# Oasis Database Documentation

- Oasis Database overview
- Data model
- Browsing
- Editing
- Search
- Import and export
- Report
- Statistics
- Database properties, capabilities, limitations
- Technical requirements

# Overview

Oasis Database is a database schema and accompanying software for management of chemical information. It provides an extensive list of features such as:

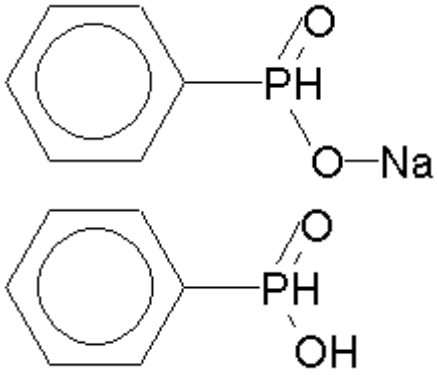
- storage and retrieval of chemical structures with 2D and 3D data
- user-defined property fields – descriptors at the levels of structures, conformers, or atoms
- import/export compatibility with a number of known formats, such as SDF, MOL, SMILES
- highly developed search subsystem – CAS numbers, names, 2D and 3D fragment search, structure, conformational, or atomic properties, structure similarity, logical combinations, saved queries, search result exports
- database browser including inspection and editing of all data items contained in the database
- visual structure editor
- database statistics
- report generator with visual report template editor

Oasis Centralized QSAR Database is the largest Oasis Database to date. Presently it contains approx. 179,000 structures and 3,400,000 conformers with hundreds of millions field data items.

The Database Manager is a part of the Oasis software providing all functionality needed to work with Oasis databases.

# Data Model

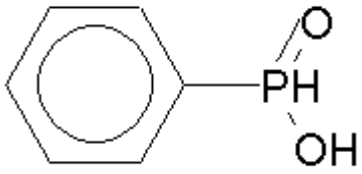
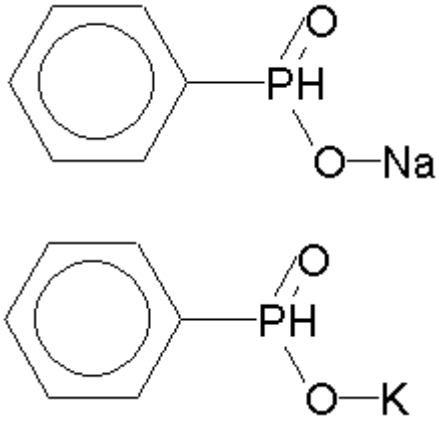
## Structures

| **The following information is stored for each chemical structure (some items are optional)** | **Example** |
|---|---|
| CAS number | **4297954** |
| Representative 2D structure (R-structure)<br>Calculation 2D structure (C-structure)<br><br>Usually the R- and C-structures are the same. They differ if | **Representative 2D structure**<br>**Calculation 2D structure**<br>  |
| A structure may have one or more names. | **Phosphinic_acid,_phenyl-,_sodium_salt**<br>**Sodium phenylphosphinate** |
| Observed descriptors (numerical) | |
| Observed descriptors (character string) | |
| Database affiliations: The chemical databases this structure is known to belong to. | **Danish EPA, DSL, HPVC, IUCLID** |

## Structure sets

Structures that have the same C-structure form a structure set. Such structures share also a number of other properties.

**Example.**

| Structure set | Structures |
|---|---|
|  |  |

None of the items in the next table is obligatory. The only required data that every structure set has is the C-structure.

| All structures of a set share: | Example |
|---|---|
| Calculated descriptors (numerical) | |
| Calculated descriptors (character string) | |
| Conformers<br>Each conformer has<br>3D data (atom coordinates)<br><br>Conformational descriptors<br><br>Atomic descriptors (also called local parameters)<br><br>Functional descriptors<br><br>Boolean descriptors | |

Every Oasis database can define its own set of descriptors at all levels – structure, conformational, or atomic, of any kind – observed or calculated.
The names, types, and values of descriptors can be managed by the user.

## Models

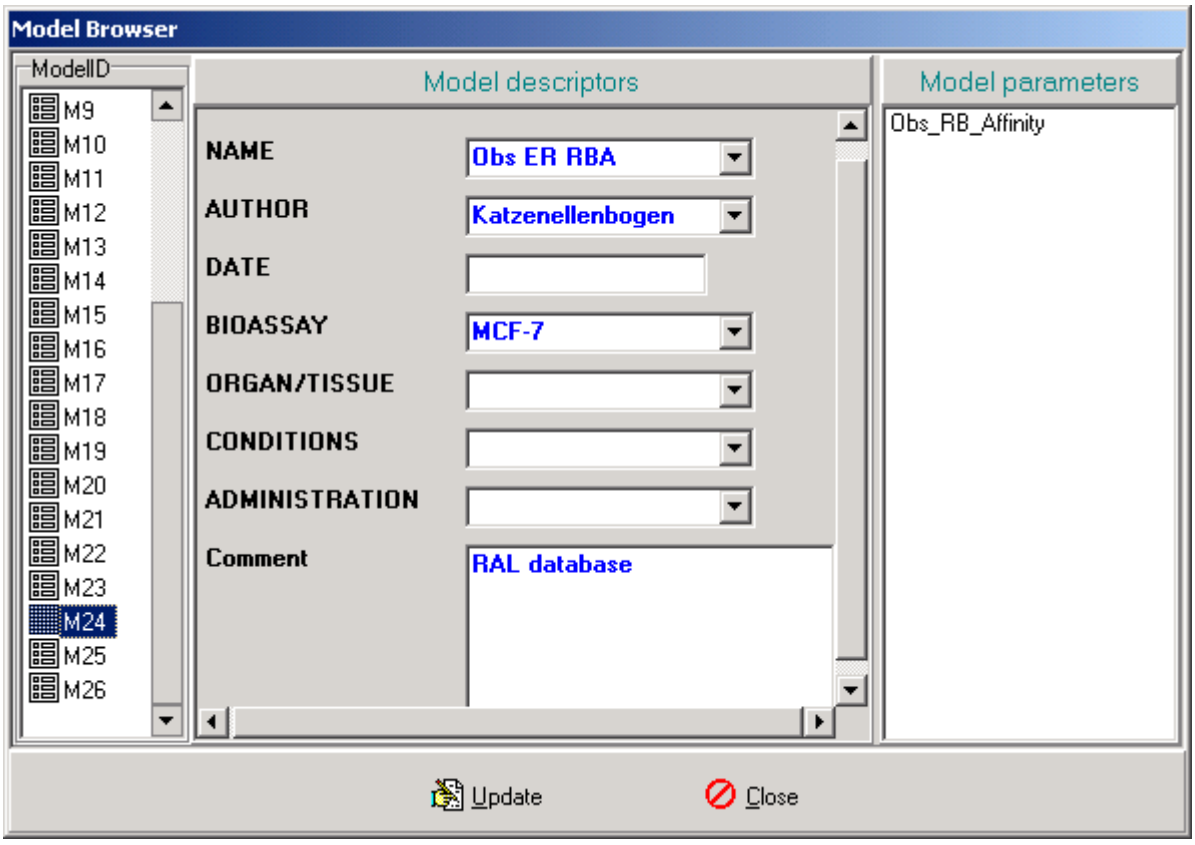
A model is a collection of logically related descriptors (global numeric parameters). Model definitions also comprise model's name, author, date, bioassay, organ/tissue, conditions, administration, and comment text.

Models can be defined using the Model declaration window. You can call this feature before an import operation – see chapter **Import** for details.

Model declaration may be inspected and edited from the Model browser. The Model browser is accessible through a click on View | Model Browser on the main menu:


Another way to call he Model browser is to double-click on a parameter that is registered to some model. Do this in the Observed or String parameters box in the Database browser.

The Model browser is where you can view details on the registration of any model or edit them:




Click on a Model Id (the leftmost box) and the corresponding model's details will appear. Edit them by typing or selecting from the drop-down lists. The list of all parameters currently registered with the model can be seen in the rightmost box.

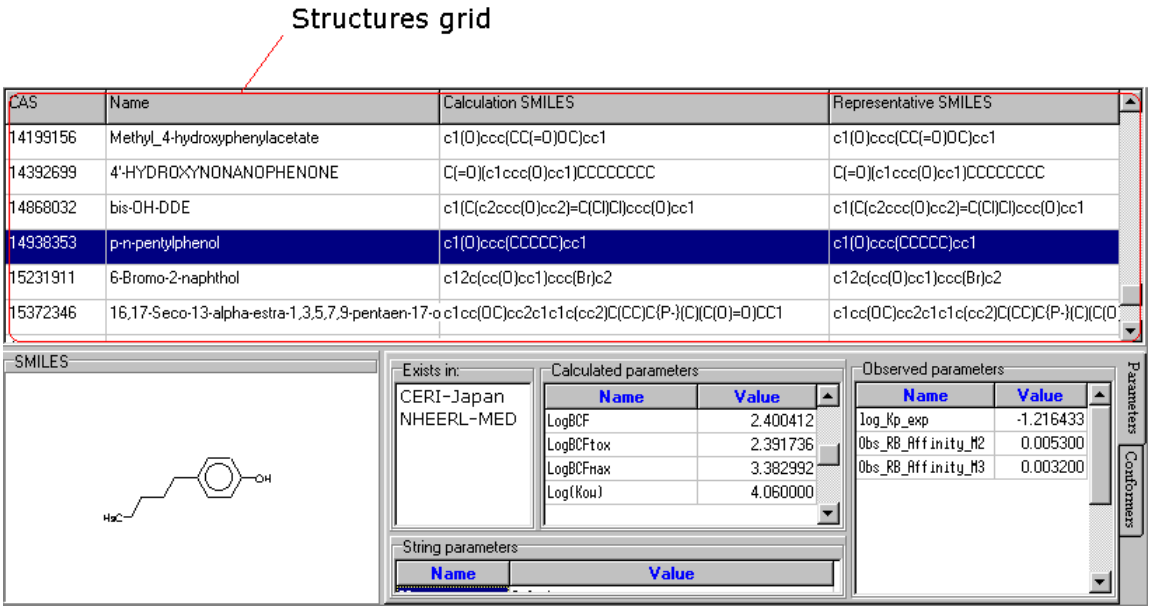
# Database browser

The Database browser is where you monitor your data interactively. A new browser window is displayed for each database you open. Database browsers can display either a whole database or a set of structures resulting from search.

A database browser window can function in two modes called Flat view and Group view, which are covered in the two respective sections below. The default viewing mode is Flat view.

1. Flat view

The central component of a browser window is the Structures grid:




Navigate through the list of structures by the mouse, or using the following key combinations:

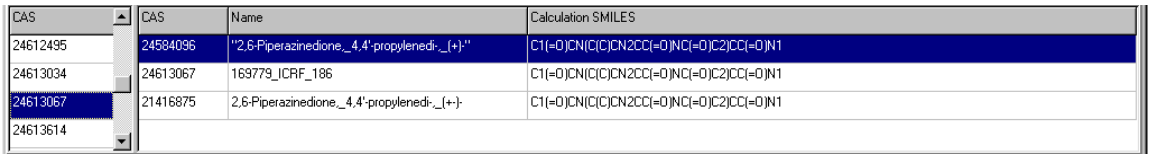
| Key | Function |
|---|---|
| Up arrow, left arrow | Moves one structure up |
| Down arrow, right arrow | Moves one structure down |
| PgUp | Scrolls one page up |
| PgDn | Scrolls one page down |
| Ctrl-PgUp | Moves to the topmost visible row |
| Ctrl-PgDn | Moves to the lowest visible row |
| Home | Moves to the beginning of file |
| End | Moves to the end of file |

The structures in the Structures grid are ordered by their CAS number.

The boxes below the Structures grid show more information about a structure. They are refreshed when a structure is selected.

2. Grouped view

Grouped view is a way to see the structures that share the same (calculation) SMILES. In this view there are two grids: the groups grid and the structures grid:




The Groups grid (to the left) lists all CAS numbers found in the database. You can navigate through that list with the mouse or with the same key combinations as you can use for the Structures grid in the Flat view.

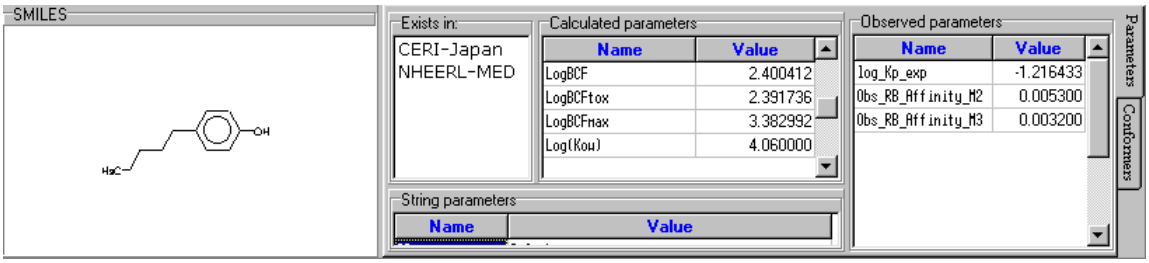
When selecting a CAS number from the Groups grid, all the structures sharing the same calculation SMILES appear in the Structures grid to the right.

The data displayed in the Structures grid is similar to the case of Flat view.

Information boxes are refreshed when changing the selected CAS number from the Groups grid.

3. Information boxes

3.1 Basic mode




2D box (SMILES). This shows a 2D picture of the structure. If the Calculation SMILES and the Representation SMILES are the same, there will be a single 2D box. If they differ, two 2D boxes appear showing 2D pictures of the two SMILES.

Database affiliation box (Exists in): This box lists the databases that contain data on the selected structure.

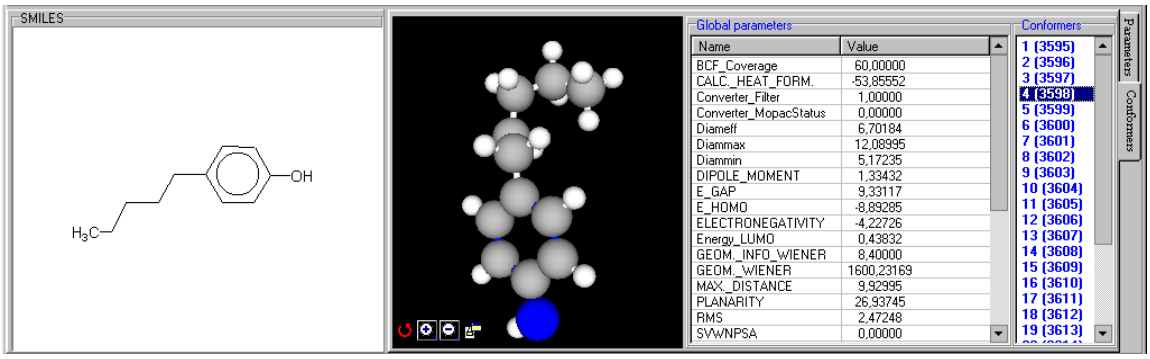
Calculated parameters. The calculated parameters defined for the structure are displayed, each on a single line.

Observed parameters. The observed parameters defined for the structure are displayed, each on a single line.

String parameters. The string parameters defined for the structure are displayed, each on a single line.

3.2 Professional mode

In Professional mode you have all the information boxes that exist in Basic mode plus a number of extra boxes that show conformational information.




The rightmost box (Conformers) contains the list of conformers stored for this structure. Click on any conformer to get detailed information about it – it appears in the next two boxes to the left.

Right-click on a conformer item invokes the conformer pop-up menu where you can choose to delete a specific conformer.

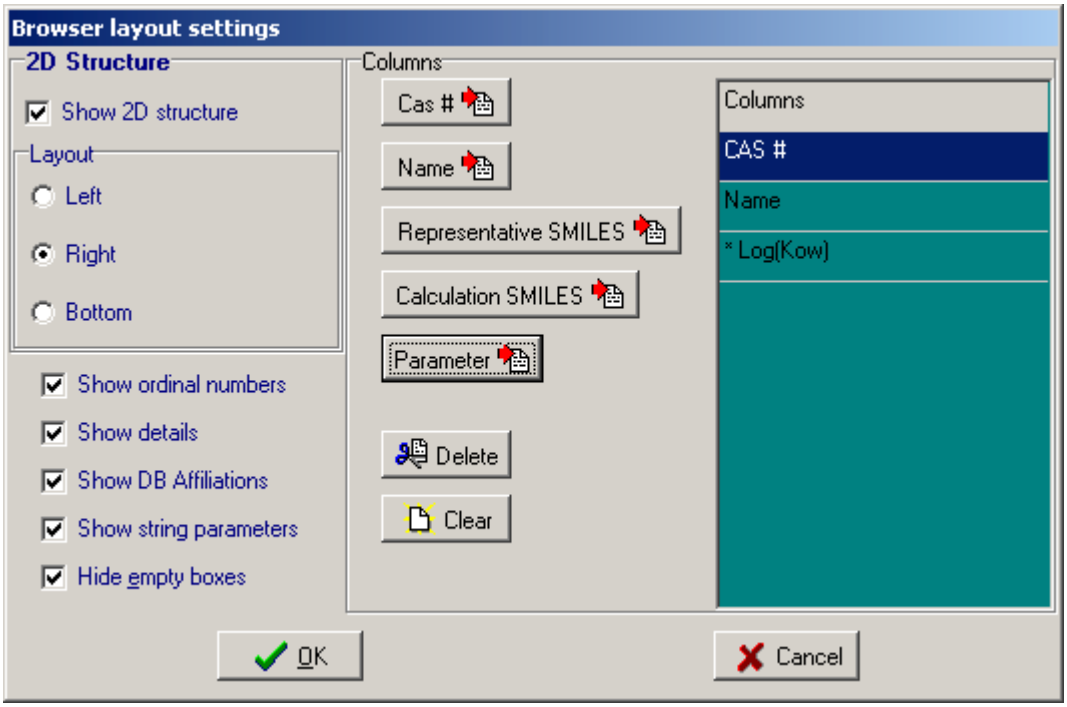
The Global parameters box displays all conformational global parameters for the selected conformer.

The 3D box (with the black background) shows a 3D picture of the selected conformer. The 3D box has the complete functionality known from other OASIS software (rotate, zoom, auto-rotate, local parameters check, atom properties, calculation of distances, angles, and torsion angles).

## Browser configurations

An important feature of the browser is the browser configurator – users can define their own sets of displayed elements, together with their positions.

Select View | Configure browser from the main menu to invoke the Browser configurator dialog:



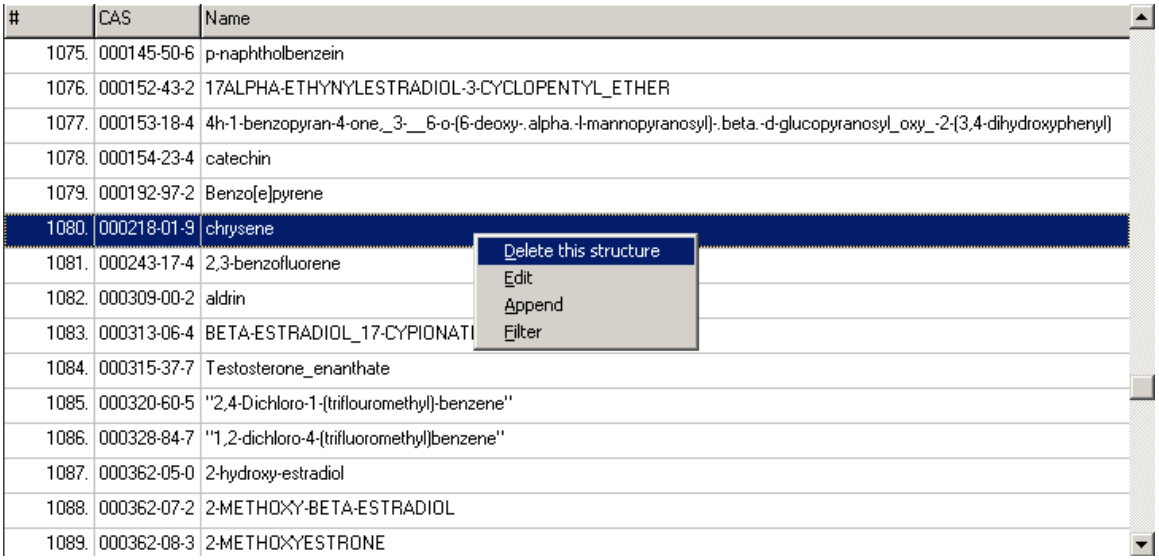

Check the '2D structure' box to toggle the display of the 2D image of the selected structure. Select the position of the 2D structure image by the radio buttons below ('Left', 'Right', 'Bottom').

The rest of checkboxes at the left control the corresponding display elements of the details of the selected structure.

The 'Columns' box controls the display of columns in the main browser grid. In addition to the ordinal number of the structure, you can have CAS, Names, Representative and/or Calculated SMILES strings, as well as any number of structure or conformational parameters. Conformational parameters are displayed as intervals showing the minimum and maximum values for each structure.

## Local menu

Click the right mouse button on any structure in the structures grid to invoke the Browser local menu. This is where you can edit, dele, insert, or filter structures:



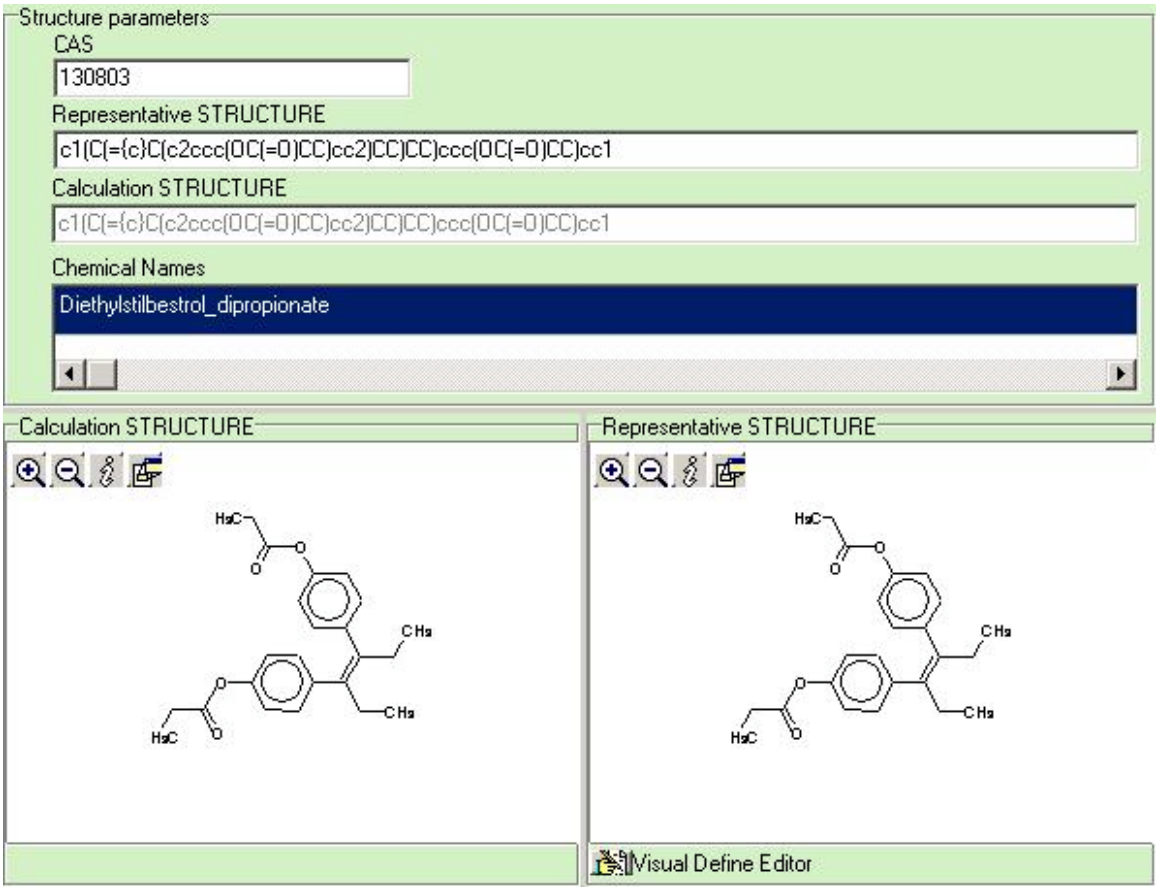

Deletion of structures will prompt you with a message displaying the CAS, Name, and 2D data of the selected structure so that you can confirm or cancel the operation.

Appending and editing structures is covered in the ‘Editor’ section below.

# Structure editor

Structure Editor is used to define or change the properties of a structure. These properties include structure's CAS number, names, 2D data, and parameters.

The left-hand part of the Editor window is where you inspect and modify the basic structure properties:



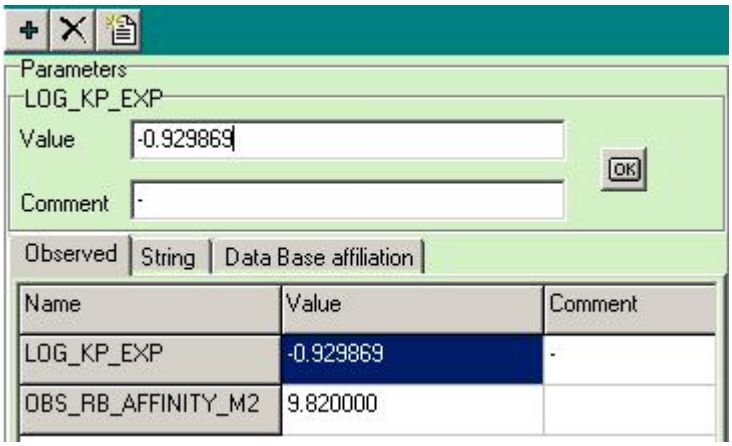

Click on CAS or Chemical Names to view or change the corresponding structure data.

Click on 'Visual Define Editor' to invoke the Visual Structure Definition module where you can draw or edit interactively the 2D representation of the structure.

The right-hand part of the Editor window is the Parameter and Database affiliation editor. Here you can:

- inspect the values of all parameters and database affiliations of the selected structure;
- change their values

- delete parameters values of the selected structure
- add new parameters from the list of existing parameters in the database and define their values for the selected structure
- add entirely new parameters (not existing so far in the database) and define their names, comments, and values for the selected structure



# Search capabilities

Oasis DB Manager can perform a variety of search queries, single, combined, or result-based.

A single search is defined by one search condition. A detailed list of the available types of search conditions is presented in section A below.

Combined search queries contain one or more queries combined with the logical operators AND, OR, or NOT. You can use single as well as other combined queries and obtain a search query of arbitrary complexity and level of nesting.

Result-based queries are executed over the results of a previous search. All requested queries and their results are stored during your DBM session. The database browser windows maintains a list of the requested queries; you can switch between them using their box tabs.
Whenever you start a new search, you will be asked if it is going to be a result-based one.
Result-based searches can use any of the previously completed searches.
Result-based queries can be themselves either single or combined.

A. Single search queries

1. Structure identification search

1.1. CAS search

You can search for CAS numbers equal to, or greater than (or equal to), or less than (or equal to), or different than a specified value, or within a specified range of values.
Alternatively, you can provide more than one value to be matched. Type the list of CAS numbers by hand into an editable list or import it from a text file.

1.2. Name search

Search for structures with names beginning with, or ending with, or containing, or exactly coinciding with a specified text string.
You can provide more than one text string to be matched. Type the list of text strings by hand into an editable list or import it from a text file.

1.3. Database affiliation search

Search for structures found in specified chemical databases. Select chemical databases from the list of available database affiliations.

1.4. SMILES search.

Search for structures having a specific SMILES. This is not the substructure search feature described below; the SMILES search is a simple option of searching for structures having exactly the specified SMILES structure.
You can provide more than one SMILES string. Type the list of SMILES strings by hand into an editable list or import it from a text file.

1.5. Structure quality search

Search for structures having specific quality data. The possible quality values for search are:

- Structure OK
- Incomplete name
- Too many symbols for molecule description
- Complex compound
- Addition compound
- Oligomer or Polymer
- Non-specified structure or Carbon Atoms Number
- No counterion specified
- No structure available
- No CAS available

2. Parameter search

2.1. Global structure parameters

Select 'Structure' from the 'GP Type' list and then a global observed parameter from the list of available global observed parameter names. Search for structures having that parameter equal to, or greater than (or equal to), or less than (or equal to), or different than a specified value, or within a specified range of values. It is also possible to search for structures having this parameter defined (i. e. having any value for this parameter).

2.2. Global structure parameters by models

This options makes it possible to search for parameters participating in models. Check the 'By models' box at the top of the search area.

Then, select a global observed parameter from the list of available model parameters. A list of models for the selected parameter will appear.

Select all or just the desired models either by pointing them directly with the mouse, or using the specially provided Model filter.

The Model filter is a feature for selecting models matching any elements of the model description – Author, Date, Bioassay, Conditions, Organ/Tissue, Administration. The possible values of all these components are listed in individual boxes so that you can form a foltering condition selecting some of them. Click 'Filter' to get the result of the filtering condition – the models matching the specified properties will appear in the model list.

Once you have done specifying models, proceed to specifying the search conditions.

Search for structures having any model value of that parameter equal to, or greater than (or equal to), or less than (or equal to), or different than a specified value, or within a specified range of values. It is also possible to search for structures having any model of this parameter defined (i. e. having any value for this parameter in any of its models).

2.3. Global conformational parameters

Select a global conformational parameter from the list of available global conformational parameter names. Search for structures having that parameter equal to, or greater than (or equal to), or less than (or equal to), or different than a specified value, or within a specified range of values. It is also possible to search for structures having this parameter defined (i. e. having any value for this parameter).

This search can be executed structure-wise or conformer-wise, depending on the selected option in Search / Options from the Main menu.
Professional mode only.

2.4. Local parameters

Select a local parameter from the list of available local parameter names. Search for structures having that parameter equal to, or greater than (or equal to), or less than (or equal to), or different than a specified value.
This search can be executed structure-wise or conformer-wise, depending on the selected option in Search / Options from the Main menu.
Professional mode only.

2.5. Text parameters

Select a text parameter from the list of available text parameter names. Search for structures having that parameter beginning with, or ending with, or containing, or exactly coinciding with a specified text string. It is also possible to search for structures having this parameter defined (i. e. having any value for this parameter).

3. Fixed fragment search

Fixed fragment is a fragment with no varying characteristics or extra conditions. The possible characteristics and conditions that can vary are described in subsection A4 below.
To start a fixed fragment search, either:

- in Basic mode, click ’Fragment’ from the main toolbar
- in Basic or Professional mode, click ’Flexible’ or ’Search’ from the main toolbar and go to ’Fragment’ section.

The Fragment Editor is where you build your fragment specification. Construct a fragment with the editor or just type the corresponding SMILES. Click OK to confirm.
The search performance depends, except on the database size, on the number of results of the search query – the fewer the results, the faster the execution. Therefore, complex searches are executed faster than simple ones.
See general recommendations for improving fragment search performance below.

4. Extended fragment search

Extended search queries allow you to specify additional conditions on the fragment to be found.
The following types of additional conditions are implemented:

- Wildcard atom. To get this, select ’Wildcard’ from the Periodic table dialog (activated when you select an atom type) and mark the corresponding atom site in the Fragment Editor. You can put as many wildcard atoms into a fragment as you wish.
- Enumerated wildcard atom. Create a list of alternatives for an atom and register it with the Enumeration box in the Fragment Editor. Mark the desired site in the fragment with the mouse.

- Local conditions. Conditions involving local parameters can be put over any atom in the fragment. Use the Local condition box to select a local parameter and a range of it desired values.
- Distance. You can search for structures having a specified distance between specified atoms or fragments.
- Atom qualifiers. Atoms in the fragment can be additionally required to fulfill various conditions specified by the so-called qualifiers.
  Select the necessary qualifiers (if necessary, more than one) from the Atom qualifiers box in the Fragment Editor.

Example: To make sure an atom does not participate in a ring, qualify it as 'acy' – acyclic.

To start a fragment search, either:

- in Basic mode, click 'Fragment' from the main toolbar
- in Basic or Professional mode, click 'Flexible' or 'Search' from the main toolbar and go to 'Fragment' section.

The Fragment Editor is where you build your fragment specification. Construct a fragment with the editor or just type the corresponding SMILES. Click OK to confirm.

B. Combined search queries

The Flexible search dialog is where you construct combined queries (you can also build single queries there).

Invoke the corresponding single query constructor(s) to build the single queries that will participate in the search. Use the Negation, Conjunction, or Disjunction buttons to combine these in larger queries.

The logical operators can be applied to single and combined nodes alike.

Negation. Right-click on the node you wish to negate, then click 'NOT'. A new node appears visually linked to the negated node.

Conjunction. Right-click on the nodes you wish to conjugate, then click 'AND'. A new 'AND' node appears visually linked to the combined nodes

Disjunction. Right-click on the nodes you wish to combine, then click 'OR'. A new 'OR' node appears visually linked to the combined nodes

Double-click any node to execute search on it. The node needs not be the topmost node of a tree.

C. General recommendations to achieve best fragment search performance.

Search performance depends, except on the database size, on the number of results of the search query – the fewer the results, the faster the execution. Therefore, complex

searches can be executed faster than simple ones if the result set they return is smaller.

Since non-fragment queries are the fastest type of queries, try to use them in conjunction or as result providers for subsequent fragment queries.

Since fixed fragment queries are faster than extended fragment queries, try to use fixed fragment queries in conjunction or as result providers for subsequent extended fragment searches.

Since combined queries are executed left-to-right, specify the fastest and stronger-filtering queries first.

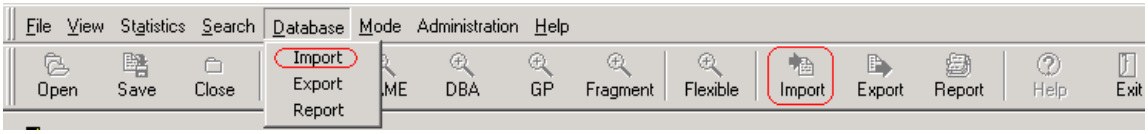
# Data import and export

## Import

Oasis Database Manager provides import of structural (2D and 3D) and field (parameter) data stored in the following types of files:

- **SMI.** SMI files are text files containing information on CAS numbers, chemical name and SMILES 2D notation for each chemical.
- **SDF.** SDF files store CAS numbers, chemical name, 2D and 3D structure data and numeric parameters.
- **MOL -**
- **XYZ –**
- **CMP.** A CMP file is a binary file in OASIS format that stores series of chemicals and all pertinent data describing them. Each logical record of the file describes a separate chemical or a particular conformer of a chemical.
- **DBF.** A DBF is a Dbase database that stores CAS numbers of structures and various numeric and character string parameters.
- **DAT.** This is a text file that includes columns (tab- or blank- delimited) with CAS numbers of structures and various numeric and string parameters.

To import a file, either select *Database* and then *Import* from the main menu or click the *Import* button on the main toolbar:

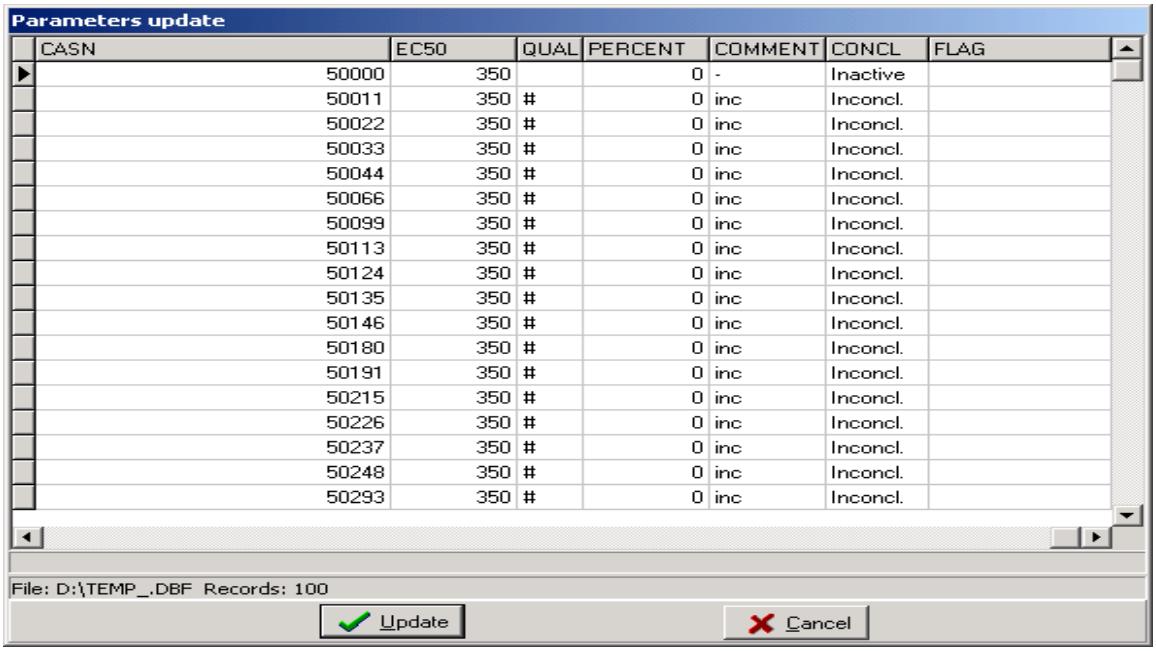



A warning message appears informing you that browsing is temporarily disabled during import. Also, search results from your current session will not be available after making the import.

Click *Yes* to continue.

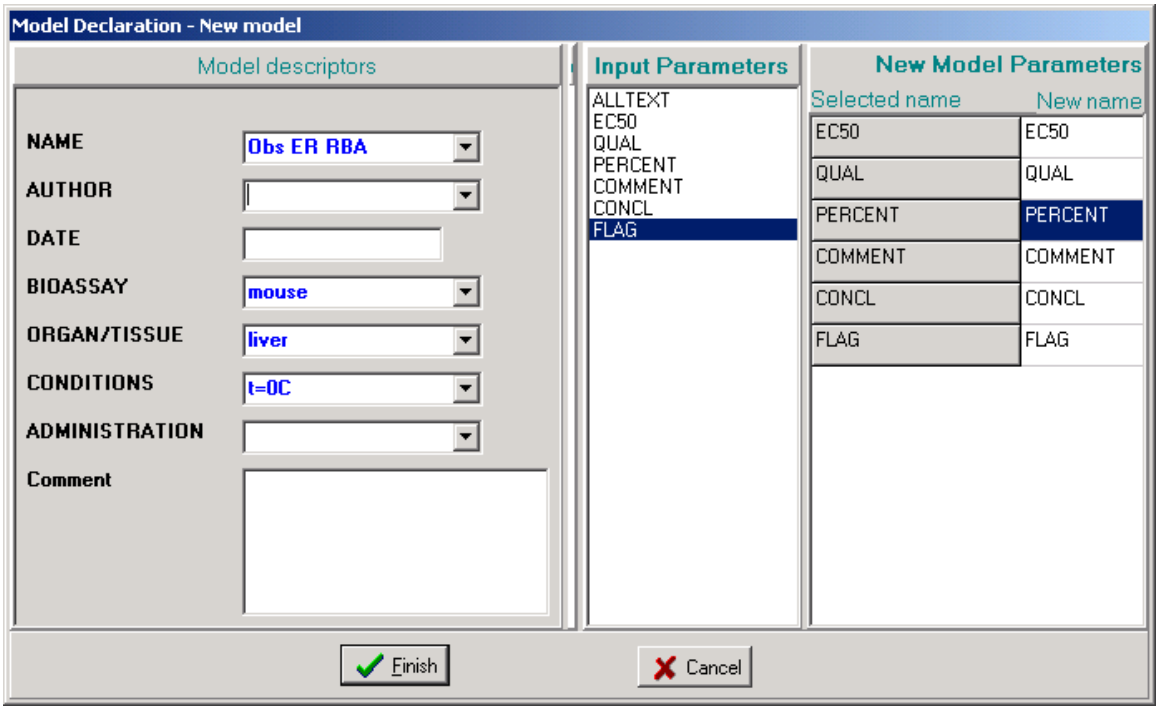
Select your file name and type in the file opening dialog that follows.

The file is opened in a new window so that you may browse through its contents and confirm the import action. Click Update to continue or Cancel to stop.




The Model selection form appears. If you choose Default Model, the parameters will retain the column (field) names from the imported file and will not be assigned to any model. When importing DBF files, it is advised to create a new model for each new DBF file although you may choose not to do so).

If you decide to create a new model, the Model definition window will appear:

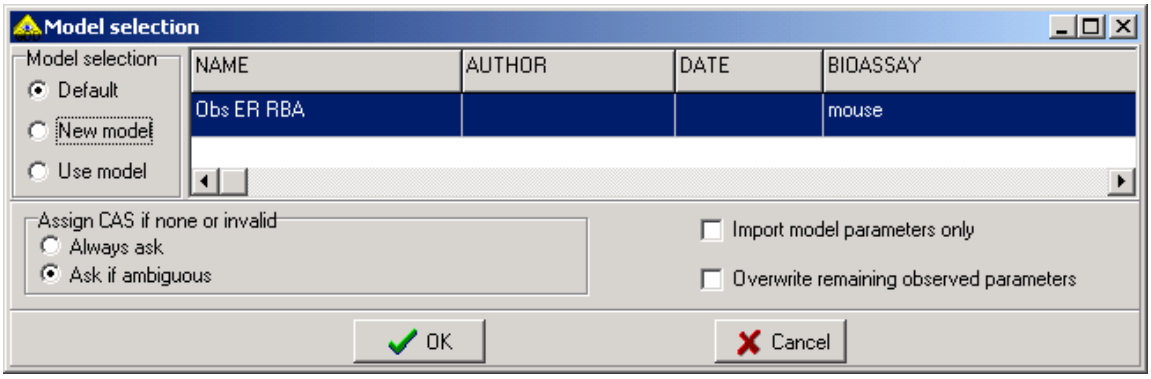


At least one Model descriptor is required. If the combination of model descriptors you have entered specifies an existing model, you will be warned to modify it.
Double-click the Input parameter names that you wish to be imported. Optionally, define new names for them in the New Model Parameters box.

Click Finish when ready with the model definition.

Repeat the procedure if you are aware that the input file contains more models (this is the case with some CMP files).

When you have finished defining models, click OK in the Model selection form to start importing the file:




Note that you may later delete the imported data with the Administration | Delete parameter function.

## 1. Export

The whole database or any subset of it can be exported in the following file formats:

- **Excel.** A tab-delimited text file is generated that contains user-defined combination of CAS number, 2D structure, name, or any collection of parameters (calculated, observed, or conformational).
- **HTML.** An HTML file is generated that contains user-defined combination of CAS number, 2D structure, name, or any collection of parameters (calculated, observed, or conformational). 2D structures can be exported as images. Image files are compressed and appropriate links to them are created.
- **Oasis Database.** Any selection of structures can be saved to a new Oasis database. This saves all possible data associated with the structure – 2D, 3D, DB affiliations, conformer, etc. The database structure is optimized during this process.

# Report generation and usage

In addition to being able to export data into various file formats, Oasis Database Manager offers a powerful subsystem for generation of reports to be viewed on the screen, printed pages, or in turn converted into HTML or other popular formats. The report subsystem is based on the concept of report template – a layout of the page as it would be printed or appear on the screen viewer. This layout comprises the desired data for a single structure. For example, you may wish to produce a report displaying just the CAS, Names, 2D picture and the LogKOW value for each structure. With the Oasis Database Manager Report Designer, you can easily design the corresponding template, save it into a template file and then use it later with any possible collection of structures – e. g. the whole database or some part of it yielded by a search query.

The report subsystem includes a Report Generator, which performs the report generation on the specified subset or the whole database using a given report template, and a Report Designer which makes possible the creation and editing of report templates.


## Report Generator

Click ‘Report’ from the main toolbar of the Oasis Database Manager or select the corresponding item from the main menu of the program. The Report Designer window will appear; select ‘Template’, then ‘Load’ from the menu. When the template loads, Click ‘Generate report’ to produce the report for the currently selected browser window.

## Report Designer

Click ‘Report’ from the main toolbar of the Oasis Database Manager or select the corresponding item from the main menu of the program. The Report Designer window will appear.

A report template is a collection of items, each of which can be of the following types:

- fixed text
- CAS number
- Calculation structure 2D image
- Representative structure 2D image
- Names
- Calculation SMILES string

- Representative SMILES string
- Parameter box
- Table
- Horizontal line

Drag and drop the corresponding components from the toolbar to the design canvas. Click the right mouse button on a component to edit or inspect its current details. For instance, select a parameter box to represent a parameter value and then click the right mouse button on it to specify a particular parameter.

Tables are usually included to represent models or other collections of parameter boxes. Table cells usually contain parameter values of the structure. Tables may have a fixed row and column to represent table headers. The height, width, numbers of rows and columns can be altered at any time.

# Database statistics functions

The Database Statistics provided charts about the quality of chemicals.
Click *Statistics* from the main menu or the *Statistics* button on the main toolbar to work with database statistics.

The following features are included:

- Database affiliation
- Database affiliation unique
- Quantum chemistry quality (MOPAC) by records
- Record quality filter
- Conformer multiplication
- Model correlation
- Parameter distribution (across structure)
- Parameter distribution (across dataset)

# Database properties

## Files

The main file of an Oasis database has the extension '.ODB'. It is the file that represents the database and the one you quote in open/save dialogs.

There may be other files that are required during operation. These files have always the same name as the main database file and may have the following extensions:

| File extension | Description | Required in |
|---|---|---|
| .SDB | Conformational data. | Professional mode |
| .C02, .C03 | Extensions of the conformational part. | If exist, are required in Professional mode |
| .Q02, .Q03,… | Extensions of the main database file (ODB) | If exist, are required in all modes |

The SDB part (and its extension files C02, …) are only required in Professional mode. Basic mode users only need the ODB file (and its extensions .Q01, .Q02, … if ones exist). Note that if you save a selection into a new database when in Basic mode, the conformational part (SDB and its extensions) is not created and such a database can be used only in Basic mode. To create a new database that can be used in both modes, you should be running Professional mode.

Thus, the main file (.ODB) allows for a standalone use (in the Basic mode). The conformer data file(s) are automatically attached and can be dropped if Professional mode is not necessary.

## Limitations

The following table lists the maximum possible numbers of data items. The actual limitations are usually imposed by hardware considerations.

| Item | Maximum number |
|---|---|
| Structures | Unlimited |
| Structure sets | Unlimited |
| Conformers | Unlimited |
| Conformers belonging to one structure set | Unlimited |
| Calculated descriptors (numerical) | 65535 |
| Calculated descriptors (character string) | 65535 |
| Observed descriptors (numerical) | 65535 |
| Observed descriptors (character string) | 65535 |
| Conformational descriptors | 65535 |
| Atomic descriptors | 65535 |

| | |
|---|---|
| SMILES string length, characters | 704 |
| Descriptor name length, characters<br>(all kinds of descriptors) | 192 |
| Chemical name length, characters | 192 |
| Chemical names of a structure | Unlimited |
| Models | Unlimited |

# Hardware & software requirements

## Software requirements

- MS Windows 98/ME, NT, 2000, XP: Oasis database runs on all 32-bit MS Windows platforms. On Windows NT/2000/XP, NTFS file system is recommended for performance reasons.

- Interbase Server & client libraries 6.0.2 (Open Source).

## Hardware requirements

Generally, CPU / RAM requirements depend on the database size, with Professional mode requiring more memory.

For Oasis QSAR Centralized Database, 128 MB RAM are required for Basic mode and 256 MB RAM for Professional mode. While these are the minimum requirements, the system can be user-adjusted for optimum performance if more RAM is present.

The current version of Oasis QSAR Centralized Database requires 1 GB disk space in Basic mode and 20 GB in Professional mode.