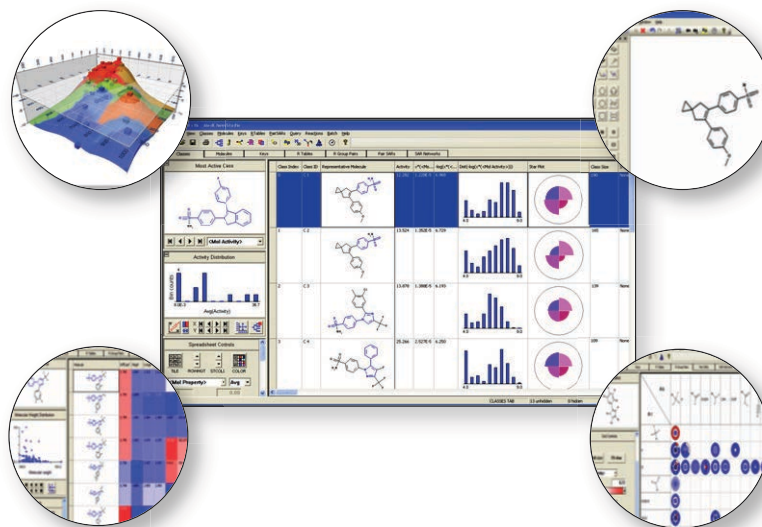




## Advanced data mining and molecule design

*Redefining your cheminformatics experience...*

- **NEW!** Advanced optical structure recognition (OSR) feature – extract structures from images
- **NEW!** 64-bit version that supports analysis of very large data sets
- Data visualization – easily browse very large data sets
- Queries: quickly search compounds by substructure or property
- Organize compounds into families using scaffolds
- R-Tables allow convenient extraction of SAR
- **NEW!** Re-number the R groups within an R table
- Standardize tautomers, fix errors, find duplicates, etc.
- Create databases in MedChem Studio proprietary format for super fast searches
- **NEW!** 10-fold increased speed for database substructure searches
- Hot URL links in spreadsheet cells
- User-qualified 2D coordinate templates when generating structure images
- Save/load equations defining molecule attributes



MedChem Studio

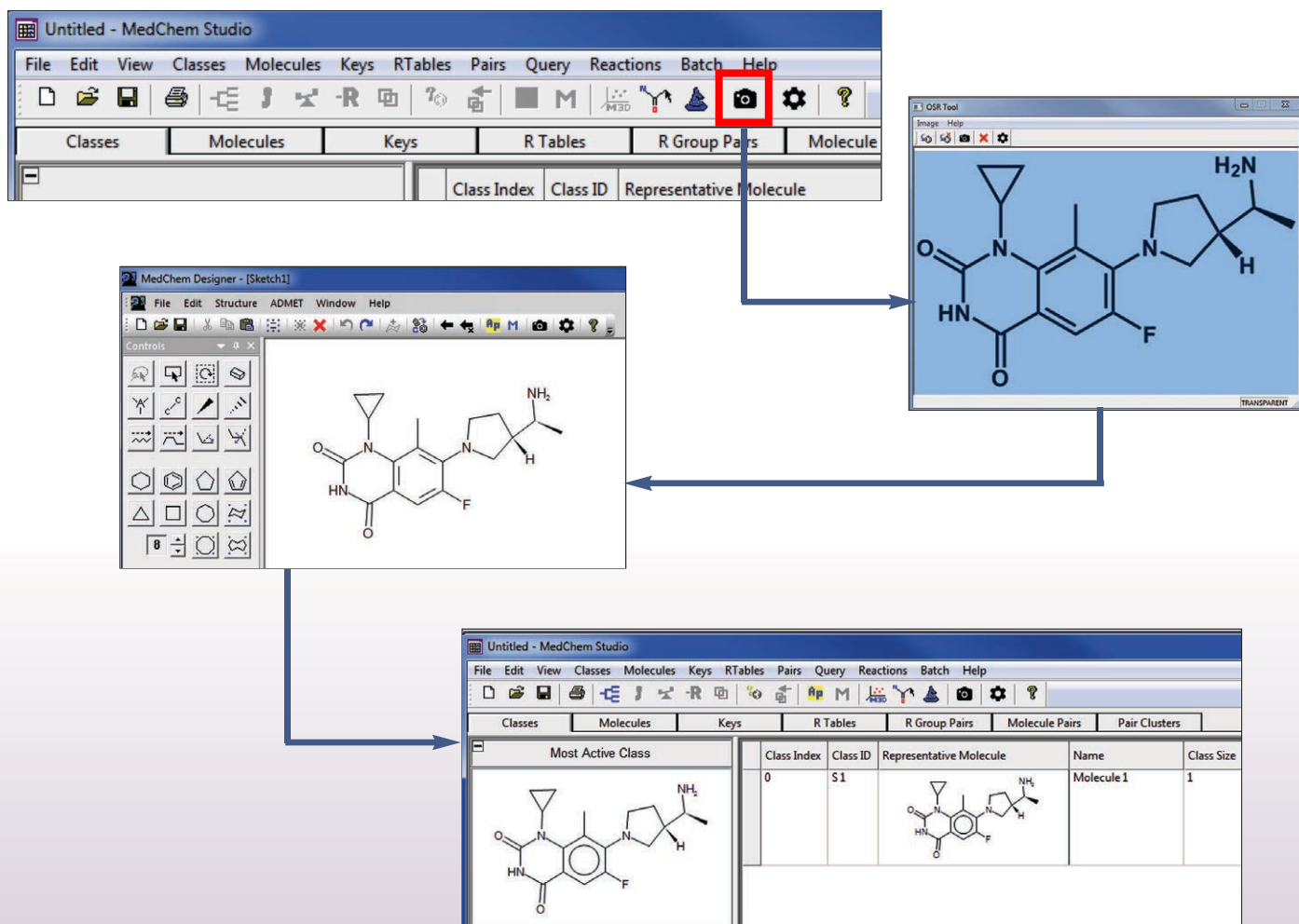


## What is MedChem Studio?

MedChem Studio is a complete cheminformatics software package containing all the tools you need for high throughput screening analysis, lead identification and prioritization, *de novo* design, scaffold hopping, and lead optimization. The "VIEWER" mode of MedChem Studio does not require a license. It allows an unlimited number of users to import structure files, including MedChem Studio's proprietary file format, in order to foster collaboration among scientists with different skill sets. For example, a medicinal chemist can view the results of a computational chemist's analyses. MedChem Designer™ is a free molecular drawing tool that can be accessed from MedChem Studio. It can be used to input or modify structures, define structure queries, visualize metabolites, and much more!

## NEW! Optical structure recognition (OSR)

Images from Word documents, PowerPoint slides, web pages, etc. can be extracted and imported into MedChem Studio using the new OSR tool. Click on the OSR icon in MedChem Studio, center the tool over the image to be extracted, send the image to MedChem Designer for editing (if needed), and then transfer to MedChem Studio.



Structure Input

MedChem Studio offers several options to input structures:

- Reads SD, SMILES, or MOL formats
- Sketch molecules in MedChem Designer™, ISIS/Draw, or ChemDraw™
- Send spreadsheet structures to MedChem Designer to be edited and re-imported

The screenshot shows the MedChem Studio interface with a table of molecules. The 'MedChem Designer' button is highlighted with a red box. A blue arrow points from this button to a separate MedChem Designer sketcher window.

Class Index	Class ID	Representative Mole
114	S 115	
117	S 118	
68	S 69	

The screenshot shows the MedChem Designer sketcher window. It features a toolbar with various drawing tools and a central canvas where a chemical structure is being drawn. The structure consists of a benzimidazole ring system connected to a side chain containing a carboxylic acid group and a methyl group. The label 'STD\_0045' is visible below the structure.

Click MedChem Designer button to launch sketcher

Sketch molecule and transfer into MedChem Studio

Data Import

MedChem Studio has a very easy-to-use interface for importing data associated with molecules in the spreadsheet:

- Data can be read from SD or tab-delimited file
- Fields can be either text or numeric
- Missing values are set to user-defined value

In this example, IC50 values for several receptor subtypes of sphingosine-1-phosphate along with rat PK data\* are read into MedChem Studio. Greater-than-sign (>) for specific IC50 values are read in without trouble.

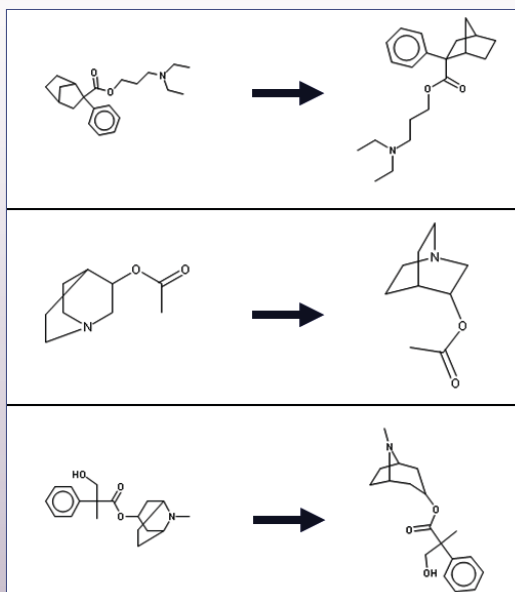
IC50_rh	IC50_Modifier	IC50_rh	IC50_Modifier	IC50
900	=	14000.0000	=	49.00000000
1	S2	0.400	=	13000.0000
2	S3	0.600	=	12000.0000
3	S4	0.800	=	4018.0000

\* Li et. al. J. Med. Chem. 48(20), 6169 - 6173, 2005

NEW! Template Library

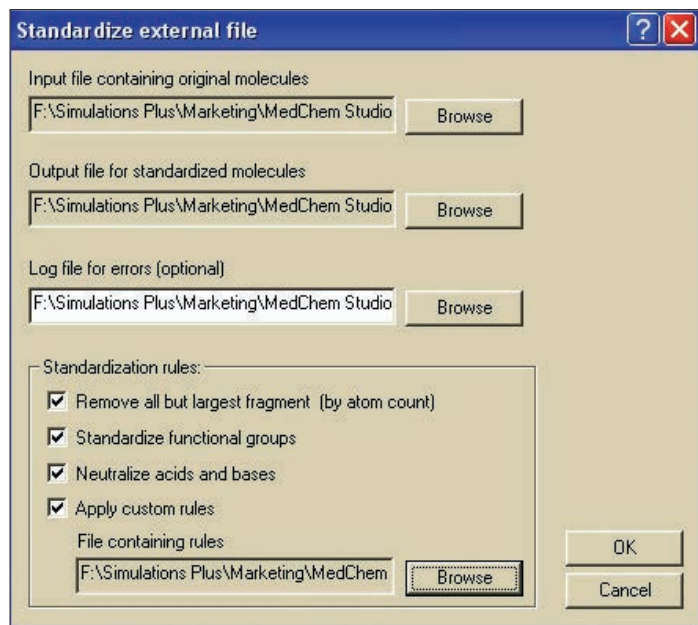
Users can define a template library to customize the way that molecules are drawn.

Without template library      With template library

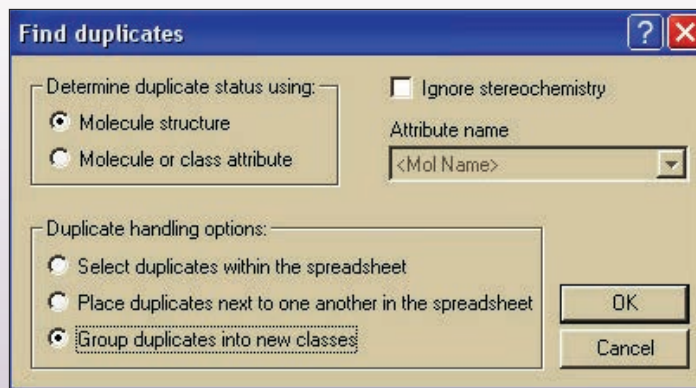
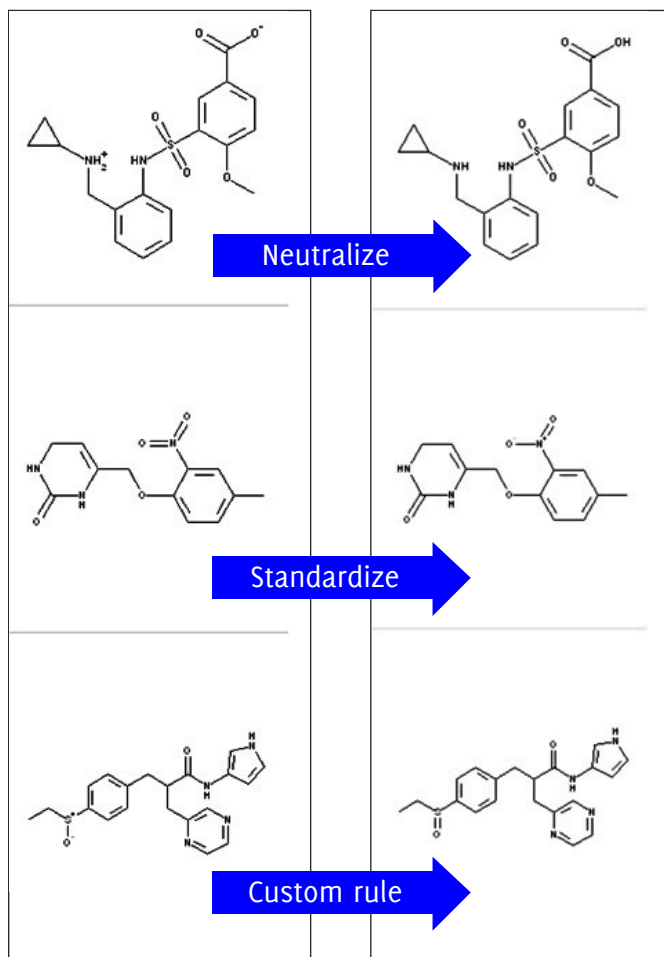


Structure Standardization

MedChem Studio can quickly “clean-up” structures so that they are represented in a conventional and consistent manner:



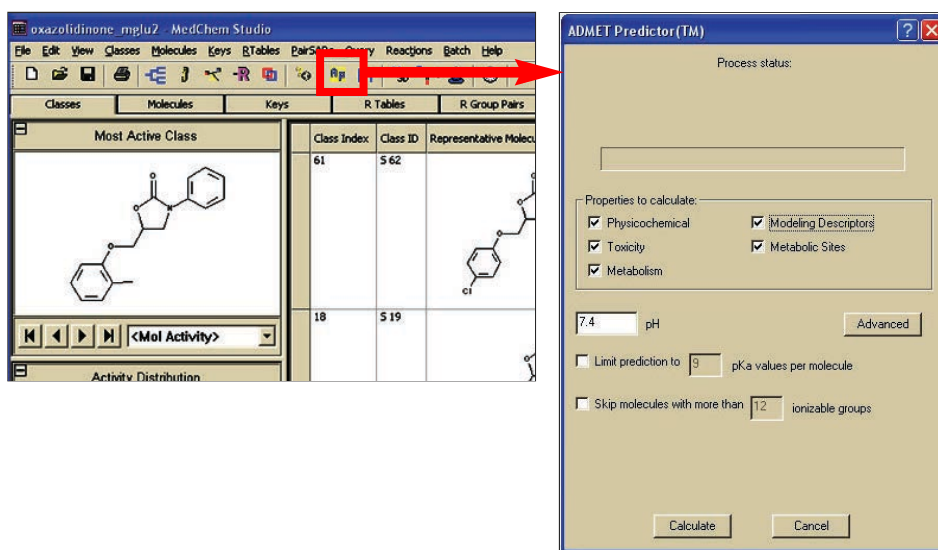
- Remove counter ions and salts from multi-component molecules
- Transform functional groups using built-in or user-provided rules
- Generate canonical SMILES strings
- Generate and export Kekule SMILES strings
- Find duplicate structures
- Perform comparison with or without stereochemistry
- Find duplicates of attributes, e.g. molecule name, CAS No.



ADMET Predictor™ Properties

MedChem Studio is integrated with ADMET Predictor, enabling rapid estimation of ADMET properties:

- Calculate ~140 ADMET properties with the click of a button

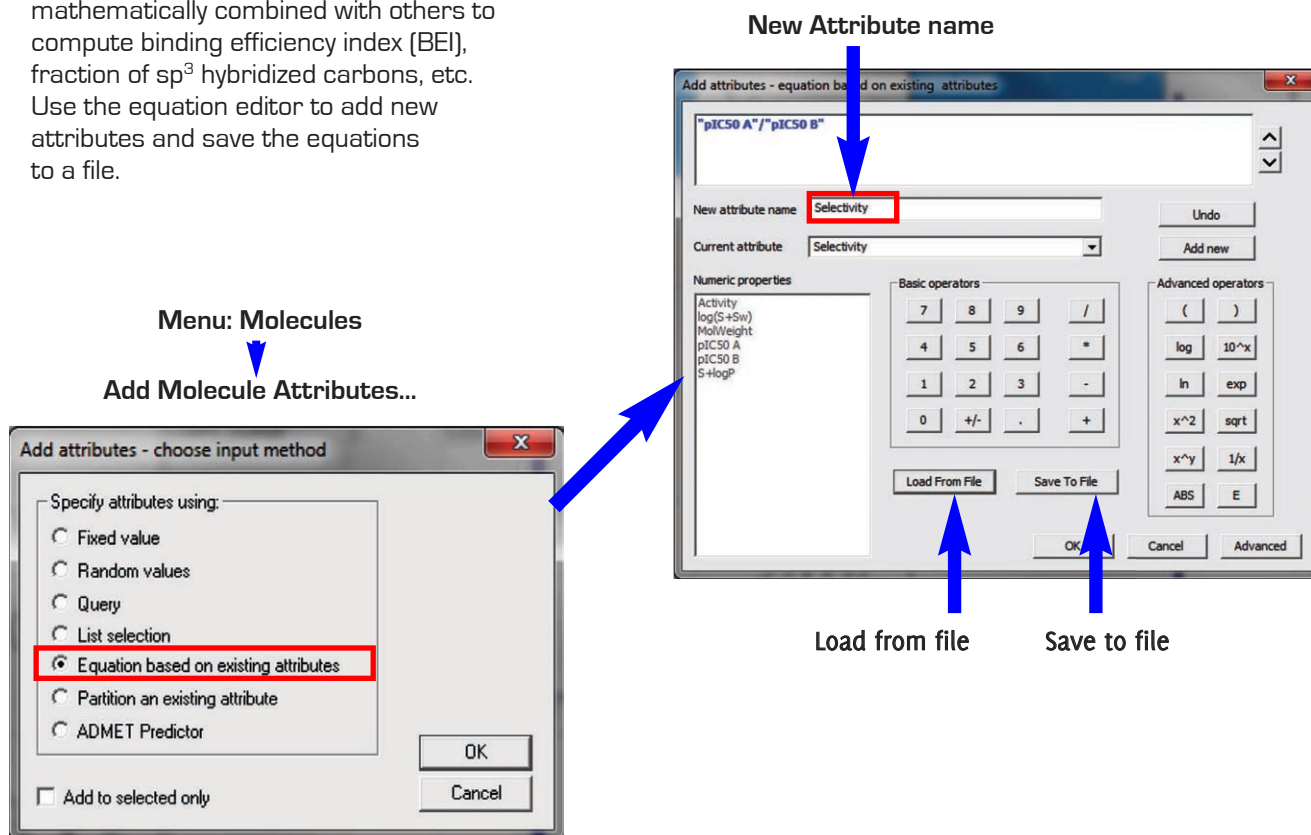


Predictions for logD (S+logD), human jejunal permeability (S+Peff), aqueous solubility (S+Sw), blood brain barrier (S+BBB\_Filter), percent unbound in plasma (S+PrUnbnd), and volume of distribution (S+Vd) were calculated in ADMET Predictor and returned to the MedChem Studio spreadsheet. (Over 130 additional columns are not visible in this screen capture.)

Representative Molecule	pEC50	S+logD	S+Peff	S+Sw	S+BBB_Filter	S+PrUnbnd	S+Vd
<chem>CC1=CC=C(C=C1)OC2=CC(=C(C=C2)C3=CC=CC=C3)C4=CC=CC=C4</chem>	4.62	4.52	High	High	High	High	1.45
<chem>CC1=CC=C(C=C1)OC2=CC(=C(C=C2)C3=CC=CC=C3)C4=CC=CC=C4</chem>	4.62	4.82	High	High	High	High	1.59
<chem>CCCCCCCC1=CC=C(C=C1)OC2=CC(=C(C=C2)C3=CC=CC=C3)C4=CC=CC=C4</chem>	4.60	4.15	High	High	High	High	1.25

Equations to calculate new attributes by combining existing ones

MedChem Studio offers several built-in attributes such as a heavy atom, bond, and ring counts. These attributes can be mathematically combined with others to compute binding efficiency index (BEI), fraction of sp<sup>3</sup> hybridized carbons, etc. Use the equation editor to add new attributes and save the equations to a file.



Matched Molecular Pairs

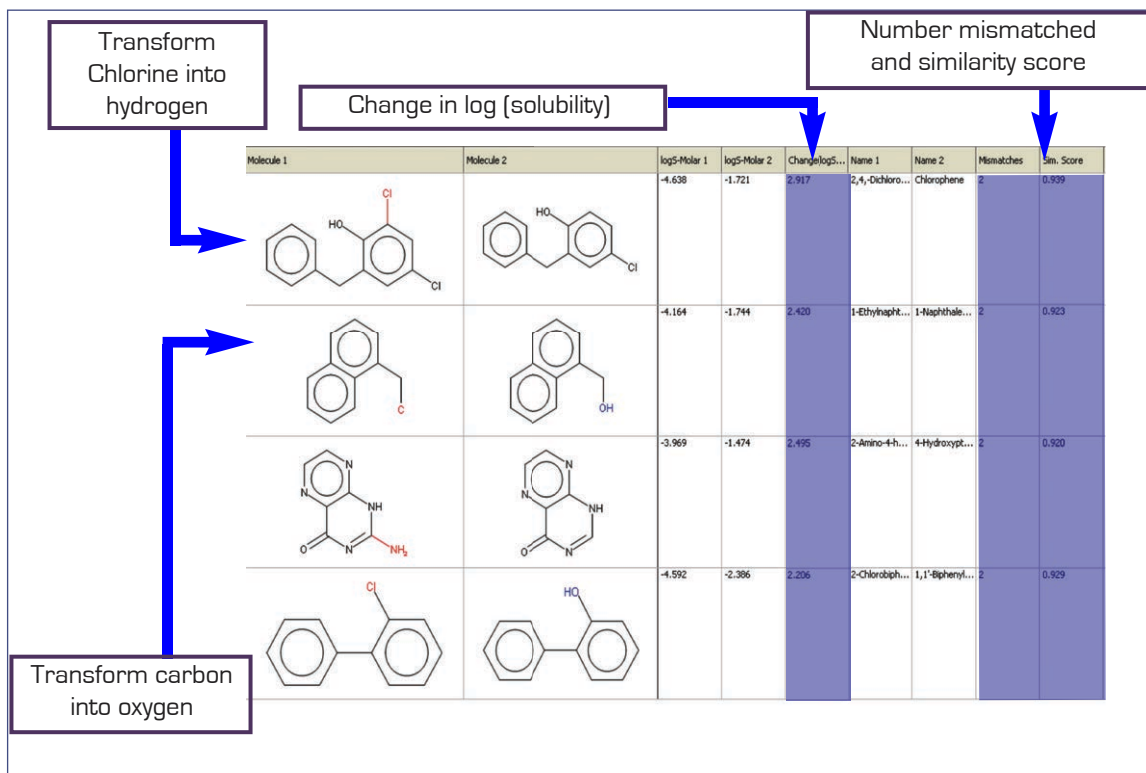
Previous versions of MedChem Studio contained a "Pair SAR" feature that, among other things, allowed one to identify activity/property cliffs. This feature has been expanded with new capabilities and renamed to "Molecule Pairs". The expanded capabilities allow matched molecular pair (MMP) analysis:

- MMPs are pairs of molecules with small, specific structural transformations
- MMP analysis (MMPA) is the automated identification of matched molecule pairs and analysis of the change in property values between the pairs
- The average change in property value is the expected size of the property value when the transformation is applied to another molecule
- MedChem Studio automatically creates a SMIRKS string that encodes each transformation rule
  - These rules can be added to our "Combinatorial Transform..." feature in order to generate novel molecules potentially containing the desired change in property value

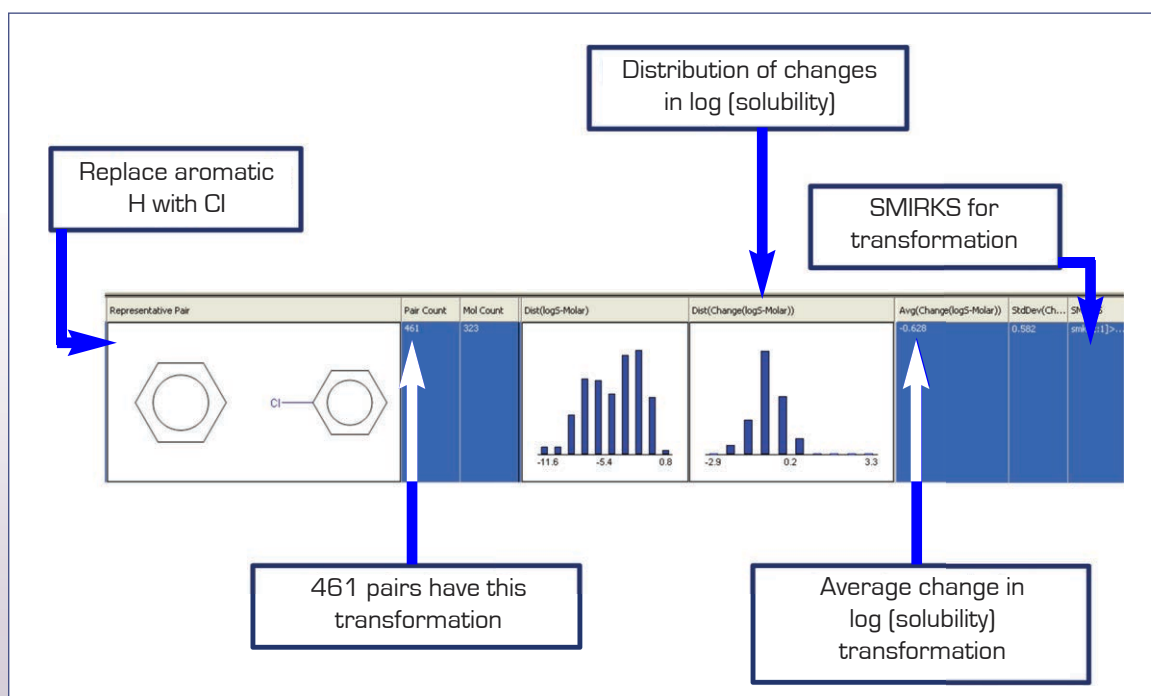
Example:

- Start with 3,485 compounds and their native water solubility
- MedChem Studio generates ~300k MMPs in ~15 minutes on a quad-core Intel Core i7 processor
- A few example matched molecular pairs are shown on the following page

Matched Molecular Pairs *continued*



Molecule pairs with the same transformation can be combined in order to make generalizations about the property change. One can also examine molecule pairs that violate the general trend.



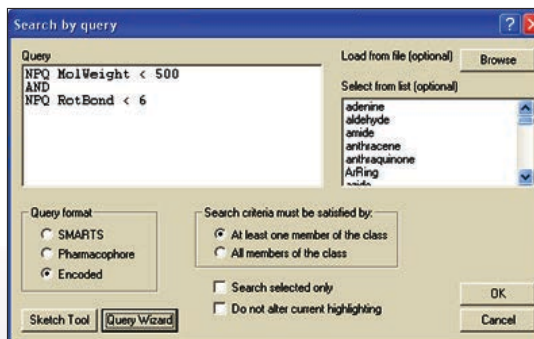
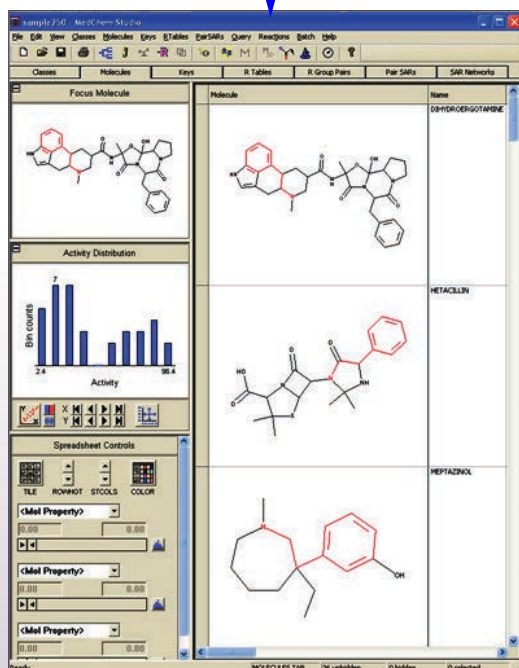
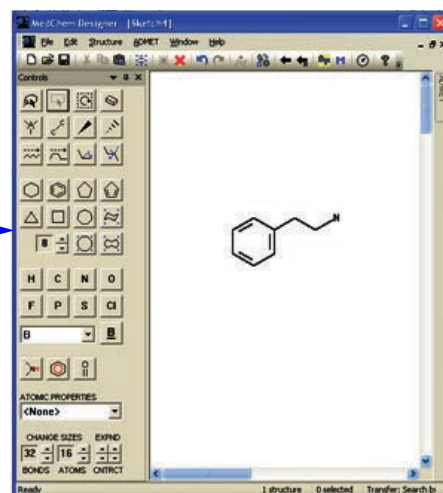
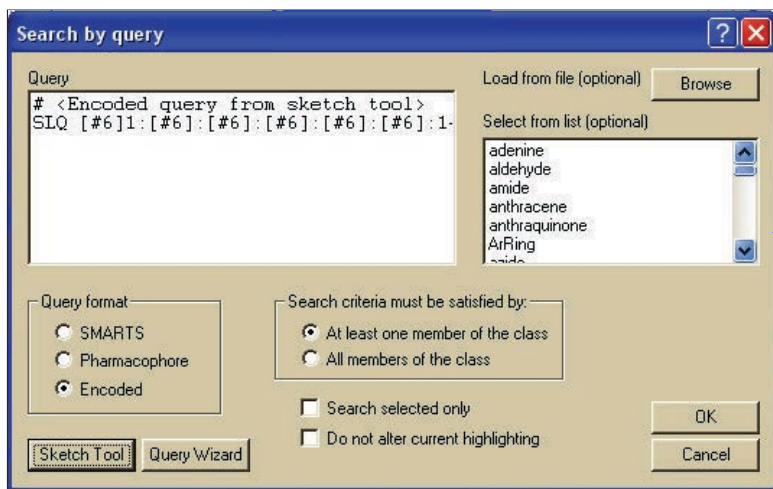
Example of solubility changes when an aromatic hydrogen is replaced with Cl.



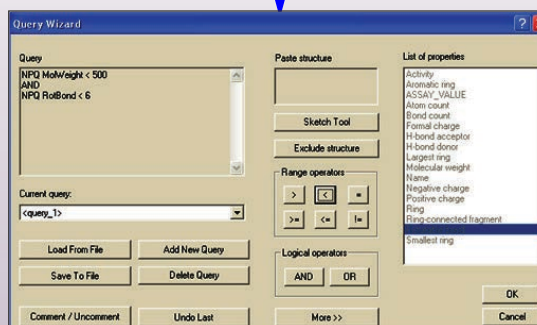
Queries

MedChem Studio contains a powerful and robust query language that offers multiple options for selecting molecules containing specific molecular features and properties:

- Substructure queries can be sketched using our MedChem Designer™ software or popular sketchers
- Property queries can be encoded with our easy-to-use Query Wizard
- MedChem Studio has an internal list of special named queries such as aromatic ring count, atom count, and largest ring size
- Queries can be combined using logical operators

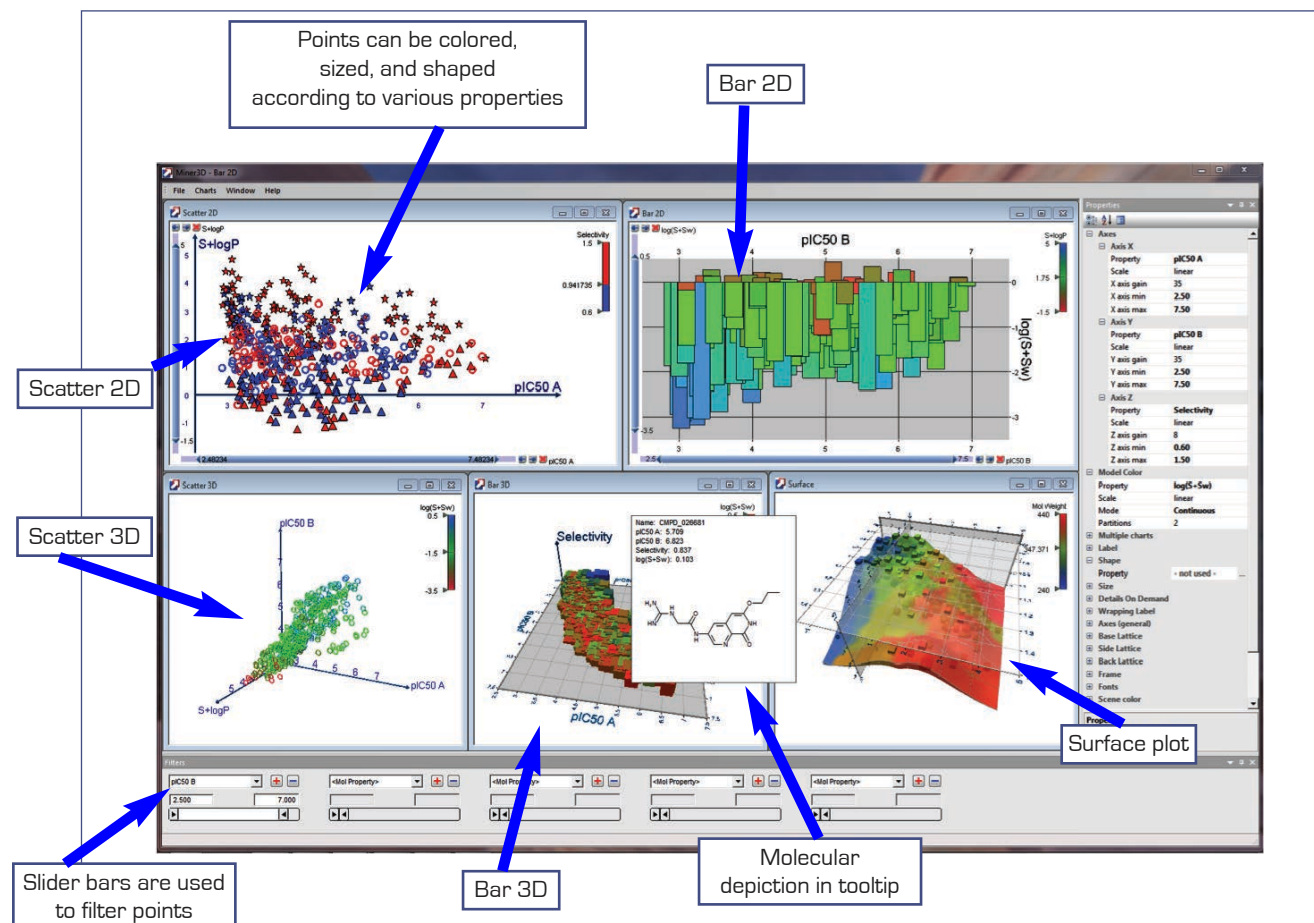


This query is used to find molecules having a molecular weight less than 500 and fewer than 6 rotatable bonds



## Enhanced Miner 3D

MedChem Studio's powerful plotting package has been expanded with 2D scatter and 3D bar and surface charts.

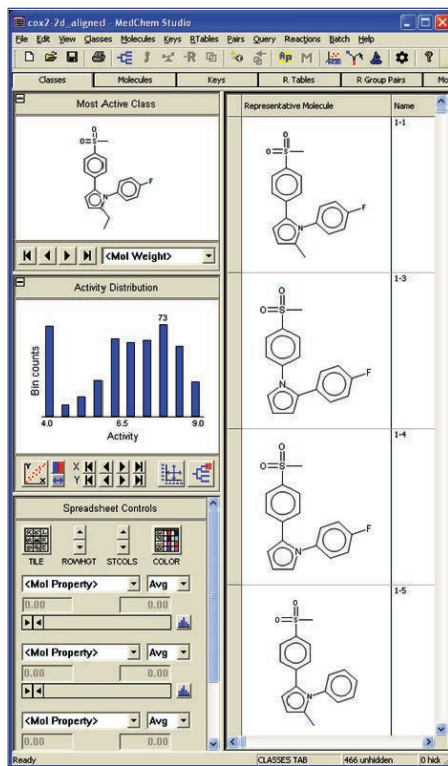


## Generating Classes

MedChem Studio's class generation clustering technology reproduces the chemist's reasoning by automatically organizing molecules into chemical families based on shared scaffolds:

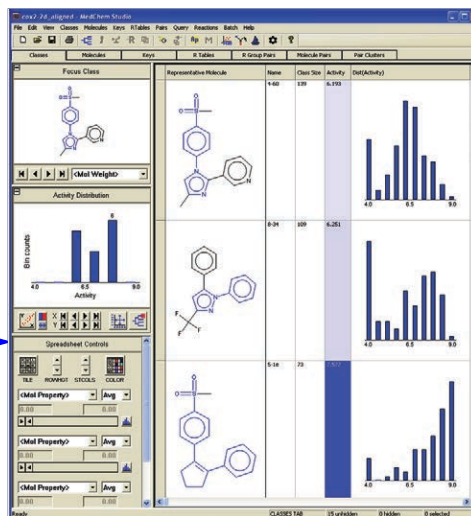
- Uses maximum common substructures (MCSs) rather than fingerprints, so results are chemically intuitive
- Generates scaffolds from the data rather than reading them from a predefined list, so novel structural motifs can be discovered
- Provides an ideal starting point for local QSAR generation, molecule design, and other analysis tasks
- User-defined scaffolds can also be used to generate classes
- "Frameworks" class generation option similar to the Murcko assemblies method, and it's fast!

Intuitive Class Generation



466 ligands with COX-2 IC<sub>50</sub> values

Example: MedChem Studio was used to group 466 COX-2 inhibitors with IC<sub>50</sub> values between 1 nM and 111 μM (pIC<sub>50</sub> between 3.9 and 9.0) into 15 families containing common scaffolds. The largest class contains 139 compounds and two classes are singletons.

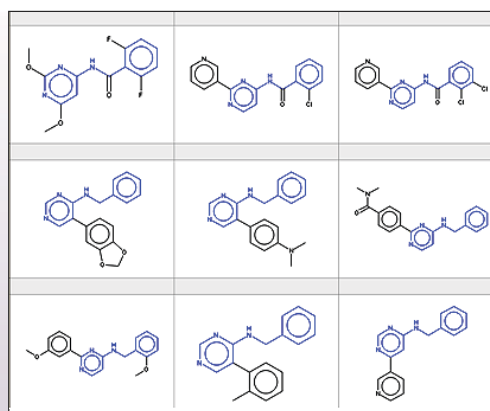
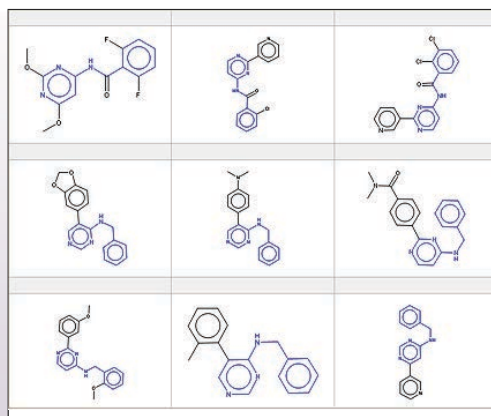


Each row corresponds to a unique class. A representative molecule is displayed with atoms denoting the scaffold highlighted in blue. Column values can be set to average, minimum, maximum, or a variety of other statistical functions.

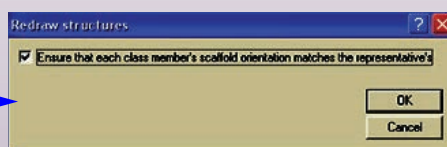
Aligning Molecules

MedChem Studio aligns molecules in a class relative to a reference molecule:

- Easily identify structural changes from one molecule to another
- The structures are redrawn – not just rotated



Example: All molecules are redrawn and oriented relative to the one in the upper left hand corner

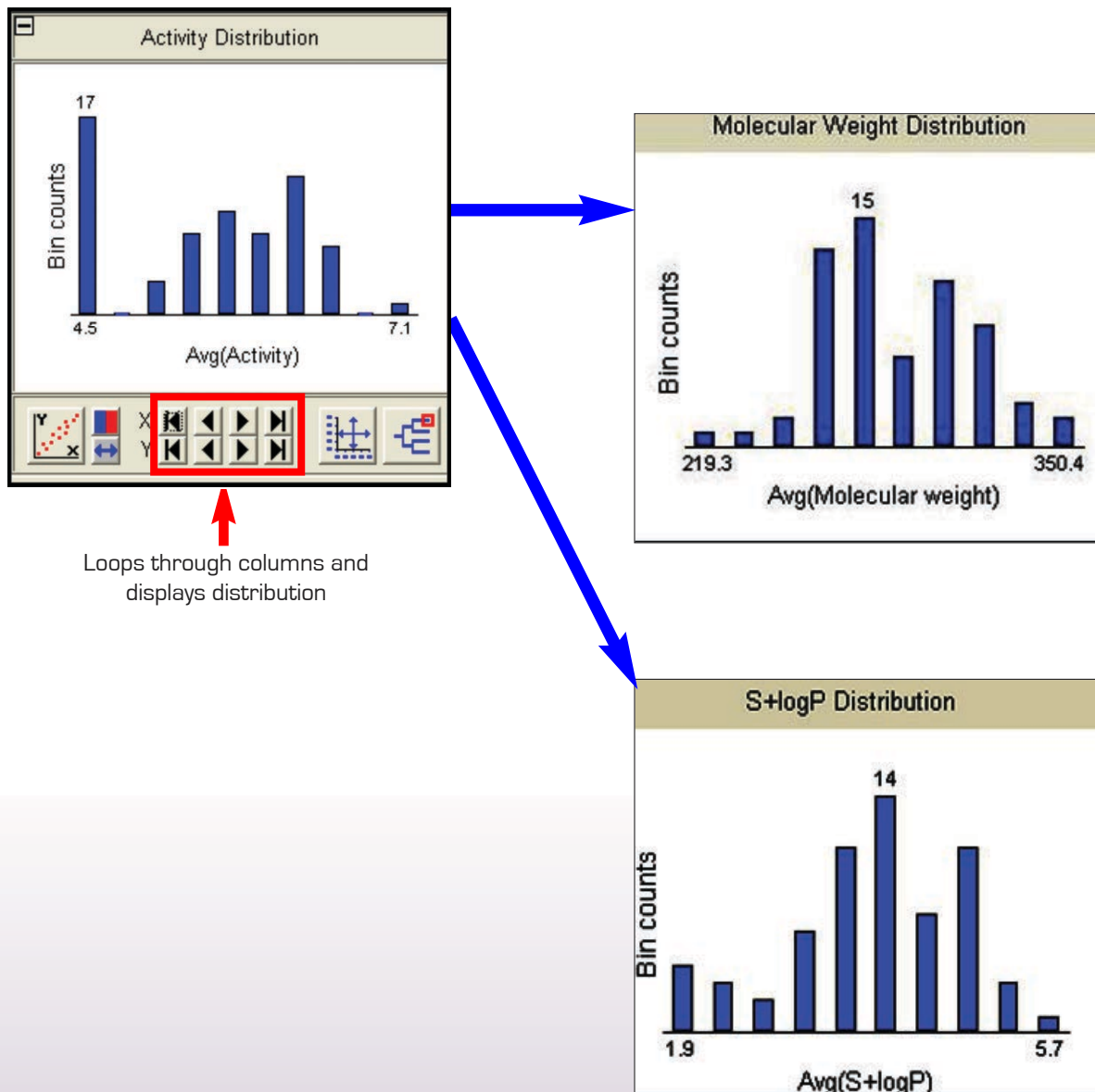


Note how the torsion angle between the two rings in blue has been modified from 0° to 180°

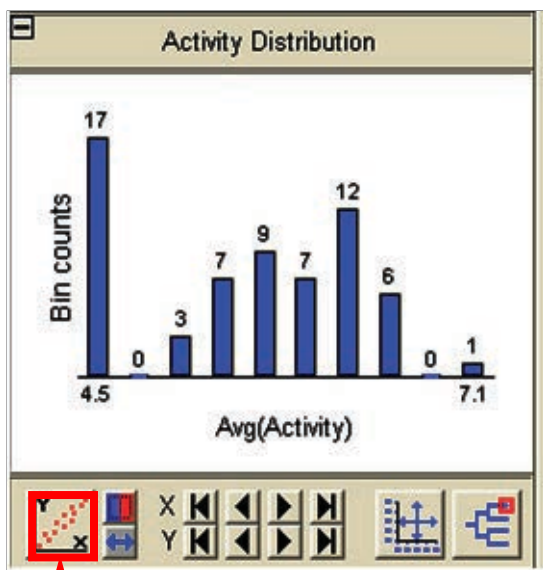
Analyzing Classes - Distribution, Heat Maps, Scatter and Star Plots

Quickly create property distributions and scatter plots

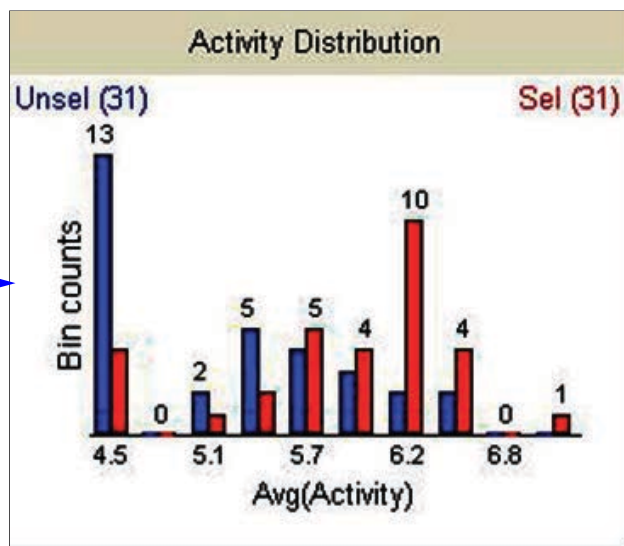
- Distribution plots to explore property ranges and distributions
- Color coding to see values of subsets



Distributions, Heat Maps, Scatter and Star Plots *continued*

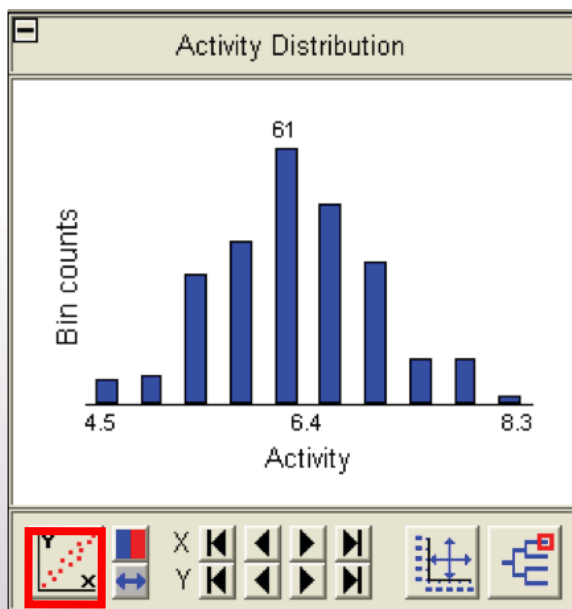


Two-color plots help uncover SAR trends

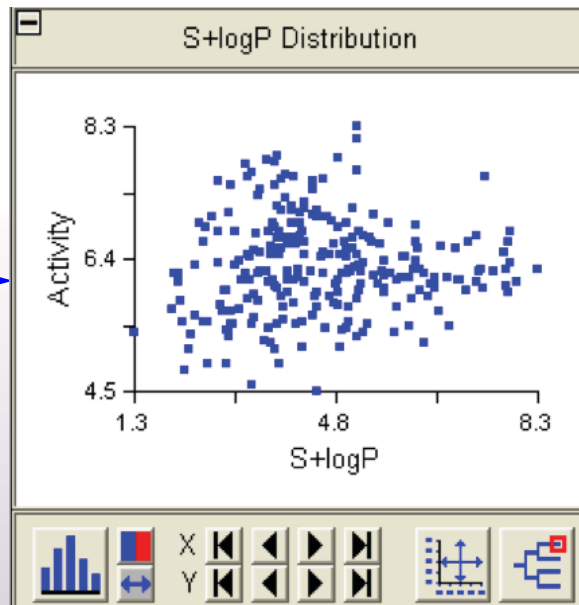


logP values  $\geq 4$  are selected. The distribution plot indicates that activity values are skewed towards compounds with higher logP.

Switch between distribution and scatter plots with a single mouse click.



Click on this button to change to scatter plots



Distributions, Heat Maps, Scatter and Star Plots *continued*

Heat maps can be created quickly

Click on "Color" button to bring up "Attribute color selection dialog box"

Representative Molecule	mGluR2-pEC50	S+logP	MlogP	S+logBBB	S+Sw	S+Sp	S+Printbind	S+logD	S+logP
	7.056	5.572	460.475	-0.451	0.031	0.311	11.270	1.743	0.704
	8.301	5.214	317.431	0.435	0.000	0.010	21.084	1.402	0.751
	6.630	4.988	384.407	-0.368	2.530E-5	2.520E-5	6.373	4.372	0.631
	6.119	4.188	172.349	-0.202	0.003	0.003	21.843	1.284	0.681
	5.369	1.119	269.305	-0.418	0.499	0.499	17.450	1.089	0.954
	5.996	7.799	475.636	1.406	0.004	0.419	1.567	0.646	0.542

Specify attributes to color

Star Plots

Star plots can depict up to 16 dimensions. This allows one to quickly visualize relative values of various properties

In this example, four properties are graphed in the star plot. Starting in the upper right hand quadrant and proceeding clockwise, the properties depicted are pEC50, S+logP, log(S+Sw) (water solubility), and T\_PSA. The first row has the highest pEC50 value, a logP between the 2nd and 3rd rows, the lowest water solubility, and highest T\_PSA.

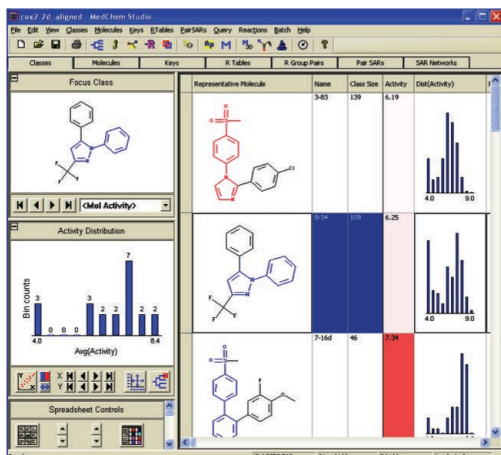
Star Plot

- pEC50
- S+logP
- log(S+Sw)
- T\_PSA

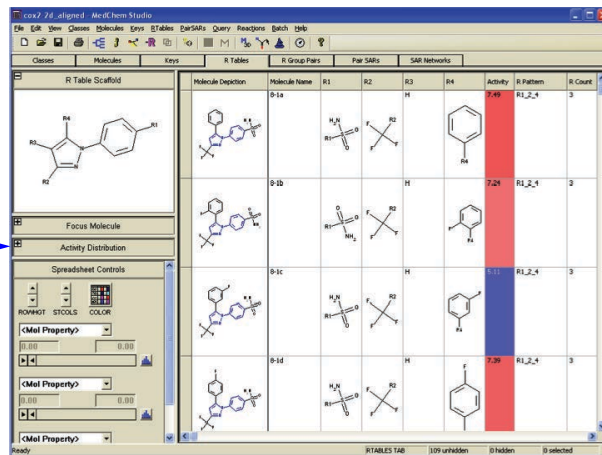
R Table Generation

MedChem Studio can build R tables after compounds are grouped into classes:

- A single mouse click is all that is needed to generate an R Table
- R groups are determined automatically based on the class scaffold
- R group columns can be sorted for easier SAR detection



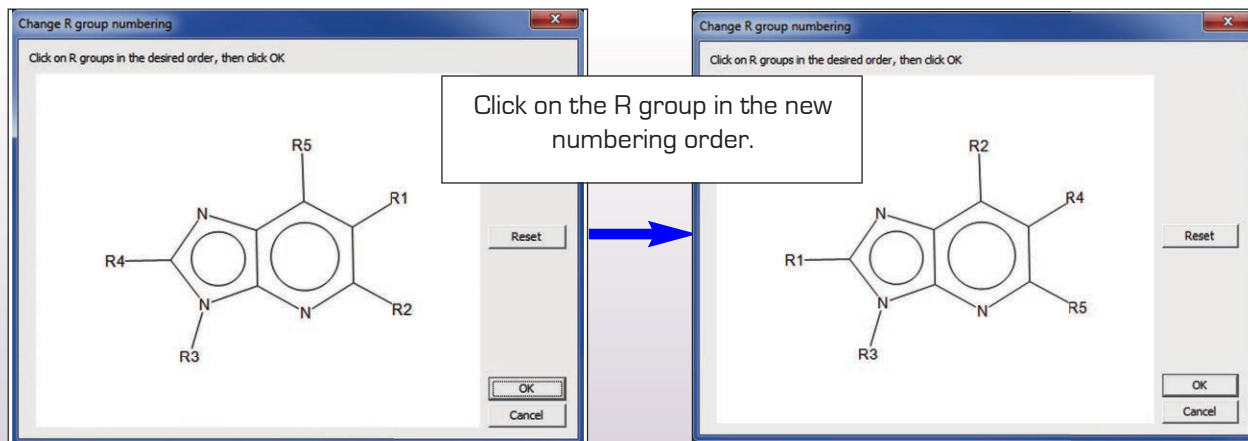
An R Table is created for the selected row



A column is created for each of the four variable positions. The R group column can be sorted to aid analysis

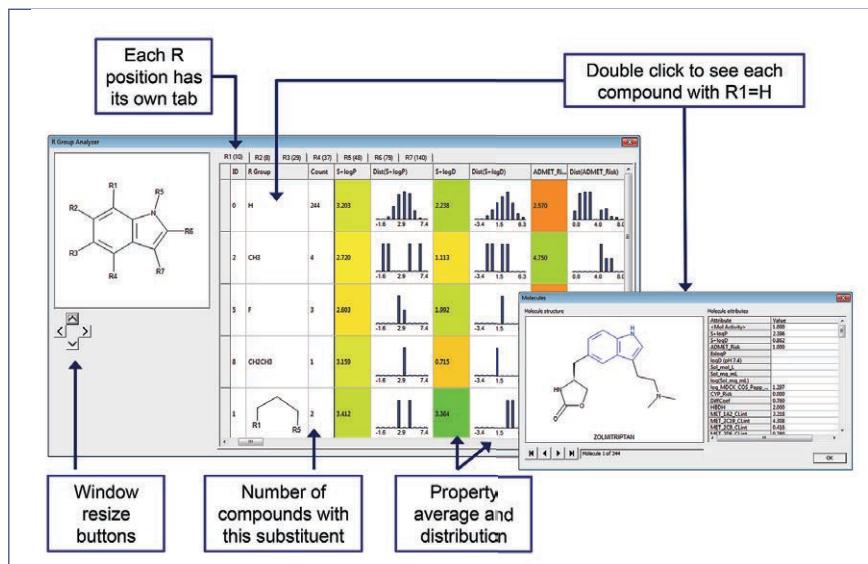
NEW! Re-numbering of R groups

R groups are automatically generated based on the scaffold and structures within a class. The R group numbers, e.g., R1, R2, etc. are also automatically generated. MedChem Studio allows users to enter their own numbering scheme. Select "Change R group numbering" from the "R tables" pulldown menu.



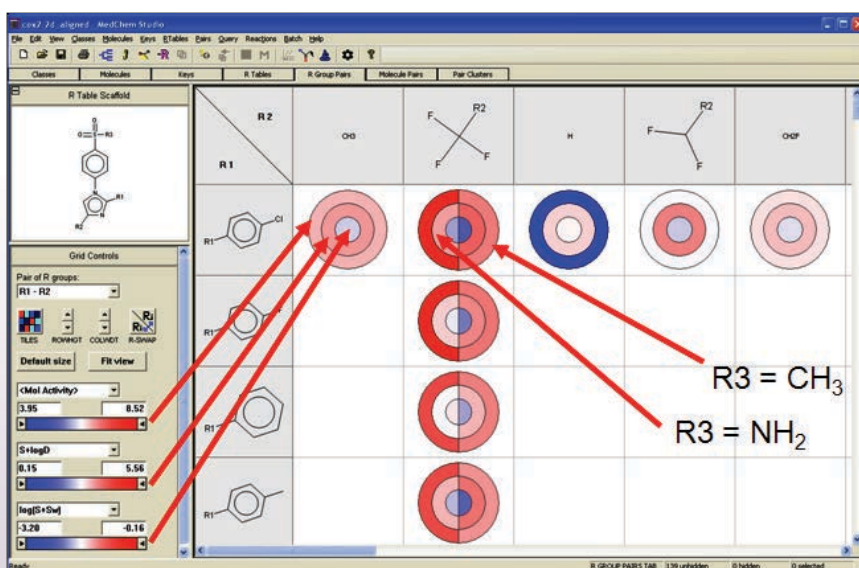
## R Group Analyzer

MedChem Studio's new "R Group Analyzer" feature converts a traditional R table spreadsheet (R's as columns) into one that shows unique substituents at each R position along with the average substituent properties (e.g., activity, logP) and distribution charts.



## R Group Pairs

- R Group Pairs analysis highlights potent pairs and unexplored territory
- Molecule property information is represented by circle charts



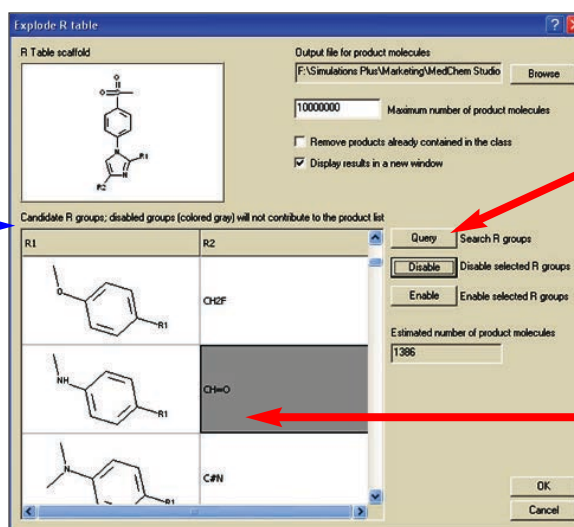
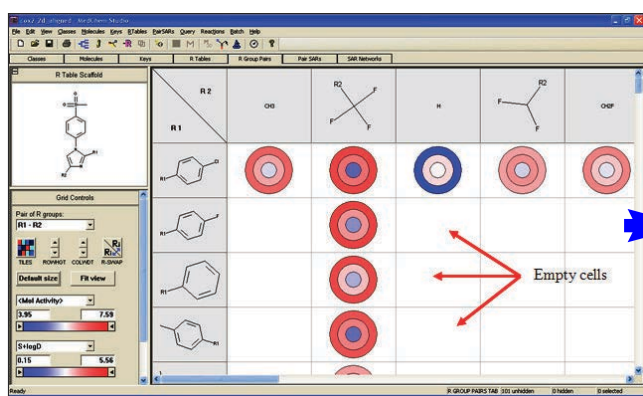
Color coded circle charts quickly confer property information. For example, the first three cells in row one (R1 = p-chlorophenyl) indicate that R2 = CH<sub>3</sub> and R2 = CF<sub>3</sub> are much more potent (outer shell) than R1 = H and their predicted logD values (middle shell) are higher. The R2 = CF<sub>3</sub> compound has the lowest predicted solubility (inner circle). Also note that when R2=CF<sub>3</sub> then R3 is either CH<sub>3</sub> or NH<sub>2</sub>.



De novo design - Generating Structural Analogs: R Table Explosion

MedChem Studio can combine every R group with all other R groups to create a large number of virtual products:

- Empty cells in an R table represent potential new molecules - the R Table Exploder can generate them automatically
- Because all new molecules use only groups that have already been synthesized at each R position, the new molecules are likely to be synthetically feasible
- An activity model based on the original class of actual molecules is likely to be accurate for molecules in the same chemical space
- Potencies and properties of these new molecules can be predicted and filtered to prioritize chemical synthesis

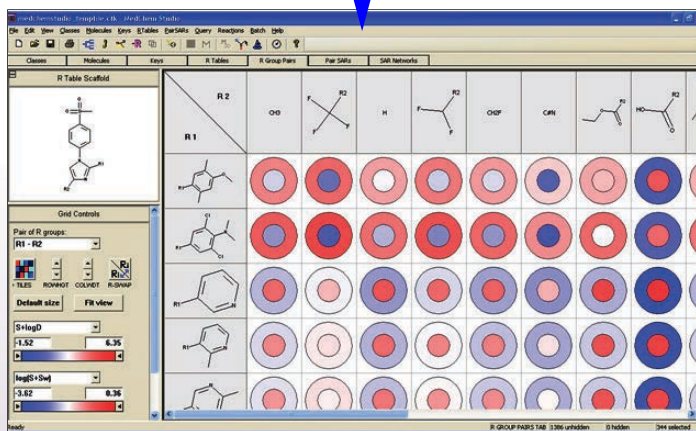


Can search for specific R-groups

User can specify which R-groups to explode

This group has been disabled

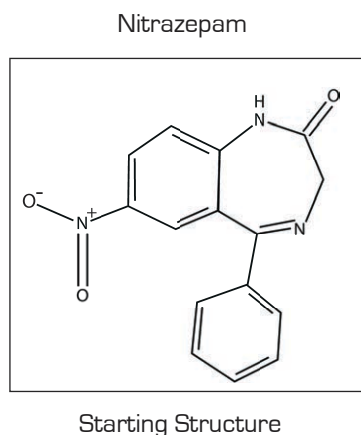
New virtual molecules have been created and their predicted properties (from ADMET Predictor™) are shown



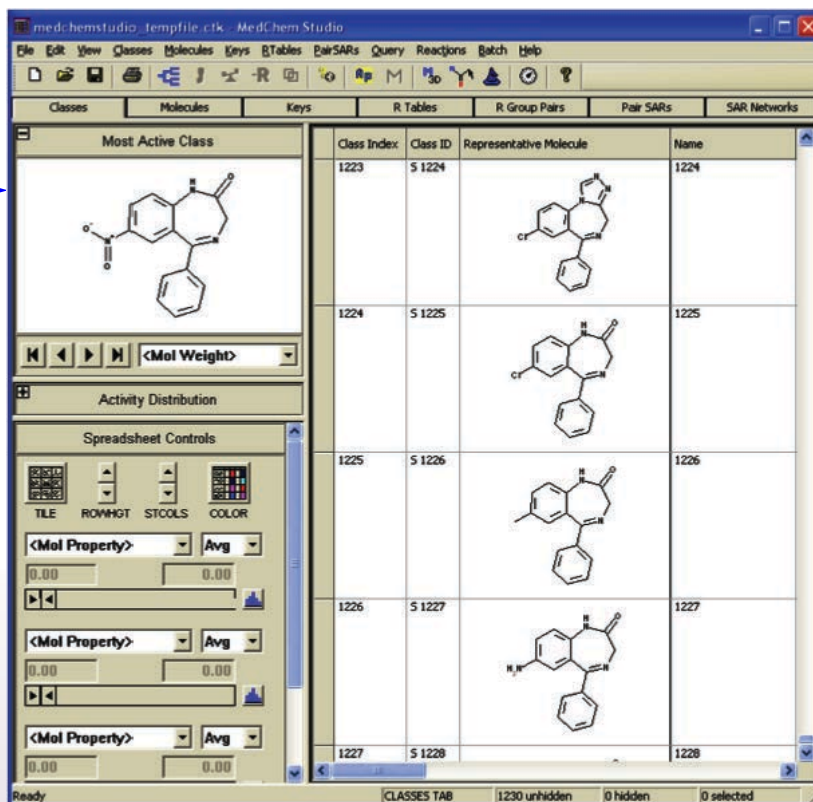
Generating Structural Analogs: Combinatorial Transformations

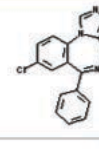
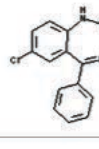
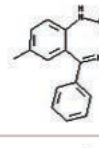
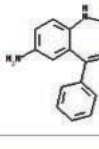
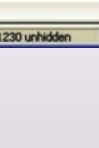
MedChem Studio provides functionality to perform *de novo* generation of new structures by applying rule-based transforms:

- Common transformations provided with software
- Common bioisoteric rules
- Fragment substitutions learned from literature or proprietary libraries
- User defined rules
- Rules can be applied multiple times
- Structural filters eliminate unwanted products



Combinatorial Transformations

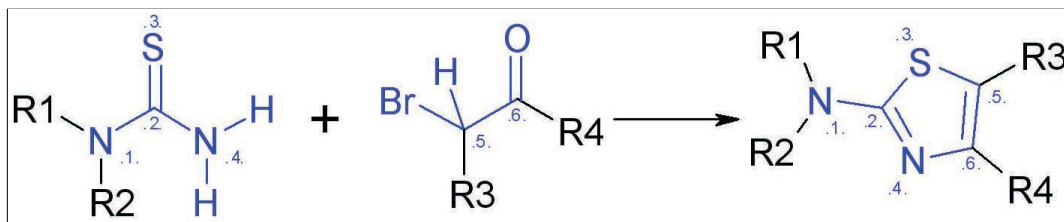


Class Index	Class ID	Representative Molecule	Name
1223	S 1224		1224
1224	S 1225		1225
1225	S 1226		1226
1226	S 1227		1227
1227	S 1228		1228

Reaction Processing

MedChem Studio provides functionality to encode and explode combinatorial chemistry reactions:

- Powerful, customizable engine allows specification of complex reactions and transformations with SMIRKS and ISIS/Draw formats
- Generate structural analogs of a lead molecule to help guide lead evolution



This example illustrates enumeration of an aminothiazole library by reacting thiourea and bromo-ketone reagents

Database Screening

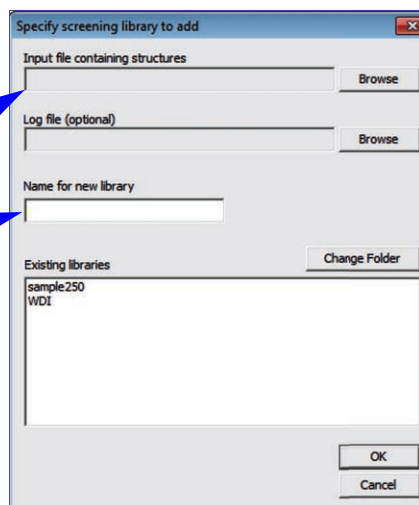
Create libraries in a MedChem Studio proprietary format for very fast MMP generation and searches:

- One-time procedure to format any in-house or commercial compound library
- Search libraries using Name/ID, structure, property, or MMP similarity queries
- Perform MMP screen against one or more desired target libraries
  - Compounds with sufficient MMP similarity are imported back into MedChem Studio
  - Structural differences are conveniently displayed, with highlighting, in the “Molecule Pairs” tab
  - Useful for assessing compound novelty and potentially for predicting side effects
  - Screen hundreds of thousands of compounds in less than a minute!

Library creation

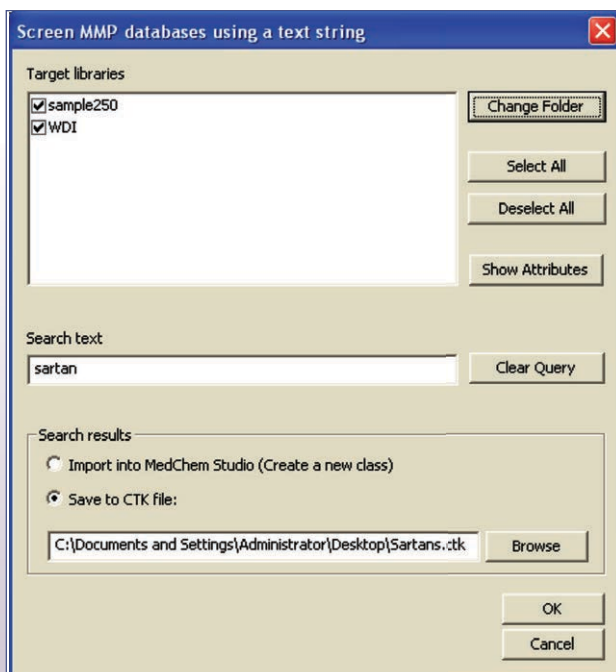
Input file can be SDF, SMILES, or .ctk format. Specify attributes in the file to store in the library.

Name for the library



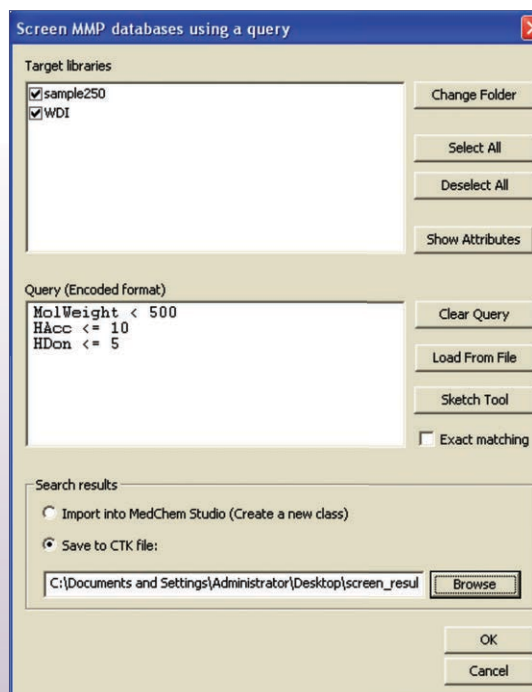
Name/ID search

Example: Find all compounds whose name contains the phrase “sartan”



Structure/Property queries

Example: Retrieve all compounds with molecular weight below 500, less than 10 acceptors and less than 5 donors



Database Screening *continued*

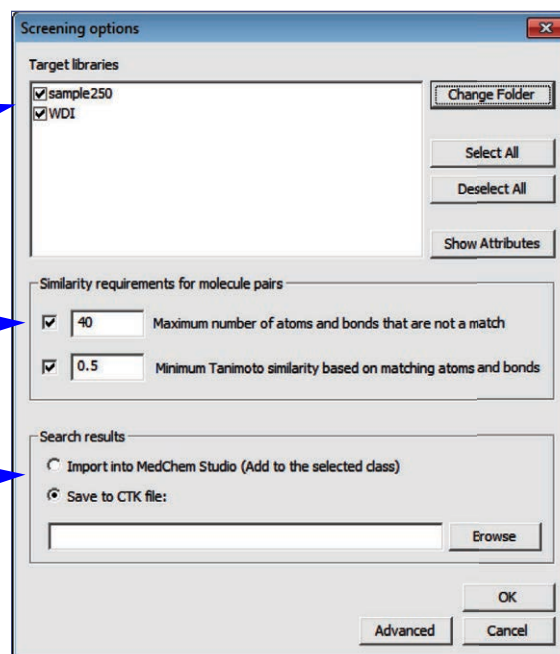
**MMP query**

Example: Retrieve molecules that have fewer than 40 atom/bond mismatches and greater than 50% similarity to a query molecule highlighted in the spreadsheet.

Libraries to search

Similarity criteria

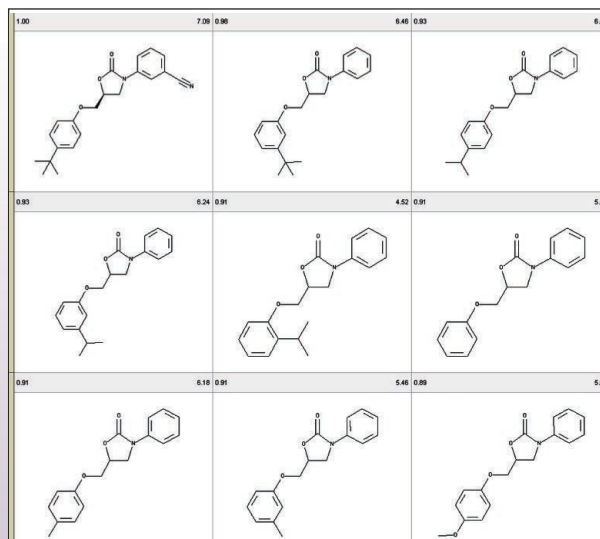
Import results into MedChem Studio or save to a file



Similarity Searching and Diversity Selection

MedChem Studio performs similarity searching or diversity selection:

- Customize scores by modifying the similarity metric (Tanimoto, shared keys, or Euclidean) or fingerprint type
- Calculate similarities using molecules in the spreadsheet or an external file
- Find molecules similar to the query molecule
- Diversity selection finds suitable screening compounds



The molecule in the upper left hand corner of the spreadsheet was used as the query for a similarity search.

Pipeline Pilot Integration

MedChem Studio's class generation is available as a Pipeline Pilot™ component:

- Generate classes while processing molecules in Pipeline Pilot
- Runs like a calculator component: each molecule in the pass port is tagged with a list of the classes to which it belongs
- Input parameters specify algorithm settings and optional additional output
- Results can be viewed using MedChem Studio, Pipeline Pilot's HTML Molecular Cluster Viewer, or other software applications

The screenshot displays the Pipeline Pilot Professional Client interface. The main window title is "Pipeline Pilot Professional Client - [MedChem Studio Class Generation (On Client) run example\*]". The interface includes a menu bar (File, Edit, View, Tools, Window, Help), a toolbar, and a "Favorites" pane. The central workspace shows a workflow diagram with three components: "SD Reader" (labeled 1), "MedChem Studio Class Generation (...)" (labeled 250), and "HTML Molecular Cluster Viewer" (labeled 250). A status bar at the bottom indicates "Finished" and "Elapsed Time: 17 Seco...".

Below the main workspace, there is a "Parameters" pane for the "MedChem Studi..." component. It lists various output destinations:

- Destination
- PropertyContainingMolName
- PropertyContainingMolActivity
- RetainAuxiliaryOutputFiles
- OverwriteExistingResultFiles
- MultipleRecordsClassMembers
- AggregateFieldClassMembers
- VariableFieldClassMembers
- MultipleRecordsClassList
- AggregateFieldClassList
- VariableFieldClassList
- ClassesByIndex
- ClassGenerationMethod

At the bottom of the interface, there is a "Help" pane with the following text:

**Structure classification of molecules using MedChem Studio**  
 MedChem Studio is intuitive software for the analysis of screening data through the identification of novel structural classes, hit-to-lead development and lead optimization. This component allows MedChem Studio to be used within Pipeline Pilot to classify chemical compounds. Benefits from MedChem Studio's proprietary classification algorithms include the following:

- Learn novel scaffolds
- Rationalize library design with scaffold diversity
- Identify mechanisms of action from the classes

**Ports:**

- **Input:** Generic Data
- **Pass:** Generic Data

Keys

MedChem Studio can be used to view molecule fragments (keys) that have been generated for a particular class. Analysis of the molecule fragments provides SAR information about the class.

This example shows a class of 139 imidazole containing COX-2 inhibitors. The selected 38 rows contain a sulfonamide highlighted in red on the molecules and the distribution graph. The unselected compounds contain a sulfone. The distribution chart shows that in general, sulfonamide-containing molecules are more active than sulfone-containing molecules.

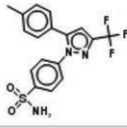
Hot URL Links in Spreadsheet

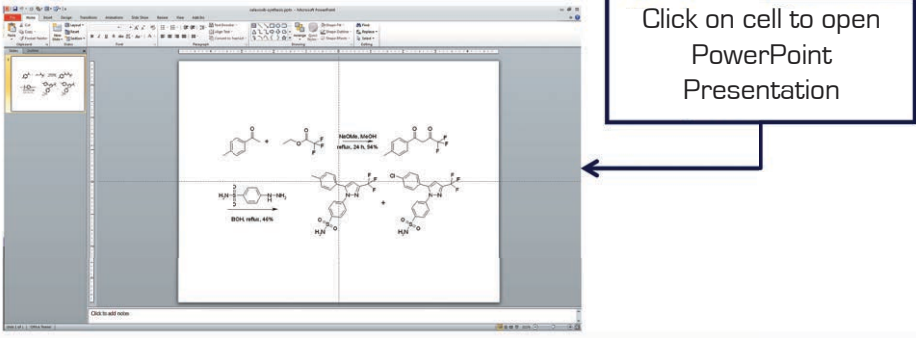
One may specify a “hot URL” link in a spreadsheet cell. In the example below, the hot link points to the FDA website containing the datasheet for belviq.

Representative Molecule	Name	API	DataSheet
<chem>C1=CC=C(Cl)C=C1</chem>	Belviq	Locaserin	<a href="http://www.accessdata.fda.gov/drugsatfda_docs/label/2012/022529bl1...">http://www.accessdata.fda.gov/drugsatfda_docs/label/2012/022529bl1...</a>

## Hot URL Links in Spreadsheet

In the example below, the hot link points to a PowerPoint file containing a synthesis scheme for celecoxib.

Representative Molecule	Name	Synthesis
	Celecoxib	<a href="file:///C:/Documents and Settings/Administrator/Desktop/celecoxib-synthesis.pptx">file:///C:/Documents and Settings/Administrator/Desktop/celecoxib-synthesis.pptx</a>

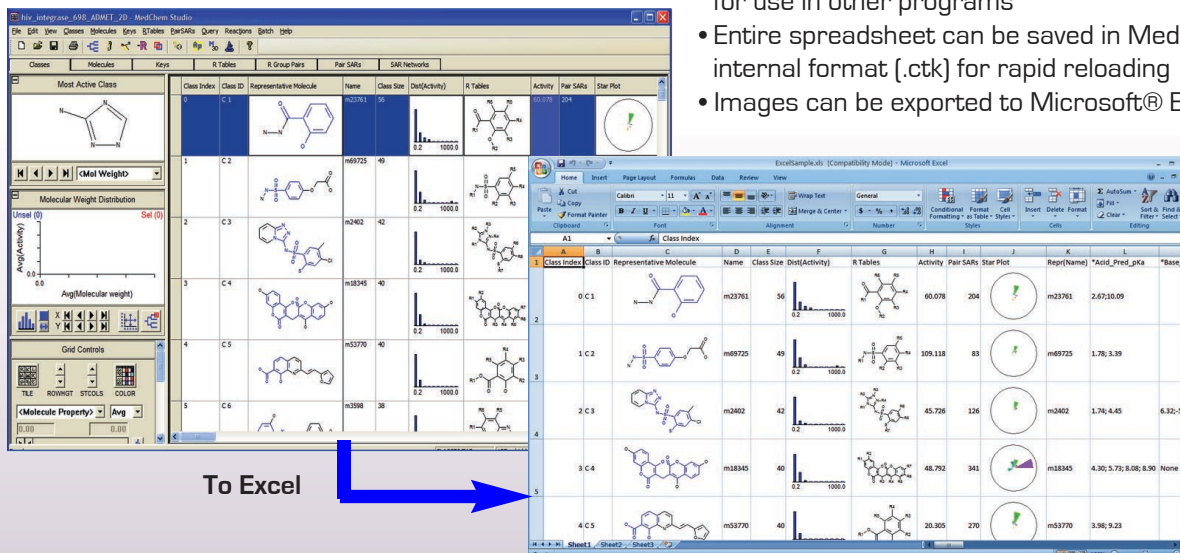
  


Click on cell to open PowerPoint Presentation









## Data Export

MedChem Studio has a large number of options for data export:

- Molecules and data can be exported as SD or SMILES files for use in other programs
- Entire spreadsheet can be saved in MedChem Studio internal format (.ctk) for rapid reloading
- Images can be exported to Microsoft® Excel®



To Excel

Class Index	Class ID	Representative Molecule	Name	Class Size	Dos(Activity)	R Tables	Activity	Pair SAs	Star Plot
1	C-1		m23761	56	66.078	204		m23761	2.67;10.09
2	C-3		m69725	49	109.118	83		m69725	1.78; 3.39
3	C-4		m38345	40	48.792	341		m38345	4.30; 5.71; 6.08; 6.90
4	C-5		m53770	40	20.305	270		m53770	3.98; 9.23

MedChem Studio™, MedChem Designer™ and ADMET Predictor™ are trademarks of Simulations Plus, Inc. • Microsoft® Excel™ is a trademark or registered trademark of Microsoft Corporation • ISIS/Draw™ is a trademark of MDL Information Systems, Inc. Pipeline Pilot™ is a trademark of SciTegic, Inc.