

User Manual for pChem

Version 1.0

Contents

1. Requirements	3
2. Download	4-5
3. Configuration	6-9
4. Run	10
5. Output	11-15
6. Supporting Protocol 1: Protein sequence database	16
7. Supporting Protocol 2: MSconvert	17-19
8. Supporting protocol 3: ChemCalc	20

1. Requirements

1) Computing system

pChem search requires a computer with recommended configuration as follows:

- Microsoft Windows 64-bit
- Intel Core i7/i9/Xeon Processor
- > 32GB of RAM or more

Note: pChem v1.0 is NOT supported by non-Windows operating systems (incl. MacOS, Linux and so on).

2) MS Data

Data dependent acquisition (DDA) with BOTH MS1 and MS/MS spectra recorded in the <u>High-Resolution</u> mode

Note: 1) For automatic performance assessment of chemoproteomic probes, it is recommended to acquire MS data from probe-labeled samples with isotope-coding. 2) MS data from non-isotope-labeled samples can also be processed by pChem, but the search result may be subjected to manual inspection.

2. Download

 pChem can be freely downloaded from the website: http://pfind.org/software/pChem/index.html

pFind Studio: a computational solution for mass spectrometry-based proteomics

Home Members Publications Applications Software Links About us

pChem

Introduction - Cite us - Downloads

Introduction

Chemical probe coupled with mass spectrometry (MS)-based proteomics, herein termed chemoproteomics, offers versatile tools to globally profile protein features and to systematically interrogate the mode of action of small molecules in a native biological system. Nonetheless, development of an efficient and selective probe for chemoproteomics can still be challenging. Besides, it is also difficult to unbiasedly assess its chemoselectivity at a proteome-wide scale. Here we present pChem, a modification-centric blind search and summarization tool to provide a pipeline for rapid and unbiased assessing of the performance of ABPP and metabolic labeling probes. This pipeline starts experimentally by isotopic coding of PDMs, which can be automatically recognized, paired, and accurately reported by pChem, further allowing users to score the profiling efficiency, modification-homogeneity and proteome-wide residue selectivity of a chemoproteomic probe.



Cite us

pChem: a modification-centric assessment tool for performance of chemoproteomic probes. Ji-Xiang He, Zheng-Cong Fei, Fu-Chu He, Si-Min He, Hao Chi, Jing Yang. Under review

Downloads

pChem version 1.0 is currently free to use. click to download.

For source code, please refer to github.

For detailed usage, please refer to user guide.

2) click "click to download" to get the zipped software.

📑 pChem.zip

 Un-zip the "pChem.zip" package into a specified file folder (e.g., Local disk C).



3. Configuration

1) Double click *pChem* - p^{Chem} to open the main folder.

📕 🛃 📜 👻 pChem							
File Home Share View							
\leftarrow \rightarrow \checkmark \uparrow \blacksquare > This PC > Local Disk (C:) > pChem							
📌 Quick access		Name ^	Date modified	Туре			
Desktop	*	📕 bin	11/30/2021 9:14 AM	File folder			
		📜 Protein_seq_database	11/30/2021 9:14 AM	File folder			
	~	results	11/30/2021 9:14 AM	File folder			
Documents	×	pChem.cfg	11/30/2021 9:14 AM	CFG File			
Pictures	*	Pchem.exe	11/30/2021 9:15 AM	Application			
🧢 This PC	*						

- Open configuration file "pChem.cfg" using a text editor, e.g., Microsoft Notepad or Notepad++ (<u>https://notepad-plus.en.softonic.com/</u>).
- 3) Setting "pChem.cfg".

```
# If isotope coding is adopted to facilitate the discovery of unknown modifications (True or False)
     isotope labeling=True
    #.Path.to.the.output.file
     output_path=D:\pchem\pChem_new\results
    #.Path.to.the.protein.sequence.database.
8
    fasta_path=D:\pChem\pChem_new\Protein_seg_database\Homo_sapiens_uniprot_canonical_20395_entries_20210516.fasta
    # · Format · of · MS · data, · RAW · or · MZML ·
    msmstype=RAW
    #·The·number·and·path·of·MS·data
13
     msmsnum=1
14
15
    msmspathl=D:\pchem\pChem_data\QE_Plus_YangJing_FL_ALK_50per_20170531.raw
18
    #.Type.of.MS.dissociation.method.
19
    activation_type=HCD-FTMS
20
21
    #·Usage.of.open.search.(True/.False),.against.Unimod,.the.common.modification.can.be.set.if.not.
22
    open_flag=False
    common_modification_number=2..
    common_modification_list=Carbamidomethyl[C];Oxidation[M];
24
25
26
27
    #.Mass.tolerance.of.the.mass.shift.between.light.isotope.and.heavy.isotope
28
    mass_of_diff_diff=6.020132
29
    #.Isotopic.mass.difference.within.empirically.defined.tolerance(Da)
    mass diff diff range=0.005
34
    #·Mass·range.of.unknown.modification.(Da)
35
    min_mass_modification=200
36
    max_mass_modification=1000
     #. Isotopic pairs of mass shifts with PSMs less than X% of that of overall PDMs were neglected
    filter frequency=5
40
41
    #. If . consider. the . N-side. or . C-side. for . amino. acid. localization. (True. or . False)
42
    side_position=True
43
44
    #.P-value.threshold.enabling.confident.amino.acid.localization.
45
    p_value_threshold=0.001
46
    #. If . report . the . statistical . information . (True . or . False)
47
48 report_statistics=False
```

General Note 1:

For the first-time users, custom settings are required for (1-5), (8) default settings can be adopted for (6), (7), (9-(14)).

General Note 2:

All parameters (shown in red below) are case sensitive.

General Note 3:

The blank space should be avoided.

① # If isotope coding is adopted to facilitate the discovery of unknown

modifications (True or False)

Isotope_labeling=True

illustration: default

Note: Choose 'False', if pChem is adopted to search endogenous modifications from probe-free and/or label-free protein samples,

2 # Path to the output file

output_path=C:\pChem\results

Note: If the output file folder does not exist, an error will be reported.

③ # Path to the protein sequence database

fasta_path=C:\pChem\Protein_seq_database\Homo_sapiens_uniprot_can onical_20395_entries_20210516.fasta

Note: The protein *.fasta database databases of several commonly used species (*e.g.*, *home sapiens*) are included in the subfolder (named as Protein_seq_database) of pChem. Note that the databases of other species can be downloaded from Uniprot as described in **Supporting Protocol 1**.

PC > Local Disk (C:) > pChem > Protein_seq_database

	Name	Date modified	Туре
	Arabidopsis_thaliana_uniprot_canonical_16043_entries_20210516.fasta	5/17/2021 12:07 PM	FASTA File
	Caenorhabditis_elegans_uniprot_canonical_4226_entries_20210516.fasta	5/17/2021 12:23 PM	FASTA File
	Drosophila_melanogaster_uniprot_canonical_3632_entries_20210516.fasta	5/16/2021 11:44 PM	FASTA File
	Escherichia_coli_uniprot_canonical_4518_entries_20210516.fasta	5/17/2021 12:15 PM	FASTA File
6	Homo_sapiens_uniprot_canonical_20395_entries_20210516.fasta	6/4/2021 9:23 PM	FASTA File
	Mus_musculus_uniprot_canonical_17073_entries_20210516.fasta	5/17/2021 12:18 PM	FASTA File
	Pseudomonas_syringae_uniprot_canonical_5431_entries_20210516.fasta	7/27/2021 9:56 PM	FASTA File
	Rattus_norvegicus_uniprot_canonical_8126_entries_20210516.fasta	5/17/2021 12:22 PM	FASTA File

Format of MS data (RAW or MZML)

msmstype=RAW

Note: Non-Thermo MS data need to be converted into mzML files before pChem search. The users can refer to **Supporting Protocol 2**.

(5) # The number and path of MS data

```
msmsnum=N
msmspath1=X:\XXX\XXX.raw
msmspath2=X:\XXX\XXX.raw
```

.....

msmspathN=X:\XXX\XXX.raw

Note: The suffix of MS data files MUST be input.

Example:

msmsnum=1

msmspath1=D:\pChem\IPM\QE_Plus_YJ_FL_50per_20190501_F1_R1.raw



6 # Type of MS dissociation method

activation_type=HCD-FTMS

illustration: default

Note: 1) Users can adopt this default setting if their MS data is generated by TOF (Time-of-flight) instruments implementing CID (Collision-induced dissociation)-type of fragmentation (e.g., SCIEX 5600, SCIEX 6600, Bruker TimsTOF); 2) pChem v1.0 can NOT support MS data generated under electron-transfer dissociation ETD, electron-transfer/higher-energy collision dissociation EThcD, and the likes.

⑦ # Usage of open search (True/ False) against Unimod, the common

modification can be set if not open_flag=False common_modification_number=2 common_modification_list=Carbamidomethyl[C];Oxidation[M]; illustration: default Note: The names of common modifications should be the same as those appeared in <u>Unimod</u> database.

8 # Mass tolerance of the mass shift between light isotope and heavy isotope

mass_of_diff_diff=6.020132

Note: This default value is calculated based on the isotopic mass shift between six heavy and light carbons encoded in probe-derived modifications (PDMs). Users can set any other values based on their different isotope labeling strategies.

Troubleshooting: One needs to confirm this value being correctly input.

(9) # Isotopic mass difference within empirically defined tolerance (Da)

mass_diff_diff_range=0.005 illustration: default **Troubleshooting:** If the pChem search mis-identified the targeted PDMs or even report nothing, one might want to loose the defined mass tolerance (e.g., 0.01Da).

Mass range of unknown modification (Da)

min_mass_modification=200 max_mass_modification=1000 illustration: default **Note:** The PDMs generated from the use of bioorthogonal cleavable linkers typically possess masses higher than 200 Da and less than 1000Da.

(1) # Isotopic pairs of mass shifts with PSMs less than X% of that of overall PDMs were neglected

filter_frequency=5

illustration: default

Note: This parameter can be set as 0 if one wants to retrieve all PDMs including those with just a few PSMs.

- # If consider the N- or C-termini for amino acid localization (True or False) side_position=True illustration: default
- # P-value threshold enabling confident amino acid localization p_value_ threshold=0.001 illustration: default
- # if report the statistical information (True or False) report_statistics=False illustration: default

4. Run

Once all parameters have been set, double click "*pChem.exe*" **Phase** press any key to continue" means that program runs to completion.

Note: pChem search will generate several intermediate files in the main folder. do NOT open those files during program running.

🔀 Select Windows PowerShell		-		Х
######################################				
0xb97040 Init OK Load MGF OK Load pChem Result OK ====================================	cting			
Calculate the ion of PFIND_DELTA_252 100% Calculate the ion of PFIND_DELTA_258 100% Calculate the ion of PFIND_DELTA_268 100% Calculate the ion of PFIND_DELTA_274 100% [probe evaluation metric] profiling efficiency: 96.55 PDM homogeneity: 92.53 residue selectivity: 92.53 Find search cost time (s): 930.5 postpieted search cost time (s): 629.3 Press any key to continue	4111/4111 [00:50<00:00, 3237/3237 [00:33<00:00, 352/352 [00:03<00:00, 293/293 [00:02<00:00,	80. 611 97. 593 103. 723 112. 973	it/s] it/s] it/s] it/s]	

5. Output

- 2) Double click "reporting summary".



3) There are three major output documents.



Note: Users are recommended to copy these output documents and paste into another file. Otherwise, they can be covered by those generated from the next search event.

1 pChem.summary

pChem.summary is a tab-delimited text file contains the details of every PDM.

						5125
1 PFIND_DELTA_2	52 252.122339	C 0.988		13876	7368 6508	
2 PFIND_DELTA_2	68 268.116411	C 0.487	M(0.291); N-SIDE(0.212);	1578	872 706	302.104737, 301.103528, 320.113184

PDM: Probe-derived modifications

#PSM: The number of PSMs corresponding to modified peptides identified by targeted search

#PSM L|H: The number of PSMs assigning to light and heavy bearing the corresponding PDM, respectively

Note: For data from on-isotope-labeled samples, this information will NOT be shown.

Top1 site | Top1 Probability: The amino acid most likely to be modified with the corresponding localization probability.

Others: Other amino acid sites that may also be labeled by probes and their corresponding localization probability values

DFLs: Diagnostic fragment losses **Note:** For data from on-isotope-labeled samples, DFLs will NOT be provided.

② Heat_map.pdf



Horizontal coordinate: The Amass of each PDM

Longitudinal coordinate: The types of amino acids

Color gradient: The localization probability that the modification occurs at each potential site.

Note: 1) Those amino acids with p-value higher than 0.001 are considered mis-localized sites. As such, their localization probability values are defined to be null. 2) For data generated from non-isotope-labeled samples, heatmap will NOT be provided.

③ Radar.pdf



Radar.pdf contains a radar plot whose radial axes correspond to the three scores as indicated.

Profiling efficiency (%) that evaluates whether a probe enables the efficient identification of modified peptides for chemoproteomics.

Modification homogeneity (%) that evaluates whether a probe forms a uniform modification.

Residue selectivity (%) that evaluates whether a probe selectively targets specific amino acid:

Note: 1) *PDM homogeneity* and *Residue selectivity* are calculated based on the blind search results, while *Profiling efficiency* is calculated according to restricted search; 2) For data generated from on-isotope-labeled samples, radar plot will NOT be provided.

④ PSM-level results

C:\pChem\results\source\blind\pFind-Filtered.spectra

This file contains the PSM-level information regarding all possible modifications on peptides.

1 File_Name	Scan_No	Exp.MH+	Charge Q-value	Sequence C	c.MH+	Mass_Shift	Raw_Score	Final_Sc	ne Modification	pecificity Proteins	Position: Label	Target/D	ec Miss Clv Sib Avg F	rag.Mass.Sh	Others	Accurate modification mass
2 QE_Plus_Yangling_TCI	34997	3323.517	3	0 NUNDQVLFIDQGNRPLFEDMTDSDCR	323.503	0.013332	32.11487	7.05E	2 20.0xidation(M) 25.PFIND_DELTA_252.12	3 sp)Q14116(IL18.H	UF49.R.D/ 1 CIPFING	target 1.	1	4,89814	96	252 120478
3 QE_Plus_VJ_FL_50per_	19072	3450.497	3	0 SSSQPSSCCSDPSKPGGNVEGATQSLAEC	50.502	-0.00423	33.12794	1.51E	0 8.PRIND_DELTA_309 15:30.Oxidation(M):	3 spi013501/SOST	1 281 K.K/ 1 PFIND	Détarget	1	-3.22096	36	309.132429
4 QE_Plus_Yangling_TCI	34275	3198.554	3	D ASSDLSIASSEEDKLSQNACILESVSEK	98.549	0.005356	26.01278	2.79E	0 20.PFIND_DELTA_258.15	3 spiO9H2G2ISUK, H	U 338.R.T. 1 PFIND.	DEtarget	1	-0.65771	36	258 142982
5 OE Plus Yanaling TCI	34834	3458.562	3	0 CPEALFOPSFLGMESCGIHETTFNSIMK	8458.55	0.011821	33.4101	2.95E	0 1 PRIND, DELTA 252 12:13 Oxidation(M):1	3 spiP60709(ACTB)	U 256 R.C. 1IPFIND	DEtarget	0	2.01364	100	252 118449
6 OE Plus Yangling TC	34272	3192 535	3	D ASSOLSIASSEEDKLSONACILESVSEK	92.529	0.005925	25,48116	2.96E	0 20 PFIND DELTA 252 13	3 spiO9H2G2ISLK H	UT338 R.T/ 1PFIND	DEtarget	1	-1.67882	36	252 123574
7 OE Plus Yanaling TCI	25059	2666.205	3	0 VVMALGDYMGASCHACIGGTNVR	566 192	0.01266	30,79532	6.01E	0 3 Oxidation[M]:9 Oxidation[M]:13 Carbam	3 spiP60842/IF4A1	U118KA 1000P	tarpet	0	2 26348	96	252 122352
8 OE Plus Yanaling TCI	28056	2345.168	2	0 KPTDGASSSNCVTDISHLVR	845.168	-0.00077	30.04682	6.30E	0 11 PFIND DELTA 258 15:	3 spiP49321INASP	HU 697 R.K/ 1 PFIND	Détarget	1	-0.94686	32	258 140161
9 OE Plus Yangling TCI	30373	2684,348	3	D ILGLOVOQAEHCSIODADAAMR	884.351	-0.00316	31,96724	7.85E	0 12 PFIND DELTA 258 16:21 Oxidation[M]	3 splQ9GZR2IREXO	4.1370.KL/1PFIND	DEtarget	0	-219117	32	258 146453
10 QE Plus Yangling TCI	35575	2495.251	3	D LLDVLSGHEGPISGLCFNPMK	495.25	0.001428	28.07684	9.10E	0 16 PFIND_DELTA_252 13:20 Oxidation[M]	3 spiQ15269(PWP2	HL492.R.S/1/PFIND	Di target	0	-2.2878	32	252 121778
11 OE Plus Yangling TCA	25045	2672.221	3	0 VVMALGDYMGASCHACIGGTNVR	572.212	0.009384	29.89181	1.02E-	9 3.Oxidation[M]:9.Oxidation[M]:13.Carbarr	3 spiP60842/iF4A1	HU118KA/1000P	Therest diff	0	3.11740	96	258.139053
12 OE Plus Yangling TCI	22491	2908.322	3	D NPOVASTWNEVNSHSNAQCVSNNR	908.312	0.010883	25.1849	1.035	9 19 PFIND DELTA 252 12:	3 spl015154/PCM1	HI 549 R.D. 1 PFIND	Détarget	0	2.5937	100	252 119635
13 OE Plus Yangling TCI	25692	2993.364	3	D ELFOTPGHTEESMTDDNITEVSCK	993.348	0.015681	24.47987	1.26E	9 13 Oxidation[M] 23 PFIND_DELTA_252 12	3 spiP46013iKi67_H	UN 2441 K.S 1 ICIPPING	target [1	1.50220	96	252.124106
14 OE Plus Yangling TCI	28336	2980.322	3	D EGICALGGTSELSSEGTOHSYSEEEK	880.319	0.003476	21.01752	1.30E	9 4 PRIND DELTA 252 13	3 solP13797IPLST +	UT100.K.Y/1IPFIND	Détarget	0	-0.18048	36	252 121951
15 OE Plus Yangling TCI	28826	2650.209	3	0 VVMALGDYMGASCHACIGGTNVR	\$50 197	0.011875	27 98035	1.526	9 9 Ovidation[M]:13 Carbamidomethy[C]:14	3 spiP608420F4A1	HU 118 K AV 1 ICICIPEI	Ctarget	0	2.47523	96	252 121628
16 OE Plus Yangling TCI	26229	3434,505	3	D SSSOPSSCCSDPSKPGGNVEGATOSLAEC	34.507	-0.00218	26,26885	1.85E	9 9.PRND DELTA 309.15	3 spi013501/505T	1 1281 K.K/ 1IPFIND	DEtarget	1	0 19416	36	309 134533
17 OE Plus Yangling TCI	23967	2451.174	2	D VGVGTCGIADKPMTOYODTSK	51 172	0.002028	25,43208	2.14E	9 6 PRIND DELTA 252 13	3 spi075940(SPF30	HI 208 K Y/ 1 PFINO	DEtarget	1	-1.80833	32	252 122549
18 OE Plus VI FL 50per	35837	4250.127	4	D IIPTLEEGLOLPSPTATSOLPLESDAVECUN	250.13	-0.00272	25.80814	2.42E	9 29 PFIND DELTA 252 12	3 spiP61976HNRP	H 103 K G/ 1 PFIND	DEtarpet	0	-156726	36	252 100837
19 OE Plus Yangling TCI	15312	2680.217	3	0 NGLGNOLSSGSHTSAASOCDSASSR	880.205	0.012494	21 11898	3.60E	9 19 PFIND DELTA 258 14:	3 spiO5VV42ICDKA	+ 619.R.M 1 PFIND	Détarget	0	-1.08752	96	258 142127
20 OE Plus Yangling TCI	35611	2789.418	3	D LEVDAMNAANSSLIGGGGVDGCIHR	89 398	0.02055	22,67217	4.38E	9 23 PFIND DELTA 252 12	3 solO9BO69MACC	1 163 K A/ 1 PFIND	DEtarget	0	0.97621	100	252 129757
21 OE Plus Yangling TCI	29513	2650.211	3	0 VVMALGDYMGASCHACIGGTNVR	850 197	0.013714	25,2873	4.80E	9 3 Oxidation[M]:13 Carbamidomethyl[C]:11	3 spiP6084211F4A1	HU118 KAV100PFI	C target	0	2 5457	96	252 123467
22 OE Plus Yangling TCI	29511	2656.23	3	0 VVMALGDYMGASCHACIGGTNVR	556.217	0.012958	25.26673	6.05E	9 3.Oxidation[M]:13.Carbamidomethyl[C]:14	3 spiP60842/iF4A1	HU118 KA/ 1 DIDIPPI	C target	0	2.56769	96	258.142688
23 OE Plus Yangling TCI	35823	3190,508	3	D SGDAAIVDMVPGKPMCVESFSDYPPLGR	90.508	-0.0001	24,22125	6.28E	9 16 PFIND DELTA 252 13:	3 sciP68104/EF1A1	HU395 K.F/ 1PFIND	Distarget	1	-0.03768	36	252,11756
24 OE Plus Yangling TCI	26068	2242.098	2	D INOVPHOSCITEGNELTK	2242.09	0.008404	27.60092	7.48E	9 9.PRIND_DELTA_252.12	3 spiP78527 PRKDC	H 1895.K.T 1 PFIND	DEtarget	0	1.49560	100	252 119734
25 OE Plus YJ FL Soper .	32471	2859.302	3	D PSWADOVEEEGEDDKCVTSELLK	859.297	0.004956	25,85762	7.64E	9 16 PFIND DELTA 252 12:	3 spiO75821/EIF3G	HL9.K.G/ 1PPIND	DEtarget	1	-1.76907	32	252.1139
26 OE Plus VI R. SOper 1	22104	3108.404	3	D KEGICALGGTSELSSEGTOHSYSEEEK	08.404	0.00005	34,2139	8.20E	9 5 PRND DELTA 252 12	3 spiP13797iPLST +	UT99.R.Y/ 1IPFIND	DEtarget	1	-2.73900	36	252 108028
27 OE Plus Yangling TCI	33938	2088.077	2	D KPANDITSOLEINFGDLGR	088.072	0.004788	24,23487	9.15E	9	3 spiO8NC51IPAIRE	H 345.R.P/1	target	1	0,4989	0	
28 OE Plus Y/ RL 50per :	25688	2345.159	2	D KPTDGASSSNCVTDISHLVR	845 158	0.000861	30,14683	1.055	8 11 PFIND DELTA 258 14:	3 spiP49321/NASP	HU 697 R.K. 1 IPPIND	DEtarget	1	-1.46101	32	258.13179
29 OE Plus Yangling TCI	26392	2996.312	3	D EGICALGGTSELSSEGTOHSYSEEEK	996.309	0.003507	1914234	1.16E-	8 4 PRND. DELTA 268 12	3 spiP13797iPLST_F	UT100 K.Y/ 1PPIND	DEtarget	0	1 73435	36	268 111921
30 QE_Plus_Yangling_TCI	23718	3209.427	3	D AGEPNSPDAEEANSPDVTAGCDPAGVHP	209.416	0.010716	22.30986	1.20E	8 21.PFIND_DELTA_252.12;	3 sp(Q08)23(NSUN)	+ 737.R/ 1 PFIND_	Détarget	0	3.35816	100	252 118304
31 QE Plus YI R SOper	4465	1783.745	2	D GGGGGGGGGGGGGGGGSGSASR	183.744	0.000579	27.09277	1.28E	8 12 PFIND_DELTA_258 14	3 spiP42356(PI4KA	HUT.R.G/ 1PPIND	DEtarget	0	-1.47209	32	258.133584
32 QE Plus Yangling TCI	25499	1954.071	2	D VAPEEHPVLLTEAPLNPK	954.064	0.007161	24.01486	1.32E	8	3 spiP60709(ACTB)	U 95.R.A/61	target	0	1,43206	64	
33 QE_Plus_Yangling_TCI	35345	2919.447	3	0 LTTPTYGDLNHLVSATMSGVTTCLR	19.432	0.015846	26.16512	1.32E	8 17.Oxidation[M] 23.PFIND_DELTA_252.12	3 spiP07437[T885_F	UI 216 K.F/ 1 DIPFIND	l target	0	1 28033	100	252 124552
34 OE Plus Yangling TCI	22257	2330.104	3	D ISCGGGINVDVNQHPDGGPGGK	30 102	0.002161	18,24762	1.32E	8 3.PRIND. DELTA 252.13	3 spiQ96F45iZN503	H 283 R.A. 1 PFIND	DEtarget	0	-1.86942	32	252.12315
35 OE Plus Yangling TCI	31201	2493.195	2	D OGEYGLASICNGGGGASAMLIOK	93194	0.001113	25.01269	1.35E	8 10.PFIND_DELTA_252 13:19.Oxidation(MS	3 spiP24752ITHIL H	UN 403 K.L/ 1 PFIND	Détarget	0	0.68453	36	252.12147
36 OE Plus Yangling TC	29491	2702.278	2	D VMTIPYOPMPASSPVICAGGODR	702.271	0.006783	24,2682	1.35E-	8 2.Oxidation[M]:9.Oxidation[M]:17.PFIND.1	3 spiQ15365iPC8P1	H 177.R.C. 1000PFM	Ctarget	0	1.92746	36	252.116332
37 OE Plus Yangling TCI	24400	3107.48	3	0 GTEAGOVGEPGIPTGEAGPSCSSASDKLPI	07.467	0.012457	21,53915	1.385	8 21 PFIND DELTA 252 12	3 spi015355/PPM1/	F 220 R.V/ 1PPIND	Détarget	1	1.33006	100	252 120436
38 OE Plus Yangling TCI	28553	2789.301	3	0 SPDEALPGGLSGCSSGSGHSPYALER	189.297	0.004013	19.655	1.38E	8 13 PFIND. DELTA 258.15	3 spi05/525iPRC28	HI 1469 R A 1 PPIND	DEtarget	0	0.94006	36	258.143227
39 QE_Plus_Yangling_TCI	33407	2594.17	2	D VCEDLDTSVNLAWTSGTNCTR	594 169	0.001555	23.26515	1.41E	8 2 PRIND_DELTA_252 13 19 Carbamidomet	3 spiP45880/VDAC2	H 208 K.F/ 1 PFIND	Détarget	0	-2.80033	32	252 121526
40 QE Plus Yangling TC	4684	1777.727	2	D GGGGGGGGGGGGGGGGSGSSASR	777.724	0.002579	22.88986	1.56E	8 12.PFIND_DELTA.252.12	3 spiP42356IPI4KA	HUT.R.G/ 1PFIND.	DEtarget	0	1.16012	36	252 115707
41 QE Plus Yangling TCH	35971	3190.51	3	D SGDAAIVDMVPGKPMCVESFSDYPPLGR	90.508	0.00133	23.51976	1.56E	8 16.PFIND_DELTA_252.13:	3 spiP68104[EF1A1	HU395 K.F/ 1 PFIND	Détarget	1	-118682	36	252.11899
42 QE Plus Yangling TCI	29398	2995,364	3	D SYDPPCPGHWTPEAPGSGTTCPGLPR	95.353	0.010671	24.18949	1.65E	8 6 Carbamidomethy [C]:21 PRND_DELTA_2	3 spi060936INOL3	HU 103 R.A. 1 DIPPING	[target	0	0.75267	100	258 139094
43 QE Plus YJ RL SOper	25613	2339.139	2	D KPTDGASSSNCVTDISHLVR	339.138	0.00085	28.67104	1.70E	8 11.PFIND. DELTA 252 12:	3 spiP49321[NASP]	U 697.R.K 1 PFIND	DEtarget	1	-1.33616	32	252 111802
44 QE Plus VI R. Soper_	25926	2980.31	3	D EGICALGGTSELSSEGTQHSYSEEEK	80.309	0.001464	20.55071	1.79E	8 4.PRIND_DELTA_252.12	3 spiP13797[PLST_H	UT 100 K Y/ 1 PFIND	Détarget	0	-2.64583	36	252.109939

6. Supporting protocol 1: Protein sequence database

This protocol is used to download protein *.fasta files for database search.

1) Open <u>https://www.uniprot.org/</u>, enter the Latin name of the species (e.g., *home sapiens*), then click search.

			pr -				t. ()		10 Mar 191	-	an A
UniProt		homo sapiens						×	Advanced 🗸	4	Search
		K			1 miles		Carlos Maria	7.			N 26532.
BLAST Align Retrieve	e/ID mapping	Peptide search	SPARQL		and the		A	20	He	lp	Contact
The mission of UniDrot is	to provide the	scientific comm	unity with a c	omprohonsiva	biab-aupli	ty and from	ly accossible	0 1000	urco of pr	atalı	-

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

2) Click "Reviewed" (Swiss-Prot).

UniProt	UniProtKB - homo sapiens AND reviewed:yes							
BLAST Align Retrieve/ID mapping Peptide search SPARQL								
UniProtKB 2021_03 results								
UniProtKB consists of two section	UniProtKB consists of two sections:							
Reviewed (Swiss-Prot) - Manua Records with information extracted from	Reviewed (Swiss-Prot) - Manually annotated Records with information extracted from literature and curator-evaluated computational analysis.							
Unreviewed (TrEMBL) - Comput Records that await full manual annotation	ationally analyzed m.							
Filter by ⁱ	Section 3 Section 3 Section 3 Section 3 Section 3 Section 3							
Reviewed (26,570) X Quote terms: "homo sapiens"								

3) Select "Uncompressed", then Click "Download" and "Go".

Filter by ⁱ	🔧 В	LAST E Align		🛨 Download 🔒 Add to basket	🔎 Column	ns ≻
Reviewed (26,570) * Swiss-Prot	Q	uote terms: '	"h	 Download selected (0) Download all (26570) 	×	
Popular organisms		Entry 🖨	Е	Format: FASTA (canonical)	~	
Human (20,386)		P18440	А	○ Compressed Uncompressed	c	etyltra
Mouse (7)		P01920	D	Preview first 10 ⁱ	Go	stocomp
4) Get the *.fasta file.						
📋 uniprot-homo sapiens-filtered-reviewed_yes.fas	ta			2021/9/17 14:24 FASTA 文件	17	7,137 KB

7. Supporting protocol 2: MSconvert

This protocol is used to convert non-Thermo MS data into mzML format files for pChem search.

1) Download MSconvertGUI that is embedded in the ProteoWizard platform from: <u>https://proteowizard.sourceforge.io/download.html</u>.

wizard	Projects	Download	Publications	Team	Documentation -	Support
Download						
Project	Pleas	e cite us in yo	ur publications			
ProteoWizard	~					
Platform	A cro	ss-platform too	olkit for mass sp	ectrometr	ry and proteomics.	
Windows 64-bit installer (able to convert vendor files except T2D)		ibers, M.C., M	acLean, B., N	Mallick, P.	Nature	
Email (optional) ?	Diole	Shinology 50, 8	910-920 (2012).	Article -	→	
john.appleseed@gmail.com	lf you	would like to	support the effo	rts of Pro	teoWizard,	
I have read and accepted the license agreements	dona	ions are happ	ily accepted. Th	ank you f	or your support!	
DOWNLOAD			DONAT	E		

2) Install ProteoWizard according to the following instruction.

🛃 ProteoWizard 3.0.21236 64-bit Setup	_		×
Windows Explorer Integration		C	Ð
Add MSConvertGUI to the Windows Explorer right-dick menu			
✓ Add SeeMS to the Windows Explorer right-dick menu			
<u>B</u> ack Next	t	Cano	cel

🙀 ProteoWizard 3.0.21236	64-bit Setup — 🗆 🗙
Ready to install Proteov	Vizard 3.0.21236 64-bit
Click Install to begin the inst installation settings. Click Ca	allation. Click Back to review or change any of your incel to exit the wizard.
	Back Install Cancel
🖟 ProteoWizard 3.0.21236	64-bit Setup – 🗆 🗙
S	Completed the ProteoWizard 3.0.21236 64-bit Setup Wizard
	Click the Finish button to exit the Setup Wizard.
	Back Einish Cancel

3) Open MSconvertGUI

🖳 MSConvertGUI (64-bit)		- 🗆 ×
List of Files File of file names File: 2 Add Remove E:\HJX\MZML\9204_IFM_Slot1-21_1.d Dutput Directory:	Browse network resour	About MSConvert Scan Summing Frecursor tolerance m/z Scan time tolerance seconds Ion mobility toleran5 ms or vs/cm ² 2 Sum MS1 scans also:
E:\HJX\MZML Browse 3		6 Add Remove
Options Output forma mzML VExtension	Filter	Parameters
Binary encoding precis 64-bit O 32-bit	titleMaker	<pre><runid>. <scannumber>. <scannumber>. <chargestate> File: "<sourcepat< pre=""></sourcepat<></chargestate></scannumber></scannumber></runid></pre>
Write index: 🗹 Use zlib compression: 🗌	scanSumming	precursorTol=0.05 scanTimeTol=5 ionMobilityTol=5 sumMs1=0
TPP compatibility: 🗹 Package in grip: 🗌		
Use numpress linear compression:		
Use numpress short logged float compression		
Use numpress positive integer compression		
Combine ion mobility scans:		
CTTF . CTTFF		
SLM as spectra: [SNM as spectra:]		8

- ①-②Browse and add MS data (e.g., *.d, *.WIFF files)
- 3 Define output route
- 4 Choose *.mzML as the output data format
- (5) -6 Define parameters for Scan Summing
- 6 -8 Save and run

8. Supporting protocol 3: ChemCalc

This protocol is used to estimate candidate molecular formulas from the pChem-determined accurate masses.



1) Open https://www.chemcalc.org/mf-finder .

2) Click **2**, check the element composition.

MOLECULAR FORMULA	CHO	and a second	FRDM	MONOISOTOPIC	MASS Ø Ø	ChemCalc
MONUCSCIOPIC MASS BIO / POLYMERS ~ NEW DERICISES ~ WEB SERVICES REFERENCE DATA ~ NEWS CITE US	54	AT -	Range C0-100 H0-200 N0-20 C0-2 Accuracy 5000 (ppm) Unsaturation filters Min 0 Only integer •	onizations Max 999 Only non-integer		
TA TA	B	MF	Monoisotopic mass	PPM	mDa may	unsaturation

3) Input the monoisotopic mass of each PDM shown in *pChem.summary* file. The candidate molecular formulas will immediately appear below.

Rank	PDM	Accurate Mass	Top1 Site Probability	Others	#PSM	#PSM L H	DFLs
	1 PFIND_DELTA_252	252.1223	391C 0.988		13876	/368 6508	
ChemCalc molecul	ar formula and isotopic distribution - Find molecula	ar formula from a monoisotopic mass					: – ø ×
T	75 5 9"						
~ \	0/50		CONV		00		Grienicaic
MOLECIUM	FORMULA		FIXIIM		aa	-	
MOLECODA			252.122339		9	•	
		04			7 -	2, •.	
NEW EVER		L	A las	· · /			
WER SERVIC		# MF	Monoisotopic mass	PPM	mDa		unsaturation
DECEDENCE	FI IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	ilter data 👘 Filter data	min 🔹 max	mín • max	min 🖬 m	ax min	• max
NEVER	Ball	1 C ₁₁ H ₁₆ N ₄ O	252.1222	0.39	0.10	-/	6
CITCUE	amen	2 C ₁₃ H ₁₈ NO,	252.1236	-4.93	-1.24		5.5
CHE US		3 C ₁₀ H ₂₀ O ₇	252.1209	5.70	1.44		1
74	IT.	4 C ₉ H ₁₄ N ₇ O;	252.1209	5.72	1.44		6.5
-		5 C ₁₄ H ₁₄ N ₅	252.1249	-10.24	-2.58		10.5
2.1		6 C ₈ H ₁₈ N ₃ O	252.1196	11.02	2.78		1.5
100	and the second	7 C ₇ H ₁₂ N ₁₀ C	252.1196	11.04	2.78		7
	CH CH CH	8 C ₁₆ H ₁₆ N ₂ C	252.1263	-15.56	-3.92	/	10
	A COMPANY OF A COMPANY	9 C ₅ H ₁₆ N ₆ O	252.1182	16.35	4.12		2
the second	and the second se	10 C ₅ H ₁₀ N ₁₃	252.1182	16.37	4.13		7.5