

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

Giang T. H. Nguyen,^{1§} Jack L. Bennett,^{1§} Sherrie Liu,¹ Sarah E. Hancock,² Daniel L. Winter,³ Dominic J.
Glover,³ and William A. Donald^{1*}

¹School of Chemistry, University of New South Wales, Sydney, New South Wales 2052, Australia

²School of Medical Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia

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

[§]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

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.

INTRODUCTION

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.¹ 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.^{1,2} 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.³ 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.⁴⁻⁷ 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.⁸ Consequently, native MS experiments are typically employed for hit validation.⁹ 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.¹⁰ However, owing to historical limitations with native MS in resolving individual components within polydisperse biomolecular systems,¹¹ 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.¹² 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,^{13,14} 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.¹⁵ 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.^{16,17} 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 (MSⁿ) 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.

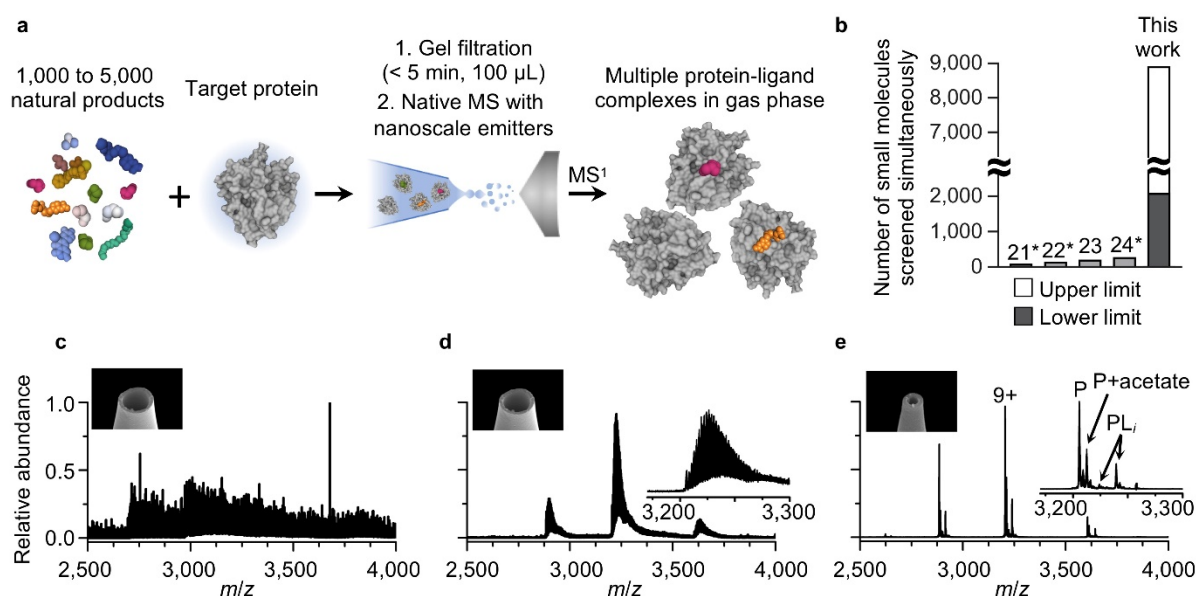


Figure 1. Multiplexed detection of small molecule binders from crude natural product extracts in native MS by integrating gel filtration with nanoscale ion emitters. **a**, Natural product extracts containing thousands of identified small molecules are incubated with a target protein. Following a rapid gel filtration step, native MS measurements are conducted using nanoscale ion emitters. Using this method, multiple protein-ligand complexes (PL_i) 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 MS¹ 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**.

RESULTS AND DISCUSSION

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,^{18,19} 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 (Fig. S1). 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 (Fig. S2);²⁰ *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^{21–24} (Fig. 1b).

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 CO₂, facilitating pH homeostasis and enabling the transfer of CO₂ between blood vessels and surrounding tissue.²⁵ Isozyme-specific CA inhibitors are currently being developed to treat several disease conditions including cancer, with a recent emphasis on improving their selectivity.²⁶ 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 μM) observed across all samples (Fig. 2, Table S1, S2). 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).

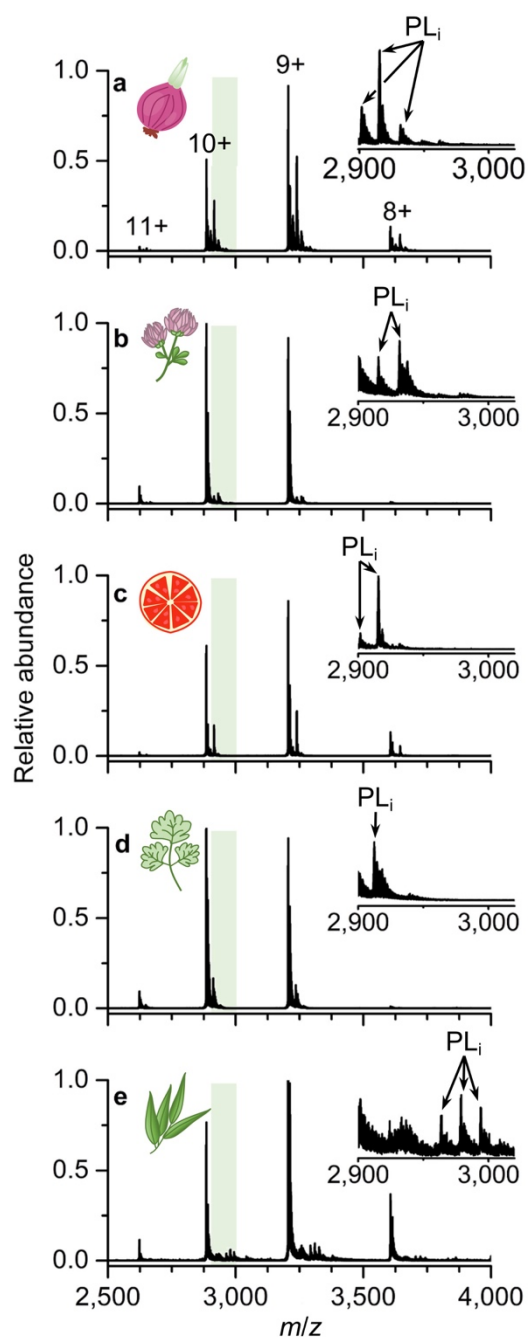


Figure 2. High-resolution native MS analysis of complexes formed between human carbonic anhydrase I and ligands from crude natural product extracts. Native mass spectra of human carbonic anhydrase I and crude natural product extracts from **a**, red onion peels (838-3,336 compounds), **b**, red clover (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_{*i*}) 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 (Table S1). Thus, additional experimental approaches were required to confidently deduce ligand identity. In one approach, quantitative native MS experiments¹⁷ using the isolated compounds were used to confirm that the ligands bind and to obtain their binding affinities (Table S2, S3). 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^n) measurements. Multistage ion activation approaches in native MS have been recently used for the identification of: (i) lipids co-purified with membrane proteins²⁷ 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).²¹⁻²⁴ Using a multistage approach, a previously unknown ligand (1,2,3,4,6-penta-O-galloyl- β -D-glucose) of hCAI was identified from the eucalyptus extract and found to bind to the target with micromolar affinity (Table S4). 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- β -D-glucose (Fig. S3); and (ii) both the metabolomics data and MS^2 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 measurement²⁷ 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.

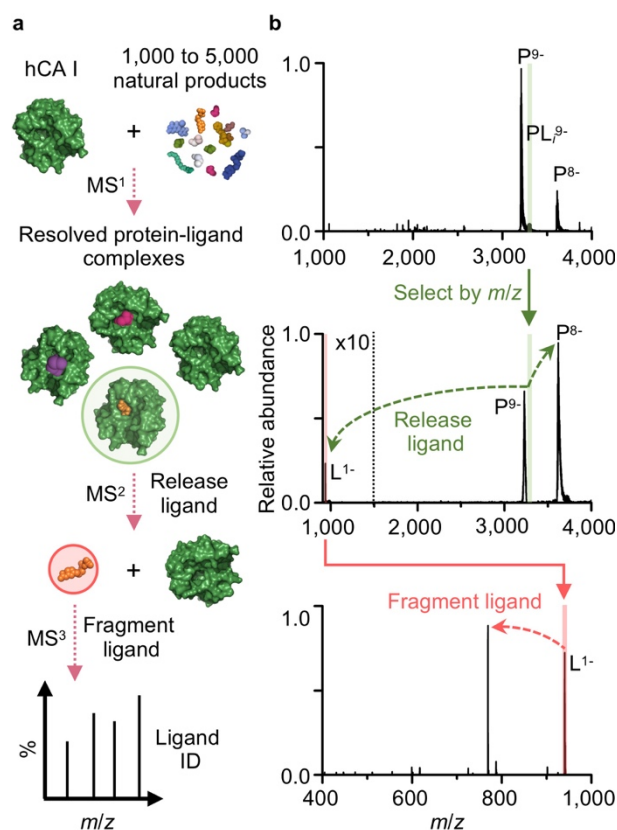


Figure 3. Multistage ion activation (MS³) enables direct identification of protein-bound small molecules in native mass spectrometry.

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 (MS¹), isolated, and released using ion-neutral collisions (MS²). Subsequent isolation and fragmentation of the ligand (MS³) 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_{*n*}) 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- β -*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 β -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 (Fig. S4, S5, Table S1). 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 (Fig. S6, Table S1), 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 (Fig. S6, S7).

Fundamental detection limits

To examine some fundamental limits of the screening method, an equilibrium binding model was developed (based on analytically derived equations, see Experimental Section) 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 μM . Given that multiple ligands with K_d values far higher than 19 μM were detected (i.e. up to 370 μM), 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.

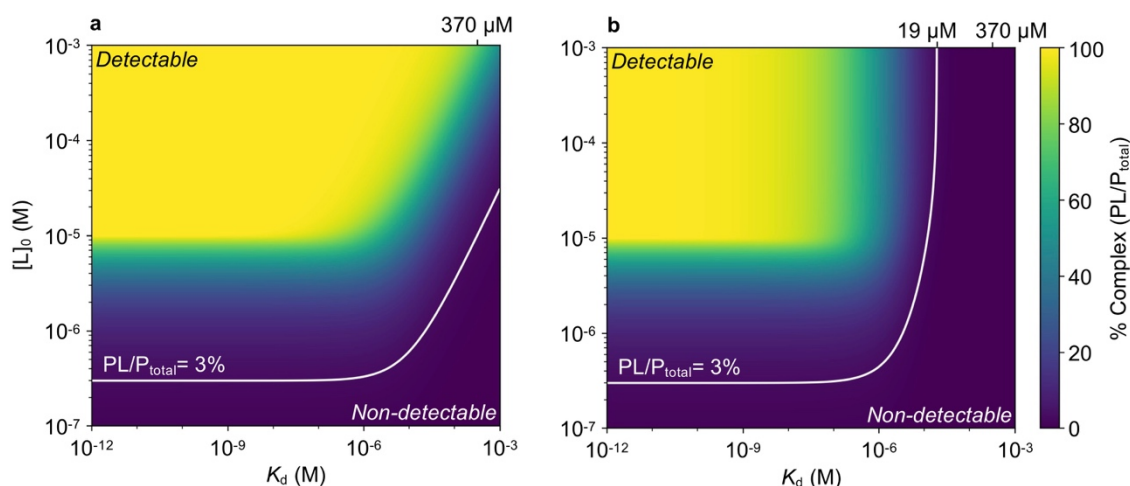


Figure 4. Equilibrium binding model demonstrates nanomolar limit of detection for micromolar and sub-micromolar affinity ligands.

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_{total} , 10 μ 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 ~ 19 μ M cannot be detected. Ligands with K_d values up to 370 μ M were detected experimentally.

CONCLUSIONS

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^{16,17} 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 ~ 8.5 -fold increase in efficiency.^{21–24} Furthermore, due to the improved resolution over previously reported measurements, protein-ligand complexes could be directly resolved in the MS¹ 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.^{28,29} The improved resolution offered by this approach should also enhance studies using high-resolution native MS for structural biology,¹¹ in which increases in resolution can potentially reveal surprising effects of small modifications (e.g. post-translational modifications, salt/ligand binding) on macromolecular structure.³⁰

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.⁴ 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.^{31,32} 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.

EXPERIMENTAL SECTION

Materials

All reagents used for protein expression and purification were purchased from Sigma-Aldrich and used without further purification. Analytical reagent-grade ethanol was purchased from Ajax Finechem. HPLC-grade acetonitrile, methanol, formic acid (~98% purity) and ammonium acetate (≥99.0% purity) were purchased from Honeywell. Deionised water with a resistivity of 18.2 MΩ/cm was obtained from an ultrapure water purification system (Milli-Q® IQ 7000; Millipore).

Preparation of Natural Product Extracts

5-15 g of red onions (*Allium cepa*), dried red clover (*Trifolium pratense*), mature parsley (*Petroselinum crispum*), navel oranges (*Citrus sinensis*) and tallowwood eucalyptus (*Eucalyptus microcorys*) were each extracted with 100 mL ethanol in a Soxhlet extractor for 6 h. Extracts were dried *in vacuo* at 37 °C, then stored as solids at -20 °C until use. For more details regarding each extract, see the SI text.

Preparation of Proteins for Screening

All proteins were purchased from commercial suppliers, excluding human carbonic anhydrase VII which was obtained by heterologous expression in *Escherichia coli* (see SI text for details). Lysozyme from chicken egg

white (Sigma L6876), carbonic anhydrase I from human erythrocytes (Sigma C4396), carbonic anhydrase II from human erythrocytes (Sigma C2522) and bovine carbonic anhydrase II (Sigma C7749) were reconstituted in water at approximately 0.5 mg/mL. Aqueous stock solutions for commercial proteins were subsequently diluted into 50 mM ammonium acetate, pH 7.0 to a concentration of 40 μ M as measured by a microvolume UV-Vis spectrophotometer at 280 nm (DS-11; DeNovix) and stored at -20 °C until use.

Metabolomics

The dried natural product extracts were prepared for LC-MS/MS analysis by reconstituting the crude extract in a 1:1 (v/v) water:acetonitrile solution to a final concentration of 1 mg/mL. Residual solids were removed from the mixture using centrifugation (15,000 g, 15 min, 4 °C) before the supernatant was transferred to autosampler vials and stored at 4 °C. LC-MS/MS measurements were performed using an ultra-high performance liquid chromatography system (UltiMate 3000; Thermo Fisher Scientific) coupled to a hybrid quadrupole-Orbitrap mass spectrometer (QExactive HF; Thermo Fisher Scientific). Samples (2 μ L) were injected onto a Acquity CSH C18 Column (pore diameter 130 Å, particle diameter 1.7 μ m, column inner diameter 3.0 mm, column length 100 mm; Waters). Chromatography was performed at 45 °C using 0.1% (v/v) formic acid in water (A) and acetonitrile (B) at 250 μ L/min with the following gradient: 0-2 min, 0% B; 2-22 min, 0-50% B; 22-24 min, 50-100% B; 24-26 min, 100% B. A 5 min equilibration at 0% B was employed between sequential injections to ensure measurement reproducibility.

Following analyte separation, the column eluate was infused into a heated electrospray ionisation probe (50 °C; HESI-II; Thermo Fisher Scientific). Ions were generated by the application of either a +4 or -3.5 kV potential to the ESI probe relative to the instrument sampling interface for measurements conducted in positive and negative mode, respectively. To facilitate sensitive detection of ions across the expected m/z range, instrument source conditions were set as follows: capillary temperature 280 °C, sheath gas flow 25, auxiliary gas flow 10, and S-lens RF level 50 (arbitrary units). Survey scans were acquired over the m/z range 100-1,500, with a resolution of 120,000 at m/z 200. MS/MS spectra were acquired for the twenty most abundant ions detected in each survey scan using higher energy collisional dissociation (HCD) fragmentation (1.4 m/z isolation window, normalised collision energy 22.5). The resolution for HCD spectra was set to 15,000 at m/z 200 with a maximum ion injection time of 28 ms.

Data from triplicate LC-MS/MS runs in both positive and negative mode were processed and analysed using vendor software (Compound Discoverer 3.1; Thermo Fisher Scientific). Briefly, scans were read without filtering prior to alignment using an adaptative curve regression model with a maximum retention time shift of 0.5 min and a mass tolerance of 5 ppm for feature matching. Unknown compounds were subsequently detected

from the processed data using the following parameters: mass tolerance 3 ppm, intensity threshold 1%, signal-to-noise ratio threshold 3, minimum peak intensity 250,000, charge states $\pm 1-2$. Sodium and ammonium adducts were considered, in addition to analyte dimers and fragment ions formed through the neutral loss of either ammonia or water. The maximum element count was set to C90 H190 Br3 Cl4 K2 N10 Na2 O15 P2 S5. Following compound detection, grouping of ions formed from the same analyte was performed with a mass tolerance of 3 ppm and a retention time tolerance of 0.2 min. Any compounds that were also present in a solvent blank were removed prior to further analysis. To identify metabolites, the MS/MS spectra were compared to the mzCloud reference library (date accessed 29/01/21) using the Similarity Forward algorithm (match factor threshold of 60). It is important to note that estimates of the number of compounds in the sample likely represent the upper limit of the number of compounds with polar functionalities present in the sample due to the possibility of in-source ion fragmentation.

To assess the suitability of the extracts for application in drug discovery, we analysed the physicochemical properties of the identified compounds using the chemoinformatic tool Instant JChem (<https://chemaxon.com/>). Briefly, the names of all compounds identified by LC-MS/MS were converted to SMILES format under operator supervision prior to the automated calculation of physicochemical properties using predefined functions. The structures of FDA-approved drugs were sourced from the ZINC compound database using the 'fda' filter.³³

Fabrication of Nanoscale Ion Emitters

Single barrel borosilicate capillaries (1.2 mm o.d./0.68 mm i.d., Harvard Apparatus) were pulled to form tips with different internal diameters using a Flaming/Brown micropipette puller (P-97; Sutter Instruments) equipped with a box filament (FB255B; Sutter Instruments)¹⁷. The parameters used to pull tips of different size are as follows: tip internal diameter $\sim 2,000$ nm, heat 500, pull 10, velocity 15, time 250, pressure 500; tip internal diameter 250 nm, heat 500, pull 30, velocity 50, time 250, pressure 500. The inner diameters of the fabricated emitters were measured using scanning electron microscopy (Nova NanoSEM 450; Thermo Fisher Scientific). These measurements are performed at least every three months to ensure that the tip sizes remain constant over time, and this process may necessitate slight changes in instrument parameters, particularly after general maintenance of the puller (e.g. changing drying agent and changing the filament). We have also observed that the degree of tip taper is important in ensuring the nanoscale emitters function reliably and are not damaged during handling. The degree of the tip taper can be readily evaluated by optical microscopy (Fig. S8). To make the nanoscale ion emitters conductive, the tips were sputter coated for 30 s with a mixture of gold and palladium (1.25 kV; 30 mA, 0.1 mbar; Scancoat Six; Edwards). The resulting metal coating is relatively weakly bound to the glass

so care must be taken when handling the tips. For example, the tips are usually handled with gloves to prevent oils from disrupting the coating.

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 μL of analyte solution were positioned \sim 2-3 mm from the heated (200 $^{\circ}\text{C}$) 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.³⁴ Data were processed using vendor software and analysed using both the mMass software package (<http://www.mmass.org/>), and UniDec.³⁵

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 μM of the target protein of interest (50 μL) and 2 mg/mL of the natural extract (50 μL) in 50 mM ammonium acetate (pH 7.0) were combined and incubated at 4 $^{\circ}\text{C}$ for 24 h, then 25 $^{\circ}\text{C}$ 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 (ZebaTM 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 μL) were loaded into the cartridge and centrifuged at 1,500 g at 4 $^{\circ}\text{C}$ for 1.5 min.

For MS¹ 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 2.0×10^5 ions, and a maximum injection time of 500 ms. For MSⁿ 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 (≤ 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.

Native MS Screening Model

Consider a model system comprising of a protein at concentration $[P]_0$ and a natural product extract containing one small molecule binder at concentration $[L]_0$. The small molecule will associate with the protein by the reversible reaction $P+L \rightleftharpoons PL$, where P is the protein, L is the small molecule ligand, and PL is the protein-ligand complex. The degree of association depends on the protein-ligand dissociation constant (K_d) which is defined as the equilibrium constant for the reverse, dissociation reaction.

$$K_d = \frac{[P][L]}{[PL]} \quad (1)$$

It is also known that,

$$[P]_0 = [P] + [PL] \quad (2)$$

$$[L]_0 = [L] + [PL] \quad (3)$$

By rearranging and substituting Equations 2 and 3 into 1, we obtain Equation 4 can be obtained.

$$K_d = \frac{([P]_0 - [PL])([L]_0 - [PL])}{[PL]} \quad (4)$$

Equation 4 can be rearranged into the quadratic Equation 5,

$$[PL]^2 - ([P]_0 + [L]_0 + K_d)[PL] + [P]_0[L]_0 = 0 \quad (5)$$

Equation 5 can be solved for [PL] by application of the quadratic formula, yielding Equation 6.

$$[PL]_1 = \frac{[P]_0 + [L]_0 + K_d - \sqrt{([P]_0 + [L]_0 + K_d)^2 - 4[P]_0[L]_0}}{2} \quad (6)$$

If the system is then perturbed, and all free ligand is removed (as is the case for gel filtration), the total concentration of ligand in the system (i.e. $[L]_0$) is reduced to $[PL]_1$. Thus, if the system is left to re-equilibrate following one round of gel filtration, the concentration of the protein-ligand complex can be calculated using Equation 7, which is obtained by substituting $[PL]_1$ for $[L]_0$ into Equation 6.

$$[PL]_2 = \frac{[P]_0 + [PL]_1 + K_d - \sqrt{([P]_0 + [PL]_1 + K_d)^2 - 4[P]_0[PL]_1}}{2} \quad (7)$$

This equation can be generalised to Equation 8, which describes the concentration of the protein-ligand complex following N-1 rounds of gel filtration and re-equilibration.

$$[PL]_N = \frac{[P]_0 + [PL]_{N-1} + K_d - \sqrt{([P]_0 + [PL]_{N-1} + K_d)^2 - 4[P]_0[PL]_{N-1}}}{2} \quad (8)$$

Thus, following any number of gel filtration cycles, the proportion of protein bound to ligand can be expressed by Equation 9

$$\frac{[PL]_N}{[P]_0} = \frac{[P]_0 + [PL]_{N-1} + K_d - \sqrt{([P]_0 + [PL]_{N-1} + K_d)^2 - 4[P]_0[PL]_{N-1}}}{2[P]_0} \quad (9)$$

The plots in Figure 5 were generated for $N = 1$ and 4 using in-house Python scripts. $N = 1$ corresponds to the case in which the gel filtration occurs sufficiently fast that none of the initially bound ligand dissociates. In contrast, $N = 4$ corresponds to the extreme case in which protein-ligand binding is allowed to fully re-equilibrate between each gel filtration cycle. The limit of detection was defined as $[PL]_N/[P]_0 = 3\%$, as protein-ligand complexes with this abundance should be readily distinguishable above background/chemical noise based on experimental observations.

Quantitative Native MS Measurements

All authentic standards were purchased from Sigma-Aldrich. Stock solutions of the standard compounds were prepared in methanol at ~ 10 mM and stored at -20 °C until analysis. For binding experiments, solutions containing 5 μ M of the target protein and 5-50 μ M of the individual candidate compound of interest in a 99:1 (v/v) 50 mM ammonium acetate:methanol solution were incubated at 4 °C for 24 h. Following incubation, solutions were allowed to re-equilibrate for 1 h at 25 °C before native mass spectra were acquired in triplicate. The dissociation constant (K_d) for each compound was subsequently calculated by examining the relative integrated abundances of the protein (P) and protein-ligand complexes (PL) as measured by native MS (Equation 10).¹⁷

$$K_d = \frac{\sum_n P^{n+}}{\sum_n PL^{n+}} \left([L]_0 - \frac{\sum_n PL^{n+}}{\sum_n P^{n+} + \sum_n PL^{n+}} [P]_0 \right) \quad (10)$$

where $[L]_0$ is the initial concentration of ligand and $[P]_0$ is the initial concentration of protein. K_d values were benchmarked against previously reported solution-phase measurements.³⁶⁻³⁹

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.; Zotchev, S. B.; Dirsch, V. M.; Orhan, I. E.; Banach, M.; Rollinger, J. M.; Barreca, D.; Weckwerth, W.; Bauer, R.; Bayer, E. A.; et al. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discov.* **2021**, *20* (3), 200–216. <https://doi.org/10.1038/s41573-020-00114-z>.
- (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.* **2015**, *14* (2), 111–129. <https://doi.org/10.1038/nrd4510>.
- (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.* **2017**, *16* (8), 531–543. <https://doi.org/10.1038/nrd.2017.111>.
- (4) Bennett, J. L.; Nguyen, G. T. H.; Donald, W. A. Protein–Small Molecule Interactions in Native Mass Spectrometry. *Chem. Rev.* **2021**. <https://doi.org/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.* **2017**, *10* (1), 25–44. <https://doi.org/10.1146/annurev-anchem-061516-045414>.
- (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.* **2020**, *11*, 566. <https://doi.org/10.1038/s41467-019-14179-x>.
- (7) Cong, X.; Liu, Y.; Liu, W.; Liang, X.; Russell, D. H.; Laganowsky, A. Determining Membrane Protein–Lipid Binding Thermodynamics Using Native Mass Spectrometry. *J. Am. Chem. Soc.* **2016**, *138* (13), 4346–4349. <https://doi.org/10.1021/jacs.6b01771>.
- (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.* **2010**, *53* (7), 2719–2740. <https://doi.org/10.1021/jm901137j>.
- (9) Sacco, M. D.; Ma, C.; Lagarias, P.; Gao, A.; Townsend, J. A.; Meng, X.; Dube, P.; Zhang, X.; Hu, Y.; Kitamura, N.; 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.* **2020**, *6* (50), eabe0751. <https://doi.org/10.1126/sciadv.abe0751>.
- (10) Nguyen, G. T. H.; Nocentini, A.; Angeli, A.; Gratteri, P.; Supuran, C. T.; Donald, W. A. Perfluoroalkyl Substances of Significant Environmental Concern Can Strongly Inhibit Human Carbonic Anhydrase Isozymes. *Anal. Chem.* **2020**, *92* (6), 4614–4622. <https://doi.org/10.1021/acs.analchem.0c00163>.
- (11) Tamara, S.; den Boer, M. A.; Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* **2021**. <https://doi.org/10.1021/acs.chemrev.1c00212>.
- (12) Yang, B.; Feng, Y. J.; Vu, H.; McCormick, B.; Rowley, J.; Pedro, L.; Crowther, G. J.; Van Voorhis, W. C.; Forster, P. I.; Quinn, R. J. Bioaffinity Mass Spectrometry Screening. *J. Biomol. Screen.* **2016**, *21* (2), 194–200. <https://doi.org/10.1177/1087057115622605>.
- (13) Gault, J.; Donlan, J. A. C.; Liko, I.; Hopper, J. T. S.; Gupta, K.; Housden, N. G.; Struwe, W. B.; Marty, M. T.; Mize, T.; Bechara, C.; et al. High-Resolution Mass Spectrometry of Small Molecules Bound to Membrane Proteins. *Nat. Methods* **2016**, *13* (4), 333–336. <https://doi.org/10.1038/nmeth.3771>.
- (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.* **2018**, *10* (2), 139–148. <https://doi.org/10.1038/NCHEM.2908>.
- (15) Lössl, P.; Snijder, J.; Heck, A. J. R. Boundaries of Mass Resolution in Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (6), 906–917. <https://doi.org/10.1007/s13361-014-0874-3>.

- (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.* **2017**, *56* (27), 7912–7915. <https://doi.org/10.1002/anie.201702330>.
- (17) Nguyen, G. T. H.; Tran, T. N.; Podgorski, M. N.; Bell, S. G.; Supuran, C. T.; Donald, W. A. Nanoscale Ion Emitters in Native Mass Spectrometry for Measuring Ligand-Protein Binding Affinities. *ACS Cent. Sci.* **2019**, *5* (2), 308–318.
- (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.* **2004**, *74* (4), 553–568.
- (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.* **2012**, *64* (SUPPL.), 4–17. <https://doi.org/10.1016/j.addr.2012.09.019>.
- (21) Park, H.; Jung, J.; Rodrigues, E.; Kitova, E. N.; MacAuley, M. S.; Klassen, J. S. Mass Spectrometry-Based Shotgun Glycomics for Discovery of Natural Ligands of Glycan-Binding Proteins. *Anal. Chem.* **2020**, *92* (20), 14012–14020. <https://doi.org/10.1021/acs.analchem.0c02931>.
- (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.* **2014**, *25* (1), 111–119. <https://doi.org/10.1007/s13361-013-0752-4>.
- (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.* **2012**, *84* (1), 50–58. <https://doi.org/10.1021/ac202760e>.
- (24) Gao, J.; Cheng, X.; Chen, R.; Sigal, G. B.; Bruce, J. E.; Schwartz, B. L.; Hofstadler, S. A.; Anderson, G. A.; Smith, R. D.; Whitesides, G. M. Screening Derivatized Peptide Libraries for Tight Binding Inhibitors to Carbonic Anhydrase II by Electrospray Ionization-Mass Spectrometry. *J. Med. Chem.* **1996**, *39* (10), 1949–1955. <https://doi.org/10.1021/jm960013g>.
- (25) Supuran, C. T. Carbonic Anhydrases: Novel Therapeutic Applications for Inhibitors and Activators. *Nat. Rev. Drug Discov.* **2008**, *7* (2), 168–181. <https://doi.org/10.1038/nrd2467>.
- (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.* **2012**, *112* (8), 4421–4468. <https://doi.org/10.1021/cr200176r>.
- (27) Gault, J.; Liko, I.; Landreh, M.; Shutin, D.; Bolla, J. R.; Jefferies, D.; Agasid, M.; Yen, H. Y.; Ladds, M. J. G. W.; Lane, D. P.; et al. Combining Native and 'Omics' Mass Spectrometry to Identify Endogenous Ligands Bound to Membrane Proteins. *Nat. Methods* **2020**, *17* (5), 505–508. <https://doi.org/10.1038/s41592-020-0821-0>.
- (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.* **2014**, *25* (12), 2060–2068. <https://doi.org/10.1007/s13361-014-0928-6>.
- (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.* **2015**, *137* (28), 9128–9135. <https://doi.org/10.1021/jacs.5b04628>.
- (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.* **2021**, *143*, 4085–4089. <https://doi.org/10.1021/jacs.0c11837>.
- (31) Vidal, M.; Cusick, M. E.; Barabási, A. L. Interactome Networks and Human Disease. *Cell* **2011**, *144* (6), 986–998. <https://doi.org/10.1016/j.cell.2011.02.016>.
- (32) Szklarczyk, D.; Santos, A.; Von Mering, C.; Jensen, L. J.; Bork, P.; Kuhn, M. STITCH 5: Augmenting Protein-Chemical Interaction Networks with Tissue and Affinity Data. *Nucleic Acids Res.* **2016**, *44* (D1), D380–D384. <https://doi.org/10.1093/nar/gkv1277>.

- (33) Sterling, T.; Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>.
- (34) Hernández, H.; Robinson, C. V. Determining the Stoichiometry and Interactions of Macromolecular Assemblies from Mass Spectrometry. *Nat. Protoc.* **2007**, *2* (3), 715–726. <https://doi.org/10.1038/nprot.2007.73>.
- (35) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015**, *87* (8), 4370–4376. <https://doi.org/10.1021/acs.analchem.5b00140>.
- (36) Beyza Öztürk Sarıkaya, S.; Gülçin, I.; Supuran, C. T. Carbonic Anhydrase Inhibitors: Inhibition of Human Erythrocyte Isozymes i and Ii with a Series of Phenolic Acids. *Chem. Biol. Drug Des.* **2010**, *75* (5), 515–520. <https://doi.org/10.1111/j.1747-0285.2010.00965.x>.
- (37) Innocenti, A.; Beyza Öztürk Sarıkaya, S.; Gülçin, I.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Inhibition of Mammalian Isoforms I-XIV with a Series of Natural Product Polyphenols and Phenolic Acids. *Bioorganic Med. Chem.* **2010**, *18* (6), 2159–2164. <https://doi.org/10.1016/j.bmc.2010.01.076>.
- (38) Karioti, A.; Ceruso, M.; Carta, F.; Bilia, A. R.; Supuran, C. T. New Natural Product Carbonic Anhydrase Inhibitors Incorporating Phenol Moieties. *Bioorganic Med. Chem.* **2015**, *23* (22), 7219–7225. <https://doi.org/10.1016/j.bmc.2015.10.018>.
- (39) Ekinci, D.; Karagoz, L.; Ekinci, D.; Senturk, M.; Supuran, C. T. Carbonic Anhydrase Inhibitors: In Vitro Inhibition of α Isoforms (HCA I, HCA II, BCA III, HCA IV) by Flavonoids. *J. Enzyme Inhib. Med. Chem.* **2014**, *28* (2), 283–288. <https://doi.org/10.3109/14756366.2011.643303>.