Charge Assignment of Isotopically Resolved Direct Mass Data by 2D Voting

Ping F. Yip[1], Jared O. Kafader[2], Pei Su[2], John P. McGee[2], Philip D. Compton[2], Michael W. Senko[1] and Neil L. Kelleher[2]

[1]Thermofisher Scientific, 355 River Oaks Pkwy, San Jose, CA 95134

[2]Departments of Molecular Biosciences, Chemistry, Chemical and Biological Engineering, and the Feinberg School of Medicine, Northwestern University, Evanston, IL, 60208

ABSTRACT

Purpose: A 2D voting algorithm is developed to reduce the variability (~5% CV) of charge estimates by the Direct Mass Technology Mode to produce unique charge assignments for isotopically resolved data.

Methods: For isotopically resolved data, "charge related" ions – ions of neighboring isotopes, and charge states, are connected in the 2D data space of (m/z, Z). A voting method, making use of this precise connection among the charge related ions, is developed. Incorrect charge assignments from the initial estimates will be iteratively voted out, resulting in unique charge assignments.

Results: Three Direct Mass data sets; SILu™Lite SigmaMAb Universal Antibody Standard (SILu, MilliporeSigma, Burlington, MA) (native and denatured) and native Enbrel® (Etancercept), were analyzed by the 2D voting method. After 8 iterations, unique charge assignments were achieved. The charge assigned data result in isotopically resolved mass spectra, capturing not just the main profile of the analytes but also other less abundant analytes as well.

INTRODUCTION

The Direct Mass Technology mode adds an estimate of the charge state for each ion in addition to its m/z value.[1] Specifically, a STORI plot is constructed, tracking the evolution of an ion during acquisition in an Orbitrap analyzer. The slope of the STORI plot is proportional to the charge of the ion. After calibrating against slopes of known charge states, an estimate of the charge state is arrived at. Analyzing all the single ions over multiple spectra yields a 2D data set of m/z vs Slope (and after calibration, a putative charge state Z). The process of data generation is illustrated in Figure 1. Trajectory variation within the Orbitrap leads to approximately 5% CV in the charge estimate for ions with the same m/z and charge state. In order to obtain the molecular weight (MW) of an ion unambiguously, a "charge assignment" step is necessary to remove the 5% uncertainty to obtain a unique charge state for the ion (Figure 2). We describe here a 2D voting method for the step of charge assignment.

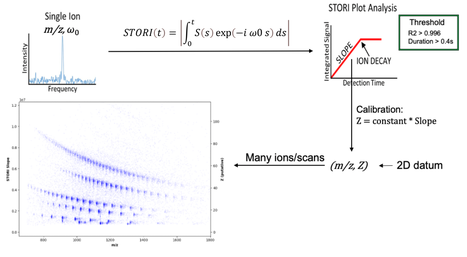
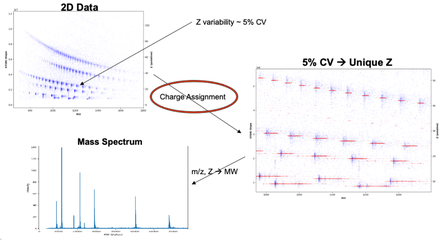


Figure 2. Charge Assignment--- removing uncertainty in charge estimate



MATERIALS AND METHODS

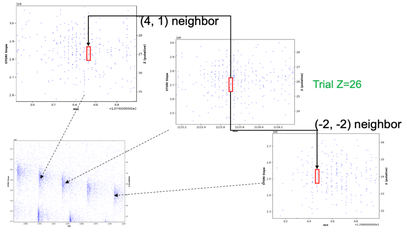
Algorithm Description

The voting algorithm starts with binning the 2D (m/z, Z) data. Here Z stands for the nearest integer charge state obtained by applying the calibration regression to the STORI slope, m/z is the centroid mass to charge of a single ion peak. The bin width in the m/z dimension is between 1-5 ppm and the Z dimension 1.

Charge Related Neighbor --- for a bin with coordinate (m/z^0, Z^0), and a trial charge state Z', we can obtain the coordinates (m/z^1, Z^1) of its (m(isotope), n(charge)) neighbor by Equation 1.

$$(m/z)^1 = 1.007 + (m/z)^0 + m * 1.003 / Z' - 1.007) * Z' / (Z' + n);\ Z^1 = Z^0 + n \qquad \textbf{Equation 1.}$$

Figure 3. Charge Related Neighbors. For the bin shown in the middle, at the trial charge state of 26, its (4, 1) and (-2, -2) neighbors are shown. A different trial charge state will define neighbor with different coordinates.



The voting proceeds stepwise as follows,

1. Define the size of charge related neighborhood (M, N), typically (10, 2).
2. Each bin (m/z^0, Z^0) starts with equal probability of possible trial Z's (+/- 5% of Z^0)
3. For each trial Z, cast a vote equal to the bin occupancy times its probability, for the charge state Z + n, for each (m, n) of (2M+1) x (2N+1) neighbors. m: (-M, -M+1,.. +M), n: (-N, -N+1,.. +N)
4. Tally up all votes for each trial Z, for each ion.
5. Keep top 3 most likely trial Zs. Normalize the votes to probability (summing to 1).
6. Iterate steps 2 to 4 (6-9 iterations)
7. Assign the trial Z with the highest probability as the final charge state. (usually filter out winners with low probability (<0.5))

Sample Preparation

SILu and Enbrel were buffer exchanged into 100mM ammonium acetate (AmAc) using Amicon Ultra-0.5 centrifugal filter unit (100 kDa MWCO, MilliporeSigma, Burlington, MA). The experimental details have been described elsewhere.[2] Briefly, the filter unit was first equilibrated with 500 µL water and spun for 3 min at 11,000 × g. 1 mg SILu was dissolved in 100 mM AmAc, loaded into the filter unit, and spun at 4° C for 3 min at 11,000 × g for 15 times. In each time, the solution was spun down to <100 µL and the filter unit was filled up to 500 µL with AmAc. The final solution was concentrated to 30 µM, aliquoted, and stored in -80 ° C freezer before further analysis.
MS analysis was performed using a Thermo Scientific™ Q Exactive™ MS with Ultra High Mass Range (UHMR).[1] Specifically, buffer-exchanged SILu and Enbrel were diluted to 1µM and 200 nM, respectively, in AmAc and loaded into a static NSI tip (Thermo Fisher Scientific). For SILu, 100 mM AmAc or acetonitrile/water/acetic acid = 40%/59%/1% were used as the solvent for native and denatured conditions, respectively. The source voltage was held at 1.1-1.4 kV and the inlet capillary was heated to 300° C. Relevant MS instrument settings are as followings: RF level: 150%; in-source trapping: off; averaging 0; eFT off; extended trapping: 15 eV, HCD pressure was between 0.1 and 0.5. STORI files were collected and saved directly through the Tune/UHMR software interface. Injection times were manually tuned down to the individual ion regime on a per sample basis. Individual ion spectra were collected between 45 – 90 minutes to acquire enough individual ions to fill out isotopic channels in the output MASS spectrum.

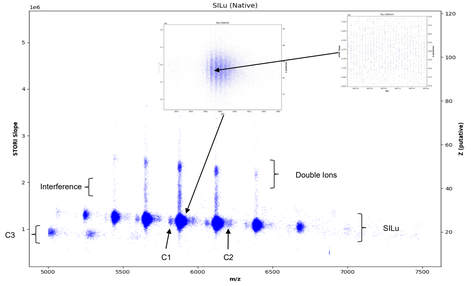
*A prototype version of the voting algorithm, written in Python, was used to analyze the data. The algorithm is incorporated in the STORIboard software.

RESULTS

1. SILu (Native)

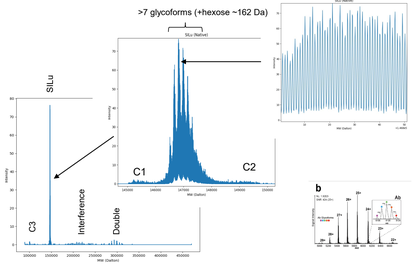
Approximately 2,000 spectra were acquired resulting in about 100,000 ions passing the threshold criterion (see Figure 1.)

Figure 4. Scatter plot of m/z vs STORI slope. Secondary y axis shows the putative charge state after calibration regression. Successive zooms reveal finer and finer details (glygoforms to isotopes). Minor components C1, C2 and C3 are also evident.



After eight iterations of voting, ions with assigned charge states passing the probability threshold of 0.5 are kept. The resulting molecular weight distribution is shown in Figure 5.

Figure 5. Molecular Weight Distribution after charge assignment by voting. At least 7 glycoforms, consistent with additions of hexose, are observed. All glycoform peaks are isotopically resolved. As a comparsion, insert b (Fig S4b, Reference 2), shows the low-resolution distribution from an ensemble data set.



2. SILu (Denatured)

Approximately 500,000 ions (from ~4000 spectra) passed the threshold criterion. The data (Figure 6) shows much higher complexity than that of the SILu native data set. More than 30 charge states of SILu, with significant overlapping distributions, are discernible. Other components (c1, c2 and c3) are evident. Charge states of component c1 also overlap with those of SILu. Charge assignment by voting was carried as in SILu Native.

Figure 6. First zoom reveals different glycoforms. Isotopic separation is less evident. Minor components C1, C2 and C3 as well as the dimer of SILu are also observed.

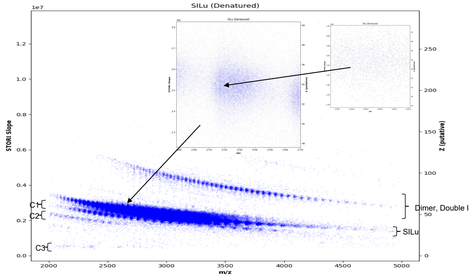
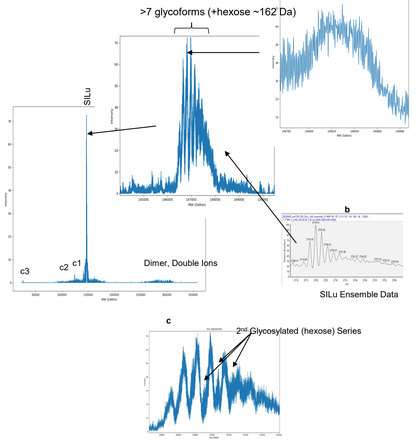


Figure 7. Molecular Weight Distribution after charge assignment by voting. At least 7 glycoforms, consistent with additions of hexose, are observed. Isotopes are discernable, though not baseline resolved. Insert b shows a low-resolution ensemble m/z spectrum for a single charge state. An expanded view (insert c) of the main profile shows a second glycosylated series.



3. Enbrel (Native)

Approximately 2.5e6 ions (from ~4000 spectra) passed the threshold criterion. Data show isotopic resolution but with significantly overlapping charge states.

Figure 8. Charge state overlap is evident, while isotopes are discernible.

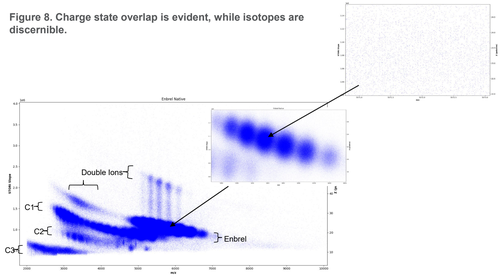
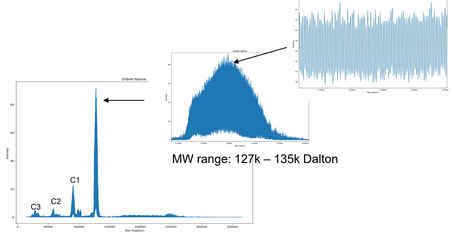


Figure 9. Molecular Weight Distribution. Isotopes are clearly resolved across the whole molecular weight range (127k – 135kk Dalton) of Enbrel. Other minor components are visible.



CONCLUSIONS

The 2D voting algorithm utilizes the precise relationship among charge related ions for charge assignment of isotopically resolved data collected in the Direct Mass Mode. It produces well resolved and detailed molecular weight distributions even for data with high level of complexity.

REFERENCES

1. Kafader, J. O.; Beu, S. C.; Early, B. P.; Melani, R. D.; Durbin, K. R.; Zabrouskov, V.; Makarov, A. A.; Maze, J. T.; Shinholt, D. L.; Yip, P. F.; Kelleher, N. L.; Compton, P. D.; Senko, M. W. *Journal of the American Society for Mass Spectrometry* **2019,** *30*, 2200-2203
2. Schachner, L. F.; Ives, A. N.; McGee, J. P.; Melani, R. D.; Kafader, J. O.; Compton, P. D.; Patrie, S. M.; Kelleher, N. L. *Journal of the American Society for Mass Spectrometry* **2019**, *30*, 1190-1198.

TRADEMARKS/LICENSING

PO66170-EN0422S