ELSEVIER

World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches

Peter Ertl

Agro Research Computing Group, Novartis Crop Protection AG, Basel, Switzerland

Easy to use, interactive, and platform-independent WWW-based tools are ideal for development of chemical applications. By using the newly emerging Web technologies such as Java applets and sophisticated scripting, it is possible to deliver powerful molecular processing capabilities directly to the desk of synthetic organic chemists. In Novartis Crop Protection in Basel, a Web-based molecular modelling system has been in use since 1995. In this article two new modules of this system are presented: a program for interactive calculation of important hydrophobic, electronic, and steric properties of organic substituents, and a module for substituent similarity searches enabling the identification of bioisosteric functional groups. Various possible applications of calculated substituent parameters are also discussed, including automatic design of molecules with the desired properties and creation of targeted virtual combinatorial libraries. © 1998 by Elsevier Science Inc.

Keywords: substituent parameters, substituent constants, bioisosterism, QSAR, World Wide Web

INTRODUCTION

The Hansch–Fujita correlation analysis proved to be one of the major breakthroughs in the process of understanding the relation between properties of molecules and their biological activity.[1,2] In this approach electronic, hydrophobic, and steric substituent parameters are used to quantify the properties of molecules under study. Currently, when performing QSAR analysis, substituent constants are usually manually extracted from various data tables. This approach, however, has numerous disadvantages. Actually looking up constants, especially for large data sets, may be quite tedious and it is not easy to avoid errors caused by manual inputting of large columns of numbers. The quality of data for uncommon substituents is not good, which requires laborious checking in the original literature. And the most serious drawback of experimentally determined substituent constants is the fact that the data for many important substituents, despite the comprehensive tables,[3] are simply not available. It would be therefore useful to have a system for automatic estimation of substituent parameters.

In this article a program is presented that enables fast and easy calculation of important substituent parameters and their processing by using World Wide Web technology.

The capabilities of the World Wide Web in providing textual and graphic information are well known. Various newly emerging techniques, such as sophisticated scripting and especially the possibility of adding short Java programs—applets[4]—directly into the HTML page added another functionality to the Web and changed it from a static to a fully dynamic environment. Easy-to-use, interactive, and platform-independent WWW-based tools became in this way ideal for the development of chemical applications.[5,6]

WWW-BASED MOLECULAR MODELING SYSTEM

At Novartis Crop Protection AG in Basel we have been using Web technology to deliver powerful and easy-to-use modeling tools directly to the desk of synthetic organic chemists since 1995. A Web-based chemical information and molecular modeling system,[7] developed in-house and used currently by about 200 users, enables

- Easy retrieval of molecules and corresponding data from the corporate database
- Creation and editing of molecules by using a molecular editor written in Java
- Sophisticated molecular and 3-D property visualization
- Automatic generation of hydrophobic, electronic, and steric molecular descriptors
- Interfacing of quantum chemical calculations and visualization of results

The Color Plate for this article is on page 36.

Address reprint requests to: P. Ertl, Agro Research Computing Group, Novartis Crop Protection AG, CH-4002 Basel, Switzerland.

Received 19 February 1998; revised 27 April 1998; accepted 5 May 1998.

655 Avenue of the Americas, New York, NY 10010

1093-3263/98/$19.00
PII S1093-3263(98)00012-6

- Interactive QSAR analysis
- Molecular similarity searches

In this article two new modules are presented: a program for interactive calculation of important substituent parameters, and a module for substituent similarity searches, enabling a set of bioisosteric functional groups to be found.

Calculation of substituent parameters

The importance of substituent parameters in characterizing molecular properties was discussed in the Introduction. To enable bench chemists access to such data in an easy way, a Web-based module for this purpose has been developed.

This module enables interactive calculation of basic substituent constants for any organic functional group. Properties calculated currently include hydrophobic, electronic, and steric properties. Hydrophobic properties are represented by octanol–water partition coefficient (log*P*) and molar refractivity. Both these parameters are calculated according to the well-known methodology of Ghose and Crippen,[8,9] based on the sum of atomic hydrophobicity contributions. The electron-donating and -withdrawing power of substituents is characterized by theoretical parameters compatible with the Hammett σ constants. These are calculated according to the methodology developed in-house[10] from simple quantum chemical data. The agreement of calculated σ constants with experimental is good (r^2 = 0.89 for σ_{meta} and 0.93 for σ_{para}, for 63 of the most common organic substituents). Steric properties of substituents are represented simply by their topological size (number of nonhydrogen atoms) and maximal topological length. In our experience these parameters are sufficient to characterize steric requirements of substituents. The addition of more sophisticated parameters (i.e., STERIMOL), however, would be straightforward.

Users interact with the system through a simple Web interface. In the entry part of the program (Color Plate 1) the substituent for which data should be calculated is created with the help of our molecular editor, written in Java.* After pressing the [Calculate Parameters] button a complex chain of processes is launched. The editor creates a SMILES[11] code for the substituent, which is passed to the CORINA[12] 3-D geometry builder. A 1SCF AM1 calculation[13] is then run, using the standard Mopac93 package. Other in-house programs extract atomic charges from the Mopac output, calculate the σ constants from them, and estimate other substituent properties. Despite this complex "behind the scenes" processing, the response is fast and data are delivered within 2–3 s (Color Plate 2).

The processing engine behind the program may also be called directly (without the graphic interface) just by referencing to the http address of the cgi-bin script with substituent's SMILES as a parameter. In this way it is possible to calculate data for a large number of substituents in a "batch" mode. By using this technique, data for more than 80 000 functional groups used in substituent similarity searches have been generated.

*Our Java Molecular Editor applet enables easy construction and modification of organic molecules direct within the HTML page and generation of SMILES of the processed molecule. Editor is freely available for noncommercial use. Interested parties should contact the author at peter.ertl@cp.novartis.com

Substituent similarity searches

During the process of designing new drugs or agrochemicals the concept of bioisosterism[1,2] is often used. Functional groups are considered to be bioisosteric when they have similar physicochemical properties and therefore induce the same or similar biological response.

Synthetic organic chemists are trained to think mainly in terms of electronic similarity of substituents. The successful binding to the substrate, however, is determined not only by electrostatic interactions but also by hydrophobic forces, the steric requirements of ligands, and by hydrogen bonding. Even the most experienced chemist (or modeler) cannot identify substituents with an optimal balance of electronic, steric, and hydrophobic properties just by comparing the 2-D structural formulas.

It seems, therefore, to be highly desirable to have an easy-to-use tool that would provide automatic selection of functional groups that are similar in their important physicochemical properties. Therefore a module for the interactive calculation of substituent similarity has been developed.

Using this program substituents physicochemically compatible with the target may be identified and retrieved from a database of more than 80 000 organic functional groups with precalculated properties. As input (Color Plate 3) a substituent is drawn, search criteria are chosen (from electronic, hydrophobic, steric, and H-bond capabilities) and the search is then started. The similarity of two substituents is calculated simply as a Euclidean distance of two points in the space of (properly scaled) physicochemical properties. Properties considered in searching are those described in the previous section, with the addition of hydrogen-bonding capabilities of the substituent. These are generated simply from the molecular topology, on the basis of the well-known definition of H-bond donor and acceptor sites.[14] As output (Color Plate 4) a set of bioisosteric functional groups is provided.

To illustrate the Web technology behind this module a schematic outline of the program flow is shown in Figure 1. Users interact with the program through the HTML page with

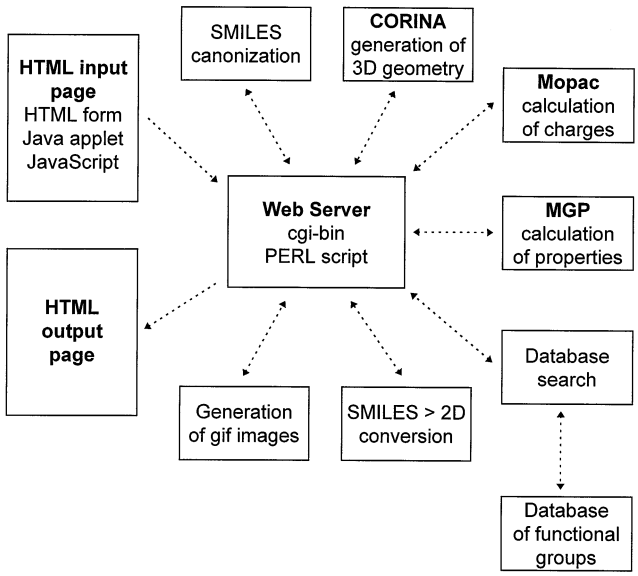

Figure 1. Schematic outline of the program flow.

an incorporated molecular editor applet and a set of menu buttons. After drawing the target substituent and choosing the proper options for searching, the [Start Search] button is pressed. JavaScript included in the HTML gets the SMILES from the editor and sends it together with the search options to the server. The PERL cgi-bin script running there then takes responsibility for further processing and coordination of various support programs. The substituent's SMILES is first canonized (creation of unique SMILES). CORINA then generates the 3-D geometry and Mopac calculates the atomic charges. An in-house program, MGP, calculates the properties of the target substituent. A database search is then started that results in a set of SMILES's of similar substituents. They are converted into 2-D chemical formulas, out of which the gif images are generated. Finally, an HTML page with the answer is generated on the fly. Practically all programs used in this module have been developed in-house, mostly in Java and C. Mopac93 and CORINA are the only commercial programs used. The whole system runs on a Silicon Graphics Origin 200 server and typical response time for one search is 8 s.

CONCLUSIONS

The Web-based programs described here allow the easy and convenient calculation of important hydrophobic, electronic, and steric properties of organic functional groups, as well as the identification of bioisosteric substituents.

Substituent parameters, despite their "classic" use in QSAR studies, have many other application possibilities. One promising application area is combinatorial chemistry. It is well known that the proper design of combinatorial libraries (i.e., selection of optimal building blocks) can increase considerably the chances of finding new lead structures. The successful use of a Web interface to select building blocks for array synthesis has been reported.[15] A large database of calculated substituent parameters with nearly 90 000 entries provides a good foundation for such effort. By selecting substituents with the desired properties it is possible to plan libraries that regularly cover particular parts of physicochemical space in the search for leads in yet unexplored areas, or focus on the complete coverage of a known "island of activity."

Another interesting application field for a database with calculated substituent parameters is the automatic design of molecules with desired physicochemical properties, either as a mimic of a known lead, or based on the QSAR model. Such a procedure using genetic algorithm optimization is under development in our group.

REFERENCES

1 Hansch, C. and Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. ACS, Washington, D.C., 1995

2 Hansch, C. and Fujita, T. (eds.). *Classical and Three-Dimensional QSAR in Agrochemistry*. ACS, Washington, D.C., 1995, pp. 13–44

3 Hansch, C., Leo, A., and Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*. ACS, Washington, D.C., 1995

4 Sun Microsystems. *What Is Java*? http://java.sun.com, 1995

5 Taylor, N.R. and Smith, R. The World Wide Web as a graphical user interface to program macros for molecular graphics, molecular modelling and structure based drug design. *J. Mol. Graphics* 1996, **14,** 291–296

6 Bachrach, S.M., Murray-Rust, P., Rzepa, H.S., and Whitaker, B.J. *Publishing Chemistry on the Internet*. Network Science, http://www.awod.com/netsci/Issues/Mar96/feature4.html. 1996

7 Ertl, P. and Jacob, O. WWW-based chemical information system. THEOCHEM 1997, **419,** 113–120; see also http://www.elsevier.com/homepage/saa/eccc3/paper6

8 Ghose, A.K., Pritchett, A., and Crippen, G.M. Atomic physicochemical parameters for three dimensional structure directed quantitative structure–activity relationships. III. *J. Comput. Chem*. 1988, **9,** 80–90

9 Viswanadhan, V.M., Ghose, A.K., Revankar, G.R., and Robins, R.K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure–activity relationships. IV. *J. Chem. Inf. Comput. Sci*. 1989, **29,** 163–172

10 Ertl, P. Simple quantum chemical parameters as an alternative to the Hammett sigma constants in QSAR studies. *Quant. Struct.–Act. Relat*. 1997, **16,** 377–382

11 Weininger, D. SMILES, a chemical language and information system. *J. Chem. Inf. Comput. Sci*. 1988, **28,** 31–36

12 Sadowski, J. and Gasteiger, J. From atoms and bonds to three-dimensional atomic coordinates: Automatic model builders. *Chem. Rev*. 1993, **93,** 2567–2581

13 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. 1985, **107,** 3902–3909

14 Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I., and Pavlik, P.A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided. Mol. Design* 1993, **7,** 83–102

15 Leach, A.R. Structure-based selection of building blocks for array synthesis via the World-Wide Web. *J. Mol. Graphics Modelling*. 1997, **15,** 158–160