



ICClassify

The InfoChem Reaction Classification Program

InfoChem GmbH

www.infochem.de

Version 3.2

InfoChem GmbH
Dr. Troll-Str. 14
82194 Gröbenzell
Tel: +89 58 30 02
Fax: +89 58 03 839
email: infochem@t-online.de

InfoChem GmbH
Landsberger Str. 408
81241 München
Tel: +89 58 30 02
Fax: +89 58 03 839

Document History:

Date	Name	Document-Version	Comment
11.06.07	HM	2.9	First Draft
24.08.2007	HM	2.9.1	Visualization of ClassCodes
28.09.2007	HM	2.9.2	Corrections in chapter 2.1.2 Adding examples in chapter 2.1.3
02.04.2008	HM	3.0	Revised
21.05.2008	HM	3.1	Revised
08.10.2009	HM	3.2	Addition of Visualization Option

1.	INTRODUCTION	6
2.	REACTION CLASSCODES.....	6
2.1	Overall ClassCodes.....	6
2.1.1	Functionality	6
2.1.1.1	Level 1: BROAD	7
2.1.1.2	Level 2: MEDIUM	7
2.1.1.3	Level 3: NARROW	7
2.1.2	Generalizations	8
2.1.3	Examples.....	9
2.1.3.1	Inclusion of atoms in the Immediate Environment (Spheres).....	9
2.1.3.2	Standardization of Elements	10
2.1.3.2.1	Elimination of Hydrogen Halide	10
2.1.3.2.2	Aromatic Substitution of a Halogen by an Acetylene Group	11
2.1.3.3	Multiple occurrences of "Atom Clusters"	12
2.1.3.4	Classification of Reactions from Literature	12
2.2	Discrete ClassCodes.....	15
2.2.1	Functionality	15
2.2.2	Examples.....	15
2.2.2.1	Reaction with two Individual Transformations	15
2.3	Visualization of ClassCode Substructure	18
2.3.1	Highlighting of the ClassCode Substructure	18
2.3.1.1	Examples.....	18
2.3.1.1.1	Example 1 (see 2.1.3.1)	18
2.3.1.1.2	Example 2 (see 2.1.3.2)	19
2.3.1.1.3	Example 3	21
2.3.2	Extraction of the ClassCode Substructure	23
2.3.2.1	Examples.....	23
2.3.2.1.1	Example 1 (see 2.1.3.1)	23
2.3.2.1.2	Example 2 (see 2.1.3.2)	24
2.3.2.1.3	Example 3	26
3.	TECHNICAL PROVISION OF <i>ICCLASSIFY</i>	28
3.1	Batch Version.....	28
3.1.1	System Requirements	28
3.1.2	Contents of the Delivered ZIP File	28
3.1.3	Installation	29
3.1.3.1	Windows 2000 / XP	29
3.1.3.2	SUN Solaris	29
3.1.3.3	Linux.....	30

3.1.4	Using the Program	30
3.1.4.1	Command Line Parameters	30
3.1.4.2	Controlling the Program via XML Configuration File	31
3.1.4.2.1	Parameters in Configuration File	31
3.1.4.2.1.1	Parameter RDFDataField	32
3.1.4.2.1.2	HighlightClassCode	32
3.1.4.2.1.3	RemoveUnmarkedStructures	33
3.1.4.2.1.4	ExtractClassCodeSubstructure	33
3.1.4.3	I/O via standard input /output	33
3.1.4.4	Windows 2000 / XP	33
3.1.4.5	SUN Solaris	34
3.1.4.6	Linux	34
3.2	Windows DLL Version	34
3.2.1	System Requirements	34
3.2.2	Installation	35
3.2.3	Calling the DLL	35
3.2.3.1	Exported DLL Functions	35
3.2.3.1.1	Function for Calculation of Overall Hashcodes for BROAD, MEDIUM and NARROW	35
3.2.3.1.1.1	Description of Input Parameters	35
3.2.3.1.1.2	Description of Output Parameters	35
3.2.3.1.2	Function for Version Display	37
3.2.3.1.2.1	Description of Output Parameter	37
3.3	ISIS/Draw Addin	38
3.3.1	System Requirements	38
3.3.2	Installation	38
3.3.3	Configuration of the Addin	38
3.3.3.1	Description of Parameters in XML file	38
3.3.3.1.1	Map Options	40
3.3.3.1.1.1	VersionOperationMode	40
3.3.3.1.1.2	CenterType	41
3.3.3.1.1.3	ProductMappingMode	41
3.3.3.1.1.4	ReactantMappingMode	41
3.3.3.1.1.5	MechanismPath	42
3.3.3.1.2	Classify Options	42
3.3.3.1.2.1	VersionOperationMode	42
3.3.3.1.2.2	CenterType	42
3.3.3.1.2.3	ProductMappingMode	43
3.3.3.1.2.4	ReactantMappingMode	43
3.3.3.1.2.5	MechanismPath	43
3.3.3.1.2.6	TypeMode	43

3.3.3.1.2.7	MaxSpheres	43
3.3.3.1.2.8	AlkylProgress.....	44
3.3.3.1.2.9	NormHash.....	44
3.3.3.1.2.10	StandardizeElements.....	44
3.3.3.1.2.11	IgnoreDuplicateReactants	44
3.3.4	Using the Addin.....	44
4.	COMMERCIAL APPLICATIONS OF <i>ICCLASSIFY</i>	47

1. Introduction

InfoChem's Reaction Classification Program *ICClassify* categorizes reactions according to the type of chemical transformation they represent. The resulting reaction "Overall ClassCodes" represent the overall reaction, which may consist of several "discrete" transformations. Additionally calculated "Discrete ClassCodes", represent those "discrete" transformations.

These ClassCodes may be used to cluster large reaction databases and to select subsets of reactions or to define groups of analogous reactions in large sets. In particular, the "Overall ClassCodes" may be used to link different reaction databases.

2. Reaction ClassCodes

2.1 Overall ClassCodes

ICClassify assigns a set of numeric values called "Overall ClassCodes" to each processed reaction, characterizing the chemical transformation that occurs in the reaction.

Information on the chemical transformation is generated in three different levels of precision. *ICClassify* produces an "Overall ClassCode" for each precision level.

2.1.1 Functionality

ICClassify first determines atom-atom-mappings and reaction centers (sites) by its own built-in algorithm and uses this information to define the reaction transformation taking place.

2.1.1.1 Level 1: BROAD

For the first level *ICClassify* uses information on the reaction center (site) atoms and bonds only, and encodes them into the first ClassCodes.

A classification of reactions based on this level of ClassCodes results in a BROAD selection, i.e. all reactions having only equal reaction center atoms are grouped together.

2.1.1.2 Level 2: MEDIUM

The second level includes information on the reaction center (site) atoms and bonds and, in addition, on neighboring atoms of each reaction center (alpha position to the reaction center atoms). These combined data are encoded into the second ClassCodes.

These ClassCodes yield a MEDIUM selection, by selecting all reactions with equal reaction center atoms and equal immediate neighbor atoms.

2.1.1.3 Level 3: NARROW

On this level *ICClassify* encodes information on the reaction center (site) atoms and bonds, their immediate neighbor atoms, and on the neighbors' neighboring atoms (alpha and beta position to the reaction center atoms). This procedure results in a third ClassCode.

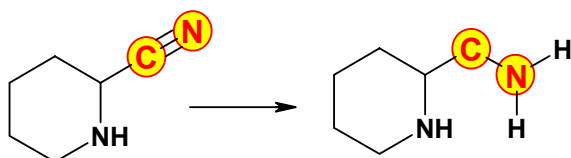
Using these ClassCodes, the selection of reactions becomes NARROW, since all selected reactions are required to have the reaction center atoms and two spheres of neighboring atoms in common.

2.1.2 Generalizations

- Within all spheres all elements of the alkali metal group (Li, Na, K, Rb, Cs, Fr) are totally standardized, i.e. all alkali metal elements are treated as one type of "alkali metal atom".
- Within all spheres all elements of the alkaline earth metal group (Be, Mg, Ca, Sr, Ba, Ra) are totally standardized, i.e. all alkaline earth metal elements are treated as one type of "alkaline earth metal atom".
- Within all spheres all elements of the noble gas group (He, Ne, Ar, Kr, Xe, Rn) are totally standardized, i.e. all noble gas group elements are treated as one type of "noble gas atom".
- Within all spheres all transition group elements are totally standardized, i.e. all transition group elements are treated as one type of "transition metal atom".
- Within all spheres for all elements of other groups only the atom types of the second period of the periodic table of the elements (B to F) are treated differently from the other members of the corresponding element group. I.e. all higher homologues within one element group are considered as one of eight "group atoms".
- All atoms, excluding hydrogen atoms, of the first sphere around reaction center atoms are used to form the "atom clusters" of the MEDIUM level.
- Building the "atom clusters" for the NARROW level by adding the neighbor atoms of the second sphere to the clusters, hydrogen atoms and regular tetravalent carbon atoms (sp^3 -atoms) are ignored. All hetero atoms and carbon atoms with multiple bonds or in aromatic rings are considered to be different from regular atoms.
- All the resulting "atom clusters" of the reactant and product site are used to generate the Overall ClassCodes for the reaction type.
- Multiple occurrences of "atom clusters" are treated as only one occurrence. That means reactions showing several identical chemical transformations will be classified as equivalent to the reaction showing the chemical transformation only once.
- If a reaction contains two or more product molecules, one set of Overall ClassCodes is generated for each product molecule. The Overall ClassCodes refer to the transformation leading to this particular product.

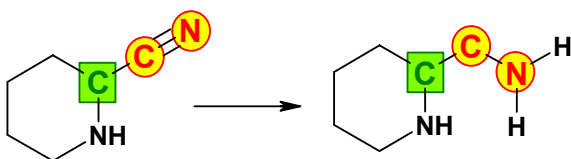
2.1.3 Examples

2.1.3.1 Inclusion of atoms in the Immediate Environment (Spheres)



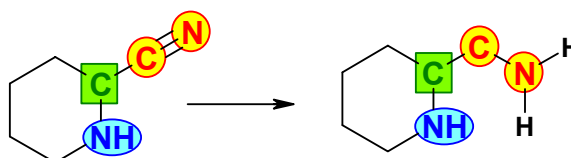
0-Sphere (BROAD)

Reaction centers only



1-Sphere (MEDIUM)

Reaction centers plus alpha atoms,
excluding hydrogens



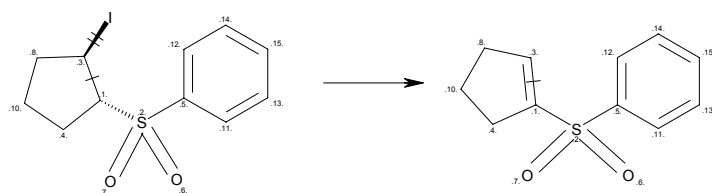
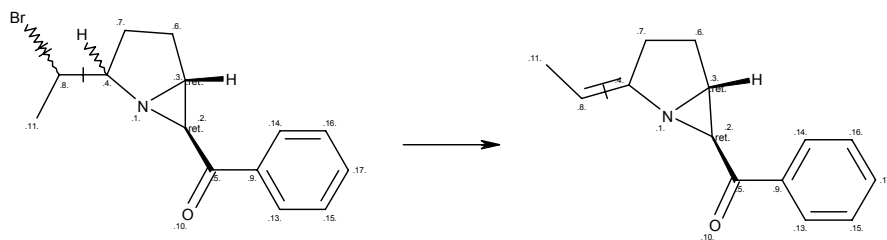
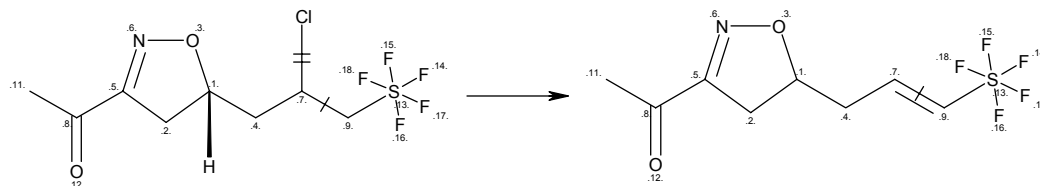
2-Sphere (NARROW)

Reaction centers plus beta atoms,
excluding hydrogens and
consecutive sp³-atoms

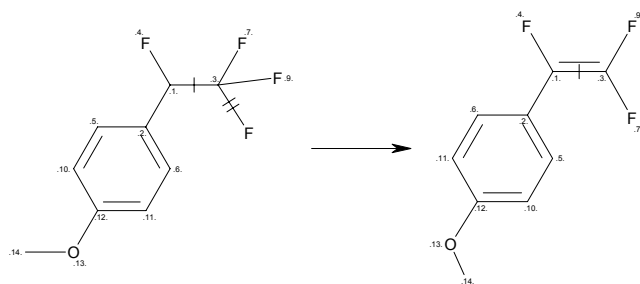
2.1.3.2 Standardization of Elements

2.1.3.2.1 Elimination of Hydrogen Halide

In the following examples from the literature, the elimination of hydrogen halides to form a C-C-double bond leads to an identical BROAD ClassCode for Chlorine, Bromine, and Iodine (BROAD Value = 256089746133396):

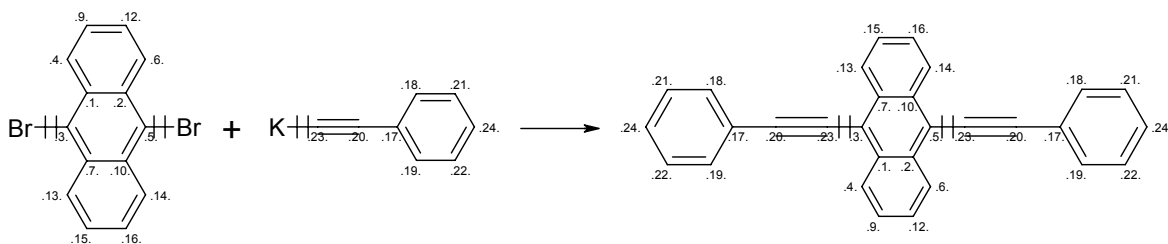
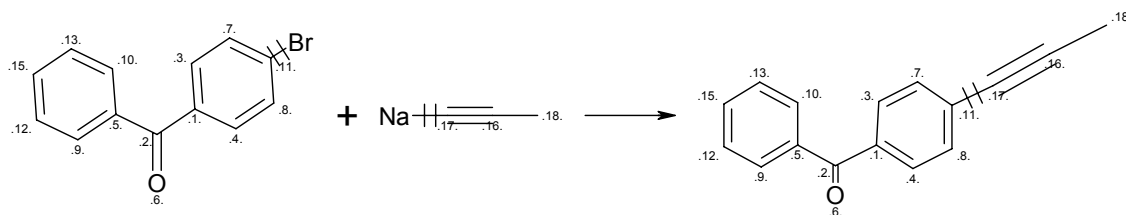
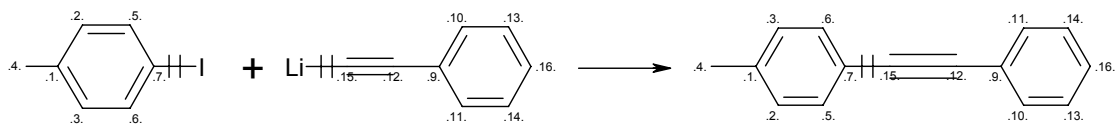


But not in case of Fluorine (BROAD Value = 256089745619557)



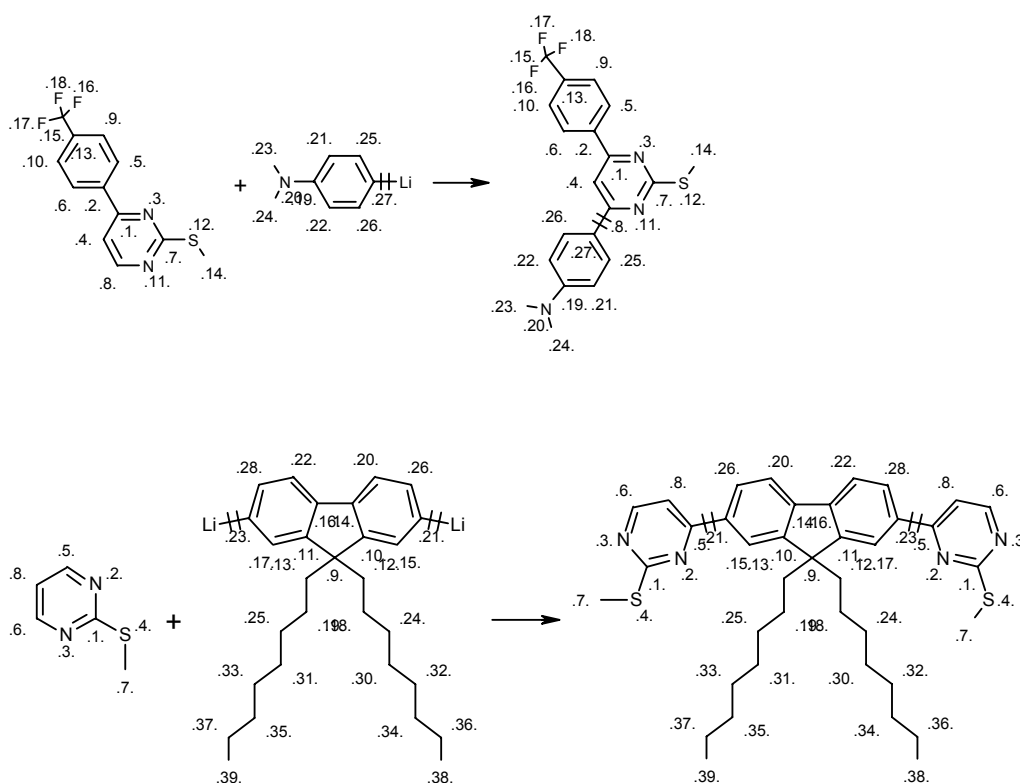
2.1.3.2.2 Aromatic Substitution of a Halogen by an Acetylene Group

In the following examples from the literature all aromatic substitutions of a halogen by an acetylene group using an alkali acetylide yields the identical BROAD ClassCode (BROAD Value = 261583166534268):



2.1.3.3 Multiple occurrences of "Atom Clusters"

In the following examples from the literature the coupling of aromatic rings with elimination of LiH occurring once or two times, yields the identical BROAD ClassCode (BROAD Value = 250137129778699)

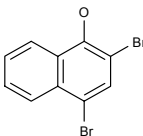
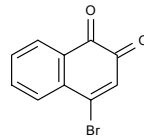
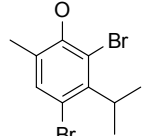
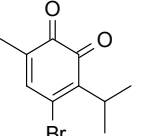
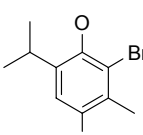
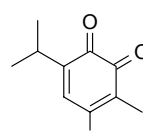
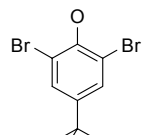
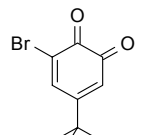
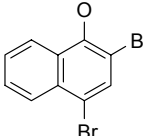
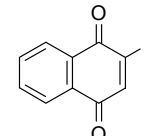
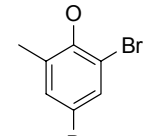
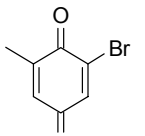
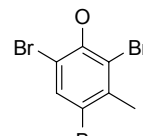
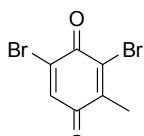
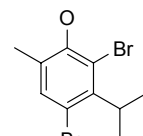
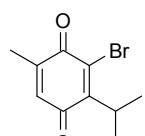
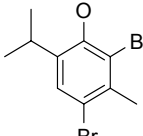
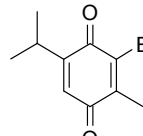
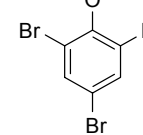
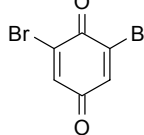


2.1.3.4 Classification of Reactions from Literature

As an example a set of 10 reactions from the literature has been processed using *ICClassify* in order to detect groups of closely related, analogous reactions, and to rearrange the list according to similarity in chemical transformations.

The classification results on three levels (BROAD, MEDIUM, NARROW) are given for each reaction in the following diagrams. For simplicity reasons the numeric ClassCodes have been replaced by one-character symbols (A, B, C, ...) to increase the readability of the diagrams.

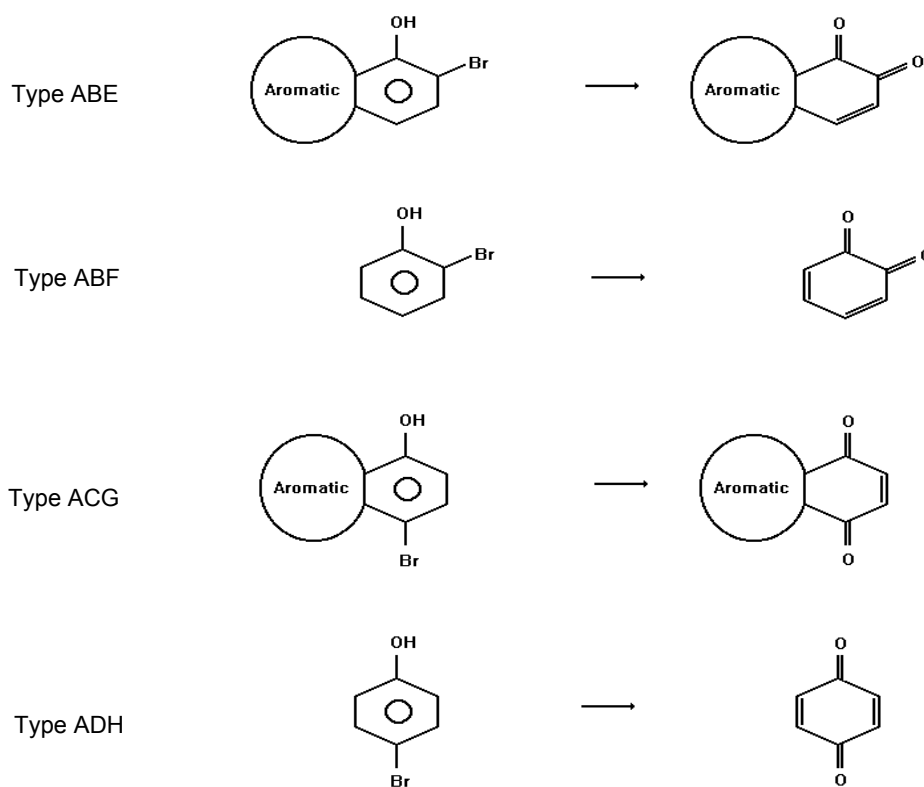
Obviously, the same symbols represent the same numeric values calculated. The example has been taken from the literature: 'A New Method for the Conversion of Halophenols and Halonaphthols to Quinones', Perumal P. T. and Bhatt M. V., *Synthesis* 1979, 205-206.

			Broad	Medium	Narrow	
1.)		→		A	B	E
2.)		→		A	B	F
3.)		→		A	B	F
4.)		→		A	B	F
5.)		→		A	C	G
6.)		→		A	D	H
7.)		→		A	D	H
8.)		→		A	D	H
9.)		→		A	D	H
10.)		→		A	D	H

The symbols A ... H represent complex numeric classification ClassCodes. For the example given above the following results are obtained:

BROAD: 1 Cluster (A)
MEDIUM: 3 Cluster (B, C, D)
NARROW: 4 Cluster (E, F, G, H)

Evaluating the classification ClassCodes for the level 'NARROW' four different reaction types are identified:



2.2 Discrete ClassCodes

For the calculation of discrete ClassCodes the same rules apply as for calculating overall ClassCodes, except that for the discrete ClassCodes, the resulting reaction center clusters will not be combined to form one ClassCode for a product of a reaction, but will be treated individually, to give a set of ClassCodes.

2.2.1 Functionality

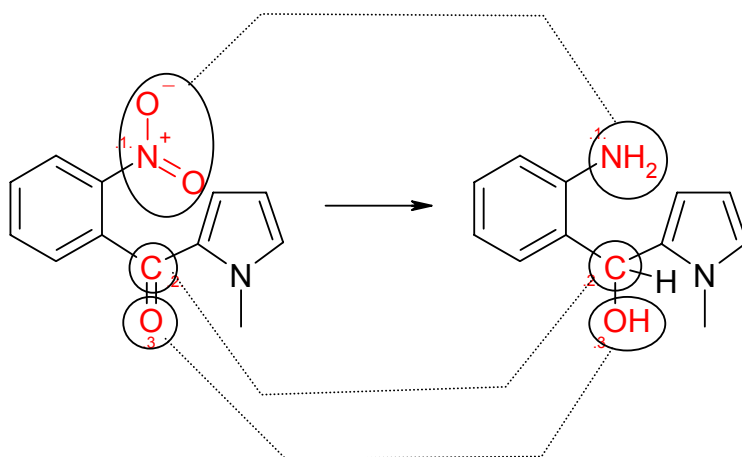
See 2.1.1.

Exception:

- All the resulting "atom clusters" of the reactant and product site are used individually to generate a set of "Discrete ClassCodes".

2.2.2 Examples

2.2.2.1 Reaction with two Individual Transformations



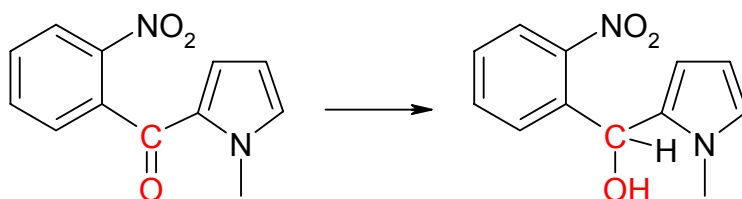
Overall ClassCode (BROAD): K

Discrete ClassCodes (BROAD):
 a (sp^2 - to sp^3 -carbon, adding Hydrogen)
 b (double to single bonded O, adding Hydrogen)
 c (Nitro group to primary amine)

In this example a carbonyl group and a nitro group are reduced simultaneously. The resulting Overall ClassCode (K) represents the combination of both transformations, by combining all reaction centers.

The resulting Discrete ClassCodes (a, b, c) each represent only one reaction center cluster, consisting of one reaction center at the product site and the corresponding reaction center(s) at the reactant site. The correspondence is derived from the atom-atom-mappings.

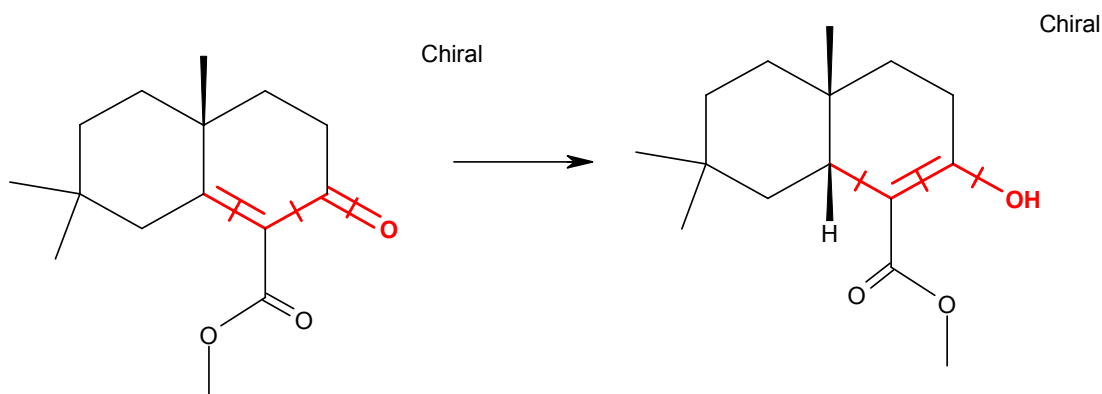
Transformation 1:



Overall ClassCode (BROAD): L

Discrete ClassCodes (BROAD): a (sp^2 - to sp^3 -carbon, addition of Hydrogen)
b (double to single bonded O, addition of Hydrogen)

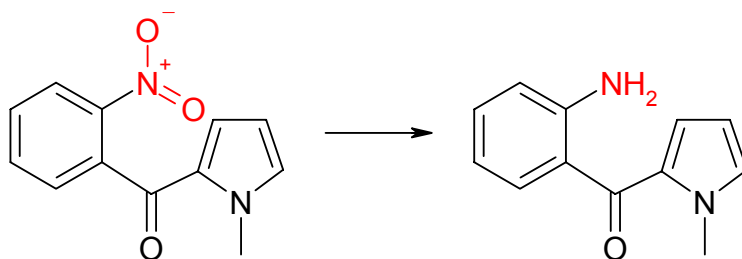
The reduction of the carbonyl group leads to two discrete ClassCodes (a, b), because the bond between carbon and oxygen is not broken and therefore not treated as reaction center. In fact not only the reduction of a carbonyl group but also the addition of hydrogen in the following example leads to an identical set of Broad ClassCodes (not for MEDIUM and NARROW):



Overall ClassCode (BROAD): L

Discrete ClassCodes (BROAD): a (sp^2 - to sp^3 -carbon, addition of Hydrogen)
b (double to single bonded O, addition of Hydrogen)

Transformation 2:



Overall ClassCode (BROAD): M (identical to Discrete ClassCode c)

Discrete ClassCode (BROAD): c (Nitro group to primary amine)

The reduction of the nitro group leads to one discrete ClassCode c. The value of this discrete ClassCode c is identical with the value of the overall ClassCode M.

A RTS-Search (Reaction Type Search) using Overall ClassCode K will only retrieve reactions, where the reduction of the carbonyl group and the reduction of the nitro group occur simultaneously without reduction of other functional groups.

A RTS-Search using Overall ClassCode L will only retrieve reactions, where solely a carbonyl group is reduced and using Overall ClassCode M will only retrieve reactions where solely a nitro group is reduced to a primary amine.

Performing a search on discrete ClassCodes will produce different results.

A RXCCluster-Search (Reaction Center Cluster Search) using discrete ClassCodes a, b and c will retrieve all reactions with Overall ClassCode K and additionally all reactions where additional functional groups will be reduced.

A RXCCluster-Search using discrete ClassCodes a and b will retrieve all reactions where at least a sp^2 -carbon atom and a double bonded oxygen atom are hydrogenated and a RXCCluster-Search using discrete ClassCode c will retrieve all reactions, where at least a nitro group is reduced to a primary amine.

2.3 Visualization of ClassCode Substructure

InfoChem offers a visualization tool to automatically highlight the ClassCode substructure for a selected product of a chemical reaction. The tool uses as input a reaction in MDL-CTFile format and returns a highlighted reaction in MDL-CTFile format (V3000).

2.3.1 Highlighting of the ClassCode Substructure

The visualization tool first highlights all reaction center atoms and bonds determined by the mapping algorithm for the selected product.

For the BROAD level, no additional atoms or bonds will be highlighted.

For the MEDIUM level additionally all atoms and bonds in the first sphere around the reaction center atoms will be highlighted, excluding Hydrogen atoms.

For the NARROW level additionally all atoms and bonds in the second sphere around the reaction center atoms will be highlighted, excluding Hydrogen atoms and sp³-Carbon atoms.

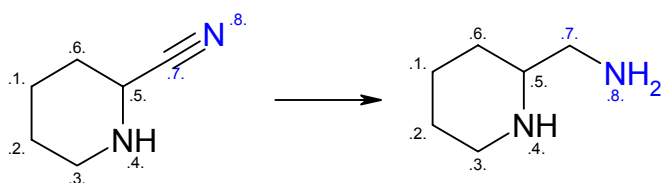
Only the highlighted atoms and bonds contribute to the ClassCodes of the reaction transformation.

The highlighted reaction will be returned without structural modification.

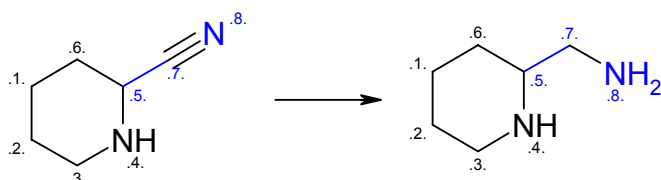
2.3.1.1 Examples

2.3.1.1.1 Example 1 (see 2.1.3.1)

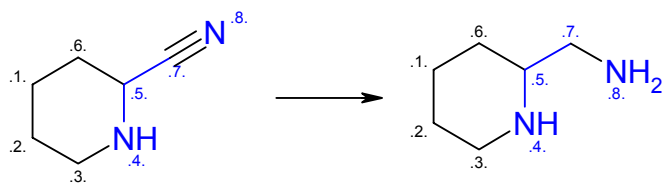
BROAD



MEDIUM

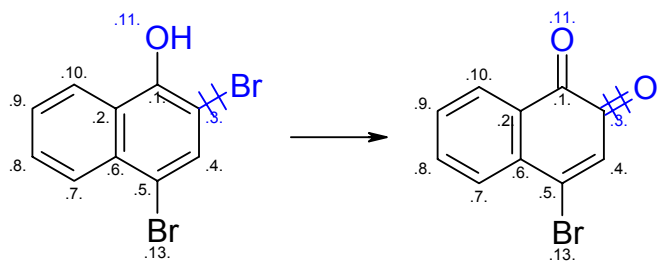


NARROW

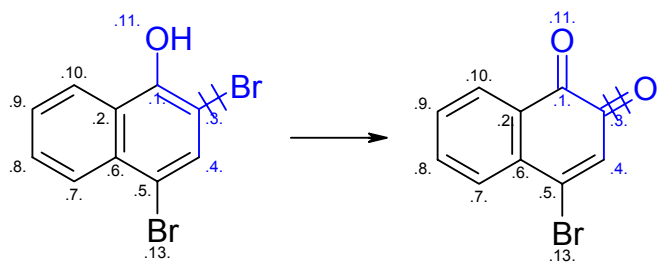


2.3.1.1.2 Example 2 (see 2.1.3.2)

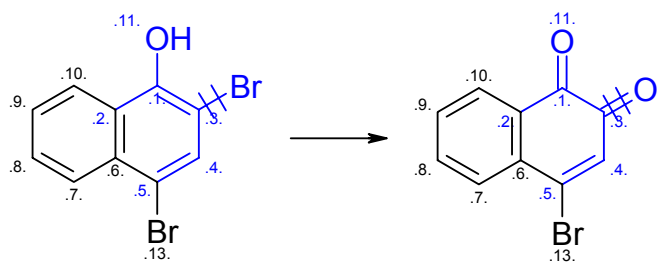
Type ABE (BROAD)



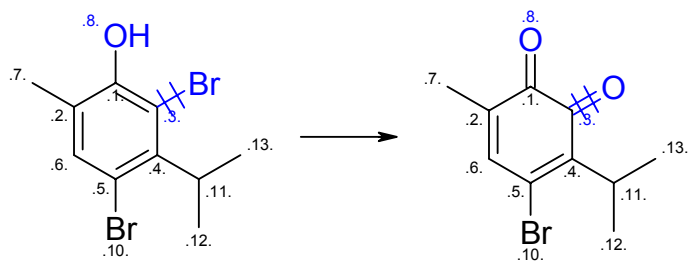
Type ABE (MEDIUM)



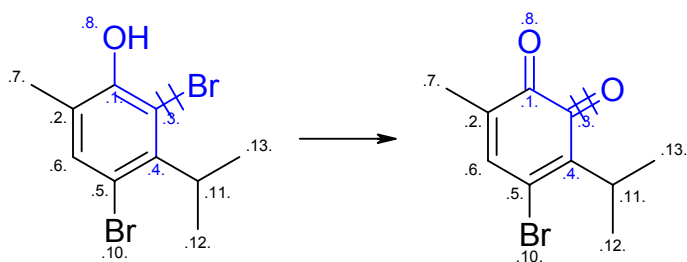
Type ABE (NARROW)



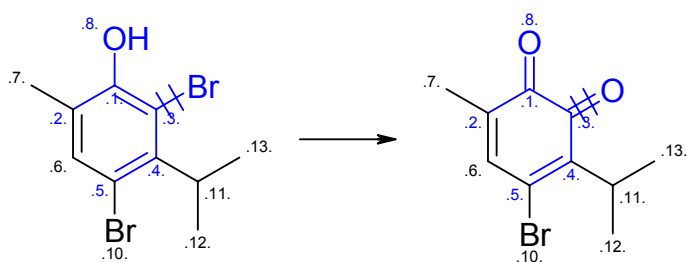
Type ABF (BROAD)



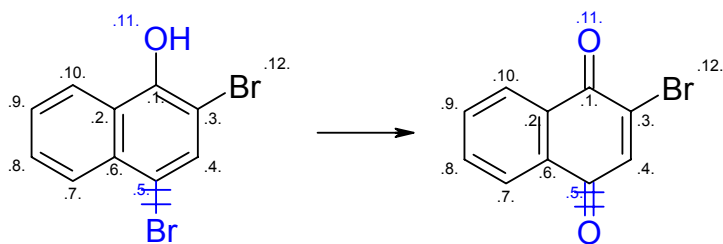
Type ABF (MEDIUM)



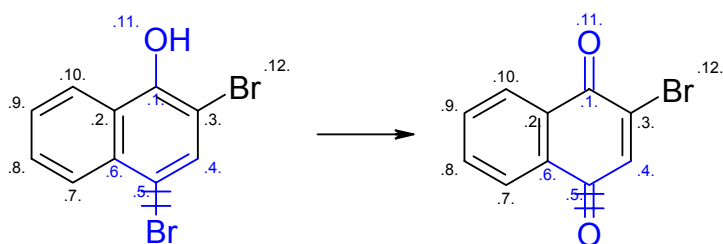
Type ABF (NARROW)



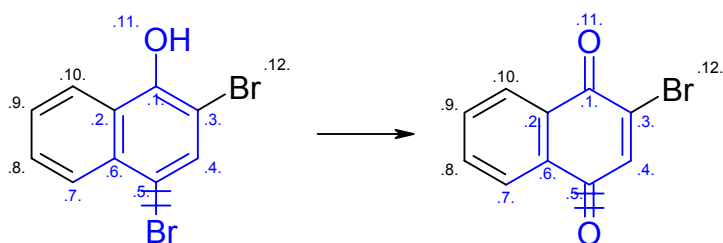
Type ACG (BROAD)



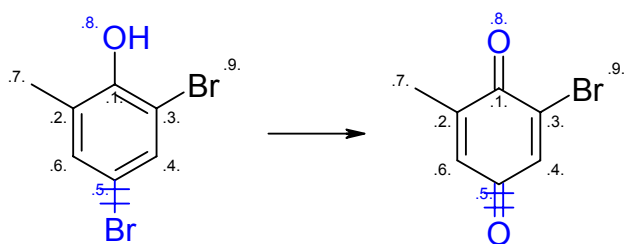
Type ACG (MEDIUM)



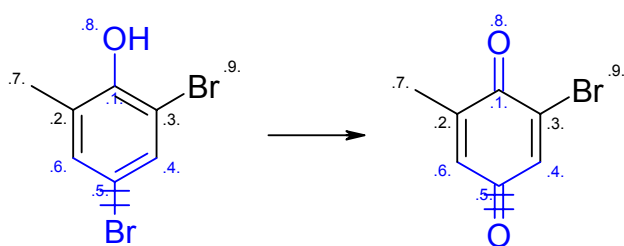
Type ACG (NARROW)



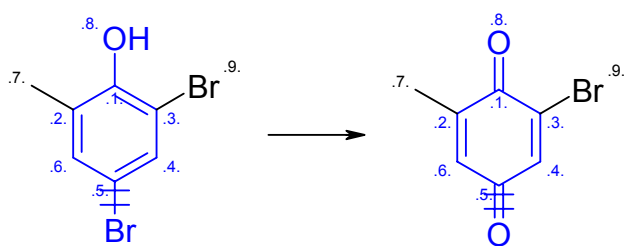
Type ADH (BROAD)



Type ADH (MEDIUM)



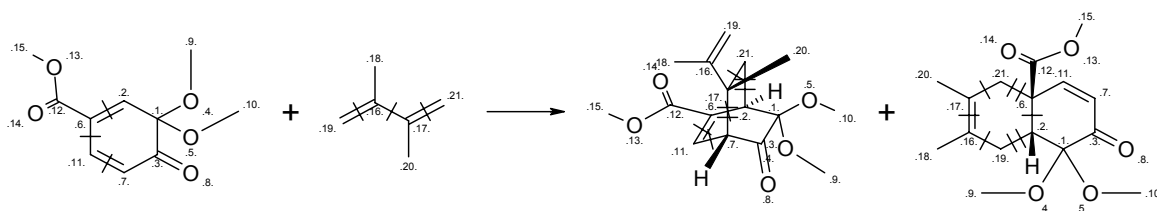
Type ADH (NARROW)



2.3.1.1.3 Example 3

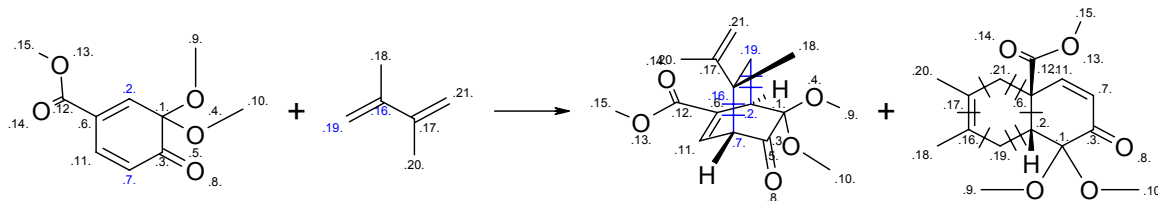
In this example the ClassCodes of the BROAD level are identical for the two products, but the ClassCodes of the MEDIUM and NARRROW level are different.

Original reaction

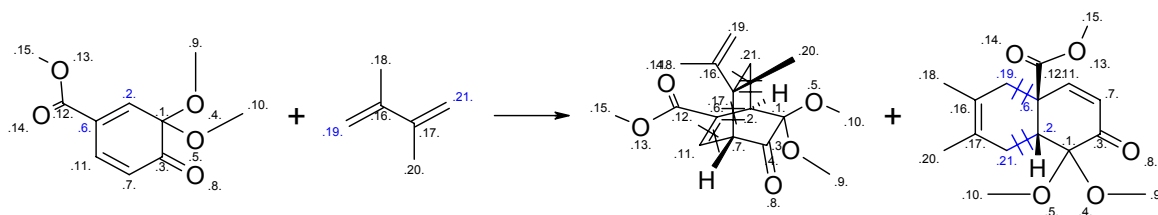


Highlighted ClassCode BROAD substructures

Product 1

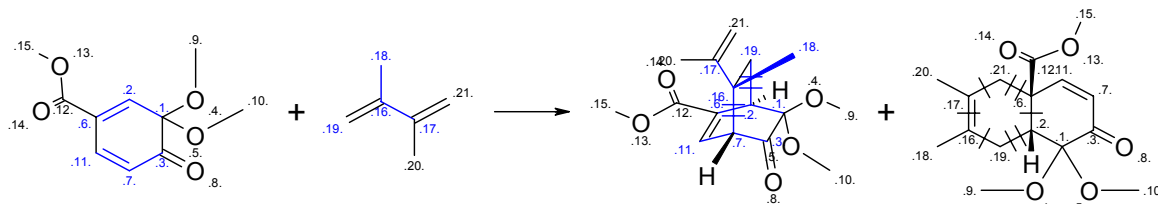


Product 2

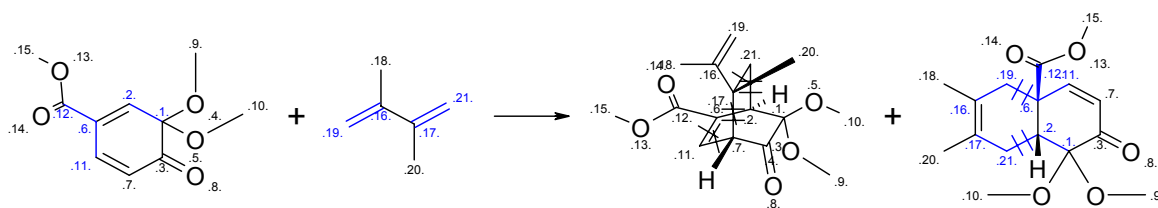


Highlighted ClassCode MEDIUM substructures

Product 1

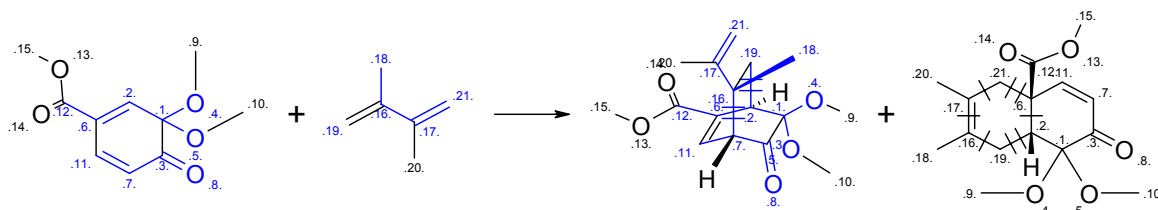


Product 2

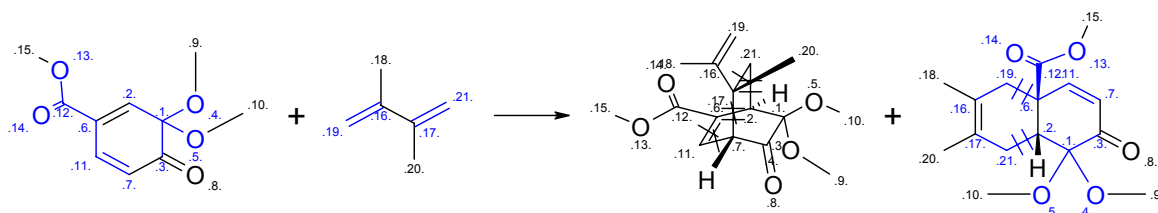


Highlighted ClassCode NARROW substructures

Product 1



Product 2



2.3.2 Extraction of the ClassCode Substructure

By setting the option for extracting the ClassCode substructure to TRUE, the highlighted part of the reaction will be extracted and most atoms not part of the reaction transformation will be removed.

Not all atoms outside the ClassCode substructure will be removed to keep some information of the atoms being part of the reaction transformation:

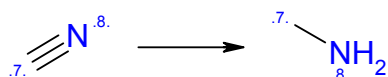
- If an atom of the ClassCode substructure is part of an aromatic ring, this ring will not be removed to indicate the aromatic property of the atom.
- If an atom of the ClassCode substructure is connected to an atom outside the ClassCode substructure by a multiple bond, this atom and the connecting multiple bond will also not be removed to indicate the π -state of the atom.
- Reactants not taking part in the reaction transformation and additional products will be removed completely from the reaction scheme.

Only the highlighted atoms and bonds contribute to the ClassCodes of the reaction transformation.

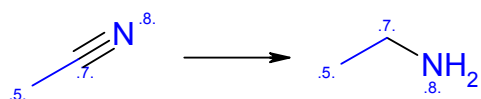
2.3.2.1 Examples

2.3.2.1.1 Example 1 (see 2.1.3.1)

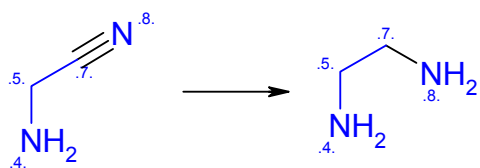
BROAD



MEDIUM

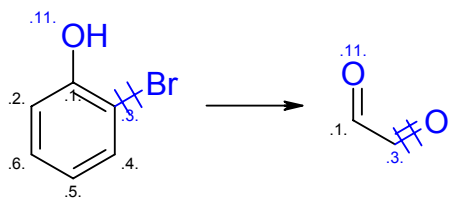


NARROW

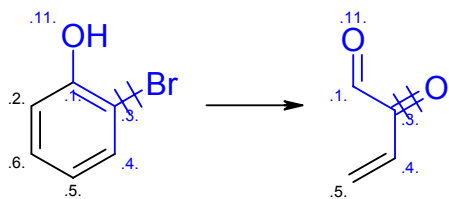


2.3.2.1.2 Example 2 (see 2.1.3.2)

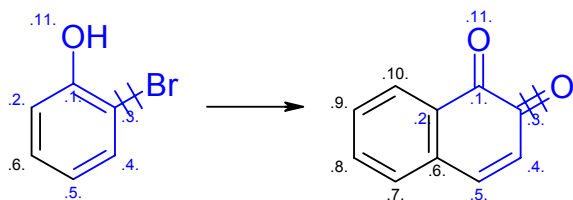
Type ABE (BROAD)



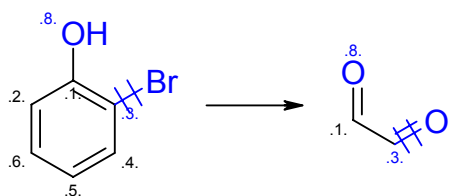
Type ABE (MEDIUM)



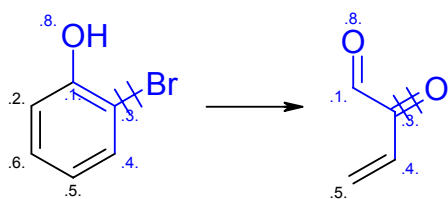
Type ABE (NARROW)



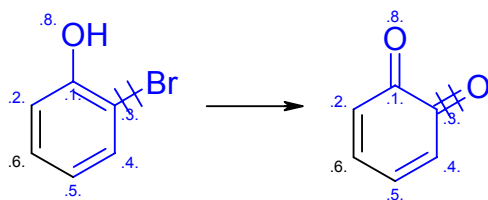
Type ABF (BROAD)



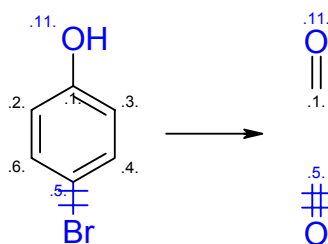
Type ABF (MEDIUM)



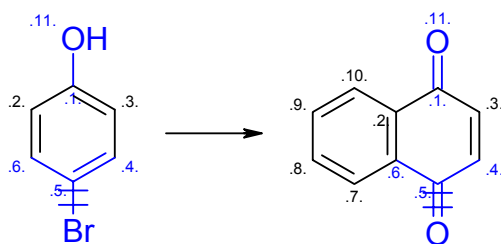
Type ABF (NARROW)



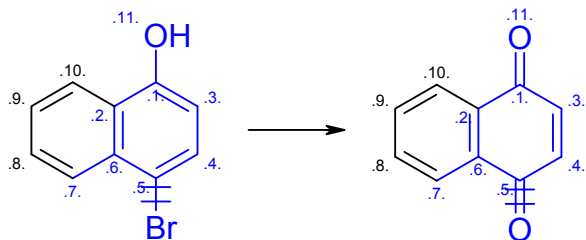
Type ACG (BROAD)



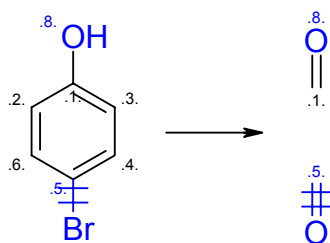
Type ACG (MEDIUM)



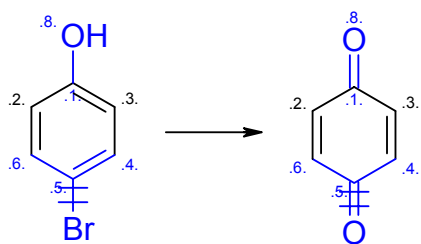
Type ACG (NARROW)



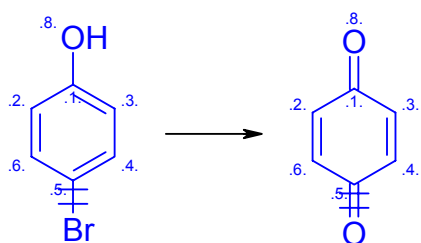
Type ADH (BROAD)



Type ADH (MEDIUM)



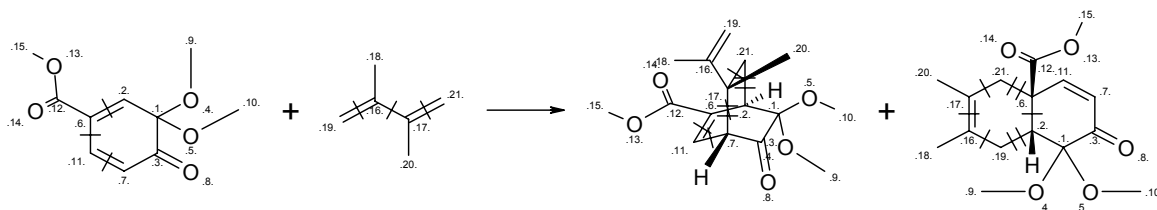
Type ADH (NARROW)



2.3.2.1.3 Example 3

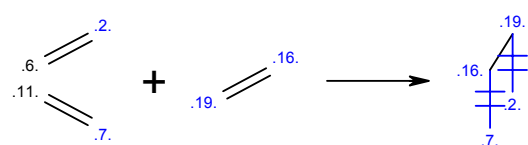
In this example the ClassCodes of the BROAD level are identical for the two products, but the ClassCodes of the MEDIUM and NARROW level are different.

Original reaction

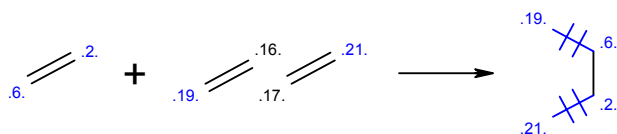


Extracted ClassCode BROAD substructures

Product 1

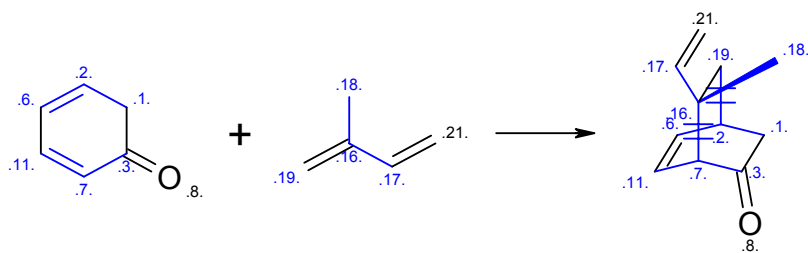


Product 2

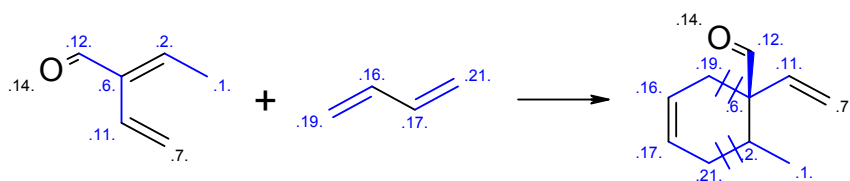


Extracted ClassCode MEDIUM substructures

Product 1

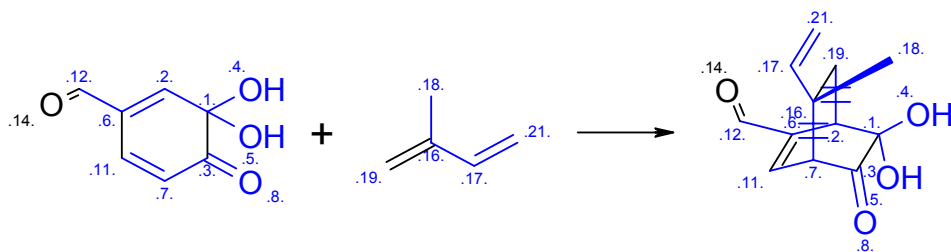


Product 2

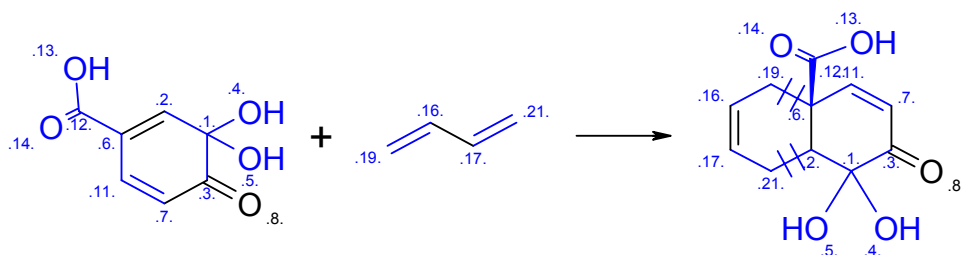


Extracted ClassCode NARROW substructures

Product 1



Product 2



3. Technical Provision of *ICClassify*

ICClassify reads reactions from an MDL RD file (ISIS / Host) and writes reaction numbers (RIREGs) and the determined ClassCodes to an output RD file.

The program *ICClassify* is available for the use as **batch application** on MS-Windows desktops, Sun Solaris or Linux.

In addition, a separate **Windows-DLL** for PC Windows systems is provided which may be integrated into your applications or as an **Add-In** function into ISIS/Draw (MDL Information Systems). *ICClassify* is started with an additional button added to ISIS/Draw allowing the fast, interactive determination of the ClassCodes of individual reactions drawn in ISIS/Draw at the speed of a mouse click.

3.1 Batch Version

3.1.1 System Requirements

Disk Space:	3 MB
Hardware:	PC-Windows / Intel; Sun;
Operating Systems:	Windows 2000 / XP; Solaris; Linux

3.1.2 Contents of the Delivered ZIP File

Executable	Windows: <i>ClassifyV4.02.31.exe</i> Sun/Linux: <i>classifyv4.02.31</i>
ICClassify Configuration file	<i>ICClassifyConfig.xml</i>
ICCheck Configuration file	<i>ICCheckConfig.xml</i>
ICCheck Alias Database file	<i>ICAliasDB.xml</i>
ICCheck Transform Rxn files	<i>TransformAzideGroup.rxn</i> <i>TransformCOMolecule.rxn</i> <i>TransformDiazoGroup.rxn</i> <i>TransformEnolatGroup.rxn</i> <i>TransformFulminateGroup.rxn</i> <i>TransformThioEnolatGroup.rxn</i> <i>TransformThioFulminateGroup.rxn</i>

ICMap Mechanism Rxn files *C_MgBrMap.rxn*
C_MgBrMech.rxn
C_O_EtherMap.rxn
C_O_EtherMech.rxn
N_N01_Map.rxn
N_N01_Mech.rxn
NeberMap.rxn
NeberMapComplete.rxn
NeberMapSave.rxn
NeberMech.rxn
NeberMechComplete.rxn

3.1.3 Installation

3.1.3.1 Windows 2000 / XP

- Unpack the ZIP file *WinClassifyV4.02.32.ZIP* on a PC running Windows to your product directory
(The Executable is *ClassifyV4.02.32.exe*)

3.1.3.2 SUN Solaris

- Unpack the ZIP file *SunClassifyV4.02.32.zip* on a PC running Windows.
- Transfer the binary **executable classifyv4.02.32** via FTP in **binary mode** to the product directory on the SUN Solaris system.
- Transfer all other files in text mode to the product directory on the SUN/Solaris system.
- Set executable privileges for this file
\$ `chmod a+x classifyv4.02.32`

3.1.3.3 Linux

- Unpack the ZIP file *LinuxClassifyV4.02.32.zip* on a PC running Windows.
- Transfer the binary **executable classifyv4.02.32** via FTP in **binary mode** to the product directory on the Linux system.
- Transfer all other files in text mode to the product directory on the Linux system.
- Set executable privileges for this file

```
$ chmod a+x classifyv4.02.32
```

3.1.4 Using the Program

Input: Filename of configuration file (e.g. "ICClassifyConfig.xml")
 Or filename of RDFFile (e.g. "TestRxn.rdf")

Output: The output of the program will be written to the file with the name built from the original filename by replacing the file extension with "CRV40232" (e.g. "TestRxn.CRV40232"). All program errors and messages will be written to a trace file with a name built from the original filename by replacing the file extension with "CRV40232trc" (e.g. "TestRxn.CRV40232trc").

3.1.4.1 Command Line Parameters

Available parameters:

- | | |
|-------------------------|--|
| -i, --inputfile: | Path and name of the input RDFFile (optional) |
| -c, --configfile: | Path and name of the configuration file containing all parameters to control the program (optional) |
| -p, --mode: | mode-parameter (optional: OVERALL(default), DISCRETE, ALL)

OVERALL: only overall ClassCodes will be calculated.

DISCRETE: only discrete ClassCodes will be calculated

ALL: overall and discrete ClassCodes will be calculated |
| -d, --datafield: | Parameter specifying a data field of the input RDFFile to identify the reaction in the output RDFFile (optional). |
| -m, --mapmechanismpath: | Path containing Rxn files for controlling the ICMaP module (optional) |

-k, --checkstructures:	check parameter (optional: ON(default), OFF). Parameter for switching on or off the checking of reactants and products.
-f, --checkconfigpath:	Path containing the configuration files for controlling the ICCheck module (optional).
-s, --highlighting:	switch for activating highlighting of the ClassCode substructure.
-r, --remove_unmarked:	switch for elimination of unmarked Reactants and Products for highlighting of ClassCode substructure
-e, --extract_substructure:	switch for the extraction of the substructure for highlighting of ClassCode substructure
-l, --tracelevel:	Parameter for controlling the trace output (optional: 0(default), >0). 0: only error messages (default) >0: error messages, warnings and informational messages.
-v, --version:	Version parameter (optional). If specified the program version will be prompted without starting the classification process
-h, --help:	Help parameter (optional). If specified the program help will be prompted without starting the classification process

3.1.4.2 Controlling the Program via XML Configuration File

You may control the program by using a XML configuration file. In that case you only have to enter the name of the XML configuration file as command line parameter (-c, or --configfile). Before using the configuration file you have to edit the XML file and specify the appropriate values for the parameters.

3.1.4.2.1 Parameters in Configuration File

InputFile:	Path and name of the input RD file (required)
ClassCodeTypes:	mode-parameter (optional, OVERALL(default), DISCRETE, ALL) OVERALL: only overall ClassCodes will be calculated. DISCRETE: only discrete ClassCodes will be calculated ALL: overall and discrete ClassCodes will be calculated
MechanismPath:	Path containing Rxn files for controlling the ICMAP module (optional)

CheckStructures:	check parameter (optional: ON(default), OFF). Parameter for switching on or off the checking of reactants and products.
CheckConfigPath:	Path containing the configuration files for controlling the ICCheck module (optional).
TraceLevel:	Parameter for controlling the trace output (optional: 0(default), >0). 0: only error messages (default) >0: error messages, warnings and informational messages.
RDFDataField	Parameter specifying a data field of the input RDFFile to identify the reaction in the output RDFFile (optional).
HighlightClassCode	Parameter specifying the highlighting of ClassCode substructure (NO(default), YES) (optional).
RemoveUnmarkedStructures	Parameter specifying the elimination of unmarked Reactants and Products for highlighting of ClassCode substructure (NO(default), YES) (optional).
ExtractClassCodeSubstructure	Parameter specifying the extraction of the substructure for highlighting of ClassCode substructure (NO(default), YES) (optional).

3.1.4.2.1.1 *Parameter RDFDataField*

For identifying the reactions referring to the resulting ClassCodes in the output RDFFile usually the RIREG of the input RDFFile is used. But if you need a unique reaction identifier, you may specify a data field of the input RDFFile identifying the reactions. The content of that data field will be copied to the output RDFFile. This option is more important if you choose the option for highlighting the ClassCode substructure (see below).

3.1.4.2.1.2 *HighlightClassCode*

When the optional parameter *HighlightClassCode* is set to "YES" the substructure used for calculation of a ClassCode will be highlighted and the reaction with highlighted substructure will be written to the output RDFFile in MDLCT V3000 format. Therefore every reaction of the input RDFFile will produce several reactions in the output RDFFile, one reaction for every ClassCode. The reactions in the output RDFFile will not have been

modified, i.e. all reactants and products will be written to the output RDFFile. Structures will not be changed, as long as the parameter *CheckStructures* has not been set to YES.

3.1.4.2.1.3 *RemoveUnmarkedStructures*

When the optional parameter *RemoveUnmarkedStructures* is set to "YES" the substructure used for calculation of a ClassCode will be highlighted and all reactants and products not being reaction partners, i.e. having no mappings assigned, will be removed from the reaction written to the output RDFFile. Structures will not be changed, as long as the parameter *CheckStructures* has not been set to YES.

3.1.4.2.1.4 *ExtractClassCodeSubstructure*

When the optional parameter *ExtractClassCodeSubstructure* is set to "YES" the ClassCode substructure will be extracted and written to the output RDFFile as reaction only. All atoms and bonds not being part of the ClassCode substructure will be removed, with a few exceptions (s. [Extraction of the ClassCode Substructure](#)).

3.1.4.3 **I/O via standard input /output**

In case there has not been specified an input RDFFile or a configuration file, the program reads the input from standard input (stdin) and writes the output to standard output (stdout). All messages of the program will be written to standard error (stderr).

3.1.4.4 **Windows 2000 / XP**

Navigate to your product directory (see installation)

Start executable:

```
>ClassifyV4.02.32 -i "RDFilePathAndName" -p "mode" ...
```

or

```
>ClassifyV4.02.32 -c "XMLConfigurationFilePathAndName"
```

3.1.4.5 SUN Solaris

Navigate to your product directory (see installation)

Start executable:

```
>./ClassifyV4.02.32 -i "RDFilePathAndName" -p "mode" ...
```

or

```
>./ClassifyV4.02.32 -c " XMLConfigurationFilePathAndName "
```

3.1.4.6 Linux

Navigate to your product directory (see installation)

Start executable:

```
>./ClassifyV4.02.32 -i "RDFilePathAndName" -p "mode" ...
```

or

```
>./ClassifyV4.02.32 -c " XMLConfigurationFilePathAndName "
```

3.2 Windows DLL Version

3.2.1 System Requirements

Hardware:	PC-Windows / Intel;
Operating Systems:	Windows 2000 / XP
Software:	none

3.2.2 Installation

- Unpack the ZIP file *ICClassifyV4.02.32.ZIP* on a PC running Windows to your product directory. To integrate the DLL into your application you have to copy the DLL *ICClassify.DLL* and the associated library *ICClassify.lib* to your project directory. For C++-Projects you have to link your application with the library.

3.2.3 Calling the DLL

3.2.3.1 Exported DLL Functions

3.2.3.1.1 Function for Calculation of Overall Hashcodes for BROAD, MEDIUM and NARROW

The exported function for calculating overall hashcodes has been declared in C++ as

```
__declspec(dllexport) int ICTYPEV2 ( char *rxnFileName,  
                                     char *options,  
                                     char *errorMessage,  
                                     char *shell0,  
                                     char *shell1,  
                                     char *shell2 );
```

3.2.3.1.1.1 Description of Input Parameters

char *rxnFileName: Path and filename of the RxnFile containing the Reaction Type Query.
In case of an invalid filename the parameter *errorMessage* will indicate the problem.
Minimum size: string length of path/ filename.

char *options: Control options - for future use
Recommended Size: -

3.2.3.1.1.2 Description of Output Parameters

Important: Please choose the sizes of the output parameter strings in your calling functions for *errorMessage* and *shells* as described below! If the strings are too short the DLL functions might crash !!

char *errorMessage: Return message containing error status

Minimum size: 100

The string errorMessage will be filled, if at least one of the products causes an error. In this case an error code will be added for every product of the reaction.

(Example: "Error(s): 0, 31001001, 0")

Currently available error codes:

Error Code:	Associated Message:
0:	No error occurred.
1:	Too many breaking bonds (no longer supported).
2:	No reaction centers found (no longer supported).
3:	Product mapping less than 55% (no longer supported).
31nnnnnn:	Error in mapping module (values >= 31000000).
41nnnnnn:	Error in RxnFile read module (values >= 41000000).

In case of a file read error the following string is returned:

"Error 99 openCTfile: Cannot open file *rxnFileName* for input"

char *shell0: 15 digit hashcode of sphere 0 (BROAD)
If the reaction has more than one product, the sphere 0 hashcodes of the products are concatenated and separated by a BLANK (' ').
In case of any error a zero value hashcode is concatenated consisting of 15 0's:

Example:

0000000000000000

Recommended minimum size: 256 in case of a maximum of 15 products
(16 * MAXProduct)

char *shell1: 15 digit hashcode of sphere 1 (MEDIUM)
For more than one product the same as for shell 0.
Recommended minimum size: 256 in case of a maximum of 15 products
(16 * MAXProduct)

char *shell2: 15 digit hashcode of sphere 2 (NARROW)
For more than one product the same as for shell 0.

Recommended minimum size: 256 in case of a maximum of 15 products
(16 * MAXProduct)

3.2.3.1.2 *Function for Version Display*

The exported function for displaying the version of the DLL has been declared in C++ as

```
__declspec(dllexport) int GetICClassifyVersion (char *version);
```

3.2.3.1.2.1 **Description of Output Parameter**

The version of the DLL *ICClassify* will be returned in the character array *version*.

char *version: Character array containing the version string
Recommended minimum size: 256

The returned character string looks like below

```
ICClassifyDll 4.01.30 19-Oct-2007 WIN_RELEASE OS32Bit ( Rebuild date:Oct 19 2007. time:14:54:24. )
```

3.3 ISIS/Draw Addin

3.3.1 System Requirements

Hardware:	PC-Windows / Intel;
Operating Systems:	Windows 2000 / XP
Software:	ISIS / Draw from MDL

3.3.2 Installation

- Unpack the ZIP file *ICClassifyAddin.ZIP* on a PC running Windows to your product directory.
- To integrate the Addin into ISIS/Draw you have to copy the DLLs *icclassify.dll*, *icmapcpp.dll*, *IVAddXil.dll* and the XML file *icmapcpp.xml* to the installation directory of your ISIS installation (e.g. "F:\programme\MDL ISIS Desktop 2.5")
- Copy the EPL files *ICTYPE.EPL* and *IVAddDrw.epl* to the ISIS/Draw start directory (e.g. "F:\programme\MDL ISIS Desktop 2.5\Drwstart")

3.3.3 Configuration of the Addin

The ISIS Addin may be configured by changing the content of the XML file *icmapcpp.xml*.

3.3.3.1 Description of Parameters in XML file

```
<!-- ICMAP Configuration File - Comments in one Line only please-->
<XML>
  <ICMapOptions>
    <VersionOperationMode>
      <!-- ICMAP_VERSION_V40_MDL_CLASSCODE_2000 -->
      <!-- ICMAP_VERSION_V41 -->
      <!-- ICMAP_VERSION_V50 -->
      <!-- ICMAP_VERSION_CURRENT -->
      ICMAP_VERSION_CURRENT
    </VersionOperationMode>
    <CenterType>
      <!-- INFOCHEM_CENTER, MDL_CENTER, JSM_CENTER, ICFS_CENTER_TYPE -->
      MDL_CENTER
```

```
</CenterType>
<ProductMappingMode>
  <!-- MAP_ONE_PRODUCT, MAP_ALL_PRODUCTS -->
  MAP_ALL_PRODUCTS
</ProductMappingMode>
<ReactantMappingMode>
  <!-- ICMAP_VERSION_V50 or later -->
  <!-- INPUT_REACTANT_MODE, COPY_REACTANT_MODE -->
  INPUT_REACTANT_MODE
</ReactantMappingMode>
<MechanismPath>
  <!-- ICMAP_VERSION_V50 or later -->
  F:\Programme\MDL ISIS Desktop 2.5\MapMechanisms\
</MechanismPath>
</ICMapOptions>
<ICTypeOptions>
  <TypeMapVersion>
    <!-- ICMAP_VERSION_V40_MDL_CLASSCODE_2000 (Default) -->
    <!-- ICMAP_VERSION_V41 -->
    <!-- ICMAP_VERSION_V50 -->
    <!-- ICMAP_VERSION_CURRENT -->
    ICMAP_VERSION_CURRENT
  </TypeMapVersion>
  <TypeCenter>
    <!-- INFOCHEM_CENTER (Default) -->
    <!-- MDL_CENTER -->
    <!-- JSM_CENTER -->
    <!-- ICFS_CENTER_TYPE -->
    INFOCHEM_CENTER
  </TypeCenter>
  <TypeProductMappingMode>
    <!-- MAP_ONE_PRODUCT -->
    <!-- MAP_ALL_PRODUCTS (Default) -->
    MAP_ALL_PRODUCTS
  </TypeProductMappingMode>
  <TypeReactantMappingMode>
    <!-- ICMAP_VERSION_V50 or later -->
    <!-- INPUT_REACTANT_MODE -->
    <!-- COPY_REACTANT_MODE (Default) -->
    INPUT_REACTANT_MODE
  </TypeReactantMappingMode>
  <TypeMechanismPath>
    <!-- ICMAP_VERSION_V50 or later -->
    F:\Programme\MDL ISIS Desktop 2.5\MapMechanisms\
  </TypeMechanismPath>
  <TypeMode>
```

```
        <!-- OVERALL (Default) -->
        <!-- CLUSTER -->
        <!-- OVERALL_AND_CLUSTER -->
        OVERALL_AND_CLUSTER
    </TypeMode>
    <MaxSpheres>
        <!-- 0-9 (Default = 2) -->
        2
    </MaxSpheres>
    <AlkylProgress>
        <!-- 0-9 (Default = 1) -->
        1
    </AlkylProgress>
    <NormHash>
        <!-- YES/NO (Default = YES) -->
        YES
    </NormHash>
    <StandardizeElements>
        <!-- YES/NO (Default = YES) -->
        YES
    </StandardizeElements>
    <IgnoreDuplicateReactants>
        <!-- YES/NO (Default = YES) -->
        NO
    </IgnoreDuplicateReactants>
</ICTypeOptions>
</XML>
```

3.3.3.1.1 Map Options

All map options are enclosed by the tag <ICMapOptions>.

3.3.3.1.1.1 VersionOperationMode

The tag <VersionOperationMode> defines the type of mapping used.

Values

- ICMAP_VERSION_V40_MDL_CLASSCODE_2000 (Default)
- ICMAP_VERSION_V41
- ICMAP_VERSION_V50
- ICMAP_VERSION_CURRENT

3.3.3.1.1.2 *CenterType*

The tag <CenterType> defines the type of assigned reactions centers

Values

- INFOCHEM_CENTER (Default): atom and bond centers, no bond changes, no ambiguous mappings
- MDL_CENTER: atom and bond centers, bond changes, ambiguous mappings
- JSM_CENTER: atom and bond centers, bond changes, ambiguous mappings (special version)
- ICFS_CENTER_TYPE: atom and bond centers, bond changes, ambiguous mappings (optimized for RSS)

3.3.3.1.1.3 *ProductMappingMode*

The tag <ProductMappingMode> defines how reactions are treated when there is more than one product.

Values

- MAP_ONE_PRODUCT
- MAP_ALL_PRODUCTS (Default)

3.3.3.1.1.4 *ReactantMappingMode*

The tag <ReactantMappingMode> defines how the reactants of the reaction are treated

Values

- INPUT_REACTANT_MODE (Default): No modification of the reactants
- COPY_REACTANT_MODE: Reactants taking part more than once in the reaction will be multiplied.

3.3.3.1.1.5 *MechanismPath*

The tag <MechanismPath> defines the path to the reactions defining some reaction mechanisms used by the mapping program to improve the results.

3.3.3.1.2 *Classify Options*

All classification options are enclosed by the tag <ICTypeOptions>.

3.3.3.1.2.1 *VersionOperationMode*

The tag <VersionOperationMode> defines the type of mapping used.

Values

- ICMAP_VERSION_V40_MDL_CLASSCODE_2000 (Default)
- ICMAP_VERSION_V41
- ICMAP_VERSION_V50
- ICMAP_VERSION_CURRENT

3.3.3.1.2.2 *CenterType*

The tag <CenterType> defines the type of assigned reactions centers

Values

- INFOCHEM_CENTER (Default): atom and bond centers, no bond changes, no ambiguous mappings
- MDL_CENTER: atom and bond centers, bond changes, ambiguous mappings
- JSM_CENTER: atom and bond centers, bond changes, ambiguous mappings (special version)
- ICFS_CENTER_TYPE: atom and bond centers, bond changes, ambiguous mappings (optimized for RSS)

3.3.3.1.2.3 *ProductMappingMode*

The tag <ProductMappingMode> defines how reactions are treated when there is more than one product.

Values

- MAP_ONE_PRODUCT
- MAP_ALL_PRODUCTS (Default)

3.3.3.1.2.4 *ReactantMappingMode*

The tag <ReactantMappingMode> defines how the reactants of the reaction are treated.

Values

- INPUT_REACTANT_MODE (Default): No modification of the reactants
- COPY_REACTANT_MODE: Reactants taking part more than once in the reaction will be multiplied.

3.3.3.1.2.5 *MechanismPath*

The tag <MechanismPath> defines the path to the reactions defining some reaction mechanisms used by the mapping program to improve the results.

3.3.3.1.2.6 *TypeMode*

The tag <TypeMode> defines which ClassCodes will be created

Values:

- OVERALL (Default): Only Overall ClassCodes will be generated
- CLUSTER: Only Discrete ClassCodes will be generated
- OVERALL_AND_CLUSTER: Overall and Discrete ClassCodes will be generated

3.3.3.1.2.7 *MaxSpheres*

The tag <MaxSpheres> defines the maximum number of spheres limiting the expansion of the reaction centers

Values: 0 - 9 (Default = 2)

3.3.3.1.2.8 *AlkylProgress*

The tag `<AlkylProgress>` defines the maximum number of spheres limiting the expansion of the reaction centers over sp^3 carbon atoms.

Values: 0 – MaxSpheres (Default = 1)

3.3.3.1.2.9 *NormHash*

The tag `<NormHash>` defines, whether Multiple occurrences of "atom clusters" are treated as only one occurrence or not.

Values: YES/NO (Default = YES).

3.3.3.1.2.10 *StandardizeElements*

The tag `<StandardizeElements>` defines, whether the elements of the PSE (Periodic System of the Elements) will be standardized (refer to 2.1.2) or not.

Values: YES/NO (Default = YES).

3.3.3.1.2.11 *IgnoreDuplicateReactants*

The tag `<IgnoreDuplicateReactants>` defines whether duplicate reactants from input will be ignored or not.

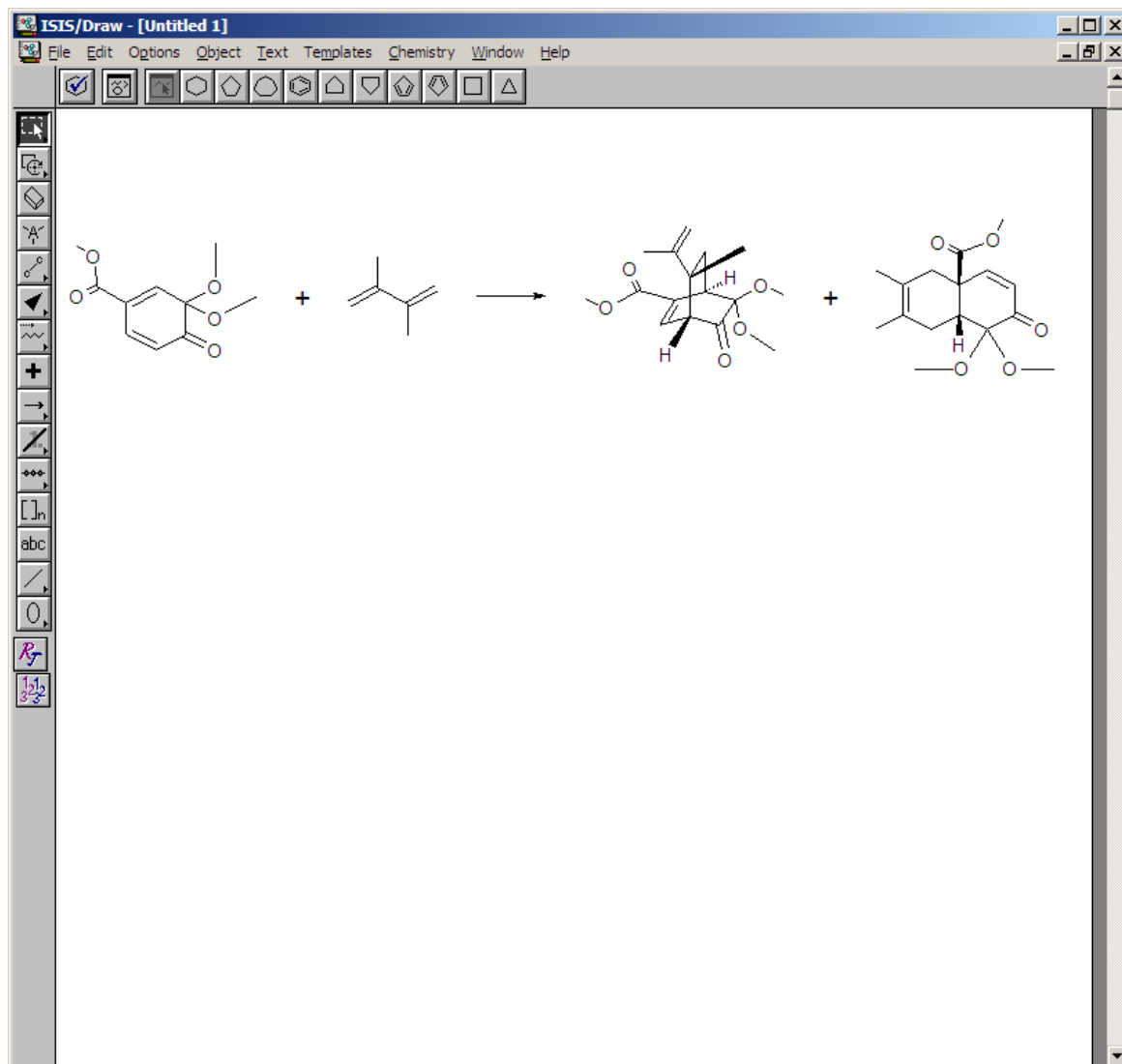
Values: YES/NO (Default = YES).

3.3.4 Using the Addin

After installation of the Addin and starting ISIS/Draw you will find new buttons on bottom of the left toolbar of ISIS/Draw.

First draw a fully defined reaction (do not use query features!).

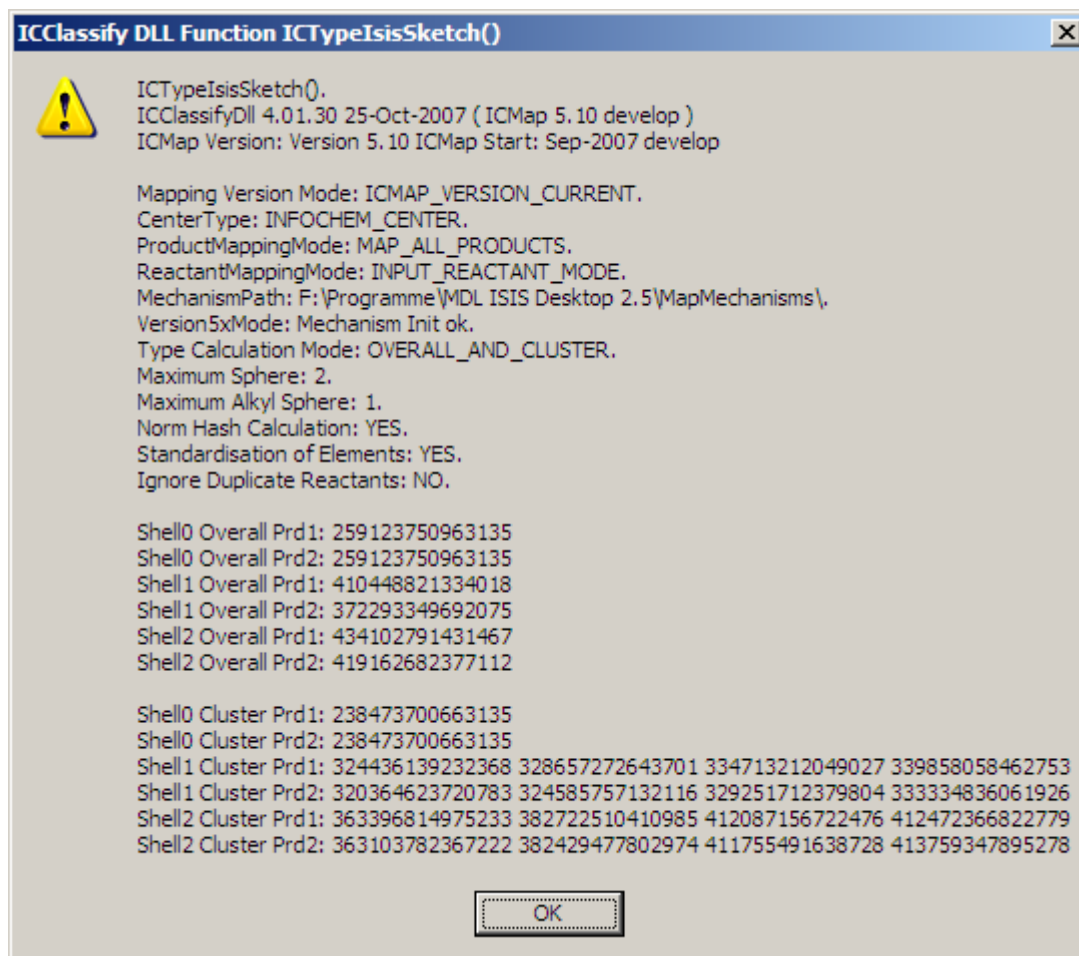
Example from 2.3.1.1.3



Then press the button for calculating the ClassCodes.



The resulting ClassCodes will be displayed in a message box.



4. Commercial Applications of *ICClassify*

- *ICClassify* is used by MDL (now Symyx) in the RXL-Browser to cluster the results in the hit lists of reaction queries.
- All reaction databases distributed by MDL (now Symyx) are classified and include the InfoChem ClassCodes.
- All reactions included in the Beilstein database will be classified to contain the InfoChem ClassCodes.
- The InfoChem ClassCodes are the only way to link different reaction databases. The product "*integrated* Major Reference Works" distributed by MDL (now Symyx) is based on this concept and architecture.
- *ICClassify* is integrated as an optional package (ICART_RXNTYPE) in the InfoChem Chemistry Cartridge.
- Several chemical and pharmaceutical companies are using *ICClassify* to process their internal proprietary reaction databases (e.g. BASF, Bayer, Merck, Boehringer, Novartis, a.o.m) to determine the InfoChem ClassCodes.