

ProteinPilot (Paragon) Search Guide

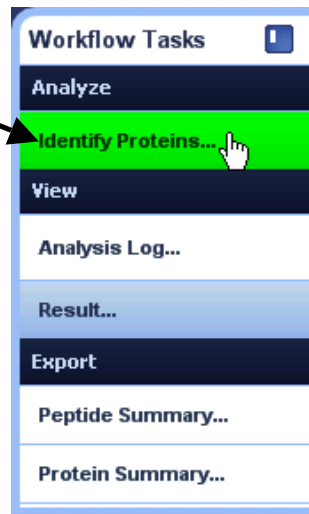
Introduction :

Paragon is part of ProteinPilot suite of search algorithms from Applied Biosystems. The search algorithm uses sequence tag algorithm to calculate Sequence Temperature Values for identification of peptides from a database. After peptide identification, the ProGroup Algorithm performs a statistical analysis on the identified peptides to determine the minimal set of identifications. The algorithm also offers an opportunity to carry out quantification analysis using its Protein Quant tab. Average ratios for each protein along with associated P-values and error factors offers an output that can be used to determine differentially expressed proteins in an experiment.

Getting Started :

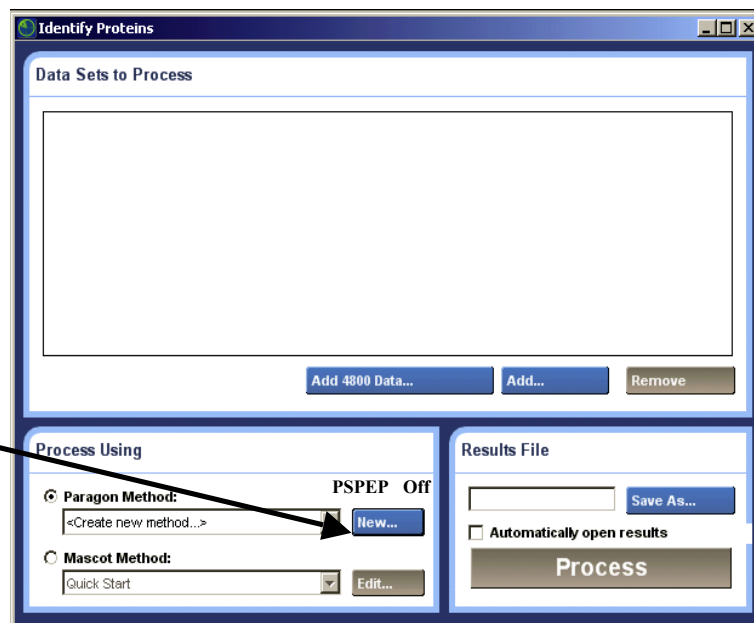
Click onto icon “identify proteins” to open the identification (and optional quantification) window.

1



Click onto “New” icon to open the method window.

2



3 Select sample type, Cys alkylation, enzyme used for digestion and instrument used for acquiring spectra depending on the experimental set up.

4 Click on “Quantitate” and modifications depending upon focus of investigation.

5 Enter the database that you want to search the dataset against and mode of search. (Thorough recommended)

6 Check and make changes to Correction factor values if required. Set the Conf threshold according to requirements.

7 Name Paragon Method

8 Save Method

Reagent	% of -2	% of -1	% of 0	% of +1	% of +2
ITRAQ114	0.00	1.00	92.90	5.90	0.20
ITRAQ115	0.00	2.00	92.30	5.60	0.10
ITRAQ116	0.00	3.00	92.40	4.50	0.10
ITRAQ117	0.10	4.00	92.20	3.50	0.10

3. The Paragon method interface offers a variety of choices with respect to a) sample types (iTRAQ (peptide-based, protein-based, 4-plex, 8-plex), SILAC, ICAT, etc.); b) mode of cysteine alkylation; c) digestion (type of proteolytic enzyme used); d) instrument used for analysis.

4. In case you chose to use any of the quantitative isotopic-labeling methods (iTRAQ, SILAC, ICAT) for your study, then Paragon offers an option of quantification of the dataset through their analysis software.

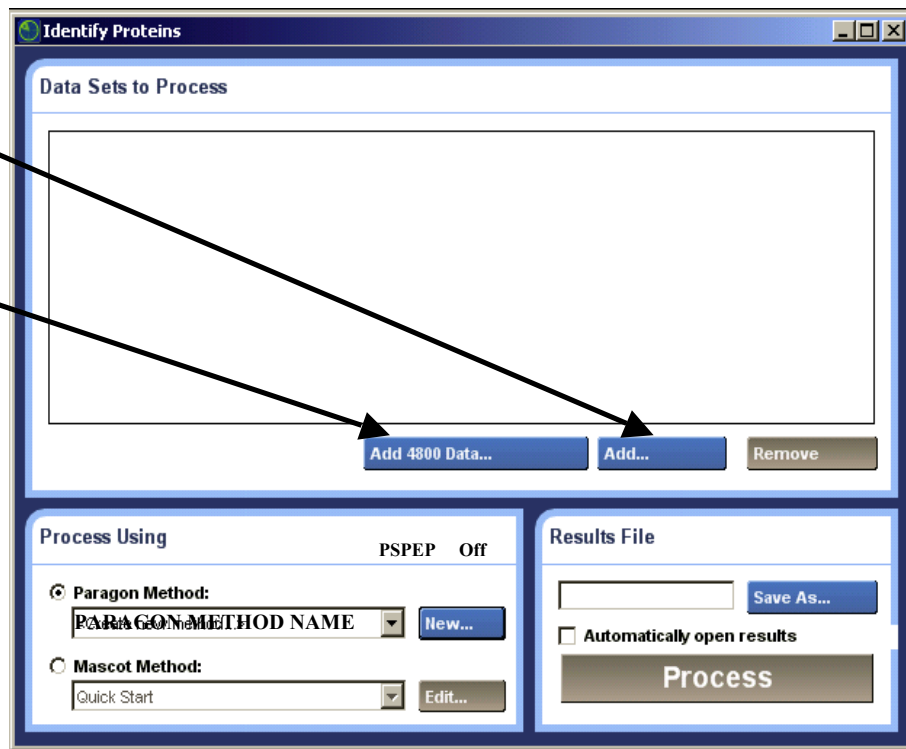
5. The database with protein sequences in FASTA format can be either in ‘target’ or ‘target-decoy’ version depending on whether you are interested in false positive search or not. However, Paragon also offers another option wherein you can use their PSPEP feature (See 11) to generate “reverse” or “random” sequences for FPR analysis (See 11 for details).

6. The correction factors can be changed by looking at the correction factor table that comes long with the kit – if the values are different from that shown in default format. The protein threshold can be set as per convenience to set a high or low stringency for identification.

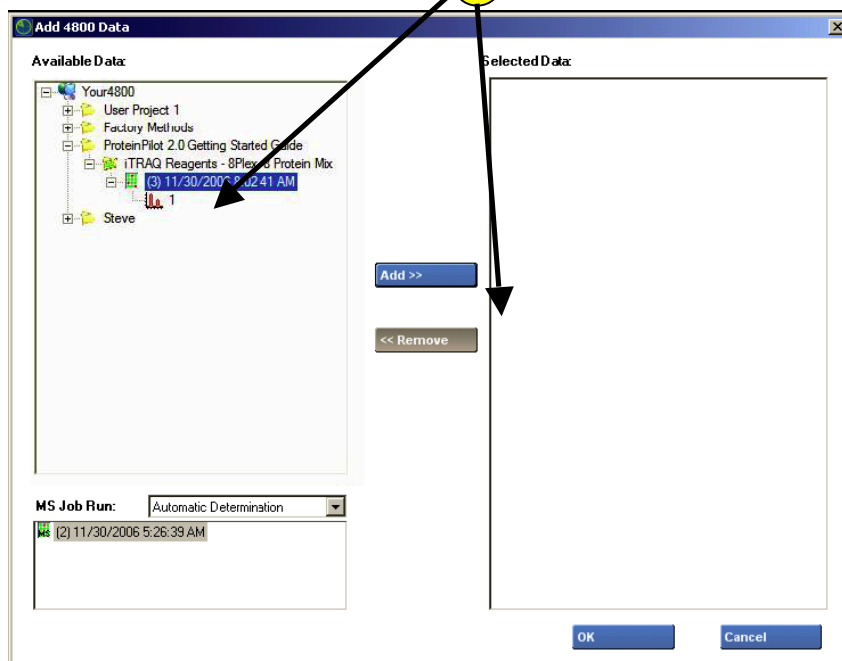
Click on “Add” to add spectral data if it is in mgf file format. **9B**

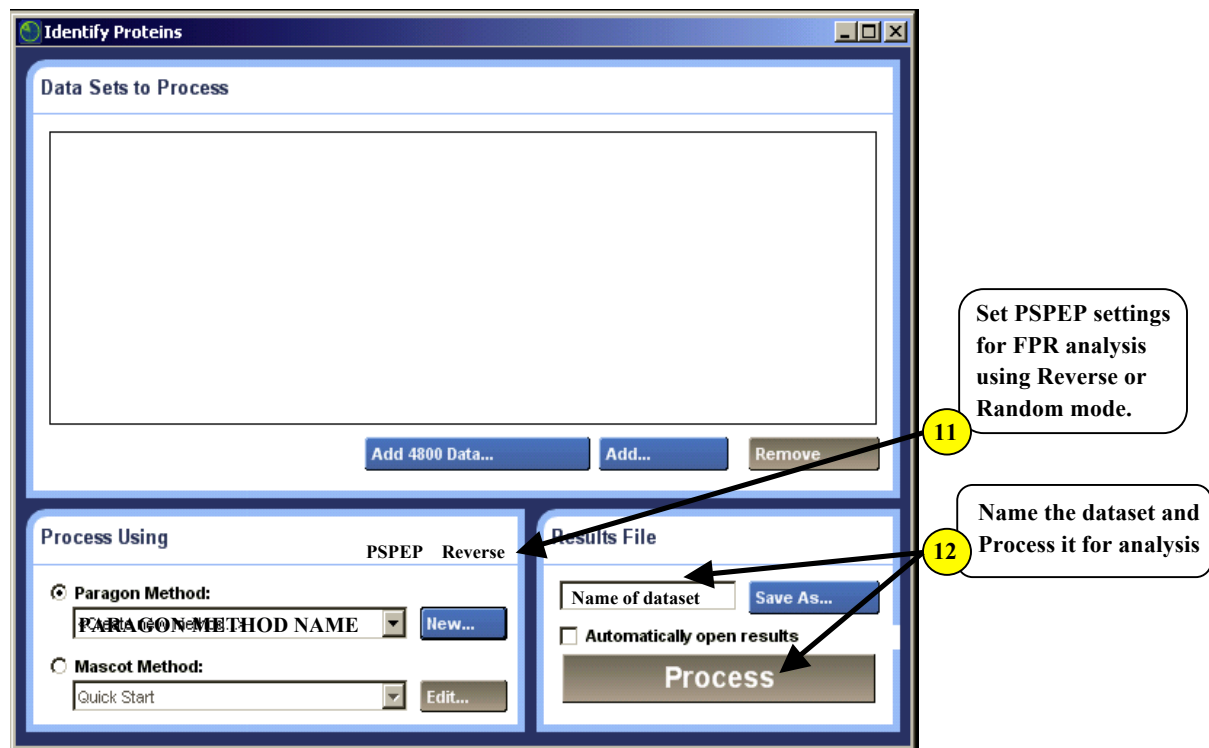
Click on “Add 4800 data” to add spotsets from ABI instrument for analysis. **9A**

9. Paragon offers a choice in terms of the format that you can submit your data for search. However, quantification can be performed only when spotsets are directly added from the instrument (9A).



Choose appropriate spotset and add to submit for analysis **10**





11. The new PSPEP feature allows you to perform false positive rate analysis using reverse sequence or random sequences. Using random or reverse option in the box creates a concatenated “target-decoy” database from your “target” FASTA database (See point 5). The PSPEP analysis provides an analysis report in Excel format which is independent from Paragon analysis. This Excel output format contains FPR analysis at protein, unique peptide and spectrum level along with ROC curves and other tables / graphs that can be of great use in analysis the quality of dataset.

Protein Quant tab gives quantification information at the peptide and protein level.

Protein ID tab gives information about protein identification.

Spectra tab gives information about spectra in the dataset.

Summary Stats tab summarizes statistical information for the dataset.

Set denominator according to analysis requirements.

The screenshot displays a software interface with four tabs: Protein Quant, Protein ID, Spectra, and Summary Statistics. Callout boxes 13-17 point to specific features. The Protein Quant tab shows a table of proteins detected. The Protein ID tab shows a table of peptide quantitation. The Spectra tab shows a peptide quantitation information graph and a precursor MS region graph. The Summary Statistics tab shows a dropdown menu for the denominator.

N	Unused	Total	% Cov	Accession #	Name	Species	114:113	115:113	116:113
1	86.95	86.95	73.2	spt P02789 TRFE_CHICK	Ovotransferrin precursor (Conalbumin) (Allergen Gal d 3) (G...	Gallus gallus	0.6855	1.4157	1.3023
2	62.55	65.08	61.5	spt P02787 TRFE_HUMAN	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-...	Homo sapiens	1.0993	1.0942	0.8837
3	51.75	51.75	61.0	spt P02769 ALBU_BOVIN	Serum albumin precursor (Allergen Bos d 6) (BSA)	Bos taurus	0.9427	1.0091	0.8564
4	28.85	28.85	66.4	spt P00921 CAH2_BOVIN	Carbonic anhydrase 2 (EC 4.2.1.1) (Carbonic anhydrase II) (...)	Bos taurus	1.3209	0.7500	0.6462

Used	Annotation	Conf	Sequence	Modifications	Theor m/z	Theor z	Spectrum	114:113	115:113
<input checked="" type="checkbox"/>	outo	86	AIANNEADAISLDGGQAFEAGLAPYK	ITRAG8plex@N-term Deamidated(N)@4 Ala->Val@17 ITRAG8plex(K)@26	3243.6003	1	1.1.3.123.1		0000.0000
<input checked="" type="checkbox"/>	auto	5	ANVMDYR	No ITRAG8plex@N-term ITRAG8plex(Y)@6	1172.6035	1	1.1.3.23.3	1.2938	2.1389
<input checked="" type="checkbox"/>	auto	99	AQSDFGVDIK	ITRAG8plex@N-term ITRAG8plex(K)@10	1675.9111	1	1.1.3.23.5	0.6712	1.4203
<input checked="" type="checkbox"/>	auto	9	CDRNSVVSNGDVECTVVDETK	ITRAG8plex@N-term Methylthio(C)@1 Deamidated(N)@9	3042.4137	1	1.1.3.103.1	9999.0000	

Peptide Quantitation Information

Precursor MS Region

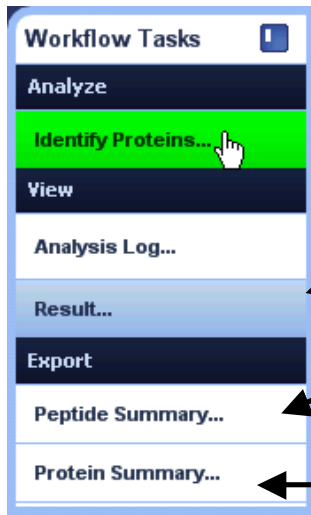
13. Protein Quant tab has information about proteins detected, peptide quantitation and change the denominator for quantification if required.

14. Protein ID tab has information on number of proteins detected, along with detailed view of the grouping information of each protein. Moreover, the tab also has information and viewable format about protein sequence coverage.

15. Spectra tab gives a spectrum-centric perspective of results. This tab shows precursor region and MS/MS data with matching fragment ions.

16. Summary Statistics Tab has information regarding proteins, peptides and spectra identified at different confidence thresholds and information about search parameters.

17. The denominator can be changed and the protein report and peptide report can be saved for the denominator format (See 19 and 20).



Click on “Result” to open the results section with Protein Quant, Protein ID, Spectra and Summary Stats tab.

18

After opening results, click on Peptide Summary and save in text format.

19

Click on Protein Summary and save in text format. The Protein Summary format can be saved with different denominator values (see 17 above).

20

Protein Report

Accession number and Protein name from FASTA database.

Average iTRAQ ratio for 114 iTRAQ label as a denominator.

P-value associated with the estimated average iTRAQ ratio.

Error factor associated with the estimated average iTRAQ ratio.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Unused	Total	%Cov	Accession	Name	Species	115:114	PVal 115:114	EF 115:114	116:114	PVal 116:114	EF 116:114	117:114	PVal 117:114	EF 117:114
2	13.7	13.7	7.6	IPI00793409.1	ACACA 265 kDa protein	9606	1.3	0.0	1.1	0.9	0.2	1.1	1.3	0.0	1.1
3	6.0	6.0	6.1	IPI00021290.5	ACLY ATP-citrate synthase	9606	1.5	0.0	1.4	1.0	0.0	1.0	1.6	0.0	1.6
4	29.0	29.0	42.0	IPI00020599.1	CALR Calreticulin precursor	9606	1.3	0.0	1.1	1.0	0.7	1.1	1.3	0.0	1.1
5	22.4	22.4	26.2	IPI00020984.1	CANX Calnexin precursor	9606	1.3	0.0	1.1	1.0	1.0	1.1	1.4	0.0	1.1
6	2.0	2.0	17.5	IPI00792087.1	CTTN 26 kDa protein	9606	1.4	0.0	1.2	1.0	0.9	11.3	1.4	0.0	1.0
7	79.0	79.0	32.7	IPI00645907.3	FASN fatty acid synthase	9606	2.0	0.0	1.1	1.0	0.9	1.1	2.0	0.0	1.1
8	6.4	6.4	7.6	IPI00022228.1	HDLBP Vigilin	9606	1.2	0.0	1.2	1.0	0.8	1.4	1.2	0.1	1.2
9	50.2	50.2	50.4	IPI00003362.2	HSPA5 HSPA5 protein	9606	1.3	0.0	1.1	0.9	0.0	1.1	1.2	0.0	1.1
0	23.1	23.2	17.0	IPI00019502.3	MYH9 Myosin-9	9606	1.3	0.0	1.1	1.1	0.2	1.1	1.3	0.0	1.1
1	22.4	22.4	47.2	IPI00783586.1	NDRG1 43 kDa protein	9606	2.4	0.0	1.1	1.0	0.4	1.1	2.6	0.0	1.2
2	33.1	33.1	36.5	IPI00843748.1	VCP 89 kDa protein	9606	1.3	0.0	1.1	1.0	1.0	1.1	1.3	0.0	1.1
3															
4															

The protein report provides elaborate information about detected proteins, along with quantification data. It is a good starting point to sort proteins in Excel format according to relevance of your study. For example, for an iTRAQ analysis experiment, it provides protein IDs, accession numbers and names along with iTRAQ ratios, P-values and error factors. The denominator of the protein report can be changed according to which label was used as a “control” to estimate the fold changes. Moreover, you can sort the proteins with decreasing P-values and later by iTRAQ ratios (115:114 etc.), so as to determine proteins that show statistically significant differential expression.

Peptide Report

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	N	Unused	Total	%Cov	Accessions	Names	Used	Contrib	Conf	Sequence	Modifications	dMass	Prec m/z	Sc	Spectrum
2	260	6.0	8.0	25.5	IP100000816.1	YWHAE 14-3-3 protein epsilon	1.0	2.0	99.0	AAFDDAIAELDTLSEESYK	ITRAQ4plex@N-term	-0.8	2375.4	34.0	1.3.7.146.2
3	260	6.0	8.0	25.5	IP100000816.1	YWHAE 14-3-3 protein epsilon	1.0	0.0	99.0	AAFDDAIAELDTLSEESYK	ITRAQ4plex@N-term	0.1	2376.3	32.0	1.2.7.122.1
4	260	6.0	8.0	25.5	IP100000816.1	YWHAE 14-3-3 protein epsilon	1.0	0.0	99.0	AAFDDAIAELDTLSEESYK	ITRAQ4plex@N-term	0.2	2376.4	32.0	1.3.7.145.1
5	5	50.2	50.2	50.4	IP100003362.2	HSPA5 HSPA5 protein	1.0	2.0	99.0	TFAPEEISAMVLTGK	ITRAQ4plex@N-term	0.1	1825.1	23.0	1.2.7.99.8
6	5	50.2	50.2	50.4	IP100003362.2	HSPA5 HSPA5 protein	1.0	0.0	99.0	NGLTSPNPENTVFDK	ITRAQ4plex@N-term	0.1	1966.1	23.0	1.3.7.56.5
7	5	50.2	50.2	50.4	IP100003362.2	HSPA5 HSPA5 protein	1.0	0.0	99.0	DNHLLGTFDLTGIPPAPR	ITRAQ4plex@N-term	-0.9	2077.2	23.0	1.4.7.85.9
8	114	11.2	11.2	50.8	IP100007427.2	AGR2 AGR2	1.0	0.0	99.0	GWVGDQLWVTQTYEALYK	ITRAQ4plex@N-term	0.4	2489.6	16.0	2.6.7.108.3
9	114	11.2	11.2	50.8	IP100007427.2	AGR2 AGR2	1.0	2.0	99.0	IMFVDPPLSLTVR	ITRAQ4plex@N-term	0.0	1421.8	15.0	1.1.7.94.1
10	114	11.2	11.2	50.8	IP100007427.2	AGR2 AGR2	1.0	0.0	99.0	IMFVDPPLSLTVR	ITRAQ4plex@N-term	0.0	1421.8	14.0	1.2.7.89.4
11	114	11.2	11.2	50.8	IP100007427.2	AGR2 AGR2	1.0	0.0	99.0	IMFVDPPLSLTVR	ITRAQ4plex@N-term	0.0	1421.8	13.0	1.3.7.81.7
12	218	7.3	7.3	16.1	IP100027252.6	PHB2 Prohibitin-2	1.0	0.0	99.0	IGGVQQDTILAEGLHFR	ITRAQ4plex@N-term	0.1	1998.2	25.0	1.7.7.88.1
13	218	7.3	7.3	16.1	IP100027252.6	PHB2 Prohibitin-2	1.0	2.0	99.0	IVQAEGEAEAAK	ITRAQ4plex@N-term	0.0	1503.8	22.0	1.2.7.45.9
14	218	7.3	7.3	16.1	IP100027252.6	PHB2 Prohibitin-2	1.0	0.0	99.0	IVQAEGEAEAAK	ITRAQ4plex@N-term	0.0	1503.8	22.0	1.4.7.46.2
15	218	7.3	7.3	16.1	IP100027252.6	PHB2 Prohibitin-2	1.0	2.0	99.0	IGGVQQDTILAEGLHFR	ITRAQ4plex@N-term	0.1	1998.2	21.0	1.8.7.87.6
16	218	7.3	7.3	16.1	IP100027252.6	PHB2 Prohibitin-2	1.0	0.0	99.0	IGGVQQDTILAEGLHFR	ITRAQ4plex@N-term	0.3	1998.4	19.0	1.6.7.89.1
17															

The peptide report provides information about peptide identification. It is a good starting point to sort peptides in Excel format to determine number of peptides that are identified above a particular Conf threshold. For example, you can sort the peptides with increasing “spectrum” column, decreasing “Conf” column and ‘Sc” column and select for unique spectra using advanced filter option. In this manner, now you can determine the number of spectra that are assigned to peptides at a particular threshold (95 or 90 Conf).