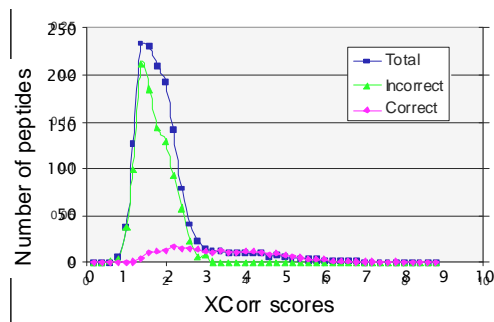


## Issues in Achieving In-depth Proteomic Profiling

by Katheryn A. Resing, *University of Colorado at Boulder*

Recently, there has been increased need for improved protein and peptide analysis of complex samples using mass spectrometry (MS). A typical approach utilizes collision induced dissociation of peptides to generate fragment ions. The information in the resulting “MSMS” spectra is used to identify the parent peptide sequence by searching protein databases for a sequence with best fit to the MSMS spectrum. When analyzing peptide digests of purified proteins on a reverse phase column interfaced with MS (LC/MSMS), sequence assignments for >75% of the high quality MSMS are typically achieved, yielding >80% protein sequence coverage. When applied to complex samples, the MS scan rate becomes limiting, requiring fractionation of the sample. A popular approach is the MuDPIT approach (MultiDimensional Protein Identification Technology), where peptides are fractionated by strong cation exchange before LC/MSMS; very complex samples require prefractionation of the proteins. The quandary is that in-depth profiling of complex samples by MuDPIT is very difficult to achieve, because protein sequence coverage is low (1-35%), and usually no more than 30% of the high quality MSMS spectra can be assigned to peptides with high confidence.

Most work in this area has focused on improving search program scoring and validation (reviewed in 1). When the number of MSMS is small, manual analysis can be used to evaluate the chemical plausibility of the observed fragmentation, but that method is prohibitive for complex samples. Instead, investigators often rely on search program-generated scores, using an acceptance threshold equal to the highest score obtained by chance alone. The latter can be determined by searching databases where protein sequences are randomized or inverted, and all assignments are false positive.



*Figure 1. Distribution of Sequest XCorr scores for MSMS spectra of a complex sample. Comparison of scores for a full MuDPIT dataset (blue), correctly identified, manually validated “hits” (pink) and the remaining MSMS spectra (including alternative charge assignments).*

Fig. 1 shows the histogram of scores for MH2+2 ions in a small MuDPIT dataset of human proteins, where manual analysis was used to validate the assignments made by the Sequest search program. The inverted database search shows that a false discovery rate (FDR) ~0.5% can be achieved using an acceptance threshold of XCorr=3.3 (high confidence threshold = HCT). However, this occurs at the expense of false negative rate ~45%, thus rejecting many correct assignments. In order to capture more data, researchers often lower acceptance thresholds, then try to filter false positive cases. A method commonly used requires a minimal difference in scores between the top two assignments, which validates 10-20% additional MSMS than those accepted by the HCT.

An alternative validation method uses independent inputs to increase statistical discrimination. For example, peptide physicochemical properties can be used to filter false positives, by using exact mass measurements of the parent peptide, or by comparing observed vs. calculated elution times from RP

chromatography. We extended the use of independent inputs by evaluating consensus between three MSMS scoring algorithms and applying several physicochemical filters, and found that 77% of the manually validated data could be accepted, yielding 35% greater data capture than the HCT (2).

Another approach simulates the intensities and m/z of fragment ions, then scores for similarity between resulting theoretical MSMS and observed spectra. Early studies used statistical analyses of large datasets to predict ion intensities, but the best results have been obtained by simulating gas phase fragmentation chemistries. The feasibility of the latter approach was shown by Z. Zhang (3), who modeled rate constants for peptide fragmentation based on proton affinity at the peptide bond, availability of protons, intrinsic cleavage kinetics determined by the adjacent amino acids, and other reactions such as C-terminal rearrangement and loss of H<sub>2</sub>O, NH<sub>3</sub> or CO. We have found that similarity scoring between observed and theoretical spectra is able to confirm >95% of the manually validated sequence assignments in a MuDPIT dataset (4).

Despite these advances, when applied to a wide range of datasets, 30-70% of the high quality MSMS still cannot be assigned. Many can be identified as MSMS of "source" fragment ions, generated inside the MS as ions are accelerated through a regions containing gas (e.g. at the source, or as ions enter, cycle in, or exit from the helium-filled ion trap). Others are revealed as "chimera spectra", when two or more parent ions are trapped together in the isolation step, and a composite MSMS spectrum is produced. The Aebersold lab has estimated that 28% of MSMS from SCX fractionated peptides of a yeast extract are chimeras (5), and our own studies show as many as 45% MSMS are chimeras in mammalian datasets (4). Newer MS instruments may not solve the chimera problem, because even with high resolution instruments, 2-3 Da width in the isolation window is required. Generation of chimera spectra with high intensity source fragment ions from abundant proteins is probably an important reason why lower abundance proteins and low intensity fragment ions are so difficult to detect; this would account for the difficulty in achieving in-depth profiling and high sequence coverage.

Improving peak capacity during peptide fractionation is a possible solution, and may be viable as ultra high pressure HPLC on 75 micron ID columns becomes robust enough for 24/7 data collection. Another solution is to fractionate ions in the gas phase, for example by ion mobility spectrometry (IMS), where ions are resolved by through a gas-filled drift tube. Recent reports indicate that sensitivity problems that compromise the usefulness of IMS are being solved (6), although the generation of source fragment ions generated in the drift tube has not been addressed. In the meantime, utilizing judicious protein and peptide chromatography protocols or focusing on specific subcellular fractions may provide the best means of side-stepping the complications of instrument generated fragment ions and chimera spectra.

1. Sadygov RG, Cociorva D, Yates JR 3rd. (2004) *Nat Methods* 1:195-202.
2. Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old WM, Cheung HT, Russell S, Wattawa JL, Goehle GR, Knight RD, Ahn NG. (2003) *Anal Chem* 76: 3556.
3. Zhang Z. (2004) *Anal Chem* 76:3908-3922
4. Sun S, Meyer-Arendt K, Eichelberger B, Brown R, Yen CY, Old WM, Pierce K, Cios K, Ahn NG, Resing KA. *Mol Cell Proteomics*, in press (on-line Oct 2006).
5. Zhang N, Li Z-J L, Ye M, Pan S, Schwikowski B, Aebersold R. (2005) *Proteomics* 5: 4096-4106.
6. Valentine SJ, Plasencia MD, Liu X, Krishnan M, Naylor S, Udseth HR, Smith RD, Clemmer DE *J Prot Research*, in press (on-line Oct 2006).

Dr. Katheryn Resing  
Chemistry & Biochemistry Dept  
University of Colorado  
Boulder, CO  
[Katheryn.Resing@colorado.edu](mailto:Katheryn.Resing@colorado.edu)

