JOURNAL OF CHEMICAL EDUCATION

Activity

pubs.acs.org/jchemeduc

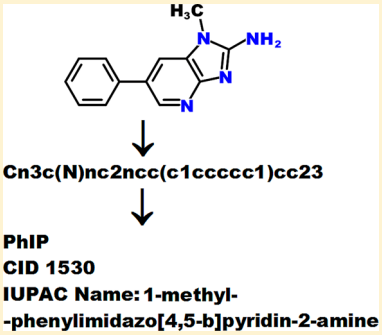
Using Internet Databases for Food Science Organic Chemistry Students To Discover Chemical Compound Information

Piotr Minkiewicz,* Anna Iwaniak, and Małgorzata Darewicz

University of Warmia and Mazury in Olsztyn, Chair of Food Biochemistry, Plac Cieszyński 1, 10-726 Olsztyn-Kortowo, Poland

Supporting Information

ABSTRACT: Internet databases serve as an important source of information on chemical compounds that students can readily investigate, including those studying food science and food technology. This Activity provides a brief introduction to the application of chemical information resources with a focus on conducting structure-based searches.



KEYWORDS: *Agricultural Chemistry, Bioorganic Chemistry, Chemoinformatics, Computer-Based Learning, Consumer Chemistry, First-Year Undergraduate, Interdisciplinary/Multidisciplinary, Internet/Web-Based Learning, Organic Chemistry*

Courses in food technology, human nutrition, and other areas of study offered by faculties of food science and related fields utilize information about low-molecular-weight organic compounds, in particular, bioactive food components. Internet resources provide quick access to the relevant information. Unlike in conventional handbooks, the information found on the Internet can be continuously updated, which is a major advantage. However, a Web search by compound name has certain limitations, in particular, for non-native English speakers. The number of compounds described in English is greater than in other languages. Molecular structures are the universal language of chemistry that are used in, for example, chemical databases. Molecular structures can be drawn in computer applications known as "molecular editors". State-of-the-art compound structures developed in molecular editors have been reviewed by Ertl.[1]

Specialized databases are the most reliable sources of information about chemical compounds. Their emerging role in food and nutrition science has been emphasized in several reviews.[2−6]

This paper proposes the simplest possible approach to using chemoinformatics databases in the process of educating food technologists in view of the limited time allocated to organic chemistry courses. Databases such as PubChem[7,8] are presented as a source of information on the properties of the searched compounds. First-year students will learn to use chemoinformatics databases during a basic course in organic chemistry. The relevant exercise is detailed in this paper.

BRIEF DESCRIPTION OF THE EXERCISE

During the exercise, students will draw molecular structures in the molecular editor on the Chemical Structure Lookup Service Web site,[9,10] and they will translate the structures into SMILES.[11] Molecular structures translated into SMILES will be used as queries in the PubChem database[7,8] by running a structure similarity search. By default, compounds identical to the query will be searched. Alternative applications include the molecular editor in the ChemMapper program[12,13] and the ChemSpider database.[14,15] All of the above tools are available on the MetaComBio Web site;[16] they are freely accessible and do not require registration. Before the exercise, the instructor delivers a presentation and provides students with general information about databases, SMILES language, and structure search. The exercise includes supervised training in compound structure development and the use different descriptors for database screening.

The report summarizes the information on the physicochemical properties and biological activity of compounds, translated from English (in our case, into Polish). An example of classroom workflow is available in the supplement. The organic chemistry course at the Faculty of Food Sciences of the University of Warmia and Mazury in Olsztyn covers 30 h of lectures and 30 h of tutorials. Students will have completed a chemistry course in secondary school and a general chemistry course (30 h of lectures and 30 h of tutorials) at university. Tutorials concerning chemical information last 2 h, including database presentation and training supervised by an instructor. Students develop final reports outside class. The exercise was first introduced in the 2013/2014 academic year. In autumn 2014, the exercise was expanded by allowing students to search

DOI: 10.1021/ed5006739
J. Chem. Educ. XXXX, XXX, XXX−XXX

for compound information using any query (e.g., SMILES, InChI, or InChiKey) and any database.

DISCUSSION

We would like to discuss some issues that are related to the proposed exercise and affect its application for educational purposes.

Brief Explanation Of The Workflow Applied

The proposed activity contains intentionally inserted problems that need to be solved by the students. Both PubChem and ChemSpider databases have their own molecular editors. Those editors and the Chemical Structure Lookup search engine are not used in the exercise. Students are asked to develop the compound structure and translate it into SMILES with the use of one chemoinformatics tool (Chemical Structure Lookup Service). They will use another tool (PubChem) to obtain information about compound properties. This search scheme demonstrates that a compound's chemical structure may be used to screen multiple databases. Students may also learn that the Chemical Structure Lookup Service can be used to screen multiple resources via a common search engine. The discussed tool provides a wealth of information that cannot be fully interpreted during the described exercise (due to short time), but it may be recommended for future application. Flexible search options are required to make the most effective use of more than 300 open-access databases and other tools supporting the identification and description of low-molecular-weight compounds.[16] SMILES is the most commonly used machine-readable chemical code, and its rules are well described. PubChem and ChemSpider contain information about chemical compounds' systematic and common names, SMILES, Chemical Abstract Service Registry, InChi (IUPAC International Chemical Identifiers), as well as InChi Key. The latter one is recommended for Google searches.[17] The above identifiers constitute a "query set" for screening specialized chemical and biological databases containing information about chemical compounds.

Web Site Facilitating the Access to Databases

Links to other databases of low-molecular-weight compounds are available on the MetaComBio Web site.[16] MetaComBio tools can be used during several courses in food technology or relevant fields of study. For instance, databases of biochemical reactions may be recommended for courses in biochemistry, enzymology, or biotechnology, databases of metabolic pathways—for courses in biochemistry or nutrition, and databases of toxic compounds—for courses in toxicology. There are also specialized databases of food components. The information found in various database categories should be sufficient for writing a diploma thesis. It is still too early to evaluate the significance of chemical databases for other courses. PubChem citations are found in laboratory reports written by students enrolled in chemistry or organic chemistry courses.

Student's Experience with Exercise

Conflict management is a problem that is often reported by students. Molecular editors require the most recent version of Java. Problems with Java-based molecular editors were observed between March and December 2014. Java blocks windows in selected molecular editors (excluding PubChem and ChemSpider), and security extensions should be added in Java configuration options.

Observations indicate that computer exercises are relatively easy for the students. Students performing the exercise received higher grades (average grade of 4.19 for 119 reports submitted in the spring semester of the 2013/2014 academic year, on a scale of 3 (satisfactory) to 5 (excellent) for reports with a passing grade) in comparison with other exercises (average grade of 3.72 for 476 reports). Those results are not surprising because contemporary students belong to a generation that develops computer skills in childhood. It should be noted that most students used basic primitives to generate compound structures for the first time. Compound structures developed from those elements are commonly used in specialized databases and other Internet resources[18] but not in handbooks for secondary schools.

Students encountered certain problems when describing the biological activity of compounds in Polish. The above can probably be attributed to the fact that chemical and biological terms are not taught during English language courses in primary and secondary school.

Final Remarks

In the described exercise, the use of chemoinformatics and bioinformatics tools is presented in the simplest possible way (with some intentionally inserted problems that need to be solved by the students). Research into the biological activity of food components involves more advanced in silico methods.[19] Chemical space screening[19] and modeling protein interactions require advanced skills. Sophisticated chemoinformatic methods are introduced during chemistry courses.[20] It should be noted, however, that most students of food technology and related areas will not need advanced algorithms that are designed for chemists or biologists. More advanced methods may be introduced as part of special, optional courses at postgraduate level. The proposed exercise is addressed to future users of chemoinformatics and bioinformatics tools at the simplest level, including students with limited time for chemistry courses. Students of food technology and other fields of food sciences, such as gastronomy, culinary art, and commodities, can take part in such exercises to expand their knowledge of Internet resources concerning organic compounds and the relevant data search options.

Students learn about databases of chemical compounds, such as PubChem, and the relevant data search options. It should be noted that search options may be used without previous knowledge of specialist chemical terminology in English.

ASSOCIATED CONTENT

Supporting Information

The supplement contains an example of classroom workflow, a summary of the instructor's presentation, trouble management protocol, and a brief description of the MetaComBio Web site. This material is available via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: minkiew@uwm.edu.pl.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been financed by the University of Warmia and Mazury in Olsztyn.

DOI: 10.1021/ed5006739
J. Chem. Educ. XXXX, XXX, XXX−XXX

REFERENCES

(1) Ertl, P. Molecular Structure Input on the Web. *J. Cheminform.* **2010**, *2* (1), No. 10.1186/1758-2946-2-1, DOI: 10.1186/1758-2946-2-1.

(2) Martínez-Mayorga, K., Peppard, T. L., Medina-Franco, J. L., Software and Online Resources: Perspectives and Potential Applications. In *Foodinformatics. Applications of chemical information to food chemistry*; Martínez-Mayorga, K., Medina-Franco, J. L., Eds.; Springer International Publishing AG: Cham, Switzerland, 2014; pp 233−248.

(3) Scalbert, A.; Andres-Lacueva, C.; Arita, M.; Kroon, P.; Manach, C.; Urpi-Sarda, M.; Wishart, D. Databases on Food Phytochemicals and their Health-Promoting Effects. *J. Agric. Food Chem.* **2011**, *59*, 4331−4348.

(4) Malkaram, S. A.; Hassan, Y. I.; Zempleni, J. Online Tools for Bioinformatics Analyses in Nutrition Sciences. *Adv. Nutr.* **2012**, *3*, 654−665.

(5) Minkiewicz, P.; Miciński, J.; Darewicz, M.; Bucholska, J. Biological and Chemical Databases for Research into the Composition of Animal Source Foods. *Food Rev. Int.* **2013**, *29*, 321−351.

(6) Holton, T. A.; Vijayakumar, V.; Khaldi, N. Bioinformatics: Current Perspectives and Future Directions for Food and Nutritional Research Facilitated by a Food-Wiki Database. *Trends Food Sci. Technol.* **2013**, *34*, 5−17.

(7) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217−224.

(8) PubChem website. http://PubChem.ncbi.nlm.nih.gov/ (accessed Mar−Dec 2014).

(9) Sitzmann, M.; Filippov, I. V.; Nicklaus, M. C. Internet Resources Integrating Many Small Molecular Databases. *SAR QSAR Environ. Res.* **2008**, *19*, 1−9.

(10) Chemical Structure Lookup Service website. http://cactus.nci.nih.gov/cgi-bin/lookup/search (accessed Mar−Dec 2014).

(11) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(12) Gong, J.; Cai, C.; Liu, X.; Ku, X.; Jiang, H.; Gao, D.; Li, H. ChemMapper: a Versatile Web Server for Exploring Pharmacology and Chemical Structure Association Based on Molecular 3D Similarity Method. *Bioinformatics* **2013**, *29*, 1827−1829.

(13) ChemMapper website. http://lilab.ecust.edu.cn/chemmapper/index.html (accessed Mar−May 2014).

(14) Pence, H. E.; Williams, A. Chemspider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87*, 1123−1124.

(15) ChemSpider website. http://www.chemspider.com/ (accessed Mar−Dec 2014).

(16) MetaComBio website. http://www.uwm.edu.pl/metachemibio/index.php/about-metacombio (accessed Mar−Dec 2014).

(17) Southan, C. 2013, InChI in the Wild: an Assessment of InChIKey Searching in Google. *J. Cheminform.* **2013**, *5* (10), No. 10.1186/1758-2946-5-10, DOI: 10.1186/1758-2946-5-10.

(18) Clark, A. M. Basic Primitives for Molecular Diagram Sketching. *J. Cheminform.* **2010**, *2* (8), No. 10.1186/1758-2946-2-8, DOI: 10.1186/1758-2946-2-8.

(19) Martínez-Mayorga, K.; Peppard, T. L.; López-Vallejo, F.; Yongye, A. B.; Medina-Franco, J. L. Systematic Mining of Generally Recognized as Safe (GRAS) Flavor Chemicals for Bioactive Compounds. *J. Agric. Food Chem.* **2013**, *61*, 7507−7514.

(20) Price, G. W.; Gould, P. S.; Marsh, A. Use of Freely Available and Open Source Tools for in Silico Screening in Chemical Biology. *J. Chem. Educ.* **2014**, *91*, 602−604.

DOI: 10.1021/ed5006739
J. Chem. Educ. XXXX, XXX, XXX−XXX