

# **Multiplexed screening of thousands of natural products for protein-ligand binding in native mass spectrometry**

Giang T. H. Nguyen,<sup>1§</sup> Jack L. Bennett,<sup>1§</sup> Sherrie Liu,<sup>1</sup> Sarah E. Hancock,<sup>2</sup> Daniel L. Winter,<sup>3</sup> Dominic J. Glover,<sup>3</sup> and William A. Donald<sup>1\*</sup>

<sup>1</sup>*School of Chemistry, University of New South Wales, Sydney, New South Wales 2052, Australia*

<sup>2</sup>*School of Medical Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia*

<sup>3</sup>*School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia*

<sup>§</sup>These authors contributed equally

\*Correspondence to be addressed to:

William A. Donald  
Science & Engineering Building (E8), Room 601  
School of Chemistry  
UNSW Sydney  
Sydney NSW 2052  
AUSTRALIA  
Ph: +61 (2) 9385 8827  
E: [w.donald@unsw.edu.au](mailto:w.donald@unsw.edu.au)

## **ABSTRACT**

The structural diversity of natural products offers unique opportunities for drug discovery, but challenges associated with their isolation and screening can hinder the identification of drug-like molecules from complex natural product extracts. Here we introduce a mass spectrometry-based approach that integrates untargeted metabolomics with multistage, high-resolution native mass spectrometry to rapidly identify natural products that bind to therapeutically relevant protein targets. By directly screening crude natural product extracts containing thousands of drug-like small molecules using a single, rapid measurement, novel natural product ligands of human drug targets could be identified without fractionation. This method should significantly increase the efficiency of target-based natural product drug discovery workflows.

## MAIN

Natural products (NPs) have been historically renowned as useful starting points for drug development due to their inherent structural diversity, serving as an accessible source of complex chemical scaffolds that can be adapted to afford effective pharmacotherapies<sup>1</sup>. However, the utilisation of NPs in major drug discovery and development pipelines has significantly declined across the past thirty years, partly because of the technical challenges associated with screening and isolating bioactive small molecules from NP extracts<sup>1,2</sup>. Traditional approaches to NP drug discovery require bioactivity-guided fractionation, which is a laborious process that cannot be readily implemented alongside standard high-throughput drug discovery platforms. Furthermore, the identification of bioactive NPs often relies on the application of large-scale phenotypic screens. Lead compounds identified from such cell- or animal-based screens are often difficult to optimise, in part due to an initial lack of understanding regarding their underlying molecular mechanisms<sup>3</sup>. Thus, the development of novel, target-based biophysical assays for the rapid screening of NP extracts should facilitate the efficient discovery of bioactive NPs that are amenable to rational optimisation.

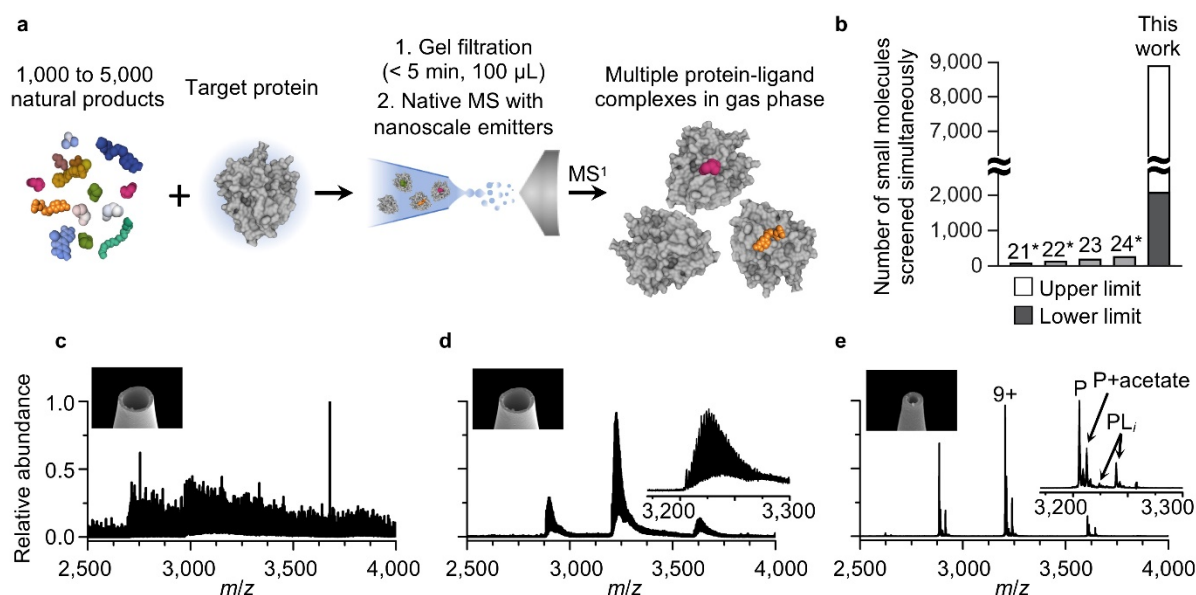
Native, non-denaturing mass spectrometry (MS) has served as a valuable approach for the rapid, direct, and sensitive measurement of protein-small molecule complexes, revealing valuable information regarding the stoichiometry, thermodynamics and kinetics of ligand binding<sup>4–7</sup>. Due to the ability to directly detect protein-ligand complexes, native MS measurements are highly resistant to common assay interferences encountered during pharmaceutical screening such as cytotoxicity, compound aggregation, and redox cycling<sup>8</sup>. Consequently, native MS experiments are typically employed for hit validation<sup>9</sup>. Unlike most established biophysical methods, native MS-based screening can be accomplished in a multiplexed manner, whereby multiple compounds within a mixture are evaluated simultaneously<sup>10</sup>. However, owing to historical limitations with native MS in resolving individual components within polydisperse biomolecular systems<sup>11</sup>, such analysis has been typically restricted to low complexity mixtures that contain only a few potential small molecule binders, hindering highly multiplexed analysis of protein-ligand interactions. This limitation has greatly restricted the screening of NP extracts by native MS, with reported methods using extensive prefractionation of extracts to simplify compound mixtures<sup>12</sup>. We therefore set out to develop a native MS-based method that was tolerant of highly complex samples and could be employed for the multiplexed screening and identification of small molecule binders from NP extracts without the need for fractionation.

Although there has been significant advances in high-resolution MS platforms<sup>13,14</sup>, native MS approaches using electrospray ionization (ESI) are generally considered incompatible with high complexity mixtures. In ESI, the adduction of non-volatile components to the protein during droplet desolvation results in significant peak broadening, effectively lowering measurement resolution<sup>15</sup>. One recent strategy in native MS to reduce the extent of adduction is to use nanoscale ion emitters, which effectively increase resolution without the requirement for gas-phase ion activation and the associated possibility of in-source ligand dissociation<sup>16,17</sup>. Here, we utilise this approach to enable the direct multiplexed screening of crude NP extracts containing thousands of small molecules against specific protein targets in native MS. A simple, yet effective, workflow was developed involving: (i) incubation of the target protein with a crude natural extract, (ii) rapid (< 5 min) gel filtration to enrich bound compounds, and (iii) native MS measurement of intact protein-ligand complexes using nanoscale ion emitters (Fig. 1a). To confidently identify novel NP ligands without fractionation, a multistage ion activation (MS<sup>n</sup>) approach was used to isolate individual intact protein-ligand complexes, dissociate the ligand through collisional activation, and subsequently fragment the released ligand for identification by comparison to a metabolomics database.

## RESULTS

### Highly complex natural product extracts can be directly screened using native MS

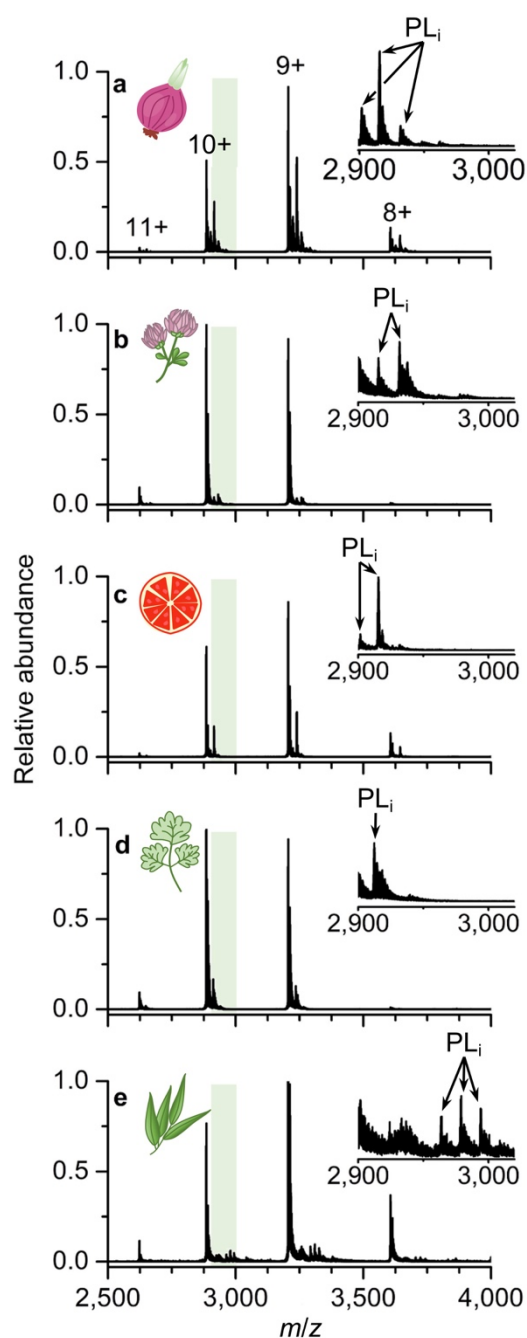
To demonstrate the feasibility of the screening approach, five distinct ethanolic plant extracts were screened against multiple soluble protein targets. Red onion peel, red clover, parsley, eucalyptus leaves and orange peels were selected as they are rich in flavonoids<sup>18,19</sup>, which is a group of phenolic NPs that demonstrate broad spectrum biological activity. A total of 4,436 compounds were identified by high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS) from the five ethanolic extracts, with up to an additional 9,547 compounds that could not be readily annotated (Supplementary Fig. 1). A comparison of the physicochemical properties of the annotated compounds to small molecule drugs approved by the United States Food and Drug Administration indicates that the NP libraries were drug-like, with 78% satisfying Lipinski's rule-of-five (Supplementary Fig. 2)<sup>20</sup>; *i.e.* the NP extracts served as relevant examples of drug-like molecules that are commonly encountered within NP drug discovery. Moreover, the samples contained at least ~8.5 times the number of distinct small molecules that were reported to be screened simultaneously in other leading native MS studies<sup>21–24</sup> (Fig. 1b).



**Figure 1. Multiplexed detection of small molecule binders from crude natural product extracts in native MS by integrating gel filtration with nanoscale ion emitters.** **a**, Natural product extracts containing thousands of identified small molecules are incubated with a target protein. Following a rapid gel filtration step, native MS measurements are conducted using nanoscale ion emitters. Using this method, multiple protein-ligand complexes (PL<sub>i</sub>) can be resolved and detected from a single measurement. **b**, Column graph displaying the highest number of potential small molecule binders screened by use of a single native MS measurement in leading studies. In-text reference numbers are indicated above each column. Asterisks indicate previous studies that did not result in individually resolved protein-ligand complexes from MS<sup>1</sup> measurements alone. Native mass spectra of human carbonic anhydrase I and crude orange peel extract acquired using **c**, ~2,000 nm nESI ion emitters, **d**, ~2,000 nm nESI ion emitters following gel filtration, and **e**, ~250 nm nESI ion emitters following gel filtration. Insets are representative scanning electron micrographs of emitter tips and spectral expansions of the 9+ charge state. Scale bars correspond to 3,000 nm for **c** and **d**, and 1,000 nm for **e**.

As a proof-of-concept, crude NP extracts were first screened against human carbonic anhydrase I (hCAI). This is a soluble ~29 kDa metalloenzyme that catalyses the hydration of CO<sub>2</sub>, facilitating pH homeostasis and enabling the transfer of CO<sub>2</sub> between blood vessels and surrounding tissue<sup>25</sup>. Isozyme-specific CA inhibitors are currently being developed to treat several disease conditions including cancer, with a recent emphasis on improving their selectivity<sup>26</sup>. Native MS analysis of hCAI incubated with each NP extract (5 mg/mL) using conventional ion emitters resulted in highly complex spectra in which adjacent protein charge states and individual protein-ligand complexes could not be resolved (Fig. 1c). High concentrations of small, non-volatile molecules in native MS samples can result in extensive adduction and prevent the resolution of bound ligands. Thus, to remove any non-bound compounds from the mixture and effectively enrich the sample with protein-ligand complexes, a low-volume gel filtration step was used before native MS experiments. Compared to alternate desalting approaches that were evaluated, gel filtration was more rapid, thereby minimising the dissociation of

NP-protein complexes during sample preparation. This rationale was supported by a comparison of the experimental data to a binding model which suggested that dissociation was minimal (*vide infra*). The use of gel filtration with conventional ion emitters resulted in spectra in which peaks corresponding to individual charge states could be resolved (Fig. 1d). However, the observed peaks were broad with an elevated baseline and thus, distinct ligand-bound protein complexes could not be resolved (Fig. 1d). In contrast, the use of rapid, native gel filtration and nanoscale ion emitters resulted in well-resolved *apo* and ligand-bound protein ions for all charge states (Fig. 1e). Similar results were obtained for all five NP extracts tested, with up to three distinct protein-bound ligands identified from a single extract and a total of 11 protein-bound ligands ( $K_d$  values from ~20 to 370  $\mu$ M) observed across all samples (Fig. 2, Supplementary Table 1, 2). Thus, it is evident that this approach can be used to significantly increase the number of small molecules that can be screened simultaneously by native MS (Fig. 1b).



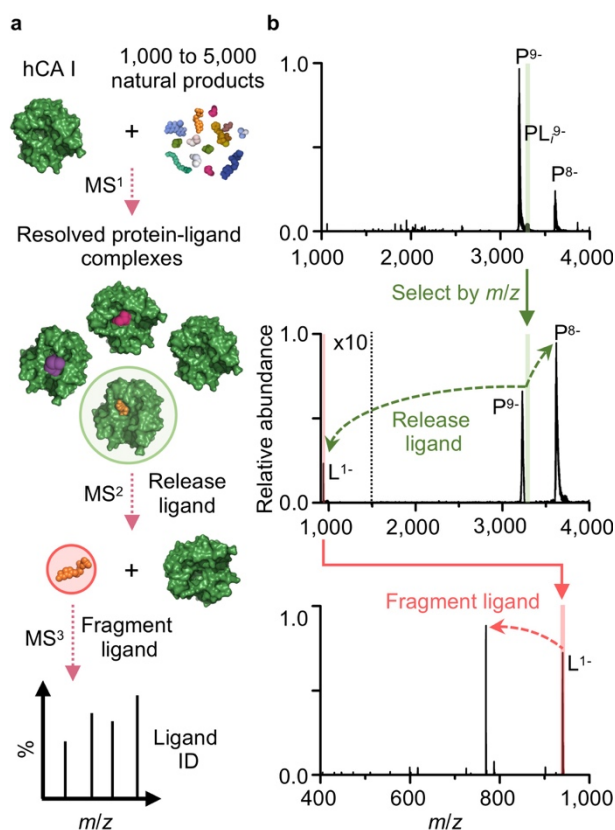
**Figure 2. High-resolution native MS analysis of complexes formed between human carbonic anhydrase I and ligands from crude natural product extracts.** Native mass spectra of human carbonic anhydrase I and crude natural product extracts from **a**, red onion peels (838-3,336 compounds), **b**, red clover (1,580-4,699 compounds), **c**, orange peels (2,097-8,919 compounds), **d**, parsley leaves (1,611-5,376 compounds), and **e**, eucalyptus leaves (1,800-6,605 compounds) obtained using ~250 nm emitters and native gel filtration. Insets correspond to an x-axis expansion of the 9+ charge state for the ligand-bound protein complexes (PL<sub>i</sub>) as indicated by the shaded green boxes.

### Bound ligands can be directly identified using multistage ion activation

To identify small molecule binders, the molecular weights of the bound ligands were obtained from the measured mass-to-charge ratio ( $m/z$ ) and charge state ( $z$ ) of the *apo* and *holo* protein ions. Comparison of these masses to a candidate list generated from metabolomics experiments resulted in

one or two plausible assignments for each ligand (Supplementary Table 1). Thus, additional experimental approaches were required to confidently deduce ligand identity. In one approach, quantitative native MS experiments<sup>17</sup> using the isolated compounds were used to confirm that the ligands bind and to obtain their binding affinities (Supplementary Table 2, 3). In a second approach, some of the ligands could be effectively identified during the direct analysis of the NP extracts by leveraging the capacity of a linear ion trap to perform multistage tandem mass spectrometry (MS<sup>n</sup>) measurements. Multistage ion activation approaches in native MS have been recently used for the identification of: (i) lipids co-purified with membrane proteins<sup>27</sup> and (ii) ligands of peptide- and glycan-binding proteins upon screening synthetic and natural libraries containing up to 247 structurally-related small molecules (Fig. 1b)<sup>21-24</sup>. Using a multistage approach, a previously unknown ligand (1,2,3,4,6-penta-O-galloyl- $\beta$ -D-glucose) of hCAI was identified from the eucalyptus extract and found to bind to the target with micromolar affinity (Supplementary Table 4). This involved isolating the ligand-bound protein (9- charge state,  $m/z$  3310.3), dissociating and isolating the ligand (1- charge state,  $m/z$  939.1), and subsequently activating the ligand using ion-neutral collisions to obtain a ligand fragmentation spectrum (Fig. 3). The resulting fragmentation pattern closely matched those obtained directly from: (i) an authentic standard of 1,2,3,4,6-penta-O-galloyl- $\beta$ -D-glucose (Supplementary Fig. 3); and (ii) both the metabolomics data and MS<sup>2</sup> database for the assigned compound. Although this multistage ion activation approach could not be readily used for all hits owing to the low mass cut-off of the ion trap, a recently reported MS instrument capable of isolating, activating and detecting ions across a broader range of  $m/z$  values within a single measurement<sup>27</sup> should be useful for addressing this limitation. These results indicate that multistage ion activation can be used to directly identify small molecules bound to druggable protein targets in native MS, thereby rapidly increasing the efficiency of target-based screening in NP drug discovery.





**Figure 3. Multistage ion activation (MS<sup>3</sup>) enables direct identification of protein-bound small molecules in native mass spectrometry.** **a**, Schematic of the key steps that were used to identify protein-bound ligands by native mass spectrometry. A protein-ligand complex of interest is resolved from the unbound protein and other protein-ligand complexes (MS<sup>1</sup>), isolated, and released using ion-neutral collisions (MS<sup>2</sup>). Subsequent isolation and fragmentation of the ligand (MS<sup>3</sup>) produces fragment ions that can be matched to a database or authentic standard to confirm ligand identity. **b**, Application of this workflow to protein-ligand complexes (PL<sub>i</sub>) formed between human carbonic anhydrase I (P) and unknown molecules from an ethanolic eucalyptus leaf extract facilitated the identification of a protein-bound tannin, 1,2,3,4,6-penta-*o*-galloyl- $\beta$ -D-glucose (L).

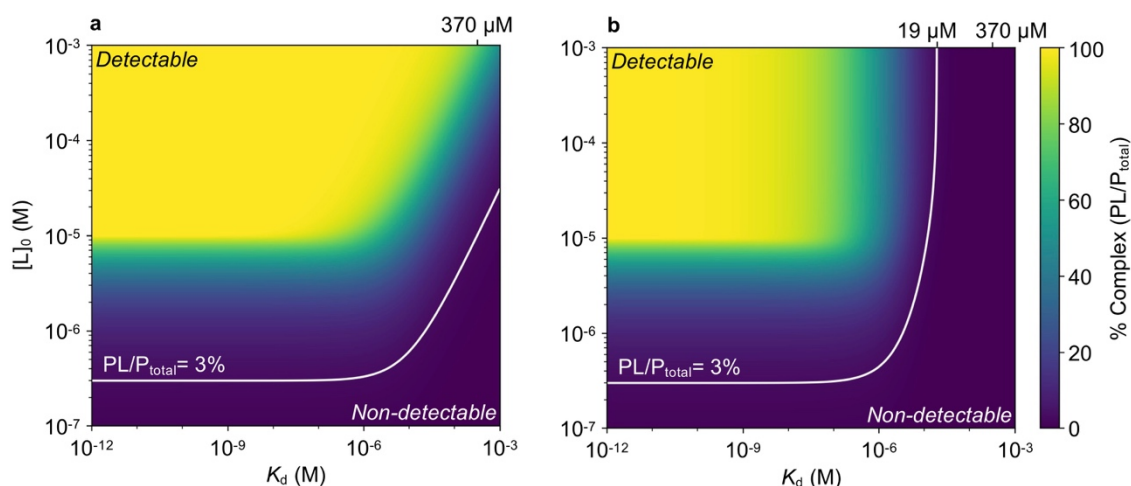
### Native MS-based screening to identify novel natural product ligands

To further demonstrate the utility of the multiplexed approach, the natural product extracts were screened against three other soluble proteins. These included two more carbonic anhydrase isoforms, bovine CAII and human hCAVII (hCAVII), as well as the functionally distinct enzyme lysozyme, which catalyses the hydrolysis of  $\beta$ -1,4-glycosidic linkages within bacterial cell walls and acts as a component of the innate immune response. Altogether, 30 protein-ligand complexes were detected and assigned across 16 measurements, corresponding to 14 distinct small molecule binders (Supplementary Fig. 4, 5, Supplementary Table 1). Most of the hits against the CA isoforms were flavonoids, with 12 previously reported as inhibitors of at least one CA isozyme. In addition to the pentagalloyl hexose, ambocin was identified as a novel bovine CAII ligand that had not been previously reported to bind to any CA isozyme. For lysozyme, screening of the red onion extract led to the identification of isoquercetin ligand of this

enzyme (Supplementary Fig. 6, Supplementary Table 1), which has not been previously reported. Quantitative native MS analysis using the purified compound resulted in a measured dissociation constant of  $66.0 \pm 5.0 \mu\text{M}$ . A multistage ion activation approach was used to confirm the ligand identity, resulting in fragment ions that closely matched that of a reference standard (Supplementary Fig. 6, 7).

### **Fundamental detection limits**

To examine some fundamental limits of the screening method, an equilibrium binding model was developed (based on analytically derived equations, see supplementary methods) to predict the abundances of protein-ligand complexes following sample preparation as a function of the initial ligand concentration and the protein-ligand dissociation constant (Fig. 4). Two extreme scenarios were considered in which: (i) the protein-ligand complex undergoes no dissociation during gel filtration and prior to native MS (Fig. 4a), and (ii) the system fully re-equilibrates after gel filtration (Fig. 4b). In both cases, it was observed that ligands with micromolar or sub-micromolar affinity should be detected if their concentration in the initial mixture is greater than 300 nM. By assuming that all ~5000 small molecules in the extracts are present at the same concentration, the average concentration of each small molecule is calculated to be ~3,300 nM. Thus, the majority of high-affinity binders present in the extracts should be detectable. Moreover, in the extreme case in which full equilibrium occurs after filtration, the highest detectable  $K_d$  is predicted to be 19  $\mu\text{M}$ . Given that multiple ligands with  $K_d$  values far higher than 19  $\mu\text{M}$  were detected (i.e. up to 370  $\mu\text{M}$ ), it is evident that the initial equilibrium concentrations of the bound complexes and unbound protein can be partially 'kinetically trapped' by use of rapid separation and detection methods.



**Figure 4. Equilibrium binding model demonstrates nanomolar limit of detection for micromolar and sub-micromolar affinity ligands.** Contour plots of the predicted relative abundance of the protein-ligand complex (PL) as a function of both the initial ligand concentration ( $[L]_0$ ) and the protein-ligand dissociation constant ( $K_d$ ). The white isoline indicates all sets of conditions at which PL constitutes 3% of the total protein population ( $P_{\text{total}}$ , 10  $\mu\text{M}$ ). Ligands that have concentrations and  $K_d$  values that lie to the upper left of this curve should be theoretically detectable using the detailed native MS screening method. Two different sets of assumptions were applied to represent a 'best-of' and 'worst-of' situation. For **a**, it was assumed that protein-ligand complexes undergo no dissociation after rapid gel-filtration. For **b**, it was assumed that the mixture is allowed to fully re-equilibrate following gel filtration. In the extreme case of **b**, ligands with  $K_d$  values greater than  $\sim 19 \mu\text{M}$  cannot be detected. Ligands with  $K_d$  values up to  $370 \mu\text{M}$  were detected experimentally.

## DISCUSSION

The work described here provides a powerful strategy to efficiently screen complex mixtures of small molecules for binding to a protein target. By combining nanoscale ion emitters<sup>16,17</sup> with a rapid, low-volume gel filtration step, protein-small molecule interactions can be directly observed from mixtures containing up to  $\sim 8,900$  potential small molecule ligands in a measurement that takes a few minutes. By leveraging the capability of an ion trap to perform multistage MS experiments, bound ligands can be confidently identified without the need for sample fractionation and/or compound isolation. Compared to previously reported approaches using native MS for compound screening, this workflow offers over an  $\sim 8.5$ -fold increase in efficiency.<sup>21–24</sup> Furthermore, due to the improved resolution over previously reported measurements, protein-ligand complexes could be directly resolved in the MS<sup>1</sup> spectra thus, gas-phase dissociation is not necessarily required to assign bound ligands. This would be advantageous in the case of ligands that tend to dissociate in the gas-phase as neutral species, or in studies aiming to integrate the screening workflow with ion activation methods capable of localising ligand binding.<sup>28,29</sup> The improved resolution offered by this approach should also enhance studies using high-resolution native MS for structural biology<sup>11</sup>, in which increases in resolution can potentially reveal

surprising effects of small modifications (e.g. post-translational modifications, salt/ligand binding) on macromolecular structure.<sup>30</sup>

It is expected that this method can be readily applied within target-based NP drug discovery workflows to confidently identify small molecule binders without the need for the fractionation of extracts prior to (or following) screening<sup>4</sup>. This should facilitate highly efficient hit dereplication and potentially enable the rapid and rational prioritisation of specific NPs for further investigation based on their structural features. Using this approach, it is anticipated that large groups of NP extracts that collectively contain more than 100,000 compounds could be screened per day, accelerating target-based drug discovery campaigns without the resource intensive collection, curation, and storage of large libraries of pure natural products or fractionated NP extracts. Publicly available libraries such as the National Cancer Institute's Natural Products Repository consist of over 260,000 extracts that represent an untapped source of chemical diversity. Using direct, multiplexed approaches such as native MS to screen these libraries against druggable targets should enable drug leads to be confidently identified without access to the specialized robotics and detection systems used for high-throughput screening. Furthermore, related workflows should be beneficial in directly linking the metabolome and proteome by facilitating the identification of endogenous protein-ligand complexes to expand our knowledge of protein-small molecule interaction networks<sup>31,32</sup>. Ultimately, the integration of nanoscale ion emitters with rapid gel filtration in native MS should enable highly efficient, ultrahigh throughput screening of bioactive compounds and substantially accelerate biochemical discovery.

## **METHODS**

### **Native Mass Spectrometry: General Procedures**

All native mass spectrometry experiments were performed using a hybrid linear ion trap-Orbitrap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific) equipped with a nanoESI source. Prior to MS measurements, instrument calibration was performed using the Pierce Positive and Negative Ion Calibration Solutions (Thermo Fisher Scientific). For nESI-MS measurements, emitter tips containing 3-5  $\mu$ L of analyte solution were positioned ~2-3 mm from the heated (200 °C) capillary entrance of the mass spectrometer. Application of a  $\pm 0.7$ -1.5 kV potential between the emitter tip and the instrument sampling interface was used to generate gas-phase ions by ESI. To minimise in-source dissociation of gas-phase protein-ligand complexes, differences between the potentials applied to adjacent rf-ion guides and/or lenses were minimised throughout the instrument while still facilitating

sufficient ion transmission<sup>33</sup>. Data were processed using vendor software and analysed using both the mMass software package (<http://www.mmass.org/>), and UniDec<sup>34</sup>.

### Natural Product Screening

Crude natural extracts were reconstituted in methanol at a concentration of 100 mg/mL and sonicated for 10 min to assist in dissolution. Stock solutions were subsequently diluted in 50 mM ammonium acetate (pH 7.0) to a final concentration of 10 mg/mL.

Aqueous solutions containing 20  $\mu$ M of the target protein of interest (50  $\mu$ L) and 2 mg/mL of the natural extract (50  $\mu$ L) in 50 mM ammonium acetate (pH 7.0) were combined and incubated at 4 °C for 24 h, then 25 °C for a further hour. No precipitate could be observed at this time, suggesting most compounds present in each of the extracts were soluble under these conditions. Gel filtration (Zeba™ Spin Desalting Columns, 7K MWCO, 0.5 mL; Thermo Fisher Scientific) was subsequently used to remove unbound small molecules, including those that are non-volatile and can adversely impact the quality of native mass spectra. The solutions containing the target protein and natural extracts were desalted using three spin desalting columns. In each cycle, solutions (100  $\mu$ L) were loaded into the cartridge and centrifuged at 1,500 *g* at 4 °C for 1.5 min.

For MS<sup>1</sup> measurements used to identify protein-ligand complexes, the instrument was operated in positive ionization mode. The Orbitrap was set to scan over a high *m/z* range (1,000–4,000) with a resolution of 100,000 at *m/z* 200, automatic gain control (AGC) target of  $2.0 \times 10^5$  ions, and a maximum injection time of 500 ms. For MS<sup>*n*</sup> experiments to identify unknown ligands, ions were detected in the ion trap in both positive and negative modes. Measurement of both ion polarities ensures that both acidic and basic ligands can be detected as charged species following gas-phase dissociation. Furthermore, analysing both the cationic and anionic ligand ions increases the number of fragment ions that can be used to assist in identifying the ligand. Following ESI and ion focusing, complexes were isolated (100 *m/z* isolation window) in the linear ion trap and subsequently activated using collision-induced dissociation (CID). The normalised collision energy (NCE) was selected (30–70%) to almost completely deplete the protein-ligand complex and form the charge-stripped *apo* protein. By lowering the resonance excitation frequency (corresponding to an activation *q* value of 0.2), dissociated ligands at low *m/z* could also be trapped. The dissociated ligand was subsequently isolated in the ion trap (10 *m/z* isolation window) and fragmented by CID until the precursor ion abundance was depleted by ~50% using an NCE of ~10–30%. At this stage, detection of ligand fragments typically required a long maximum ion injection time (1,000 ms) and increased AGC targets to increase ion abundances. For

CID of the authentic ligand standards, the ionized ligand was isolated in the ion trap using the same mass isolation and ion activation parameters as in the native MS experiment.

To assign the identity of bound ligands, the charge states ( $z$ ) of all peaks in the spectra with a signal to background chemical noise ratio greater than 3 were determined based on the spacing of adjacent isotopes. The mass of individual species could then be calculated as follows: mass (Da) =  $m/z \times z - z \times 1.00727647$ . Ligand masses were calculated as the difference in mass between the unbound and ligand-bound protein. At least three replicate measurements produced 95% confidence intervals for the ligand mass which were used to search the complete list of compounds identified by untargeted metabolomics. Multiple compounds ( $\leq 3$ ) were found to be plausible for some spectral peaks. In this case, a multistage ion activation approach could be employed to determine the ligand identity. However, due to the lower mass limit of the linear ion trap after ion isolation this approach was not achievable for small ligands bound to larger proteins. In these cases, a search of the ChEMBL database was used to identify if any of the plausible compounds had been previously identified as inhibitors of CA isozymes, facilitating dereplication of known CA inhibitors.

## ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Australian Research Council (DP190103298, FT200100798); the Bioanalytical Mass Spectrometry Facility and Electron Microscope Unit of the UNSW Mark Wainwright Analytical Centre for access to analytical instrumentation; Dr Russell Pickford (Bioanalytical Mass Spectrometry Facility) for assistance with early investigations; and the UNSW Recombinant Products Facility for assistance in the production and purification of proteins used in this study. Professors Claudiu T. Supuran (University of Florence), Clemente Capasso (IBBR-CNR), Paul Ashworth (UNSW Sydney), and James M. Hook (UNSW Sydney) are thanked for useful discussions. Professor David A. Jacques (UNSW Sydney) is thanked for providing SuperTEV protease. Yiling Liu, Daniel Loo, and Manatsu Nose are thanked for assistance in some of the experiments.

## COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

## REFERENCES

1. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).

2. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111–129 (2015).
3. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
4. Bennett, J. L., Nguyen, G. T. H. & Donald, W. A. Protein–Small Molecule Interactions in Native Mass Spectrometry. *Chem. Rev.* (2021). doi:10.1021/acs.chemrev.1c00293
5. Eschweiler, J. D., Kerr, R., Rabuck-Gibbons, J. & Ruotolo, B. T. Sizing Up Protein–Ligand Complexes: The Rise of Structural Mass Spectrometry Approaches in the Pharmaceutical Sciences. *Annu. Rev. Anal. Chem.* **10**, 25–44 (2017).
6. Marchand, A., Czar, M. F., Eggel, E. N., Kaeslin, J. & Zenobi, R. Studying biomolecular folding and binding using temperature-jump mass spectrometry. *Nat. Commun.* **11**, 566 (2020).
7. Cong, X. *et al.* Determining Membrane Protein-Lipid Binding Thermodynamics Using Native Mass Spectrometry. *J. Am. Chem. Soc.* **138**, 4346–4349 (2016).
8. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).
9. Sacco, M. D. *et al.* Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. *Sci. Adv.* **6**, eabe0751 (2020).
10. Nguyen, G. T. H. *et al.* Perfluoroalkyl Substances of Significant Environmental Concern Can Strongly Inhibit Human Carbonic Anhydrase Isozymes. *Anal. Chem.* **92**, 4614–4622 (2020).
11. Tamara, S., den Boer, M. A. & Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* (2021). doi:10.1021/acs.chemrev.1c00212
12. Yang, B. *et al.* Bioaffinity Mass Spectrometry Screening. *J. Biomol. Screen.* **21**, 194–200 (2016).
13. Gault, J. *et al.* High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods* **13**, 333–336 (2016).
14. Li, H., Nguyen, H. H., Loo, R. R. O., Campuzano, I. D. G. & Loo, J. A. An integrated native mass spectrometry and topdown proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **10**, 139–148 (2018).
15. Lössl, P., Snijder, J. & Heck, A. J. R. Boundaries of mass resolution in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* **25**, 906–917 (2014).
16. Susa, A. C., Xia, Z. & Williams, E. R. Native Mass Spectrometry from Common Buffers with Salts That Mimic the Extracellular Environment. *Angew. Chemie Int. Ed.* **56**, 7912–7915 (2017).
17. Nguyen, G. T. H. *et al.* Nanoscale Ion Emitters in Native Mass Spectrometry for Measuring Ligand-Protein Binding Affinities. *ACS Cent. Sci.* **5**, 308–318 (2019).
18. Bhagwat, S., Haytowitz, D. B. & Holden, J. M. *USDA Database for the Flavonoid Content of Selected Foods.* (2014).
19. Moore, B. D., Wallis, I. R., Wood, J. T. & Foley, W. J. Foliar Nutrition, Site Quality, and Temperature Influence Foliar Chemistry of Tallowood (*Eucalyptus microcorys*). *Ecol. Monogr.* **74**, 553–568 (2004).
20. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **64**, 4–17 (2012).
21. Park, H. *et al.* Mass Spectrometry-Based Shotgun Glycomics for Discovery of Natural Ligands of Glycan-Binding Proteins. *Anal. Chem.* **92**, 14012–14020 (2020).
22. Han, L., Kitova, E. N., Tan, M., Jiang, X. & Klassen, J. S. Identifying carbohydrate ligands of a

- norovirus P Particle using a catch and release electrospray ionization mass spectrometry assay. *J. Am. Soc. Mass Spectrom.* **25**, 111–119 (2014).
23. El-Hawiet, A., Shoemaker, G. K., Daneshfar, R., Kitova, E. N. & Klassen, J. S. Applications of a catch and release electrospray ionization mass spectrometry assay for carbohydrate library screening. *Anal. Chem.* **84**, 50–58 (2012).
  24. Gao, J. *et al.* Screening derivatized peptide libraries for tight binding inhibitors to carbonic anhydrase II by electrospray ionization-mass spectrometry. *J. Med. Chem.* **39**, 1949–1955 (1996).
  25. Supuran, C. T. Carbonic anhydrases: Novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov.* **7**, 168–181 (2008).
  26. Alterio, V., Di Fiore, A., D'Ambrosio, K., Supuran, C. T. & De Simone, G. Multiple binding modes of inhibitors to carbonic anhydrases: How to design specific drugs targeting 15 different isoforms? *Chem. Rev.* **112**, 4421–4468 (2012).
  27. Gault, J. *et al.* Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nat. Methods* **17**, 505–508 (2020).
  28. Li, H., Wongkongkathep, P., Van Orden, S. L., Ogorzalek Loo, R. R. & Loo, J. A. Revealing ligand binding sites and quantifying subunit variants of noncovalent protein complexes in a single native top-down fticr ms experiment. *J. Am. Soc. Mass Spectrom.* **25**, 2060–2068 (2014).
  29. Cammarata, M. B., Thyer, R., Rosenberg, J., Ellington, A. & Brodbelt, J. S. Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **137**, 9128–9135 (2015).
  30. Agasid, M. T., Sørensen, L., Urner, L. H., Yan, J. & Robinson, C. V. The Effects of Sodium Ions on Ligand Binding and Conformational States of G Protein-Coupled Receptors - Insights from Mass Spectrometry. *J. Am. Chem. Soc.* **143**, 4085–4089 (2021).
  31. Vidal, M., Cusick, M. E. & Barabási, A. L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
  32. Szklarczyk, D. *et al.* STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **44**, D380–D384 (2016).
  33. Hernández, H. & Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2**, 715–726 (2007).
  34. Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).