Adding colour to mass spectra: Charge Determination Analysis (CHARDA) assigns charge state to every ion peak

Yaroslav Lyutvinskiy,¹ Konstantin O. Nagornov,² Anton N. Kozhinov,² Natalia Gasilova,³ Laure Menin,³ Zhaowei Meng,¹ Xuepei Zhang,¹ Amir Ata Saei,^{1,4} Yury O. Tsybin,² Alexander Makarov,⁵ Roman A. Zubarev^{1,6,7*}

¹Division of Physiological Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17 177 Stockholm, Sweden

²Spectroswiss, 1015 Lausanne, Switzerland

³Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

⁴Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

⁵ThermoFisher Scientific, Bremen, Germany

⁶SciLifeLab, Stockholm, Sweden

⁷Department of Pharmacological & Technological Chemistry, I.M. Sechenov First Moscow State Medical University, Moscow, Russia

*Correspondence and requests for materials should be addressed to R.A.Z. (email: Roman.Zubarev@ki.se)

Abstract

Traditionally, mass spectrometry (MS) output is the ion abundance plotted versus ionic mass-to-charge ratio m/z . While employing only commercially available equipment, Charge Determination Analysis (CHARDA) adds a third dimension to MS, estimating for individual peaks their charge states z , starting from $z=1$, and colour-coding z in m/z spectra. CHARDA combines the analysis of ion signal decay rates in the time-domain data (transients) in Fourier transform (FT) MS with the interrogation of mass defects of biopolymers. Being applied to individual isotopic peaks in a complex protein tandem (MS/MS) dataset, CHARDA facilitates charge state deconvolution of large ionic species in crowded regions, estimating z even in the absence of isotopic distribution (*e.g.*, for monoisotopic mass spectra). CHARDA is fast, robust and consistent with conventional FT MS and FT MS/MS data acquisition procedures. An effective charge state resolution $R_z \geq 6$ is obtained, with potential for further improvements.

Introduction

Traditionally, mass spectrometry (MS) measures mass-to-charge ratio m/z of ions, providing two-dimensional mass spectra, where the ion abundance is plotted against the ionic m/z . At the same time, the main parameters of interest are the compound's neutral mass m and its abundance. During the first seven decades of MS development, the absolute majority of mass spectra contained only singly charged ions, and with the experimental mass error being much larger than the electron mass, the m/z values were in practice equal to the neutral mass. The situation has changed with the advent of electrospray ionization (ESI) and similar techniques producing multiply-charged ions of polypeptides¹. In response to that, various approaches emerged aimed at estimating m from multi-charge MS data. Up to now, the two main strategies of charge state deconvolution are both based on using several ion peaks originating from the same molecule. In lower-resolution mass spectra of biopolymers, the m/z difference between different charge states is employed². In high-resolution mass spectra, where individual isotopic peaks are resolved, z can be deduced as the closest integer to $(1.003*k)/\Delta m/z$, where $\Delta m/z$ is the distance on the m/z scale between the N -th and $(N+k)$ th peaks in an isotopic cluster, with N being a positive integer. However, in crowded mass spectra the isotopic clusters often overlap, making it difficult to determine which ones of the multitude of closely spaced ion peaks belong to the same isotopic cluster³.

To address this issue, several methods have emerged to estimate the charge states of individual ion peaks independently from the other ions in a mass spectrum. Most of these methods are based on measuring the charge of individual ions. An early approach used a secondary ion detector, in which the number of secondary ions created by an incident analyte ion depended upon the ion kinetic energy. As the kinetic energy of an ion accelerated by a potential difference ΔU is $z\Delta U$, the number of secondary ions depends upon the ionic charge z ⁴. Similarly, a time-of-flight cryodetector could measure the kinetic energy of the ions impacting the detector, which correlated with the ionic charge⁵. Charge detection mass spectrometry (CDMS) measures the charge of individual ions by the signal induced by them on an image current detector that records a time-domain signal, or transient. CDMS then applies the Fourier transform (FT) or a similar approach to these time-domain signals to derive z and m/z of ions⁶. The recent technique called individual ion MS (I^2 MS) implements CDMS in commercially available OrbitrapTM mass spectrometers⁷. In the I^2 MS approach, ≤ 120 individual ions are obtained per spectrum acquisition, with thousands of acquisitions obtained to generate a full mass spectrum. The ionic charge state is estimated based on the

fact that ion peak's abundance increases with time-domain transient duration in a charge-state dependent manner. I²MS can assign the charge state even in the absence of isotopic resolution, which is particularly useful for very large ions, such as protein complexes. On the other hand, many thousands of acquisitions may be required to generate the charge calibration curve by the I²MS approach, and it is limited by the lowest charge state $z \approx 10$. Increasing time-domain transient duration in I²MS may help to improve the charge state limit of detection, albeit at the expense of a proportionally increased experimental time. In measuring $z=22+$ ions, Full Width at Half Maximum (FWHM) charge state resolving power $R_z \approx 2$ has been obtained ⁷, with higher charge resolution for higher charge states.

Here we present the Charge Determination Analysis (CHARDA) method that is applicable in FT MS for all charge states starting from $z=1$. In CHARDA, a calibration curve can be obtained from a single mass spectrum acquired with commercially available FTMS equipment in a conventional data acquisition mode, allowing for post-processing already acquired MS and MS/MS data. While not requiring single ion detection, or detection of very small number of ions per mass spectrum, CHARDA still adds a third dimension to MS by estimating the charge state z for every ion peak in the FT mass spectrum. For compact representation of 3D information, the ion charge state is displayed by colour-coding the ion peaks.

Method

CHARDA combines the complementary approaches of time-domain transient decay analysis (TDA) characteristic for FTMS with mass defect interrogation (MDI) that can be implemented using any high-resolution mass spectra (Figure 1). TDA utilizes the well-known fact that in FTMS, the time-domain transient, theoretically being nearly sinusoidal, is actually more complex, often close to a sinusoidal wave with exponentially decaying amplitude (Figure 1, upper left). The decay is usually attributed to collisions with the residual gas ⁸. The exponent parameter known as decay constant $1/\tau$ is thus linked to the collisional cross section (CCS) ⁹. For ions with a similar m/z , larger mass (and thus also larger charge) tends to imply larger CCS, and thus faster transient decay. Indeed, the ions in the m/z interval 772-780 have the following experimental CCS values: ubiquitin 11+ (8.6 kDa) - 2394 Å², whereas apomyoblobin 22+ (17.0 kDa) - 4887 Å² ¹⁰. In ion mobility MS, very similar CCS values (2340 and 4920 Å², respectively) have been obtained for these ions ¹¹. Thus we concluded that decay constants in FTMS must carry charge-predictive information. Indeed, using the

linear regression of the above values for 11+ ubiquitin and 22+ apomyoblobin, the cross section of $3528 \text{ \AA}^2$ corresponds to the charge state $z=16.004\pm0.010$, a very accurate estimate for the 16+ cytochrome C ions with such an experimental CCS value determined by FTMS¹⁰.

Note that the charge state being a discrete value, the accuracy of its estimation in the above example is greatly excessive. Monte Carlo simulations showed that, as long as the standard deviation of both calibration CCSs is kept below 2.4%, more than 95% of the charge state estimates approximated to the nearest integer were correct, i.e. 16+. With a $\leq 6.7\%$ standard deviation, >95% cases gave the correct charge state within a ± 1 charge state margin, and with a $\leq 10.6\%$ standard deviation - within ± 2 charge state units. Even such approximate charge state assignment can be useful in practice.

In TDA, the transient is first converted using the Fast Fourier Transform (FFT) to a frequency domain spectrum, and the latter – to an m/z spectrum (Figure 1, 1). The ion peaks are distinguished from the background by an appropriate algorithm. Narrow m/z areas around every individual peak are then converted to a time-domain partial transient using inverse FFT, and the decay parameter of this partial transient is estimated. There are several known ways to estimate the decay constant $1/\tau$, including STORI plots¹² and FIR filtering¹³, but we have found experimentally that fitting a decaying exponent $\exp(-t/\tau)$ to the time interval of 0.2-1.0 s (more general, from 1/8 to 5/8 of the transient) provides the best charge resolution.

In parallel, FFT of the two halves of the full transient, A and B, is performed, each half-transient providing an ion peak at the same position as with full transient but at half the m/z resolution. The ratio of the intensities of the peak in FFT(B) versus FFT(A), $\eta = \frac{FFT(B)}{FFT(A)}$ (Figure 1, 3), the decay constant $1/\tau$, the peak m/z and abundances are then combined in a single z -predictive TDA model.

This model is calibrated as follows. First, a conventional algorithm of charge state deconvolution based on the isotopic distribution of ions is applied to the whole m/z spectrum, generating a list of identified isotopic clusters with reliably determined charge states. These usually encompass <10% of the ion peaks in a mass spectrum. Then a logistic regression model is built based on these reliably identified charge states predicting z from the transient parameters for any ion peak.

The position of ions on the m/z scale also carries information of the molecular mass, and thus about the charge state of ions z . A large fraction of all FTMS analyses performed to

date concerns polypeptides, and thus our discussion will be limited to these biopolymers. The mass defect interrogation (MDI) investigates the ionic mass defect d_m , which is defined here as a difference between the accurate neutral mass of a compound and its nearest lower-mass integer approximation: *e.g.*, mass defect of $m=2083.8753$ Da is $d_m=0.8753$ Da. For monoisotopic polypeptide molecular masses, mass defect is a periodic function with a period of 1.9-2.2 kDa¹⁴⁻¹⁷. MDI uses the well-known fact that the monoisotopic molecular mass defects are not evenly distributed over the mass scale but are clustered around their average values¹⁴. For instance, for the monoisotopic polypeptide masses between 1000 and 1001 Da, the mass defect distribution is centered around 0.55 Da, with a majority of molecules falling within the ± 0.1 Da interval around this value¹⁶. For other isotopic peaks of proteins, mass defects increase by ≈ 3 mDa for every Dalton away from the monoisotopic peak. This is because ^{13}C is the main contributor to the isotopic distribution, with the mass difference between ^{13}C and ^{12}C of ≈ 1.0033 Da, while the presence of other elements makes the average inter-isotope distance in polypeptides closer to 1.003 Da.

Since the difference between the average isotopic mass and monoisotopic one increases by 1 Da for every 1.6-1.8 kDa of molecular mass increase¹⁴, and the mass of the most abundant isotopologue is always ≤ 1 Da below the average isotopic mass¹⁸, for polypeptides below 25 kDa the most abundant isotopologue has a mass defect within 0.05 Da of the monoisotopic mass defect (disregarding the discontinuity occurring when d_m reaches 1.0). Therefore, knowing the theoretical distribution of mass defects as a function of polypeptide mass, one can obtain a probability estimate for different charge states of that mass. As an example, for a polypeptide ion with m/z 1000.200, the charge state 1+ is highly unlikely, as the mass defect of 200 mDa falls within the forbidden “band gap” between the mass distribution peaks. According to Mann¹⁵, the molecular mass at which the interval encompassing $\geq 95\%$ of all amino acid compositions exceeds 1 Da is ≈ 8 kDa. Thus around and beyond that critical mass the band gap disappears and the distribution of monoisotopic protein masses becomes pseudo-continuous (there are still gaps on a micro-scale). However, the differences between the regular maxima and minima of the distribution remain significant for much larger masses, and thus MDI analysis is still plausible.

Molecular mass defects have previously been used to estimate the probability of a given mass to belong to a peptide¹⁹ or to a modified peptide²⁰. But the use of d_m to augment charge state determination in polypeptides appears to be novel²¹, while the link between the mass defect and the charge state for synthetic polymers has recently been explored²². In our

implementation MDI uses not only monoisotopic molecular masses, but also poly-isotopic masses as charge state is predicted for any ion peak regardless of its position in the isotopic cluster.

When CHARDA is applied to an individual m/z peak resolved in a mass spectrum (Figure 1, 4), both the TDA model and the MDI model predict the ion's charge state and its probability density distribution (Figure 1, 5-6). The MDI model is universal and only needs to be adjusted if a different class of molecules than unmodified polypeptides is considered (*e.g.*, specifically modified polypeptides, polymers, RNAs, sugars, *etc.*). To create such a model from scratch, theoretical mass distribution needs to be simulated for a given class of molecules, including not only monoisotopic masses but also those of other isotopologues. The discrete nature of this distribution²³ can cause inconveniences, and thus the distribution has to be averaged with a narrow window (*e.g.*, 0.01 Da), providing smooth periodic envelopes. The envelopes have distinct peaks at low molecular masses with “band gaps” between them. The abundance of a peak at each mass unit is normalized, so that its area is unity. Such normalization converts the mass distribution into a mass-dependent probability distribution function. The probability of a certain discrete z for a specific m/z value is established assuming that the ionic charge is due to protonation. Therefore, the value of $(m/z - m_p) \cdot z$ with its error band ($m_p=1.007277$ is proton's m/z) is mapped onto the probability distribution, and the area under the curve (AUC) is established. In a similar manner, AUCs are obtained for other plausible z values, and all thus obtained AUCs are normalized such that their sum is unity. In the MDI model, thus normalized AUCs represent the probabilities for a given m/z to correspond to respective z values.

The TDA and MDI probability charge state density distributions for a given ion peak are merged, giving rise to a single distribution (Figure 1, 7). The most probable charge state value (not necessarily an integer) as well as the interval of charge state values corresponding to >95% certainty ($p<0.05$) are thus determined. This charge state value is then colour-coded into the ion peak envelope in the mass spectrum, providing a third dimension of information (Figure 1, 8).

RESULTS

Proof of principle

As a proof of principle, we analyzed an MS/MS spectrum of ubiquitin 11+ ions acquired on a Q Exactive HF Orbitrap (Thermo) as a sum of 350 individual transients, each 1.5 s duration

²⁴. The mass spectrum contains 33,796 ion peaks, of which 7208 peaks (21 % of the total) representing 2323 isotopic clusters were charge-state deconvolved using the Hardklör algorithm ²⁵. In total 1167 peaks in 212 clusters were recognized as being due to ubiquitin b/y product ions, and 533 most reliable of them (see Methods section) representing 159 isotopic clusters with z ranging from 1+ to 9+ were used for the TDA model calibration. Figure 2a demonstrates that the TDA-extracted decay constants scale linearly with the monoisotopic molecular mass of the calibration ions. Figure 2b shows the distribution of the CHARDA-predicted charge states of these peaks and their Gaussian fits versus the charge state determined by the Hardklör algorithm.

In order to optimize the CHARDA algorithm, a quantitative measure of its performance was needed. Such a measure was charge state resolution determined as the maximum value of the predicted charge state divided by FWHM of the fitted bell curves as in Figure 2b averaged over all charge states. The thus determined TDA resolution was 5.6, while the addition of MDI improved the resolution to 6.0 (Figure S1). As expected, most of the improvements were observed for the lower charge states, starting from $z=1$.

Figure 3a shows the application of CHARDA to a busy region of the MS/MS spectrum between m/z 811.0 and 813.5. The colour-coded peaks allow one to easily discern separate isotopic clusters. To reduce the complexity, the spectrum can be divided into three subspectra of overlapping charge state intervals, e.g. $z \leq 4$, $3 \leq z \leq 6$ and $z \geq 5$ (Figure 3b-d).

Hardklör deconvolution of the whole MS/MS spectrum presented in an expanded view in Figure 3 gave 212 isotopic clusters. To estimate the effect of CHARDA on isotopic deconvolution, the entire MS/MS spectrum was divided into eight subspectra, one for every predicted charge state, with peaks having a >1% probability to be in a given charge state included in the corresponding subspectrum. In each subspectrum, Hardklör deconvolved the isotopic clusters, producing additional 19 isotopic clusters (a 9% improvement) that were missed in the full-spectrum deconvolution. The masses of all these clusters corresponded to expected fragment ions of ubiquitin. This result confirms that CHARDA can improve isotopic cluster deconvolution for busy tandem mass spectra.

Manual spectra inspection

Another possible CHARDA application is the manual inspection of mass spectra with overlapping charge state distributions of molecular or fragment ions. This is particularly useful when proteins exist in different proteoforms ²⁶, and thus there is no clear “charge state ladder” in a mass spectra. As an example, Figure 4a shows a colour-coded mass spectrum

with a 1 s transient duration of a mixture of three proteins. Colour-coding was made by TDA without accurate calibration of charge states, by instructing the algorithm to recognize different charge states using the maximum likelihood approach. For clarity the charge state ladders are connected by lines. While many peaks obviously do not belong to any ladder, their tentative attribution to this or that protein molecule can be facilitated by peak colour coding. This attribution can be verified by other approaches, such as charge state deconvolution by Hardklör or MS/MS.

Absence of isotopic resolution

In the absence of isotopic resolution, the ion peak represents a weighted average of the whole isotopic envelope. We performed numerous CHARDA on a transient shortened from to 0.5 s that gave isotopic distributions merged into a single peak, but were unable to predict charge states correctly at such conditions. Thus, unlike I²MS, CHARDA requires isotopic resolution for charge state determination.

Charge state of individual isotopologues

Close inspection of the CHARDA-predicted z values for individual isotopologues revealed a systematic increase with the isotopologue distance from the monoisotopic peak (see as an example ubiquitin y_{45}^{6+} ion in Figure 4b). To verify this effect, we split the isotopic distributions of the peaks with $z=5+$, $6+$, and $7+$ with similar signal-to-noise ratios into the left and right halves (Figure S2a) and measured the average decay rate for the ion peaks in each half separately. For all three charge states, the right (heavier) parts showed significantly faster decay (Figure S2b). The origin of this effect could be the increased multiplicity of the isotopic fine structure (IFS) for peaks more distant from the monoisotopic one²⁷. Indeed, a mixture of two close frequencies with equally decaying amplitude gives a pattern with a complex amplitude envelope (Figure 4c), from which the extraction of the original decay constant is nontrivial²⁸. We calculated the IFS of the 5th isotopic peak of the y_{24} ubiquitin fragment and simulated its transient as a weighted sum of the equally decaying exponents of the individual IFS components (Figure 4d). The TDA of that summed transient gave a 47% larger decay constant than that of each component. As an effective compensation method for this artifact has not been found so far²⁸, we hypothesized that CHARDA should work best for the molecules with dominant monoisotopic masses.

MS/MS of monoisotopic proteins

Proteins are large molecules, with an average size of bacterial proteins >30 kDa, while the monoisotopic ion peak is usually indiscernible in the noise at MW>10 kDa. Thus to create proteins with abundant monoisotopic ion peaks and test CHARDA applicability on them, we grew *E. coli* bacteria in minimal media with 3-30 fold depleted ^2H , ^{13}C , ^{15}N and ^{18}O isotopes, as well as in normal media for control. At the same amount of protein loaded, LC-MS of the soluble proteome gave 3-5 times more abundant signal for the depleted proteins, as the signal concentrated in fewer isotopic peaks (Figure 5a, S3). This resulted in 283 proteoforms selected for MS/MS in depleted proteome versus 120 proteoforms for normal proteome. Hardklör could not process depleted proteome data in the automatic mode, and had to be supplemented with manual analysis. At least nine proteoforms were identified by MS/MS in both normal and monoisotopic datasets, and 5 proteoforms with 413 monoisotopic peaks in total were used for CHARDA calibration. Consistent with our expectations, CHARDA provided higher resolution (Figure 5b, c), namely $R\approx 5.4$ with both TDA and MDI, and 4.5 with TDA only for the monoisotopic proteome versus $R\approx 3.2$ and 3.4, respectively, for the normal proteome.

After calibration, CHARDA could easily identify the charge states of ions for which conventional deconvolution algorithms provided no results. In total, CHARDA identified 1120 backbone fragments of 5 proteins and estimated charge states for them in the monoisotopic MS/MS dataset, providing on average 81% sequence coverage (Table S1). Of these, only 321 fragments were identified by Hardklör in the corresponding full-isotope MS/MS dataset, giving on average 41% sequence coverage. The performance of the CHARDA models for the normal (656 peaks) and monoisotopic (413 peaks) MS/MS data are presented in Fig 5b-c, providing $R=3.4$ for the normal and $R=5.4$ for the monoisotopic case.

Simplified CHARDA

Full-scale CHARDA analysis requires inverse FT of hundreds and thousands of individual peaks, which is calculation-intensive. Thus we tested the simplified CHARDA version (sCHARDA, Figure 6) which compared to the normal FTMS spectrum processing adds to the normal FT procedure only FTs of the two half-transients, maximum doubling the FT calculation time. Yet the charge state resolution obtained with sCHARDA is comparable with that of the full-scale CHARDA ($R\approx 4.0$ for ubiquitin 11+ MS/MS spectrum in Figure 3), which is sufficient in many cases.

CONCLUSIONS

Here we demonstrated that CHARDA can determine charge states of individual ion peaks in the presence of the isotopic resolution and/or isotopic distribution. CHARDA has no limitation on the lowest charge state unlike other methods^{7,26} and requires no special acquisition technique except for transient recording. The colour-coding of the ionic charge states resulting from CHARDA provides a 3rd dimension to mass spectra, facilitating analysis of crowded MS and MS/MS data and hinting at possible peak interferences and overlaps. The simplified CHARDA version, sCHARDA, can be implemented in a real-time, online with data acquisition on most FT MS instruments.

Acknowledgements

This work was supported by the Swedish Research Council (grants 2017-04303 to RAZ and 2020-00687 to AAS), as well as EU Horizon 2020 program (FET OPEN project TopSpec, grant agreement 829157).

Author contributions

RAZ has conceived the study and drafted the manuscript; YOT and AM contributed to the concept and implementation; YL implemented CHARDA, created models and software and processed the data, as well as created figures together with AASL; KON and ANK performed transient acquisition and data processing, as well as provided modeling; NG and LM collected LC-MS data, XZ generated monoisotopic samples; ZM provided technical support. All authors critically reviewed the manuscript.

Data Availability statement

All the data files underlying the figures are available with CodeOcean computing capsule [DOI identifier will be added upon potential acceptance of the manuscript].

Code Availability statement

The source code implementing CHARDA is provided as CodeOcean capsule [submitted directly to Nature Communications Editor for the peer-review process] with no restrictions. The same capsule contains all necessary data and source code for application of CHARDA to form the article figures.

333 **Competing interests**

334 The authors declare the absence of competing interests.

335 **Materials & Correspondence**

336 Correspondence and material requests should be addressed to R.A.Z.

337 (Roman.Zubarev@ki.se).

References

1. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* (1989) doi:10.1126/science.2675315.
2. Covey, T. R., Bonner, R. F., Shushan, B. I., Henion, J. & Boyd, R. K. The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid Commun. Mass Spectrom.* (1988) doi:10.1002/rcm.1290021111.
3. Horn, D. M., Zubarev, R. A. & McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* (2000) doi:10.1016/S1044-0305(99)00157-9.
4. Zubarev, R. A., Bitensky, I. S., Demirev, P. A. & Sundqvist, B. U. R. Multiplicities of secondary ions emitted in macromolecular ion impact events. *Nucl. Inst. Methods Phys. Res. B* (1994) doi:10.1016/0168-583X(94)96094-1.
5. Wenzel, R. J., Matter, U., Schultheis, L. & Zenobi, R. Analysis of megadalton ions using cryodetection MALDI time-of-flight mass spectrometry. *Anal. Chem.* (2005) doi:10.1021/ac0482054.
6. Fuerstenau, S. D. & Benner, W. H. Molecular weight determination of megadalton DNA electrospray ions using charge detection time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* (1995) doi:10.1002/rcm.1290091513.
7. Kafader, J. O. *et al.* Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nat. Methods* (2020) doi:10.1038/s41592-020-0764-5.
8. Wobschall, D., Fluegge, R. A. & Graham, J. R. Collision cross sections of hydrogen and other ions as determined by ion cyclotron resonance. *J. Chem. Phys.* (1967) doi:10.1063/1.1701581.
9. Yang, F., Voelkel, J. E. & Dearden, D. V. Collision cross sectional areas from analysis of fourier transform ion cyclotron resonance line width: A new method for characterizing molecular structure. *Anal. Chem.* (2012) doi:10.1021/ac300379a.
10. Sanders, J. D. *et al.* Determination of Collision Cross-Sections of Protein Ions in an Orbitrap Mass Analyzer. *Anal. Chem.* (2018) doi:10.1021/acs.analchem.8b00724.
11. 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.* (2010) doi:10.1021/ac1022953.
12. Kafader, J. O. *et al.* STORI Plots Enable Accurate Tracking of Individual Ion Signals. *J. Am. Soc. Mass Spectrom.* (2019) doi:10.1007/s13361-019-02309-0.
13. Mao, L. *et al.* Collision cross section measurements for biomolecules within a high-resolution fourier transform ion cyclotron resonance cell. *Anal. Chem.* (2015) doi:10.1021/acs.analchem.5b00102.
14. Pomerantz, S. C. & McCloskey, J. A. Fractional mass values of large molecules. *Org. mass Spectrom.* **22**, 251–253 (1987).

15. Mann, M. Useful tables of possible and probable peptide masses. in *PROCEEDINGS OF THE ASMS CONFERENCE ON MASS SPECTROMETRY AND ALLIED TOPICS* 639 (AMERICAN SOCIETY FOR MASS SPECTROMETRY (ASMS), 1995).
16. Zubarev, R. A., Håkansson, P. & Sundqvist, B. Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements. *Anal. Chem.* (1996) doi:10.1021/ac9604651.
17. Zubarev, R. A. & Bonddarenko, P. V. An A priori relationship between the average and monoisotopic masses of peptides and oligonucleotides. *Rapid Commun. Mass Spectrom.* (1991) doi:10.1002/rcm.1290050606.
18. Zubarev, R. A., Demirev, P. A., Håkansson, P. & Sundqvist, B. U. R. Approaches and Limits for Accurate Mass Characterization of Large Biomolecules. *Anal. Chem.* (1995) doi:10.1021/ac00116a028.
19. Toumi, M. L. & Desaire, H. Improving mass defect filters for human proteins. *J. Proteome Res.* (2010) doi:10.1021/pr100291q.
20. Bruce, C., Shifman, M. A., Miller, P. & Gulcicek, E. E. Probabilistic enrichment of phosphopeptides by their mass defect. *Anal. Chem.* (2006) doi:10.1021/ac060046w.
21. Sleno, L. The use of mass defect in modern mass spectrometry. *J. Mass Spectrom.* (2012) doi:10.1002/jms.2953.
22. Ishitsuka, K., Kakiuchi, T., Sato, H. & Fouquet, T. N. J. An arsenal of tools based on Kendrick mass defects to process congested electrospray ionization high-resolution mass spectra of polymers with multiple charging. *Rapid Commun. Mass Spectrom.* (2020) doi:10.1002/rcm.8584.
23. Conrads, T. P., Anderson, G. A., Veenstra, T. D., Paša-Tolić, L. & Smith, R. D. Utility of accurate mass tags for proteome-wide protein identification. *Anal. Chem.* (2000) doi:10.1021/ac0002386.
24. Nagornov, K. O., Zennegg, M., Kozhinov, A. N., Tsybin, Y. O. & Bleiner, D. Trace-Level Persistent Organic Pollutant Analysis with Gas-Chromatography Orbitrap Mass Spectrometry-Enhanced Performance by Complementary Acquisition and Processing of Time-Domain Data. *J. Am. Soc. Mass Spectrom.* (2020) doi:10.1021/jasms.9b00032.
25. Hoopmann, M. R., MacCoss, M. J. & Moritz, R. L. Identification of peptide features in precursor spectra using hardklör and krönik. *Curr. Protoc. Bioinforma.* (2012) doi:10.1002/0471250953.bi1318s37.
26. Smith, L. M. & Kelleher, N. L. Proteoform: A single term describing protein complexity. *Nature Methods* (2013) doi:10.1038/nmeth.2369.
27. Yergey, J. A. A general approach to calculating isotopic distributions for mass spectrometry. *Int. J. Mass Spectrom. Ion Phys.* (1983) doi:10.1016/0020-7381(83)85053-0.
28. Hofmann, A. E. *et al.* Using Orbitrap mass spectrometry to assess the isotopic compositions of individual compounds in mixtures. *Int. J. Mass Spectrom.* (2020) doi:10.1016/j.ijms.2020.116410.

FIGURE CAPTIONS

Figure 1. Charge determination analysis (CHARDA): 1. In transient decay analysis (TDA), time-domain signal (transient) obtained from an FT mass analyser is converted by FFT to a frequency spectrum. 2. The latter is converted to an m/z scale spectrum using known ion peaks or external calibration. 3. Narrow m/z areas around individual peaks are converted to time-domain transient using inverse FFT. The TDA model is calibrated with parameters of the transients and charge states of the known ions. 4. Peak-by-peak TDA of the whole mass spectrum is performed, and for each peak a charge state probability density distribution is obtained. 5. This distribution is then transferred to mass defect interrogation (MDI). 6. MDI provides its own charge state probability density distribution. 7. The TDA and MDI charge state probability density distributions are merged, with the most probable charge state z_0 and the charge interval $\Delta z_{p<0.05}$ corresponding to $p<0.05$ determined. 8. The peak is labelled with $z_0 \pm \Delta z_{p<0.05}$ and colour-coded by the CHARDA-estimated charge.

Figure 2. TDA proof of principle. **a**, The TDA-extracted decay constants of the calibration ions of ubiquitin 11+ fragments versus their molecular mass. **b**, The distribution of the CHARDA-predicted charge states of the calibration peaks and their Gaussian fits versus the charge determined by the deconvolution algorithm (Hardklör).

Figure 3. CHARDA of a busy MS/MS spectrum region. Charge prediction for individual ion peaks **(a)**. Subsets of ion peaks with predicted charges $z \leq 4$ **(b)** $3 \leq z \leq 6$ **(c)**, and $z \geq 5$ **(d)**.

Figure 4. CHARDA implementation and features. **a**, Manual inspection of busy mass spectra of polypeptide mixtures can be facilitated by CHARDA color coding, with charge ladders connected by lines for better visibility. CAH – carbonic anhydrase; MYO – myoglobin; NEO – unidentified protein, possibly a hydrolysis product. **b**, A monotonous increase of the CHARDA-predicted charge state for high order isotopologues (ion peaks more distant from the monoisotopic peak). **c**, An interference of two close frequencies with equally decaying amplitude gives a pattern with a complex amplitude envelope. **d**, The isotopic distribution of the y_{24} ubiquitin fragment and the calculated isotopic fine structure of its 5th isotopic peak.

Figure 5. Comparison of CHARDA performance in top-down analysis of *E. Coli* bacteria grown in the normal and isotopically depleted (monoisotopic) media. a, An example of a 6+ molecular ion. b, Performance of CHARDA (TDA+MDA) models for MS/MS fragments of 5 different proteins for the full isotope dataset, R=3.4. c, The same for the monoisotopic dataset, R=5.4.

Figure 6. A simplified CHARDA (sCHARDA) version. Compared to the normal FT MS spectrum processing, sCHARDA adds only FTs of the two half-transients to the normal FT procedure, which maximum doubles the FT calculation time.

SUPPLEMENTARY MATERIALS

Table S1. Comparison of top-down LC-MS/MS analysis of E. Coli proteome with normal isotopic composition (full isotope) and isotopically depleted (monoisotopic).

		Full Isotope spectra			Monoisotopic spectra - CHARDA	
Proteoform	Mono-isotopic Mass, Da	Clusters	Nr. of fragments	Sequence coverage	Nr. of fragments	Sequence coverage
PFR244733: Cold shock-like protein CspC	7266.72	393	38	39%	199	91%
PFR244781: DNA-binding protein HU-alpha	9529.19	494	75	54%	197	81%
PFR279339: Protein YciI	10595.42	211	24	20%	133	67%
PFR244833: DNA-binding protein HU-beta	9219.99	525	110	54%	295	86%
PFR245160: 50S ribosomal protein L7/L12	12198.48	448	74	40%	296	82%

Supplementary Figures

Figure S1. The CHARDA (TDA + MDI) model based on ubiquitin 11+ MS/MS data; the charge resolution is 6.0.

Figure S2. a, An example of splitting the isotopic distribution of the y_{37}^{6+} ubiquitin ions into the left and right halves; **b**, decay rate comparison for the left and right halves of the isotopic distributions of ions with $z=5+$, $6+$ and $7+$. Two-sided Student's t-test for equal variance samples (Center line, median; box limits contain 50%; upper and lower quartiles, 75 and 25%; maximum, greatest value excluding outliers; minimum, least value excluding outliers; outliers, more than 1.5 times of upper and lower quartiles).

Figure S3. Comparison of the same part of the MS/MS spectrum of *E. Coli* cold shock-like protein CspC from (a) the monoisotopic proteome and (b) the normal (full-isotope) proteome.