

# 蛋白定性结果查看说明

由于做蛋白定性的客户经常在拿到检测结果后，需要对表格进行信息解读。因此，本文对定性相关的蛋白检测结果做个基本的说明。

一般此项目样品类型包括蛋白溶液、胶条、泳道、Co-IP 样品等，下机数据经过合适的数据库搜库，会得到一些检测结果表格。搜库结果表格通常有三种：Proteins、Peptide、PSMs，分别对应鉴定到蛋白的表格、匹配到肽段的表格、匹配到二级图谱的表格。



## proteins 表格对应含义



不同的搜库软件得到的 proteins 表格略有不同，基本信息如下：

**Accession:** 数据库中蛋白对应的 ID 号，是不同蛋白特征性的编号；

**Description:** 蛋白名称的详细描述；

**Coverage:** 鉴定到的肽段在蛋白上的覆盖度；

**# Peptides:** 鉴定到不同肽段的总数目，即有多少肽段匹配到此蛋白上；

**# PSMs:** 肽段匹配到二级谱图的数目；

**# Unique Peptides:** 此蛋白包含的特异性肽段数目；

**# Protein Groups:** 该蛋白出现在几个 group 里，一般为 1；

**# AAs:** 该蛋白包含的氨基酸数目；

**MW [kDa]:** 理论分子量；

**calc. pI:** 理论等电点；

**Score Sequest HT:** 对蛋白匹配度的打分，分数越高可信度越高；

**-10lgP:** 蛋白可信度得分，值越大表明该蛋白包含的可信肽段越多；

**PTM:** 修饰类型；

Avg. Mass: 平均分子量;

说明:

关于 Accession 号:

不同的软件对不同数据库搜库结果中显示的 Accession 号形式可能会不同, 下面以 uniprot 数据为例。

搜库软件 PEAKS Studio 8.5 得到的 Accession 号一般是 **J9TAY5|J9TAY5\_IPOB** 或者 **sp|P10965|SPORB\_IPOBA** 格式。则可以用 “|” 前面的编号 **J9TAY5** 或者 “sp|” 和 “|” 之前的 **P10965** 在 uniprot 网站 <http://www.uniprot.org/uniprot> 搜取此蛋白编号对应具体蛋白、基因、氨基酸序列等信息。

搜库软件 ProteomeDiscoverer 2.1 得到的 Accession 号如 P10809。则客户直接用 P10809 在对应数据库网站搜索对应蛋白详细信息, 如图:



UniProtKB P10809

BLAST Align Retrieve/ID mapping Peptide search

From June 20, 2018 all traffic will be automatically redirected to HT

## UniProtKB - P10809 (CH60\_HUMAN)

Display

- Entry
- Publications
- Feature viewer
- Feature table

BLAST Align Format Add to basket History

**Protein** | 60 kDa heat shock protein, mitochondrial

**Gene** | HSPD1

**Organism** | *Homo sapiens (Human)*

**Status** |  Reviewed - Annotation score: ●●●●● - Experimental evidence

**Isoform 1** (identifier: **P10809-1**) [UniParc] [FASTA](#) [Add to basket](#)

*This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.*

[« Hide](#)

**Length:** 573

**Mass (Da):** 61,055

**Last modified:** August 1, 1990

- v2

**Checksum:**<sup>1</sup>

E51E1BAD9615899C

```
      10      20      30      40      50
MLRLPTVFRQ MRPVSRVLAP HLTRAYAKDV KFGADARALM LQGVDLLADA
      60      70      80      90     100
VAVTMGPKGR TVIIEQSWGS PKVTKDGVTV AKSIDLKDKY KNIGAKLVQD
     110     120     130     140     150
VANNTNEEAG DGTTTATVLA RSIAKEGF EK ISKGANPVEI RRGVMLAVDA
     160     170     180     190     200
VIAELKKQSK PWTTPPEIAQ VATISANGDK EIGNIISDAM KKVGRKGVIT
     210     220     230     240     250
VKDGKTLNDE LEIIEGMKFD RGYISPYFIN TSKGQKCEFQ DAYVLLSEKK
     260     270     280     290     300
ISSIQSIVPA LEIANHRKP LVIIAEDVDG EALSTLVLNR LKVGLQVVAV
     310     320     330     340     350
KAPGFGDNRK NQLKDMAIAT GGAVFGEEGL TLNLEDVQPH DLGKVGEVIV
     360     370     380     390     400
TKDDAMLLKG KGDKAQIEKR IQEIIIEQLDV TTSEYEKEKL MERLAKLSDG
     410     420     430     440     450
VAVLKVGGS DVEVNEKKDR VTDALNATRA AVEEGIVLGG GCALLRCIPA
     460     470     480     490     500
LDSLTPANED QKIGIEIIEKR TLKIPAMTIA KNAGVEGSLI VEKIMQSSSE
     510     520     530     540     550
VGYDAMAGDF VNMVERGIID PTKVVRTALL DAAGVASLLT TAEVVVTEIP
     560     570
KEEKDPGMGA MGGMGGGMGG GMF
```

## 关于 Description:

例如，对 P10809 蛋白编号的 Description 如下：

**60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1  
PE=1 SV=2**

60 kDa heat shock protein, mitochondrial : 蛋白名称

OS=Homo sapiens: 物种的拉丁文名称

GN: 表示 gene name, 基因的名称

PE: ProteinExistence, 即蛋白的可靠性, 分 5 个等级, 数据越小代表可靠性越高。

SV: SequenceVersion, 蛋白序列版本;

可参考上图显示。

### 关于 Coverage:

表示该蛋白鉴定肽段的氨基酸个数占该蛋白序列氨基酸总数的比例。一般越高可以表示蛋白可信度越高。

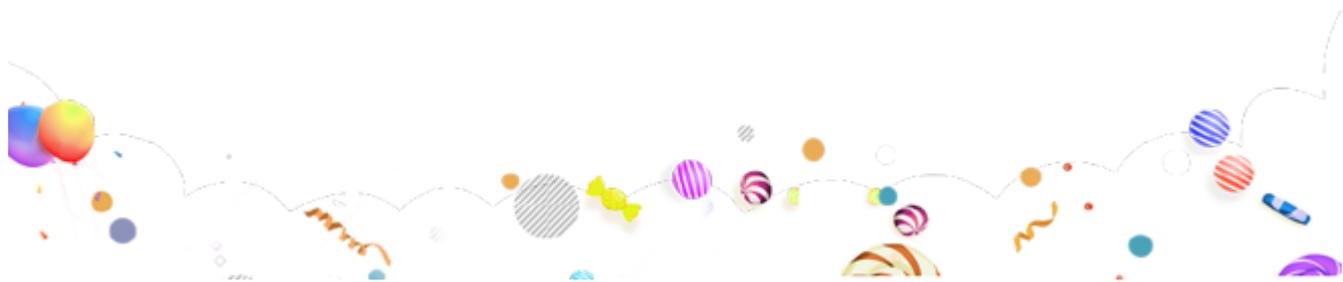
### 关于# Unique Peptides:

表示蛋白中含有的特异肽段数(或唯一肽段数)。文章一般要求 2 个及以上;是蛋白的特征肽段,可以用此肽段的二级图谱作为鉴定到某蛋白的标志。有些搜库软件得到的 protein-peptide 表格中会注明哪些是唯一肽段,或者可以参考在 protein groups 中数目为 1 的肽段。

### 关于 Protein Groups:

当某些肽段匹配到多个蛋白时,软件会根据蛋白中是否存在唯一肽段、匹配肽段的多少等权重进行打分,得到一个 protein group 蛋白,其余的蛋白会被 group 掉。因此经常看到鉴定到的蛋白数为上千个,protein group 数目只有几百个,这几百个即实际鉴定到的蛋白。

最后,如果想确定蛋白可信度,可根据蛋白匹配到的肽段数目、唯一肽段数目、肽段对应的 PSM 数目、及覆盖度和蛋白的打分值等来综合考虑。肽段和 PSM 数目越多越好,覆盖度越高越好,Score Sequest HT/-10lgP 值越大越好(值要看相对大小,没有最高最低之说)。当然这些都基于数据库合适的情况下。



### Peptide 表格对应含义

Peptide/Sequence: 肽段序列;

Modifications: 修饰类型;

# Protein Groups: 肽段归属的 protein group 数;

Master Protein Accessions: 该肽段归属的主要蛋白名称;

# Proteins: 该 protein group 内蛋白数;

-10lgP: 该肽段对应谱图鉴定的可信度, 值越大表明匹配结果越好;

Mass: 此肽段单同位素质量;

Length: 此肽段氨基酸数目;

Ppm: 此肽段质量与理论值偏差;

m/z: 该肽段质荷比;

RT: 保留时间, 以 min 为单位;

Area Sample 1: 肽段在实验结果中所有谱峰的峰面积加和;

#Spec/# PSMs: 该肽段匹配谱图数;

PTM: 肽段修饰类型;

Percolator q-Value Sequest HT: 该肽段 q 值, 越小越好;

# Missed Cleavages: 漏切位点 (一般 $\leq 2$ ), 过大会导致错误匹配;

**说明:** 可以根据此表, 查找感兴趣的蛋白对应的肽段序列、唯一肽段序列。有些文章中, 会添加目标蛋白对应肽段的一级或二级谱图, 一般质谱图上也会出现保留时间和质荷比信息。



PSMs 表格对应含义



Annotated Sequence: 二级谱图匹配到的肽段序列;

Modifications: 修饰类型;

# Protein Groups: 归属的 protein group 数;

# Proteins: 该 protein group 内蛋白数;

Master Protein Accessions: 该肽段归属的主要蛋白名称;

Protein Accessions: 该肽段归属的所有蛋白的名称;

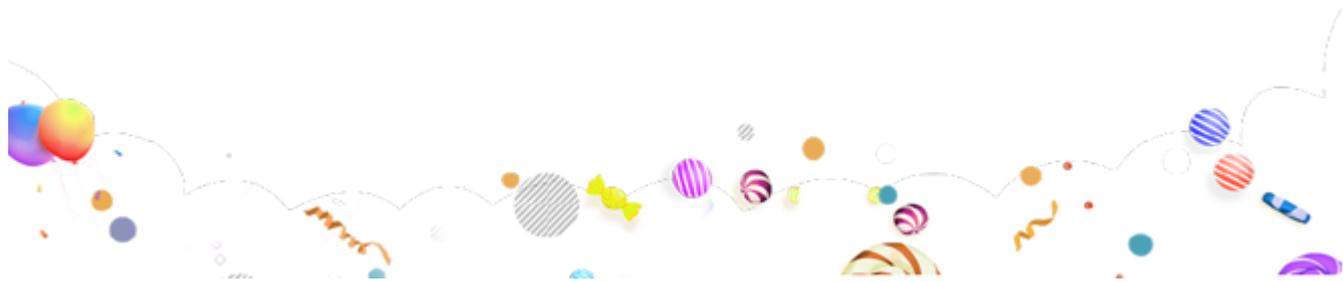
Charge: 该肽段序列带电荷数;

m/z [Da]: 该肽段的质荷比;

RT [min]: 该肽段对应的保留时间;

XCorr: PSM 可信度的打分, 仪器设备有一个可信度标准, 带不同电荷得分标准不同, 此得分高低会直接体现在肽段和蛋白的可信度上。

**说明:** 此表格主要是二级谱图 PSM 对应肽段的信息, 由于个别的搜库软件 Peptide 表格信息较少, 因此可参考此表格的结果说明。另外, 在 Annotated Sequence 此列搜取肽段序列, 可以得到某个肽段序列对应的全部 PSM 信息。



一般来说, 想了解鉴定到蛋白的信息, 只参考 proteins 表格即可。想了解关注的肽段的信息, 需要参考 peptide 表格。而 PSMs 表格一般用不到, 可看做是 peptide 表格的补充。

以上信息, 是做蛋白定性项目的客户经常咨询的问题。仅希望此文能满足检测表格的解读需求。