

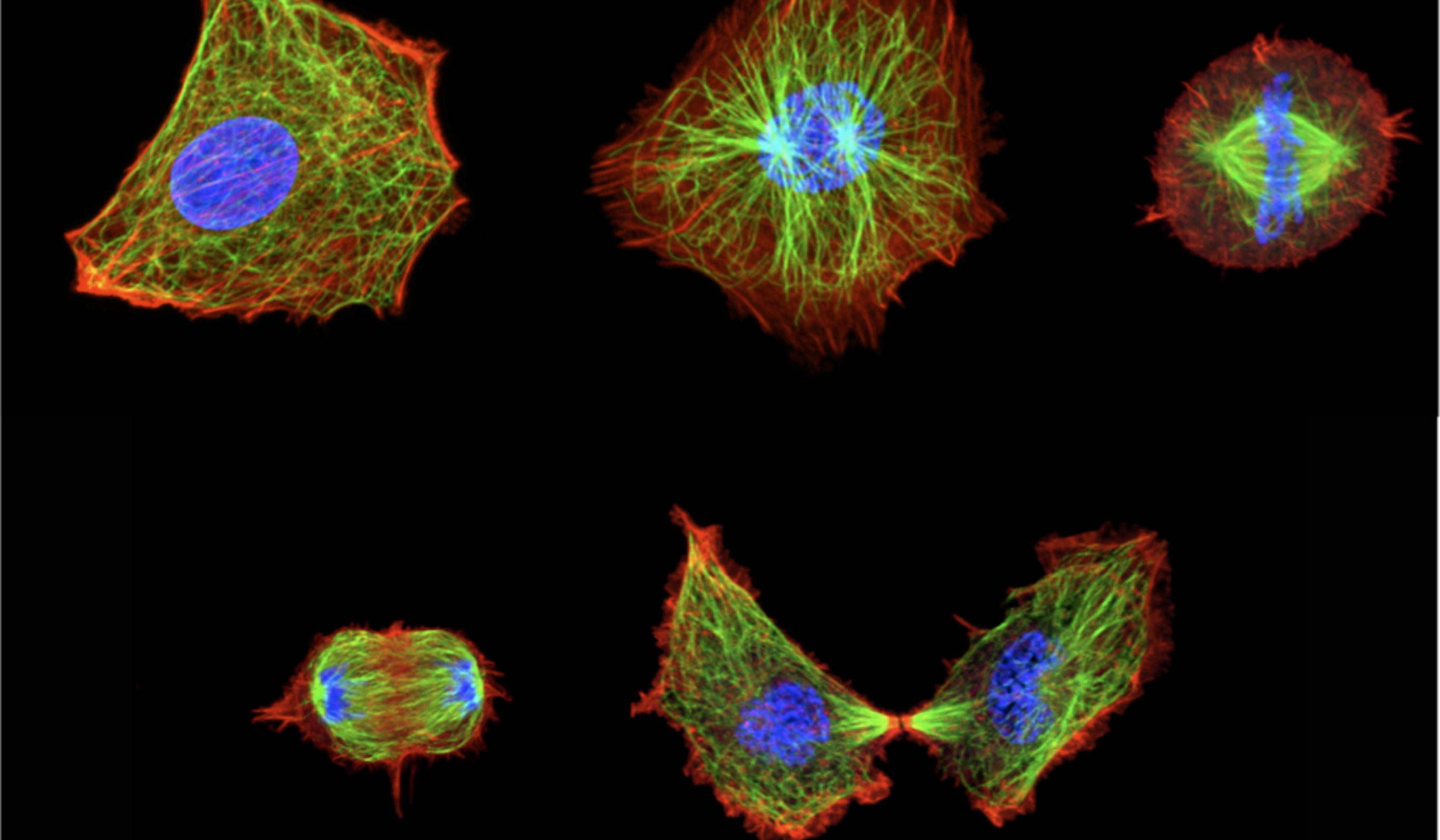
# Introduction to Proteomic Data Analysis

---

University College Dublin, Conway Institute

Matthias Wilm, MD PhD,  
Conway Institute, UCD,  
Belfield, Dublin 4

<https://conwaymserver.ucd.ie/guides/proteomics/index.html>

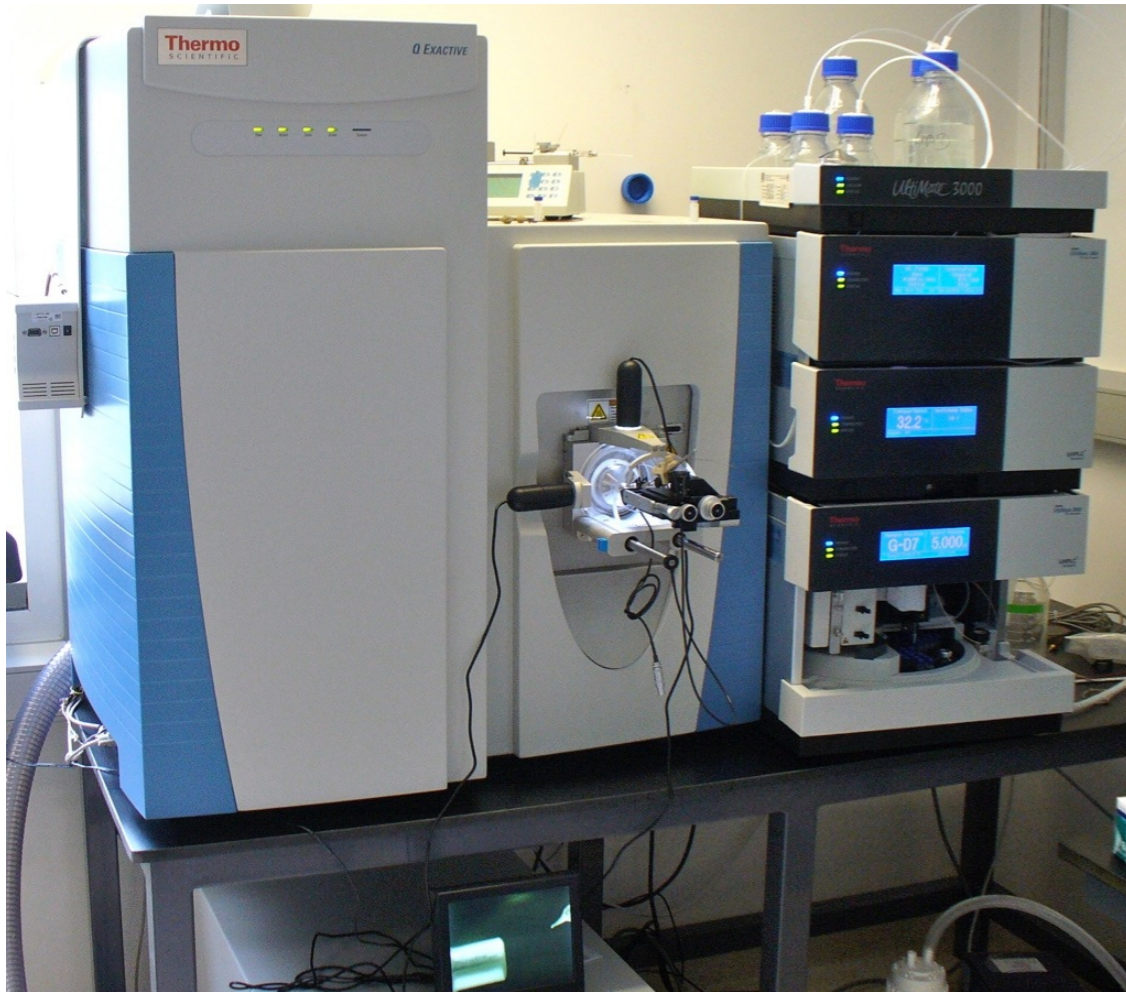


Bio-Molecular  
Research

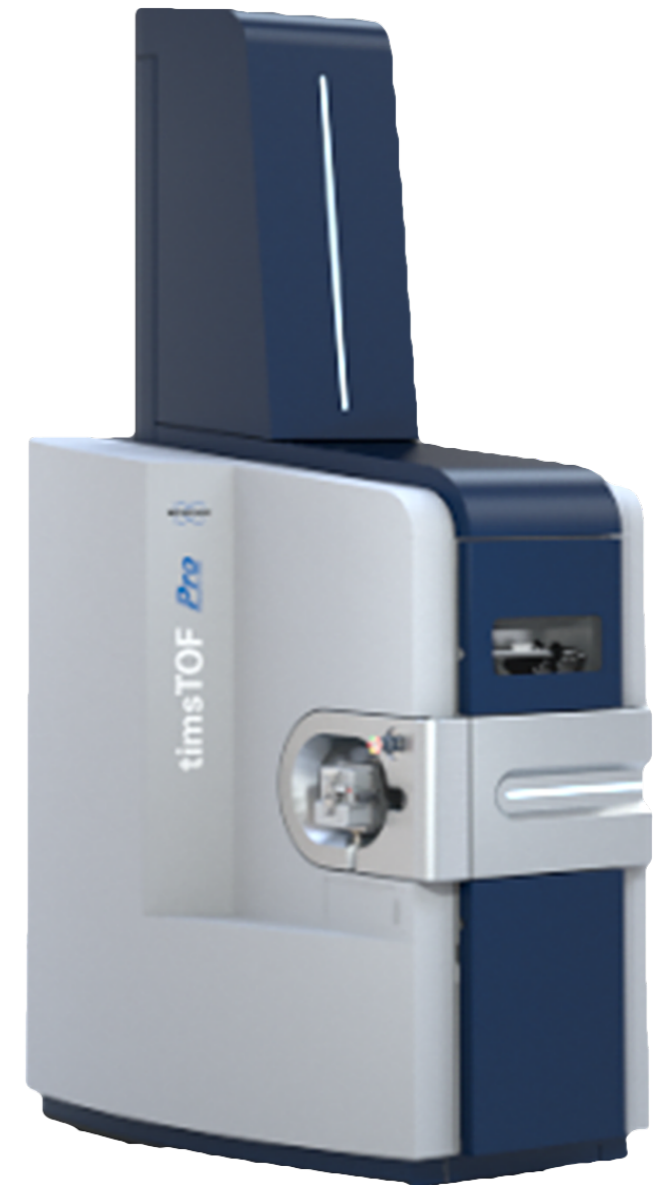
Understanding of complex  
biological processes

# Mass Spectrometry based Protein Identification & Quantification

---



Orbitrap Tandem MS Machine



**t**rapped **i**on **m**obility **s**pectrometry -  
**T**ime **o**f **F**light (timsTOF)  
Mass Spectrometer



# Protein Analysis

---

## MaxQuant



- Protein Identification
- Quantification

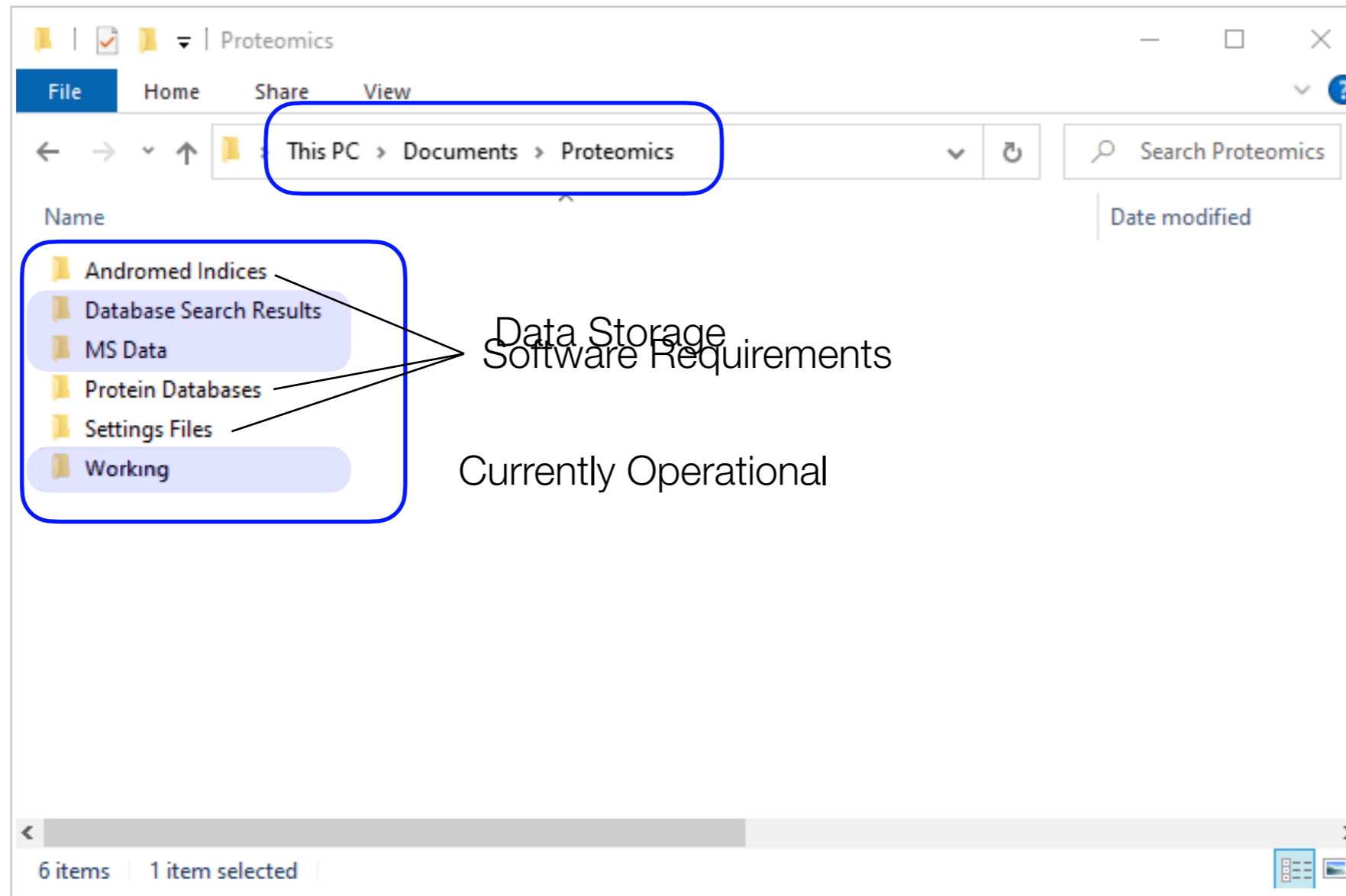
## Perseus

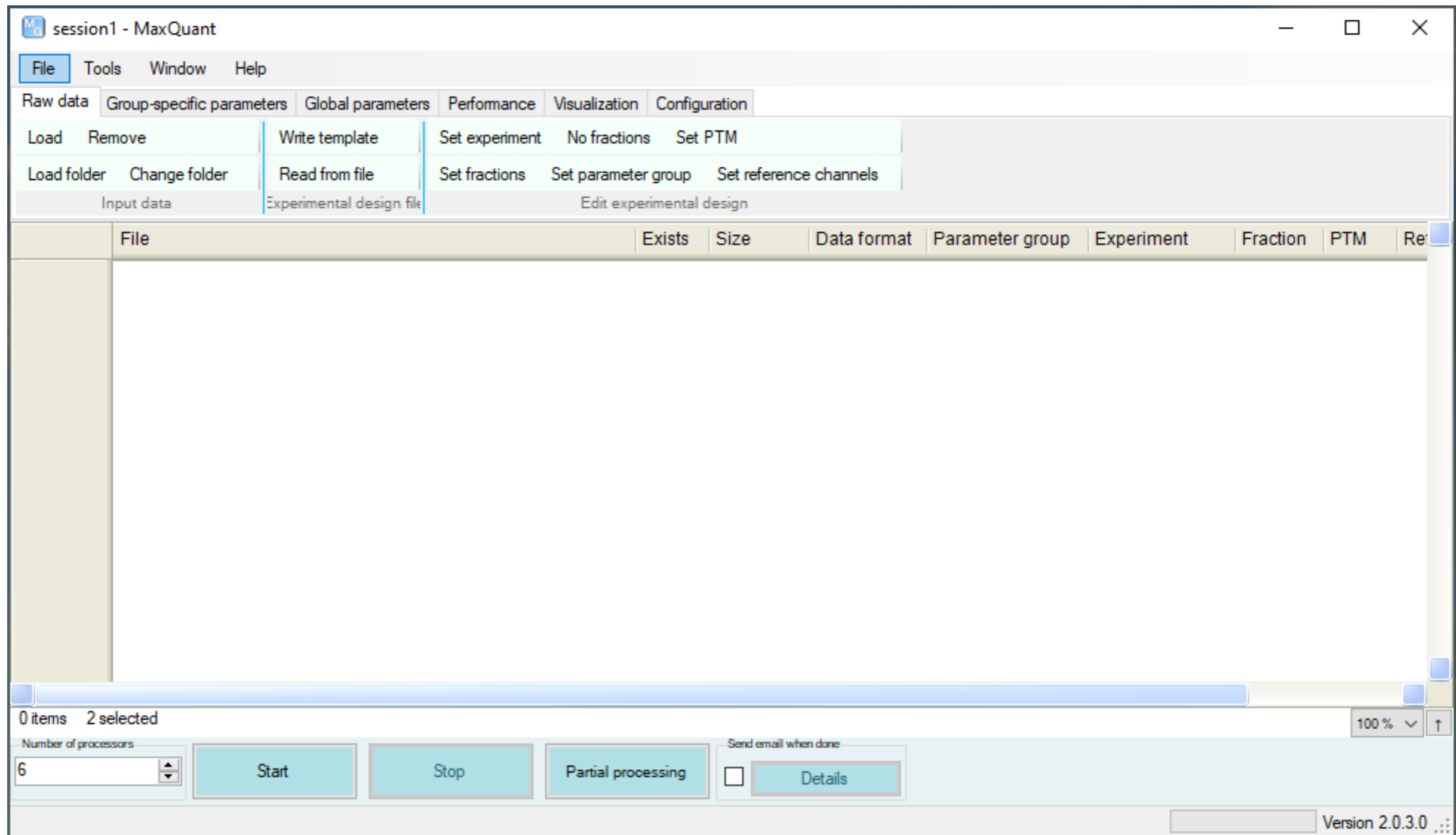


- Statistical Analysis of Quantitative Results

# Data File Organisation

---

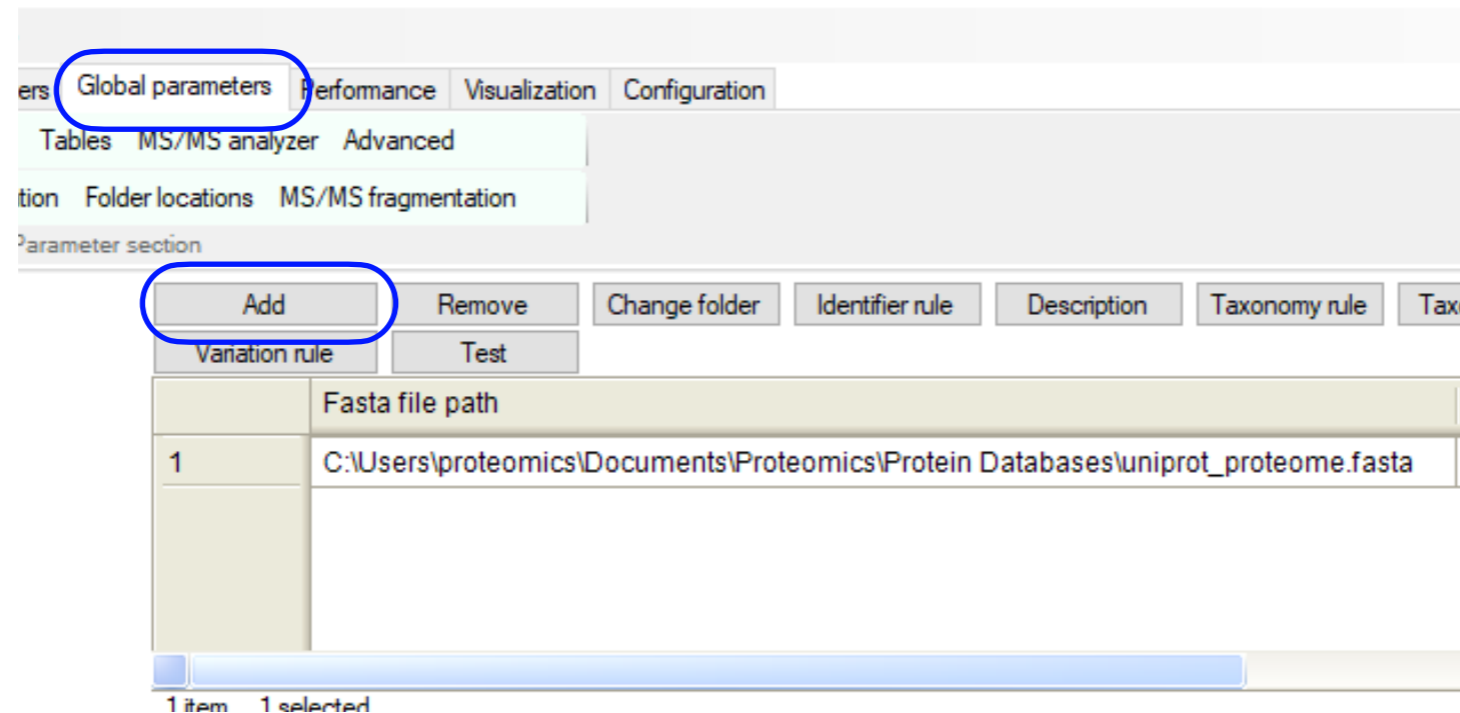






## Database Search Programme

1. Set the database: **Global Parameters -> Add**



- Andromed Indices
- Database Search Results
- MS Data
- Protein Databases
- Settings Files
- Working

## Database Search Programme

### 2. Configure Protein Modifications: **Configuration > Modification**

The screenshot shows the 'Configuration > Modification' window in MaxQuant. The window title is 'session1 - MaxQuant'. The menu bar includes 'File', 'Tools', 'Window', and 'Help'. The 'Configuration' tab is active, and the 'Modifications' sub-tab is selected. The 'Add' button is highlighted. The table below lists various protein modifications with columns for Name, Description, Composition, Position, Type, New terminus, Specificities, Creation date, Last modified, and User.

	Name	Description	Composition	Position	Type	New terminus	Specificities	Creation date	Last modified	User
17	Deamidation 18O (N)	Deamidation 18O (N)	H(-1) N(-1...	Anywhere	Standard	None	[N]	10/18/2010...	10/18/2010...	cox
18	Deamidation (NQ)	Deamidation (NQ)	H(-1) N(-1...	Anywhere	Standard	None	[N],[Q]	6/15/2011 1...	6/15/2011 1...	cox
19	Hydroxyproline	4-hydroxyproline	O	Anywhere	Standard	None	[P]	5/6/2014 1...	5/6/2014 1...	cox
20	Carbamyl (N-term)	Carbamylation	C H N O	Any N-term	Standard	None	[-]	2/16/2015 5...	2/16/2015 5...	cox
21	Delta:H(2)C(2) (N-term)	Acetaldehyde +26	C(2) H(2)	Any N-term	Standard	None	[-]	2/16/2015 5...	2/16/2015 5...	cox
22	Dioxidation (MW)	Dioxidation	O(2)	Anywhere	Standard	None	[M],[W]	2/16/2015 5...	2/16/2015 5...	cox
23	Trioxidation (C)	Trioxidation	O(3)	Anywhere	Standard	None	[C]	2/16/2015 5...	2/16/2015 5...	cox
24	Dethiomethyl (M)	Prompt loss of side chain from...	H(-4) C(-1...	Anywhere	Standard	None	[M]	2/16/2015 6...	2/16/2015 7...	cox
25	Cation:Na (DE)	Sodium adduct	H(-1) Na	Anywhere	Standard	None	[D],[E]	2/16/2015 7...	2/16/2015 7...	cox
26	Methyl (E)	Methylation	C H(2)	Anywhere	Standard	None	[E]	2/16/2015 7...	2/16/2015 7...	cox
27	Dehydrated (ST)	Prompt loss of phosphate from...	H(-2) O(-1)	Anywhere	Standard	None	[S],[T]	2/16/2015 8...	2/16/2015 8...	cox
28	Oxidation (P)	Oxidation	O	Anywhere	Standard	None	[P]	2/16/2015 8...	2/16/2015 8...	cox
29	Dimethyl (K)	di-Methylation	H(4) C(2)	Anywhere	Standard	None	[K]	2/16/2015 8...	2/16/2015 8...	cox
30	Amidated (Protein C-term)	Amidation of the protein C-term...	H N O(-1)	Protein C-...	Standard	None	[-]	2/16/2015 8...	3/12/2015 1...	cox
31	Sulfo (STY)	O-Sulfonation	S	Anywhere	Standard	None	[S],[T],[Y]	2/16/2015 8...	2/16/2015 9...	cox
32	Acetyl (N-term)	Acetylation of every peptide N...	C(2) H(2) O	Any N-term	Standard	None	[A]	2/12/2015 1...	2/12/2015 1...	cox

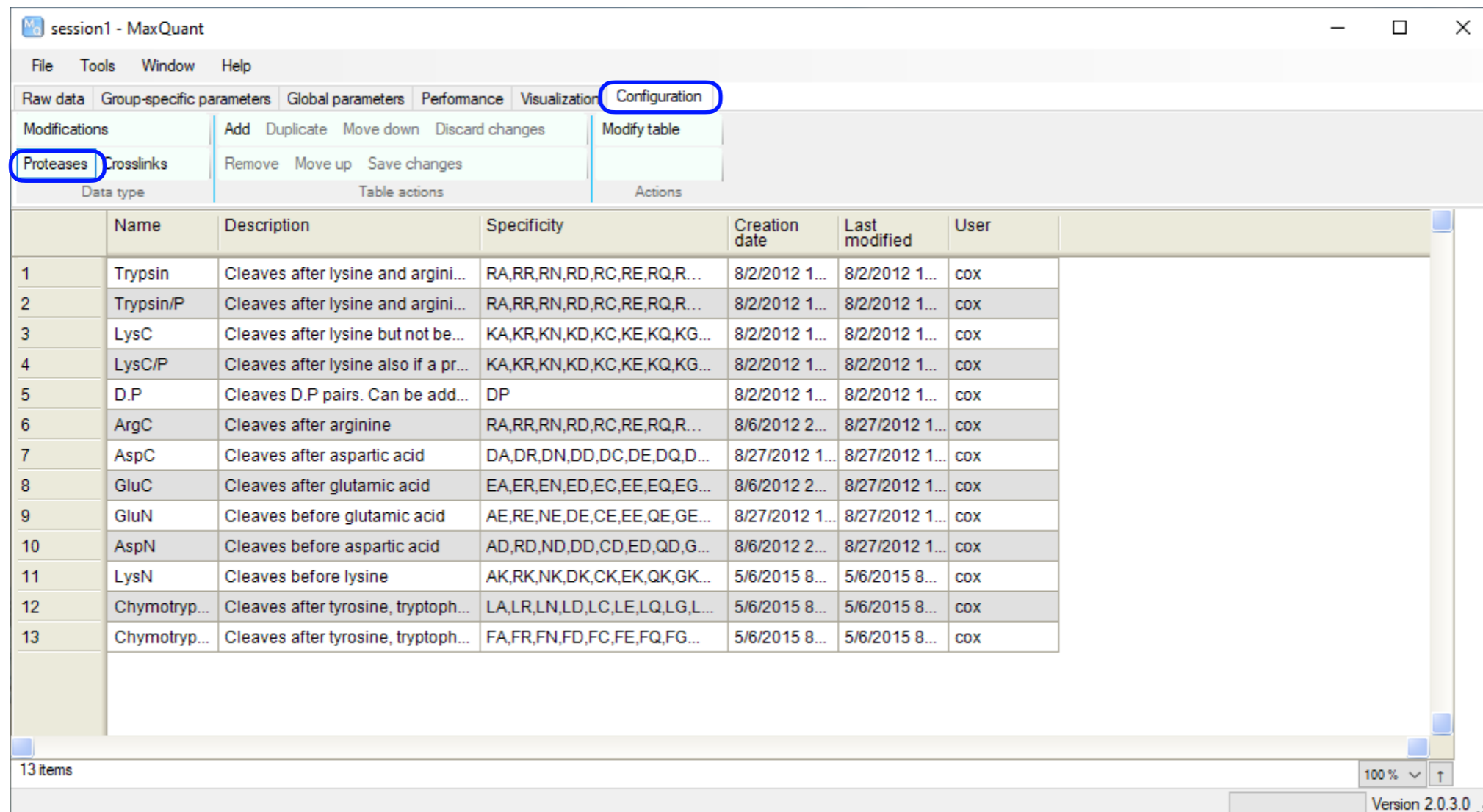
551 items

100% ↑

Version 2.0.3.0

## Database Search Programme

### 3. Configure Protease: **Configuration > Proteases**



The screenshot shows the MaxQuant interface with the 'Configuration' tab selected. The 'Proteases' sub-tab is active, displaying a table of proteases. The table has the following columns: Name, Description, Specificity, Creation date, Last modified, and User. There are 13 items listed in the table.

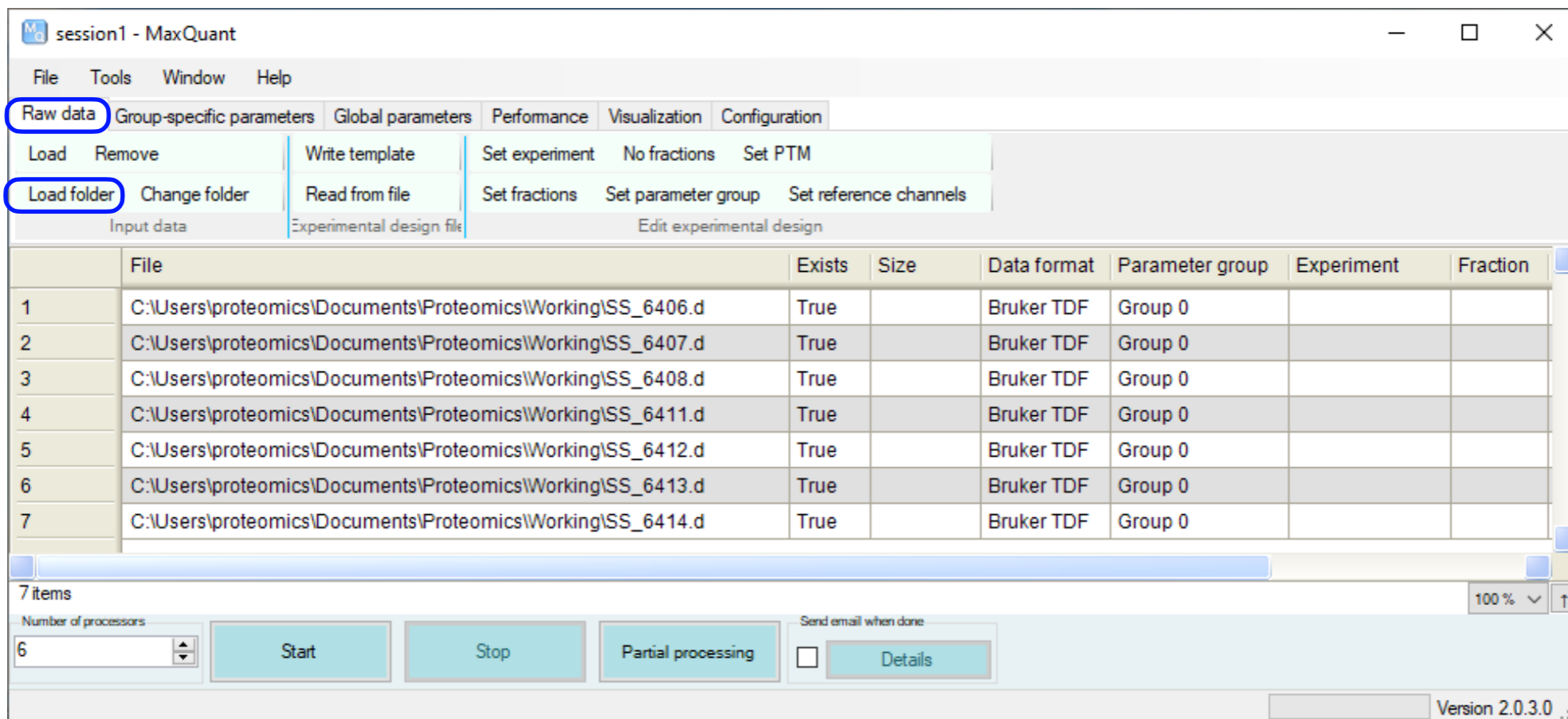
	Name	Description	Specificity	Creation date	Last modified	User
1	Trypsin	Cleaves after lysine and argini...	RA,RR,RN,RD,RC,RE,RQ,R...	8/2/2012 1...	8/2/2012 1...	cox
2	Trypsin/P	Cleaves after lysine and argini...	RA,RR,RN,RD,RC,RE,RQ,R...	8/2/2012 1...	8/2/2012 1...	cox
3	LysC	Cleaves after lysine but not be...	KA,KR,KN,KD,KC,KE,KQ,KG...	8/2/2012 1...	8/2/2012 1...	cox
4	LysC/P	Cleaves after lysine also if a pr...	KA,KR,KN,KD,KC,KE,KQ,KG...	8/2/2012 1...	8/2/2012 1...	cox
5	D.P	Cleaves D.P pairs. Can be add...	DP	8/2/2012 1...	8/2/2012 1...	cox
6	ArgC	Cleaves after arginine	RA,RR,RN,RD,RC,RE,RQ,R...	8/6/2012 2...	8/27/2012 1...	cox
7	AspC	Cleaves after aspartic acid	DA,DR,DN,DD,DC,DE,DQ,D...	8/27/2012 1...	8/27/2012 1...	cox
8	GluC	Cleaves after glutamic acid	EA,ER,EN,ED,EC,EE,EQ,EG...	8/6/2012 2...	8/27/2012 1...	cox
9	GluN	Cleaves before glutamic acid	AE,RE,NE,DE,CE,EE,QE,GE...	8/27/2012 1...	8/27/2012 1...	cox
10	AspN	Cleaves before aspartic acid	AD,RD,ND,DD,CD,ED,QD,G...	8/6/2012 2...	8/27/2012 1...	cox
11	LysN	Cleaves before lysine	AK,RK,NK,DK,CK,EK,QK,GK...	5/6/2015 8...	5/6/2015 8...	cox
12	Chymotryp...	Cleaves after tyrosine, tryptoph...	LA,LR,LN,LD,LC,LE,LQ,LG,L...	5/6/2015 8...	5/6/2015 8...	cox
13	Chymotryp...	Cleaves after tyrosine, tryptoph...	FA,FR,FR,FD,FC,FE,FQ,FG...	5/6/2015 8...	5/6/2015 8...	cox

13 items

Version 2.0.3.0

## Database Search Programme

### 4. Load Mass Spec files: **Raw data > Load Folder / Load**



The screenshot shows the MaxQuant software interface. The 'Raw data' menu is open, highlighting the 'Load folder' option. Below the menu is a table listing 7 mass spec files. The table has columns for File, Exists, Size, Data format, Parameter group, Experiment, and Fraction. The 'File' column contains paths to .d files in the 'Working' directory. The 'Exists' column is 'True' for all files. The 'Data format' is 'Bruker TDF' for all. The 'Parameter group' is 'Group 0' for all. The 'Experiment' and 'Fraction' columns are empty.

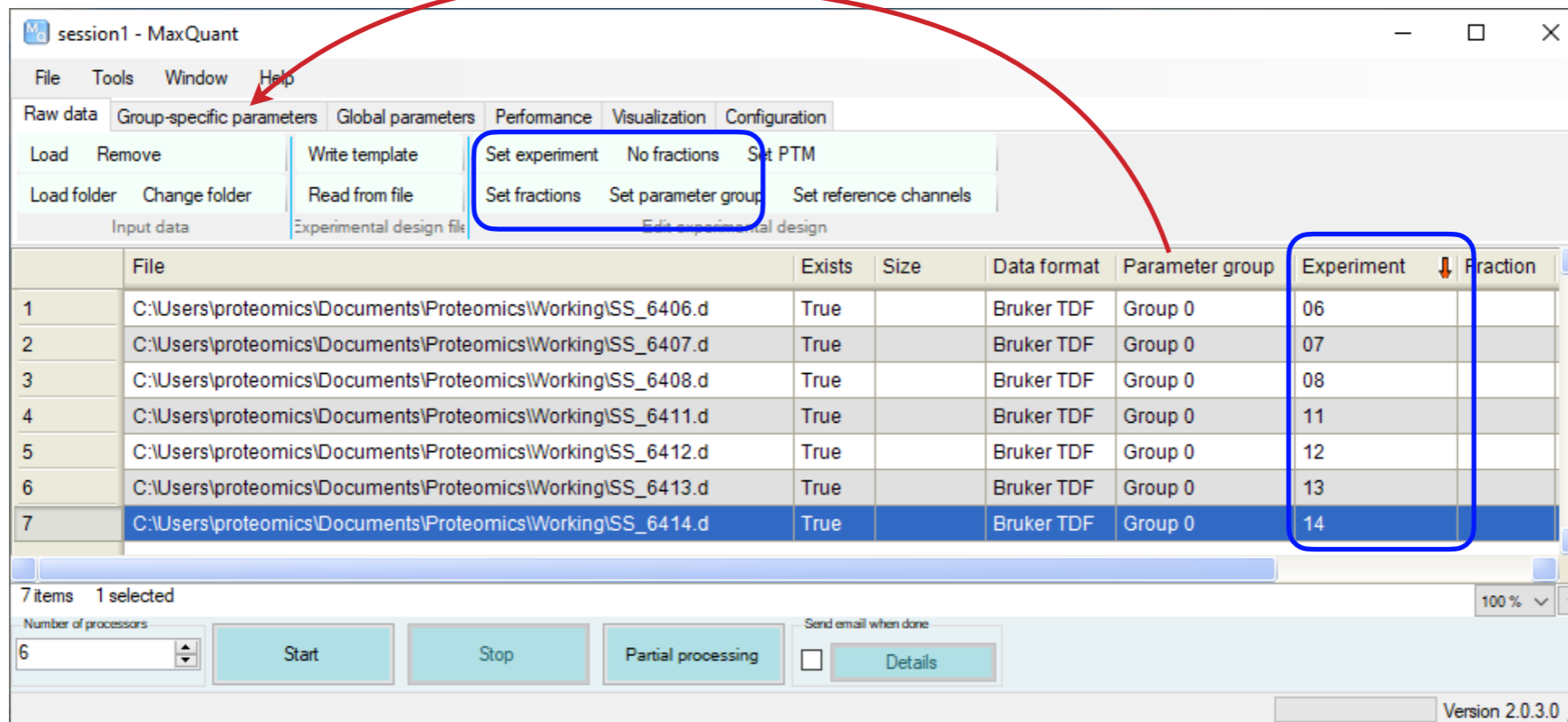
	File	Exists	Size	Data format	Parameter group	Experiment	Fraction
1	C:\Users\proteomics\Documents\Proteomics\Working\SS_6406.d	True		Bruker TDF	Group 0		
2	C:\Users\proteomics\Documents\Proteomics\Working\SS_6407.d	True		Bruker TDF	Group 0		
3	C:\Users\proteomics\Documents\Proteomics\Working\SS_6408.d	True		Bruker TDF	Group 0		
4	C:\Users\proteomics\Documents\Proteomics\Working\SS_6411.d	True		Bruker TDF	Group 0		
5	C:\Users\proteomics\Documents\Proteomics\Working\SS_6412.d	True		Bruker TDF	Group 0		
6	C:\Users\proteomics\Documents\Proteomics\Working\SS_6413.d	True		Bruker TDF	Group 0		
7	C:\Users\proteomics\Documents\Proteomics\Working\SS_6414.d	True		Bruker TDF	Group 0		

7 items  
Number of processors: 6  
Start Stop Partial processing Send email when done:  Details  
Version 2.0.3.0

- Andromed Indices
- Database Search Results
- MS Data
- Protein Databases
- Settings Files
- Working**

## Database Search Programme

5. Define Experimental Order: **Raw data > Set Experiment**  
**Set Fraction**  
**No Fraction**



The screenshot shows the MaxQuant software interface. The 'Raw data' menu is open, highlighting the 'Set experiment' option. A red arrow points from the 'Set experiment' option to the 'Experiment' column in the table below. The table lists 7 items with columns for File, Exists, Size, Data format, Parameter group, Experiment, and Fraction. The 'Experiment' column values are 06, 07, 08, 11, 12, 13, and 14. The 'Fraction' column is empty. The 'Set experiment' option is circled in blue, and the 'Experiment' column header is also circled in blue.

	File	Exists	Size	Data format	Parameter group	Experiment	Fraction
1	C:\Users\proteomics\Documents\Proteomics\Working\SS_6406.d	True		Bruker TDF	Group 0	06	
2	C:\Users\proteomics\Documents\Proteomics\Working\SS_6407.d	True		Bruker TDF	Group 0	07	
3	C:\Users\proteomics\Documents\Proteomics\Working\SS_6408.d	True		Bruker TDF	Group 0	08	
4	C:\Users\proteomics\Documents\Proteomics\Working\SS_6411.d	True		Bruker TDF	Group 0	11	
5	C:\Users\proteomics\Documents\Proteomics\Working\SS_6412.d	True		Bruker TDF	Group 0	12	
6	C:\Users\proteomics\Documents\Proteomics\Working\SS_6413.d	True		Bruker TDF	Group 0	13	
7	C:\Users\proteomics\Documents\Proteomics\Working\SS_6414.d	True		Bruker TDF	Group 0	14	

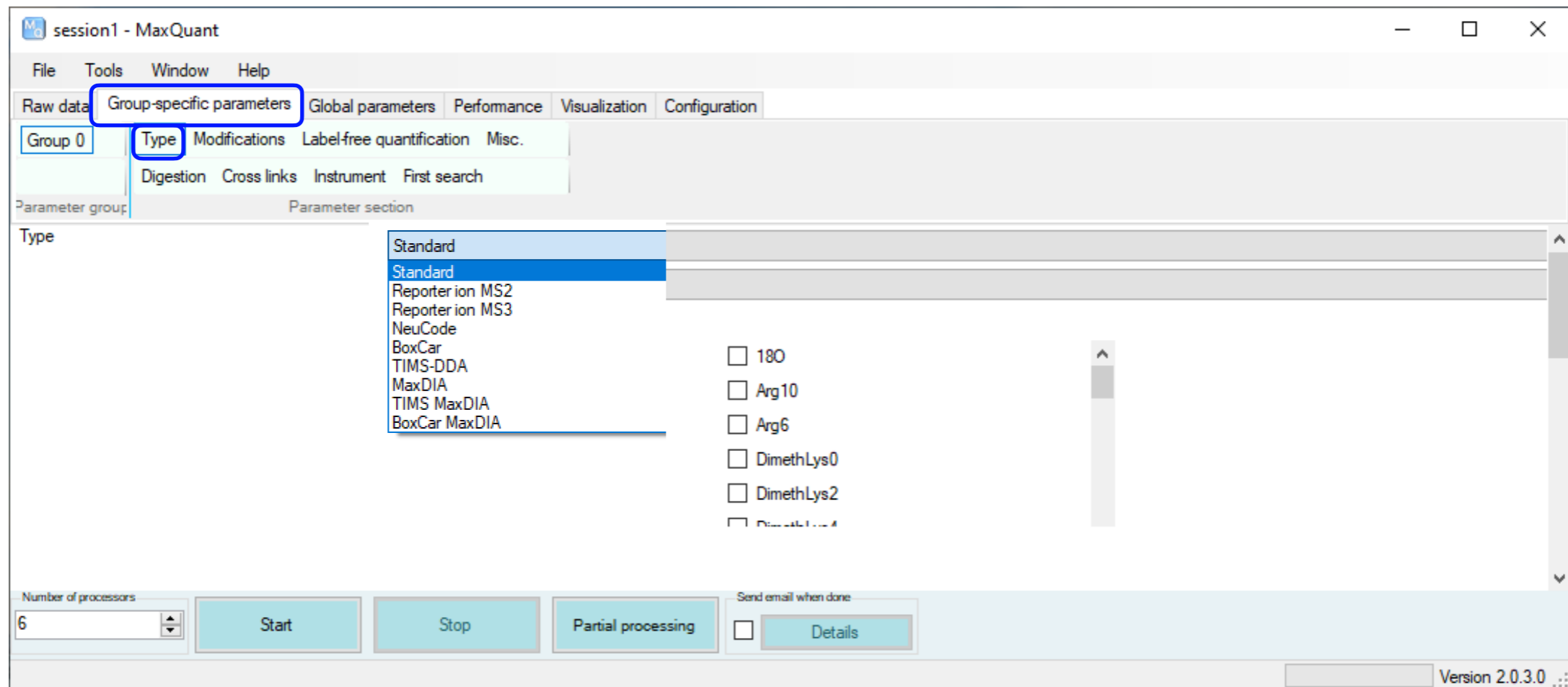
7 items 1 selected  
 Number of processors: 6  
 Start Stop Partial processing Send email when done:  Details  
 Version 2.0.3.0

## Database Search Programme

6. Define Details of MS Data and Sample Content:

**Group-specific parameters > Type > Standard**

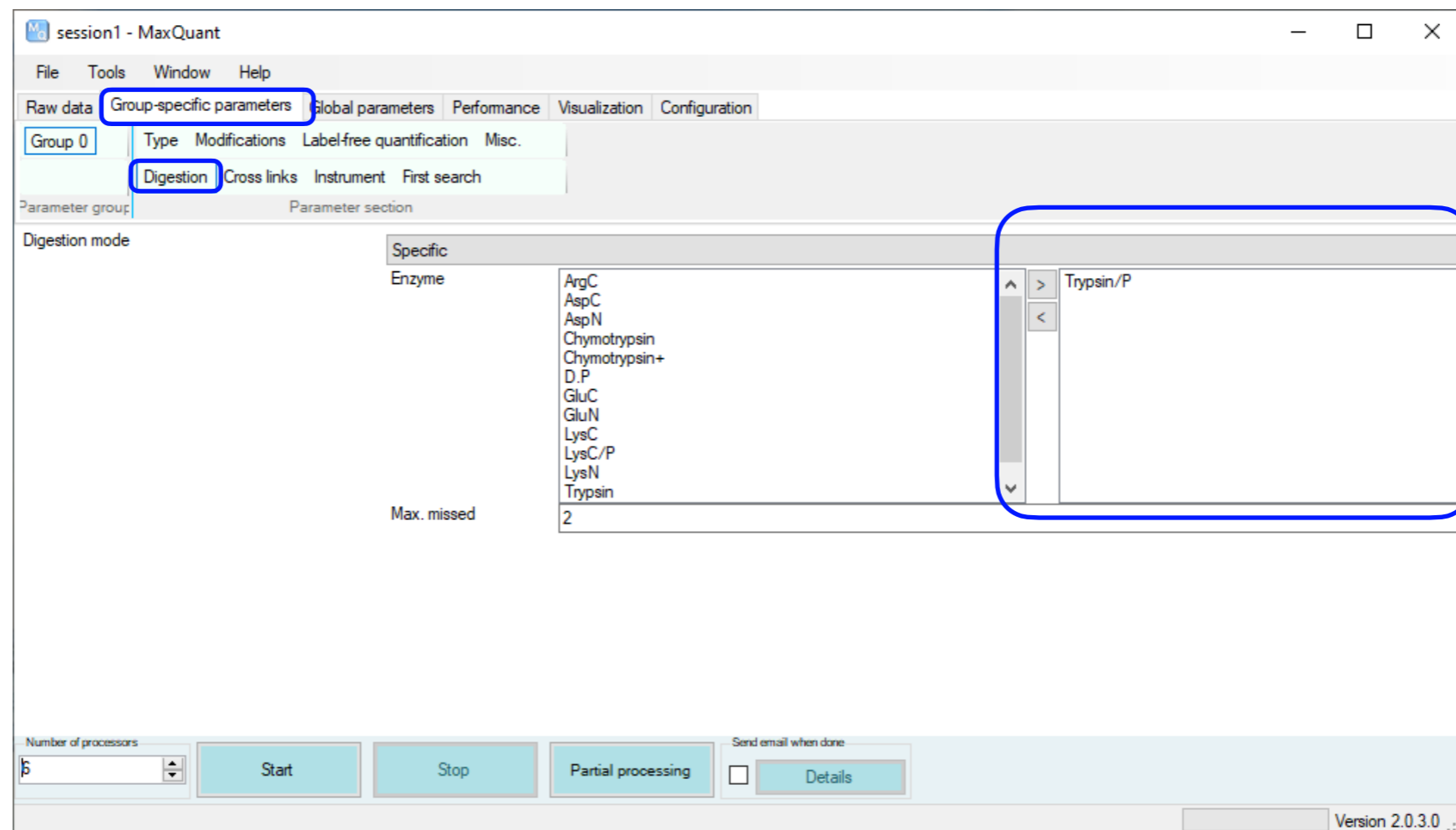
Where should the quantification come from ?



## Database Search Programme

6. Define Details of MS Data and Sample Content:

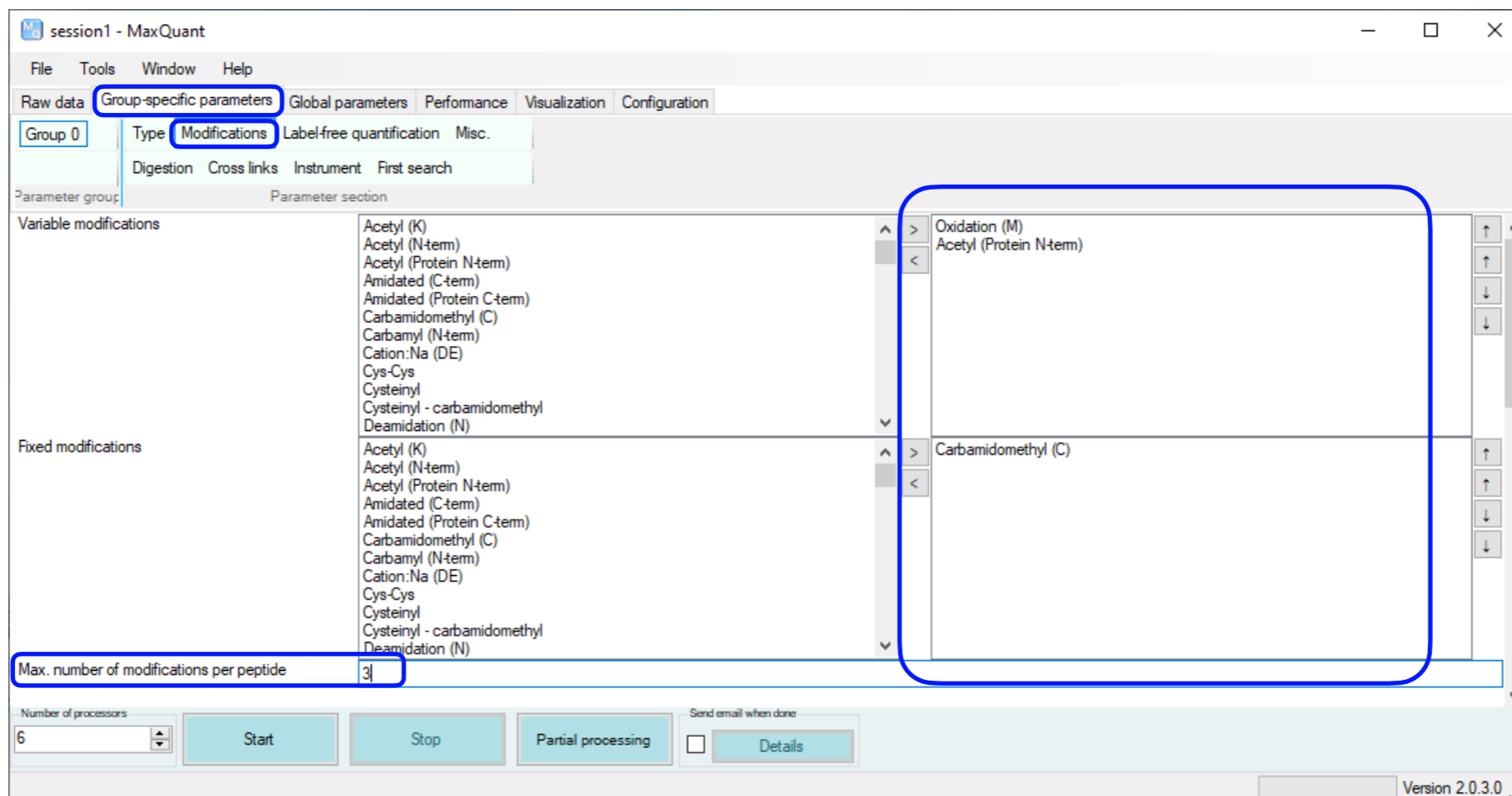
### Group-specific parameters > Digestion



## Database Search Programme

6. Define Details of MS Data and Sample Content:

### Group-specific parameters > Modifications

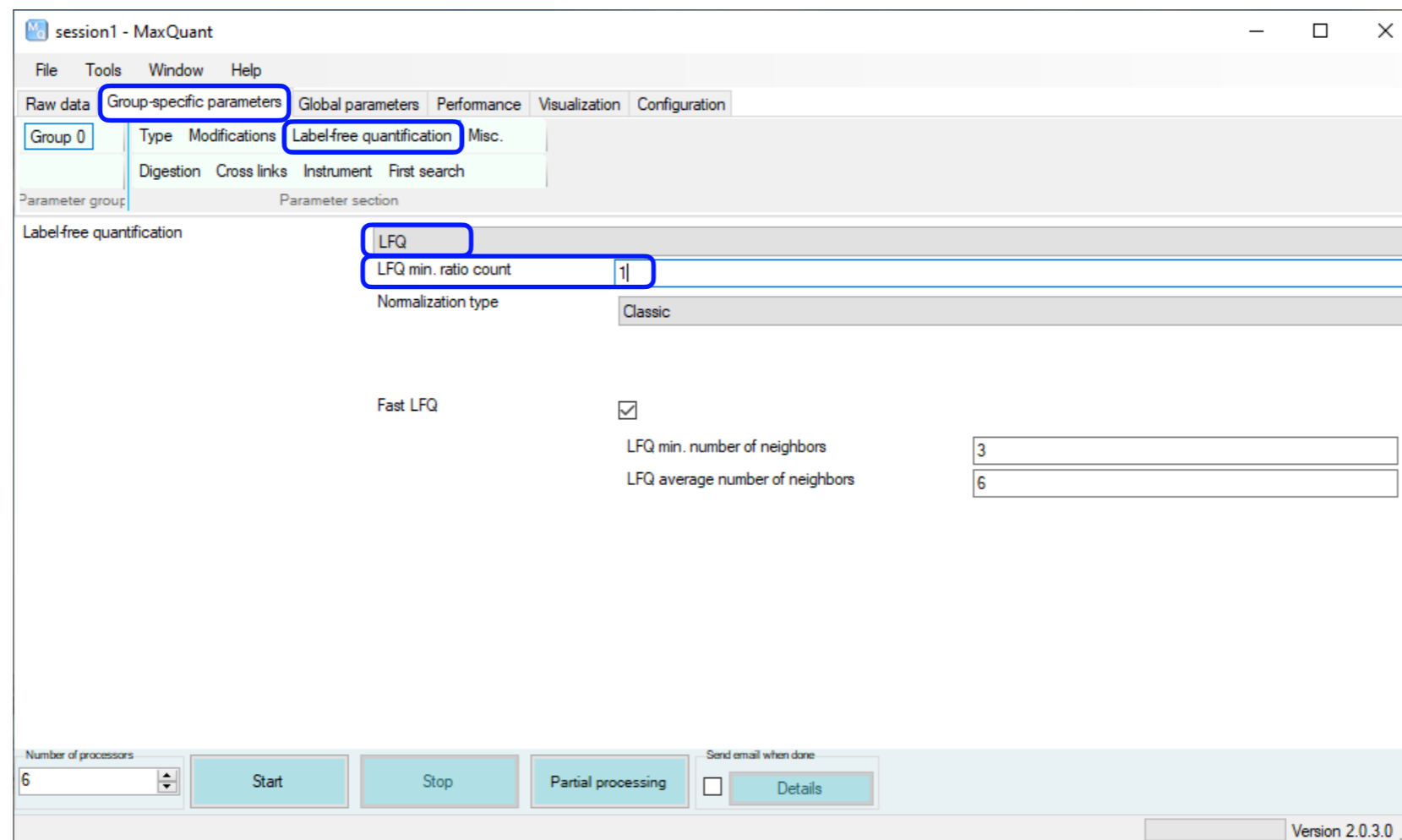




## Database Search Programme

6. Define Details of MS Data and Sample Content:

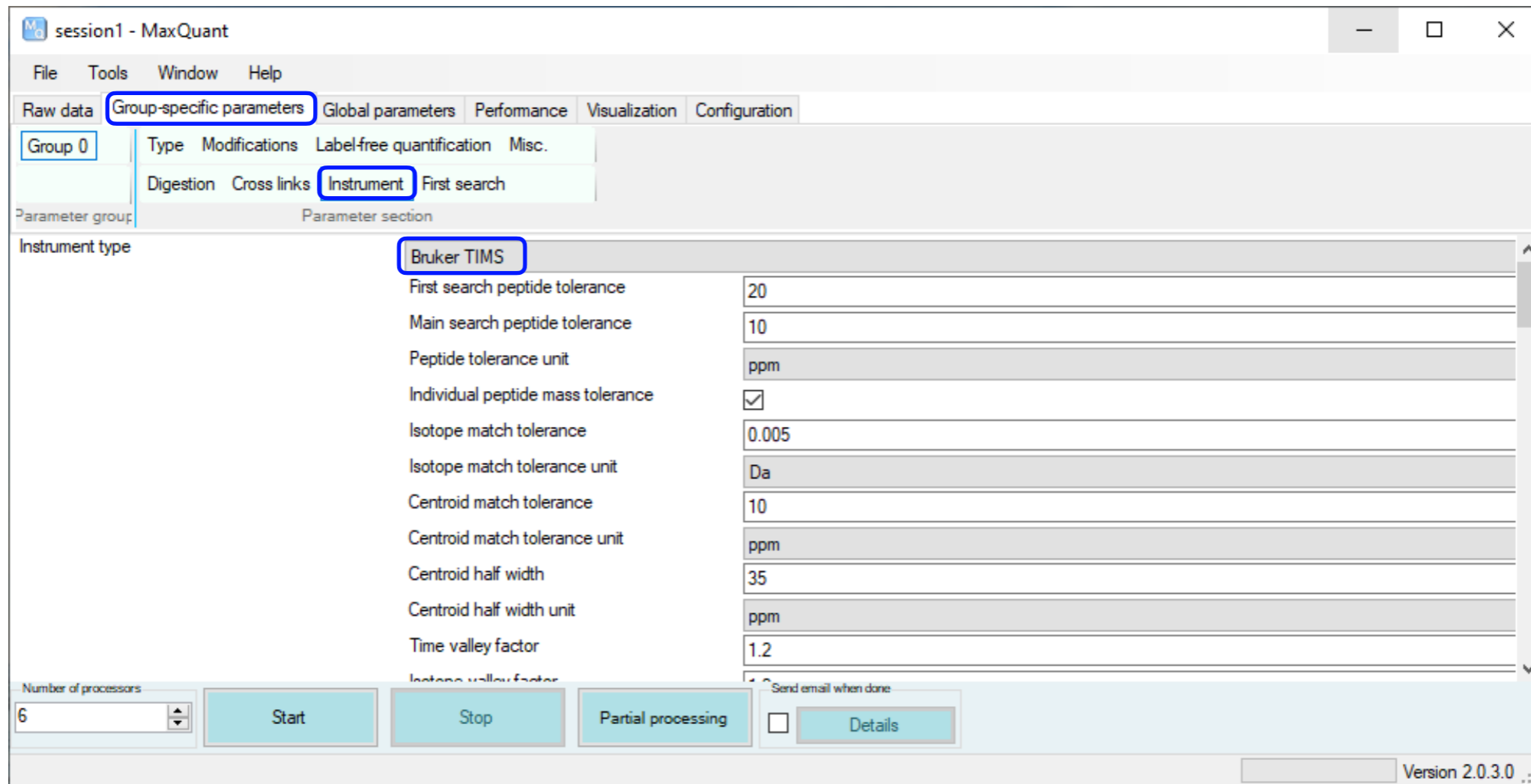
### **Group-specific parameters > Label-free quantification**



## Database Search Programme

### 6. Define Details of MS Data and Sample Content:

### Instrument



session1 - MaxQuant

File Tools Window Help

Raw data **Group-specific parameters** Global parameters Performance Visualization Configuration

Group 0 Type Modifications Label-free quantification Misc.  
Digestion Cross links **Instrument** First search

Parameter group: Parameter section

Instrument type **Bruker TIMS**

First search peptide tolerance	20
Main search peptide tolerance	10
Peptide tolerance unit	ppm
Individual peptide mass tolerance	<input checked="" type="checkbox"/>
Isotope match tolerance	0.005
Isotope match tolerance unit	Da
Centroid match tolerance	10
Centroid match tolerance unit	ppm
Centroid half width	35
Centroid half width unit	ppm
Time valley factor	1.2

Number of processors: 6

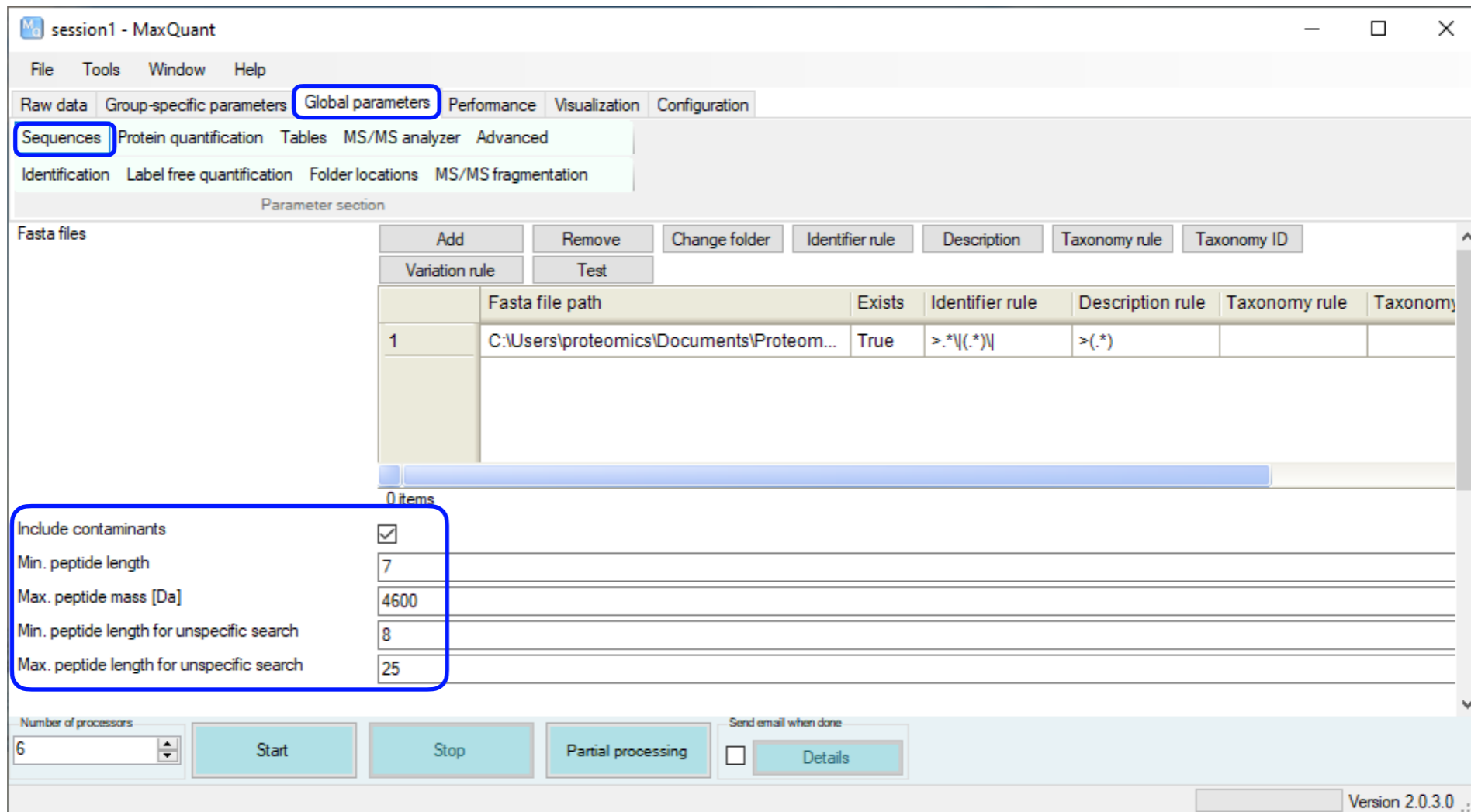
Start Stop Partial processing  Send email when done Details

Version 2.0.3.0

## Database Search Programme

7. Define How the Database Search Should be Done:

### Global parameters > Sequences



session1 - MaxQuant

File Tools Window Help

Raw data Group-specific parameters **Global parameters** Performance Visualization Configuration

**Sequences** Protein quantification Tables MS/MS analyzer Advanced

Identification Label free quantification Folder locations MS/MS fragmentation

Parameter section

Fasta files

Add Remove Change folder Identifier rule Description Taxonomy rule Taxonomy ID

Variation rule Test

	Fasta file path	Exists	Identifier rule	Description rule	Taxonomy rule	Taxonomy ID
1	C:\Users\proteomics\Documents\Proteom...	True	>.*\(.*)\	>(.*)		

0 items

Include contaminants

Min. peptide length 7

Max. peptide mass [Da] 4600

Min. peptide length for unspecific search 8

Max. peptide length for unspecific search 25

Number of processors 6

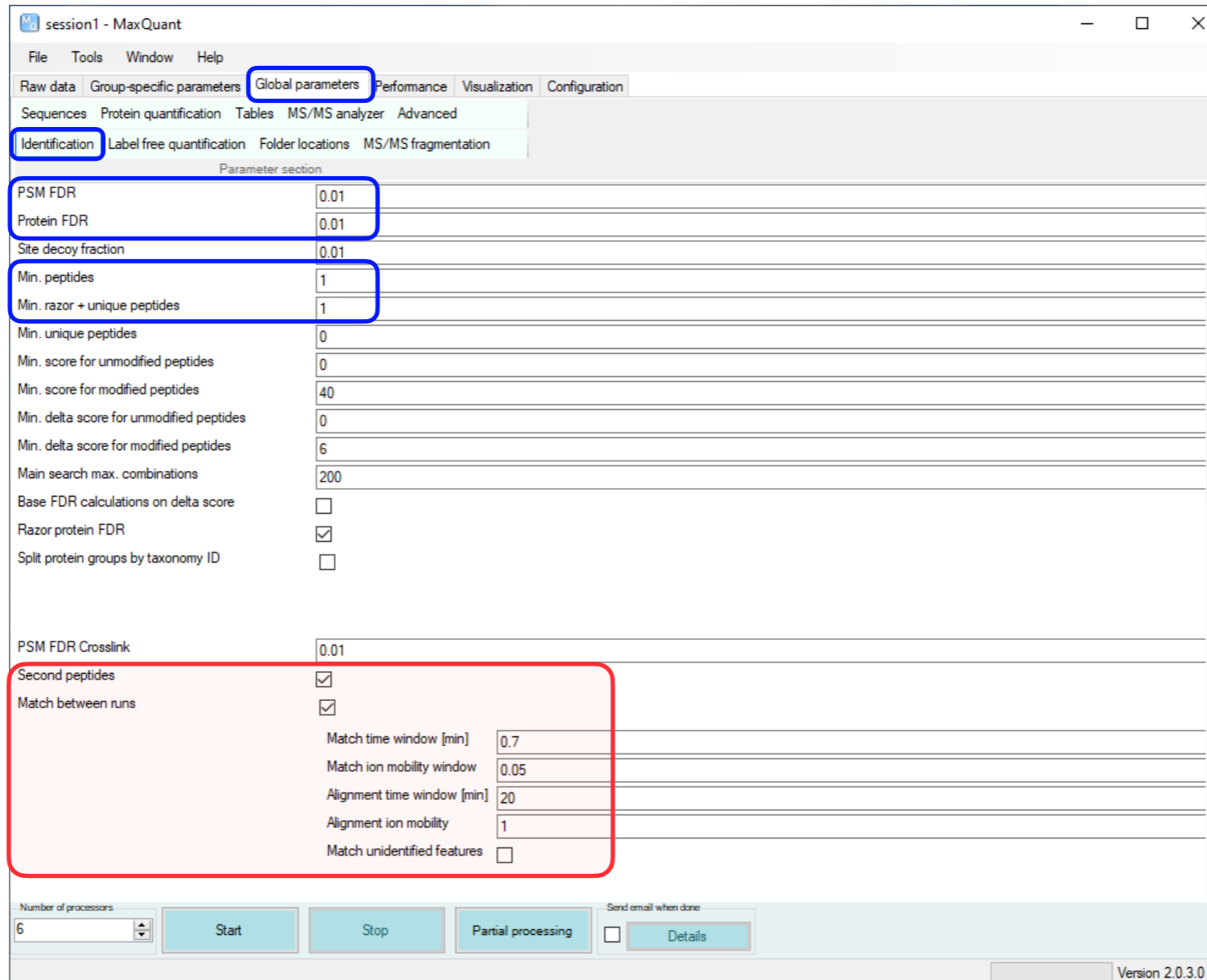
Start Stop Partial processing  Send email when done Details

Version 2.0.3.0

## Database Search Programme

7. Define How the Database Search Should be Done:

### Global parameters > Identification



session1 - MaxQuant

File Tools Window Help

Raw data Group-specific parameters **Global parameters** Performance Visualization Configuration

Sequences Protein quantification Tables MS/MS analyzer Advanced

**Identification** Label free quantification Folder locations MS/MS fragmentation

Parameter section

PSM FDR	0.01
Protein FDR	0.01
Site decoy fraction	0.01
Min. peptides	1
Min. razor + unique peptides	1
Min. unique peptides	0
Min. score for unmodified peptides	0
Min. score for modified peptides	40
Min. delta score for unmodified peptides	0
Min. delta score for modified peptides	6
Main search max. combinations	200
Base FDR calculations on delta score	<input type="checkbox"/>
Razor protein FDR	<input checked="" type="checkbox"/>
Split protein groups by taxonomy ID	<input type="checkbox"/>
PSM FDR Crosslink	0.01
Second peptides	<input checked="" type="checkbox"/>
Match between runs	<input checked="" type="checkbox"/>
Match time window [min]	0.7
Match ion mobility window	0.05
Alignment time window [min]	20
Alignment ion mobility	1
Match unidentified features	<input type="checkbox"/>

Number of processors: 6

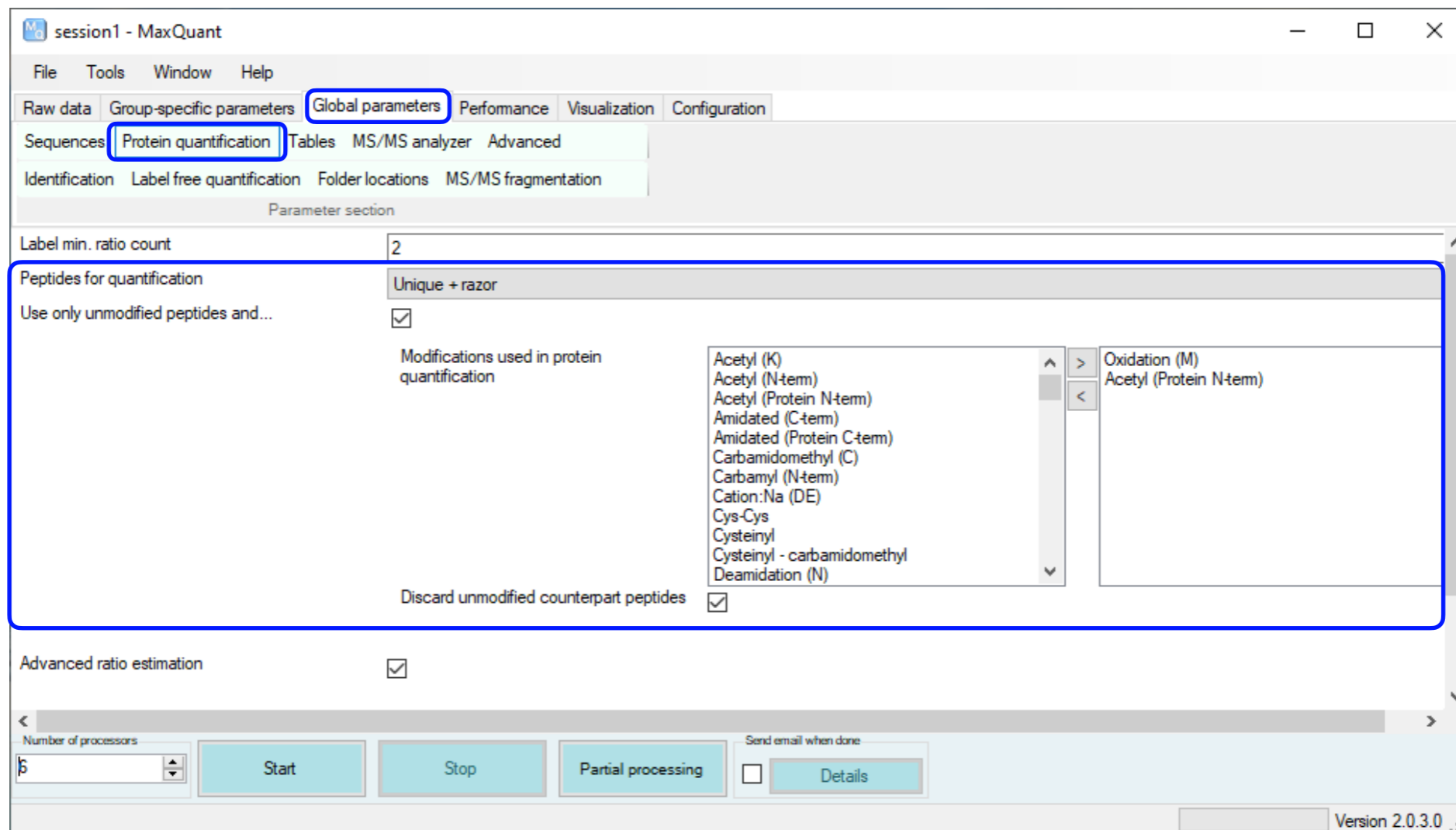
Start Stop Partial processing Send email when done:  Details

Version 2.0.3.0

## Database Search Programme

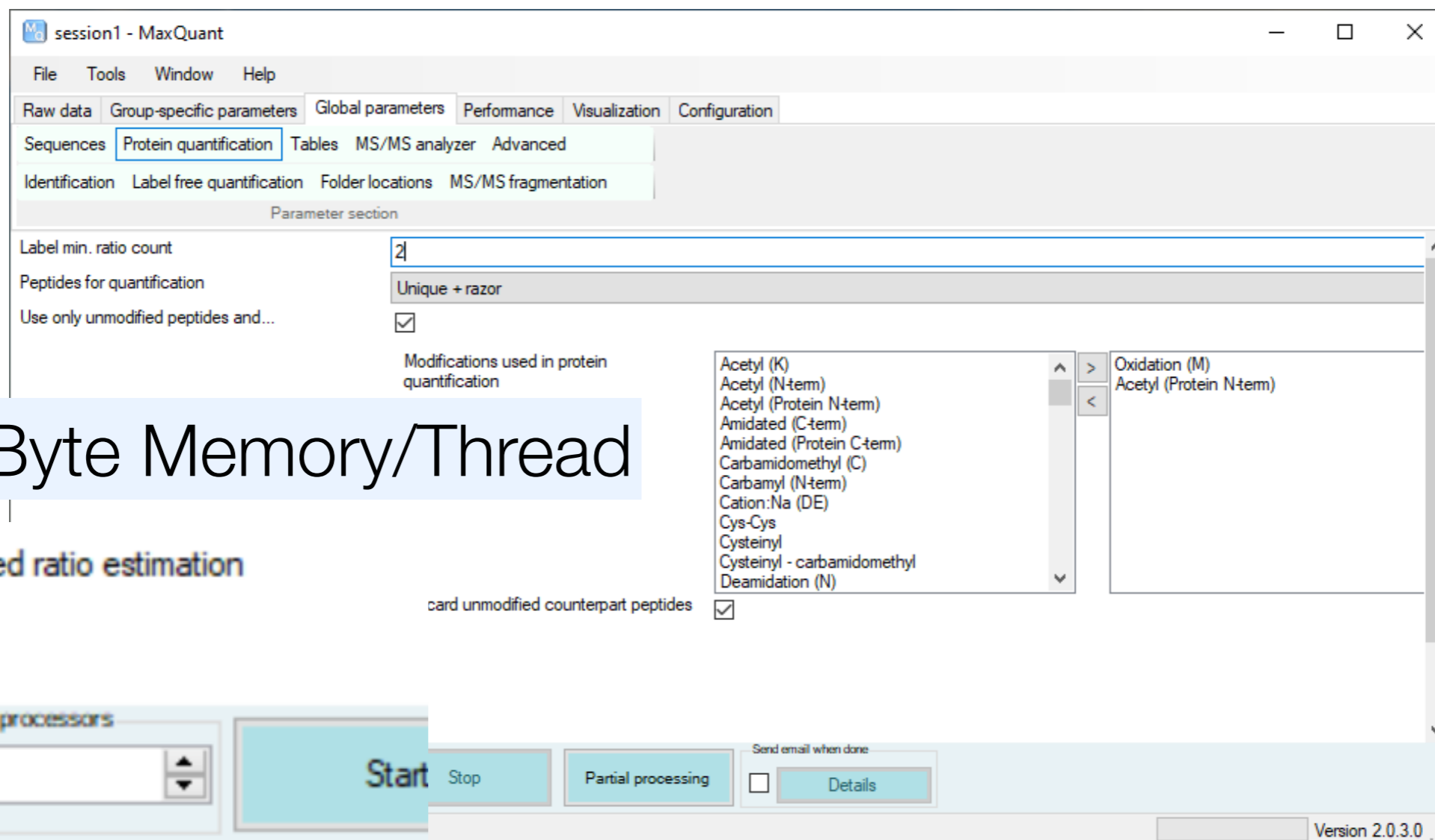
7. Define How the Database Search Should be Done:

### **Global parameters > Protein Quantification**



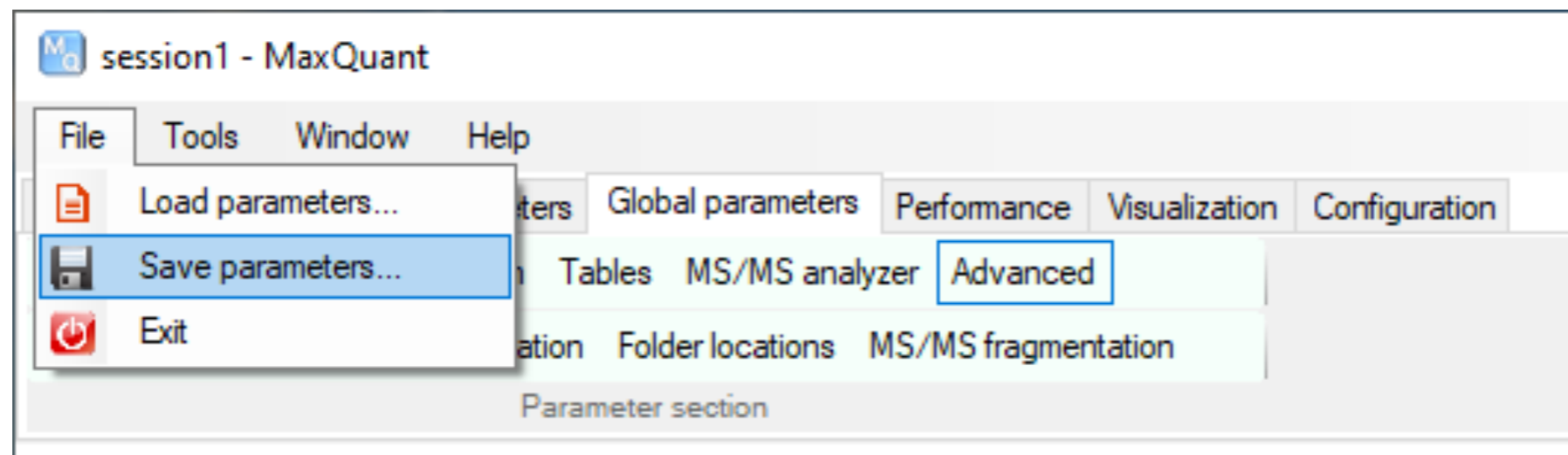
## Database Search Programme

### 8. Define Number of Threads Available:



## Database Search Programme

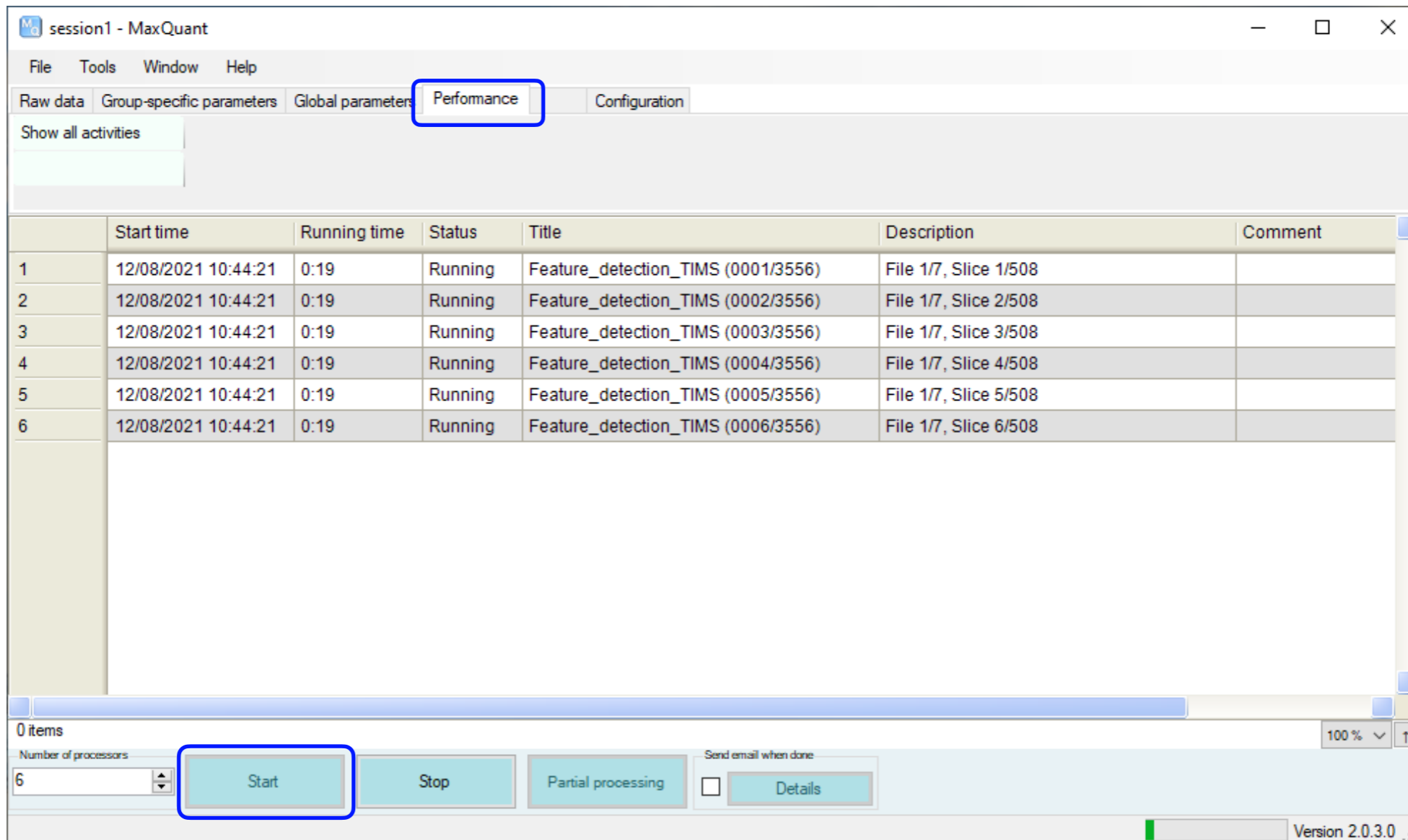
8. Save your Parameters: **File > Save parameters...**



- Andromed Indices
- Database Search Results
- MS Data
- Protein Databases
- Settings Files
- Working

## Database Search Programme

### 8. Start the Search: **Start**



The screenshot shows the MaxQuant Performance tab. The 'Performance' tab is selected and highlighted with a blue box. Below the menu bar, there is a 'Show all activities' button. The main area contains a table with the following data:

	Start time	Running time	Status	Title	Description	Comment
1	12/08/2021 10:44:21	0:19	Running	Feature_detection_TIMS (0001/3556)	File 1/7, Slice 1/508	
2	12/08/2021 10:44:21	0:19	Running	Feature_detection_TIMS (0002/3556)	File 1/7, Slice 2/508	
3	12/08/2021 10:44:21	0:19	Running	Feature_detection_TIMS (0003/3556)	File 1/7, Slice 3/508	
4	12/08/2021 10:44:21	0:19	Running	Feature_detection_TIMS (0004/3556)	File 1/7, Slice 4/508	
5	12/08/2021 10:44:21	0:19	Running	Feature_detection_TIMS (0005/3556)	File 1/7, Slice 5/508	
6	12/08/2021 10:44:21	0:19	Running	Feature_detection_TIMS (0006/3556)	File 1/7, Slice 6/508	

At the bottom of the window, there is a control panel with the following elements:

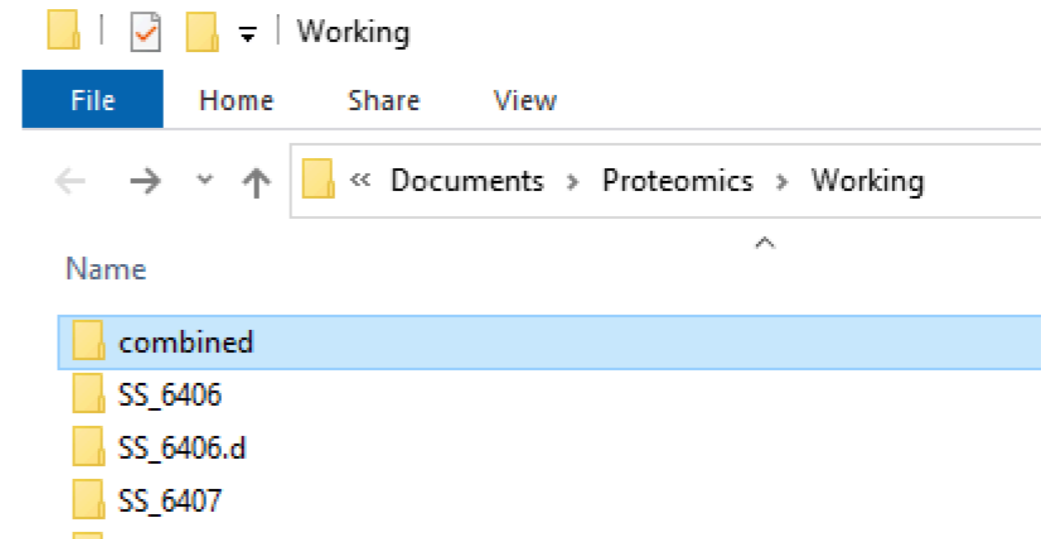
- 0 items
- Number of processors: 6
- Start** button (highlighted with a blue box)
- Stop button
- Partial processing button
- Send email when done checkbox (unchecked)
- Details button
- 100% zoom and scroll controls
- Version 2.0.3.0



## Database Search Programme

Results:

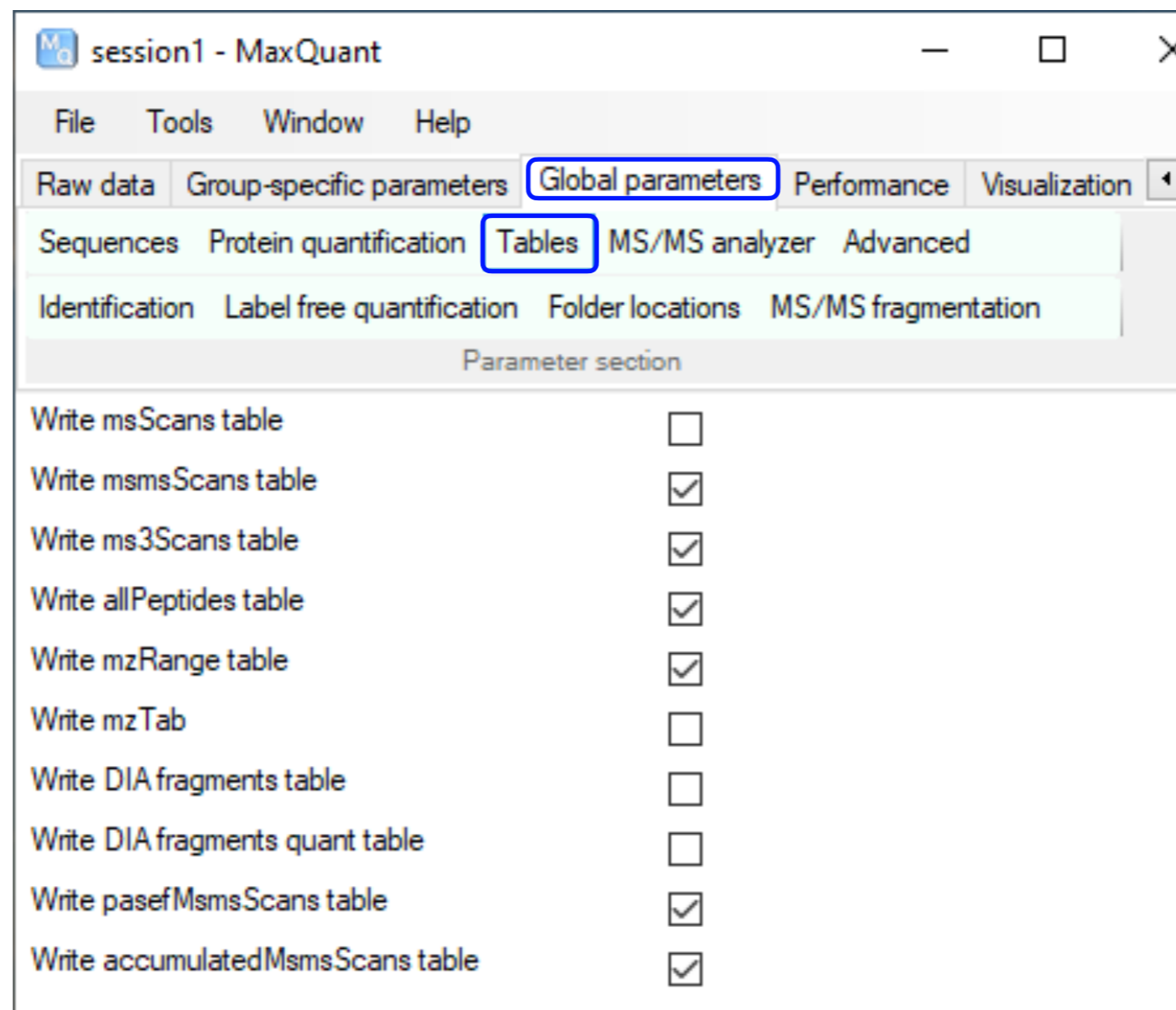
- Andromed Indices
- Database Search Results
- MS Data
- Protein Databases
- Settings Files
- Working**



Name	Date modified	Type	Size
accumulatedMsmsScans	11/26/2021 10:51 PM	Text Document	156,214 KB
allPeptides	11/26/2021 10:51 PM	Text Document	283,127 KB
evidence	11/26/2021 10:51 PM	Text Document	102,459 KB
matchedFeatures	11/26/2021 10:51 PM	Text Document	0 KB
modificationSpecificPeptides	11/26/2021 10:51 PM	Text Document	11,090 KB
msms	11/26/2021 10:51 PM	Text Document	305,656 KB
Oxidation (M)Sites	11/26/2021 10:51 PM	Text Document	161 KB
parameters	11/26/2021 10:50 PM	Text Document	4 KB
pasefMsmsScans	11/26/2021 10:51 PM	Text Document	385,235 KB
peptides	11/26/2021 10:51 PM	Text Document	13,305 KB
proteinGroups	11/26/2021 10:51 PM	Text Document	6,552 KB
summary	11/26/2021 10:51 PM	Text Document	7 KB

## Database Search Programme

Results: Choose written Columns: **Global parameters > Tables**



## Database Search Programme

Results: **summary.txt** :

Raw file	Experiment	Enzyme
211111kw_Simeng_Cntr_1_S1-C1_1_6427	Cntr_1_S1-C1_1_6427	Trypsin/P
211111kw_Simeng_Cntr_1_S1-C2_1_6428	Cntr_1_S1-C2_1_6428	Trypsin/P
211111kw_Simeng_Cntr_2_S1-C3_1_6429	Cntr_2_S1-C3_1_6429	Trypsin/P
211111kw_Simeng_Cntr_2_S1-C4_1_6430	Cntr_2_S1-C4_1_6430	Trypsin/P
211111kw_Simeng_Cntr_3_S1-C5_1_6431	Cntr_3_S1-C5_1_6431	Trypsin/P
211111kw_Simeng_Cntr_3_S1-C6_1_6432	Cntr_3_S1-C6_1_6432	Trypsin/P
211111kw_Simeng_Diff1_S1-A7_1_6409	Diff1_S1-A7_1_6409	Trypsin/P
211111kw_Simeng_Diff1_S1-A8_1_6410	Diff1_S1-A8_1_6410	Trypsin/P
211111kw_Simeng_Diff2_S1-A10_1_6412	Diff2_S1-A10_1_6412	Trypsin/P
211111kw_Simeng_Diff2_S1-A9_1_6411	Diff2_S1-A9_1_6411	Trypsin/P
211111kw_Simeng_Diff3_S1-A11_1_6413	Diff3_S1-A11_1_6413	Trypsin/P
211111kw_Simeng_Diff3_S1-A12_1_6414	Diff3_S1-A12_1_6414	Trypsin/P
211111kw_Simeng_DiffHyp1_S1-A1_1_6403	DiffHyp1_S1-A1_1_6403	Trypsin/P
211111kw_Simeng_DiffHyp1_S1-A2_1_6404	DiffHyp1_S1-A2_1_6404	Trypsin/P
211111kw_Simeng_DiffHyp2_S1-A3_1_6405	DiffHyp2_S1-A3_1_6405	Trypsin/P
211111kw_Simeng_DiffHyp2_S1-A4_1_6406	DiffHyp2_S1-A4_1_6406	Trypsin/P
211111kw_Simeng_DiffHyp3_S1-A5_1_6407	DiffHyp3_S1-A5_1_6407	Trypsin/P
211111kw_Simeng_DiffHyp3_S1-A6_1_6408	DiffHyp3_S1-A6_1_6408	Trypsin/P
<b>Total</b>		



Build your own meta-data file

Experiment	Background	Date	Biological Repeat	Technical Repeat	Raw File Name

## Database Search Programme

Results: **proteinGroups.txt** :

The image shows a dense table of search results from a MaxQuant database search. The table is organized into several columns, including protein names, accession numbers, and search scores. The text is too small to read in detail, but the structure is clear. The table is divided into sections by protein names, with each section containing multiple rows of data. The columns include protein names, accession numbers, and search scores. The table is very long, spanning many rows and columns.

*End*



## Statistical Analysis & Visualization

