⌘ *Author's Choice*

Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation*[S]

Tamar Geiger‡§, Juergen Cox‡, and Matthias Mann¶

The orbitrap mass analyzer combines high sensitivity, high resolution, and high mass accuracy in a compact format. In proteomics applications, it is used in a hybrid configuration with a linear ion trap (LTQ-Orbitrap) where the linear trap quadrupole (LTQ) accumulates, isolates, and fragments peptide ions. Alternatively, isolated ions can be fragmented by higher energy collisional dissociation. A recently introduced stand-alone orbitrap analyzer (Exactive) also features a higher energy collisional dissociation cell but cannot isolate ions. Here we report that this instrument can efficiently characterize protein mixtures by alternating MS and "all-ion fragmentation" (AIF) MS/MS scans in a manner similar to that previously described for quadrupole time-of-flight instruments. We applied the peak recognition algorithms of the MaxQuant software at both the precursor and product ion levels. Assignment of fragment ions to co-eluting precursor ions was facilitated by high resolution (100,000 at *m/z* 200) and high mass accuracy. For efficient fragmentation of different mass precursors, we implemented a stepped collision energy procedure with cumulative MS readout. AIF on the Exactive identified 45 of 48 proteins in an equimolar protein standard mixture and all of them when using a small database. The technique also identified proteins with more than 100-fold abundance differences in a high dynamic range standard. When applied to protein identification in gel slices, AIF unambiguously characterized an immunoprecipitated protein that was barely visible by Coomassie staining and quantified it relative to contaminating proteins. AIF on a benchtop orbitrap instrument is therefore an attractive technology for a wide range of proteomics analyses. ***Molecular & Cellular Proteomics 9:2252–2261, 2010.***

Mass spectrometry (MS)-based proteomics is commonly performed in a "shotgun" format where proteins are digested to peptides, which are separated and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (1, 2). Many peptides typically co-elute from the column and are selected for fragmentation on the basis of their abundance ("data dependent acquisition"). The precursor mass, which can be determined with high mass accuracy in most current instruments, together with a list of fragment ions, which are often determined at lower mass accuracy, are together used to identify the peptide in a sequence database. This scheme is the basis of most of current proteomics research from the identification of single protein bands to the comprehensive characterization of entire proteomes. To minimize stochastic effects from the selection of peptides for fragmentation and to maximize coverage in complex mixtures, very high sequencing speed is desirable. Although this is achievable, it requires complex instrumentation, and there is still no guarantee that all peptides in a mixture are fragmented and identified. Illustrating this challenge, when the Association of Biomolecular Resource Facilities (ABRF)[1] and the Human Proteome Organisation (HUPO) conducted studies of protein identification success in different laboratories, results were varying (4, 5).[2] Despite using state of the art proteomics workflows, often with extensive fractionation, only a few laboratories correctly identified all of the proteins in an equimolar 49-protein mixture (ABRF) or a 20-protein mixture (HUPO).

As an alternative to data-dependent shotgun proteomics, the mass spectrometer can be operated to fragment the entire mass range of co-eluting analytes. This approach has its roots in precursor ion scanning techniques in which all precursors were fragmented simultaneously either in the source region or in the collision cell, and the appearance of specific "reporter ions" for a modification of interest was recorded (6–8). Several groups reported the identification of peptides from MS scans in conjunction with MS/MS scans without precursor ion selection (9–12). Yates and co-workers (13) pursued an intermediate strategy by cycling through the mass range in 10 *m/z* fragmentation windows. The major challenge of data-independent acquisition is that the direct relationship between precursor and fragments is lost. In most of the above studies, this problem was alleviated by making use of the fact that precursors and fragments have to "co-elute."

In recent years, data-independent proteomics has mainly been pursued on the quadrupole TOF platform where it has

From the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

⌘ *Author's Choice*—Final version full access.

Received, June 3, 2010

Published, MCP Papers in Press, July 7, 2010, DOI 10.1074/mcp.M110.001537

[1] The abbreviations used are: ABRF, Association of Biomolecular Resource Facilities; AIF, all-ion fragmentation; LTQ, linear trap quadrupole; SILAC, stable isotope labeling by amino acids in cell culture; HCD, higher energy collisional dissociation; HUPO, Human Proteome Organisation; UPS, Universal Proteomics Standard; Bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

[2] P. Andrews, D. Arnott, M. Gawinowicz, J. Kowalak, W. Lane, K. Lilley, L. Martin, and S. Stein, unpublished data.

been termed MS^E in analogy to MS^2, MS^3, and MS^n techniques used for fragmenting one peptide at a time. Geromanos and co-workers (14–16) applied MS^E to absolute quantification of proteins in mixtures. Another study showed excellent protein coverage of yeast enolase with data-independent peptide fragmentation where enolase peptide intensities varied over 2 orders of magnitude (17). In a recent comparison of data-dependent and -independent peptide fragmentation, the authors concluded that fragmentation information was highly comparable (18, 19).

Recently, the orbitrap mass analyzer (20–23) has been introduced in a benchtop format without the linear ion trap that normally performs ion accumulation, fragmentation, and analysis of the fragments. This instrument, termed Exactive, was developed for small molecule applications such as metabolite analysis. It can be obtained with a higher energy collisional dissociation (HCD) cell (24), enabling efficient fragmentation but no precursor ion selection. This option is called "all-ion fragmentation" (AIF) by the manufacturer, and this is the term that we use below. We reasoned that the high resolution (100,000 compared with 10,000 in quadrupole TOF) and mass accuracy of this device in both the MS and MS/MS modes might facilitate the analysis of the complex fragmentation spectra generated by dissociating several precursors simultaneously. The simplicity and compactness of this instrumentation platform would then make it interesting for diverse proteomics applications.

EXPERIMENTAL PROCEDURES

BSA Sample Preparation—Bovine serum albumin fraction V (Sigma) was solubilized in 8 M urea solution (6 M urea + 2 M thiourea in 10 mM HEPES, pH 8) to a concentration of 350 μg/ml. The protein was reduced with 1 mM DTT, alkylated with 55 mM iodoacetamide, and digested with Lys-C (ratio of enzyme to BSA, 1:70; Wako Chemicals) for 3 h at room temperature to initiate the digestion in denaturing conditions followed by 1:4 dilution of the urea and overnight digestion with trypsin (ratio of enzyme to BSA, 1:70; Promega). The peptide mixture was then acidified with TFA and desalted on C_{18}-Empore disc StageTips (25) (5 pmol per StageTip). For each LC-MS run, 50 fmol of BSA digest were used, and we analyzed the results of triplicate runs.

Universal Proteomics Standard (UPS) Sample Preparation—Universal Proteomics Standard (UPS1; Sigma) and Proteomic Dynamic Range Standard (UPS2; Sigma) were solubilized in 8 M urea solution (6 M urea + 2 M thiourea in 10 mM HEPES, pH 8). Proteins were reduced with 1 mM DTT and alkylated with 50 mM iodoacetamide followed by 3-h digestion with Lys-C (ratio of enzyme to protein, 1:50). The protein mixture was diluted 1:4 in ammonium bicarbonate and digested overnight with trypsin (ratio of enzyme to protein, 1:50). The peptide mixture was acidified and loaded on C_{18}-Empore disc StageTips. The digest from one UPS kit was split to 12 StageTips; each of them was then used for three LC-MS runs. This resulted in the analysis of 166 ng of UPS1 (140 fmol of each protein) or 290 ng of UPS2 (ranging from 14 amol to 1400 fmol of each protein) in each of the three runs.

Immunoprecipitation and In-gel Digest—HeLa cells were lysed with a buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 5% glycerol, 1% Nonidet P-40, and protease inhibitors. 25 μl of Protein G Dynabeads (Invitrogen) were washed with PBS and incubated for 1 h with 5 μg of anti-vinculin antibody (Sigma catalog number V9131) at room temperature. After PBS washes, the beads were incubated with 1 mg of HeLa lysates for 2 h at 4 °C. Beads were then washed with lysis buffer, and proteins were eluted from the beads with lithium dodecyl sulfate sample buffer. The protein sample was split, and 20% was used for Western blot, and 80% was used for further MS analysis. In both cases, proteins were separated on a NuPAGE Bis-Tris gel (Invitrogen). For Western blot analysis, proteins were transferred to a PVDF membrane, and the membrane was incubated with the anti-vinculin antibody. For MS analysis, the gel was stained with Novex Colloidal Blue staining kit (Invitrogen). A very faint band at the molecular weight of vinculin was detected and excised. The gel band was subjected to in-gel digestion with trypsin (26). The resulting peptides were extracted from the gel with 30% acetonitrile in 3% TFA. After using a SpeedVac, peptides were concentrated and desalted on C_{18}-Empore disc StageTips.

LC-MS Analysis—Peptides were analyzed by on-line nanoflow LC-MS on an EASY-nLC system (Proxeon Biosystems) connected to an Exactive instrument with HCD option (Thermo Fisher Scientific) via a Proxeon nanoelectrospray ion source. Peptides were separated on an in-house packed 15-cm column (75-μm inner diameter with 3-μm C_{18} beads; Reprosil-AQ Pur, Dr. Maisch). BSA peptides were separated using a 40-min linear gradient of 2–30% buffer B (80% acetonitrile in 0.5% acetic acid). UPS peptides were separated using a 100-min linear gradient of 5–28% buffer B, and vinculin peptides were separated using a 190-min linear gradient of 2–30% buffer B.

The Exactive mass spectrometer was operated in positive ion mode with alternating MS scans of the precursor ions and AIF scans in which the peptides were fragmented by HCD. Both scan types were performed with 100,000 resolution (at *m/z* 200) with each scan taking 1 s, and the maximal fill time was set to 1 s as well. The m/z range for MS scans was 300–1600, and the m/z range for AIF scans was 100–1600. The target value for the MS scans was 10^6 ions, and the target value for the AIF scans was 3×10^6 ions. We used a higher target value for the AIF to compensate for loss of charges and overfragmentation of ions. HCD collision energy was 30 eV. For stepped collision energy, a new feature was implemented where during filling of the HCD collision cells the energy is stepped between values that are set at specified percent values around the chosen middle energy. We used a middle energy of 30 eV with 20% steps above and below this value.

MaxQuant Data Analysis—All data analysis was performed in the MaxQuant environment version 1.1.0.37 (27–29). Three-dimensional peaks were detected as described before (27) except that this procedure was applied separately for the MS and the AIF scans. The total number of fragment ions per 100 *m/z* was limited to 25 or 15 for low and medium complexity protein mixtures, respectively. Precursor ion masses were associated with possible fragment ion candidates on the basis of a cosine correlation value of at least 0.7.

Database search was performed in MaxQuant against the human International Protein Index version 3.68 (87,083 entries) to which contaminants and reverse sequences were added. In some searches, a small database consisting only of BSA and potential contaminants or of the proteins of the UPS mixture and contaminating proteins was used as indicated in the text. The search was performed with an initial precursor mass tolerance of 7 ppm and fragment mass tolerance of 15 ppm. We included carbamidomethylcysteine as a fixed modification and oxidized methionine and *N*-acetylation as variable modifications. The minimal peptide length was six amino acids, and we allowed two miscleavages. For peptide and protein identifications, the 1% false discovery rate was determined by accumulating 1% of reverse database hits. To avoid apparent misidentifications that are only due to protein name discrepancies (4), we manually examined gene names and UniProt ids and compared them to the list of proteins given by the manufacturer.

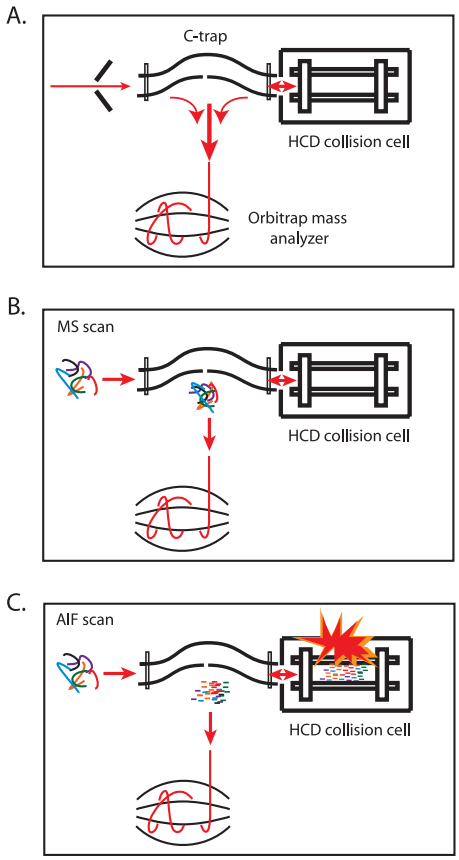


FIG. 1. **Schematic depiction of all-ion fragmentation on Exactive instrument.** *A*, the instrument consists of ion optical elements, the C-trap, the orbitrap mass analyzer, and the HCD cell. *B*, in MS scan mode for precursors, electrosprayed ions are accumulated in the C-trap, typically to 1 million ions, and are then injected in defined packages into the orbitrap analyzer. *C*, in the AIF mode, all electrosprayed ions are fragmented by collisions with nitrogen gas in the HCD cell (indicated by the *red explosion*). Fragments are then moved back to the C-trap from where they are injected into the orbitrap analyzer.

RESULTS

All-ion Fragmentation on Standalone Orbitrap Instrument—In proteomics, orbitrap mass analyzers have been used so far only as a part of hybrid instruments in combination with linear ion traps (LTQ-Orbitrap and LTQ-Orbitrap Velos). This combination enables high resolution and high mass accuracy analysis of the precursor peptides in the orbitrap and isolation of peptides in the LTQ. In these hybrid instruments, fragmentation can be done in the LTQ with CID or in the HCD collision cell. The Exactive benchtop instrument consists of a single orbitrap mass analyzer and no LTQ; therefore, peptides cannot be isolated for fragmentation (Fig. 1*A*). We wished to examine the applicability of this instrument to proteomics by basing peptide identification on HCD fragmentation without precursor peptide selection, which is termed AIF. Peptides are separated on a reverse-phase column on a nano-LC system and directly electrosprayed into the mass spectrometer. The peptides pass at low kinetic energy through a number of ion guides into the C-trap where their energy is dampened by collision with background gas and where they accumulate. In MS scans, the peptides are injected from the C-trap into the orbitrap analyzer for precursor mass measurement (Fig. 1*B*). In AIF scans, peptides are injected into the HCD cell for fragmentation, and then fragment ions are moved back into the C-trap and into the orbitrap for analysis (Fig. 1*C*).

We acquired the data with alternating MS scans and AIF scans, both performed with very high resolution (100,000 at *m/z* 200). Using this resolution, there is a scan every second, and each cycle is composed of two scan events. This guarantees that each eluting peptide peak will be detected and fragmented several times. A unique feature of the Exactive, as compared with the LTQ-Orbitrap, is its ability to fill the C-trap in parallel to transient acquisition in the orbitrap analyzer. Therefore, we set the maximal injection time to 1 s, to gain maximal sensitivity, which still does not affect cycle times. The total number of ions that can be analyzed is limited by the capacity of the C-trap as is the case for the LTQ-Orbitrap. In the MS scans, 10^6 ions accumulate in the C-trap, typically within some tens of milliseconds, and are then analyzed in the orbitrap. In the AIF scans, 3×10^6 ions are fragmented by collisions with nitrogen background gas upon entry into the HCD cell. Milliseconds before analysis the fragments are moved to the C-trap and then into the orbitrap analyzer.

Visualization and Interpretation of AIF Results in MaxQuant—Because of the co-fragmentation of multiple peptides, AIF poses the challenges of efficient peptide fragmentation, interpretation of spectra, and peptide identification. We used the MaxQuant environment (27) to address these computational and data interpretation tasks. First, MaxQuant creates two separate contour plots of all the MS and of the AIF scans consisting of *m/z* values as the *x* axis, retention time as the *y* axis, and a color code for the intensity of the signal. Examination of the MS scans of a simple peptide mixture created by BSA digestion shows the distribution of the peptides as vertical strips along the retention time (Fig. 2*A*). Zooming in on a single peptide shows its isotope cluster and illustrates the high resolution (Fig. 2*B*). Three-dimensional visualization of this isotope cluster further renders the intensities of the peaks as a function of chromatographic elution time.

In contrast to data-dependent MS/MS, in AIF, each eluting peptide is fragmented multiple times, creating an elution profile of each of its fragments, which have to match precisely with the elution profiles of the precursor. Thus, AIF peaks can be visualized in the same way as the precursor peaks. The two-dimensional contour map shows stripes of "co-eluting" fragments belonging to a single or multiple co-eluting precursor peptides (Fig. 2*C*). Zooming in on one of these stripes shows the co-eluting fragments and their elution profile (Fig. 2*D*).

In MaxQuant, the three-dimensional shape of each peak is determined independently for the MS and for the AIF scans

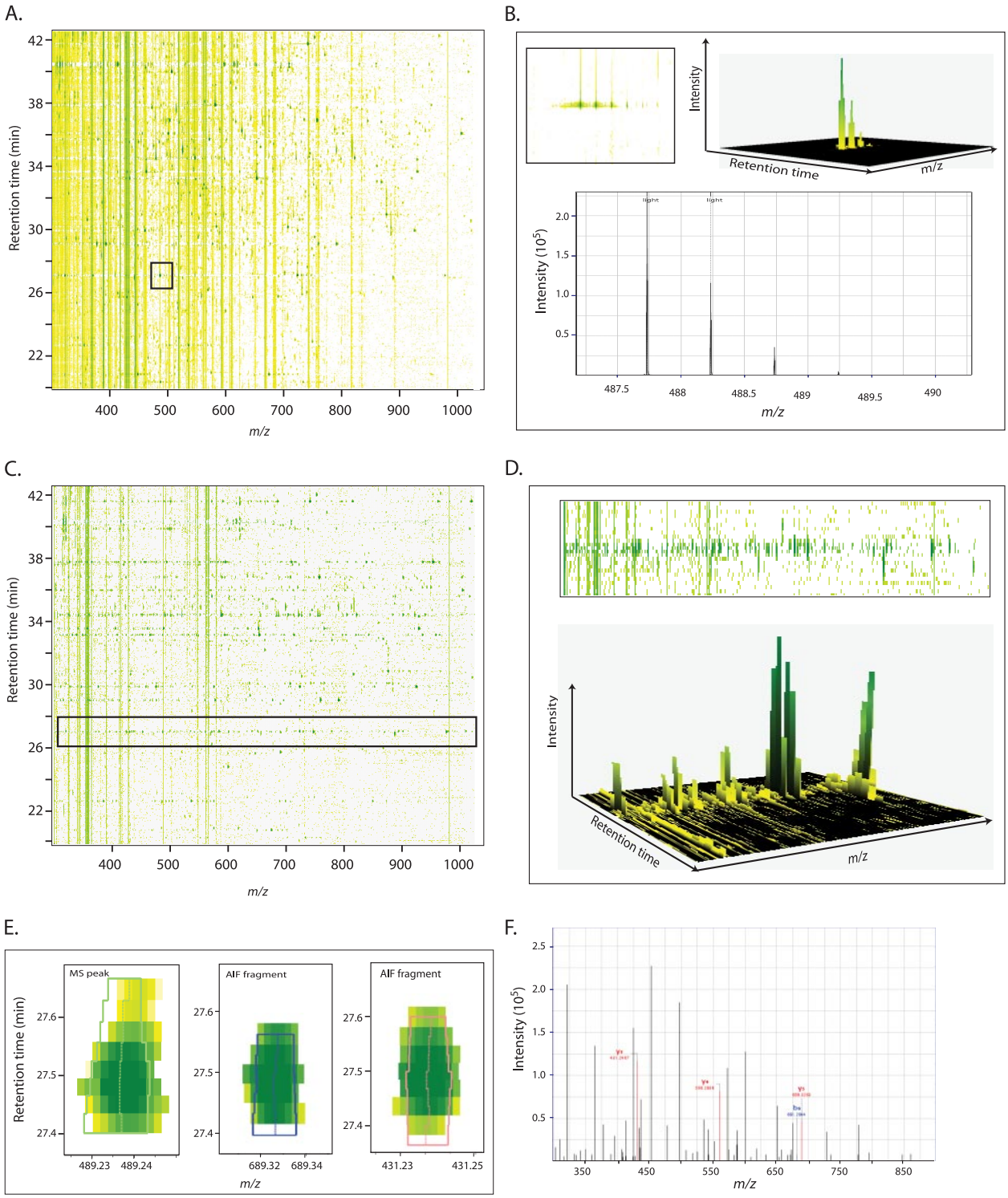


FIG. 2. **Visualization and peak determination in AIF.** *A*, contour plot of a BSA peptide mixture LC MS/MS run in which the MS scans are shown. The elution profile of one BSA peptide is indicated by a *rectangle*. *Vertical stripes* are due to chemical noise from laboratory air. *B*, visualization of the BSA peptide indicated in *A* in three different ways: zoom of the contour plot in *A*, as a three-dimensional peak, and integration of the MS signal over the peak, shown as a part of a mass spectrum. *C*, contour plot of the AIF scans indicating the fragmentation pattern of the BSA peptides. Fragments of peptides co-eluting with the one indicated in *A* and *B* are indicated by a *rectangle*. *D*, zoom of the fragmentation pattern of the BSA peptide in *C* in two and in three dimensions. *E*, illustration of peak determination in MaxQuant. The calculated peak outline and centroids for the fourth isotope of the BSA peptide indicated in *A* and for two of its fragments are shown. *F*, AIF MS/MS spectrum of the BSA peptide. Identified fragment ions are marked in *blue* (b-ions) and *red* (y-ions).

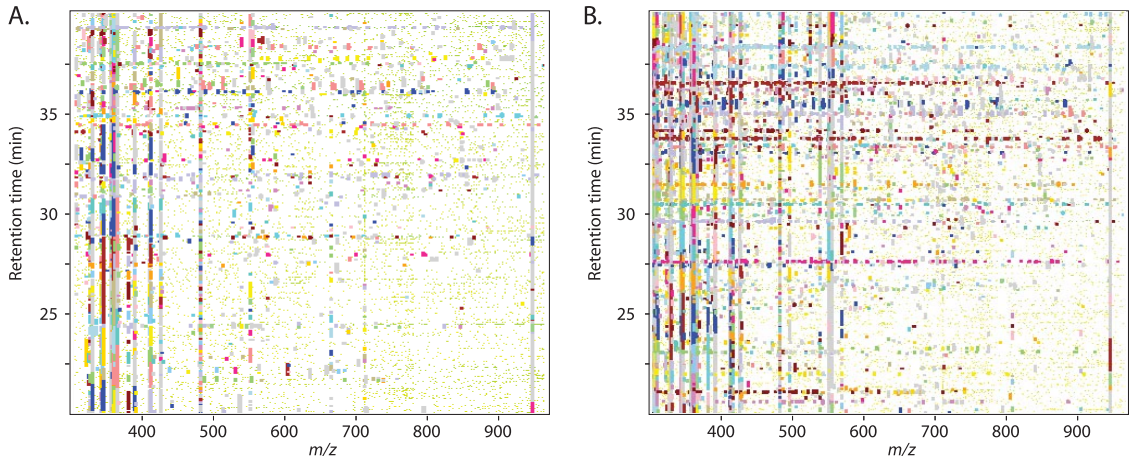


FIG. 3. **AIF contour plot with and without stepped collision energy.** *A*, same as Fig. 2*C* except that all fragment peaks are color-coded in accordance with the precursor ions that they were assigned to by MaxQuant. *B*, same as *A* except for implementation of stepped collision energies. Each spectrum combines fragments generated at low, medium, and high collision energies. Note that many more prominent fragmentation patterns appear (*horizontal stripes*).

(Fig. 2*E*). Then the peaks forming isotope clusters are deisotoped as done normally in MaxQuant analysis. Residues of unfragmented peptides in the AIF plane are found by matching their mass and retention time profile to precursors, and they are discarded in further analysis. AIF peaks whose elution times overlap with a given MS isotope pattern are grouped together. For the fragments, isotope patterns as well as single peaks not clustered into isotope patterns are considered because low intensity fragments have a high likelihood of being only detectable in the monoisotopic form. For each MS and AIF feature that has an overlap in retention time, the correlation coefficient of the elution profile is determined. Based on this correlation, AIF peaks are then assigned to their potential precursors. Three-dimensional fragment peaks assigned to a precursor are collapsed to a conventional MS/MS spectrum by summing intensities over the retention time profile. Subsequently, peptide identification is performed as in data-dependent peptide fragmentation (Fig. 2*F*) including a search against a database of an *in silico* digest of protein sequences.

Stepped Collision Energies Improve Fragmentation Behavior—In standard MS analysis, different peptides are fragmented with different energies according to their masses and charge states. This is obviously not the case in AIF where all co-eluting peptides are fragmented simultaneously and with the same energy. For this reason, in Q-TOFs, the collision energy has been ramped when acquiring fragmentation data in the full mass range (14, 17).

As expected, when we analyzed a mixture of BSA peptides, we found that the single collision energy inefficiently fragmented many of the peptides. In the contour plot, we color-coded all the peaks that belonged together and found that in many cases only a few fragments could be associated with their precursors (Fig. 3*A*). To more efficiently fragment peptides with different *m/z* and charge, stepped collision energy was therefore implemented into the mass spectrometer. This feature enables a shift between three collision energies during filling of the HCD cell. We opted against acquisition of separate spectra at each collision energy. Although potentially informative, this would have lengthened the cycle time or decreased the resolution. Instead, we fragmented peptides for a third of the fill time with the low, medium, and high collision energies, respectively, and accumulated all fragments in the HCD cell for combined orbitrap analysis.

Using this feature, we detected multiple extra peaks in the AIF spectra, creating easily visible horizontal stripes of fragments of each peptide (Fig. 3*B*). Without stepped energy we identified BSA, which had been added to the human International Protein Index, with 22 peptides (supplemental Table 1). Analysis of the data showed that these extra peaks translated into an increase in the number of identified peptides (to 26 different peptides).

We suspected that information about even more peptides might be present in the data but that these peptides might not be statistically significant. To test this hypothesis, we created a small database consisting of BSA and 254 sequences of potential contaminants, mainly keratins. When we analyzed the BSA runs, we now found 27 and 33 BSA peptides without and with stepped collision energy, respectively. Together, these results clearly showed the advantage of the improved fragmentation by the stepped energy; therefore, in all subsequent experiments, we used these fragmentation parameters.

Analysis of Equimolar Mixture of 48 Proteins with AIF—Large efforts have been made in the past decade to evaluate and standardize proteomics analyses by organizations such as ABRF and HUPO. Both perform comparative studies and create protein standards for evaluation of qualitative and quantitative proteomics methods. One of the products of the ABRF efforts is a commercially available protein mixture, the

TABLE I
Proteins identified in 48-protein UPS

Triplicate runs were analyzed against a database of the UPS proteins and contaminants and against the complete human database and contaminants. PEP, posterior error probability; Percent seq., percent protein sequence coverage; Cyt, cytochrome; PPIA, peptidyl-prolyl cis-trans isomerase A; BID, BH3-interacting domain death agonist; HARS, histidyl-tRNA synthetase; CK-M, creatine kinase M-type.

Protein name	UniProt accession no.	UPS database				Human database			
		Intensity ($\times 10^7$)	No. peptides	PEP	Percent seq.	Intensity ($\times 10^7$)	No. peptides	PEP	Percent seq.
Annexin A5	P08758	35.2	17	0e+00	57.4	24.0	8	7e−70	29.4
Antithrombin-III	P01008	40.9	20	0e+00	46.1	26.2	10	0e+00	30.8
BID	P55957	13.1	8	0e+00	52.8	12.5	7	0e+00	41.5
CAH- 2	P00918	22.6	11	0e+00	52.9	21.7	9	4e−73	49.6
CAH-1	P00915	25.9	12	0e+00	63.1	19.9	8	0e+00	46.4
Catalase	P04040	53.0	29	0e+00	62.9	44.0	19	0e+00	44.6
CK-M	P06732	35.6	21	0e+00	67.7	25.9	13	0e+00	42.0
Complement C5	P01031	2.9	2	2e−42	28.4	1.9	1	1e−21	0.8
CRP	P02741	6.5	5	1e−46	22.6	6.4	5	1e−18	21.0
Cyt-b_5	P00167	18.4	7	0e+00	55.7	14.1	5	8e−70	49.3
Cyt-*c*	P99999	10.1	8	8e−32	55.8	5.0	2	1e−19	23.8
EGF	P01133	0.4	1	2e−02	24.1				
FABP3	P05413	20.4	10	6e−94	76.5	12.8	5	2e−57	45.9
Gelsolin	P06396	32.0	23	0e+00	40.7	20.6	13	0e+00	26.6
GST-A1	P08263	17.2	11	5e−18	44.8	5.6	2	8e−05	8.9
GST-P	P09211	15.9	11	0e+00	57.4	10.6	6	0e+00	35.7
HARS	P12081	47.9	26	0e+00	57.9	36.4	18	0e+00	48.1
Hemoglobin-α	P69905	5.4	5	6e−45	43.3	3.1	3	2e−27	30.3
Hemoglobin-β	P68871	9.3	9	6e−86	69.9	7.9	7	1e−49	55.1
H-Ras	P01112	14.1	10	0e+00	58.7	14.0	9	4e−69	55.0
IFN-γ	P01579	6.2	4	0e+00	44.1	6.2	4	0e+00	38.0
IGF-2	P01344	5.3	3	9e−83	73.1	4.5	2	4e−11	10.6
IL-8	P10145	0.5	1	7e−03	22.2				
Lactalbumin	P00709	10.9	7	2e−16	82.1	7.4	3	2e−07	18.3
Lactotransferrin	P02788	59.1	41	0e+00	67.3	47.2	29	0e+00	44.7
Leptin	P41159	19.8	5	0e+00	49	19.7	5	0e+00	43.1
Lysozyme C	P61626	10.3	9	0e+00	78.5	8.2	5	3e−78	44.6
Microglobulin-β2	P61769	20.8	3	0e+00	42.4	20.8	3	2e−105	34.4
Myoglobin	P02144	15.0	7	0e+00	54.2	12.7	5	0e+00	50.0
NEDD8	Q15843	7.0	4	8e−52	45.7	6.8	3	5e−24	34.6
NQO1	P15559	12.3	8	8e−39	24.9	6.7	4	2e−19	17.5
NQO2	P16083	27.8	10	0e+00	55.7	20.6	7	0e+00	44.2
PDGF-B	P01127	2.1	3	5e−10	34.9				
Peroxiredoxin 1	Q0683	26.7	12	9e−80	73.2	18.2	7	2e−40	41.2
PPIA	P62937	15.2	8	0e+00	60.5	11.4	5	4e−36	40.6
RETbp-4	P02753	5.7	6	1e−37	30.2	3.3	3	9e−14	15.4
Serotransferrin	P02787	46.9	35	0e+00	64.2	36.7	26	0e+00	46.8
Serum albumin	P02768	84.7	37	0e+00	71.9	64.0	29	0e+00	63.1
SOD1	P00441	11.9	9	0e+00	75.2	8.2	5	0e+00	59.7
SUMO1	P63165	29.2	15	0e+00	46.4	3.4	3	1e−07	19.9
Synuclein-γ	O76070	26.2	12	0e+00	89.8	18.0	7	0e+00	63.0
Tau	P10636	45.4	18	0e+00	60.5	40.7	14	0e+00	57.1
Thioredoxin	P10599	8.9	5	2e−32	46.4	7.7	3	3e−17	29.5
TNF-α	P01375	18.9	7	0e+00	63.7	16.2	6	0e+00	39.5
UB2E1	P51965	2.5	3	2e−12	18.6	1.9	2	2e−09	8.3
UBC9	P63279	10.2	7	2e−38	50	4.6	3	4e−19	16.8
UBE2C	O00762	14.0	8	0e+00	67	15.8	8	4e−70	67.6

Universal Protein Standard (UPS1), which contains 48 different proteins in equimolar concentrations. The mixture includes proteins ranging in size from polypeptides as small as 6 kDa (EGF) to large proteins of 83 kDa (gelsolin), resulting in large differences in the number of potential tryptic peptides that can be identified from each protein.

This standard mixture allowed us to examine whether our new algorithm enables protein identification in more complex samples. We digested the proteins and separated the peptides using a 100-min gradient on a reverse-phase column. Each chromatographic run contained peptides obtained by digestion of 160 ng of total protein (140 fmol/protein). In the mass spectrometer, peptides were fragmented with step collision energy as described above. Remarkably, in triplicate runs, we identified 348 different peptides in the human database, corresponding to 45 UPS proteins (Table I). We identified only two peptides that did not correspond to any of the UPS proteins or contaminants. This is as expected from

the 1% false discovery rate that we specified in MaxQuant. The median number of peptides per protein was 5, and the median sequence coverage was 41%. The three proteins that we could not identify were EGF, PDGF, and IL-8, which are all low molecular weight proteins. EGF has only one tryptic peptide that can be efficiently characterized by MS, making the protein difficult to identify in a large database. IL-8 is very basic, containing 14 lysines and arginines (18% of the amino acids). As a result, efficient digestion with trypsin creates only three peptides with more than the minimal six amino acids (without incomplete digestion) that we used as a requirement for identification.

To check whether more peptides were present in the analysis, we searched the data against a small database containing only the UPS mixture and common contaminants. In this database, we were able to identify all 48 proteins (Table I). Sequence coverage of these proteins improved to a median of 56% and ranged between 19 and 90%.

Dynamic Range of AIF on Orbitrap Analyzer—The Protein Dynamic Range Standard (UPS2) is another commercial protein mixture. It contains the same proteins as UPS1, but these 48 proteins are present in different concentrations, ranging over 5 orders of magnitude. The proteins in the mixture are divided into six abundance classes with a 10-fold difference in concentration between each class, ranging from 500 amol to 50 pmol. UPS2 analysis tests both the sensitivity of a work flow and its dynamic range. We analyzed the UPS2 mixture in triplicates, and in each run, we separated peptides from 290 ng of total protein (14 amol to 1.4 pmol/protein).

In the human database, we identified 16 proteins with 123 peptides in total, spanning three abundance classes (supplemental Table 2). EGF was in the middle abundance class but was nevertheless not identified-presumably because of the reasons discussed above. Likewise, four proteins in the lowest abundance class of the three were not identified. When searching only against the UPS and contaminant database, despite the challenging dynamic range, we were able to identify 21 different proteins (excluding known contaminants). We identified all 16 proteins from the most abundant groups and five proteins from the lower abundance group where only 14 fmol of each protein were injected into the mass spectrometer. Summed peptide intensities correlate well with the absolute protein amount in proteomics measurements (30, 31), and data-independent fragmentation has been reported to have specific advantages for absolute quantification (16). Indeed, when we plotted summed peptide intensities for each protein, the different abundance classes were clearly distinguished, and the overall -fold change between the classes was correctly represented (Fig. 4*A*). Taking other parameters such as the theoretical number of tryptic peptides into consideration will likely further increase the accuracy of the estimate of absolute protein amount.

Because the same proteins are present in the equimolar UPS1 mixture and the dynamic range UPS2 mixture, we

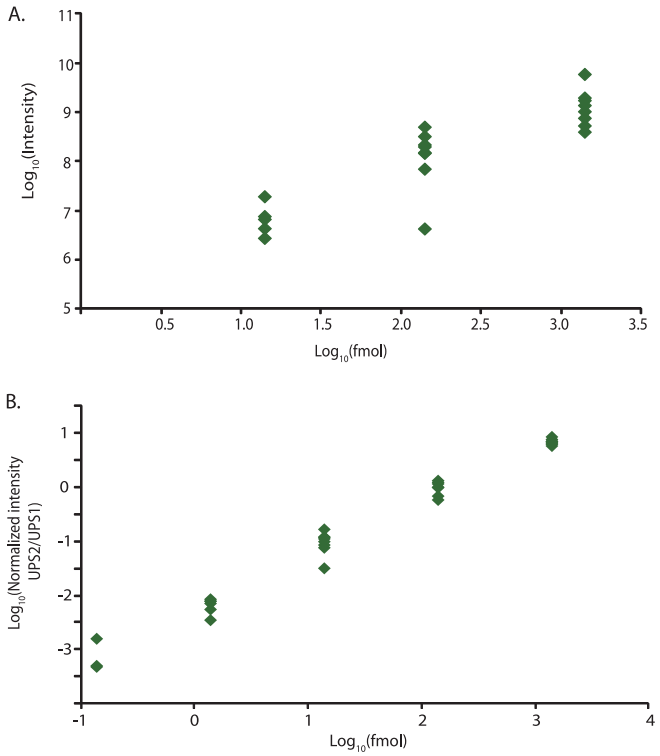
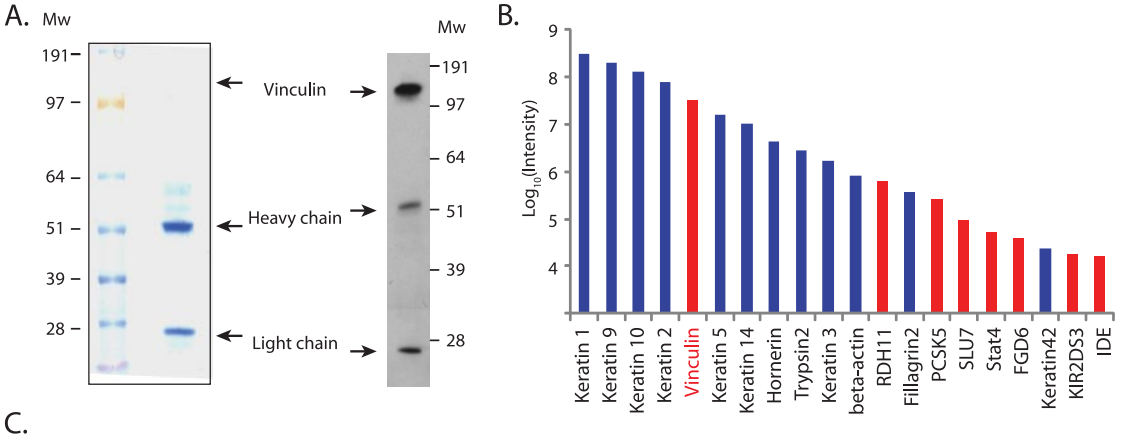


FIG. 4. **Quantitation of dynamic range universal protein standard.** *A*, absolute protein amount as estimated by added peptide signal. Values of the *x axis* are the reported amounts in the UPS2 mixture in fmol. *B*, relative amount of the same protein in the dynamic range protein standard compared with the equimolar protein standard. Values of the *x axis* are the reported amounts in the UPS2 mixture in fmol.

reasoned that combined analysis of the two data sets would represent a good assessment of the ability of relative protein quantification. In this analysis, identities of peptides in the UPS1 run, where they are more abundant, can be transferred to the UPS2 run, potentially increasing the dynamic range of quantification. Such matching is based on the exact mass, obtained in the orbitrap, and the retention time of the peptide ("match between runs" feature in MaxQuant). This matching enabled identification of 32 proteins in five abundance classes (supplemental Table 3), demonstrating that the Exactive instrument is able to detect peptides ranging over 4 orders of magnitude. We next examined the quantification accuracy of the proteins by using the label-free algorithm of MaxQuant (32) to calculate the normalized intensities of UPS1 proteins and UPS2 proteins and then calculated the ratio between UPS2 and UPS1 (Fig. 4*B*). We found a linear correlation between the protein mixtures, covering the five abundance groups. As expected, the ratio of the proteins in the two standards changed in 10-fold increments, ranging from 10-fold higher to 1000-fold lower amounts in UPS2 compared with UPS1. The ratios were very accurate in the more abundant protein classes and less accurate in the lower abundance classes as anticipated from the number of peptides



C.

```
1    MPVFHTRTIESILEPVAQQISHLVIMHEEGEVDGKAIPDLTAPVAAVQAAVSNLVRVGKETVQTTEDQILKRDMPPAFIKVENACTKLVQAAQMLQSDPY
101  SVPARDYLIDGSRGILSGTSDLLLTFDEAEVRKIIRVCKGILEYLTVAEVVETMEDLVTYTKNLGPGMTKMAKMIDERQQELTHQEHRVMLVNSMNTVKE
201  LLPVLISAMKIFVTTKNSKNQGIEEALKNRNFTVEKMSAEINEIIRVLQLTSWDEDAWASKDTEAMKRALASIDSKLNQAKGWLRDPSASPGDAGEQAIR
301  QILDEAGKVGELCAGKERREILGTCKMLGQMTDQVADLRARGQGSSPVAMQKAQQVSQGLDVLTAKVENAARKLEAMTNSKQSIAKKIDAAQNWLADPNG
401  GPEGEEQIRGALAEARKIAELCDDPKERDDILRSLGEISALTSKLADLRRQGKGDSPEARALAKQVATALQNLQTKTNRAVANSRPAKAAVHLEGKIEQA
501  QRWIDNPTVDDRGVGQAAIRGLVAEGHRLANVMMGPYRQDLLAKCDRVDQLTAQLADLAARGEGESPQARALASQLQDSLKDLKARMQEAMTQEVSDVFS
601  DTTTPIKLLAVAATAPPDAPNREEVFDERAANFENHSGKLGATAEKAAAVGTANKSTVEGIQASVKTARELTPQVVSAARILLRNPGNQAAYEHFETMKN
701  QWIDNVEKMTGLVDEAIDTKSLLDASEEAIKKDLDKCKVAMANIQPQMLVAGATSIARRANRILLVAKREVENSEDPKFREAVKAASDELSKTISPMVMD
801  AKAVAGNISDPGLQKSFLDSGYRILGAVAKVREAFQPQEPDFPPPPPDLEQLRLTDELAPPKPPLPEGEVPPPRPPPPEEKDEEFPEQKAGEVINQPMMM
901  AARQLHDEARKWSSKPGIPAAEVGIGVVAEADAADAAGFPVPPDMEDDYEPELLLMPSNQPVNQPILAAAQSLHREATKWSSKGNDIIAAAKRMALLMAE
1001 MSRLVRGGSGTKRALIQCAKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTNISDEESEQATEMLVHNAQNLMQ
1101 SVKETVREAEAASIKIRTDAGFTLRWVRKTPWYQ
```

FIG. 5. **Identification of proteins in gel bands.** *A*, Coomassie stain of a one-dimensional gel of immunoprecipitated vinculin. A faint band had been visible by eye but not after scanning. The *right-hand panel* is a Western blot with the anti-vinculin antibody. *B*, proteins identified by AIF on the Exactive are ordered by added peptide signal. Note that contaminants (*blue*) are up to 10-fold more abundant than the target protein (*red*) and that the target protein is 100-fold more abundant than the next possible candidate protein. *C*, sequence coverage of vinculin by the 16 peptides identified by AIF on the Exactive instrument.

that can be used for quantification. Nevertheless, the ratios of almost all proteins reflected the change in the protein amounts between the two standards very well.

Identifying Gel Bands by AIF on Orbitrap Analyzer—We next examined whether our work flow can be applied to a common analytical task in proteomics laboratories, the identification of gel bands. We immunoprecipitated the focal adhesion protein vinculin (33) from HeLa cells, separated the proteins by gel electrophoresis, and analyzed the sample by Western blot and by Coomassie staining. The amount of precipitated vinculin was very low as we could only see a very faint band by eye in the Coomassie gel that did not even appear in the scan, which mostly detected the antibody bands in the gel lane (Fig. 5*A*). However, we could verify that the faint band is indeed vinculin based on the clear band in the Western blot. We excised the gel band, performed an in-gel digest, and analyzed the peptides on the Exactive with AIF. In total, we identified 120 peptides originating from 20 proteins (supplemental Table 4). We ordered these proteins according to their estimated abundance and color-coded them by their occurrence in the human database or the contaminant database to distinguish between background and potential true hits (Fig. 5*B*). The four most intense proteins were different keratins, which are known contaminants. The next protein was our target, vinculin, which we identified with 16 peptides. All other non-contaminants showed much lower intensities by at least 100-fold and could therefore not be the origin of the very faintly stained band. The 16 peptides of vinculin covered 14.4% of the protein sequence and originated from the different regions of the protein (Fig. 5*C*). Therefore, by excluding contaminants and sorting based on the protein intensity, the true protein of interest was easily revealed even in this challenging situation.

DISCUSSION

Here we have developed methods that enable protein identification through fragmentation of all co-eluting ions by HCD in a standalone orbitrap instrument. HCD, with injection of ions with three different energies, ensured good fragmentation of peptides. We made use of the high resolution and mass accuracy of the orbitrap analyzer and combined this with the sophisticated peak recognition features of MaxQuant. High accuracy fragmentation peaks were assigned to precursors on the basis of "co-elution" profiles.

This technology easily characterized a digest of a single model protein and unexpectedly identified nearly all proteins in a 48-protein standard. This compares favorably with results of multilaboratory studies in which many participants reported fewer proteins and misidentification of proteins was common (4).[2] Even more striking was the ability of AIF on the Exactive to accommodate a relatively large dynamic range of analyte, which we initially thought would be a weakness of AIF. Pro-

teins that were less than 1% of the abundance of the major protein were readily identified, and when combining results from an equimolar protein mixture with a high dynamic range protein mixture, up to 4 orders of magnitude of abundance were covered. This ability would be useful in quantitative studies where the protein of interest is more abundant in one cellular state and where its peptides can then be quantified in the other states despite a large expression difference. Although not shown here, AIF on the Exactive should be readily amenable to relative quantification based on stable isotope methods such as SILAC (3). The information encoded in SILAC peptide fragments should further help in assigning fragments to precursors as well as in MS/MS-based quantification.

Much of proteomics work involves the analysis of gel bands from low or medium complexity protein mixtures. In this study, we have demonstrated straightforward identification of an immunoprecipitated protein, a typical gel band identification application. Absolute quantification (in the form of added peptide signals) differentiated the target protein from more abundant background proteins that were also contained in the contaminant database and proteins of much lower abundance that were not the protein of interest. We anticipate that gel-separated protein mixtures of moderate complexity (at least up to several hundred proteins) would also be amenable to the AIF-Exactive analysis, and we will investigate this in the future. Likewise, it would be interesting to compare AIF on the Exactive directly against existing quadrupole TOF instrumentation. Based on the high resolution and the high mass accuracy even for small peaks, we expect data from the Exactive to perform favorably in such a comparison.

In conclusion, the AIF-Exactive technology combines highly confident identification of simple to moderate protein mixtures and excellent quantitative analysis capabilities in a simple, robust, and economical format. Although it does not reach the depth of analysis of a high end hybrid instrument, it appears to perform at least equally well on many mainstream proteomics work flows. Therefore, we expect the technology described here to be a useful addition to the proteomics toolbox.

Acknowledgments—We thank Nadin Neuhauser for technical help, Oliver Lange for implementing the stepped collision energy feature, and Alexander Makarov for helpful discussions.

* This work was supported in part by the European Commission's 7th Framework Program Proteomics Specification in Time and Space (PROSPECTS; Grant HEALTH-F4-2008-021,648).

S This article contains supplemental Tables 1–4.

‡ Both authors made equal contributions to this work.

§ Supported by the Humboldt Foundation.

¶ To whom correspondence may be addressed. Tel.: 49-89-8578-2557; Fax: 49-89-8578-2219; E-mail: mmann@biochem.mpg.de.

REFERENCES

1. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19,** 242–247
2. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
3. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386
4. Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E., Beavis, R., Sechi, S., Nilsson, T., and Bergeron, J. J. (2009) A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Methods* **6,** 423–430
5. Mann, M. (2009) Comparative analysis to guide quality improvements in proteomics. *Nat. Methods* **6,** 717–719
6. Huddleston, M. J., Bean, M. F., and Carr, S. A. (1993) Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* **65,** 877–884
7. Wilm, M., Neubauer, G., and Mann, M. (1996) Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **68,** 527–533
8. Bateman, R. H., Carruthers, R., Hoyes, J. B., Jones, C., Langridge, J. I., Millar, A., and Vissers, J. P. (2002) A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass Spectrom.* **13,** 792–803
9. Kosaka, T., Yoneyama-Takazawa, T., Kubota, K., Matsuoka, T., Sato, I., Sasaki, T., and Tanaka, Y. (2003) Protein identification by peptide mass fingerprinting and peptide sequence tagging with alternating scans of nano-liquid chromatography/infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **38,** 1281–1287
10. Purvine, S., Eppel, J. T., Yi, E. C., and Goodlett, D. R. (2003) Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3,** 847–850
11. Williams, J. D., Flanagan, M., Lopez, L., Fischer, S., and Miller, L. A. (2003) Using accurate mass electrospray ionization-time-of-flight mass spectrometry with in-source collision-induced dissociation to sequence peptide mixtures. *J. Chromatogr. A* **1020,** 11–26
12. Ramos, A. A., Yang, H., Rosen, L. E., and Yao, X. D. (2006) Tandem parallel fragmentation of peptides for mass spectrometry. *Anal. Chem.* **78,** 6391–6397
13. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., and Yates, J. R. (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1,** 39–45
14. Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77,** 2187–2200
15. Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G. Z., Richardson, K., Wall, D., and Geromanos, S. J. (2006) Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. *Mol. Cell. Proteomics* **5,** 589–607
16. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5,** 144–156
17. Chakraborty, A. B., Berger, S. J., and Gebler, J. C. (2007) Use of an integrated MS-multiplexed MS/MS data acquisition strategy for high-coverage peptide mapping studies. *Rapid Commun. Mass Spectrom.* **21,** 730–744
18. Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., Bateman, R. H., and Langridge, J. I. (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9,** 1683–1695
19. Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V., and Geromanos, S. J. (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9,** 1696–1719
20. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **72,** 1156–1162
21. Hardman, M., and Makarov, A. A. (2003) Interfacing the orbitrap mass

analyzer to an electrospray ion source. *Anal. Chem.* **75,** 1699–1705
22. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., and Horning, S. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78,** 2113–2120
23. Scigelova, M., and Makarov, A. (2006) Orbitrap mass analyzer—overview and applications in proteomics. *Proteomics* **6,** Suppl. 2, 16–21
24. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **4,** 709–712
25. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670
26. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1,** 2856–2860
27. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372
28. Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* **4,** 698–705
29. Cox, J., and Mann, M. (2009) Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *J. Am. Soc. Mass Spectrom.* **20,** 1477–1485
30. de Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Fröhlich, F., Walther, T. C., and Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455,** 1251–1254
31. Malmström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R. (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460,** 762–765
32. Luber, C. A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., and Mann, M. (2010) Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **32,** 279–289
33. Geiger, B. (1979) A 130K protein from chicken gizzard: its localization at the termini of microfilament bundles in cultured chicken cells. *Cell* **18,** 193–205