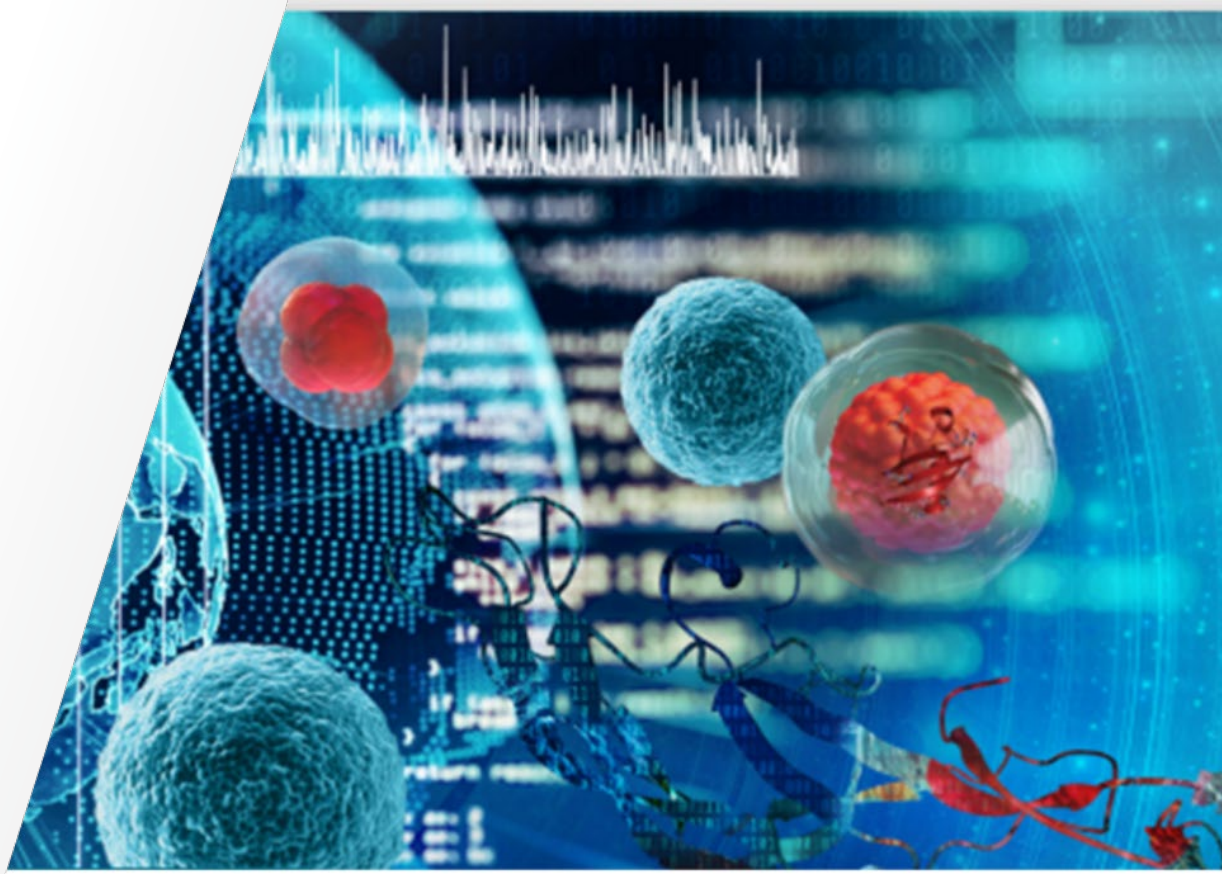
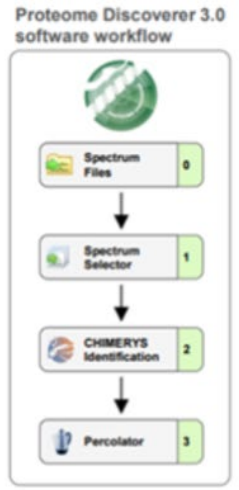
Boosting Proteome Coverage by a Combination of the Next Generation UHPLC and a Novel Search Node in Proteome Discoverer

Tabiwang N. Arrey[1]

Eugen Damoc[1] Bernard Delanghe[1]; Runsheng Zheng[2]; Christopher Pynn[2]; Oleksandr Boychenko[2]; Sean Woodmansey[3]; Wim Decrop[2]; Paula McBride[3]; Alexander Harder[1]

[1]Thermo Fisher Scientific, Bremen, Germany; [2]Thermo Fisher Scientific, Germering, Germany; [3]Thermo Fisher Scientific, Hemel Hempstead, United Kingdom





The world leader in serving science

Abstract

- Due to the substantial number of MS and MS/MS scans collected during shotgun proteomics experiments, data interpretation is only possible using advanced database search engines. However, despite using advanced search algorithms, many fragment ions in the spectra remain unassigned. Some of these fragment ions may belong to peptides carrying modifications that are not considered during database search, but also to peptides that are co-isolated and co-fragmented. Here, we evaluate a novel search algorithm, CHIMERYS™, which is available in Proteome Discoverer™ 3.0 software and that can identify multiple peptides in a single spectrum.

Introduction

Primary Challenge

- In bottom-up proteomics MS\MS spectra often contain fragment ions from multiple co-isolated peptides, which in most cases not taken into account when doing data interpretation with existing search algorithms.

Prior efforts in the field

- Implementation of Iterative database search algorithms or by the use of combinations of search algorithms. However, since most of these search algorithms do not take into account that the measured intensities of fragment ions could be a sum of multiple peptides, thus, valuable information is not used during data processing. This leads to fewer peptides and protein being identification.

Novel approach

- High-throughput methods using the Vanquish Neo UHPLC system and a PepMap Neo column in combination with a novel search algorithm, CHIMERYS to boost proteome coverage.

Overview content

- Reproducible and consistent chromatography
- Comparison of CHIMERYS to Sequest HT, Sequest HT with INFERYS rescoring for short and long gradients.

Summary

- CHIMERYS outperforms Sequest HT and Sequest HT algorithms with INFERYS rescoring in both semi-complex and a complex proteome for both long and short gradient. However, the highest benefits are seen in shorter gradients.

Materials and Methods: Sample preparation

- Thermo Scientific™ Pierce™ HeLa Digest Standard (20 µg/vial) was reconstituted by adding 200 µL of 5% ACN in 0.1% formic acid (FA) in water. Sample was aspired and pipetted approximately 10 times and then transferred to an autosampler vial. 2 ul of this solution was injected on to the trapping column.
- Yeast protein Digest (Promega, 100ug) was reconstituted in 200µl of 5% ACN in 0.1% formic acid (FA) in water. It was further diluted with 0.1 % formic acid to obtain a concentration of 500 ng/ul. Sample was aspired and pipetted approximately 10 times and then transferred to an autosampler vial. 1 ul of this solution was injected on to the trapping column.

Materials and Methods: MS Methods

The analytical column was connected to an Easy-Spray Source and maintained at 50 °C. Eluting peptides were sprayed on to the Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer and measured.

180,100 and 60 Samples Per Day, SPD

Trap and Elute Mode

Full Scan	
Orbitrap resolution	45,000
Scan Range	350-1,200
AGC Target (%)	300
injection time	20ms
RF Lens (%)	50
Intensity threshold	
Min. Intensity	5.00E+03
Charge state filter	
Charge state	2-5
Dynamic Exclusion	
Exclude after n times	1
Exclusion duration	25
Exclude isotopes	TRUE
Data dependent properties	
TopN	40
Isolation width	2
NCE (%)	28
Orbitrap Resolution	7,500
Injection time	12 ms
AGC Target (%)	50

30 and 24 SPD

Trap and Elute Mode

Full Scan	
Orbitrap resolution	60,000
Scan Range	350-1,200
AGC Target (%)	300
injection time	20ms
RF Lens (%)	50
Intensity threshold	
Min. Intensity	5.00E+03
Charge state filter	
Charge state	2-5
Dynamic Exclusion	
Exclude after n times	1
Exclusion duration	45
Exclude isotopes	TRUE
Data dependent properties	
TopN	30
Isolation width	2
NCE (%)	28
Orbitrap Resolution	15,00
Injection time	Auto
AGC Target (%)	50

60 and 90 min gradients

Direct injection

Full Scan	
Orbitrap resolution	120,000
Scan Range	350-1,200
AGC Target (%)	300
injection time	20ms
RF Lens (%)	50
Intensity threshold	
Min. Intensity	8.00E+03
Charge state filter	
Charge state	2-5
Dynamic Exclusion	
Exclude after n times	1
Exclusion duration	45
Exclude isotopes	TRUE
Data dependent properties	
TopN	20
Isolation width	2
NCE (%)	28
Orbitrap Resolution	15,000
Injection time*	Auto
AGC Target (%)	50

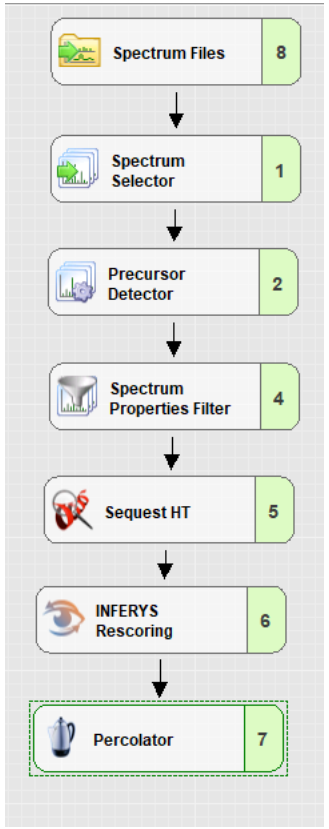
*40 ms for 90 min gradient

Materials and Methods: Data processing method

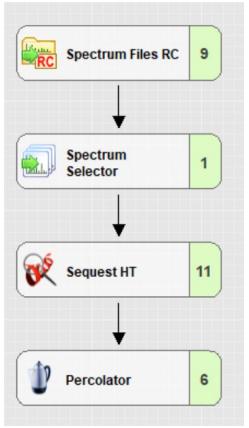
The acquired raw data files were processed with Thermo Scientific™ Proteome Discoverer™ 3.0 using Sequest HT, Sequest HT with INFERYS rescoring and CHIMERYS. PSM and Peptide validation is performed using Percolator.

Processing Workflows

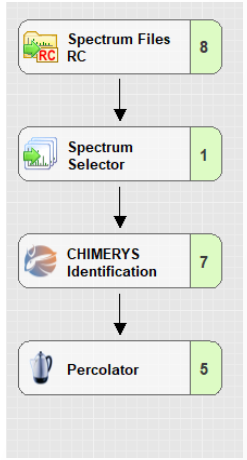
INFERYS_Rescoring_SequestHT_Percolator



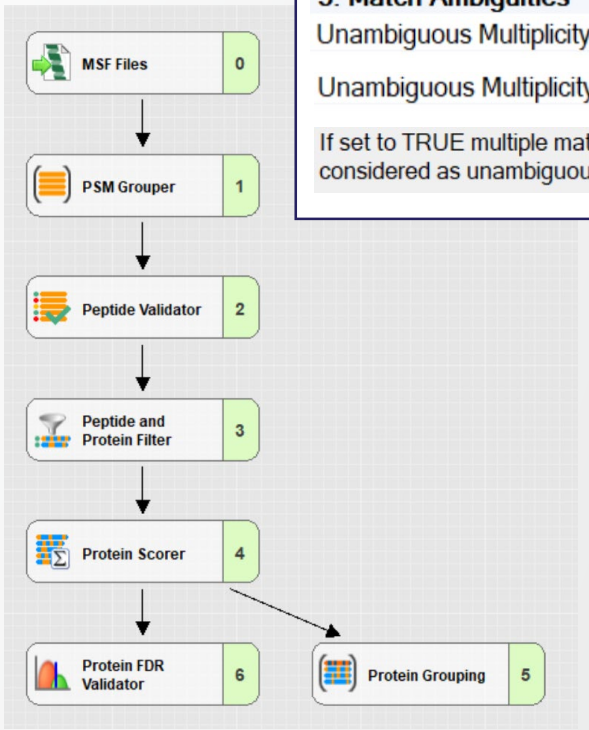
Sequest HT_Percolator



Chimerys_Percolator



Basic Consensus Workflow



5. Match Ambiguities

Unambiguous Multiplicity False → Sequest HT

Unambiguous Multiplicity True → Inferys & Chimerys

If set to TRUE multiple matches per spectrum (chimeric matches) are considered as unambiguous by definition.

Materials and Methods: LC parameters

LC parameters: Direct Injection mode

- Column: Thermo Scientific™ Easy-Spray™ PepMap™ Neo 2 µm C18 75 µm X 500 mm
- Flow rate: 250 nl\min
- Gradient
  - 60 min (80 min run time)
  - 90 min (120 min run time)

60 min gradient: 8-25 % B in 50 min, 10 min to 40 %B

90 min gradient: 8-25 % B in 80 min, 10 min to 40% B

LC parameters: Trap and Elute mode

- Easy column: Thermo Scientific™ Easy-Spray™ PepMap™ Neo 2 µm C18 75 µm X 150 mm
- Trapping Column: Thermo Scientific™ PepMap™ Neo 5 µm C18 300 µm X 5 mm Trap Cartridge
- Flow rate: 1.3 to 0.4 ul/min
- Gradients:
  - 5.5 min (180 Samples Per Day, SPD),
  - 11 min (100 SPD),
  - 20.1 min (60 SPD),
  - 44.4 min (30 SPD),
  - 56.4 min (24 SPD)

Solvent A: 0.1 % FA

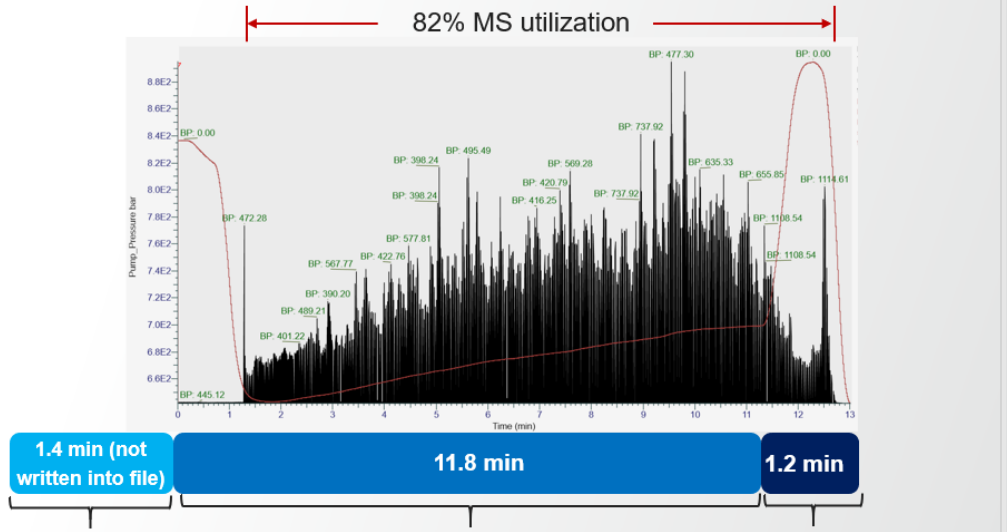
Solvent B: 80 % ACN, 0,1%FA

Weak wash solution: 0.1 % FA

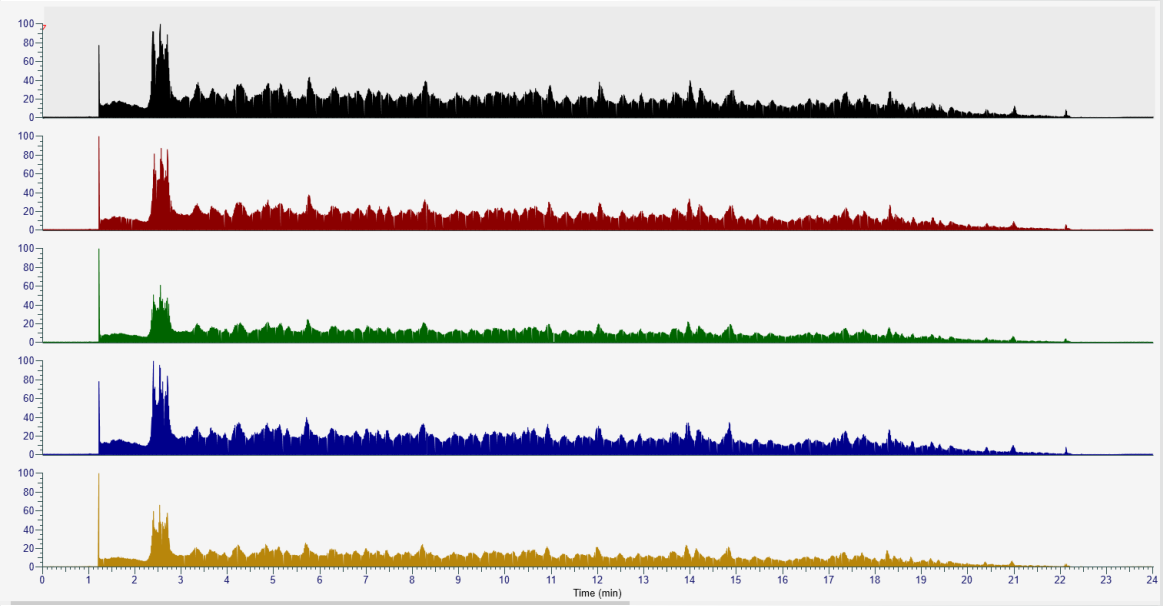
Strong wash solution: 80 % ACN, 0,1%FA

Results: Productivity and Repeatability

- Maximizing the gradient is a key to obtaining the most out of every gradient in bottom-up proteomics. Through parallel gradient separation and sampler washing procedures along with fast sample loading on the Vanquish Neo up to 82% of the time is used to acquire useful MS spectra.



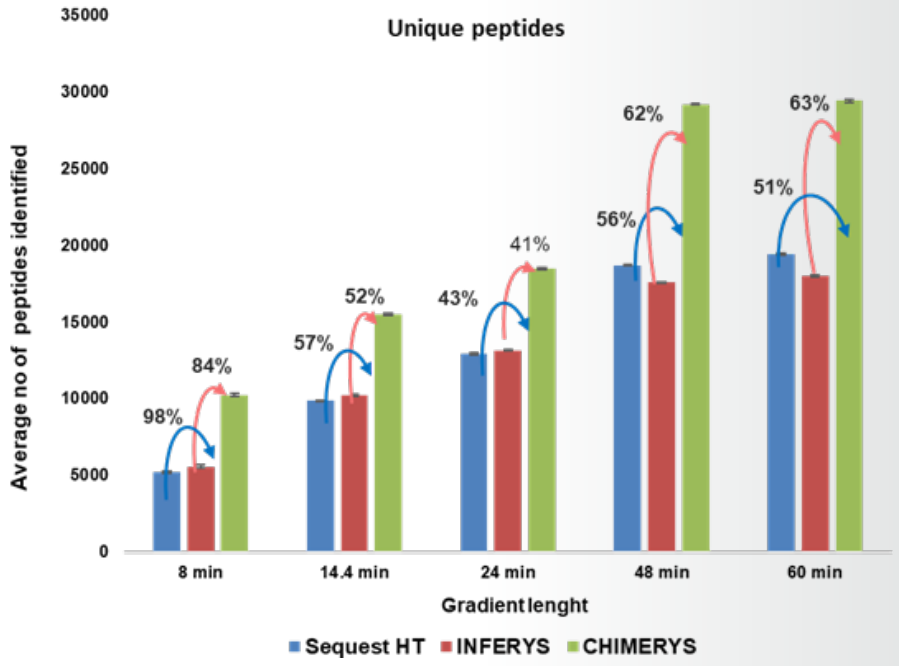
- Reproducible chromatographic and maximum MS utilization times.



Results: Comparison of different search algorithms

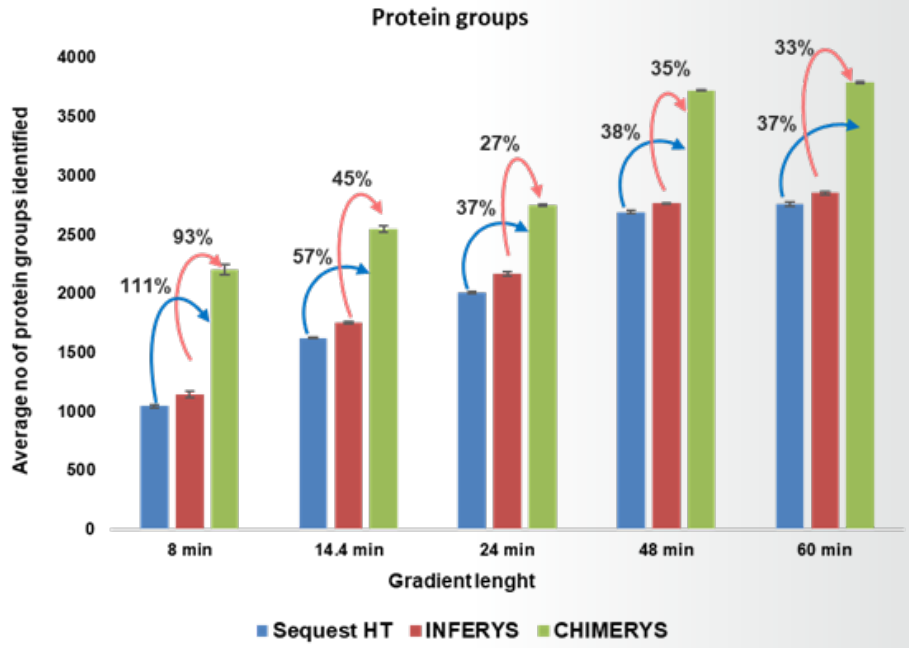
Yeast protein digest (500ng): Unique peptides

- Minimal variation between injections of the same gradient.
- The average number of peptides identified increased with gradient length for all the search algorithms, with CHIMERYS showing the highest identification.



Yeast protein digest (500ng): Protein groups

- The percentage increase in the average number of protein groups identified decreases with gradient length, nevertheless CHIMERYS still outperforms the other search algorithm used in these studied.

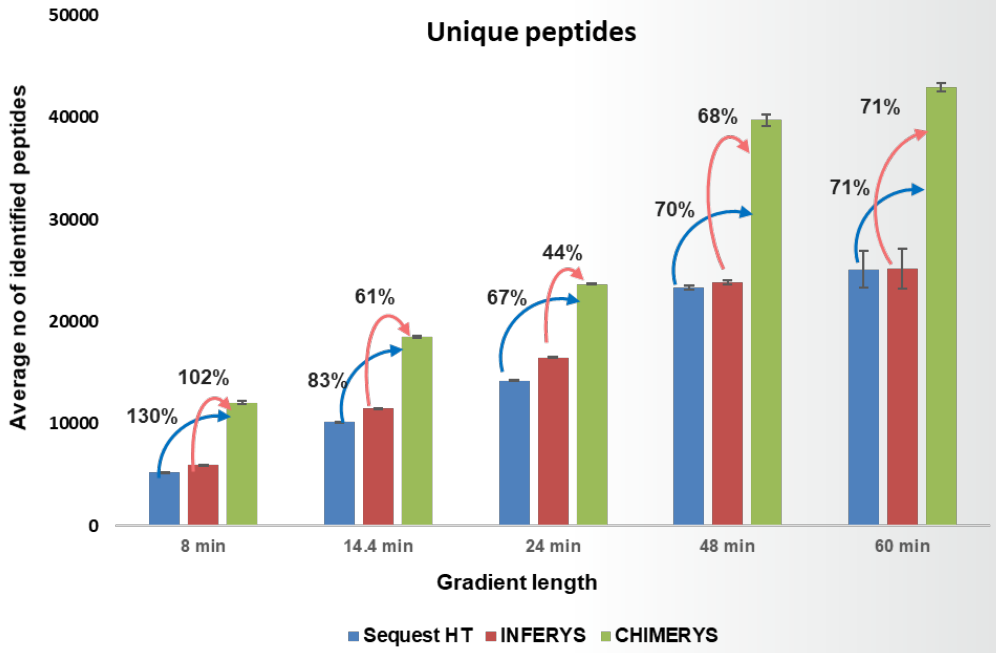


Trap and Elute mode

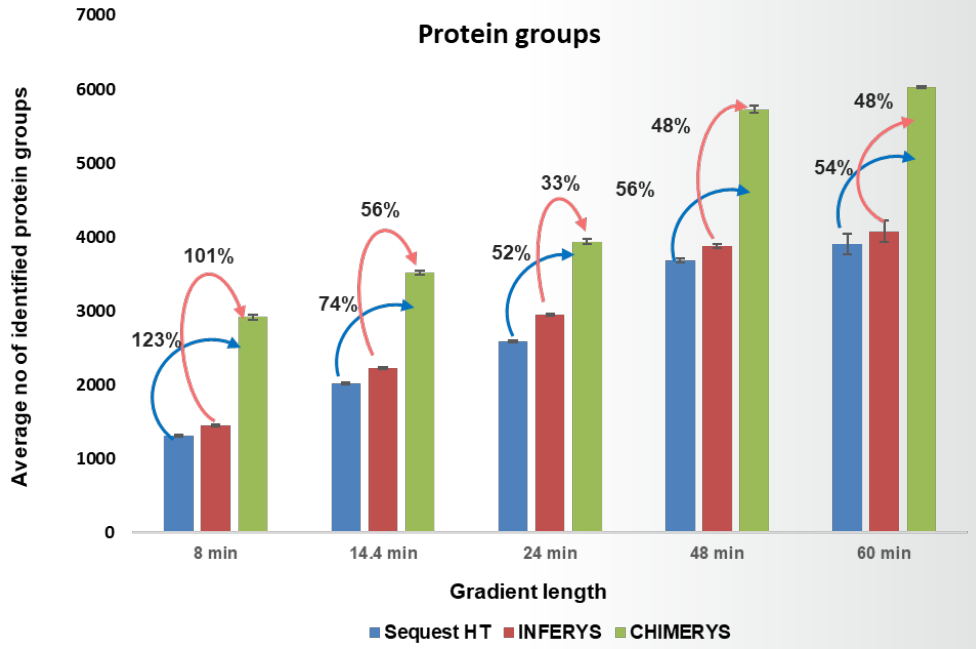
Results: Comparison of search algorithm

HeLa digest (200ng): Trap and Elute mode

- Also, for a complex proteome, CHIMERYS shows its superiority to the other search algorithms.



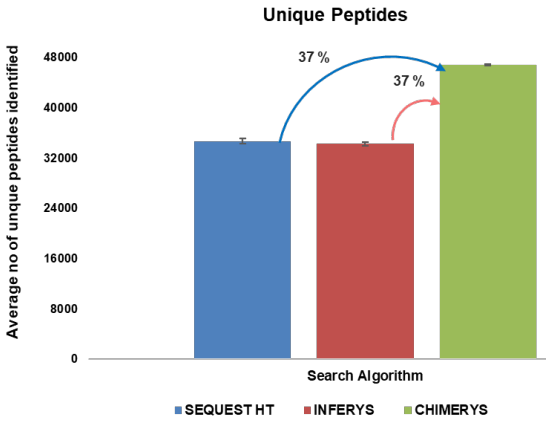
HeLa digest (200ng): Trap and Elute mode

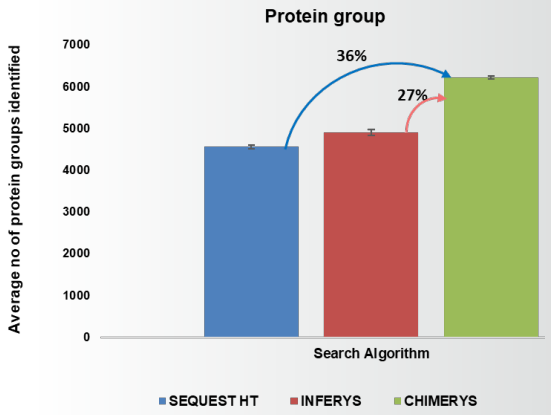


Results: Comparison of search algorithm

HeLa digest (200ng): Direct Injection mode, 60 min

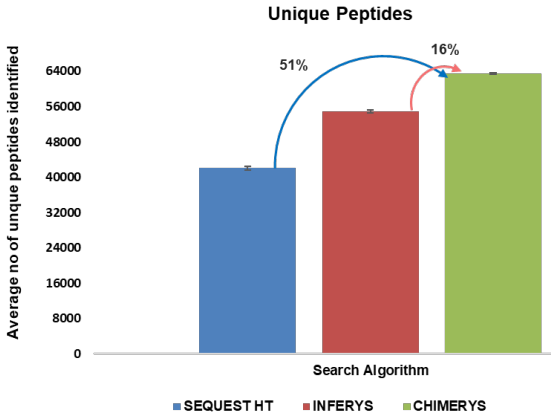
- CHIMERYS still perform well with longer gradients

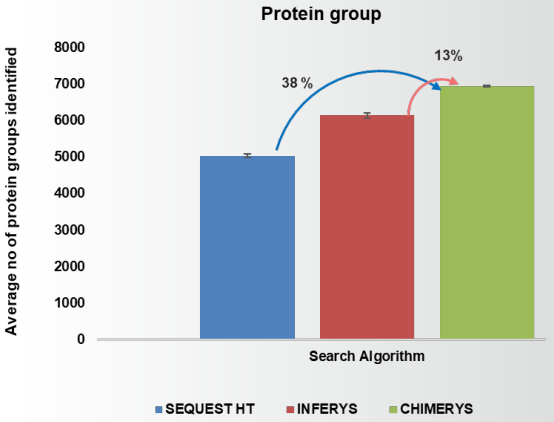




HeLa digest (200ng): Direct Injection mode, 90 min

- Even though spectra complexity is reduced by extending the gradient length, CHIMERYS is still able to confidently pick up co-fragmented precursors and identify them.





Conclusions

- Due to its parallelized operation, up to 82 % MS utilization time is obtained on the Vanquish Neo with reproducible gradients and consistence results.
- Compared to Sequest HT and Sequest HT with INFERYS rescoring, the results obtained using the CHIMERYS search node show tremendous increase in the number of peptides and proteins identified from yeast- and HeLa digest.
- An increase of 30% in protein\peptide identifications for the longest gradient and of more than 80% for the shortest gradients.
- Going from 2- 4Th isolation width there observed a slight increase in the number of identified peptides. However, the protein group numbers did not change much.