

## Structural bioinformatics

# NaViA: a program for the visual analysis of complex mass spectra

Daniel Quetschlich <sup>1,2</sup>, Tim K. Esser<sup>1</sup>, Thomas D. Newport<sup>2,†</sup>,  
Francesco Fiorentino <sup>1</sup>, Denis Shutin<sup>1</sup>, Siyun Chen<sup>1</sup>, Rachel Davis<sup>3</sup>, Silvia Lovera<sup>4</sup>,  
Ildir Liko<sup>5</sup>, Phillip J. Stansfeld <sup>2,6,\*</sup> and Carol V. Robinson<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry, University of Oxford, Oxford OX1 3TF, UK, <sup>2</sup>Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK, <sup>3</sup>UCB Pharma, Slough, Berkshire SL1 3WE, UK, <sup>4</sup>UCB Pharma, Chemin du Foriest, 1420 Braine-l'Alleud, Belgium, <sup>5</sup>OMass Therapeutics, Oxford OX4 4GE, UK and <sup>6</sup>School of Life Sciences & Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

\*To whom correspondence should be addressed.

Present address: Oxford Nanopore Technologies, Oxford OX4 4DQ, UK.

Associate Editor: Pier Luigi Martelli

Received on February 23, 2021; revised on April 28, 2021; editorial decision on May 23, 2021; advance access date 19 June 2021

## Abstract

**Motivation:** Native mass spectrometry is now a well-established method for the investigation of protein complexes, specifically their subunit stoichiometry and ligand binding properties. Recent advances allowing the analysis of complex mixtures lead to an increasing diversity and complexity in the spectra obtained. These spectra can be time-consuming to tackle through manual assignment and challenging for automated approaches.

**Results:** Native Mass Spectrometry Visual Analyser is a web-based tool to augment the manual process of peak assignment. In addition to matching masses to the stoichiometry of its component subunits, it allows raw data processing, assignment and annotation and permits mass spectra to be shared with their respective interpretation.

**Availability and implementation:** NaViA is open-source and can be accessed online under <https://navia.ms>. The source code and documentation can be accessed at <https://github.com/d-que/navia>, under the BSD 2-Clause licence.

**Contact:** [phillip.stansfeld@warwick.ac.uk](mailto:phillip.stansfeld@warwick.ac.uk) or [carol.robinson@chem.ox.ac.uk](mailto:carol.robinson@chem.ox.ac.uk)

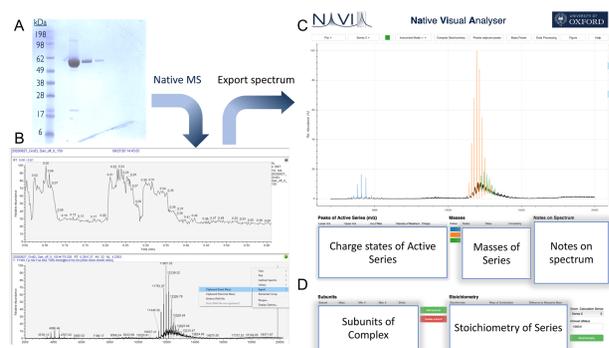
**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

Native mass spectrometry (nMS) is an established technique that allows a label-free analysis of intact proteins and their complexes in a mass spectrometer (Marty *et al.*, 2015). It has proven to be a powerful tool in investigating the stoichiometry and interactions of a wide variety of proteins with unprecedented resolution (Bolla *et al.*, 2019). In nMS, protein complexes are transferred into the mass spectrometer by means of electrospray ionization (Dole *et al.*, 1968). A series of ion selection steps and application of energy through accelerating voltages and collisions with inert gas molecules allow the transmission of folded proteins and interaction partners into a mass analyser (Gupta *et al.*, 2018). Soluble proteins and protein complexes with molecular weights up to the mega-Dalton mass range are amenable to study and examples include ribosomes (McKay *et al.*, 2006) and intact viruses (Snijder *et al.*, 2008). Also, more recently membrane proteins have been studied following their release from their solubilizing agent (e.g. detergent micelles, bicelles, nanodiscs or vesicles) through application of energy prior to mass

analysis (Laganowsky *et al.*, 2013). Mass spectra are then analysed manually or by using automated approaches such as UniDec (Bern *et al.*, 2018; Marty *et al.*, 2015). Interpretation of these spectra requires a priori knowledge of the molecules included in the complex from gel electrophoresis, western blot or 'omics techniques (Gault *et al.*, 2016).

Recent progress in sample preparation paved the way for the investigation of increasingly complex environments such as vesicles obtained from cell membranes. Following this procedure, samples are injected directly into the mass spectrometer and analysed through nMS without previous purification or treatment with detergents (SoLve-MS) (Chorev *et al.*, 2018), to preserve the native conditions for membrane proteins. These experiments yield highly complicated spectra containing a plethora of charge state series corresponding to the different protein complexes present in the vesicle. Analysing these spectra through high-level automated approaches is challenging given the extent of overlapping peak series. Thus, performing a manual analysis of these spectra is necessary yet incredibly time-consuming.



**Fig 1.** Example of the integration of NaViA into native mass spectrometry workflow. (A) In this example purified GroEL purchased from Sigma Aldrich was used as a sample. (B) The native mass spectrometry spectrum can be exported as a CSV file. (C) Import CSV file into NaViA and analyse by visually selecting peaks for each molecular species. (D) Assess stoichiometry for identified species through ‘Complex stoichiometry’ in NaViA

Motivated by these necessities, we created Native Mass Spectrometry Visual Analyser (NaViA), with the aim of occupying the middle ground between manual and automated analysis. Therefore, NaViA is complementary to existing software (e.g. UniDec). NaViA is a web service that allows a semi-automated analysis of native mass spectra through an easy-to-use interface (Fig. 1). Charge states corresponding to a particular protein are selected visually and masses are assigned automatically following known algorithms. The focus on augmenting a manual assignment is also instructive for scientists entering the field of nMS, since it provides key information on where peaks are to be expected for a certain complex and highlights peak series for this complex. In addition, sessions may be saved comprising both raw data and also processed data with annotations. This permits the nMS community to share both, spectra and the analysis, in a standardized format online as part of a collaboration or publication.

## 2 Materials and methods

### 2.1. Implementation

The web service is implemented in Python using the Bokeh visualization package. It is a standalone HTML file without any server in the background. It is implemented in Python 3.6 and JavaScript using the Bokeh framework (version 1.4.0). According to Wellcome and the UK research and innovation’s Common Principles on Data Policy on data, software and materials management and sharing, all data supporting this study will be openly available from the software repository.

### 2.2. Documentation and tutorials

The documentation of NaViA is available on its GitHub repository as well as an introductory video (<https://www.youtube.com/channel/UCRvzQxegz0WNkjvvaRJ8PpQ/>).

### 2.3. Input

NaViA currently loads files in a CSV format of the two columns  $m/z$  and intensity. These can be exported by commonly used software such as MassLynx for Waters Instruments or Xcalibur for ThermoFisher Scientific instruments such as OrbiTraps (Makarov, 2000). Upon loading, the intensities are normalized.

## 3 Results

### 3.1 Processing of mass spectra

NaViA enables smoothing and background subtraction of native mass spectra using the same algorithm as standard software, e.g. UniDec. Smoothing is performed through a Gaussian filter and can

be performed multiple times. In addition, different modes of background subtraction are facilitated and a filter for ‘minimum intensity’ can be applied. The background subtraction modes include subtracting a fixed value, a linear interpolation between the intensities of the smallest and largest  $m/z$  values and a curved subtraction as described previously (Morgner and Robinson, 2012).

### 3.2 Peak selection and mass calculation

Peaks are selected through a range on the  $m/z$ -axis using the mouse cursor. Once multiple charge states are selected, the corresponding mass is calculated by the MacSED algorithm (McKay et al., 2006). This algorithm systematically creates a matrix of possible charge states assuming there are no missing charge states in the series. Based on this series a matrix of masses is created where each  $m/z$  value of a peak is deconvoluted into a mass value. The variance between the masses is then calculated for each series of charges. The series of masses with the minimal variance is then selected to be the true series of charge states. Finally, the mean of the measured mass is set as the true mass and the standard deviation set to be the uncertainty value.

### 3.3 Tools for augmenting peak assignment

For the analysis of complex native mass spectra two tools have been implemented: a ‘Mass Finder’, which allows the user to highlight  $m/z$  regions for a specific mass of a complex, and a ‘Peak Prediction’ tool, which highlights adjacent peaks for a known series. The ‘Mass Finder’ makes it possible to see whether a complex with known mass is present. This is especially important for SoLVE spectra where many molecular species are present. For these spectra, ‘Peak prediction’ augments the differentiation of overlapping peak series.

### 3.4 Comparing complex mass to subunit masses

The ‘Complex Stoichiometry’ function makes it possible to see whether a complex with known mass is present. This is especially important for cases where many molecular species are present.

### 3.5 Sessions in NaViA

NaViA allows sessions to be saved, including raw data, data processing and peak assignment, in a *navia* file in JSON format. This feature allows users to share spectra including assignments as well as collaborating or deposition of mass spectra for enabling raw data sharing. In addition, JSON formats are easily readable for various programming languages and software packages, e.g. Python or Origin. Therefore, *navia* files are well suited for consecutive data analysis. The format is described in the documentation on the GitHub page.

### 3.6 Compatibility with UniDec

NaViA was developed to be complementary to the current software packages available for native MS data analysis. As part of this, an option to import NaViA sessions has been incorporated into the current UniDec version (4.4.1) (Experimental → Import from NaViA).

### 3.7 Exporting images

NaViA provides the possibility to export images in 4k quality sufficient for A4 pages in publications as well as vector graphics in the SVG format.

## Acknowledgements

We would like to thank the Robinson, Stansfeld and Benesch groups for their support and feedback. In addition, we would like to thank Bryan Van de Ven and Eugene Pakhomov from the Bokeh forum for their extensive support on technical issues. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## Funding

Research in the C.V.R. laboratory is supported by a Medical Research Council program grant [MR/N020413/1], a European Research Council Advanced Grant ENABLE [695511] and a Wellcome Senior Investigator Award [104633/Z/14/Z]. Research in P.J.S.'s lab is funded by Wellcome [208361/Z/17/Z], the MRC [MR/S009213/1] and BBSRC [BB/P01948X/1, BB/R002517/1 and BB/S003339/1]. D.Q. and F.F. hold a SABS CDT studentship supported by the EPSRC and MRC [EP/L016044/1].

*Conflict of Interest:* none declared.

## References

- Bern, M. *et al.* (2018) Parsimonious charge deconvolution for native mass spectrometry. *J. Proteome Res.*, **17**, 1216–1226.
- Bolla, J.R. *et al.* (2019) Membrane protein–lipid interactions probed using mass spectrometry. *Annu. Rev. Biochem.*, **88**, 85–111.
- Chorev, D.S. *et al.* (2018) Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science* (New York, N.Y.), **362**, 829–834. [10.1126/science.aau0976](https://doi.org/10.1126/science.aau0976) [30442809](https://pubmed.ncbi.nlm.nih.gov/30442809/)
- Dole, M. *et al.* (1968) Molecular beams of macroions. *J. Chem. Phys.*, **49**, 2240–2249.
- Gault, J. *et al.* (2016) High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods*, **13**, 333–336.
- Gupta, K. *et al.* (2018) Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.*, **13**, 1106–1120.
- Laganowsky, A. *et al.* (2013) Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.*, **8**, 639–651.
- Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry*, **72**, 1156–1162. [10.1021/ac991131p](https://doi.org/10.1021/ac991131p) [10740853](https://pubmed.ncbi.nlm.nih.gov/10740853/)
- Marty, M.T. *et al.* (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.*, **87**, 4370–4376.
- McKay, A.R. *et al.* (2006) Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *J. Am. Chem. Soc.*, **128**, 11433–11442.
- Morgner, N. and Robinson, C.V. (2012) Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.*, **84**, 2939–2948.
- Snijder, J. *et al.* (2008) Studying 18 mega Dalton virus assemblies with native mass spectrometry. *Bone*, **23**, 1–7.