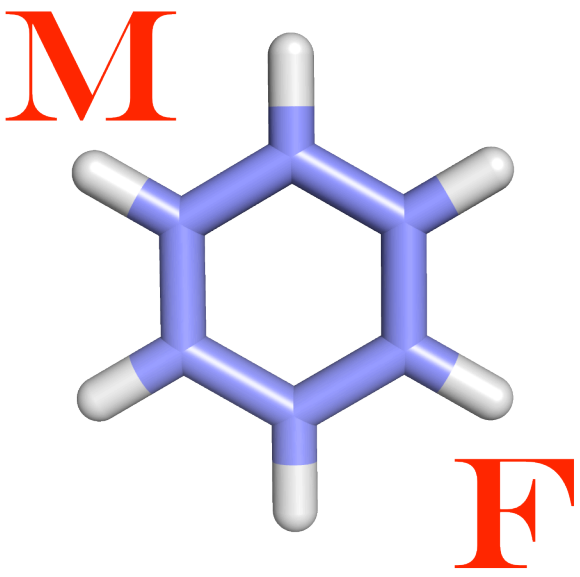
MolFind User Manual

MolFind 1.9

06-18-2014



Lochana C. Menikarachchi Ph.D.
Department of Pharmaceutical Sciences
School of Pharmacy
University of Connecticut

1. Introduction

MolFind is a java based software package for identifying unknown chemical structures in complex mixtures using HPLC/MS data. Identifying an unknown involves matching orthogonal experimental features measured for the unknown (RI, ECOM50, drift time and CID spectra) with computationally predicted values for candidate compounds contained in chemical or biochemical databases. The program features an easy to use graphical user interface and a highly multi threaded pipeline for identifying unknowns.

2. System Requirements

Operating System: MolFind should work on any operating system (Windows XP, Windows Vista, Windows 7, 8, 8.1, Mac OSX, Any version of Linux or Solaris) provided **java 1.6 or higher** is installed. Java standard edition (SE) run time (jre 1.6 / jre 1.7) can be freely downloaded from
http://www.oracle.com/technetwork/java/javase/downloads/index.html

Memory: We recommend 1-2 GB of RAM for java virtual machine, however, more RAM may be required depending on the type of calculation. Bundled program execution scripts (**MolFind.bat** for windows and **MolFind.sh** for linux/osx/solaris) will allocate 2 GB of RAM for the java virtual machine. The amount of RAM allocated for java virtual machine can be changed by editing MolFind.bat or MolFind.sh file. (See FAQ for more details)

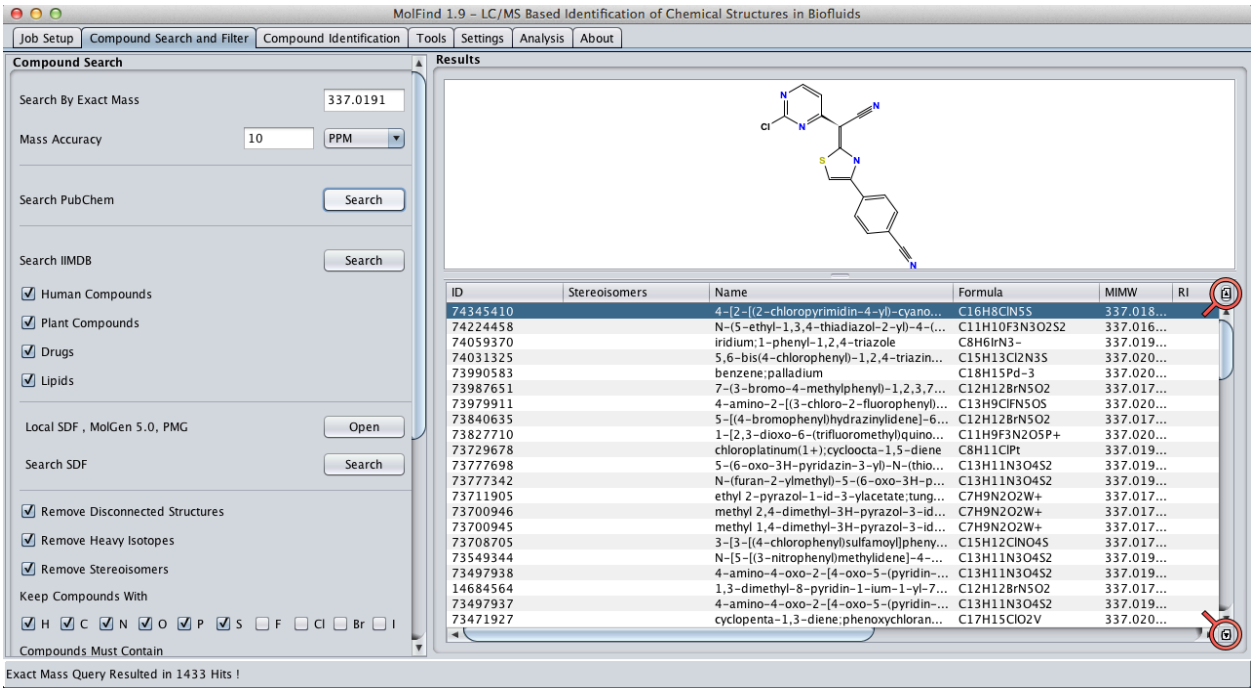
Processor: There is no processor requirement. However, faster multicore processors will certainly help. MolFind can take advantage of modern multicore architectures.

3. Installation Instructions

No installation required. Simply extract the downloaded zip file into a folder. If you are on Windows (Windows XP, Windows Vista or Windows 7, 8, 8.1), double click on the MolFind.bat file to start MolFind. If you are on Mac OSX, Linux or Solaris, double click on the MolFind.sh file to start MolFind. Double clicking MolFind.jar should also work.

4. User Interface

MolFind comes with a tabbed graphical user interface (TGUI).

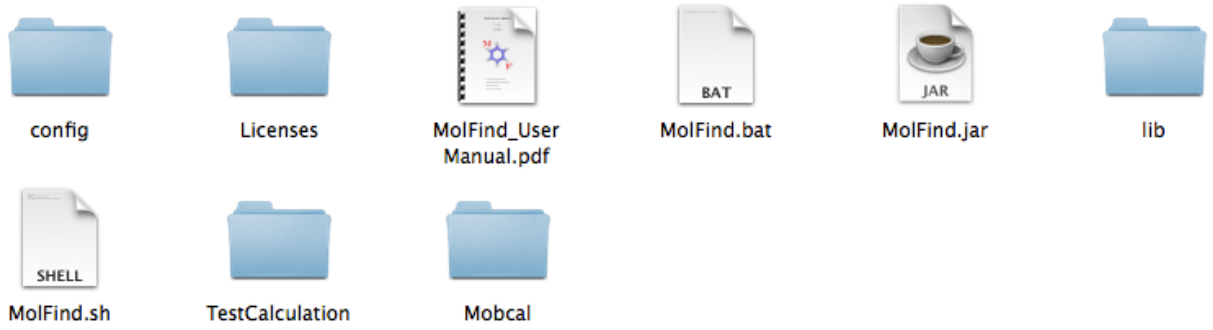


Tab	Function
Job Setup	Setup job name & working directory
Compound Search and Filter	Download and filter candidates
Compound Identification	Rank/identify candidates using MetFrag algorithm
Tools	Various tools to generate Gaussian, Mobcal input files etc.
Settings	Various settings such as QSPR model windows, passwords etc.
Analysis	Analyze/visualize MolFind output file (_out.txt)
About	Licensing terms and references

By default, MolFind loads first 100 candidates into the table on the “Compound Search” panel. If

your filters resulted in more than 100 structures, the rest will not show up in the first page. Please use the page down/up buttons (buttons located under the magnifying glass icons in the figure above) to load next/previous 100 candidates. The table on the “Compound Identification” panel behaves similarly. Both tables (one on the search panel and the one on the identification panel) do not allow sorting. The table on the “Analysis” panel allows sorting.

5. Contents of the MolFind Folder



config: configuration files

Licenses: licenses for third party libraries

MolFind_UserManual.pdf: this document

MolFind.bat: program startup script for windows

MolFind.jar: MolFind (platform independent executable)

lib: third party libraries

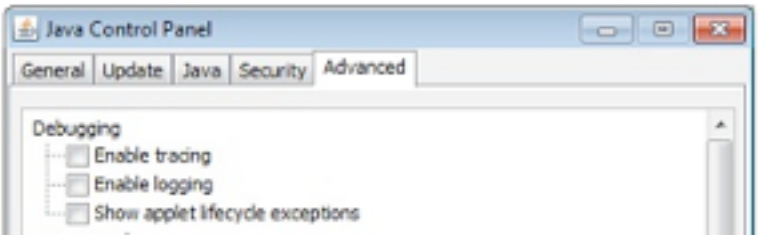
MolFind.sh: program startup script for osx, linux or solaris

TestCalculation: data for test calculations

Mobcal: Mobcal source files and compilation instructions

MolFind.log: This file is created at the startup. MolFind.log contains logging information for the current MolFind session.

Check the log file for any errors. If logging is disabled, the MolFind.log file will not be created.



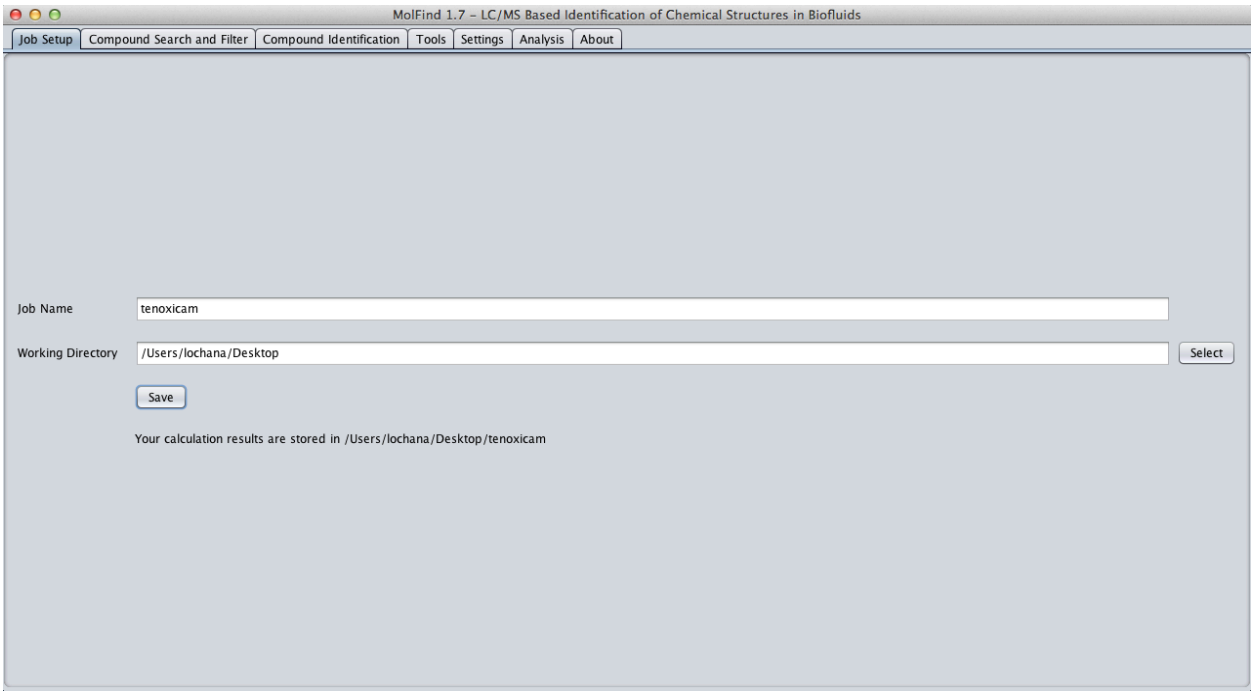
Follow the instructions on http://www.java.com/en/download/help/javaconsole.xml to locate java control panel. Click on the “Advanced” tab and check the “Enable logging” option.

6. Running Calculations

1. Go to “Job Setup” tab

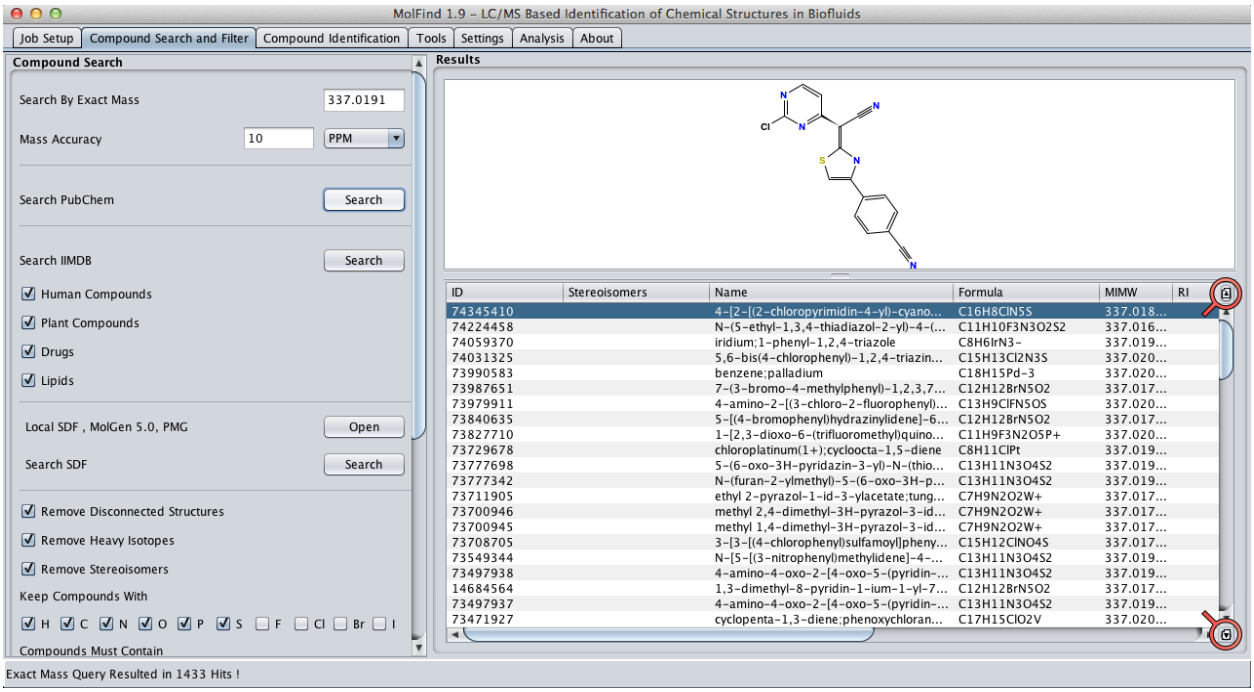
2. Type in a name for the job. All data for the job will be saved in a sub directory

3. Select a working directory for MolFind calculations. All jobs will be stored inside the working directory.



4. Click Save (“**Current Job...**” label will change to your job folder)

5. Go to Compound Search and Filter tab



6. Type in the neutral exact mass (or accept the default value for test calculation provided)

Note that MolFind currently allows users to search for a single unknown. Future versions will allow batch processing of multiple unknowns.

7. Select the mass accuracy in ppm or ppb.

MolFind is designed for high mass accuracy data (± 1-20 ppm), and thus there is a limit to the number of candidate compounds that can be downloaded for any given mass.

8. Click on the appropriate search button to download compounds from PubChem, IIMDB or a local SD file.

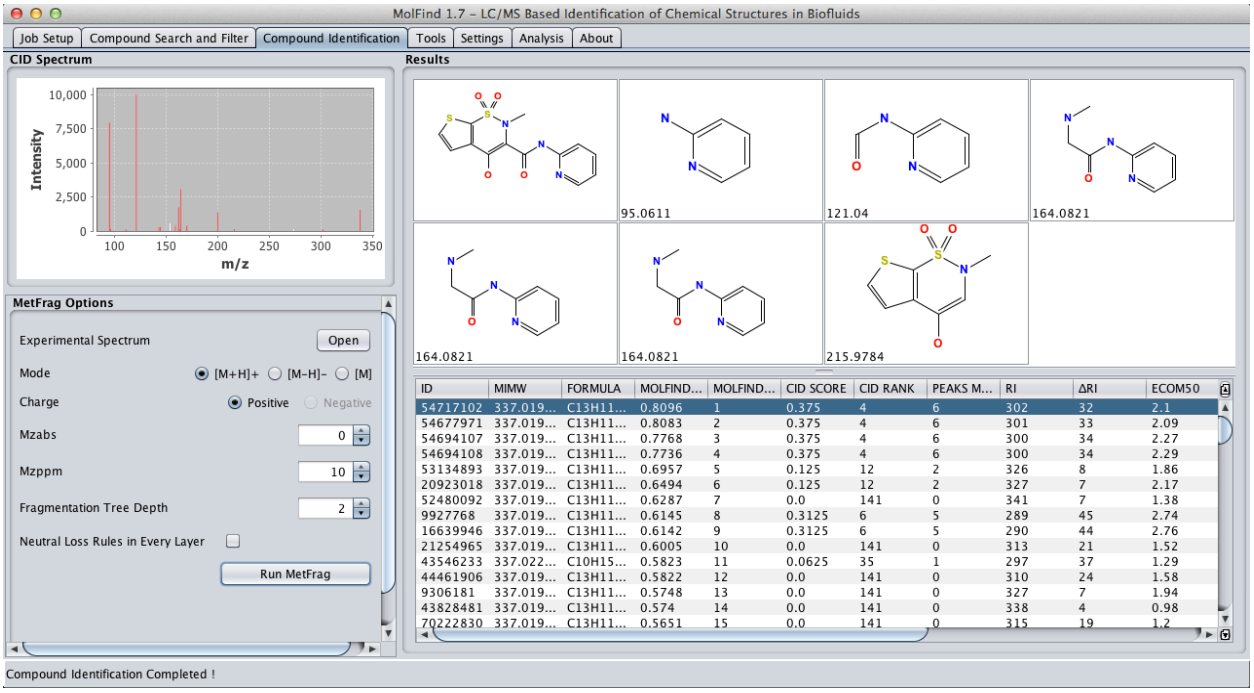
9. The **Status bar** at the bottom of the screen will show the progress of the calculation and whether any errors occurred during the calculation. Real time status updates will be delivered through the status bar.

Please note that no filters are used until after the compounds have been downloaded. The selected filters will be applied as soon as you hit the "Apply Selected Filters" button.

10. Apply filters (Elements, Disconnected Structures, RI, ECOM50, Drift Time) once the download is complete.

11. Click on "**Proceed to Identification**" button once the filtering process is over.

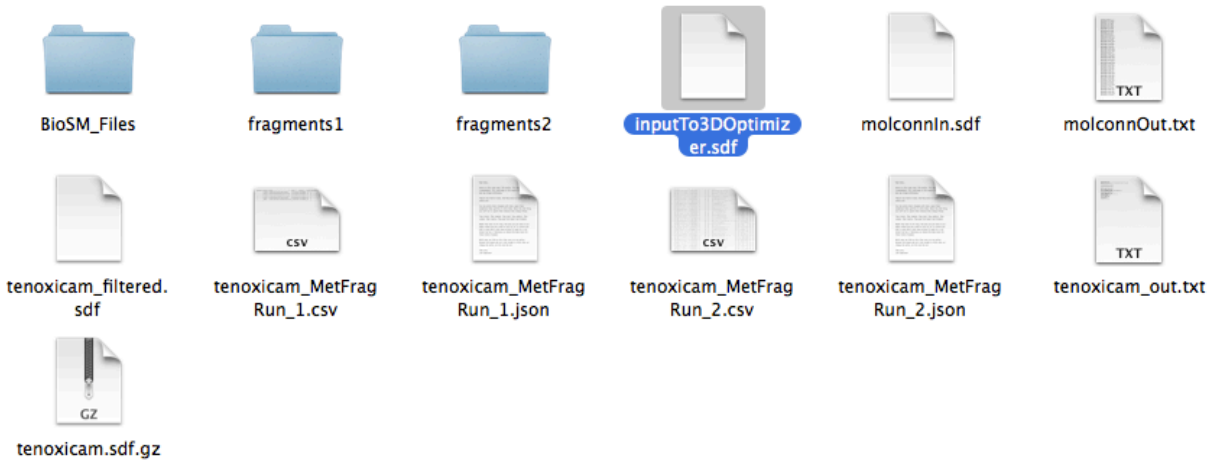
12. Switch to "Compound Identification" panel.



13. Select the experimental spectrum located under "**TestCalculations**" folder.

14. Click on "Run MetFrag" button to rank candidates

7. MolFind Ouput Files



BioSM_Files: Input/Output files for BioSM program (predicts whether a compound is biological or not)

fragments1/fragments2: MetFrag predicted fragments for two MetFrag runs

inputTo3DOptimizer.sdf: 2D structure input file for geometry optimizations

molconnIn.sdf: Input file for Molconn

molconnOut.txt: Calculated RI and $ECOM_{50}$ values

tenoxicam_filtered.sdf: filtered candidates for tenoxicam bin

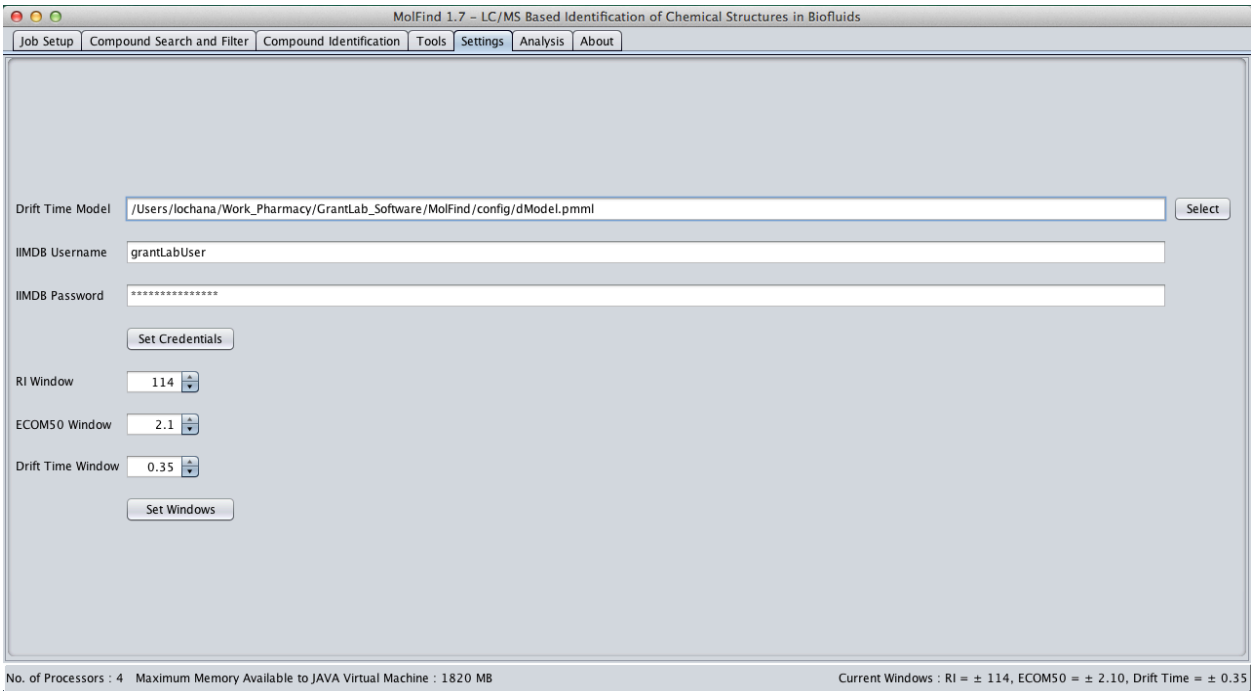
tenoxicam_MetFragRun_1.csv/tenoxicam_MetFragRun_2.csv: Results for two MetFrag runs in csv format (does not include MetFrag fragments)

tenoxicam_MetFragRun_1.json/tenoxicam_MetFragRun_2.json: Results for two MetFrag runs in json format (Include MetFrag fragments)

tenoxicam_out.txt: Output file for the MolFind job (can be visualized with Analysis Panel)

tenoxicam.sdf.gz: candidates downloaded from PubChem

8. Settings Panel



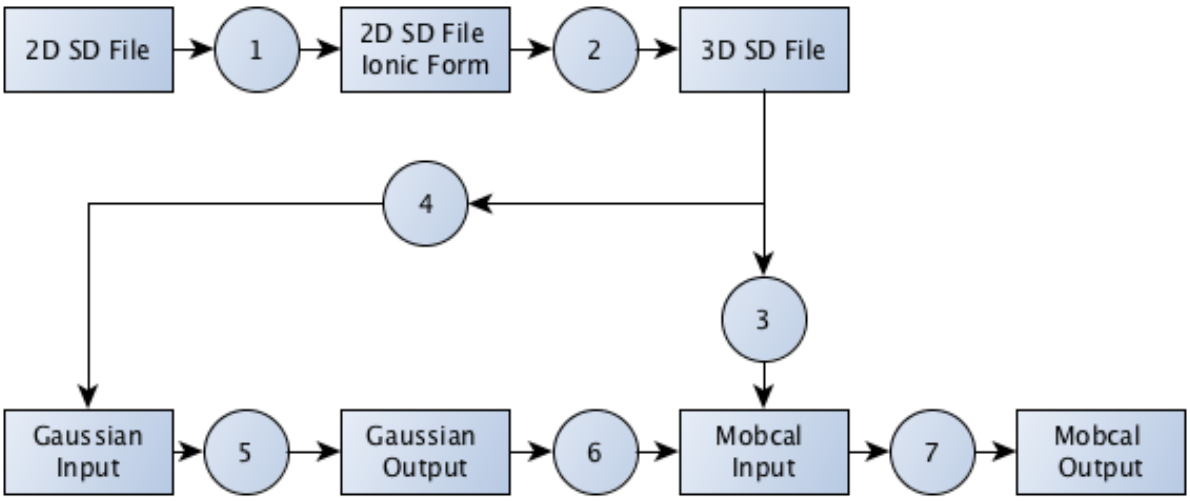
Drift Time Model: PMML (Predictive Modeling Markup Language) file containing drift time model - this file is usually located under config folder.

IIMDB Username and Password: Current Lhasa members can request a username and password to access IIMDB. Please contact Mr. Scott McDonald at Lhasa Ltd. for a user account. Scott.McDonald@lhasalimited.org

RI, ECOM50 and Drift Time Windows: Model Windows for filtering candidate compounds

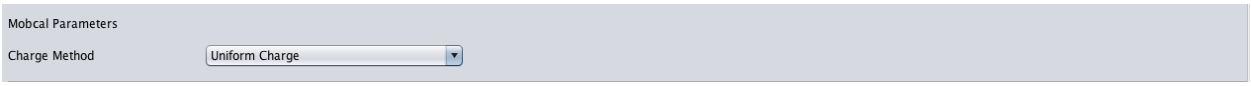
9. Running Mobcal Calculations

MolFind's "Tools" panel provides several utilities to prepare Mobcal/Gaussian input files and run Mobcal calculations in parallel. Mobcal input files can be prepared from 2D SD files or Gaussian (G03 or G09) output files. Please refer to the numbered steps in the following figure.

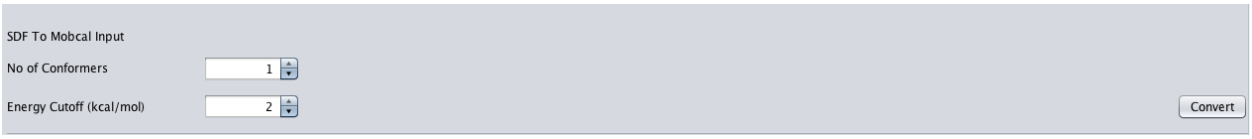


Step-1: Convert the SD file to [M+H]+, [M-H]- or [M+Na]+ form

Step-2,3: Generate Mobcal input file(s) using molecular mechanics optimized structure(s)



Select Mobcal charge method.

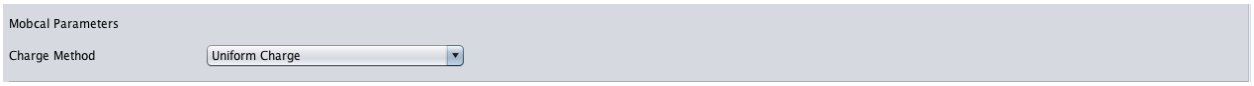


Select the number of conformers and energy cutoff. If number of conformers (this is the maximum possible number of conformers) is set to more than 1, multiple conformers will be used. An energy cutoff of 0 will use however many structures that conformer generation comes up with. An energy cutoff of "N" will remove any conformer that is "N" kcal/mol higher than the lowest energy conformer. When you click on the "Convert" button, a new directory named

"mobcalData" will be created in the directory of the SD file. Mobcal input files will be stored in the mobcalData directory.

Steps-2,4: Generate Gaussian input files from 2D SD file(s)

Step 6: Gaussian output to Mobcal input



Select Mobcal charge method.

Select a directory to store Gaussian input files. Click on the "Convert" button to generate Gaussian input files.

Step 7: Run Mobcal

Compile (Following the compilation instructions in the "ReadMe.txt" file) the Mobcal source code located in the Mobcal directory. Select the Mobcal executable and Mobcal input files directory. Click on the "Run" button to run Mobcal in parallel. Mobcal output files will be found in the mobcalData directory.

Please Note: Both ionic form generation and MM charge generation algorithms require separate ChemAxon licenses. Make sure to have valid ChemAxon licenses (for pKa and charge modules) installed.

Citing MolFind and Mobcal:

Please cite Mobcal, N2-Optimized Mobcal, and MolFind by adding a statement similar to this:

"All structure manipulations (generation of ionized forms and molecular mechanics based conformers) and input file preparations (for gaussian09 and Mobcal) were done using MolFind's tools panel[1]. A modified version of Mobcal[2-4] optimized for room temperature N_2-based trajectory method (TM) was used for calculating average collision cross-sectional areas."

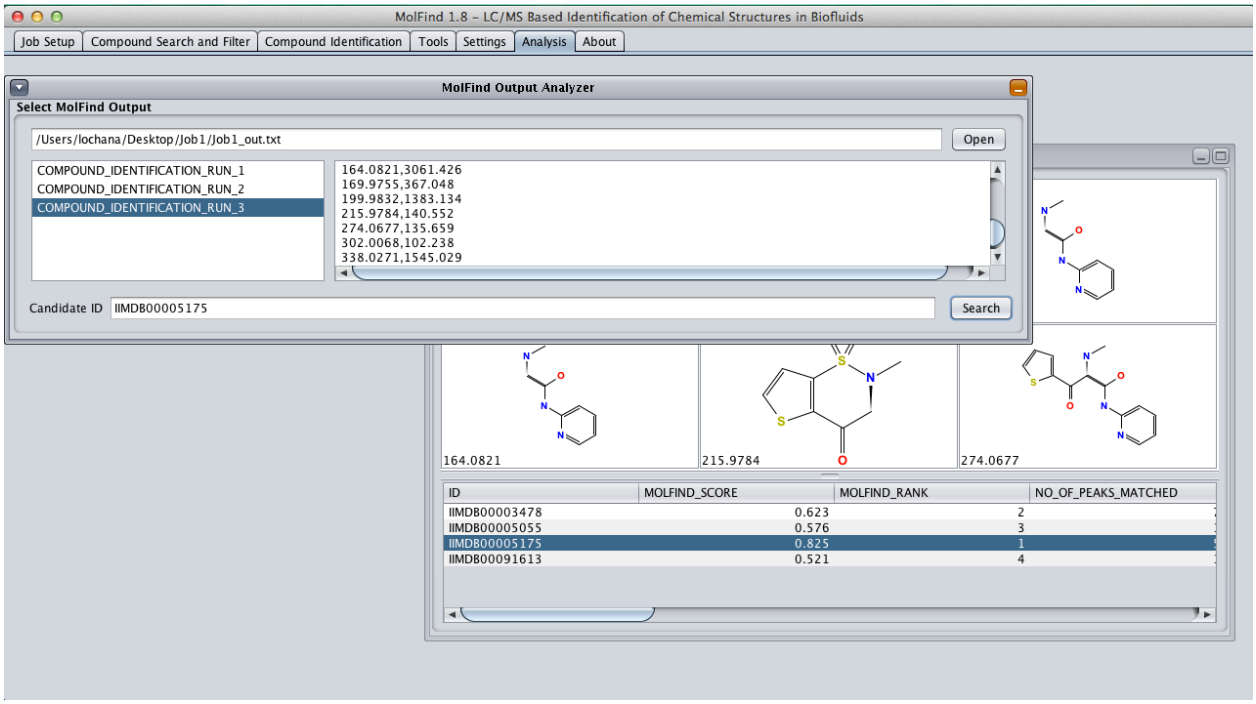
(1) Menikarachchi, L. C.; Cawley, S.; Hill, D. W.; Hall, L. M.; Hall, L.; Lai, S.; Wilder, J.; Grant, D. F. *Anal. Chem.* **2012**, *84*, 9388–93394.

(2) Campuzano, I.; Bush, M. F.; Robinson, C. V; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. *Anal. Chem.* **2012**, *84*, 1026–1033.

(3) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082–16086.

(4) Shvartsburg, A. *Chem. Phys. Lett.* **1996**, *261*, 86–91.

10. Analysis Panel



Analysis panel allows you to analyze results from a previously ran MolFind job. Simply load the MolFind output ("_out.txt") file using "Open" button. If your MolFind job has multiple MetFrag runs, they will show up as a list (as shown in figure above). The data for currently selected MetFrag run will be shown on the tables.

11. FAQ

1. How do I change the amount of memory allocated to java virtual machine?

Open the startup script (MolFind.bat or MolFind.sh) with a text editor. Change the highlighted number to desired amount –Xmx1g (1g = 1 GB; g stands for giga bytes)

Recommended free text editors:

Windows – Notepad++ (http://notepad-plus-plus.org/), Komodo Edit

OSX – TextEdit, TextWrangler, Komodo Edit

Linux: gedit

2. What happens when I double click on the startup script (not the jar file) to start the program?

Memory allocation information is passed on to java virtual machine via startup scripts.

3. What happens if I don't select a working directory for my calculations?

By default, output files will be saved in tempMolFind folder located under user's home directory. However, using "tempMolFind" is not recommended except for testing (files in tempMolFind folder are deleted at the startup).

4. I don't see the CID number of my target in the identification panel?

Sometimes MolFind can list a different stereoisomer than your target. Look Under PubChem StereIsomers column as well.

5. Nothing happens when I double click MolFind.bat

This happens when operating system cannot locate the java run time (jre)

Add

C:\Program Files (x86)\Java\jre7\bin;

or

C:\Program Files\Java\jre7\bin; to windows path. (jre might be found at a different location; Look for jre6\bin or jre7\bin directories on your system)

6. How do I know java runtime is correctly installed and java path is correctly set?

Type **java –version** at the command prompt (dos prompt in windows or terminal in osx / linux)
This command should return the version number of java runtime

7. I have a question? How do I contact the authors?

You can use either the contact form on the web site or MolFind google+ page for submitting questions.