

UCSF

UC San Francisco Previously Published Works

Title

A Liquid Chromatography/Mass Spectrometry Method for Screening Disulfide Tethering Fragments

Permalink

<https://escholarship.org/uc/item/5qb786xx>

Journal

SLAS DISCOVERY, 23(2)

ISSN

2472-5552

Authors

Hallenbeck, Kenneth K
Davies, Julia L
Merron, Connie
[et al.](#)

Publication Date

2018-02-01

DOI

10.1177/2472555217732072

Peer reviewed



Published in final edited form as:

SLAS Discov. 2018 February ; 23(2): 183–192. doi:10.1177/2472555217732072.

A Liquid Chromatography/Mass Spectrometry Method for Screening Disulfide Tethering Fragments

Kenneth K. Hallenbeck¹, Julia L. Davies¹, Connie Merron¹, Pierce Ogden¹, Eline Sijbesma², Christian Ottmann², Adam R. Renslo¹, Christopher Wilson¹, and Michelle R. Arkin¹

¹Department of Pharmaceutical Chemistry and Small Molecule Discovery Center, University of California, San Francisco, San Francisco, CA, USA ²Department of Biomedical Engineering, Laboratory of Chemical Biology, and Institute for Complex Molecular Systems, Eindhoven University of Technology, Eindhoven, The Netherlands.

Abstract

We report the refinement of a high-throughput, liquid chromatography/mass spectrometry (LC/MS)-based screening method for the identification of covalent small-molecule binders to proteins. Using a custom library of 1600 disulfide-capped fragments targeting surface cysteine residues, we optimize sample preparation, chromatography, and ionization conditions to maximize the reliability and flexibility of the approach. Data collection at a rate of 84 s per sample balances speed with reliability for sustained screening over multiple, diverse projects run over a 24-month period. The method is applicable to protein targets of various classes and a range of molecular masses. Data are processed in a custom pipeline that calculates a percent bound value for each compound and identifies false positives by calculating significance of detected masses (signal significance). An example pipeline is available through Biovia's ScienceCloud Protocol Exchange. Data collection and analysis methods for the screening of covalent adducts of intact proteins are now fast enough to screen the largest covalent compound libraries in 1 to 2 days.

Keywords

mass spectrometry; tethering; disulfide trapping; fragment screening; covalent binding

Introduction

The past decade has seen an increase in the development of covalent inhibitors as potential therapeutic agents. This interest has been driven by an appreciation of the advantages of covalent mechanisms of inhibition,^{1,2} including the ability to overcome resistance, such as in EGFR gatekeeping mutations,³ the opportunity to increase affinity for otherwise “undruggable” targets, and distinct pharmacokinetic properties due to very long target-

Corresponding Author: Michelle R. Arkin, Department of Pharmaceutical Chemistry and Small Molecule Discovery Center, University of California, San Francisco, 1700 4th Street, Box 2552, San Francisco, CA 94143, USA. michelle.arkin@ucsf.edu.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

residency times.^{4,5} A barrier to the pursuit of such compounds has been the perception that electrophilic drugs present greater risk due to nonspecific binding to off-targets, formation of reactive metabolites, or rapid inactivation by reaction with glutathione or other endogenous nucleophiles.⁶⁻⁸ However, the design and synthesis of covalent inhibitors, particularly targeting cysteine residues, have proven an effective discovery approach for select targets and therapeutic areas.⁹ Furthermore, covalent inhibitors have been used as chemical probes of proteins with native or engineered cysteine residues. These success stories have used reversible adduct formation, such as disulfides¹⁰ and cyanoacrylamides,¹¹ or irreversible electrophiles.¹²⁻¹⁴

As interest in covalent drug discovery has grown, so have analytical techniques to screen for adduct formation, as well as chemical methods to prepare disulfide-based and electrophilic compound libraries.^{12,15,16} Despite these improvements, the largest reported screen of an electrophile library involved just 1000 compounds,¹³ similar in size to our 1600-member disulfide-fragment library. Library sizes reflect several challenges inherent to the goal of discovering selective covalent inhibitors. First, adduct-forming libraries are generally custom synthesized^{12,14-16} to normalize chemical reactivity and optimize structural diversity. Ideally, covalent ligand binding involves initial noncovalent recognition of the protein surface, followed by reaction with a proximal nucleophilic residue on the protein. If a compound is too reactive, binding is dominated by the energetics of covalent bond formation and is insensitive to molecular recognition (such chemotypes are unfortunately ubiquitous in many high-throughput screening [HTS] libraries and act as “pan-assay interference compounds,” or PAINS).¹⁷ At the same time, small changes to compound structure can affect chemical reactivity through electronic or steric effects, obscuring underlying structure-activity relationships that derive from molecular recognition of the target. Well-designed libraries therefore seek to normalize reactivity, either by selecting electrophiles with lower functional-group sensitivity¹⁴ or by separating the diverse structure elements from the reactive group using linkers.¹⁰ The design of covalent compound libraries and the development of effective covalent screening conditions must therefore control for the differing reactivity of screening compounds and/or include counterscreens to establish selectivity.²

When identifying covalent ligands is the goal, it is reasonable for the primary screen to detect the formation of a covalent bond, with secondary screens for biochemical and cellular activity. Methods for measuring covalent protein modification are usually based on liquid chromatography/mass spectrometry (LC/MS), analyzing either intact protein or proteolytic peptides (LC/MS/MS). The chromatographic step in tandem MS generally takes >10 min and is therefore incompatible with demands of HTS, where seconds per sample is ideal. Intact protein detection has been reported at ~3 min/sample in LC formats that take advantage of ultra-pressure liquid chromatography (UPLC)¹⁸ and as quickly as 1.5 min/sample at high concentrations (>10 μ M) with flow injection analysis.¹⁹ Solid-phase extraction MS (SPE-MS) has been shown to be a viable alternative to LC/MS with reported speeds of 20 s/sample.¹³ While fast, SPE-MS does not allow fractionation of complex samples through chromatography. Typically, only the expected masses—rather than a full spectrum—are recorded, which can lead to false positives for noisy spectra and loss of information about multiple adduct formation.¹³ Finally, SPE-MS is a relatively insensitive

MS method, using high ng/low μg amounts of protein/injection; screening is therefore done with micromolar concentrations of protein, limiting the ability to distinguish high-affinity binders and measure apparent binding affinities.

Here we report an intact protein LC/MS method for the rapid (84 s/sample) screening of covalent small molecules using a custom 1600-compound library of disulfide-bearing fragments. While 4-fold slower than available SPE-MS methods,¹³ our approach takes advantage of efficient UPLC desalting to inject less sample. Across 31 proteins of various molecular weights (MWs), our method has detection limits of 0.2 to 20 ng, with screening injections of 12 to 120 ng (6 μL of 100–500 nM). This enables screening of our library, including assay development, with as little as 20 μg of purified protein. We have applied the method to a range of protein classes, collecting high-quality spectra at a speed capable of sustainably screening 1000 compounds/day. Custom pipelines facilitate data processing and analysis. Since this method uses commonly available equipment, has low protein consumption, and is analyzed with publicly available computational tools, it can be readily adopted in other laboratories.

Materials and Methods

Protein Expression and Purification

Desired WT sequences of target proteins were cloned from their respective complementary DNA (cDNA) into a pET15b plasmid containing a 6xHis affinity tag followed by a tobacco etch virus (TEV) protease cleavage site at the *N*-terminus. Cysteine mutations were made via the Megawhop PCR²⁰ or QuikChange Site-Directed Mutagenesis Kit (Agilent, Santa Clara, CA). All constructs were verified by DNA sequencing.

Recombinant protein expression protocols for targets in Table 1 varied to obtain optimal yield. For example, Lfa1, Mac1, and 14–3–3 σ were grown in *Escherichia coli* Rosetta 2(DE3) at 37 °C until OD₆₀₀ reached 0.3. The temperature was reduced to 25 °C, and at OD₆₀₀ = 0.6, expression was induced with 0.25 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) followed by overnight culture. Cells were harvested by centrifugation; resuspended in 50 mM HEPES (pH 7.5), 500 mM NaCl, 10 mM MgCl, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 10 mM imidazole, and 5% w/v glycerol; and lysed by microfluidization (Microfluidics, Westwood, MA). The soluble lysate fraction was incubated with HisPur Cobalt resin (Thermo Fisher, Waltham, MA), washed, and eluted by gravity flow in lysis buffer containing 150 mM imidazole. To remove the 6xHis affinity tag, purified protein was incubated overnight at 4 °C with 0.5 mg recombinant TEV protease with its own 6xHis affinity tag and dialyzed with an excess of 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl, 0.25 mM TCEP, and 5% w/v glycerol. TEV protease and uncleaved protein were removed by repress over a HisPur Cobalt resin column equilibrated in lysis buffer. Cleaved and repressed protein was further purified by size exclusion chromatography on a Superdex 75 16/600 column (GE Healthcare, Little Chalfont, UK) in 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl, and 5% w/v glycerol. Protein purity was confirmed via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). WT protein identity and cysteine mutation presence were confirmed by intact protein LC/MS on a Xevo

G2-S (Waters, Milford, MA). Pure protein was concentrated to >5 mg/mL, flash frozen in LN₂, and stored at -80 °C.

Compound Library

A custom library of 1600 disulfide exchangeable compounds available at the UCSF Small Molecule Discovery Center (SMDC) was synthesized using parallel methods as previously described.^{15,16} For screening, the compounds were arrayed in 384-well plates as 50 mM solutions in DMSO.

Disulfide Tethering

Protein constructs containing target cysteines were diluted to screening concentration in 20 mM Tris (pH 8.0) (Table 1). Then, 15 µL of the dilute protein was plated into columns 3 to 22 of a 384-well low-volume V-well Greiner Bio plate, with water in rows 1 to 2 and 23 to 24. Next, 30 nL of disulfide-capped fragments was pinned into the 320 wells containing protein with a Biomek FX (Beckman Coulter, Indianapolis, IN), and the reaction mixture was incubated for 1 to 3 h at room temperature (RT) (depending on experimental determination of time to equilibrium). Two plates of compounds were prepared simultaneously for overnight data collection.

Liquid Chromatography

UPLC used an I-Class Acquity UPLC (Waters) with a BEH C4, 300 Å, 1.7-µm × 2.1-mm × 50-mm column. A flow rate of 0.4 mL/min was used with the gradient scheme outlined in Supplemental Figure S1, operating at pressures of 8000 to 10,000 pounds per square inch (PSI). Mobile phase A was H₂O + 0.5% formic acid (FA), and B was acetonitrile + 0.5% FA. Then, 6 µL of sample was drawn from 384-well low-volume plates and injected, a 12-s process. A postinjection wash of 50:50 MeOH:H₂O added 6 s to yield a total experiment time of 84 s. The UPLC was diverted to waste from time = 0 to 0.60 min and again after 0.90 min; eluent from 0.60 to 0.90 min was routed to the mass spectrometer for detection. UV absorbance at 280 nm was collected for troubleshooting purposes during the experiment time of 0.60 min to 0.90 min.

Mass Spectrometry

Mass spectrometry data were acquired on a Xevo G2-XS Quadrupole Time of Flight mass spectrometer with a ZSpray ion source (Waters). Electrospray ionization (ESI) conditions were optimized for *m/z* signal intensity of a leucine enkephalin dimer (LeuEnk) (Waters) peak at 1111.6 Da by direct infusion of a 200-pg/µL solution of MeOH:H₂O with 0.1% FA. The dimer peak was used because it falls in the typical *m/z* range of analyzed protein charge envelopes (1000–2000 Da). In addition, 2 ng/µL LeuEnk was used as a detector control with the ZSpray LockSpray system. Screening experiments were done at a capillary voltage of 3.20 kV, cone voltage of 40 V, source temperature of 150 °C, desolvation temperature of 650 °C, cone gas at 50 L/h, and desolvation gas at 1200 L/h. Data were collected at 1 spectra/s from 50 to 5000 *m/z*.

Limit of Detection Experiments

Limit of detection (LOD) experiments were run using the LC/MS conditions reported above. Protein samples were twofold serial diluted from 500 to 5 nM in 10 mM Tris (pH 8.0) using Optima LC/MS-grade water (Fisher Scientific, Hampton NH). Injections of increasing concentration were monitored by manual inspection of the chromatogram until a protein peak began to appear (between 0.75 and 0.90 min). The LOD was defined as the first concentration at which the processing parameters below yielded the expected deconvoluted mass.

Data Processing

Raw LC/MS data files were batch processed with Waters OpenLynx within a MassLynx v4.1 environment. A maximum entropy algorithm for mass deconvolution, MaxEnt1, was used on background subtracted m/z spectra from the portion of the LC chromatogram containing protein signal. Peak picking of the chromatogram was performed with parameters noted in Supplemental Figure S2 and always fell between 0.75 and 0.90 min. As noted in previous work,¹³ rare peak-picking errors in noisy data can be manually inspected and combined prior to deconvolution. The OpenLynx processing parameters subtracted m/z background between 750 and 2000 Da, with back-ground defined as 1% maximum intensity. The 750- to 2000-Da range for m/z subtraction and deconvolution was chosen for general application to a range of target MW but can be varied to match a target protein charge envelope. Deconvolution was performed with a range of ± 6000 Da around the target's expected mass, a target resolution of 0.5 Da, with 20 iterations of MaxEnt1 (Suppl. Fig. S3). The 384-well plates were batch-processed into one large .rpt file at an analysis rate of ~ 30 s per sample. The resulting .rpt text file was inspected for data quality within MassLynx.

The expected highest abundance monoisotopic adduct masses were calculated for all compounds using Pipeline Pilot (Biovia) via a systematic transformation using a defined virtual reaction (Suppl. Fig. S4A). Once the expected adduct structure was verified, the highest abundance monoisotopic masses were registered using an adduct mass registration system through the Pipeline Pilot WebPort into a MySQL database. The protocol code for the adduct mass registration system has been uploaded with an example compound set to the publicly accessible ScienceCloud Protocol Exchange (Biovia) as "Adduct Highest Abundance Monoisotopic Mass Registration." The mass of the protein- β -mercaptoethanol (β ME) conjugate (cap) was calculated analogously. Protein and cap masses were registered via HiTS, a custom web application, into a MySQL database. Finally, a separate Pipeline Pilot algorithm used equation (1) (see Results and Discussion) to report adduct formation and equation (2) to provide a measure of data quality; the output was recorded in a MySQL database (Suppl. Fig. S4B).

Results and Discussion

Tethering Screening Technology

Figure 1 describes the tethering screening methodology. Library compounds are built from structurally diverse fragment moieties (commonly < 200 Da), joined via amides, 1,2,3-triazoles, or other more extended linkers to a common aliphatic disulfide terminated with a

basic amine to afford good solubility (Fig. 1A). The common aliphatic disulfide moiety roughly normalizes library members' intrinsic reactivity in disulfide exchange reactions. Fragments are mixed with proteins containing native or engineered disulfides under conditions (pH, reduction potential) that favor thiolate-disulfide exchange. Once equilibrium is reached, the reaction mixture is injected onto a UPLC/MS system; UPLC offers partial purification, and ESI-time-of-flight (TOF) mass spectrometry allows determination of protein and protein + adduct masses. Sample data are provided in Figure 2.

Method Optimization

The UPLC step was optimized for speed, signal/noise, and consistency by varying solvent flow rate (0.2–1.0 mL/min), column chemistry (C4, C8, C18), and elution strategy. A 0.4-mL/min flow over a 50-mm C4 column with a rapid (10-s) gradient provided the fastest desalting, which still afforded separation of proteins from postelution noise (Fig. 2A). A second “wash” elution immediately followed the detected gradient to reduce carryover of compounds and proteins on the C4 column (Suppl. Fig. S1). Flow diversion to waste before 0.6 min and after 0.9 min minimized contamination of the Xevo ion source.

We optimized the Xevo G2 LC/MS ionization conditions for detection of various proteins between 500 and 5000 *m/z*. Varying cone voltage (80–200 V), desolvation temperature (350–650 °C), the source capillary proximity to the cone, and angle toward the cone led us to the settings described in the Materials and Methods. We then performed a LOD test on a series of proteins with varying molecular weight, without modifying the experimental or analysis parameters (Fig. 3A). LOD was defined as the lowest concentration at which a given sample could be successfully processed in the data analysis pipeline; LOD values varied from 5 to 10 nM (ca. 1–5 ng per 6- μ L injection; 12 proteins) to 250 nM (5 proteins). Representative chromatograms, *m/z* spectra, and deconvoluted masses from a range of protein classes and MWs are shown in Figure 3B–E.

Assay Development

Assay development for screens followed a three-step process. First, protein concentration was selected to be 2- to 10-fold LOD. For example, various cysteine mutants of adapter protein 14–3–3 σ had detection limits of 10 to 50 nM (0.2–2.5 ng; Fig. 3), and we selected a screening concentration of 100 nM. Second, tethering constructs were probed for reactivity with a titration of β ME, a thiol capable of forming a disulfide with an available cysteine thiolate, to confirm solvent accessibility and chemical reactivity of the target cysteine.¹⁰ Screens were run from 100 to 1000 μ M β ME, and screening conditions were selected where a minor β ME peak (ca. 20%) was present. Higher β ME concentration resulted in a more stringent screen by providing competitor and increasing reduction potential of the mixture; selecting an appropriate screening concentration allowed tuning of the signal/noise and hit rate. Notably, some cysteines showed no β ME labeling during assay development but resulted in normal screening data sets.

The time to equilibrium for the selected protein and β ME concentration was tested by incubation at RT for 1 to 3 h before analysis. As previously shown,¹⁰ disulfide labeling assays are thermodynamically (vs. kinetically) controlled, balancing chemical reactivity with

specific small-molecule/protein interactions. One benefit of directly injecting the biochemical reaction (vs. methods requiring sample preprocessing) is access to time course and washout experiments to test labeling equilibrium and reversibility. For example, by repeated injection of 2 μL from the same well containing a 100- μL reaction of 100 μM compound, 100 nM 14–3–3 σ , and buffer, the reaction reached equilibrium in 5 min and remained stable for 45 min (Suppl. Fig. S5). An aliquot was then diluted 1000-fold with reaction buffer and assayed to confirm reversibility. For screening, a time was selected where the signal intensity was stable and no change in signal or percent βME labeling was observed, indicating thermodynamic equilibrium.

Primary Screening

The library of 1600 disulfide fragments was stored in a 384-well format in DMSO at 50 mM. Then, 30 nL of the compound library was pinned into a reaction mixture of protein diluted into 20 mM Tris or ammonium acetate (pH 8.0), with the high pH chosen to increase the concentration of thiolate and therefore facilitate thiolate/disulfide exchange. The exchange reaction was incubated until reaching equilibrium before beginning analysis.

The Acquity UPLC was equilibrated at initial conditions until PSI over 1 min was 1% of total system PSI (2–3 min) before beginning injections. Two plates of 320 compounds were queued simultaneously, with water in the first two and last two columns. Four dummy injections of high-performance liquid chromatography–grade H_2O were included to remove impurities in the UPLC before injections. The experiment cycle time was 84 s, a rate that allowed us to complete two 384-well plates overnight (18 h) and was sustainable over long periods of use. In 24 months of operation at 84 s/sample, we performed 184,301 injections over 6317 h of experimental time, consuming 134 L of mobile phase. Including idle time, regular maintenance, and intermittent instrument repair, these values translated to 8.75 h, 251 experiments, and 0.18 L of solvent per day for 2 years. During this time, we performed screens of several target proteins; representative screens are shown in Table 1. The method was broadly applicable and agnostic to target class or construct size. While we have not attempted to screen a protein >50 kDa, the method detected proteins ranging from 8 to 90 kDa (Fig. 3).

The speed of the LC step relies on minimizing the amount of labeling reaction injected per sample. Injecting more than twofold the LOD of a target protein can lead to detectable carryover between samples. Conversely, screening too close to the LOD results in low signal/noise and increases the false-positive rate from compound ion suppression. We find that twofold LOD allows for rapid, sustainable LC desalting. The sustainability of the method for screening also relies on the LC step being long enough to remove buffer salts. We have obtained tractable MS data sets with cycle times as fast as 50 s/sample but found that this method suffered from column pressure buildup and salt residue deposits on the ion source, leading us to extend cycle time to decrease maintenance requirements.

Data Processing

Raw screening data were processed with the Waters OpenLynx program, software designed to apply a single Waters algorithm across large data sets. The m/z data were combined

across the total ion count (TIC) peak, subtracted, and analyzed with MaxEnt1, a maximum entropy algorithm for deconvoluting intact protein mass (Figs. 2, 3). These data were reported as mass versus percent, in .rpt format.

Due to the volume of data and the varying quality of individual spectra, we developed a high-throughput analysis algorithm to quantify adduct formation. OpenLynx output files were read and processed using a custom Pipeline Pilot (Biovia, San Diego, CA) protocol to quantify binding and indicate the quality of each experiment (Suppl. Fig. S4; supplemental materials). Spectra were divided into small mass bins surrounding the expected masses for free protein, β ME-capped protein, and protein bound to adduct, as well as one large bin for unexpected masses (Suppl. Table S1). “Expected mass” bins included ± 5 amu from the expected mass to accommodate resolution fluctuations due to signal/noise or drift of mass lock. The bin width could be varied from screen to screen to match sample quality, from ± 2 to ± 5 amu from target peaks. If bin overlap occurred due to a larger bin size (possible in lower quality data) or a similarity in mass between the adduct and the reductant (possible for small fragments), bins were adjusted by dividing the difference between the cap and adduct mass by 2, rounding down to the nearest integer. Within each bin, the intensities were summed and used to calculate the percent bound as in equation (1):

$$\% \text{ bound} = \frac{\sum i_{\text{adduct}} + \sum i_{\text{double adduct}}}{\sum i_{\text{protein}} + \sum i_{\text{adduct}} + \sum i_{\text{double adduct}}} \quad (1)$$

where the % of β ME-protein adduct is included with “protein.” The protocol also checked for double-adduct formation in constructs that had alternative nucleophilic residues (e.g., two exposed cysteine residues near compound-binding sites). The algorithm also identified unanticipated species and adducts by reporting a maximum intensity found outside of the expected mass ranges as a secondary peak. In fact, these data were used in one study to identify and correct incorrectly drawn structures in the database.

Screening hits could be identified by plotting percent bound versus compound number (e.g., as shown in Fig. 1D). However, this strategy was sensitive to false positives; during the ± 5 amu binning step, experiments with low signal/noise could report high percent labeling. To provide indicators of data quality, a “signal significance number,” analogous to a signal-to-noise ratio, was generated by calculating the percentage of the sum of intensities in meaningful bins versus the sum of all intensities:

$$\text{signal significance} = 100 \times \left(\frac{\sum i_{\text{protein}} + \sum i_{\text{adduct}} + \sum i_{\text{double adduct}} + \sum i_{\text{secondary}}}{\sum i_{\text{protein}} + \sum i_{\text{adduct}} + \sum i_{\text{double adduct}} + \sum i_{\text{secondary}} + \sum i_{\text{noise}}} \right) \quad (2)$$

False positives with high percent bound but low signal significance were readily identified by plotting the results of equation (1) versus the results of equation (2) (Fig. 4). Manual

inspection of hits from the lowest 5% of the signal significance range was found to be necessary (Fig. 4B,C). The protocol code for the analysis step has been uploaded with an example data set to the publicly accessible ScienceCloud Protocol Exchange (Biovia)²¹ as “Read and Analyze HTS LCMS RPT File.” In addition, the module code is included as text in the supplemental information. The outputs from equations (1) and (2) were then loaded into the SMDC’s MySQL database for further analysis in a custom web application, HiTS.²²

In conclusion, we report an optimized LC/MS method for screening intact protein for covalent adduct formation, using a library of disulfide-capped fragments. By taking advantage of advances in UPLC and ESI-TOF technology, we have developed an LC method capable of more rapid (<90 s) and sustainable injections than previously reported.¹⁸ The method is capable of detecting proteins across a range of molecular weights and with varying amenability to electrospray ionization (Figs. 3, 4). In addition, the labeling reaction is directly injected, facilitating kinetic studies (Suppl. Fig. S5). While our approach remains slower than extraction-based methods, it benefits from a LC desalting step to increase MS data quality, requiring less than 10 ng of material per injection and 20 to 200 µg of protein for a full screen. This sample usage compares favorably to SPE-MS; for a 19.5-kDa protein, SPE-MS had a detection limit of 40 ng, and screening used 400 ng per injection (10 µL of 2 µM).¹³ The sensitivity allows the screening of low-expressing and/or poorly ionizing proteins, as well as the ability to characterize binding events over a wide affinity range. Finally, full MS spectra are collected and analyzed for unexpected adducts and for acceptable signal/noise (signal significance), allowing post hoc inspection of data quality.

A throughput of 1000 compounds per day represents an advance in LC/MS-based screening, which shifts the limiting factor in screening covalent compounds to the size of available libraries. We routinely screen and analyze our library of 1600 compounds in 3 days. Further increasing the throughput of LC/MS methods or screening compounds in mixtures will become attractive as larger libraries of electrophilic compounds become available.

Although our method is applicable to multiple target classes (Fig. 3, Table 1), some targets are intractable due to protein instability at 10 °C or in low salt, highly reducing conditions. These limitations represent inherent facets of this approach, and targets not amenable to UPLC desalting would require a reimagining of our screening conditions. In rare cases where the target protein is excessively hydrophobic and requires more robust chromatography, we have extended the elution gradient step from 15 s to 120 to 180 s, keeping all other parameters identical. While successful, the resulting screening time of 5 days could motivate the use of higher-throughput and higher-consumption methods such as SPE¹³ or matrix-assisted laser desorption/ionization.

Applications of this and other LC/MS screening of covalent molecules extend beyond drug discovery. Adduct formation is a complex reaction, where reaction rate and equilibrium report on the availability and reactivity of the nucleophile and the affinity of the probe molecule for the local environment.² Experiments that control for compound reactivity and affinity can probe surface ligandability.²³ Screens can be run in the presence and absence of a protein-protein interaction partner or an active-site ligand to identify or confirm active-site binding or allosteric regulation.²⁴ Combining the control of site-directed technologies with

the sampling size of high-through-put experiments generates compelling data about a target protein and the molecules that bind to it.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Pam England, Chimno Nnadi, John Chorba, Kevan Shokat, Ambika Bhagi-Damodaran, and Yinyan Tang, who provided proteins to be used in this manuscript.

Funding

The authors thank the National Science Foundation for a predoctoral fellowship to KKH, and the UCSF Research Resource Program for an instrumentation grant.

References

1. Bauer RA Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discov. Today* 2015, 20, 1061–1073. [PubMed: 26002380]
2. Hallenbeck K; Turner D; Renslo A; et al. Targeting Non-Catalytic Cysteine Residues through Structure-Guided Drug Discovery. *Curr. Topics Med. Chem* 2017, 17, 4–15.
3. Engel J; Richters A; Getlik M; et al. Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach. *J. Med. Chem* 2015, 58, 6844–6863. [PubMed: 26275028]
4. Lu H; Tonge PJ Drug-Target Residence Time: Critical Information for Lead Optimization. *Curr. Opin. Chem. Biol* 2010, 14, 467–474. [PubMed: 20663707]
5. Copeland RA The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat. Rev. Drug Discov* 2016, 15, 87–95. [PubMed: 26678621]
6. David-Cordonnier M-H; Laine W; Joubert A; et al. Covalent Binding to Glutathione of the DNA-Alkylating Antitumor Agent, S23906–1. *Eur. J. Biochem* 2003, 270, 2848–2859. [PubMed: 12823555]
7. Johnson DS; Weerapana E; Cravatt BF Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Med. Chem* 2010, 2, 949–964. [PubMed: 20640225]
8. Mah R; Thomas JR; Shafer CM Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorg. Med. Chem. Lett* 2014, 24, 33–39. [PubMed: 24314671]
9. Singh J; Petter RC; Baillie TA; et al. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov* 2011, 10, 307–317. [PubMed: 21455239]
10. Erlanson DA; Braisted AC; Raphael DR; et al. Site-Directed Ligand Discovery. *Proc. Natl. Acad. Sci. U. S. A* 2000, 97, 9367–9372. [PubMed: 10944209]
11. Serafimova IM; Pufall MA; Krishnan S; et al. Reversible Targeting of Noncatalytic Cysteines with Chemically Tuned Electrophiles. *Nat. Chem. Biol* 2012, 8, 471–476. [PubMed: 22466421]
12. Allen CE; Curran PR; Brearley AS; et al. Efficient and Facile Synthesis of Acrylamide Libraries for Protein-Guided Tethering. *Org. Lett* 2015, 17, 458–460. [PubMed: 25582165]
13. Campuzano ID; San Miguel T; Rowe T; et al. High-Throughput Mass Spectrometric Analysis of Covalent Protein-Inhibitor Adducts for the Discovery of Irreversible Inhibitors: A Complete Workflow. *J. Biomol. Screen* 2016, 21, 136–144. [PubMed: 26676098]
14. Kathman SG; Xu Z; Statsyuk AV A Fragment-Based Method to Discover Irreversible Covalent Inhibitors of Cysteine Proteases. *J. Med. Chem* 2014, 57, 4969–4974. [PubMed: 24870364]
15. Burlingame MA; Tom CTMB; Renslo AR Simple One-Pot Synthesis of Disulfide Fragments for Use in Disulfide-Exchange Screening. *ACS Comb. Sci* 2011, 13, 205–208. [PubMed: 21500860]

16. Turner DM; Tom CTMB; Renslo AR Simple Plate-Based, Parallel Synthesis of Disulfide Fragments Using the CuAAC Click Reaction. *ACS Comb. Sci* 2014, 16, 661–664. [PubMed: 25353066]
17. Dahlin JL; Nissink JWM; Strasser JM; et al. PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *J. Med. Chem* 2015, 58, 2091–2113. [PubMed: 25634295]
18. Everley RA; Croley TR Ultra-Performance Liquid Chromatography/Mass Spectrometry of Intact Proteins. *J Chromatogr. A* 2008, 1192, 239–247. [PubMed: 18417140]
19. Helmich F; van Dongen LJ; Kuijper PHM; et al. Rapid Phenotype Hemoglobin Screening by High-Resolution Mass Spectrometry on Intact Proteins. *Clin. Chim. Acta* 2016, 460, 220–226. [PubMed: 27417394]
20. Miyazaki K MEGAWHOP Cloning: A Method of Creating Random Mutagenesis Libraries via Megaprimer PCR of Whole Plasmids. *Methods Enzymol* 2011, 498, 399–406. [PubMed: 21601687]
21. ScienceCloud Exchange. <https://exchange.sciencecloud.com/exchange/> (accessed July 27, 2017).
22. Arkin MR; Ang KKH; Chen S; et al. UCSF Small Molecule Discovery Center: Innovation, Collaboration and Chemical Biology in the Bay Area. *Comb. Chem. High Throughput Screen* 2014, 17, 333–342. [PubMed: 24661212]
23. Hajduk PJ; Huth JR; Fesik SW Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. *J. Med. Chem* 2005, 48, 2518–2525. [PubMed: 15801841]
24. Bowman GR; Bolin ER; Hart KM; et al. Discovery of Multiple Hidden Allosteric Sites by Combining Markov State Models and Experiments. *Proc. Natl. Acad. Sci. U. S. A* 2015, 112, 2734–2739. [PubMed: 25730859]
25. de Jesus Cortez F; Suzawa M; Irvy S; et al. Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). *PLoS ONE* 2016, 11, e0159316. [PubMed: 27467220]

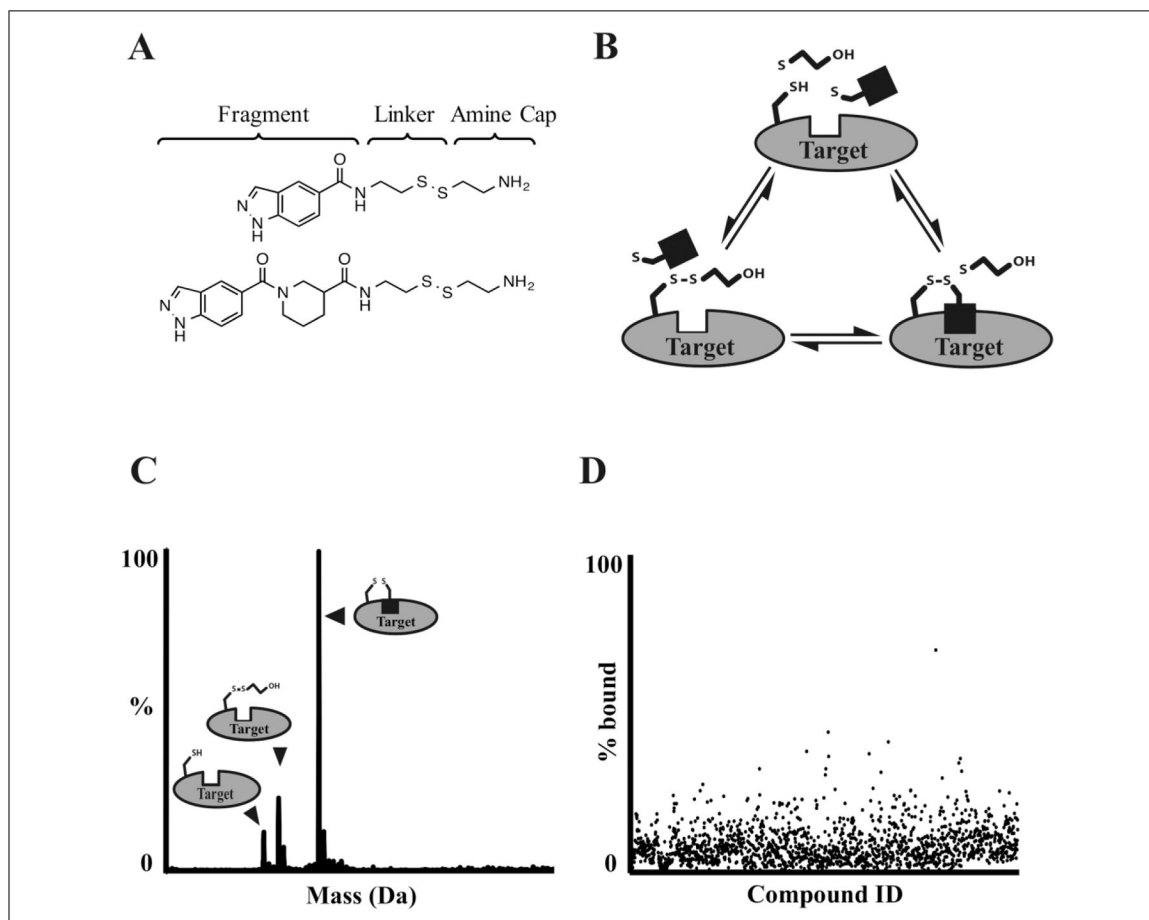


Figure 1. Liquid chromatography/mass spectrometry (LC/MS) screening workflow. **(A)** Examples of structures from the tethering library.^{15,16} **(B)** Labeling reaction scheme. Target protein, β -mercaptoethanol (β ME), and various fragments (black square) are mixed in individual wells of a 384-well plate and incubated until equilibrium is reached. **(C)** Rapid ultra-pressure liquid chromatography (UPLC) desalting, time-of-flight (TOF) detection, and m/z deconvolution identify unlabeled, β ME-capped, and fragment-bound protein species. **(D)** Detected species are checked for expected fragment adduct formation and plotted as a percentage of protein that is fragment bound. Results are checked for data quality and uploaded to an internal database where hits are selected for follow-up.

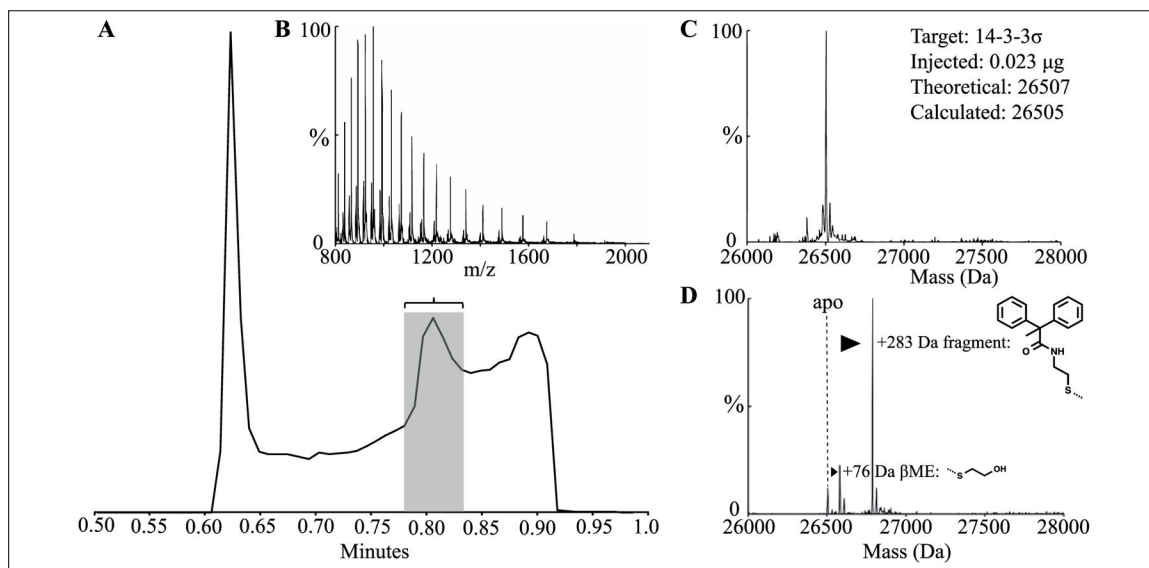


Figure 2.

Liquid chromatography/mass spectrometry (LC/MS) data and processing. **(A)** Total ion count trace of liquid chromatography step. Flow before 0.6 min and after 0.9 min is diverted to waste with Xevo G2S fluidics. **(B)** The peak corresponding to protein ions (0.78–0.84 min) is combined, background subtracted, and reported as m/z . **(C)** MaxEnt (maximum entropy) deconvolution of the m/z charge spectrum identifies the masses present in a sample containing unlabeled protein. **(D)** MaxEnt spectrum deconvoluted from m/z shown in **(B)** of a reaction containing β -mercaptoethanol (β ME) and screening compound, noting adduct formations.

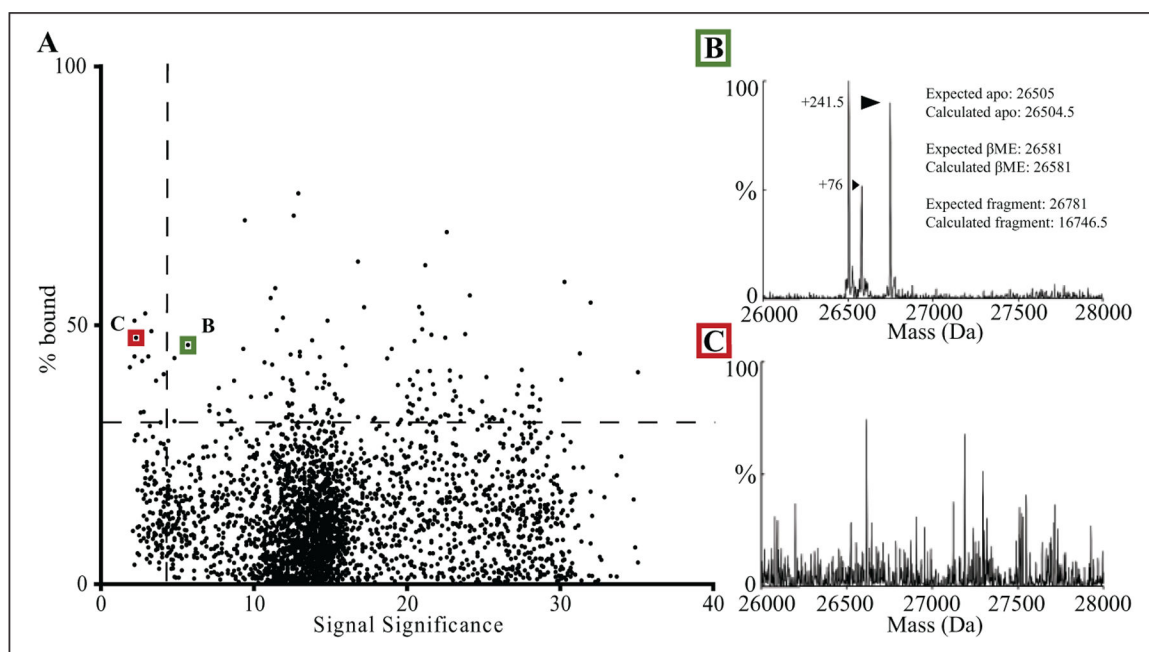


Figure 4. Data set analysis. (A) A typical data set with each of the 1600 screening compounds plotted to compare signal significance of each sample versus its calculated percent bound. The horizontal dotted line is drawn at 3 standard deviations above the mean percent bound. The vertical dotted line is drawn at an empirical cutoff for low-quality samples determined by manually inspecting the data. (B) MaxEnt spectrum for a sample (green box) with medium signal significance (<10), where adduct formation and calculated percent bound are well correlated. (C) MaxEnt spectrum for a sample (red box) with low signal significance (<5) where high noise has artificially inflated the percent bound value.

Table 1.

Screening Outcomes across Representative Targets.

Target Protein	Protein Class	Protein Mass (kDa)	Engineered/Native	Hit Rate (>3 σ), %
ATG4B	Protease	44.5	Native	0.1
			Native ^a	1.4
			Engineered	1.5
Lfa1	Integrin, I-domain	21.0	Engineered ^b	1.8
			Engineered ^b	1.6
Mac1	Integrin, I-domain	22.8	Engineered ^b	2.6
			Engineered ^b	2.3
LRH-1 ^{c,d}	Nuclear receptor	28.3	Native	0.7
Target 1	Ubiquitin ligase	8.78	Native	1.4
Target 2	Kinase	19.3	Native	0.4
Target 3 ^c	Kinase	37.3	Engineered	0.6
14-3-3 σ	Adapter protein	26.5	Native	1.8
			Engineered ^b	2.8
			Engineered ^b	2.7

^aScreened in the presence of a protein partner.^bCys-mutants targeting the same pocket on a respective target.^cScreened 1280 of 1600 compounds.^dde Jesus Cortez et al.²⁵