Native Mass Spectrometry Dissects the Structural Dynamics of an Allosteric Heterodimer of SARS-CoV-2 Nonstructural Proteins

Stephanie M Thibert\textsuperscript{1,§}, Deseree J Reid\textsuperscript{2,§}, Jesse Wilson\textsuperscript{1,§}, Rohith Varikoti\textsuperscript{3}, Natalia Maltseva\textsuperscript{4,5}, Katherine Schultz\textsuperscript{3}, Agustin Krueg\textsuperscript{9}, Gyorgy Babnigg\textsuperscript{4,6}, Andrzej Joachimiak\textsuperscript{4,5}, Neeraj Kumar\textsuperscript{3}, Mowei Zhou\textsuperscript{1,⊥, *}

1. Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, United States
2. Chemical and Biological Signature Sciences, Pacific Northwest National Laboratory, Richland, WA, United States
3. Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, United States
4. Center for Structural Biology of Infectious Diseases, Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL, United States
5. Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Argonne, IL, United States
6. Biosciences Division, Argonne National Laboratory, Argonne, IL United States

§ These authors contributed equally.

⊥ Present address: Department of Chemistry, Zhejiang University, Hangzhou, Zhejiang, China

Correspondence: moweizhou@zju.edu.cn
Abstract

Structure-based drug design, which relies on precise understanding of the target protein and its interaction with the drug candidate, is dramatically expedited by advances in computational methods for candidate prediction. Yet, the accuracy needs to be improved with more structural data from high throughput experiments, which are challenging to generate, especially for dynamic and weak associations. Herein, we applied native mass spectrometry (native MS) to rapidly characterize ligand binding of an allosteric heterodimeric complex of SARS-CoV-2 two nonstructural proteins (nsp) nsp10 and nsp16 (nsp10/16). Native MS showed that the dimer is in equilibrium with monomeric states in solution. Consistent with literature, well characterized small co-substrate, RNA substrate and product bind with high specificity and affinity to the dimer but not the free monomers. Unsuccessfully designed ligands bind indiscriminately to all forms. Using neutral gas collision, the nsp16 monomer with bound co-substrate can be released from the holo dimer complex, confirming the binding to nsp16 as revealed by the crystal structure. However, an unusual migration of the endogenous zinc ions bound to nsp10 to nsp16 after collisional dissociation was observed, and can be suppressed using an alternative surface collision method at reduced precursor charge states. This highlighted the importance of careful optimization of experimental techniques. Overall, with minimal sample input (~µg), native MS can rapidly detect ligand binding affinities and locations in dynamic multi-subunit protein complexes, demonstrating the potential of an “all-in-one” native MS assay for rapid structural profiling of protein-to-Al-based compound systems to expedite drug discovery.
Structure-based drug design (SBDD) is one approach for accelerated discovery of new compounds\(^1\) with therapeutic potential that focuses on finding small molecule inhibitors binding to target proteins’ active sites. The recent progress in artificial intelligence (AI) and machine learning (ML) has greatly elevated the efficiency of in silico screening for potential compounds, offering considerable savings in time and resources compared to traditional experimental screening methods. However, this approach can be prone to errors if not executed thoughtfully.\(^2\) While AI/ML both benefit from training with large datasets, current high-throughput experimental screening methods only provide activity and phenotypical data.\(^3\)-\(^5\) Structural details that can be quantitatively interfaced with first-principle calculations such as docking and molecular dynamics (MD) simulations are often lacking. Therefore, classical structural biology methods remain the primary source of high-quality structural data, but they are resource- and labor-intensive and are usually applied only at a later stage of the screening on a limited number of protein-ligand complexes.

Native mass spectrometry (native MS) characterizes noncovalent complexes under nondenaturing conditions, and can provide fast assessment (typically in minutes) of protein-protein and protein-ligand interactions using small quantities (~µg) of often precious samples.\(^6\)-\(^9\) Herein, we show the types of critical information relevant to SBDD provided by native MS for an allosteric complex of the two nonstructural proteins (nsp) nsp10 and nsp16 heterodimer from SARS-CoV-2, including protein dynamics, allostery, and ligand binding specificity gleaned from these measurements.

The nsp10/16 heterodimer is part of the SARS-CoV-2 RNA replication and transcription complex.\(^10\) Nsp10/16 heterodimer is an enzyme - 2’-O methyltransferase with nsp16 serving as a catalytic unit and nsp10 functioning as an allosteric activator essential for enzyme activity. The complex methylates the ribose 2’-O of the first adenosine ribonucleotide of the nascent mRNA Cap-0-RNA to form Cap-1-RNA (\(\text{m}^7\text{GpppAm}2’-\text{O-RNA}\)) using SAM as the methyl donor and helps evade the host’s immune response.\(^11\)-\(^15\) Nsp10 is essential to activate the methyltransferase (MTase) of nsp16\(^15\) via allosteric effect.\(^16\)-\(^18\)
Inhibition of the methyltransferase activity may reduce viral proliferation, making the complex an attractive pan-coronavirus drug target considering its essential function and conservation in coronaviruses.\textsuperscript{19} We separately expressed and purified nsp10 and nsp16 from SARS-CoV-2 in \textit{E. coli} cells and mixed them at equimolar ratios in HEPES buffer to form the heterodimer nsp10/16 as previously described.\textsuperscript{11} The proteins were then buffer exchanged into 100 mM ammonium acetate (AA) solution for native MS. The native MS spectrum (Figure 1A) showed a mixed population of nsp10 monomer, nsp16 monomer, and nsp10/16 heterodimer. The mass of nsp10 was 127.9 Da larger than the mass measured under the denaturing state (Figure S1) due to presence of two zinc ions (Zn\textsuperscript{2+}) that are known to be essential for nsp10 function.\textsuperscript{13, 20, 21} The relative abundance of the dimer increased with higher protein concentration (Figure 1B), allowing us to plot a binding curve based on the native MS data and estimate a \(K_d\) of 1.4 \(\mu\)M (Figure S2A). Despite the different signal response (e.g., ionization efficiency) between nsp10 and nsp16 that could complicate the analysis (Figure S2B),\textsuperscript{22} our estimation is similar to the apparent \(K_d\) of MERS nsp10/16 (2 \(\mu\)M) reported by activity assay.\textsuperscript{17} Methods to account for differences in signal response during binding affinity studies\textsuperscript{23} by native MS, such as slow mixing mode (SLOMO),\textsuperscript{23} could be implemented in future studies.

The free monomers detected in native MS prompted further investigation into the dynamic nature of the dimer assembly. First, our molecular dynamics (MD) simulations of the heterodimer crystal structure in water solvent corroborated the reported allosteric effect of nsp10 binding to nsp16 (Figure S3), and also revealed many transient interfacial hydrogen bonds and salt bridges (Figure S4). Second, we mixed the heterodimer with another nsp10 construct with a 6.9 Da mass shift from different purification tag sequences (Figure S5A, sequence difference in Table S1). The mixing in solution prior to MS resulted in the appearance of a 6.9 Da mass shift on nsp10/16 in the native MS spectrum (Figure S5B). The transient interfacial interactions in MD and the exchange of subunits in solution suggested a dynamic equilibrium between the monomeric and dimeric states. Interestingly, the crystal structure of nsp16 monomer has
not been reported. Isolated nsp10 showed minimal rearrangement compared to its form in the heterodimer structure, complementing our MD results. Self-dimerization of nsp10 was also noted.\textsuperscript{24} Taken together, these results highlight that besides the known active form of nsp10/16, nsp10 and nsp16 monomers are likely relevant and should not be ignored in SBDD. Indeed, a recent MD study identified a cryptic pocket in the inactive monomeric state of nsp16 that could serve as a conserved feature for developing pan-coronavirus antivirals that stabilize the inactive form of the enzyme.\textsuperscript{16}

Figure 1. (A) Native MS spectrum showing a mixture of nsp10, nsp16, and nsp10/16 at 7 µM. The inset shows the magnified display of the dimer peaks. (B) Dimer formation can be directly tracked by the changing intensities as a function of protein concentration.

We then probed the binding specificity of several known ligands to nsp10, nsp16, and nsp10/16 by native MS. The S-adenosyl-L-methionine (SAM) co-substrate binds to nsp16, stabilizes nsp10/16, and
serves as the methyl donor for MTase activity during viral RNA capping. Cap-0 RNA substrate binds to the active site and is converted to the mature Cap-1 RNA by methyl transfer from the SAM, and generation of S-adenosyl-L-homocysteine (SAH) product. A pan-MTase inhibitor sinefungin (SFG) occupies the SAM and SAH binding site in co-crystals with nsp10/16 and has been used as a model to inform inhibitor design. We performed native MS with increasing concentrations of co-substrate SAM, product SAH and inhibitor SFG to 5 µM protein.

The relative abundances of the apo vs. holo populations for nsp10, nsp16, and nsp10/16 were plotted in Figure 2. All compounds bound primarily to nsp10/16, consistent with the isothermal titration calorimetry (ITC) binding assays using individual proteins. The proteins’ ionization efficiencies did not change significantly within these ligand concentrations (Figure S6). Therefore, we used a single-site binding model to calculate the apparent binding affinities of the ligands for nsp10/16. Binding affinities to SAM and SAH were similar with $K_d$ of ~2 µM, while SFG showed much weaker binding at ~25 µM (Figure 2E and Figure S7), similar to the values reported previously using ITC. At high ligand concentrations, we observed more than one compound binding per nsp10/16 heterodimer for all compounds, with calculated $K_d$ values at least an order of magnitude higher than the first binding event (see supplemental excel file) and were not discussed here due to the likelihood of nonspecific binding.

Next, we tested a potential inhibitor designed computationally through our high throughput virtual screening (HTVS) and ordered it from a drug library MCULE database (Mcule-ID: MCULE-8583086513-0). The specificity of binding was quickly assessed from native MS data with weak binding of the compound to not only nsp10/16, but also the monomers observed (Figure S8A). In contrast, binding of SAM/SAH/SFG displayed significant binding to nsp10/16 and only negligible binding to the monomers (Figure S8B-D). The recent MD study proposed that the SAM binding pocket was in a more open state in the heterodimer than in the nsp16 monomer, and thus any compound that binds indiscriminately to
monomer and dimer forms is unlikely to follow the same binding mode as SAM,\textsuperscript{16} possibly outside the predicted pocket.

Figure 2. Deconvolved mass spectra for SAM binding to (A) nsp10, (B) nsp16, and (C) nsp10/16 complex. The heat maps (D) show the percentage of the total peak area that is ligand-bound for each species as the ligand concentration was increased from 0 – 20 μM for SAM and SAH, and from 0 – 60 μM for SFG. Estimated $K_d$ curves (E) are shown for each ligand when fit to a single-site binding model.

Native “top-down” MS can inform binding location of ligands in multimeric protein complexes, via the release of ligand bound subunits after gas-phase “heating” of the intact complex.\textsuperscript{6} We m/z-isolated the SAM-bound holo dimer away from the monomeric states and dissociated the dimer in the mass
spectrometer via gas collision (collision induced dissociation, CID). SAM was readily released in the low m/z region (Figure 3). The released nsp16 partially retained SAM, and high collision energy led to complete loss of substrates from the protein. Similar behavior was seen for SFG inhibitor (Figure S9). In contrast, cap-0 RNA bound strongly to nsp10/16 and nsp16 (Figure S10). The RNA bound to heterodimer also showed complete preservation of bound RNA in the released nsp16, regardless of collision energy (Figure S11), likely due to the enhanced electrostatic interactions between RNA and nsp16 in the gas phase. The detection of SAM and cap-0 RNA in the released nsp16 was consistent with the binding pockets within nsp16 observed in the crystal structures, suggesting that such dissociation experiments can be used to probe which subunit the ligand binds to under optimized conditions. This crucial information can serve as feedback for affirming that the AI-designed compounds are targeting the appropriate component of a heterocomplex without having to solve the high-resolution structure. Additionally, we monitored enzyme activity, and inhibition in the presence of SFG inhibitor, by directly measuring formation of cap-1 RNA over time (Figure S12). Complementary SAM-to-SAH conversion was also measured in the same native MS experiment, highlighting the potential of native MS to simultaneously quantify multiple species relating to reaction kinetics within a single experiment using low sample quantities.
Figure 3. (A) Nsp10/16 was isolated from monomers and subjected to collisional dissociation in 2V increments from 4 – 60V. (B) Peak areas were extracted as a function of collision voltage and plotted for SAM-bound nsp10/16 (pink triangles), apo nsp10/16 (blue circles), SAM-bound nsp16 (green triangles), apo nsp16 (black circles), SAM-bound nsp10 (orange triangles), and apo nsp10 (purple circles). Additional peaks visible in the nsp10 and nsp16 spectra (A) are due to either 1 or 2 bound Zn ions.

Although co-substrate binding was successfully probed by CID, we noticed an unusual transfer of Zn$^{2+}$ on dissociated nsp10 and nsp16. The free monomers in native MS only showed Zn$^{2+}$ binding on nsp10 (Figure S1B), consistent with the crystal structure where Zn$^{2+}$ bind in two discrete cysteine rich motifs in nsp10. The metal transfer is reminiscent of the ligand migration previously observed in CID of cholera toxin B homopentamer, and is possibly related to the gas-phase structure and charge configuration unique to nsp10/nsp16. We therefore tested an alternative method implemented via custom modification, surface induced dissociation (SID), which generally results in less structural rearrangement than the widely accessible CID method, particularly with low charge state precursors. As expected, SID after reducing precursor charge (Figure S13) largely eliminated Zn$^{2+}$ transfer. While further experiments are needed to understand the mechanism of Zn$^{2+}$ transfer, the results demonstrated that
the dissociation experiment must be optimized to minimize structural rearrangements for obtaining the most relevant ligand/metal binding information to native-like states.

In summary, we demonstrated that native MS could measure many structural aspects of the allosteric heterodimeric enzyme nsp10/16. Co-existing populations of inactive monomeric states and the active dimeric state can be resolved, and their ligand binding properties separately examined. Monitoring time-resolved changes in enzymatic reactions is also feasible because of non-denaturing measurement. The “all-in-one” native MS assay offers molecular structure details that can serve as rapid feedback to computer aided drug design. The additional capability of isolating and dissecting selected species for more structural information requires more optimization to precisely define the connection with solution states. Nonetheless, with many ongoing technical developments in dissociation methods, native MS is anticipated to generate even more intricate structural data that will augment existing high throughput screening methods, and foster the creation of a mechanistic knowledge database for AI/ML models that will lead to improved success rates and throughput for drug discovery.
Acknowledgement

We thank Ryan Choi and Wesley Van Voorhis at University of Washington for providing the alternative nsp10 construct, supported by the Seattle Structural Genomics Center for Infectious Diseases under Contract No. HHSN272201700059C. We also thank Dalton Snyder, Benjamin Jones, and Vicki Wysocki for providing the SID device under funding by the NIH National Institute of General Sciences using grant P41GM128577 (Wysocki). This research was supported by the EBSD Mission Seed, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL), performed at the Environmental Molecular Sciences Laboratory (EMSL) under project 60268 (doi.org/10.46936/staf.proj.2021.60268/60008436) and in part by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract 75N93022C00035 (AJ) and by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act (AJ). 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.
References

nsp16/nsp10 Protein Complex. *PLOS Pathogens* 2011, 7 (10), e1002294. DOI: 10.1371/journal.ppat.1002294.


