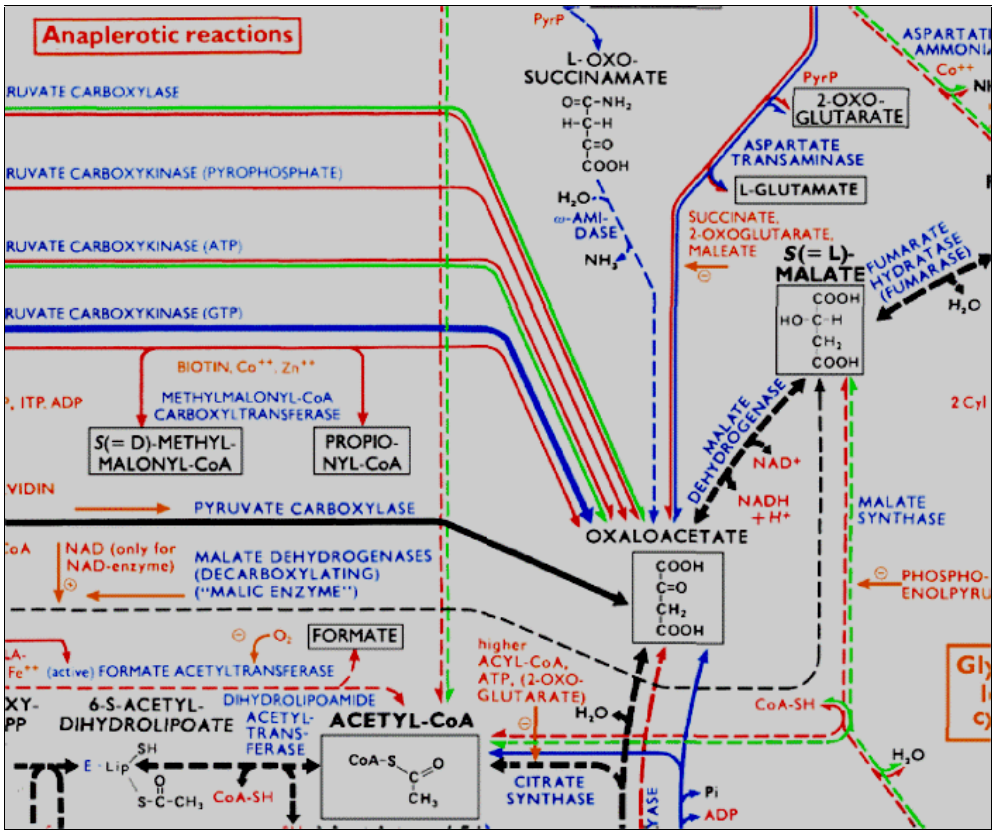
1. Foundations of Chemical Databases

Computer-Representation of Chemical Structures and Reactions

More than 100 years ago, chemists developed an internationally understood graphical language of its own: structural formulas and reaction equations. This language had to be translated into an electronic form in order to utilize computer technology for the processing of chemical information. A series of problems had to be solved such as unique and unambiguous representation of chemical structures, ring perception, treatment of π - systems, perception of aromaticity, treatment of tautomers, representation of stereochemistry, coding of molecular chirality.

Applications

One of the proudest products to which we could contribute is the Beilstein-Database with factual data for 7 million organic compounds. Clearly, many people have contributed to this monumental work but without the far-sighted and efficient leadership through *Clemens Jochum*, the first graduate student of our research group, this database would not have come to life. He made best use of the knowledge learnt in our group and our consulting laid the foundation to the data structures.
Heinz Saller and *Peter Löw* dared already 15 years ago to found a software company, *ChemoData*, to make use of the knowledge gained in our group. *InfoChem*, the successor of ChemoData, distributes chemical structure and reaction databases and develops innovative chemical information systems, particularly for publishing houses such as Springer, Wiley-VCH and Thieme Verlag.
The expertise of our research group in computer processing of chemical structures and reactions was instrumental in the building of the ChemInform Reaction Database. *Johann Gasteiger* was the project manager for FIZ Chemie, Berlin and ChemoData performed the essential software development. Presently, the ChemInform Reaction Database is the most widely used reaction database. It is produced by FIZ Chemie, Berlin and distributed by MDL Information Systems, Inc., San Leandro, CA, USA.
Wolf-Dietrich Ihlenfeldt has developed the **CACTVS** system, a comprehensive package for the processing of chemical structures, reactions, and spectra. On this basis the database of the National Cancer Institute has been implemented on the internet. With 250,000 structures it is presently the largest chemistry database accessible free of charge.



In recent years, we have converted the information on the Poster *"Biochemical Pathways"*, originally produced by Boehringer Mannheim (now Roche) into a reaction database. This project was pursued in cooperation with Spektrum Akademischer Verlag, LionBioscience, and the Universities of Mannheim and Passau. Biochemical Reactions can now be analyzed with all modern structure and reaction search methods.
A few years ago we founded the company *Molecular Networks GmbH*, to develop and distribute software for the analysis of chemical information and for drug design. This is done in cooperation with the University of Erlangen-Nuernberg building on software developed in our research group.

Coworkers

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, University of Erlangen-Nürnberg

2. Knowledge Based Systems

Simulation of Chemical Reactions and Synthesis Design

From the very beginning our research group developed program systems for the simulation of chemical reactions and for the design of organic syntheses. A major characteristics of this endeavor is that chemical reactions are treated in a mechanistic manner. Therefore, we had to develop models and methods for the calculation of physicochemical effects to model the driving forces of chemical reactions and of reaction mechanisms. The development resulted in three systems: **EROS** (Elaboration of Reactions for Organic Synthesis) for the simulation and prediction of organic reactions, and **WODCA** (Workbench for the Organization of Data for Chemical Applications) for the design of organic syntheses. With the availability of reaction databases it became increasingly interesting to derive knowledge on chemical reaction from this information source. The system **CORA** (Classification of Organic Reactions for Applications) was developed for this purpose.

Applications

The synthesis design system WODCA is distributed by *Molecular Networks* and is in practical use at a few chemical companies.

Coworkers

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, University of Erlangen-Nürnberg
Jörg Marusczyk, University of Erlangen-Nürnberg
Markus Sitzmann, University of Erlangen-Nürnberg

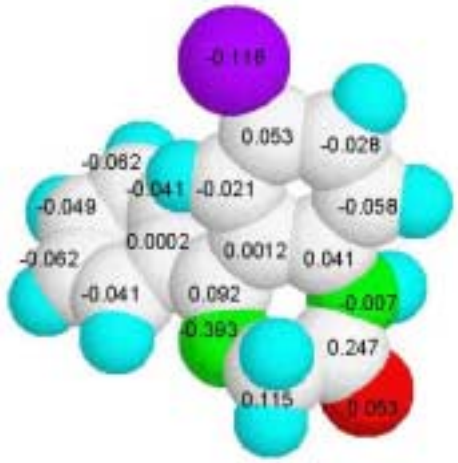
3. Chemical Knowledge

Physicochemical Effects

The modeling of reaction mechanisms asked for an evaluation of chemical reactivity. To this effect, empirical methods for the calculation of fundamental physicochemical effects such as charge distribution, inductive, steric, resonance and polarizability effects as well as heats of formation and bond dissociation energies. Great attention was devoted to come up with methods that require short computation times in order to be able to process large datasets of millions of structures such as those encountered in combinatorial libraries.

Applications

The **PEOE** method for the calculation of charge distributions is incorporated in basically all molecular modeling systems. The program package **PETRA** (Parameter Estimation for the Treatment of Reactivity Applications) can be applied to the calculation of physicochemical effects in large datasets and is distributed by *Molecular Networks*.



Coworkers

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

4. Drug Design

Methods for the Discovery and Optimization of Lead Structures

The same physicochemical effects causing the breaking and making of bonds in chemical reactions are also responsible for the binding of a biologically active ligand to its protein receptor. Thus, it was quite natural that we turned our attention to modeling the relationships between chemical structure and biological activity. The automatic 3D structure generator **CORINA** (COoRdINAtes) was developed to access the three-dimensional world of molecules. Furthermore, we developed methods for the treatment of conformational flexibility of molecules (**ROTATE** and **GAMMA**).
The development of methods for the discovery of lead structure, the establishment of structure-activity relationships, the optimization of drugs, the definition of similarity and diversity of combinatorial libraries are since several years in the center of our attention. All these methods critically hinge on structure coding methods (**SURFACE**, **AUTOCORR**, **ARC**).

Applications

From the very beginning these developments have been funded by the Federal Ministry of Education and Research (BMBF) in close cooperation with pharmaceutical companies such as *Merck KGaA, BASF, Byk Gulden* and *Solvay Pharmaceuticals*. The exchange of experience with our partners in industry and in the Universities of Tübingen and Marburg was decisive for our progress. Additional cooperations were established with *Pfizer,* USA, *Organon,* NL and *Janssen Pharmaceuticals,* Belgium.

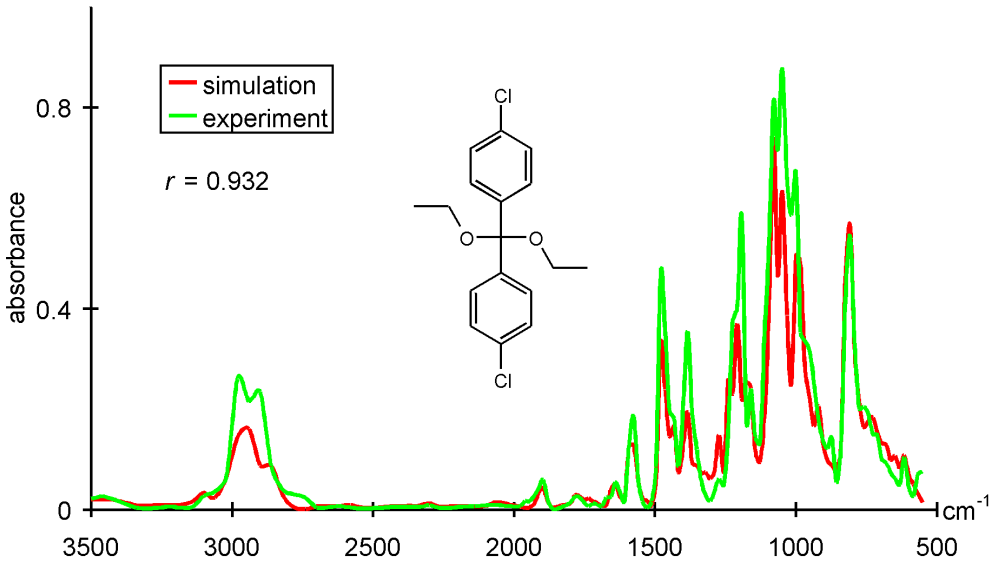
CORINA is distributed by *Molecular Networks GmbH*. Presently, we have more than 60 installations worldwide, particularly in all major pharmaceutical companies. Furthermore, Molecular Networks develops and distributes software for the discovery and optimization of lead structures.

Coworkers

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

5. Chemical Analysis

Simulation and Analysis of Spectra



The interpretation of spectra asks for chemical experience. Therefore, the relationships between the chemical structure of a compound and its spectra is a typical area for applying automatic learning methods.
We put much efforts into the simulation of mass spectra but have not yet been able to conclude this project. However, we have obtained nice results in the simulation and analysis of infrared spectra. And in recent work we obtained promising results in the simulation of ^{1}H NMR spectra.

Applications

This work was also initially funded by the Federal Ministry of Education and Research (BMBF) and then by the German Research Network (DFN-Verein) in cooperation with *Chemical Concepts,* Weinheim. The method for the simulation of infrared spectra has been made generally accessible over the internet.

Coworkers

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

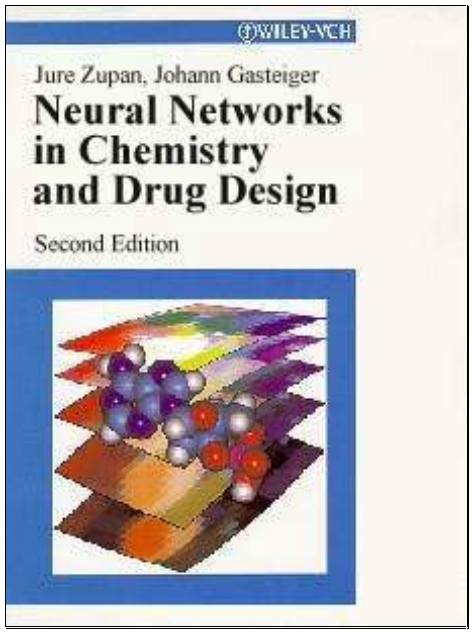
6. Learning from Data

Neural Networks and Genetic Algorithms

A central approach in chemistry is to learn from data. From the outset we used statistical and pattern recognition methods. With the advent of powerful artificial neural networks we have studied these inductive learning methods. Presently, neural networks play a major role in our research: basically each group member will utilize them sooner or later. The establishment of relationships between chemical structure and biological activity, models of chemical reactivity, the comparison of combinatorial libraries, knowledge acquisition of reaction databases, the definition of strategic bonds in synthesis design, the modeling of the relationships between chemical structure and spectra all are problems that we study by neural networks. Furthermore, we use genetic algorithms for difficult optimization problems such as the finding of pharmacophores to establish the structural requirements for biological activity without having a detailed knowledge of the structure of the receptor.

Applications

Nearly 10 years ago we made our first steps into the exciting world of neural networks together with *Jure Zupan*, National Institute of Chemistry, Ljubljana, Slovenia. A coauthored book *(J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design)* is now in its second edition thus underlining the importance of this field.



Coworkers

Xinzhi Li, Canada
Vera Simon, AKR Computer, Ismaning
Lothar Terfloth, University of Erlangen-Nürnberg
Giorgi Lekishvili, University of Erlangen-Nürnberg
Ulrike Burkard, University of Erlangen-Nürnberg

7. Learning with New Media

Networked Study in Chemistry

Our experience in the application of computers in chemistry and the use of the internet destined us to also employ computers in our teaching assignments.

Applications

We develop multimedia tools for teaching chemistry together with 15 research groups at 14 German universities under the administration of FIZ Chemie, Berlin. Our group prepares material for teaching chemical information, chemoinformatics, and chemistry for medical students.

Coworkers

University of Erlangen-Nürnberg:
Thomas Engel, Angelika Hofmann, Axel Schunk, Simon Spycher

Acknowledgements

It asked for a lot of courage and conviction to establish the field of Chemoinformatics in the last 25 years. I want to thank my coworkers for embarking with me on this endeavor and believing in our success.
The financing of our research came nearly exclusively from outside, first from the German Research Council (DFG). The funding by the German Ministry of Education and Research (BMBF) allowed us to pursue our research on a larger scale. Funds from chemical industry have always been important: quite early on ICI, UK, put trust in our research, then followed by Sumitomo Chemical Industries, Japan, later by Shell Research, Schwabenheim, and finally Pfizer, USA.
BMBF projects allowed us to establish fruitful collaborations with Merck KGaA, BASF, Byk Gulden and Solvay Pharmaceuticals. Quite flexible support came from the Association of Chemical Industries in Germany (VCI). The state of Bavaria and the University of Erlangen-Nürnberg rented rooms that provide our research group with excellent working conditions.
Again, however, my sincere thanks go to my coworkers. And, last not least, my secretaries Sabine Sell, Jill Seidlitz, Angela Döbler and Ulrike Scholz made life easier for me.

Prof. Dr. Johann Gasteiger
Computer-Chemie-Centrum and
Institute for Organic Chemistry
University of Erlangen-Nürnberg
Nägelsbachstr. 25
D-91052 Erlangen
Tel.: +49 9131 / 85-26570
Fax: +49 9131 / 85-26566
E-Mail: gasteiger@chemie.uni-erlangen.de
http://www2.chemie.uni-erlangen.de/