Accord for Excel

Knut Baumann

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604

Received October 24, 1996

Introduction. Synopsys' Accord 2.0[1] is an add-on transforming Microsoft Excel into a "chemical" spreadsheet. The package provides a set of tools to manage and analyze chemistry objects, such as reactions, molecules, and substituents, and related data in Microsoft Excel. The set of tools comprises reaction, structure, and substructure searching, similarity ranking, substituent−property (R-group) analyses, management of substituent libraries and retrieval of the associated property data, cell-based chemical calculations, and reading and writing of standard chemical structure files. Being "chemically aware" the program recognizes and differentiates stereoisomers and aromatic and nonaromatic rings during structure and substructure search. Tautomers are considered to be different molecules.

As an add-on Accord does not principally change the Excel environment but does extend it significantly with respect to chemical structure handling facilities. With Accord it becomes possible, that Excel worksheets can contain embedded chemistry objects. Chemistry objects are displayed by default with their name only. Choosing the "draw" command, from either the newly added chemistry menu, or the additional toolbar underneath the main Excel toolbar displays the chemical structure. Applying the "delete" command switches back to the former representation of the chemistry object. Double-clicking an Excel worksheet-cell containing a chemistry object launches an OLE-enabled chemical structure drawing program[2] and loads the respective object for editing purposes. Double clicking an empty cell will also launch the structure drawing program enabling the user to create a new object. On return, the worksheet will contain the updated or newly created chemistry object.

System Requirements. Windows edition: The hardware required is an IBM compatible Personal Computer using an Intel 386 or higher processor, running in enhanced mode, at least 2.5 MB of free hard disk space, and 2 MB available memory. The software required is Microsoft Windows 3.1 or Windows 95, Microsoft Excel 4.0 or higher installed on the system, and a standard chemical drawing package. Accord can be locally installed or on a network.

Macintosh edition (not reviewed): The hardware required is a 68020 or higher Macintosh, at least 2.5 MB of free hard disk space, and 2 MB available memory. The software required is System 7, Microsoft Excel 4.0, or higher installed on the system and a standard chemical drawing package.

The program has been evaluated on three IBM compatible PC systems. The first using an Intel Pentium 75 MHz processor, 8 MB of RAM running under Windows 3.11, the second using an Intel Pentium 100 MHz processor, 16 MB of RAM running under Windows 3.10, and the third using an Intel Pentium Pro 200 MHz processor, 64 MB of RAM running under Windows NT 3.51. Microsoft Excel 5.0 and ISIS Draw 1.2 were teammates of Accord on all systems. Running Accord for Excel on the first system demands some patience from the user.

Installation, Help, and Documentation. The installation was friendly, flawless, and easy to follow without the manual. Accord's installation routine needs to update some Dynamic Link Libraries (DLL). If the installation routine finds these DLL's to be present it prompts the user to back up these files before replacing them.

After installation one may work through the printed tutorial guide (30 pages) to learn the package. The tutorial is concise and introduces each feature of the program. Example worksheets are supplied facilitating the first steps in the new environment. However, the introduction to the R-group analysis was too short. Seeking advice in the well-written user's guide (130 pages) or the on-line help finds the way out. On-line help is available through Excel's standard mechanisms and through an additional help topic in the chemistry menu. If the directory containing Excel is not included in the path-statement (autoexec.bat) Excel will not find Accord's on-line help file. Even without having studied the tutorial guide the new Accord user will accomplish much aided only by the clarity of the menus and the toolbar tips.

Chemical Analyses with Accord. Structure and substructure search as well as similarity ranking are well implemented and reasonably fast operating tools. They use the Excel 5.0 filter tool to selectively display the result of the procedure. The documentation of stereochemical geometries that Accord recognizes and differentiates during exact structure and substructure searches is excellent. The same applies to the documentation describing how substituents are matched during R-group analyses (see below). Therefore it is astonishing why a similar documentation does not exist for Accord's similarity ranking algorithm. Generally, similarity is context dependent. A chemist interested in structure−activity relationships of drugs will have a different understanding of similarity than a chemist dealing with structure−spectra correlations. Systematically changing substituents of a core structure reveals the way Accord behaves performing similarity searches. However, this should be part of a comprehensive documentation and should not be a user's task.[3]

The R-group analysis tool will be very helpful for all chemists dealing with structure−property correlations to quickly reveal trends and relationships in the data under scrutiny. It relies on the substructure search tool and on the supplied[4] or user-defined substituent libraries. Figure 1 shows a result of an R-group analysis. After the user has defined a generic core structure with attached R-group labels, Accord searches a selected list of chemical structures in the worksheet for this core. In case a match is found, it automatically perceives the substituents (R-groups) and displays the entire structure containing the core; additionally it separately displays the perceived substituents. If a substituent is not contained in any library the user can choose from various options to display these substituents (e.g.,

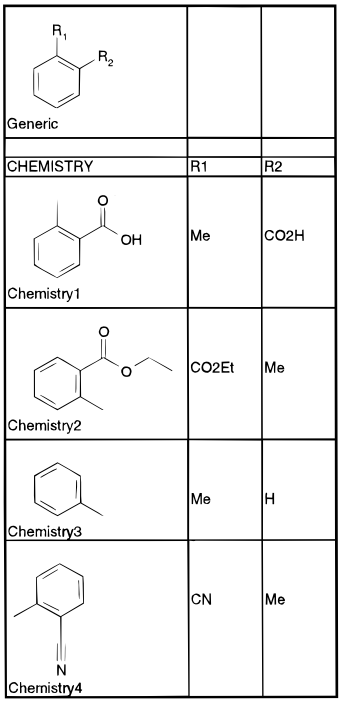

Figure 1. R-group analysis with symmetric core and possible ambiguities in the assignment of the substituents. The user is unable to control the outcome (e.g., Me or alkyl chains, respectively, should always be assigned R_1).

Smiles-strings). Creating and maintaining the substituent libraries is straightforward and well documented.

If the core displays symmetry with respect to the attached R-labels (Figure 1), or, more generally, if there are possible ambiguities with the R-group assignment, no unequivocal pattern as to how substituents are assigned was found. The User's guide and on-line help do not give the particulars for these cases. It seems that the program relies on the numbering scheme of the molecule (i.e., on the way the molecule is inputted).

After searching, sorting, and filtering the chemical structures with Accord's structure handling tools, the very strong point of the concept of Accord comes into play, because the property data associated with the structures are right in place to be analyzed with Excel's mathematical, statistical, and charting tools, making it unnecessary to switch between several applications. Another software package that has similar functionality and interfaces to Excel is MDL's ISIS SAR Table, which is part of ISIS/Base.

Data Exchange, Data Handling. Accord for Excel reads and writes various chemical standard data formats, making the transfer to and from chemistry and biology database systems easy. Particularly, Accord imports and exports the following file-formats for collection of molecules: MDL's SD- and RD-files and SMD-files (standard molecular data). Data formats for single molecule data that can be read and written by the program are MDL's Mol-files (*.mol), Rxn-files (*.rxn), and ISIS sketch files (*.skc), Smiles strings (*.smi), standard molecule data (*.smd), and ChemDraw-files (*.chm). Additionally, the following cut and paste formats to read to and from the clipboard are supported: ISIS Draw sketch file, MDLCT, ChemDraw, and the native Accord format.

Synopsys does not state an upper limit for the number of structures that can be handled. This will, e.g., depend on the hardware being used. It was found, that files with up to several hundred structures are convenient to handle.

Conclusion. There is much to recommend Accord both for casual users and QSAR/QSPR-scientists. It is a flexible and easy to use structure handling and analysis tool framed within Excel. Therefore it is a good complement to a chemical drawing package and a structure database. As noted, a drawback is some missing options for the R-group analysis which would allow a more controlled assignment of the substituents. According to Synopsys, this is an area currently under development. Furthermore, Synopsys intends to extend Accord's functionality with Combinatorial Chemistry tools (reverse R-Group analysis) and will interface it with particular Daylight packages (CLogP/CMR, Daylight Chemical Information Systems, Inc.). I look forward to seeing the results.

Summing up, Accord is a good and useful tool, with further steps necessary to reach perfection.

REFERENCES AND NOTES

(1) Accord for Excel 2.0 is available from Synopsys Scientific Systems Ltd., 175 Woodhouse Lane, Leeds, West Yorkshire, United Kingdom, LS2 3YY, (Phone: +44(0)113 245 3339, Fax:+44(0)113 243 8733, WWW: http://www.synopsys.co.uk) for £495 (£295 for academics). Free evaluation copies are also available.

(2) OLE-enabled (Object linking and embedding) structure drawing packages that are compatible with Accord 2.0 are: ISIS Draw 1.2 (MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577, WWW: http://www.mdli.com) and ChemDraw (CambridgeSoft Corporation, 875 Massachusetts Avenue, Cambridge, MA 02139, WWW: http://www.camsci.com). Accord for Excel 2.0 is not yet fully compatible with ISIS Draw 2.0 according the Synopsys support department (which immediately responded to all questions occurring during the review).

(3) When asked, Synopsys stated that the similarity ranking is based on the Tanimoto coefficient of mutual presence/absence of a set of fragments within molecule pairs.

(4) Accord comes with a substituent library of approximately 300 common substituents with various substituent-specific data.

CI9501960