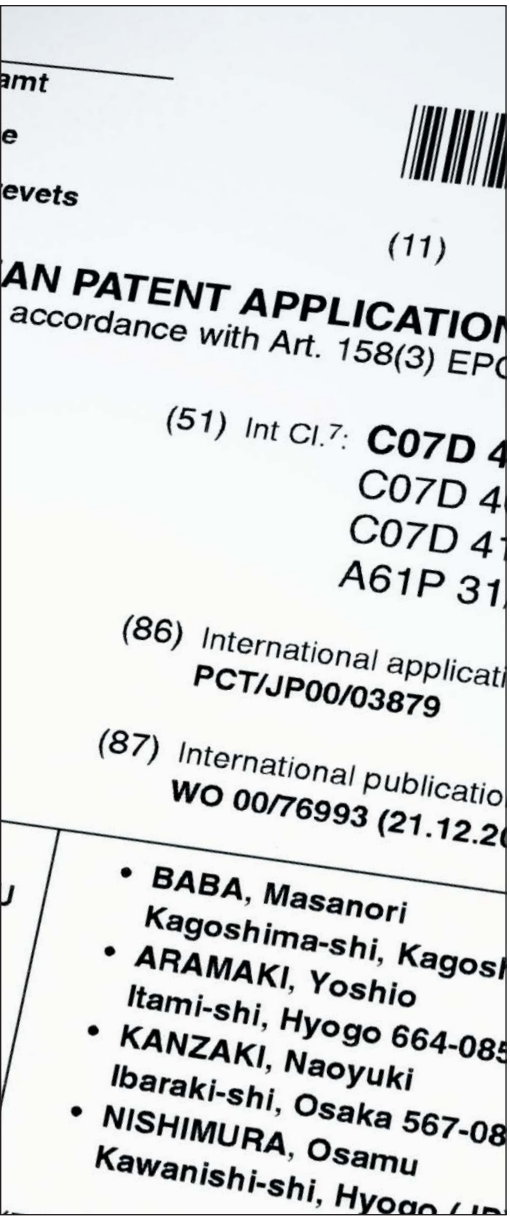
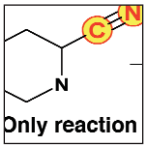


MDL® Patent Chemistry Database

Featuring "At the Bench" articles:

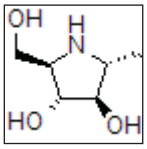
Accessing three decades of patent chemistry
(2005 Vol 23 No 1)

Page 2



More effective synthesis planning
(2005 Vol 23 No 2)

Page 4



Better bioactivity profiling
(2005 Vol 23 No 3)

Page 6

The structure-searchable MDL® Patent Chemistry Database provides scientists with indexed chemical reactions, substances and substance data from chemistry and life sciences patent documents (World, European since 1978, US since 1976). This series of reprinted articles from *Molecular Connection* newsmagazine demonstrates typical workflow scenarios in which the Patent Chemistry Database can improve the speed and efficiency of synthesis planning and drug discovery efforts.

ELSEVIER

MDL

Access to three decades of patent chemistry

Patents are an important and under-used source of information in chemistry and life sciences research. While many text-based systems exist for accessing patent information, structure-based searching offers a more powerful and flexible way for scientists to mine this vast pool of important information.

To support this need, Elsevier MDL released the new structure-searchable MDL® Patent Chemistry Database, specifically designed for research scientists andinformation professionals.

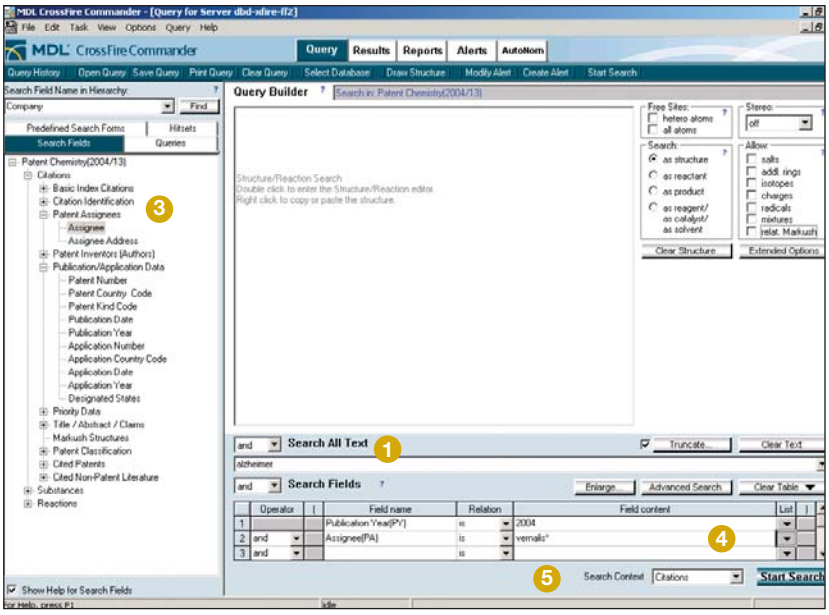


Figure 1 illustrates the steps for querying the Patent Chemistry Database using the MDL® CrossFire® Commander browser. The database can also be accessed through the DiscoveryGate® environment.

1. **Search All Text:** Enter **Alzheimer** to search across the complete citation record (title, abstract, claims, etc.). The search system automatically applies left and right truncation (note **Truncate** check box) to find variants such as “Alzheimers”.
2. **Search Field Name:** Enter **Company** and click **Find**. There is no field name Company, but the database reference guide suggests using the search field **Patent Assignee**. Use **Locate in Tree** from the guide menu.
3. **Select Field Name:** Double click on **Patent Assignee** in the data field tree to copy this field to the **Search Fields** grid.
4. **Enter Field Content:** Enter **Vernalis*** (use the asterisk to set right truncation and find expressions like “Vernalis Limited”), or select specific company names using the LIST button. Repeat steps 2-4 analogously to specify Publication Year=2004.
5. **Search Context:** Select Citations from the **Search Context** drop-down menu to get citation records, which show the front-page data, claims text and all indexed structures and reactions as a hit set. Click **Start Search**.

Updated every two weeks, **MDL Patent Chemistry Database** contains about 2.4 million reactions, about 3 million organic, inorganic, organometallic (and polymeric*) compounds and related information from approximately 400,000 organic chemistry and life sciences patent publications (covering U.S. patents since 1976, World and European patents since 1978). The database includes:

- **Complete reaction texts** for easily checking a reaction’s relevance
- **InfoChem ClassCodes** to find similar reactions in other databases and group them according to similarity
- **Markush reaction display*** together with claims text for exploring the scope of a patent and the claimed reactions
- **Location (page)*of the reaction** in the original patent document for quick reference

The MDL Patent Chemistry Database can assist researchers in monitoring competitors and industry trends, designing new synthetic methods, developing drug profiles and selecting and optimizing leads. The database helps researchers quickly understand the scope and relevance of patents, allowing more effective synthesis planning and better bioactivity profiling.

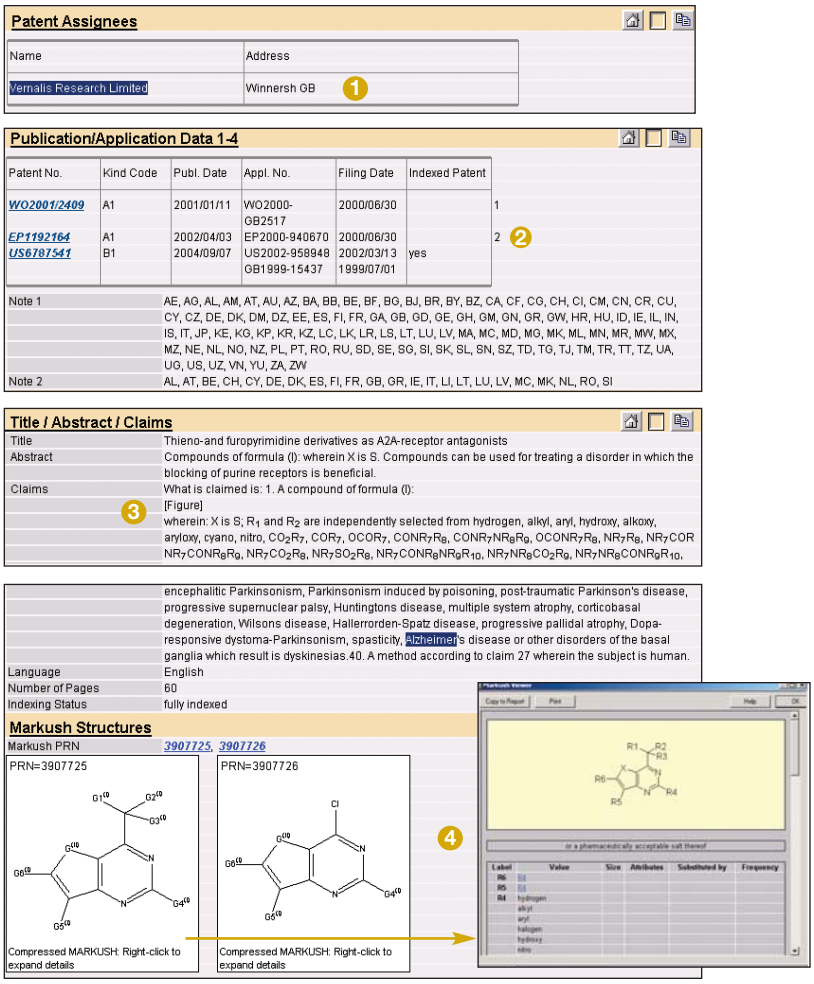
The MDL Patent Chemistry Database brings patent claim texts and Markush structures/reactions[1] to the desktop in an easy-to-view format, enabling researchers to check the relevance of located patents quickly and easily.

The workflows illustrated on the following pages demonstrate the search capabilities of the database in typical work scenarios such as using reaction searching for more effective synthesis planning (pages 4-5) and structure searching for better bioactivity profiling (pages 6-7).

The scenerio beginning on this page introduces a simple text search to see if a company is active in a particular therapeutic area, in this case whether Vernalis Research had any patent applications on anti-Alzheimer agents published in 2004.

Figure 2 displays the results of searching the Patent Chemistry Database for Alzheimer's drug patents assigned to Vernalis Research in 2004. From this, scientists can quickly assess all the relevant information and determine whether this particular Vernalis patent has a bearing on the project under consideration.

1. The database highlights the retrieved hit term, the patent assignee name, for clear viewing.
2. The database includes a table of the main patent equivalents of a patent family (including the designated states for World and European patents) with each patent number linked to the original patent document.
3. To facilitate more effective searches, the complete claim text for the patent can be searched in combination with a structure/reaction search.
4. By seeing the claim text alongside the main Markush[1] structures/reactions, scientists can easily check the relevance of a located patent. Right-clicking on the structure formula shows the expanded Markush formula with the complete substituent list. Markush structure search is not possible today.



More effective synthesis planning

Reaction searching in MDL® Patent Chemistry Database

The previous exercise showed how scientists and information professionals can use a text search of the new MDL Patent Chemistry Database to quickly determine if a company is active in a particular therapeutic area.

This scenerio illustrates a reaction search in Patent Chemistry Database, and then show how reaction classification with InfoChem ClassCodes enables scientists to rapidly find similar reactions in MDL and other databases that might be useful in synthesis planning.

(The following examples illustrate the use of MDL® CrossFire Commander to search Patent Chemistry Database. The database can also be accessed through DiscoveryGate®.)

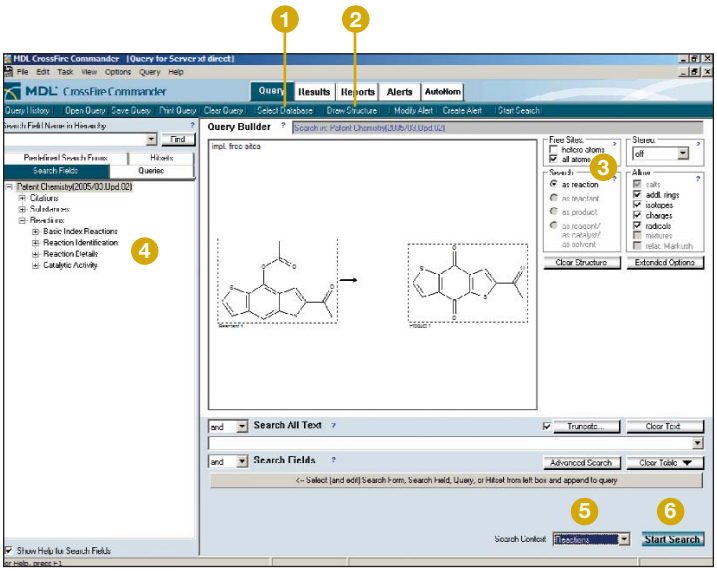


Figure 1 illustrates how to query for reactions in the Patent Chemistry Database.

1. Click **Select Database** and choose Patent Chemistry Database from the menu.
2. Click on **Draw Structure** and then enter the reaction query.
3. To perform a substructure search, select **all atoms** under **Free sites**.
4. To modify the query, select data fields from the hierarchy tree (click to add).
5. Next to **Search Context**, select **Reactions**.
6. Click **Start Search**.

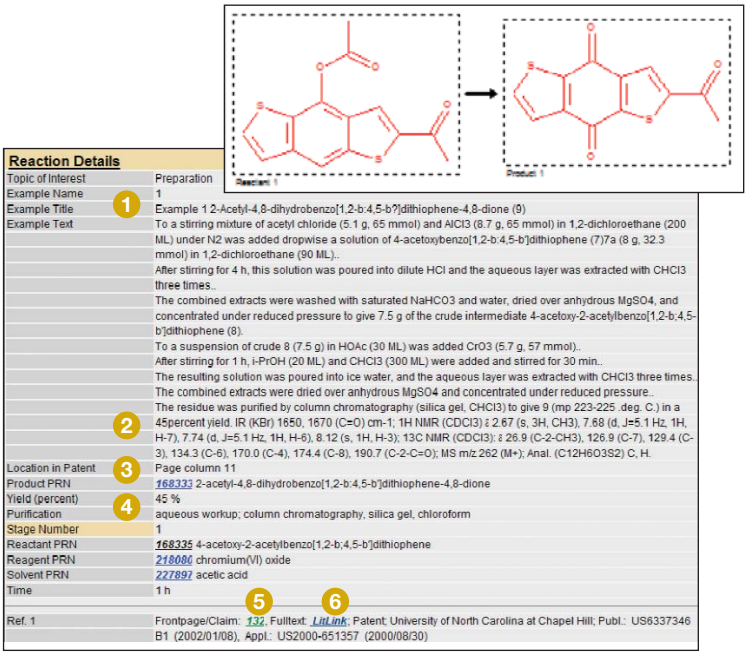


Figure 2 shows a typical hit record from the reaction search described in Figure 1. From the database record, researchers can quickly evaluate the relevance of a patent.

1. **Complete example text** including example number and title as given in the original patent document
2. **Spectral data** (including peaks) of the product
3. **Location in Patent** specifies the page on which this reaction appears in the original patent publication (available from December 2003)
4. **Substance dossier:** a product registry number (PRN) links to the product record with all substance data and related reactions
5. **Frontpage/Claim** link leads to the corresponding database citation record including indexed structures and reactions
6. **Link to full-text document:** Clicking on **LitLink** connects to the original patent document at a patent provider (e.g., EspaceNet, US Patent Server, MicroPatent, Delphion)

The DiscoveryGate reaction collection

MDL Patent Chemistry Database—and all DiscoveryGate reaction databases—are indexed with InfoChem ClassCodes. This classification relates 17 million reactions from different databases, giving researchers the ability to immediately compare a reaction to similar reactions and group similar reactions.

With the simple steps outlined below, scientists planning a synthesis can acquire the entire reaction text of a published reaction and locate an overview of the complete methodology from another database.

In the next issue: Learn how to use the MDL Patent Chemistry Database for better bioactivity profiling.

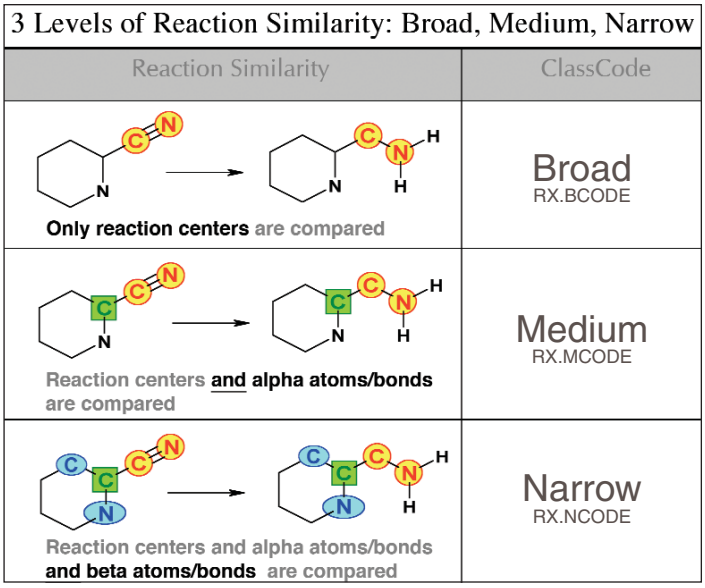


Figure 3 displays the principles of InfoChem ClassCodes.

Three levels of reaction similarity are defined—Broad, Medium and Narrow—depending on whether only reaction centers or also alpha and beta atoms/bonds are compared.

Reactions which are similar with respect to the similarity level have the same ClassCode.

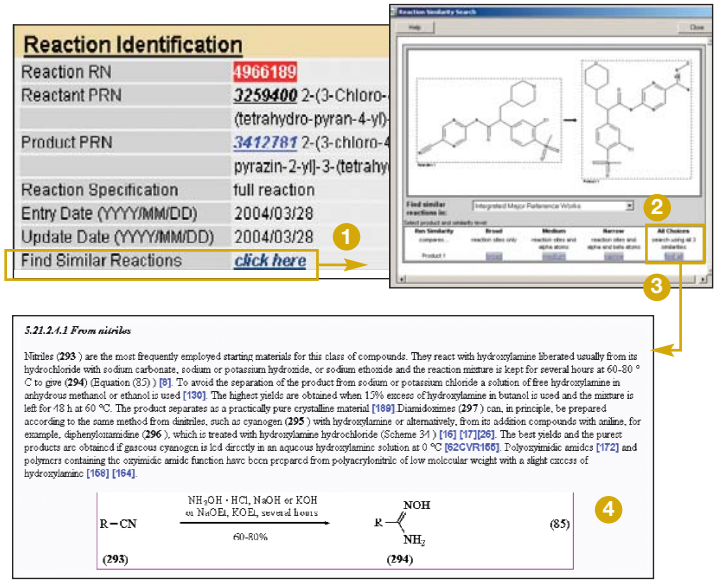


Figure 4 shows how to find similar reactions using InfoChem ClassCodes.

1. Click the **Find Similar Reactions** link in a hit record from a reaction search in the Patent Chemistry Database.
2. Select the database you want to search (for example, ChemInform Reaction Library, other CrossFire databases, Current Synthetic Methodology, etc.)—in this case, Integrated Major Reference Works.
3. Select similarity level: Broad, Medium, Narrow or All Choices.
4. The hit record in Integrated Major Reference Works is an expert dossier evaluating the methodology with respect to highest yields and best reaction conditions for specific transformations.■

Better bioactivity profiling

Export bioactivity data from MDL® Patent Chemistry Database to SAR tables

The ability to quickly find property data for substances of interest and create structure-activity relationship (SAR) tables is important to researchers in the lead discovery process.

This scenerio demonstrates this type of bioactivity profiling with MDL Patent Chemistry Database, using the MDL® CrossFire® Commander browser.

How to search substances and their data

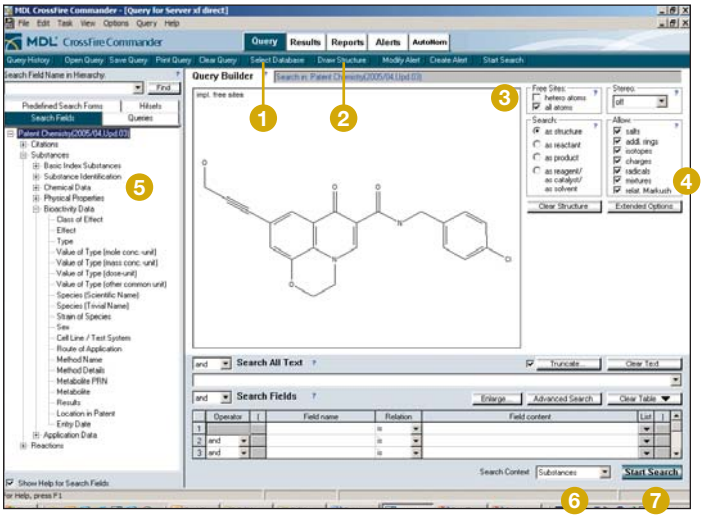


Figure 1 illustrates how to find substance data in the Patent Chemistry Database.

1. Click **Select Database** and choose **Patent Chemistry Database** from the menu.
2. Click **Draw Structure** and then enter the substance query.
3. For a substructure search, select **all atoms** under **Free Sites**.
4. Check **Allow related Markush** to get Markush structures that are related to your specific hit structures included in the hit set (this is not a Markush structure search).
5. **Data Search:** See the fields in the hierarchy tree, which are indexed in the area of bioactivity/application data.
6. Next to **Search Context**, select **Substances**.
7. Click **Start Search**.

Viewing the bioactivity and application data of a specific hit substance

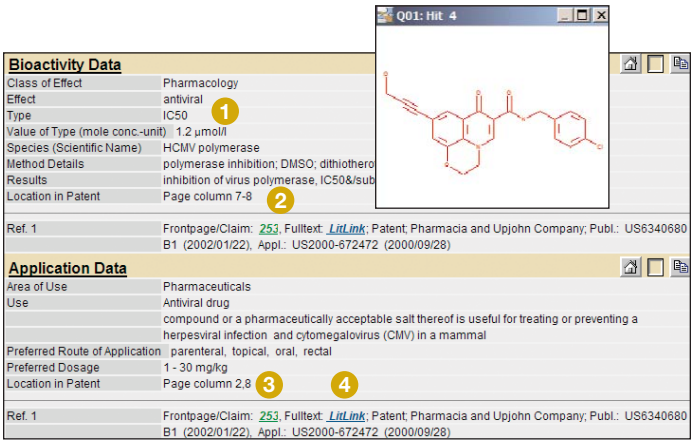


Figure 2 shows a substance hit record from the search in Figure 1, including measured bioactivity and application data described in patents published since December 2003.

Besides the inhibition concentration shown here [1: Type=IC50], the bioactivity data field also may include effective concentrations (EC50), binding constants for drug-target interactions (K_i), lethal doses (e.g. LD50) and other parameters.

The record shown includes the **Location in Patent** 2, which is the page of the bioassay in the original publication, a **frontpage/claim** link 3 to the database citation record and a link 4 to the original document at a patent provider (e.g. EspaceNet, US Patent Server, MicroPatent, Delphion).

Including the related Markush structures in a hit set

From December 2003 onwards specific substances—real or "prophetic" compounds—are linked to the corresponding "Related Markush Structure" in a given patent publication during the indexing process. Selecting the query option **Allow related Markush** includes these related Markush structures in the hit set. *Note: This is not a Markush structure search.*

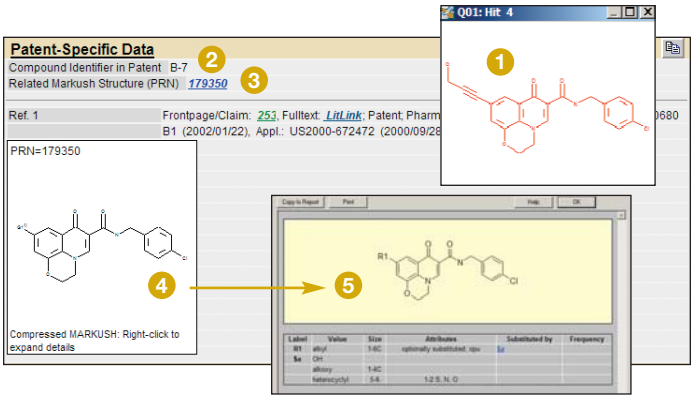


Figure 3

In addition to the bioactivity data, the hit substance of Figure 2 shows **Patent-Specific Data** illustrating the **Compound Identifier in Patent** 2 label. In this case it is B-7, which the inventor used in the document to describe the specific compound 1.

Also shown is the **Related Markush Structure** 3, of which the hit substance is a representative. The corresponding Markush structure image appears in compressed form 4. Right clicking on the compressed form opens an expanded form 5.

Viewing the Markush Family

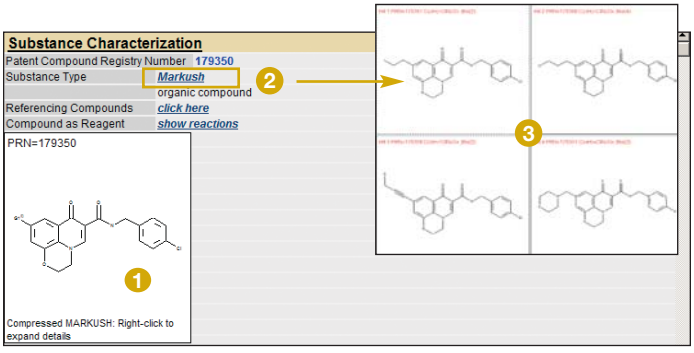


Figure 4

To view other specific examples (real or prophetic) that are representative of the Markush structure 1, click on the Markush link 2 to view the Markush Family 3.

Exporting bioactivity data to a SAR table

All four compounds shown in Figure 4 have bioactivity data (indicated by the red text "Bio" on top of the structure window) that can be easily exported to a structure-activity-relationship table.

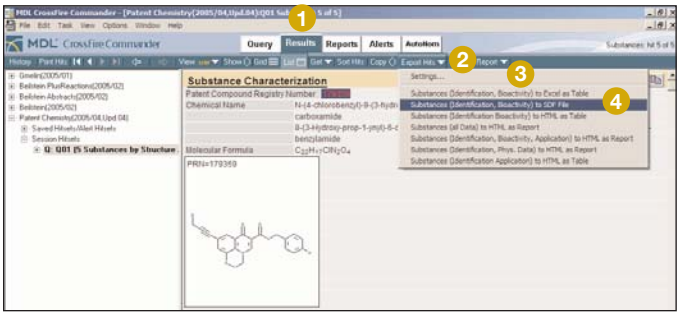


Figure 5 shows how bioactivity data can be exported to a SAR table from the **Results** menu 1 using **Export Hits** 2. You can create your own export formats by clicking on **Settings** 3, or use the predefined export formats 4 delivered with CrossFire Commander 7.0 SP2 covering export to structure-data files (SDfiles), Microsoft Word, Microsoft Excel and HTML. To create a SAR table, use the export format **"Substance (Identification, Bioactivity) to SDfile"** 4 and import the SDfile into MDL® ISIS for EXCEL using the menu options **ISIS** ⇨ **Import** ⇨ **SDfile to worksheet**.

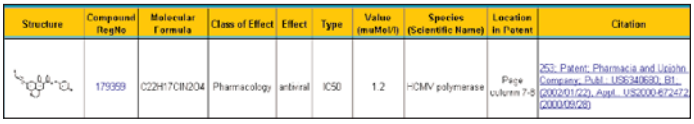


Figure 6 shows one line of the SAR table created for the four Markush Family structures in Figure 4.

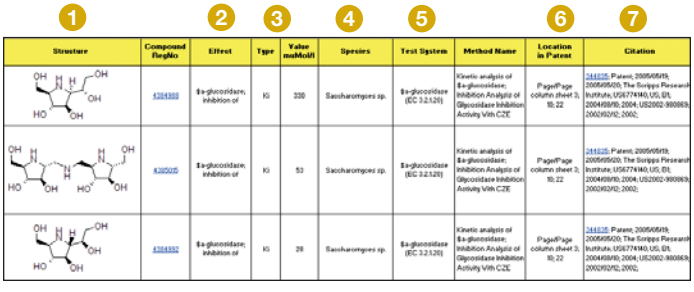


Figure 7 shows another SAR table with drug-target interactions showing the structure 1, the test system 5 (here proteins) and the species 4 together with the effect 2 and the corresponding drug-target interactions constants K_i 3. The table also includes the **Location in Patent** 6 where this bioassay appears and the citation 7, including patent assignee and publication/application number and date.

W W W . M D L . C O M

Elsevier MDL

2440 Camino Ramon, Suite 300

San Ramon, CA 94583

Phone: +1 (925) 543-5400

Fax: +1 (925) 543-5401

About Elsevier MDL

Elsevier MDL provides informatics, database and workflow solutions that accelerate the discovery and development of successful new drugs by improving the speed and quality of scientists' decision making. Academic and industrial life sciences researchers around the world depend on Elsevier MDL for innovative and reliable discovery informatics software solutions augmented by 400 Elsevier chemistry and life sciences journals and related products. For more information, visit www.mdl.com.

Elsevier is a world-leading publisher of scientific, technical and medical information products and services. Working in partnership with the global science and health communities, Elsevier publishes more than 1,800 journals and 2,200 new books per year, in addition to offering a suite of innovative electronic products and online reference works. For more information visit www.elsevier.com.

PAT DBWF/02-06/5k