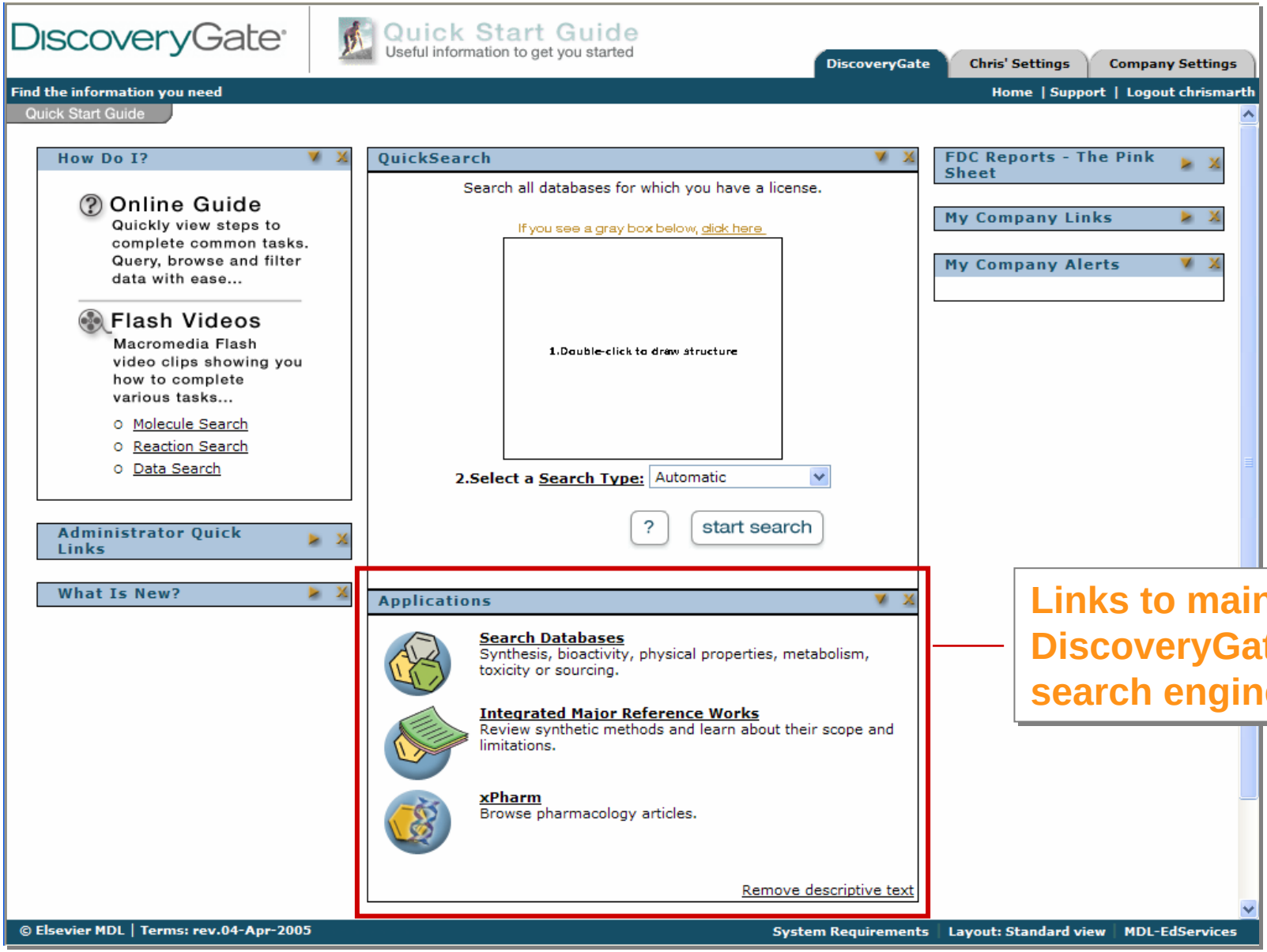
Introduction to DiscoveryGate

Presented to Chulalongkorn University

What is DiscoveryGate?

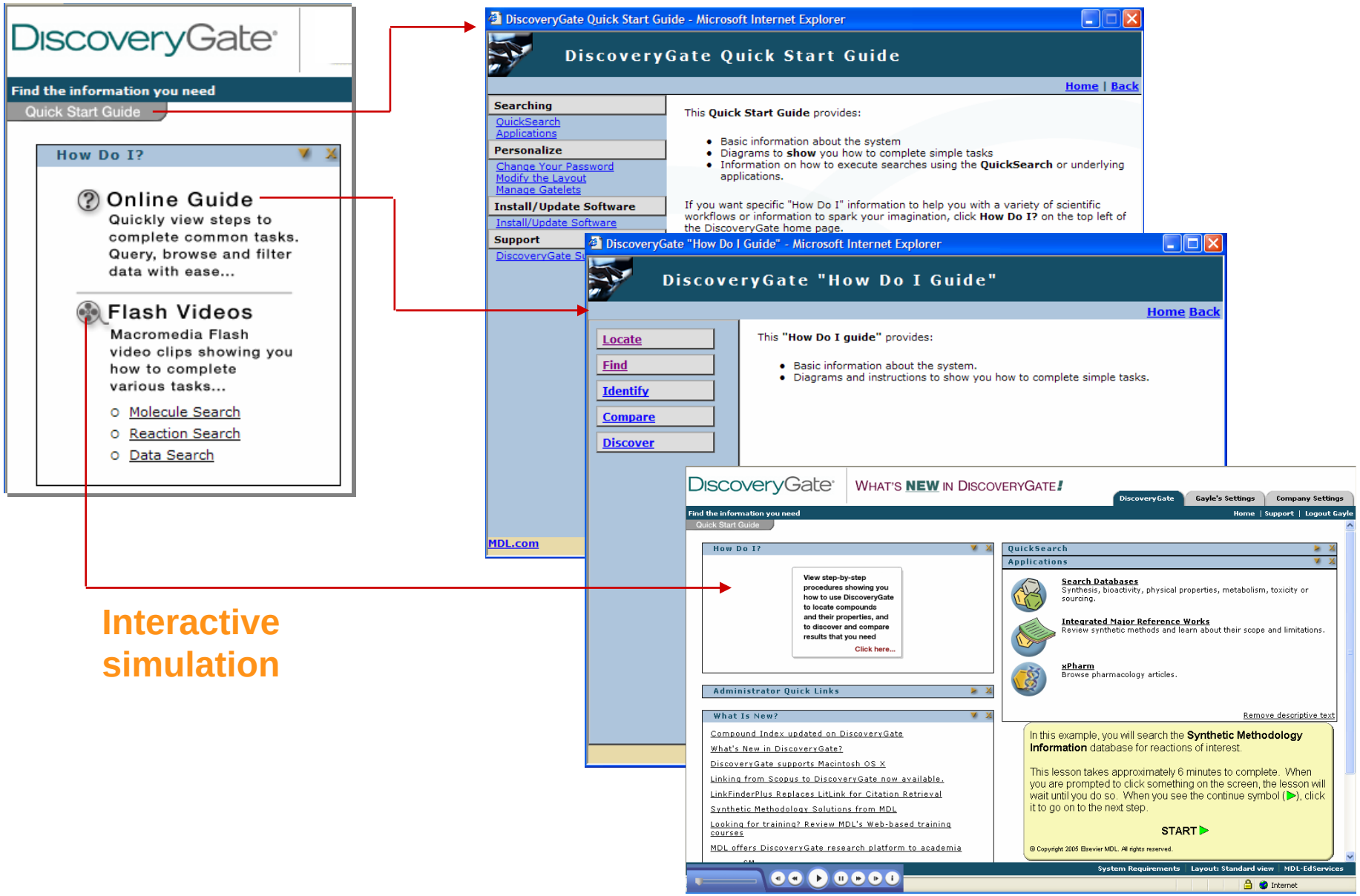
- An internet service that increases the productivity of the drug discovery process.
- Allows one-query access to over 27 million compounds found in synthesis, bioactivity, physical property, and sourcing databases.
- Provides links to authoritative major reference works in synthetic chemistry.
- Provides links to over 15,000 journals and patent archives.

DiscoveryGate home page



Links to main DiscoveryGate search engines

Help for DiscoveryGate



Today's objectives

- Conduct structure and data searches using the MDL Compound Index and individual databases
- View the search results using the Grid, Database, and Properties views
- Save search results and combine lists
- Link to literature
- Create reports
- Modify query forms and save as default forms
- Introduce synthetic methods

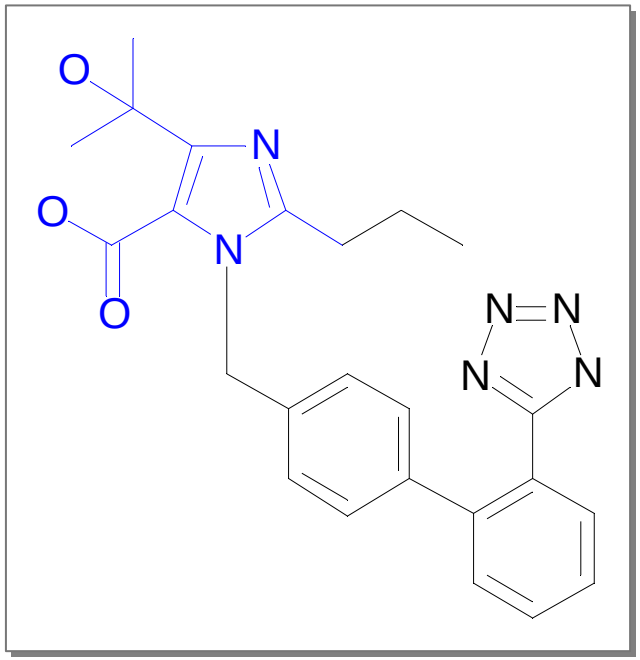
Structure searching

In this section, you will learn how to:

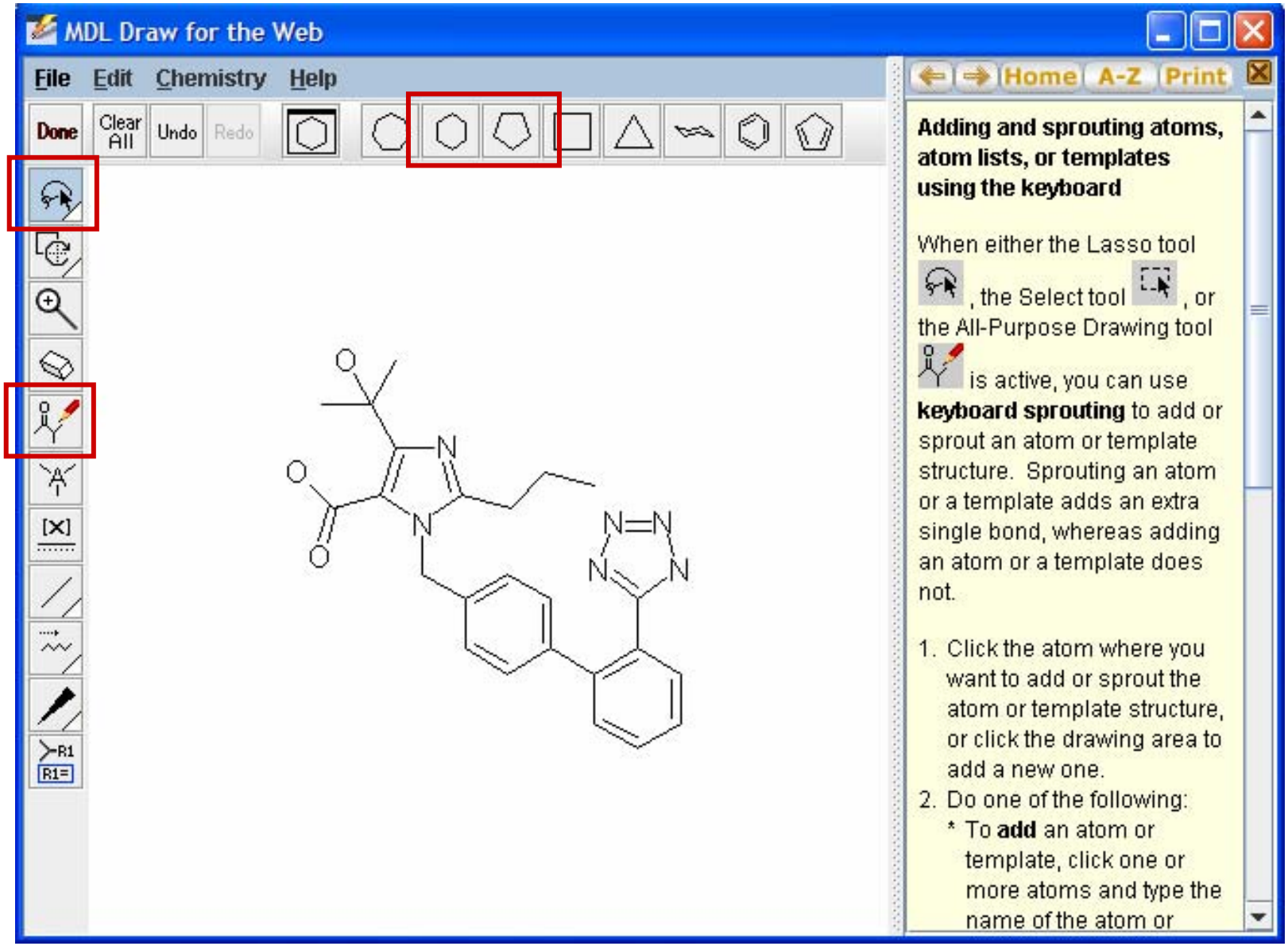
- Conduct an exact match and substructure search using the MDL Compound Index
- View search results using the Grid, Database, and Properties views
- Conduct a combination structure/data search

Search scenario

Olemesartan is an antihypertensive agent, acting as an angiotensin II receptor agonist. You want to retrieve available information for the compound, as well as determine other known 2-alkyl-, 4-carboxy-, 5-alkylhydroxy-imidazole derivatives.

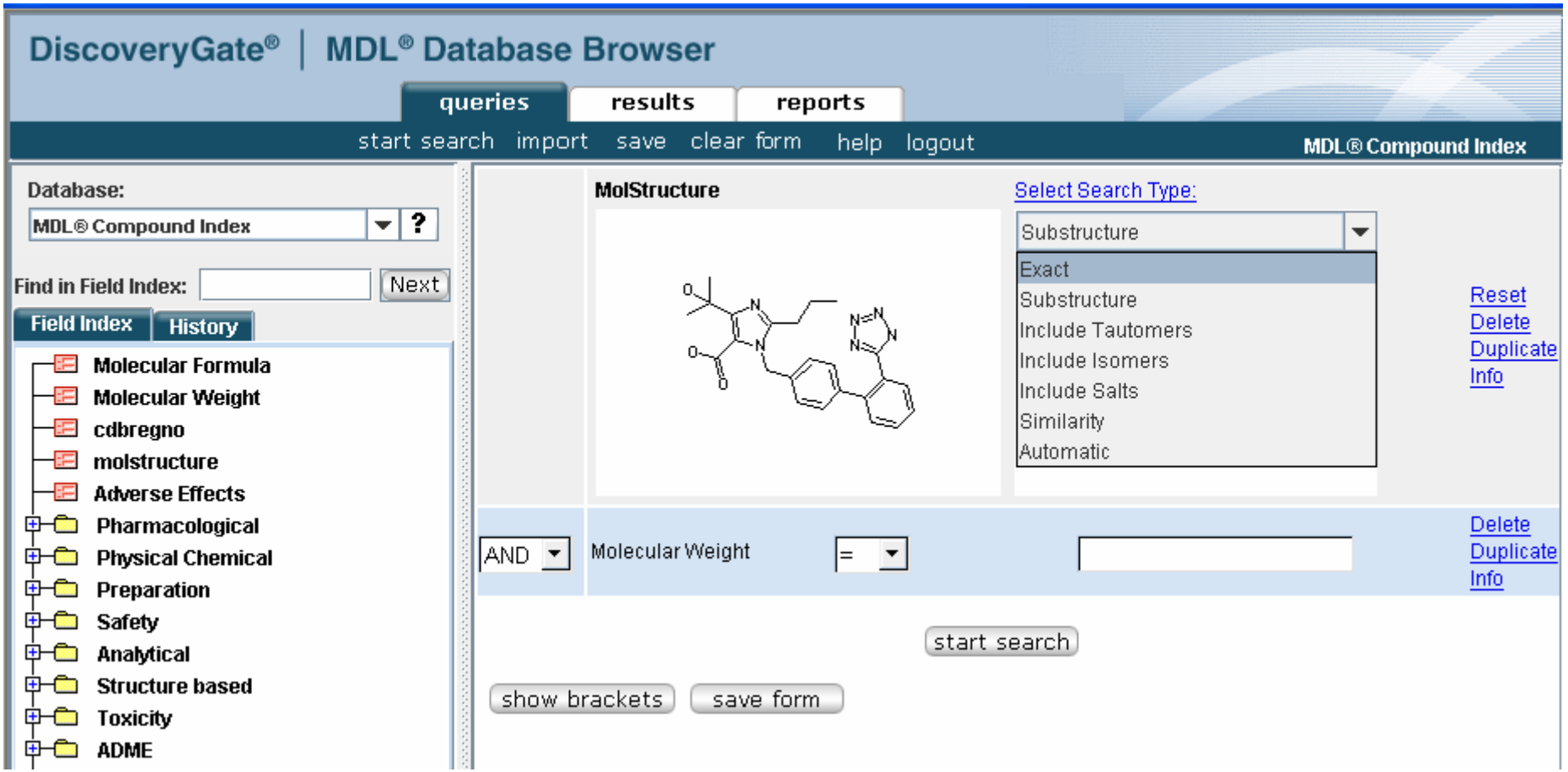


Draw the structure query

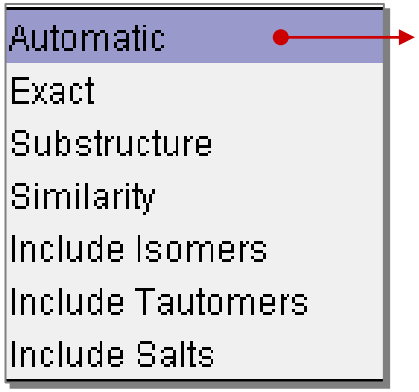


Conduct an exact match search

Compound Index looks at an index of molecules for all databases



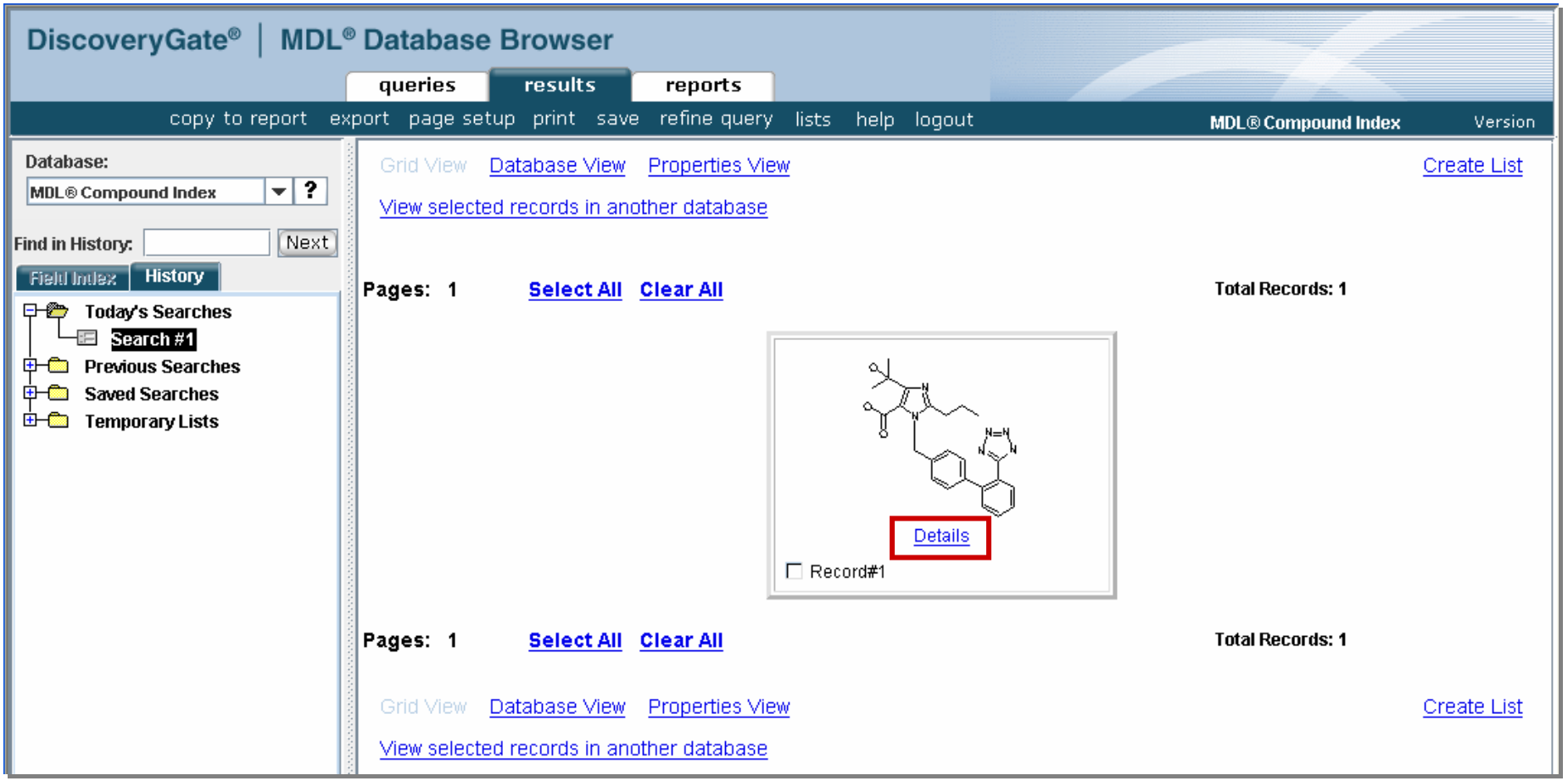
Search types



DiscoveryGate conducts a series of searches until it retrieves at least one compound that matches your query. Each subsequent search is more general than the preceding search. The order is listed below.

- **Exact Match:** Search for substances that match your structural query exactly.
- **Include Isomers:** Search for all stereoisomers (all R and S configurations, all E and Z configurations, and so on).
- **Include Tautomers:** Search for all tautomers, and also for structures with different charges and different isotopes.
- **Include Salts:** Search for substances that include different salts.
- **Substructure:** Search for substances that contain your query wholly embedded within them.
- **Similarity:** Search for substances that are structurally similar to your query.

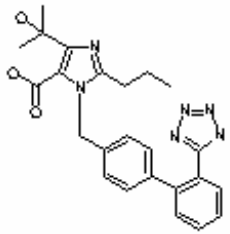
View the retrieved compound



View the detailed information

Return to Search Results View selected records in another database Record # 1 **Total Records:** 1

Also found in: ACD CMC DWPI Index Chemicus MDDR Metabolite Patent Chemistry



Use as Query

☐ Select current record

Torsional degree of freedom	5
clogP (Syracuse Res. Center)	3.63
Molecular Formula	C24 H26 N6 O3
Mol. Weight/Largest Fragment	446.508
Proton Acceptors (O+N)	9
Molecular Weight	446.508
Total Polar Surface Area	129.81
Rule of Five	0
Proton Donors	3

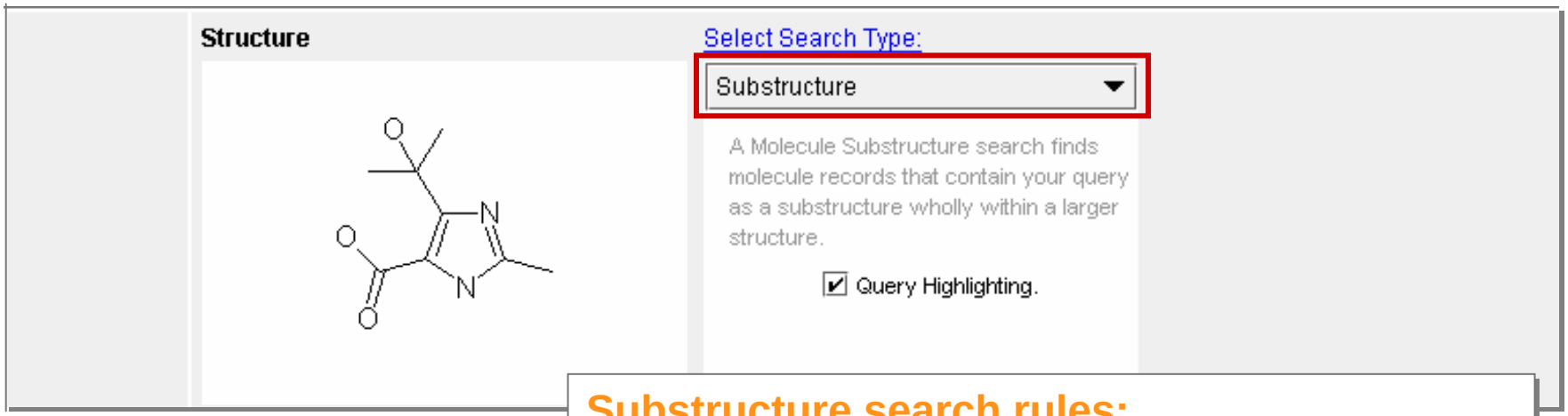
Return to Search Results View selected records in another database Record # 1 **Total Records:** 1

Also found in: ACD CMC DWPI Index Chemicus MDDR Metabolite Patent Chemistry

Information for this compound can also be found in these databases.

Conduct a substructure search

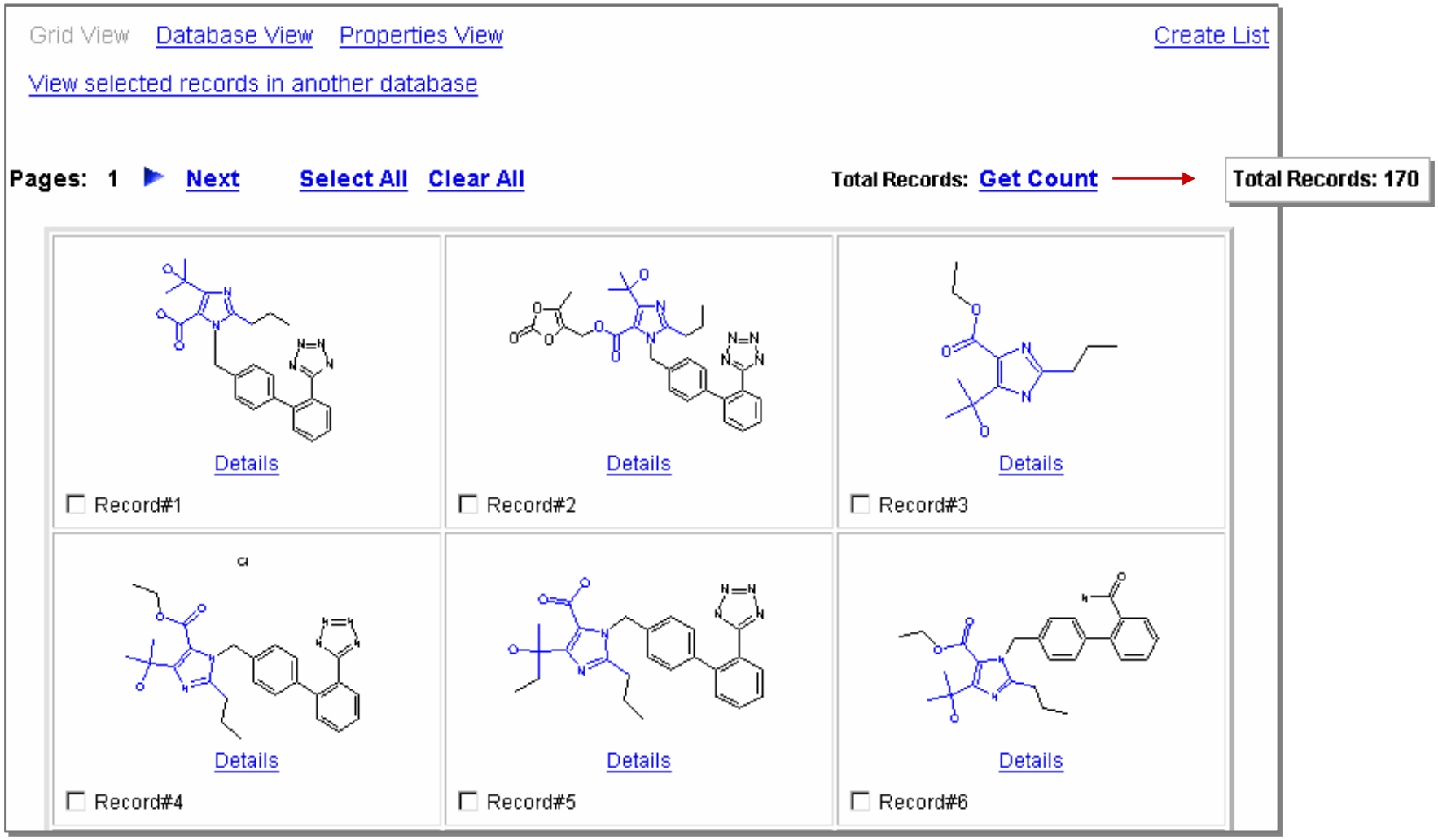
Determine other known 2-alkyl-, 4-carboxy-, 5-alkylhydroxy-imidazole derivatives.



Substructure search rules:

- **The core structure must be embedded in all compounds retrieve**
- **Substitution can occur at any open valence**

Substructure results – Grid View

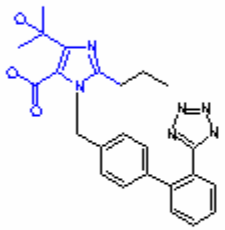
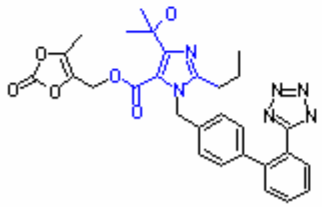


Database View

Grid View | Database View | Properties View | Create List

View selected records in another database

Pages: 1 2 3 4 5 6 7 8 9 10 ▶ **Next** **Next 10** **Select All** **Clear All** **Total Records: 158**

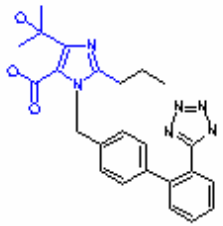
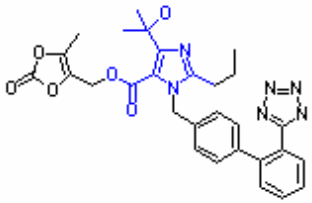
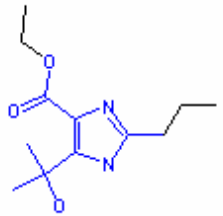
1	 Record#1	ACD, CMC, DWPI, Index Chemicus, MDDR, Metabolite, Patent Chemistry
2	 Record#2	ACD, DWPI, Index Chemicus, MDDR, Metabolite

Properties View

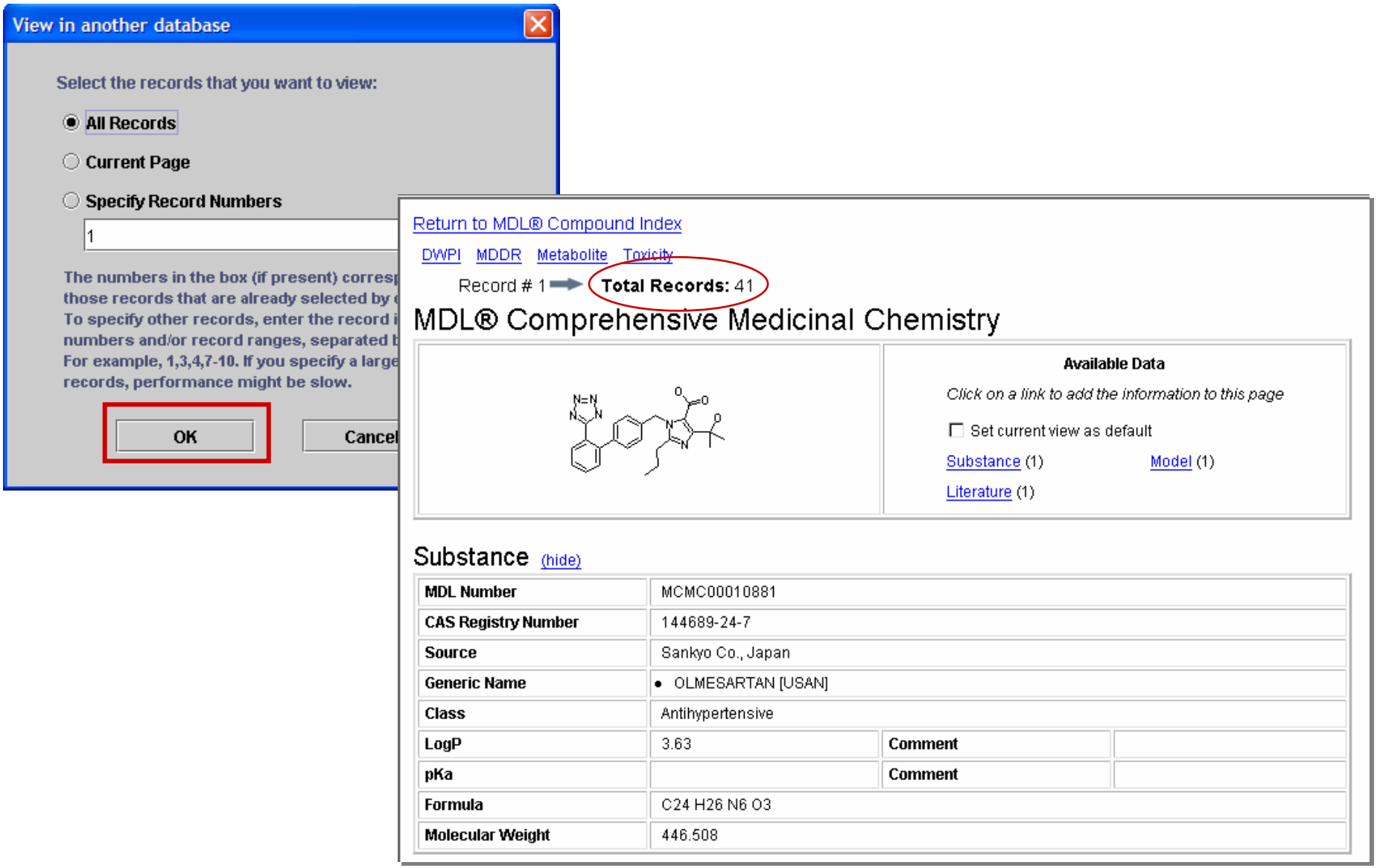
Grid View | Database View | Properties View | Create List

View selected records in another database

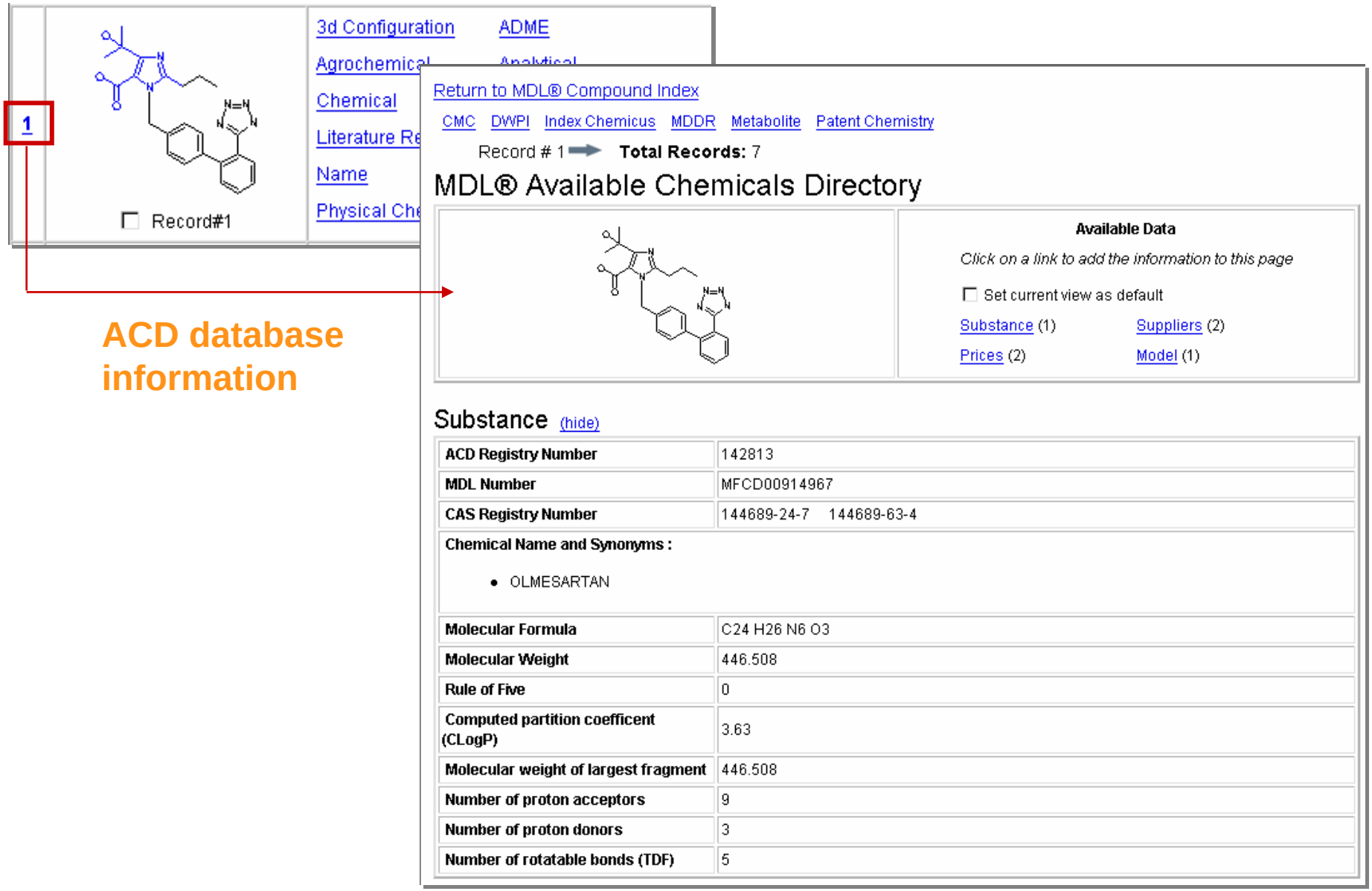
Pages: 1 2 3 4 5 6 7 8 9 10 ▶ **Next** **Next 10** **Select All** **Clear All** **Total Records: 170**

1	 Record#1	3d Configuration, ADME, Agrochemical, Analytical, Chemical, Commercially Available, Literature Reference, Metabolism, Name, **Pharmacological**, Physical Chemical, Preparation
2	 Record#2	3d Configuration, ADME, Agrochemical, Chemical, Commercially Available, Literature Reference, Metabolism, Name, Pharmacological, Preparation, Toxicity
3	 Record#3	Chemical, Literature Reference, Name

Viewing compounds by property



Viewing data for a single record



ACD database information

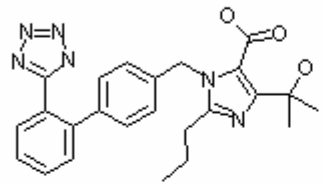
CMC

Return to MDL® Compound Index

ACD DWPI Index Chemicus MDDR Metabolite Patent Chemistry

Record # 2 **Total Records:** 7

MDL® Comprehensive Medicinal Chemistry



Available Data

Click on a link to add the information to this page

Set current view as default

Substance (1) Model (1)

Literature (1)

Substance (hide)

MDL Number	MCMC00010881		
CAS Registry Number	144689-24-7		
Source	Sankyo Co., Japan		
Generic Name	• OLMESARTAN [USAN]		
Class	Antihypertensive		
LogP	3.63	**Comment**	
pKa		**Comment**	
Formula	C24 H26 N6 O3		
Molecular Weight	446.508		

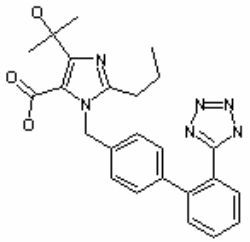
DWPI

Return to MDL® Compound Index

ACD CMC Index Chemicus MDDR Metabolite Patent Chemistry

Record # 3 **Total Records:** 7

Derwent World Patents Index®



DCR:	121764-0-0-0
Name:	OLMESARTAN
Molecular Formula:	C24 H26 N6 O3
Molecular Weight:	446.508
Description:	The active form of CS-866
Description:	5-(1-Hydroxy-1-methyl-ethyl)-2-propyl-3-[2'-(1H-tetrazol-5-yl)-biphenyl-4-ylmethyl]-3H-imidazole-4-carboxylic acid
No. of refs:	29

Patent references in Derwent World Patents Index®

2003-015683	New method for determining and utilizing the circulating blood of a living being over a range of shear rates, useful for e.g. the treatment of peripheral arterial disease

Derwent Classification

(A96) Medical, dental, veterinary, cosmetic.

(B07) General - tablets, dispensers, catheters (excluding drainage and angioplasty), encapsulation etc, but not systems for administration of blood or saline or IV feeding etc.

(C07) Apparatus, formulation, general. including veterinary syringes, general formulations where the active compound is not central to the invention (e.g. wettable powders) and analysis.

Compound Index Classification

Agrochemical

Analytical > Separation Methods

Literature Reference > Patent

Literature Reference > Patent > Analysis and Purification

Literature Reference > Patent > Pharmaceutical/Agrochemical

Literature Reference > Patent > Therapeutic Effect

Pharmacological

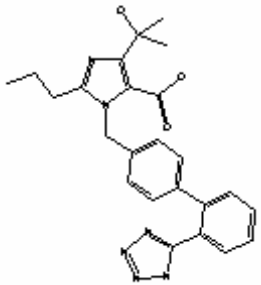
Index Chemicus

Return to MDL® Compound Index

ACD CMC DWPI MDDR Metabolite Patent Chemistry

Record # 4 **Total Records:** 7

Index Chemicus®



Compound Number:	538321-072
Molecular Formula:	C24 H26 N6 O3
Molecular Weight:	446.513

ISI Index Chemicus(R) citation details	
Abstract Number:	538321
Compounds abstracted:	126

Title
NONPEPTIDE ANGIOTENSIN-II RECEPTOR ANTAGONISTS - SYNTHESIS, BIOLOGICAL-ACTIVITIES, AND STRUCTURE - ACTIVITY RELATIONSHIPS OF IMIDAZOLE-5-CARBOXYLIC ACIDS BEARING ALKYL, ALKENYL, AND HYDROXYALKYL SUBSTITUENTS AT THE 4-POSITION AND THEIR RELATED-COMPOUNDS.

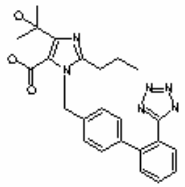
MDDR

Return to MDL® Compound Index

ACD CMC DWPI Index Chemicus Metabolite Patent Chemistry

Record # 5 **Total Records:** 7

MDL® Drug Data Report



Available Data

Click on a link to add the information to this page

☐ Set current view as default

Substance (1) Model (1)

Biology (1) Identification (1)

Literature and Patent (1)

Substance (hide)

External Registration Number	188868
Prous Entry Number	188868
Preview Number	
CAS Registry Number	144689-24-7
Derivative	

Chemical Name :

- 4-(1-HYDROXY-1-METHYLETHYL)-2-PROPYL-1-[2'-(1H-TETRAZOL-5-YL)BIPHENYL-4-YLMETHYL]IMIDAZOLE-5-CARBOXYLIC ACID

Generic Name	OLMESARTAN < USAN >
Formula	C24 H26 N6 O3
Molecular Weight	446.508
Development Phase:	Launched
Year	
Active Investigation	

Index	Activity
16000	ANTIGLAUCOMA
31000	ANTIHYPERTENSIVE
31432	ANGIOTENSIN II AT1 ANTAGONIST

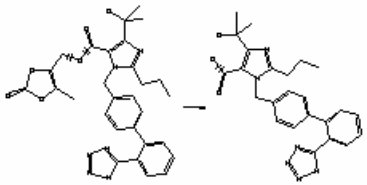
Metabolite

Return to MDL® Compound Index

ACD CMC DWPI Index Chemicus MDDR Patent Chemistry

Record # 6 **Total Records:** 7

MDL® Metabolite Database



Available Data

Click on a link to add the information to this page

Set current view as default

View results as **transformations:**

Transformation (1)
Reference (1)
Species (1)
Enzyme (1)

View results as **molecules:**

Parent (1)
Species (1)
Substrate (1)
Metabolite (1)

Transformation Results -- Transformation (hide)

MDL number	RMTB00058294
Path	MTB9122-A
Step	1 Step
Scheme	MTB9122

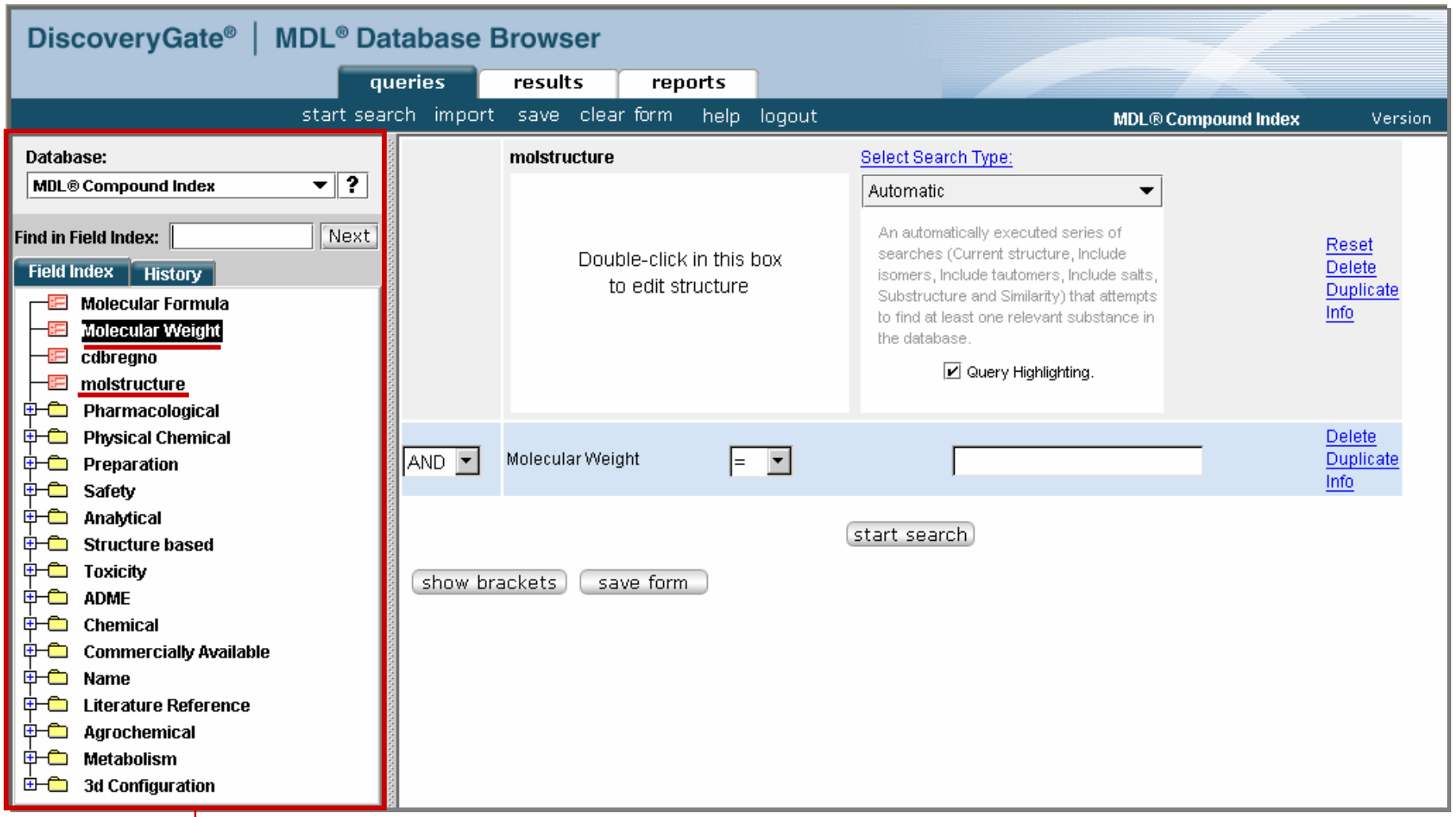
Chemical Name and Synonyms :

- CS-866
- (5-Methyl-2-oxo-1,3-dioxol-4-yl)methyl 4-(hydroxy-1-methylethyl)-2-propyl-1-((2'-(1H-tetrazol-5-yl)-1,1'-biphenyl-4-yl)methyl)-1H-imidazole-5-carboxylate

Reaction Class :

- Hydrolysis
- O-Dealkylation

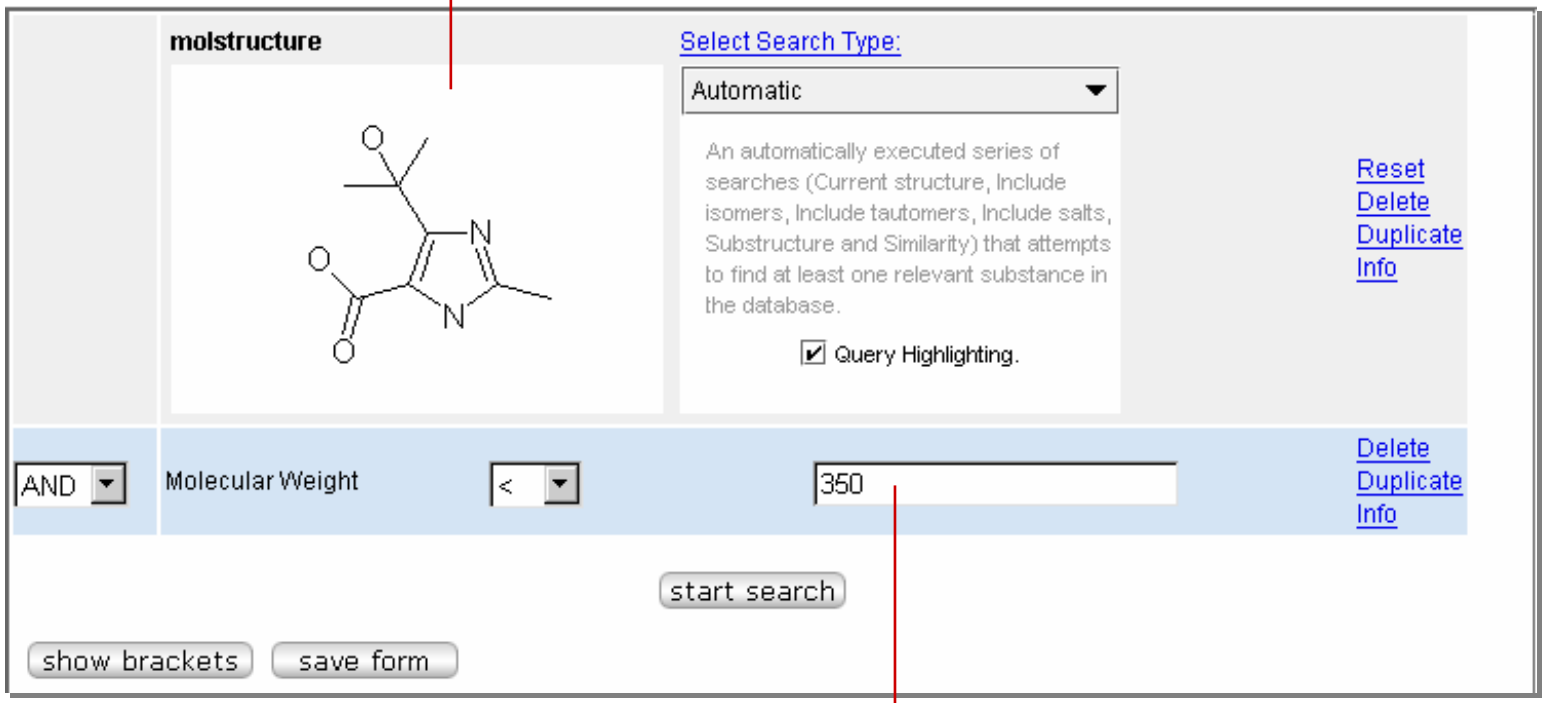
Choose data fields



Double-click a data field to create a custom form.

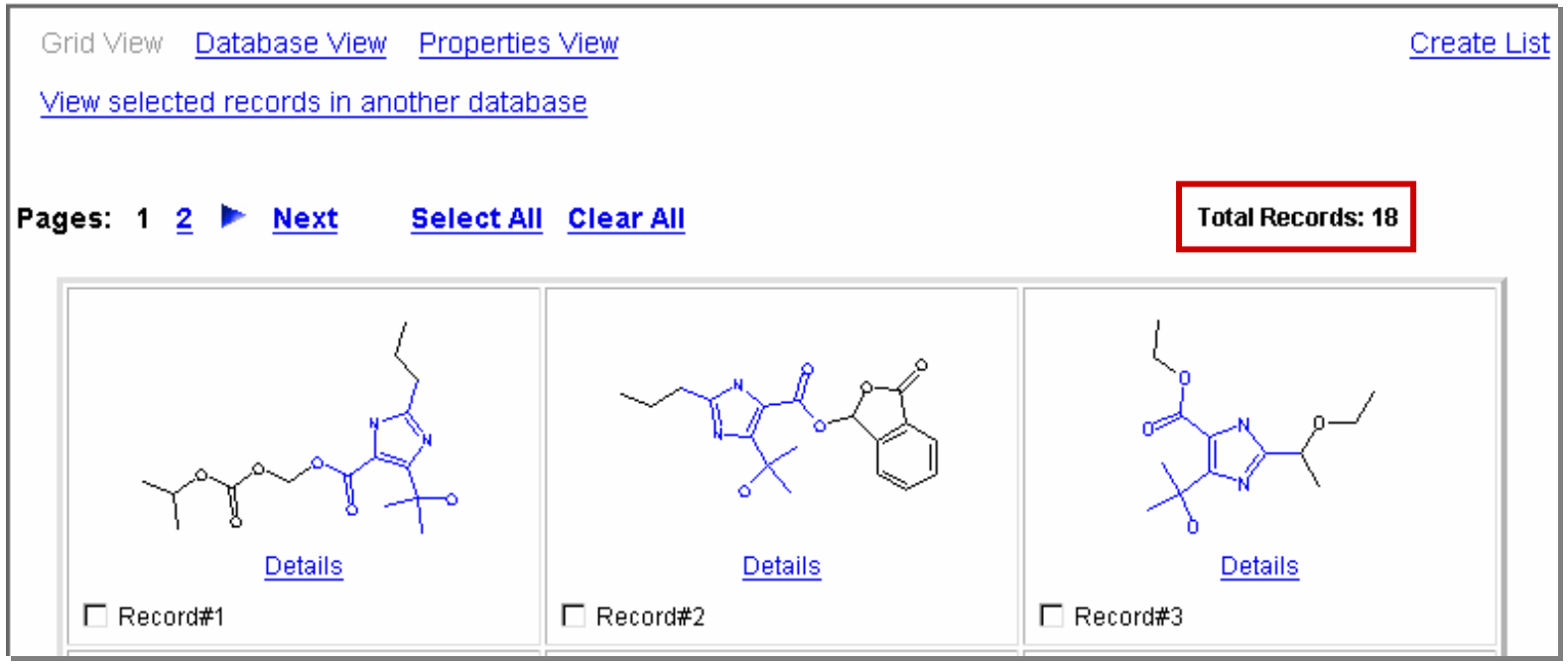
Structure/Data search

Enter the core structural fragment



Enter data and specify an operator

Structure/Data results



Substructure search results – 170 compounds.

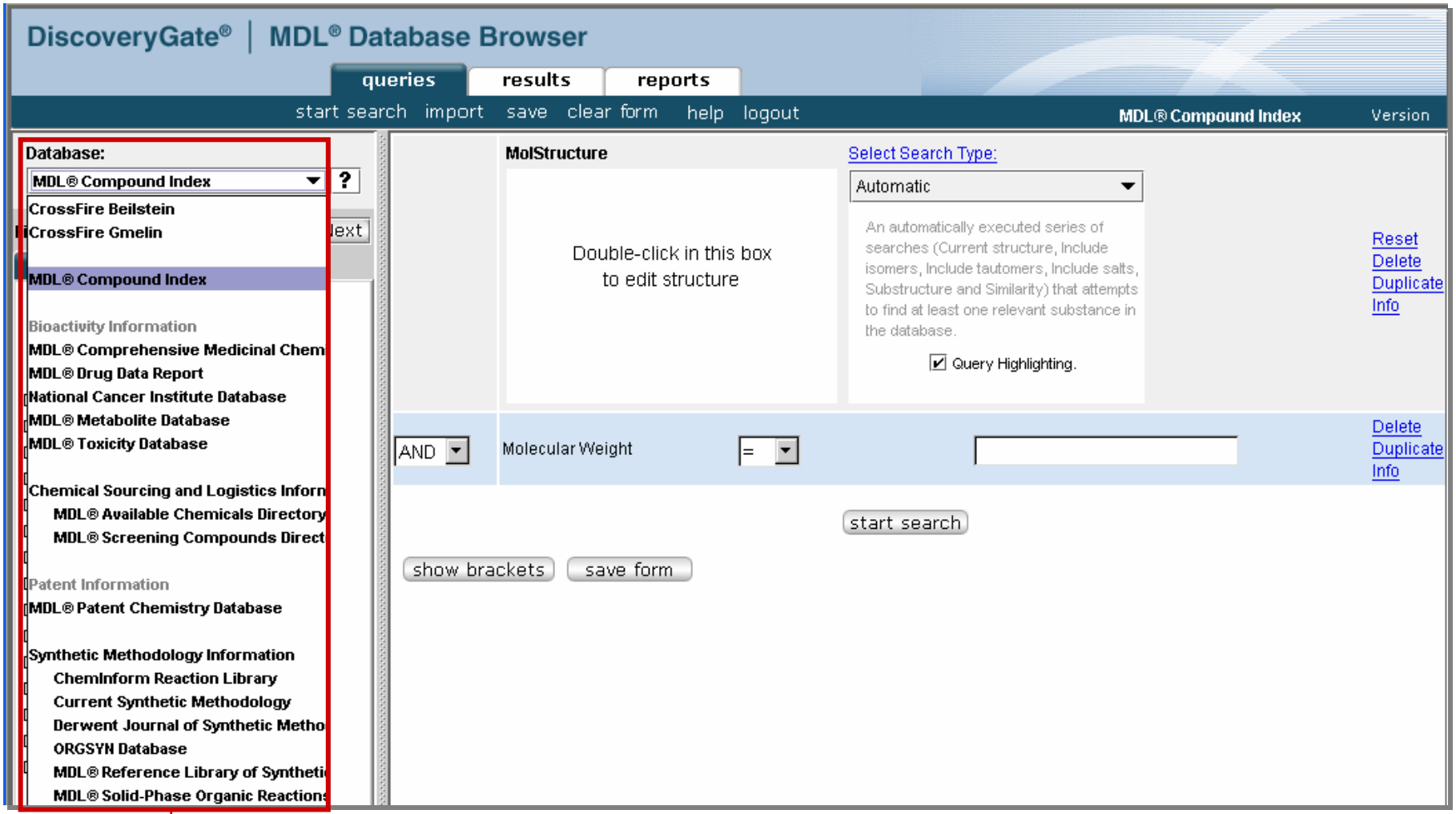
Substructure search with data constraint results – 18 compounds.

Data searching

In this section, you will learn how to:

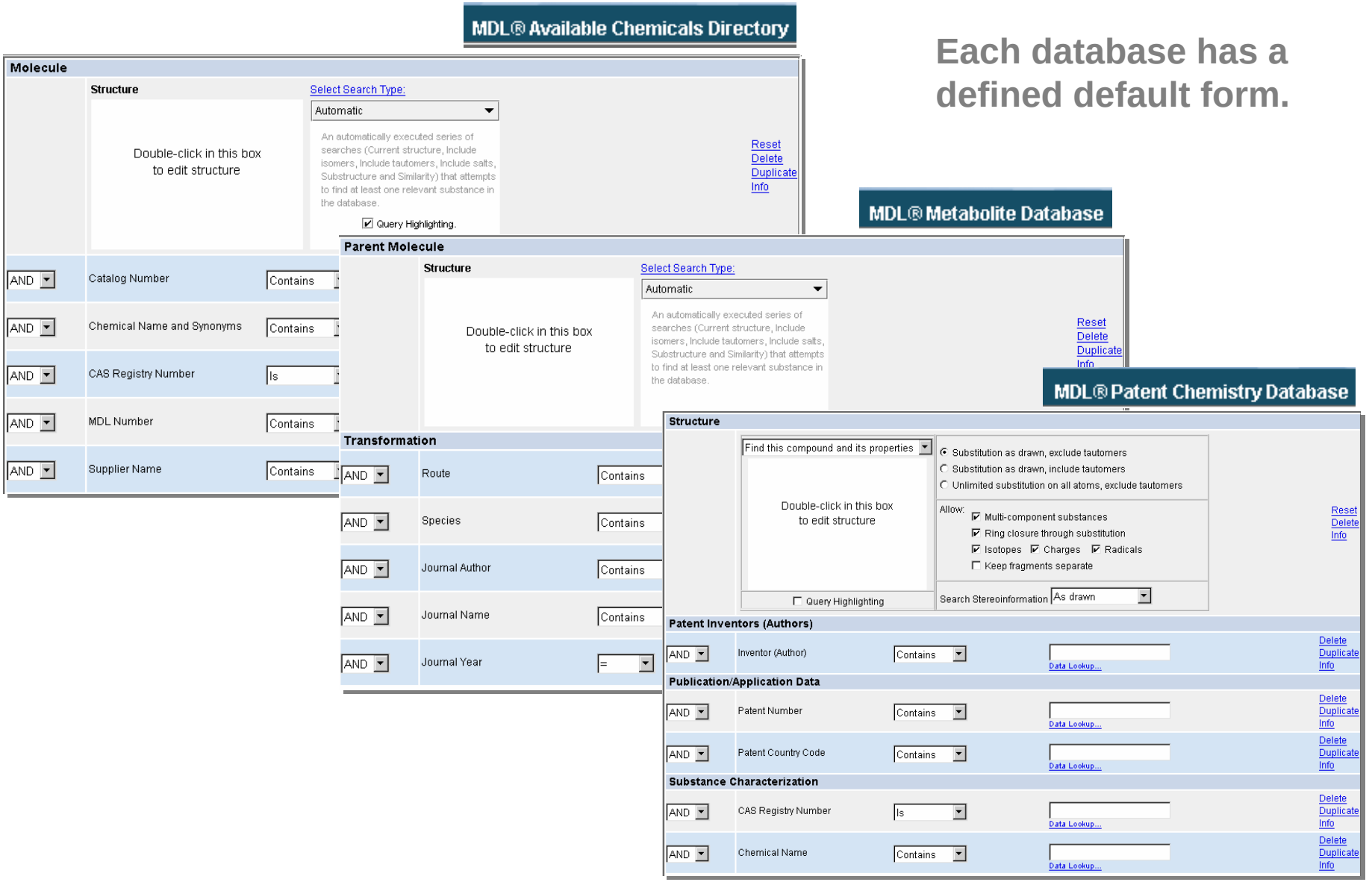
- Search individual databases
- Select the data fields that you want to search
- Use the Data Lookup table to build your query
- Save the modified query form as your default form for the current database
- Use predefined search forms
- Display hidden data in Details view
- Create a report

Selecting an individual database



Select an individual database from the drop-down list

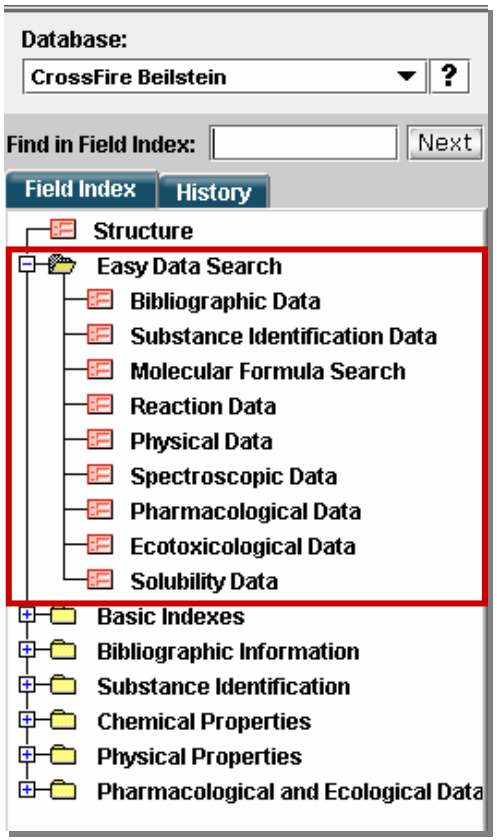
Default forms

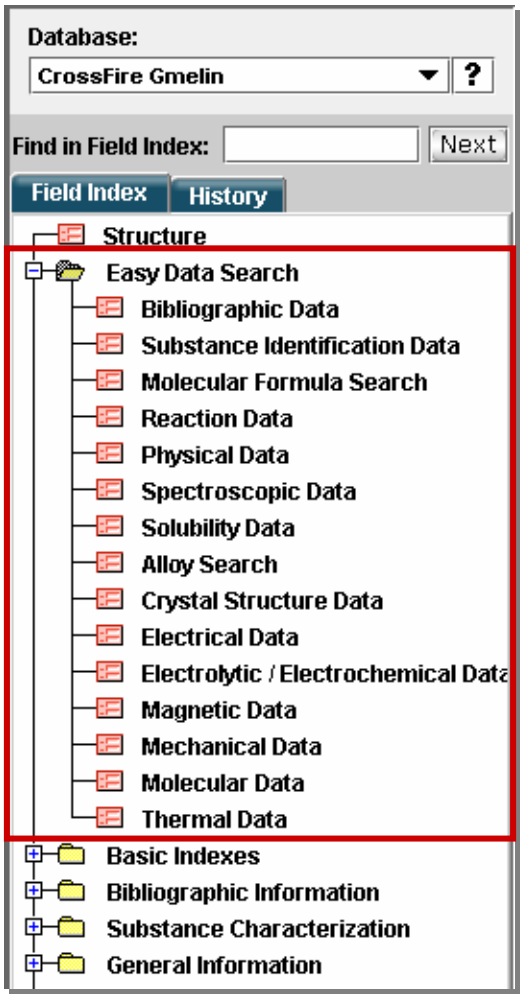


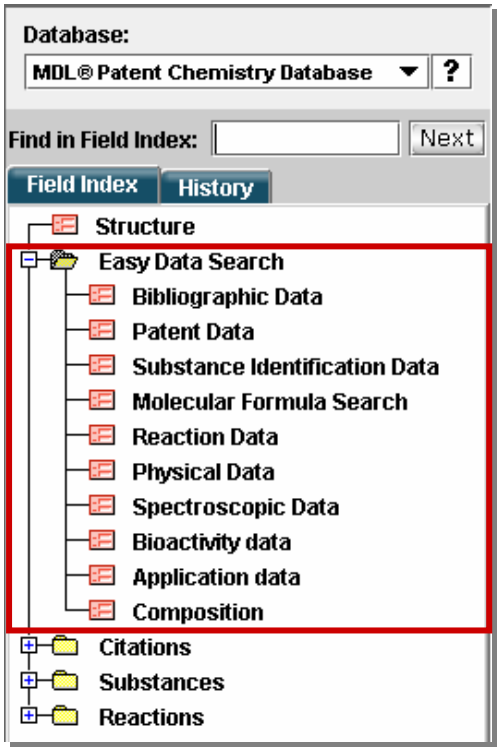
Each database has a defined default form.

Predefined search forms

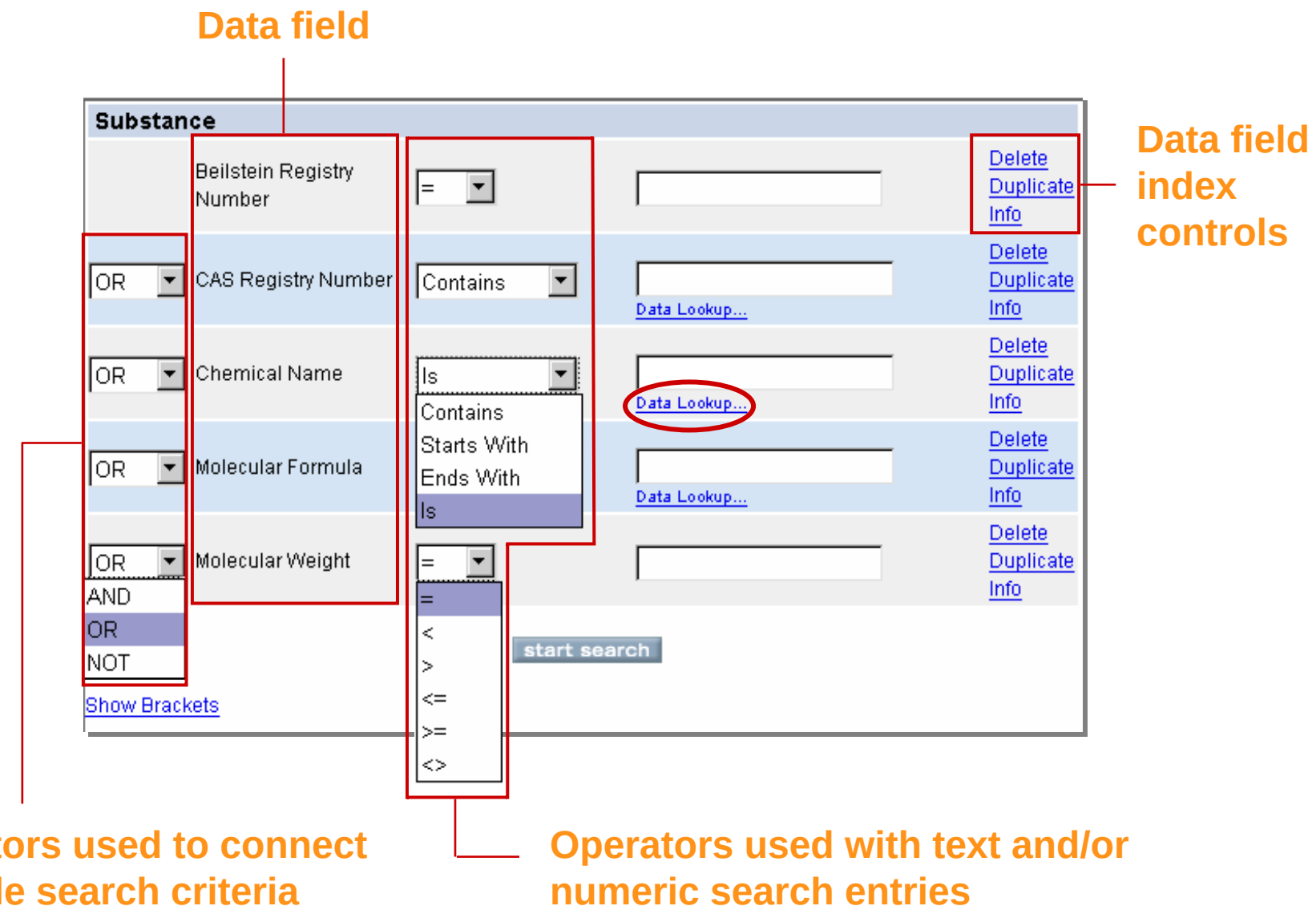
In addition to the default forms, the CrossFire Beilstein, CrossFire Gmelin, and the MDL Patent Chemistry databases have a series of predefined forms.







Data search features



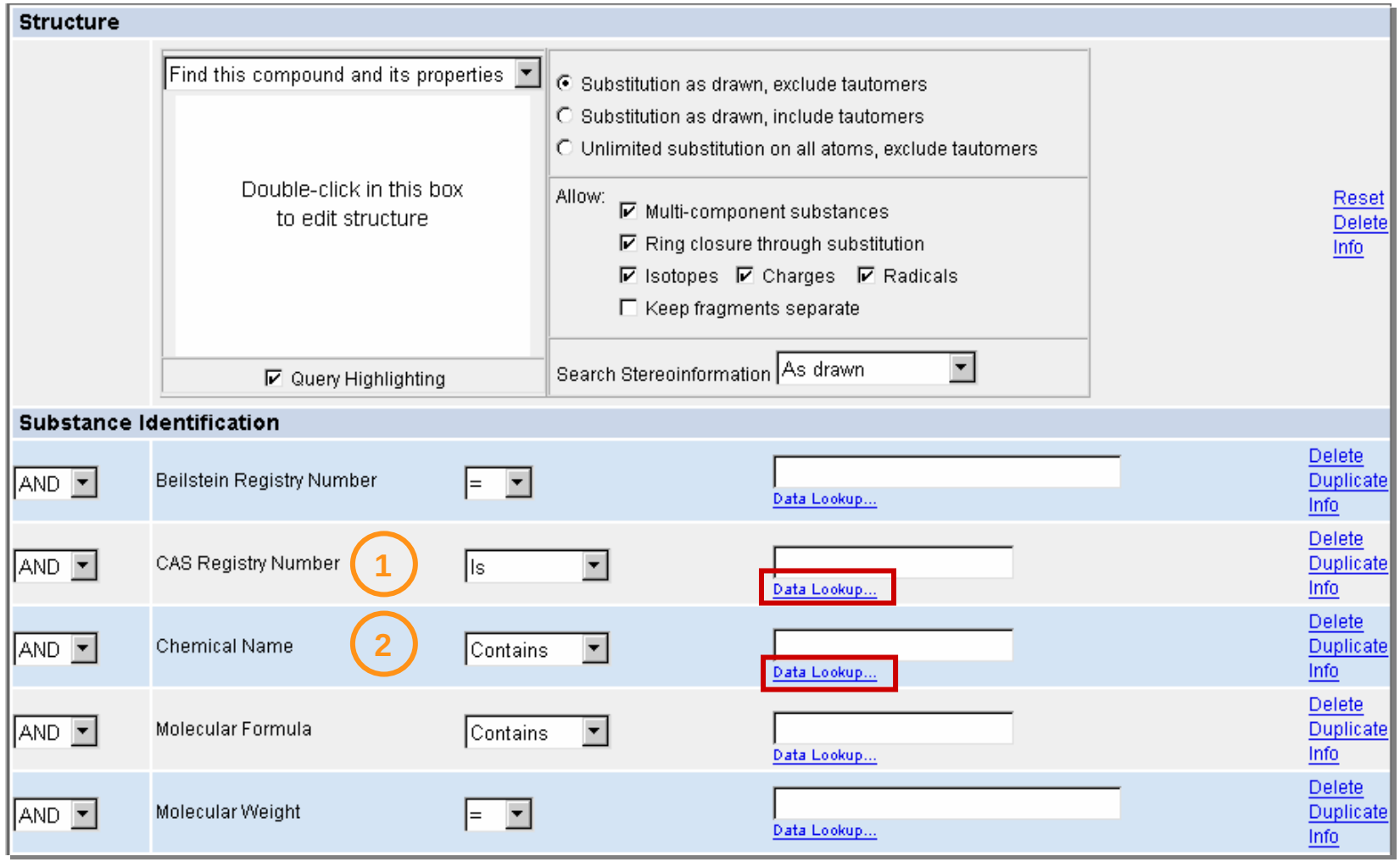
Data operators and wildcards

Operators	Use to retrieve records that:
AND	satisfy both criteria
OR	satisfy either criterion
NOT	do not satisfy the criterion
=	satisfy the exact entry
<	are less than the specified entry
<=	are less than or equal to the specified entry
>	are greater that the specified entry
>=	are greater than or equal to the specified entry
<>	are not included as an entry
Is	have the exact entry
Starts with	begin with the entry
Ends with	end with the entry
Contains	have the term contained within

Wildcards	Use in text string to represent
?	One character
??	Two characters

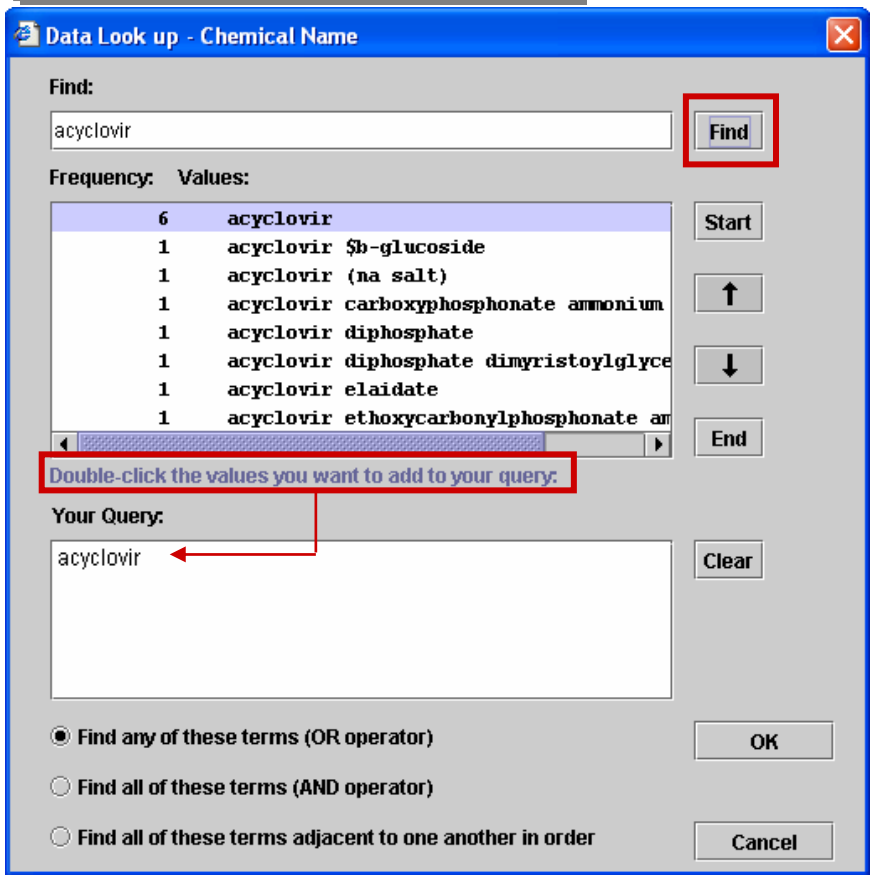
Search for a specific compound

Conduct two different data searches to retrieve the compound acyclovir (CAS number 59277-89-3) from the CrossFire Beilstein database.

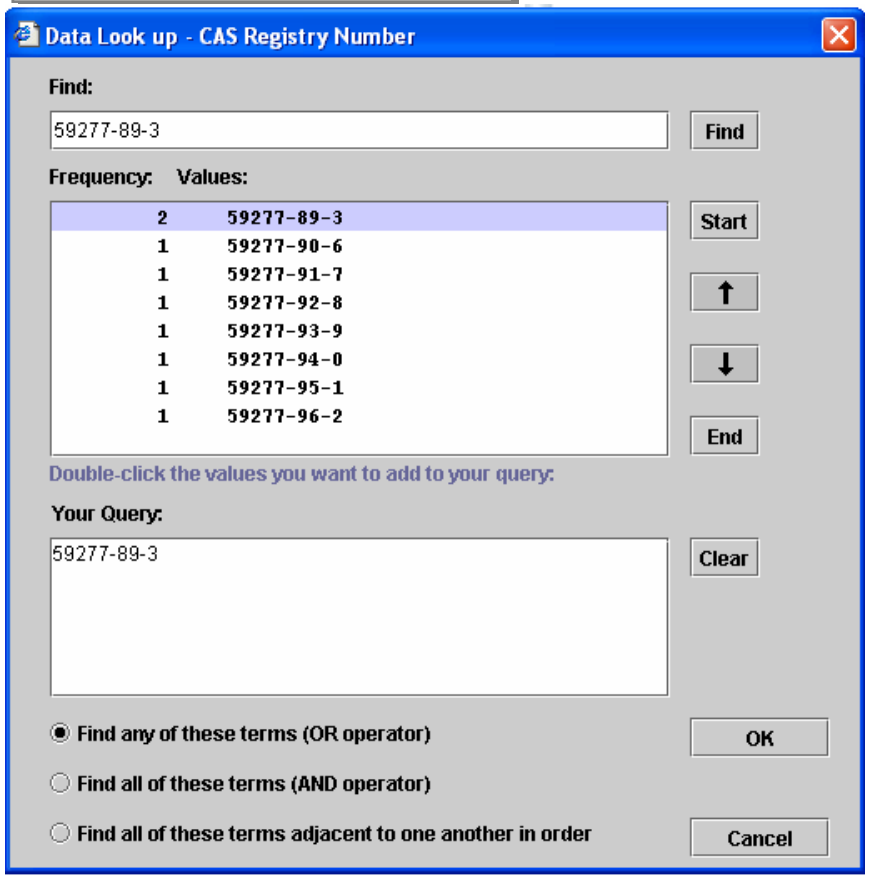


Data Lookup

Chemical Name search



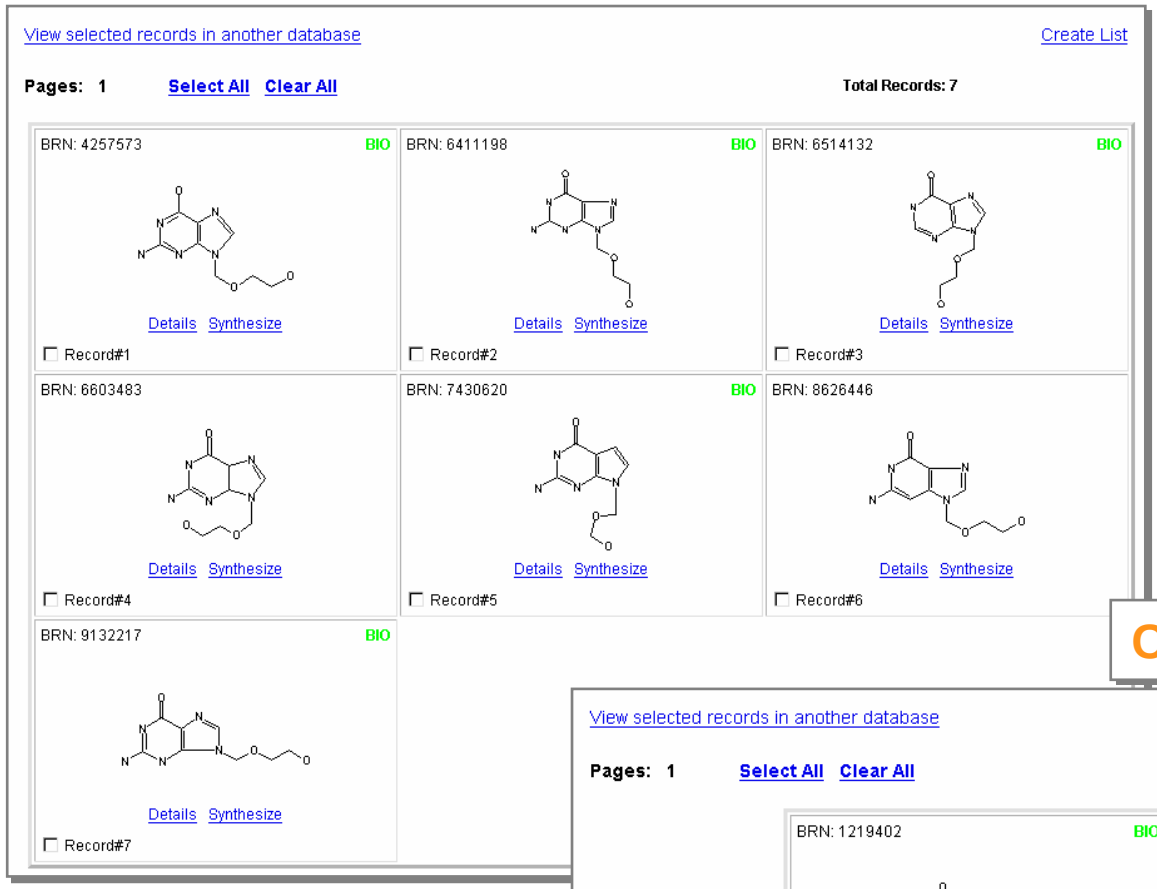
CAS number search



Data Lookup tables display the specified value and the number of occurrences found in the database.

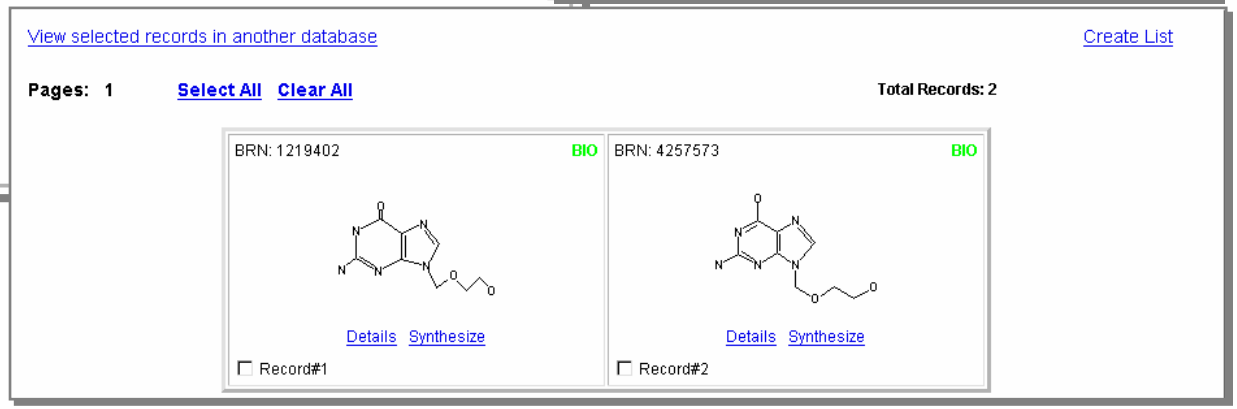
Search results

Chemical Name search results

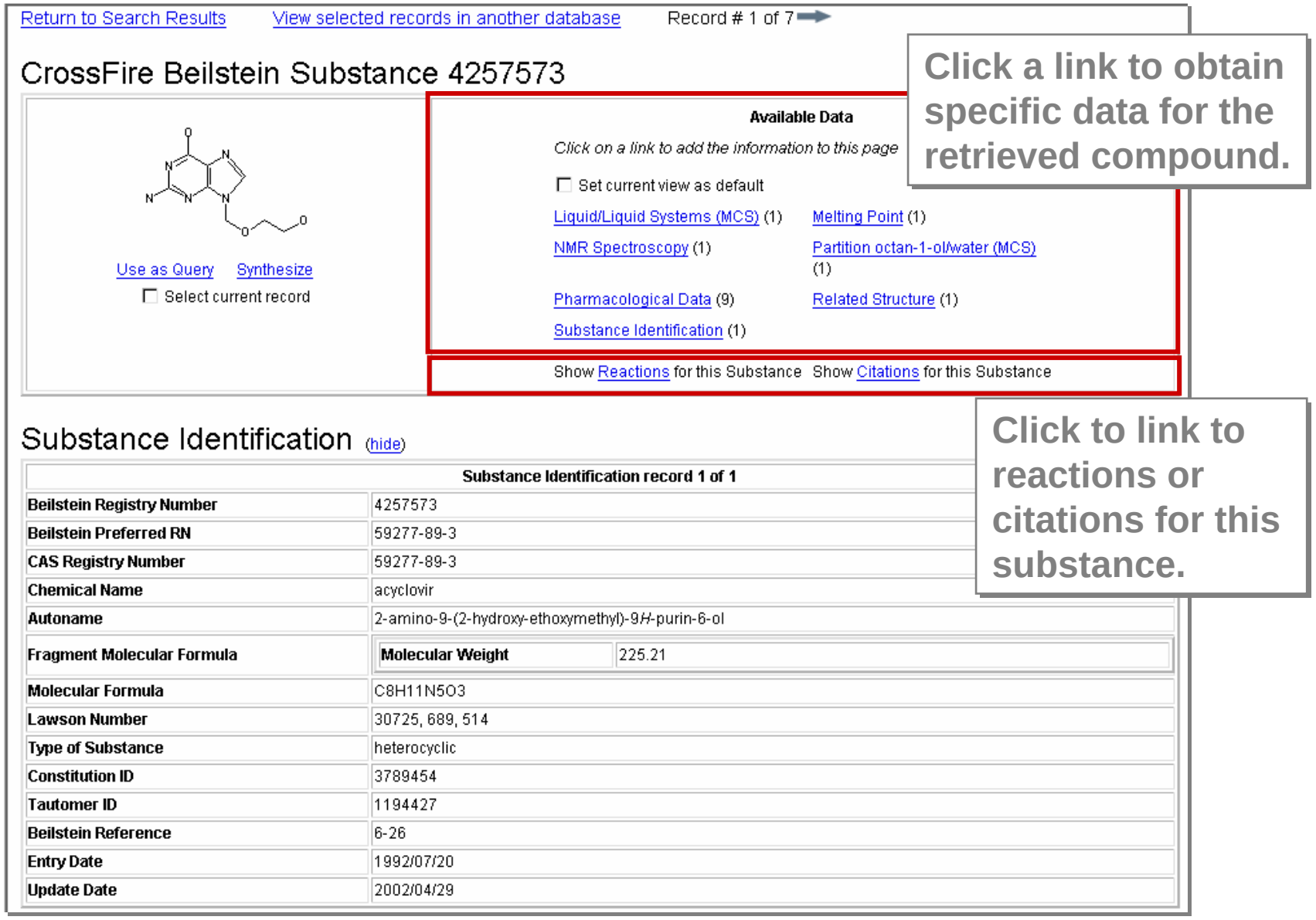


The “BIO” label indicates that the compound has pharmacological or toxicological data associated with it.

CAS number search results



Link to additional data



Available data

Displays or hides the data

Links to the literature reference

Pharmacological Data (hide)

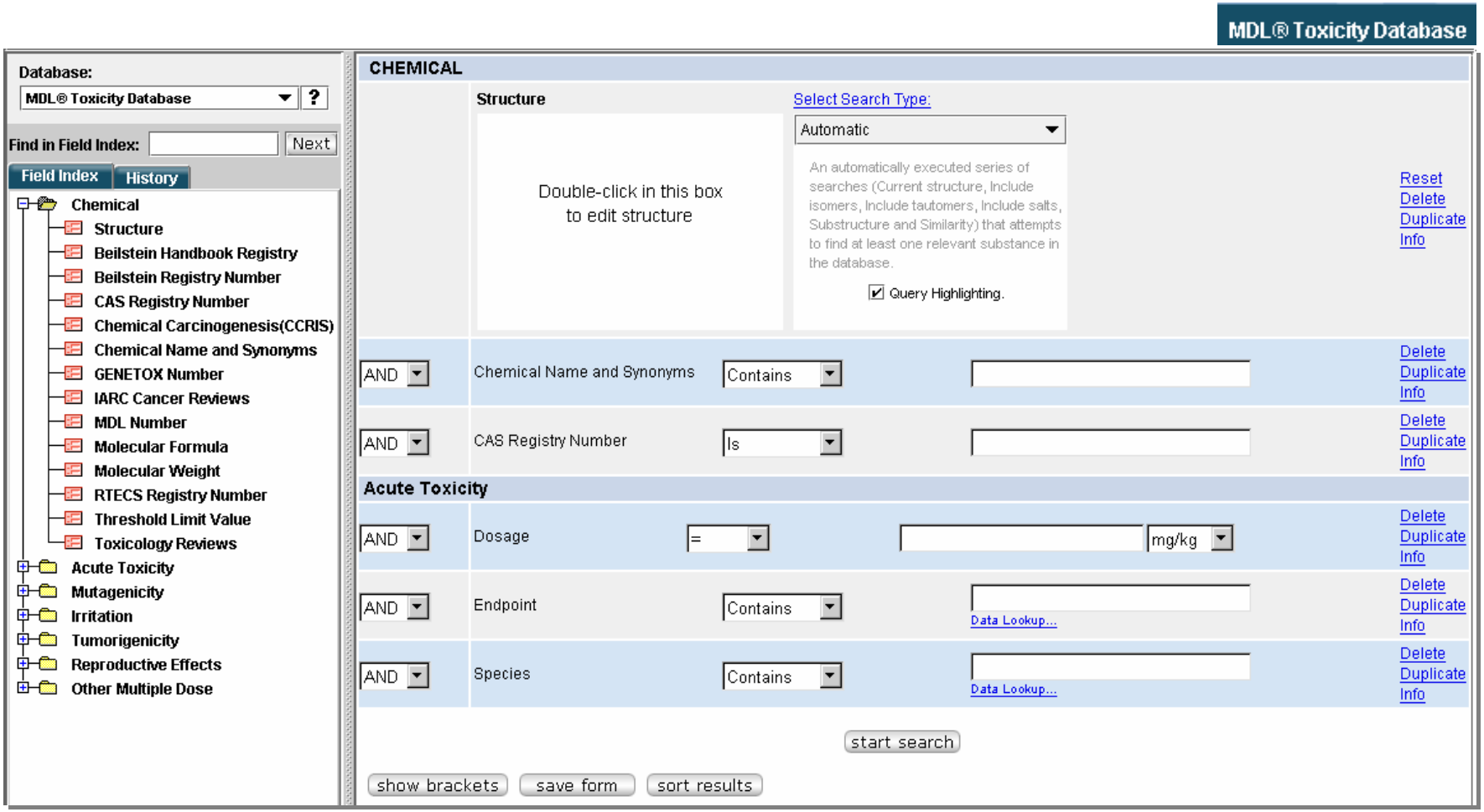
Pharmacological Data record 1 of 9	
Citation Pointer	5960932
Pharmacological Data Citations	• Journal; Sato, Yoshiko; Maruyama, Tokumi; CPBTAL; Chem. Pharm. Bull.; EN; 43; 1; 1995; 91-95;
Entry Date	1995/12/31
Comment	antiviral activity against HSV-1 in Vero cell (EC50: 0.41 µg/ml) and HSV-2 in Vero cell (0.18 µg/ml)

Pharmacological Data record 2 of 9	
Citation Pointer	5911291
Pharmacological Data Citations	• Journal; Winkelmann, E.; Winkler, I.; Rolly, H.; Roesner, M.; Jaehne, G.; ARZNAD; Arzneim. Forsch.; EN; 38; 11; 1988; 1545-1548;
Entry Date	1988/06/27
Comment	Chemoterapeutic activity against Herpes simplex (HSV-1) infection in mouse i.p.

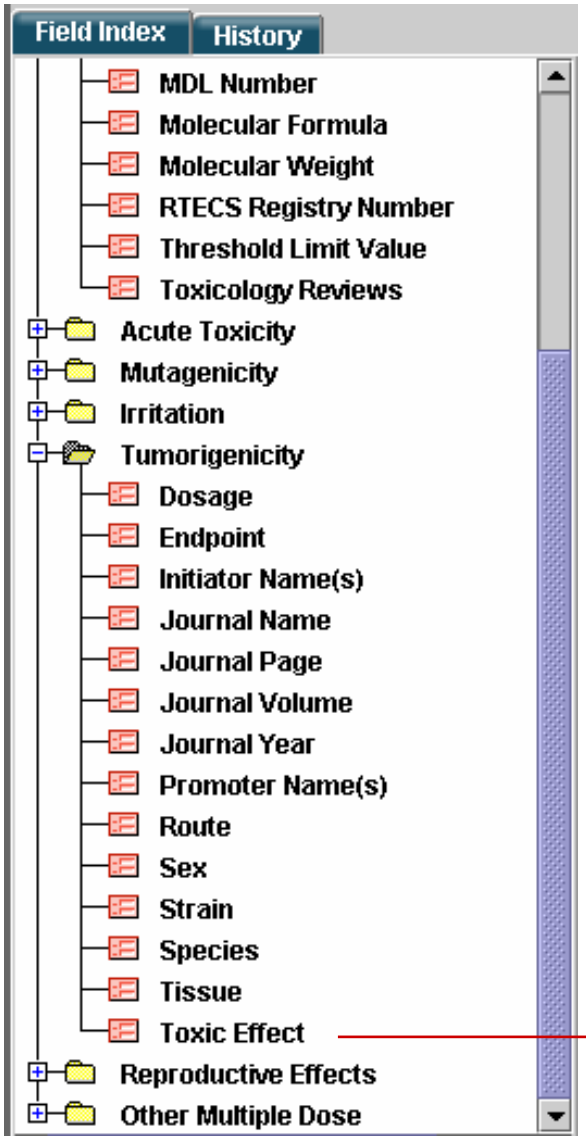
Pharmacological Data record 3 of 9	
Citation Pointer	5904365
Pharmacological Data Citations	• Journal; El-Kousy, S.; Pedersen, E. B.; Nielsen, C.; MOCMB7; Monatsh. Chem.; EN; 125; 6/7; 1994; 713-722;
Entry Date	1994/12/31
Comment	activity against Herpes Simplex Virus, type 1, strain McIntyre, in SIRC cells: ED 50 = 0.5 µM; cytoxicity against normal uninfected SIRC cells: CD 50 = >50 µM

Search the MDL Toxicity database

Conduct a search, using the MDL Toxicity database to retrieve compounds that have tumorigenicity data associated with fatty liver degeneration.

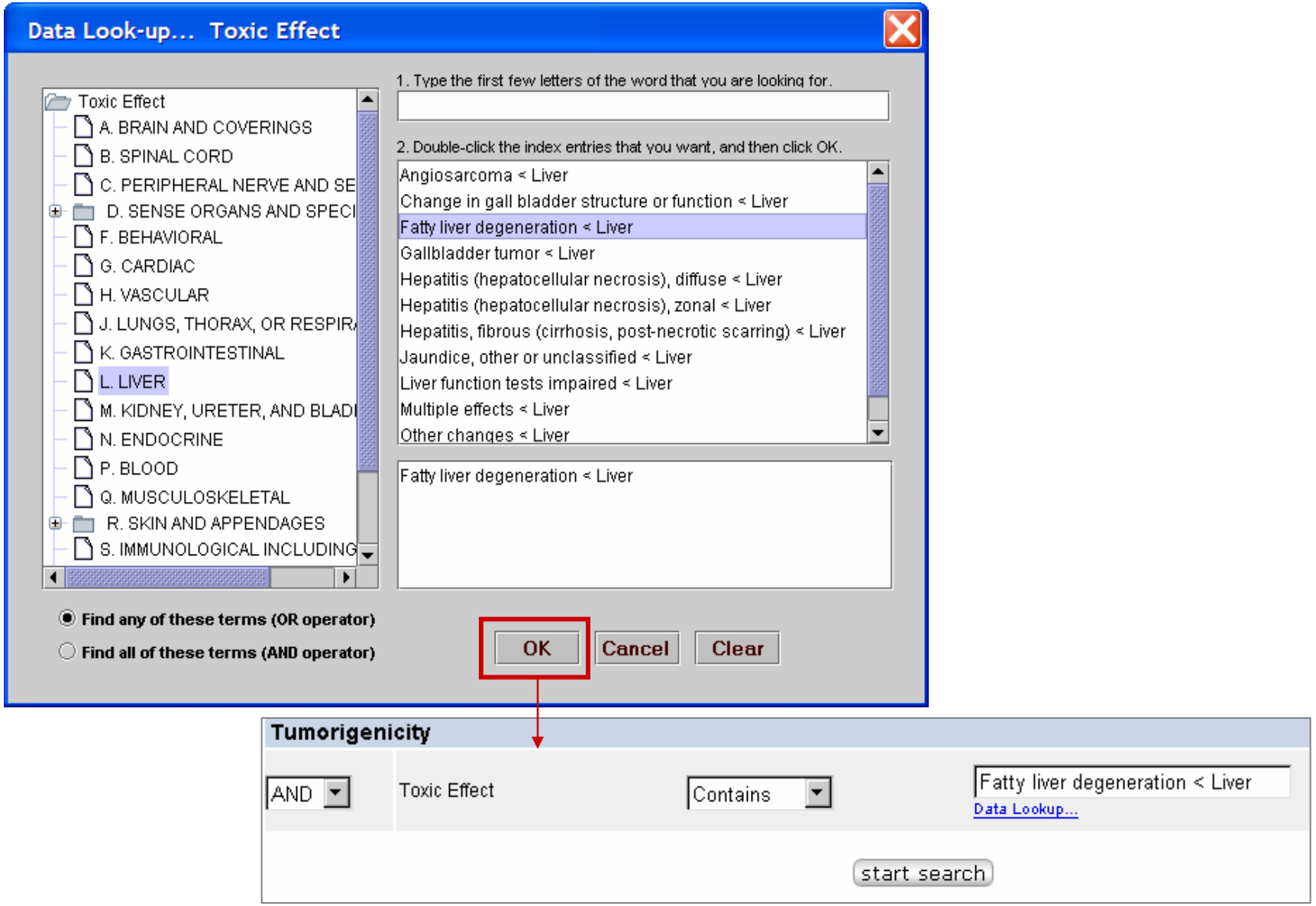


Create a custom form

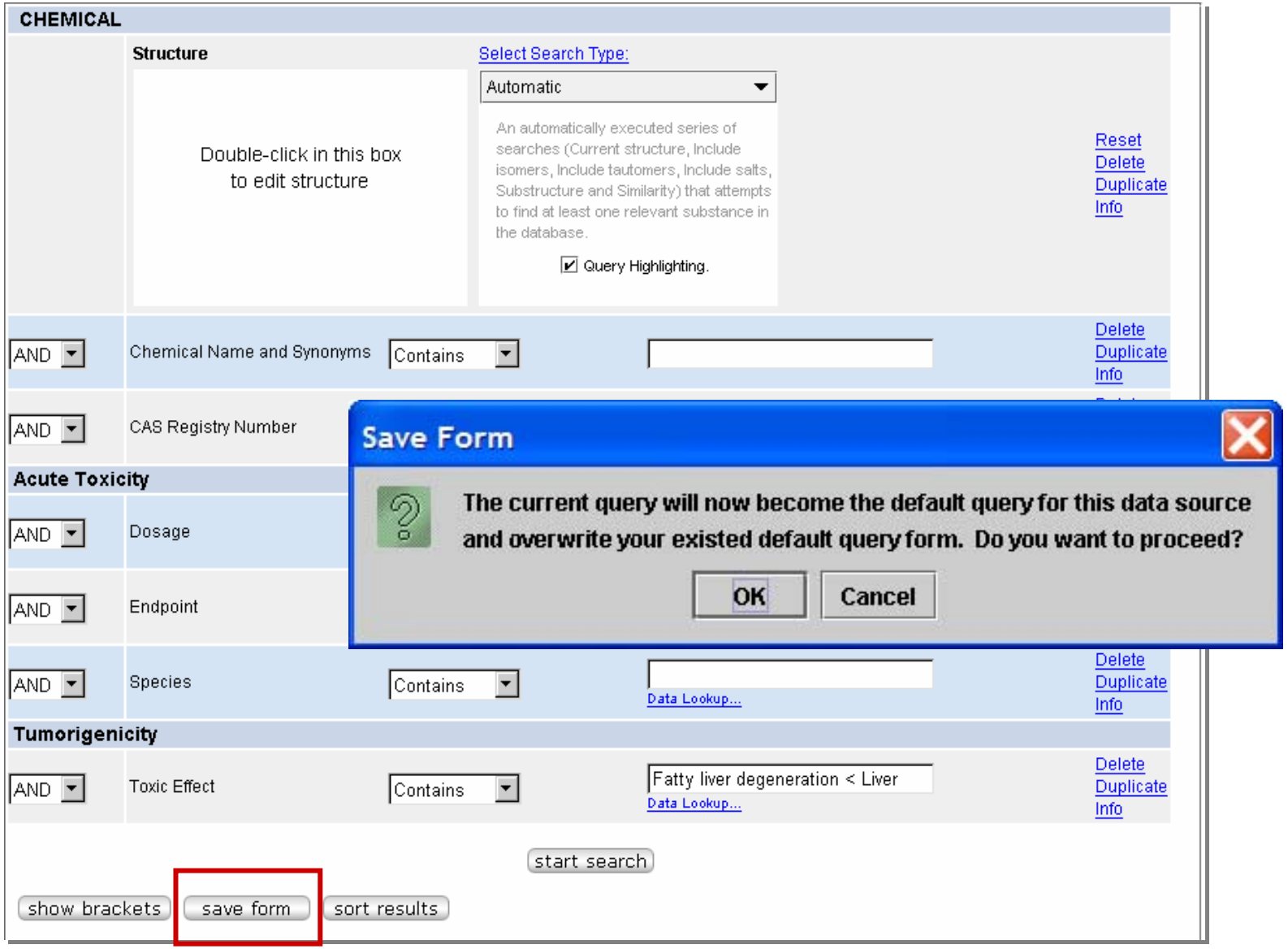


Double-click to add field to query form

Enter the ‘toxic effect’ data constraint

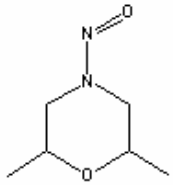


Saving a form



Tumorigenicity data

MDL® Toxicity Database



Use as Query

Select current record

Available Data

Click on a link to add the information to this page

Set current view as default

Chemical (1) Acute Toxicity (4)

Mutagenicity (35) Tumorigenicity (35)

Review (1) Model (1)

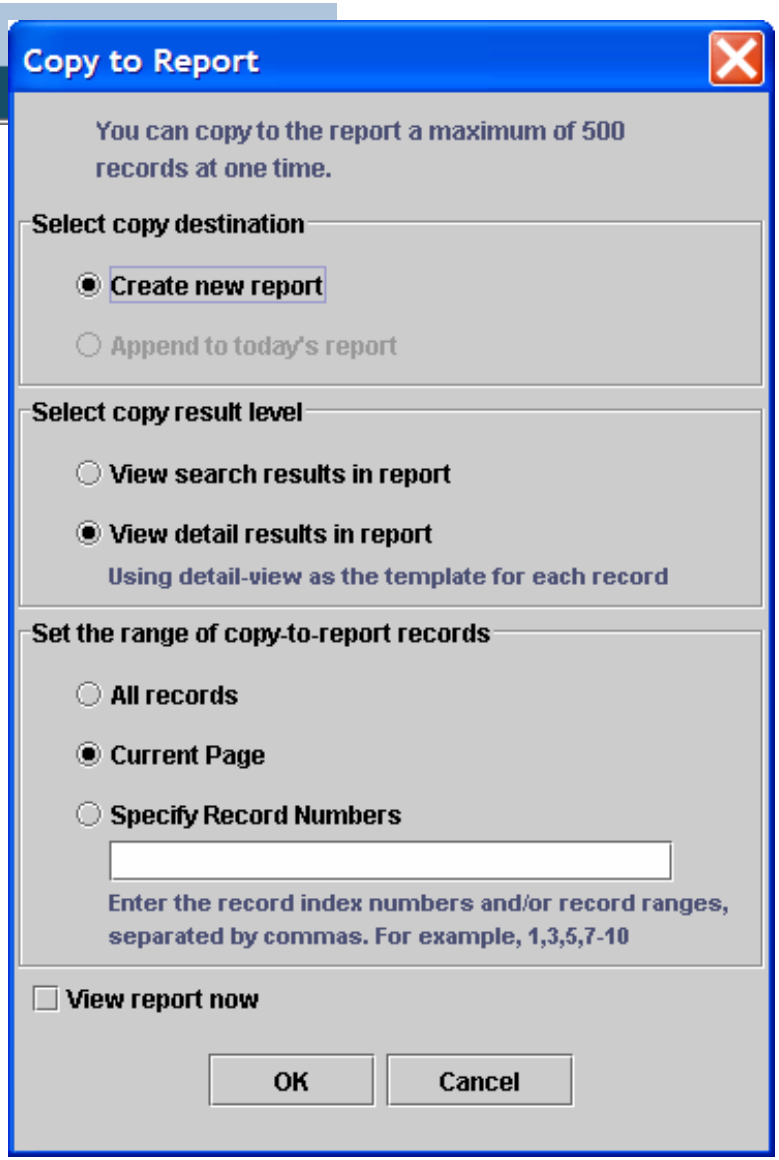
Tumorigenicity (hide)

Full Citation 1 of 35	
Journal of the National Cancer Institute 64,529,1980	
Source ID : QE2150000	**Source :** RTECS
Chemical Name	MORPHOLINE, 2,6-DIMETHYL-N-NITROSO-
Species	guinea pig
Route	ORAL
Dosage	400 MG/KG/23W-I
Endpoint :	TD
Endpoint (Incidence) :	(Carcinogen, Carcinogen+Inhibitor, %Diff., Significance)
Endpoint (Multiplicity) :	
Endpoint (Latency) :	(Carcinogen, Carcinogen+Inhibitor, Significance)

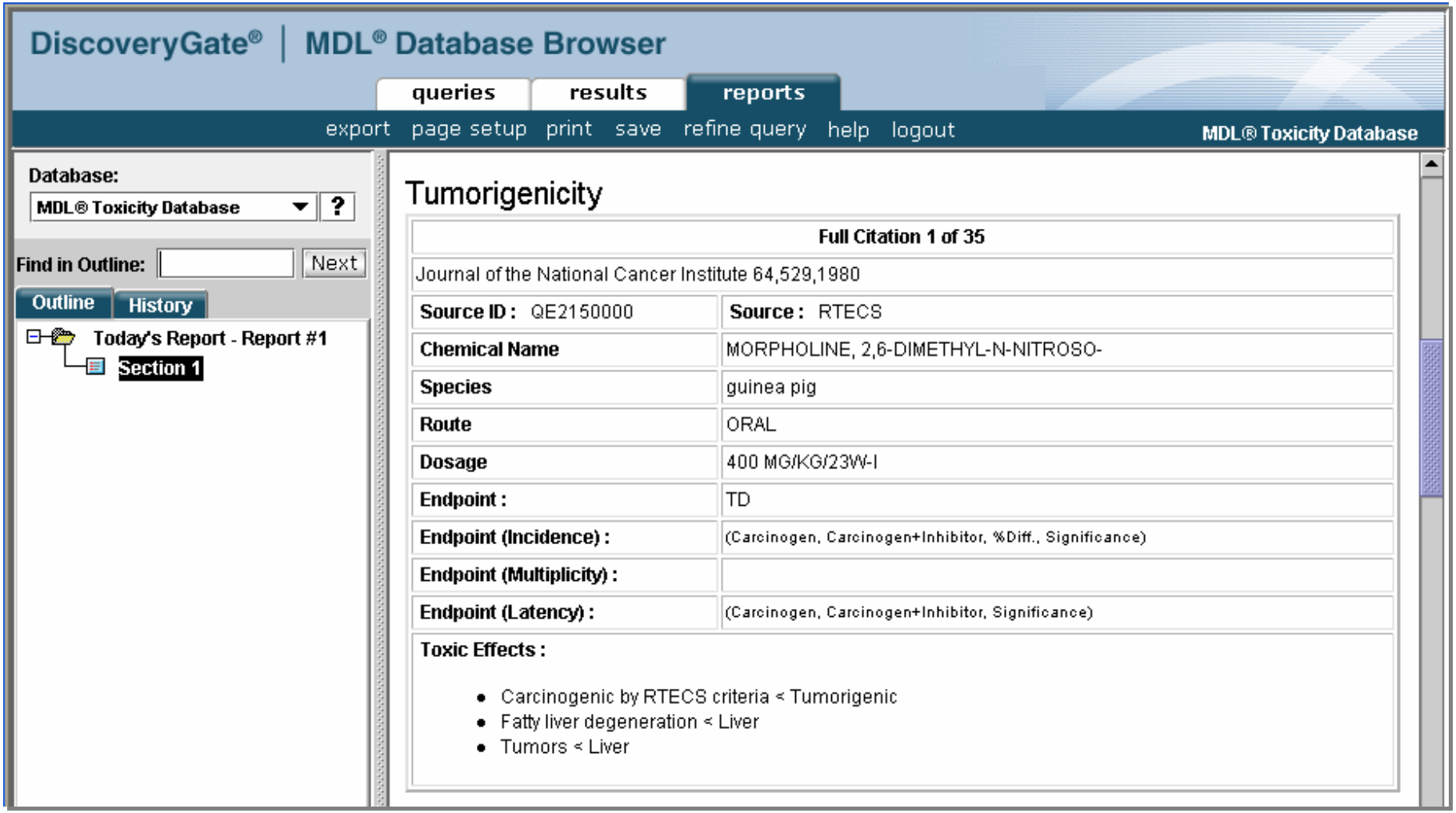
Toxic Effects :

- Carcinogenic by RTECS criteria < Tumorigenic
- Fatty liver degeneration < Liver
- Tumors < Liver

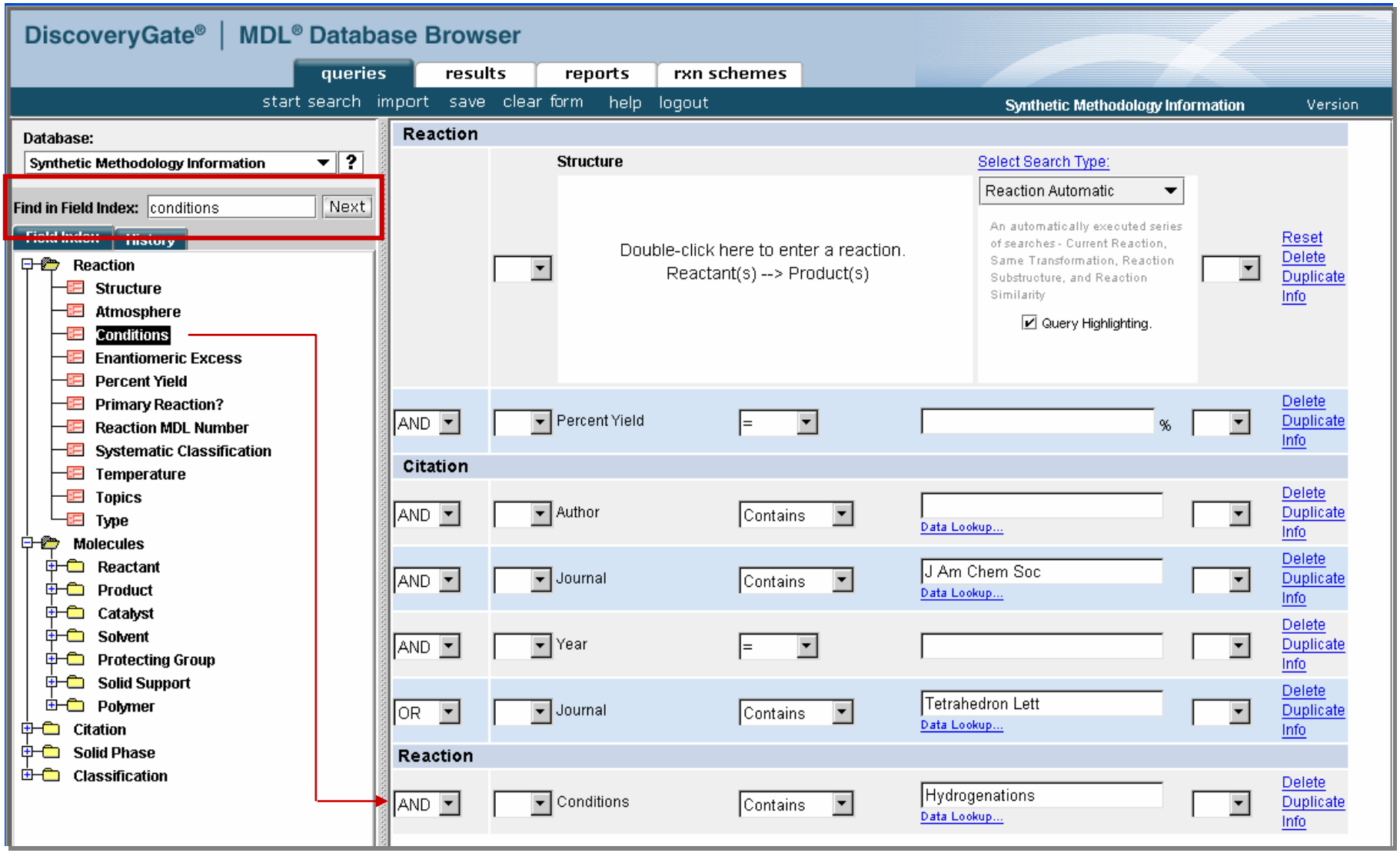
Creating a report



Reports tab


DiscoveryGate® | MDL® Database Browser
queries
results
reports
export
page setup
print
save
refine query
help
logout
MDL® Toxicity Database
Database:
MDL® Toxicity Database
Find in Outline:
Next
Outline
History
Today's Report - Report #1
Section 1
Tumorigenicity
Full Citation 1 of 35
Journal of the National Cancer Institute 64,529,1980
Source ID : QE2150000
Source : RTECS
Chemical Name
MORPHOLINE, 2,6-DIMETHYL-N-NITROSO-
Species
guinea pig
Route
ORAL
Dosage
400 MG/KG/23W-I
Endpoint :
TD
Endpoint (Incidence) :
(Carcinogen, Carcinogen+Inhibitor, %Diff., Significance)
Endpoint (Multiplicity) :
Endpoint (Latency) :
(Carcinogen, Carcinogen+Inhibitor, Significance)
Toxic Effects :
Carcinogenic by RTECS criteria < Tumorigenic
Fatty liver degeneration < Liver
Tumors < Liver

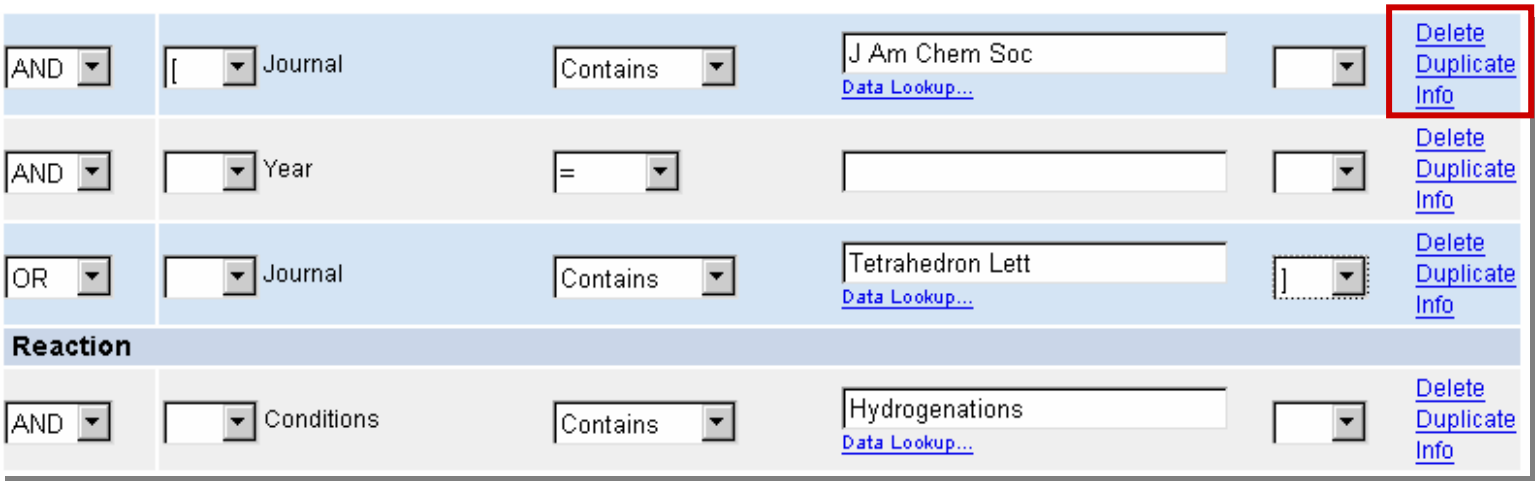
Manually add a data field



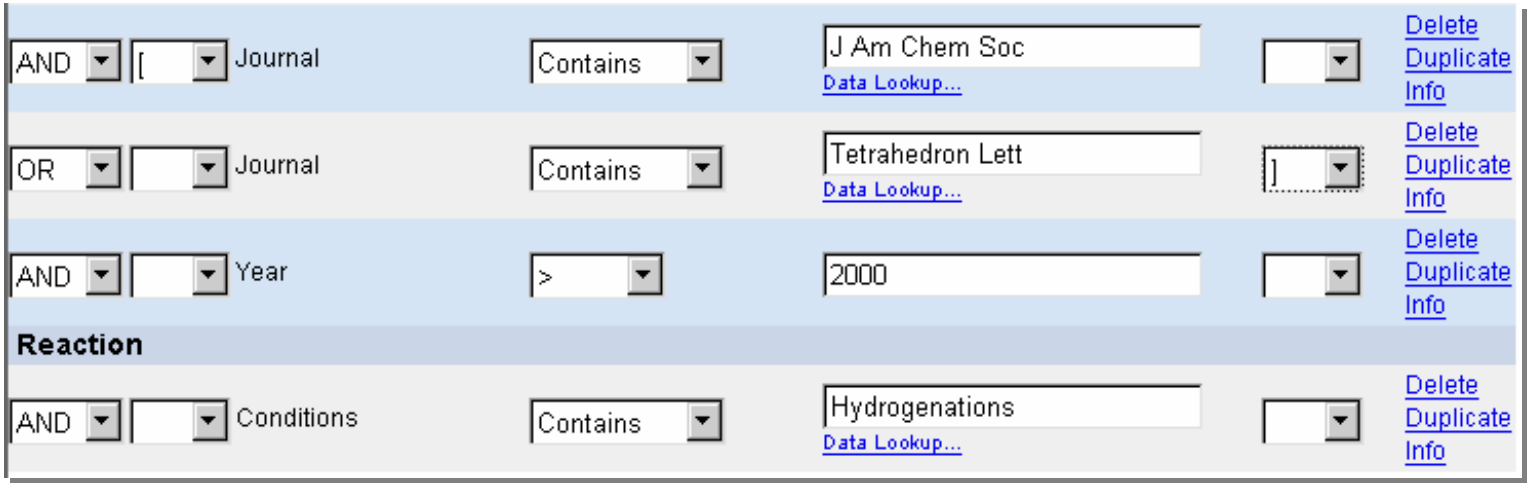
Use the Field Index to find the data field. Double-click to add it to the form.

Setting parenthetical statements

Initial search



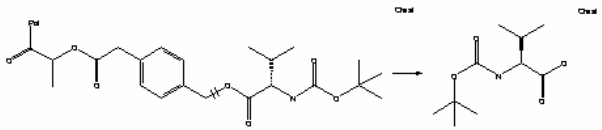
Modified search



Manually adjust fields to create the correct parenthetical statements.

Search results

MDL® Solid-Phase Organic Reactions



Use as Query

Select current record

Available Data

Click on a link to add the information to this page

Set current view as default

Reaction Details (4) Reactant (1)

Product (1)

Reaction Details (hide)

RSPO69000650 **Reaction Variation 1 of 4**

Product no.	% Yield	Grade	%cs	%ds	%de	%ee
1	50	(S)-isomer	100			

Reactant no.	Reactant Grade
1	(S)-isomer

Step no.	Catalysts and Solvents	Catalyst ID click ID for details	Solvent ID click ID for details
1	H2 Pd(O-Ac)2 (cat.) DMF	18137 48	9

Detailed Data	Conditions Solid Support References Protecting Group	External Registry No.	Path	Step
		698000501	I	4 OF 4

TAM, J. P.; TJOENG, F. S.; MERRIFIELD, R. B.; J Am Chem Soc [JACSAT] 1980, 102 (19), 6117-6127.
TAM, J. P.; TJOENG, F. S.; MERRIFIELD, R. B.; Tetrahedron Lett [TELEAY] 1979, 51, 4935-4938.
WHITNEY, D. B.; TAM, J. P.; MERRIFIELD, R. B.; Tetrahedron [TETRAB] 1984, 40 (21), 4237-4244.

Conditions

Step	1
Hours	17
Temp (C)	50
Conditions	Hydrogenations Transition Metal Salts Thermal

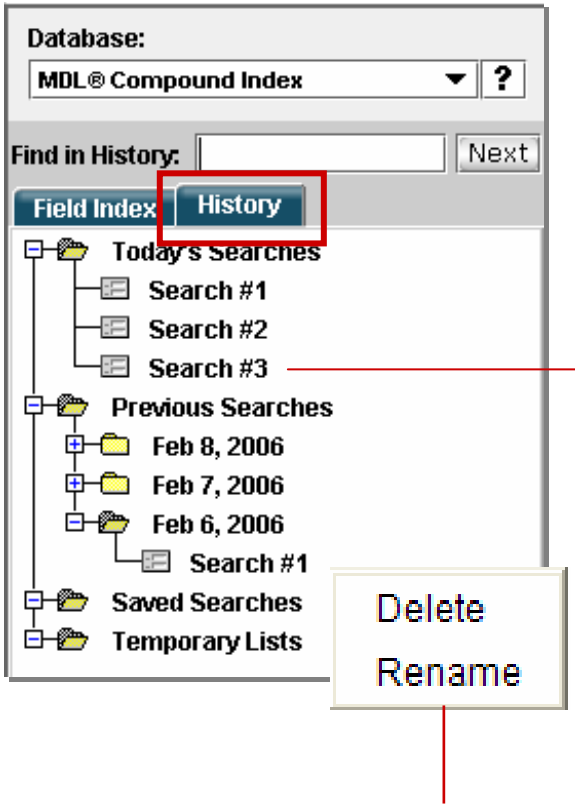
Manipulating search results

In this section, you will learn how to:

- Work with the History tab
- Create a temporary list
- Save a permanent list
- Combine temporary lists
- Export results

Working with the History tab

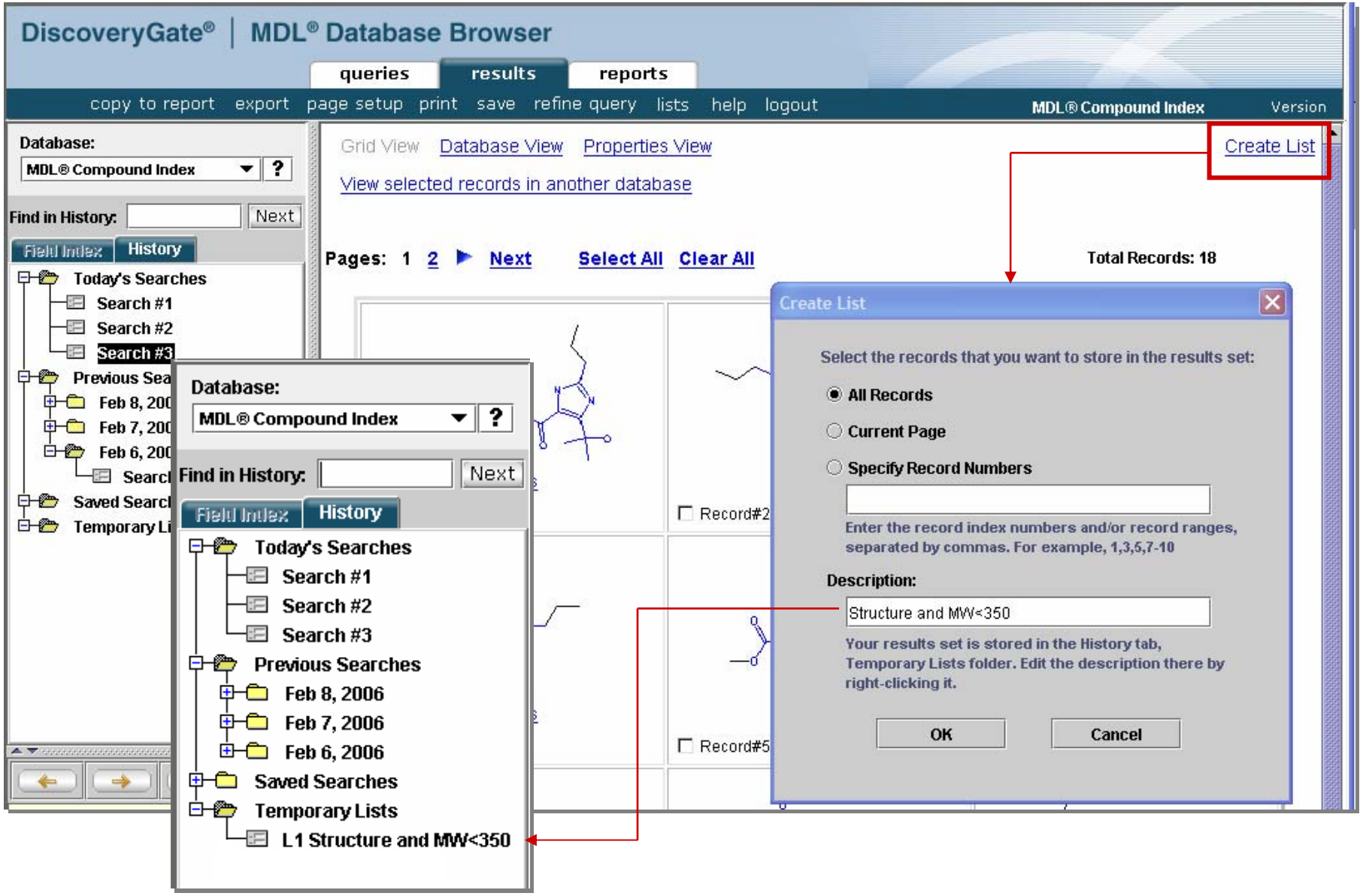
The History tab is unique to the database you are using.



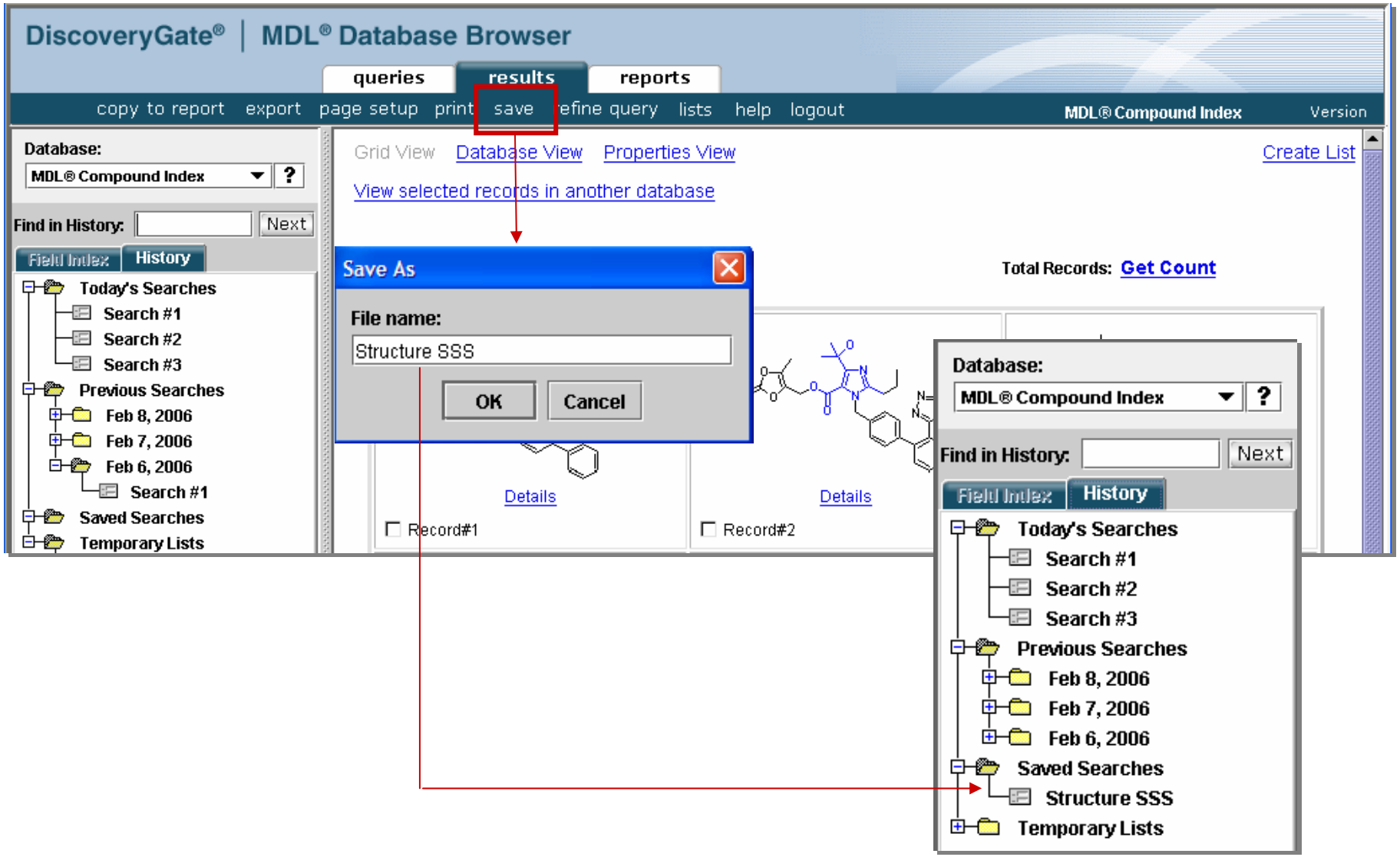
Using the “results” tab, double-click to open

Right-click a list to Delete or Rename

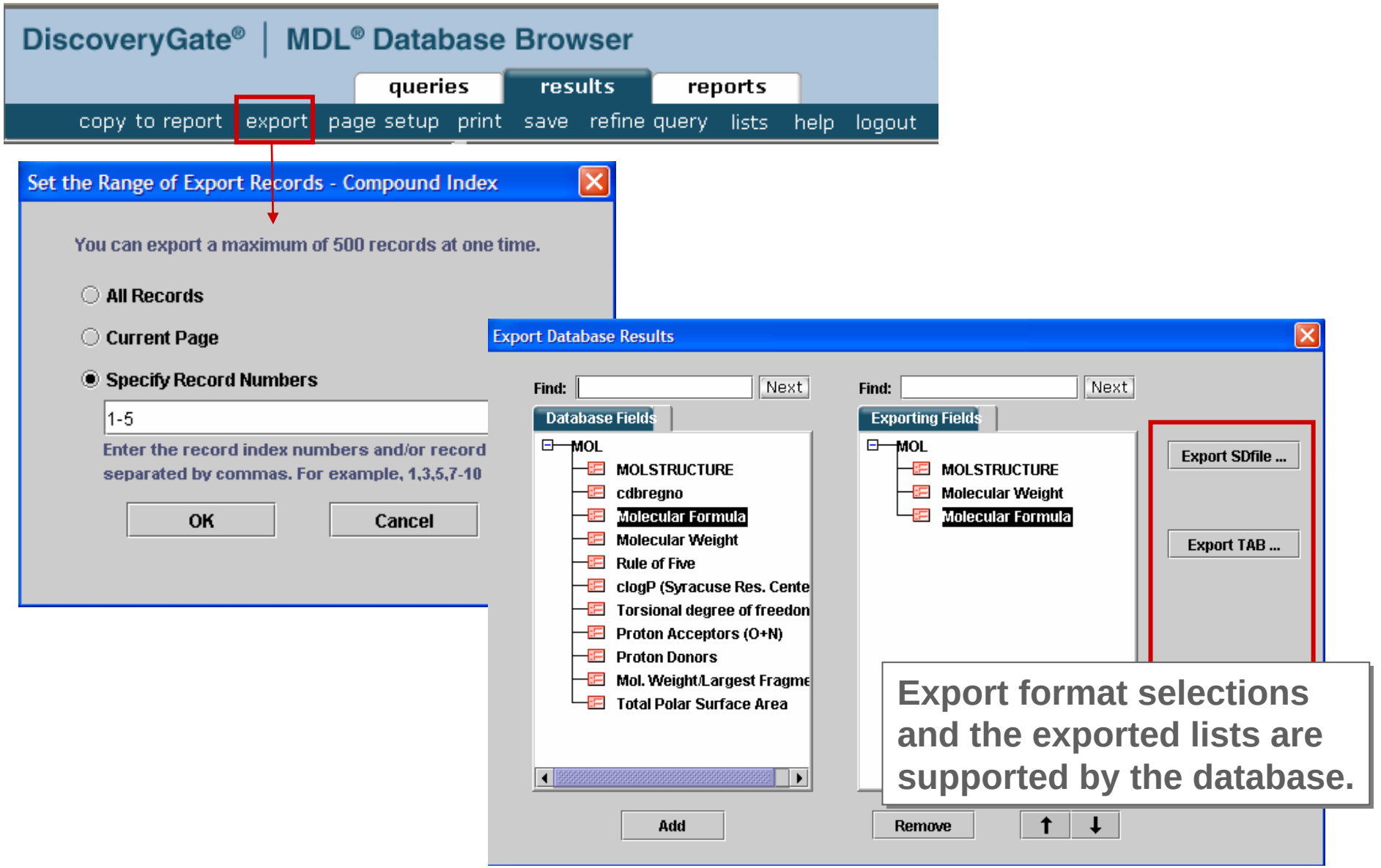
Create a temporary list



Create a permanent list



Export a list



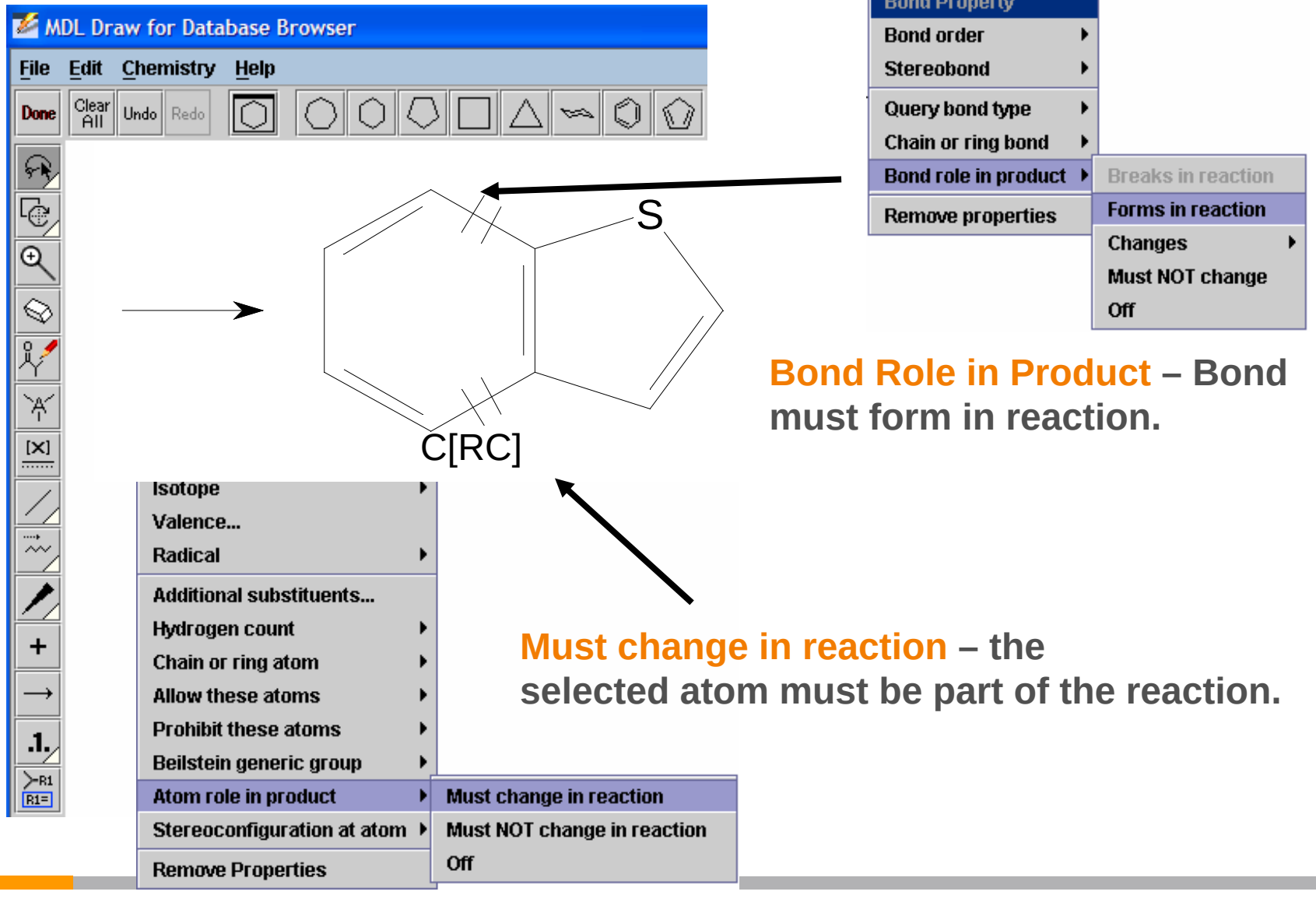
Export formats

You can export results, both structures and data, in the following formats:

.sdf	▪ **This is a flat file format and can store structures, but not reactions.** ▪ **This format is suitable for files to be exported to ISIS for Excel and to ISIS/Base databases.** ▪ **The “results” tab needs to be active.**
.rdf	▪ **This is a hierarchical file format and can store reactions.** ▪ **This format is suitable for files to be exported to ISIS/Base databases.** ▪ **The “results” tab needs to be active.**
.tab	▪ **This format is suitable for files to be exported to ISIS for Excel.** ▪ **The “results” tab needs to be active.** ▪ **You require Chime software to view the structures.**

Searching for reactions

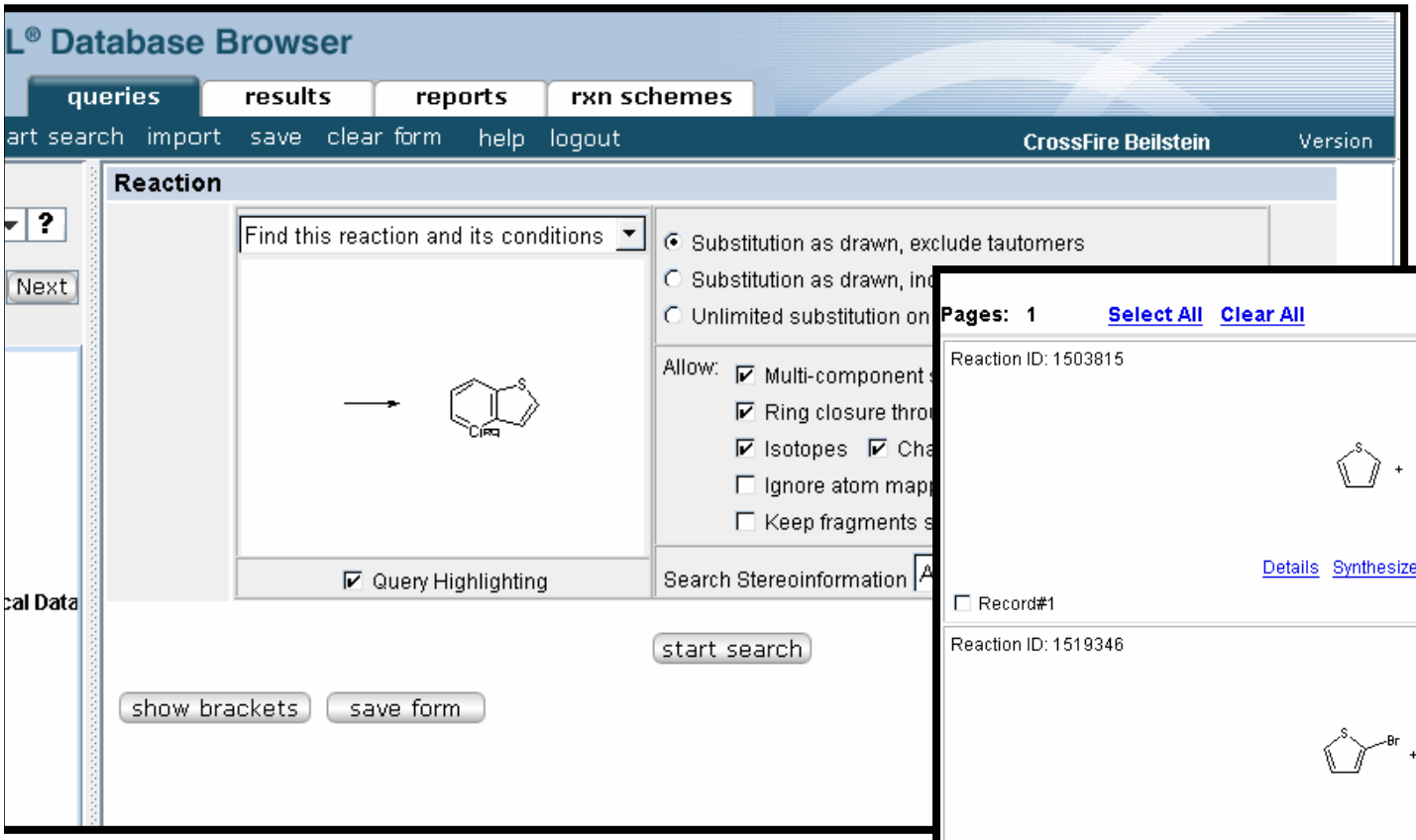
Reaction Query



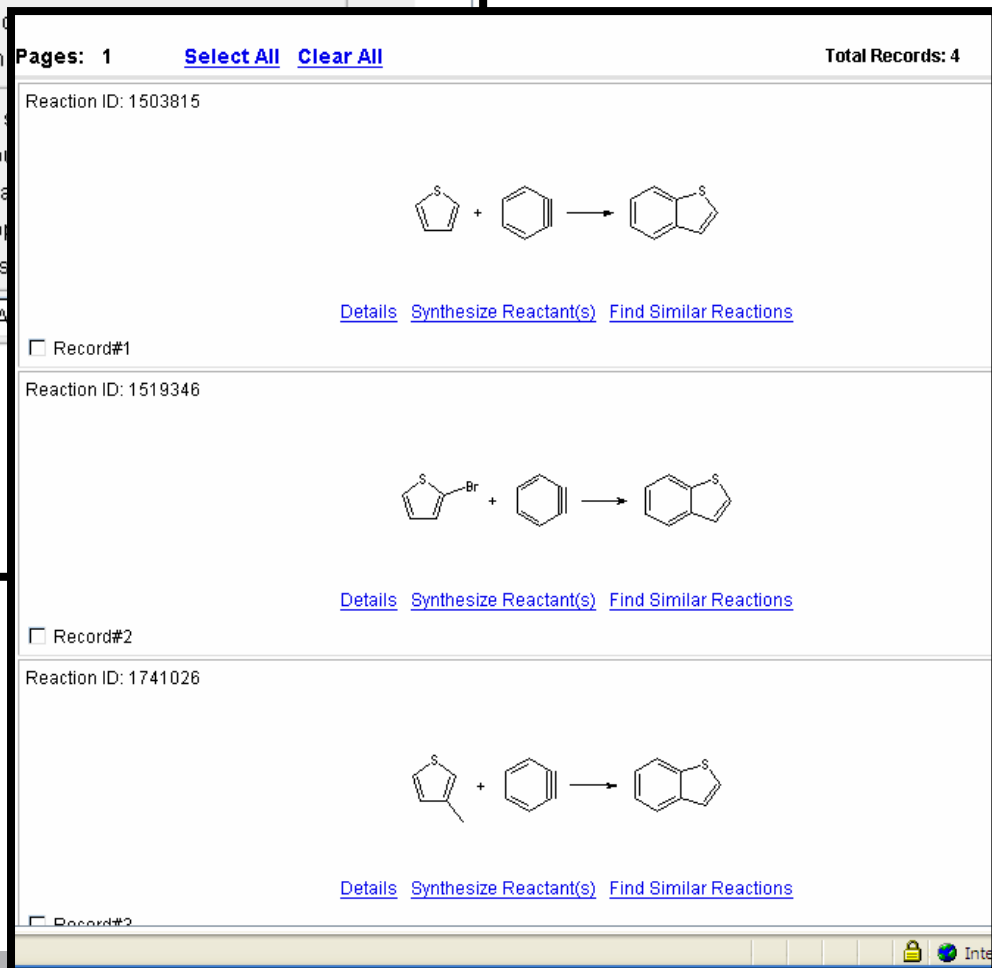
Bond Role in Product – Bond must form in reaction.

Must change in reaction – the selected atom must be part of the reaction.

Query

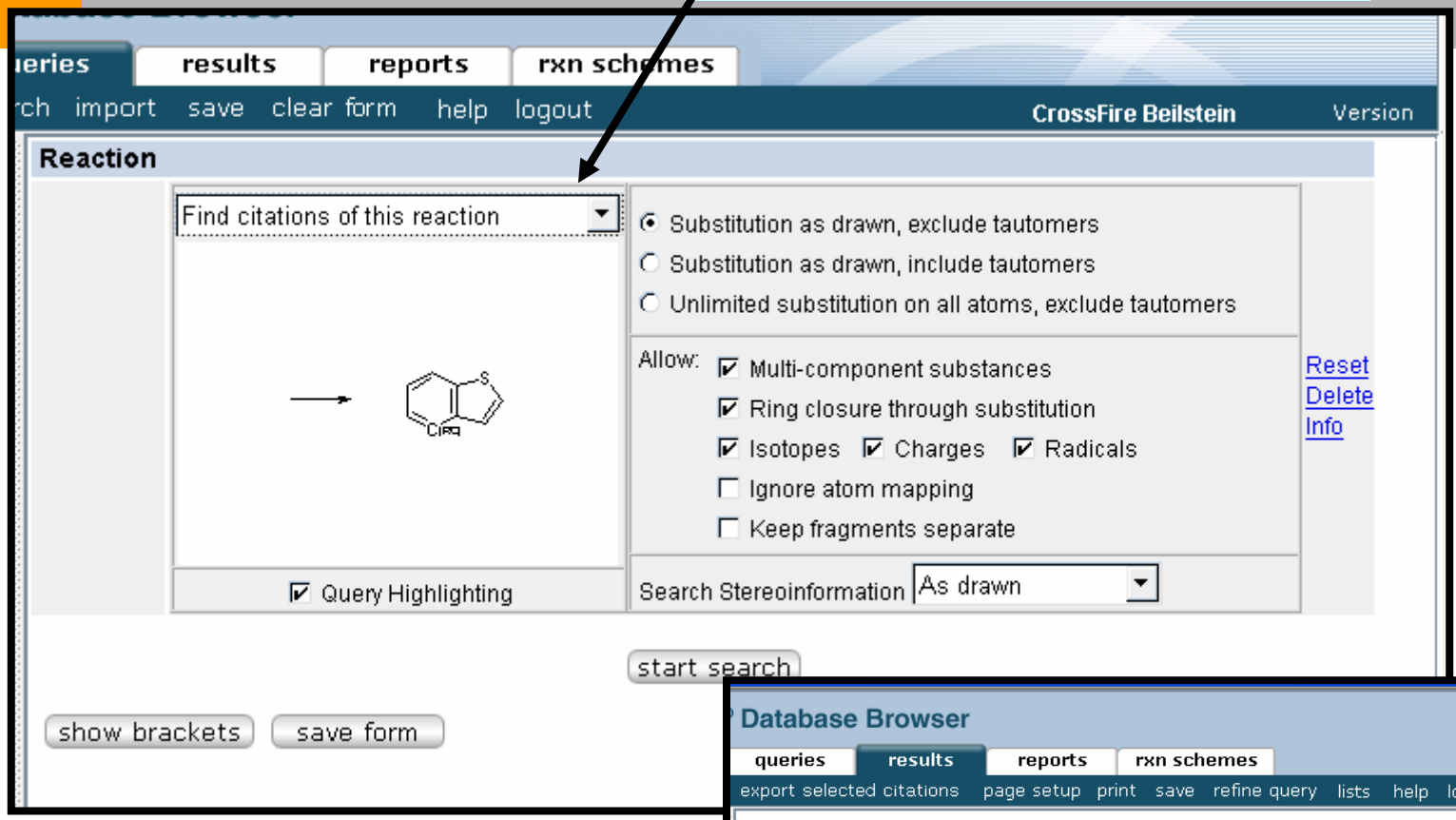


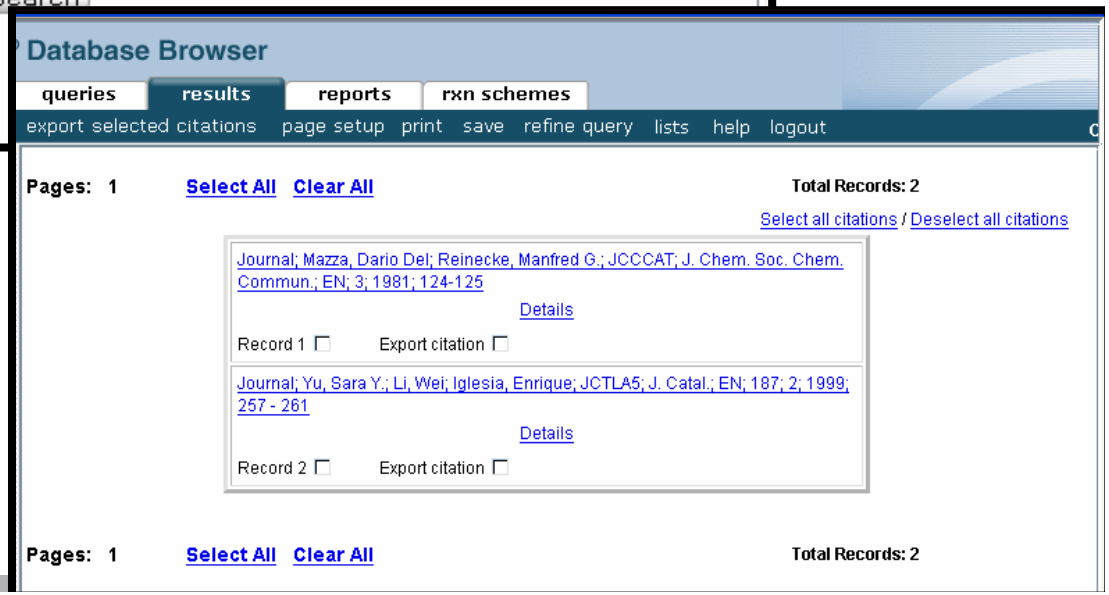
Results



Return to **Query Tab**

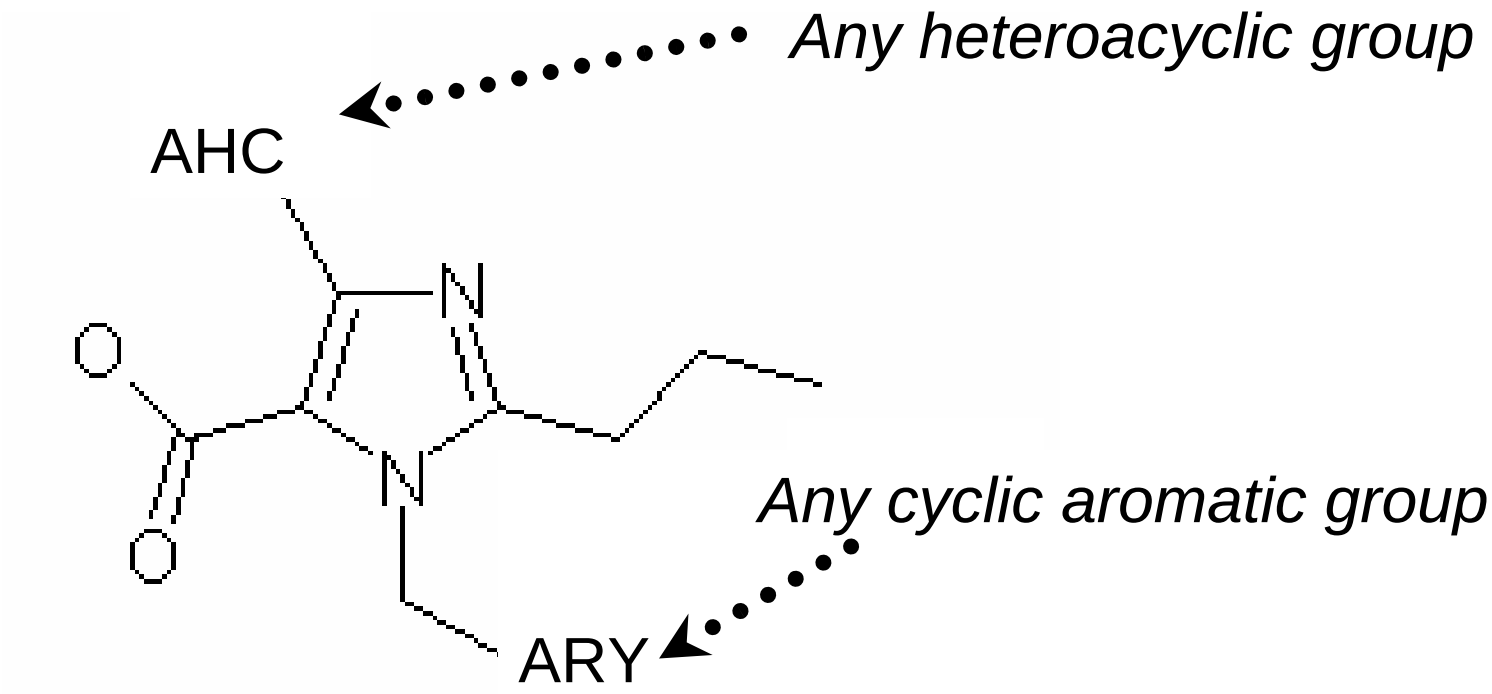
Choose **Find Citations**



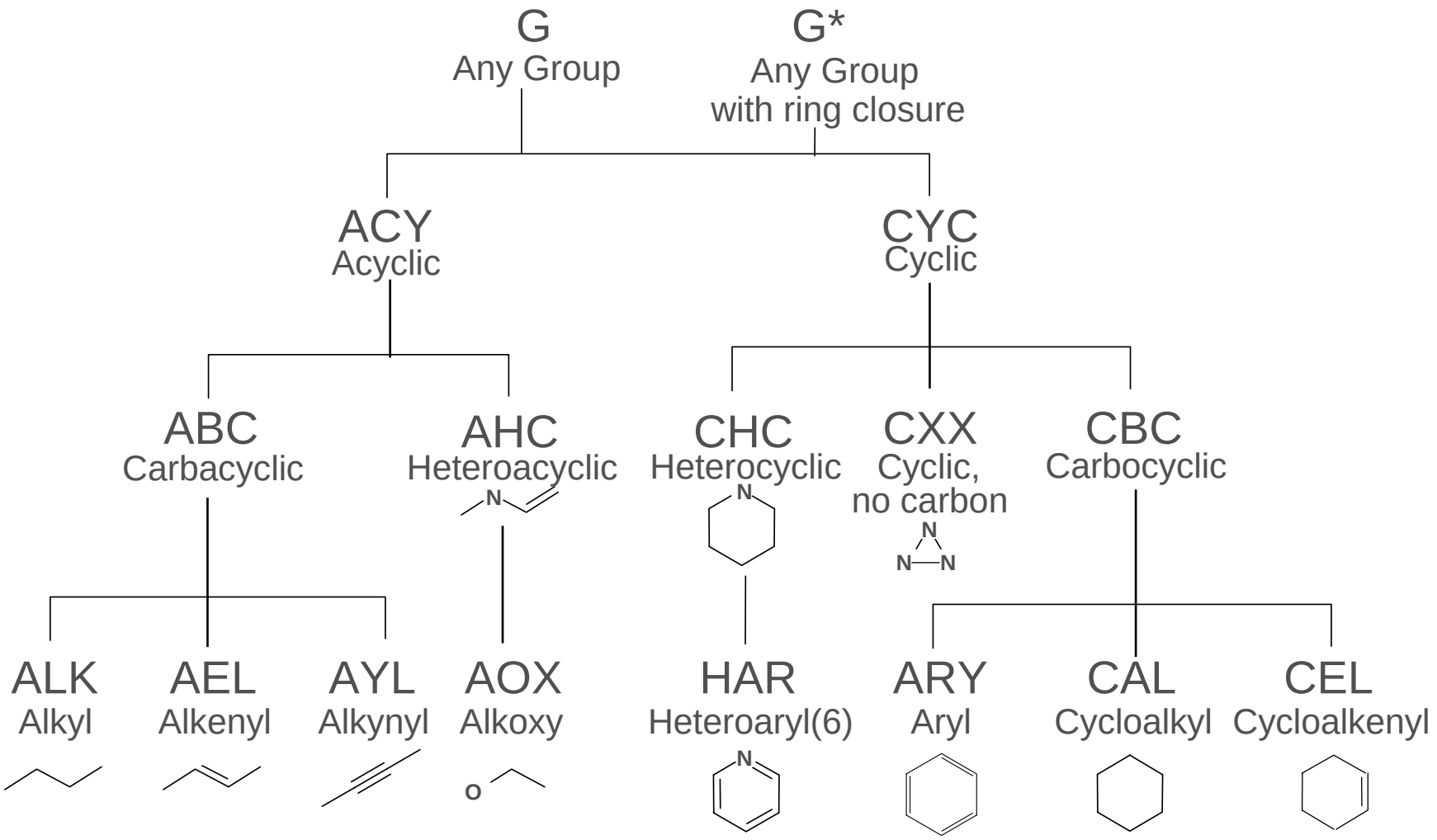


Citations

Query

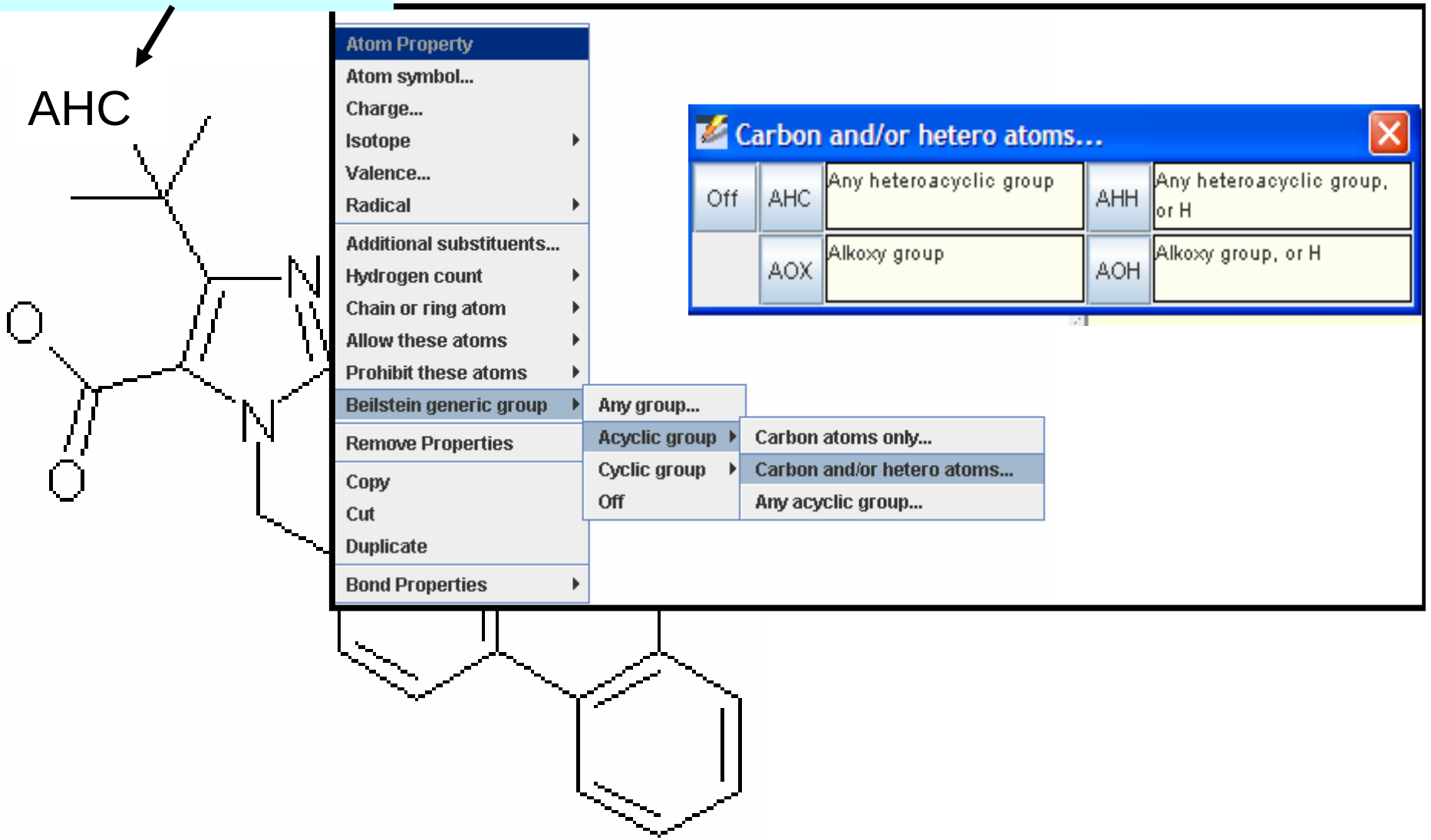

Any heteroacyclic group
AHC
O
N
O
N
Any cyclic aromatic group
ARY

Beilstein generic groups

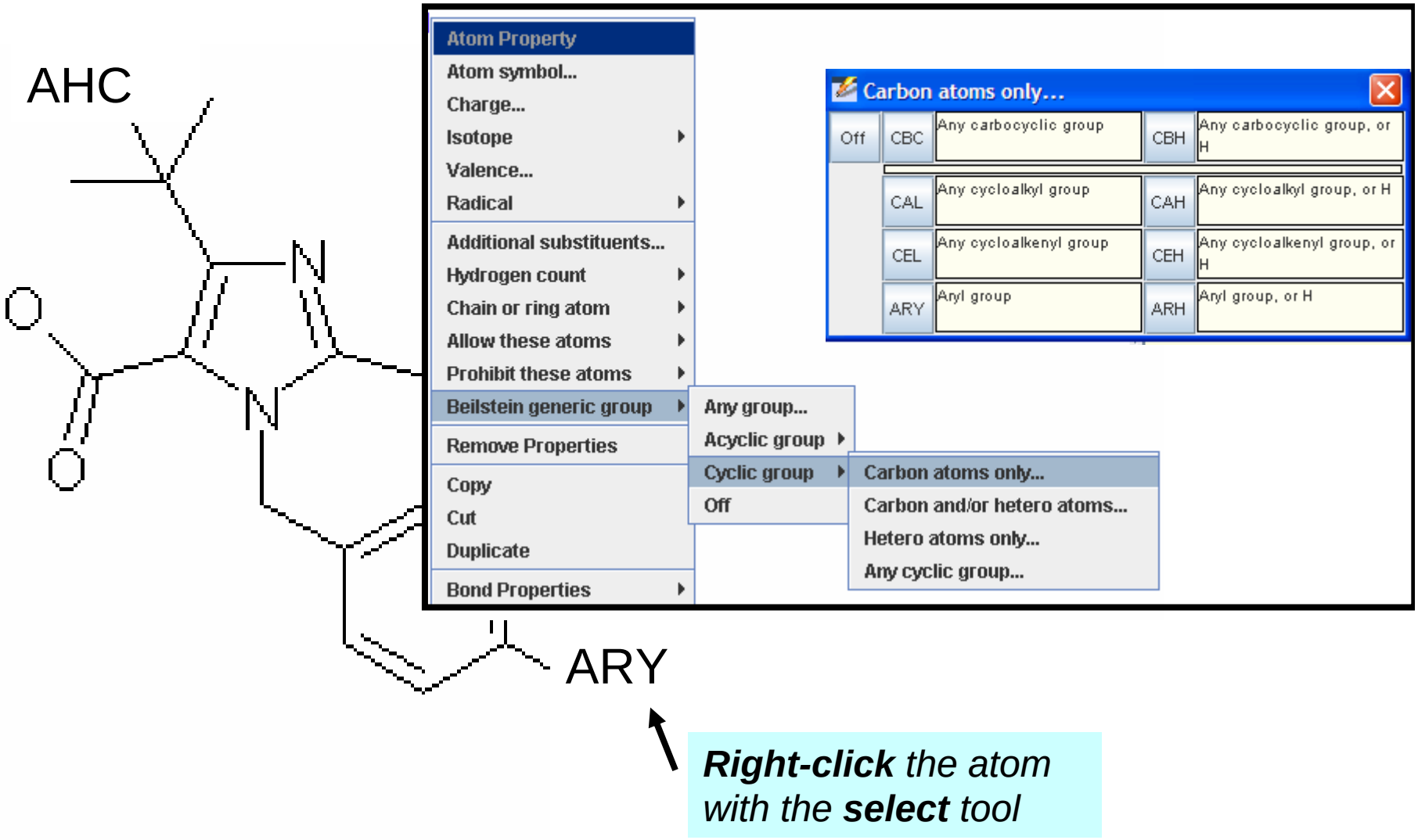


Add Beilstein Generic Group

***Right-click** the atom with the **select** tool*



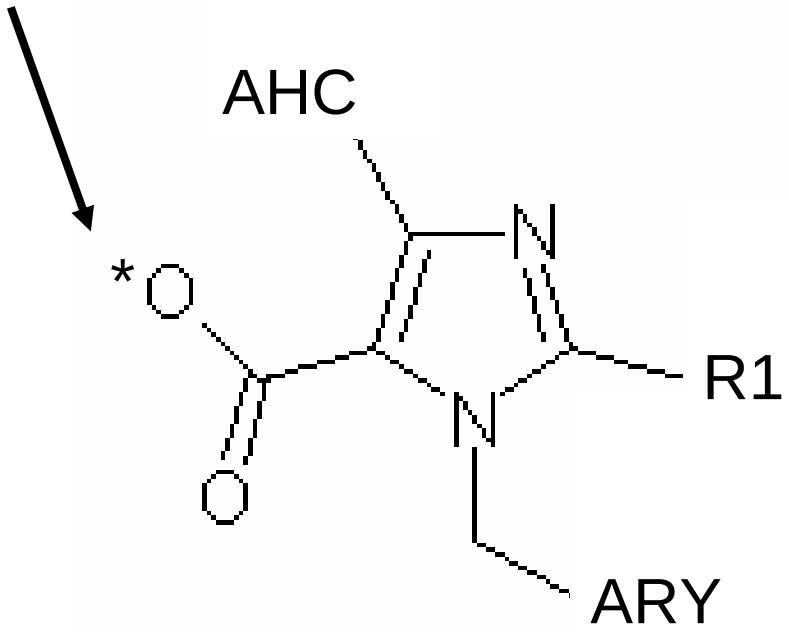
Add Beilstein Generic Group



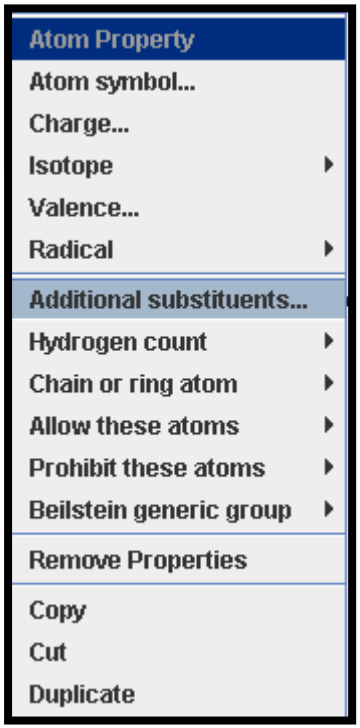
Right-click *the atom with the* ***select*** *tool*

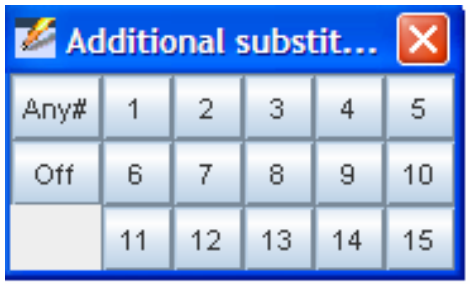
Force Substitution

***Right-click** the atom using the **select** tool*



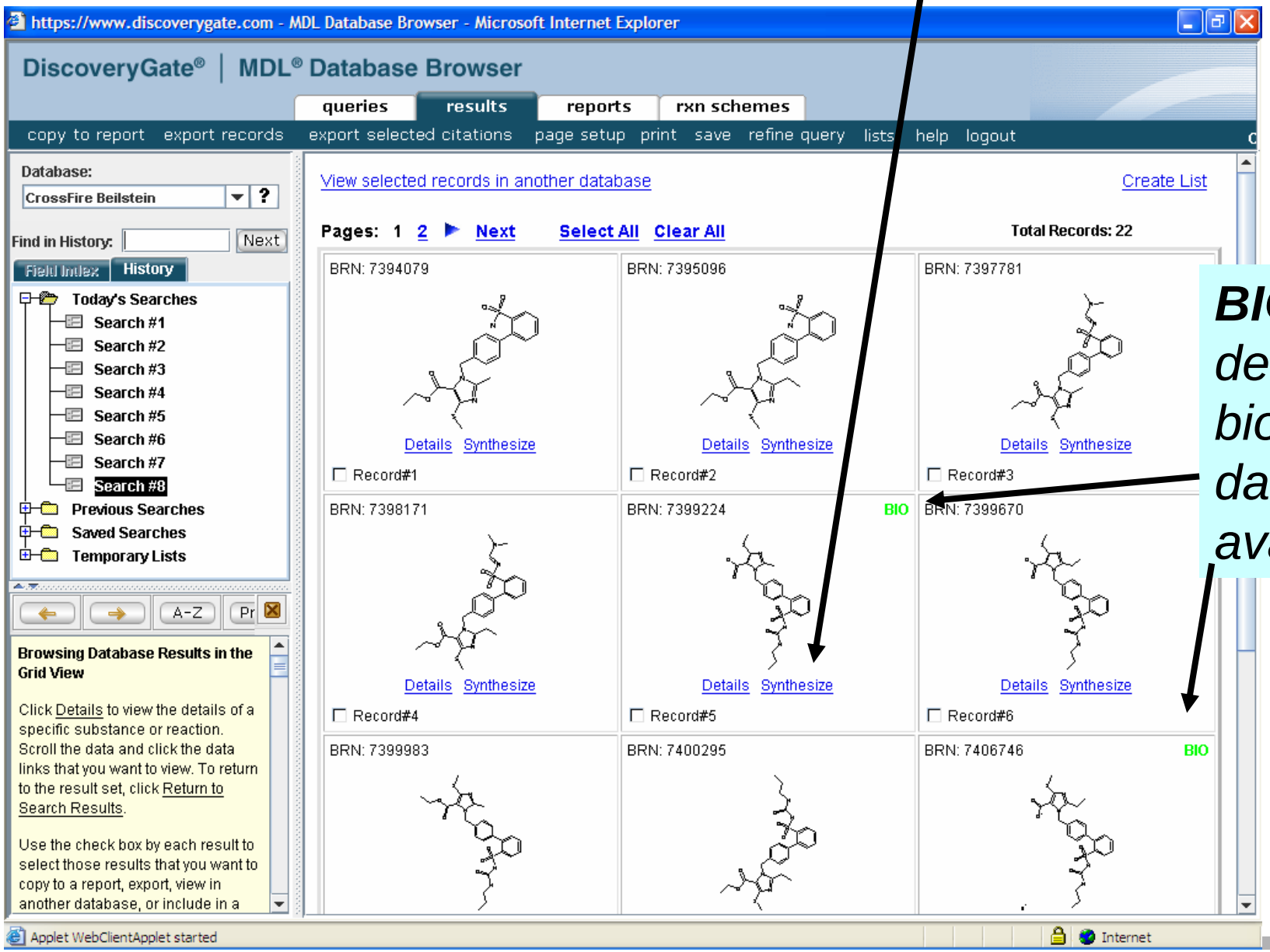
*Choose **Additional substituents>Any#***





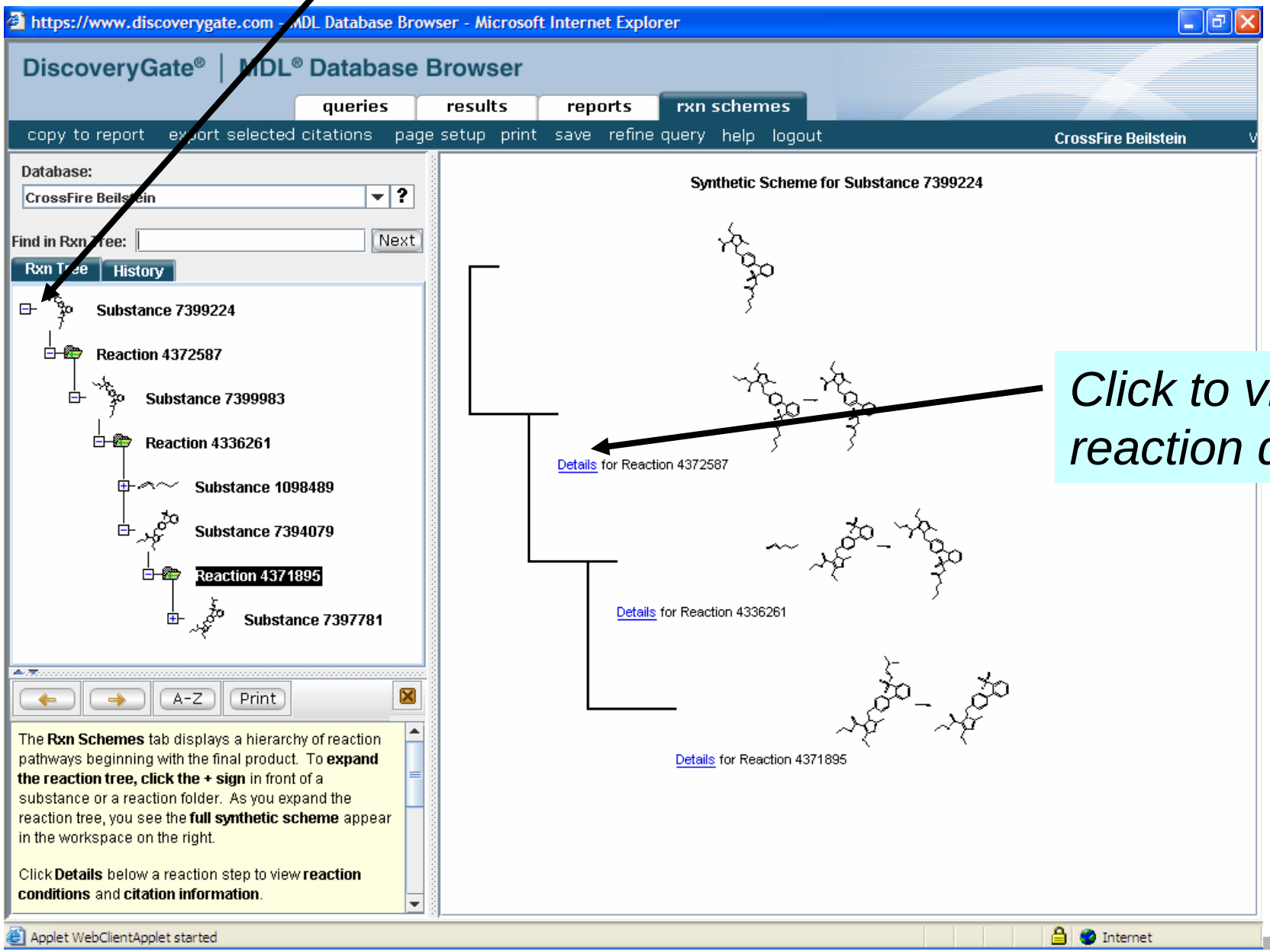
R1 = * * Br * C

*Click **Details** or **Synthesize***



***BIO** denotes bioactivity data is available*

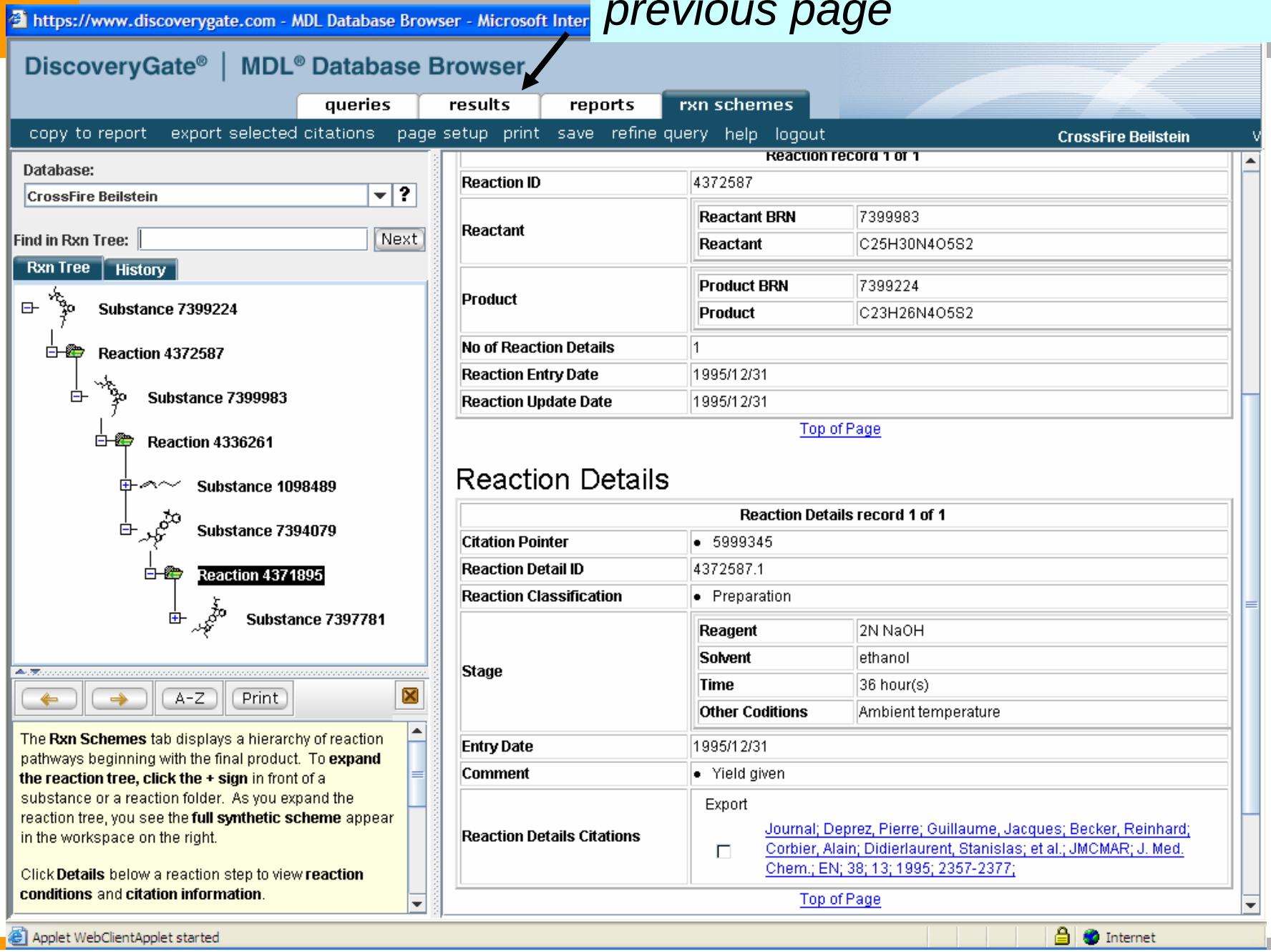
*Click on the **"+"** to uncover the* ***retrosynthetic pathway***



Click to view reaction details

View reaction details

*Click **Results Tab** to return to previous page*



DiscoveryGate

- We covered a lot of material!
- DiscoveryGate is an extremely valuable tool.
- Beilstein is the largest and most powerful source of data on DiscoveryGate for organic chemists. The additional databases create a very strong single repository of data.

- Thank you for your time and attention