# Increased Dynamic Range of DDA-based label-free quantification using the CHIMERYS algorithm

David M. Horn,[1] Kai Fritzemeier,[2] Katja Tham,[2] Frank Berg,[2] Daniel Hermanson,[1] Tobias Schmidt,[3] Daniel Zolg,[3] Martin Frejno,[3] and Bernard Delanghe,[2] [1]Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose, CA 95134, [2]Thermo Fisher Scientific, Hannah-Kunath-Strasse 11, Bremen, Germany, [3]MSAID, GmbH, Lichtenbergstr. 8, Garching b. München, Germany

## Abstract

**Purpose:** Here we show how the CHIMERYS™ intelligent search algorithm significantly increases the number of identified and quantified peptides for label-free quantification. We also show that CHIMERYS combined with other search approaches in the Thermo Scientific™ Proteome Discoverer™ software framework can further improve results.

**Methods:** Human and E coli proteomes were mixed and run on a Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer, acquiring MS/MS spectra using ion trap HCD MS/MS. Data were searched using Proteome Discoverer 3.0 software using three different processing workflows using Sequest™ HT, CHIMERYS, and a combined 3 search node strategy, each of which also performed label-free quantification with the Minora Feature Detector.

**Results:** CHIMERYS produced a significant increase in the number of identified and quantified peptides and proteins. A combined search strategy using three different search methods further improved results, producing 10,045 proteins and 144,709 unique peptides. The human and E coli protein ratios were calculated with high precision.

## Introduction

Current DDA-based approaches for label-free quantification match peptide identifications from a search engine to features detected independently by a quantification algorithm. With higher resolution instruments with MS1 resolution up to 480,000, it is quite common for quantification algorithms to detect substantially more than 100,000 features in a 2-hour LC/MS run. However, typical search engines can usually only identify about half of these features. Peptides may remain unidentified due to unexpected post-translation modifications of enzyme cleavage specificity, but quite frequently peptides are simply not identified because they are co-isolated and fragmented with several other peptides to produce a chimeric spectrum.

The CHIMERYS intelligent search algorithm helps address this issue by identifying significantly more peptides in chimeric spectra, dramatically increasing the number of peptide identifications in complex proteomics LC-MS/MS runs. For high resolution MS/MS data, CHIMERYS can identify more than 10 PSMs in the same MS/MS spectrum. We show here that the pairing of CHIMERYS with the Minora feature detection algorithm in Proteome Discoverer 3.0 software leads to an increase in the number of quantified peptides and proteins in complex proteomics data compared to a Sequest HT-based approach. We further show that CHIMERYS can be combined with other search approaches including INFERYS™ Rescoring to further increase the number of identified and quantified peptides.

## Materials and methods

Increasing amounts of an *E. coli* tryptic digest sample (0, 25, 50, 100, and 200 ng) were mixed with a human tryptic digest to make a total of 1 µg sample per mixture. Each sample was run in triplicate using an Orbitrap Eclipse Tribrid mass spectrometer equipped with a Thermo Scientific™ FAIMS Pro™ interface, using a 2-hour gradient and two compensations voltages (-55 and -70 V). MS scans were detected at 240K resolution and HCD MS/MS scans were collected in the ion trap using a 0.7-m/z isolation window with a 1 s cycle time.

All data analysis was performed using Proteome Discoverer 3.0 software. The first analysis used a Sequest HT-based Processing workflow for label-free quantification (Figure 1). The Spectrum Files RC node performs mass recalibration for both the precursors masses for the Sequest HT search and the LCMS Features created by the Minora Feature Detector. The nodes used a combination of the March 2022 canonical human and E coli proteome FASTA databases downloaded using the Proteome Discoverer annotation service.

Data were also analyzed using a CHIMERYS-based Processing workflow (Figure 2). The CHIMERYS workflow includes a “Top N Peaks Filter” node prior to the search, passing through the top 20 peaks per 100 m/z window for MS/MS spectra. The filter significantly improves CHIMERYS results for ion trap MS/MS data but is not recommended for high resolution MS/MS data. The CHIMERYS search node used a fragment tolerance of 0.3 Da, fixed Carbamidomethyl cysteine and oxidized methionine, and the inferys_2_1_fragmentation prediction model.

A standard label-free quantification Consensus workflow was used for both workflows using default parameters except requiring a channel occupancy of 50% in the Precursor Ions Quantifier node. p-values were calculated using the background-based method and ratios were calculated using the pairwise ratio approach. All protein, peptide group, and PSM counts are reported at a 1% FDR threshold, using Percolator for PSM-level FDR for all 3 search approaches, qvality for peptide group FDR, and the Protein FDR Validator node for Protein FDR.

**Figure 1. Sequest HT-based processing workflow for label-free quantification.**

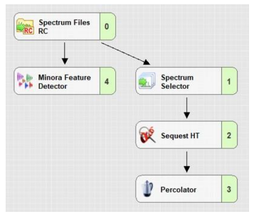
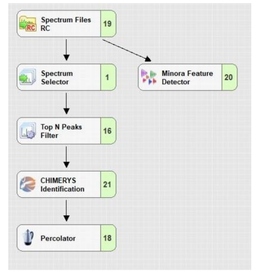
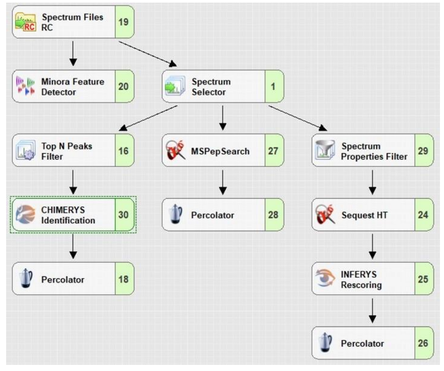



**Figure 2. CHIMERYS-based processing workflow for label-free quantification.**




Data were subsequently analyzed using a branched workflow. While CHIMERYS currently supports enzyme-specific cleavage and peptides with fixed carbamidomethyl cysteine, TMT™, or TMTpro modifications, other nodes can be used to identify peptides produced by semi-specific cleavage and peptides with post-translational modifications. Sequest HT with INFERYS™ Rescoring[1] has been shown to significantly increase the number of peptide identifications for peptides produced by non-specific cleavage. To speed up the search, the FASTA databases for the human and *E. coli* proteomes were pre-indexed prior to search. Another branch of the workflow included the MSPepSearch algorithm using the NIST OT HCD library to identify proteins commonly detected in human samples. The combined processing workflow with all 3 search nodes is shown in Figure 3.

**Figure 3. Branched processing workflow for label-free quantification with 3 search nodes and the Minora Feature Detector. The Sequest HT and INFERYS Rescoring search were used to identify semi-tryptic peptides for both the human and *E coli* proteomes, while the MSPepSearch node was used to identify commonly detected peptides with post-translational modifications.**




## Results

**Figure 4. Comparison of proteins, unique peptides, and PSMs between the Sequest HT and CHIMERYS searches.**

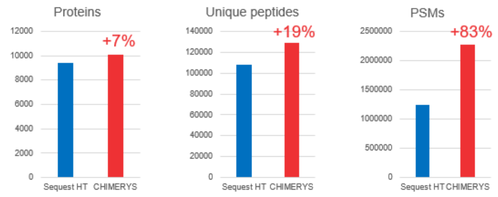
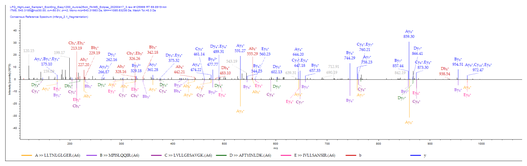



Figure 4 shows a comparison of the protein and peptide identification results from the Sequest HT and CHIMERYS workflows. CHIMERYS produces 19% more unique peptides and 7% more proteins compared to Sequest HT, with an equal increase in protein identifications for both the human and *E coli* proteomes. While this increase is less than the 50-70% observed for high resolution MS/MS data, the improvement is still significant given the nominal resolution and mass accuracy of ion trap MS/MS data. The CHIMERYS search produced up to 6 PSMs in a single MS/MS spectrum while the Sequest HT search can only identify 1 PSM. Figure 5 shows one example of an MS/MS spectrum where 5 peptides were simultaneously identified by CHIMERYS.

**Figure 5. Fragment Match Spectrum for an MS/MS scan with 5 PSMs identified using CHIMERYS. The lower spectrum in the mirror plot shows the INFERYS-predicted combined MS/MS spectrum for all 5 peptides. Almost every abundant peak in the spectrum has been assigned as a fragment.**



The number of quantified *E coli* proteins by search engine is shown in Figure 6. For each concentration level, the CHIMERYS results produce about the same 7% increase in the number of quantified proteins as is seen for the number of protein IDs.

**Figure 6. Number of quantified *E. coli* proteins by amount of sample. There are about 6% more quantified proteins using CHIMERYS at each concentration level. The number of quantified human proteins are similar across all samples and, thus, not displayed here. CHIMERYS quantified about 7% more human proteins for each sample compared to Sequest HT.**

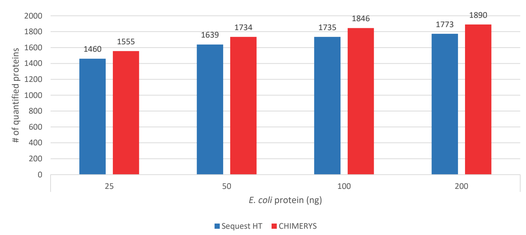
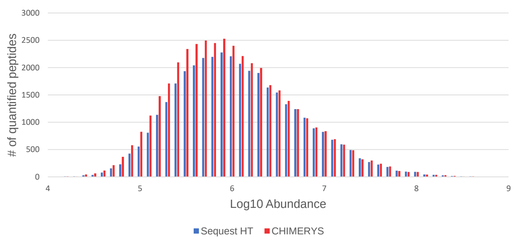


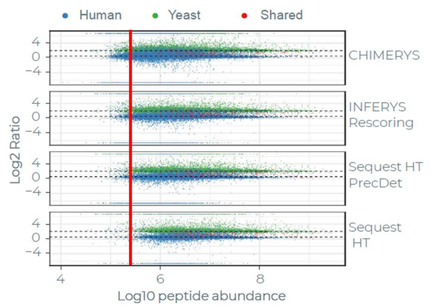

Figure 7 shows the distribution of *E. coli* peptide abundances for the three 200 ng *E. coli* datasets. The higher abundance peptides are quantified by both workflows, while the lower abundance peptides are more frequently quantified using the CHIMERYS-based workflow.

**Figure 7. Distribution of peptide group abundances for the 200 ng *E. coli* samples. CHIMERYS identifies more low abundance peptides with Log10 intensities below 6.**




The dynamic range is improved due to the ability of CHIMERYS to identify lower abundance peptides in mixed spectra. For nominal mass, this improvement is relatively subtle, but can be seen more clearly when searching high resolution MS/MS data. Figure 8 shows the improvement in the dynamic range for identification using a high-resolution MS/MS dataset with human and yeast proteomes. The CHIMERYS results show the high dynamic range of quantification.

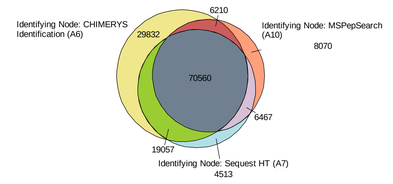
**Figure 8. Distribution of peptide group abundances for a human/yeast proteome mixture acquired using Orbitrap-based HCD MS/MS and processed with different search strategies.**




**Results from combined 3 search node method**

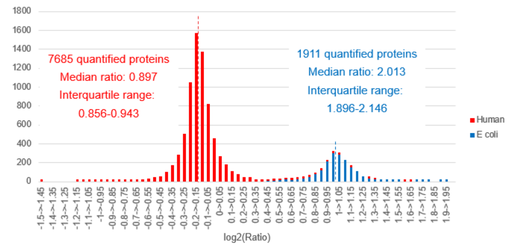
The combined workflow produced a similar number of protein identifications (10045) as the CHIMERYS search alone (10120), but there was a moderate increase in the number of unique peptide identifications (+12%). The overlap in identifications between the three different search strategies is shown in Figure 9. Many of the identifications that are unique to Sequest HT and INFERYS Rescoring correspond to semi-tryptic peptides and many unique to MSPepSearch results include modified peptides.

**Figure 9. Venn diagram showing the overlap in unique peptide IDs for the 3 different search nodes. While CHIMERYS produced by far the most identifications, both MSPepSearch and Sequest HT identified several thousand peptides not identified by CHIMERYS.**




The Minora Feature Detector and Feature Mapper nodes were improved for the Proteome Discoverer 3.0 software release to better support CHIMERYS search results. As a result, there is substantially reduced false quantification in Proteome Discoverer 3.0 software than past releases for label-free quantification. Figure 10 shows the distribution of quantification values for the human and *E. coli* proteins for the comparison of the 200 ng and 100 ng *E. coli* samples.

**Figure 10. Histogram of log2 ratios for the human and *E. coli* proteins for the 3 search node results. The dashed lines show expected ratios.**




Of the 144700 identified peptides, 104916 had measured abundances for the 200 ng samples, 106738 had measured abundances for the 100 ng samples, 106193 had measured abundances for the 50 ng samples, 104112 for the 25 ng samples, and 97247 for the 0 ng samples. The lower abundance samples are expected to have fewer quantified peptides due to the decreasing amounts of *E. coli* peptides.

## Conclusions

CHIMERYS produces a substantial increase in identifications compared to a standard Sequest HT search. CHIMERYS is not only able to identify more peptides in these mixed spectra, but also the additional peptides in the same spectra are usually lower in abundance than the peptide identified by Sequest HT. Thus, when CHIMERYS is paired with the Minora Feature Detector, the workflow will be able to identify lower abundance peptides than a traditional search engine. Subsequent work after these data were acquired shows that CHIMERYS can identify even more peptides from ion trap MS/MS data of human samples using a wider isolation window (1.5-m/z) and a higher maxIT (15 ms) than what was used for these data (0.7-m/z and 10 ms). We would thus expect another substantial increase in identified and quantified peptides using this updated wide window acquisition method that takes better advantage of the CHIMERYS algorithm.

## References

1. “INFERYS rescoring: Boosting peptide identifications and scoring confidence or database search results”, Rapid Commun. Mass Spectrom., 2021

## Trademarks/licensing



Learn more at thermofisher.com/ProteomeDiscoverer