

IsoX: Iso^{topolog} data eXtraction from Orbitrap RAW files

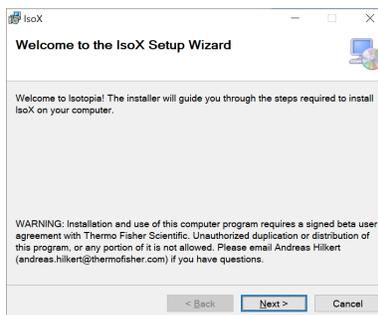
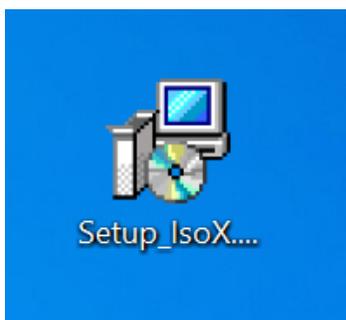
(Version 2022; 03-MAY-2022)

For questions, bug reports, feature requests, please contact: Caj Neubauer (123caj@gmail.com)

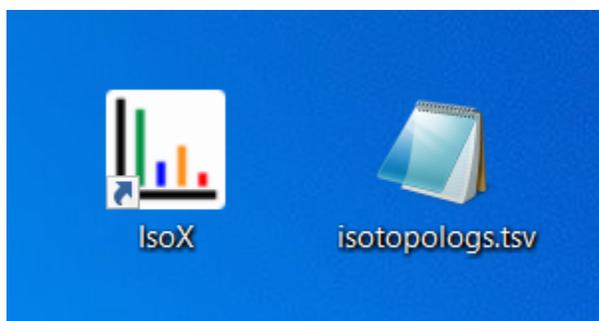
For beta testing licenses, please contact: Andreas Hilkert (andreas.hilkert@thermofisher.com)

This software runs on **Microsoft Windows** (utilizing the **.Net Framework**).

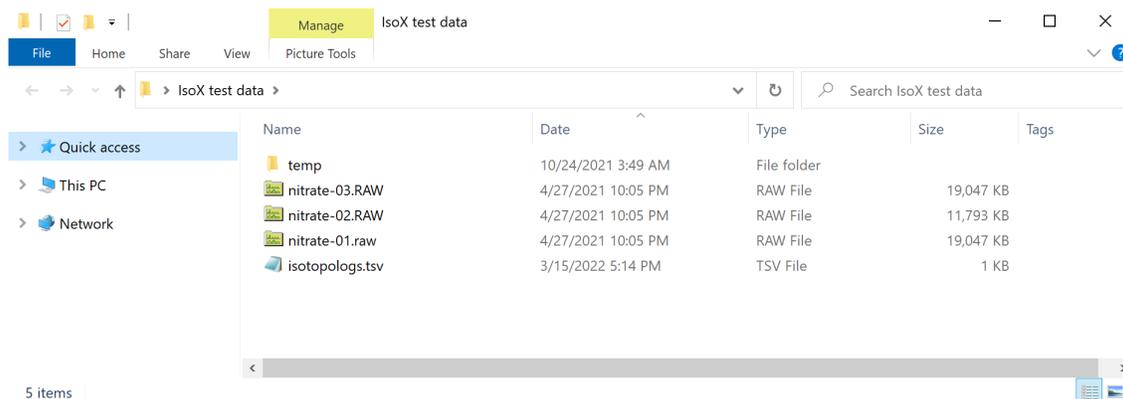
- To install **IsoX**, double click the **IsoX** installer (`Setup_IsoX.msi`) and follow the installation prompts.



- After the installation two new files should appear on the desktop.



- Copy `isotopologs.tsv` into the directory that contains RAW files you want to process.



- Open and edit `isotopologs.tsv` in **Notepad** (preferred). If you edit the file in **Excel**, make sure to save it again in the tab-delimited file format with the file ending `.tsv` (eg, not as `isotopologs.tsv.txt`)

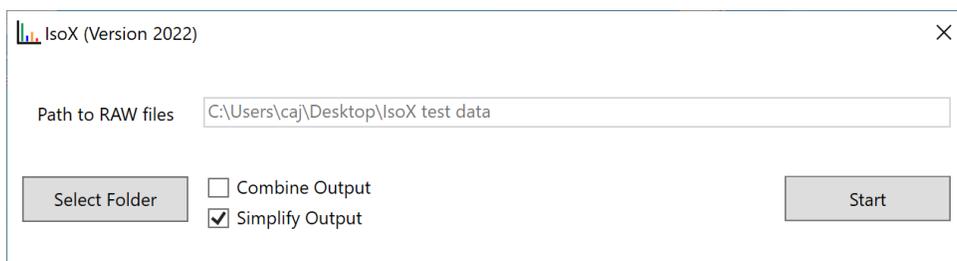
The input file has five columns:

- 1) **Compound**: name of the analyte
- 2) **Isotopolog**: name of the isotopolog
- 3) **m/z**: mass-to-charge ratio (m/z) of the isotopolog
- 4) **Tolerance [mmu]**: +/- tolerance for the mass-to-charge ratio (in mmu)
- 5) **z**: the charge state of the ion (typically 1). This value is used when calculating the number of ions.

#Compound	Isotopolog	m/z	Tolerance [mmu]	z
NO3-	M0	61.98670	0.1	1
NO3-	15N	62.98360	0.1	1
NO3-	17O	62.99080	0.1	1
NO3-	18O	63.99090	0.1	1
NO3-	15N18O	64.98789	0.1	1
NO3-	17O18O	64.99520	0.1	1
NO3-	18O18O	65.99518	0.1	1
HSO4-	M0	96.96010	0.1	1
HSO4-	33S	97.95934	0.1	1
HSO4-	17O	97.96382	0.1	1
HSO4-	34S	98.95589	0.1	1
HSO4-	18O	98.96349	0.1	1
HSO4-	34S17O	99.95976	0.1	1
HSO4-	33S18O	99.96362	0.1	1
HSO4-	36S	100.9549	0.1	1
HSO4-	18O34S	100.9600	0.1	1

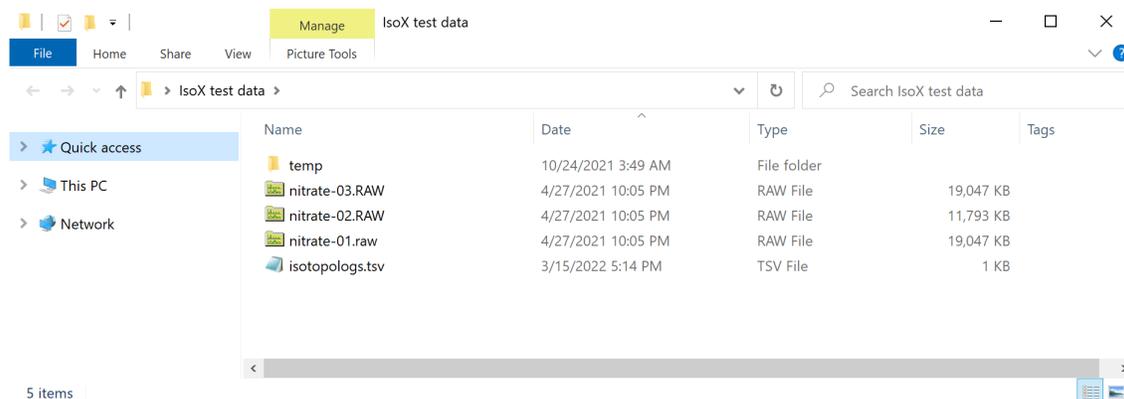
IsoX will look for all isotopologs defined in the input file (in the example above, NO_3^- and HSO_4^- ions). However, in practice for many data acquisitions the mass range is restricted by the isolation window of the quadrupole mass filter. It is therefore feasible to use 'isotopologs.tsv' as a simple database for analytes of interest in different experiments. (Ideally data is collected by using a lockmass for the target analyte, so that observed m/z are not varying between data collected on different days. Then the input file `isotopologs.tsv` does not have to be manually edited.)

- Open **IsoX** by double clicking the icon on the desktop (the program, as well as an original version of `isotopologs.tsv` can also be found in the program folder: `C:\Program Files\IsoX`).



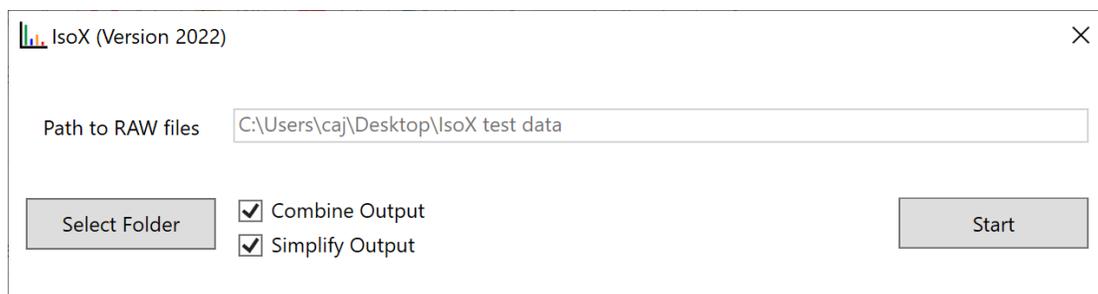
- Click 'Select Folder' and select a RAW file in the folder that contains your data and `isotopologs.tsv`.

In this example, the folder `IsoX test data` contains 3 RAW files with MS data for nitrate (NO_3^-) and `isotopologs.tsv` for quantifying its isotopologs.

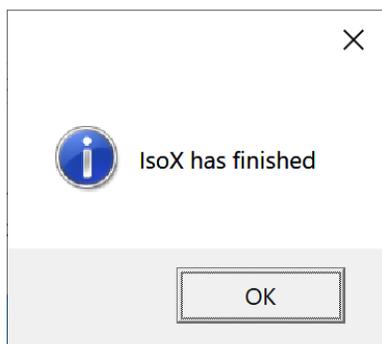


IsoX offers two options:

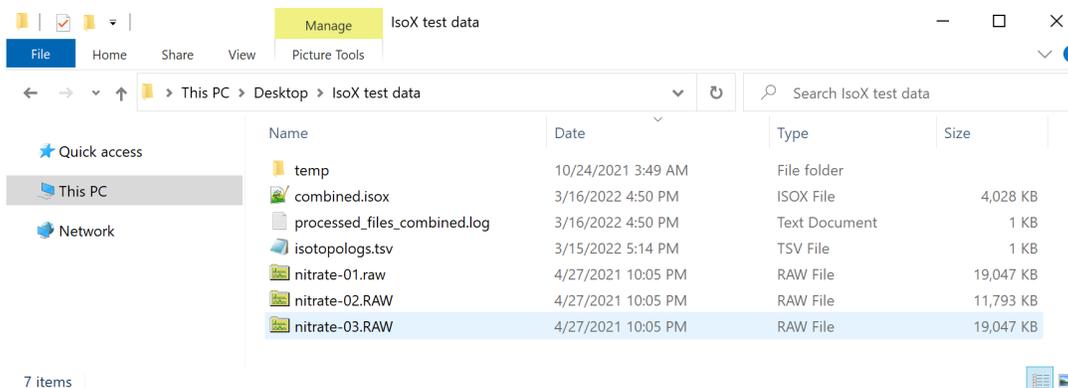
- 1) **Combine Output:** Combines data from all RAW files into one output `.isox` file (`combined.isox`). If unchecked, **IsoX** creates one `.isox` output file per RAW file (`nitrate-01.isox`, `nitrate-02.isox`, etc.)
- 2) **Simplify Output:** If checked, **IsoX** reports only basic data that is needed for isotope ratio analysis in the `.isox` file. If unchecked, **IsoX** reports additional columns that are not strictly necessary for isotope ratio analysis, but may be useful for some advanced users (eg, for data quality control or to remove outliers).



- When clicking 'Start' **IsoX** will extract information relevant to isotopic analysis from the RAW files. When the program is finished (usually after a few seconds) it shows the following prompt.



The data folder now contains two new files, `combined.isoX` contains the extracted information and `processed_files_combined.log` provides a record of which RAW files have been processed so far (this is useful in case more RAW files are added to the folder at a later time). The `.log` files also keeps track which version of IsoX was used.



The file `combined.isoX` also is a simple tab-delimited file and looks like this when opening in *Notepad++* (it can be opened also in *Excel*, *Notepad*, etc):

1	filename	scan.no	time.min	compound	isotopolog	ions.incremental	tic	it.ms
2	nitrate-01	1	0	NO3-	M0	74439.04302556	841165500	0.657
3	nitrate-01	1	0	NO3-	15N	398.91818322	841165500	0.657
4	nitrate-01	1	0	NO3-	170	91.71172582	841165500	0.657
5	nitrate-01	1	0	NO3-	180	654.04592494	841165500	0.657
6	nitrate-01	2	0.001	NO3-	15N	360.27838347	893524160	0.677
7	nitrate-01	2	0.001	NO3-	180	723.81580842	893524160	0.677
8	nitrate-01	3	0.002	NO3-	M0	77763.10759807	846024510	0.69
9	nitrate-01	3	0.002	NO3-	15N	390.95298857	846024510	0.69
10	nitrate-01	3	0.002	NO3-	180	684.55697767	846024510	0.69
11	nitrate-01	4	0.002	NO3-	M0	76695.73422788	872473600	0.649
12	nitrate-01	4	0.002	NO3-	15N	381.16973742	872473600	0.649
13	nitrate-01	4	0.002	NO3-	170	103.45469564	872473600	0.649
14	nitrate-01	4	0.002	NO3-	180	748.76647904	872473600	0.649
15	nitrate-01	5	0.003	NO3-	M0	79319.17785306	878924860	0.688
16	nitrate-01	5	0.003	NO3-	15N	409.26447993	878924860	0.688
17	nitrate-01	5	0.003	NO3-	170	103.52314164	878924860	0.688
18	nitrate-01	5	0.003	NO3-	180	723.09691251	878924860	0.688
19	nitrate-01	6	0.003	NO3-	M0	73994.58657285	858651650	0.667
20	nitrate-01	6	0.003	NO3-	15N	352.56652361	858651650	0.667
21	nitrate-01	6	0.003	NO3-	180	736.82299052	858651650	0.667
22	nitrate-01	7	0.004	NO3-	15N	341.36126495	848477310	0.661
23	nitrate-01	7	0.004	NO3-	180	684.78426567	848477310	0.661
24	nitrate-01	8	0.005	NO3-	M0	78934.85433692	925770880	0.69
25	nitrate-01	8	0.005	NO3-	15N	390.07573621	925770880	0.69
26	nitrate-01	8	0.005	NO3-	170	127.99291999	925770880	0.69

The output is provided in a '[tidy data](#)' format, ideally suited for import into data science software such as *R* or *Python*.

The basic output contains 8 column:

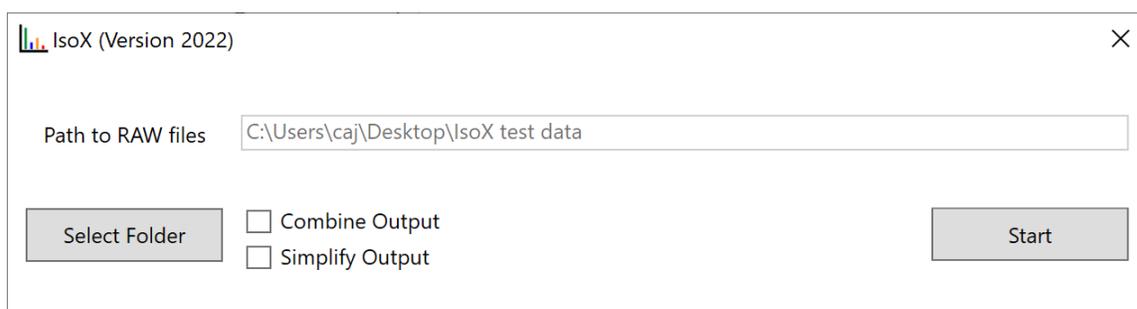
- 1) **filename**: name of the processed RAW file
- 2) **scan.no**: number of the processed scan
- 3) **time.min**: time in minutes

- 4) **compound**: name of the compound as defined in `isotopologs.tsv`
- 5) **isotopolog**: name of the isotopolog as defined in `isotopologs.tsv`
- 6) **ions.incremental**: a calculated number of ions for the signal. This value is calculated using the following approximation:

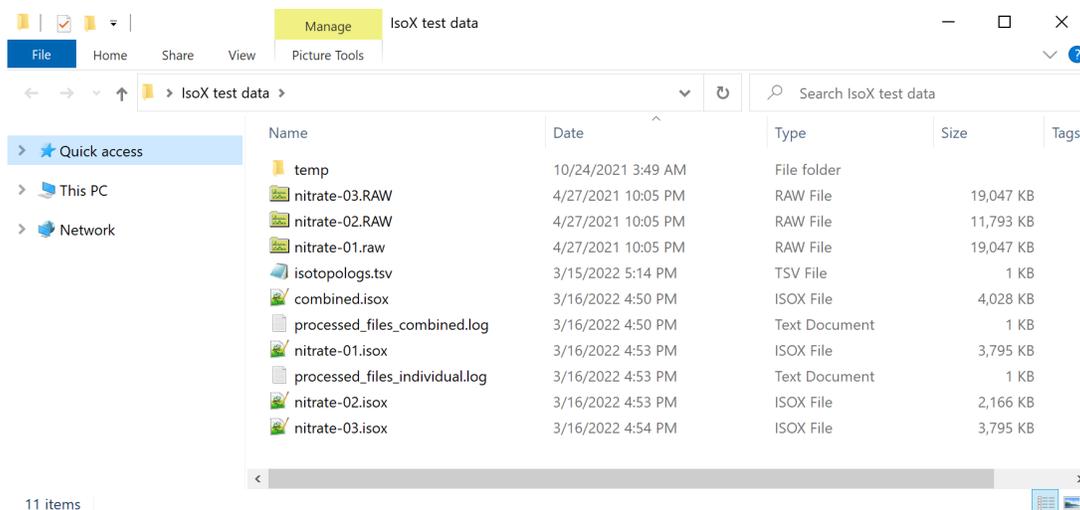
$$\text{ions.incremental} = (\text{Intensity}/\text{Peak.Noise}) * 3 * \text{sqrt}(240000/\text{FT.Resolution}) * \text{sqrt}(\text{Microscans})$$

- 7) **tic**: total ion current
- 8) **it.ms**: injection time in milliseconds (`tic*it.ms` is an estimate of the ions per scan. This is often a useful filter to remove scans that are outliers due to an unusually small or large number of ions entering the Orbitrap.)

For illustration, we here rerun the program in the same data folder with the following settings:



Four new files have been created. One new log file (`processed_files_individual.log`), and three new `.isox` files (one for each RAW file):



Since the option ‘Simplify Output’ has been deselected, new `.isox` files contains additional data columns. These data columns can be explored by advanced users to better understand aspects of the experiment that may affect isotope ratios.

- **intensity**: the measured intensity of the isotopolog signal
- **resolution**: the mass resolution setting of the Orbitrap

- **peakResolution:** the mass resolution of the isotopolog signal
- **peakNoise:** the noise of the isotopolog signal
- **mzMeasured:** the noise of the isotopolog signal
- **basePeakIntensity:** the intensity of the most intense signal in the scan
- **rawOvFtT:** TIC estimation done with the Orbitrap
- **intensCompFactor:** intensity compensation factor
- **agc:** AGC setting during acquisition
- **agcTarget:** AGC target used
- **microscans:** number of micro scans
- **numberLockmassesFound:** number of lock masses found in scan
- **analyzerTemperature:** temperature of the Orbitrap
- **baseline:** Baselines

1	filename	scan.no	time.min	compound	isotopolog	ions.incremental	tic	it.ms	intensity	resolution	peakResolution	peakNoise	mzMeasured	basePeakIntensity	rawOvFtT
2	nitrate-01 1	0	0	NO3-	MO	74439.04302556	841165500	0.657	806705408	15000	32002	130045.531	61.98666	806705408	650148.3
3	nitrate-01 1	0	0	NO3-	15N	398.91818322	841165500	0.657	4320667	15000	31102	129971.523	62.98362	806705408	650148.3
4	nitrate-01 1	0	0	NO3-	17O	91.71172582	841165500	0.657	993322	15000	26300	129970.992	62.99083	806705408	650148.3
5	nitrate-01 1	0	0	NO3-	18O	654.04592494	841165500	0.657	7079870	15000	30702	129896.75	63.99089	806705408	650148.3
6	nitrate-01 2	0.001	0.001	NO3-	15N	360.27838347	893524160	0.677	3916856.75	15000	31202	130461.008	62.98362	857524608	697854.6
7	nitrate-01 2	0.001	0.001	NO3-	18O	723.81580842	893524160	0.677	7898247	15000	30802	130943.484	63.99089	857524608	697854.6
8	nitrate-01 3	0.002	0.002	NO3-	MO	77763.10759807	846024510	0.69	811200640	15000	32102	125180.281	61.98666	811200640	679728.5
9	nitrate-01 3	0.002	0.002	NO3-	15N	390.95298857	846024510	0.69	4049628.5	15000	31102	124300.219	62.98363	811200640	679728.5
10	nitrate-01 3	0.002	0.002	NO3-	18O	684.55697767	846024510	0.69	7040160	15000	30702	123411.086	63.99088	811200640	679728.5
11	nitrate-01 4	0.002	0.002	NO3-	MO	76695.73422788	872473600	0.649	827335360	15000	32102	129446.891	61.98662	827335360	664396.3
12	nitrate-01 4	0.002	0.002	NO3-	15N	381.16973742	872473600	0.649	4130089.25	15000	31202	130023.625	62.98358	827335360	664396.3
13	nitrate-01 4	0.002	0.002	NO3-	17O	103.45469564	872473600	0.649	1120999.25	15000	27000	130027.844	62.99087	827335360	664396.3
14	nitrate-01 4	0.002	0.002	NO3-	18O	748.76647904	872473600	0.649	8149471.5	15000	30902	130606.352	63.99089	827335360	664396.3
15	nitrate-01 5	0.003	0.003	NO3-	MO	79319.17785306	878924860	0.688	842968384	15000	32102	127530.578	61.98665	842968384	698904.1

The `.isoX` files are best viewed and analyzed in a data science software such as **R** or **Python**.

A note of caution:

Please be careful when rerunning **IsoX** in the same data folder. RAW files that have previously been processed (recorded in the `.log` file) will not be re-processed. This enables efficient processing of long sample queues with many RAW files.

This default can cause inconsistencies when the input file `isotopologs.tsv` has been modified in the meantime. Other inconsistencies can be caused when **IsoX** is re-run with additional RAW files but different settings for the 'Simplify Output' option are used. In such a scenario, some data will have basic output columns only while others contain extra columns.

To avoid inconsistencies it is good practice to delete old output files created by **IsoX** and rerun the program once again with all data and the preferred settings in **IsoX** and `isotopologs.tsv`.

Why are sometimes several rows reported per scan & isotopolog?

IsoX extracts data from RAW files for all signals that occur within the m/z range specified in `isotopologs.tsv` (m/z range = $m/z \pm \text{Tolerance [mmu]}$). If there are multiple signals in the m/z range **IsoX** will simply report multiple rows. All data are reported, since there could be an artifact, contaminant or issue with the settings used in `isotopologs.tsv`. Looking at the respective scan in **XCalibur QualBrowser** or **Freestyle** will be helpful to identify what is going on.

Often the signal of interest is the most abundant isotopolog. Appropriate filters can be readily applied in data science software. For example in **R** a command like the following can be used:

```
df.filtered <- df %>% group_by(filename, compound, isotopolog) %>%  
filter(ions.incremental = max(ions.incremental))
```

Notepad++

This text viewer is particularly useful to view tab delimited files on Windows PCs. It can be downloaded for free (<https://notepad-plus-plus.org/downloads/>). To visualize all characters use View-> Show Symbols -> Show All Characters. To align tab-separated files neatly with columns, consider adding the plugin `Elastic Tabstops` via the menu option Plugins -> Plugins Admin. **Notepad++** can then be set as the default program to open files with the `.isox` extension.