

Published in final edited form as:

Nat Protoc. 2020 February 01; 15(2): 236–265. doi:10.1038/s41596-019-0233-8.

Direct Characterization of Overproduced Proteins by Native Mass Spectrometry

Shay Vimer[#], Gili Ben-Nissan[#], Michal Sharon^{*}

Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

[#] These authors contributed equally to this work.

Abstract

Proteins derived by recombinant technologies must be characterized to ensure quality, consistency and optimum production. These properties are usually assayed following purification procedures that are time-consuming and labor-intensive. Here we describe a native mass spectrometry approach, known as direct-MS, for rapid characterization of intact overexpressed proteins immediately from crude samples. In describing this protocol, we discuss the multiple applications of the method, and outline the necessary steps required for sample preparation, data collection and interpretation of results. We begin with the sample preparation workflows, which are relevant for either recombinant proteins produced within bacteria, those analyzed straight from crude cell lysate, or secreted proteins generated in eukaryotic expression systems that are assessed directly from the growth culture medium. We continue with the mass acquisition steps that enable immediate definition of properties such as expressibility, solubility, assembly state, folding, overall structure, stability, post-translational modifications, and associations with biomolecules. We demonstrate the applicability of the method through the characterization of a computationally designed toxin-anti toxin heterodimer, activity and protein interaction determination of a regulatory protein and detailed glycosylation analysis of a designed intact antibody. Overall, we describe a simple and rapid protocol that is relevant to both prokaryotic and eukaryotic expression systems that can be carried out on multiple mass spectrometers such as Orbitrap and QTOF-based platforms that enable intact protein detection. The entire procedure takes between 30 minutes to several hours, from sample collection to data acquisition, depending on the depth of MS analysis.

A key contribution to our understanding of how cells work arose from the ability to produce active recombinant proteins for structural and functional investigations. Similarly, the production of recombinant proteins revolutionized industry, due to the wide variety of enzymes that are used today in food processing, agriculture, leather production, paper and detergent manufacture¹. Clinical applications of recombinant proteins have also grown tremendously, with the development of biological and biosimilar therapeutics². To date, protein production has become far easier than ever before, due to advances in computational tools that enable the design of proteins with tailored activities, increased stability and yield³. Efforts to overcome challenges in protein expression have also led to improvements in

^{*}Corresponding author: Michal Sharon, Tel: (+972)-8-934-3947, Fax: (+972)-8-934-6010, michal.sharon@weizmann.ac.il.

Data availability statement

The datasets generated during the current study are available from the corresponding author on reasonable request.

vectors, DNA manipulation techniques, growth media, and expression platforms, together facilitating the task of protein overproduction⁴.

The production of recombinant proteins generally encompasses four major steps: gene cloning, protein expression, protein purification and characterization. Here, we will focus on the protein characterization aspect, which is critical to the quality control assessment that ensures proper production of the target protein. Characterizing the generated protein is also essential for selecting the ideal host system, optimizing codon usage and yield, as well as providing input for iterative redesign and optimization. Such analysis is also relevant for ensuring batch-to-batch consistency, and selecting lead candidates for further optimization. Multiple methods are available for protein characterization, such as SDS-PAGE analysis⁵, circular dichroism (CD)⁶, small-angle X-ray scattering (SAXS)⁶, dynamic light scattering (DLS)⁶ and nuclear magnetic resonance (NMR)⁷. Such measurements, however, are usually undertaken with purified proteins, with significant costs in time and labor invested in product purification. Here, we provide a simple and rapid protocol for in-depth analysis of overproduced proteins directly from crude samples with minimal purification, using native mass spectrometry (MS)^{8–10} (Fig. 1).

The main advantage of the direct-MS method that we describe is that it overcomes the need for protein purification. Thus, proteins produced within bacteria are analyzed directly from the crude lysate^{9–11}, while assessment of secreted proteins generated by eukaryotic expression systems is performed straight from the crude culture medium⁸ (Fig. 1). Native MS measurement, which allows the analysis of intact protein assemblies under non-denaturing conditions, then provides in-depth structural characterization of the overexpressed protein(s). Properties such as solubility, molecular weight, folding, assembly state, stability and topological arrangements are immediately revealed. The high resolution afforded by the intact protein mass measurement also facilitates assessment of sequence variations, and binding to relevant biomolecules. The molecular heterogeneity imposed by post-translational modifications^{12,13}, as acetylation¹⁴, phosphorylation¹⁵ and glycosylation forms¹⁶ (see Table 1) can also be identified and assigned. In terms of production efficiency, such analysis enables identification of necessary micronutrient supplementation, and capture of the optimal harvesting time. Furthermore, as we demonstrate in the ANTICIPATED RESULTS section, the workflow is not only limited to quality assessment, but may also be expanded for addressing different biological questions such as the strength of amino acid interactions, protein interactions, and the in-vitro activity of the overproduced protein.

Traditionally, a prerequisite for acquiring native MS data entails prior purification of the protein of interest¹⁷. During this multiple-step process, non-covalent associations of the protein with ligands and cofactors are likely to be lost. The method we describe here makes it possible to overcome the requirement for purification, not only reducing the time lag between production and characterization, but also preserving the apo-protein forms. The direct-MS approach relies on a weakness of mass spectrometry; i.e., limited dynamic range¹⁸, which in this instance is turned into a strength. Typically, the high dynamic range of cellular protein levels is beyond the intrinsic limitations of mass spectrometer sensitivity, resulting in the masking of proteins found at lower abundance, by those found at higher abundance. Thus, taking into consideration the high protein expression levels typically

achieved in overproduction systems, the analysis is intrinsically biased towards the detection of recombinant proteins, overlooking proteins of lower abundance. This inherent property of MS, will lead to the detection of multiple proteoforms of the overproduced protein, though low abundant recombinant species may be under-sampled.

The multiple systems (e.g., bacteria, yeast, insect and human cells)^{8,10} that we have been studying enabled us to establish general guidelines, as described below. The protocol, which can be readily extended to throughput analysis, can be adapted to multiple mass spectrometers that enable intact protein detection, such as Orbitrap or QTOF-based platforms that encompass extended mass range^{19–21}. The workflows we outline are relevant to both proteins expressed intracellularly by bacterial systems, and those secreted by eukaryotic or bacterial host systems. Notably, although we have chosen to emphasize MS sample preparation and data acquisition, screening for optimal expression conditions such as post-induction time, temperature, inducer concentration, type of cultivation medium, expression host system and choice of signal peptide for secreted proteins are important, as such attributes are critical for maximizing the productivity of the expression systems. Prior to detailing the experimental workflow, we begin by addressing basic questions that highlight the key features of the direct-MS methodology.

Experimental design

1 Is the direct-MS method suitable for any recombinant protein?

The fundamental requirement for this approach is that the overexpressed target becomes the dominant protein, such that it outperforms the levels of endogenous background proteins. It is difficult to provide exact numbers for the overproduction fold, or the protein concentration required for direct-MS analysis, as the method is highly dependent on the abundance, molecular weight and ionization efficiency of both generated and background proteins. As a rule of thumb, following induction, if the produced protein appears as a dominant band in coomassie-stained SDS-PAGE gels, or is far greater or smaller, in terms of size, from other proteins, it is likely to be detected by direct-MS (Fig. 2). However, as described below, dilution of the sample might be required prior to acquisition (Fig. 3).

In bacterial expression systems, high expression levels are typically achieved in *E. coli*²⁹; therefore, unless the protein is compartmentalized into insoluble inclusion bodies, its detection would be expected (Fig. 2). Production quantities are usually lower in eukaryotic hosts, though they provide a greater capacity for proper protein folding, assembly and post-translational modifications, in comparison to the bacterial platform^{10,30}. Such eukaryotic expression systems are predominantly designed for secretion of the recombinant protein into the growth medium. This attribute lends itself to direct-MS analysis, as when protein-free growth medium is used, the secreted recombinant protein is expected to become the dominant protein in the culture medium, despite the presence of background endogenous proteins (Fig. 2). When proteins are produced intracellularly, however, this is not the case. Based on our experience, the upper limit for recombinant protein accumulation inside eukaryotic cells is insufficient to outperform the background endogenous proteins, thus preventing the detection of the unpurified recombinant protein. For this type of investigation, other approaches, which rely on prior protein purification, are still needed.

2 What is the minimal amount of protein that can be detected?

Signal intensity in MS analysis is largely determined by the effectiveness of gas-phase ion production from analyte molecules in solution; namely, ionization efficiency³¹. This property depends on numerous physical and chemical properties of the protein that are difficult to predict. Moreover, ionization efficiency is also influenced by the presence of species that compete for ionization; in this case, background endogenous proteins and interfering compounds. Thus, the minimum amount of generated protein that is likely to be detected by the direct-MS method, will be specific for each protein produced in a particular expression system. Nevertheless, we could roughly estimate this value by titrating known protein concentrations into both bacterial lysate and insect cell growth media from non-expressing cells (Fig. 4). The data suggest that in both bacterial and eukaryotic host systems, 1 μ M of proteins within crude samples can be detected.

3 How is it known that the detected charge series corresponds to the overexpressed protein?

Distinguishing between background charge series and those corresponding to the recombinant protein produced is a critical step in the analysis. We therefore recommend first acquiring a reference spectrum from a sample prepared under conditions in which protein expression does not occur. Specifically, for bacterial expression systems, a sample of a cell lysate prior to induction of expression should be measured, while in the case of secretion systems, the growth medium from non-expressing cells, grown under the same conditions, should be analyzed. This step will enable mapping of background signals, which on the one hand might have masses similar to that of the target protein, and should therefore be distinguished from it, and on the other hand, may serve as indicators for cell state and integrity, as discussed below.

The presence of background peaks that exhibit molecular masses highly similar to that of the overexpressed protein is shown in Figure 5. In this example, a background charge state series with a mass of 146 kDa was identified in the growth medium of HEK293F cells grown to confluence in suspension, which corresponds to the tetrameric lactate dehydrogenase B (LDH) enzyme. This protein complex leaked into the medium from damaged cells. Due to the similarities in mass, the LDH charge series could easily have been confused with the secreted recombinant antibody, unless detected in the absence of protein expression.

Another important validation step involves monitoring the change in charge series intensity along the expression time course. Figure 6A-B shows the amplification of the signal-to-noise ratio as protein expression progresses, providing evidence for protein production. Over long expression periods, a reduction in signal intensity might occur, due to aggregate formation that affects protein solubility and its ability to ionize, or due to extensive masking of the signal by endogenous proteins that accumulate in the growth medium due to continuous cell death. Addition to the growth medium of antibodies or proteins that specifically bind to the target protein could also offer a means of identifying the overexpressed protein charge series. The shift in mass that is expected to occur following addition of the interacting protein would confirm expression of the target protein. Similarly, ligands and cofactors, or

chelating agents and processing enzymes that are known to specifically interact or react with the target protein, may be mixed with the crude sample to measure differences in mass, before and after their addition.

4 How can the optimal protein expression conditions and harvesting time be determined?

Monitoring the change in protein expression patterns over time using the direct-MS method provides a valuable tool for optimizing expression conditions and harvesting times, preventing protein misfolding and proteolysis, as well as determining the necessity of cofactor supplements to ensure proper folding and assembly. To this end, in order to assess the quality of protein production, it is important to compare the measured masses and the intensity ratio between the generated protein and a selected background peak at each time point⁸. At short time points post-induction/transfection/infection, it is expected that charge state series would not be easily resolved, due to the low relative abundance of the produced protein in comparison to the levels of certain endogenous background proteins (Fig. 6A-B). As protein production progresses, the relative intensity of the protein peaks will gradually increase, a feature that will reach a limit of detection linearity, due to the low abundance of the reference peak. The absence of a discernible protein signal during the time course analysis suggests that the protein is either not expressed, or it is not soluble (Fig. 2A). Similarly, a reduction in the relative abundance of the produced protein at longer growth periods hints towards the generation of a non-soluble fraction (aggregates/inclusion bodies), eliminating their detection by native MS, thereby suggesting the optimal time point for harvesting¹⁰.

Exhaustion of trace elements, cofactors and/or nutrients from the growth medium are identified by means of reduction in the mass of the recombinant protein as measured along the production time. The difference in mass between the holo- and apo-protein, will point towards the element that was originally associated with the protein, needing supplementation⁸. Such exhaustion of nutrients resulting in generation of the apo-protein form, may lead to structural destabilization and partial unfolding, features that will be reflected in a larger distribution of charge states with lower m/z values, in comparison to the folded protein. Thus, easy and informative time course measurements can be used for optimization of expression conditions.

The relative abundance of background peaks also provides essential information on the culture state. In particular, the accumulation of cytosolic proteins in the growth culture of secretion systems is attributed to ongoing cell death, which causes the cellular contents to spill into the medium. For example, the green fluorescent protein (GFP) that is often used as a reporter for infection efficiency, is an intracellular protein. Therefore, the detection of relatively high levels of GFP in the medium indicates that the growth cultures are exhausted and more pronounced cell death is occurring (Fig. 6B). Such is also the case when LDH is detected in human cells (Fig. 5A-B). High levels of LDH may serve as an indicator of the level of cell death during extensive culturing of mammalian cells. In summary, direct MS monitoring may be amenable to optimizing production yields by selecting the optimum harvesting point.

5 What is the minimum amount of culture required?

The bottleneck that defines the minimal culture volume is found in the sample preparation step, as only 1-2 μl are required per nanoflow capillary per experiment. For intracellular expression of recombinant proteins in bacterial cultures, the minimal volume that can be handled through the lysis process delineates the limit. The lowest volume that we have used is 1 ml of bacterial culture, lysed by sonication using a 3 mm diameter tip probe. Microtip sonication probes, however, will enable further reduction in the sample volume.

Nevertheless, in order to maintain the quaternary structure of the protein produced, care should be taken to prevent heat buildup during the sonication process, a scenario that is significantly reduced upon increasing the density of the bacterial suspension¹⁰. Similarly, lysis by sample boiling in the presence of detergents/reagents will disrupt the structure of the protein produced, and suppress the MS signal. Therefore, the method chosen for cell disruption should be such that the native protein structure, assembly state, and non-covalent interactions are preserved.

In the case of eukaryotic secretion systems, the growth medium needs to be buffer-exchanged into MS-compatible solutions. The minimum volume applicable to buffer exchange procedures defines the lower limit of culture. Most buffer exchange devices, such as microcentrifuge gel filtration columns and dialysis cassettes, require a minimal load volume of about 20-30 μl . To maintain complexes intact in solution a pH range of 6-8 is most commonly used, using aqueous volatile solutions such as ammonium acetate, triethylammonium acetate (TEAA) and ethylenediammonium diacetate (EDDA). Care must be taken that ionic strength concentrations are not too high, so as to preserve electrostatic interactions (Fig. 7). As can be seen in Figure 6C, in our lab, highly resolved spectra were obtained from 100 μl cultures grown in a 96-well plate format. Thus, the protocol can be extended to throughput screening procedures.

6 Can the amounts of overproduced proteins be quantified?

In general, MS does not constitute a quantitative approach, due to differences in ionization efficiency, charging and transmission across the instrument of protein ions³². However, the relative abundance of the produced protein may be monitored by selecting a reference background peak, and calculating the ratio of peak intensities for the overproduced and reference proteins. An upward trend line is expected to be generated over the growth period, reflecting the increase in protein accumulation; this may be then followed by a decrease, if the protein falls out of solution⁸. If a distinct background peak cannot be used as a reference, due to low intensities or fluctuations in appearance, a reference protein can be added externally (Fig. 4). For absolute quantification of production, an isotope-labeled (¹³C and/or ¹⁵N) form of the recombinant protein with a known concentration should be spiked into the sample, followed by calculation of the intensity of the produced protein, divided by that of the labeled reference, as per the AQUA method³³.

7 How are samples prepared for direct-MS?

To preserve the natural state of protein complexes and to ensure compatibility with native-MS requirements, analysis is generally carried out using millimolar concentrations of aqueous solutions composed of volatile molecules at physiological pH's. Owing to their

close-to-neutral pH and high volatility, the most commonly used solutions are ammonium acetate, ammonium carbonate, TEAA and DEAA, with minimal adducts to prevent ionization suppression of protein complexes.

For analyzing intracellularly produced proteins in bacterial expression systems, lysis of harvested cells should be performed in a volatile solution, supplemented with protease inhibitors. Following a centrifugation step, and removal of the insoluble pellet, spectra can be directly recorded from the generated supernatant (Fig. 1). Typically, due to the high levels of protein production, the sample should be serially diluted. As can be seen in Figure 3, in the absence of dilution, highly resolved peaks could not be detected; however, after 20-fold dilution, the quality of the spectrum was significantly improved, a feature that was maintained even after 140-fold dilution. At low protein production yields, a concentration step might be required to enhance the signal of the produced proteins above the background noise (though concentration will similarly affect the endogenous background proteins, the threshold for detection might be reached). The effect of increasing ammonium acetate concentrations³⁴, in combination with enhancement of higher energy collision-induced dissociation (HCD) voltages on the generated spectrum, can be seen in Figure 7. In this example, at low ammonium acetate concentrations and HCD voltage, only a partially resolved charge state series is obtained, whereas at a concentration of 1 M ammonium acetate and 150 V, a narrow charge state series is clearly resolved.

For secreted recombinant protein analyses, the crude medium should be cleared of cells and insoluble debris by centrifugation, the resulting cleared medium is then buffer-exchanged into a volatile solution at physiological pH. We found experimentally that performing 1–2 cycles of buffer exchange into high salt ammonium acetate using a microcentrifuge gel filtration column, efficiently replaces the bulk non-volatile compounds in the sample, and results in high quality spectra. However, other buffer exchange devices can be similarly effective. In some instances, an additional buffer exchange cycle, into a low concentration of ammonium acetate (150–200 mM) is required, to prevent weakening of ionic-strength interactions. This, for example, can be seen in the case of antibody production (Fig. 7), in which the presence of 1 M, as opposed to 150 mM, of ammonium acetate, promotes the dissociation of the light chain, following sample reduction and activation of the complex in the mass spectrometer. In summary, a starting range of 150 mM to 1 M ammonium acetate is suggested, though lower concentrations may be also used. A serial dilution into ammonium acetate (or other volatile solutions) will enhance the signal in cases of high protein production, while if the protein concentration is too low, concentration of the sample will be beneficial.

8 What types of mass-based experiments may be conducted?

The direct-MS method is not dependent on a specific mass spectrometer; rather it is a general approach applicable to any mass spectrometer that provides extended mass range; e.g., the Orbitrap- or QTOF-based platforms^{19–21}. The high mass range provided by these instruments not only enables the analysis of monomeric large proteins, but it also ensures the detection of high-mass oligomers formed by the recombinant protein, an assembly characteristic that is common to many proteins³⁵.

Once the sample is prepared, multiple layers of information can be obtained by integrating various types of native MS-based experiments, as undertaken with standard purified samples (Fig. 9). These include intact mass measurements, tandem MS (MS/MS), collision-induced dissociation (CID), top-down fragmentation by pseudo-MS³ experiments, collision cross-section (CCS) and collision-induced unfolding (CIU) measurements. As excellent protocols describing these approaches are already available^{17,36–39}, our detailed protocol, described below, will focus only on aspects that differ from those methods.

In general, the following method provides researchers focused on expressing recombinant proteins, tools to characterize production quality, without investing time and effort in protein purification. Moreover, as described in detail in the ANTICIPATED RESULTS section, the method can be extended to assess the activity, non-covalent protein interactions, and structural constraints of overproduced proteins. Figure 1 provides an overview of the various characteristics that may be revealed by direct MS measurements, as briefly outlined below:

Expression – The appearance of charge state series corresponding to the expected m/z range, which are enhanced over time, indicates that the target protein is expressed.

Solubility – Insoluble proteins will not give rise to an apparent signal. Therefore, detection of charge state peaks corresponding in mass to the expressed protein denotes that the protein is soluble, i.e. not accumulated in inclusion bodies and aggregates.

Folding – The folding condition of the protein produced is determined by the distribution of charge states that it acquires⁴⁰. A partially or fully unfolded protein would give rise to higher charge states with a broader distribution, in comparison to a folded protein.

Harvest – Multiple signs detected in the spectra can be used as hints for determining the optimal harvesting time. These include: decline in protein signal, increase in intensity of background peaks, increase in the protein charge series distribution and formation of an apo protein.

Molecular weight – The measured mass provides evidence for the production of the target protein. It can reflect the incorporation of mutations, binding of biomolecules, and assembly state.

Assembly state – Given an intact mass of the generated species, the assembly state encompasses the sum of individual subunit masses.

Overall fold – In IM-MS experiments, the measured drift time is converted to CCS values, which provide information on the protein(s) overall shape.

Stability – The generated protein's ability to tolerate increasing collision energy voltages reflects its stability. This is especially relevant when drawing comparisons between wild-type and mutated or designed proteins, and can be measured by generating IM-MS CIU profiles.

Biomolecule associations and PTMs – The shift between the protein's theoretical and measured mass may reflect non-covalent association with ligands, cofactors and

biomolecules, or covalently attached PTMs. Measurement under denaturing conditions, which liberates non-covalent-associated molecules from the generated proteins, distinguishes between the two possibilities. In addition, chelating or other specific agents that remove the associated biomolecule may be used to confirm the biomolecule's identity. Similarly, specific enzymes that remove the PTM may be used for validation.

Activity and protein interactions – By spiking relevant reactants into the crude sample, and monitoring the shifts in mass, overall signal intensity or peak area that consequently appear, the activity and/or interactions of the produced protein may be examined.

Sequence – Top-down pseudo-MS³ experiments that generate peptide fragments of the produced proteins enable sequence analysis.

Pairwise interactions – by adapting the double mutant cycle method to MS analysis, the strength of pairwise interactions may be determined directly from crude samples^{9,41}.

9 What are the current limitations of the direct-MS method?

The simplicity and feasibility of the approach across different platforms, coupled with the depth of information that is provided, are valuable features of the direct-MS method. However, like any methodology, it has its limitations. The method may be applied to bacterial hosts and secretion eukaryotic expression systems. Nonetheless, it is challenged by the limitations in intracellular production of recombinant proteins, particularly in eukaryotic hosts. In the latter case, standard methods of protein purification prior to native MS analysis are still needed. Moreover, the protocol that we describe here is only relevant to soluble proteins, as opposed to membrane proteins. However, an impressive recent study focusing on MS analysis of intact membrane proteins underscores the progress being made towards minimal sample preparation of these challenging complexes⁴².

Reagents

- Ammonium acetate solution, 7.5 M (Sigma-Aldrich, St. Louis, MO, USA, cat. no. A2706)
- Benzamidine Hydrochloride (Sigma-Aldrich, cat. no. 434760)
- BMGY and BMMY medium for *P. pastoris* cultures, containing 1% yeast extract yeast extract (BD Bacto™, cat. no. 212750), 2% pepton (BD Bacto™, cat. no. cat. no. 211677), 100 mM Potassium Phosphate (J.T. Baker, cat. no. 3252-01), 1.34% yeast nitrogen base (BD Bacto™, cat. no. 2191940), 0.4 µg/ml of biotin (Sigma-Aldrich, cat. no. B4639) and 1% glycerol (Sigma-Aldrich, cat. no. G5516).
- Rapid PNGase-F (NEB, Ipswich, MA, USA cat. no. P0711S)
- PNGase-F (NEB cat. no. P0704S)
- Model proteins: Cytochrome C (Sigma-Aldrich, cat. no. C2506), β-lactoglobulin (Sigma-Aldrich, cat. no. L7880), Concanavalin A (Sigma-Aldrich, cat. no.

- C2010), Alcohol dehydrogenase (Sigma-Aldrich, cat. no. A7011), Albumin from human serum (HSA) (Sigma-Aldrich, cat. no. A1653)
- PMSF - Phenylmethanesulfonyl fluoride (Sigma-Aldrich, cat. no. P7626)
- Pepstatin A (Sigma-Aldrich, cat. no. P4265)
- Ethanol medical, 70% BP (Gadot, Netanya, Israel, cat. no. 830107411)
- Ultra pure (type 1) water from the Direct-Q® 3 UV water purification system (Merck, Kenilworth, NJ, USA, cat. no. ZRQSVP3WW)
- LB medium for *E. coli* cultures, composed of an autoclaved medium containing 10 gr/L Tryptone (BD Bacto™, Franklin Lakes, NJ, USA, cat. no. 211705), 10 gr/L NaCl (J.T. Baker, cat. no. 3624-01) and 5 gr/L Yeast extract (BD Bacto™, cat. no. 212750)
- ESF 921 medium for insect cell suspension cultures (Expression Systems, Davis, CA, USA cat. no. 96-001-01)
- FreeStyle™ 293 Expression Medium for HEK293 cell suspension cultures (Thermo Fisher Scientific, Waltham, MA, USA, cat. no. 12338018)

Instruments and Equipment

- Mass spectrometers. In this study, we used Synapt G2 and G1 HDMS instruments (Waters MS Technologies, Manchester, UK), adapted for the measurement of high-mass proteins, and a Q Exactive Plus EMR Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany), modified for pseudo-MS³ top-down analysis⁴³. The latter instrument is also equipped with an electron capture dissociation (ECD) device (e-MSion, Inc., Corvallis, OR, USA), positioned in place of the original transfer octupole, and connecting the mass selection quadrupole with the C-trap.
- Flaming/Brown micropipette puller (Sutter Instruments Co. Novato, CA, USA), model P-97.
- Sputter coater (Electron Microscopy Sciences, Hatfield, PA, USA), cat. No. EMS550X, equipped with a gold target, 60 mm diameter by 0.1 mm thick (Electron Microscopy Sciences, Hatfield, PA, USA), cat. no. 91010.
- Vibra Cell sonicator (Sonics, Newtown, CT, USA), cat. no. VCX750, connected to a standard probe, 138 mm length with a 3 mm diameter tip, cat. no. 630-0422.
- Benchtop refrigerated mini-centrifuge (Eppendorf, Hauppauge, NY, USA, cat. no. 5418 R).
- Benchtop refrigerated centrifuge (Eppendorf, Hauppauge, NY, USA, cat. no. 5810 R), with a fixed angle rotor (F-34-6-38) and adaptors for 15 ml tubes (cat. no. 5814.776.003).
- 96-well plates for yeast and mammalian cell cultures (Corning®, NY, USA 14831 , cat. no. 3596).

- AeraSeal™ Sealing Film (Sigma-Aldrich, cat. no. BS-25).
- 1.7 mL MaxyClear snaplock polypropylene microcentrifuge tubes (Axygen®, Corning, NY, USA, cat. no. MCT-175-C).
- 2 ml graduated microtubes (SSIBio Lodi, California, USA.,cat. no. 1310-00).
- 15ml conical polypropylene centrifuge tubes, (Miniplast, Ein Shemer, Israel, cat. no. 835-015-40-111).
- Bio-Rad Micro Bio-Spin chromatography columns (Bio-Rad, Hercules, California, USA, cat. no. 732-6222).
- Borosilicate thin wall glass capillaries with filament (Warner Instruments, Hamden, CT 06514, USA.cat. no. G100TF-4). Alternatively, use commercially available capillaries such as borosilicate emitters (Thermo Fisher Scientific, Bremen, Germany), cat. no. ES380, or from other suppliers.

Software

MassLynx, version 4.1 (Waters, Hertfordshire, UK).

- Driftscope™ HDMSTM, version 2.8 (Waters).
- Protein Unfolding for Ligand Stabilisation and Ranking (PULSAR), version 2⁴⁴, <http://pulsar.chem.ox.ac.uk/>
- Thermo Tune, version 2.9 Build 2926 (Thermo Fisher Scientific).
- Thermo Xcalibur, version 4.1.31.9 (Thermo Fisher Scientific).
- Protein Deconvolution, version 4 (Thermo Fisher Scientific).

Reagent setup

The choice of MS-compatible lysis solution, including concentration, pH, as well as addition of protease inhibitors and other components such as cofactors or chelating agents, is dependent on the protein being expressed, and its downstream applications. Below we highlight critical points that are relevant for lysis solution of recombinant proteins accumulated in cells or buffer exchange solution for secreted proteins.

- pH – Before choosing the pH of the MS-compatible solution, check the pI of the protein, and adjust the pH of the solution accordingly.
- Ammonium acetate concentration – We typically use 1 M ammonium acetate at pH ~7. For most proteins, this solution results in minimal accumulation of adducts and high signal-to-noise ratios during measurements (Figure 8). However, lower concentrations may also be used, and additional buffer exchange steps may be performed after lysis.
- Protease inhibitors – We suggest adding protease inhibitors to the solution to protect the recombinant proteins from endogenous proteases, as follows: 1 mM Benzamidine, 1 mM PMSF, and 1.4 µg/ml Pepstatin A. You may omit the

protease inhibitors if they interfere with downstream applications in the lysate; for example, if the protein produced is a protease, and its activity is being analyzed, protease inhibitors may interfere with its function. Alternatively, if your protein contains intrinsically disordered regions, in the absence of protease inhibitors it is susceptible to degradation. In this case, keep in mind that it may not remain intact for more than a few hours. Make sure to keep your lysate on ice throughout the entire procedure.

- Cofactors – If your protein is associated with cofactors or other small molecules, it is possible to supplement the lysis solution with them. Since the concentration of the recombinant protein in the lysate is not known, we typically add around 100 - 500 μM of the cofactor, preferably as ammonium or acetate salts. Note that these concentrations are lower than those typically used in biochemical assays; nevertheless, non-specific binding to the protein can still occur, due to desolvation that takes place during the ionization process. In such cases, reduce the cofactor concentration in a stepwise manner, to eliminate non-specific adduct binding and monitor the level of binding saturation.
- Chelating agents – To probe metal binding properties of the recombinant protein, chelating agents such as EDTA or EGTA may be added to the MS-compatible lysis solution, or directly into the crude lysate, following lysis. The amount of chelating agent required depends on the concentration of the metals in the sample, and the affinity between the protein and the metal ions. In our lab, the amount of chelating agent required to strip off bound Ca^{2+} from the secreted CBM3a expressed in *P. pastoris*, was 50 mM^8 . After addition of the chelater, incubate the lysate for 15-30 min on ice. Since high concentrations of chelating agents are usually not compatible with native MS measurements, perform a subsequent buffer exchange step without the chelater.
- Reducing agents – some proteins may require a reducing environment. In general, reducing agents can be added to the MS-compatible solutions, typically up to a concentration of approximately 1– 5 $\text{mM}^{17,45}$. We recommend to use DTT or TCEP and not β -mercaptoethanol, since the latter may covalently bind to proteins and lead to a mass increase of 76 Da^{46} . In case you need to use higher concentrations of reducing agents (such as for example in the case of antibody reduction, where we used 20 mM TCEP), a subsequent step of buffer exchange might be required.

Equipment setup

typical instrument parameters

Q Exactive Plus EMR Orbitrap			
Parameter	Glycans (0-2 kDa)	Proteins 10-80 kDa	Proteins >80 kDa
Scan range (m/z)	250 - 2000	1500 - 8000	2000 - 12000

Q Exactive Plus EMR Orbitrap			
Parameter	Glycans (0-2 kDa)	Proteins 10-80 kDa	Proteins >80 kDa
Resolution	70,000	17,500	10,000
Inject time (ms)	250	250	500
Trapping pressure	1	1.5	2 - 4
Capillary temp (°C)	160	160	160
Spray voltage (kV)	1.3 - 1.7	1.3 - 1.7	1.3 - 1.7
Bent flatpole DC bias (V)	1.5	1.5 - 1.8	2 - 2.4
Bent flatpole gradient (V)	5	10 - 15	25 - 40
HCD (V)	0	0 - 50	10 - 150
Central Electrode Inject (V)	3800	3200	3200

Parameter	Synapt G2		Synapt G1	
	10-80 kDa	>80 kDa	10-80 kDa	>80 kDa
Protein mass	10-80 kDa	>80 kDa	10-80 kDa	>80 kDa
Backing pressure (mbar)	4	6 - 8	4	6 8
Capillary (kV)	1.2 - 1.9	1.2 - 1.9	1.2 - 1.9	1.2 - 1.9
Sampling Cone (V)	20	25	10	25
Extraction Cone (V)	2	5	2	5
Temperature (°C)	25	25	25	25
Cone gas (L/h)	30	50		
Purge Gas (L/h)	50	100		
Trap collision energy (V)	10	35	10	20
Trap DC bias (V)	45	45	22	22
Transfer collision energy (V)	4	4	4	10
Trap gas (ml/min)	4	8	4	6
Helium cell (ml/min)	120	120		
IMS gas (ml/min)	60	40	24	24
Trap wave velocity (m/s)	160	160	160	300
Trap wave height (V)	4	4	4	4
IMS wave velocity (m/s)	300	250	200	250
IMS wave height (V)	20	15	10	15

Preparation of samples from bacterial cultures

Timeline: ~ 30 min

1. Use ~10 ml of induced bacterial culture expressing the desired recombinant protein, depending on your lab's working procedures.

Critical step

The bacterial strain used for the production of recombinant proteins may have implications on the acquired direct MS spectra. For example, if you work with bacterial strains that co-express chaperones that assist in increasing the overall yield of your recombinant protein, such as ArcticExpress cells (Agilent technologies), or the Chaperone Competent Cell BL21 Series (Takara), the chaperons may also appear in your spectra.

2. Transfer the culture to a 15 ml tube and spin for 5 min at 5,000 *g*, at ambient temperature.
3. Discard the liquid and resuspend the cell pellet in 1 ml of 150 mM ammonium acetate.

Caution

Discard the biological waste according to your institute's safety regulations.

Critical step

Washing the cells prior to lysis will clear the sample from contaminating materials present in the culture medium, and improve the signal-to-noise ratio of the downstream measurements. We typically wash cells in MS-compatible ammonium acetate, at a concentration of 150 mM, pH~7, to keep the solute close to physiological conditions. Washing the cells in water only or in 1 M ammonium acetate (a solution we subsequently employ during the procedure) should be avoided, because it will generate a hypotonic or hypertonic environment, respectively, and may result in cell membrane disruption, and spilling of cellular contents into the medium prior to lysis.

4. Transfer the cells to a 2 ml tube and spin for 3 min at 5,000 *g*, at ambient temperature. Discard the liquid.

Pause point

At this stage, cells can be stored in tubes at -80°C, or processed directly.

5. Resuspend the cells in 2 ml of MS-compatible lysis solution, composed of 1 M ammonium acetate and protease inhibitors.
6. Transfer 1.5 ml of the resuspended cells into a new 2 ml tube, and store on ice.

Critical step

For sonication, it is recommended to use a 2 ml tube with a minimal cone at the bottom. A conical tube would increase the chances that the sonication tip would touch the tube walls during the procedure, and interfere with cell lysis. A 2 ml tube will provide sufficient volume for 1.5 ml of resuspended cells, and proper immersion of the sonication tip without spillage during the process. If smaller volumes of resuspended cells are used, a micro-tip may be employed, and sonication performed in smaller tubes.

Caution

Sonication is a process that generates high-frequency sound waves, which convert electrical signals into physical vibrations to disrupt cells. The exposure to these high frequency waves

may be hazardous due to hearing damage. To avoid this hazard, wear earphone-type sound mufflers to protect your ears while sonicating, locate the sonicator in a sound-proof cabinet and do not sonicate in a room with people that are not wearing ear protection.

Critical step

When less than 1.5 ml of resuspended cells are used for sonication with the standard probe, described in the Instruments and Equipment section, care should be taken as to prevent heating up and foaming of the cell suspension.

7. Prepare a small ice bucket, filled with tightly packed ice. Make sure that it can comfortably fit into the sonication chamber, and allow for free movement of the sonication tip, up and down the tube.
8. Place the tube in the ice bucket so that its cap levels with the ice surface. Press the ice around the tube to stabilize it firmly within the ice.

Critical step

Make sure that the tube is firmly placed in the ice, in a vertical position. Otherwise, the tube might wiggle during sonication and cause the sonication tip to touch the tube wall during the procedure. The entire tube must be covered with ice, to minimize overheating of the sample during sonication.

9. Clean the sonication tip with 70% ethanol and then with water. Wipe the tip gently after each wash.
10. Open the tube and insert the ice bucket into the sonication chamber. Adjust the height of the bucket such that the tip is almost completely immersed in the tube. Make sure that the tip does not touch the tube walls or bottom.
11. Sonicate the sample until full lysis is achieved, depending on the instrument type and model. Our sonication parameters are set to 10 min, at cycles of 5 sec ON and 25 sec OFF, at an amplitude of 35%.

Critical step

We recommend monitoring the sonication process and inspecting the tube in between sonication cycles. If the sample starts to foam, stop sonication immediately. Readjust the tip position, make sure that it is immersed deep enough in the solution, and prevented from touching any tube wall. During sonication, the tube might gradually sink down into the ice. Therefore, readjust the height of the ice bucket between sonication cycles, to make sure that most of the tip is immersed in the solution at all times.

12. Spin the sonicated sample at 4°C, for 10 min at 16,900 *g*.
13. Transfer the supernatant to a new tube. Freeze the pellet at -20°C for further analysis, if troubleshooting is required.
14. Divide the supernatant into 50 µl aliquots. If analysis is not performed immediately, snap freeze in liquid nitrogen. Store at -80°C.

Caution

Liquid nitrogen can cause severe health hazards. The vapors of liquid nitrogen can quickly freeze skin tissue and eye fluids, and result in cold burns and even permanent eye damage. Proper handling of liquid nitrogen includes working with appropriate containers in a well-ventilated area, gentle handling to reduce the danger of boiling and splashing, and the use of appropriate body protection, as recommended by your institute's safety unit.

Pause point

Once stored in -80°C , samples can be processed at any time.

Preparation of samples from eukaryotic cells

Timeline: ~ 30 min

15. Grow your culture of choice (e.g., yeast, insect cells, mammalian cells) according to the specific requirements for overexpression (after transfection / infection / transformation of cells, or after appropriate induction). Culture growth time may vary, depending on the type of cells used (Fig. 6A-B). Culture volumes can be kept to a minimum, depending on the cell type, and your growth chamber facility (Fig. 6C).

Critical step

The composition of the culture medium may have significant effects on the outcome of direct-MS measurements. The growth medium most commonly used in mammalian cell cultures typically comprises ~ 10% bovine serum containing ~ 50 mg/ml albumin⁴⁷. Such a high concentration typically interferes with direct-MS measurements, as well as with classic purification processes. To overcome this difficulty, the use of protein-free growth media, which nowadays are broadly used, particularly in pharmaceutical applications, is highly recommended^{48,49}.

Critical step

Cells can be grown in minimal culture volumes. When utilizing 96-well plate formats, the risk of cross-contamination, spillage and evaporation is high. To avoid these problems, seal the culture plate with a sealing film that allows uniform gas exchange within the wells.

16. Transfer 100 μl - 1 ml of culture into a 1.5 ml tube and spin for 5 min at 1,000 *g*, at ambient temperature, to pellet out the cells.

Critical step

When handling intact cells for direct-MS analysis, perform the centrifugation at ambient temperature in order to reduce cell damage, which can result in spillage of cellular content into the growth medium. Similarly, do not spin the culture medium at higher speeds than 600 - 1000 *g*, since cells, particularly those lacking a cell wall, may rupture at high centrifugation forces and spill their contents into the growth medium.

Caution

Discard the biological waste, according to your institute's safety regulations.

17. Collect the growth medium and spin again at 16,900 *g* for 10 min at 4°C, to pellet out insoluble materials and debris.
18. Collect the cleared growth medium and supplement it with protease inhibitors, as described in the Reagent Setup section.
19. Divide the growth medium into 50 µl aliquots. If analysis is not performed immediately, snap freeze in liquid nitrogen. Store at -80°C.

Pause point

Once stored in -80°C, samples can be processed at any time.

20. On the day of the measurement, equilibrate a buffer exchange column with the appropriate MS-compatible solution of choice, and perform buffer exchange according to the manufacturer's instructions.

Critical step

The choice of MS-compatible solution, pH, concentration, as well as addition of components such as cofactors, is dependent on the expressed protein and its downstream applications, as detailed in the Reagent Setup section. We typically buffer exchange the growth medium into 1 M ammonium acetate. In many cases, a second buffer exchange step is required, and a third cycle can further improve the measurement quality (Fig. 7B). The number of buffer exchange cycles typically does not affect the relative ratio between the recombinant antibody and the other proteins in the crude sample (Fig. 7A). In specific instances, the concentration of the MS-compatible solution can also affect the measurement; such is the case, for example, with antibody stability assays (Fig. 7C).

Direct-MS measurements of proteins from crude samples

Timeline: ~ 30 min – several hours

In the next section, we will describe measurements performed on our high-mass instruments either the modified Q Exactive Plus EMR Orbitrap⁴³ or Synapt G1 and G2 mass spectrometers¹⁵. General lists of recommended parameters to be used in mass measurements are shown in the section of Equipment Setup.

21. Load 2-3 µl of crude sample into a capillary (pulled and gold-coated using a micropipette puller and a sputter coater)⁵⁰ and spray into the instrument.

Critical step

In order to avoid miss-assignment of the spectra, always measure as control a lysate / growth medium from cells that do not express your target protein. Thus, you will detect the major background charge states in the sample, and will easily distinguish between them, and those of your protein of interest. However, inducible bacterial expression systems are often leaky

and exhibit basal expression of the recombinant protein even before induction¹⁰. In such cases, the charge state series of the recombinant protein may already be visible before induction, but will typically be at intensities close to those of other endogenous proteins.

22. Optimize MS conditions as detailed in the section of equipment setup, shown above. In Orbitrap-based instruments, set the averaging to a low value (around 10) and the resolution to 10,000. In most cases, a spectrum of the recombinant protein will be easily obtained.
23. The major requirement for an accurate measurement of crude samples is that the desired protein outperforms the endogenous background proteins in the sample. A low concentration of the recombinant protein, however, is not a limitation *per se*. We found that even low amounts of the overexpressed proteins, present in a background environment of relatively low complexity, are sufficient to obtain well-resolved spectra (Fig. 2B). In case a spectrum is not obtained, refer to the TROUBLESHOOTING TABLE below.

Critical step

Mass measurement of crude samples at high resolution can result in a reduction in the signal-to-noise ratio, due to an increase in overall background signals from other endogenous proteins and contaminants (Fig. 10). Therefore, in our modified Q Exactive Plus EMR Orbitrap we typically measure crude samples at a resolution of 10,000, and increase when required, according to the sample characteristics.

24. Increase the averaging to 100, wait for about 30 sec for the measurement to average, and record spectra for 1-5 min, depending on the signal intensity. In general, in both QTOF and Orbitrap platforms, longer acquisition times and higher averaging will improve peak resolution and signal-to-noise ratio⁵¹.

Glycoprotein analysis directly from the crude sample

Timeline: 3-10h

Here we describe a simple approach for characterization of protein glycan modifications in crude samples. We tend to perform such experiments on our modified Q Exactive Plus EMR Orbitrap, due to its high sensitivity and resolution capabilities, however QTOF based platforms are suitable as well.

25. Collect 100 μ l of growth medium from a culture of cells secreting your protein of choice, at the appropriate time for harvest.
26. Add protease inhibitors and buffer exchange twice into 1 M ammonium acetate, and a third time into 150 mM ammonium acetate.
27. Spray the sample into the instrument. Set the instrument parameters for gentle conditions (we typically work with an HCD energy of between 10 to 50 V), and acquire well-resolved spectra of the protein at a range of resolutions, starting at 10,000 and up to 50,000. Initially use a broad mass range (as 1,000 to 30,000

m/z), however, after an initial spectrum is recorded, this mass range can be narrowed accordingly.

Critical step

Gentle activation energy (HCD energy, in the case of the Q Exactive Plus EMR Orbitrap) is important to keep modifications such as glycans intact, and attached to the protein.

Inspection for backbone modifications and glycosylations

28. To gain insights into the protein's multiple co-existing states, carefully inspect the spectra measured in step 27. A spectrum of the intact protein will display the major forms of the differentially glycosylated species, as shown in Figure 5B.
29. Deconvolute the spectrum using your software of choice. We use Deconvolution 4 (Thermo Fisher Scientific), and apply the following parameters for the deconvolution of antibody spectra: peak model – intact protein; minimum adjacent charges 4-10; noise rejection 95% confidence; m/z range 5500-8000; output mass range 143,500 – 150,000; target mass 147,000; mass tolerance 20 ppm; number of iterations 3.
30. Calculate the theoretical mass of the protein based on its amino acid sequence, and compare it to the measured mass, taking into account major modifications such as phosphorylations, acetylations, glycosylations, and others (Table 1). Accordingly, attempt to assign the potential modifications of your protein. In Figure 5, we demonstrate glycan characterization of a designed antibody against lysozyme. The theoretical mass of the antibody backbone is 144,464 Da, and the major mass we measured was $147,032 \pm 4$ Da. The 2,569 Da mass difference may be explained by the loss of two heavy chain C-terminal lysines (-128 Da x2), formation of 16 disulfide bounds (-2 Da x16) and the addition of two G0F glycan moieties (1,455 Da x2)^{22,52} (see BOX 1). Deeper sequence analysis, and confirmation of the type and position of post-translational modifications, can be further validated by MS³ top-down analysis (see ANICIPATED RESULTS).
31. To validate that the mass shift indeed corresponds to the modifications described above and to N-glycosylations, perform a deglycosylation step using an enzyme that specifically removes N-glycans. Take 50 μ l of crude sample, add 12.5 μ l Rapid PNGase-F reaction buffer and 1 μ L of the MS-compatible Rapid PNGase-F enzyme (New England Biolabs), and incubate at 50°C for 10 min, according to the manufacturer's instructions.
32. Buffer exchange the sample into 150 mM ammonium acetate, and measure the sample as described in step 28.

Critical step

For the purpose of deglycosylation and inspection of the antibody backbone, we recommend the use of MS-compatible enzymes; e.g., the MS-compatible Rapid PNGase-F, according to the manufacturer's instructions (i.e. using the supplied reaction buffer, the recommended

enzyme amount and incubation times). We found that deglycosylation in crude samples is quite efficient and results in a high level of deglycosylation. In such cases, one-step of buffer exchange following the enzymatic treatments is typically sufficient to remove buffer components that may interfere with ionization. Note that in these cases, you may detect the PNGase-F enzyme (~ 35 kDa, charge state series around 3,000 – 3,500 m/z) in your sample⁸. If no MS-compatible enzymes are available, standard enzymes may be used. In such cases, we add far lower amounts of the enzyme than recommended, sometimes even without addition of the recommended reaction buffer, to ensure compatibility with native-MS requirements. Compensate for the low enzyme amount by incubating the reaction for a prolonged period, before performing the buffer exchange. In these cases, lower efficiency of the enzymatic activity may occur.

Critical step

In addition to PNGase-F, crude samples may be treated with additional enzymes, such as those that cleave O-glycans or other sugar moieties, alkaline phosphatase that remove phosphorylations or phosphatase inhibitors that protect the produced proteins from dephosphorylation by cellular enzymes.

33. After deglycosylation, the major charge state series obtained will correspond to the non-glycosylated protein (However, it will include the other covalent modifications, described in step 30).

Glycan analysis directly from the crude sample

34. To characterize the major glycans of the protein, we rely on the assumption that the recombinant protein is the major glycosylated protein in the crude growth medium. Therefore, the majority of the glycans, released by the PNGase-F treatment, will originate from this protein. A control spectrum obtained from a medium of non-expressing cells, grown under the same conditions, prior to and following PNGase-F treatment will validate this assumption.
35. We typically measure glycans on our Q Exactive Plus EMR Orbitrap instrument, but measurements can also be performed on other instrumental platforms. Start by measuring a sample of the crude medium from step 27.
36. Focus on the low range of the m/z region, where intact glycans are detected (in our lab, N-glycans are typically measured in the form of singly charged ions, in the range of 800 – 3,000 m/z). Set instrumental parameters for the analysis of glycans and do not apply any HCD energy. Acquire spectra at a resolution of 70,000. Note that prior to deglycosylation, we do not expect to detect free glycans in this measurement. Peptides and other contaminants, however, may be detected (Fig. 5A).

Critical step

Make sure that no HCD energy is applied during this measurement, since elevated acceleration energies can fragment intact glycans.

37. To cleave the glycans from the protein, add 0.2 μ l of PNGase-F to a 50 μ l sample, and incubate for 5-8 hours at 37 $^{\circ}$ C.
38. Spray the deglycosylated sample into the instrument and measure again, as described in steps 35 – 36.

Critical step

If you wish to examine the intact glycans that were cleaved off the protein, do not perform an additional buffer exchange step after deglycosylation, since glycans are typically smaller than the size cutoff for the buffer exchange columns used for intact proteins, and will be lost during the process. In addition, do not use the deglycosylating enzyme together with its reaction buffer. Instead, use a minimal amount of the regular PNGase-F enzyme, and compensate for the low enzyme amount and suboptimal reaction conditions with a long incubation time.

Note that in this case, the enzymatic reaction will not be complete; rather, only partial deglycosylation will occur. Therefore, this sample should not be used to measure the intact, deglycosylated protein.

39. Start by inspecting the spectrum obtained in step 38, and search for species corresponding in mass to common glycans (BOX 1). In proteins expressed in mammalian cells, the most common glycan is typically an N-linked branched glycan⁵³, whereas in proteins expressed in insect cells, the most common glycan is a core mannose containing one fucose moiety, corresponding in mass to 1,039 Da⁵⁴. Common glycans, typically found in this type of experiment, are protonated ions $[M+H^+]$ or sodiated ions $[M+Na^+]$. In our case, the major glycosylation moieties found in the analysis of the intact antibody were sodiated G0F and G1F (Figs. 5 and 11).

Troubleshooting

The tables describes problems that become evident during the direct-MS measurement (step 22).

Problem	Possible reason	Solution
The desired charge state distribution is not detected and only background signals are obtained.	Above-optimal concentration of the sample, particularly when overexpression levels are very high (Fig. 3). High volume of bacterial cultures, > 10 ml, (step 1), or sonication in low volumes of lysis solution (step 5) may also lead to this problem.	Begin by diluting the sample, using the MS-compatible buffer of choice. In our lab, dilution of 5- to 10-fold typically solves the problem; however, dilutions up to 140-fold have also been proven to yield excellent results in highly concentrated lysates.
	The recombinant protein is not expressed, or its levels are similar to that of the endogenous proteins.	Run an SDS-PAGE gel and monitor the level of the recombinant protein in comparison to the endogenous proteins levels, by means of Coomassie staining. In the case of proteins expressed in bacteria, run equivalent samples of both the soluble and insoluble fractions of the cell lysate (Fig. 2).

Problem	Possible reason	Solution
		Solubility problems and low expression levels of recombinant proteins can be tackled by optimizing the expression conditions ^{29,56} .
The spectrum of bacterial lysates contains additional dominant charge state series, which do not correspond to the protein of interest.	Overexpression is insufficient, either due to sub-optimal induction conditions, or to low solubility of the recombinant protein (step 1). Alternatively, it is possible that your culture was over-induced, resulting in gradual shifting of the recombinant protein into insoluble inclusion bodies (step 1).	Monitor soluble and insoluble bacterial fractions by SDS-PAGE and Coomassie staining, and identify the expression level and solubility of the protein. Screen for optimal induction conditions, such as concentration of the inducer, induction time, growth temperature etc.
	The concentration of the inducer (such as IPTG) is not optimal. In leaky expression systems, such as the lactose operon-based system, background levels of the recombinant protein will be detected even in the complete absence of the inducer.	Screen for optimal concentration of the inducer.
The spectrum of secreted proteins from the crude growth medium contains additional dominant charge state series, which do not correspond to the protein.	Over-culturing of the cells (step 15), is often accompanied by continuous cell death, which results in spillage of the cellular contents into the growth medium. In such cases, highly expressed endogenous proteins, such as LDH (146 kDa), or even GFP (27.5 kDa), will be detected. The latter is commonly used, as a marker for infection / transfection ⁸ .	Screen for shorter induction times. Choose the induction time in which you achieve the optimal tradeoff between expression level of the recombinant protein and levels of the contaminating endogenous proteins.
Additional charge state series overlap with the charge state series of the recombinant protein of interest.	This problem may occur due to over-culturing, as detailed above in the troubleshooting table.	When analyzing a protein complex, it is possible to distinguish between the recombinant protein and contaminating proteins, by means of tandem MS at elevated collision energies. Another option is to add charge-reducing agents such as TEAA, at a ratio of 0.9/0.1 (v/v) between the ammonium acetate concentration in the sample, and the concentration of the reducing agent. Following charge reduction, the distance between each two adjacent ions will increase, thus making it easier to distinguish between them. Alternatively, IM-MS spectra can be recorded which spread the data into a third dimension, reducing peak overlap.
The desired charge state series is detected, but the signal-to-noise ratio is too low.	The recombinant protein is not sufficiently over-expressed.	Troubleshoot as described above.
	The measurement is performed at too high resolution, resulting in a loss of signal (step 22).	Reduce the resolution of your measurement. In Orbitrap instruments, you can also increase the maximum injection time, to increase the accumulation time of ions in the C-trap per scan. Increase averaging.
The protein of interest does not appear bound to its cofactor.	The measurement is conducted at instrumental conditions that result in dissociation of the cofactor from the protein (equipment setup).	Measure at more gentle conditions, such as reduced pressure, lower capillary voltages, lower collision energies, etc.
	There are no sufficient endogenous sources for the required cofactor (reagent setup).	Supplement the culture or the crude sample with the cofactor.
The recombinant protein does not adopt the expected assembly state.	The recombinant protein is not expressed to sufficient levels (Figure 6A, Hsp32 induced for 1h)	Increase the induction time.
	A cofactor, essential for assembly, is missing.	Add the required cofactor to the growth medium, and to the lysis solution.

Problem	Possible reason	Solution
	The ionic strength of the lysis solution is too high, and disrupts ionic-strength interactions (Figure 7C).	Perform a round of buffer exchange to reduce the ionic strength of the solution.

Anticipated results

To illustrate the applicability of the direct-MS method, we demonstrate results from three different systems, each of which were analyzed according to the described protocol. We begin with a computationally designed heterodimer, and illustrate how direct analysis of the crude lysate enables rapid validation of the *in silico* prediction, generating input on the assembly state, sequence and pairwise interactions of distinct amino acids. Next, we demonstrate how the activity of a produced protein, and its interactions with other proteins, may be assessed by inserting the relevant components into the crude cell lysate. Finally, we exemplify analysis of an intact antibody produced in human cells, focusing in particular on glycan analysis.

Multilevel analysis of a computationally designed heterodimer

In *E. coli* bacteria, we co-expressed a pair of computationally designed toxin / anti-toxin proteins, the colicin endonuclease (colE) and immunity (Im) heterodimers⁵⁷, using a single mRNA⁹. Native MS analysis of the crude *E. coli* lysate indicated that the colE and Im variants were expressed, and that they form a soluble and folded heterodimer (Fig. 12A). The assembly was then trapped and activated in the front-end trap of a high-mass range Orbitrap instrument, to induce the heterodimer's dissociation into the constituent colE and Im variants (Fig. 12B). We then applied pseudo-MS³ analysis by selecting either the colE or Im variants within the quadrupole mass analyzer, followed by fragmentation in the ECD and HCD cells. The multiply charged fragments generated were then detected at high mass resolution, enabling sequence identification (Fig. 12C-D).

Thus, as demonstrated here for an *in silico* designed protein pair, we expect that this method would be applicable to rapid screening of libraries of engineered or randomly mutated proteins. Initially, protein characterization under native conditions would be used for selecting the pool of proteins that forms the desired interaction. This pool would then be immediately subjected to sequencing of the relevant proteins. Such a workflow would alleviate the need for combining gene sequencing with protein-based characterization.

Such an analysis can then be taken a step further, by determining the strengths of intermolecular pairwise interactions between residues at the interfaces of the interacting proteins. The approach is based on the double mutant cycle method, wherein the two target residues are mutated both separately and in combination, usually to alanine, and the energetic effects of the mutations are determined. We previously showed that pairwise interaction energies may be determined from a single high-resolution native mass spectrum directly from crude lysates, by measuring the intensities of the complexes formed by the two wild-type proteins, the complex of each wild-type protein with a mutant protein, and the complex of the two mutant proteins^{9,41}. Figure 12G-H exemplifies an analysis of the

interactions between Gln31 in Im, and Asn78 in colE, in the designed version of the complex. The pair of genes coding for the WT and mutant forms of each of the variants, were cloned in tandem into the pRSF-Duet expression vector; the expression itself was performed in *E. coli*. Direct-MS measurements indicated that the charge series of the four individual complexes could be well-resolved in the crude lysates (Fig. 12G), enabling rapid calculations of the coupling energy (Fig. 12H; details of the calculation procedure may be found in ref⁴¹). Thus, direct-MS analysis not only obviates the need for prior protein purification, but also yields estimates for coupling energies under semi-crowding conditions that mimic those of the cellular environment.

Determination of protein activity and protein-protein interactions

Direct-MS can be utilized for “quick and dirty” determination of the functions and interactions of proteins, as shown here for RAB1A. This protein was suspected to be ‘moonlighting’ as a 20S proteasome inhibitor, while being enzymatically active in other cellular pathways. We therefore overexpressed RAB1A in *E. coli*. As control, we added α -synuclein, a 20S proteasome substrate, to a non-expressing crude bacterial lysate and monitored its levels before and after spiking in the 20S proteasome complex. We repeated this experiment using the cell lysate expressing RAB1A and quantified the levels of α -synuclein over time (Fig. 13A-B). α -synuclein remained stable in the absence of the 20S proteasome; however, after the proteasome’s addition, it was rapidly degraded. On the other hand, the degradation rate of α -synuclein in the lysate of RAB1A-expressing cells showed a marked decrease. Our results suggest that RAB1A reduces the rate of α -synuclein degradation by the 20S proteasome, over the course of the experiment, and that isolation and purification of RAB1A is not a prerequisite for performing the degradation assay.

Next, we explored whether the ability of RAB1A to reduce the proteolytic capacity of the 20S proteasome constitutes a direct effect; that is, whether RAB1A physically binds the complex. To this end, bacterial cultures over-expressing RAB1A were lysed, and incubated with purified 20S proteasomes. The charge states corresponding to the 20S proteasome complex in the MS spectrum were subjected to tandem MS analysis, at high HCD energy. Comparison of the subunits that were ejected from the free 20S spectrum with those that were ejected from the 20S proteasome incubated with crude lysate of RAB1-expressing cells revealed additional peaks that corresponded in mass to RAB1A (Fig. 13C). By extrapolation, we can therefore conclude that prior to tandem MS analysis, RAB1A was bound to the 20S proteasome.

From intact antibodies to glycosylation characterization

The use of antibodies as therapeutic modalities is rapidly expanding⁵⁸. Considering that the development of antibodies for clinical applications requires optimization of multiple attributes such as binding affinity, specificity, folding stability, and solubility, an efficient assessment procedure of the designed antibodies is required⁵⁹. Such an analytical approach is critical during every step of the antibody-generating process, to validate the safety and efficacy of the product, as well as to ensure batch-to-batch consistency. Similarly, the assessment step is also necessary, for drawing comparisons between biosimilar antibodies

and their original reference products. We anticipate that the protocol we outlined here may facilitate such analyses, as exemplified in Figure 11.

Initially, we co-transfected suspension-grown HEK293F cells, with plasmids for the expression of light and heavy chains of the monoclonal, designed variant of the anti-lysozyme antibody D44.1_{des}⁸. We grew the cells for three days in FreeStyle™ 293 Expression Medium, then collected the medium, acquiring data directly after performing a buffer exchange step into a MS-compatible solution. Under these growth conditions, no concentration steps were required, and a nicely resolved mass spectrum of the intact antibody was present with no free heavy or light chain ions, demonstrating efficient antibody assembly (Fig. 5B).

Moreover, mass measurement indicated a ~3 kDa shift between the measured and calculated masses, suggesting that the antibody is glycosylated. To validate this assumption, and test whether the mass shift resulted from glycosylations, we treated the medium with PNGase-F. A shift in molecular mass measured in a spectrum recorded after PNGase-F treatment, confirmed the release of glycans from the antibody (Fig. 5B-C). Examination of the low m/z region of the spectrum prior to and following PNGase-F treatment indicated the appearance of two main glycans; namely, the sodiated G0F and the G1F species (Fig. 11A-B), which were only detected in the spectrum of the treated sample. We then applied the tandem mass spectrometry (MS/MS) approach, to further validate the composition of the glycans. Figure 11 shows the isolation and collisional activation of the G0F glycan to induce its fragmentation. Assignment of the fragments of the glycan confirmed its structural composition (Figure 11D-E). We conclude that laborious antibody purification can thus be bypassed through high-resolution native mass spectrometry analysis.

Acknowledgments

We would like to thank Ely Morag and Ed Bayer for providing us with the pPICK9 plasmid for CBM3a expression, Yoav Peleg for providing pPICK9 plasmid for HSA expression. We are grateful to Olga Kersonsky and Sarel Fleishman for providing us with the plasmid for the expression of MBP-PTE. We also want to thank Shira Warszawski, Aliza Katz, Ron Diskin and Sarel J. Fleishman for providing us with the growth media containing secreted antibodies, and Ron Diskin, Hadas Cohen-Dvashi, Meital Yona and Tamar Unger for providing us with the growth media containing the secreted TfR1. We are also grateful for the support of a Starting Grant from the European Research Council (ERC) (Horizon 2020)/ERC Grant Agreement No. 636752, and for an Israel Science Foundation (ISF) Grant 300/17. M.S. is the incumbent of the Aharon and Ephraim Katzir Memorial Professorial Chair.

References

1. Adrio JL, Demain AL. Microbial enzymes: tools for biotechnological processes. *Biomolecules*. 2014; 4:117–139. DOI: 10.3390/biom4010117 [PubMed: 24970208]
2. Carter PJ. Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp Cell Res*. 2011; 317:1261–1269. DOI: 10.1016/j.yexcr.2011.02.013 [PubMed: 21371474]
3. Gainza-Cirauqui P, Correia BE. Computational protein design-the next generation tool to expand synthetic biology applications. *Curr Opin Biotechnol*. 2018; 52:145–152. DOI: 10.1016/j.copbio.2018.04.001 [PubMed: 29729544]
4. Assenberg R, Wan PT, Geisse S, Mayr LM. Advances in recombinant protein expression for use in pharmaceutical research. *Curr Opin Struct Biol*. 2013; 23:393–402. DOI: 10.1016/j.sbi.2013.03.008 [PubMed: 23731801]
5. Holtzhauer, M. *Basic Methods for the Biochemical Lab*. Springer: 2006.


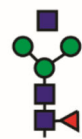
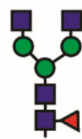


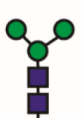







6. Miles, AJ, Wallace, BA. *Biophysical Characterization of Proteins in Developing Biopharmaceuticals*. Elsevier; 2015.
7. Kay LE. NMR studies of protein structure and dynamics. *J Magn Reson*. 2005; 173:193–207. DOI: 10.1016/j.jmr.2004.11.021 [PubMed: 15780912]
8. Ben-Nissan G, et al. Rapid characterization of secreted recombinant proteins by native mass spectrometry. *Commun Biol*. 2018; 1:213.doi: 10.1038/s42003-018-0231-3 [PubMed: 30534605]
9. Cveticanin J, et al. Estimating interprotein pairwise interaction energies in cell lysates from a single native mass spectrum. *Anal Chem*. 2018; 90:10090–10094. DOI: 10.1021/acs.analchem.8b02349 [PubMed: 30106564]
10. Gan J, et al. Native mass spectrometry of recombinant proteins from crude cell lysates. *Anal Chem*. 2017; 89:4398–4404. DOI: 10.1021/acs.analchem.7b00398 [PubMed: 28345863]
11. Ghaemmaghami S, Oas TG. Quantitative protein stability measurement in vivo. *Nat Struct Biol*. 2001; 8:879–882. DOI: 10.1038/nsb1001-879 [PubMed: 11573094]
12. Larsen MR, Trelle MB, Thingholm TE, Jensen ON. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques*. 2006; 40:790–798. DOI: 10.2144/000112201 [PubMed: 16774123]
13. Eyers, CE, Gaskel, SJ. *Mass spectrometry to identify posttranslational function and types of posttranslational modifications*. John Wiley & Sons; 2008.
14. Christensen DG, et al. Mechanisms, Detection, and Relevance of Protein Acetylation in Prokaryotes. *MBio*. 2019; 10doi: 10.1128/mBio.02708-18
15. St-Denis N, Gingras AC. Mass spectrometric tools for systematic analysis of protein phosphorylation. *Prog Mol Biol Transl Sci*. 2012; 106:3–32. DOI: 10.1016/B978-0-12-396456-4.00014-6 [PubMed: 22340712]
16. Zhou Q, Qiu H. The Mechanistic Impact of N-Glycosylation on Stability, Pharmacokinetics, and Immunogenicity of Therapeutic Proteins. *J Pharm Sci*. 2019; 108:1366–1377. DOI: 10.1016/j.xphs.2018.11.029 [PubMed: 30471292]
17. Hernandez H, Robinson CV. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat Protoc*. 2007; 2:715–726. DOI: 10.1038/nprot.2007.73 [PubMed: 17406634]
18. Wu L, Han DK. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Rev Proteomics*. 2006; 3:611–619. DOI: 10.1586/14789450.3.6.611 [PubMed: 17181475]
19. Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJ. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat Methods*. 2012; 9:1084–1086. DOI: 10.1038/nmeth.2208 [PubMed: 23064518]
20. Chernushevich IV, Thomson BA. Collisional cooling of large ions in electrospray mass spectrometry. *Anal Chem*. 2004; 76:1754–1760. DOI: 10.1021/ac035406j [PubMed: 15018579]
21. Giles K, et al. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom*. 2004; 18:2401–2414. DOI: 10.1002/rcm.1641 [PubMed: 15386629]
22. Liu H, et al. In vitro and in vivo modifications of recombinant and human IgG antibodies. *MABs*. 2014; 6:1145–1154. DOI: 10.4161/mabs.29883 [PubMed: 25517300]
23. Qiu H, et al. Engineering an anti-CD52 antibody for enhanced deamidation stability. *MABs*. 2019; doi: 10.1080/19420862.2019.1631117
24. Jung SY, et al. Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins. *Anal Chem*. 2008; 80:1721–1729. DOI: 10.1021/ac7021025 [PubMed: 18247584]
25. Afjehi-Sadat L, Garcia BA. Comprehending dynamic protein methylation with mass spectrometry. *Curr Opin Chem Biol*. 2013; 17:12–19. DOI: 10.1016/j.cbpa.2012.12.023 [PubMed: 23333572]
26. Brown CW, et al. Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. *BMC Genomics*. 2017; 18:301.doi: 10.1186/s12864-017-3676-8 [PubMed: 28412930]
27. Raftery MJ. Determination of oxidative protein modifications using mass spectrometry. *Redox Rep*. 2014; 19:140–147. DOI: 10.1179/1351000214Y.0000000089 [PubMed: 24702029]

28. Wingfield PT. N-Terminal Methionine Processing. *Curr Protoc Protein Sci.* 2017; 88doi: 10.1002/cpps.29
29. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* 2014; 5:172.doi: 10.3389/fmicb.2014.00172 [PubMed: 24860555]
30. Dumont J, Euwart D, Mei B, Estes S, Kshirsagar R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol.* 2016; 36:1110–1122. DOI: 10.3109/07388551.2015.1084266 [PubMed: 26383226]
31. Kebarle P, Verkerk UH. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev.* 2009; 28:898–917. DOI: 10.1002/mas.20247 [PubMed: 19551695]
32. Cech NB, Enke CG. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev.* 2001; 20:362–387. DOI: 10.1002/mas.10008 [PubMed: 11997944]
33. Menetret JF, et al. Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol Cell.* 2007; 28:1083–1092. DOI: 10.1016/j.molcel.2007.10.034 [PubMed: 18158904]
34. Flick TG, Cassou CA, Chang TM, Williams ER. Solution additives that desalt protein ions in native mass spectrometry. *Anal Chem.* 2012; 84:7511–7517. DOI: 10.1021/ac301629s [PubMed: 22881839]
35. Hashimoto K, Panchenko AR. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci U S A.* 2010; 107:20352–20357. DOI: 10.1073/pnas.1012999107 [PubMed: 21048085]
36. Dixit SM, Polasky DA, Ruotolo BT. Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Curr Opin Chem Biol.* 2018; 42:93–100. DOI: 10.1016/j.cbpa.2017.11.010 [PubMed: 29207278]
37. Ruotolo BT, Benesch JLP, Sandercock AM, Hyung S-J, Robinson CV. Ion mobility–mass spectrometry analysis of large protein complexes. *Nature Protocols.* 2008; 3:1139.doi: 10.1038/nprot.2008.78 [PubMed: 18600219]
38. Toby TK, et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nature Protocols.* 2019; 14:119–152. DOI: 10.1038/s41596-018-0085-7 [PubMed: 30518910]
39. van de Waterbeemd M, et al. Dissecting ribosomal particles throughout the kingdoms of life using advanced hybrid mass spectrometry methods. *Nature Communications.* 2018; 9doi: 10.1038/s41467-018-04853-x
40. Kaltashov IA, Mohimen A. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal Chem.* 2005; 77:5370–5379. DOI: 10.1021/ac050511+ [PubMed: 16097782]
41. Sokolovski M, et al. Measuring inter-protein pairwise interaction energies from a single native mass spectrum by double-mutant cycle analysis. *Nat Commun.* 2017; 8doi: 10.1038/s41467-017-00285-1
42. Chorev DS, et al. Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science.* 2018; 362:829–834. DOI: 10.1126/science.aau0976 [PubMed: 30442809]
43. Ben-Nissan G, et al. Triple-stage mass spectrometry unravels the heterogeneity of an endogenous protein complex. *Anal Chem.* 2017; 89:4708–4715. DOI: 10.1021/acs.analchem.7b00518 [PubMed: 28345864]
44. Allison TM, et al. Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nat Commun.* 2015; 6doi: 10.1038/ncomms9551
45. Marzahn MR, et al. Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *EMBO J.* 2016; 35:1254–1275. DOI: 10.15252/embj.201593169 [PubMed: 27220849]
46. Xing G, Zhang J, Chen Y, Zhao Y. Identification of four novel types of in vitro protein modifications. *J Proteome Res.* 2008; 7:4603–4608. DOI: 10.1021/pr800456q [PubMed: 18767873]
47. Francis GL. Albumin and mammalian cell culture: implications for biotechnology applications. *Cytotechnology.* 2010; 62:1–16. DOI: 10.1007/s10616-010-9263-3 [PubMed: 20373019]

48. Darfler FJ. A protein-free medium for the growth of hybridomas and other cells of the immune system. *In Vitro Cell Dev Biol.* 1990; 26:769–778. [PubMed: 2118498]
49. Valdés R, González M, Geada D, Fernández E. Assessment of a protein-free medium performance in different cell culture vessels using mouse hybridomas to produce monoclonal antibodies. *Pharmaceut Anal Acta.* 2012; 3
50. Kirshenbaum N, Michaelevski I, Sharon M. Analyzing large protein complexes by structural mass spectrometry. *J Vis Exp.* 2010; doi: 10.3791/1954
51. Lossl P, Snijder J, Heck AJ. Boundaries of mass resolution in native mass spectrometry. *J Am Soc Mass Spectrom.* 2014; 25:906–917. DOI: 10.1007/s13361-014-0874-3 [PubMed: 24700121]
52. Liu H, May K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs.* 2012; 4:17–23. DOI: 10.4161/mabs.4.1.18347 [PubMed: 22327427]
53. Maverakis E, et al. Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *J Autoimmun.* 2015; 57:1–13. DOI: 10.1016/j.jaut.2014.12.002 [PubMed: 25578468]
54. Manneberg M, Friedlein A, Kurth H, Lahm HW, Fountoulakis M. Structural analysis and localization of the carbohydrate moieties of a soluble human interferon gamma receptor produced in baculovirus-infected insect cells. *Protein Sci.* 1994; 3:30–38. DOI: 10.1002/pro.5560030105 [PubMed: 8142896]
55. Morelle W, Michalski JC. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc.* 2007; 2:1585–1602. DOI: 10.1038/nprot.2007.227 [PubMed: 17585300]
56. Lorence, A. Recombinant gene expression. Humana Press; 2012.
57. Netzer R, et al. Ultrahigh specificity in a network of computationally designed protein-interaction pairs. *Nat Commun.* 2018; 9doi: 10.1038/s41467-018-07722-9
58. Walsh G. Biopharmaceutical benchmarks 2018. *Nat Biotechnol.* 2018; 36:1136–1145. DOI: 10.1038/nbt.4305 [PubMed: 30520869]
59. Tiller KE, Tessier PM. Advances in antibody design. *Annu Rev Biomed Eng.* 2015; 17:191–216. DOI: 10.1146/annurev-bioeng-071114-040733 [PubMed: 26274600]

Box 1

Types of common glycosylations and their masses

Core Mannose-F	G0F - GlcNac	G0F	G1F	G2F
1056.39	1259.47	1462.55	1624.60	1786.65
				
Core Mannose	G0-GlcNac	G0	G1	G2
910.33	1113.41	1316.49	1478.54	1640.59
				
Fucose	Mannose	GlcNac	Sodium	Water
146.06	162.06	203.01	22.99	18.02
			Na ⁺	H ₂ O

40. Using the quadrupole mass filter, isolate the most dominant glycan ion in the spectrum. Usually on the Exactive Plus Orbitrap instrument we set a narrow isolation window of about 4 m/z units. Record the spectrum, while adjusting the m/z range to ~ 200 at the low end of the scale, and up to slightly higher than the m/z value of the isolated glycan ion, at the high end of the scale.
41. Gradually increase the HCD energy in 5 V increments, and record spectra at each step. Under these experimental conditions, fragmentation of singly charged glycans [M+H⁺] will occur at around 15 V, whereas the sodiated glycans [M+Na⁺], which are more stable, will fragment at around 50 V.
42. Repeat steps 40 - 41 for the other glycan species in the spectrum.
43. Carefully inspect the fragmentation spectra, and identify the parent ion on the right end of the m/z scale. Starting from this position, systematically calculate the mass differences between the monoisotopic m/z values of every two adjacent ions. In some cases, mass differences between non-adjacent peaks should also be considered (Figure 11D and BOX 1). Since glycans typically bear a single charge, the m/z shift between two fragments will reflect the mass difference between them. In order to identify the different moieties that were cleaved during fragmentation that give rise to peaks with lower m/z values, refer to BOX 1. Typically, the majority of the in the spectrum can be explained using this method. Note that glycans, similar to peptides, may lose water molecules during the fragmentation process, therefore species with a mass difference of 18 Da are also frequently observed. For example, in Figure 11D we start with the parent ion (1485.5 m/z), and calculate the mass difference between it and the adjacent peak (1467.5 m/z). The calculated mass

difference between these two peaks equals 18 Da, reflecting a neutral loss of a water molecule. The next meaningful peak in the spectrum (1339.5 m/z) results from the loss of a fucose moiety (146 Da) and the next peak (1282.5 m/z) reflects the loss of a GlcNac moiety (203 Da).

44. Accordingly, assign the composition of the parent glycan and all the measured fragments⁵⁵ (Fig. 11E).

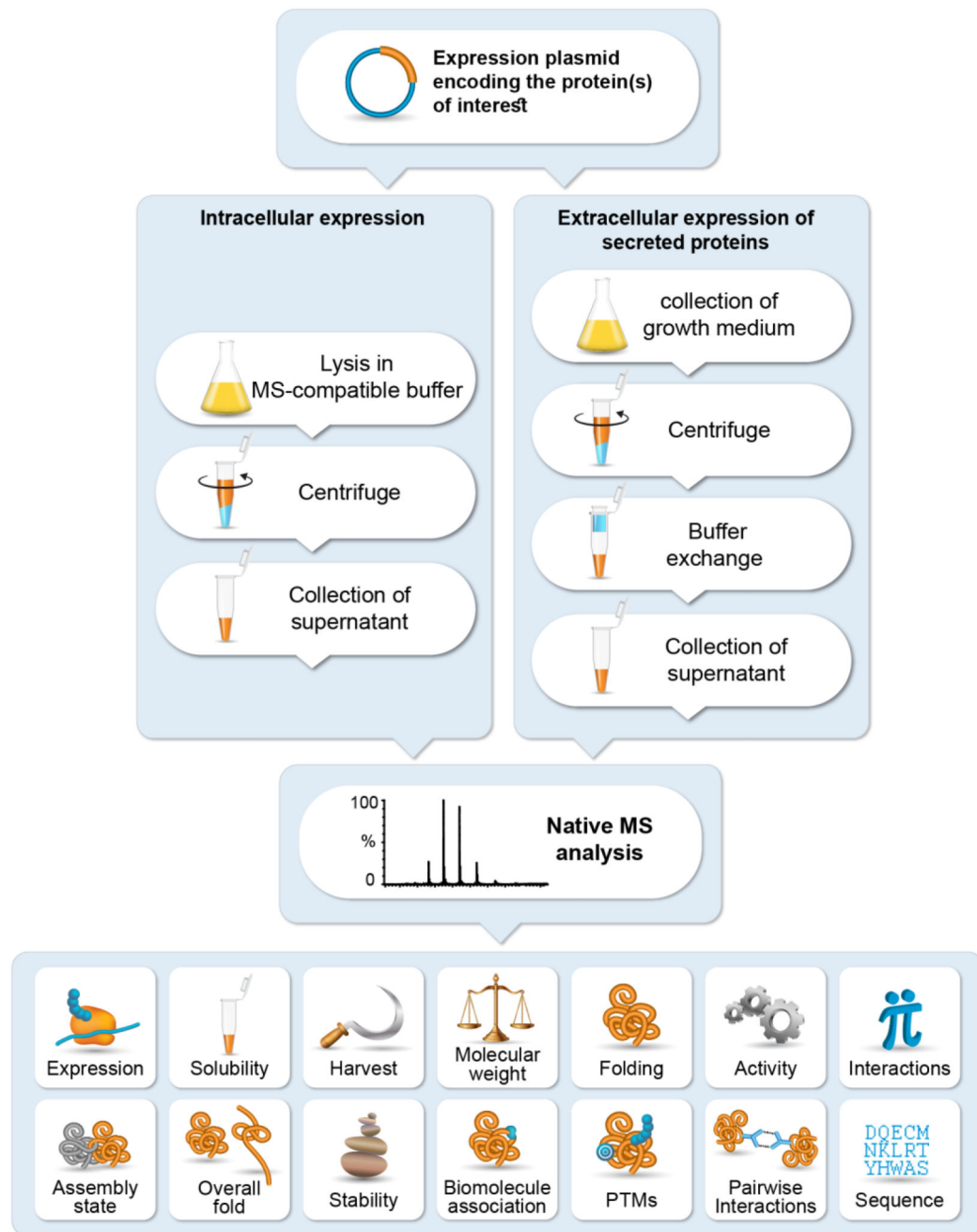


Figure 1. An overview of the direct-MS workflow for analysis of recombinant proteins from crude samples.

Initially, expression of the protein of interest is induced. Harvesting is then followed by sample preparation for direct-MS analysis. In cases of intracellular expression in bacterial cells, the cellular lysate is cleared out by centrifugation, and the supernatant is directly used for MS analyses. Alternatively, when protein expression is performed in eukaryotic secretion systems, the growth medium is collected, cells and insoluble debris are cleared by centrifugation and buffer exchanged into a MS-compatible solution; the supernatant is then collected for MS acquisition. The high resolution afforded by mass measurement of the

intact protein enables immediate assessment of the expressibility, identity, solubility, assembly and folding state, overall structure, and stability of the protein produced. In addition, the method provides immediate information on the optimal harvesting time of the protein, sequence variations, binding of biomolecules, post-translational modifications, associations with other proteins, and activity.

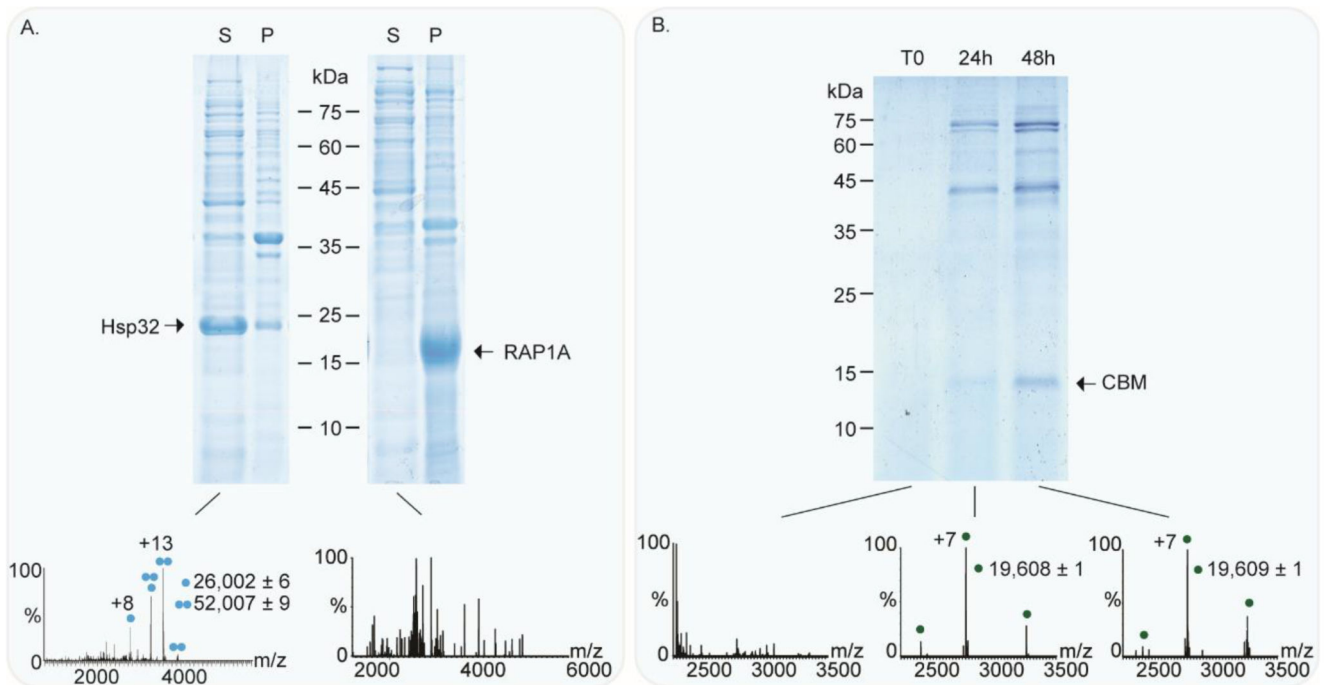


Figure 2. Analysis of proteins in crude samples from prokaryotic and eukaryotic expression systems.

(A) Soluble (S) and insoluble (P) fractions from bacterial cultures overexpressing two proteins. In the case of Hsp32, the protein accumulates in the soluble fraction and is easily measured by direct MS analysis, using the method described in this protocol (monomeric and dimeric Hsp32, are designated by one or two cyan circles, respectively). However, when the protein is insoluble and accumulates in the pellet within inclusion bodies, as in the case of RAP1A, only background signals of endogenous bacterial proteins are detected in the soluble fraction. (B) The absolute expression level of a protein in the crude sample is not a critical limitation for direct MS measurements. As in the case of CBM3a, which is secreted from the yeast *P. pastoris*, the low amounts of the protein that accumulated after 24 h are sufficient to obtain good measurements. This is achieved due to the fact that the crude growth medium displays a relatively low complexity background, which does not mask the signals of CBM3a. Charge series corresponding to CBM3a are labeled by green circles.

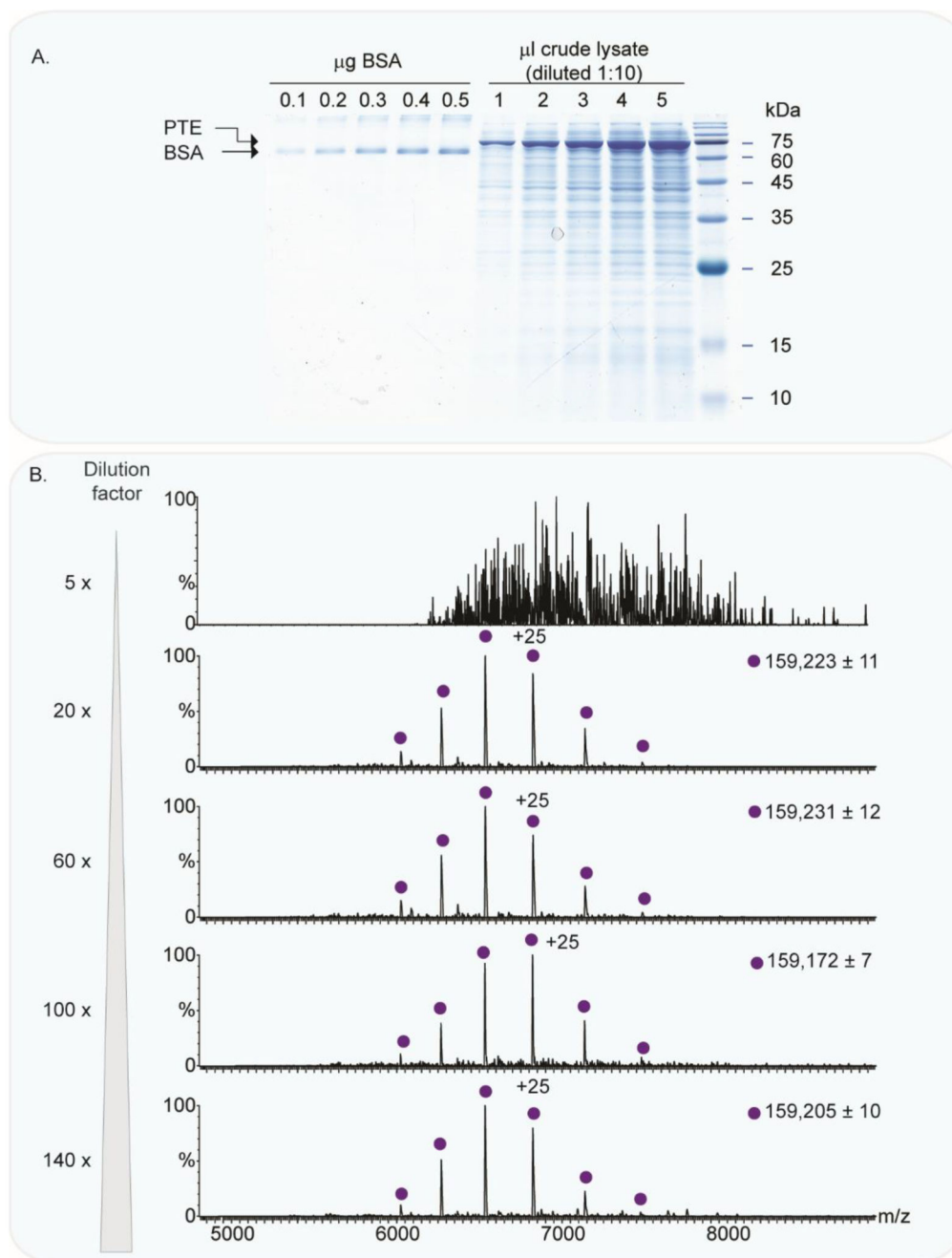


Figure 3. Crude samples may require dilution for efficient MS measurements.

(A) The recombinant dimeric protein phosphotriesterase (PTE), fused to the maltose-binding protein, was overexpressed in 10 ml of bacterial culture and analyzed by direct MS, according to the protocol described here. To evaluate the level of PTE in the lysate, known amounts of BSA were loaded onto SDS-PAGE gels. Following staining, the intensity of PTE was compared to that of BSA; accordingly, we calculated that the concentration of PTE in the lysate is 5.7 mg/ml, equaling 36 µM. (B) The high concentration of the lysate hindered

direct-MS measurements. Only after a series of dilutions could the clearly resolved charge state series be detected. Charge series corresponding to PTE are labeled by purple circles.

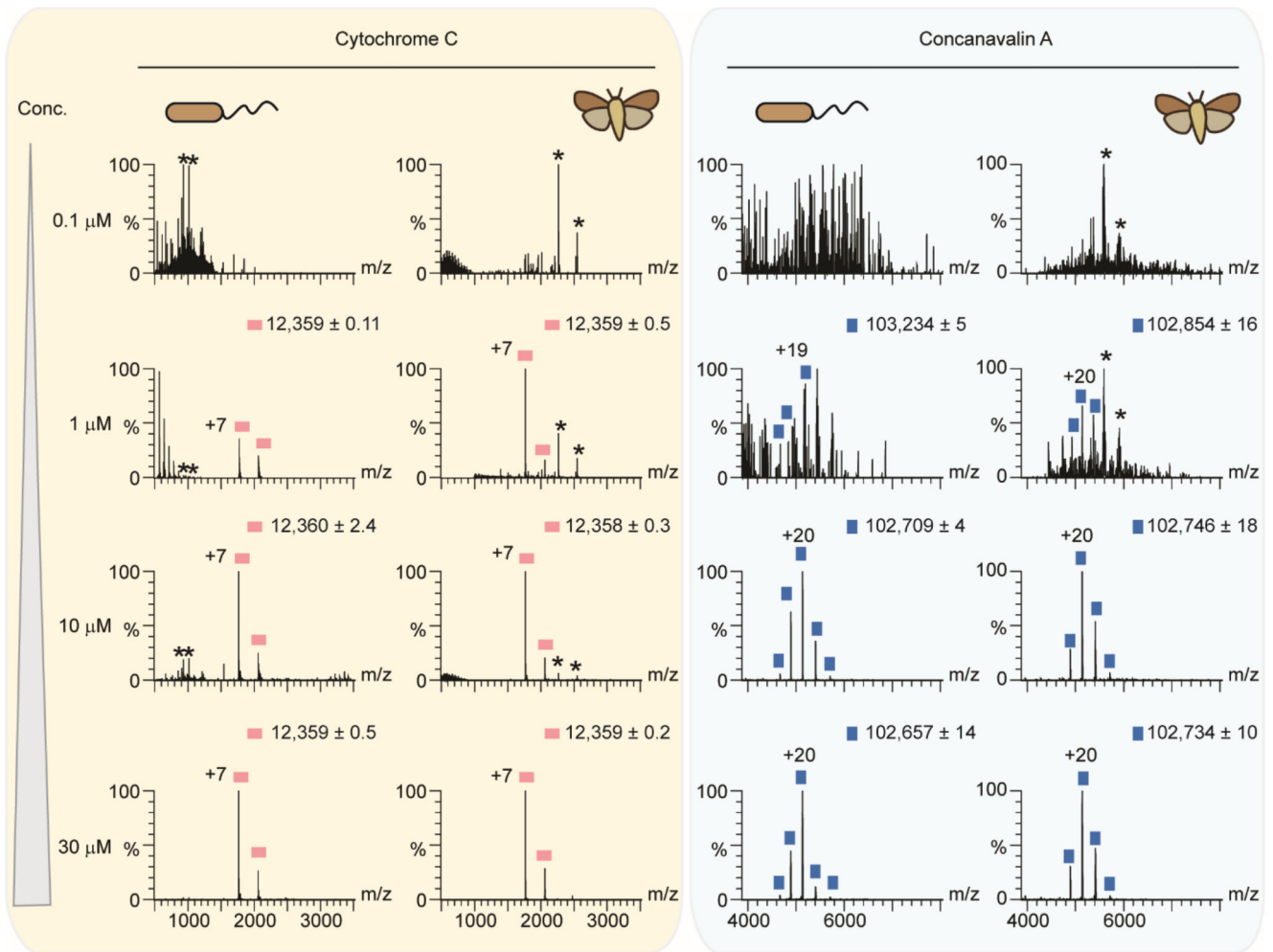


Figure 4. The detection limit of a protein in crude samples is around 1 μM . Different concentrations of cytochrome C and concanavalin-A, ranging from 0.1 - 30 μM , were externally added to non-expressing *E. coli* lysate or insect cell growth medium. Proteins at a concentration as low as 1 μM could be detected in these crude samples. Asterisks denote the most dominant background peaks. Charge series corresponding to cytochromeC and concanavalin-A are labeled by pink and blue squares, respectively.

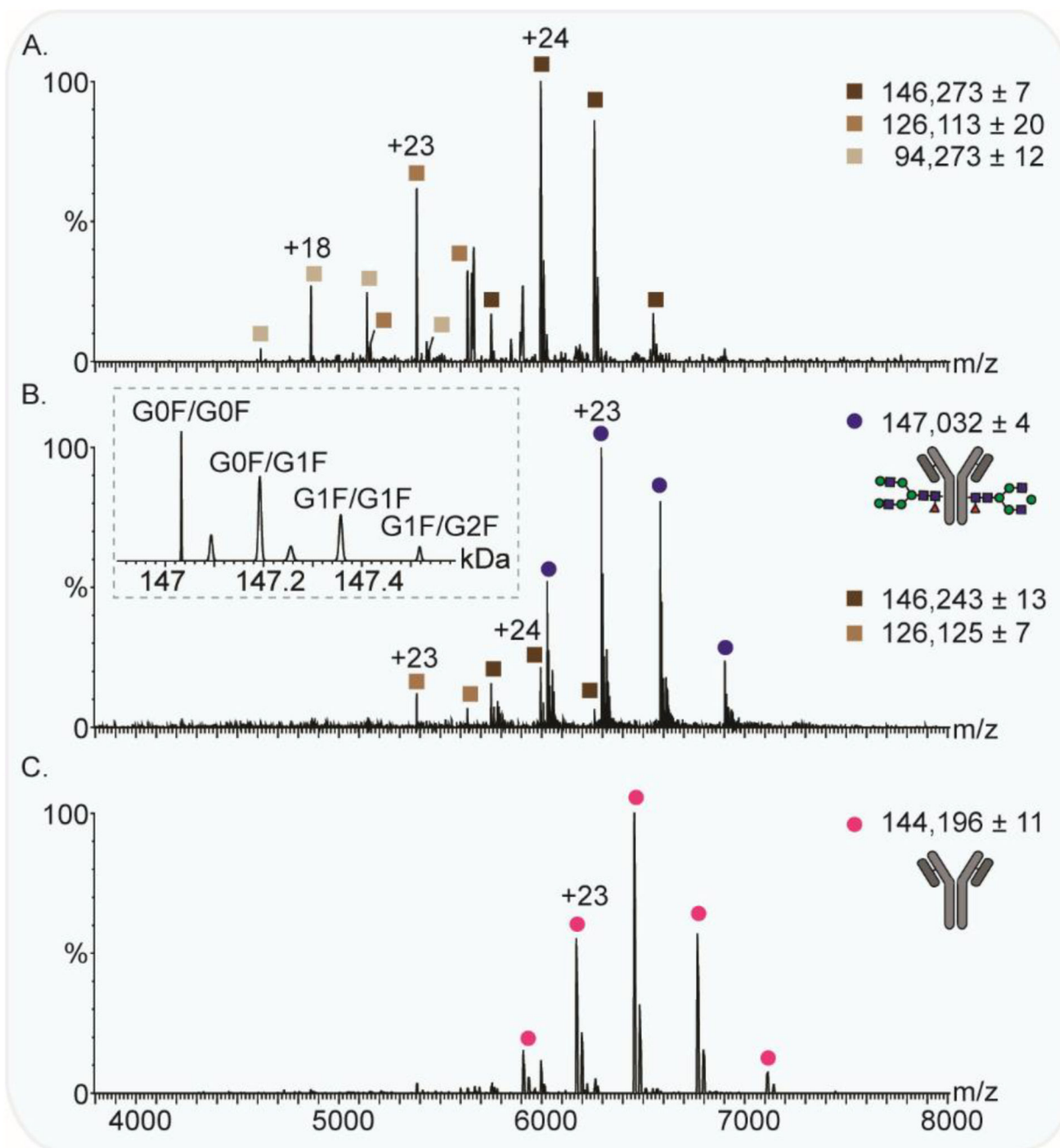


Figure 5. Direct-MS analysis of a secreted antibody unravels its multiple co-existing forms.

(A) The growth medium of confluent, suspension-grown, non-expressing HEK293F cells contains different proteins, which originate from spillage of the cellular contents of dead cells into the growth medium (labeled by brown squares). The 146 kDa protein corresponds to the endogenously expressed tetrameric LDH. (B) When a secreted antibody was expressed in these cells, it became the major charge state series in the spectrum (blue circles). However, without pre-analyzing the control medium, the charge state series of LDH could have easily been mistaken for the specific signal of a proteoform of the antibody, due

to very close similarities in total mass. The accurate and highly resolved measurement of the antibody enabled us to assign different backbone modifications, and demonstrate the relative abundance of different glycosylations on the antibody. Inset shows the deconvoluted spectrum of the antibody, with the differentially modified glycosylations. (C) Following deglycosylation by PNGase-F in the growth medium, the major charge state series in the spectrum corresponds to the non-glycosylated antibody (magenta circles).

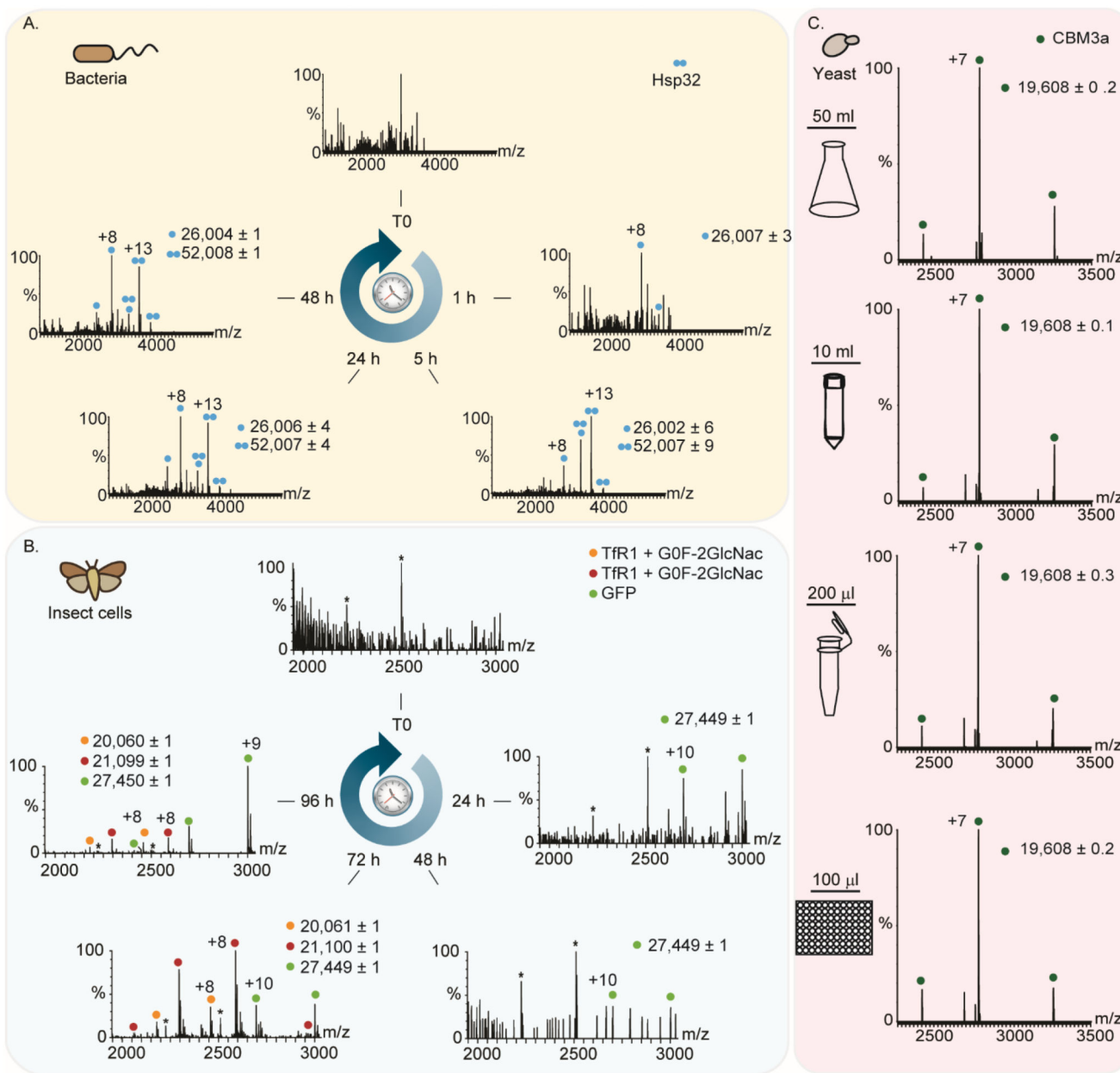


Figure 6. The effect of induction time and culture volume on the analysis of recombinant proteins in crude samples

(A) Lysates from *E. coli* cells expressing the Hsp32 homodimer, were analyzed at different time points over a course of 48 h, following induction with IPTG. By 1 hr after the addition of IPTG, a low charge state series of the Hsp32 monomer could be detected (cyan circles). At this time point, the assembled dimer was not observed, possibly due to its low expression level. After 5 h of induction (the typical harvest time of Hsp32), the major charge state observed was that of the assembled dimer (designated by two cyan circles). Longer induction times (24- and 48 h) did not result in any observed benefit, as exemplified both by similar overall signal-to-noise ratios between background signals, and a slight tendency

towards accumulation of monomers over the longer induction times. (B) Growth medium from insect cells expressing the TfR1 protein was examined at various time points over a course of 96 h post- infection. Different proteoforms of TfR1 accumulated to detectable levels after 72 h (red and orange circles). GFP, used as a reporter for infection efficiency, leaked into the growth medium from dead cells in the culture, and was detected as early as 24 h after infection (light green circles). After 96 h, GFP became the major charge state series in the spectrum, due to increasing levels of cell death. (C) CBM3a was expressed in *P. pastoris* cells for 24 h in different culture volumes, starting from 50 ml of culture in a 250 ml flask, down to 100 μ L cultures in 96-well plates. In all the different growth volumes, the CBM3a charge series could be clearly detected (dark green circles).

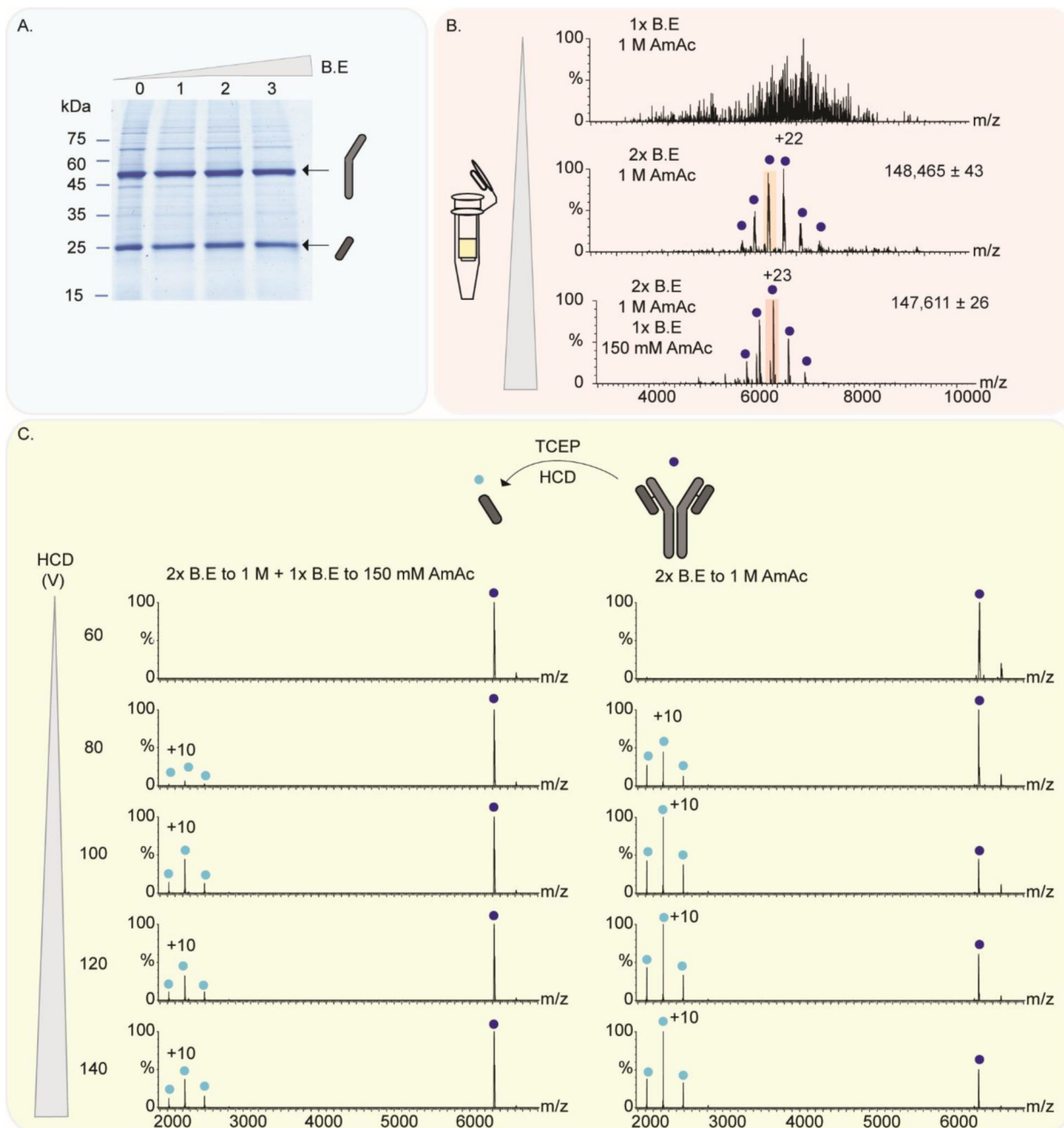


Figure 7. Effect of ionic strength on the spectra of a recombinant antibody.

Growth medium from HEK293F cells secreting the designed anti-human VEGF antibody G6des13, was buffer exchanged twice into 1M ammonium acetate, and a third time into 150 mM ammonium acetate. (A) 15 μ l of each sample were separated by SDS-PAGE and stained with coomassie staining, demonstrating that the number of buffer exchange cycles did not significantly affect the relative ratio between the recombinant antibody subunits and the other proteins in the sample. In this example (B), a single round of buffer exchange was not sufficient to remove all the contaminants from the growth medium, such that the antibody

did not resolve into a clear spectrum. Following a second step of buffer exchange into 1 M ammonium acetate, a nicely resolved spectrum of the antibody was detected. A third round of buffer exchange into 150 mM ammonium acetate further improved the results, yielding measurements of higher ion intensities and narrower peak widths. (C) Tandem MS experiments, performed at different collision energies following antibody reduction with TCEP, demonstrated that high ionic strength affects protein-protein interactions and weakens the association between the antibody's light and heavy chains. Blue and light blue labels designated charge states of the intact antibody ($147,676 \pm 9.0$ Da) and the stripped light chain ($23,291 \pm 0.6$ Da), respectively.

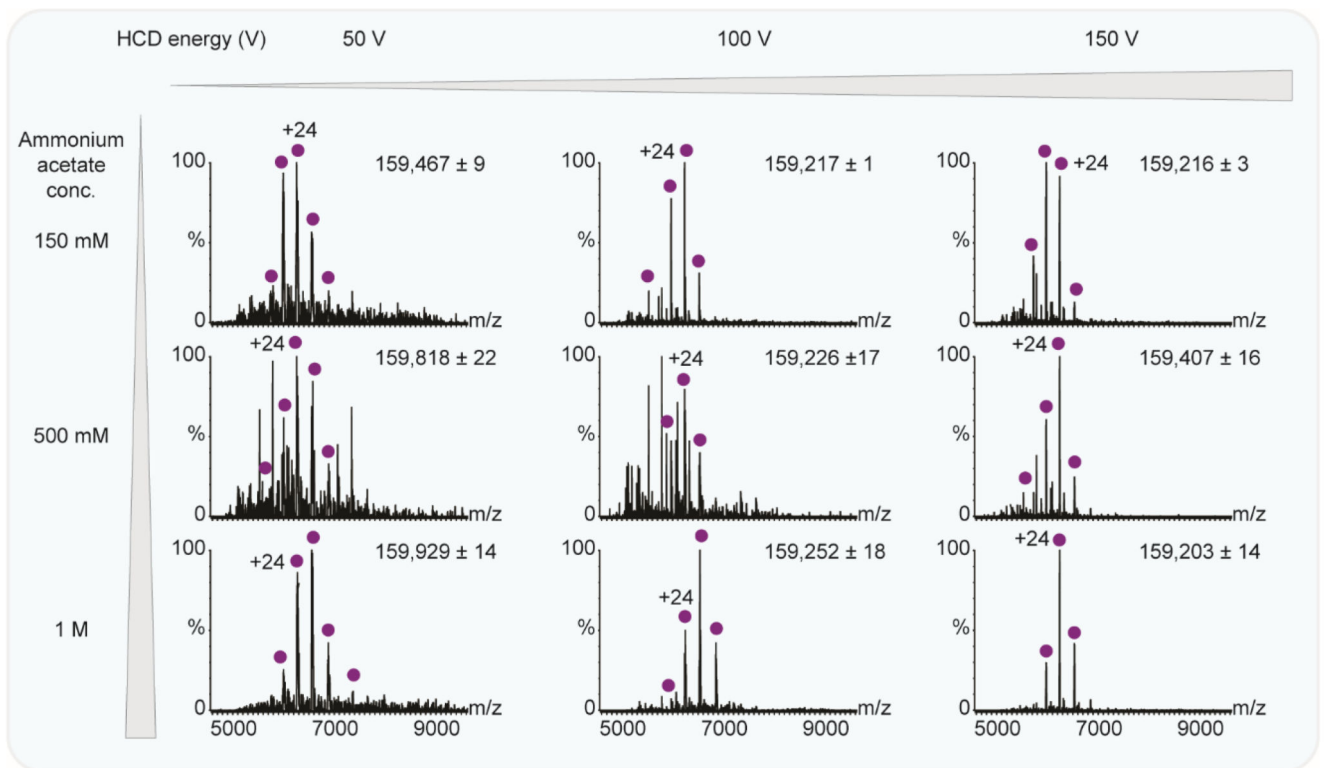


Figure 8. Effects of ammonium acetate concentration and collision energy on the spectra of the protein phosphotriesterase.

The dimeric protein phosphotriesterase (PTE), fused to the maltose-binding protein, was overexpressed in *E. coli*, and lysed in different concentrations of ammonium acetate. Lysates thus prepared were all amenable to direct-MS analysis; however, high ammonium acetate concentrations and elevated HCD energy increased the signal-to-noise ratio, and reduced both peak width and the level of background signals. All measurements were performed after a 50-fold dilution of the lysis solution. Charge states corresponding to PTE are labelled by purple circles.

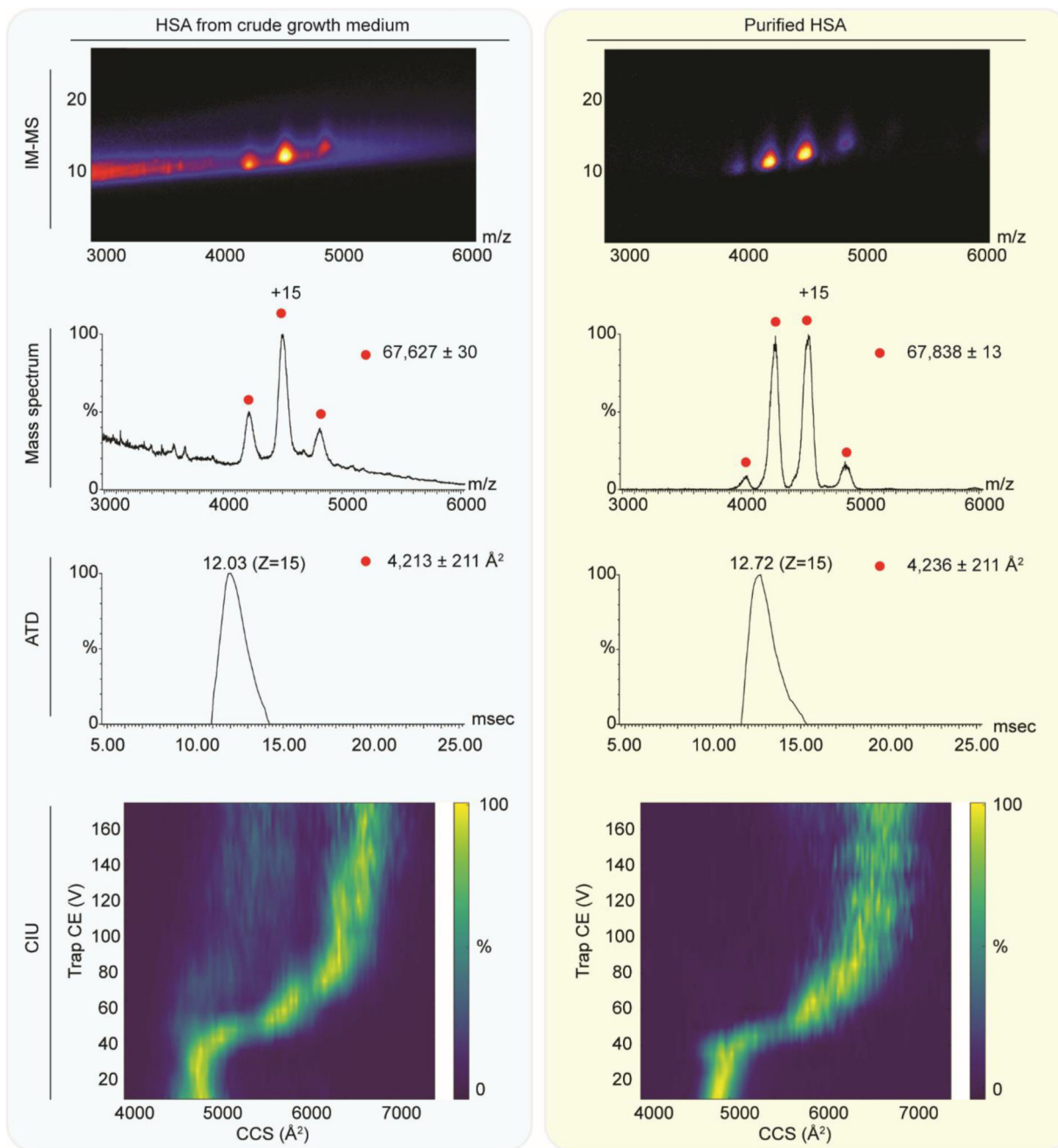


Figure 9. Ion mobility measurements and representative mass spectra from purified and crude samples of human serum albumin (HSA).

The IM-MS plots, mass spectra, arrival time distribution, and CIU profiles of the 16^+ charge state of HSA secreted from a *P. pastoris* culture (left panel), and those measured from the purified protein sample (right panel) are highly similar. All samples were measured on a Synapt G2 instrument, modified for the measurement of high masses. Charge states corresponding to HSA are labeled by red circles.

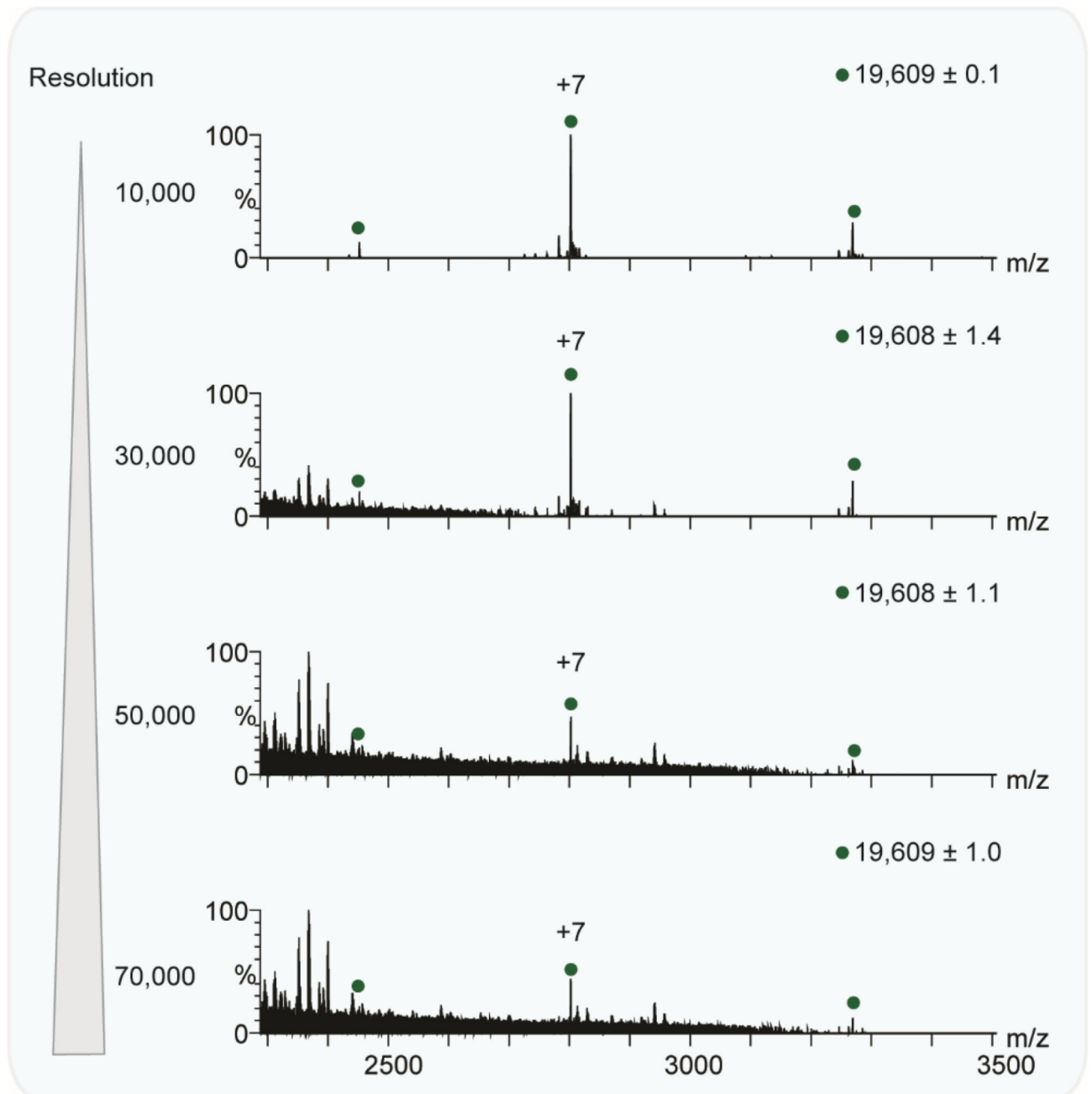


Figure 10. Effect of resolution on the measured spectra of CBM3a.

The growth medium of *P. pastoris* cells secreting the CBM3a protein was measured on a Q Exactive plus EMR Orbitrap mass spectrometer, at different resolutions, ranging from 10,000 to 70,000. At the 10,000 resolution, a well-resolved CBM3a charge state distribution was detected (green circles); however, at values of 30,000 and above, the signal-to-noise ratio decreased, and background peaks in the spectrum became dominant.

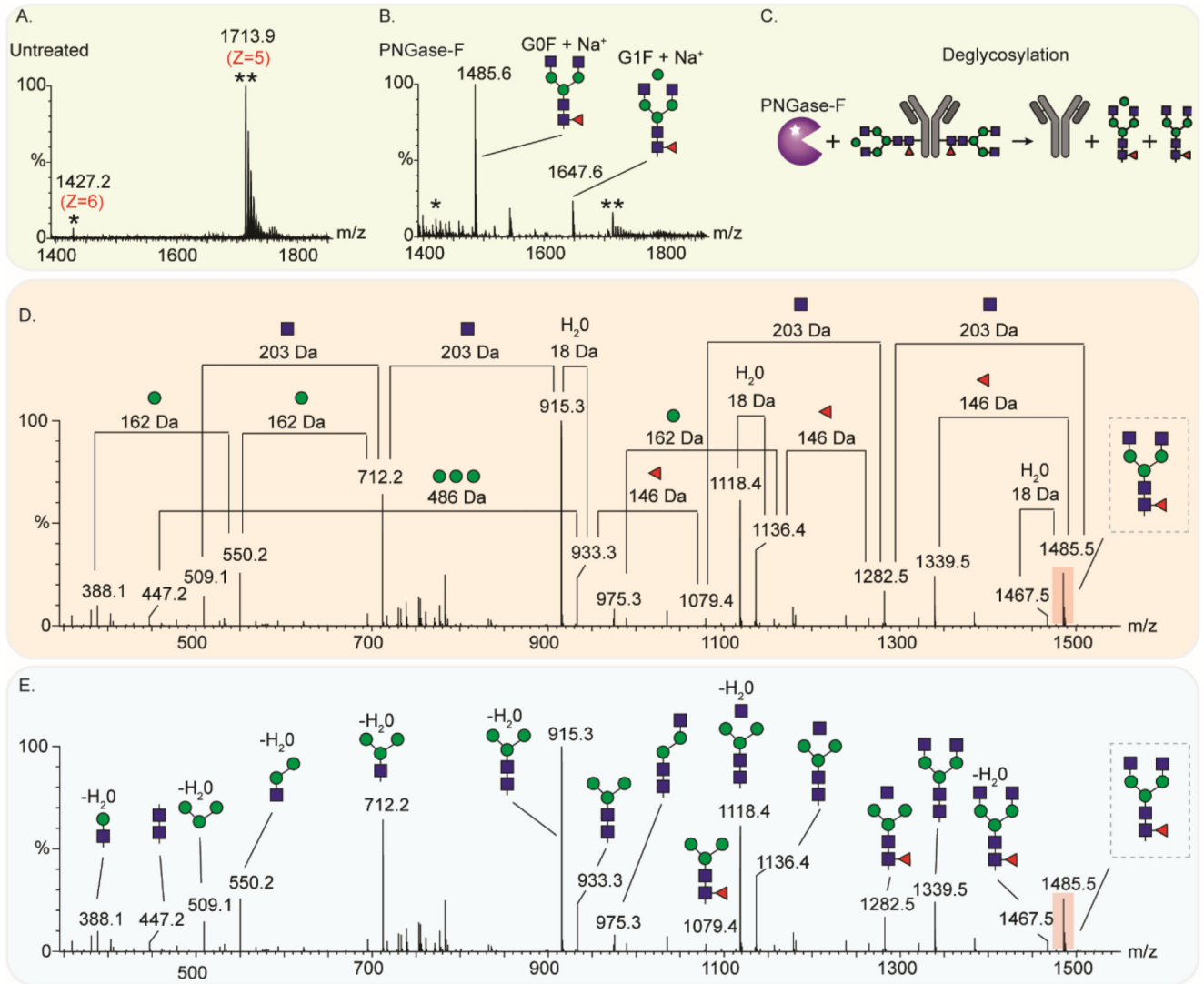


Figure 11. Detailed analysis of antibody glycosylation

The growth medium from HEK293F cells expressing a designed form of the anti-lysozyme antibody was measured before (A) and after (B) PNGase-F treatment. (A) Examination of the low m/z region showed that the growth medium acquired prior to deglycosylation contains two major peptides carrying 6+ and 5+ charges (m/z values and charges are labeled). (B) The spectrum after PNGase-F treatment revealed sodiated G0F (1485.6 Da) and G1F (1647.6 Da) glycans. (C) PNGase-F reaction scheme, showing that following incubation of the glycosylated antibody with this enzyme, the glycans are released into the growth medium, and a fully deglycosylated antibody is generated. (D-E) Tandem MS and ion fragmentation of the sodiated G0F. The composition of G0F can be assigned by calculating the mass difference of the successive (and non-successive) peaks (masses of the different monosaccharides and relevant labels were taken from BOX 1).

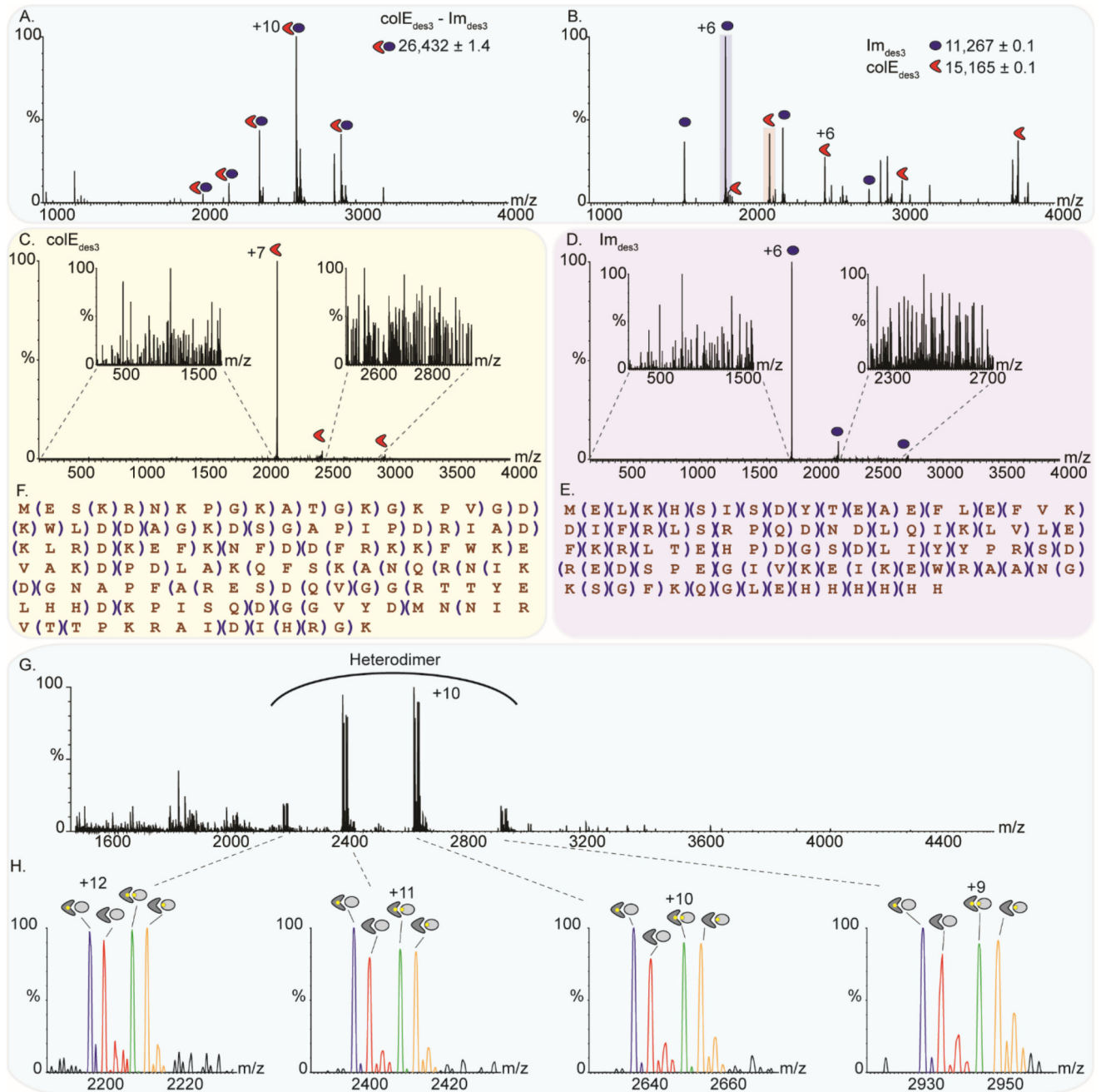


Figure 12. Pairwise interactions strength and sequence mapping may be performed in crude samples.

Crude lysates offer a platform for multilevel analysis of protein complexes. (A) The computationally designed heterodimer composed of the $\text{colE}_{\text{des3}}$ and Im_{des3} proteins was coexpressed in *E. coli*. Direct MS analysis indicated that the majority of the recombinant proteins in the lysate form heterodimers. (B) Following activation in the front end of the instrument, the dimer dissociated into its constituent building blocks. A single charge state of both the colE (C) and Im (D) proteins was isolated in the quadrupole, and further activated by a combination of ECD and HCD energies. The generated fragments were

measured, and further subjected to top-down proteomic analysis, resulting in a sequence coverage of 62% and 84% for the colE and Im proteins, respectively. (G-H) The native-MS double mutant cycle method then took the analysis a step further, determining the strengths of intermolecular pairwise interactions between residues at the interfaces of the interacting proteins. To this end, we overexpressed, in the same bacterial cells, two WT colE_{des3} and Im_{des3} proteins together with two of their mutants, in which two target residues were mutated to alanine (N83A in colE_{des3}, and N31A in Im_{des3}). The crude lysate contained all four different complexes comprising the WT and mutated proteins (G). We then calculated the pairwise interaction energies from a single high-resolution native mass spectrum directly from crude lysates, by measuring the intensities of the complexes formed by the two wild-type proteins (red peaks), the complex of each wild-type protein with a mutant protein (blue and orange peaks), and the complex of the two mutant proteins (green peaks). The coupling energy value obtained from the crude measurement (-0.14 ± 0.03 kcal / mol) was essentially similar to that obtained for the purified proteins (-0.02 ± 0.02 kcal / mol), and indicated not only that these residues do not contribute to the overall binding energy between the two proteins, but also that hydrogen bond strengths are not affected by the more crowded conditions in cell lysates. Measurements represent averages of four repeats. Error bars represent standard deviation. Measured masses of the four different complexes comprising the WT and mutated proteins are as follows: colE_{des3}(N83A)-Im_{des3}(WT) – $26,390 \pm 1.0$ Da (peaks labeled in blue), colE_{des3}(WT)-Im_{des3}(WT) – $26,432 \pm 1.5$ Da (peaks labeled in red), colE_{des3}(N83A)-Im_{des3}(WT) – $26,518 \pm 1.4$ Da (peaks labeled in green), colE_{des3}(N83A)-Im_{des3}(WT) – $26,561 \pm 1.4$ Da (peaks labeled in orange).

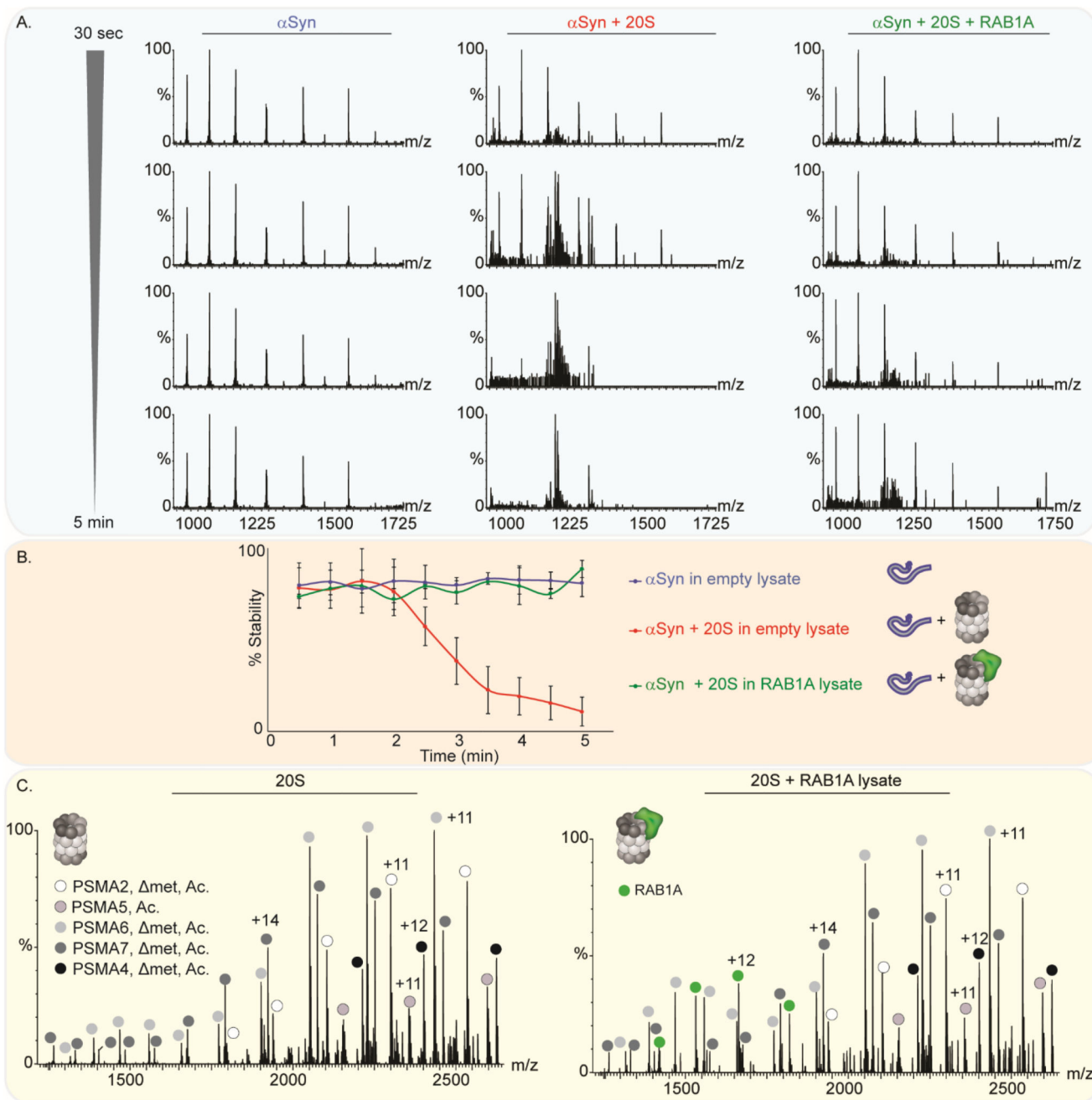


Figure 13. Determining RAB1A activity and interactions in crude lysates by native-MS.

Enzymatic activities can be probed in the context of crude lysates, as demonstrated for the 20S proteasome. (A) α -synuclein (α Syn), a substrate of the 20S proteasome, was spiked into a control bacterial lysate, not supplemented with protease inhibitors, and remained stable over a period of 5 minutes. When the 20S proteasome was added to the medium, the levels of α -synuclein decreased to background levels within 3.5 minutes. However, if the 20S proteasome was preincubated in a lysate expressing the putative 20S proteasome regulator RAB1A, α -synuclein levels remained stable, indicating that RAB1A reduced the activity of

the 20S proteasome. (B) Quantification of the relative levels of α -synuclein in the different lysates. In this experiment, intensities of all the peaks of α -synuclein were normalized to the highest peak in each spectrum, and averaged. The graph represents an average of five experiments. Error bars represent standard deviation. (C) The inhibitory effect of RAB1A on the 20S proteasome suggested direct interactions between the two. To probe this hypothesis, the 20S proteasome was preincubated with a control lysate, or a lysate from cells overexpressing RAB1A, and subjected to tandem MS analysis. Following isolation and acceleration of the charge state series corresponding to the 20S proteasome, different α -subunits dissociated from the complex (gray circles). However, in the lysate expressing the recombinant RAB1A, another charge state series was detected, corresponding in mass to the RAB1A protein (green labels). This finding demonstrates that prior to isolation, the 20S proteasome was physically bound to its regulator, RAB1A. The measured masses of the different proteins are as follows: PSMA2 missing the initial methionine (- met) and acetylated (Ac.) - $25,838 \pm 1.0$ Da, PSMA5, Ac. - $26,453 \pm 0.4$ Da, PSMA6, - met, Ac. $27,311 \pm 1.1$ Da, PSMA7, - met, Ac. $27,609 \pm 0.9$, PSMA4, - met, Ac. - $29,409 \pm 0.6$ Da, RAB1A - $20,547 \pm 1.8$ Da.

Table 1
Common post-translational modifications

Modification	Mass Shift (Da)	Chemical formula	Reference example
*C-terminal proline amidation	1.00	H	22
Disulfide bond formation	-2.01	H ₂	23
Methylation	14.01	CH ₃	24,25
Oxidation	16.02	OH	26,27
*Pyroglutamic acid formation from N-terminal glutamine	17.03	NH ₃	22
*Pyroglutamic acid formation from N-terminal glutamic acid	18.01	H ₂ O	22
Acetylation	42.01	C ₂ H ₃ O	15,26
Phosphorylation	79.96	PO ₃ H ₂	15,26
*C-terminal lysine removal	-128.08	C ₆ H ₁₄ N ₂ O ₂	22
*C-terminal lysine and glycine removal	-185.13	C ₈ H ₁₉ N ₃ O ₄	22
N-terminal methionine removal	-131.04	C ₅ H ₁₁ NO ₂ S	28

* Typical antibody modifications.