

Mini-Tutorial: Understanding Mascot Reports (MS/MS Ion Search Results)



The screenshot shows a Mascot search results report. It includes a table with columns for 'Protein Name', 'Score', 'Molecular Weight', 'pI', and 'Accession'. The table lists several protein matches, with the top results having higher scores. The report also includes a summary of the search parameters and a list of the proteins identified.

Mascot is a software package from Matrix Science (www.matrixscience.com) that interprets mass spectral data into protein identities. Mascot compares the observed spectra to a database of known proteins and determines the most likely matches. When you submit a protein sample for mass spectrometry analysis at the OHRI Proteomics Core Facility you will receive a report in this format.

Mascot has three main search modes: Peptide Mass Fingerprint, Sequence Query, and MS/MS Ion Search. This tutorial gives an overview of the reports produced by MS/MS Ion Search. This search is used to analyse data from tandem mass spectrometry experiments. The reports you receive may vary from the illustrations in this tutorial, depending on specific options used when your report was generated.

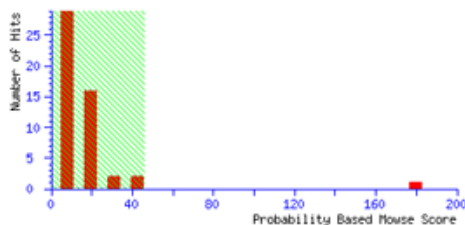
Protein Score

For each protein match, Mascot calculates an overall Protein Score. This number reflects the combined scores of all observed mass spectra that can be matched to amino acid sequences within that protein. A higher score indicates a more confident match.

A mass spectrometry experiment never produces “perfect” data. There is always some level of noise in the data and not every ion that a protein could theoretically generate will be observed in every experiment. Therefore false or random matches can be produced by the analysis software. Some matches will arise simply by chance. To address this, Mascot calculates a “threshold” score representing a 5% confidence threshold. For a specific explanation of the statistical methods used by Mascot, please see the official Mascot documentation from Matrix Science.

Probability Based Mowse Score

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 46 indicate identity or extensive homology ($p < 0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



The threshold score for statistical significance varies from experiment to experiment. In some Mascot reports you will find a histogram, as illustrated in the example on the left. In this experiment, a cut-off score of 46 was calculated for the 5% confidence threshold. The number of protein matches at each scoring position is indicated by the height of the red bars. The non-significant area is shaded in green. Most

matches will be non-significant. In example shown, most matches are non-significant but one protein match is strongly significant.

You may receive a Mascot report that does not show this histogram. If the Mascot report parameters include the detail “Max. number of hits: AUTO”, this means that only statistically significant matches are listed in the report and non-significant matches are not shown.

The Protein Summary

The protein summary contains a lot of information. The “bottom line” is that a good match should have a high **Score** and contain **multiple** Query matches listed in **bold red** type.

The screenshot shows a protein search result for **gi11351907**, Serum albumin precursor (Allergen Bos d 6) (BSA). The protein has a mass of 69328, a score of 180, and 5 queries matched. A table lists five peptides with their observed and predicted masses, delta mass, miss score, expect rank, and peptide sequence. Annotations point to various parts of the interface:

- Click for protein detail view**: Points to the protein accession number **gi11351907**.
- Protein Score**: Points to the **Score: 180**.
- Queries Matched**: Points to **Queries matched: 5**.
- Rank order of protein matches**: Points to the rank numbers 1 through 5 in the table.
- Click query number for peptide MS/MS details**: Points to the query numbers 1 through 5 in the table.
- Observed and Predicted Peptide Masses**: Points to the **Observed** and **Mr(expt)** columns.
- Number of missed trypsin cleavage sites**: Points to the **Delta** column.
- Frequency this match would occur by chance**: Points to the **Expect** column.
- Ion Scores**: Points to the **Rank** column.

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
1	1283.70	1282.69	1282.70	-0.01	0	(21)	17	3	HPEYAVSVLLR
2	1283.70	1282.69	1282.70	-0.01	0	37	0.38	1	HPEYAVSVLLR
4	1439.90	1438.89	1438.80	0.09	1	20	30	1	RHPEYAVSVLLR
5	1479.80	1478.79	1478.79	0.00	0	47	0.048	1	LGEYGFQNALIVR
6	1567.80	1566.79	1566.74	0.06	0	75	6.3e-05	1	DAFLGSFLYEYSR

Proteins matching the same set of peptides:
gi130794280
albumin (Bos taurus)
Mass: 69358
Score: 180
Queries matched: 5

Rank Order and Accession Number

The putatively identified proteins are listed in order by Protein Score. Beside the rank number is an accession number representing the identified protein. If you are viewing reports online, this accession number can be clicked to access the Protein View which contains more information about the protein sequence and shows where the matched peptides lie within the full protein sequence.

Mass

After the accession number, the expected mass of the protein is given. If analyzing gel bands, it is often useful to compare the database mass to the observed mass on the gel. If the protein identity match is high confidence but the masses do not match, this may be evidence of a post-translational modification or cleavage event.

Score and Descriptive Name

After “Mass” appears the Protein Score, which was explained above. Below this the common name of the protein is given, and often the species/taxonomy.

Queries Matched

This number indicates the number of MS/MS spectra that were matched to this protein. Typically, a confident identification should have **two or more** matching queries. The Queries Matched value can be misleading if the same peptide was scanned multiple times. In the above example, ion 1283.7 (peptide HPEYAVSVLLR) was scanned twice so the Queries Matched number reads 5 instead of 4. It is not unusual for the same peptide to be scanned more than once, especially in LC-MS experiments.

The Query Table

Query, Observed, etc...

Below the protein summary information, a table gives details on how individual MS/MS spectra were matched to peptide sequences. In Mascot terminology, each MS/MS spectrum is “Queried” against Mascot’s databases of known sequences. *Query* numbers can be clicked online to view the MS/MS spectra and full details of which ions were observed. Beside the *Query* numbers, basic information is shown such as the observed mass/charge of the ion, the calculated uncharged mass, the theoretical mass of the closest-matching peptide, and the theoretical-observed difference.

Miss

Some peptides may originate from incomplete proteolytic cleavage of the original protein. In the above example ion 1439.9 best matches to a peptide in which one potential trypsin target site (the N-terminal arginine) failed to be cleaved by trypsin. In this situation, a value of 1 is listed under the *Miss* column for one missed cleavage.

Score

The Score value listed within the Query table is the *Ion Score*. The Ion Score is a measure of how well the observed MS/MS spectrum matches to the stated peptide. The confidence threshold for Ion Scores is the same as that for the Protein Score. In the above example, ions 1479.8 and 1567.8 were statistically significant (score >46). The *Expect* value indicates the probability that the observed match between MS/MS spectra and peptide sequence would be found by chance. Confident matches typically have Expect values <0.1. Peptide matches that are non-significant on their own still contribute to the total Protein Score.

Red and Bold Text

Frequently, MS/MS spectra match to more than one possible interpretation. If the peptide shown is the highest-scoring match for a given MS/MS spectra, the result is printed in **red**. The same peptide sequence could appear in more than one protein. When a result is shown in **bold**, this indicates the highest scoring *protein* in which this peptide sequence appears (out of the proteins observed in the experiment). Matches which are both **red and bold** are the statistically strongest assignments.

Proteins matching the same set of peptides

In many cases, multiple proteins may be listed as matching the same set of peptides. This commonly results from duplicate protein entries in the databases such as splice variants or redundant entries; however cases of ambiguous match are also possible.

Remember! Software such as Mascot aids in determining correct matches but evaluating the results with a critical eye remains essential.

This document is for education and research purposes only. Official documentation on Mascot can be obtained from www.matrixscience.com