FOR-IDENT platform

User Manual

(3. Version, Mai 2017)

Responsible authors: Sylvia Grosse, Thomas Letzel (TUM)

In cooperation with the following institutes: LfU (Marion Letzel, Manfred Sengl, Veronika Gronau), LW (Thomas Lucke, Wolfgang Schulz), BWB (Uwe Duennbier, Robert Bloch), HSWT (Frank Lesske, Tobias Placht, Marco Luthardt, August Gilg)

Cite as: Grosse S., Letzel T.: User Manual for FOR-IDENT Database, 2017 (3), 1-35

funded by:

Bundesministerium für Bildung und Forschung

Table of content

1. General information

The database 'STOFF-IDENT' was originally be developed within the research project 'RISK-IDENT', funded by the German Federal Ministry of Education and Research (BMBF) in the time period 11/2011-03/2015. The database is further be developed with new functions as well as with increased amount of compounds and proven data quality in the research project 'FOR-IDENT' funded by the German BMBF in the time period 03/15-02/17.
During the first period of the FOR-IDENT project the original idea of STOFF-IDENT was continued and extended. The basis remained the same, however new requirements and workflow strategies necessitated the following:
- modern and responsive design
- a better support of Excel as import and export format
- easier and innovative development of modules for the integration of other software services
- better focus for the user on his/her workflow and data workout
- Screen with better result presentation
- Drag'n' drop for data files in the search function
- Search-History
- Quality increase of included data and other error eliminations

Primary the compound database 'STOFF-IDENT' is used for identification purposes, more exactly for non-target screening ('Hidden Targets' and/or 'Known Unknown') and suspected-target screening analysis in different parts of water research.[1] This is why just water relevant organic molecules, their transformation products and metabolites that occur in the environment are listed.[2]

Besides common compound parameters as Name, CAS Number, Formula, InChi key, IUPAC Name, SMILES Code, monoisotopic mass, also physicochemical indications as the logP and the logD value for four different pH values are included.[2]

The database query is possible by using e.g. the compound name as well as the monoisotopic mass. Also a normalization of the retention time is integrated to use this chromatographic parameter. This tool, called 'Retention Time Index (RTI)' normalizes the retention time based on the logD value. Therefore a standard mixture with 'calibration molecules' has to be measured with the same chromatographic method as for the analytes that will be searched afterwards in the database. You can find an example of such a mixture in the excel data file 'template_RTI.zip' in the sheet 'calibration'. This also can be opened by pressing the 'Help' button in the molecular screening part. Of course also an own composition of calibration molecules can be used. Here it is important just to use neutral charged molecules that include the total logP/logD range of your method. Be aware that the corrections as well as results obtained and published so far are only strictly valid using these calibration molecules as shown on page 23.[3a,b]

During the second period of the FOR-IDENT Project other open access available databases are connected directly to the platform or Links to the appropriate website is implemented. New formats to import data with MSMS information is now available as well as information about predicted transformation products.

The manual will be updated regularly if new functions are available.

[1] T. Letzel *et al.,* Chemosphere **2015**, 137, 198-206.
[2] http://risk-ident.hswt.de/media/downloads/STOFF-IDENT%20in%20der%20Anwendung_Dr_%20Marion%20Letzel.pdf (German)
[3a] http://risk-ident.hswt.de/media/downloads/Retention%20Time%20Index%20(RTI)_Dr_%20Thomas%20Letzel(1).pdf
[3b] http://risk-ident.hswt.de/media/downloads/Statusseminar_TUM.pdf

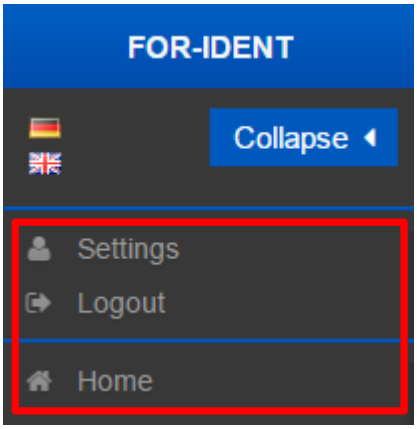
1.1 Languages, Settings, Home and Logout

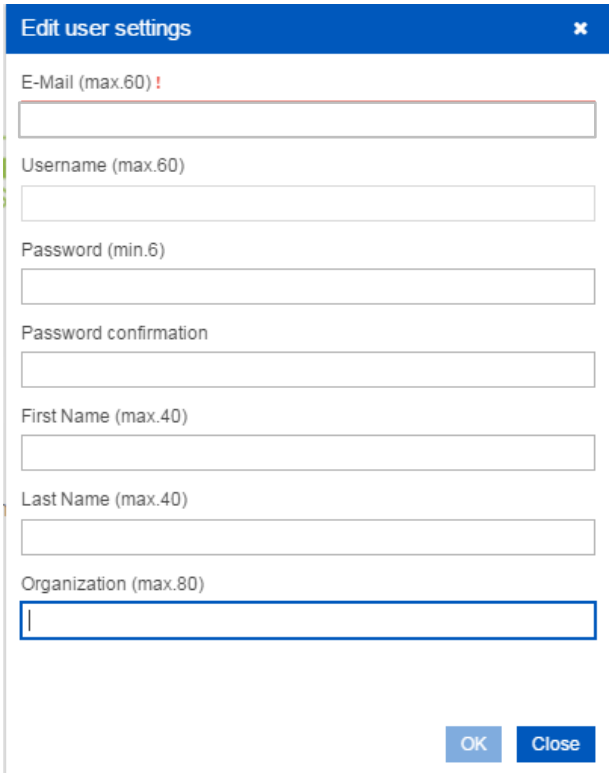
Switch between the languages: English and German

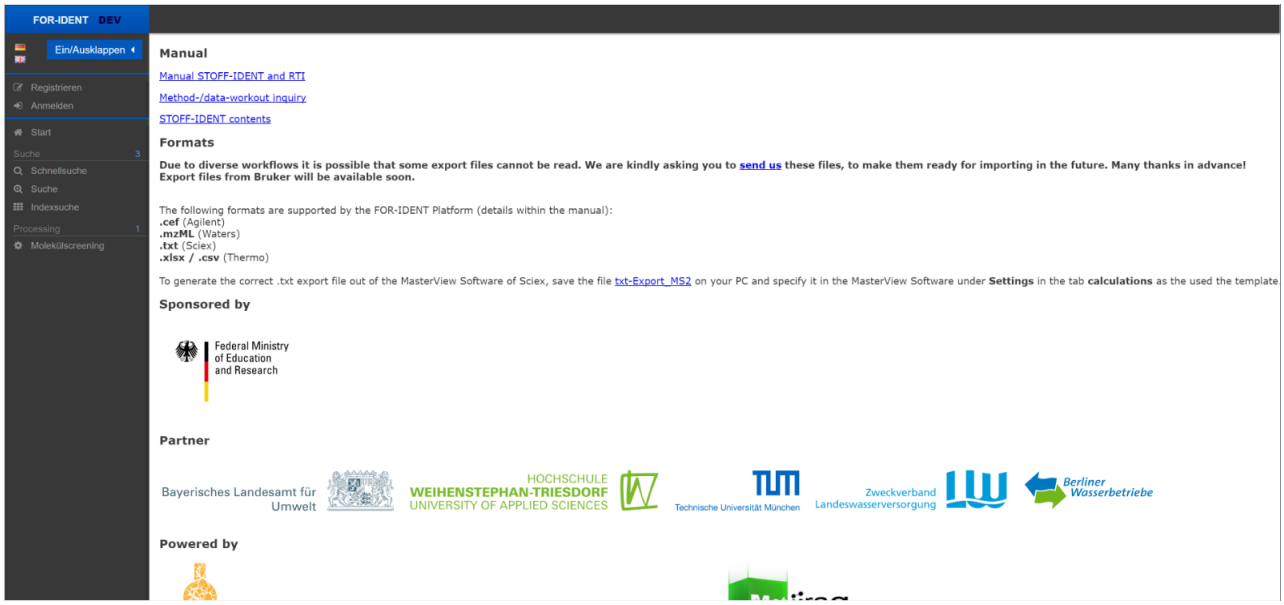
Settings: change user setting (e.g. password)

Home: Start page (e.g. FOR-IDENT manual, information of supported formats)

Logout







2. Abbreviations

Da	**Dalton**
M	**monoisotopic mass**
m/z	**mass to charge**
ppm	**parts per million**
RT	**retention time**
RTI	**Retention Time Index**
TP	**transformation product**

3. System requirements

All well-established operation software and browsers should be supported.

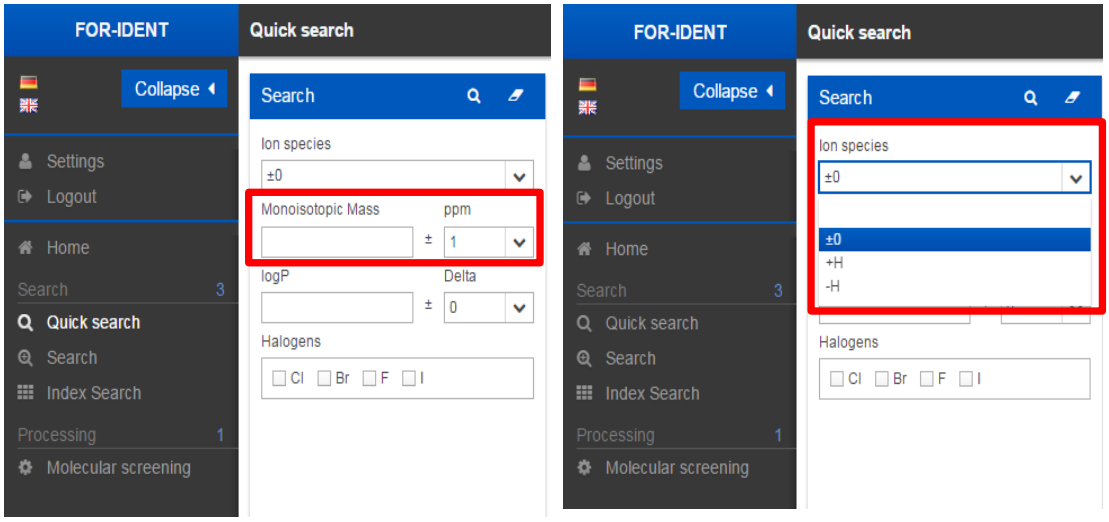
In general you may use the actual browser version, especially for Internet Explorer use version 9 or higher.

The Excel version should be at least Excel 97.

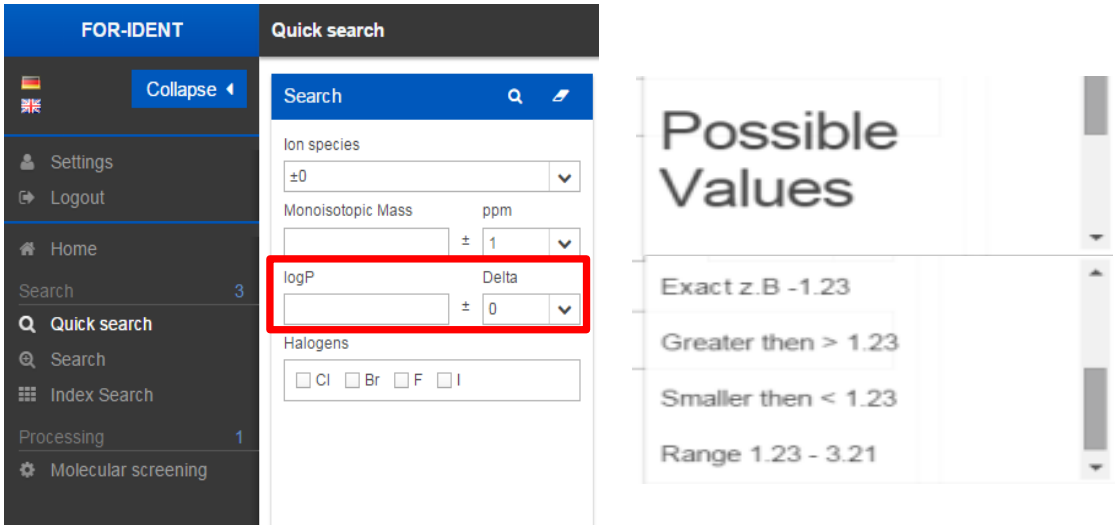
4. Search tabs

4.1 Quick search

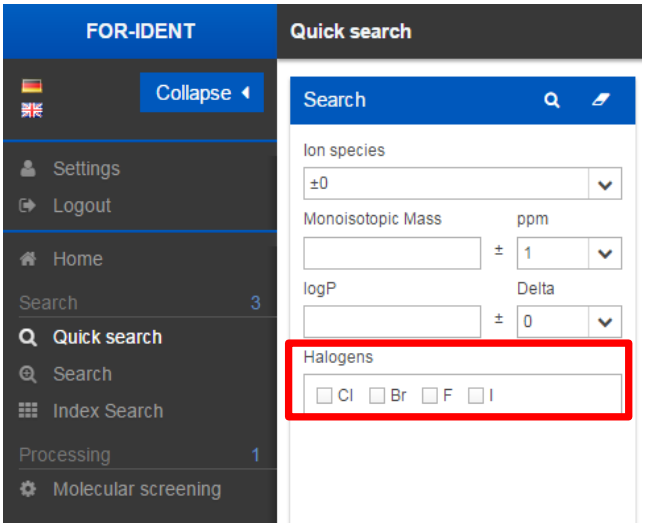
1. Input of Monoisotopic Mass as M or m/z by selection of the Ion species and determination of the mass accuracy (ppm)



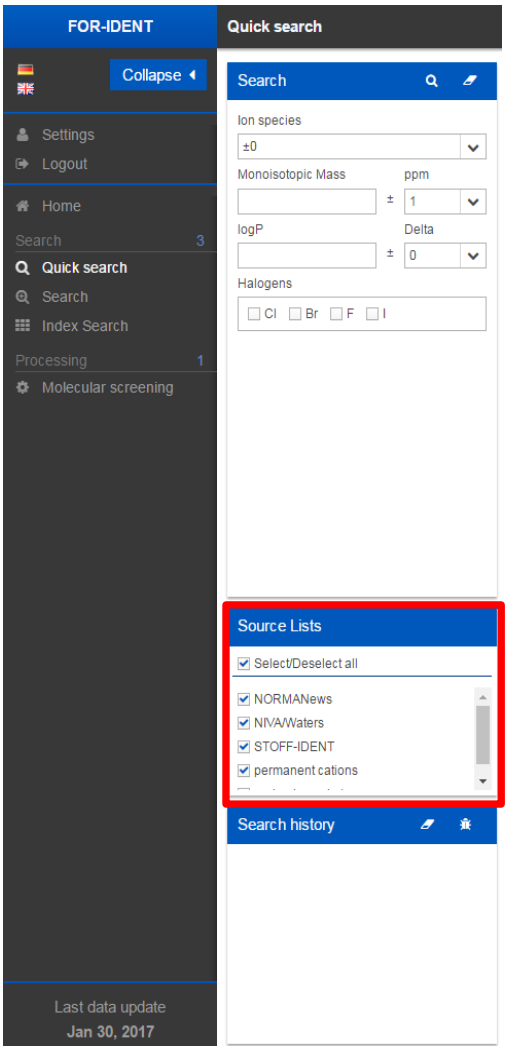
2. Entry of logP value instead of Monoisotopic Mass
   (By going with the cursor to the logP field, the possible values appear)



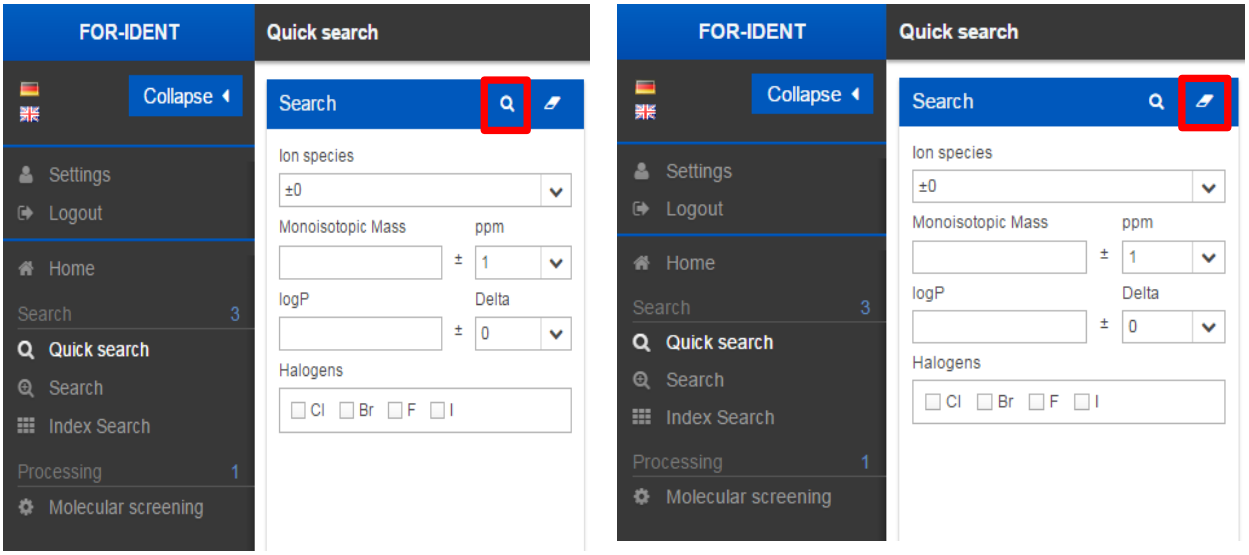
3. Selection of possible halogens, if known by isotopic pattern



4. Selecting the source lists (lists from various owners and institutions)
   (select here in which list FOR-IDENT will perform the search; this tool is still under progress and will be extended with every update)



5. Press ‘Start search’ button (magnifier)
6. By pressing the ‘Clear fields’ button (rubber) all entries will be set to default values



4.2. Search

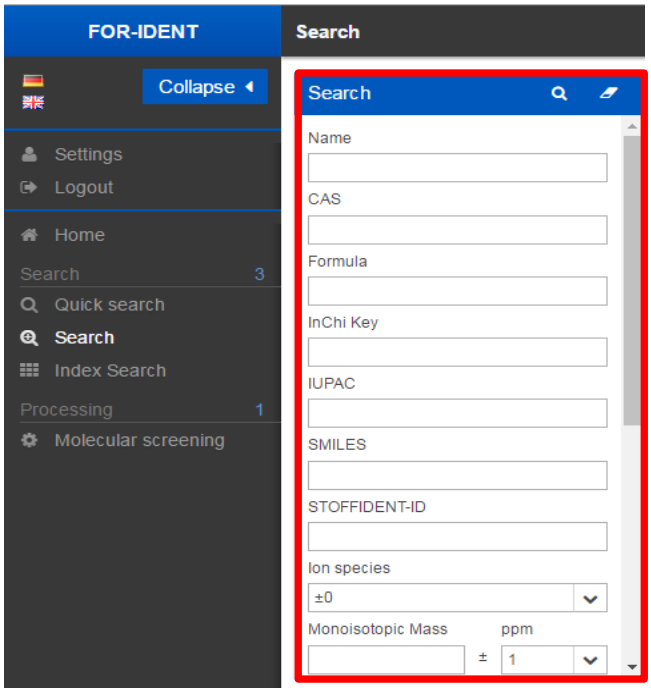
1. Entry of Name, CAS Number, Formula, InChi key, IUPAC Name, SMILES Code, STOFF-IDENT ID, Monoisotopic Mass as M or m/z by selection of the ion species and determination of the mass accuracy (ppm)
   (By going with the cursor to the Monoisotopic Mass field, the possible values are shown)

   and/ or logP value
   (By going with the cursor to the logP field, the possible values are shown)

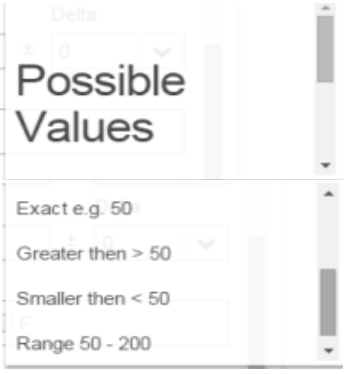
   and/ or Halogens if known

   and/ or selection of a category (this tool is still under progress and will be extended with every update)

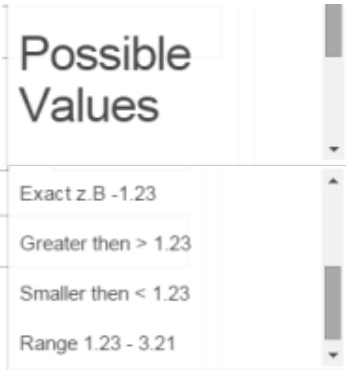
   and/ or selecting the source list
   (select here in which list FOR-IDENT will perform the search; this tool is still under progress and will be extended with every update))



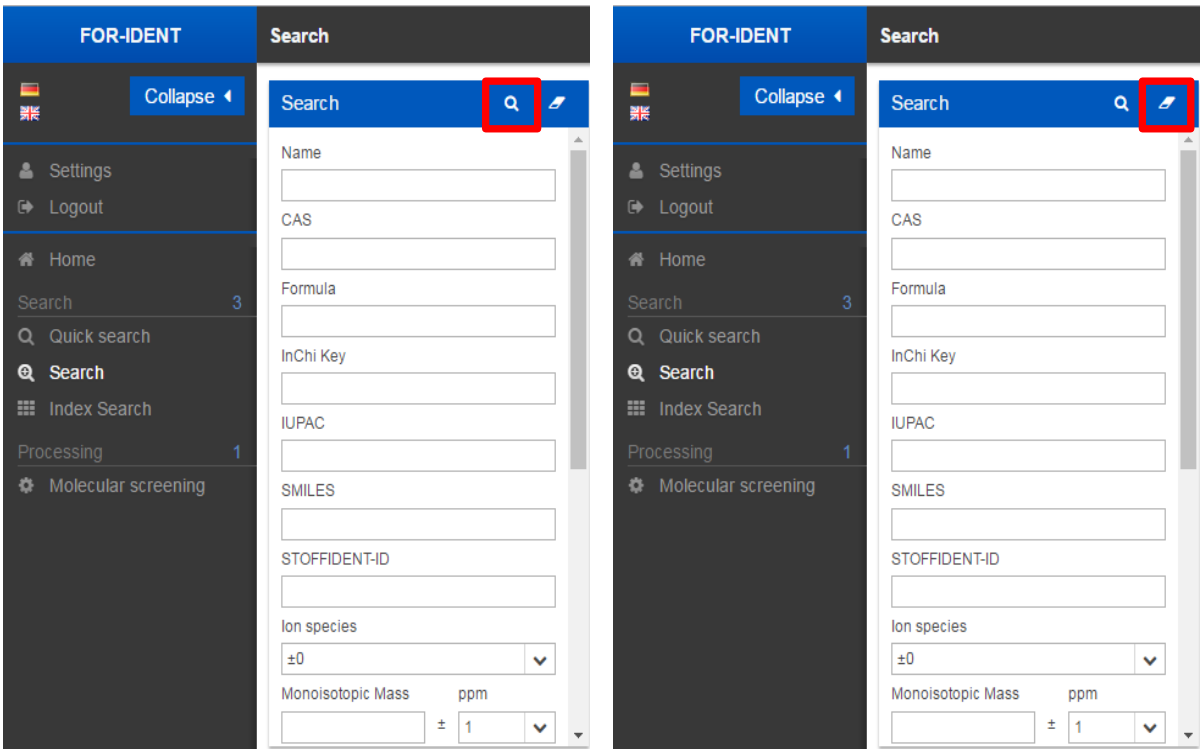
Mass values



logP values

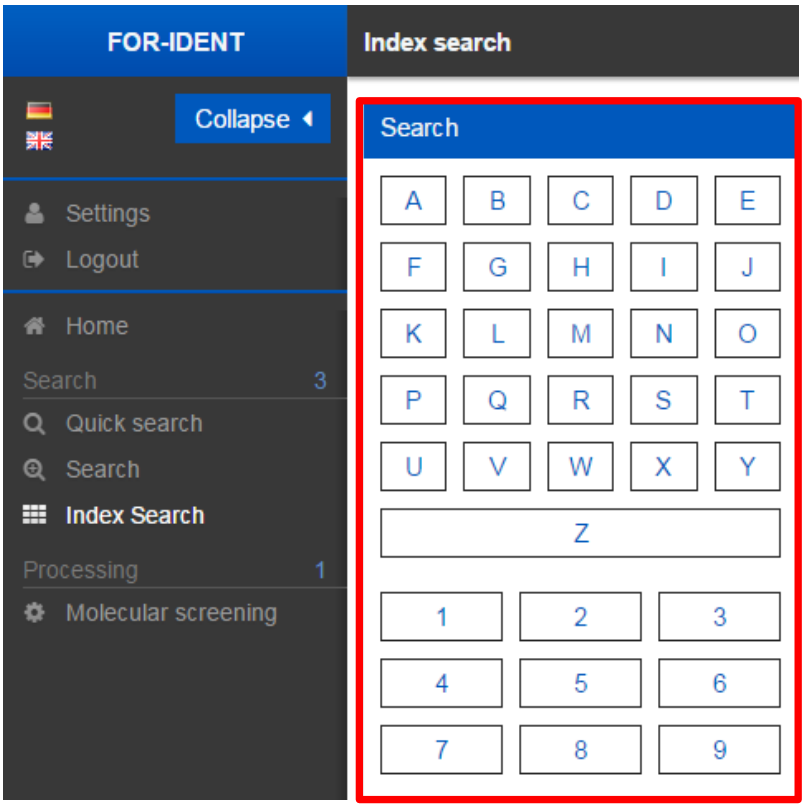


2. Press ‘Start search’ button (magnifier)
3. By pressing the ‘Clear Fields’ button (rubber) all entries will be set to default values



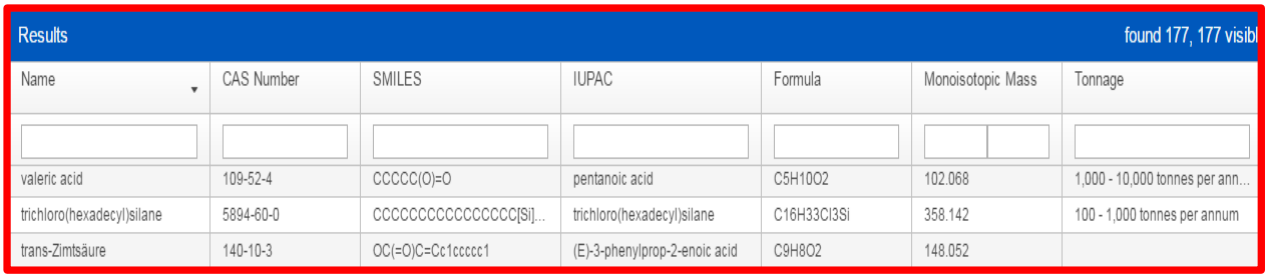
4.3. Index Search

1. Selection of search values via single letters or numbers
2. Selecting the source list
   (select here in which list FOR-IDENT will perform the search; this tool is still under progress and will be extended with every update)



4.4. Search result page

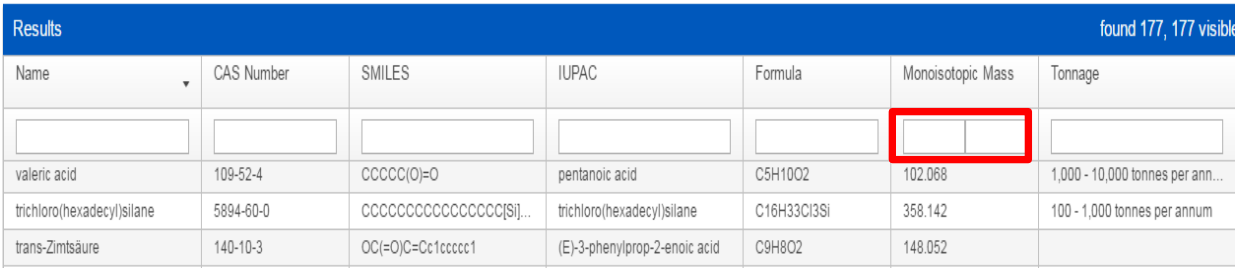
1. After pressing the ‘Start search’ button (magnifier) the result table appears



Results — found 177, 177 visibl

Name	CAS Number	SMILES	IUPAC	Formula	Monoisotopic Mass	Tonnage
valeric acid	109-52-4	CCCCC(O)=O	pentanoic acid	C5H10O2	102.068	1,000 - 10,000 tonnes per ann...
trichloro(hexadecyl)silane	5894-60-0	CCCCCCCCCCCCCCCC[Si]...	trichloro(hexadecyl)silane	C16H33Cl3Si	358.142	100 - 1,000 tonnes per annum
trans-Zimtsäure	140-10-3	OC(=O)C=Cc1ccccc1	(E)-3-phenylprop-2-enoic acid	C9H8O2	148.052	

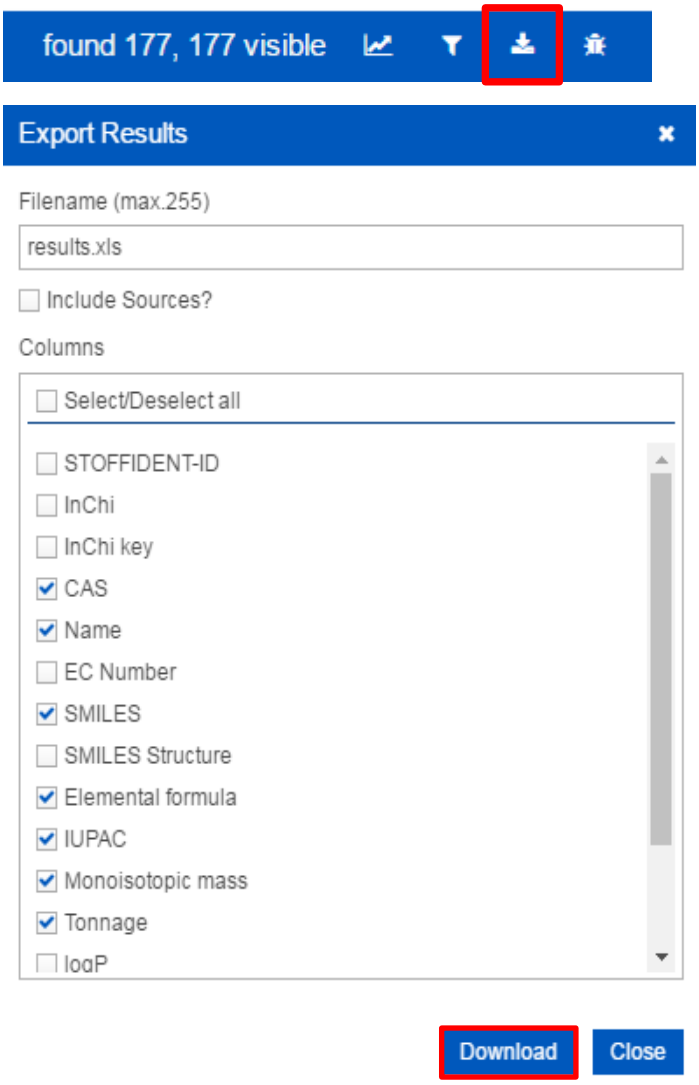
2. A filtering allows to narrow down the results e.g. for Monoisotopic Mass values (left side: min value, right side: max value)



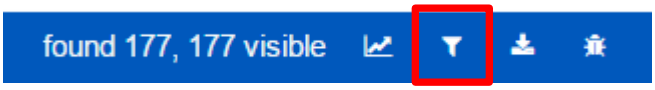
Results — found 177, 177 visible

Name	CAS Number	SMILES	IUPAC	Formula	Monoisotopic Mass	Tonnage
valeric acid	109-52-4	CCCCC(O)=O	pentanoic acid	C5H10O2	102.068	1,000 - 10,000 tonnes per ann...
trichloro(hexadecyl)silane	5894-60-0	CCCCCCCCCCCCCCCC[Si]...	trichloro(hexadecyl)silane	C16H33Cl3Si	358.142	100 - 1,000 tonnes per annum
trans-Zimtsäure	140-10-3	OC(=O)C=Cc1ccccc1	(E)-3-phenylprop-2-enoic acid	C9H8O2	148.052	

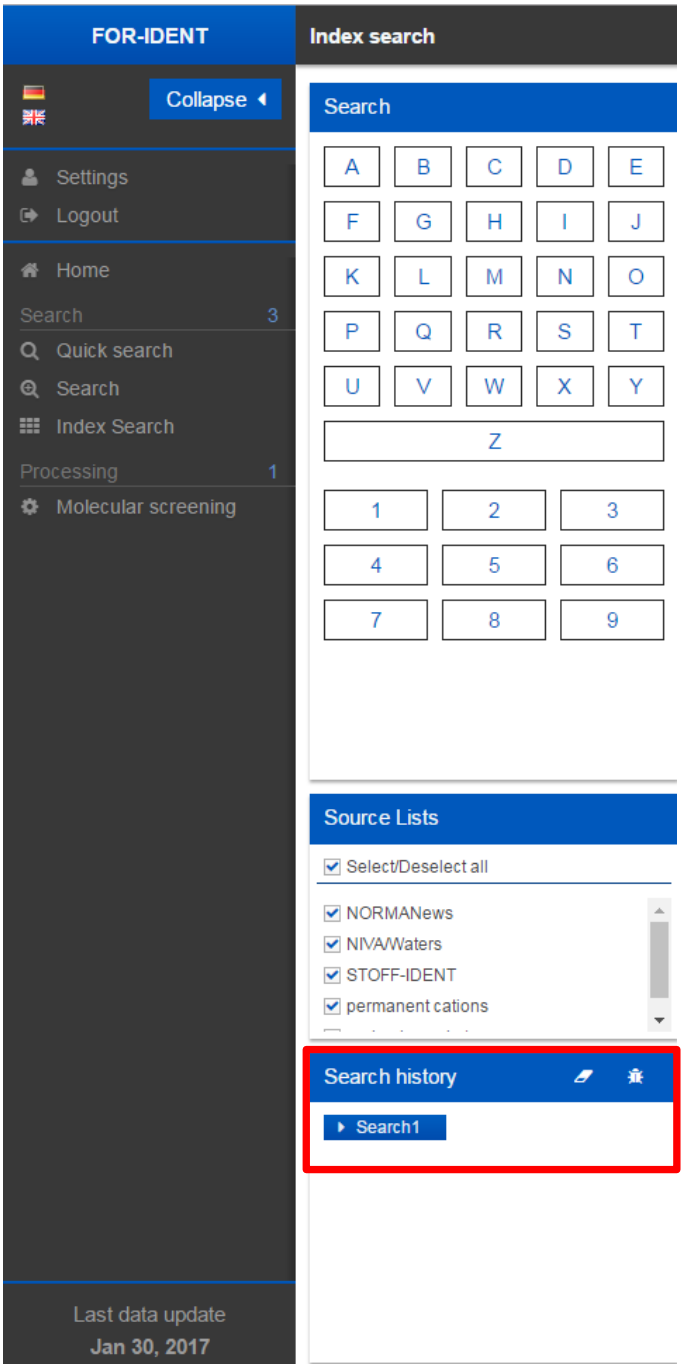
3. By pressing the ‘Download’ button, a new window appears
4. Selection of favored parameters for download to excel



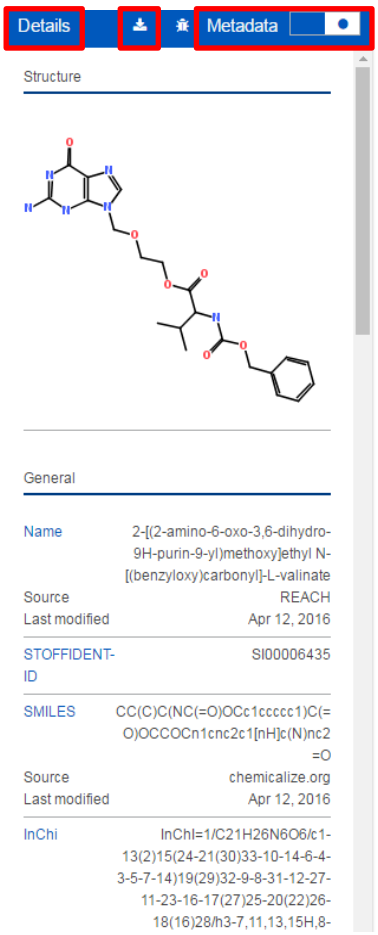
5. By pressing the ‘Clear filter’ Button all filters will be set to default values and the original result table appears

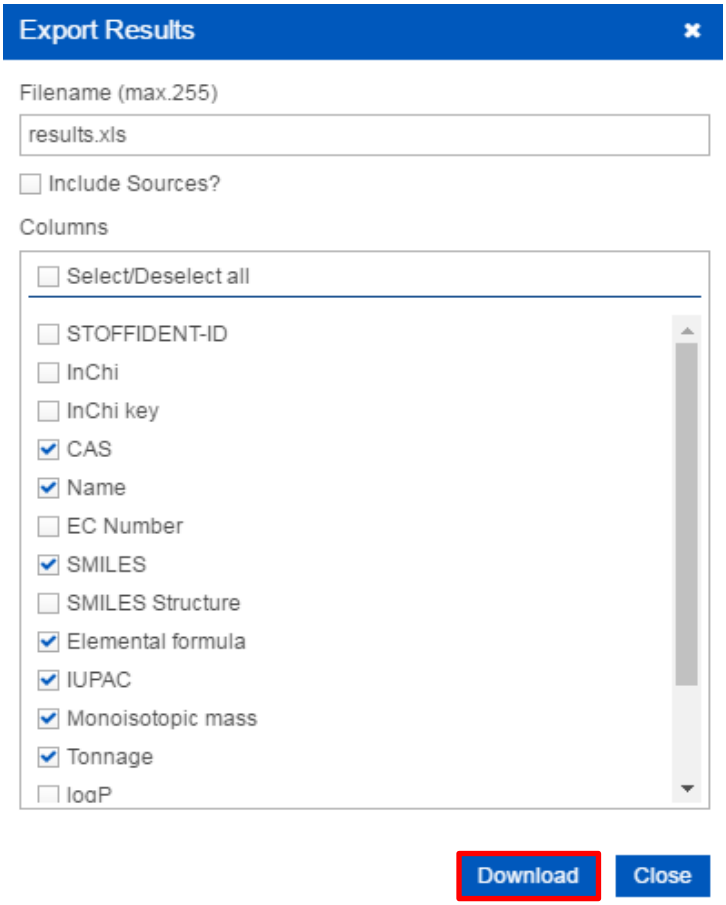


In the ‘Search history’ area all performed searches were stored and can be reloaded again by ‘doubleclick’.



In the 'Details' tab you can find specific information about the analyte. All information can be downloaded as an excel-file. Besides you have the choice to see the Metadata or not.

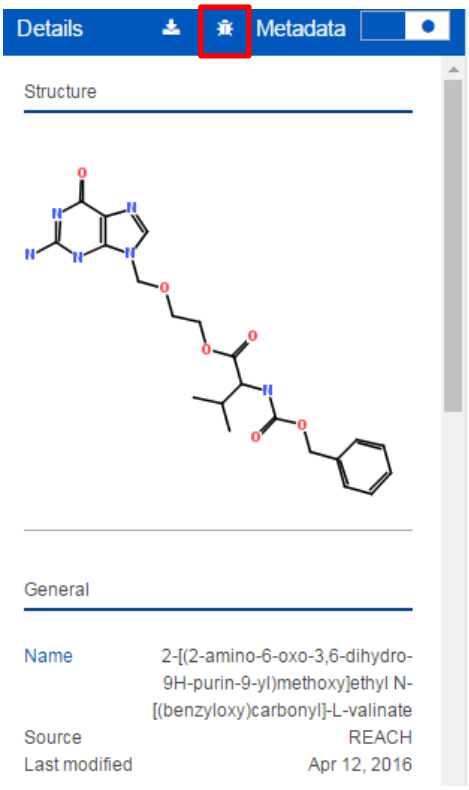


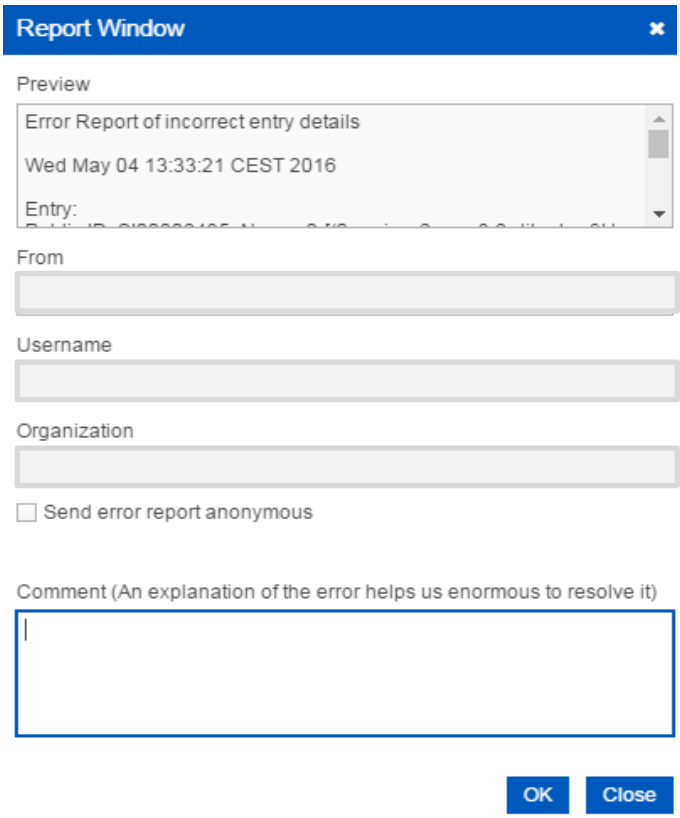


You can change the Layout with three different options

In the Details at the web search part, you can directly search the specific compound via internet or you go to the External Links to get more information of the compound

If you find errors in the parameters given, e.g. wrong sum formula, SMILES code, etc. you can use the 'report record as faulty' button to send it directly to the software developer team by putting your comments of the fault into the comment field.



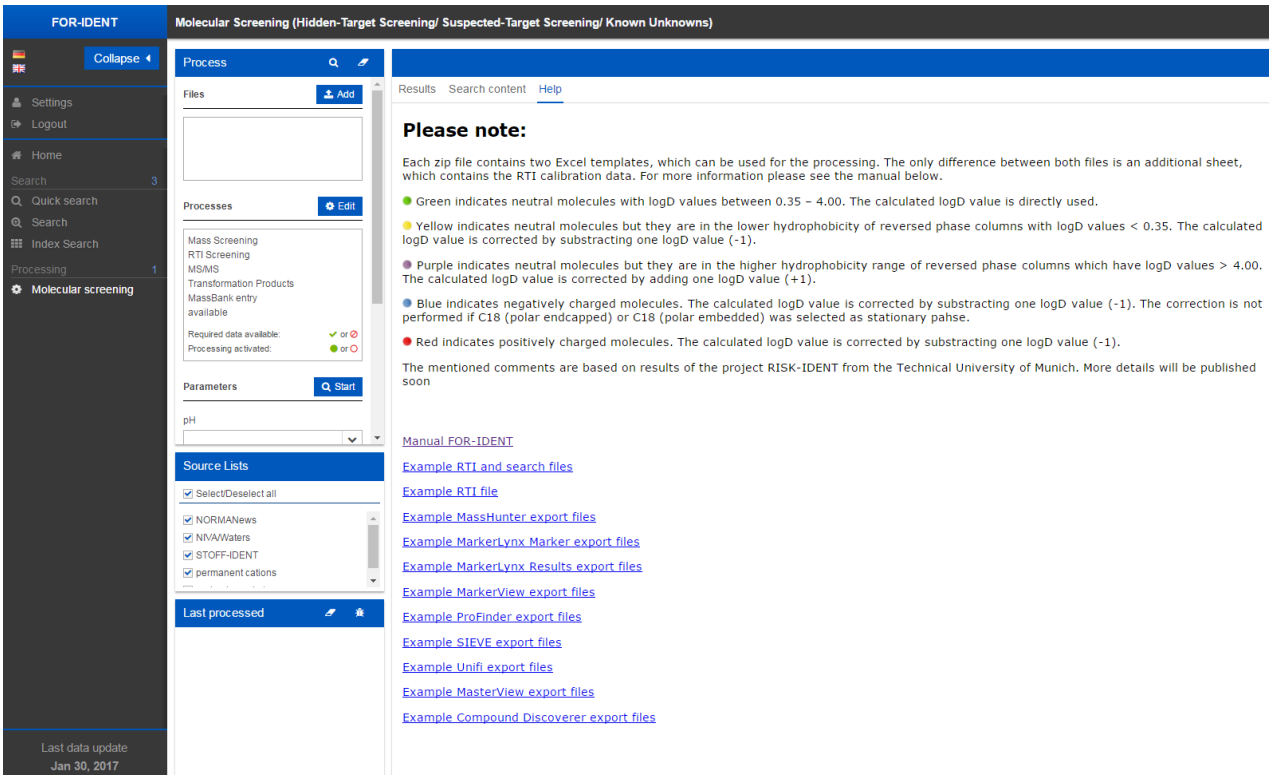


5. Processing

5.1 Molecular Screening

In the Molecular Screening area it is possible to process a list of compounds, e.g. from a non-target screening measurement, by searching the database by mass and/ or formula and if available retention time by using the retention time index function (possible for reversed phase columns) as well as MS/MS information, if available.

The software will perform the database search based on the available information (Mass, RT, RTI calibration, MS/MS) uploaded in the file. Detailed information are given in the following sections.



Possible formats

All templates can be found under the 'Help' page.

The 'template_general_search.zip' searches the database only by mass.

By using the 'template_general_RTI.zip' the database is searched by mass or formula and by retention time, including retention time index with calibration data.

Other possible Excel/ Comma separated value formats are:

- *Only mass:*

template_MH_search.zip; template_ML_marker_search.zip; template_ML_results_search.zip; template_MV_search.zip; template_MSV_search.zip; template_PF_search.zip; template_SI_search.zip; template_CD_search.zip; template_TF_search.zip; template_UN_search.zip

- *mass/formula and/or retention time*:

template_MH_RTI.zip; template_ML_marker_RTI.zip; template_ML_results_RTI.zip; template_MV_RTI.zip; template_MSV_RTI.zip; template_PF_RTI.zip; template_SI_RTI.zip; template_CD_RTI.zip; template_UN_RTI.zip

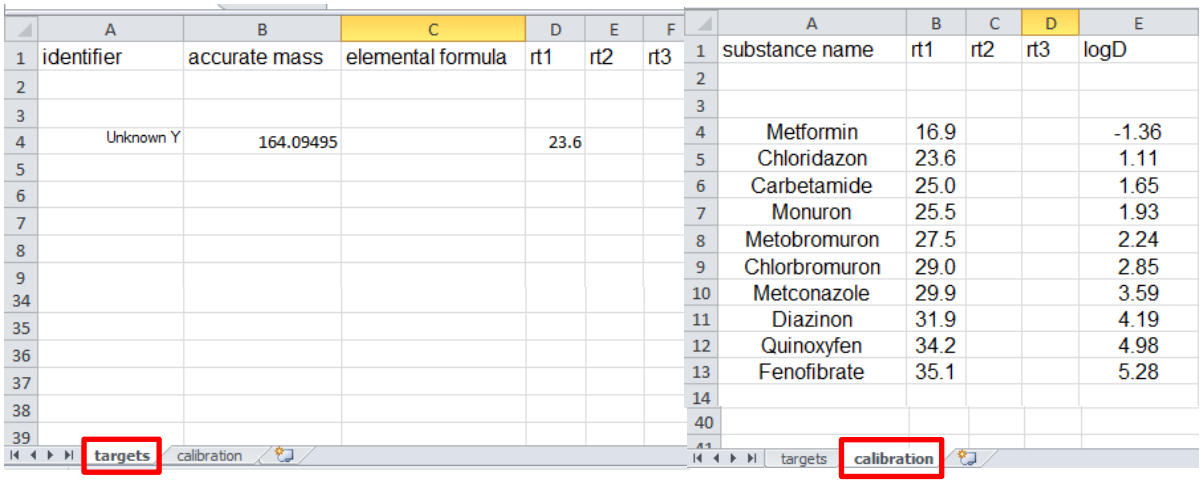
Explanation to use excel-formats:
Agilent (MH: MassHunter qualitative analysis)
Waters (ML: MassLynx Marker und ML: MassLynx results)
Sciex (MV: MarkerView)
Sciex: (MSV: MasterView)
Agilent (PF: Profinder)
Thermo Scienitifc (SI: SIEVE)
Thermo Scientific (CD: Compound Discoverer 2.0)
Thermo Scientific (TF: Tracefinder 4.1)
(For TraceFinder 4.1 and lower versions Excel files may contain duplicates, use Excel function Data-> Remove Duplicates)
Waters (UN: Unifi)

There is also the possibility to use non-excel formats:

.cef (Agilent)
.mzML (Waters)
.txt (Sciex) (please, install the template first. Instructions can be found on the Start page)
.csv (Thermo Scientific)

If you find your export file listed above, you can use it directly without any changes.
To use the retention time index function upload an extra excel-file with the calibration data like template_calibration.

Using the retention time index function requires a single measurement with a set of calibration molecules measured with the same analytical method as the sample (see below). Retention times of the calibration molecules must be filled in column 'rt1' (required), 'rt2' and 'rt3' (optional). Two options are available using the calibration data.
First option is to add a second sheet in the report file of the sample (if the report is an excel-file, otherwise use the second option), which must be named 'calibration' and follows the design showed below on the right. Templates of different vendors can be found under the 'Help' page.



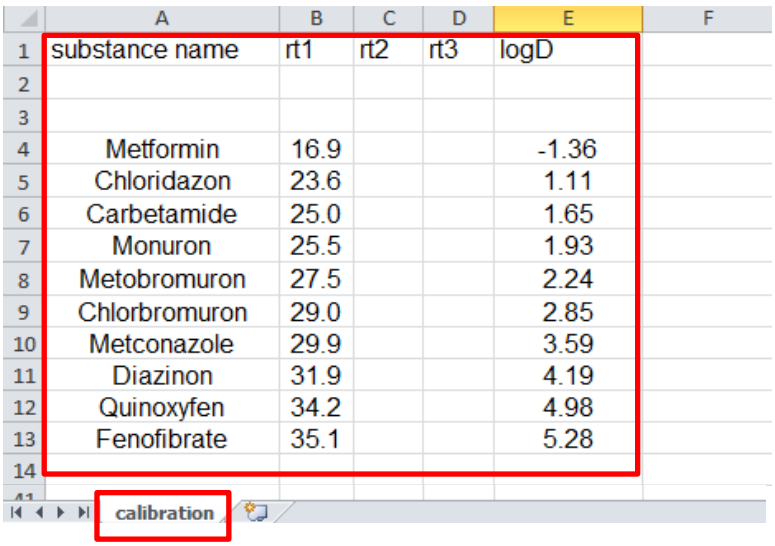
identifier	accurate mass	elemental formula	rt1	rt2	rt3
Unknown Y	164.09495		23.6		

targets | calibration

substance name	rt1	rt2	rt3	logD
Metformin	16.9			-1.36
Chloridazon	23.6			1.11
Carbetamide	25.0			1.65
Monuron	25.5			1.93
Metobromuron	27.5			2.24
Chlorbromuron	29.0			2.85
Metconazole	29.9			3.59
Diazinon	31.9			4.19
Quinoxyfen	34.2			4.98
Fenofibrate	35.1			5.28

targets | calibration

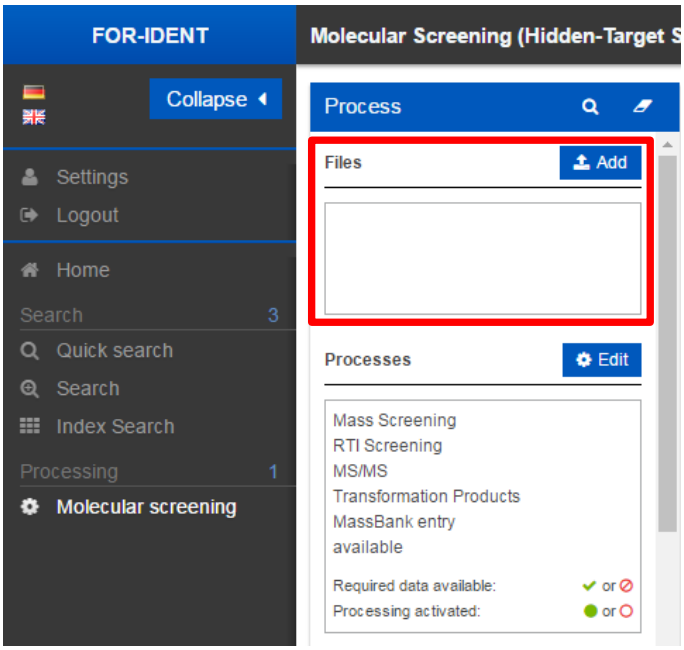
Second option is to use an extra excel-file, as can been seen below. A template can be found under the 'Help' page: 'template calibration'.



substance name	rt1	rt2	rt3	logD
Metformin	16.9			-1.36
Chloridazon	23.6			1.11
Carbetamide	25.0			1.65
Monuron	25.5			1.93
Metobromuron	27.5			2.24
Chlorbromuron	29.0			2.85
Metconazole	29.9			3.59
Diazinon	31.9			4.19
Quinoxyfen	34.2			4.98
Fenofibrate	35.1			5.28

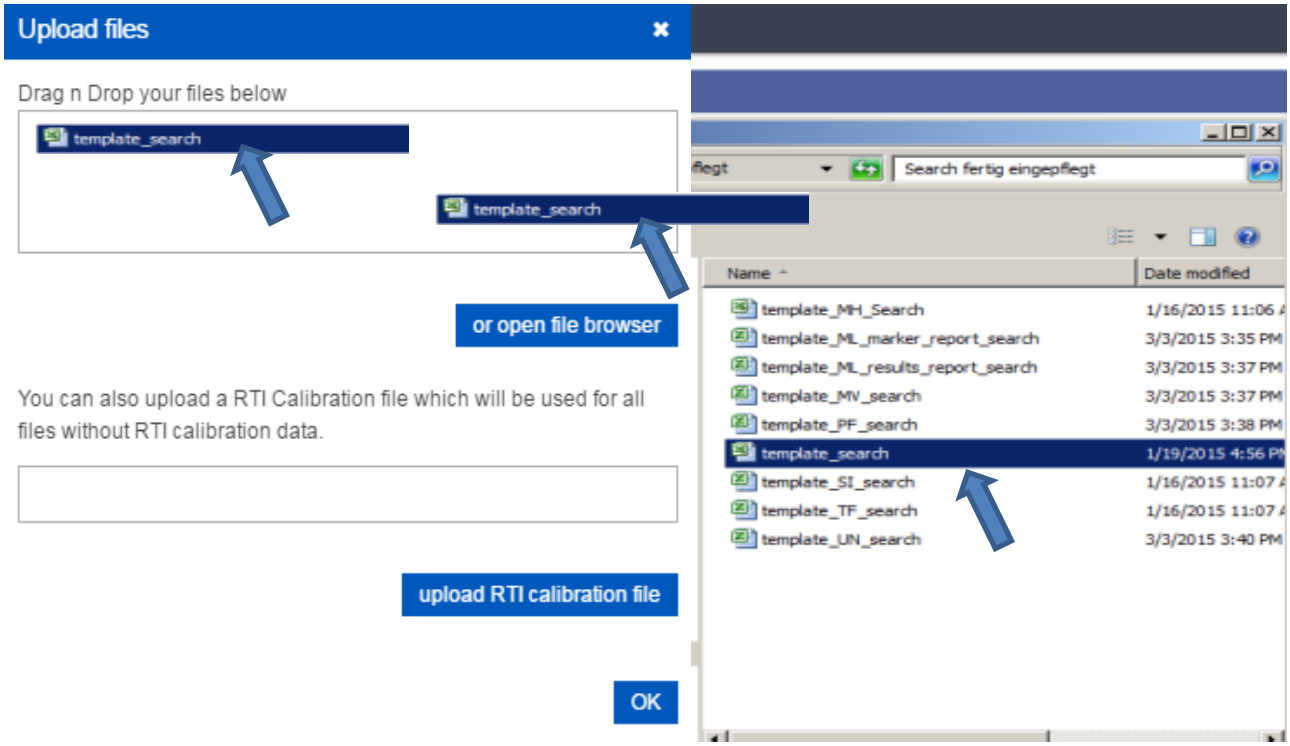
calibration

1. Press ‘Add File’ to upload data files

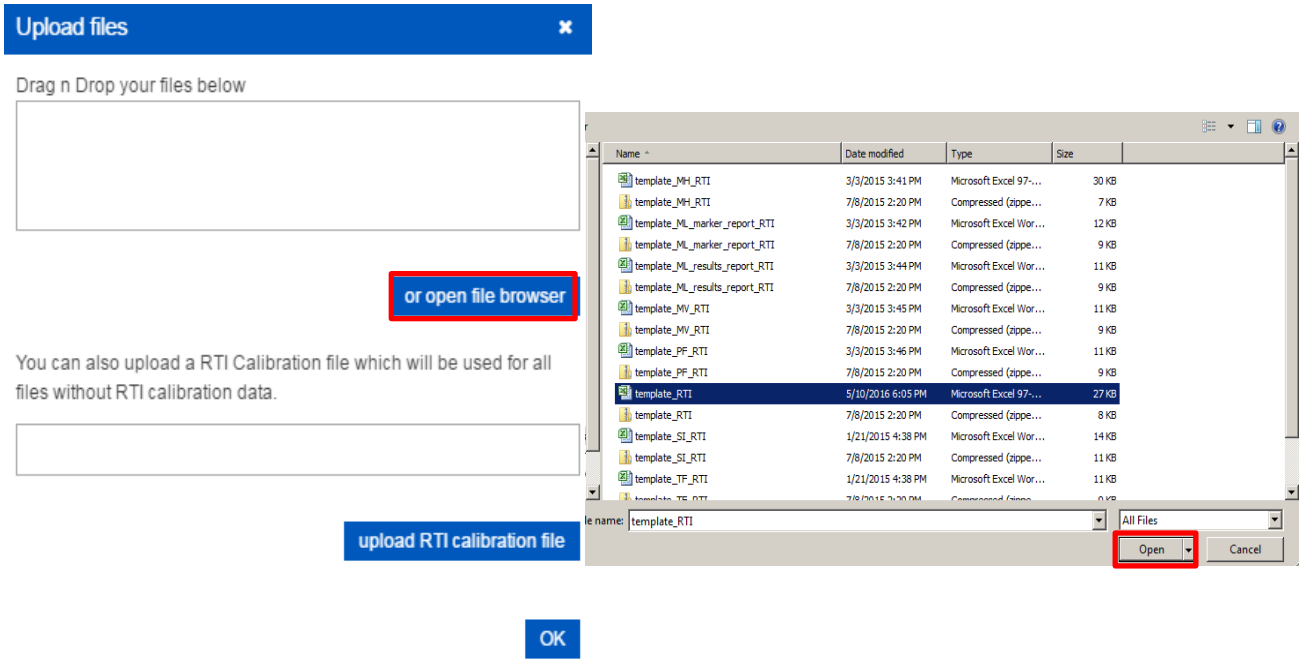


There are 2 options to import excel-files or other formats
You have the possibility to import one or more files at the same time
(refer to part ‘possible formats’ for more details on supported formats)

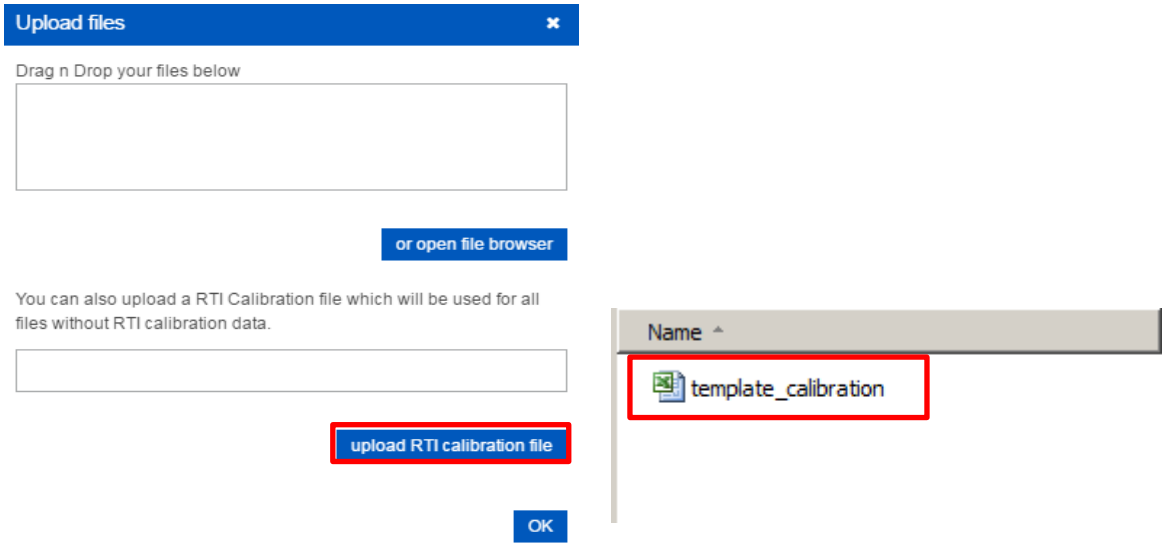
a) Drag`n Drop the file(s)



b) …or open file browser and select one or more files



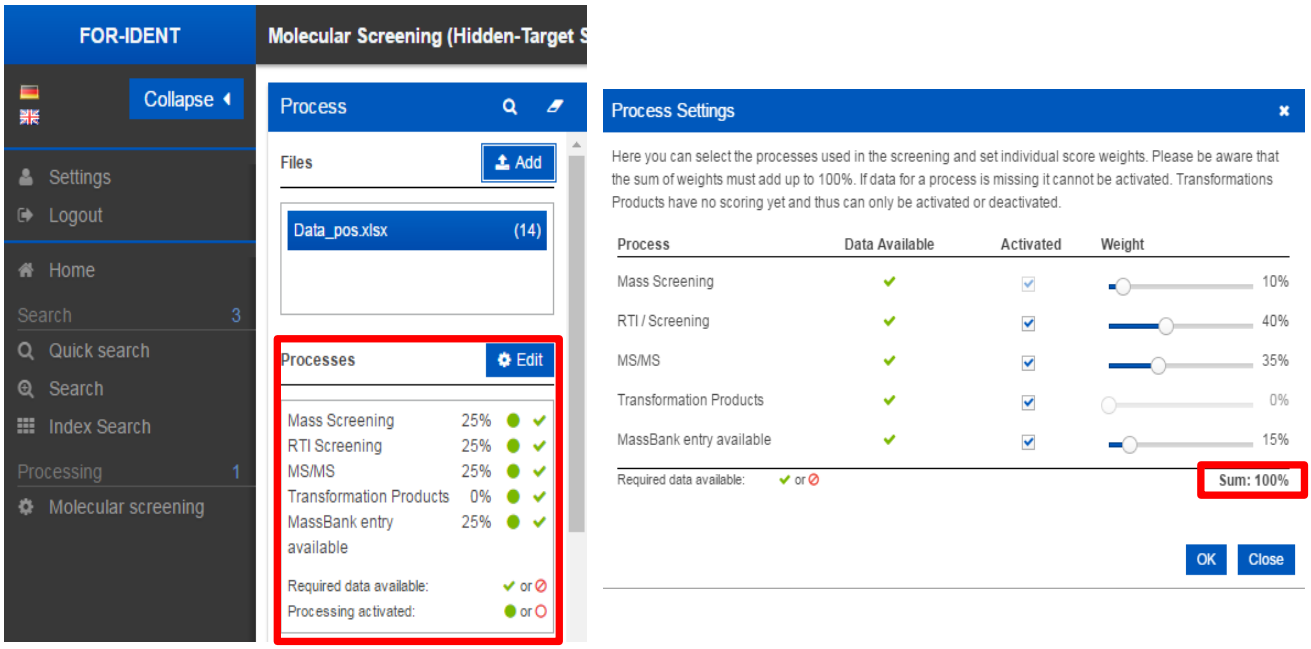
If needed it is possible to upload an extra excel-file with the RTI calibration data. This file will be used for all files uploaded in the upper part.



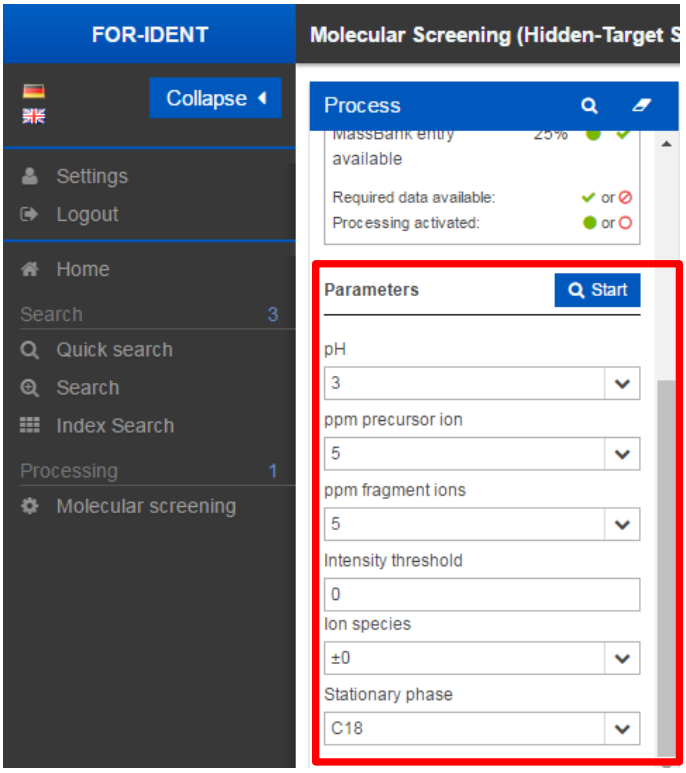
By using a series of different files in one batch, especially with and without RTI calibration data, the extra excel-file will be used, only if the uploaded file doesn´t contain the second sheet ‘calibration’.

2. Selecting the Process Settings

By clicking on ‚Edit' a new window appears. This gives you the possibility to choose the weights of the different functions by your own experience.
Note that the sum has to be 100 %.



3. Selection of chromatographic and mass spectrometric parameters, like pH value of used solvents, mass accuracy of precursor ion and fragment ions [ppm], intensity threshold of an uploaded MSMS spectra, Ion species and Stationary phase

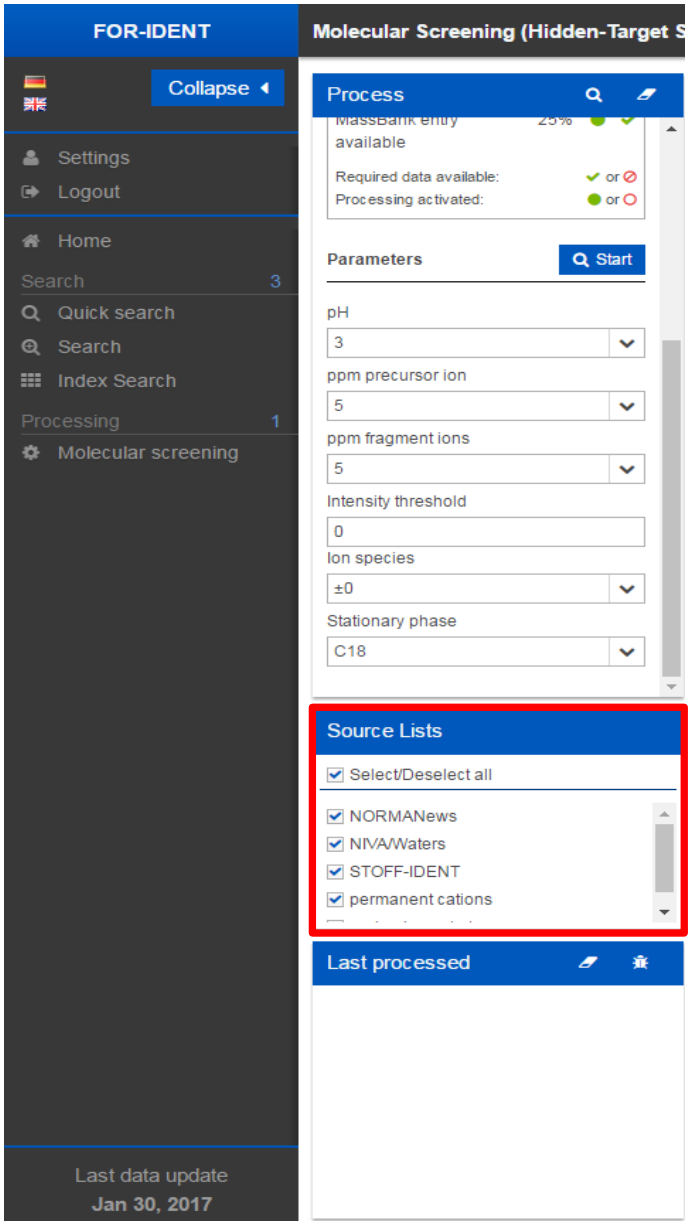


convincing results are tested for the following stationary phases so far: C18, C18 (polar endcapped), C18 (polar embedded), Phenyl and PFP (in-house column study) Other material gives probably uncertainties.

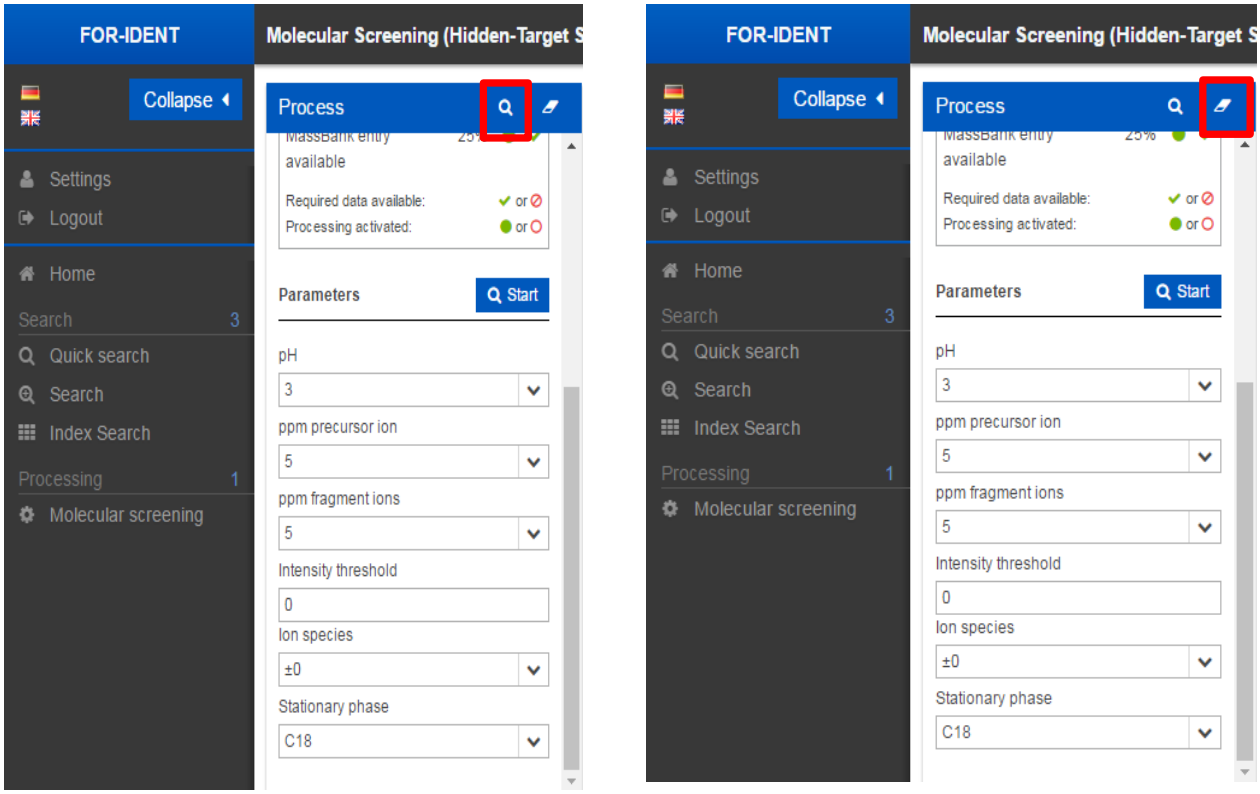
Once a selection is chosen for an uploaded file, it is kept even if you upload the next one

Besides or instead of the monoisotopic mass or m/z also the elemental formula can be used. The exact mass is then calculated by the database automatically and the accuracy is set to 0.01 ppm.
If just the monoisotopic mass or m/z is entered the accuracy is considered from 'ppm' field.

4. Selecting the source list
   (select here in which list FOR-IDENT will perform the search; this tool is still under progress and will be extended with every update)



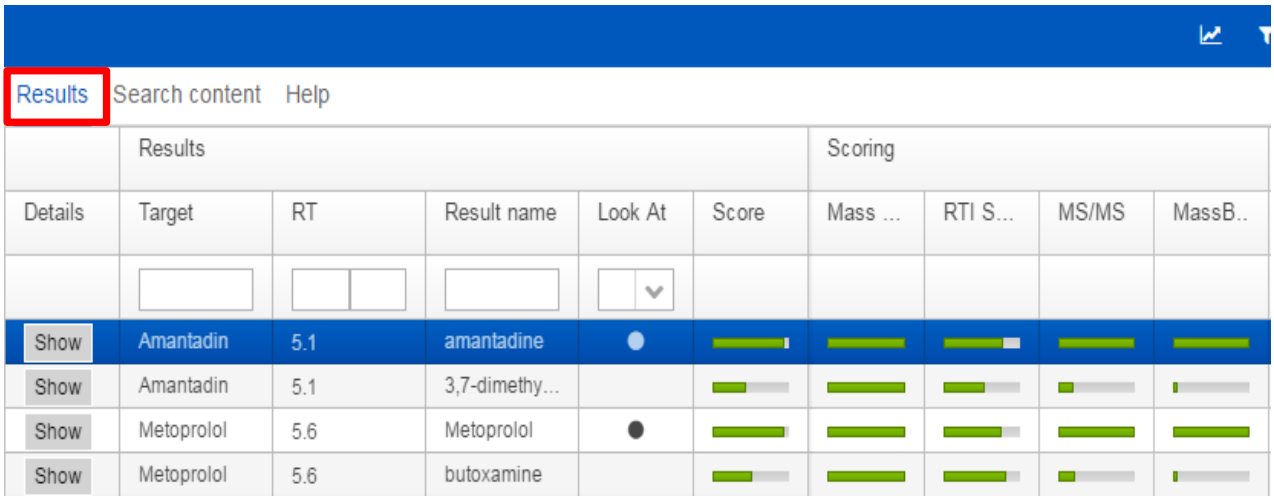
5. Press ‘Start search’ button (magnifier)
6. By pressing the ‘Clear fields’ button (rubber) all entries will be set to default values



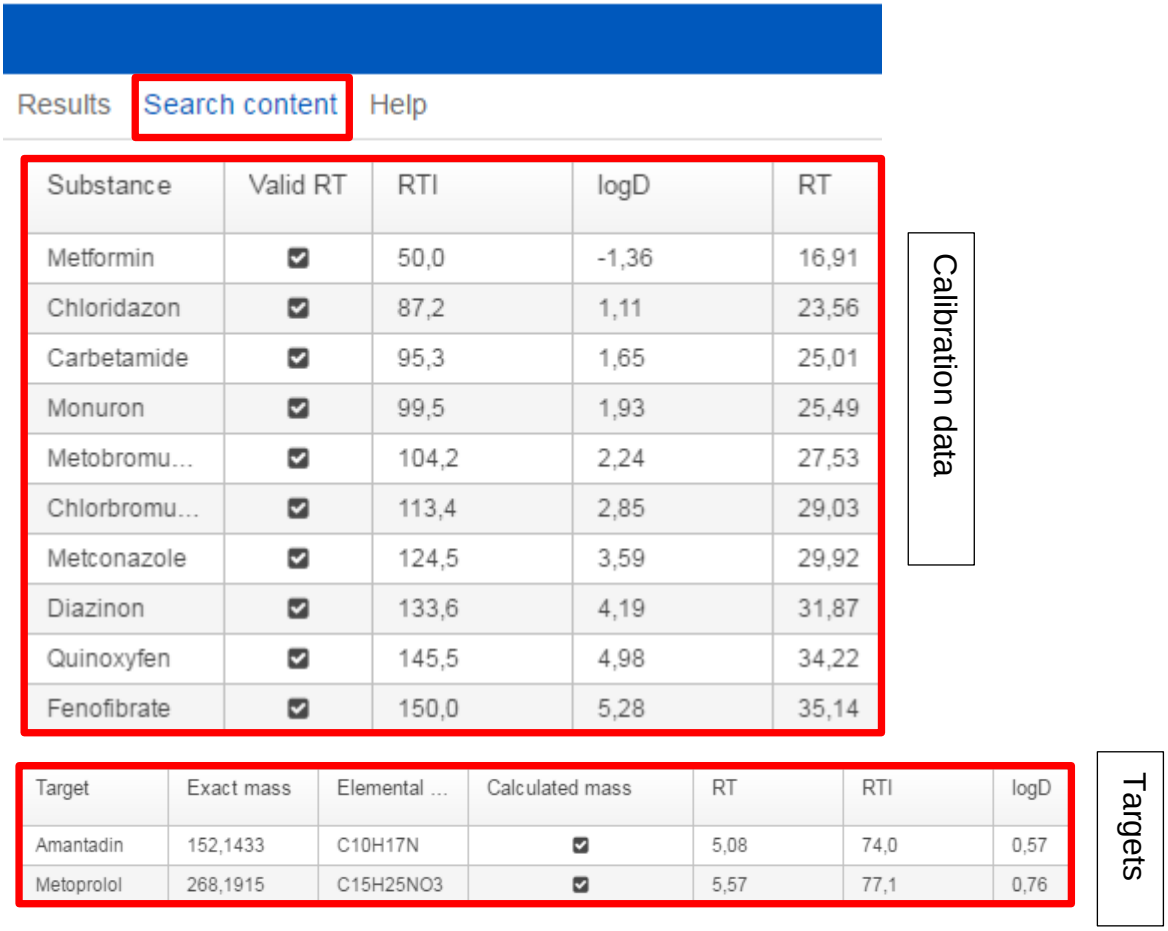
5.2 result page

5.2.1 Explanation of the result table

Under the first tab ‘Results’ the hits for the uploaded file is shown (explained in detail starting with 5.2.1.1)



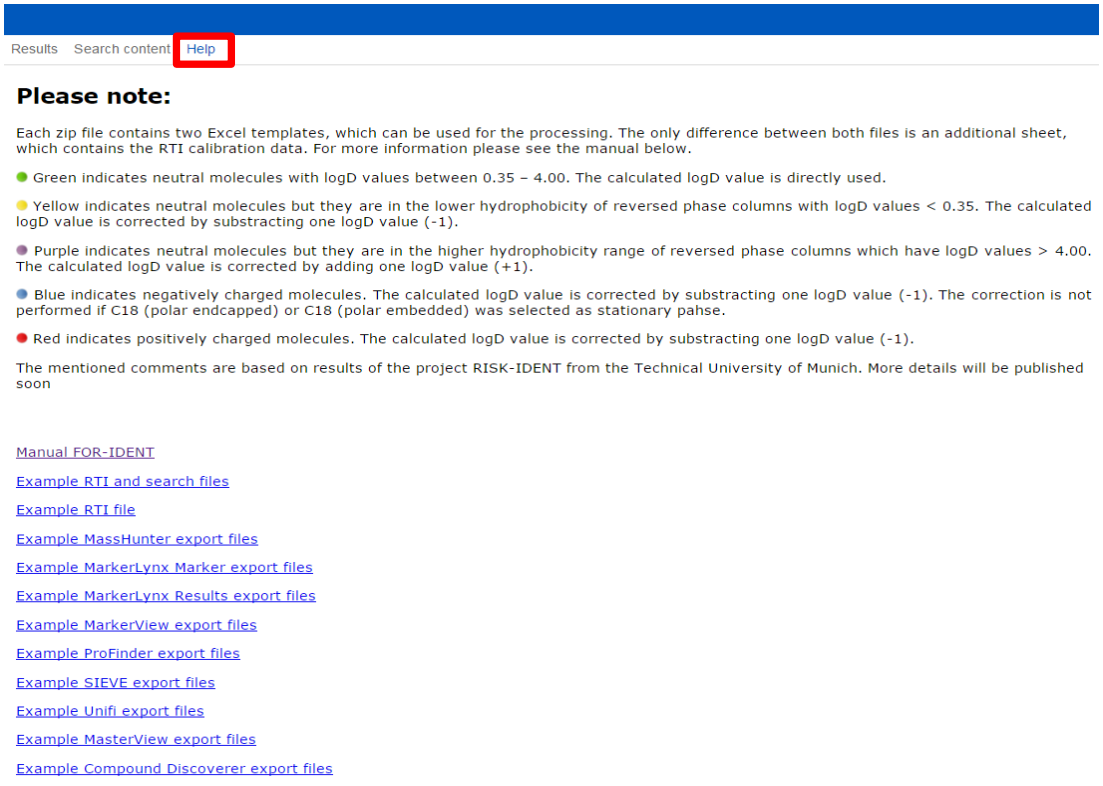
Under the second tab ‘Search content’ the data of the uploaded file is shown (in the upper part the calibration data, in the lower part the targets), with the calculation of RTI and determined logD values



Substance	Valid RT	RTI	logD	RT
Metformin	☑	50,0	-1,36	16,91
Chloridazon	☑	87,2	1,11	23,56
Carbetamide	☑	95,3	1,65	25,01
Monuron	☑	99,5	1,93	25,49
Metobromu...	☑	104,2	2,24	27,53
Chlorbromu...	☑	113,4	2,85	29,03
Metconazole	☑	124,5	3,59	29,92
Diazinon	☑	133,6	4,19	31,87
Quinoxyfen	☑	145,5	4,98	34,22
Fenofibrate	☑	150,0	5,28	35,14

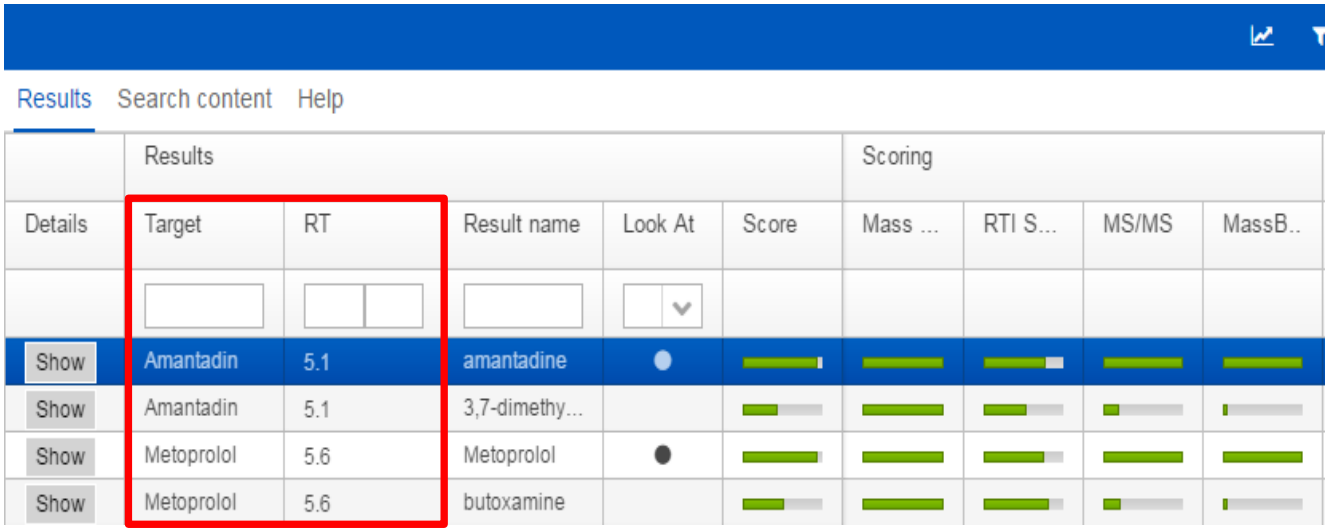
Target	Exact mass	Elemental ...	Calculated mass	RT	RTI	logD
Amantadin	152,1433	C10H17N	☑	5,08	74,0	0,57
Metoprolol	268,1915	C15H25NO3	☑	5,57	77,1	0,76

Under the third tab ‘Help’ the explanation of pH dependency is given and all templates are listed as well as the user manual can be found



5.2.1.1 columns Target and RT

In the column ‘Target’ the name is shown which was used in the uploaded report file as identifier and in the column ‘RT’ the retention time, if uploaded.



5.2.1.2 columns Result name and pH dependency

The column pH dependency shows the charge of the molecule at a definite pH value and in which range the logD value is, as the software calculates an automatic correction of the logD value depending on the color code.

Details of the color code are as follows:

Green indicates molecules, which are charged neutral for each pH value.
The calculated logD value is directly used.

Yellow indicated molecules are charged neutral for each pH value, but are in the lower hydrophobicity range of reversed phase columns with log D values < 0.35. The calculated log D value is corrected by subtracting one logD value (-1).

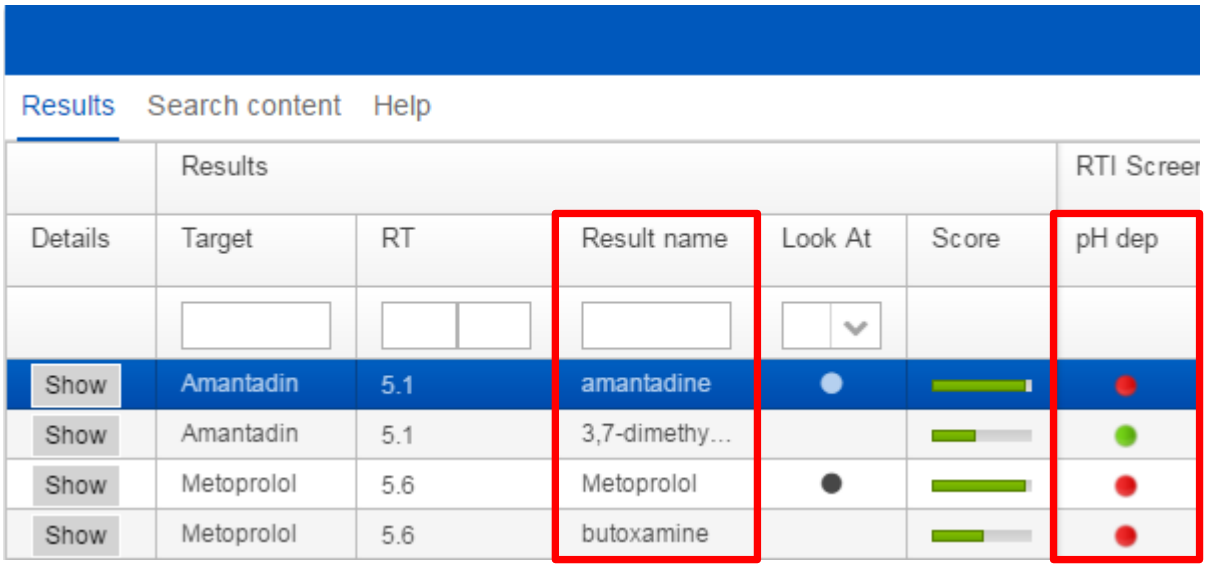
Purple indicated molecules are charged neutral for each pH value, but are in the upper hydrophobicity range of reversed phase columns with logD values > 4.00. The calculated logD value is corrected by adding one logD value (+1).

Blue indicated molecules are charged negative. The calculated logD value is corrected by subtracting one logD value (-1). The correction is not performed, if C18 (polar endcapped) or C18 (polar embedded) columns are selected as stationary phase.

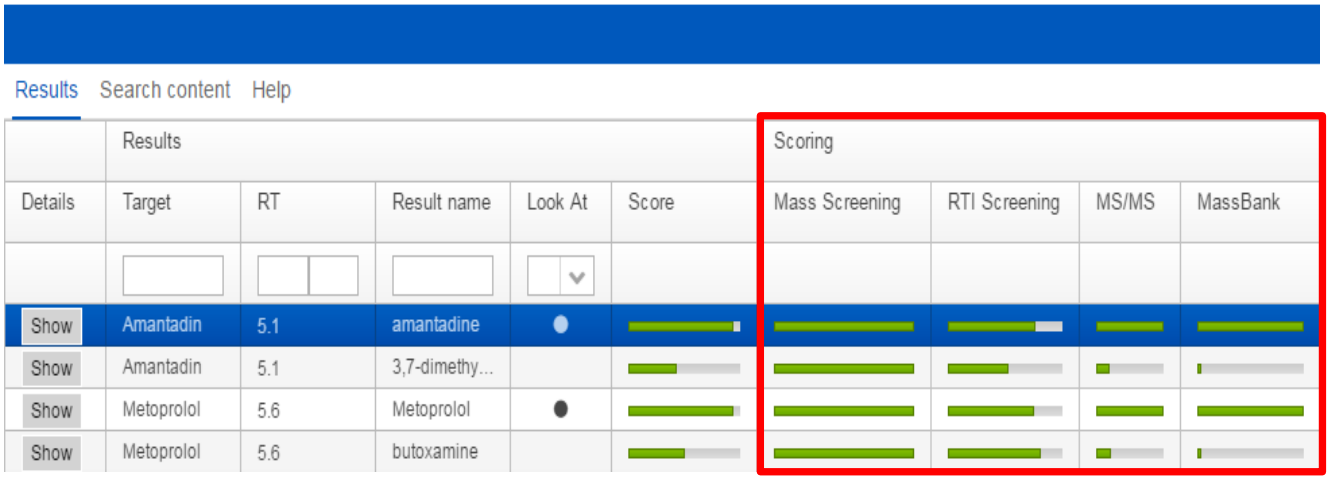
Red indicated molecules are charged positive. The calculated logD value is corrected by subtracting one logD value (-1).

The correction, if necessary, is automatically done by the database.

In the column ‘Result name’ the name of the database hit of the target analyte is shown with its pH dependency color code in column ’pH dep’.



5.2.1.3 columns Scoring: Mass Screening, RTI Screening, MS/MS and MassBank



To evaluate the hits suggested for one target analyte by STOFF-IDENT the scoring bars indicates the likelihood of each function, especially if several hits are suggested for one target analyte.

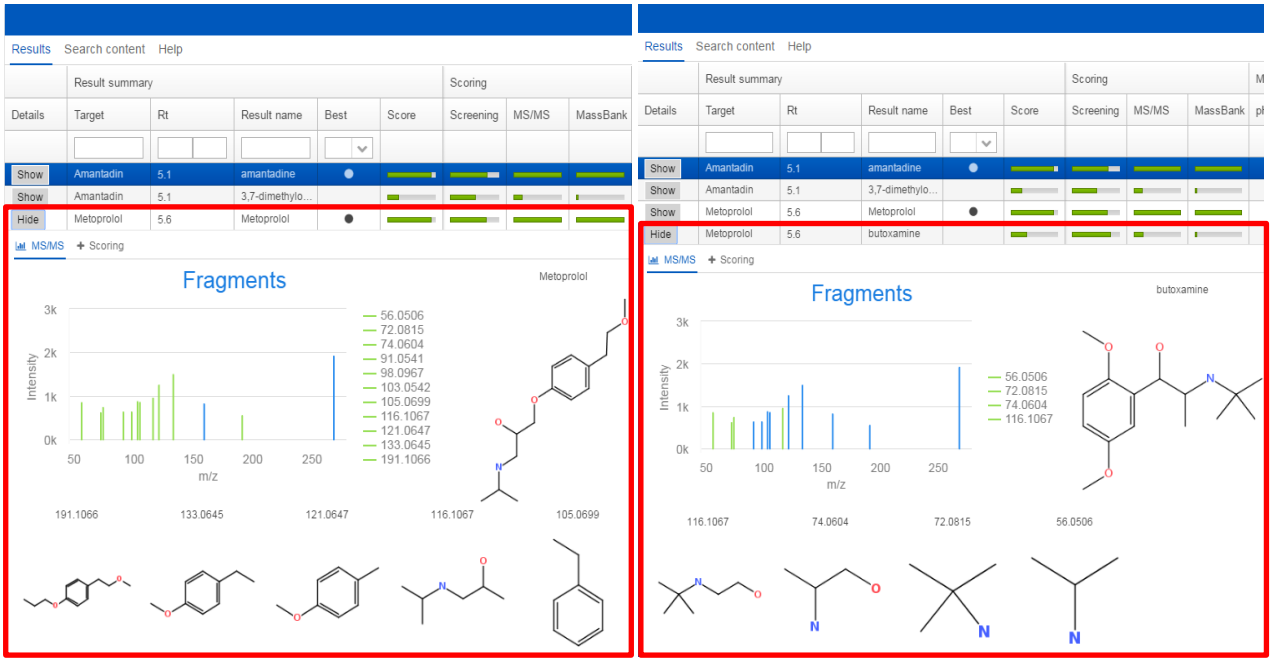
The Mass Screening Score represents the difference between the measured mass and the one listed in the database (column ‘Δ mass’).

The RTI Screening Score represents the difference between the determined logD value via RTI normalization and the one listed in the database (column ‘Δ logD (Adj.-DB’).

A full green bar indicates very low difference while as smaller the bar indicates as higher the difference.

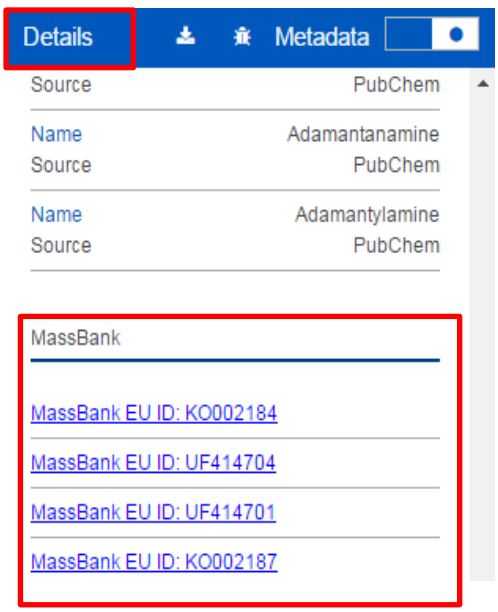
The MS/MS Score describes the differences between the uploaded MS/MS fragments and the ones, which MetFrag predicts in silico. The real MetFrag score is used, which includes some other rules and parameters. A scoring is just possible, if two or more hits are given for one target analyte. Refer to section “additional information” in this part for more details.

By pressing the ‘Show’ Button the comparison of uploaded fragments and in silico fragments of MetFrag is shown. Green indicates a fragment mass match and blue a mismatch of fragment. Below also the matched fragment structures are shown.



The MassBank Score represents, if the database hit has an experimental determined spectra in the MassBank database (Links are shown in the ‘Details’ area).

A full bar means at least one spectra available, an empty bar means no spectra available in MassBank database. No comparison of fragmentation spectra are performed.



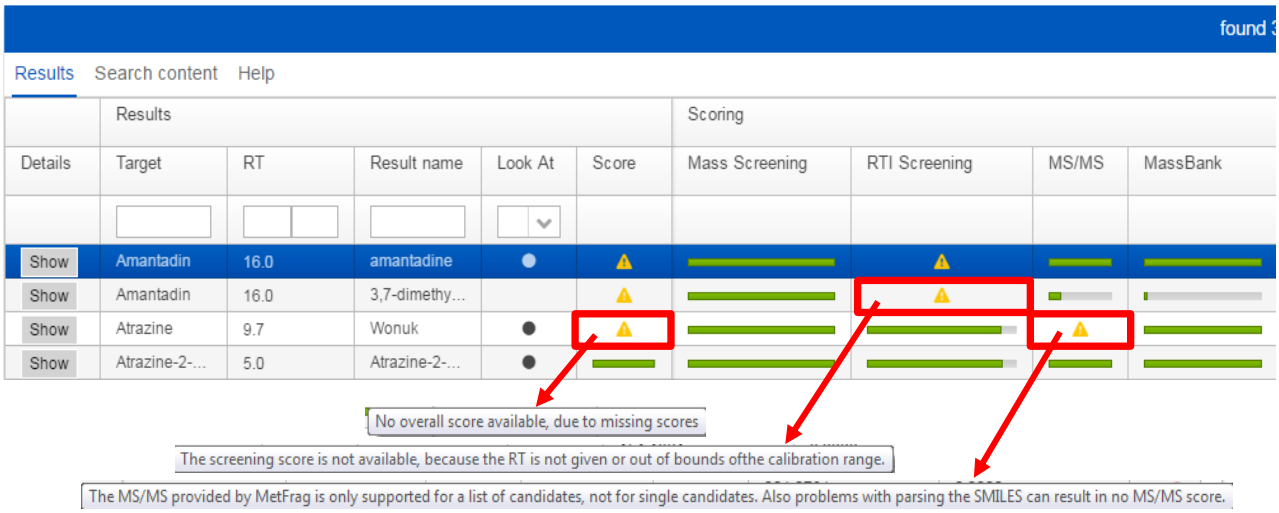
Additional information

Triangles in the Score columns

a) **missing RTI Screening Score: A logD value cannot be determined via RTI, if the RT of the target analyte is outside the calibration area of the RTI calibration molecules**

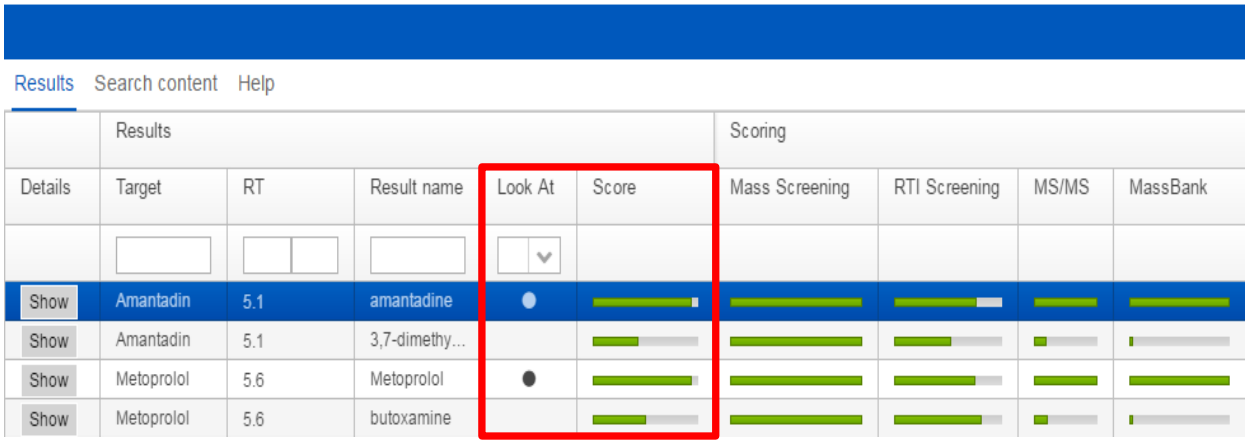
b) **missing MSMS Score: A scoring can only be given, if at least two candidates are found for one target. Or if problems by parsing SMILES code occured**

c) **missing Score: The overall Score is missing, if one of the single Score gives no value**



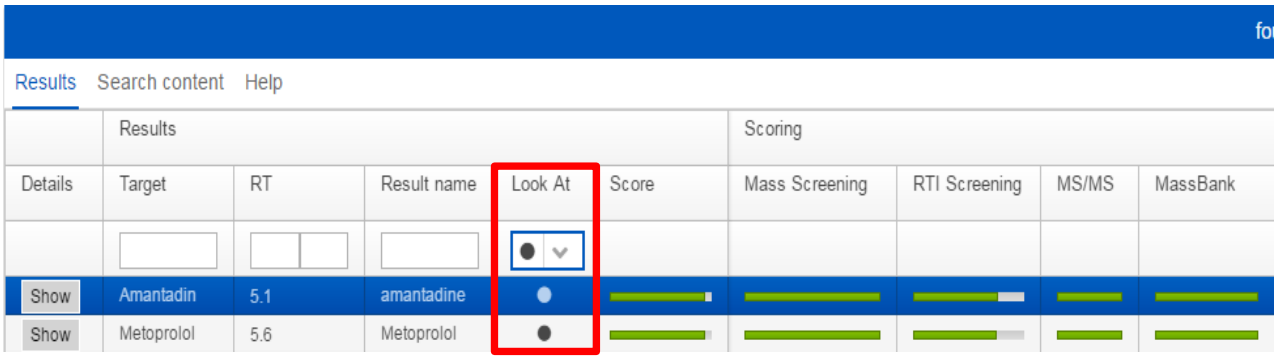
5.2.1.4 columns Look At and Score

If several hits are suggested for one target analyte by STOFF-IDENT, in the column ‘Look At’ and ‘Score’ a summary of the single Scores is shown. The one with the highest overall Score, depending on the weight chosen by the operator, is marked with ‘●’.



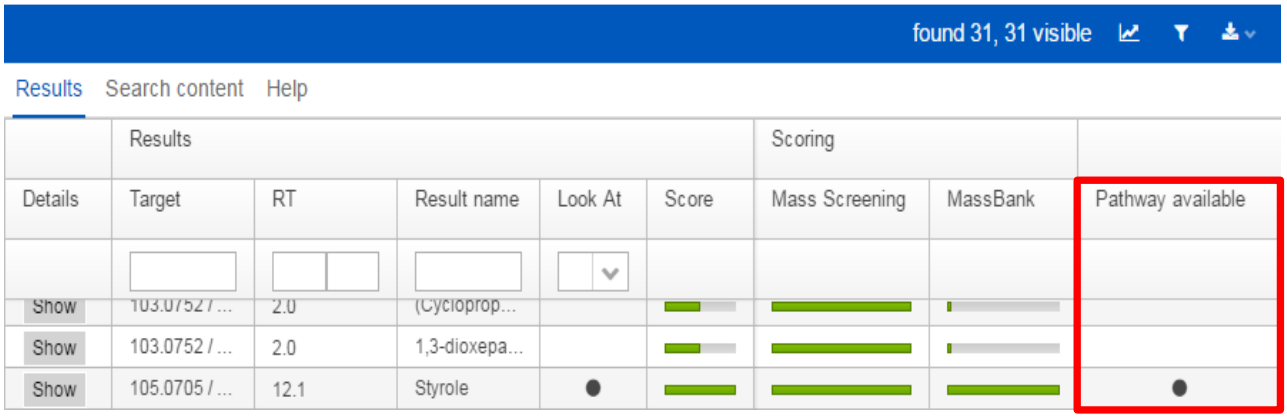
A filtering of the results on all the 'Look At' molecules is easily available by selecting '●' in the 'Look At' column.

The result can be exported separately.

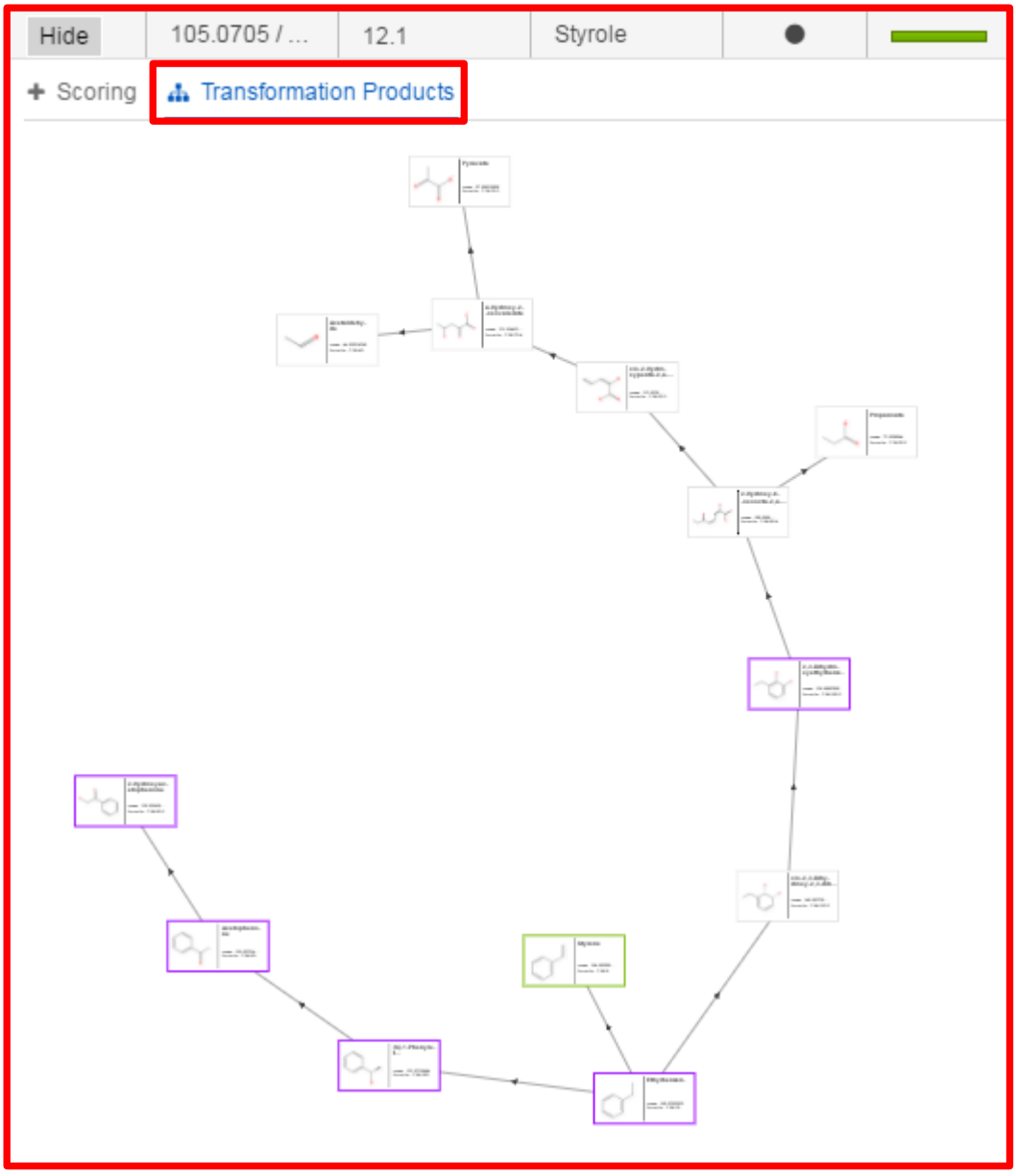


5.2.1.5 column Pathways available

Transformation products are predicted by the EnviPath database using the software tool of EAWAG-BBD. If a pathway exist for one suggested hit in the database, it is marked with '●' in the column 'pathways available'.



By pressing the ‘Show’ Button and using the Tab ‘Transformation products’ the predicted pathway is shown. <u>This tool is still under progress and will be extended soon.</u>



<u>Description of pathway colors:</u>

Green Box: selected database hit

Blue Box: parent compound

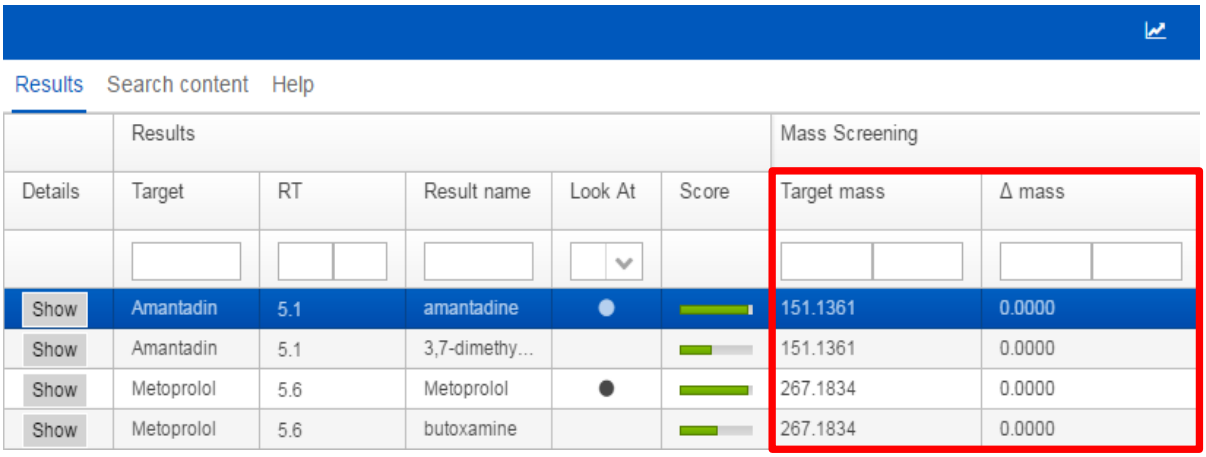
Purple Box: transformation product found in the uploaded datafile

Grey Box: transformation product not found in the uploaded datafile

5.2.1.6 columns Target mass and Δ mass

In the column ‘Target mass’ the monoisotopic mass of the uploaded target is shown.

In the column ‘Δ mass’ the difference between the mass of the database hit and the Target mass is calculated and shown automatically.



5.2.1.7 columns RTI, logD DB, logD RTI, Δ logD (RTI-DB), Adj. logD RTI und Δ logD (Adj.-DB)

The column RTI contains the RTI values normalized via the calibration standards of the target analyte.

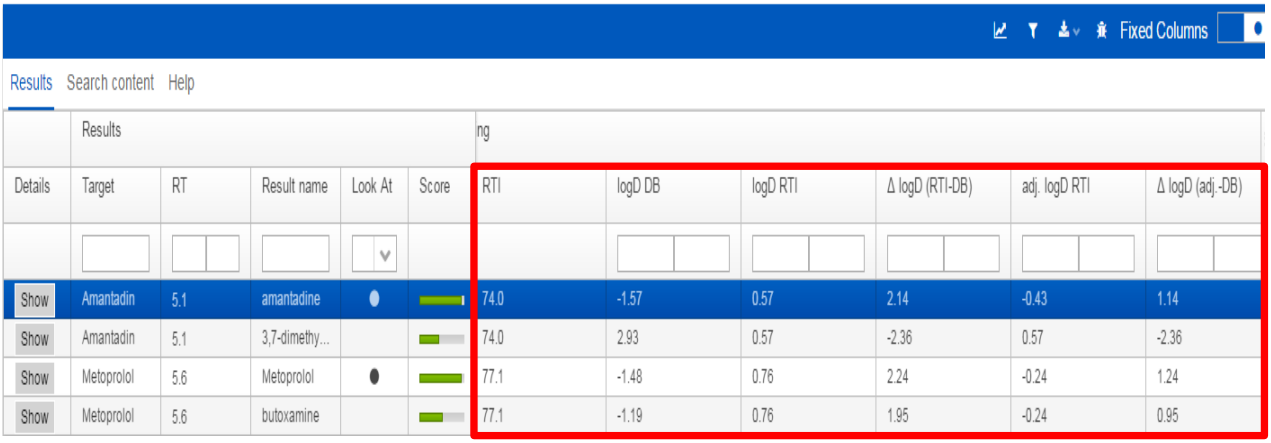
The column logD DB contains the logD value of the entered pH value of the substance, listed in the database, which was found as a hit via STOFF-IDENT.

The column logD RTI contains the logD value of the target analyte, calculated via RTI.

In the column Δ logD (RTI-DB) the difference between the logD value calculated via RTI and the one that is listed in the database is calculated and shown automatically.

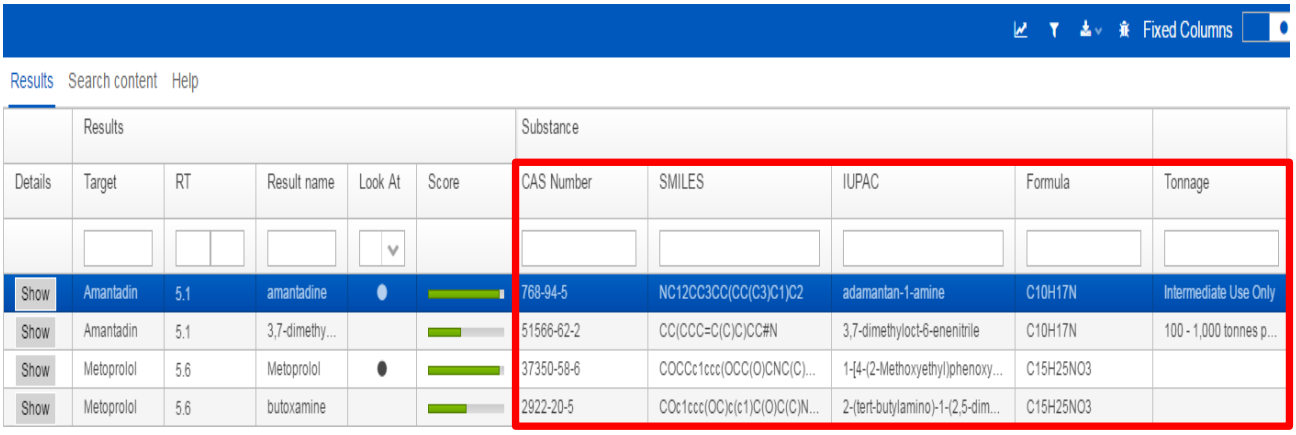
In the column Adj. logD RTI an automatic correction like written in 5.2.1.2 is done, if necessary.

In column Δ logD (Adj.-DB) the difference between the adjusted logD value and the one listed in the database is calculated and shown.

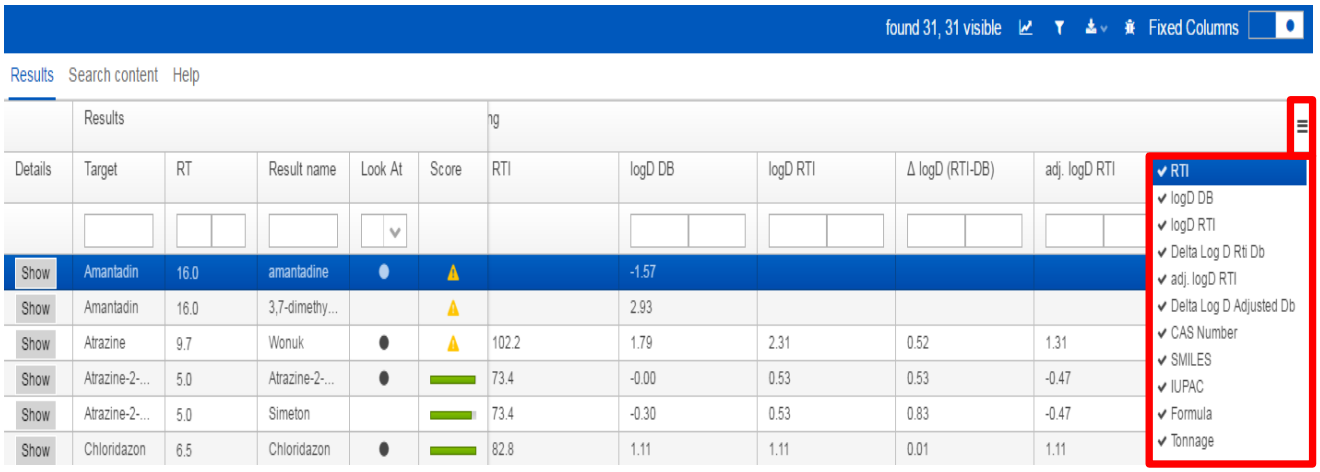


5.2.1.8 CAS number, SMILES, IUPAC, Formula and Tonnage

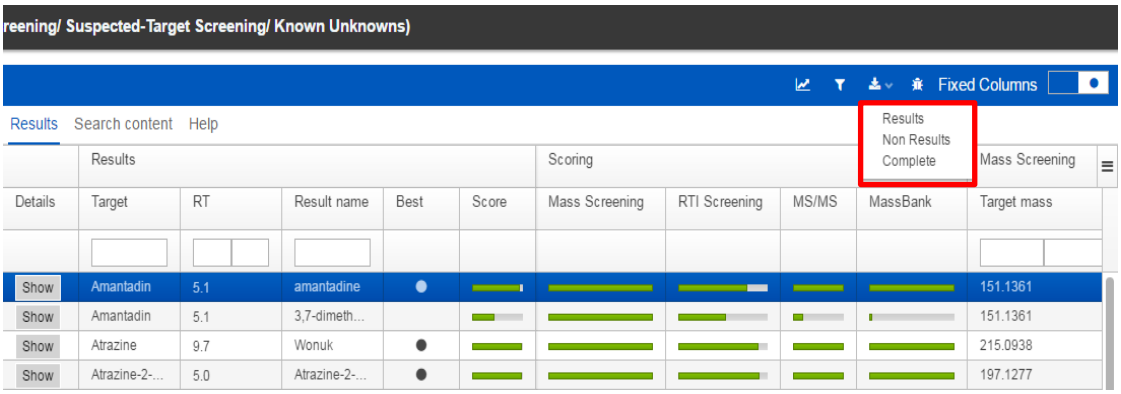
In the following columns CAS number, SMILES, IUPAC, Formula and Tonnage all in the database listed information for the found hits are shown

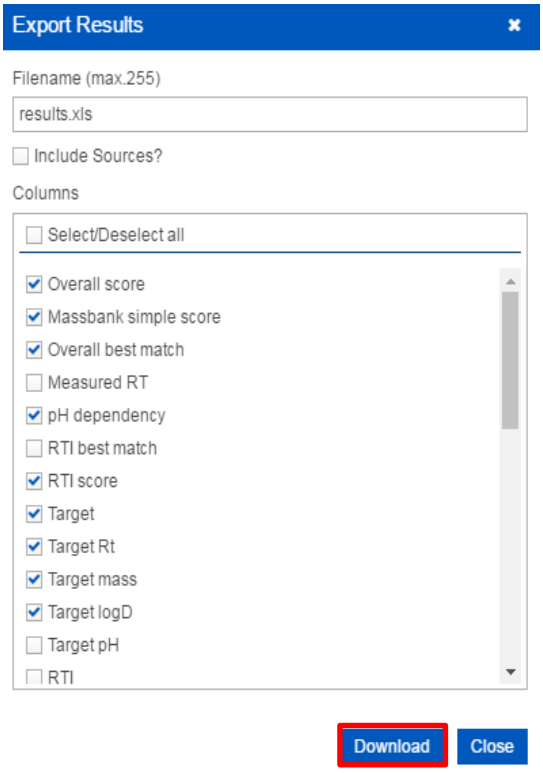


You have the possibility to check or uncheck columns of interest by using the drop down menu

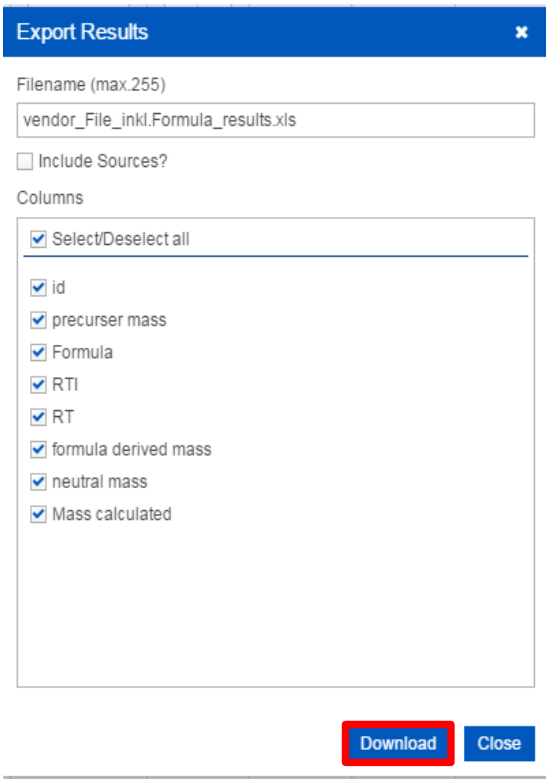


Via the 'Download' button and choosing 'Results' the results table can be exported in excel with all favored parameters.

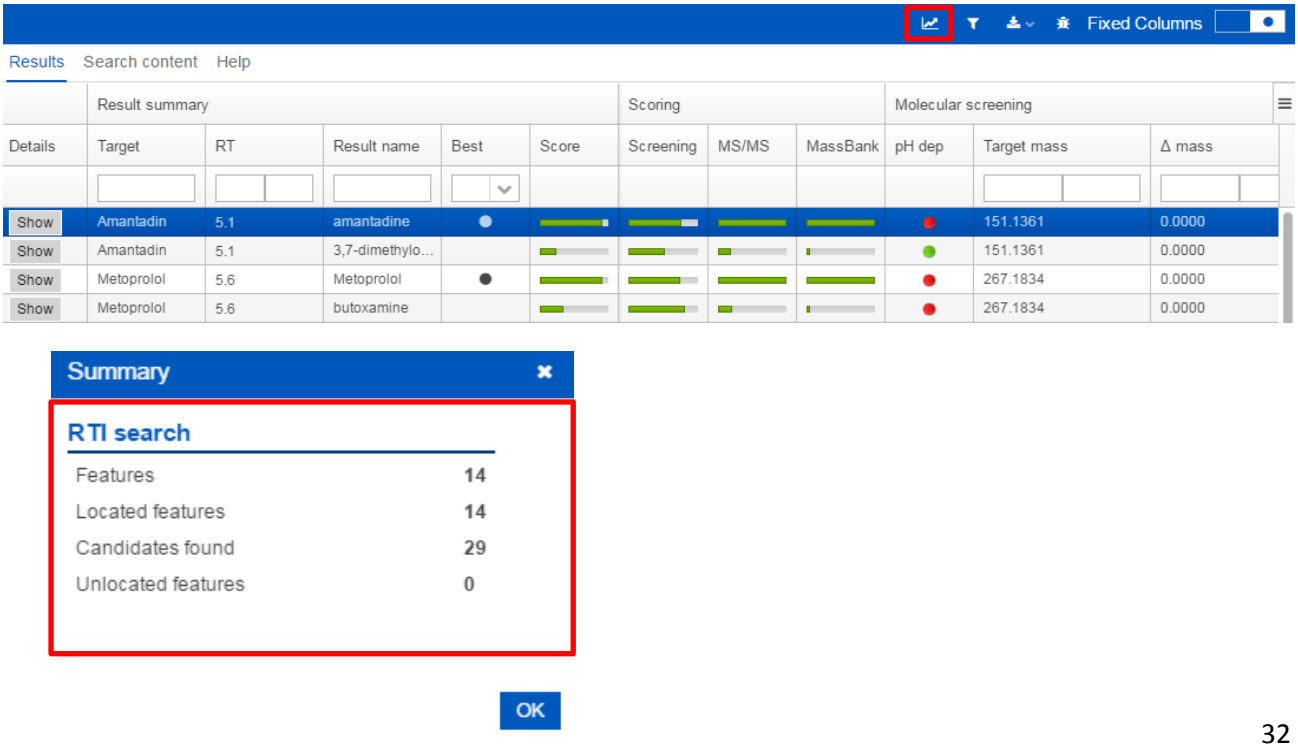




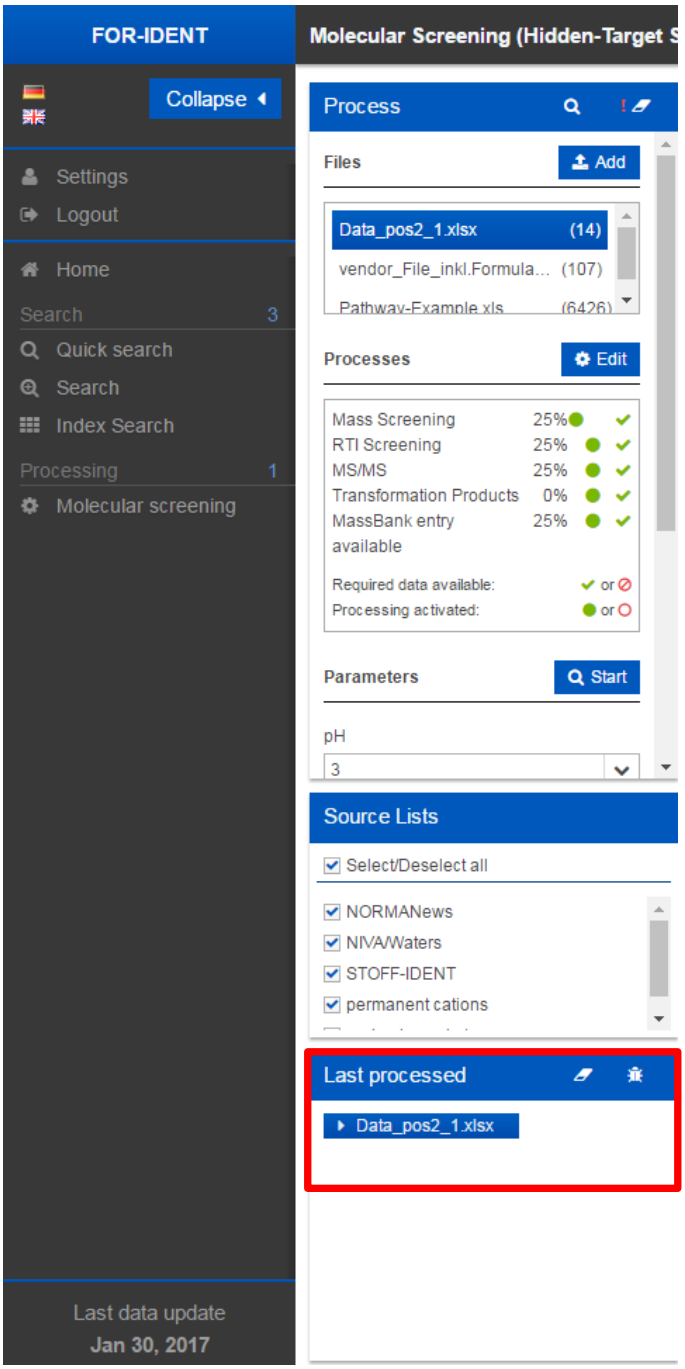
Besides exporting the FOR-IDENT results there is also the option to download the 'non results' or the complete set. By pressing the 'Download' button and choosing 'non results' the list of targets with no hit in STOFF-IDENT can be exported in an extra excel file with all favored parameters, e.g. RTI.



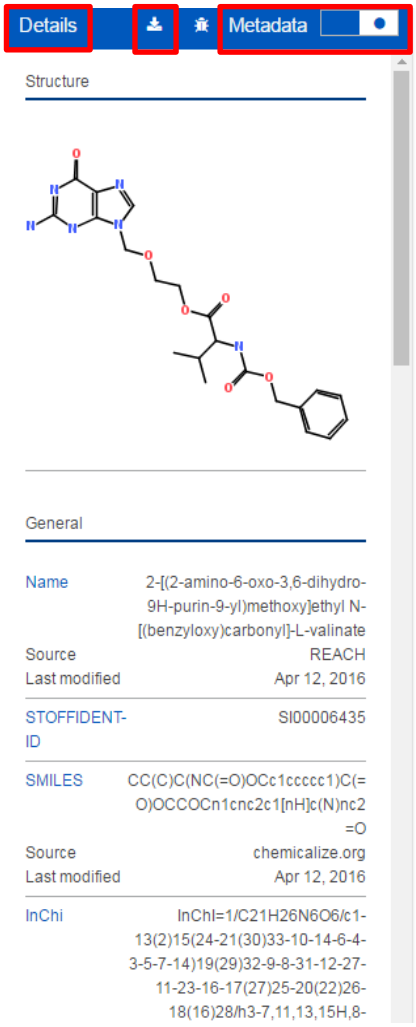
By pressing the 'Show summary' button a summary window appears and shows the values of uploaded features, database hits and non database hits

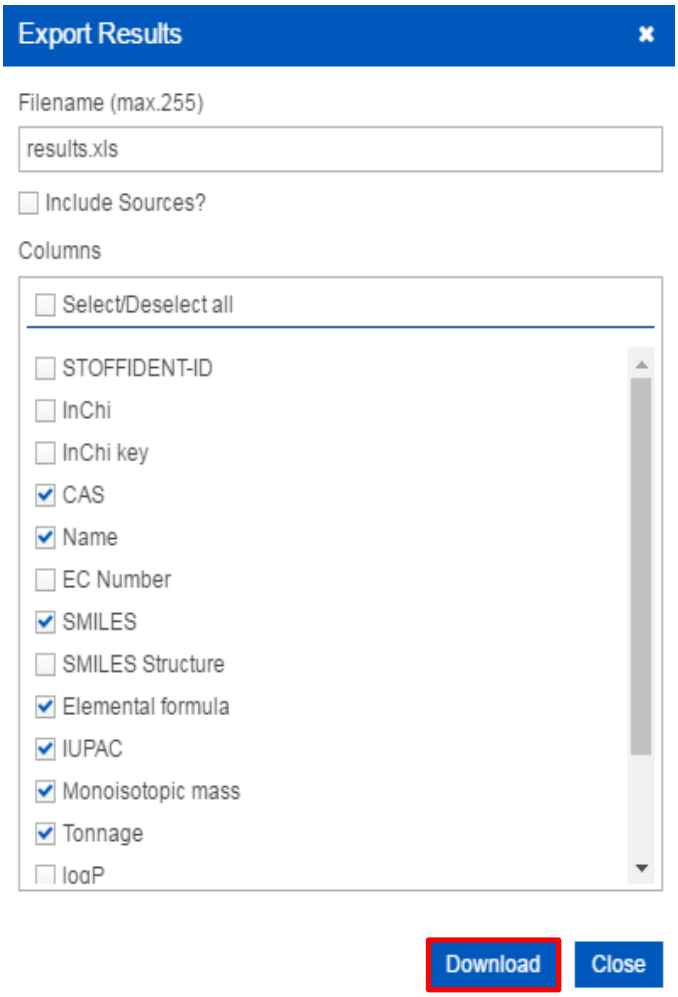


In the ‘Last processed’ area all performed searches were stored and can be reloaded again by ‘doubleclick’.

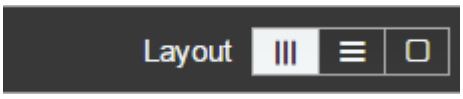


In the 'Details' tab you can find specific information about the analyte. All information can be downloaded as an excel-file. Besides you have the choice to see the Metadata or not.





You can change the Layout with three different options



In the Details at the web search part, you can directly search the specific compound via internet or you go to the External Links to get more information of the compound

If you find errors in the parameters given, e.g. wrong sum formula, SMILES code, etc. you can use the 'report record as faulty' button to send it directly to the software developer team by putting your comments of the fault into the comment field.