COMPUTER SOFTWARE REVIEWS

SARvision Plus

SARvision Plus is an effective tool to enable the medicinal chemist to rapidly examine data patterns related to chemical structure in compound collections. The strength of the program centers on the capability to intuitively categorize chemical connectivity and examine trends in data and property space in the context of these chemical scaffolds. SARvision Plus is the main component of a suite of tools by ChemApps (a project of Altoris, Inc.) which also includes Chem SABRE (in-depth Substituent Analysis, Bioisostere Replacement, and Enumeration) and ChemDOX (chemical literature and electronic chemical patent analysis). Although this review will focus on SARvision Plus, it is important to note that ChemApps has built these additional tools which provide powerful additional functionality and nicely complement the SARvision platform.

Opening SARvision brings the user to a workspace consisting of a tree view (scaffold based expandable/collapsible folders of user defined or automatically generated molecular substructures) and a table view ("MS Excel-like" main data and structure visualization space). The program accepts SD files (.sdf) as the main format for structural and data input. Importation of text (via .csv or .txt) format is possible as a means to append new data to the structural information. Additionally, SARvision has a built-in capability to add basic property space descriptors to files. Molecular weight, logP, polar surface area, hydrogen bond donors and acceptors, rotatable bonds, number of rings, and SMILES are among the properties that are readily calculated within SARvision.

The capability to rapidly create hierarchies of chemical structures in the "Tree View" is a defining feature of SARvision. Upon SD file import, it is possible to have the software automatically identify and organize scaffolds for the user. This automatic scaffold identification algorithm can be very useful for analyzing the composition buckets of new or unfamiliar compound collections. A double click on a scaffold on the tree performs an instant substructure filter allowing the user to interrogate that subset of chemical matter. With the ability to sort the scaffold tree by frequency and size, SARvision allows for the rapid understanding of the structural distribution of a compound collection. Furthermore, if the SD file contains biological data or the user is interested in filtering on the property descriptors calculated within SARvision, the scaffold tree can be sorted by value determined by a biological or property space end-point. Overall, *automatic* scaffold identification is an effective tool for characterizing and comparing data sets less familiar to the user such as high throughput screening results, library screening collections, and commercially available starting material or fragment collections, among others. However, for drug discovery project-focused files, which include property space, biological, and pharmacokinetic data, taking advantage of *user-defined* scaffold trees within SARvision can be quite enabling to a project team and to the individual scientists analyzing SAR patterns for discovery programs.

Series-based structural-activity trends are critical to informed decision making for drug discovery. Deconstructing activity and/or liability trends by R-group (or fragment) analyses provides complementary design strategies to those based simply on a whole-molecule SAR perspective. With a well constructed *user-defined* scaffold tree in SARvision, scientists on a project team can communicate in a common series language and are well equipped to understand fragment-activity relationships.

Setting up a *user-defined* scaffold tree in SARvision is relatively straightforward; however, some iterative tweaking is needed in order to get the most benefit out of the software. A simple right-click in the tree space allows the user to draw a scaffold of interest. The program allows for the option of using a licensed copy of either ChemDraw or IsisDraw in addition to the built-in ChemApps drawing program. A tree can then be manually built in a hierarchical manner with substructures recognizable to the project team. Scaffold renaming is straightforward allowing the user to name chemical classes and series in a way familiar to the project scientists. If a consistent structural hierarchy is well organized, then many of the annotation options work seamlessly. The use of explicit hydrogen atoms on the scaffold significantly simplifies the R-group analyses to provide SAR tables targeted for variables of interest and to more easily define compound classes and series. The interface allows for simple drag and drop of one scaffold to the folder of another scaffold to rapidly organize the tree. There may be a need to custom order the scaffold tree to keep related scaffolds together. Unfortunately, the program does not allow for custom scaffold ordering via drag and drop, and the user is required to use a workaround if custom scaffold ordering is desired.

After a well organized structural hierarchy is built with logical class and series naming, the user can then "correlate molecules to scaffold" to add the class, series, subseries, etc. as a multiline text in a single column. The program has a convenient algorithm for rapidly splitting multiline text into separate columns providing an effective method for series annotation. Thus, the tree can be used as a template for any SD file making it easy to classify any new compounds in a project and also allowing for instant scanning of new sources of chemical matter for compounds of interest.

The table workspace in SARvision provides the user with tabular or gridlike interface familiar to users of MS Excel. There are three subareas with the table workspace: data set, R-group analysis, and a multivariate graphing component.

The first component, the data set table view, contains many common features users might expect such as column sorting, renaming, property based conditional formatting, show/hide, and drag and drop column reordering among others. The conditional formatting (heat mapping) is useful to highlight compounds of interest; however, it falls short in that it works for numeric data only with the present version (no text-based or date-based option). The formatting is limited to continuous coloring schemes omitting more common categorical (stoplight) formatting options. The functionality to globally categorize compounds by "active properties" is very useful in that project teams can code and thus highlight or filter by active screening criteria. There is an option to categorize actives in a *subset* of the defined criteria, enabling more "fuzzy edge" decision making.

A somewhat hidden benefit of SARvision is the capability to flatten relational data using the "calculate from multiline" functionality. When data are shared by means of an SDF export from a relational database, the data may contain multiline data in some fields due to multiple values/compound. Although several methods exist to prepare flat files from a relational export, this task can sometimes be challenging to the project scientist interested in rapidly viewing summarized data. SARvision provides a very intuitive way to flatten SD files with a "one-to-many" issue. The options of using the average, geometric mean, median min, max, first, last, or average with standard deviation provides great flexibility of converting multiline SDF exports to useful flat file formats. The ability to export the modified SD file for use in other desktop applications provides a convenient workflow for rapid relational to flat file conversion for the typical project scientist.

The second component of the table workspace, the R-group table, is very powerful to the medicinal chemist who prefers to examine data trends via SAR tables. Enabling the R-group table and double clicking on the scaffold of interest will immediately bring up an SAR table with the R-groups tabulated in a column format integrated with the compound data. Property space calculations can be performed on these R-groups allowing the user to examine the activity of the chemical matter as a function of deconstructed fragments. Although the data in the main table can be examined graphically for the parent molecules within SARvision, plots of the data for *fragment*s are possible only by

10.1021/ci800152v CCC: $40.75
Published on Web 06/05/2008

means of export to another component (ChemSABRE) or a third party software such as MS Excel or Spotfire.

The third component of the table workspace is the multivariate graphing functionality. The user can perform snapshot 2D scatter plot or histogram plots of numerical data as well as 2D plots of data fields as a function of scaffold. The plots are structurally interactive in that a mouse float over a point will provide a pop-up window of the structure of interest. Additionally, by double clicking on a scaffold in the tree window, the graph adapts to the filtered chemical scaffold, displaying only those compounds of interest. SARvision allows for the export of graphical analyses as an image or to save on the clipboard for use in MS Word or MS PowerPoint for reporting purposes. The graphic capability is a welcome addition to any program designed to interrogate SAR; however, there are some limitations and opportunities for enhancement for the graphical interface in SARvision.

One very important advantage of structure/data graphical analyses is providing a means to communicate SAR information, particularly progress over time. Although SARvision's graphic capability works fine for numerical data (notably property space or activity as a function of scaffold), it lacks the ability to create plots using text as bins (for example, scaffold frequency histograms). Also, the program currently does not allow for the incorporation of date fields in graphical analyses, leaving the user to rely on other software to analyze or communicate progress over time. In addition, details such as the small text on the axes do not appear to be modifiable (neither the font nor the numerical formatting) making the graphs difficult to read on slides, thus, tempering the benefits of graphical SAR communication via SARvision.

The ability to export SD Files from SARvision allows users to utilize the program's strengths of scaffold analysis/series identification while tapping into the power of other programs, if needed. The advent of the SD file as a common currency of exchanging structure/data information has introduced this file format to a wide population of scientists who do not specialize in information technology. Manipulation or modification of SD files, however, remains primarily in the realm of information specialists. SARvision empowers the nonspecialist with a user-friendly interface to modify SD files. As described, adding data, inserting calculated fields, appending hierarchical scaffold information, flattening relational data, merging, and pruning SD files is straightforward in SARvision. The export options to MS Word and MS Excel include standard delimited text files, html, and SD files, making SD file modification as easy as editing a word document. This can be very helpful for tweaking an SD file for import to other software. Moving the data from SARvision to a chemically aware *interactive* graphical program (Spotfire) via an SD file is a very complementary workflow to derive maximum value from the structural organizational strength of SARvision.

Although much of the functionality that is built into SARvision begins to approach a data interface, there are presently several critical components missing which limits the software for use as a main front-end for interface for drug discovery programs. With no text string search capability, poor list management, and no option for single compound comprehensive form views, it is difficult to compete with relational database front-end interfaces such as Dotmatics, Instant JChem, Seurat, ChemFinder, and Isentris, among others. The program can be integrated with other in-house applications via line commands or data piping tools such as Pipeline Pilot or Knime. Alternatively, ChemASP (a server version of the automated scaffold perception algorithms) provides another dimension to the possible integration of the tree building capability with in-house software. It is important to keep in mind that the program's strength is the chemical scaffold organizational power that adds value in the form of a standalone program, or in a complementary fashion with other front-end programs.

The support for the program is solid with a detailed user guide and integrated help menu in the program. There are currently no Web-based forums for the community of users; however, ChemApps is extremely helpful with addressing any questions and aggressively making improvements to the software.

Overall, SARvision is a valuable component for organizing chemical structures and remains a powerful tool in the arsenal of chemoinformatics software. The strengths of the program are the intuitive interface to incorporate a structural hierarchy to chemical data files as well as the capability to easily work with and customize SD files. Many, but not all, of the identified limitations with the program are addressed in ChemSABRE, a related component in the suite of desktop tools by ChemApps. Many large pharmaceutical companies will likely have custom software that captures much of the functionality in SARvision and may find the program to be redundant with other tools. However, midsize pharma and smaller biotechs which operate with commercial desktop data interfaces will likely gain an advantage with SARvision, particularly groups that operate in MS Excel-focused data sharing environments. Any organization that is involved in leveraging data matrices of chemical structural information coupled with multifaceted biological data would benefit from evaluating SARvision to determine if it adds value to the organization's suite of informatics tools.

Gregory A. Reichard*
Decision Informatics, Ltd., Ann Arbor, Michigan

CI800152V
10.1021/ci800152v