



ARTICLE

DOI: 10.1038/s42003-018-0231-3

OPEN

# Rapid characterization of secreted recombinant proteins by native mass spectrometry

Gili Ben-Nissan<sup>1</sup>, Shay Vimer <sup>1</sup>, Shira Warszawski<sup>1</sup>, Aliza Katz<sup>2</sup>, Meital Yona<sup>3</sup>, Tamar Unger<sup>3</sup>, Yoav Peleg<sup>3</sup>, David Morgenstern<sup>4</sup>, Hadas Cohen-Dvashi<sup>2</sup>, Ron Diskin <sup>2</sup>, Sarel J. Fleishman<sup>1</sup> & Michal Sharon<sup>1</sup>

Characterization of overexpressed proteins is essential for assessing their quality, and providing input for iterative redesign and optimization. This process is typically carried out following purification procedures that require pronounced cost of time and labor. Therefore, quality assessment of recombinant proteins with no prior purification offers a major advantage. Here, we report a native mass spectrometry method that enables characterization of overproduced proteins directly from culture media. Properties such as solubility, molecular weight, folding, assembly state, overall structure, post-translational modifications and binding to relevant biomolecules are immediately revealed. We show the applicability of the method for in-depth characterization of secreted recombinant proteins from eukaryotic systems such as yeast, insect, and human cells. This method, which can be readily extended to high-throughput analysis, considerably shortens the time gap between protein production and characterization, and is particularly suitable for characterizing engineered and mutated proteins, and optimizing yield and quality of overexpressed proteins.

<sup>1</sup>Department of Biomolecular Sciences, Weizmann Institute of Science, 7610001 Rehovot, Israel. <sup>2</sup>Department of Structural Biology, Weizmann Institute of Science, 7610001 Rehovot, Israel. <sup>3</sup>Israel Structural Proteomics Center, Weizmann Institute of Science, 7610001 Rehovot, Israel. <sup>4</sup>The De Botton protein Profiling Institute of the Nancy and Stephen Grand Israel national Center for Personalized Medicine, Weizmann Institute of Science, 7610001 Rehovot, Israel. These authors contributed equally: Gili Ben-Nissan, Shay Vimer. Correspondence and requests for materials should be addressed to M.S. (email: [michal.sharon@weizmann.ac.il](mailto:michal.sharon@weizmann.ac.il))

The ability to produce recombinant proteins has revolutionized biochemistry, enabling new experimental, therapeutic, industrial and commercial applications. For example, to date about 90% of industrial enzymes used for paper, leather, detergents, textiles, food, biofuels and animal feed are recombinant proteins<sup>1</sup>. Similarly, recombinant protein therapeutics have increased dramatically, both in number and in frequency<sup>2</sup>. Since 2011, more than 60 recombinant protein drugs have been approved by the FDA, including therapies for various cancers, autoimmune and autoinflammatory diseases, infectious diseases and genetic disorders<sup>3–5</sup>. However, generating proteins for such applications is a highly complex process, due to their high molecular weight, three-dimensional structures, heterogeneity, and dependence on production in living cells<sup>6</sup>.

A critical step in the development of recombinant proteins is characterization, especially when codon optimization and protein-engineering approaches are employed. Characterizing the generated protein variants is essential for assessing their quality, and providing input for iterative redesign and optimization<sup>7,8</sup>. Such characterization is especially relevant when selecting the ideal host system for protein expression, optimizing product yield, or employing high-throughput investigations of lead candidates. A preliminary characterization step reduces the quantity and expense of resources wasted on inadequate protein products. It also provides a means to ensure batch-to-batch consistency. However, the characterization step is usually carried out with purified proteins, with significant costs of time and labor invested in product purification. Therefore, quality assessment of overproduced proteins with no prior purification offers a great advantage.

Recently, we described a direct native mass spectrometry approach for rapid characterization of overexpressed proteins directly from crude bacterial cell lysates, overcoming the need for protein purification<sup>9</sup>. In spite of this progress and the fact that prokaryotic systems are often the first-choice host for initial screening of recombinant protein expression, many proteins fail to express and fold properly in this platform<sup>10</sup>. Alternatively, eukaryotic secretion systems provide higher capacity for proper protein folding, assembly and post-translational modifications, in comparison to bacterial hosts<sup>10,11</sup>. Therefore, here, we developed a direct mass spectrometry (direct-MS) approach for the analysis of recombinant proteins secreted from yeast, insect and mammalian cells. In particular, we expressed the scaffoldin cellulose-specific carbohydrate-binding module (CBM3a) from *C. thermocellum*<sup>12</sup>, and serum albumin in the *P. pastoris* yeast expression system, the ectodomain of *N. albigula* Transferrin receptor 1 (TfR1) in Sf9 insect cells and several antibodies in human embryonic kidney (HEK293) cell cultures.

We show that a single measurement of the crude culture medium provides in-depth structural characterization of the overproduced protein. Specifically, the high resolution afforded by the intact protein mass measurement enables immediate assessment of sequence variations, assembly state, folding conditions, and association of ligands with the generated proteins. The method, which is suitable for high-throughput studies, also enables informative time course measurements in real time that can be used to optimize expression levels and quality. Moreover, the method considerably shortens the development timeline by enabling the selection and ranking of lead candidates directly from the culture media, avoiding the burdens associated with purifying inadequate variants.

## Results

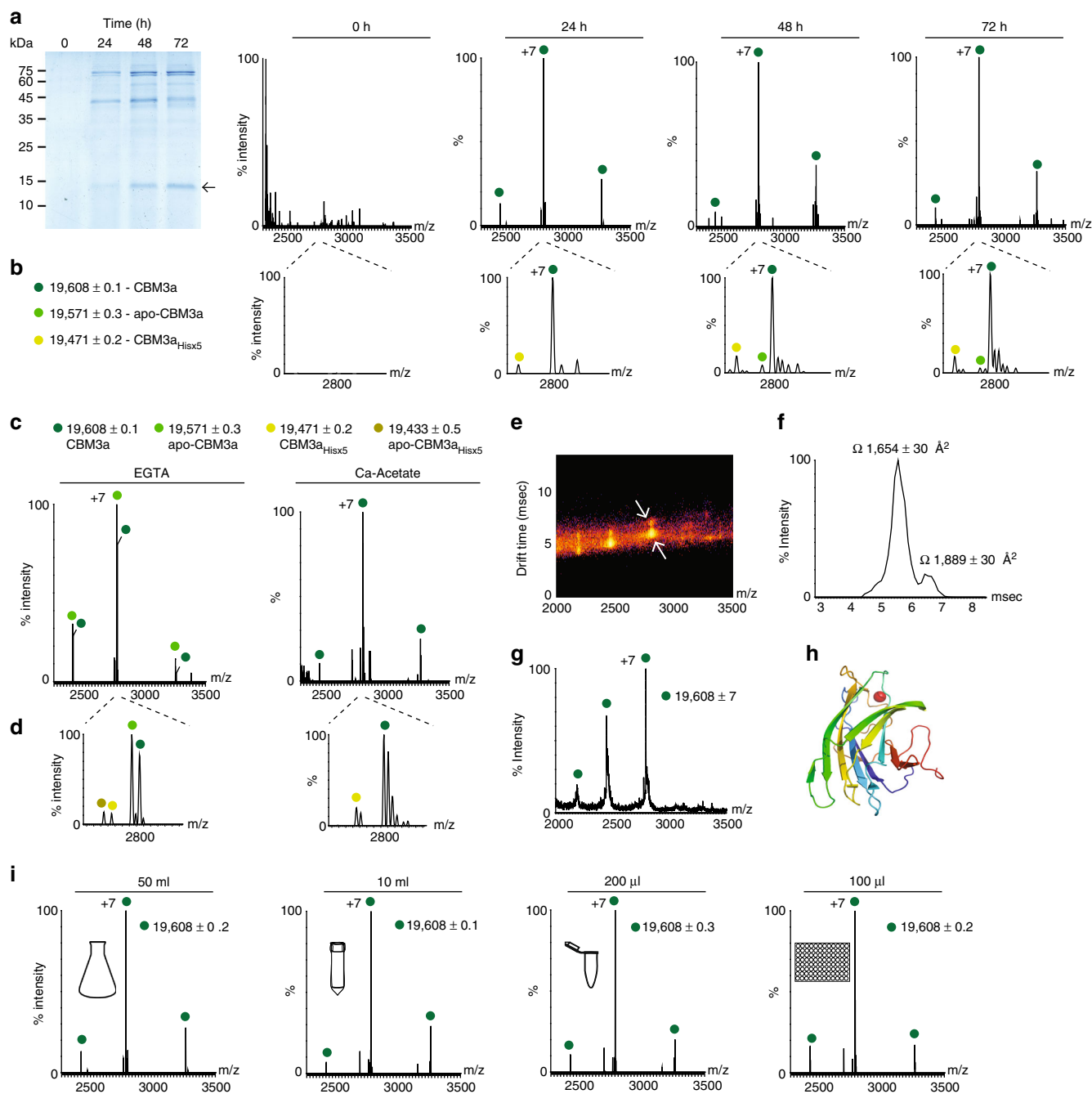
**Direct mass spectrometry characterization of recombinant proteins in yeast.** To investigate the applicability of the direct-MS

mass spectrometry approach to eukaryotic expression systems, we initially turned to *P. pastoris*, which is considered a successful heterologous expression system for the production of recombinant biopharmaceuticals and industrial enzymes<sup>13</sup>. We examined the production of two proteins: CBM3a (19.6 kDa) from *C. thermocellum*<sup>12</sup>, and serum albumin (67 kDa). We began by monitoring the expression profiles of the two proteins by SDS-PAGE at different time points, following their induction. Twenty-four hours following methanol induction, expression of both CBM3a and albumin was detected (Fig. 1a and Supplementary Fig. 1).

To test whether we could obtain similar types of information by means of native mass spectrometry, collected aliquots were buffer exchanged into 1 M ammonium acetate at pH = 7.0, to ensure compatibility with mass spectrometry analysis. Spectra were then recorded on an Orbitrap platform<sup>14</sup>, without any concentration step. Highly resolved peaks with narrow charge state distributions were detected 24 h after induction initiation, with negligible contributions from background proteins, indicating that CBM3a and albumin are expressed in soluble and folded forms (Fig. 1a and Supplementary Fig. 1). From the mass measurements, we could confirm that the signal peptides of both proteins were properly cleaved. We could also determine that CBM3a is associated with one calcium (Ca<sup>2+</sup>) ion, as would be expected from a protein containing a calcium-binding domain<sup>12</sup>. In addition, we identified another minor charge series corresponding to a CBM3a species lacking one histidine residue (CBM3a<sub>Hisx5</sub>), from the C-terminal His×6 affinity tag designed for subsequent purification. Moreover, we noticed that at longer production times (48 and 72 h), a population of apo-CBM3a lacking the associated Ca<sup>2+</sup> ion (apo-CBM3a) is detected, suggesting the exhaustion of calcium from the growth medium, and the need for supplementation (Fig. 1b). Minor peaks of the slightly larger populations, which correspond to sodium and water adducts, were also detected. Overall, the spectra acquired directly from the growth medium revealed the expression states of both CBM3a and albumin, providing useful information for optimization of growth conditions.

To confirm that Ca<sup>2+</sup> ions are indeed associated with CBM3a, we treated the growth medium with 50 mM EGTA, a calcium chelating agent. The resulting spectrum indicated the appearance of additional charge states, corresponding in mass to the apo form of the protein (Fig. 1c, d). Such populations were detected both for the full-length protein and the His truncated variant (CBM3a<sub>Hisx5</sub>). Adding 0.5 mM calcium acetate to the treated medium shifted the protein back to the holo, Ca<sup>2+</sup> bound state (Fig. 1c, d), confirming that calcium is, indeed, a CBM3a-associated cofactor. Hence, cofactor identification can be performed from crude samples, eliminating need for isolation and purification steps.

Next, we wished to examine whether direct mass spectrometry analysis of secreted proteins can also be expanded to include ion-mobility mass spectrometry (IM-MS) measurements, which provide information on the higher-order structure of the protein<sup>15,16</sup>. In an ion-mobility experiment, the time it takes a protein to traverse a weak electrical gradient in a gas-filled chamber is measured. Drift times are then converted to collision cross-sections (CCS), reflecting the overall shape of the protein<sup>15,16</sup>. The resulting data indicated that despite the presence of contaminant proteins within the culture medium, well-resolved spectra could be collected for CBM3a using the Synapt G2 instrument (Fig. 1e–g). Two distinct conformations were detected for each charge state. The predominant species gave rise to a CCS value of  $1654 \pm 30 \text{ \AA}^2$ . Theoretical CCS for CBM3a, calculated by the projection approximation algorithm, based on the crystal structure (PDB Entry 4JO5) (Fig. 1h), yielded a value of  $1489 \text{ \AA}^2$ ,



**Fig. 1** Proteins can be analyzed directly from the crude growth medium of *P. pastoris* cells. **a** SDS-PAGE and direct-MS time-course analyses of CBM3a over a period of 72 h post-induction. The band corresponding to CBM3a in the gel is denoted with an arrow. The uncropped gel image can be seen in Supplementary Fig. 11A. Mass spectra reveal that CBM3a is detected as early as 24 h after induction is initiated. **b** Enlargement of the +7 charge state at each time point (2775–2820 *m/z* range). The data reveal the presence of different CBM3a populations. The major population (dark green circles) corresponds to the Ca<sup>2+</sup> bound form of the His<sub>6</sub> tagged CBM3a. Light green circles label the population of the Ca<sup>2+</sup> bound form of CBM3a lacking one histidine (CBM3a<sub>Hisx5</sub>), and green circles label the apo-CBM3a that is not associated with the Ca<sup>2+</sup> cofactor. The non-labeled minor peaks correspond to sodium and water molecule adducts. **c** Ca<sup>2+</sup> binding to CBM3a was validated by incubation of the medium with the Ca<sup>2+</sup> chelator EGTA (left panel). Incubation with EGTA resulted in a decrease in the intensity of the Ca<sup>2+</sup>-bound form of CBM3a, and a subsequent increase in the charge series of the apo-CBM3a. To recover Ca<sup>2+</sup> binding to CBM3a, the medium was then incubated with calcium acetate. **d** Spectra showing the enlargement of the +7 charge states. **e** Analysis of the higher-order topology of CBM3a by native IM-MS measurements in crude medium. The IM-MS spectrum reveals the existence of two different arrival times of CBM3a, suggesting that the protein adopts two distinct conformations (marked with arrows on the mobilogram). **f** Arrival time distributions for the +7 charge states of the two populations are shown together with calculated CCS values. **g** A representative two-dimensional plot of *m/z* vs. intensity. **h** The crystal structure of CBM3a (PDB Entry 4J05). **i** *P. pastoris* was induced with methanol for 24 h at decreasing culture volumes, ranging from 50 mL to 100  $\mu$ L. Well-resolved charge state series of CBM3a were detected at all cell culture volumes, even at 100  $\mu$ L.

which fits well with the measured CCS value, given that the projection approximation algorithm used to calculate this value is known to underestimate CCS values by ~13%<sup>17</sup>. We noticed that in all measurements, which were performed under gentle conditions such that structural perturbations are minimized, a second, minor drift-time population was observed (Fig. 1e, f). This population (15%) displayed a CCS value of  $1889 \pm 30 \text{ \AA}^2$ , possibly reflecting the existence of a more extended conformation. Direct-MS ion-mobility measurements can therefore reflect the higher-order shape of unpurified proteins.

We then aimed to determine the minimal culture volume sufficient for direct-MS analysis. In order to achieve this we induced *P. pastoris* at decreasing amounts of culture volumes, starting from a 50 mL culture in a 250 mL flask, through a 10 mL culture grown in a 50 mL tube, a 200  $\mu\text{L}$  culture in a 1.5 mL tube, down to a 100  $\mu\text{L}$  culture in a 96-well plate format. As can be seen in Fig. 1i, well-resolved charge series of CBM3a with high signal-to-noise ratios were obtained throughout the decrease in culture volume. Even in the 100  $\mu\text{L}$  cell culture grown in a 96-well plate, charge states attributed to CBM3a could be clearly detected. Given this observation, we anticipate that the direct MS approach can be scaled up for high-throughput screening.

**Characterization of overproduced proteins in insect cells.** We continued by examining whether the direct-MS strategy may be applied to protein production characterization in insect cells, a system that produces proteins that are functionally similar to the native mammalian proteins, and can be readily adapted to high-density suspension cultures for large-scale expression<sup>18</sup>. We inoculated Sf9 cells with baculoviruses containing an expression cassette for the production of the ectodomain of *N. albigula* TfR1, fused to the N-terminal signal peptide gp67, and to a C-terminal His $\times$ 6 tag. The viruses also included an expression cassette for the green fluorescent protein (GFP), which is used as a reporter for infection efficiency.

We began by collecting the culture medium at different time points post-virus infection, in order to determine the protein expression profile. SDS-PAGE analysis indicated that 48 h after infection, TfR1 is barely detected, but it can be clearly seen after 72 h (Supplementary Fig. 2). Similarly, direct-MS analysis from crude growth media revealed the production of TfR1 72 h following virus infection, whereas GFP was already detected after 24 h. Notably, GFP accumulates intracellularly in cells, and its presence in the growth medium is attributed to ongoing cell death, which causes the cellular content to spill into the medium. As evidenced by SDS-PAGE and native mass spectrometry analyses, 96 h after inoculation, the major protein in the growth medium was GFP, though TfR1 was still clearly detected. At this time point, the growth cultures are exhausted and more pronounced cell death occurs, resulting in the accumulation of high levels of GFP in the medium. Thus, direct MS enables simple, informative time course measurements that can be amenable to optimizing production yields.

Mass analysis of the spectra enabled the identification of multiple TfR1 forms (Fig. 2a). The most abundant population corresponds to the intact protein after cleavage of the signal peptide and modification by glycosylation. The 1039 Da shift in mass suggests that the glycoform is a hexasaccharide with a  $\text{GlcNac}_2\text{Man}_3\text{Fuc}_1$  (TfR1<sup>a</sup>) composition, the most common N-linked core glycan in insect cells<sup>19</sup>. In addition, we identified a population with two  $\text{GlcNac}_2\text{Man}_3\text{Fuc}_1$  glycosylation sites (TfR1<sup>b</sup>), as well as a population of TfR1 modified with a pentasaccharide ( $\text{GlcNac}_2\text{Man}_3$ ), which is missing the fucose moiety (TfR1<sup>c</sup>). We also assessed the relative abundance of the differentially modified glycoforms, indicating that 65% of the

protein is modified with one  $\text{GlcNac}_2\text{Man}_3\text{Fuc}_1$ , while 10 and 7% are modified with two  $\text{GlcNac}_2\text{Man}_3\text{Fuc}_1$  or one  $\text{GlcNac}_2\text{Man}_3$ , respectively. Though relative quantitation for the glycoforms can be evaluated by SDS-PAGE analysis (Fig. 2c), we were not able to distinguish between the penta- and hexa-oligosaccharides.

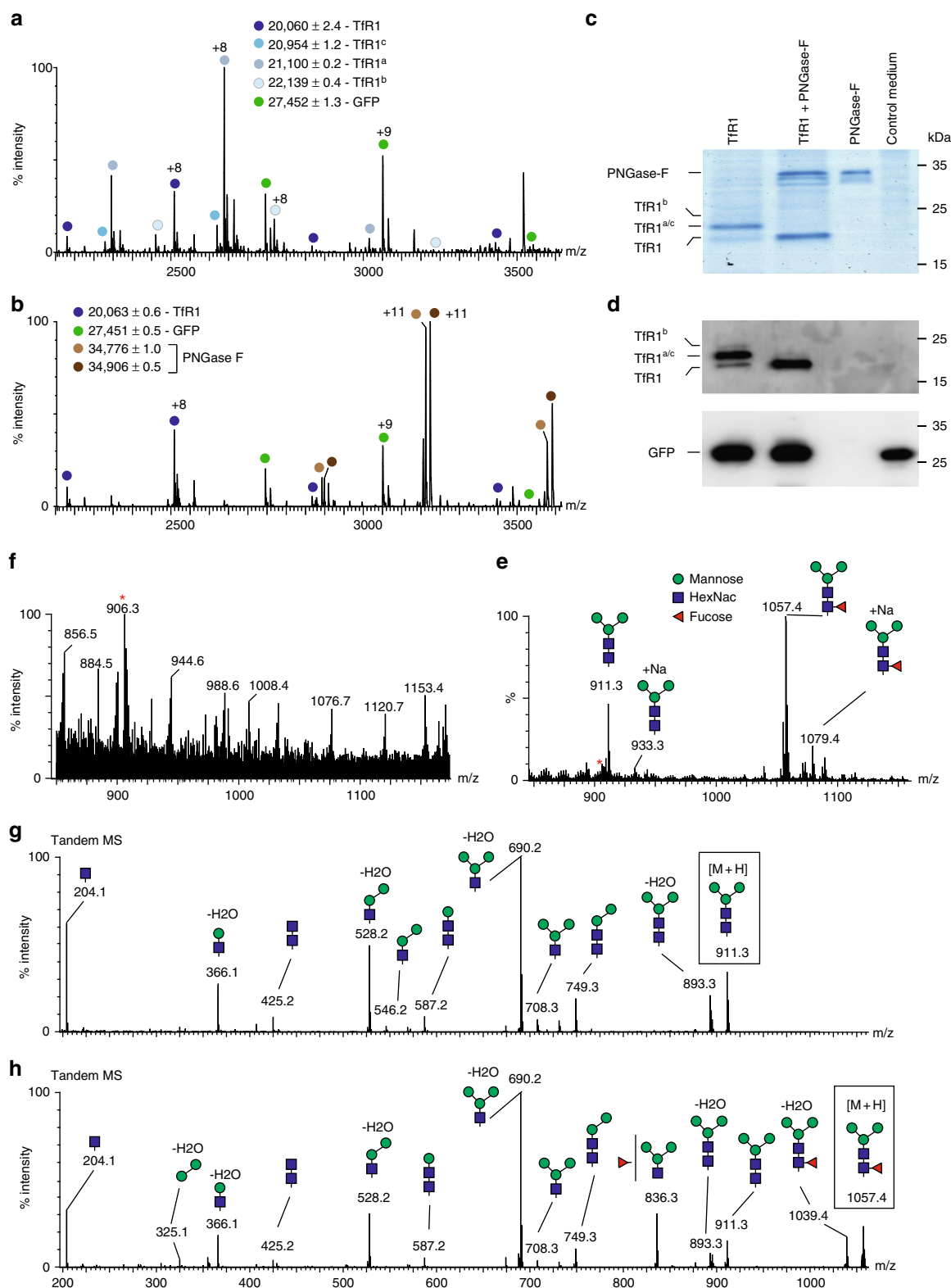
To validate that the molecular heterogeneity is indeed caused by glycosylations, we treated the growth medium with the deglycosylating enzyme PNGase F, which removes N-linked oligosaccharides at the asparagine linkage<sup>20</sup>. The enzymatic removal of the glycans gave rise to a simplified spectrum of TfR1, consisting solely of the unmodified protein (Fig. 2b). The PNGase F enzyme itself, as well as GFP, were also detected in the spectrum (Fig. 2b). SDS-PAGE followed by Western blot analysis confirmed that the heterogeneity of TfR1 results from glycosylation (Fig. 2c, d).

To characterize each glycoform in detail, we treated the culture medium with PNGase F, and recorded spectra of the low  $m/z$  region ( $1500 > m/z$ ). The data indicated the presence of two glycoforms, that in accordance with their measured masses of 1057 and 911 Da, are likely to be protonated  $\text{GlcNac}_2\text{Man}_3\text{Fuc}_1$  and  $\text{GlcNac}_2\text{Man}_3$ , respectively, containing a free reducing end (Fig. 2e). In a control measurement of the same medium prior to PNGaseF treatment, these glycans were not detected, suggesting that they indeed originate from TfR1 (Fig. 2f). We then applied the tandem mass spectrometry (MS/MS) approach, in order to further validate the composition of the two glycans. Accordingly,  $\text{GlcNac}_2\text{Man}_3\text{Fuc}_1$  and  $\text{GlcNac}_2\text{Man}_3$  were isolated and collisionally activated to induce their fragmentation. Assignment of the fragments of both the glycans confirmed their structural composition (Fig. 2g, h; Supplementary Table 1). Hence, direct-MS analysis enabled glycosylation analysis directly from growth media, overcoming the need for purification and liquid chromatography/mass spectrometry measurements.

#### Analysis of antibodies from crude media of human cell lines.

Our next goal was to examine proteins expressed in a mammalian cell culture system. We therefore turned to HEK293 cells, which are frequently used in the manufacture of proteins for therapeutic and for industrial purposes<sup>21</sup>. Initially, we co-transfected adherent HEK293 cells with plasmids for the expression of light and heavy chains of the 289 antibody, a monoclonal antibody that recognizes the envelope glycoprotein 1 from the Junin mammarena virus. We then grew the cells for three days in a medium containing a reduced amount of serum (2% instead of 10%), to decrease signal masking by the bovine serum albumin. After concentrating the growth medium by a factor of 10, and exchanging the buffer into volatile ammonium acetate, a nearly baseline resolution spectrum of the secreted antibody was acquired on the Orbitrap instrument<sup>14</sup> (Supplementary Fig. 3A). The measured mass and the low number of charge states indicated that a folded antibody is produced. Moreover, the 2925 Da mass deviation between the expected and measured mass is likely due to two GOF modifications, corresponding to core-fucosylated glycoforms ( $\text{GlcNac}_4\text{Man}_3\text{Fuc}_1$ ), one of the major modifications of recombinant antibodies<sup>22–24</sup>.

To overcome the need for sample concentration, and possible inflicted stress owing to the reduced serum levels, we turned to suspension-grown HEK293F cells, which are cultured in a chemically defined protein-free medium. We co-transfected cells with plasmids for the expression of light and heavy chains of the Mut23 antibody, which recognizes a sugar epitope. After 5 days of culture, we collected the medium, and acquired data directly after performing a buffer exchange step into a mass spectrometry-compatible solution. Under these growth conditions, no concentration steps were required,



and a nicely resolved mass spectrum of the intact antibody could be obtained (Supplementary Fig. 3B). The data indicate that the presence of intact antibodies can be detected, as in the case of the 289 and Mut23 antibodies, with no free heavy or light chain ions present, demonstrating efficient antibody assembly.

Next, we explored the applicability of this direct-MS approach for assessing the quality of designed versions of proteins. We

examined two unrelated antibodies, targeting the human vascular-endothelial growth factor (VEGF), termed G6<sup>25</sup>, the anti-lysozyme antibody D44.1<sup>26</sup>, and their designed variants, G6<sup>des13</sup> and D44.1<sup>des</sup>, respectively, specifically designed for increased stability, aggregation resistance, and expression yields. We began by recording spectra of the growth medium from non-expressing cells, which indicated the presence of several well-resolved charge state series between 4000 and 8000 *m/z*, reflecting



**Fig. 2** Multiple forms of TfR1 are produced in Sf9 cells. **a** Direct-MS analysis of secreted TfR1 from insect cell growth medium reveals four different populations of the protein. The major population corresponds in mass to the protein, modified with GlcNac<sub>2</sub>Man<sub>3</sub>Fuc<sub>1</sub> (TfR1<sup>B</sup>), leading to a 1039 Da shift in mass. In addition, three other populations of TfR1 can be detected, corresponding in mass to the intact protein (TfR1), the protein labeled with two GlcNac<sub>2</sub>Man<sub>3</sub>Fuc<sub>1</sub> hexasaccharides (TfR1<sup>C</sup>), as well as a population labeled with the pentasaccharide GlcNac<sub>2</sub>Man<sub>3</sub> (TfR1<sup>E</sup>). The charge state series labeled with green circles represents GFP, which is accumulating intercellularly, but is present due to cell death, and spillage of the cellular contents into the growth medium. **b** In order to strip the glycosylations from TfR1, we treated the growth medium with PNGase F. Direct-MS analysis indicated that after 15 min of incubation, only charge state series corresponding to the unmodified form of the protein remained, while the glycosylated TfR1 populations disappeared. The series labeled with brown circles represent PNGase F. **c** SDS-PAGE and **(d)** Western blot analysis reveal the different glycoforms of TfR1 prior to and following incubation with PNGase F. The blots in **d** were reacted with a His Probe HRP-Conjugate to detect TfR1 and an anti-GFP antibody. The uncropped gel and blot images can be seen in Supplementary Fig. 11B-C. **e** Direct-MS spectra, focused on the low *m/z* scale, were acquired, following the addition of PNGase F to the growth medium. Two major glycans were detected, corresponding in mass to protonated GlcNac<sub>2</sub>Man<sub>3</sub>Fuc<sub>1</sub> and GlcNac<sub>2</sub>Man<sub>3</sub>, respectively, with a free reducing end. These oligosaccharides were not detected in the medium prior to PNGase F treatment (**f**). A background peptide from the growth medium is labeled with a red asterisk. **g, h** MS/MS analysis of the two glycans validated their assignment. Each glycan was selected within the quadrupole mass filter, and subjected to MS/MS fragmentation. The selected glycan is enclosed within a box, and the various ion fragments generated are annotated

the presence of a background signal (Fig. 3a). In attempt to identify the nature of these proteins, we performed MS/MS and pseudo-MS<sup>3</sup> top-down analyses (Supplementary Fig. 4). MS/MS analysis indicated that the largest charge state series, with a mass of 146,269 Da corresponds to a tetramer, composed of monomers of 36,549 Da. MS<sup>3</sup> activation and top-down analysis<sup>14</sup>, revealed that this protein is lactate dehydrogenase B chain, a cytosolic enzyme known to leak into the growth medium in cell cultures due to loss of cytosolic membrane integrity of damaged cells during necrosis<sup>27,28</sup>. Similar to the presence of cytosolic GFP in insect cell cultures (Supplementary Fig. 2), lactate dehydrogenase can serve as an indicator for the level of cell death during extensive culturing of mammalian cells.

Analysis of medium from cells secreting the G6 antibody revealed that, in addition to the background medium proteins, a relatively low-intensity charge state series, corresponding in mass to the intact, modified antibody, is detected. A different scenario was observed for the designed version of the antibody, G6<sup>des13</sup>. In this case, the major charge-state envelope corresponded to the designed antibody, clearly demonstrating that the modified antibody exhibits higher expressibility, compared to G6 (Fig. 3a). Similar results were obtained for D44.1 and D44.1<sup>des</sup>, indicating that the expressibility of the designed proteins was improved (Supplementary Fig. 5).

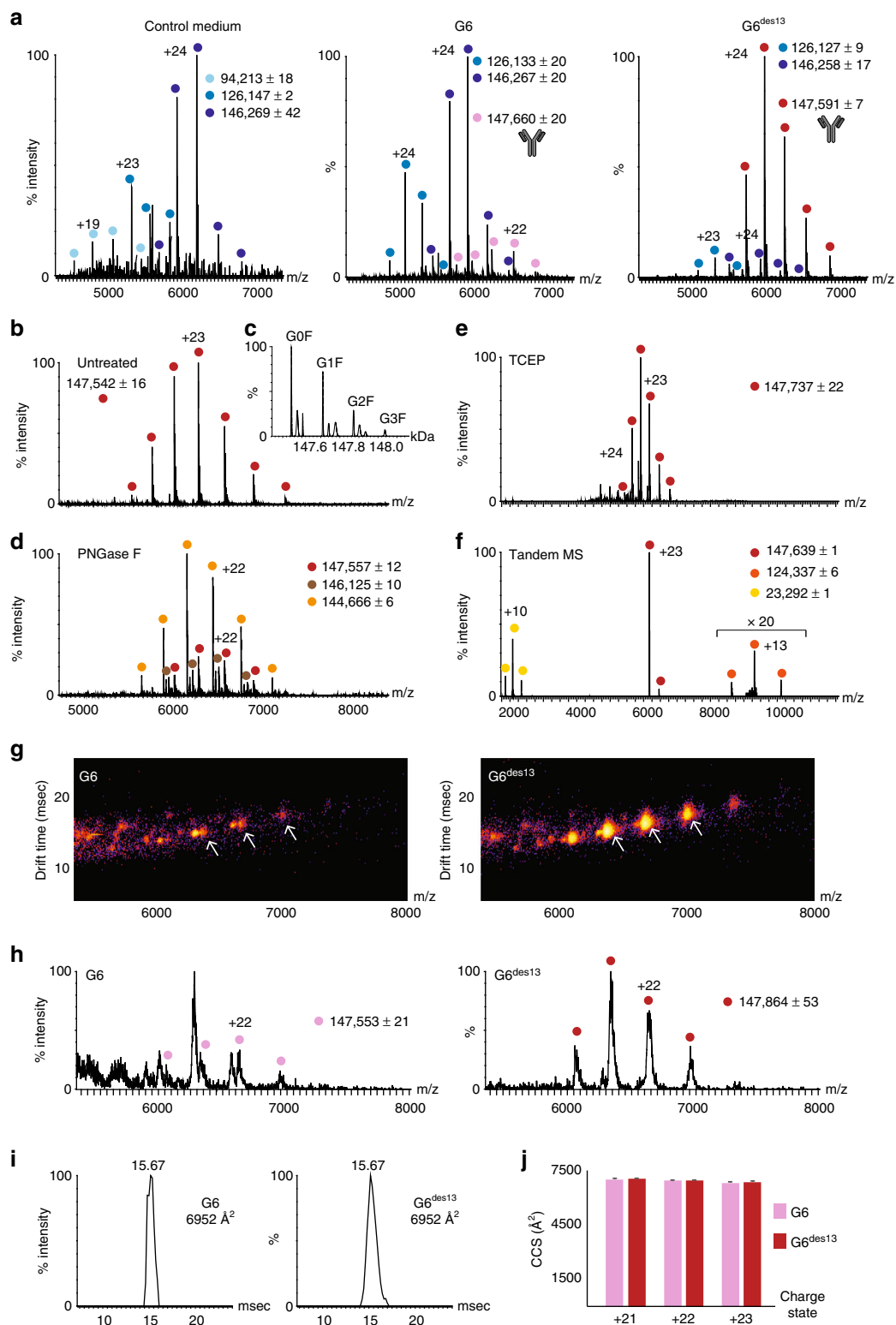
N-glycosylations are the most common type of post-translational modification on antibodies, and contribute to their structural characteristics by potentially modulating their function and pharmacokinetic properties<sup>29-31</sup>. Indeed, we noticed that all analyzed antibodies exhibit a ~3 kDa shift between their measured and calculated mass, suggesting that they are glycosylated (Fig. 3b, c, Supplementary Fig. 6). To validate this assumption, and test whether it is possible to deglycosylate antibodies directly in crude media, we treated the media with PNGase F. Molecular mass shift of spectra recorded after PNGase F treatment confirmed the release of glycans from the antibodies. (Fig. 3d; Supplementary Fig. 6). Mass measurement of the deglycosylated G6<sup>des13</sup> antibody indicated that the two heavy chain C-terminal lysines are clipped ( $-128 \times 2$  Da), and that it bears 14 disulfide bridges ( $-28$  Da). Calculating the mass difference between the deglycosylated G6<sup>des13</sup> antibody and its main glycosylated species enabled us to conclude, with an accuracy of 7 ppm, that the major glycosylated form is that of a G0F (GlcNac<sub>4</sub>Man<sub>3</sub>Fuc<sub>1</sub>) structure on each heavy chain<sup>22-24</sup> (Fig. 3c, d). In addition, three glycoforms with the characteristic 162 Da spacing, corresponding to galactose, have been defined, indicative of G1F, G2F, and G3F species, with one, two and three galactose molecules, respectively. Examination of the low *m/z* region of the spectrum prior to and following PNGase F treatment indicated the appearance of the two main glycans;

namely, the sodiated and non-sodiated G0F (GlcNac<sub>4</sub>Man<sub>3</sub>Fuc<sub>1</sub>) and G1F (GalGlcNac<sub>4</sub>Man<sub>3</sub>Fuc<sub>1</sub>), species (Supplementary Fig. 7), which were only detected in the spectrum of the treated sample. Overall, these results suggest that isolation and purification of produced antibodies is not necessarily a prerequisite for high-resolution native mass spectrometry analysis.

Full characterization of antibodies includes the analysis of the composing heavy and light chains that are generated by reduction of the disulfide bonds<sup>20,32</sup>. To assess the feasibility of in-medium reduction, we added the TCEP reductant to the growth medium of the G6<sup>des13</sup> antibody, prior to mass spectrometry recording (Fig. 3e). MS/MS experiments were then performed, in which a single peak corresponding to the intact antibody was isolated. The ions were then subjected to collisional activation in the higher energy collision-induced dissociation (HCD) cell, liberating the light chain from the antibody as it is no longer covalently attached to the heavy chain (Fig. 3f). The measured mass of the dissociated light chain was smaller than its theoretical mass (23,296 Da) by 4 Da, indicating that the TCEP treatment did not reduce the two intermolecular sulfur bridges within the light chains. In addition, the close agreement between the measured and calculated mass of the light chain implies that the light chains are not glycosylated, but rather the heavy chains are modified with glycans. The fact that heavy chains were not detected under any of the tested reduction experiments can be attributed to low reduction efficiency of the hinge region, possibly due to the structural properties of the antibody backbone scaffold. Similar results were obtained for D44.1 and its designed counterpart, D44.1<sup>des</sup> (Supplementary Fig. 8).

The baseline-resolved, direct-MS spectra encouraged us to explore whether we could also characterize the antibodies produced directly from the crude media, by means of IM-MS. Using this approach, we first collected IM-MS data for D44.1<sup>des</sup>, using the Synapt G2 instrument. The results indicate that despite the impurity of the sample, well-resolved IM-MS plots could be generated (Supplementary Fig. 9). The intact mass of the antibody and its corresponding CCS value could easily be extracted, emphasizing not only the additional experimental input yielded by this method, but also the fact that the direct-MS approach can be applied on various high-mass mass spectrometry platforms.

We next explored the G6 and G6<sup>des13</sup> pair of antibodies, examining whether the introduced mutations altered the overall structure of G6<sup>des13</sup>, in comparison to G6. IM-MS analysis demonstrated that, similar to the measurements conducted on the Orbitrap platform, the major charge series in the crude G6 sample corresponded to background medium proteins; nevertheless, the antibody signals could be clearly identified



(Fig. 3g, h). In the G6<sup>des13</sup> sample, the most intense charge envelope corresponded to the antibody, while the medium protein signals were very low, supporting the view that G6<sup>des13</sup> exhibits a gain in expressibility. Moreover, the CCS values of G6 and G6<sup>des13</sup>, averaged over 3 wave heights and 3 charge states were calculated to be  $6765 \pm 6 \text{ \AA}^2$  and  $6745 \pm 3 \text{ \AA}^2$ , respectively, suggesting that the two antibodies adopt similar conformations, and that the incorporated mutations did not alter the structural

compactness of G6<sup>des13</sup> (Fig. 3i, j). Taken together, we found that direct-IM-MS measurements can therefore reflect the higher-order shape of unpurified antibodies.

**Sensitivity evaluation of the direct-MS approach.** Finally, we examined the sensitivity of the method by assessing the limit of detection of serially diluted yeast, insect and human cultures,

**Fig. 3** Rapid characterization of recombinant antibodies, directly from the growth medium of secreting human cells. **a** Expression levels of secreted antibodies were evaluated by direct-MS. Suspension-grown HEK293 cells were transfected with plasmids for expression of light and heavy chains of chimeric antibodies. Growth media were collected and analyzed after 5 days. In comparison, a spectrum obtained from medium of control, non-expressing cells is shown, where various background proteins were detected. The results show that the G6 antibody accumulates to low levels, below the intensity of the background proteins, whereas the designed G6<sup>des13</sup> antibody accumulates to higher levels, and appears as the main charge envelope in the spectrum, due to the improved design. The antibody cartoon labels the antibodies measured mass. **b–d** Glycan analysis of unpurified antibodies in the growth medium. The high resolution spectrum of the G6<sup>des13</sup> antibody was measured before (**b**) and after (**d**) PNGase F treatment of the growth medium. **c** Deconvolution of the spectrum of the untreated antibody revealed four major glycoforms. Although minor subpopulations of partially or fully glycosylated antibodies were detected in the treated sample, comparison of the mass of the modified antibody to that of the deglycosylated form enabled us to determine that the core glycan is G0F, and the other glycans differ by the addition of a galactose molecule (G1F, G2F, and G3F). **e, f** Reduction of antibodies in the growth medium enabled us to liberate their light chains. **e** The G6<sup>des13</sup> antibody was reduced with TCEP, enabling the dissociation of the light chain during tandem MS analyses (**f**). **g–j** Analysis of the higher order topology of antibodies by whole range IM-MS measurements in crude media. Three-dimensional mobilograms (**g**) and their representative mass spectra (**h**) were measured for the G6 and G6<sup>des13</sup> antibodies, directly from the crude growth media. **i** Arrival time distributions and calculated CCS values for the +22 charge states of both antibodies, measured at wave height of 18 V. **j** Results indicate that the CCS values of the two antibodies are essentially identical, and that the mutations incorporated within G6<sup>des13</sup> did not modify the overall shape of the antibody. Standard deviations are shown. The original mean source data is shown in Supplementary Data 1

initially supplemented with 20  $\mu$ M bovine serum albumin (BSA) (Supplementary Fig. 10). The measured spectra indicated that by diluting the effective culture volume with ammonium acetate, high quality spectra of BSA can be recorded even for concentrations less than 100 nM (Supplementary Fig. 10). In addition, we examined whether it is possible to monitor changes in protein abundance within the crude media. To achieve this, a series of spectra were recorded using yeast, insect and human cultures supplemented with increasing concentrations of BSA. Figure 4 shows representative mass spectra of the three different growth media, obtained using elevated BSA concentrations. The relative abundance of BSA was then estimated in respect to selected background peaks whose intensity remained unchanged throughout the analysis. For each mass spectrum, the intensity ratio of BSA and the selected reference peaks was calculated and plotted against known BSA concentrations (Fig. 4 bottom). A linear trend line was observed for all three expression systems, however, the linearity range differed. We noticed that the range of linearity is dependent on the highest concentration level of BSA that can be quantified, which is limited by the detection of the reference peak. The low concentration of protein background in human culture (5 ng/ml), in respect to yeast and insect cells growth media (133 ng/mL and 88 ng/mL, respectively), reduces the quantification range, as seen in Fig. 4, yet, it enables the detection of low expression levels of proteins. Taken together, the data indicates that for a specified range, it is possible to measure dynamic changes in protein levels directly from culture media.

## Discussion

Here, we report a method for rapid characterization of secreted recombinant proteins, directly from crude growth media. Based on native-mass spectrometry, this method enables the assessment of multiple protein attributes in a single spectrum. We demonstrate that the expression, solubility and assembly state of overproduced proteins can be determined, as well as their associations with cofactors. The mass accuracy afforded by this strategy enables not only the identification of post-translationally modified species, but also the existence of differential modifications, and the distribution of the different forms. We also show that the overall structure of the unpurified overproduced proteins can be defined.

Here, we have focused on recombinant proteins produced in yeast, insect, and human cells; however the method can be expanded to include many other expression platforms that overproduce secreted proteins, such as Chinese hamster ovary cells (CHO)<sup>33</sup> and murine myeloma cell lines (GS-NS0)<sup>10,34</sup>. The fundamental requirement for this approach, however, is that the

overexpressed target becomes the dominant protein in the growth medium. Consequently, the in-depth structural characterization afforded by native mass spectrometry analysis can be performed. Moreover, it should be noted that the strategy is not dependent on a specific mass spectrometer platform; rather, it is a general approach that can be applied on either Orbitrap or QTOF-based platforms that provide extended mass range<sup>35–37</sup>.

A key to successful production of recombinant proteins for both industrial and pharmaceutical applications is to achieve maximum productivity at a reasonable cost<sup>10,38</sup>. Therefore, a screening program is generally performed to identify high-producing cells, optimize genetic and engineered constructs, and improve media formulation. We anticipate that the method outlined here will facilitate this labor-intensive and costly process, by overcoming the need for protein purification prior to characterization, and by providing, in a single step, in-depth quality assessment of the protein produced. In addition, the minimal amount of sample required, coupled with the simplicity and feasibility of the approach across different platforms, also make it suitable for high-throughput screening procedures that may further reduce the time gap between production and characterization.

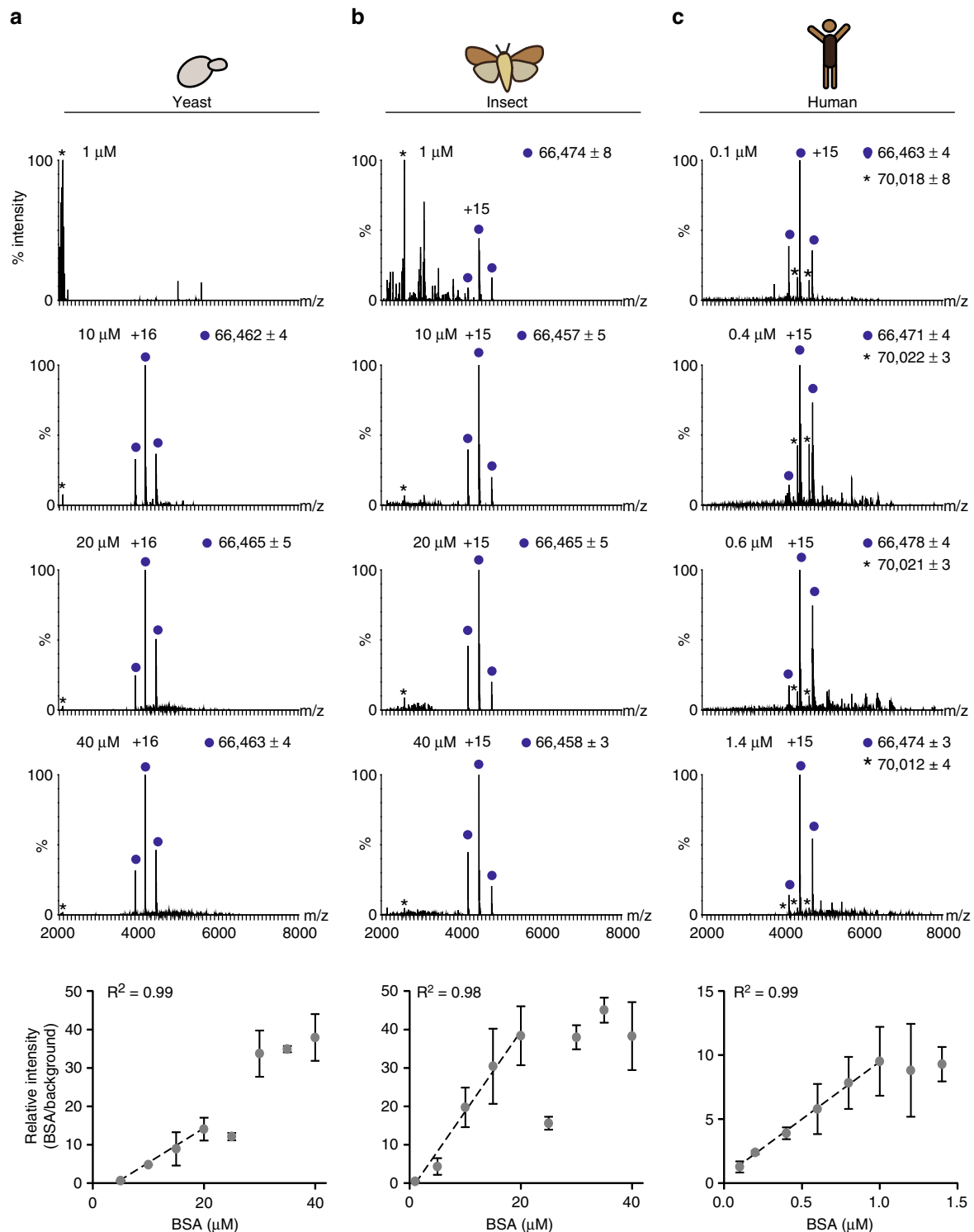
The ability to monitor, in real time, recombinant protein production is a valuable feature of this strategy. Therefore, we expect that further extension of this method will involve its coupling to microfluidics systems that will enable continued analysis of cultures grown in bioreactors. Such analyses will make it possible to capture optimal harvesting times, preventing misfolding and proteolysis, and determine the necessity of cofactor supplements to ensure proper folding and assembly. Such knowledge will improve both yield and quality of the recombinant proteins produced.

## Methods

**Plasmids.** Yeast: CBM3a, the cellulose-specific carbohydrate-binding module of the scaffoldin protein from *C. thermocellum*<sup>12</sup>, was cloned into the yeast expression vector pPIC9K (Thermo Fisher Scientific) under the regulation of the methanol inducible promoter of the *Alcohol Oxidase 1* gene. The gene was fused to the N-terminal  $\alpha$ -factor secretion signal and a C-terminal His<sub>6</sub> tag. Genomic integration into *P. pastoris* was performed at the *alcohol oxidase 1* locus, according to the manufacturer's instructions.

**Insect cells:** A modified monomeric ectodomain of the *N. albigula* Transferrin Receptor 1 (TfR1), was cloned into the pACGP67-B vector, under the regulation of the polyhedrin promoter. The gene was fused to the N-terminal signal peptide gp67, and to a C-terminal His<sub>6</sub> tag. Sf9 cells were co-transfected with 0.4  $\mu$ g of the pACGP67b vector and a 0.25  $\mu$ g linearized baculovirus genome (ABvector) using EscortIV transfection reagent (Sigma). Four days later medium was collected from transfected cells (P<sub>0</sub>), and was used to infect fresh Sf9 cells for amplification (P<sub>1</sub>–P<sub>2</sub>).





**Fig. 4** Changes in protein abundance can be monitored in crude media. A series of measurements with different concentrations of externally-added BSA, were performed in growth media from yeast (a), insect (b), and human (c) cell cultures. The intensity ratio between BSA (dark blue circles) and background reference peaks (labeled with asterisks) was plotted against known BSA concentrations (bottom graphs). A linear trend was found for BSA concentrations of up to 20 μM in yeast and insect cell cultures, and up to 1.4 μM in human growth media. Trend lines and R-squared values are shown. Error bars represent standard deviations of three measurements. The original mean source data is shown in Supplementary Data 2

Human cells: The variable regions of the heavy and light chains of several synthetic chimeric antibodies were cloned separately, upstream of the IgG1 human Ab scaffold, into p3BNC plasmids containing the human IgG secretion signal peptide. Antibodies used in this study were: (a) Antibodies G6 and G6<sup>des13</sup>, against human VEGF. The G6 antibody is based on the clinical anti-VEGF antibody avastin. The difference between G6 and its designed variant G6<sup>des13</sup> are six amino

acids in the interface between the light and heavy chains; (b) Antibodies D44.1 and D44.1<sup>des</sup>, against lysozyme. The difference between D44.1 and its designed variant D44.1<sup>des</sup> are eight amino acids in the interface between the light and heavy chains; (c) Antibody 289, recognizing the envelope glycoprotein 1 from the Junin mammarena virus; (d) Antibody Mut23, recognizing a mouse sugar epitope.

**Cells and growth conditions.** Yeast: CBM3a was expressed in *P. pastoris* GS115 cells. Serum albumin harboring its native secretion signal was expressed in *P. pastoris* GS115/His<sup>+</sup> Mut<sup>S</sup>. Cells were grown in Buffered Glycerol-Complex Medium and Buffered Methanol-Complex Medium at 30 °C, at 220 rpm. A single colony was transferred into 10 mL of Buffered Glycerol-Complex Medium, and grown until the optical density reached 4 at 600 nm (equivalent to  $2 \times 10^8$  cells/mL). Cultures were centrifuged, and the cell mass was collected and resuspended in 50 mL Buffered Methanol-Complex Medium. Expression was induced by the addition of 0.5% methanol every 24 h throughout the entire protein production period.

For minimal culture volume analysis, cells resuspended in Buffered Methanol-Complex Medium were divided into different culture volumes: 50 mL in a 250 mL flask, 10 mL in a 50 mL tube, 200  $\mu$ L in a 1.5 mL tube, and 100  $\mu$ L in a 96-well plate. Cells were grown as detailed above, except for the 96-well plate, which was covered with an AeraSeal™ breathing cover (Excel Scientific) and placed in a vibrating incubator (Heidolph Titramax 1000) at 30 °C, at 650 rpm. Growth media were analyzed 24 h post-methanol induction.

Insect cells: TfR1 was expressed in baculovirus-infected Sf9 cells. Cells were grown in ESF921 medium (Expression Systems) at 27 °C, at 140 rpm. Cultures (30 mL) containing  $1.8 \times 10^6$  cells/mL were infected with 120  $\mu$ L of TfR1 P2 viruses. Growth media were analyzed 72 h post-infection. For time point expression analysis, 1 mL culture samples were taken from a 30 mL culture of infected cells, at different time points post-infection, as indicated.

Human cells: For antibody expression and secretion, two plasmids (for light and heavy chains) were co-transfected together simultaneously into cells. Antibodies were expressed in either adherent or suspension-grown HEK293 cells. Adherent cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, penicillin-streptomycin, sodium pyruvate, nonessential amino acids (Biological Industries), and MycoZap (Lonza) in a humidified incubator at 37 °C, in a 5% CO<sub>2</sub> controlled environment. Suspension-grown HEK293F cells were grown in FreeStyle medium (Gibco), in a shaking incubator, at 37 °C, in a controlled environment of 8% CO<sub>2</sub>, at 115 rpm. Transfections to adherent cells were performed in 80% confluent 10-cm dishes, using jetPEI (PolyPlus), with 5  $\mu$ g of the light chain plasmid, and 5  $\mu$ g of the heavy chain plasmid. Post-transfection, cells were transferred to a growth medium containing 2% serum. Growth media were collected 48–72 h post-transfection. Transfections to suspension-grown cells were performed at a cell density of  $10^6$  cells/mL, using linear 40 kDa polyethylenimine (PEI) (Polysciences) at 3 mg of PEI per 1 mg of plasmid (0.5 mg of the light chain plasmid and 0.5 mg of the heavy chain plasmid) per 1 L of culture. Growth media were collected after 5–7 days.

**Sample preparation.** In general, crude media were cleared from cells and insoluble debris by centrifugation, and then buffer exchanged into 1 M ammonium acetate. We found experimentally that performing 1–2 cycles of buffer exchange into high salt ammonium acetate efficiently replaces the bulk non-volatile compounds in the sample, and result in high quality spectra. In some instances, as detailed below, we performed an additional buffer exchange cycle, into a low concentration of ammonium acetate (150–200 mM).

**CBM3a and serum albumin:** *P. pastoris* cells were cleared from the medium by centrifugation at  $14,000 \times g$  for 2 min. The growth medium was then supplemented with protease inhibitors (1 mM PMSF, 1 mM benzamidine, 1.4  $\mu$ g/mL pepstatin A), aliquoted, snap-frozen in liquid nitrogen, and stored at –80 °C. On the day of the measurements, samples were thawed, and buffer exchanged 2–3 times into 1 M ammonium acetate, pH 7, using Micro Bio-Spin™ 6 Columns (Bio-Rad) and directly sprayed into the mass spectrometer.

For cofactor binding analysis, 30–50  $\mu$ L buffer exchanged growth medium was incubated for 20 min at room temperature with 50 mM EGTA, and then buffer exchanged again into 0.2 M ammonium acetate, pH 7. Ca<sup>2+</sup> treatments were carried out by adding 0.5 mM of calcium acetate to the growth medium.

**TfR1:** Insect cells were cleared from the medium by centrifugation at  $1000 \times g$  for 5 min. The growth medium was then supplemented with protease inhibitors, as described, and cleared of insoluble materials by centrifugation at  $10,000 \times g$  for 10 min. The supernatant was aliquoted, snap frozen in liquid nitrogen, and stored at –80 °C. On the day of the measurements, samples were thawed, and buffer exchanged 2–3 times into 1 M ammonium acetate, pH 7, and directly sprayed into the mass spectrometer. For deglycosylation experiments, 50–100  $\mu$ L of buffer exchanged growth medium was incubated for 10 min at 50 °C in the presence of 1  $\mu$ L of Rapid PNGaseF (New England Biolabs) and Rapid PNGase F buffer, and then buffer exchanged again into 1 M ammonium acetate. For glycan analyses, 50–100  $\mu$ L of buffer exchanged growth medium was incubated for 5 h at 37 °C in the presence of 0.2  $\mu$ L of PNGaseF (New England Biolabs), then directly sprayed into the instrument.

**Antibodies:** Growth media were cleared from cells by centrifugation at  $600 \times g$  for 5 min, supplemented with 0.02% (w/v) sodium azide and 1 mM PMSF, and further clarified by centrifugation at  $10,000 \times g$  for 15 min. Clarified media were aliquoted, snap frozen in liquid nitrogen and stored at –80 °C. On the day of the measurements, samples were thawed and buffer exchanged 2 times into 1 M ammonium acetate, pH 7. In order to reduce cysteine bonds, we tested different concentrations of the reducing agents DTT and TCEP, added directly to the growth

medium, at different incubation times. We found that incubation of the medium of the G6<sup>des13</sup> antibody in the presence of 20 mM TCEP for 4 h at 37 °C, followed by two consecutive buffer exchanges into 1 M and 150 mM ammonium acetate, respectively, gave the best results. For deglycosylation experiments, 50–100  $\mu$ L of buffer exchanged growth medium was incubated for 10 min at 50 °C in the presence of 1  $\mu$ L of Rapid PNGaseF (New England Biolabs) and Rapid PNGase F buffer, and then buffer exchanged again into 1 M and 150 mM ammonium acetate. For glycan analyses, 50–100  $\mu$ L of buffer exchanged growth medium was incubated for 9 h at 37 °C in the presence of 0.2  $\mu$ L of PNGaseF (New England Biolabs), then directly sprayed into the mass spectrometer.

To visualize proteins on gels, aliquots from different samples were mixed with 5 $\times$  protein sample buffer, boiled for 5 min, and separated by SDS-PAGE. Gels were then stained with InstantBlue (Expedeon), or transferred to polyvinylidene difluoride membranes, and probed with a His Probe HRP-Conjugate (Thermo, 15165) at a dilution of 1:500, according to the manufacturer's instructions, or an anti-GFP antibody (Abcam, ab290) at a dilution of 1:2500.

**Native mass spectrometry.** Nano-electrospray ionization MS experiments were performed on two instrument platforms, a modified Q-Exactive Plus Orbitrap EMR<sup>14</sup> (Thermo Fisher Scientific, Bremen, Germany) and a Synapt G2 instrument modified for high mass analysis (Waters, Hertfordshire, U.K.). All measurements were performed at least three times. Spectra recorded by the Orbitrap platform were converted to Masslynx-compatible files using the Waters DataBridge software. All spectra were analyzed using the MassLynx V4.2 SCN982 software (Waters, Hertfordshire, U.K.). Reported masses (shown in Daltons) and mass errors were calculated using MassLynx. Mobilograms were generated by the Driftscope™ HDMS™ V2.8 software (Waters, Hertfordshire, U.K.). Top-down proteomic analysis was performed using the ProSightPC v.3.0 software.

Typically, an aliquot of 2  $\mu$ L protein solution was loaded into a gold-coated nano-electrospray ionization capillary prepared in-house, as previously described<sup>39</sup>, and sprayed into the instruments. The following experimental parameters were used on the Orbitrap platform: All spectra are shown without smoothing, and the instrument was calibrated externally, using cesium iodide. Conditions within the mass spectrometer were adjusted to preserve noncovalent interactions, and adjusted according to the type of protein measured. The source was operated in positive mode, capillary voltage was set to 1.7 kV, capillary temperature was 180 °C, and argon was used as the collision gas in the HCD cell. MS spectra were recorded at a resolution of 10,000–70,000, according to the complexity of the sample. Raw spectra were deconvoluted using Deconvolution 4 protein software (Thermo).

For CBM3a and albumin measurements, HCD voltage was set to 25 V and 85 V, respectively, at a trapping gas pressure setting of 1, which corresponds to an HV pressure of  $9.2 \times 10^{-6}$  mbar, and a UHV pressure of  $6.3 \times 10^{-11}$  mbar. Bent flatpole DC bias and axial gradient were set to 1.8 V and 10 V, respectively.

For TfR1 measurements, HCD voltage was set to 25 V, at a trapping gas pressure setting of 1.5, which corresponds to HV pressure of  $9.8 \times 10^{-6}$  mbar, and UHV pressure of  $6.9 \times 10^{-11}$  mbar. Bent flatpole DC bias and axial gradient were set to 1.8 V and 15 V, respectively.

For antibody measurements, HCD voltage was set to 50–100 V, at a trapping gas pressure setting of 3.9, which corresponds to HV pressure of  $1.04 \times 10^{-4}$  mbar, and UHV pressure of  $2.35 \times 10^{-10}$  mbar. Bent flatpole DC bias and axial gradient were set to 2 V and 25 V, respectively.

For the measurement of intact glycans, no HCD voltage was applied. Trapping gas was set to a pressure of 1, which corresponds to HV pressure of  $9.2 \times 10^{-6}$  mbar, and UHV pressure of  $6.7 \times 10^{-11}$  mbar. Bent flatpole DC bias and axial gradient were set to 1.8 V and 6 V, respectively.

The following experimental parameters were used on the Synapt G2 platform for IM-MS measurements: All spectra are shown with minimal smoothing, and the instrument was calibrated externally, using cesium iodide.

For CBM3a measurements, instrument parameters were set as follows: Capillary voltage of 1.5 kV, sampling cone 20 V, extraction cone 2 V. No trap and transfer collision energies were applied. Trap DC bias 45 V, and helium cell gas flow 120 mL/min. Nitrogen was used as the IMS gas, at a flow of 60 mL/min. IM wave velocity was set to 300 m/s, and wave height was set to 18, 19, or 20 V. Collision cross-sections were calculated for the +7 and +8 charge states of CBM3a.

For antibody measurements, instrument parameters were set as follows: Capillary voltage of 1.5 kV, sampling cone 10 V, extraction cone 2 V. Trap collision energy 20 V. No transfer collision energy was applied. Trap DC bias 45 V. Helium cell gas flow 120 mL/min. Nitrogen was used as the IMS gas, at a flow of 60 mL/min. IM wave velocity was set to 300 m/s, and wave height was set to 18, 19, or 20 V. Collision cross-sections were calculated for the +21, +22, and +23 charge states of the antibodies.

Proteins used for calibrations were: Ubiquitin (8.6 kDa); cytochrome-C (12.2 kDa); BSA (67.2 kDa); concanavalin A (103.1 kDa); alcohol dehydrogenase (149.3 kDa); and pyruvate kinase (227.0 kDa), all purchased from Sigma. Calibrants were dissolved in 200 mM ammonium acetate solutions to retain a native-like conformation. All measurements were conducted on the same day; the only parameters modified for the test and calibrating proteins were capillary voltages.

CCS values were calculated using Driftscope CCS calculation software (V2.8, Waters Corp., Manchester, U.K.) according to ref. 40. The theoretical CCS value for CBM3a (PDB Entry 4J05) was calculated using the Driftscope Projection Approximation algorithm (Waters).

For quantification and dilution experiments, we collected growth media from non expressing cells from the three different expression systems, yeast, insect and human cells, cultured for 24–48 h, and buffer exchanged twice into 1 M ammonium acetate. Protein concentration was determined using the Bio-Rad Protein Assay Dye Reagent (500–0006). For quantification experiments, 4  $\mu$ L of the buffer exchanged growth media were mixed with 1  $\mu$ L of BSA (Sigma A-4378) at different concentrations, as indicated. Spectra were recorded on an Orbitrap instrument, as described for serum albumin. For each mass spectrum, peak intensities of the three major charge states of BSA (+14, +15, and +16) were summed and divided by the intensity of selected background peaks, which served as references. The following reference proteins were selected: 2088, 2560, and the sum of  $m/z$  4376 and 4668  $m/z$  (corresponding to a protein with a mass of 70,000 Da) in the yeast, insect and human growth media, respectively. The relative values obtained were plotted against known BSA concentrations. For dilution experiments, yeast, insect and human cell cultures containing 20  $\mu$ M BSA were serially diluted with 150 mM ammonium acetate and spectra were recorded on the Orbitrap instrument, as described above. Each experiment was performed at least three times.

### Data availability

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

Received: 9 July 2018 Accepted: 12 November 2018

Published online: 03 December 2018

### References

- Adrio, J. L. & Demain, A. L. Microbial enzymes: tools for biotechnological processes. *Biomolecules* **4**, 117–139 (2014).
- Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug. Discov.* **7**, 21–39 (2008).
- Wells E., Robinson A. S. Cellular engineering for therapeutic protein production: product quality, host modification, and process improvement. *Biotechnol. J.* **12**, 1600105 (2017).
- Lagasse, H. A. et al. Recent advances in (therapeutic protein) drug development. *FTICR Res.* **6**, 113 (2017).
- Morrison C. Fresh from the biotech pipeline-2017. *Nat. Biotechnol.* **36**, 131–136 (2018).
- Schellekens, H. Biosimilar therapeutics-what do we need to consider? *NDT Plus* **2**, i27–i36 (2009).
- Structural Genomics, C. et al. Protein production and purification. *Nat. Methods* **5**, 135–146 (2008).
- Fazelinia, H., Cirino, P. C. & Maranas, C. D. Extending iterative protein redesign and optimization (IPRO) in protein library design for ligand specificity. *Biophys. J.* **92**, 2120–2130 (2007).
- Gan, J. et al. Native mass spectrometry of recombinant proteins from crude cell lysates. *Anal. Chem.* **89**, 4398–4404 (2017).
- Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22**, 1393–1398 (2004).
- Korajczyk, P. J., Zerbs, S., Collart, F. R. Protein production for biotechnology. In *eLS*, John Wiley & Sons, Ltd (Ed.) (2017).
- Yaniv, O. et al. Structure of a amyly 3a carbohydrate-binding module from the cellulosomal scaffoldin CipA of *Clostridium thermocellum* with flanking linkers: implications for cellulosome structure. *Acta Crystallogr. Sect. F* **69**, 733–737 (2013).
- Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* **98**, 5301–5317 (2014).
- Ben-Nissan, G. et al. Triple-stage mass spectrometry unravels the heterogeneity of an endogenous protein complex. *Anal. Chem.* **89**, 4708–4715 (2017).
- Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **6**, 281–294 (2014).
- Ben-Nissan, G. & Sharon, M. The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes. *Curr. Opin. Chem. Biol.* **42**, 25–33 (2018).
- D’Atri, V., Porrini, M., Rosu, F. & Gabelica, V. Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrometry* **50**, 711–726 (2015).
- van Oers, M. M., Pijlman, G. P. & Vlak, J. M. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *J. Gen. Virol.* **96**, 6–23 (2015).
- Manneberg, M., Friedlein, A., Kurth, H., Lahm, H. W. & 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.* **3**, 30–38 (1994).
- Beck, A., Wagner-Rousset, E., Ayoub, D., Van Dorsseleer, A. & Sanglier-Cianferani, S. Characterization of therapeutic antibodies and related products. *Anal. Chem.* **85**, 715–736 (2013).
- Liste-Calleja, L. et al. Hek293 as a recombinant protein factory: three different approaches for protein production. *BMC Proc.* **9**, P74 (2015).
- Huhn, C., Selman, M. H., Ruhaak, L. R., Deelder, A. M. & Wuhrer, M. IgG glycosylation analysis. *Proteomics* **9**, 882–913 (2009).
- Liu, H. et al. In vitro and in vivo modifications of recombinant and human IgG antibodies. *MAbs* **6**, 1145–1154 (2014).
- Beck, A., Sanglier-Cianferani, S. & Van Dorsseleer, A. Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Anal. Chem.* **84**, 4637–4646 (2012).
- Fuh, G. et al. Structure-function studies of two synthetic anti-vascular endothelial growth factor Fabs and comparison with the Avastin Fab. *J. Biol. Chem.* **281**, 6625–6631 (2006).
- Braden, B. C. et al. Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* **243**, 767–781 (1994).
- Newton, C. J. et al. Functional inactivation of the oestrogen receptor by the antiestrogen, ZM 182780, sensitises tumour cells to reactive oxygen species. *J. Endocrinol.* **161**, 199–210 (1999).
- Chan, F. K., Moriawaki, K. & De Rosa, M. J. Detection of necrosis by release of lactate dehydrogenase activity. *Methods Mol. Biol.* **979**, 65–70 (2013).
- Higel, F., Seidl, A., Sorgel, F. & Friess, W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* **100**, 94–100 (2016).
- Zhang, L., Luo, S. & Zhang, B. Glycan analysis of therapeutic glycoproteins. *Mabs* **8**, 205–215 (2016).
- Zhang, P. et al. Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discov. Today* **21**, 740–765 (2016).
- Fornelli, L., Ayoub, D., Aizikov, K., Beck, A. & Tsybin, Y. O. Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap Fourier transform mass spectrometry. *Anal. Chem.* **86**, 3005–3012 (2014).
- Kim, J. Y., Kim, Y. G. & Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* **93**, 917–930 (2012).
- Barnes, L. M., Bentley, C. M. & Dickson, A. J. Advances in animal cell recombinant protein production: GS-NS0 expression system. *Cytotechnology* **32**, 109–123 (2000).
- Lossl, P., van de Waterbeemd, M. & Heck, A. J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J.* **35**, 2634–2657 (2016).
- Sharon, M. Biochemistry. Structural MS pulls its weight. *Science* **340**, 1059–1060 (2013).
- Mehmood, S., Allison, T. M. & Robinson, C. V. Mass spectrometry of protein complexes: from origins to applications. *Annu. Rev. Phys. Chem.* **66**, 453–474 (2015).
- Ho, R. J., Gibaldi, M. Large-scale production of recombinant proteins. In: *Biotechnology and Biopharmaceuticals* (eds R. J. Ho and M. Gibaldi).
- Kirshenbaum N., Michaelevski I., Sharon M. Analyzing large protein complexes by structural mass spectrometry. *J. Vis. Exp.* **40**, e1954 (2010).
- Bush, M. F. et al. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82**, 9557–9565 (2010).

### Acknowledgements

We thank Ely Morag and Ed Bayer (Weizmann Institute of Science) for providing us with the pPICK9 plasmid for CBM3a protein expression. M.S. is 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 an incumbent of the Aharon and Ephraim Katzir Memorial Professorial Chair.

### Author contributions

G.B.N., S.V. and M.S. designed the experiments. GBN and SV performed the experiments. G.B.N., S.V. and M.S. analyzed the data with assistance from S.W. A.K., M.Y., T.U., Y.P., D.M., H.C.D., R.D. and S.J.F. G.B.N., S.V. and M.S. wrote and edited the paper.

**Additional information**

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s42003-018-0231-3>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018