MIT Open Access Articles

Identifying the proteins to which small-molecule probes and drugs bind in cells

The MIT Faculty has made this article openly available. *Please share* how this access benefits you. Your story matters.

Citation	Ong, Shao-En et al. "Identifying the proteins to which small-molecule probes and drugs bind in cells." Proceedings of the National Academy of Sciences 106.12 (2009): 4617-4622.
As Published	http://dx.doi.org/10.1073/pnas.0900191106
Publisher	National Academy of Sciences
Version	Final published version
Citable link	http://hdl.handle.net/1721.1/50251
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

DSpace@MIT

Identifying the proteins to which small-molecule probes and drugs bind in cells

Shao-En Ong[a,1], Monica Schenone[a], Adam A. Margolin[b], Xiaoyu Li[c], Kathy Do[a], Mary K. Doud[d], D. R. Mani[a,b], Letian Kuai[e], Xiang Wang[d], John L. Wood[f], Nicola J. Tolliday[c], Angela N. Koehler[d], Lisa A. Marcaurelle[c], Todd R. Golub[b], Robert J. Gould[d], Stuart L. Schreiber[d,1], and Steven A. Carr[a,1]

[a]Proteomics Platform, [b]Cancer Biology Program, [c]Chemical Biology Platform, [d]Chemical Biology Program, and [e]Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142; and [f]Chemistry Department, Colorado State University, Fort Collins, CO 80523

Contributed by Stuart L. Schreiber, January 15, 2009 (sent for review December 21, 2008)

Most small-molecule probes and drugs alter cell circuitry by interacting with 1 or more proteins. A complete understanding of the interacting proteins and their associated protein complexes, whether the compounds are discovered by cell-based phenotypic or target-based screens, is extremely rare. Such a capability is expected to be highly illuminating—providing strong clues to the mechanisms used by small-molecules to achieve their recognized actions *and* suggesting potential unrecognized actions. We describe a powerful method combining quantitative proteomics (SILAC) with affinity enrichment to provide unbiased, robust and comprehensive identification of the proteins that bind to small-molecule probes and drugs. The method is scalable and general, requiring little optimization across different compound classes, and has already had a transformative effect on our studies of small-molecule probes. Here, we describe in full detail the application of the method to identify targets of kinase inhibitors and immunophilin binders.

SILAC | small molecules | target identification

Many small-molecule (SM) probe or drug discovery efforts start by selecting a target that is expected to modulate a pathway or disease of interest. Some drug-discovery efforts optimize existing compounds so that they bind their intended targets with higher specificity and affinity. By focusing on specific protein classes (e.g., kinases), this paradigm of drug discovery routinely uses in vitro assays with recombinant proteins in binding or biochemical assays (refs. 1 and 2; also recently reviewed in ref. 3). Although such large-scale screens can provide early leads that perform well against a specific target, the absence of a biological context results in higher attrition rates in later stages of drug development arising from unanticipated or undetected off-target effects, or lack of relevance of the target protein to the underlying disease process. Furthermore, screens using purified protein substrates do not accurately represent biological levels of target proteins, potentially leading to generation of incorrect hypotheses for on- or off-target drug effects. From the standpoint of drug safety and efficacy, unbiased identification of proteins and associated molecular complexes that bind to a drug allows direct evaluation of its polypharmacology (4) and provides valuable insight into its mode of action and avenues for compound optimization.

Cell-based phenotypic screens allow the discovery of compounds that induce state transitions in cells or organisms without bias regarding specific targets, pathways or even processes. This discovery-based approach has been used with increasing frequency and success in recent years (5, 6). The ability to define cell states globally and molecularly in the context of high-throughput screens, for example, using imaged cellular features (7) and mRNA expression (8), suggests that its impact will continue to grow in the future. However, as with SMs emerging from target-based screens, there exists currently no reliable way to assess the complete set of proteins that interact with SMs discovered in phenotype-based screens. Such a capability is expected to be highly illuminating. It is especially critical with SMs identified in phenotype-based screens because even the target relevant to the induced phenotype is usually not known (the "target I.D. problem"). It could provide strong clues to the mechanisms used by SMs to achieve their recognized actions and it could suggest potential unrecognized actions.

Strategies for "target identification" have been developed that rely on genetic (9), computational (10, 11) and biochemical (12) principles. Although several key molecular targets have been identified through affinity chromatography (13–15), it has not been widely applied as a general solution to target identification for a number of reasons. It is often challenging to prepare SM affinity reagents that retain the desired cellular activity. Experiments with SM baits, even more so than antibody-based immunoaffinity reagents, require carefully chosen and effective controls as baits may vary considerably in their chemical structures and binding properties. Moreover, high stringency washes are required to minimize contamination associated with nonspecific, and bait-independent, interactions of cellular proteins with the reagents. The latter shortcoming is especially significant as it biases toward high-affinity interactions, decreasing the likelihood of identifying more weakly bound proteins or protein complexes that may play significant roles in the polypharmacology of a SM.

Classically, identifying targets of SMs through biochemical purification relied on large amounts of starting protein, extensive protein fractionation, stringent wash conditions, gel visualization and excision of specific bands to yield only the most directly and tightly bound proteins (13, 16, 17). With proteomic MS approaches (18), even affinity pull-down experiments generate large protein catalogs, inflating the list of candidate "hits" and requiring, sometimes arbitrary, prioritization of these proteins for validation. Quantitative proteomics has proven to be a powerful tool for discriminating specific protein–protein interactions from background interactions in affinity pull-downs (19, 20). Although this was recently applied to profile kinases enriched in kinase inhibitor pull-downs (21, 22), these experiments still assumed kinases as targets a priori and did not use quantitative data to define SM specific targets.

Here, we describe an analytical framework combining quantitative mass spectrometry (MS)-based proteomics (23) with affinity chromatography for unbiased, sensitive, specific and comprehensive determination of SM-protein interactions within cellular proteomes. We use SILAC to distinguish cell populations for our SM affinity enrichments (23). Cells are cultured in growth medium containing either "light," natural isotope abundance forms, or the "heavy," ^{13}C, ^{15}N-bearing versions of arginine and lysine. Growing

---

Author contributions: S.-E.O., X.L., N.J.T., A.N.K., and L.A.M. designed research; S.-E.O., M.S., X.L., K.D., M.K.D., and L.A.M. performed research; S.-E.O., A.A.M., D.R.M., L.K., X.W., and J.L.W. contributed new reagents/analytic tools; S.-E.O., M.S., A.A.M., K.D., M.K.D., T.R.G., R.J.G., S.L.S., and S.A.C. analyzed data; and S.-E.O., T.R.G., R.J.G., S.L.S., and S.A.C. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: song@broad.mit.edu, stuart_schreiber@harvard.edu, or scarr@broad.mit.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900191106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0900191106

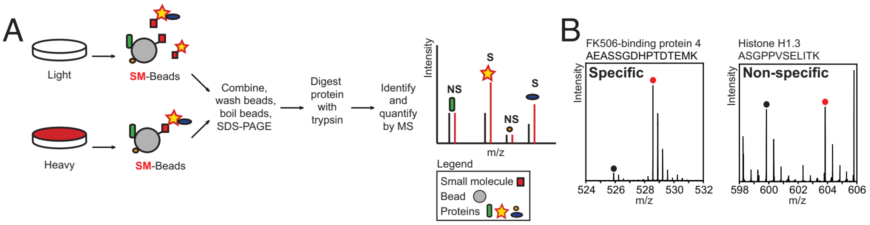

Fig. 1. Identifying specific SM-protein interactions with quantitative proteomics. (*A*) SILAC identifies specific protein interactions with SM baits. Cell populations are fully labeled with light (black) and heavy amino acids (red) and lysates incubated either with SM-loaded beads (SM-Beads) and soluble SM competitor or SM-Beads alone. Proteins interacting directly with the SM or via secondary and/or higher order interactions (marked "S" for specific) will be enriched in the heavy state over the light and will be identified with differential ratios. Nonspecific (NS) interactions of proteins will be enriched equally in both states and have ratios close to 1. (*B*) Experimental mass spectra showing specific protein interactions with the immunophilin ligand, AP1497. (*Left*) A peptide from FKBP4, a known binding partner to FK506, is observed with a highly differential ratio. (*Right*) In contrast, a histone H1.3 peptide is identified with a ratio close to 1, indicating no specific binding to the soluble SM competitor.

and dividing cells incorporate these amino acids in their proteomes, reaching full incorporation after 5 population doublings and producing the characteristic mass shift, 6 Da with $^{13}C_6$-Arg or 8 Da in $^{13}C_6{}^{15}N_2$-Lys containing peptides, observable by MS. Importantly, SILAC peptide pairs have addressable locations in mass and retention-time space. For instance, if we observe a light peptide from a SILAC pair at a given mass and retention time, we should detect its corresponding heavy partner as well, unless its absence is a direct outcome of the experiment. Using SILAC labeled lysates in pull-down experiments with SM-loaded affinity matrices to compare relative enrichment of target proteins (Fig. 1), we use relatively mild washing conditions to preserve the enrichment of weakly bound proteins, increasing sensitivity and yet retaining specificity in identifying bona fide targets. Our quantitative approach permits facile identification of direct interactors to SMs and their associated binding partners; prioritizes target proteins by SILAC ratios; and provides relative measures of binding strengths among structural variants of SMs. We anticipate that our quantitative approach will greatly improve interpretation of phenotypes in SM probe or drug discovery and facilitate downstream optimizations and development.

Results

Recognizing that a general strategy to identify interacting proteins should ideally detect protein targets across a wide range of binding affinities and protein expression in cellular samples, we evaluated our approach using 2 separate sets of well-characterized SM affinity reagents (Fig. S1*A*)—kinase inhibitors and immunophilin binders (the latter using a set of immunophilin ligands (IPL) with a range of known binding affinities). We compared, and discuss in detail below, the performance of 2 experimental designs in our quantitative approach. We also tested a range of conditions likely to affect efficiency and specificity of affinity enrichment such as the amount of SMs on bead matrices, levels of detergent in wash buffers, and amounts of soluble competitor in competition experiments and examined their impact on identification of known targets. For each SM and pull-down condition, we performed at least 2 process replicates, for a total of 98 SILAC experiments. We demonstrate the importance of appropriate controls along with highly accurate quantitative measures and statistical methods that allow the discrimination of real SM-protein interactions from nonspecific binders. We identified previously described targets, some in complexes with known partners (Table 1). Additionally, we found several protein-SM binders, and validated many with surface plasmon resonance (SPR) and Western blot analysis experiments.

Comparing 2 Experimental Designs, Bead Control (BC) and Soluble Competition (SC), for SILAC-Target I.D. We evaluated 2 experimental designs, bead control (BC) and soluble competition (SC) (Fig. 2), in our quantitative target identification approach. The BC experiment compares the relative abundance of proteins from affinity pull-downs with 2 different chemically modified bead matrices—for instance, heavy proteins enriched with the SM affinity matrix versus the population of light proteins captured by ethanol-loaded control bead. In the SC experiment, SM loaded beads are used in both light and heavy pull-downs but excess soluble SM is added to 1 set to competitively bind target proteins. We identify and compare the relative abundance of proteins bound to SM loaded beads by MS. BC experiments identify protein targets through direct enrichment on bead whereas SC infers targets through their depletion from the affinity matrix by the soluble competitor. Both modalities robustly identify known protein targets of SMs but differ significantly in their performance across all experiments. To summarize, although the BC experiment successfully identifies target proteins, it also yields a lengthy list of moderate-to-highly abundant proteins that have weak but real differential binding to the bait molecules or to the chemically modified control bead (Fig. 2*A* and Fig. S2 *A*–*C*). The SC experiment, in sharp contrast, shows much greater specificity, identifying target proteins purely by differential SILAC ratios and is largely independent of protein abundance (Fig. 2*B*). We use several examples from our kinase inhibitor datasets to illustrate the salient features of these 2 experiment designs.

Bead control experiments enrich known targets with large differential SILAC ratios in all our pull-downs. We apply mixture modeling with t-distributions to BC pull-down data, generating modeled probability distributions that describe protein binding

Table 1. Identification of target proteins

	Small Molecule	Identified targets in SILAC experiments
Kinase inhibitors	Ro-31-7549	ADK**, CAMK2D, CAMK2G, CDK2,** CRKRS**, GSK3A, GSK3B,** MYH9**, PRKCA**, **PRKCD**, RPS6KA3, SLK**, NQO2**
	SB202190	**MAPK14**, **MAPK9**, **GSK3B**, **CSNK1A1**, **CSNK1D**, **RIPK2**, TGFBR1, FAM83G, **GAK**
	K252a	AAK1, AURKA, CAMK2B, CAMK2D, CAMK2G, CCNB1, CDC2, CDC42BPB, **CDK2**, **CDK5,** CHEK1, CSNK2A1, CSNK2A2, DNAJC13/RME8, EIF2AK4, FER, GAK, GSK3A, GSK3B, IRAK4, MAPK1, MAPK8, MAPK9, MAPK10, MAP2K1, MAP2K2, MAP2K6, MAP4K4, MARK2, MARK3, MINK1, NQO2, OSGEP, **PDE1A**, PDPK1, PHKA2, PHKB, PHKG2, PKN1, PKN2, PRKAA1, **PRKACA**, **PRKACB**, PRKAG1, PRKAR1A, PRKAR2A, PRKCA, PRKCD, PRKD1, PRKD2, PRKD3, RIPK2, RBM4, RP6–213H19.1, RPS6KA1, RPS6KA3, STK3, STK4, TBK1, TP53RK, TPRKB, ULK3
IPL affinity series	AP1497	**FKBP1A**, FKBP2, FKBP4, FKBP5, FKBP9, FKBP10
	Pro-AP1497	FKBP1A, FKBP2, FKBP4, FKBP5, FKBP9
	AP1780	FKBP1A, FKBP2, FKBP4, FKBP9
	Pro-AP1780	FKBP1A, FKBP2, FKBP9

Proteins in bold type are targets known from literature. Underlined proteins are known interactors to other identified targets within the same experiment.

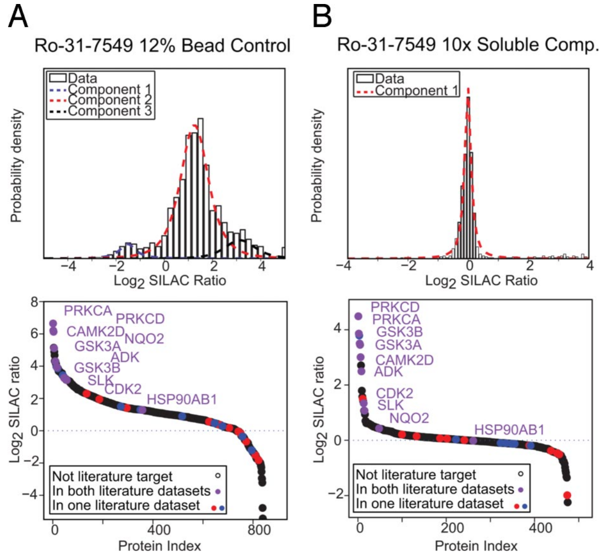

Fig. 2. Classifying SM-protein interactions with quantitative ratios. Using Ro-31-7549 BC (*A*) and SC (*B*) datasets as examples, shown are modeled ratio data distributions with mixture modeling (*Upper*) and plots of $\log_2$ SILAC ratios against proteins sorted by their ratios in descending order (*Lower*). (*Upper*) Histograms of $\log_2$ SILAC ratios for the BC and SC experiments, using mixtures of t-distributions (each modeled component is drawn in dotted lines) to highlight differences between the 2 experimental designs. Although the distribution of $\log_2$ ratios for SC experiments is tightly centered around zero, $\log_2$ ratio distributions for BC experiments vary considerably and are affected by compound loading levels and wash stringency (Figs. S2). (*Lower*) Targets of bisindolylmaleimide-type PKC inhibitors identified in other proteomic datasets are highlighted in red (17) and blue (27), and those common to both datasets are in purple. (*A*) BC experiments compare proteins enriched between SM matrices (heavy) and ethanol loaded beads (light). Ratios help classify proteins into categories of "SM binders," "control bead binders," or "undetermined/nonspecific." (*B*) SC experiments use SM matrices in pull-downs with both light and heavy lysates. Differential ratios arise through reduction of bead-bound protein in one state by competition with the soluble compound.

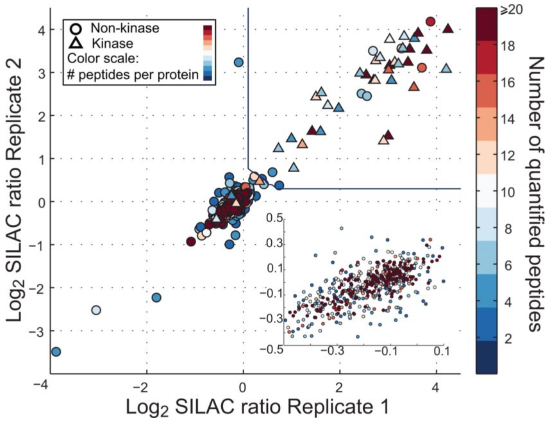

Fig. 3. Identifying significant targets of K252a in soluble competition data. Scatter plot of 2 replicate experiments of K252a 100× SC. Each data point is a single protein with kinases represented as triangles and circles denoting nonkinases. The color scale indicates the number of identified peptides per protein. The contour line demarcates a local FDR of 0.01 and all data points to the top right corner of the plot are inferred targets. Identified targets span a wide range of abundance. (*Inset*) Expanded view of the null distribution centered about $\log_2$ SILAC ratio of 0.

specificities as described above (Fig. 2*A* and Fig. S2*D*). Strikingly, BC experiments are heavily influenced by different bead loading levels (24) (6%, 12%, 25%), which we confirm with Western blot analysis, gel visualization, and modeled distributions of SILAC protein ratios (Figs. S2–S4). Although target proteins were generally found with the largest SILAC ratios in each BC experiment, large numbers of proteins were also found with SILAC ratios indicating specificity to either control EtOH-beads or SM-beads (≈100 proteins in Fig. 2*A*). This makes it difficult to apply meaningful significance thresholds as even a conservative threshold (for e.g., 1.5-fold, because differences of 20% are reliably quantified (25)) produced long lists of specific interactors. We also tested biologically inactive structural analogs of SMs as controls in BC experiments. Although providing a more similar binding profile to the SM than EtOH-beads, this approach is not generally applicable because there is usually limited SAR around screening hits.

With direct enrichment in BC experiments, SILAC ratios describe the proteins' binding preference to either the EtOH-bead or the SM-bead. The major contributors to the magnitude of a protein ratio are its abundance in the lysate and its affinity to the bait. For example, a highly abundant protein with weak affinity could have the same ratio as a low abundance protein with a very high affinity (low K_D). We mapped proteins from BC experiments to those identified and ranked by abundance in analyses of un-enriched HeLa S3 lysates (Fig. S3*B* and SI Methods, *Accompanying Text for Fig. S3*). Proteins with low to moderately differential ratios identified in BC experiments were among the most abundant proteins identified from whole lysates. We similarly analyzed SC data and found the same tendency for SM-beads to enrich abundant proteins in SC experiments (Fig. S3*C*). As we show below, however, this does not negatively impact our ability to discriminate target proteins.

Because SC experiments use the same SM affinity matrix in both light and heavy states, effectively enriching the same proteins from both cellular states, this approach is more elegant and easier to implement. Competitive binding by excess soluble SM reduces the amount of target proteins on the bead surface and generates a differential SILAC ratio between the 2 states, confidently identifying specific interactors to the soluble SM. Although abundant proteins still bind to the SM-bead, these show no differential ratios as they do not bind to soluble SM and hence have no impact on discrimination of specific SM interactors. Distributions of SILAC ratios are tightly centered about the lysate mixing ratio and show far less dependence on bead loading levels and compound used (Fig. 2*B*). We model SC protein ratios using an empirical Bayes strategy (*SI Methods*, manuscript submitted). In a typical SC dataset (Fig. 3), nonspecific binders are tightly centered about the one-to-one mixing ratio ($\log_2$ SILAC ratio of zero). SM-specific proteins are outliers to the main distribution and are easily identified by their SILAC ratios (Table 1). Abundance of a target protein neither affects its ratio nor our ability to discriminate it from the nonspecific binders and, as the color scale in Fig. 3 shows, the more abundant proteins in the sample are mostly nonspecific interactions with the SM-bead.

We compared BC and SC data for the broad spectrum kinase inhibitor, K252a. From Fig. 4, it is striking that the total number of kinases (colored red) identified from experiments is similar between all K252a experiment types, yet there is a dramatic difference in the SILAC ratios, and consequently, the ability to discriminate specific binders from nonselective interactions. Using annotated human kinases (26) as the "true positive" set, both SC and BC experiments identified approximately the same number of kinases. In K252a precision vs. recall (PRC) plots, the precision (TP/(TP+FP)) of the SC experiment was markedly better than the BC experiment and has much better specificity for kinases (Fig. S5*A*). For example, at a $\log_2$ median ratio significance threshold of 0.5, for the SC experiment 37 of the 49 identified proteins are kinases (precision 75.5%), whereas for the BC experiment only 40 of the 300 identified proteins are kinases (precision 13.3%).

Kinase Inhibitors. We included 3 kinase inhibitors in our studies, 2 selective kinase inhibitors, Ro-31-7549 and SB202190 and a broad-

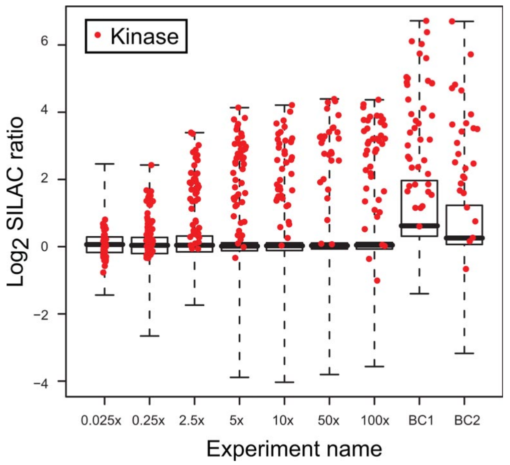

Fig. 4. Comparing K252a SC and BC experiments. Box plot of log$_2$ SILAC ratios for K252a SC (0.025×, 0.25×, 2.5×, 5×, 10×, 50×, and 100×) and BC experiments (two replicates per condition). Protein kinases (26) identified in each experiment are plotted in red, and the total number of kinases are provided in the Table 2.

specificity kinase inhibitor, K252a. Although other proteomics approaches have been applied to study the targets of the bisindolylmaleimide-based family of protein kinase C (PKC) inhibitors (17, 27), the overlap between the 2 studies is relatively small (14% of total targets). Whereas we found that 9 of these 10 common targets are specific binders in our SC experiments with Ro-31-7549 (a structural variant of Bis-III), the remaining protein, HSP90, and most of the other putative binders from these papers were found to be nonspecific in our datasets (blue and red circles in Fig. 2 *Lower*). The discrepancy in the large number of putative targets in BC and SC experiments underscores the need for a suitable affinity matrix as a control. Many examples from the literature do not use control pull-downs, probably because the cross-section of identified proteins varies dramatically depending on the SM loaded, making direct comparison difficult and consequently diminishing their usefulness. Although we have striven to generate useful control matrices by matching load levels of SM and EtOH on Affigel beads, protein binding profiles in our BC experiments were so different that interpretation of binding specificity using SILAC ratios was hampered (Fig. 2 and Figs. S2–S4). Nevertheless, it is important to note that we were still able to exclude several candidate targets of Ro-31-7549 based on our SILAC BC data.

The BC experiments with the MAPK14 (p38 MAPK) kinase inhibitor, SB202190, generated a very distinct binding profile compared with other SMs we tested. Very little protein bound to the SB202190 bead compared with EtOH-bead with the majority of identified proteins having negative log$_2$ SILAC ratios. Nevertheless, known target proteins still had high positive SILAC ratios indicating specificity to SB202190-loaded bead, a testament to the robustness of our quantitative approach. Ratio distributions with SB202190 SC experiments were similar to other SC experiments, and although the amounts of protein captured by the kinase inhibitor reagent were low to begin with, we found no difficulty in identifying target proteins by their SILAC ratios.

In BC experiments with the staurosporine analog, K252a using 2 human cell lines (HeLa S3 and H1299) and rat PC-12 cells expressing ErbB4, we find that the cell type used has a dramatic effect on the kinases identified (Fig. S5*B*). For example, we identified brain-specific isoforms, alpha and beta, of calcium/calmodulin-dependent protein kinase type from neuroblastoma PC-12 cells but not in non-CNS derived H1299 and HeLa lines. The kinases found in K252a pull-downs were predominantly Ser/Thr-kinases, and even the 3 tyrosine kinases (FER, IRAK4, RIPK2) in the list had kinase domains more closely related to dual specificity kinases.

In addition to identifying kinases that bind K252a, we also found several well-known interaction partners to primary targets with significant SILAC ratios in our SC experiments. Kinases (26) made up 37 of the 48 significant hits in our K252a 100× SC experiment (local FDR <0.01, log$_2$ SILAC ratio of 0.71, and identified by more than 2 peptides across 2 replicates), and we also identified 6 additional proteins that were either functionally coupled to kinases or well-known partners like cyclin B1 (Table 1). Of the 5 remaining candidate targets, 2 were already described as nonkinase targets: the calcium/calmodulin-dependent 3′,5′-cyclic nucleotide phosphodiesterase, PDE1A (28) and NAD(P)H:quinone oxidoreductase, NQO2 (17, 21). We also find an E3 ubiquitin ligase among the significant hits, and although we detect both AP2B1 and AP2A1 just beyond our significance threshold (local FDR 0.064 and 0.049, respectively) we believe this set of proteins involved in clathrin-mediated recycling to be likely hits as well. Therefore, our SILAC SC K252a experiments demonstrate exquisite sensitivity and specificity, with 46 of 48 significant hits directly linked to kinase biology from a background of 510 nonspecific binders.

Immunophilin Ligand Series. We previously identified the high affinity target of the widely prescribed immunosuppressant and natural product FK506 (tacrolimus) with a classical biochemical purification approach (13). To test our ability to identify weakly bound target proteins, we generated 4 structural variants in this IPL series (from ref. 29, and *SI Methods*) and determined their binding affinities (K_D: 25.7 nM to 43.8 μM) to FKBP1A-GST (Fig. 5*A* and Table S1). Affinity reagents generated with these 4 compounds were used in affinity pull-down experiments using both BC and SC formats. We successfully identified FKBP1A with SILAC ratios indicating specific interaction to the IPLs, effectively validating the SPR data and importantly, demonstrating our ability to identify a weakly bound target protein (K_D of Pro-AP1780 is 43.8 μM). We also identified other FKBP family members (FKBP2, FKBP4, FKBP5, FKBP9, FKBP10) in different pull-downs with IPLs, indicating that each IPL showed distinct specificities for members of the immunophilin family (Fig. 5*B*). We validated these with

Table 2. Total number of kinases from experiments shown in Fig. 4

	Soluble competitor[*]							Bead control[†]	
	0.025×	0.25×	2.5×	5×	10×	50×	100×	BC1	BC2
Kinase ID	50	54	47	48	42	28	42	40	26
Significant kinases	0	12	30	41	38	26	37	40	23
Significant nonkinases	4	5	4	7	9	6	11	232	116
Significant proteins	4	17	34	48	47	32	48	272	139
Total proteins identified[‡]	702	727	625	686	597	479	621	515	364
Significance threshold	1.3	1.17	1.08	0.66	0.73	0.78	0.71	0.58	0.58

[*] Log2 SILAC ratio significance threshold (FDR = 0.01) for soluble competitor experiments.

[†] Log2 SILAC ratio cut-offs for bead control experiments (>1.5-fold change).

[‡] Proteins identified by at least three peptides across two replicate experiments.

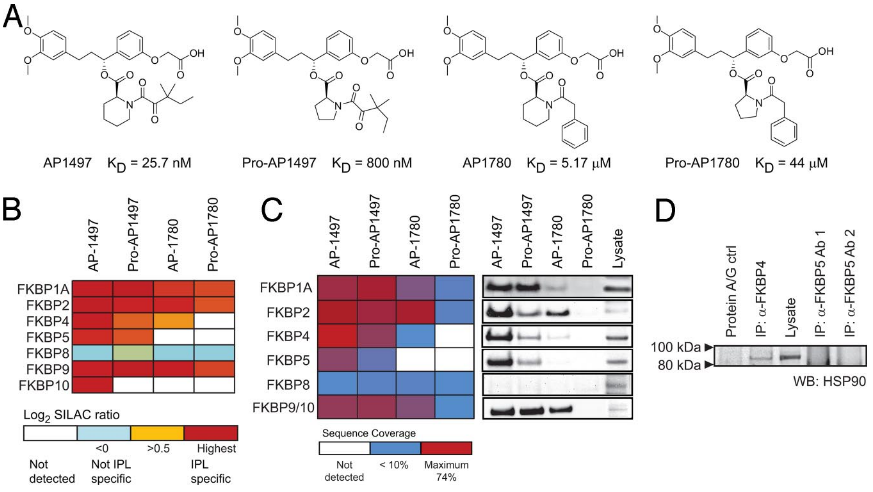

Fig. 5. Immunophilin ligand series. (*A*) Structures of immunophilin ligands and measured K_D values for FKBP1A. (*B*) Specificity of FKBP proteins for IPL ligands determined by their SILAC ratios. (*C*) Sequence coverage for FKBP proteins and validation by Western blot analysis. (*D*) Validation of HSP90-FKBP4 interaction by coimmunoprecipitation and Western blot analysis. Two FKBP5 antibodies failed to coprecipitate any significant amounts of HSP90.

Western blot analysis (Fig. 5*C*), and compared the latter to protein sequence coverage obtained from our MS analyses as an approximation of relative abundance across these samples. Both MS-based abundance estimates and Western blot analysis data show excellent agreement. Furthermore, our identification of the FKBPs with these IPLs is completely consistent with known biology for this protein family, and to our knowledge, is the first report of an unbiased proteomic survey to assess their binding specificities. FKBP8 was identified in our pull-down experiments but, consistent with structure information (30), its SILAC ratio classified it as a nonbinder to the IPLs.

Isoforms of heat shock protein 90, HSP90AA1 and HSP90AB1, were found only in AP1497 pull-downs. Because FKBP4 is known to interact with HSP90 (31), and FKBP4 was found primarily in the AP1497 pull-down, we hypothesized that HSP90 was identified in our experiments through its interaction with FKBP4. We performed co-immunoprecipitation-Western blot analysis experiments with anti-FKBP4 and anti-FKBP5 antibodies and validated FKBP4-HSP90 binding in HeLa cell lysates (Fig. 5*D*).

We identified an interactor, methylthioadenosine phosphorylase (MTAP), to all members of our IPL series in both BC and SC affinity pull-downs (Fig. S6 and SI Methods, *Accompanying Text for Fig. S6*). MTAP was found to be most highly enriched by Pro-AP1780 and we validated this with Western blot analysis and SPR experiments (K_D equil: 18 nM; K_D kinetic: 12 nM) (Fig. 6*A*, Fig. S7 and Table S2). Furthermore, we demonstrate dose-dependent inhibition of MTAP by Pro-AP1780 in a biochemical assay with endogenous MTAP from HeLa cell lysates (Fig. 6*B*).

Discussion

Our combination of SILAC with SM affinity enrichment greatly improves sensitivity and specificity of unbiased affinity purification-based target identification methods. We routinely quantify >600 proteins from single affinity pull-down experiments and still identify weakly bound targets with K_D values in the range of 40 μM (Pro-AP1780 in the IPL series) without compromising specificity. We identified targets of the IPLs, including MTAP, which we further validated in biochemical assays. Our kinase inhibitor experiments yielded known protein targets, associated protein complexes and proteins with related biology described in literature. SC experiments with the broad-spectrum kinase inhibitor K252a show that it has exquisite selectivity for kinases particularly at high concentrations of soluble competitor.

The SC experiment is clearly the experiment of choice in SILAC target identification. It rapidly yields specific interactors to SM baits; it is largely unaffected by highly abundant, but weakly bound proteins; and because it uses the same SM affinity matrix in both experiment and control samples, issues relating to compound loading or bead matrices are circumvented. We observe generally, and find unambiguous evidence with K252a and MTAP, that SC experiments should always be performed at the highest possible concentration of soluble competitor to yield the best results. Although this dependency on the level of SC may seem a limitation of the SC experiment, in our experience, SC experiments show higher specificity in identifying protein targets when compared with BC experiments. If compound solubility becomes an issue and precipitation occurs upon addition to cell lysates, the BC experiment with SILAC is then still an attractive option.

We show that compound-loading levels play a dramatic role in the binding specificities of affinity supports, SMs or controls alike. The amount of compound loaded on the bead affects the biophysical properties of the beads and is the probable cause for the change in binding characteristics observed across different bead loadings in our BC experiments (24). This effect is also highly dependent on the chemical properties of the SM loaded. We analyzed BC experiments across different bead loading levels using k-means clustering to track abundant, weakly interacting proteins (Fig. S3) and found this useful in segregating weak binders from true targets. However, performing multiple BC experiments for each molecule is time

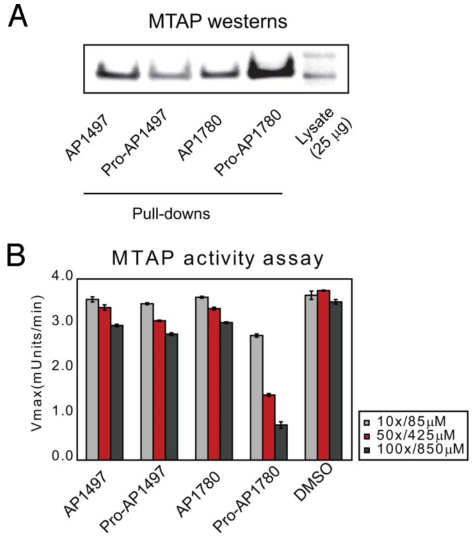

Fig. 6. MTAP is a protein target for the IPL ligand Pro-AP1780. (*A*) Validation with Western blot analysis of MTAP in IPL pull-downs and cell lysate. (*B*) In vitro MTAP activity assay with 3 dose levels of each IPL ligand. All IPL ligands show slight inhibition of MTAP at the highest levels of compound dose (850 μM). Pro-AP1780 shows the largest inhibitory effect on MTAP in comparison with DMSO treated controls.

consuming and unnecessary, particularly because using the SC experiment obviates this issue altogether.

In addition to rapid identification and prioritization of targets for further validation, using SILAC with BC or SC experiments allows unbiased evaluation of the experimental data through modeled distributions of protein ratios. SILAC ratios describe binding specificities for each protein, and because quantitative values are derived from individual peptides, it is possible to distinguish differential binding even among protein isoforms. Because identification of SM-specific binders does not depend on gel visualization, less input protein is required while permitting the use of mild buffer conditions that are generally applicable across a variety of molecules. Optimizing wash conditions to visualize differential binders is unnecessary and the sensitivity of our approach is significantly better, allowing identification of even weakly interacting protein targets. Furthermore, quantitative ratios classifying target specificity are independent of protein abundance in SC experiments, avoiding the bias of abundant proteins that would otherwise be a significant confounder in classical approaches.

We use the GeLCMS approach for MS sample processing that provides certain advantages but limits throughput and would not detect protein targets covalently coupled to SMs. Developing a gel-free sample processing workflow with on-bead enzymatic digestion of proteins would make identification of suicide inhibitor targets tractable (32, 33) and would be more amenable to automation. Although analytical throughput was not the primary focus of our studies, our current implementation of SILAC-Target I.D. already allows ≈15 target identification projects per MS instrument year. In our experience, this is significantly faster than the pace of classical affinity-based target identification. This improved throughput is afforded by the high specificity of our quantitative approach, allowing direct assessment of the quality of the data, greatly reducing the scale and number of subsequent validation experiments. Our current process has high potential for automation, and further improvements in capacity and throughput can already be achieved through implementation of liquid handling robots and an automated data analysis pipeline.

Affinity-based target identification is not routine in the pharmaceutical industry not only because of low throughput but primarily because drugs undergo several rounds of optimization and thus may lack functional handles allowing generation of affinity reagents. Diversity-oriented synthesis (DOS) can generate SM screening collections with stereochemical and skeletal diversity approaching that of natural products, yet allow facile optimization of structural variants and eventual large-scale synthesis of the optimized compound (34). We apply DOS to enrich screening decks with SMs poised for systematic conversion to affinity reagents (35). Although our method is applicable to all SM affinity reagents, we are now integrating SILAC-Target I.D. to characterize binders of bioactive molecules arising from our DOS compound-cell based screening pipeline.

SILAC-Target I.D. combines quantitative proteomics and biochemical enrichment using affinity matrices to provide unbiased, highly specific and robust identification of protein-SM targets. Our method circumvents many of the usual problems with classical affinity purification, and promises to be a general solution applicable to many affinity-based target identification projects. We expect that our target identification approach, particularly if implemented at early stages of the probe- or drug-discovery process, will transform the search for new probes or drugs because it provides a direct and unbiased interrogation of the cellular context in which a SM acts, essential in evaluating, for example, drug safety and efficacy.

Materials and Methods

SILAC labeled cell lysates were applied in affinity enrichment experiments using affinity matrices loaded with immunophilin ligands or kinase inhibitors. Two experimental designs, BC and SC, were compared. Proteins bound to solid phase were separated by SDS/PAGE and identified and quantified by high performance MS. SILAC ratios from relative abundances of proteins enriched in case vs. control pull-down experiments were modeled using Empirical Bayes-based statistical framework to identify specific protein targets interacting with SMs. Detailed methods for all experiments are provided in *SI Methods*. For analyses identifying SM protein targets, see *SI Methods*, Table S3, and Dataset S1.

ACKNOWLEDGMENTS. We thank Jeremy Duvall for assistance in synthesis of the K252a reagent; Inese Smutske for synthesis of IPL reagents; Carlos Tassa for SPR experiments; Andrew Lach, Vincent A. Fusaro, and Karl R. Clauser for software development; and Edward M. Scolnick and the members of the Proteomics platform and Chemical Biology program for helpful discussions. This research was supported by National Cancer Institute's Initiative for Chemical Genetics Contract N01-CO-12400 and National Institutes of Health Genomics Based Drug Discovery—Target ID Project Grant RL1HG004671, which is administratively linked to National Institutes of Health Grants RL1CA133834, RL1GM084437, and UL1RR024924.

1. Vesely J, et al. (1994) Inhibition of cyclin-dependent kinases by purine analogues. *Eur J Biochem* 224:771–786.
2. Fabian MA, et al. (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* 23:329–336.
3. Terstappen GC, Schlupen C, Raggiaschi R, Gaviraghi G (2007) Target deconvolution strategies in drug discovery. *Nat Rev Drug Discov* 6:891–903.
4. Roth BL, Sheffler DJ, Kroeze WK (2004) Magic shotguns versus magic bullets: Selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* 3:353–359.
5. Lindsay MA (2003) Target discovery. *Nat Rev Drug Discov* 2:831–838.
6. Stockwell BR (2004) Exploring biology with small organic molecules. *Nature* 432:846–854.
7. Carpenter AE (2007) Image-based chemical screening. *Nat Chem Biol* 3:461–465.
8. Stegmaier K, et al. (2004) Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. *Nat Genet* 36:257–263.
9. Butcher RA, Schreiber SL (2005) Using genome-wide transcriptional profiling to elucidate small-molecule mechanism. *Curr Opin Chem Biol* 9:25–30.
10. Strausberg RL, Schreiber SL (2003) From knowing to controlling: A path from genomics to drugs using small molecule probes. *Science* 300:294–295.
11. Lamb J, et al. (2006) The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935.
12. Cuatrecasas P, Wilchek M, Anfinsen CB (1968) Selective enzyme purification by affinity chromatography. *Proc Natl Acad Sci USA* 61:636–643.
13. Harding MW, Galat A, Uehling DE, Schreiber SL (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* 341:758–760.
14. Brown EJ, et al. (1994) A mammalian protein targeted by G_1-arresting rapamycin-receptor complex. *Nature* 369:756–758.
15. Bach S, et al. (2005) Roscovitine targets, protein kinases and pyridoxal kinase. *J Biol Chem* 280:31208–31219.
16. Godl K, et al. (2003) An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc Natl Acad Sci USA* 100:15434–15439.
17. Brehmer D, Godl K, Zech B, Wissing J, Daub H (2004) Proteome-wide identification of cellular targets affected by bisindolylmaleimide-type protein kinase C inhibitors. *Mol Cell Proteomics* 3:490–500.
18. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422:198–207.
19. Blagoev B, et al. (2003) A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nat Biotechnol* 21:315–318.
20. Ranish JA, et al. (2003) The study of macromolecular complexes by quantitative proteomics. *Nat Genet* 33:349–355.
21. Bantscheff M, et al. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* 25:1035–1044.
22. Daub H, et al. (2008) Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell* 31:438–448.
23. Ong SE, et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1:376–386.
24. Lee WC, Cheng CH, Pan HH, Chung TH, Hwang CC (2008) Chromatographic characterization of molecularly imprinted polymers. *Anal Bioanal Chem* 390:1101–1109.
25. Ong SE, Kratchmarova I, Mann M (2003) Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* 2:173–181.
26. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* 298:1912–1934.
27. Saxena C, Zhen E, Higgs RE, Hale JE (2008) An immuno-chemo-proteomics method for drug target deconvolution. *J Proteome Res* 7:3490–3497.
28. Matsuda Y, Nakanishi S, Nagasawa K, Iwahashi K, Kase H (1988) The effect of K-252a, a potent microbial inhibitor of protein kinase, on activated cyclic nucleotide phosphodiesterase. *Biochem J* 256:75–80.
29. Keenan T, et al. (1998) Synthesis and activity of bivalent FKBP12 ligands for the regulated dimerization of proteins. *Bioorg Med Chem* 6:1309–1335.
30. Maestre-Martinez M, et al. (2006) Solution structure of the FK506-binding domain of human FKBP38. *J Biomol NMR* 34:197–202.
31. Peattie DA, et al. (1992) Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat shock protein and is a component of steroid receptor complexes. *Proc Natl Acad Sci USA* 89:10974–10978.
32. Liu Y, Patricelli MP, Cravatt BF (1999) Activity-based protein profiling: The serine hydrolases. *Proc Natl Acad Sci USA* 96:14694–14699.
33. Cravatt BF, Wright AT, Kozarich JW (2008) Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* 77:383–414.
34. Burke MD, Schreiber SL (2004) A planning strategy for diversity-oriented synthesis. *Angew Chem Int Ed Engl* 43:46–58.
35. Wang X, Imber BS, Schreiber SL (2008) Small-Molecule Reagents for Cellular Pull-Down Experiments. *Bioconjug Chem* 19:585–587.