

web-based cheminformatics for bench chemists

Readily accessible tools with easy to use interfaces, especially integrated into regularly used desktop applications, can improve the impact of cheminformatics in pharmacological research. Starting either from a single molecule or from a larger data set, these tools provide information that can considerably improve the efficiency of drug discovery and development. The use of a standard web browser as an output tool allows the display of calculation and prediction results combined with links to additional internal and external information resources.

Computational methods get more and more involved into the modern medicinal chemist's work. This is mainly a result of an exponential growth of the amount of data needed to be processed and analysed in the process of modern drug discovery and development. Typical tasks involved in this process include the following:

- fast access to structural and bioactivity data from the corporate database
- substructure and molecular similarity searches
- calculation of physicochemical molecular properties
- SAR (structure-activity relationship) analysis
- selection of representative or diverse subset of molecules from a large set

**By Peter Ertl,
Wolfgang Miltz,
Bernhard Rohde
and Paul Selzer**

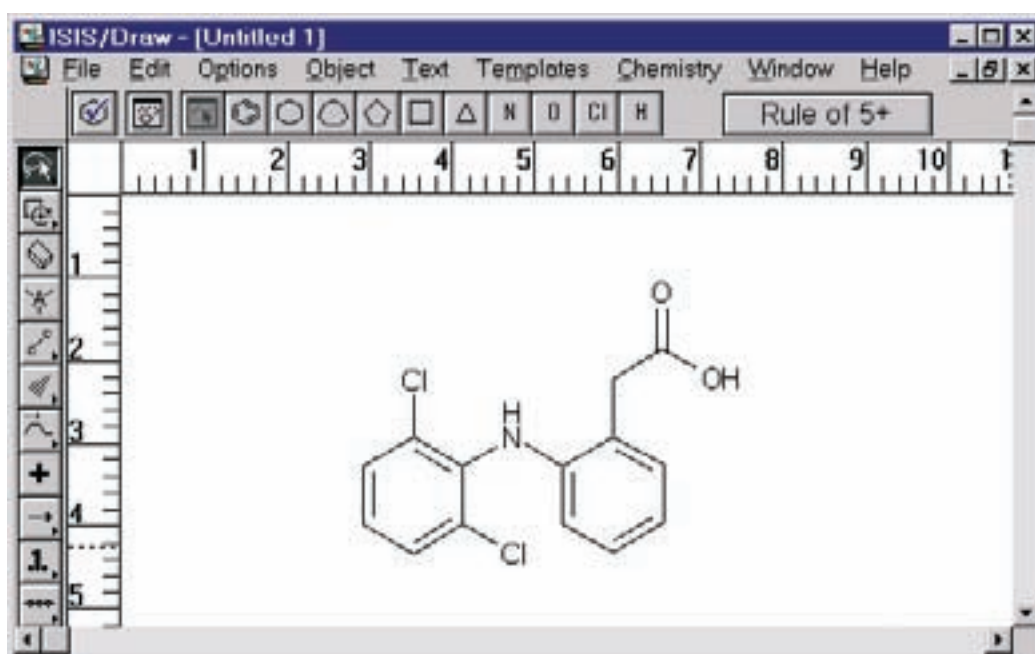


Figure 1
Link to cheminformatics engine incorporated into ISIS/Draw application. Simple click on the Rule 5+ button starts calculations of molecular properties and launches web browser with the results

Informatics

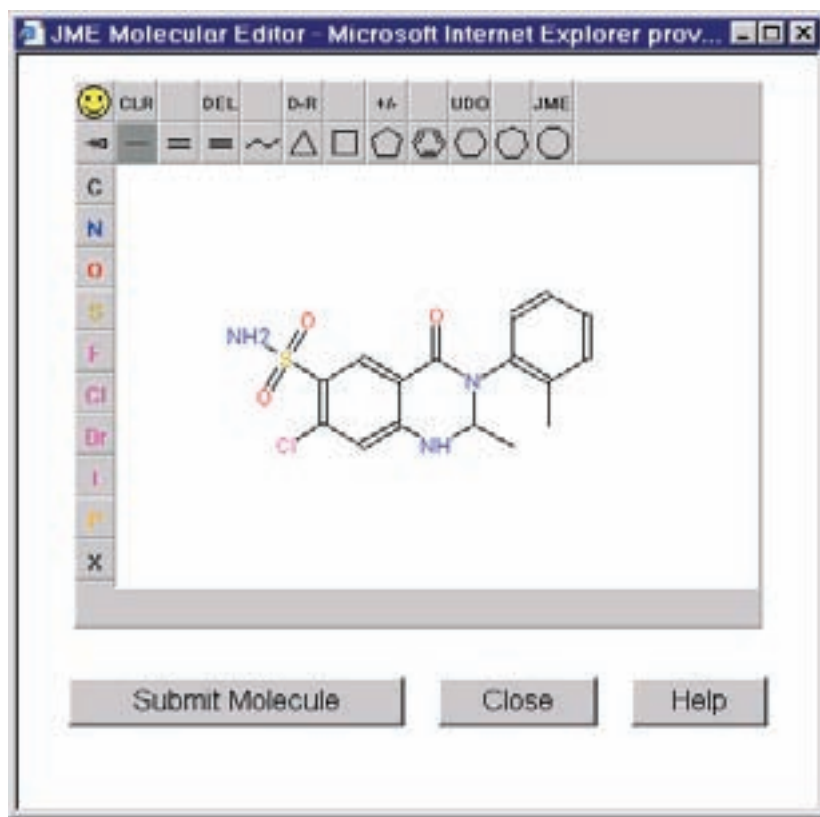


Figure 2
Example of web-based structure input. Molecule may be created or edited with help of molecular editor applet incorporated directly into the HTML page

- prediction of oral absorption, metabolic weak-points and alerts for toxicophores
- design of combinatorial libraries with optimal properties

These and similar tasks have been performed in the drug discovery process until now usually under the terms 'computational chemistry', 'chemometrics', or 'molecular modelling'. Recently a new, unifying term for all these activities has been coined, namely 'cheminformatics'^{1,2} (this term seems to be winning the race against the slightly longer 'chemoinformatics'). Several features, however, distinguish this new field from its predecessors. Probably the most important one is an enormous increase of data needed to be processed. While 10 years ago computational chemists processed datasets containing tens up to hundreds of molecules, current drug development requires property calculations or diversity analysis for hundreds of thousands, even millions of structures. This is mainly a consequence of high throughput screening and combinatorial chemistry techniques, used now routinely in the large pharma and agro companies. For further improvement of productivity, 'virtual libraries' have been introduced. Prior to the synthesis of compounds and in vitro high-

throughput screening, a 'virtual screening' is performed to select the most favoured molecules 'in silico' and reducing thus the amount of compounds to be synthesised and tested. Another imperative of modern cheminformatics is the shift to simplicity. Since a large number of molecules has to be processed, the methods used must be fast, for example the calculation of molecular properties is based usually on molecular topology only, without the necessity to perform time consuming geometry optimisation or conformational analysis. And, finally, cheminformatics is clearly moving from the hands of computational chemistry experts to the desktops of end-users – medicinal chemists.

Web-based cheminformatics

Several commercial tools to perform basic cheminformatics tasks are available on the market. These programs, however, usually run on UNIX workstations, and their usage requires a good knowledge of the special software. Therefore, these tools are used mainly by specialists – molecular modellers. Synthetic bench-chemists, however, should be involved much more in the direct cheminformatics work, calculation of molecular properties, similarity and diversity searches, design of new molecules for synthesis, since their project specific knowledge is crucial for all these tasks. Although most of them are quite interested in doing so, they are repelled by factors such as the necessity to remember UNIX commands, and to master the complicated interface and command set of commercial applications and other support programs

Recently, however, a possible solution to this dilemma has emerged – namely the World Wide Web. The enormous and still increasing popularity of web technology is due to its three great advantages – platform independence, high degree of interactivity, and ease of use. On a company network hosting various types of computers with different operating systems, the possibility of connecting all these machines in a user-friendly way is very important. Recently emerging new technologies such as Java, sophisticated web scripting, VRML (Virtual Reality Modelling Language), or chemical markup language³ have added new functionality to the web and made it a dynamic environment which is ideal for the development of user-friendly cheminformatics applications.

Leading commercial vendors are also going in this direction and are providing web interfaces to their packages, for example MSI with its WebLab suite of programs⁴ or Tripos with the ChemEnlighten database system⁵. At Novartis we have recognised the capabilities of the web

Informatics

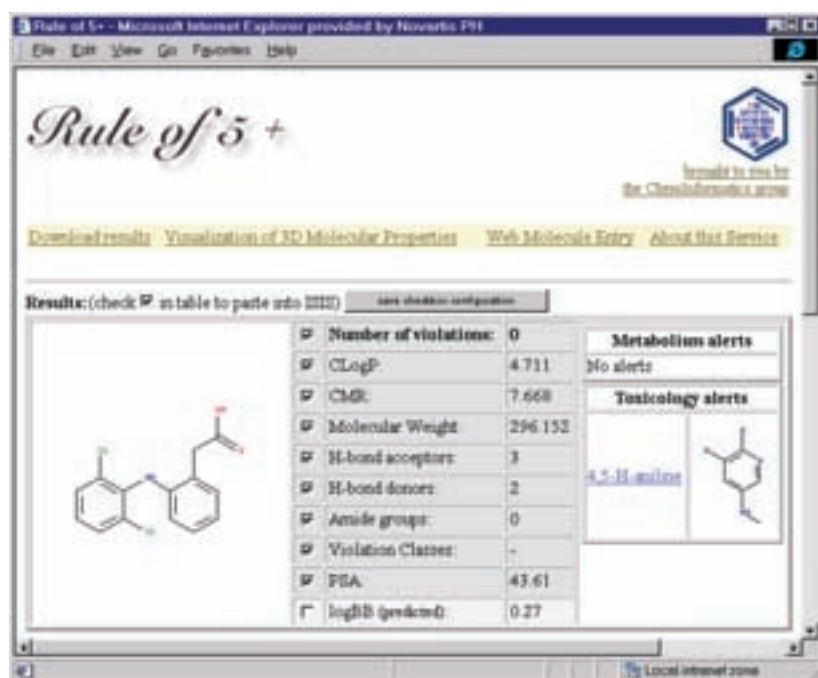


Figure 3

Display of calculated molecular properties, including octanol/water partition coefficient (logP), molar refractivity, number of hydrogen bond donors and acceptors and polar surface area. Potential metabolic and toxicity alerts are also included

technology relatively early and have been using it to deliver powerful and easy-to-use cheminformatics tools directly to the desks of synthetic organic chemists since 1995⁶. The Novartis Pharma web-based cheminformatics system, developed in-house and tailored specifically to intranet technology, supports most of the tasks which medicinal chemists need in the process of designing new drug candidates, namely:

- easy access to structural and activity data from the corporate database
- calculation of important hydrophobic, electronic and steric molecular properties
- calculation of drug-likeness and drug transport properties
- sophisticated visualisation of molecules and their surface properties
- interface to quantum chemical calculations and visualisation of results
- substituent bioisosteric searches⁷
- structure-activity (QSAR) analysis⁸
- design and generation of virtual libraries
- diversity selection and combinatorial library design⁹

Chemists access the system from their desktop PCs. All the 'heavy processing' is done on the Silicon Graphics servers. For better interactivity and specialised chemical tasks (eg molecular display and editing, interactive graphs) small Java graphics pro-

grams incorporated directly into the web page (applets) are used. In the system design special attention was paid to the construction of user interfaces, because this is actually the only part of the system the end-user sees and therefore crucially influences the acceptance of the whole product. The interface of a user-friendly system must be simple and self-explanatory (without the necessity to read extensive manuals). We have extensive experience of so-called 'single button' interfaces added into commonly used desktop applications, eg ISIS/Draw¹⁰ (Figure 1). A click on the button transfers the structure from ISIS/Draw to the server, starts the processing job and launches the web page with the result and links to further sources of information, which provide more details or refer to the scientific methodology used. This concept enables chemists to use cheminformatics tools directly from the environment they are most familiar with. Alternatively, it is also possible to use a 'pure' web solution with the in-house developed structure drawing applet for structure input and editing (Figure 2). For tools, which require the input of multiple structures (eg diversity tools), data may be transferred through the clipboard from Daylight-powered¹¹ in-house molecular database system¹² or using the SD file format, published by MDL¹⁰.

In the following paragraphs, three examples of end-user web-based cheminformatics tools on the Novartis intranet are described.

Interactive calculation of molecular properties

The estimation of drug-likeness plays an increasing role in the lead-optimisation process: the optimal candidate should show a good predicted oral absorption, certain metabolic stability and no structure-related toxicity. The 'Pfizer Rule of 5', introduced by Lipinski¹³, may give first hints, whether a compound will exhibit absorption problems.

We developed a web-based tool for the estimation of 'drug-likeness', which combines the calculation of molecular descriptors with model-based predictions of drug transport properties and hyperlinks to additional information resources. Calculation may be submitted directly from ISIS/Draw¹⁰ (Figure 1). A web browser is launched and the results are displayed within a few seconds (Figure 3). Based on the combination of the 'classical' Rule of 5 parameters and additional molecular properties such as the polar surface area, and substructure checks on the 'in silico' side, and on the pharmacokinetic data from the corporate database, an in-house model has been established for a complete 'drug-likeness' screening. Further links

are provided leading to examples for toxicology and metabolism and, for registered molecules, to databases with substance sheets and all available in-house data. Thus, the user has immediate access to archived information and predicted values related to the submitted structure by hitting just one button. A batch version of this tool allows calculation of drug transport properties for the whole virtual libraries.

Due to the nature of research, the predictions models mentioned above cannot be static. New scientific approaches and an increasing amount of experimental data offer the opportunity to improve the prediction quality of the models from time to time. This stresses the need for highly flexible and interactive tools that allow to react quickly upon new developments in computational and medicinal chemistry.

Visualisation of molecules and surface properties

This module allows chemists to generate interactively molecular images of various types. The 3D molecular structure used in this process is created from the SMILES string by the CORINA program¹⁴. The

generated 3D structure may be interactively manipulated by the Molecular Visualizer applet written in Java. After choosing a proper orientation, images are created by an in-house program and returned in a GIF format within 2-3 seconds. Images may be created in various modes (space filling, ball-and-stick, tube, dotted molecular surface, see Figure 4). Calculation and display of molecular surface properties such as electrostatic potential, lipophilicity potential or polar surface area is also possible, revealing the parts of the molecule which are involved in hydrophobic or electrostatic interactions, or may cause bioavailability problems.

This is one of the most simple modules of our system, but also one of the most popular, since generated images may be directly copied from the web page and pasted into reports or other documents.

Visualisation of structure diversity within a set of molecules

One of the most common questions chemists ask, when processing a set of molecules is: How diverse are these structures? There are, of course, many answers, depending on the definition of 'diversity'. The web tool used at Novartis provides interactive

References

- 1 Brown, FK. Cheminformatics: what is it and how does it impact drug discovery. *Annu. Rep. Med. Chem.* 33, 375-384, 1998.
- 2 Hann, M, and Green, R. Cheminformatics – a new name for an old problem. *Curr. Opin. Chem. Biol.* 3, 379-383, 1999.
- 3 Murray-Rust, P, and Rzepa, HS. Chemical Markup, XML, and the Worldwide Web. 1. Basic principles. *J. Chem. Inf. Comput. Sci.* 39, 928-942, 1999.
- 4 www.msi.com/weblab
- 5 www.tripos.com/software/cim.html
- 6 Ertl, P, and Jacob, O. WWW-based chemical information system. *J. Mol. Struct. (Theochem)* 419, 113-120, 1997.
- 7 Ertl, P. World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches. *J. Mol. Graph. Model.* 16, 11-13, 1998.
- 8 Ertl, P. QSAR analysis through the World-Wide Web. *Chimia* 52, 673-677, 1998.
- 9 Eichler, U, Ertl, P, Gobbi, A, and Rohde, B. Definition of an optimal subset of organic substituents. Interactive visual comparison of various selection algorithms. *Internet J. Chem.* 2, 1999, <http://www.ijc.com/14/paper14.html>

Continued on page 50

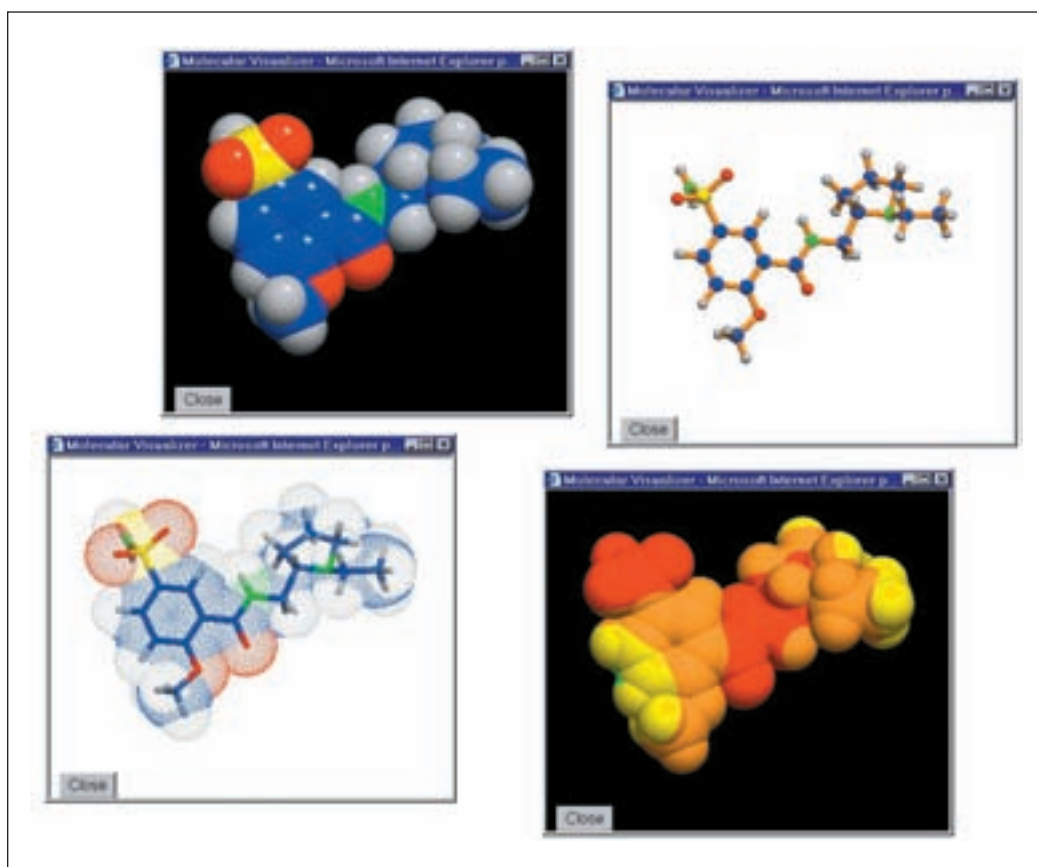


Figure 4

Display of molecular images in various modes. The advantage of space-filling images is the possibility to see the overall space requirements of the molecules. In tube models, on the other hand, the atomic connectivity is visible. The dotted surface with visible frame combines advantages of both these approaches

Informatics

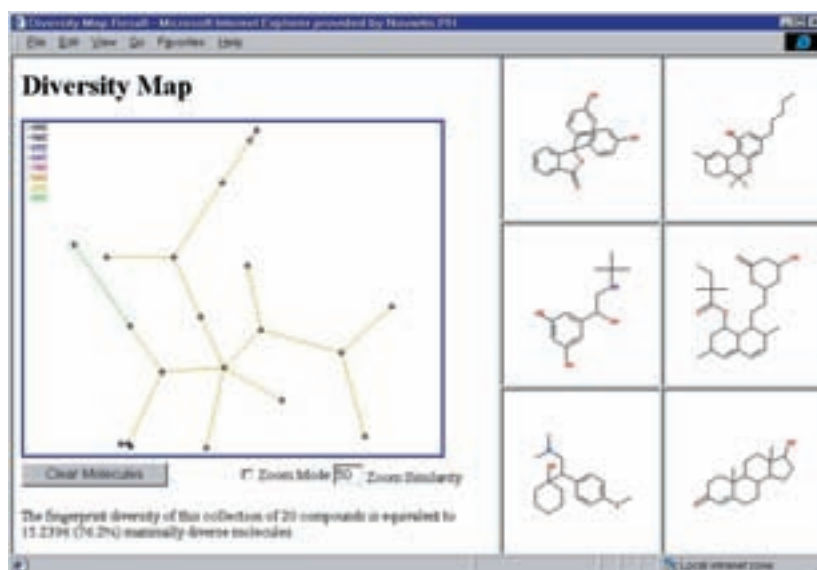


Figure 5
Example of interactive diversity map for a set of molecules. Colour of the line connecting two molecules represents their similarity. Molecules may be displayed after clicking particular graph nodes

visualisation of structural diversity (that one synthetic chemists are usually most interested in) calculated from the Daylight fingerprints¹¹. After pasting a set of molecules and launching the calculations, the program processes the molecules and displays them on a 2D map (Figure 5) with similar molecules connected together into clusters, while dissimilar molecules located far away from each other on the map. The map enables zooming and interactive display of processed structures. Such an interactive map is useful to get an idea about the overall diversity in the molecular set, about the number and type of clusters or about distribution of outliers.

Conclusions

The concept of easy-to-use web-based cheminformatics tools described in this article is highly accepted by the end users – Novartis chemists and biologists. Our in-house system is currently accessed worldwide by more than 400 users from research and development, and helps to ensure that the bench chemists are always supplied with state of the art calculation and prediction methods. The tools enable chemists really to ‘play’ with their molecules, to generate and analyse information useful in the research process. Due to the user-friendliness of web technology the system was introduced without any special training. Most of the computational modules used have been developed in-house, which assures easy maintenance and upgradeability. Other not negligible advantages are minimal licence costs (only for supporting programs) and no limitations concerning the number of users.

Another important goal, which web-based cheminformatics tools contribute to, is support for the process of excluding compounds with low drug potential from further research and development at an early stage. This helps to focus resources on more promising candidates and to increase the researcher’s efficiency, and as a result to design better drugs faster.

DDW

Peter Ertl studied organic chemistry and received his PhD at the University of Bratislava. After eight years of academic research, developing and applying methods of computational chemistry, molecular graphics and QSAR, he switched to the industry and joined the Ciba-Geigy AG in Basel in 1992. After the creation of Novartis he became an IT Project Manager in the ChemInformatics group of Pharma Research, responsible for the development of new methods for the calculation of drug properties and web-based cheminformatics tools.

Wolfgang Miltz studied organic chemistry and received his PhD at the University of Bonn, Germany, in 1988. After a postdoctoral fellowship at the University of Virginia at Charlottesville, VA, USA, he joined the Preclinical Research at Sandoz Pharma AG, Basel in 1989 as a medicinal chemist. With experience in numerous different research programmes for cardiovascular diseases, immunology and inflammation, he is now in the therapeutic area for Arthritis and Bone Metabolism at Novartis Pharma AG in Basel. His special interest is the development of predictive tools for the improvement of drug-likeness.

Bernhard Rohde received his PhD in Organic Chemistry from the University of Zurich, Switzerland in 1987. He joined Ciba-Geigy AG working on computer-assisted synthesis planning, support for combinatorial chemistry and computer-assisted research in general. After the merger of Ciba with Sandoz he took a position in Pharma Research Information Management, where he was responsible for cheminformatics. He is now concentrating on the design of a comprehensive cheminformatics data warehouse.

Paul Selzer studied organic chemistry and gained his PhD from the Computer-Chemistry-Center at the University of Erlangen. After a postdoctoral fellowship in the Research Information Management department at the Novartis Pharma AG in Basel, he is now member of the Novartis Pharma ChemInformatics group developing tools for intranet QSAR applications and data warehousing.

Continued from page 49

- 10 www.mdli.com
- 11 www.daylight.com
- 12 Gobbi, A, Poppinger, D, and Rohde, B. Developing an in-house system to support combinatorial chemistry. *Perspect. Drug Discovery Des.* 131-158, 1997.
- 13 Lipinski, CA, Lombardo, F, Dominy, BV, and Feeney, PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* 23, 3-25, 1997.
- 14 CORINA 3D structure generator is available from www.mol-net.de/products/corina