中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY , CHINESE ACADEMY OF SCIENCES
pFind Studio

pFind 3 User Guide

Version 3.1.2

pFind Group

2018.4

Contents

- Before software installation
- Software registration
- How to run pFind 3
  - Create a new search task
  - View search results

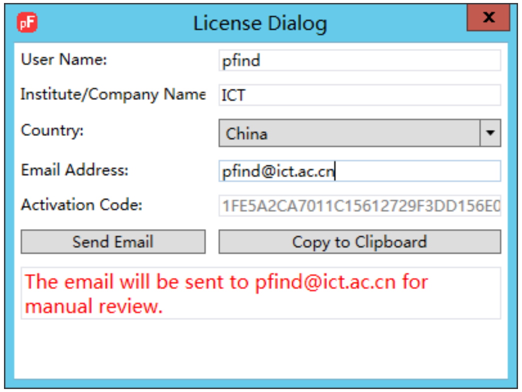
Before software installation

- **Please install 64-bit MSFileReader first**

- **Windows 7 and above**
  - **64 bit version**
  - **.NET Framework 4.5 environment**

Software registration

- **Click pFind.exe and fill in personal information in the following information panel.**


License Dialog

User Name: pfind

Institute/Company Name ICT

Country: China

Email Address: pfind@ict.ac.cn

Activation Code: 1FE5A2CA7011C15612729F3DD156E

The email will be sent to pfind@ict.ac.cn for manual review.

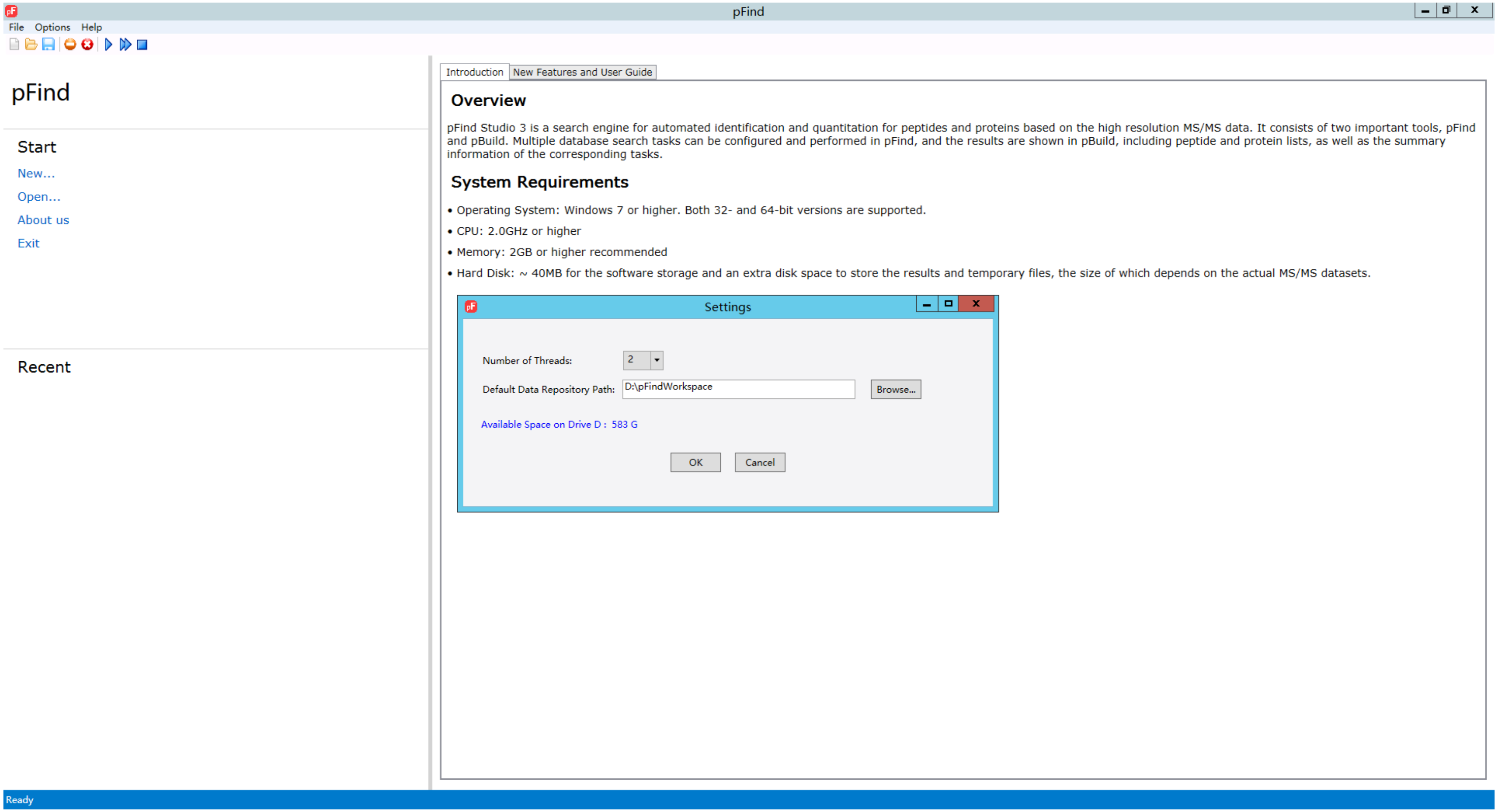
- **Send the information to pfind@ict.ac.cn**

- **Apply for pFind.license.**

- **Put the pFind.license into pFind installation directory under the \pFindStudio\bin folder.**

Main interface of pFind 3

- Select the number of threads and search task storage path according to your computer situation.



Overview

pFind Studio 3 is a search engine for automated identification and quantitation for peptides and proteins based on the high resolution MS/MS data. It consists of two important tools, pFind and pBuild. Multiple database search tasks can be configured and performed in pFind, and the results are shown in pBuild, including peptide and protein lists, as well as the summary information of the corresponding tasks.

System Requirements

- Operating System: Windows 7 or higher. Both 32- and 64-bit versions are supported.
- CPU: 2.0GHz or higher
- Memory: 2GB or higher recommended
- Hard Disk: ~ 40MB for the software storage and an extra disk space to store the results and temporary files, the size of which depends on the actual MS/MS datasets.

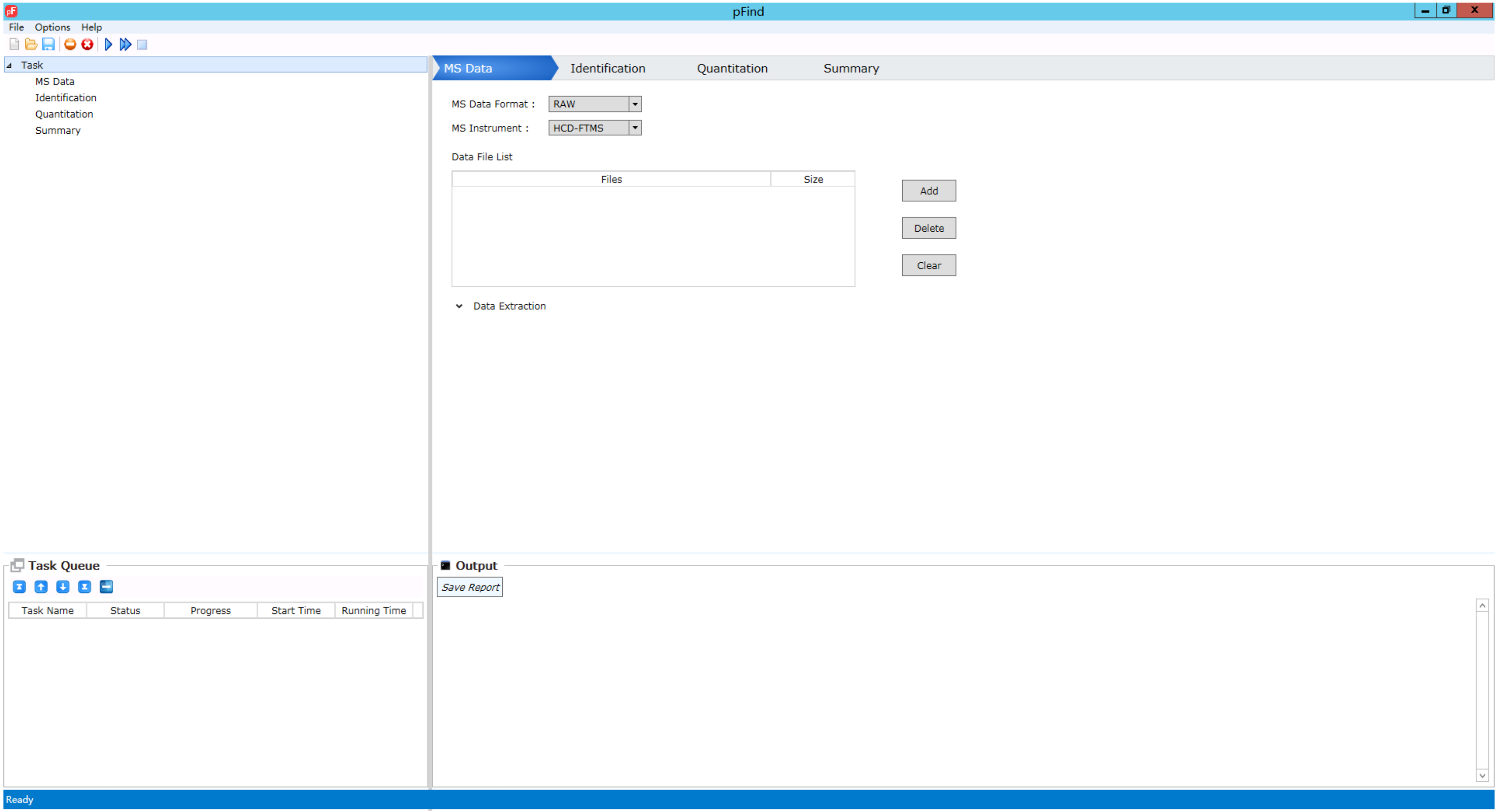
Number of Threads: 2

Default Data Repository Path: D:\pFindWorkspace

Available Space on Drive D： 583 G

Create a new search task

- **Click File→New**

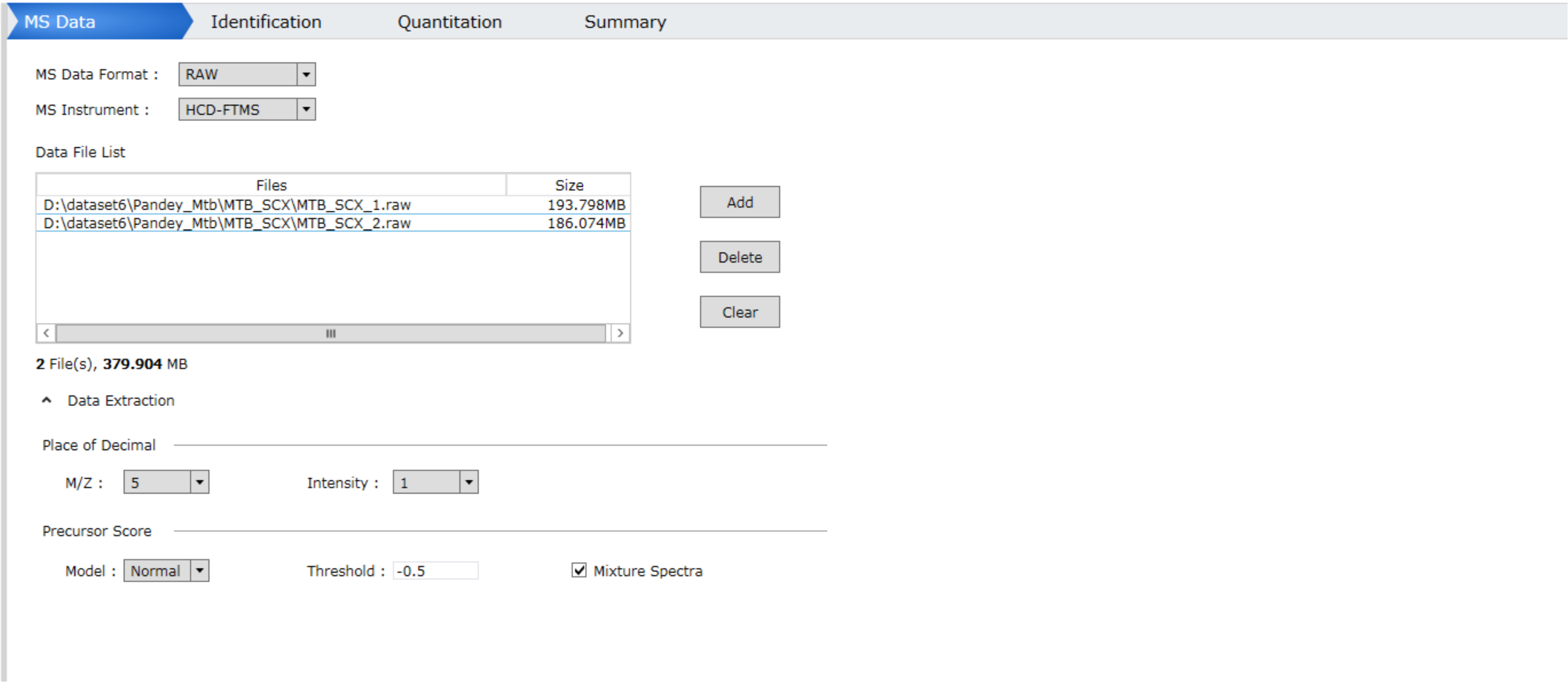


MS Data Format : RAW

MS Instrument : HCD-FTMS

1. Data import and preprocess.

- You can choose MS data format and other data preprocessing type in MS Data panel.



MS Data Format : RAW

MS Instrument : HCD-FTMS

Data File List

Files	Size
D:\dataset6\Pandey_Mtb\MTB_SCX\MTB_SCX_1.raw	193.798MB
D:\dataset6\Pandey_Mtb\MTB_SCX\MTB_SCX_2.raw	186.074MB

2 File(s), **379.904** MB

Data Extraction

Place of Decimal

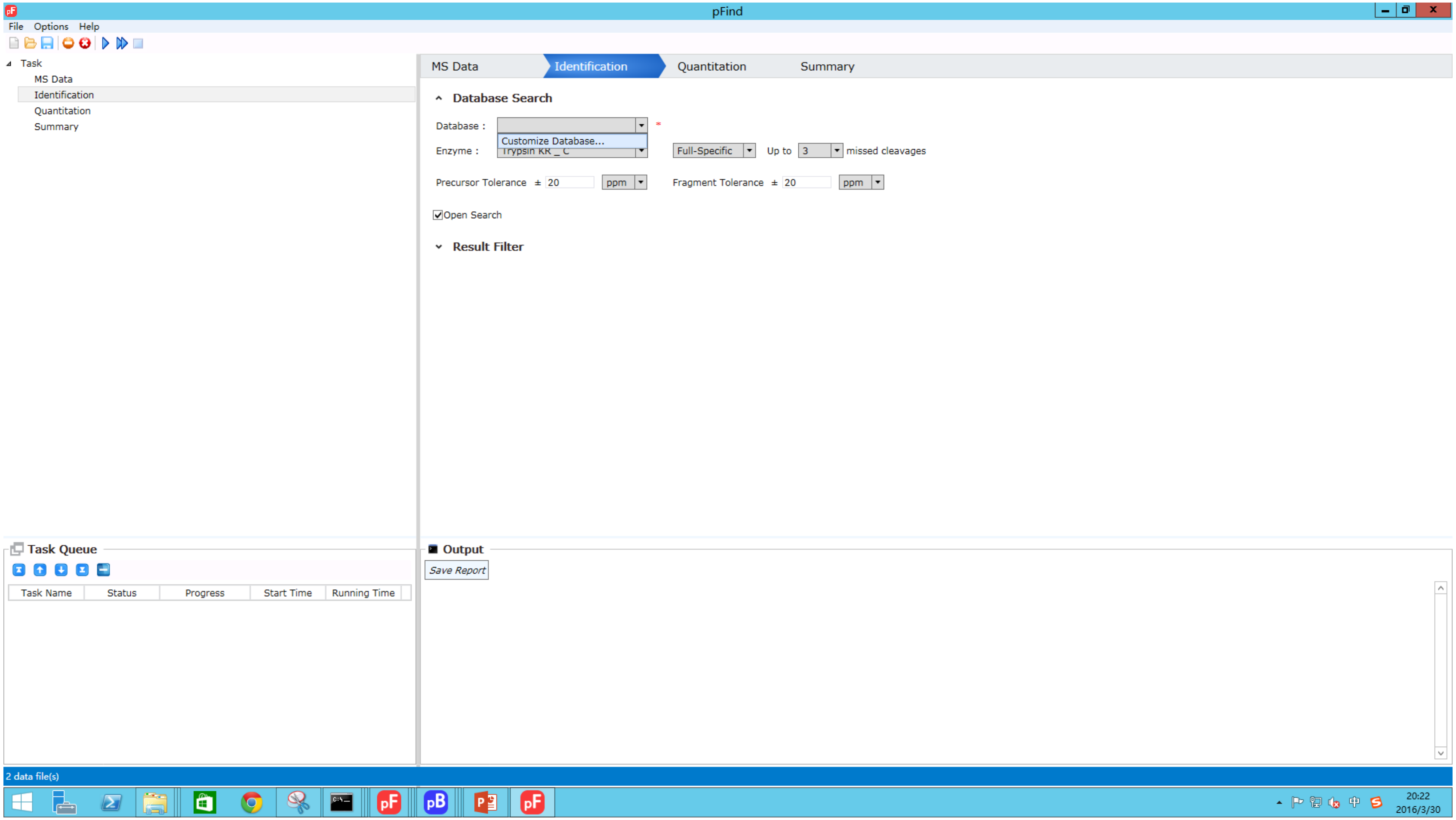
M/Z : 5　Intensity : 1

Precursor Score

Model : Normal　Threshold : -0.5　☑ Mixture Spectra

2. Set search parameters

- **A) Select and import database.**



Database : Customize Database...

Enzyme : Trypsin KR _ C   Full-Specific   Up to 3 missed cleavages

Precursor Tolerance ± 20 ppm   Fragment Tolerance ± 20 ppm

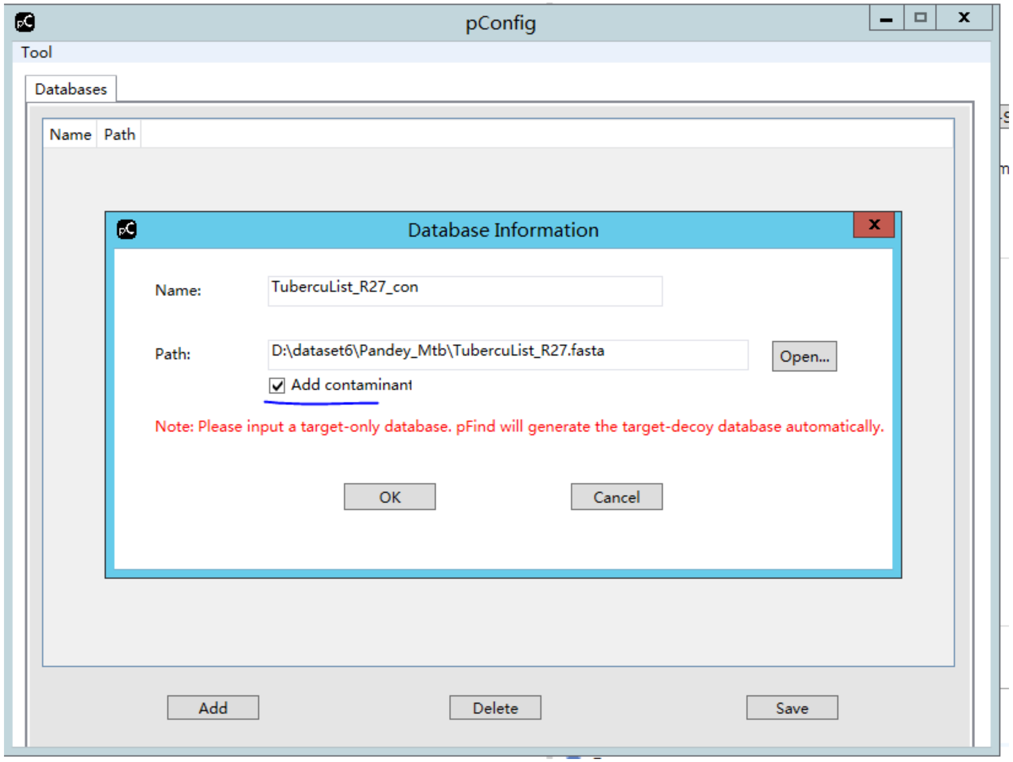
☑ Open Search

2 data file(s)

20:22
2016/3/30

2. Set search parameters

- A) Select and import database.
  - Add contaminated proteins to the database if it doesn’t contain them.



Name: TuberculList_R27_con

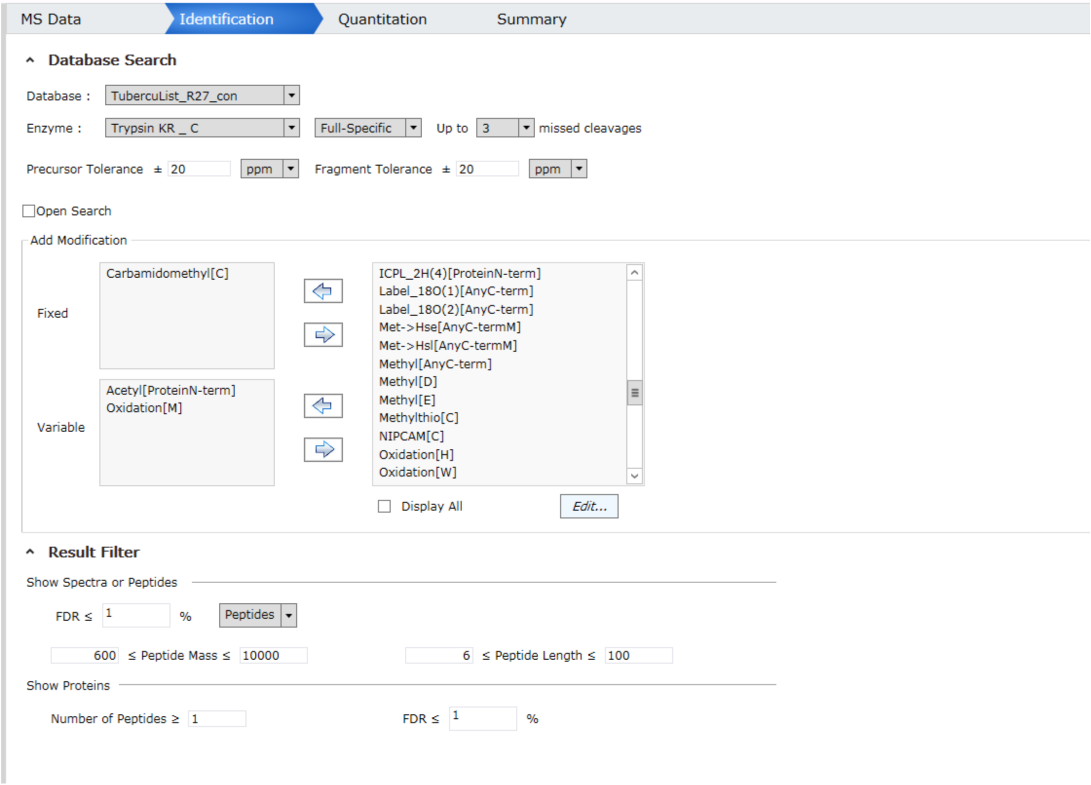
Path: D:\dataset6\Pandey_Mtb\TuberculList_R27.fasta

☑ Add contaminant

Note: Please input a target-only database. pFind will generate the target-decoy database automatically.

2. Set search parameters

- **B) Select the appropriate error range and set the suitable modification and filtering parameters.**



Database Search

Database : TubercuList_R27_con

Enzyme : Trypsin KR _ C; Full-Specific; Up to 3 missed cleavages

Precursor Tolerance ± 20 ppm

Fragment Tolerance ± 20 ppm

☐ Open Search

Add Modification

Fixed: Carbamidomethyl[C]

Variable: Acetyl[ProteinN-term]; Oxidation[M]

ICPL_2H(4)[ProteinN-term]
Label_18O(1)[AnyC-term]
Label_18O(2)[AnyC-term]
Met->Hse[AnyC-termM]
Met->Hsl[AnyC-termM]
Methyl[AnyC-term]
Methyl[D]
Methyl[E]
Methylthio[C]
NIPCAM[C]
Oxidation[H]
Oxidation[W]

☐ Display All

Result Filter

Show Spectra or Peptides

FDR ≤ 1 % Peptides

600 ≤ Peptide Mass ≤ 10000

6 ≤ Peptide Length ≤ 100

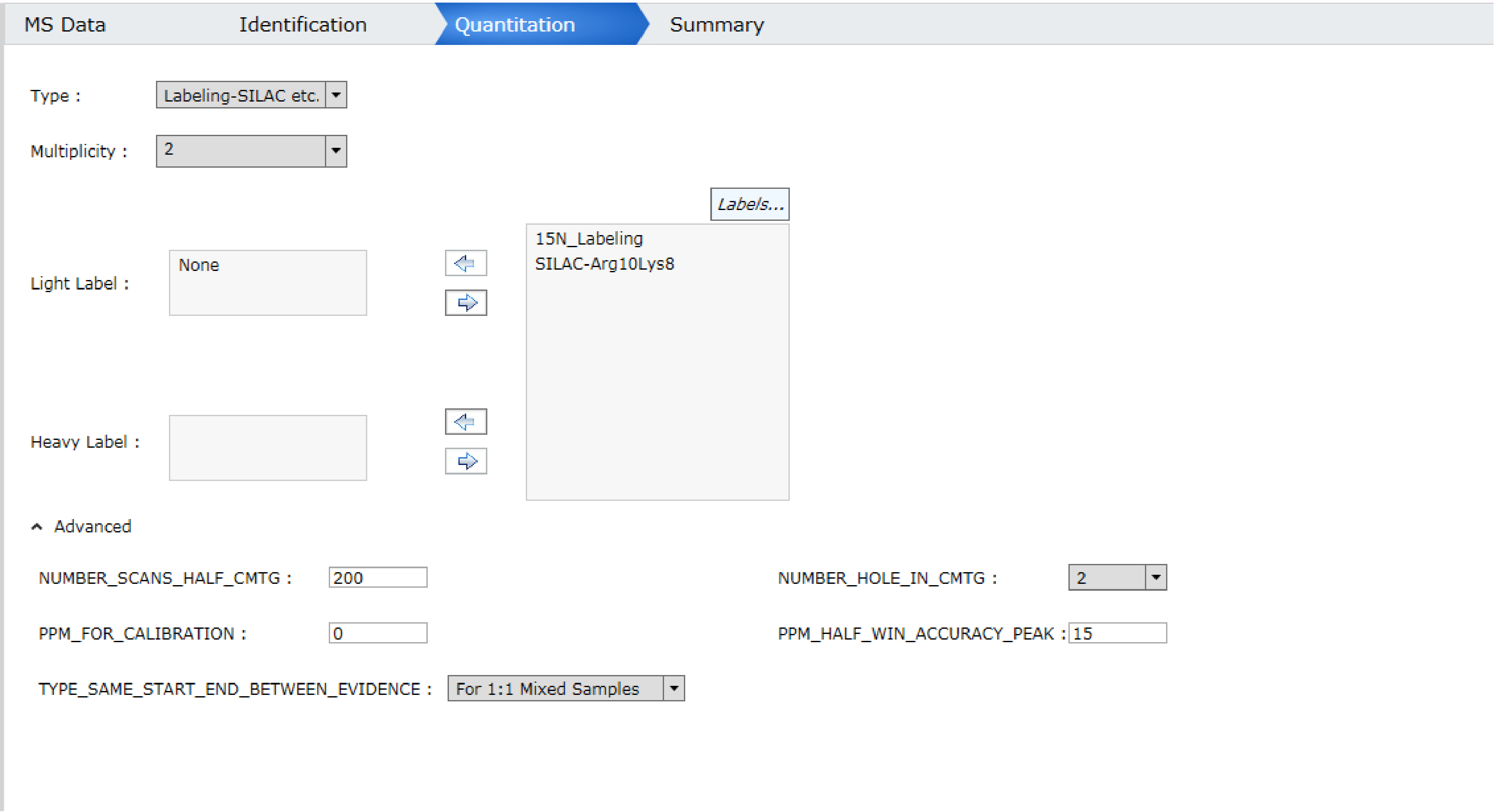
Show Proteins

Number of Peptides ≥ 1

FDR ≤ 1 %

3. Set quantification parameters – MS1

- **SILAC and 15N labeling are supported if necessary**



Type : Labeling-SILAC etc.

Multiplicity : 2

Light Label : None

Heavy Label :

15N_Labeling
SILAC-Arg10Lys8

Advanced

NUMBER_SCANS_HALF_CMTG : 200

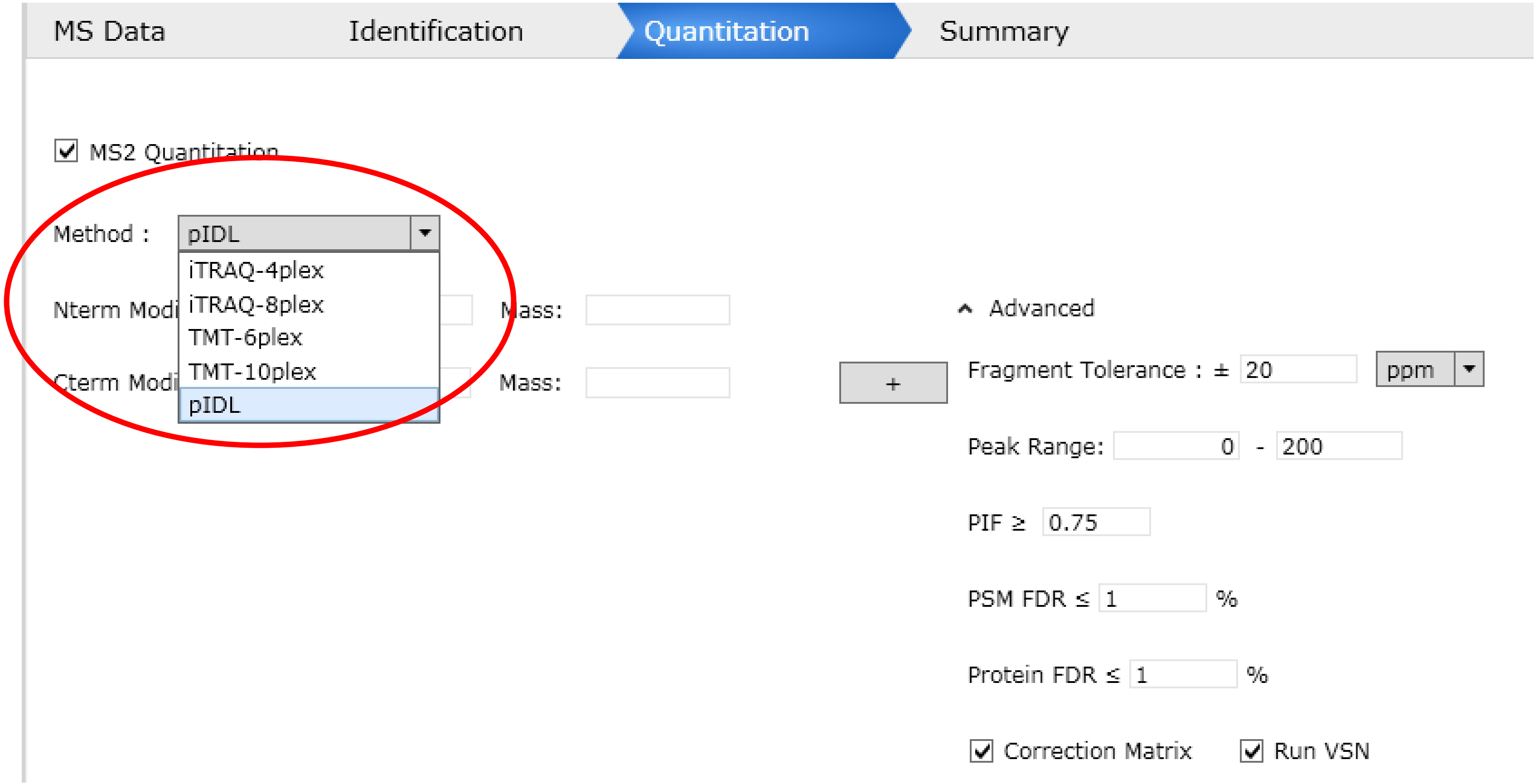
NUMBER_HOLE_IN_CMTG : 2

PPM_FOR_CALIBRATION : 0

PPM_HALF_WIN_ACCURACY_PEAK : 15

TYPE_SAME_START_END_BETWEEN_EVIDENCE : For 1:1 Mixed Samples

3. Set quantification parameters - MS2



☑ MS2 Quantitation

Method : pIDL

Fragment Tolerance : ± 20 ppm

Peak Range: 0 - 200

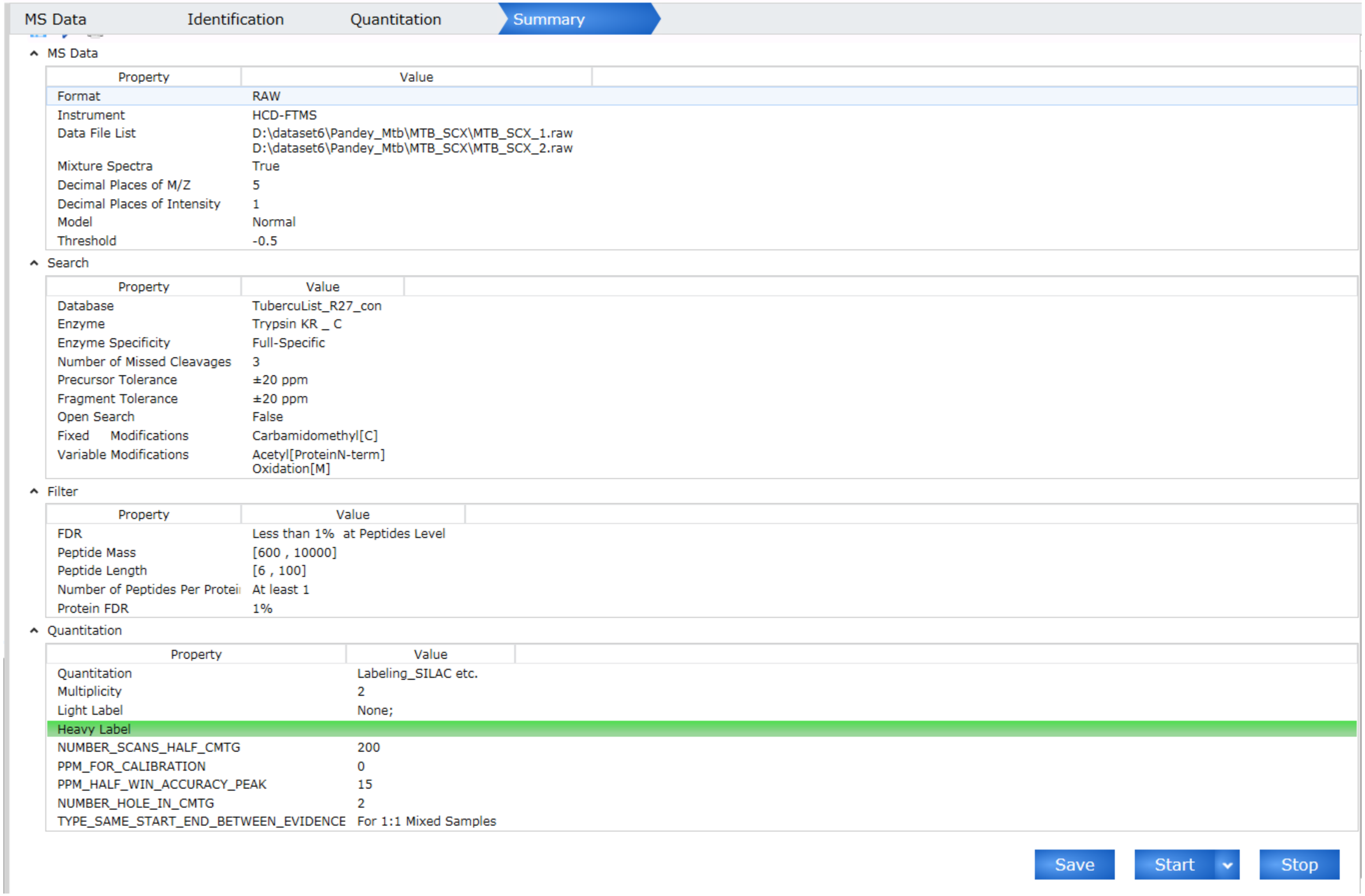
PIF ≥ 0.75

PSM FDR ≤ 1 %

Protein FDR ≤ 1 %

☑ Correction Matrix ☑ Run VSN

4. Check parameters and run tasks



MS Data

Property	Value
Format	RAW
Instrument	HCD-FTMS
Data File List	D:\dataset6\Pandey_Mtb\MTB_SCX\MTB_SCX_1.raw D:\dataset6\Pandey_Mtb\MTB_SCX\MTB_SCX_2.raw
Mixture Spectra	True
Decimal Places of M/Z	5
Decimal Places of Intensity	1
Model	Normal
Threshold	-0.5

Search

Property	Value
Database	TuberculList_R27_con
Enzyme	Trypsin KR _ C
Enzyme Specificity	Full-Specific
Number of Missed Cleavages	3
Precursor Tolerance	±20 ppm
Fragment Tolerance	±20 ppm
Open Search	False
Fixed Modifications	Carbamidomethyl[C]
Variable Modifications	Acetyl[ProteinN-term] Oxidation[M]

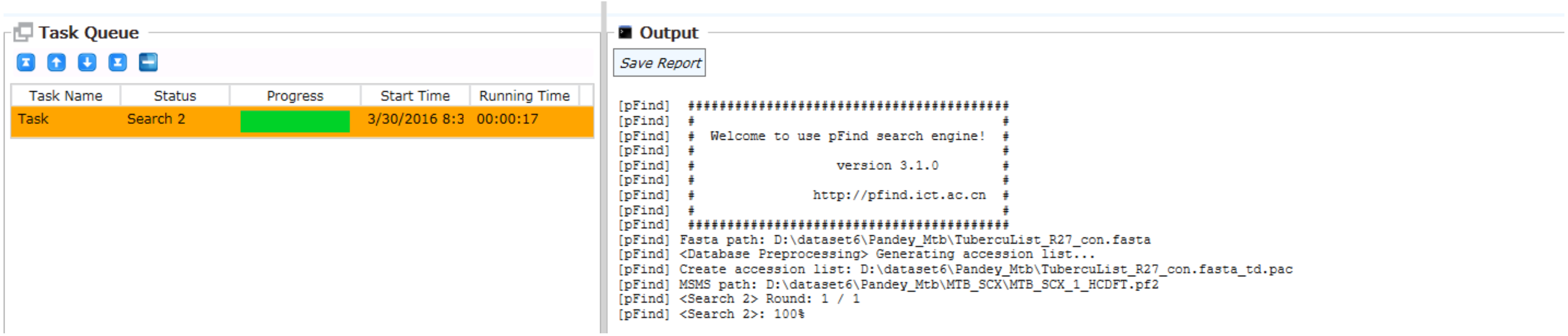
Filter

Property	Value
FDR	Less than 1% at Peptides Level
Peptide Mass	[600 , 10000]
Peptide Length	[6 , 100]
Number of Peptides Per Protei	At least 1
Protein FDR	1%

Quantitation

Property	Value
Quantitation	Labeling_SILAC etc.
Multiplicity	2
Light Label	None;
Heavy Label	
NUMBER_SCANS_HALF_CMTG	200
PPM_FOR_CALIBRATION	0
PPM_HALF_WIN_ACCURACY_PEAK	15
NUMBER_HOLE_IN_CMTG	2
TYPE_SAME_START_END_BETWEEN_EVIDENCE	For 1:1 Mixed Samples

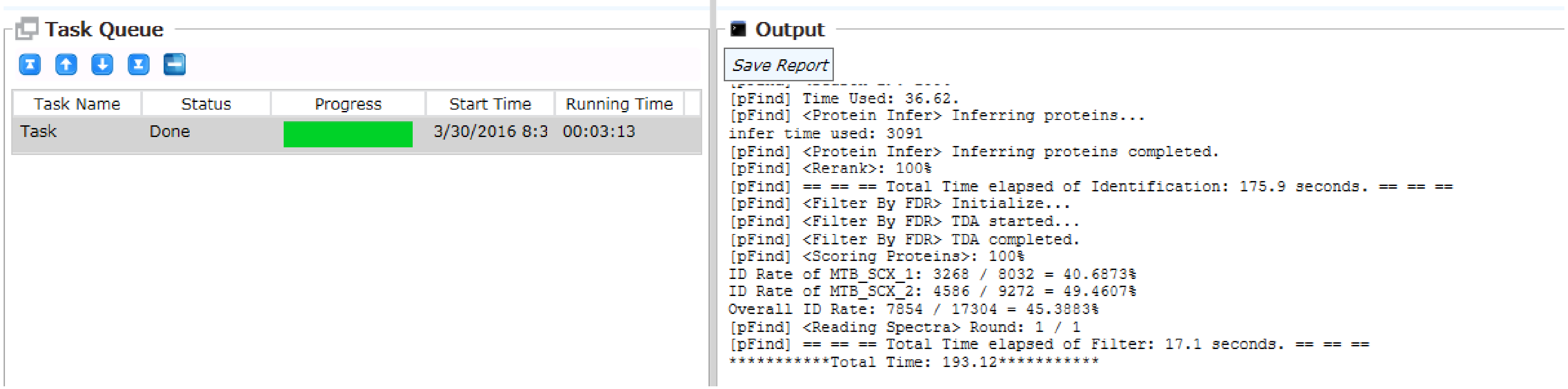
Searching…



Task Name	Status	Progress	Start Time	Running Time
Task	Search 2	🟩	3/30/2016 8:3	00:00:17

```
[pFind]  #########################################
[pFind]  #                                       #
[pFind]  #  Welcome to use pFind search engine!  #
[pFind]  #                                       #
[pFind]  #                 version 3.1.0         #
[pFind]  #                                       #
[pFind]  #               http://pfind.ict.ac.cn  #
[pFind]  #                                       #
[pFind]  #########################################
[pFind] Fasta path: D:\dataset6\Pandey_Mtb\TubercuList_R27_con.fasta
[pFind] <Database Preprocessing> Generating accession list...
[pFind] Create accession list: D:\dataset6\Pandey_Mtb\TubercuList_R27_con.fasta_td.pac
[pFind] MSMS path: D:\dataset6\Pandey_Mtb\MTB_SCX\MTB_SCX_1_HCDFT.pf2
[pFind] <Search 2> Round: 1 / 1
[pFind] <Search 2>: 100%
```

Searching completed



Task Name	Status	Progress	Start Time	Running Time
Task	Done	🟩	3/30/2016 8:3	00:03:13

```
[pFind] Time Used: 36.62.
[pFind] <Protein Infer> Inferring proteins...
infer time used: 3091
[pFind] <Protein Infer> Inferring proteins completed.
[pFind] <Rerank>: 100%
[pFind] == == == Total Time elapsed of Identification: 175.9 seconds. == == ==
[pFind] <Filter By FDR> Initialize...
[pFind] <Filter By FDR> TDA started...
[pFind] <Filter By FDR> TDA completed.
[pFind] <Scoring Proteins>: 100%
ID Rate of MTB_SCX_1: 3268 / 8032 = 40.6873%
ID Rate of MTB_SCX_2: 4586 / 9272 = 49.4607%
Overall ID Rate: 7854 / 17304 = 45.3883%
[pFind] <Reading Spectra> Round: 1 / 1
[pFind] == == == Total Time elapsed of Filter: 17.1 seconds. == == ==
***********Total Time: 193.12***********
```

Contents of search results files

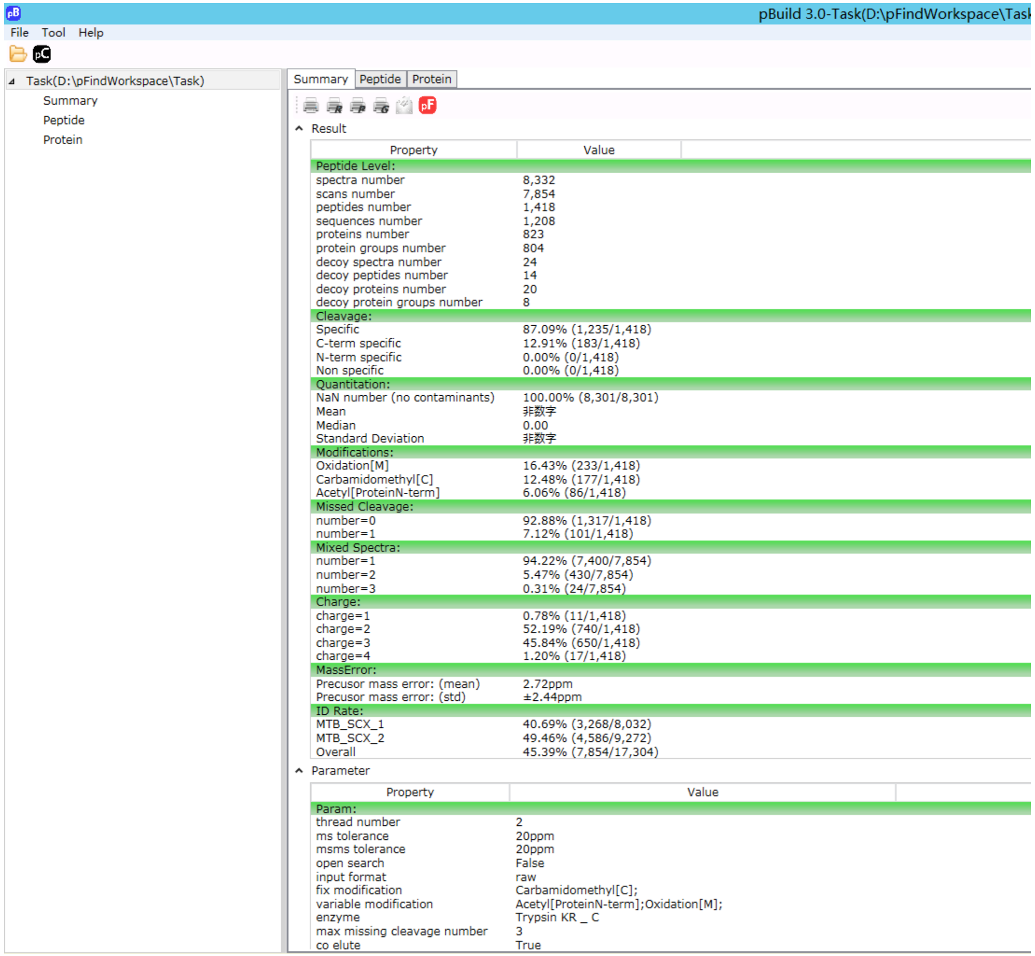
- **pFind.spectra**
  - **Search results of each PSM. Such as protein names, modifications, E-value and so on.**

- **pFind.protein**
  - **Protein group information and identified PSMs.**

- **pFind.summary**
  - **Statistical information of modifications, identification rate, missed cleavage and so on.**

View search results

❖ pBuild: Result statistics

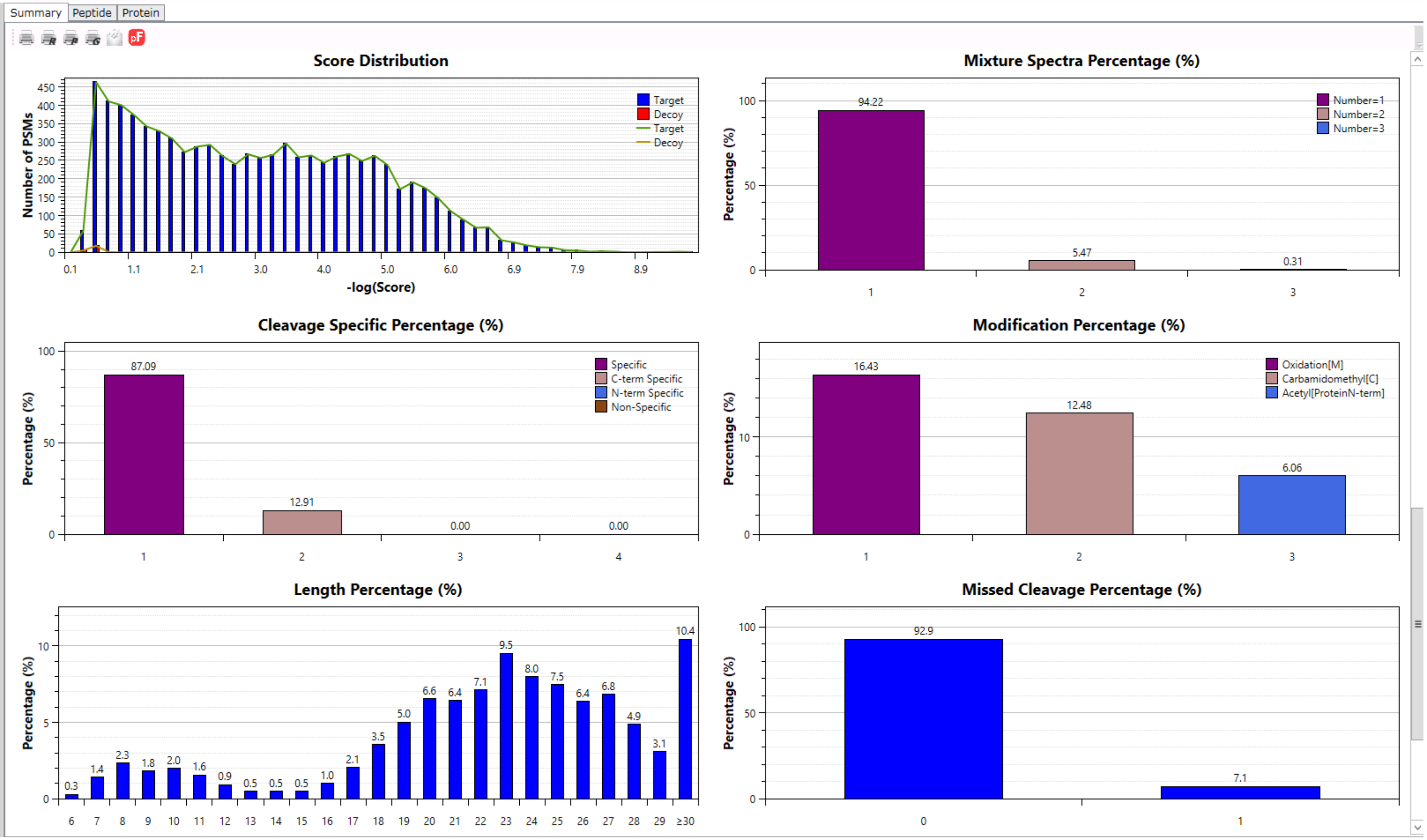


Property	Value
Peptide Level:	
spectra number	8,332
scans number	7,854
peptides number	1,418
sequences number	1,208
proteins number	823
protein groups number	804
decoy spectra number	24
decoy peptides number	14
decoy proteins number	20
decoy protein groups number	8
Cleavage:	
Specific	87.09% (1,235/1,418)
C-term specific	12.91% (183/1,418)
N-term specific	0.00% (0/1,418)
Non specific	0.00% (0/1,418)
Quantitation:	
NaN number (no contaminants)	100.00% (8,301/8,301)
Mean	非数字
Median	0.00
Standard Deviation	非数字
Modifications:	
Oxidation[M]	16.43% (233/1,418)
Carbamidomethyl[C]	12.48% (177/1,418)
Acetyl[ProteinN-term]	6.06% (86/1,418)
Missed Cleavage:	
number=0	92.88% (1,317/1,418)
number=1	7.12% (101/1,418)
Mixed Spectra:	
number=1	94.22% (7,400/7,854)
number=2	5.47% (430/7,854)
number=3	0.31% (24/7,854)
Charge:	
charge=1	0.78% (11/1,418)
charge=2	52.19% (740/1,418)
charge=3	45.84% (650/1,418)
charge=4	1.20% (17/1,418)
MassError:	
Precusor mass error: (mean)	2.72ppm
Precusor mass error: (std)	±2.44ppm
ID Rate:	
MTB_SCX_1	40.69% (3,268/8,032)
MTB_SCX_2	49.46% (4,586/9,272)
Overall	45.39% (7,854/17,304)

Property	Value
Param:	
thread number	2
ms tolerance	20ppm
msms tolerance	20ppm
open search	False
input format	raw
fix modification	Carbamidomethyl[C];
variable modification	Acetyl[ProteinN-term];Oxidation[M];
enzyme	Trypsin KR _ C
max missing cleavage number	3
co elute	True

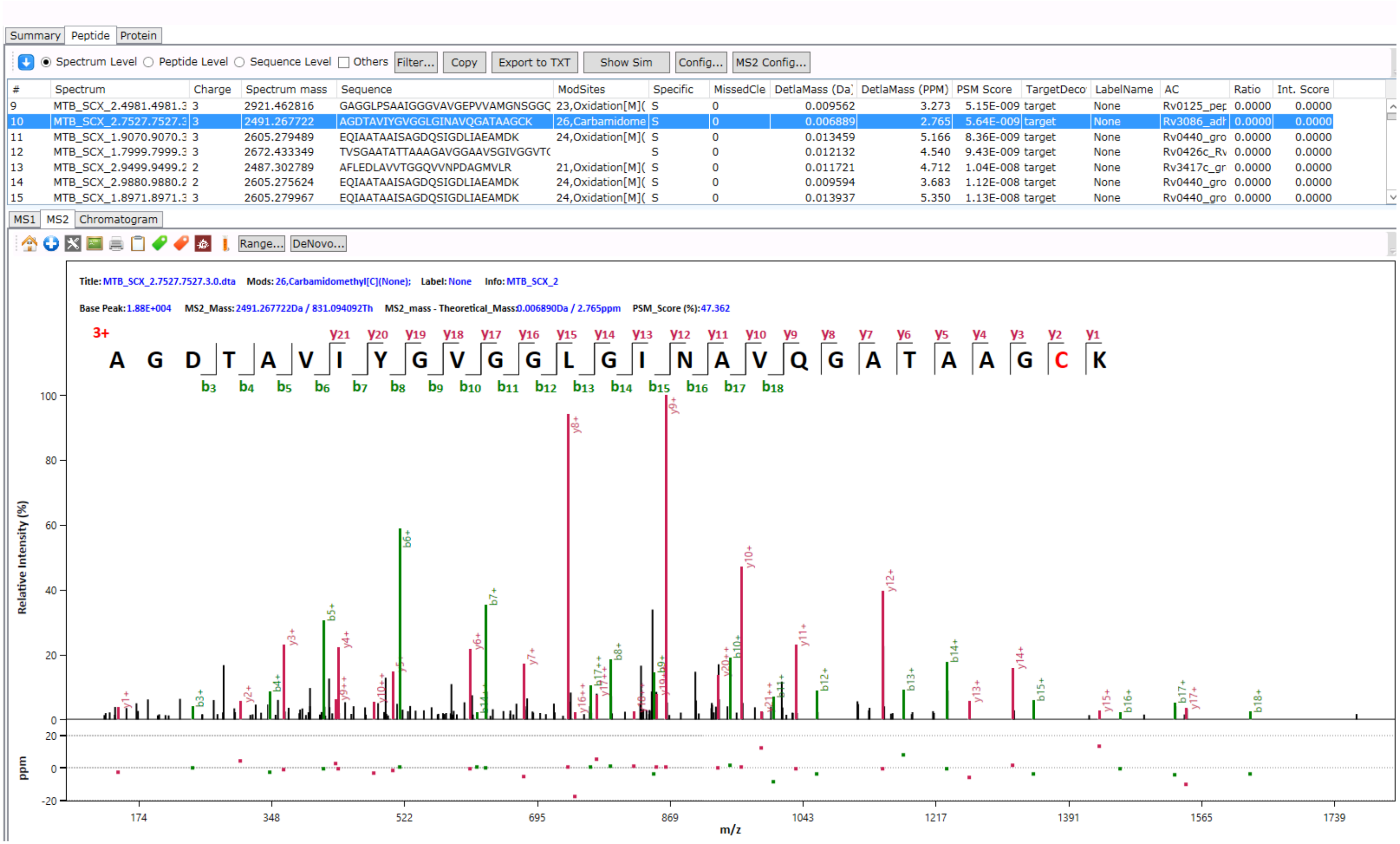
View search results

- **pBuild: The summary panel**



View search results

- **pBuild: Peptide spectrum matching**



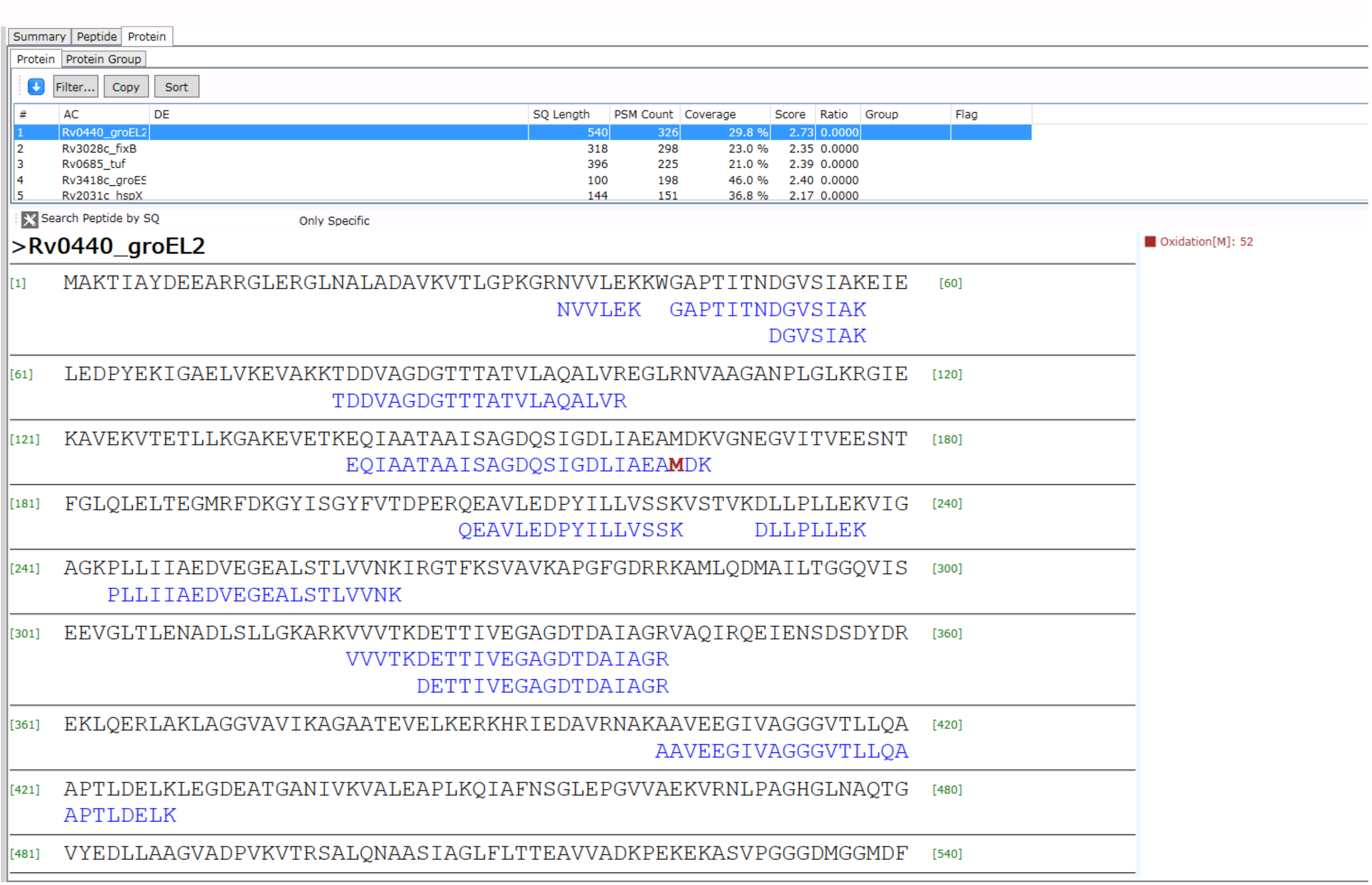
#	Spectrum	Charge	Spectrum mass	Sequence	ModSites	Specific	MissedCle	DetlaMass (Da)	DetlaMass (PPM)	PSM Score	TargetDeco	LabelName	AC	Ratio	Int. Score
9	MTB_SCX_2.4981.4981.3	3	2921.462816	GAGGLPSAAIGGGVAVGEPVVAMGNSGGC	23,Oxidation[M](	S	0	0.009562	3.273	5.15E-009	target	None	Rv0125_pep	0.0000	0.0000
10	MTB_SCX_2.7527.7527.3	3	2491.267722	AGDTAVIYGVGGLGINAVQGATAAGCK	26,Carbamidome	S	0	0.006889	2.765	5.64E-009	target	None	Rv3086_adh	0.0000	0.0000
11	MTB_SCX_1.9070.9070.3	3	2605.279489	EQIAATAAISAGDQSIGDLIAEAMDK	24,Oxidation[M](	S	0	0.013459	5.166	8.36E-009	target	None	Rv0440_gro	0.0000	0.0000
12	MTB_SCX_1.7999.7999.3	3	2672.433349	TVSGAATATTAAAGAVGGAAVSGIVGGVTC		S	0	0.012132	4.540	9.43E-009	target	None	Rv0426c_Rv	0.0000	0.0000
13	MTB_SCX_2.9499.9499.2	2	2487.302789	AFLEDLAVVTGGQVVNPDAGMVLR	21,Oxidation[M](	S	0	0.011721	4.712	1.04E-008	target	None	Rv3417c_gr	0.0000	0.0000
14	MTB_SCX_2.9880.9880.2	2	2605.275624	EQIAATAAISAGDQSIGDLIAEAMDK	24,Oxidation[M](	S	0	0.009594	3.683	1.12E-008	target	None	Rv0440_gro	0.0000	0.0000
15	MTB_SCX_1.8971.8971.3	3	2605.279967	EQIAATAAISAGDQSIGDLIAEAMDK	24,Oxidation[M](	S	0	0.013937	5.350	1.13E-008	target	None	Rv0440_gro	0.0000	0.0000

Title: MTB_SCX_2.7527.7527.3.0.dta **Mods:** 26,Carbamidomethyl[C](None); **Label:** None **Info:** MTB_SCX_2

Base Peak: 1.88E+004 **MS2_Mass:** 2491.267722Da / 831.094092Th **MS2_mass - Theoretical_Mass:** 0.006890Da / 2.765ppm **PSM_Score (%):** 47.362

View search results

❖ pBuild: Protein coverage



#	AC	DE	SQ Length	PSM Count	Coverage	Score	Ratio	Group	Flag
1	Rv0440_groEL2		540	326	29.8 %	2.73	0.0000		
2	Rv3028c_fixB		318	298	23.0 %	2.35	0.0000		
3	Rv0685_tuf		396	225	21.0 %	2.39	0.0000		
4	Rv3418c_groES		100	198	46.0 %	2.40	0.0000		
5	Rv2031c_hspX		144	151	36.8 %	2.17	0.0000		

🟥 Oxidation[M]: 52

```
>Rv0440_groEL2

[1]    MAKTIAYDEEARRGLERGLNALADAVKVTLGPKGRNVVLEKKWGAPTITNDGVSIAKEIE  [60]
                                          NVVLEK  GAPTITNDGVSIAK
                                                         DGVSIAK

[61]   LEDPYEKIGAELVKEVAKKTDDVAGDGTTTATVLAQALVREGLRNVAAGANPLGLKRGIE  [120]
                          TDDVAGDGTTTATVLAQALVR

[121]  KAVEKVTETLLKGAKEVETKEQIAATAAISAGDQSIGDLIAEAMDKVGNEGVITVEESNT  [180]
                           EQIAATAAISAGDQSIGDLIAEAMDK

[181]  FGLQLELTEGMRFDKGYISGYFVTDPERQEAVLEDPYILLVSSKVSTVKDLLPLLEKVIG  [240]
                                   QEAVLEDPYILLVSSK     DLLPLLEK

[241]  AGKPLLIIAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGGQVIS  [300]
          PLLIIAEDVEGEALSTLVVNK

[301]  EEVGLTLENADLSLLGKARKVVVTKDETTIVEGAGDTDAIAGRVAQIRQEIENSDSDYDR  [360]
                           VVVTKDETTIVEGAGDTDAIAGR
                                DETTIVEGAGDTDAIAGR

[361]  EKLQERLAKLAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLLQA  [420]
                                                 AAVEEGIVAGGGVTLLQA

[421]  APTLDELKLEGDEATGANIVKVALEAPLKQIAFNSGLEPGVVAEKVRNLPAGHGLNAQTG  [480]
       APTLDELK

[481]  VYEDLLAAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKEKASVPGGGDMGGMDF  [540]
```

- Thank you for using pFind 3!

- If you have any questions, please contact pfind@ict.ac.cn.