

pFind 3 User Guide

Version 3.1.2



pFind Group

2018.4

❖ Before software installation

❖ Software registration

❖ How to run pFind 3

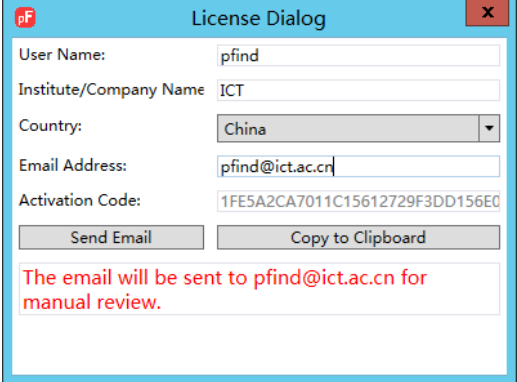
- Create a new search task
- View search results

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.
- ❖ 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.



License Dialog

User Name: pfind

Institute/Company Name: ICT

Country: China

Email Address: pfind@ict.ac.cn

Activation Code: 1FE5A2CA7011C15612729F3DD156E0

Send Email Copy to Clipboard

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

Main interface of pFind 3

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

The screenshot displays the pFind 3 main interface. The window title is "pFind" and the menu bar includes "File", "Options", and "Help". The interface is divided into a left sidebar and a main content area. The sidebar has sections for "pFind", "Start" (with links for "New...", "Open...", "About us", and "Exit"), and "Recent". The main content area shows a "New Features and User Guide" tab. Under the "Overview" section, it states: "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." Below this is the "System Requirements" section, which lists: "Operating System: Windows 7 or higher. Both 32- and 64-bit versions are supported.", "CPU: 2.0GHz or higher", "Memory: 2GB or higher recommended", and "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." A "Settings" dialog box is overlaid on the main interface, showing the "Number of Threads" set to 2, the "Default Data Repository Path" as "D:\pFindWorkspace", and "Available Space on Drive D : 583 G". The dialog box has "OK" and "Cancel" buttons.

Create a new search task

❖ Click File → New

The screenshot displays the pFind software interface. The main window is titled 'pFind' and has a menu bar with 'File', 'Options', and 'Help'. Below the menu bar is a toolbar with various icons. The interface is divided into several panels:

- Task Panel (Left):** Shows a tree view under 'Task' with sub-items: 'MS Data', 'Identification', 'Quantitation', and 'Summary'.
- MS Data Panel (Top):** Contains dropdown menus for 'MS Data Format' (set to 'RAW') and 'MS Instrument' (set to 'HCD-FTMS').
- Data File List Panel (Center):** Features a table with columns 'Files' and 'Size'. To the right of the table are three buttons: 'Add', 'Delete', and 'Clear'.
- Data Extraction Panel (Bottom):** A section labeled 'Data Extraction' with a downward-pointing arrow.
- Task Queue Panel (Bottom Left):** A table with columns 'Task Name', 'Status', 'Progress', 'Start Time', and 'Running Time'. It includes control icons for adding, deleting, and refreshing tasks.
- Output Panel (Bottom Right):** A text area for output, with a 'Save Report' button.

The status bar at the bottom of the window displays 'Ready'.

1. Data import and preprocess.

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

MS Data Identification Quantitation Summary

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

Add

Delete

Clear

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.

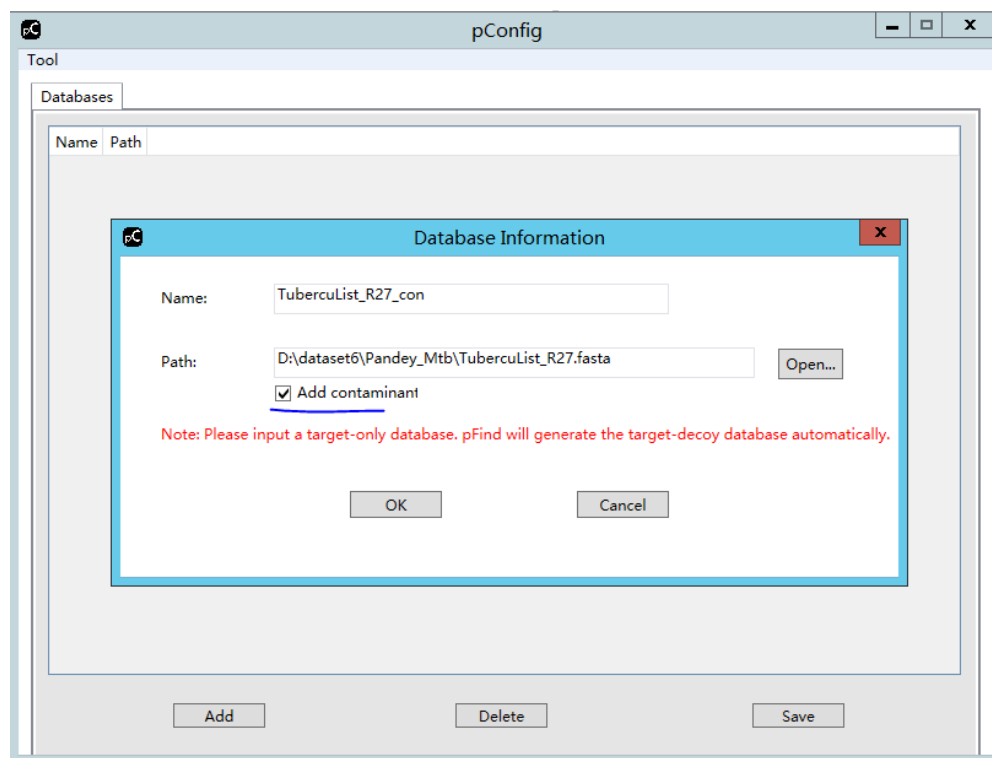
The screenshot displays the pFind software interface. The main window is titled "pFind" and has a menu bar with "File", "Options", and "Help". Below the menu bar is a toolbar with various icons. The interface is divided into several sections:

- Task Queue:** Located at the bottom left, it contains a table with columns for "Task Name", "Status", "Progress", "Start Time", and "Running Time".
- MS Data:** A sidebar on the left with a tree view containing "MS Data", "Identification", "Quantitation", and "Summary".
- Identification Tab:** The active tab in the top navigation bar, showing search parameters:
 - Database Search:** Includes a "Database:" dropdown menu, an "Enzyme:" dropdown menu (set to "Trypsin KK_C"), a "Full-Specific" dropdown, and an "Up to 3 missed cleavages" dropdown.
 - Tolerance Settings:** "Precursor Tolerance ± 20 ppm" and "Fragment Tolerance ± 20 ppm".
 - Open Search:** A checked checkbox.
 - Result Filter:** A section with a downward arrow.
- Output:** A large text area at the bottom right with a "Save Report" button.

The Windows taskbar at the bottom shows the system tray with the date and time "20:22 2016/3/30" and several application icons, including pFind, pB, p3, and pF.

2. Set search parameters

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



2. Set search parameters

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

The screenshot displays the 'Identification' tab of the pFind Studio software. The interface is divided into several sections:

- Database Search:** Includes a dropdown for 'Database' (set to 'TubercuList_R27_con'), 'Enzyme' (set to 'Trypsin KR_C'), 'Full-Specific' dropdown, 'Up to' (set to '3'), 'missed cleavages', 'Precursor Tolerance' (set to '± 20 ppm'), and 'Fragment Tolerance' (set to '± 20 ppm').
- Add Modification:** Features two columns: 'Fixed' and 'Variable'. The 'Fixed' column contains 'Carbamidomethyl[C]'. The 'Variable' column contains 'Acetyl[ProteinN-term]' and 'Oxidation[M]'. A central list of modifications includes '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]', and 'Oxidation[W]'. There are arrow buttons for moving items between columns and a 'Display All' checkbox.
- Result Filter:** Contains filters for 'Show Spectra or Peptides' (set to 'Peptides'), 'FDR ≤ 1 %', '600 ≤ Peptide Mass ≤ 10000', '6 ≤ Peptide Length ≤ 100', 'Show Proteins', 'Number of Peptides ≥ 1', and 'FDR ≤ 1 %'.

3. Set quantification parameters – MS1

❖ SILAC and 15N labeling are supported if necessary

MS Data Identification **Quantitation** Summary

Type : ▾

Multiplicity : ▾

Light Label :

Heavy Label :

15N_Labeling
SILAC-Arg10Lys8

^ Advanced

NUMBER_SCANS_HALF_CMTG : NUMBER_HOLE_IN_CMTG : ▾

PPM_FOR_CALIBRATION : PPM_HALF_WIN_ACCURACY_PEAK :

TYPE_SAME_START_END_BETWEEN_EVIDENCE : ▾

3. Set quantification parameters - MS2

MS Data Identification **Quantitation** Summary

MS2 Quantitation

Method : pIDL ▼

Nterm Mod iTRAQ-4plex Mass:

Cterm Mod iTRAQ-8plex Mass:

TMT-6plex Mass:

TMT-10plex Mass:

pIDL Mass:

^ Advanced

+ Fragment Tolerance : ± ppm ▼

Peak Range: -

PIF ≥

PSM FDR ≤ %

Protein FDR ≤ %

Correction Matrix Run VSN

4. Check parameters and run tasks

MS Data Identification Quantitation **Summary**

^ 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	TubercuList_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 Protein	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

Save Start ▾ Stop

Searching...

Task Queue



Task Name	Status	Progress	Start Time	Running Time
Task	Search 2	<div style="width: 100%; height: 10px; background-color: green;"></div>	3/30/2016 8:3	00:00:17

Output

[Save Report](#)

```
[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\TubercouList_R27_con.fasta  
[pFind] <Database Preprocessing> Generating accession list...  
[pFind] Create accession list: D:\dataset6\Pandey_Mtb\TubercouList_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 Queue



Task Name	Status	Progress	Start Time	Running Time
Task	Done	<div style="width: 100%; height: 10px; background-color: green;"></div>	3/30/2016 8:3	00:03:13

Output

[Save Report](#)

```
[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

pBuild 3.0-Task(D:\pFindWorkspace\Task)

File Tool Help

Task(D:\pFindWorkspace\Task)

- Summary
- Peptide
- Protein

Result

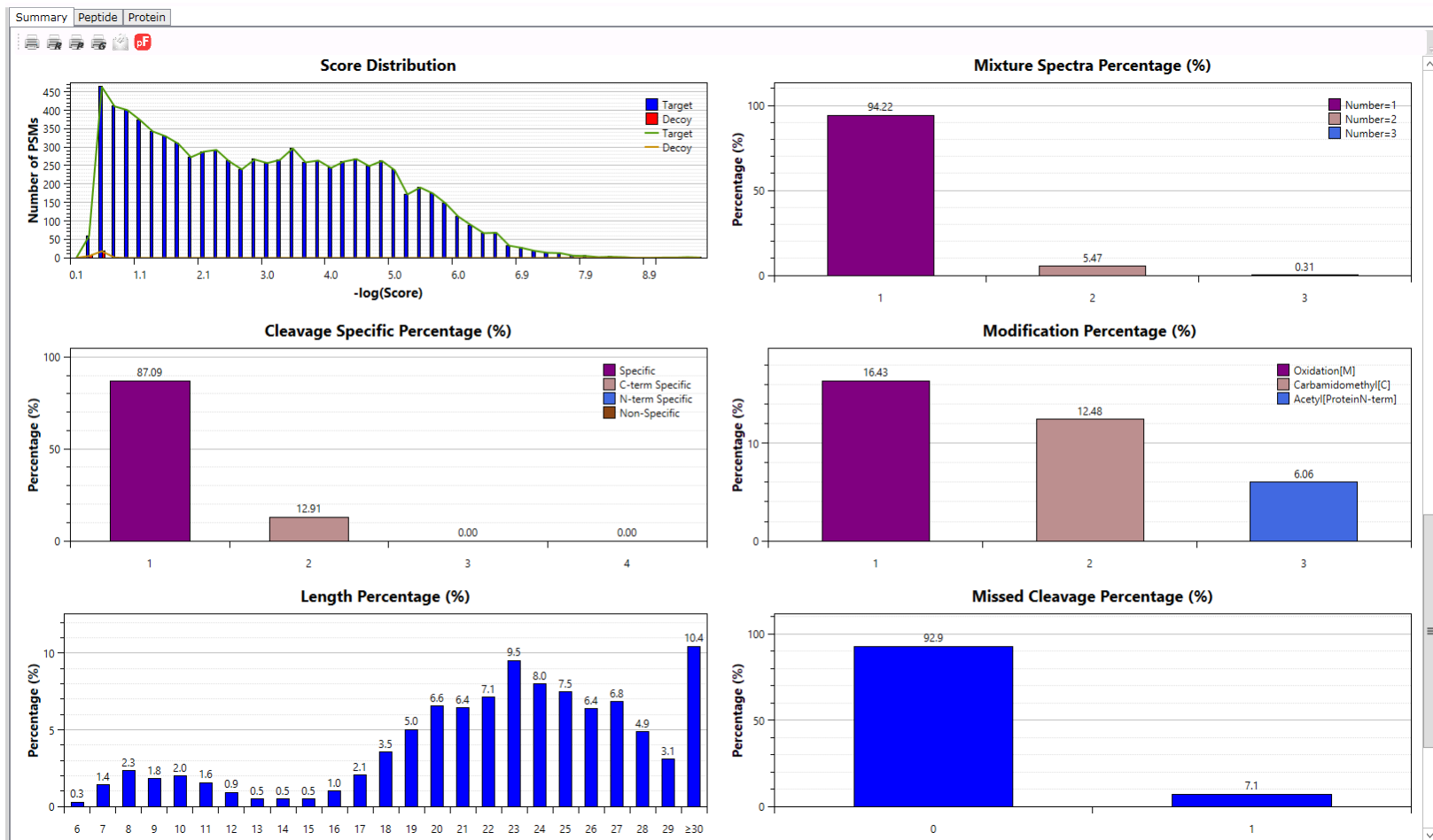
Property	Value
Peptide Level:	
spectra number	8,332
scans number	7,854
peptides number	1,418
sequences number	1,208
proteins number	823
protein groups number	804
decoy spectra number	24
decoy peptides number	14
decoy proteins number	20
decoy protein groups number	8
Cleavage:	
Specific	87.09% (1,235/1,418)
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:	
Precursor mass error: (mean)	2.72ppm
Precursor mass error: (std)	±2.44ppm
ID Rate:	
MTB_SCX_1	40.69% (3,269/8,032)
MTB_SCX_2	49.46% (4,586/9,272)
Overall	45.39% (7,854/17,304)

Parameter

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



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		

#	Peptide	Score	Ratio
[1]	MAKTIAYDEEARRGLERGLNALADAVKVTLGPKGRNVVLEKKGWAPTITNDGVSIAKEIE NVVLEK GAPTITNDGVSIAK DGVSIK	[60]	
[61]	LEDPYEKIGAEELVKEVAKKTTDDVAGDGTATVLAQALVREGLRNVAAGANPLGLKRGIE TDDVAGDGTATVLAQALVR	[120]	
[121]	KAVEKVTETLLKGAKEVETKEQIAATAAISAGDQSIGDLIAEAMDKVNEGVIITVEESNT EQIAATAAISAGDQSIGDLIAEAMDK	[180]	
[181]	FGLQLELTEGMRFDKGYISGYFVTDPERQEAVLEDPYILLVSSKSTVKDLLPILLEKVVIG QEAVLEDPYILLVSSK DLLPILLEK	[240]	
[241]	AGKPLLI IAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGGQVIS PLLI IAEDVEGEALSTLVVNK	[300]	
[301]	EEVGLTLENADLSLLGKARKVVVTKDETTIVEGAGDTDAIAGRVAQIRQEIENSDSDYDR VVVTKDETTIVEGAGDTDAIAGR DETTIVEGAGDTDAIAGR	[360]	
[361]	EKLQERLAKLAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLTLLQA AAVEEGIVAGGGVTLTLLQA	[420]	
[421]	APTLDELKLEGDEATGANIVKVALEAPLKQIAFNSSGLEPGVVAEKVRNLPAGHGLNAQTG APTLDELK	[480]	
[481]	VYEDLLAAGVADPVKVTRRSALQNAAS IAGLFLTTEAVVADKPEKEKASVPGGGDMGGMDF	[540]	



❖ **Thank you for using pFind 3!**

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