Exploring Depth and Breadth of a Protein Complex Mixture with Top-Down Data-Independent Acquisition Using an Orbitrap Fusion Tribrid Mass Spectrometer

Aaron O. Bailey, David M. Horn, Seema Sharma, Romain Huguet, Vlad Zabrouskov
Thermo Scientific San Jose, CA

Overview

Purpose: Extend dynamic range of detectable proteins in top-down proteomics

Methods: Compare data-dependent (DDA) and data-independent (DIA) MS2 acquisition using either ion trap or Orbitrap MS1 detection. Data were analyzed for DIA-based and DDA-based experiments automatically using ProSightPD in Thermo Scientific™ Proteome Discoverer™ software 2.0.

Results: MS1 detection in the Orbitrap vs. ion trap and MS2 acquisition using DDA vs. DIA methods offers complementary strengths depending on protein size, concentration, and microheterogeneity. With all methods combined we detected more than 160 proteoforms of 13 protein groups spanning 3 orders of magnitude from a 1 µg load.

Introduction

Top-down proteomics aims to complement bottom-up proteomics by identifying proteins in a mixture (via MS2) while providing data on proteoform distribution (via MS1). Protein charge state distribution presents a technical obstacle to realizing routine data-dependent (DDA) acquisition of top-down MS2 spectra. Intact proteins can produce widely ranging charge states dispersed across several hundred m/z units with a single charge state representing a mere fraction of the total ion current. For intact protein MS1 acquisition protein size, concentration, and complexity can skew detection based on chosen detector type. We present a comparison of 4 types of top-down proteomic workflows using an Orbitrap Fusion mass spectrometer. For acquiring MS1 scans we use either the high resolution Orbitrap (R=240K) or the low resolution, very high sensitivity ion trap. We also use the Orbitrap (R=240K) for MS2 acquisition, comparing conventional data-dependent acquisition **(DDA)** with a data-independent **(DIA)** approach. For DIA methods we used large isolation window sizes (100-250 m/z) to allow additional protein charge states to be included for fragmentation in order to increase the MS2 total ion current. We hypothesized that a data-independent (DIA) approach would allow us to collect and fragment multiple protein charge states for MS2 spectra to extend the dynamic range for discovery-mode experiments. We reasoned that this would have a more profound effect with larger proteins which typically have signals distributed across more charge states. We wanted to test which proteins were most amenable to detection when varying isolation window sizes to maximize analyte protein signal and minimize contaminating precursor species. In this report we demonstrate the unique strengths and weakness of top-down proteomic experiments using an Orbitrap or ion trap for MS1 acquisition and either DDA- or DIA-based MS2 acquisition: each different approach offers complimentary data.

Methods

Sample Preparation

A lyophilisate containing 10 µg Universal Proteomic Standard Dynamic Range (UPS2)[1] intact protein mixture (Sigma) was reconstituted using 40µL of 10% acetonitrile and 0.1% trifluoroacetic acid. Each vial contains 10 x 4 uL (1 µg) injections.

Liquid Chromatography

Intact protein mixture was injected directly onto a 200 µm x 25 cm Thermo Scientific™ EASY-Spray™ Pepswift monolith column and resolved with a 60 minute gradient of 10% to 60% acetonitrile and 0.1% formic acid at 4 uL/min using a Thermo Scientific™ Easy nLC1000™ HPLC.

Mass Spectrometry

Using a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass Spectrometer (Tune 1.0 software), DDA/DIA top-down data acquisition was accomplished using MS1 scans, performed in the Orbitrap (R=240K, 2e5 AGC, 2 µscans) or ion trap (normal mode, 5e4 AGC, 10 µscans) scanning from 150-2000 m/z. DDA experiments used a 5 m/z isolation width and 30 sec exclusion time. An additional charge state filter of z > +6 was used for OT-MS1-DDA run. DIA experiments used 100 or 250 m/z isolation widths serially across a range of 500-1500 m/z. Precursors were fragmented by HCD with 20% collisional energy. MS2 scans were performed in the Orbitrap (R=240K, 2.5e5 AGC, Max IT 200 ms).

Data Analysis

Top-down DDA and DIA data (MS1+MS2) were analyzed using TopDown cRAWler ("Low/High" ReSpect for IT MS1, "High/High" Xtract for OT MS1) and AbsoluteMass search in the ProSightPD node within Thermo Scientific™ Proteome Discover™ 2.0 software.

FIGURE1. UPS2 Dynamic Range intact protein mixture contains 48 proteins ranging in size (6-83 kDa) concentration (0.05-5000 fmol) for each 1 ug injection.
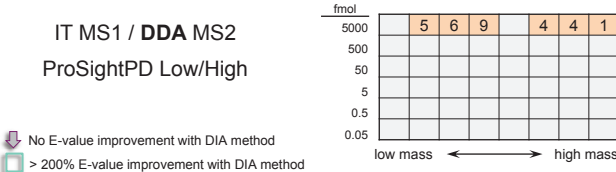
lowest difficulty									higher difficulty
5000 fmol	Complement C5 (8.6 kDa)	Ubiquitin (10.6 kDa)	Hemoglobin alpha (15.1 kDa)	Hemoglobin beta (15.8 kDa)	Leptin (16.2 kDa)	Carbonic anhydrase 1 (28.7 kDa)	Carbonic anhydrase 2 (29.1 kDa)	Serum albumin (66.4 kDa)	
500 fmol	EGF (6.3 kDa)	Cytochrome b5 (16.0 kDa)	Myoglobin (17.1 kDa)	PPCTI A (20.2 kDa)	Peroxiredoxin 1 (22.0 kDa)	NADPH DH (30.7 kDa)	SUMO-1 (38.8 kDa)	Catalase (59.6 kDa)	
50 fmol	NEDD8 (9.0 kDa)	Alpha lactalbumin (14.1 kDa)	Lysozyme C (14.7 kDa)	UBC9 (18.0 kDa)	RBP 4 (21.1 kDa)	RDHN DH (25.8 kDa)	Creatine Kinase M-type (43.1 kDa)	Histidyl-tRNA synthatase (58.2 kDa)	
5 fmol	IGF-II (7.5 kDa)	Beta-2-microglobulin (11.7 kDa)	PDGF B (12.3 kDa)	Thioredoxin (12.4 kDa)	Gamma-synuclein (15.4 kDa)	BID death agonist (22.0 kDa)	GST A1 (25.5 kDa)	Antithrombin-III (49.0 kDa)	
0.5 fmol	Cytochrome C (11.6 kDa)	Interferon gamma (16.9 kDa)	E2 C (20.5 kDa)	GTPase Hras (21.3 kDa)	E2 E1 (22.2 kDa)	GST P (23.2 kDa)	Sero-transferrin (75.2 kDa)	Gelsolin (82.9 kDa)	
0.05 fmol	IL-8 (8.4 kDa)	FABP, Heart (14.7 kDa)	Superoxide Dismutase (15.8 kDa)	TNF (17.4 kDa)	C-reactive protein (23.0 kDa)	Annexin A5 (35.8 kDa)	Tau (45.7 kDa)	Lactoferrin (76.1 kDa)	
higher difficulty	low mass ←						→ high mass		highest difficulty

Results

Data Dependent vs. Data Independent Acquisition with *Ion Trap MS1*
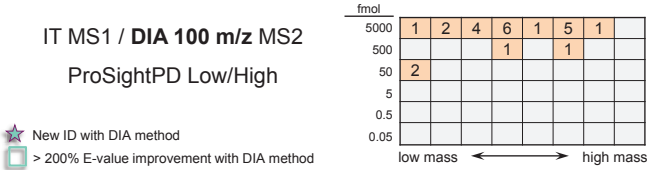
We used the ion trap for MS1 detection combined with either DDA or DIA and analyzed the results using the ProSightPD node in Proteome Discoverer software 2.0. Both approaches benefit from highly sensitive MS1 detection. A drawback to IT-MS1-DDA is that sensitivity gained in the IT is lost on an inability to differentiate charge states, which impairs data-dependency decisions. Only proteins in the top concentration tier (5 pmol each) were identified using DDA. IT-MS1-DIA was successful in extending the range of detection two orders of magnitude for small proteins as well as improving the ProSightPD quality scores (-Log E-Value) for MS2 spectra more than 2 fold. For the largest protein detected, Serum Albumin, DIA did not result in improved data. It is possible that the complexity of BSA proteoforms either did not produce additive signal increase with isolating multiple charge states or that too much non-BSA precursor contaminated MS2 spectra. Identifying protein groups which contain near-isobaric proteoforms can be a difficult challenge for top-down acquisition. Co-isolating several proteoforms with sufficient sequence commonalities, using either DDA or DIA, will produce great results due to an additive effect upon fragmentation. Likewise high microheterogeneity amongst related proteoform sequences can impede identification.

FIGURE6. Ion Trap MS1 w/ Data-Dependent MS2. Best -Log E-Values shown. Number of proteoforms per protein group detected from 1 µg load of UPS2 is plotted as a function of on-column concentration and intact molecular weight.
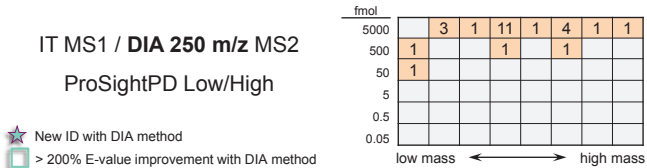

IT MS1 / **DDA** MS2

ProSightPD Low/High

⬇ No E-value improvement with DIA method

▢ > 200% E-value improvement with DIA method

Protein Descriptions	RT [min]	Ions Matched	-Log E-Value	Ion Inject Time [ms]	Form Theo. MH+ [Da]	Conc. [fmol]
Carbonic anhydrase 1	35.66	10/518	9.84	85.44	28722.35	5000
Carbonic anhydrase 2	34.76	25/516	34.12	149.21	29139.90	5000
Hemoglobin alpha	36.86	26/280	38.73	172.43	15117.89	5000
Hemoglobin beta	31.44	19/290	31.15	200.00	15858.26	5000
Serum albumin	34.97	11/1166	9.43	147.26	66314.91	5000
Ubiquitin	22.84	23/188	35.04	200.00	10591.57	5000

FIGURE7. Ion Trap MS1 w/ Data-*In*dependent MS2. Two isolation widths were tested: 100 (top panels) and 250 m/z (bottom panels). Best -Log E-Values shown.


IT MS1 / **DIA 100 m/z** MS2

ProSightPD Low/High

✩ New ID with DIA method

▢ > 200% E-value improvement with DIA method

Protein Descriptions	RT [min]	Ions Matched	-Log E-Value	Ion Inject Time [ms]	Form Theo. MH+ [Da]	Conc. [fmol]
Carbonic anhydrase 1	32.87	11/518	13.35	96.07	28789.28	5000
Carbonic anhydrase 2	31.93	34/516	37.19	0.42	29162.99	5000
Complement C5	19.50	15/148	8.31	10.54	8707.32	5000
Hemoglobin alpha	32.48	24/280	25.89	1.16	28804.00	5000
Hemoglobin beta	34.83	28/290	43.66	1.31	15871.94	5000
Leptin	42.14	20/292	9.50	17.61	16160.57	5000
NAD(P)H DH	40.31	30/546	33.63	0.84	30874.29	500
NEDD8	26.17	37/150	28.54	6.24	8562.82	50
PPCTI A	29.36	11/366	8.45	1.47	20045.81	500
Ubiquitin	23.40	56/188	71.32	5.12	10600.38	5000

IT MS1 / **DIA 250 m/z** MS2

ProSightPD Low/High

✩ New ID with DIA method

▢ > 200% E-value improvement with DIA method

Protein Descriptions	RT [min]	Ions Matched	-Log E-Value	Ion Inject Time [ms]	Form Theo. MH+ [Da]	Conc. [fmol]
Carbonic anhydrase 1	33.11	10/520	12.63	0.39	28895.41	5000
Carbonic anhydrase 2	32.11	35/516	36.73	0.57	29139.90	5000
Hemoglobin alpha	32.69	31/280	36.08	0.75	15117.89	5000
Hemoglobin beta	34.98	35/290	46.62	0.65	15900.27	5000
Leptin	42.13	17/292	8.72	0.32	16148.45	5000
NAD(P)H DH	40.23	26/546	31.59	0.62	30849.02	500
NEDD8	26.24	37/150	16.17	2.88	8555.67	50
PPCTI A	29.40	17/366	13.38	3.37	20032.84	500
EGF	25.67	9/104	7.69	0.69	6218.75	500
Serum albumin	33.79	9/1166	13.94	0.30	66314.91	5000
Ubiquitin	23.18	58/188	67.85	2.81	10591.57	5000

FIGURE8. Ubiquitin is shown in IT MS1 / OT MS2 scans acquired using DDA or DIA methods. The best-case MS1/MS2 match was from the DIA method, which resulted in 2-fold improved E-value compared to the DDA method.
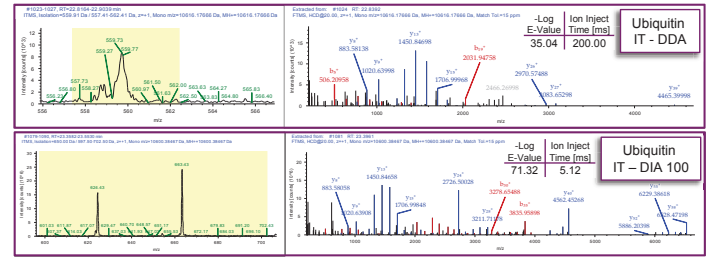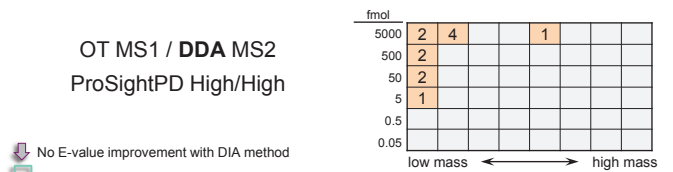

FIGURE9. IT MS1 / OT MS2 scans for NAD(P)H dehydrogenase (30 kDa, 500 fmol), detected using the DIA method, but not with the DDA method.
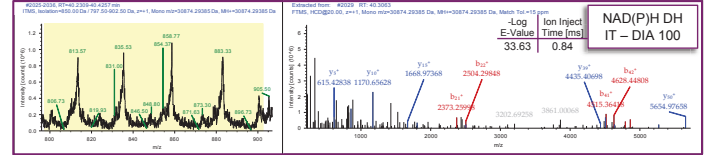
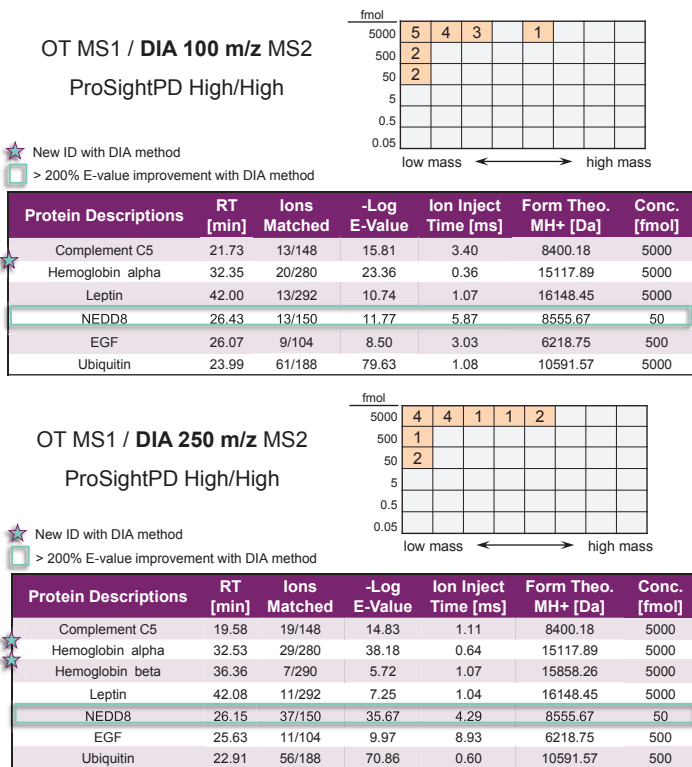### Data Dependent vs. Data Independent Acquisition with *Orbitrap MS1*

We performed a second set of DDA and DIA experiments using the Orbitrap for MS1 acquisition. We found that the high resolution of the Orbitrap, readily able to assign charge states, worked particular well for top-down detection of small proteins in very low abundance (IGF II, 5 fmol). For Orbitrap MS1 scanning, the DIA 250 m/z window method was able to increase the number of "mid-range" proteins detectable to include two additional proteins (Hemoglobin alpha/beta) which were barely visible by MS1 (required for ProSightPD match), but not in sufficient abundance to determine charge state.

FIGURE10. Orbitrap MS1 w/ Data-Dependent MS2.

OT MS1 / **DDA** MS2

ProSightPD High/High

⬇ No E-value improvement with DIA method

▢ > 200% E-value improvement with DIA method

| Protein Descriptions	RT [min]	Ions Matched	-Log E-Value	Ion Inject Time [ms]	Form Theo. MH+ [Da]	Conc. [fmol]
Complement C5	18.77	18/146	11.73	200.00	8269.14	5000
IGF II	25.16	11/132	7.91	173.86	7471.56	5
Leptin	41.42	21/292	11.55	167.56	16148.45	5000
NEDD8	26.10	19/150	8.27	200.00	8555.67	50
EGF	25.57	7/104	5.33	162.22	6218.75	500
Ubiquitin	23.33	49/188	60.14	187.39	10591.57	5000

FIGURE11. Orbitrap MS1 w/ Data-*In*dependent MS2. Two isolation widths were tested: 100 m/z (top panel) and 250 m/z (bottom panel)

OT MS1 / **DIA 100 m/z** MS2
ProSightPD High/High

☆ New ID with DIA method
▭ > 200% E-value improvement with DIA method

Protein Descriptions	RT [min]	Ions Matched	-Log E-Value	Ion Inject Time [ms]	Form Theo. MH+ [Da]	Conc. [fmol]
Complement C5	21.73	13/148	15.81	3.40	8400.18	5000
Hemoglobin alpha	32.35	20/280	23.36	0.36	15117.89	5000
Leptin	42.00	13/292	10.74	1.07	16148.45	5000
NEDD8	26.43	13/150	11.77	5.87	8555.67	50
EGF	26.07	9/104	8.50	3.03	6218.75	500
Ubiquitin	23.99	61/188	79.63	1.08	10591.57	5000

OT MS1 / **DIA 250 m/z** MS2
ProSightPD High/High

☆ New ID with DIA method
▭ > 200% E-value improvement with DIA method

Protein Descriptions	RT [min]	Ions Matched	-Log E-Value	Ion Inject Time [ms]	Form Theo. MH+ [Da]	Conc. [fmol]
Complement C5	19.58	19/148	14.83	1.11	8400.18	5000
Hemoglobin alpha	32.53	29/280	38.18	0.64	15117.89	5000
Hemoglobin beta	36.36	7/290	5.72	1.07	15858.26	5000
Leptin	42.08	11/292	7.25	1.04	16148.45	5000
NEDD8	26.15	37/150	35.67	4.29	8555.67	50
EGF	25.63	11/104	9.97	8.93	6218.75	500
Ubiquitin	22.91	56/188	70.86	0.60	10591.57	500

FIGURE12. OT MS1 / OT MS2 scans are shown for very low abundant (5 fmol) small protein (7.5 kDa) IGF II, which was detected only using the DDA method and not detected using the DIA method.
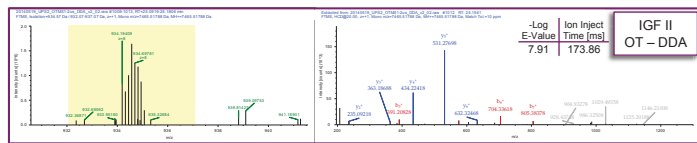
FIGURE13. Microheterogeneity (see Figures5-7) within this high concentration (5 pmol) protein group, Hemoglobin alpha (15.1 kDa), produces a highly divided, low intensity ion current. This signal is observed in OT MS1 scans, but is not robust enough for automatic charge state recognition and thus precludes DDA selection. Both the 100 and 250 m/z DIA methods produced data which could be automatically matched using ProSightPD. OT MS1 / OT MS2 scans are shown for Hemoglobin alpha acquired using the 250 m/z isolation window.
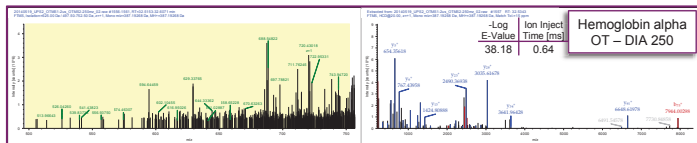
Conclusions

- The ProSightPD node with Proteome Discoverer 2.0 software is a powerful software platform for automatic analysis of each of these configurations of MS1 type and data acquisition type .

- All combinations of MS1 (IT vs. OT) and MS2 acquisition type (DDA vs. DIA) appear to offer unique strengths for top-down proteomics.

- Detection of serum albumin, a high complexity (39 proteoform IDs by IT MS1) and high mass protein (66 kDa), was not improved using the DIA methods.

- DIA window size changes detectable range of proteins. DIA method showed promise in identifying 2 low-level proteins using OT MS1 where signals observed at near-noise level could not be recognized as charge states. DDA methods are not readily able to achieve this.

- DDA method using OT MS1 detected small protein IGF II (7.5 kDa), which was the lowest concentration regime (5 fmol) detected in any configuration of MS1 type and data acquisition type. OT MS1-DDA appears to be highly sensitive to small proteins when charge state Is discernible.

- For OT MS1- and IT MS1-DIA, using a wide isolation window to capture more charge states resulted in significant quality improvements for MS2 spectra and produced some new protein group IDs. This effect is dependent on protein charge state distribution and the extent to which contaminating precursors are present. When limiting precursor isolation to a single charge state (e.g., conventional DDA) sufficient ion current may not be available to produce high quality MS2 spectra. This was observed as maxing out ion injection times (200 ms).

References

1. **Interlaboratory studies and initiatives developing standards for proteomics.** Ivanov AR, Colangelo CM, Dufresne CP, Friedman DB, Lilley KS, Mechtler K, Phinney BS, Rose KL, Rudnick PA, Searle BC, Shaffer SA, Weintraub ST. Proteomics. 2013 Mar;13(6):904-9.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Canada +1 800 530 8447
China 800 810 5118 (free call domestic)
400 650 5118
PN64437-EN 0616S

Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591

Japan +81 45 453 9100
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00

Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand