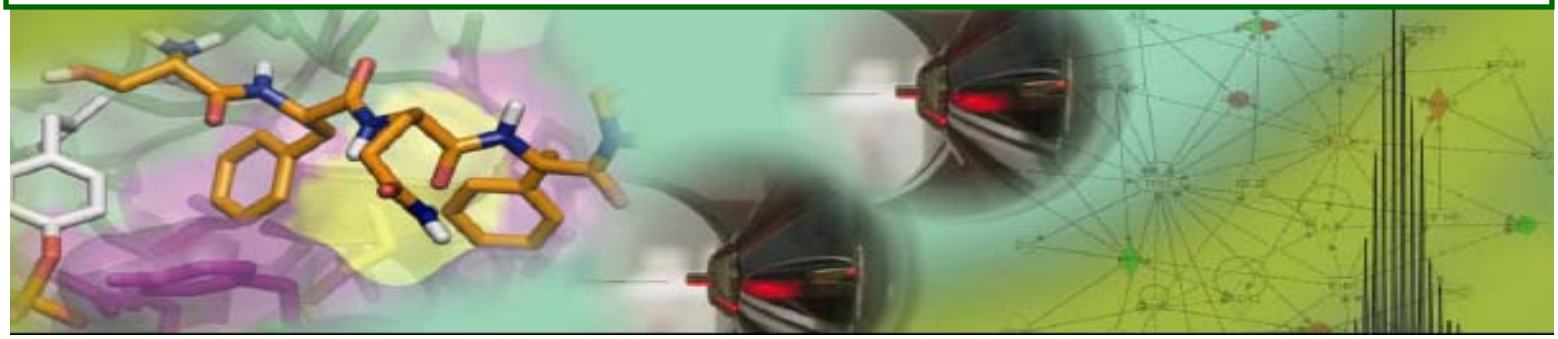
MS & Proteomics Resource

Yale School of Medicine
Keck Biotechnology Resource Laboratory



Application Note 3: iTRAQ Analysis

Background information on Isobaric Tags for Relative and Absolute Quantitation (iTRAQ):

This technique (Ross et. al., 2004) uses a multiplexed isobaric tagging reagent to label peptide mixtures which allows post label mixing of the samples (after enzymatic digestion) without adding complexity to the MS analysis. Up to 8 samples can be multiplexed with the combined group producing identical parent ion masses and MS/MS sequencing ions, except for the MS/MS reporter region of 113 -121. This greatly facilitates peptide identification due to the fact that the tag is silent in the MS analysis. The iTRAQ reagent contains a protein reactive (1) group, a reporter (2) group, and a "balance" (3) group.

1) The protein reactive group uses amine-based chemistry to specifically label N-terminal and lysine side-chains.
2) The balance group keeps all reagents at 304.2 Da (8 plex) or 144.1 Da (4 plex) and provides a distinctive neutral loss fragment in MS/MS mode corresponding to each labeled set.
3) The reporter group is used as the MS/MS quantitation ratio read-out value.

Mass spectrometric analysis is done on an AB SCIEX TripleTOF® 5600 mass spectrometer with AB SCIEX ProteinPilot™ software used for protein identification and quantitation. ProteinPilot utilizes a Paragon™ algorithm with hybrid sequence tag and feature probability database searches. Hence, specific details such as mass tolerances, specific modifications etc. are not utilized. All iTRAQ results are uploaded into the Yale Protein Expression database (YPED) for investigator viewing.

Positive identification of Proteins:

- LC MS/MS data was searched using ProteinPilot 4.2 Software. This generates a Total Protein Score which is a measurement all the peptide evidence for a protein and is analogous to protein scores reported by other protein identification software. Percent confidence can be expressed in ProtScore units, is shown below

Percent Confidence	Protein Score
99%	2.0
95%	1.3
90%	1.0
66%	0.47

- Proteins with 2 or more MS/MS spectra should be considered as a positive protein identification from the database searched
- Proteins listed as matching with only 1 significant peptide match are not considered positive identifications and may simply be present in the sample. A secondary analysis would have to be

done to confirm these (e.g. immunological reactivity and activity assays)

- Matched peptides derive from the type of enzymatic digestion performed on the protein
- Keratin matches are very likely to be the result from contamination of the sample prior to digestion and/or at some point in the sample prep.
- Enzymes used during digest such as trypsin or Lys-C (called Chain A, The Primary Structure And Structural Characteristics Of Achromobacter Lyticus Protease I, A Lysine-Specific Serine Protease) may also be observed
- Often databases contain the same or highly homologous proteins under different accession numbers – please be advised of this. Some of the proteins that contain exactly the same peptides but different accession numbers will be listed as indistinguishable meaning we cannot differentiate between these accession numbers.
- The Percent Coverage value for each protein is 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.

iTRAQ Peptide Quantitation:

- Protein Fold ratios are calculated and expressed from a pair-wise comparison of two iTRAQ channels. For each ratio the iTRAQ peak areas for each peptide are corrected for both Observed Bias Correction (as shown YPED iTRAQ result table) and Background Correction. The final ratio shown in YPED is based on a weighted averaged of the corrected peak areas.

- In addition, for each protein ratio reported, the program calculates a p-value, which can help you assess whether changes in protein expression are real or not in some experimental designs. A p-value is a standard statistical metric in hypothesis testing ranging from 0 to 1. A p-value reports the probability of observing a test statistic (in this case, a ratio) as extreme as the one observed, assuming the null hypothesis is true (in this case, that the true ratio is unity). In other words, when you observe a non-unity ratio, which is virtually always, the p-value indicates the probability that you would randomly see a ratio this different from 1. When the probability of seeing this difference randomly is low, then this can be considered grounds for rejecting the null hypothesis and for assuming the difference between the observed ratio and 1 is due to a real, and not random, variation. As the p-value gets smaller and smaller, you can be more and more confident that apparent cases of differential expression (non-unity ratios) are resulting from real measured differences. If a ratio is extremely well-determined, a real change can be detected even when the ratio is not very different from 1. For example, a protein with 38 contributing peptide ratios might have a very small uncertainty in its average ratio of 1.5 such that the p-value would be very small and thus detect this as a real change. By contrast, a protein with 3 contributing peptide ratios might have a large uncertainty in its average ratio of 1.5, leading to a large p-value (close to 1) and little evidence that the true protein ratio differs from 1. The p-value allows the results to be evaluated based on the certainty of a change in expression, not just on the magnitude of the change.
  For a general discussion on p-values, see: http://en.wikipedia.org/wiki/P-value

Ross, P. L., Huang Y. N., Marchese J. N., Williamson B., Parker K., Hattan S., Khainovski N., Pillai S., Dey S., Daniels S., Purkayastha S., Juhasz P., Martin S., Bartlet-Jones M., He F., Jacobson A.and Pappin D. J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3, 1154-69.