ACD Labs/LogP dB 3.5 and ChemSketch 3.5

Gary O. Spessard

Department of Chemistry, St. Olaf College, Northfield, Minnesota 55057

Received August 2, 1998

ChemSketch 3.5 and ACD/LogP dB 3.5 are products of Advanced Chemistry Development, Inc. of Toronto, Canada. A full-featured chemical structure drawing program, ChemSketch 3.5 serves both as the entre into and the integrator among several other modular programs such as ACD/LogP dB 3.5. Other ACD modules include programs for calculation of ^{13}C- and ^{1}H-NMR spectra, pK_as, solubilities, boiling points, and vapor pressures as well as software for generation of IUPAC names. As the name implies, ACD/logP dB 3.5 is a program for calculation of octanol/water partition coefficients (log P_{oct}) of compounds. Such values indicate the lipophilicity or fat-seeking ability of molecules, and they are of enormous importance to pharmaceutical and environmental chemists who attempt to correlate molecular properties with biological activity. The measurement of these values, however, can be tedious, and an easy-to-use and accurate program for rapid estimation of log *P*s would be a useful tool.

ACD Labs/LogP dB 3.5 and ChemSketch 3.5 run only on IBM PC-compatible computers operating under Windows 3.x or Windows 95/98. A Macintosh version of this software is not currently available. Minimum system requirements include an 80486 processor (Pentium recommended), VGA color monitor, 16 MB of RAM, at least 20 MB of free hard disk space, CD ROM drive, and a mouse. The current version of the software is 32-bit compatible, but actually 16 bit, which means that file names longer than eight characters are not acceptable. Installation of the software ran off the CD-ROM drive and was straightforward for the most part under Windows 95. I encountered some difficulty, however, in registering ACD/ChemSketch and LogP dB, since there are several long, alphanumeric serial numbers (provided by the vendor in separate documents called "Certificates of Authenticity") associated with various parts of the software, and it was not clear from the instructions which numbers should be entered.

ChemSketch 3.5. The number of features and tools included in ChemSketch 3.5 is extensive and comparable to some of the most popular commercial structure drawing programs such as ChemDraw, ChemWindow/Chemintosh, and ISIS Draw. Despite this complexity, first-time users should have little trouble drawing attractive structures with relatively little assistance. There is a full tutorial included in the documentation that is helpful to novice users, but those already acquainted with any one of the drawing programs mentioned above will probably not need to go through the entire tutorial.

The user interface of ChemSketch 3.5 differs significantly from ChemDraw and related programs, yet I found structure input intuitive and user-friendly. There are two different screens for drawing input: a "Structure" screen and another called "Draw". Users draw molecules only in the Structure screen, which has menu and button bars on the top of the drawing area. From the button bar users may choose an array of bond types and also do structure editing by selecting from several options available. Left-hand buttons consist of chemical element symbols that allow the user to select any element in the periodic table as the default atom in a structure. On the right-hand side of the drawing area are buttons to select substructures such as rings, chains of various length, common group labels such as COOH, and labels for amino acids. Once the structure is drawn, users may select the entire structure (by lasso, for example) and do operations such as resize, rotation, and translation. Users may also calculate various physical properties of the selected structure such as molecular weight, molecular formula, molar refractivity, and molar volume, to name a few. Once the structure is selected, labels can be adjusted for font type, size, and color. ChemSketch 3.5 also contains a 3D optimization feature that will take a 2D structure and find a minimum-energy geometry, which is then displayed on the screen in perspective, i.e., a flat structure for cyclohexane becomes a chair conformation. Accompanying the drawing screen is a 3D molecular modeling module. Users can import a selected structure into this module and render it as stick, ball-and-stick, space-filling, wire mesh, and space filling, etc. 3D optimization in this module, employing a molecular mechanics-based algorithm, allows users to calculate bond lengths and bond angles.

ChemSketch 3.5 comes equipped with a "Dictionary", which is a searchable database of ca. 48 000 compounds. Users may select any compound from the database and import it into the Structure screen on the fly. Several template files are available, providing an array of carbohydrate, amino acid, arene, and ring structures as well as orbital representations and reaction symbols. Users may produce their own template files for future use or create templates on the fly when performing repetitive drawing operations. Drawings may also be imported in several formats including Windows metafiles, MDL MOL (.mol), ISIS SKETCH bin (.skc), and ChemDraw (.chm) and exported as Windows meta and bitmap files, MDL MOL (.mol), ISIS SKETCH (.skc), ChemDraw (.chm), and GIF files.

The Draw screen of ChemSketch 3.5 appears at the press of a menu button, and it contains the same information as the Structure screen. Once in the Draw screen mode, users may edit their drawing by putting in text, reaction and pointer arrows, curves, and boxes. Drawing Bezier curves is quite easy, and these may be edited to virtually any shape. Editing the fill characteristics of shapes for texture and color is about as easy to do as I have seen among the several commercial drawing programs familiar to me. Access to the 3D modeling module is also possible via the Draw screen.

I noted a few characteristics of the ChemSketch 3.5 that I found to be a bit cumbersome, such as having a separate Draw screen for editing instead of just one screen for

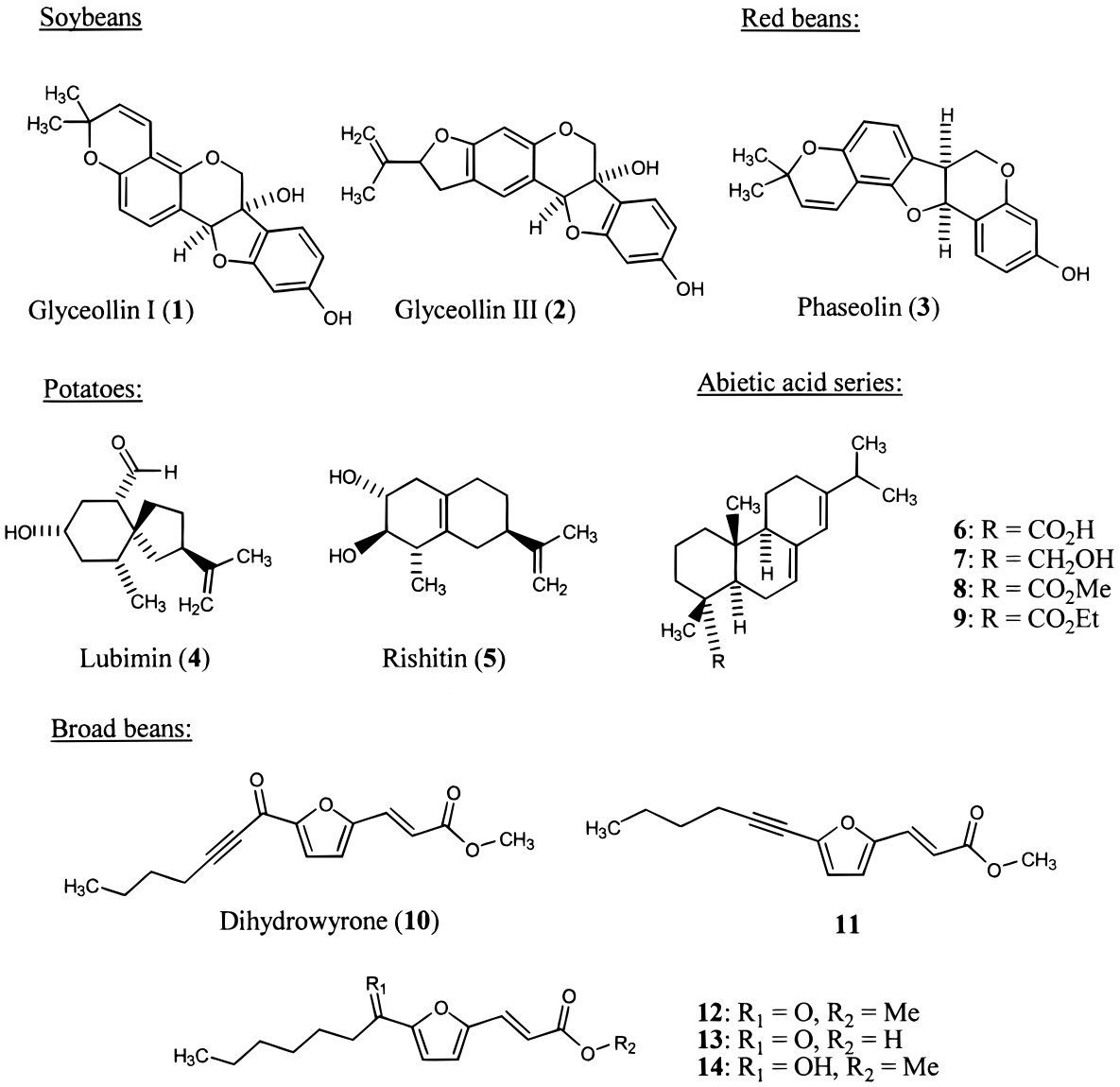

Figure 1. Phytoalexins and analogs.

composing an entire drawing. Changing font size and type for chemical labels is not as easy to do on the fly as it is with other commercial drawing software. Oddly in the Structure screen, linear molecules, which contain all possible C−H bonds explicitly shown after conversion to 3D structures, turn into Fischer projections when the 2D cleanup option is applied; these structures do not appear anything like what was previously drawn. In contrast to my overall favorable impression of ChemSketch 3.5 as a serious contender in the competition among current, commercial structure drawing programs, however, these quibbles are relatively minor. At the time of this writing, perhaps the best feature of ChemSketch 3.5 is its availability free of charge by downloading from the website of ACD Laboratories (www.acdlabs.com).

ACDLabs/LogP dB 3.5. Once a structure has been drawn in ChemSketch, it is possible to calculate log *P* values of almost any organic compound simply by loading ACDLabs/LogP dB 3.5 from the *ACD/Labs* option on the top-of-screen menu bar of ChemSketch 3.5. A **Calc LogP** button appears on the bottom of the screen, which, when pressed, places the user into the calculation mode for determining log *P*. (Users may switch readily from the LogP mode to the ChemSketch mode or back again at any time simply by pressing the **ChemSk** or **Calc LogP** button, respectively, on the bottom of the screen.) A dialogue box appears prompting the user to enter the melting point of the compound drawn if known (it is not necessary to enter a melting point value for calculation of log *P*). Simply pressing the **OK** button starts the calculation, and the result appears on a new screen. I found that all the operations required for calculating log *P* values worked straightforwardly and smoothly.

ACDLabs/LogP db 3.5 uses what the vendor calls an "additive-constitutive" algorithm somewhat similar to the approaches of Hansch and Rykker to calculate log *P* values for almost any compound drawn in ChemSketch. The calculation engine adds up contributions from separate atoms, structural fragments, and intramolecular interactions between fragments, all derived from a database of over 5000 experimentally-obtained log *P* values of over 3600 different compounds. Over 500 different functional group fragmental contributions are used based upon the chemical structure and environment of the fragment. Atom group contributions for

carbon atoms not involved in a functional group are based upon hybridization state, number of hydrogens attached, and structural environment. Over 2000 types of intermolecular contributions involving pairwise group interactions are also included in the calculation. The program checks for tautomeric forms and advises users that log *P* values of compounds having tautomeric forms may be difficult to obtain experimentally. The program will calculate log *P* values of all tautomeric forms of a compound that are generated automatically in ChemSketch, however.

There are limitation to the program, some of which include the following:

•ALLOWED ELEMENTS FOR DRAWN STRUCTURES INCLUDE ONLY THE ATOMS C, H, O, S, P, N. F, CL, BR, I, SE, SI, GE, PB, SN, AS, AND B.

•LOG *P* VALUES FOR CHARGED STRUCTURES, EXCEPT ZWITTERIONIC AMINO ACIDS AND PEPTIDES AND STRUCTURES CONTAINING TETRAVALENT NITROGEN BONDED TO OXYGEN, CANNOT BE OBTAINED.

•STRUCTURES MUST CONTAIN 255 ATOMS OR LESS.

•THE PROGRAM DOES NOT RECOGNIZE STEREOCHEMISTRY SUCH AS *E* VS *Z* C═C BONDS.

Each time the program calculates a log *P* value, it is stored along with the corresponding structure in a "History" file (*.lp) that may be saved and later retrieved. History files display up to 99 structures and log *P* values, and these may be examined several at a time by pressing the **History** button at the bottom of the screen. After opening a History file, structures may be added or deleted. History files may also be saved as database files (*.sdf) for incorporation into the "DataBase" module of the software (hence the designation "dB" in the program name). ACDLabs/LogP dB 3.5 comes equipped with its own internal database of over 3600 compounds whose log *P* values have been measured in the laboratory. Users may search this compendium according to name, formula, log *P* value range, molecular weight, substructure, and literature reference. As a search becomes more narrow, DataBase maintains two parallel search lists: a broader list and one narrower, which is a subset of the broader list. A mouse click allows users to switch between lists. Users may also import *.sdf files generated from the History module to create their own searchable databases. I found using the internal database easy, but encountered difficulty in searching the user database I created from an imported *.sdf file. Lack of documentation on setting up and using user-designed databases also exacerbated my troubles.

How good is ACDLabs/LogP dB 3.5 at predicting log *P* values of compounds outside the data set it uses in the computation process? It is difficult to answer that question in all cases, but it seemed appropriate that I should put the program to the test by calculating log *P* values of some compounds I have worked with over the last several years. These compounds are shown in Figure 1 (drawn using ChemSketch) and then also listed in Table 1. They are all classified as plant phytoalexins (compounds produced by plants to ward off microbial infection) or are known to have phytoalexin-like activity (i.e., the abietic acid series). Log *P* values for many of these compounds have been measured, sometimes by two different research groups.

Table 1. Comparison of Calculated Log *P* Values with Experiment

compd (plant source)	log P_{exp}[a]	ACDLabs/ LogP	CLogP[e]	CACheLogP[f]
phenol	1.42 (1.49)[b]	1.48	1.47	1.38
o-cresol	2.03 (1.95)[b]	1.94	1.97	1.66
p-chlorophenol	2.28 (2.39)[b]	2.43	2.48	2.87
bromobenzene	3.18 (2.99)[b]	2.99	3.01	3.26
naphthalene	3.46 (3.37)[b]	3.45	3.32	3.45
biphenyl	3.84 (4.04)[b]	3.98	4.03	4.08
1 (soybeans)	2.74 (2.62)[c]	3.15	3.17	3.28
2 (soybeans)	2.44[d]	2.82	2.49	3.37
3 (kidney beans)	3.59 (3.58)[c]	4.22	3.96	3.64
4 (potatoes)	3.13[d]	3.27	2.63	2.76
5 (potatoes)	3.26[d]	3.13	2.46	2.59
6	2.88	6.51	6.24	3.78
7	2.90	6.50	6.33	4.40
8	3.37	6.93	6.65	4.27
9	3.34	7.47	7.18	4.61
10 (broad beans)		3.08	2.97	3.73
11		3.83	4.03	3.87
12		3.07	3.72	3.50
13		3.31	3.35	2.99
14		3.16	3.56	3.52

[a] Data taken from Spessard, G. O.; Matthews, D. R.; Nelson, M. D.; Rajtora, T. C.; Fossum, M. J.; Giannini, J. L. *J. Agric. Food Chem.* **1995**, *43*, 1690; measured by reverse phase HPLC. [b] Data taken from Nys, G.G; Rekker, R. F. *Eur. J. Med. Chem. Chim. Ther.* **1974**, *9*, 361. [c] Data taken from Arnoldi, A; Merlini, L. *J. Agric. Food Chem.* **1990**, *38*, 834; measured by reverse phase HPLC. [d] Data taken from Fossum, M. J. Research report to Gary O. Spessard, 1993. [e] CLogP is a program produced by BioByte Corp., Claremont, CA. [f] CAChe LogP is a protocol for calculating log *P* values using CAChe ProjectLeader software developed by the Oxford Molecular Group, Beaverton, OR.

The first six compounds listed in Table 1 are standards whose log *P* values have been carefully measured and range from ca. 1.5 to 4. Within this spread of values are the log *P*s of a large number of naturally occurring molecules. All three programs seem to calculate log *P*s in good agreement with experiment. For compounds **1**−**5**, all three programs also give reasonable agreement with each other and with experiment, the only exception being the value for compound **3** calculated by ACDLabs/LogP, which seems rather high.

The abietic acid series (compounds **6**−**9**) shows wide disparity between calculation and experiment. ACDLabs/LogP and CLogP both give values of log *P* that are up to three logarithmic units higher than experimental results, while CACheLogP gives results that are much closer to experiment. That ACDLab/LogP and CLogP give similar numbers is perhaps not surprising, since both use similar computational approaches. In contrast, the CACheLogP protocol is based upon SCF semiempirical quantum mechanical computations and not incremental contributions of fragments and atoms. CACheLogP values are determined as a function of the difference in heat of formation of solute in the gas phase and water phase as well as the calculated solvent-accessible surface area. A training set of over 200 compounds, whose log *P* values are known from experiment, was used in developing the protocol.

Compounds **10**−**14** are either naturally occurring (**10**) or analogues of phytoalexins isolated from the broad bean plant that have been synthesized by my research group (**11**−**14**). Although the log *P* values of these compounds have not yet

been determined experimentally, it is nice to see that all three software packages give similar results.

ACDLabs/LogP db 3.5 is an excellent program for calculating log *P* values, in most cases providing reliable estimates of this most important chemical property. Like all such programs, however, its effectiveness must always be judged in comparison with experiment.

CI980264T