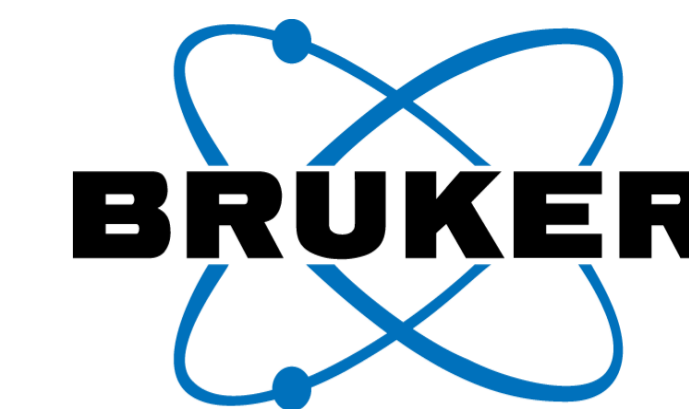


# PASEF and Bolt: enabling comprehensive sample analysis via high MS/MS acquisition and MS/MS sequencing through vast protein databases in minutes

Amol Prakash<sup>1</sup>, Swetaketu Majumder<sup>1</sup>, Jeifie Tong<sup>2</sup>, Shenheng Guan<sup>2</sup>, Matt Willets<sup>3</sup>, Frederic Lamoliatte<sup>4</sup>, Tharan Srikumar<sup>3</sup>, Sven Brehmer<sup>4</sup>, Bin Ma<sup>5</sup>, Michael Moran<sup>2</sup>



ASMS 2020, ThP290

1. Optys Tech Corporation, Shrewsbury, MA, USA
2. Sick Kids Research Center, Toronto, ON, Canada
3. Bruker Scientific, Billerica, MA, USA
4. Bruker Daltonik GmbH, 28359 Bremen, Germany
5. Univ. of Waterloo, Waterloo, ON, Canada

## Introduction

In a standard 120 min LC run, more than 800,000 fragmentation spectra are easily achievable, which can be used for near exhaustive precursor selection in complex mixtures or resequencing weak precursors. Unfortunately, a sub optimal data processing often hinders the full potential of this increase in acquisition speed and sensitivity. In almost all MS/MS peptide search engines, users must limit their search space to a canonical database due to time constraints and q value considerations. In addition, engines will nearly always assume the presence of only fully tryptic peptides and limit PTMs to a handful. Even on high-performance servers, these search engines are computationally expensive, and most users decide to dial back their search parameters.

We recently presented Bolt, a new cloud-based search engine that can search more than 2.4 million protein sequences (canonical, isoform, mutations, and contaminants) with 41 post-translation modifications and N-terminal and C-terminal partial tryptic search in minutes on a standard configuration laptop. Here we extend Bolt and Pinnacle to natively support timsTOF Pro (PASEF<sup>®</sup>) datasets fully utilizing the ion-mobility space for label-free quantitation.

## Methods

- 1) 200ng K562 – 90 min LC
- 2) 1ng, 10ng and 50ng – 60 min LC in duplicate
- 3) 200ng patient derived xenograft lysate – 90min LC

All data were processed locally with the three software packages on a high-performance server with 128GB DDR4 RAM, dual 20-core Xeon processors and raid-0 NVMe SSD drives.

- 1) MaxQuant (v1.6.10.43)
- 2) Fragpipe (v12.2; MSFragger v 2.3; Philosopher v2.1.2)
- 3) Pinnacle (v 1.0.97.0; Bolt v 1.0.97.0)

Search parameters: Human (96,817 entries) or Human and mouse (67,671 entries) fasta files. Trypsin with 2 missed cleavages, Oxidation (M), Protein Nterm Acetylation as variable and CAM (C) as static.  
- Bolt Large-db: append **2.3 million** known human proteoforms; Semi-tryptic (N and C-term); **41 PTMs** to above

## Results

Table 1: Processing times (in minutes) for the three software and Bolt Large-db on 200ng K562 and the six-data dilution series.

Data set	MaxQuant	MSFragger	Bolt	Bolt Large-db
200ng K562 File# 4786	279	95	82	94
Dilution Series (6 data file)	292	190	186	212

Figure 1 : (a) Peptide count and (b) protein groups compared between MaxQuant, MSFragger, Bolt and Bolt Large-db for the 200ng injection as well as each of the six data files from the dilution series. (c & d) Illustrates the peptide overlap for the 50ng data file (both replicates combined)

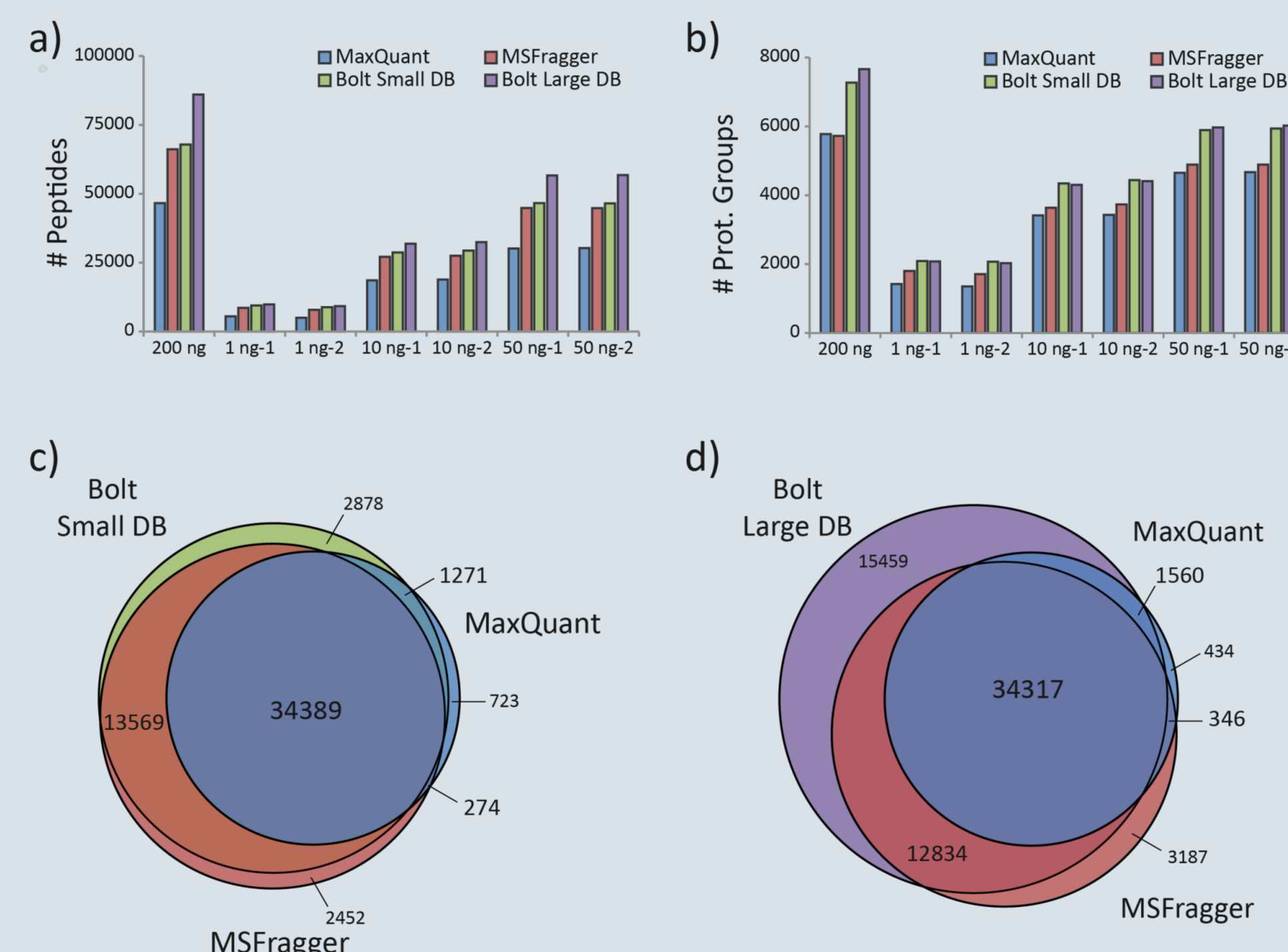
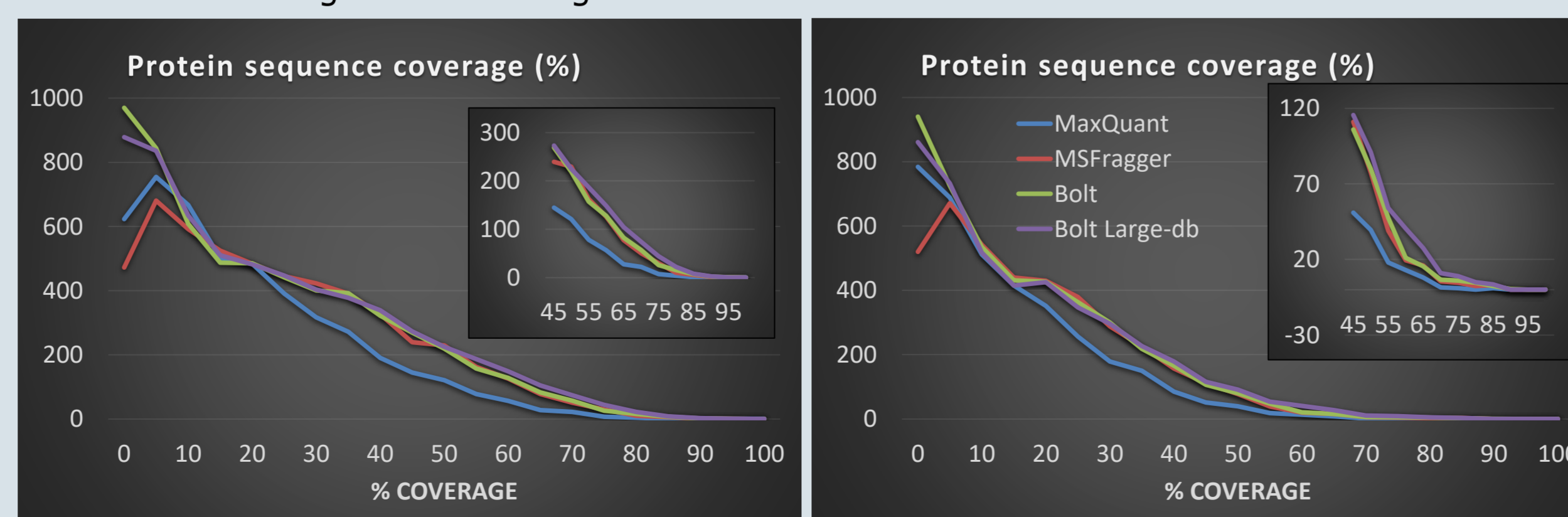


Figure 2 : Protein sequence coverage (%) compared for the (a) 50ng injection and (b) 10ng injections between each of the software. For both the panels, the inset image shows the zoomed in region for coverage >45%.



## Single 200ng injection of patient-derived xenograft (PDX)

A single 200ng injection of a PDX lysate analyzed with a 90min gradient yielded 6599 (33687), 7163 (57707) and 10629 (71060) protein groups (peptide counts) with MaxQuant, MSFragger and Bolt Large-db respectively.

Below two examples of some unique peptides identified by Bolt. (a) T75I mutation on protein ECHM\_HUMAN along with the wild type peptide (both forms observed). (b) EIAQDFKTDLR (protein: H3C\_HUMAN) observed in both methylated and dimethylated forms. (c) Shows distribution of the classes of the unique peptides identified by Bolt in the PDX sample.

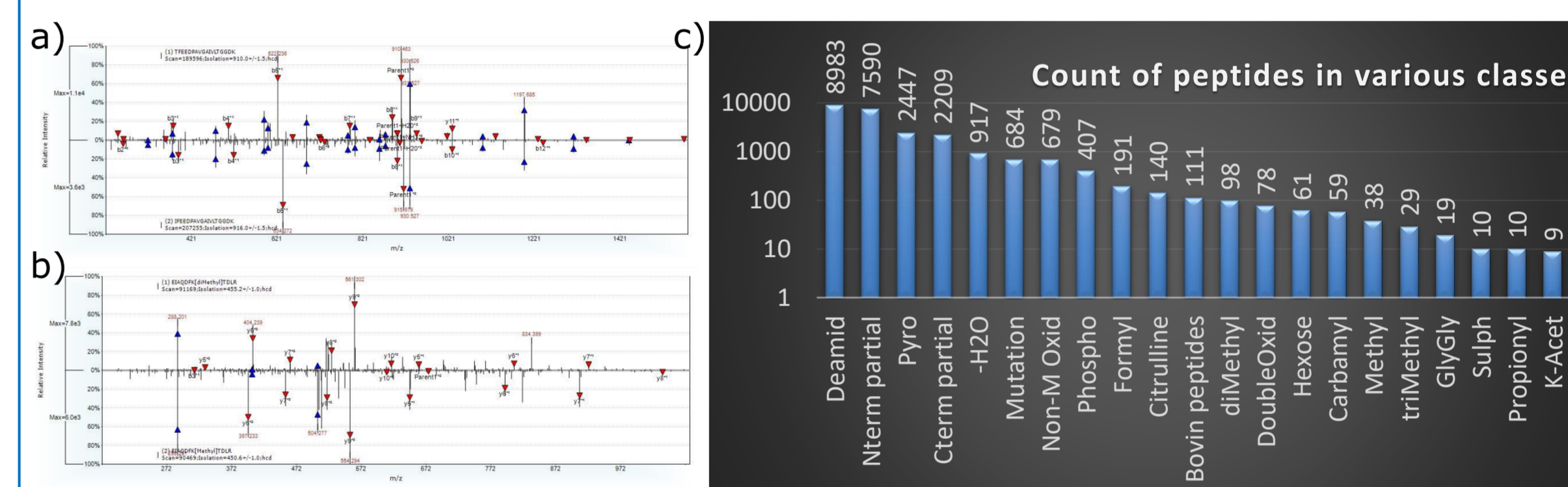
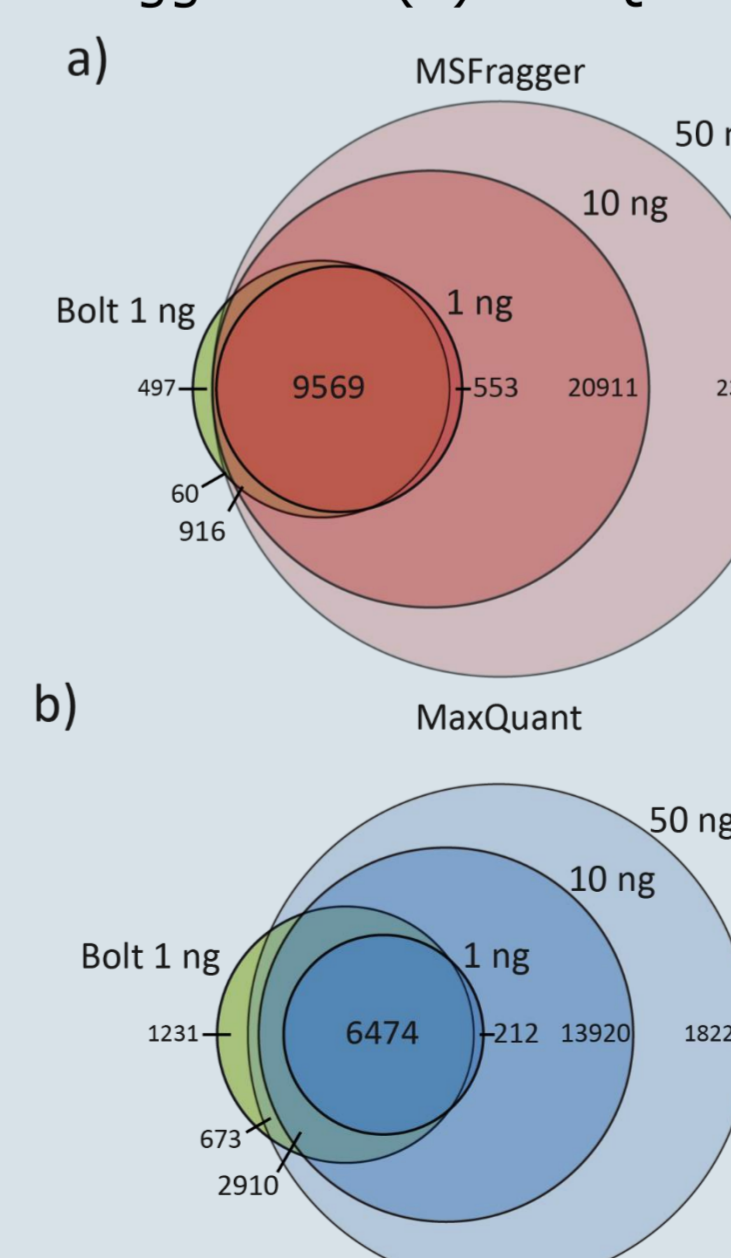


Figure 3 : Unique identifications by Bolt from the 1ng injection compared to higher load injections by (a) MSFragger and (b) MaxQuant.



## Conclusions

- Bolt identifies more peptides
  - 21% more than MSFragger
  - 81% more than MaxQuant
  - More than 2/3rd of Bolt's unique IDs are also confirmed by these software on the higher load injections
- Bolt uses significantly less processing time
- A combined PASEF<sup>®</sup> and Bolt workflow derives larger depth of biological information

timsTOF Pro