bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Kenneth K Hallenbeck[a], Julia Davies[a], Connie Merron[a], Pierce Ogden[a], Adam R. Renslo[a], Christopher Wilson[a], Michelle R. Arkin[a,*]

[a]Small Molecule Discovery Center and Department of Pharmaceutical Chemistry, University of California, San Francisco, California, USA

[*]Address correspondence to this author at the Department of Pharmaceutical Chemistry, University of California, San Francisco, Box 2552, San Francisco, CA 94158, USA; Tel/Fax: ++1-415-514-4313, +1-415-514-4504; michelle.arkin@ucsf.edu

Keywords: Liquid Chromatography, Time-of-flight Mass Spectrometry, Disulfide Tethering, High-throughput, Fragment Screening

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

We report the refinement of a high-throughput, liquid-chromatography mass spectrometry-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 90 seconds per sample balances speed and reliability for sustained screening over multiple, diverse projects run in 24 months. The method is applicable to protein targets of various class and molecular mass. Data are processed in a custom pipeline that calculates a % bound value for each compound and detects false-positives by calculating significance of detected masses (‘signal significance’). 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 1-2 days.

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The last decade has seen an increase in the development of covalent inhibitors as potential therapeutic agents. An initial barrier in the pursuit of such compounds was the perception that electrophilic drugs lacked selectivity and presented greater risk from a development perspective[1]. This was in-part due to studies which showed reactive drug metabolites could induce organ damage or evoke an immune response through off-target protein binding[1, 2]. A further concern was that electrophilic compounds could be rapidly inactivated due to adduct formation with native nucleophiles (*e.g.* GSH) resulting in their systemic clearance[3]. However, the design and synthesis of covalent irreversible inhibitors has proven an effective discovery approach for select targets and therapeutic areas[4].

Interest in covalent drugs is driven by an appreciation of the advantages of covalent mechanisms of inhibition. These advantages are well reviewed[5, 6] and include the ability to overcome resistance, such as in EGFR gatekeeping mutations[7], a method to increase affinity for otherwise 'undruggable' targets, and distinct pharmacokinetic requirements owing to very long target residency times of covalent drugs[8, 9].

As interest in covalent drug discovery has grown, so have analytical techniques for the screening of covalent small molecules and the size of electrophilic compound libraries. Despite these improvements, the largest reported screen of a covalent library is of just 1000 compounds[10]. This reflects several limitations inherent to the nature of small molecule covalent modification. First, while large diversity libraries of small molecules are commercially available, they are generally curated to avoid chemical moieties capable of covalent modification, and when such molecules are not actively avoided they remain a small minority of the hundreds of thousands of library members[11]. To obtain a covalently-focused library, groups must synthesize custom libraries[10, 12-

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[14]. Second, binding of covalent screening molecules follows a two-part binding mode, where non-covalent affinity for the protein surface brings the electrophilic compound into proximity of a nucleophilic residue on the protein. However, if a screening compound is too reactive, the initial molecular recognition step becomes a minor component, leading to non-specific binding that is dominated by the energetics of covalent bond formation (some of these chemotypes are ubiquitous in HTS libraries, and are included in the category of 'pan assay interference compounds', or PAINs)[15, 16]. 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 counter-screens to establish selectivity[6]. Finally, when identifying selective, covalent ligands is the goal, the ideal primary screen detects 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 several minutes (>10min) and is therefore incompatible with demands of high-throughput screening (HTS), where seconds per sample is ideal. Intact protein detection has been reported at ~3min/sample in LC formats which take advantage of Ultra-Pressure Liquid Chromatography (UPLC)[17] and as quickly as 1.5 min/sample at high concentrations (>10uM) with flow-injection analysis[18]. Solid-phase extraction MS (SPE-MS) is an excellent alternative to LC-MS with reported sampling speeds of 20s/sample[10]. However, not all target proteins are compatible with SPE.

Here we report an intact protein LC-MS method for the rapid (1.5m/sample) screening of covalent small molecules using a custom 1600 compound library of disulfide–bearing fragments (Fig 1). As previously shown, use of disulfide labeling assays are thermodynamically (vs

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

kinetically) controlled, balancing chemical reactivity with specific small-molecule/protein interactions. Using the Waters Acquity UPLC system in 384-well low-volume format, we inject ≤1 picomole of protein (2uL of ≤500nM) pre-incubated with a disulfide-capped fragment and reductant. An ultra-pressure desalting over a BEH-C4 column before injection provides the high-quality spectra which are a hallmark of LC-MS at a speed capable of screening 1000 compounds/day.

Materials and Methods

Protein Expression and Purification

Desired WT sequences of target proteins were cloned from their respective cDNA into a pET15b plasmid containing a 6xHis affinity tag followed by a TEV protease cleavage site at the N-terminus. Cysteine mutations were made via Megawhop PCR[19] or QuikChange™ Site-Directed Mutagenesis Kit (Agilent). All constructs were verified by DNA sequencing.

Recombinant protein expression protocols for targets in Table 1 varied to obtain optimal yield. For example, Lfa1 (Table1) was grown in *E. coli* Rosetta 2(DE3) at 37 °C until OD_{600} reached 0.3. The temperature was reduced to 25 °C and at OD_{600} = 0.6 expression was induced with 0.25 mM 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 TCEP, 10 mM Imidazole and 5% w/v glycerol, and lysed by microfluidization (Microfluidics). The soluble lysate fraction was incubated with HisPur™ Cobalt resin (Thermo), 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 repass

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

over a HisPur™ Cobalt resin column equilibrated in lysis buffer. Cleaved and repassed protein was further purified by size exclusion chromatography on a Superdex 75 16/600 column (GE Healthcare) in 20 mM HEPES pH 7.5, 250 mM NaCl, 10 mM MgCl, and 5% w/v glycerol. Protein purity was confirmed via SDS-PAGE. WT protein identity and cysteine mutation presence were confirmed by intact protein LC-MS on a Xevo G2-S (Waters). Pure protein was concentrated to >5 mg/mL, flash frozen in LN_2 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[20, 21]. Library compounds were 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 2D). The common aliphatic disulfide moiety shared by all library members serves to roughly normalize their intrinsic reactivity in disulfide exchange reactions. For screening, the compounds were arrayed in 384w plates as 50 mM solutions in DMSO.

Disulfide Tethering

Protein constructs containing target cysteines were diluted to screening concentration (Table 1) in 20 mM Tris pH 8.0. 15 uL of the dilute protein was plated into columns 3-22 of a 384-well Low Volume V-Well Greiner Bio plate, with water in columns 1-2 and 23-24. 30 nL of disulfide-capped fragments were pinned into the 320 wells containing protein with a Biomek FX (Beckman), and the reaction mixture was incubated for 3 hours at RT. Two plates of compounds were prepared simultaneously for overnight data collection.

Liquid Chromatography

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

UPLC was done on an I-Class Acquity UPLC (Waters) using a BEH C4, 300 Å, 1.7 µm x 2.1 mm x 50 mm column. A flow rate of 0.4 mL/min was used with the gradient scheme outline in SI Fig 1. Mobile phase A was H_2O + 0.5% formic acid and B was acetonitrile + 0.5% formic acid. 6 uL of sample was drawn from 384 well, low volume plates and injected, a 12 s process. Post-injection need wash of 50:50 MeOH:H_2O 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, with the interim routed to the Mass Spectrometer for detection. UV absorbance at 280 nM was collected for troubleshooting purposes during the experiment time of 0.30 min to 0.90 min.

Mass Spectrometry

Mass Spectrometry data was acquired on a Xevo G2-XS Quadropole Time of Flight (Waters) with a ZSpray ion source. ESI conditions were optimized for m/z signal intensity of Leucine Enkacephalin (Waters) at 1111.6 da, which was additionally used as a detector control with the ZSpray LockSpray system. Screening experiments were done at a capillary voltage 3.20 kV, cone voltage 40 V, source temperature 150 °C, desolvation temperature 650 °C, cone gas 50 L/hr, desolvation gas 1200 L/hr. Data was collected at 1 spectra/second from 50-5000 m/z.

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 (Fig 2B). The resulting .rpt text file was inspected for data quality within MassLynx. Theoretical adduct masses were calculated for all compounds using Pipeline Pilot (BIOVIA) in a defined virtual reaction using highest abundance isotope mass (SI Fig 3A). The mass of the protein-βME conjugate (cap) was calculated analogously. Protein and cap masses were registered

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

via HiTS, a custom web application. Finally, a separate Pipeline Pilot algorithm used Eq. (1) to report adduct formation and Eq. (2) to provide a measure of data quality, and the output was recorded in an SMDC MySQL database (SI Fig 3B).

Results & Discussion

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 (10s) gradient provided the fastest desalting which still afforded separation of proteins from post-elution noise (Fig 2A). A second 'wash' elution immediately followed the detected gradient to reduce carry-over of compounds and proteins on the C4 column (SI Fig 1). Flow diversion to waste before 0.3 min and after 0.9 min avoided contamination of the Xevo ion source.

We then optimized the Xevo G2 LC/MS ionization conditions for detection of various proteins between 500-5000 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 & Methods. We then performed a limit of detection (LOD) test on series of proteins with varying origin and molecular weight (SI Fig 2).

Assay Development

Assay development for screens followed a 3-step process. First, protein concentration was selected to be 2-fold LOD, where LOD was the lowest concentration at which a 2uL injection of a given sample could be successfully processed in the data analysis pipeline. For example, for Target 4 the detection limit ranged from 10-50 nM for wild type and engineered cys-mutants (SI Fig 2). In this case, we selected a screening concentration of 100 nM. Second, tethering

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

constructs were probed for reactivity with a titration of β-mercaptoethanol (βME), a thiol capable of forming a disulfide with an available cysteine thiolate to confirm solvent-accessibility and chemical reactivity of the target cysteine[22]. Screens were run from 100-1000 μM βME, and conditions where a minor βME peak (%20) was present. Higher βME concentration resulted in a more the 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 show no βME labeling during assay development but result in normal screening datasets. Finally, the stability and of the target at selected protein and βME concentration was tested by incubation at room temperature for ≥3 hours before analysis. A time was selected where the signal intensity was stable, and no change in signal or % βME labeling was observed, indicating thermodynamic equilibrium.

Primary Screen

The library of ~1,600 disulfide fragments was stored in 384-well format at 50 mM. 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, the high pH chosen to favor thiolate/disulfide exchange. The exchange reaction was incubated until reaching equilibrium (1-3 hours) before beginning analysis (Fig 1A).

The Acquity UPLC was equilibrated at initial conditions for 5 minutes 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 HPLC-grade H_2O were included to elute build-up from system equilibration. The experiment cycle time is 84 seconds, a rate which allows us to complete two 384-well plates overnight (15 hours) and was sustainable over long periods of use. In 24 months of operation we performed 184,301 injections over 6317 hours of experimental

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

time, consuming 134 L of mobile phase. Including idle time, regular maintenance, and intermittent instrument repair, these values translated to 8.75 hours, 251 experiments and 0.18 L of solvent per day for two years.

During this time, we performed screens of many target proteins. A representative list is shown in Table 1. The method is broadly applicable and agnostic of target class or construct size. While we have not attempted to screen a protein >50 kDa, the method is expected to be capable of detecting proteins >150 kDa.

Data Processing

Raw screening data were processed with Waters OpenLynx program, software designed to apply a single Waters algorithm across large datasets. 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 (Fig 2B-D). These data were reported as mass vs %, 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) protocol to quantify binding and indicate quality of each experiment (SI Figure 3). Spectra were divided into small mass bins surrounding the expected masses for free protein, βME-capped protein, protein bound to adduct, as well as one large bin for unexpected masses (SI Table 1). Expected mass bins run +/- 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, 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

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

calculate the percent bound as in Eq. (1); the % of βME-protein adduct is included with 'protein'. The protocol also checked for double-adduct formation in constructs that have alternative nucleophilic residues, e.g., two exposed cysteine residues near compound-binding sites.

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

In order to provide indicators of data quality a signal significance number was generated by calculating the percentage of intensity in the expected bins versus the overall range Eq. (2).

Eq. (2) $$signal\ significance = 100\ x\ \left(\frac{\sum i_{protein} + \sum i_{adduct} + \sum i_{double\ adduct} + \sum i_{secondary}}{\sum i_{protein} + \sum i_{adduct} + \sum i_{double\ adduct} + \sum i_{secondary} + \sum i_{other}}\right)$$

The signal significance value is used to identify false-positives during hit selection. During the +/- 5 amu binning step, experiments with low signal/noise could report high % labeling (Fig 3C). Compounds with high % bound but low signal significance were false-positives that could be readily identified by plotting the results of Eq. (1) vs. the results of Eq. (2) (Fig 3). Wells with high labeling sometimes also reported a low signal significance; manual inspection of hits from the lowest 5% of the signal significance range is necessary (Fig 3B-C). The algorithm additionally 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. Results were then loaded into the SMDC's MySQL database for further analysis in a custom web application, HiTS[23].

In conclusion, we report a 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 developed an LC method capable of more rapid (<90s) and sustainable injections than previously reported[17]. While our approach remains slower than extraction-based methods, it benefits from a LC desalting step to increase MS data quality,

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

therefore requiring less protein and compound per sample. A throughput of 1000 compounds per day represents an advance in LC/MS-based screening and 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 2 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.

Though our method is widely applicable to target classes (Table 1), some targets are intractable due to protein stability at 10 °C or in low salt, highly reducing conditions or poor ionization. These limitations represent inherent facets of this approach, and targets not amenable to UPLC desalting would require a re-imagining of our screening conditions or an alternative detection method such as SPE[10] or matrix-assisted laser desorption/ionization.

Applications of this and other LC/MS screening of covalent molecules extends beyond ligand discovery. Adduct formation is a complex reaction, where reaction rate and equilibrium report on availability and reactivity of the nucleophile and the affinity of the probe molecule for the local environment[6]. Experiments that control for compound reactivity and affinity can probe surface ligandability. Screens can be run in the presence and absence of a PPI partner or an active-site ligand to identify or confirm active-site binding or allosteric regulation. Combining the control of site-directed technologies with the sampling size of a high-throughput experiments generates compelling data about a target protein and the molecules which bind to it.

Acknowledgements

The authors thank Jeffrey Satkofsky (Waters Technologies Corporation) for extensive advice regarding the optimization and maintenance of the Xevo G2-S system, as well as the National Science Foundation for a predoctoral fellowship (KKH).

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Conflict of Interest

The authors confirm this article content has no conflicts of interest.

References

1. Johnson, D. S.; Weerapana, E.; Cravatt, B. F., Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* **2010,** *2* (6), 949-64.
2. Mah, R.; Thomas, J. R.; Shafer, C. M., Drug discovery considerations in the development of covalent inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2014,** *24* (1), 33-39.
3. David-Cordonnier, M.-H.; Laine, W.; Joubert, A.; Tardy, C.; Goossens, J.-F.; Kouach, M.; Briand, G.; Thi Mai, H. D.; Michel, S.; Tillequin, F.; Koch, M.; Leonce, S.; Pierre, A.; Bailly, C., Covalent binding to glutathione of the DNA-alkylating antitumor agent, S23906-1. *European Journal of Biochemistry* **2003,** *270* (13), 2848-2859.
4. Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A., The resurgence of covalent drugs. *Nat Rev Drug Discov* **2011,** *10* (4), 307-317.
5. Bauer, R. A., Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discovery Today* **2015,** *20* (9), 1061-1073.
6. Hallenbeck, K.; Turner, D.; Renslo, A.; Arkin, M., Targeting Non-Catalytic Cysteine Residues Through Structure-Guided Drug Discovery. *Current Topics in Medicinal Chemistry* **2017,** *17* (1), 4-15.
7. Engel, J.; Richters, A.; Getlik, M.; Tomassi, S.; Keul, M.; Termathe, M.; Lategahn, J.; Becker, C.; Mayer-Wrangowski, S.; Grütter, C.; Uhlenbrock, N.; Krüll, J.; Schaumann, N.; Eppmann, S.; Kibies, P.; Hoffgaard, F.; Heil, J.; Menninger, S.; Ortiz-Cuaran, S.; Heuckmann, J. M.; Tinnefeld, V.; Zahedi, R. P.; Sos, M. L.; Schultz-Fademrecht, C.; Thomas, R. K.; Kast, S. M.; Rauh, D., Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach. *Journal of Medicinal Chemistry* **2015,** *58* (17), 6844-6863.
8. Lu, H.; Tonge, P. J., Drug–target residence time: critical information for lead optimization. *Current Opinion in Chemical Biology* **2010,** *14* (4), 467-474.
9. Copeland, R. A., The drug-target residence time model: a 10-year retrospective. *Nat Rev Drug Discov* **2016,** *15* (2), 87-95.
10. Campuzano, I. D.; San Miguel, T.; Rowe, T.; Onea, D.; Cee, V. J.; Arvedson, T.; McCarter, J. D., High-Throughput Mass Spectrometric Analysis of Covalent Protein-Inhibitor Adducts for the Discovery of Irreversible Inhibitors: A Complete Workflow. *J Biomol Screen* **2016,** *21* (2), 136-44.
11. Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S., Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **2011,** *10* (3), 188-195.
12. Kathman, S. G.; Xu, Z.; Statsyuk, A. V., A Fragment-Based Method to Discover Irreversible Covalent Inhibitors of Cysteine Proteases. *Journal of Medicinal Chemistry* **2014,** *57* (11), 4969-4974.
13. Allen, C. E.; Curran, P. R.; Brearley, A. S.; Boissel, V.; Sviridenko, L.; Press, N. J.; Stonehouse, J. P.; Armstrong, A., Efficient and Facile Synthesis of Acrylamide Libraries for Protein-Guided Tethering. *Organic Letters* **2015,** *17* (3), 458-460.

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

14. Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frodin, M.; Taunton, J., Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat Chem Biol* **2012,** *8* (5), 471-6.
15. 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. *Journal of Medicinal Chemistry* **2010,** *53* (7), 2719-2740.
16. Dahlin, J. L.; Nissink, J. W. M.; Strasser, J. M.; Francis, S.; Higgins, L.; Zhou, H.; Zhang, Z.; Walters, M. A., PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *Journal of Medicinal Chemistry* **2015,** *58* (5), 2091-2113.
17. Everley, R. A.; Croley, T. R., Ultra-performance liquid chromatography/mass spectrometry of intact proteins. *J Chromatogr A* **2008,** *1192* (2), 239-47.
18. Helmich, F.; van Dongen, J. L. J.; Kuijper, P. H. M.; Scharnhorst, V.; Brunsveld, L.; Broeren, M. A. C., Rapid phenotype hemoglobin screening by high-resolution mass spectrometry on intact proteins. *Clinica Chimica Acta* **2016,** *460*, 220-226.
19. Miyazaki, K., MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods Enzymol* **2011,** *498*, 399-406.
20. Turner, D. M.; Tom, C. T. M. B.; Renslo, A. R., Simple Plate-Based, Parallel Synthesis of Disulfide Fragments using the CuAAC Click Reaction. *ACS Combinatorial Science* **2014,** *16* (12), 661-664.
21. Burlingame, M. A.; Tom, C. T. M. B.; Renslo, A. R., Simple One-Pot Synthesis of Disulfide Fragments for Use in Disulfide-Exchange Screening. *ACS Combinatorial Science* **2011,** *13* (3), 205-208.
22. Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A., Site-directed ligand discovery. *Proceedings of the National Academy of Sciences* **2000,** *97* (17), 9367-9372.
23. Michelle, R. A.; Kenny, K. H. A.; Steven, C.; Julia, D.; Connie, M.; Yinyan, T.; Christopher, G. M. W. a. A. R. R., UCSF Small Molecule Discovery Center: Innovation, Collaboration and Chemical Biology in the Bay Area. *Combinatorial Chemistry & High Throughput Screening* **2014,** *17* (4), 333-342.
24. de Jesus Cortez, F.; Suzawa, M.; Irvy, S.; Bruning, J. M.; Sablin, E.; Jacobson, M. P.; Fletterick, R. J.; Ingraham, H. A.; England, P. M., Disulfide-Trapping Identifies a New, Effective Chemical Probe for Activating the Nuclear Receptor Human LRH-1 (NR5A2). *PLOS ONE* **2016,** *11* (7).

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1. Screening Outcomes Across Targets[a]

Target Protein	*Protein Class*	*Protein Mass (kda)*	*Engineered/ Native*	*Screening Concentration (nM)*	*Hit rate (>3sigma)*
ATG4B	Protease	44.5	Native	500	0.1 %
ATG4B			Engineered	500	1.5 %
Lfa1	Integrin	21.0	Engineered	500	1.8 %
Mac1	Integrin	22.8	Engineered	500	2.6 %
[b]LRH-1[24]	Nuclear Receptor	28.3	Native	250	0.7 %
Target 1	Ubiquitin Ligase	8.78	Native	1000	1.4 %
Target 2	Kinase	19.3	Native	500	0.4 %
[c]Target 3	Kinase	37.3	Engineered	500	0.6 %
Target 4	Adapter Protein	26.5	Native	100	1.8 %
Target 4			Engineered	100	2.8 %

[a]These targets vary in class, size, and whether the cysteine was native or engineered. Generally, screens yielded 10-30 fragments (hit rate = 0.1-1.8%) with labeling above 3 standard deviations from the mean.

[b,c]Screened only 1280/1600 compounds

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

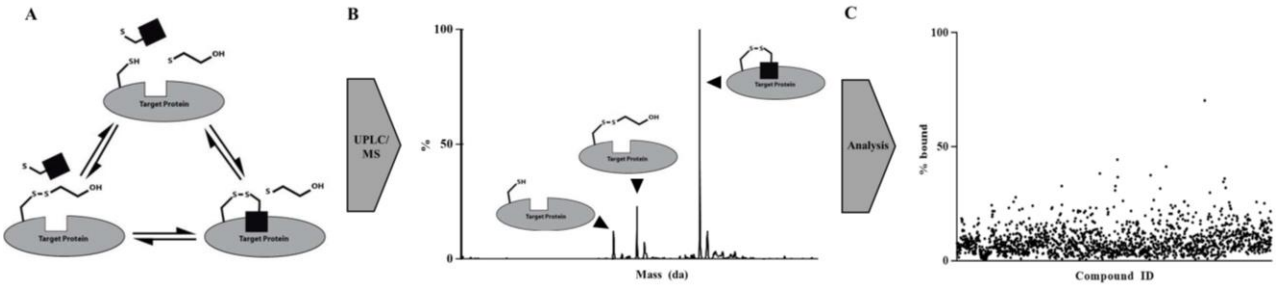

Figure 1. LC/MS Screening Workflow. A) Labeling reaction scheme. Target protein, βME, and various fragments (black square) are mixed in individual well of a 384 well plate and incubated until equilibrium. B) Rapid UPLC desalting, TOF detection and m/z deconvolution identifies unlabeled, βME capped, and fragment-bound protein species. C) Detected species are checked for expected fragment adduct formation and plotted as a % of protein which is fragment-bound. Results are checked for data quality and uploaded to an internal database where selection of hits for follow-up.

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

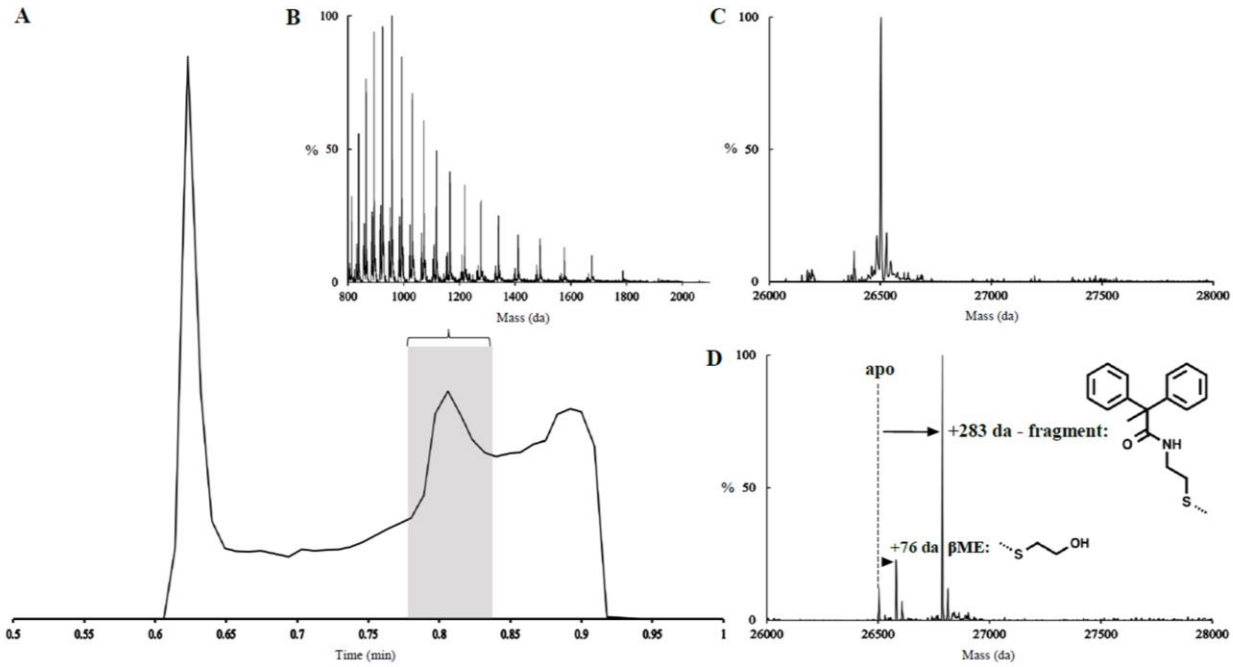

Figure 2. LC/MS Data and Processing. A) Total ion count trace of liquid chromatography step. Flow before 0.6 min and after 0.9 min are diverted to waste with Xevo G2S fluidics. B) The peak corresponding to protein ions (0.78-0.83 min) is combined, background subtracted, and reported as m/z. C) MaxEnt (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 βME and screening compound, noting adduct formations.

bioRxiv preprint doi: https://doi.org/10.1101/138941; this version posted May 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

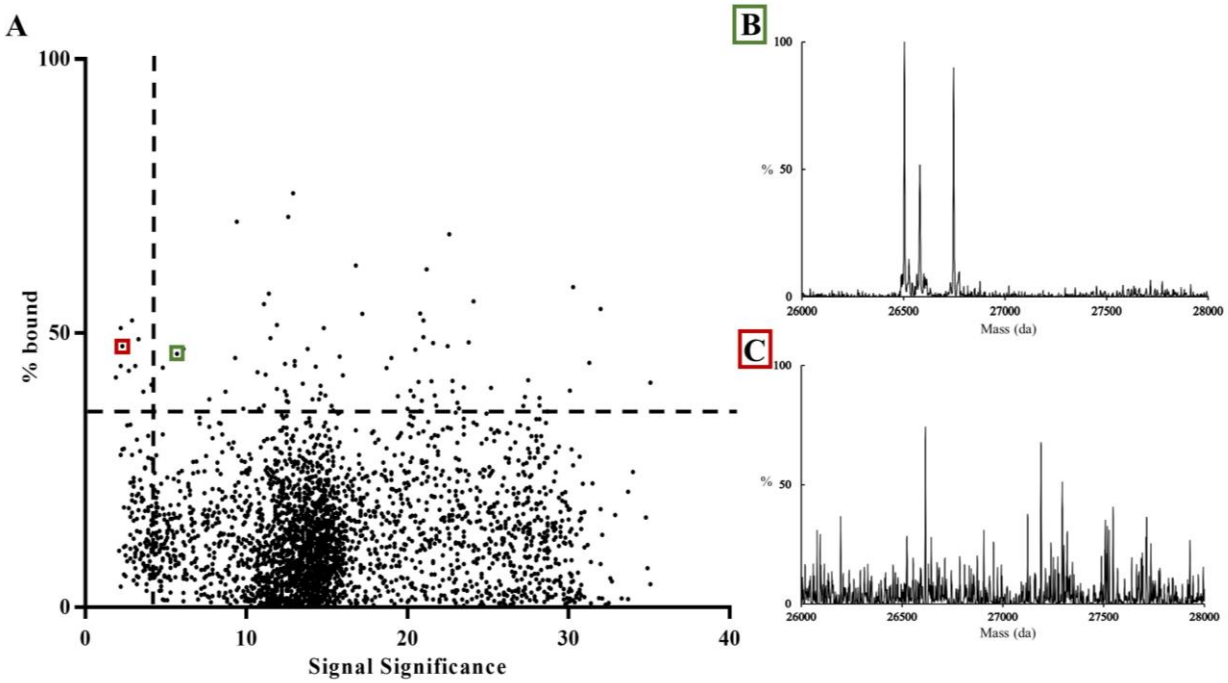

Figure 3. Dataset Analysis. A) A typical dataset with each of the 1600 screening compounds plotted to compare signal significance (Eq. 2) of each sample versus its calculated % bound (Eq. 1). The horizontal dotted line is drawn at 3 standard deviations above the mean % bound. The horizontal line is drawn at an arbitrary cut-off for low quality samples determined by manual inspecting the data. B) MaxEnt spectrum for a sample (green box) with medium signal significance (<10), where adduct formation and calculated % bound are well correlated. C) MaxEnt spectrum for a sample (red box) with low signal significance (<5) where high noise has artificially inflated the % bound value.