PROTEIN SUMMARY file

N- The rank of the specified protein relative to all other proteins in the list of detected proteins.

Unused (ProtScore) - A measure of the protein confidence for a detected protein, calculated from the peptide confidence for peptides from spectra that are not already completely "used" by higher scoring winning proteins.

Total (ProtScore) - A measure of the total amount of evidence for a detected protein. The Total ProtScore is calculated using all of the peptides detected for the protein. The Total ProtScore does not indicate anything about the confidence that a protein has been detected, because some or even all of the spectra contributing to the Total ProtScore may be better explained by higher ranked proteins.

% Cov (Coverage) - The percentage of matching amino acids from identified peptides having confidence greater than 0 divided by the total number of amino acids in the sequence.

% Cov (50) - The percentage of matching amino acids from identified peptides having confidence greater than or equal to 50% divided by the total number of amino acids in the sequence.

% Cov (95) - The percentage of matching amino acids from identified peptides having confidence greater than or equal to 95% divided by the total number of amino acids in the sequence.

Accession # - The accession number for the protein.

Name - The name of the protein.

Species - The species for this protein. Not all FASTA files have species, so this column may be blank.

Peptides (95%) - The number of distinct peptides having at least 95% confidence. Multiple modified and cleaved states of the same underlying peptide sequence are considered distinct peptides because they have different molecular formulas. Multiple spectra of the same peptide, due to replicate acquisition or different charge states, only count once.

PEPTIDE COLOR in the Protein Sequence Coverage pane

Color	Peptide Confidence	Included in Percent Coverage
Gray	No match	
Red	> 0 and < 50	% Cov, % Cov (50), and % Cov (95)
Yellow	>= 50 and < 95	% Cov (50), and % Cov (95)
Green	>= 95	% Cov (95)

PEPTIDE SUMMARY file

N- The rank of the specified protein relative to all other proteins in the list of detected proteins.

Unused (ProtScore) - A measure of the protein confidence for a detected protein, calculated from the peptide confidence for peptides from spectra that are not already completely "used" by higher scoring winning proteins.

Total (ProtScore) - A measure of the total amount of evidence for a detected protein. The Total ProtScore is calculated using all of the peptides detected for the protein. The Total ProtScore does not indicate anything about the confidence that a protein has been detected, because some or even all of the spectra contributing to the Total ProtScore may be better explained by higher ranked proteins.

% Cov (Coverage) - The percentage of matching amino acids from identified peptides having confidence greater than 0 divided by the total number of amino acids in the sequence.

% Cov (50) - The percentage of matching amino acids from identified peptides having confidence greater than or equal to 50% divided by the total number of amino acids in the sequence.

% Cov (95) - The percentage of matching amino acids from identified peptides having confidence greater than or equal to 95% divided by the total number of amino acids in the sequence.

Accession # - The accession number for the protein.

Name - The name of the protein.

Used - When this check box is selected, the ratios for the selected peptide are used in the calculation of the average ratio for the protein ratio (shown in the Proteins Detected table)

Annotation - provides information as to why the peptide is or is not used for quantitation

Contrib - shows how each peptide contributes to the identification of the winner protein for the group. This column displays the contribution of the peptide to the winner protein's Unused ProtScore. Low confidence peptides never have very large values for Contrib

Conf - The confidence that the peptide identification is correct, expressed as a percentage. For more information, see "The Peptide Score and Confidence" below

Sequence - The amino acid sequence of the peptide identified by the search

Modifications - Modifications associated with a peptide identification. For more information, see Notation in the Modifications Column

Cleavage - Indicates any atypical or missed cleavage sites. Missed cleavages are shown as "Missed Residue1-Residue2@Position". Atypical cleavages are shown as "Cleaved Residue1-Residue2@Position", where "Position" is either "N-term" or "C-term".

dMass - The difference in mass between the precursor MW and the theoretical MW of the matching peptide sequence

Prec MW - The experimentally measured monoisotopic mass for the precursor ion fragmented. When possible, this value is estimated for data types by looking at all survey spectra across the elution of the peptide. This method provides higher accuracy. When this method is not possible, the precursor mass is estimated from the survey spectrum that triggered the product ion acquisition.

Prec m/z - The experimentally measured monoisotopic m/z for the precursor ion fragmented. When possible, this value is estimated for data types by looking at all survey spectra across the elution of the peptide. This method provides higher accuracy. When this method is not possible, the precursor m/z is estimated from the survey spectrum that triggered the product ion acquisition

Theor MW - The molecular weight of the peptide calculated from the sequence and any modifications

Theor m/z - The theoretical m/z for the peptide, calculated from the theoretical MW and z

Theor z - The charge of the precursor ion assumed for the reported peptide identification.

Sc - The score for the peptide; this is based on matching ions of various charge states. For more information, see The Peptide Score and Confidence

Spectrum - Denotes a particular MS/MS spectrum in this processing run. For .wiff data, Spectrum is: Data Set.Sample.Period.Cycle.Experiment.

Specific

Time - The retention time when the fragmentation spectrum was acquired

PrecursorSignal - the intensity of the apex of elution for a peptide in a given charge state

PrecursorElution - the apex retention time

The Peptide Score and Confidence

Peptide Score

The peptide score is a count of the MS/MS peaks that match to a theoretical ion, for those ion types considered by the Paragon Algorithm. The b and y ions are always considered, with additional ions in certain cases (for example, higher charge states of b and y ions for electrospray data). The tolerance used for matching is based on information about the mass accuracy of the instrument chosen in the Paragon Method.

Sometimes the peptide score (shown as "Sc" in the Peptides in Group table in the Protein ID tab and the Peptide ID Hypotheses table in the Spectra tab) doesn't match the number of matching ions in the Fragment Ions table. There are several reasons for this:

- The Paragon Algorithm filters out some of the peaks used to score peptides, while the Fragment Ions table shows matches between the theoretical ions and the unfiltered peak list.
- Theoretical ions other than b and y are used for matching but are not shown in the Fragment Ions table. For example, phosphorylation neutral loss fragments are considered when you search the Phospho Emphasis special factor.
- A single MS/MS peak can match more than one theoretical ion. The Fragment Ions table shows all of those theoretical ions as matching, but the peak only counts once toward the score.
- In rare cases, a theoretical ion might match an MS/MS peak of indeterminate charge when the score is calculated but will not appear as a match in the table.

Peptide Confidence

Peptide confidence is based on the score, which is the number of matches between the data and the theoretical fragment ions. In general, a higher score leads to a higher confidence, but the factors listed below also influence the confidence.

- Hypothesized modifications – Rare, unexpected modifications tend to decrease confidence.
- Delta mass – A larger delta mass tends to decrease confidence.
- Peptide cleavage – Cleavages inconsistent with the digest agent specified in the analysis method tend to decrease confidence.
- Alternative hypotheses for the same spectrum – A peptide with a high score could receive a low confidence because there is another peptide hypothesis for this spectrum with an even higher score. Unlike some search engines, Paragon's confidence calculation is competitive, which means that the confidence of a peptide can be diluted by the presence of another viable answer for a spectrum, if it is not closely related to that answer.

About the Total ProtScore and the Unused ProtScore

For each detected protein, the Pro Group Algorithm calculates the Total ProtScore and the Unused ProtScore. (The Total ProtScore and the Unused ProtScore are abbreviated as “Total” and “Unused” respectively, in the Pro Group Algorithm results.)

The Total ProtScore is measurement of all the peptide evidence for a protein and is analogous to protein scores reported by other protein identification software.

The Unused ProtScore is a measurement of all the peptide evidence for a protein that is not better explained by a higher ranking protein. It is the true indicator of protein confidence.

About the Total ProtScore for a Protein

Here is how the Total ProtScore for a protein is calculated:

1. Convert the peptide confidence for each peptide to ProtScore units. You can convert this exactly for the first protein in the list only. After the first protein in the list, peptide confidences will be adjusted downward from the initial values reported. For more information, see The ProtScore Unit.
2. Sum the ProtScore for all peptides identified for the protein to get the Total ProtScore, taking into account the rules below.
    - Count only the maximal confidence instance of each distinct peptide sequence toward the ProtScore.
      When there are many observations of the same underlying peptide sequence, only the highest confidence instance of the sequence can contribute. This is because the use of cumulative probability assumes that the observations are independent, and multiple observations of matches to the exact sequence or even different modified states of the same sequence are clearly related events due to the database that was searched. Multiple MS/MS acquisitions of the same precursor, MS/MS acquisitions of multiple charge states of the same peptide and different modification states of the same peptide sequence are not considered to be independent evidence. For example, if two peptides are identical except for the presence of a C-terminal modification, only the peptide with the higher confidence is included in the calculation. Similarly, for peptides which are identical except for particular amino acid substitutions (I/L, D/N, or E/Q), only the highest confidence peptide is counted.
    - Each peptide contributes no more than 99% confidence.
      This limitation is necessary because of the limit in accuracy in reporting peptide confidence. The normal margin of error in percent confidence is 1 to 4% so it does not make sense to allow peptide confidences that exceed the limit of accuracy in the errors. For peptide confidence >99%, the program uses 99% when calculating the peptide confidence in ProtScore units. A side effect of this rounding is that a protein reported with confidence >99% (Unused ProtScore >2.0), will always have at least two peptide IDs (although the second ID can be very weak).

About the Unused ProtScore for a Protein

The calculation of the Unused ProtScore for a protein is too complex to provide the exact steps here. The Unused ProtScore depends on where the protein is ranked in the list of detected proteins, what spectra the protein explains, and which of those spectra are already explained by higher-ranked proteins. Tracking how spectral evidence for a protein has already been used by higher-ranked proteins is complicated (see The Contrib Column and the Unused ProtScore), but there are some basic rules for the Unused ProtScore.

The rules below are true for the Unused ProtScores within a group:

- For a group with only one winner protein (just one bold black protein), generally Unused ProtScore = Total ProtScore for the winner protein. (The Unused ProtScore can be slightly less than the Total ProtScore in certain cases. For more information, see The Contrib Column and the Unused ProtScore.)
- For a group with equivalent winners (one with more than one bold black protein), the first winner is arbitrarily chosen as the "representative winner" and the others are considered "equivalent winners". The representative winner is awarded all the Unused ProtScore. The other equivalent winners have Unused ProtScore = 0 because all the evidence is assigned to the representative winner. However, all of the winners are equally likely to be the protein that is detected and it must be understood that there is ambiguity between all of these accessions.
- For a multi-detection protein group (one with at least one bold blue protein), there are some additional rules about the Unused ProtScores:
    - The primary detected protein form is the one with the highest Unused ProtScore. It is the most obvious winner and generally has Unused ProtScore = Total ProtScore. (The Unused ProtScore can be slightly less than the Total ProtScore in certain cases. For more information, see The Contrib Column and the Unused ProtScore.)
    - The secondary detected proteins (all the detected proteins in the group other than the primary form) have an Unused ProtScore < Total ProtScore. This is because the primary form has already used some of the spectra cited by the secondary forms.
    - The Unused ProtScore for any particular protein is the same in all instances of a multi-detection protein group. This score is a fixed quantity, even though each instance of a multi-detection group is shown differently in the Pro Group Algorithm results. The bold blue proteins shown in a multi-detection group enable you to see how the multiple detected isoforms are related.

Notation in the Modifications Column

Notation	Meaning
Name(AA) @ Position	The modification Name is found on the amino acid AA at Position in the peptide. Example: Oxidation(M)@7
Name@Terminus	The modification Name is found on the peptide terminus specified by Terminus (either C- or N-). Example: iTRAQ®@N-term
Protein terminal Name@Terminus	The modification Name is found on the protein terminus specified by Terminus (either C- or N-). Example: Protein terminal Acetyl@N-term
AA1 -> AA2@ Position	The amino acid AA1 at Position was substituted with amino acid AA2. Example: I->V@7
No Name@Position	The modification Name was expected to be located at Position. Instead, the unmodified form of the amino acid was found. Examples: No iTRAQ@N-term, No Propionamide(C)@3 Note: If a different modification occurs on the target residue, that modification is reported, and the absence of the expected modification is not indicated using this notation.
(Incorrect mod mass from prior version. Error: X)	This comment precedes the modifications when one or more modifications for this peptide were reported in a prior version of ProteinPilot™ Software with an incorrect delta mass. X is the total mass error caused by the incorrect modification masses. The current version of ProteinPilot Software has a different delta mass. (See the note below for more information.) Example: (Incorrect mod mass from prior version. Error:-14.0923)

The modification names follow the HUPO-PSI Modification Nomenclature for Mass Spectrometry standard. For more information about this standard, see http://www.unimod.org.