Selective *N*-terminal mass defect labelling of proteins for improved *de novo* top-down sequencing

Lavrentis Dimitrios Galanopoulos, Lenette Kjaer, Sam Hughes, Faye Cruickshank, C. Logan Mackay, David J. Clarke

The School of Chemistry, The University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ

Introduction

In recent years, there has been increasing attention in the development of top-down mass spectrometry strategies for de novo protein sequencing, i.e. the complete elucidation of protein sequence (and any posttranslational or chemical modifications) without the requirement for database searches. Here, we use a strategy for selective chemical labelling of the protein *N*-terminus using a bromine-based mass defect tag (BrMDT; delta mass: C_6H_4NOBr) to the *N-terminus* of proteins.[1-2] Thus, after gas phase top-down fragmentation, all *N*-terminal fragments would be tagged, while all *C-terminal* ions would not. The large mass defect accompanying tagging will then allow for easy distinction between *N* and *C*-terminal fragments thereby aiding fragment assignment (Fig 1).

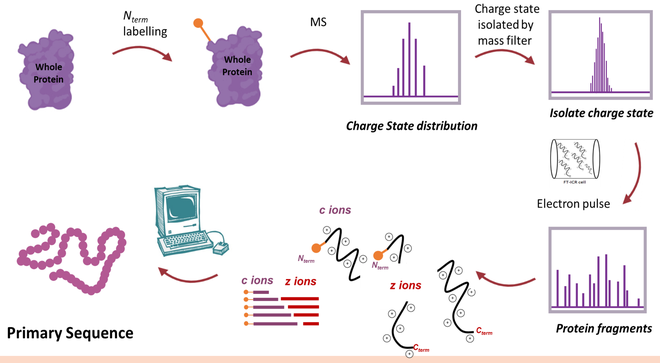
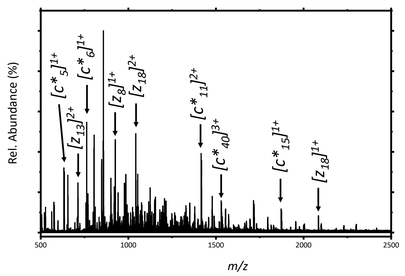


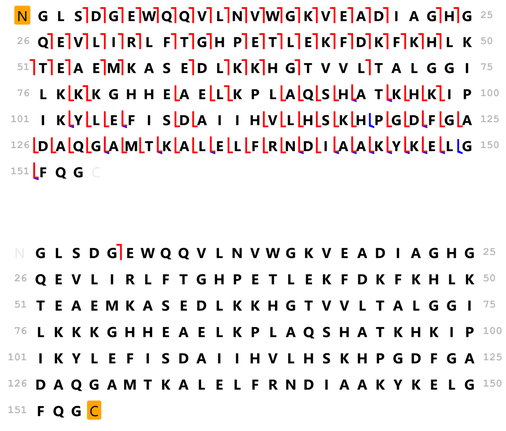
Fig. 1: The Top-down workflow in combination with a mass defect labelling. *Protein is selectively labelled on the N terminus with a covalent mass defect label before top-down fragmentation. The large mass defect allows classification of fragment ions into N-terminal (c-ions) and C-terminal (z-ions) fragment series without prior sequence information.*

Evaluating the Selectivity of Chemical Modification

A.



B.



The level of selectivity for *N*-terminal amines over non-specific modification of sidechain lysines is arguably the most important parameter for assessing the suitability of the reaction for the intended top-down workflow. ECD fragmentation was performed on the isolated 17^+ charge state of the singly modified Myo achieving 72 % sequence coverage of the protein (Fig. 4a). Of the peaks that could be assigned from the fragmentation spectrum the majority were BrMDT tagged *c*-ions (45 ions) and untagged *z*-ions (37 ions) both of which suggest the tag is located at the *N*-terminus (Fig. 4b). Importantly, The covalent modification appeared to be stable to ECD and no fragment ions where observed to suggest modification at any sidechain lysine residue.

Fig 4: A. ECD fragmentation spectrum of Myo modified with 5-bromo-2-pyridinecarboxaldehyde. B. Sequence of the Myo with BrMDT on the N-terminus (top) or on C-terminus (bottom) matched against the ECD fragmentation data.

Conclusions

A BrMDT was successfully incorporated into Myo *via* alkylation. Top down experiments indicated that the reaction was highly selective for the *N*-terminus over lysine. The utility of the BrMDT for *de novo* protein sequencing was demonstrated with an averagine-scaled mass defect plot, which allowed for clear and easy distinction of *N* and *C*-terminal fragments. After BrMD analysis the protein sequence could be determined in a de novo manner to produce a sequence coverage of 72 %. Future work will focus on achieving high conversion and selectivity when the method is applied to larger proteins and, eventually, to middle-down preparations of antibodies.

Materials and Methods

Selective N-terminal labelling was performed by adapting a published transamination strategy.[1] Bromo-substituted 2-pyridinecarboxaldehyde was stored as a 20-mM solution in ddH_2O at 4 °C. A 50-µL aliquot was taken from a 100-µM solution of protein and added to 30 µL of 0.1 M phosphate buffer at pH 7.9. A 20-µl aliquot from a 20-mM solution of 2-pyridinecarboxaldehyde reagent was the added to the resulting solution. The reaction was incubated at 37 °C overnight with gentle mixing. After labelling, proteins were desalted using ZipTip. Top-down mass spectrometry was carried out using a 12T SolariX 2XR FT-ICR instrument (Bruker Daltonics, Billerica, MA, USA) with ESI ionisation. Fragmentation was achieved using electron capture dissociation (ECD).

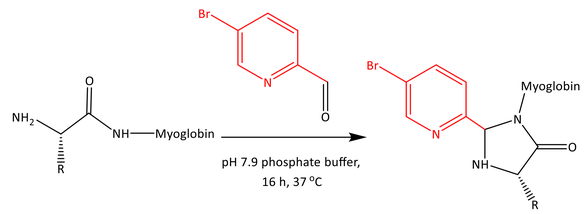


Fig. 2: Reaction scheme for the N-terminal selective incorporation of 5-Bromo-2-pyridinecarboxaldehyde.

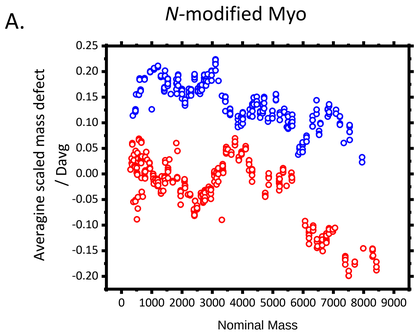
Mass Defect Analysis and De novo sequencing

Next, we used these datasets to explore the benefits of *N*-terminal mass defect labelling in assigning top-down fragment ions in *de novo* sequencing workflows. If we consider proteins as polymers of repeating monomer units (the 'poly-averagine' model), an approximate linear trend of increasing mass defect exists as protein length increases. The diversity of the twenty amino acid sidechains inevitably results in differences in the elemental composition of the monomeric units within proteins. Hence, the trend of mass defect with increasing nominal mass of proteins deviates from the theoretical straight line. However, because the elemental diversity of the natural amino acids is limited to CHNOS, this deviation is relatively minimal. As halogens have large mass defects, it should be possible to analyse ECD spectra containing halogens fragment ions distinguishable by their significantly increased mass defect.

A BrMDT plot was constructed using an averagine scaling analysis for the assigned ions from ECD fragmentation data set (Fig. 5A). By applying averagine mass rescaling, the fragments separate into two series - *N*-terminal fragments (e.g. *a-*, *b-*, *c*-ions) contain the mass defect label (blue; centered around mass defect ≈ 0.16); and *C*-terminal fragments (e.g. *z-*, *y-*, *x*-ions) are unmodified (red; centered around mass defect ≈ 0). Thus, individual fragment ions can be assigned as N-terminal or C-terminal using this plot. In a *de novo* sequencing context, the BrMDT analysis of Myo allows over 70% sequence to be assigned without assuming prior knowledge about the peptide sequence (Fig. 5B).

This workflow allows more confident assignment of sequence tags and increased sequence coverage in *de novo* top-down analyses.

A.



B.

Fig 5: A. Mass defect plot constructed using averagine scaling analysis of the assigned ions from the ECD fragmentation data of Myo modified. B. The protein sequence of Myo determined de novo using BrMDT top-down sequence. Ambiguous sequences are highlighted in Red.

Chemical Tagging of Myo

To develop our BrMDT strategy, the 13.7 kDa protein Myoglobin (Myo) was used. After optimisation of chemical labelling, mass spectra of unmodified and modified Myo were obtained (Fig 3). The mass defect tag (185.955451 Da) was successfully incorporated into Myo as demonstrated by MS analysis.

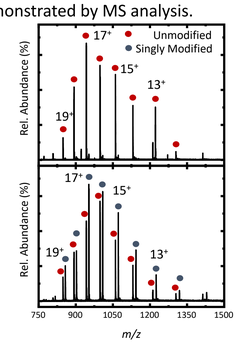


Fig 3: Mass spectra of unmodified Myo (top) and modified Myo (bottom).

References

1. Gilmore, J.; Scheck, R.; Esser-Kahn, A.; Joshi, N.; Francis, M. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angewandte Chemie* 2006, *118* (32), 5433-5437.
2. Hall, M.; Ashrafi, S.; Obegi, I.; Petesch, R.; Peterson, J.; Schneider, L. Mass Defect Tags For Biomolecular Mass Spectrometry. *Journal of Mass Spectrometry* 2003, *38* (8), 809-816.