Pacific Northwest
NATIONAL LABORATORY

PNNL-33361

High-throughput native mass spectrometry as experimental validation for in silico drug design

September 2022

Mowei Zhou, Katherine Schultz, Jesse Wilson, Agustin Kruel, Stephanie Thibert, Carter Bracken, Daniel Orton, Bryson Gibbons, Rosalie Chu, Andrew McNaughton, Ljiljana Pasa-Tolic

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-33361

High-throughput native mass spectrometry as experimental validation for in silico drug design

September 2022

Mowei Zhou, Katherine Schultz, Jesse Wilson, Agustin Kruel, Stephanie Thibert, Carter Bracken, Daniel Orton, Bryson Gibbons, Rosalie Chu, Andrew McNaughton, Ljiljana Pasa-Tolic

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

In this project, we developed automated workflows for both experimental validation and computational prediction of protein-ligand interactions. The ultimate goal is to establish an integrated pipeline for high throughput design of inhibitors to enzymes relevant to all areas of biological research. Our experimental approach is based on native mass spectrometry (native MS), which measures accurate masses and quantify the relative abundance of protein-ligand complexes to define binding affinity. We set up an in-house built autosampler with highly flexible configurations to minimize the manual steps for high throughput native MS. In parallel, we also performed manual native MS to characterize the binding of substrates and inhibitors of SARS-Cov-2 nonstructural protein nsp10/16 in order to optimize the experimental parameters for future automation. On the computational side, we streamlined the pipeline to achieve minimal manual intervention for predicting enzyme inhibitors via simulation, using the same nsp10/16 system as an example. Using the native MS method we examined 8 top-ranked designed compounds, 2 of which showed weak binding of ~50 μM. The information from native MS experiment provided critical insights and the foundation for a fully integrated workflow for enzyme inhibitor design.

Summary

We aim to develop an automated experimental screening method to validate computationally designed small molecule inhibitors against selected protein targets. The proposed workflow will tremendously increase the throughput and accuracy for structure-based drug design. The general methodology can also be expanded to tuning functions of important enzymes in bioenergy and environmental applications.

The experimental screening method will be based on native mass spectrometry (native MS), which directly measures binding of small molecules to proteins. Higher-order structure and kinetic information can also be extracted from the native MS data with advanced experimental design. Unfortunately, native MS has largely been performed manually and cannot easily be scaled up for high throughput screening. We developed a custom robotic system that is robust, reproducible, and flexible for advanced experimental designs.

At the same time, we interfaced native MS with cutting-edge computational methods to streamline data analysis and advance virtual screening technology. Large-scale in silico screening of compounds against a target of interest has the ability to greatly reduce time and cost in the search for promising lead compounds; however, current computational methods are prone to high failure rates that can only be discerned after time- and resource-intensive in vitro and in vivo experiments. By coupling our virtual screening pipeline to automated, high throughput native MS, we can substantially reduce the time and cost required to verify computationally designed small molecules. The much larger volumes of collected structural data produced by high throughput native MS will enable the rapid optimization of in silico technologies in an automated pipeline.

Acknowledgments

We thank our collaborators Andrzej Joachimiak, Gyorgy Babnigg, and Natalia Maltseva at the Argonne National Laboratory for providing the protein samples; Weijing Liu at ThermoFisherScientific for advising on experiment optimization. This research was supported by the **EBSD Mission Seed**, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Contents

Figures

1.0 Introduction

Enzyme functions can often be modulated via interaction with small molecules. For example, inhibitors are actively sought in the pharmaceutical industry to block protein targets and treat diseases. A similar concept is applicable for treating other organisms (e.g., bioenergy plants and microbes) with designed compounds for desired outcomes. The challenge of finding a small molecule that binds to a protein with high affinity and high specificity is quite well-known and costly. Existing high throughput assays are limited to proteins with at least some prior knowledge of function and are costly and prone to high failure rates. Thus, *in silico* design holds significant potential for high-throughput, low-cost drug screening without the need for prior knowledge of protein function.(Gupta, Srivastava et al. 2021) However, given the current state of the computational chemistry, the effectiveness of *in silico* drug design lags in affinity and specificity behind experimental approaches.

Native mass spectrometry (native MS) is an emerging proteomics technique to study protein complexes, including protein-drug binding.(Boeri Erba, Signor et al. 2020, Bennett, Nguyen et al. 2022, Tamara, den Boer et al. 2022) Compared to the more established methods such as immunoprecipitation, affinity probe, or crosslinking, native MS is unique in that it measures binding directly without requiring labels/probes. Data collection for native MS is also fast, on the order of minutes for confirming ligand binding. In addition, native MS can readily resolve apo/holo proteins and protein oligomers. Since native MS operates under non-denaturing conditions, protein dynamics and enzymatic reactions can be monitored by time-resolved measurements, allowing us to interrogate protein functions in great detail. Such molecular information is extremely useful as feedback for iterative computational compound design and further improving success rate. However, most native MS studies have been performed manually. Automated systems available commercially are not very flexible for complex experimental designs and significant failure rates. To address these challenges, we aim to develop an automated, scalable native MS platform that is compatible with various experimental designs. In parallel, we will develop a computational workflow that is compatible with native MS data. The integrated platform will provide the critical foundation for advanced high-throughput functional proteomics analysis of protein complexes and protein-ligand interactions.

2.0 Experimental Development of Native MS Methodology

The current standard native MS protocol in academic labs uses static nano-electrospray (nano ESI) setup. Briefly, microliters of samples are loaded into glass capillaries. Electric potential is applied either by metal coating or metal wire that is in contact with the sample. Such workflow is highly robust and low sample consumption, but cost prohibitive for large scale studies due to tedious, repetitive manual labor. Our goal is to establish an automated platform for maximum throughput and robustness. In parallel, we also used the standard manual native MS workflow to interrogate the SARS-Cov-2 nsp10/16 heterodimer and provide feedback to the computational design of inhibitors to pave the way for full integrated workflow in the near future.

2.1 Automation of the Native MS Workflow

We established an automated nano ESI system by improving the previous design built in-house.(Orton, Tfaily et al. 2018) The performance of the system has been tested using several protein standards with different sizes: carbonic anhydrase (23 kDa), concanavalin A (52 kDa dimer, 103 kDa tetramer), monoclonal antibody (150 kDa). We achieved good performance of 100-300 nL/min which is similar to regular nanoflow liquid chromatography conditions. In addition, we obtained stable signal down to 50 nL/min. Lower flow rate of ~10 nL/min, which is in the same regime as manually operated static nanospray, is possible but not highly reproducible. The limit of detection on a Waters Synapt mass spectrometer is estimated to be ~50 nM (roughly an order of magnitude higher than manual operation), although this can be reduced on other instruments and can vary depending on the sample. Nonetheless, we established an automated platform for flow injection for native MS, which can be further optimized for more complex experimental designs.

In parallel, we also implemented an online buffer exchange (OBE) chromatography system,(VanAernum, Busch et al. 2020) which operates at high flow rate but can eliminate the manual desalting step for fast screening. We tested different flow rates and different tubing to optimize the performance, with the goal to have good desalting while minimizing analysis time. After careful tuning, the system allows for quick separation of proteins from buffer salts (Figure 1), effectively allowing high throughput analysis of samples without having to perform manual desalting steps.

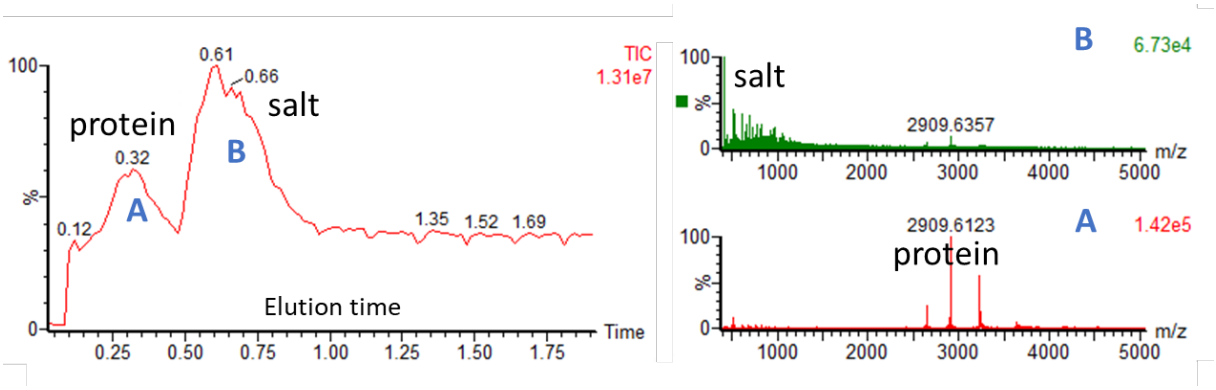

Figure 1. Exemplary ion chromatogram showing the fast online desalting of proteins (A) in salt (B) buffer at flow rate of 150 µL/min. Size exclusion separation was achieved in less than 1 min for high-speed operation. Manual desalting can take ~15 min or more.

2.2 Native MS Characterization of Nsp10-16 Heterodimer

SARS-CoV-2 forms a replication transcription complex (RTC) composed of non-structural proteins (nsp). Among those, nsp10/16 heterodimer is a methyltransferase enzyme involved in the final capping process of nascent-viral RNA, which can serve as targets for development of pan-coronavirus antiviral treatments. X-ray crystallography structures of nsp10/16 have provided critical insights into the protein's function.(Wilamowski, Sherrell et al. 2021) Herein, we performed native MS to study the binding of substrate RNA, cofactor (S-adenosyl-L-methionine), and inhibitors to the nsp10/16 heterodimer (Figure 2). The ligand binding behavior in native MS was consistent with reported information, which validates the native MS method. Interestingly, we also observed fast equilibrium of the heterodimer assembly and disassembly in solution, highlighting the potential role of dynamics in protein function and inhibitor design. In addition, we also showed that native MS can be used to monitor the "live" methyltransferase reaction of nsp10/16 on Cap-0 RNA substrate. The change of substrates and proteins can be monitored simultaneously over time. Native MS can serve as an all-in-one technique to perform fast biophysical and biochemical measurements on heterocomplexes to inform structure-based drug design.

While the method needs to be further developed for high throughput feedback to computational design, we manually tested 16 designed compounds for nsp10/16. Our results suggest that 4 of them bind weakly to the protein at Kd of ~50 μM. Furthermore, some of them appear to bind to both the heterodimer and free monomer subunits in solution. This suggests that some of the compounds may not be binding to the heterodimer interface where the active site is. Such information is critical for further *in silico* optimization of the inhibitors.

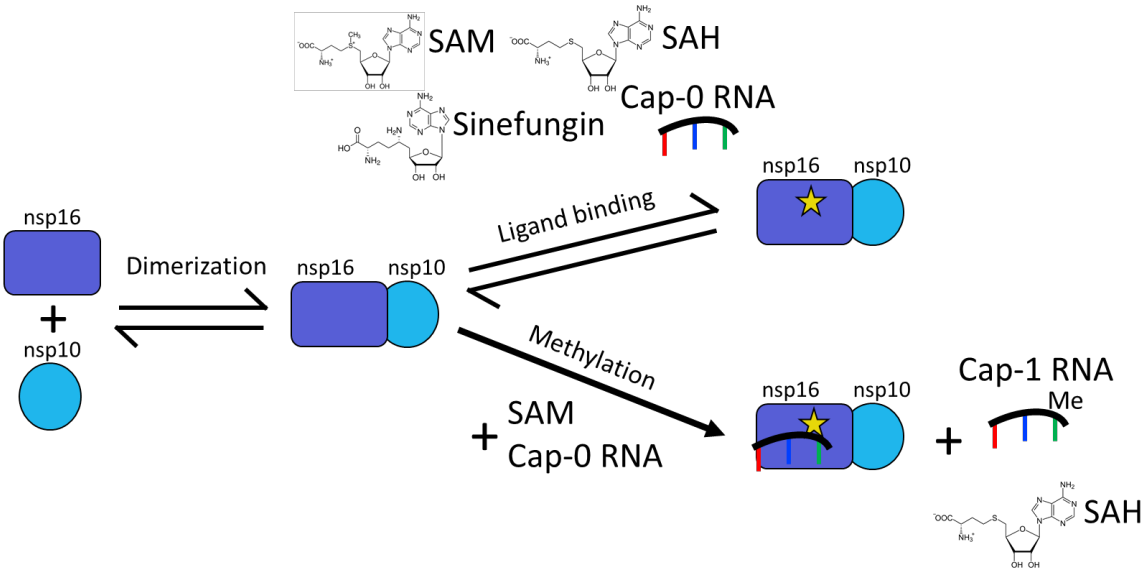

Figure 2. A summary of the native MS experiments performed to study nsp10/16 interaction with substrates and inhibitors. Various species, including protein-protein, protein-ligand, protein-RNA complexes can be resolved simultaneously in the mixture by their unique masses. In addition, we monitored the methylation reaction by monitoring the change of abundances of the species over time. Overall, these results demonstrated the potential of native MS to rapidly obtain high precision structural data for characterizing molecular interactions.

3.0 Streamlining the Computational Pipeline

To enable the efficient discovery of new NSP10/16 inhibitors, we tailored a Python-based computational analysis approach that utilizes a range of established methodologies to the NSP10/16 system. The pipeline begins with a compound database represented in simplified molecular input line entry system (SMILES) notation as input and features the following primary components: i) chemical property prediction; ii) protein-ligand docking simulations; and iii) inhibitor down-selection. The initial compound database was creating by selecting the subset of the compounds in the Mcule database with structural similarity to known NSP10/16 inhibitors. Throughout the stages of the pipeline, candidate compounds were assessed against known inhibitors at the target binding pocket (aka the SAM pocket), notably SAM (S-adenosyl-l-methionine) and sinefungin. Once each piece of the NSP10/16 *in silico* pipeline was in place, we automated it to the greatest extent possible (there are a couple of points that require human decision making to proceed). The first set of candidate inhibitors we identified with the pipeline was unfortunately not orderable, nor were there time and resources to synthesize the compounds. We then limited the compounds under consideration to only those that were immediately orderable and produced a set of 8 candidates for experimental testing. Overall, this *in silico* pipeline allows a large compound database to be assessed for promising leads targeted to a particular binding site efficiently and with minimal user input required. This work lays the groundwork to further develop boundary-pushing high throughput closed loop computational-experimental platforms.

4.0 References

Bennett, J. L., G. T. H. Nguyen and W. A. Donald (2022). "Protein-Small Molecule Interactions in Native Mass Spectrometry." Chem Rev **122**(8): 7327-7385.
Boeri Erba, E., L. Signor and C. Petosa (2020). "Exploring the structure and dynamics of macromolecular complexes by native mass spectrometry." Journal of Proteomics **222**: 103799.
Gupta, R., D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta and P. Kumar (2021). "Artificial intelligence to deep learning: machine intelligence approach for drug discovery." Molecular Diversity **25**(3): 1315-1360.
Orton, D. J., M. M. Tfaily, R. J. Moore, B. L. LaMarche, X. Zheng, T. L. Fillmore, R. K. Chu, K. K. Weitz, M. E. Monroe, R. T. Kelly, R. D. Smith and E. S. Baker (2018). "A Customizable Flow Injection System for Automated, High Throughput, and Time Sensitive Ion Mobility Spectrometry and Mass Spectrometry Measurements." Analytical Chemistry **90**(1): 737-744.
Tamara, S., M. A. den Boer and A. J. R. Heck (2022). "High-Resolution Native Mass Spectrometry." Chemical Reviews **122**(8): 7269-7326.
VanAernum, Z. L., F. Busch, B. J. Jones, M. Jia, Z. Chen, S. E. Boyken, A. Sahasrabuddhe, D. Baker and V. H. Wysocki (2020). "Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry." Nature Protocols **15**(3): 1132-1157.
Wilamowski, M., D. A. Sherrell, G. Minasov, Y. Kim, L. Shuvalova, A. Lavens, R. Chard, N. Maltseva, R. Jedrzejczak, M. Rosas-Lemus, N. Saint, I. T. Foster, K. Michalska, K. J. F. Satchell and A. Joachimiak (2021). "2′-O methylation of RNA cap in SARS-CoV-2 captured by serial crystallography." Proceedings of the National Academy of Sciences **118**(21): e2100170118-e2100170118.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov