HHS Public Access

Author manuscript

Nat Chem. Author manuscript; available in PMC 2022 March 25.

Published in final edited form as:

Nat Chem. 2021 November ; 13(11): 1081–1092. doi:10.1038/s41557-021-00765-4.

A Proteome-Wide Atlas of Lysine-Reactive Chemistry

Mikail E. Abbasov[1,2,*], **Madeline E. Kavanagh**[1], **Taka-Aki Ichu**[1], **Michael R. Lazear**[1], **Yongfeng Tao**[1], **Vincent M. Crowley**[1], **Christopher W. am Ende**[3], **Stephan M. Hacker**[4], **Jordan Ho**[5], **Melissa M. Dix**[1], **Radu Suciu**[1], **Matthew M. Hayward**[3], **Laura L. Kiessling**[5], **Benjamin F. Cravatt**[1,6,*]

[1]Department of Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[2]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA.

[3]Medicine Design, Pfizer Worldwide Research and Development, Eastern Point Road, Groton, CT 06340, USA

[4]Department of Chemistry, Technische Universität München, Lichtenbergstraße 4, 85748 Garching, Germany

[5]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Lead Contact

Abstract

Recent advances in chemical proteomics have begun to characterize the reactivity and ligandability of lysines on a global scale. Yet, only a limited diversity of aminophilic electrophiles have been evaluated for interactions with the lysine proteome. Here, we report an in-depth profiling of >30 uncharted aminophilic chemotypes that greatly expands the content of ligandable lysines in human proteins. Aminophilic electrophiles showed disparate proteomic reactivities that range from selective interactions with a handful of lysines to, for a set of dicarboxaldehyde fragments, remarkably broad engagement of the covalent small molecule-lysine interactions captured by the entire library. We used these latter “scout” electrophiles to efficiently map ligandable lysines in primary human immune cells under stimulatory conditions. Finally, we show that aminophilic compounds perturb diverse biochemical functions through site-selective modification of lysines in proteins, including protein-RNA interactions implicated in innate

*Correspondence: mikail.abbasov@cornell.edu (M.E.A.), cravatt@scripps.edu (B.F.C.).

Author Contributions. M.E.A. and B.F.C. conceived the research, designed experiments and analysed data. M.E.A. performed mass spectrometry experiments and data analysis. M.E.A. designed and synthesized compounds, cloned, expressed and purified proteins, and conducted RIDA and Ku70-Ku80 studies. M.E.A., M.R.L., R.S. and M.M.D. compiled and analysed mass spectrometry data. M.E.K. and M.E.A conducted IFIT1 and IFIT5 studies and data analysis. T.A.I. and M.E.A conducted LPCAT1 studies and data analysis. M.E.A., Y.T., and V.M.C. characterized synthetic compounds. V.M.C. conducted reactivity studies with a model amine nucleophile. C.W.E. and M.M.H. designed and synthesized compounds **17a-l**,**r**, **18a-c**, and **19a-g**. S.M.H. designed and synthesized compounds **17m-q**,**s**,**t**. J.H. and L.L.K. designed and synthesized compounds **33a,b**. M.E.A. and B.F.C. wrote the manuscript.

Competing interests. B.F.C. is a founder and scientific advisor to Vividion Therapeutics, a biotechnology company interested in developing small-molecule therapeutics.

Supplementary Information. Included are Extended Data Figures 1–10, Supplementary Datasets 1–4, Supplementary Table 1, and Supplementary Note – Synthetic and Analytical Chemistry.

Code availability. Custom code used for proteomic data processing have been deposited in Github.

immune responses. These findings support the broad potential of covalent chemistry for targeting functional lysines in the human proteome.

Small-molecule probes are critical for illuminating the biological functions of proteins and serve as leads for the discovery of therapeutics[1]. At present, the vast majority of human proteins lack selective chemical probes and certain categories of proteins are considered potentially undruggable[2]. Historical strategies for discovering chemical probes, such as high-throughput screening of large compound libraries[3], have been more recently complemented by alternative approaches that include fragment-based drug discovery[4] and covalent ligand development[5], which have been applied proteome-wide by leveraging reactive chemical probes and quantitative mass spectrometry (MS) methods[6–8]. By combining features of recognition and reactivity, electrophilic compounds can engage more shallow or dynamic binding pockets on proteins, thereby expand the scope of proteins targeted by small molecules, and these irreversible interactions may also produce extended pharmacological effects that are maintained until protein targets physically turnover in the cells[5].

Original covalent probes mainly targeted catalytic serine/threonine[9] or cysteine[10] residues located within the active sites of enzymes. More recently, however, covalent probes and drugs have been developed that target non-catalytic cysteine residues in functional sites of proteins, such as the ATP-binding pockets of kinases[11], the substrate recognition groove of the nuclear export receptor XPO1[12], and the oncogenic G12C variant in KRAS[13]. The propensity of the cysteine thiol group to react with covalent probes and drugs is not surprising, given its greater relative nucleophilicity compared to other amino acid side chains under physiological conditions. Productive covalent binding to other amino acids, such as lysine, typically depends on the surrounding microenvironment, which may either perturb the pK_a of the lysine amino group or support high effective molarity of the reactive compound through tight reversible binding. Given these constraints, examples of site-specific covalent binding to non-catalytic residues other than cysteine remain limited.

Expanding the scope of covalent probes also depends on understanding the reactivity and chemoselectivity of candidate electrophiles. Diverse electrophilic groups can form covalent adducts with proteinaceous lysines, including sulfonyl fluorides[14], fluorosulfates[15], dichlorotriazines[16], activated esters[17], activated sulfonamides[18], vinyl sulfonamides[19], imidoesters[20], isothiocyanates[21], salicylaldehydes[22], iminoboronates[23], and α,β-unsaturated carbonyls[24]. These “aminophilic” chemotypes show varying degrees of selectivity for lysine over other amino acids, but very few have been evaluated for reactivity on a proteome-wide scale, and, only in rare cases, have these chemotypes been leveraged to create chemical probes that can site-specifically target individual lysines to perturb protein function[7,24,25]. Despite this, there remains great potential for aminophilic electrophiles to progress to more advanced chemical probes and drugs, as exemplified, for instance, by the recent development of voxelotor (GBT440). This salicylaldehyde drug is used to treat sickle cell disease and engages the *N*-terminal amine of hemoglobin in a reversible covalent bond to increase hemoglobin affinity for oxygen[22].

We have previously described a chemical proteomic strategy for quantifying the site-specific reactivity and small-molecule interactions of nucleophilic amino acid residues in native biological systems[6,26,27]. In this activity-based protein profiling (ABPP) approach, libraries of electrophilic compounds are evaluated for their ability to "compete" or block proteomic interactions of a chemical probe that displays broad, chemoselective reactivity with a specified amino acid. Such ABPP probes were originally developed for cysteine[27] and lysine[7,8], and have, more recently, been extended to aspartate/glutamate[28–30], methionine[31], and tyrosine[32]. Initial studies with the lysine-directed probe sulfotetrafluorophenyl pentynoate (STPyne), however, only assessed a limited diversity of aminophilic electrophiles (activated esters and *N,N'*-diacyl-pyrazolecarboxamidines) in the human proteome[7]. Consequently, our understanding of the types of electrophiles that can react site-selectively with lysine residues to afford functional outcomes remains limited.

Motivated by these findings, we hypothesized that a more thorough understanding of the ligandability of lysines in the human proteome could be achieved by profiling a much broader array of aminophilic chemotypes in diverse human cell types and states. With this goal in mind, we report herein the chemical proteomic analysis of lysine reactivity for ~180 compounds distributed across >30 aminophilic chemotypes, including both covalent reversible and irreversible electrophiles and hitherto uncharacterized terpene natural products. Across >14,000 total lysines quantified in human cancer cell line and primary human immune cell proteomes, we identified numerous sites showing preferential reactivity with distinct aminophilic chemotypes. These liganded lysines are found in structurally and functionally diverse proteins, and we show, in several cases, that site-specific engagement by aminophilic compounds affects protein function. We furthermore report the discovery of a remarkable set of dicarboxaldehydes that broadly landscape the ligandable lysine proteome, thus constituting versatile "scout" compounds for efficient mapping of small molecule-lysine interactions in diverse biological systems. Finally, we show that cyanomethyl acyl sulfonamide compounds serve as cell-active probes that perturb the RNA-binding interactions of the IFIT family of innate immune proteins by targeting a conserved lysine residue. These results underscore how the proteome-wide analysis of lysine-reactive chemistries can uncover new chemical tools for heretofore unliganded proteins.

Results

Design of a lysine-reactive library of aminophilic compounds.

We synthesized an ~180-member compound library composed of 34 distinct aminophilic chemotypes tethered to structurally diversified molecular recognition (or binding) elements intended to promote interactions with distinct proteins and afford initial structure-activity relationships (SARs) both *across* and *within* chemotypes. (Figure 1a, b; see Supplementary Dataset 1 for structures of aminophilic compounds). The compounds, which can be subcategorized based on their predicted modes of reactivity (Figure 1a), had an average molecular weight of 312 Da and were prepared using three or fewer synthetic steps. Most library members were low molecular weight fragments, but a subset had more elaborated structures, representing structural modifications to natural products (**28n**, **33h**)

and drugs (e.g., **7a**, a derivative of glibenclamide; **13d**, a derivative of celecoxib)[33,34] (Figure 1b and Supplementary Dataset 1). Attention was also paid in the library design to installation of polarizable groups proximal to the aminophilic center, which we hoped would promote reactivity with lysines at compound-protein interaction sites (e.g., hydroxyl groups positioned adjacent to aminophilic centers with potential to displace water molecules in the hydration shell of solvent-exposed lysines (**26a-q, 28a-u, 29a-g**)). Most of the library was compliant with the Lipinski rule-of-five values for lead- and drug-like compounds[37] (Figure 1c and Extended Data Fig. 1).

As an initial qualitative assessment of the proteome-wide reactivity and chemoselectivity of each compound, we performed gel-ABPP experiments in human cancer cell lysates with fluorescent, broad-spectrum probes targeting individual nucleophilic amino acids - a lysine-directed probe Alexa-Fluor® 488 (**P3**)[7], a cysteine-directed probe iodoacetamide-rhodamine (**P7**)[6], and a serine-directed probe fluorophosphonate-rhodamine (**P8**) (Supplementary Dataset 1)[36]. Representative gel-ABPP results are shown in Figure 1d–f for heterocyclic aldehydes **32a**-**i**, which blocked several **P3**-protein interactions (Figure 1d), but did not show evidence of cross-reactivity with **P7**-labeled (Figure 1e) or **P8**-labeled (Figure 1f) proteins. Dicarboxaldehyde **32i** was notable in that it impaired **P3**-reactivity with many proteins, while still preserving strong apparent chemoselectivity (negligible blockade of **P7**- or **P8**-reactive proteins) (Figure 1d–f). The atypically broad lysine reactivity of dicarboxaldehydes will be revisited below. Other aminophilic chemotypes blocked **P3**-labeled proteins with a selectivity that ranged from exclusive to preferential over **P7** and **P8**-labeled proteins (Supplementary Dataset 2). These initial gel-ABPP experiments suggested that the aminophilic compound library engages diverse lysine residues in the human proteome while showing limited cross-reactivity with other amino acids.

Proteomic analysis of aminophilic compound-lysine interactions.

We next screened aminophilic compounds for blockade of probe **1** (**P1**; Supplementary Dataset 1) labeling events of lysines by the mass spectrometry (MS)-based proteomic method isoTOP-ABPP (isotopic tandem orthogonal proteolysis-activity-based protein profiling)[7]. Experiments were performed using two human cancer cell line proteomes representing suspension hematological (Ramos, Burkitt's lymphoma) and adherent epithelial (MDA-MB-231, breast) cancer cells, respectively, which we have previously found to display complementary protein content[6]. Cancer cell proteomes were pre-treated with aminophilic compounds (10–100 μM, 1 h, 23 °C) or DMSO followed by **P1** (100 μM, 1 h, 23 °C), after which **P1**-labeled proteins in compound- and DMSO-treated proteomes were conjugated to isotopically differentiated azide-biotin tags (heavy and light, respectively) by copper-catalyzed azide-alkyne cycloaddition (CuAAC)[37], combined, enriched by streptavidin, proteolytically digested on-bead by sequential exposure to trypsin and TEV protease, and the TEV-released **P1**-labeled peptides analyzed by liquid chromatography-mass spectrometry (LC-MS) (Figure 2a). Lysine residues were considered 'liganded' if they showed substantial reductions (≥ 75%) in enrichment by **P1** in the presence of compounds compared to DMSO (MS1 chromatographic peak ratios (R), of ≥ 4 for DMSO/compound).

Most compounds were screened against both Ramos and MDA-MB-231 proteomes and against at least one of these proteomes in duplicate, resulting in >460 total isoTOP-ABPP data sets (Supplementary Dataset 3). A median number of 2,593 lysines was quantified per data set (Extended Data Fig. 2a), and we required that a lysine was quantified in at least 5 independent data sets for assessment of small-molecule interactions, or 'ligandability'. From an aggregate tally of 13,785 quantified lysines on 3,552 unique proteins, we identified 818 lysines on 581 proteins that were liganded by one or more aminophilic compound (Figure 2b and Supplementary Dataset 3). About half of the liganded lysines (55%) were engaged by a single compound, with the remaining lysines showing distributed interactions profiles ranging from two to many (> 5) aminophilic compounds (Extended Data Fig. 2b, left panel). The majority of proteins contained zero or one liganded lysine, indicating that lysine ligandability may often be site-specific within proteins[7,8] (Extended Data Fig. 2b, right panel). Each aminophilic chemotype was found to engage a discrete set of liganded lysines and displayed marked differences in their overall lysine reactivity, ranging from the least reactive benzoxazinones (**9**), heterocyclic sulfamates (**23**), and fluorosulfates (**18**) to the most reactive diacylphloroglucinols (**28**) and phthalaldehydes (**27**) (Figure 2c).

We also compared the reactivity of representative members of each aminophilic chemotype with a model amine (*N*α-acetyl-*L*-lysine-OMe) and found a generally good correlation with proteomic reactivity for the compounds (i.e., aminophilic compounds with strong proteomic reactivity also tended to show strong model amine reactivity; Extended Data Fig. 3). We noticed that diacylphloroglucinols (**28j, l, n**) tended to show greater proteomic reactivity versus model amine reactivity, which could reflect better stabilization of presumably reversible covalent adducts with proteinaceous lysines. And, there were also some exceptional compounds that showed strong model amine reactivity, but limited proteomic reactivity, such as the acyl pyridazinone **8a** and multiple acylcyclohexadione-containing polyketones (**26a-d**). We are unsure of the basis for these differences, but, for **26a-d**, possible steric hindrance surrounding the most electrophilic ketone could result in more limited reactivity with proteinaceous lysines.

The extent of lysine engagement for chemotypes and individual compounds did not correlate with cLogP or MW (Extended Data Fig. 4a). For instance, compounds **27d** and **28f**, which contained simple *ortho*-phthalaldehyde and elaborate 2-hydroxybenzaldehyde cores, respectively, engaged comparable numbers of lysines, despite varying considerably in their MWs (184 and 480 Da, respectively). Likewise, representative members of the carbonate chemotype - **14a** and **14f** displayed similar overall lysine engagement, but substantially different cLogP values (1.18 and 4.96, respectively). We also found that the relative lysine ligandability values of aminophilic compounds aligned with the frequency their representation across HBA, HBD, and RB categories (Extended Data Fig. 4b), indicating the potential for compounds of differing structures to engage lysines in the proteome.

Liganded lysines that showed broad cross-reactivity with diverse aminophilic chemotypes (Extended Data Fig. 4c) tended to correspond to residues that were also found to be liganded in our previous study that relied on activated esters as competitor compounds[7], suggesting that they might represent hot spots in the proteome for aminophilic compound reactivity. Even within this category, as well as more broadly across the entire set of

liganded lysines, we found evidence for a substantial recognition component directing small molecule interactions, as reflected in individual liganded lysines displaying markedly distinct SARs (Figure 2d) that, in some cases, opposed the overall reactivity profiles of the chemotypes. For instance, K153 in Hsc70-interacting protein (ST13) was preferentially targeted by squarate **33** over other compounds that showed much greater proteome-wide reactivity (Figure 2d, upper panel). Within-chemotype SAR was also apparent for this lysine, as it was engaged preferentially by **33e** over other squarates (Figure 2d, lower panel). Another clear example of distinct SAR was observed for **17r**, which showed a broader and more selective engagement of kinase active-site lysines compared to other sulfonyl fluoride compounds (Extended Data Fig. 4d, e), reflecting a recognition element that preferentially binds to the ATP-binding pocket of kinases[14].

The vast majority (~89%) of liganded lysines had not been previously identified to engage aminophilic small molecules (Figures 2e)[7], likely reflecting the much broader array of chemotypes used in the current study. The proteins harboring liganded lysines originated from diverse structural and functional classes (Figure 2f), a modest fraction (~23%) of which, consisting primarily of enzymes, have established interactions with small molecules as reflected by their presence in the DrugBank database (Figure 2f, left panel). The much larger fraction (~77%) of proteins with liganded lysines that were not represented in DrugBank showed a broad functional class distribution that included transcription factors/regulators and scaffolding/modulator/adaptor proteins (Figure 2f, right panel). Additionally, only a small fraction (~21%) of proteins with liganded lysines were found in other chemical proteomic studies to contain liganded cysteines (Figure 2g)[6].

Approximately a quarter (21%) of the liganded lysines represented established “functional” sites, including residues that undergo post-translational modification (e.g., acetylation ubiquitination/SUMOylation) or participate in substrate- or cofactor-binding (Figure 2h), including, for instance, active-site lysines in enzymes like GLUD1/2 (K183) and UGP2 (K396). A survey of the OMIM (Online Mendelian Inheritance in Man) database identified several human disease-relevant proteins with liganded lysines (Figure 2i) and an interesting subset of these cases where disease-causing missense mutations occurred in the liganded lysine residues themselves, including PMVK (K69→E) in individuals with porokeratosis 1[38], RPL10 (K78→E) in individuals with X-linked (MRXS35) microcephaly[39], and CPOX (K404→E) in patients with harderoporphyria form of hereditary coproporphyria[40] (Extended Data Fig. 4f–h). Such convergence of ligandability and human genetic data point to functional lysines with the potential to be targeted by chemical probes.

Characterization of aminophilic compound-lysine interactions.

We next aimed to verify and understand the functional consequences of representative aminophilic compound-lysine interactions. We noted that K404 of CPOX, which catalyzes the aerobic oxidative decarboxylation of coproporphyrinogen-III to protoporphyrinogen-IX during heme biosynthesis[40], was liganded by only a single member of the aminophilic compound library - sulfonyl fluoride **17b** (*R* value = 12.6; Figure 2a). This conserved lysine is enclosed within a cavity binding the heme precursor, coproporphyrinogen III (Figure 2b, left panel), and other quantified lysines located elsewhere in CPOX (K347, K370,

K371) were unaffected by **17b** (Figure 2a). Among a panel of commercially available aminophilic fluorescent probes used for convenient gel-based analysis of ligandable lysines in recombinantly expressed proteins[7] (Supplemental Dataset 1), we found that probe **P4** labeled wild-type (WT) CPOX, but not K404R or K404E mutants in transfected HEK293T cell proteomes (Figure 2b, right panel), and confirmed that **17b**, but not the structurally related sulfonyl fluoride **17c**, blocked **P4** labelling of WT-CPOX. Thioimido ester **5a** and formyl phloroglucinol **28p**, but not with structurally analogoues **5b** and **28j** (Figure 2b, right panel), partially blocked **P4** labeling of K404 of CPOX, matching the SAR profile acquired by chemical proteomics (*R* values of 3.9 and 3.3 for **5a** and **28p**, respectively).

A similarly strict SAR was observed by chemical proteomics for K53 of the glutathione S-transferase GSTT2B.[41] K53 is an active-site proximal residue (Figure 2c, left panel) and was found by chemical proteomics to be engaged by ammoniumsulfonyl carbamate **22b**, *N*-hydroxyphthalimide **12a**, and diketone **26h**, but not by analogues **22c**, **12c**, and **26j** (Figure 2c, right panel). Recombinant GSTT2B showed a similar SAR, as readout by site-selective profiling of K53 with probe **P1** (Figure 2c, right panel). The most active compound **22b** engaged K53 of GSTT2B with an apparent IC_{50} value of 3.7 μM (Figure 2d).

Compelling SAR profiles were also observed for lysines in scaffolding proteins, as exemplified by the conserved and homologous lysines K155 and K73 in the transcriptional repressors SIN3A and SIN3B, respectively. We previously found that K155, located in the first paired amphipathic helix (PAH1) domain of SIN3A, was liganded by an activated ester compound and this interaction blocked SIN3A interactions with binding partner TGIF1[7]. Here, we found that probe **P5** site-specifically modified the N-terminal PAH1 and PAH2 domains of SIN3A and SIN3B, but not their corresponding K155R and K73R mutants (Figures 2e and 2f). and confirmed that diformyl phloroglucinol **28i**, *N*-succinimidyl ester **4a**, and squarate **33e**, but not other compounds (**28h**, **4d**, and **33f**), liganded both SIN3A and SIN3B, generally matching the SAR profiles for endogenous forms of these proteins. We also found that *N*-hydroxyphthalimide **12a,** which possesses the same 3,5-bis(trifluoromethyl)phenyl recognition element found in an activated ester that engages K155 of SIN3A[7], liganded K155 and K73 with apparent IC_{50} values of 0.54 μM and 0.92 μM, respectively (Figure 2g). Taken together, these results demonstrate that diverse types of proteins possess lysines that can be liganded by aminophilic small molecules with interpretable SAR assignments that are preserved in the recombinantly expressed forms of the proteins.

For functional analysis, we prioritized proteins that lack chemical probes and/or were site-specifically liganded on lysines located at protein-protein interfaces. The liganded lysine K117 in the metabolic enzyme RIDA, which catalyzes the hydrolytic deamination of toxic enamine/imine intermediates, lines the cleft of the putative substrate-binding site[42] (Extended Data Fig. 5a). Recombinantly expressed WT-RIDA, but not a K117R mutant reacted with probe **P2**, and this interaction was blocked by pre-treatment with diketone **26l** and diformyl phloroglucinol **28h**, but not structural analogs **26k** and **28g**, respectively (Extended Data Fig. 5b). This SAR matched the chemical proteomic data acquired for K117 of endogenous RIDA. We were particularly interested in **26l**, which showed low micromolar activity (Extended Data Fig. 5c, d) and limited cross-reactivity with lysines across the

proteome (Extended Data Fig. 5e). We found that both **26l** and **28h** blocked RIDA catalytic activity with similar IC_{50} values to those measured by ABPP with probe **2** (Extended Data Fig. 5f, g). Neither compound blocked substrate hydrolysis mediated by a K117I mutant of RIDA, which retained near-wild type levels of activity (Extended Data Fig. 5f, g) despite being unreactive with **P2** (Extended Data Fig. 5h). In contrast, **26l**, but not **28h** retained inhibitory activity when tested against a K117R mutant (Extended Data Fig. 2g), which may indicate that the 1,3-dicarbonyl reactive group of **26l** can form covalent adducts with both lysine and arginine residues[43] (Extended Data Fig. 5i). Finally, we noted that the most common natural variant for K117 in the Exome Aggregation Consortium database is K117E, and the testing of this RIDA mutant revealed that it shows substantial reductions in catalytic activity and reactivity with **P2** (Extended Data Fig. 5h). These data thus demonstrate how chemical proteomics can identify residues for which engagement by small molecules or natural genetic mutation affect protein function.

Our chemical proteomic experiments furnished a rich map of quantified lysines in the DNA helicase XRCC6, one of which (K351) was liganded by diverse aminophilic compounds, including isatoic anhydride **11e** and squarate **33e** (Figure 3a). XRCC6, also known as Ku70, together with XRCC5 (or Ku80), forms the Ku70/Ku80 heterodimer that plays a pivotal role in non-homologous end-joining (NHEJ), an important pathway for repairing DNA double strand breaks in human cells[44]. K351 is located at the heterodimer interface of the Ku70/Ku80 complex and engages in a salt bridge with D475 of Ku80 (Figure 3b), and we found that pre-treatment of recombinantly expressed Ku70 with **11e** or **33e** blocked co-immunoprecipitation with Ku80 in a concentration-dependent manner (Figure 3c–e), with **11e** displaying greater potency (IC_{50} = 3.2 μM; Figure 3e, right panel). Neither **11e** nor **33e** blocked the co-immunoprecipitation of a K351R mutant of Ku70 with Ku80 (Figure 3d, e). On the other hand, pre-formation of the Ku70/Ku80 heterodimer prevented **11e** from disrupting the complex and its DNA-binding ability (Extended Data Fig. 6a, b). These results demonstrate that aminophilic compounds targeting K351 in Ku70 can block the formation of, but not disrupt pre-assembled Ku70-Ku80 heterodimers.

Dicarboxaldehydes as scout fragments for mapping lysine ligandability.

An overview of our chemical proteomic data identified three dicarboxaldehyde fragments **27c**, **28o**, and **32i** that liganded a remarkably high fraction (~58%) of the protein targets of the aminophilic compound library as a whole, as well as by activated ester compounds previously profiled for lysine reactivity[7] (Figure 4a, b). These compounds represent fluorogenic reagents used for analyzing amine metabolites and for traceless chemoselective bioconjugations with lysines[45,46] (**27c, 32i**), as well as a natural product with a reactive diformylphloroglucinol core (**28o**).

The dicarboxaldehyde fragments each liganded >200 lysine residues, which greatly exceeded the more limited engagement profiles of other aminophilic compounds in the library (Figure 4c), and showed overlapping, but distinct lysine interaction profiles (Extended Data Fig. 7a). Previous chemical proteomic studies evaluating cysteine-directed electrophilic fragments identified rare compounds that showed similarly broad patterns of reactivity[6], and these fragments have since been used as ‘scouts’ to efficiently survey the

cysteine ligandability of diverse biological systems[47], as well as to discover E3 ligases that support small molecule-mediated protein degradation[48]. We were therefore interested in understanding whether dicarboxaldehyde fragments could also be deployed as scouts for profiling lysine ligandability and functionality.

To explore the potential utility of **27c**, **28o**, and **32** as scout fragments for further expanding the fraction of lysines that can be targeted by aminophilic small molecules, we evaluated the reactivity of these compounds in primary human immune cell proteomes, specifically, the proteomes of human T cells activated by anti-CD3/CD28 antibodies and human peripheral blood mononuclear cells (PBMCs) ± stimulation with bacterial lipopolysaccharides (LPS). From a total of 7,881 quantified lysines on 2,495 unique proteins across the immune cell proteomes treated with scout fragments, we identified 1,439 liganded lysines on 867 proteins (Figure 4d and Supplementary Dataset 3). These liganded lysines were found in several immune-relevant proteins, defined as proteins with immune cell-enriched expression profiles and/or mutations that cause immune-related disorders in humans[49] (Figure 4d). Gene ontology (GO) term analysis confirmed the enrichment of diverse immune processes for proteins harboring liganded lysines (Figure 4e). A subset of liganded, immune-relevant proteins also showed heightened expression in LPS-stimulated PBMCs compared to quiescent PBMCs (Figure 4f), underscoring the importance of studying human immune cells in activated states in order to more broadly capture immune-relevant proteins.

We selected a liganded lysine (K221) in the immune-relevant protein LPCAT1, a lipid acyltransferase involved in phospholipid synthesis and remodeling[50], for further study due to its conservation among other LPCATs, as well as more distantly related AGPAT (acylglycerol-3-phosphate O-acyltransferase) enzymes (Extended Data Fig. 7b, c). We found that mutation of K221 to arginine blocked LPCAT1 activity (Figure 4g), suggesting that K221 may be involved in catalysis (three-dimensional structures of LPCAT1 and related LPCATs have not yet been determined). Consistent with this premise, each scout fragment inhibited LPCAT1 activity to variable extent, with **28o** showing the highest apparent potency (Figure 4g) and matching the reduction in activity of HEK239T cells transfected with the inactive K221R-LPCAT1 mutant (Figure 4g and Extended Data Fig. 7d). We next screened compounds from the **28** chemotype (Extended Data Fig. 7e–g) and found that **28f** showed the strongest LPCAT1 inhibitory activity (Extended Data Fig. 7e) with an IC_{50} value of 38 nM (Figure 4h and Extended Data Fig. 7h). Previous studies also indicated that K221 in mouse LPCAT1 was ubiquitinated[51]; however, we did not find evidence of ubiquitin modification of human LPCAT1 in a K221-dependent manner (Extended Data Fig. 7i–k).

We finally noted, in our scout fragment isoTOP-ABPP data sets, cases of differential ligandability of lysines in stimulated immune cells that occurred on proteins that did not show apparent alterations in expression. For example, K252 in the porphobilinogen synthase ALAD, showed substantially weaker interactions with scout fragments in LPS-stimulated vs control PBMCs (e.g., $R_{control \rightarrow LPS}$ 9.3→1.5 for scout fragment **27c**) (Extended Data Fig. 8). LPS treatment had little effect on the reactivity of a different lysine in ALAD (K159) (Extended Data Fig. 8).

Cyanomethyl acyl sulfonamides inhibit IFIT RNA-binding proteins by engaging a conserved lysine.

Proteins possessing scout fragment-sensitive lysines were found in 44 of the 47 immune cell-resolved functional modules (ME) established in a previous proteomic analysis of protein expression across diverse human immune cell types[52] (Figure 5a). Modules lacking liganded proteins mainly corresponded to rare immune cell types, such as plasmacytoid dendritic cells (ME28), plasma blasts (ME33), and basophils (ME35), which may not be sufficiently represented in PBMC preparations for evaluation by chemical proteomics. Modules strongly correlated with T-cell subtypes, such as ME4, harbored several proteins with liganded lysines (Figure 5a), and GO annotation further revealed an enriched network of RNA-related cellular functions (Figures 4e and 5a, inset), including a conserved lysine in the interferon-induced RNA-binding proteins IFIT1 (K151), IFIT3 (K148), and IFIT5 (K150) (Extended Data Fig. 9a, b), which suppress viral replication, in part, by binding to viral-specific RNA structures[53]. IFIT1–3 were identified in LPS-treated, but not control PBMCs, consistent with their induced expression by immunostimulatory agents, while IFIT5, which is thought to have broader functions beyond antiviral immunity[54], was quantified in both LPS-treated and control PBMCs (Extended Data Fig. 9c). Previous literature has pointed to the liganded lysine in IFIT1 and IFIT5 playing an important role in binding viral RNA, based on both mutagenesis and structural studies[55], where the lysine appears to directly interact with the 5’-triphosphate (5’-PPP) group of the RNA (Figure 5b and Extended Data Fig. 9d). Considering further that, to our knowledge, chemical probes are lacking for IFITs, we pursued the further characterization of aminophilic compounds that engage the conserved lysine in IFITs.

We first confirmed that probe **P2** labeled recombinantly expressed WT-IFIT5, but not the K155R mutant of this protein (Figure 5c), and that **P2** reactivity with recombinant WT-IFIT5 was blocked by aminophilic compounds with an SAR that generally matched our chemical proteomic data for the endogenous protein (Figure 5c). Using an *in vitro* RNA pulldown assay, we established that recombinant WT-IFIT1 and IFIT5 were selectively pulled down by a biotinylated 5’-PPP-RNA probe, but not a 5’-hydroxyl-RNA (5’-OH-RNA) control probe (Figure 5d and Extended Data Fig. 10a). The yield from pulldown of the corresponding K151R and K150R mutants of IFIT1 and IFIT5 by the 5’-PPP-RNA probe was considerably lower, requiring greater input loading for detection, and comparable in signal to the interactions of these mutant proteins with the 5’-OH-RNA control probe (Figure 5d and Extended Data Fig. 10a). Among the aminophilic ligands, we found that **7a** and **32i** showed strong blockade of WT-IFIT5, but not K151R mutant, binding to the 5’-PPP-RNA probe, while other ligands blocked both WT and mutant protein interactions (Figure 5d and Extended Data Fig. 10a), possibly indicating that they engage additional lysines on IFIT5. Notably, we observed divergent SARs for blockade of IFIT1 and IFIT5 binding to 5’-PPP-RNA, pointing to the potential to create subtype-selective IFIT chemical probes (Extended Data Fig. 10b). Consistent with this premise, using fluorescent probes that label each recombinantly expressed WT-IFIT, but not their corresponding lysine-to-arginine mutants (K151R for IFIT1, K148R for IFIT3, and K150R for IFIT5) (Figure 5e, f), we found that **7a** blocked probe labeling of IFIT5 with an IC_{50} value of ~0.2 μM, while not inhibiting probe labeling of IFIT1 and IFIT3 up to 50 μM (Figure 5e, f, right

panel). Compound **32i** also preferentially blocked fluorescent probe labeling of IFIT5, but cross-reacted with IFIT1 and IFIT3 at higher concentrations (Figure 5e, f, left panel). The potency of inhibition of probe **3** labeling by **7a** was greater than that originally observed for blockade of IFIT5 interactions with 5'-PPP-RNA (Figure 5d and Extended Data Fig. 10b); however, the latter assay contained nonionic detergent, which we surmised might slow the rate of engagement of IFIT5 by **7a**. Consistent with this hypothesis, we found that the potency of **7a** blockade of 5'-PPP-RNA interactions with IFIT5 improved considerably when the pre-incubation time was extended from 1 to 4 hours before performing the 5'-PPP-RNA pulldown (Extended Data Fig. 10c). We next synthesized an alkyne analogue of **7a**, compound **7e** (Figure 5g),for targeted labeling of IFIT5 using CuAAC conjugation to azide reporter tags[37,56]. We found that **7e** labeled WT-IFIT5 expressed in HEK293T cells both *in vitro* (Extended Data Fig. 10d–f) and *in cellulo* (Figure 5g) at concentrations as low as 0.1 μM and showed limited cross-reactivity with other proteins in HEK293T cells below 1 μM (Figure 5g and Extended Data Fig. 10e, f). Negligible labeling was observed for **7e** with the K150R mutant of IFIT5 (Figure 5g and Extended Data Fig. 10e, f). We leveraged probe **7e** to measure a cellular (*in situ*) IC_{50} for **7a** of 1.3 μM (Figure 5h and Extended Data Fig. 10g). We also found that **7a** exhibited good selectivity in cells, where the compound (1 μM, 2 h) engaged few additional lysines beyond K150 of IFIT5 (Extended Data Fig. 10h, i). Taken together, these findings demonstrate that aminophilic compounds targeting a conserved lysine in human IFIT proteins with subtype selectivity can pharmacologically disrupt specific RNA-protein interactions implicated in viral replication and immune response.

Discussion

Several conclusions can be drawn from this large-scale study of the proteomic reactivity of aminophilic compounds that address both the opportunities and challenges facing the development of covalent ligands targeting lysines residues in proteins, First, we note that, despite identifying > 800 liganded lysines, we still consider such events to be rare across the proteome, considering that > 14,000 lysines were quantified in our studies. it is, however, important to qualify that the total lysines quantified herein represent a small fraction of all lysine residues in the human proteome, and it is therefore possible that our ligandability estimates may not reflect the broader potential for aminophilic compounds to engage lysines across the entirety of human proteins. Regardless, we are encouraged by the discovery of liganded lysines in structurally and functionally diverse proteins, including many that lack chemical probes, underscoring the potential of aminophilic compounds to expand the scope of the human proteome that can be targeted by small molecules. Indeed, our follow-up studies verified the ligandability and functionality of lysines not only at traditional druggable locations like enzyme active sites, but also at protein-protein (K351 in XRCC6) and protein-RNA (K150 in IFIT5) binding interfaces. In each case, we observed SARs that point to unique and substantial contributions of both the reactivity and recognition elements of aminophilic compounds. These findings highlight the potential for future optimization of potency and selectivity based on matching ligandable lysines with preferred aminophilic chemotypes and increasing binding affinity through modifications to the recognition elements. We are particularly intrigued by the discovery of conserved, ligandable lysines

involved in RNA binding, as targeting protein-RNA interactions with small molecules has, to date, proven challenging[57]. Considering the high prevalence of lysines at protein-RNA interfaces, where these residues often bind to negatively charged RNA backbone phosphates, we speculate that aminophilic compounds may offer a privileged type of chemical probe for perturbing protein-RNA interactions.

Considering the large number of lysines that preferentially or exclusively interacted with a single aminophilic chemotype, our data emphasize the value of continued exploration of different types of aminophilic compounds to fully assess the ligandability of lysines in the human proteome. Across the chemotypes tested herein, some stood out as potentially attractive starting points for broader library construction and focused chemical probe development. We call attention to both the squarates (**33e-i**) and cyanomethyl acyl sulfonamides (**7a-d**) as showing atypical lysine reactivity profiles that furnished functional compounds targeting protein-protein and protein-RNA interfaces, respectively. A review of within-chemotype SAR further underscored certain features that may enhance lysine reactivity with specific compound classes. We note, for instance, that squarate **33e**, as well as **33b**, showed a broader lysine ligandability profile compared to other squarates, which could reflect the presence of a small, sterically unhindered methoxy leaving group that favors lysine modification by aza-Michael addition, along with a vicinal recognition scaffold bearing electron-withdrawing substituents that further activate the electrophilic β-carbon. Other aminophilic compounds served different purposes. The reversible-covalent dicarboxaldehydes showed broad reactivity with ligandable lysines and were subsequently deployed as scout fragments to efficiently map covalent small molecule-lysine interactions in primary human immune cells under different stimulation states. We anticipate that these dicarboxaldehyde scout fragments will offer versatile tools for the future surveying of lysine ligandability in diverse biological systems. Finally, a recent and complementary study that explored the direct proteomic reactivity of diverse electrophilic groups also evaluated some of the same aminophilic compounds studied herein, providing additional evidence for preferential reactivity with lysine over other proteinaceous amino acids for several chemotypes (activate esters, cyanomethyl acyl sulfonamides, squarates), while for others showing a capacity to react with lysines and additional amino acids (sulfonyl fluorides)[58].

In considering limitations of our studies, as well as future directions, we should note that some aminophilic compound-lysine interactions may be overlooked by our approach of assessing these interactions in native proteomes, followed by confirmation with recombinant proteins (and lysine mutants of these proteins), if, for instance, the interactions require an intact cellular environment or involve proteins that are unstable in cell lysates or not straightforward to recombinantly express in heterologous systems. Future efforts to address these items could include using alternative lysis buffers, as well as establishing protocols for the *in cellulo* profiling of aminophilic compound-lysine interactions. Also, as the recognition element of aminophilic compounds is more extensively elaborated, we may encounter instances where reversible rather than covalent binding blocks lysine reactivity in our chemical proteomic experiments. Indeed, this possibility should even be considered for **17r**, which is a sulfonyl fluoride bearing an ATP pocket-directed recognition element that we found to engage many more active-site lysines in protein kinases than other sulfonyl

fluorides. While we currently assume that the blockade of active-site lysine reactivity by **17r** reflects covalent modification, it is also possible that reversible binding by **17r** could disrupt interactions between probe **1** and kinases. Of course, this outcome would point to another intriguing utility of lysine reactivity profiling, namely, as a way to discover reversible small molecule interactions that competitively disrupt probe **1** labeling of lysines in druggable pockets in proteins. We further acknowledge that the aminophilic ligands discovered herein require improvements in potency and selectivity to furnish advanced chemical probes, and this optimization would benefit from a deeper understanding of the SARs for aminophilic compound-lysine interactions, including measurements of not only their concentration-dependency, but also time-dependency, as well as generating alkyne analogues of hit ligands, which allow for confirmation of direct and site-specific labeling of lysine residues on proteins (as we have shown for K150 in IFIT5) and provide tailored probes for assaying such lysines in more diverse experimental settings. We are encouraged by the initial potency and selectivity observed for interactions such as compound **7a** with K150 of IFIT5, which may provide a path to the first chemical probes to study the contributions of this IFIT to anti-viral immunity and other biological processes. Finally, the conservation of K150 across the broader IFIT family, combined with our initial evidence of divergent SARs for aminophilic compound interactions with K150 and K151 in IFIT5 and IFIT1, respectively, indicates the potential to create covalent probes with subtype selectivity for individual IFITs.

In summary, our in-depth chemical proteomic analysis of structurally diverse aminophilic chemotypes has uncovered many hundreds of ligandable lysines, including those that reside at functional sites on proteins historically considered challenging to target with small molecules. We have also shown how integrating these ligandability maps with human genetic information and cell activation state profiling can further refine our knowledge of lysines for which covalent modification by small molecules is likely to affect the activity of proteins. By defining the aminophilic chemotypes that prefer to react with such ligandable and functional lysines, our study provides attractive starting points for chemical probe development for a diverse array of proteins in the human proteome.

Methods

Cell lines.

All cell lines were purchased from ATCC, tested negative for mycoplasma contamination, and used without further authentication. HEK293T (CRL-3216) and MDA-MB-231 (HTB-26) cells were maintained at 37 °C with 5% CO_2 in DMEM (Corning, 15–013-CV) supplemented with 10% (v/v) fetal bovine serum (FBS, Omega Scientific, FB-11, Lot #441224), penicillin (100 U/mL), streptomycin (100 μg·mL^{-1}) and *L*-glutamine (2 mM). Ramos (CRL-1596) cells were grown at 37 °C in a humidified 5% CO_2 atmosphere in RPMI-1640 medium (Corning, 15–040-CV) supplemented with 10% (v/v) FBS, penicillin (100 U·mL^{-1}), streptomycin (100 μg·mL^{-1}) and *L*-glutamine (2 mM). All cell lines were maintained at a low passage number (≤10 passages).

Isolation of primary human T cells and peripheral blood mononuclear cells (PBMCs).

All studies with primary human cells were performed with samples from human volunteers followed by protocols approved by The Scripps Research Institute Institutional Review Board. Blood from healthy donors (age 18 to 65) was obtained after informed donor consent. Peripheral blood mononuclear cells (PBMCs) were isolated over Lymphoprep (STEMCELL Technologies, 07851) gradient using slightly modified manufacturer's instructions. Briefly, 25 mL of freshly isolated blood was carefully layered on top of 12.5 mL of Lymphoprep in a 50 mL Falcon tube minimizing mixing of blood with Lymphoprep. The tubes were centrifuged (931 g, 20 minutes, 23 °C with brakes off) and the plasma with Lymphoprep layers containing PBMCs were transferred to new 50 mL Falcon tubes and diluted (2:1) with Dulbecco's Phosphate-Buffered Saline (DPBS, VWR, 45000–434). The cells were pelleted (524 g, 8 minutes, 4 °C) and washed with DPBS (20 mL). T cells were isolated by negative selection from freshly isolated PBMCs using EasySep™ Human T Cell Isolation Kit (STEMCELL Technologies, 17951) according to manufacturer's instructions.

Preparation of human cancer cell proteome for gel- and MS-based ABPP analysis.

Cells were grown to 95% confluence for MDA-MB-231 or until cell density reached 2 × 10^6 cells·mL^{-1} for Ramos. Cells were washed and scraped with cold DPBS, and cell pellets were isolated by centrifugation (1,400 g, 3 minutes, 4 °C). Cell pellets were either directly processed or kept frozen at −80 °C until further use. Cell pellets were next lysed using a Branson Sonifier probe sonicator (14 pulses, 30% duty cycle, output setting = 4) and fractionated (100,000 g, 45 min) to yield soluble (supernatant) and membrane (pellet) fractions, which were then adjusted to a final protein concentration of 1.8 mg·mL^{-1} for competitive isoTOP-ABPP experiments. Membrane pellets were resuspended in cold DPBS after separation by sonication. For gel-based ABPP experiments, protein concentration was adjusted to 1.0 mg·mL^{-1} for MBA-MB-231 and Ramos cell lysates, or HEK293T cell lysates expressing target proteins. The lysates were prepared fresh from frozen cell pellets directly before each experiment. Protein concentration was determined using the DC™ Protein Assay (Bio-Rad) and absorbance read using a Tecan, Infinite F500 plate reader following manufacturer's instructions.

Activation of primary human T cells for MS-based ABPP analysis.

Non-tissue culture treated 6-well plates were pre-coated with αCD3 (5 μg·mL^{-1}, BioXCell) and αCD28 antibodies (2 μg·mL^{-1}, BioXCell) in DPBS (2 mL·$well^{-1}$) and kept at 4 °C overnight. The plates were then transferred to an incubator (37 °C in a humidified 5% CO_2 atmosphere) for 1 hour and washed with DPBS (2 × 5 mL·$well^{-1}$). Freshly isolated T cells were resuspended in RPMI-1640 medium supplemented with 10% FBS, penicillin (100 U·mL^{-1}), streptomycin (100 μg·mL^{-1}) and *L*-glutamine (2 mM) at 1 × 10^6 cells·mL^{-1}, plated into pre-coated 6-well plates (8 mL·$well^{-1}$) and kept at 37 °C in a humidified 5% CO_2 atmosphere for 3 days. Activated T cells were then combined into 50 mL Falcon tubes, pelleted (524 g, 8 minutes, 4 °C), washed with DPBS (10 mL), and cell pellets were flash-frozen and stored at –80 °C until in vitro treatments with lysine-reactive electrophiles.

Stimulation of human PBMCs for MS-based ABPP analysis.

Freshly isolated PBMCs were resuspended in RPMI-1640 medium supplemented with 10% FBS, penicillin (100 U·mL^{-1}), streptomycin (100 μg·mL^{-1}) and *L*-glutamine (2 mM) to a cell density of 2 × 10^6 cells·mL^{-1}. PBMCs were then treated with bacterial lipopolysaccharides (100 ng·mL^{-1}, Sigma-Aldrich, L2630, from *Escherichia coli* O111:B4) over a period of 18 hours at 37 °C in a humidified 5% CO_2 atmosphere. Stimulated PBMCs were next combined into 50 mL Falcon tubes, pelleted (524 g, 8 minutes, 4 °C), washed with DPBS (10 mL), and cell pellets were flash-frozen and stored at –80 °C until *in vitro* treatments with lysine-reactive compounds.

In vitro treatment of cell lysates with lysine-reactive compounds.

Lysine-reactive compounds were prepared as either 2, 5, or 10 mM stock solutions in dimethyl sulfoxide (DMSO, Sigma-Aldrich, D8418) and were used at a final concentration of 20, 50, or 100 μM, respectively. For each profiling sample, 500 μL of soluble or membrane proteomes (1.8 mg·mL^{-1}) were treated with 5 μL of the 2-, 5-, or 10-mM fragment stock solutions or 5 μL of DMSO vehicle for 1 hour at 23 °C. Samples were next labelled with 100 μM of lysine-reactive probe **1** (5 μL of 10-mM stock in DMSO) for 1 hour at 23 °C. Samples were then conjugated by copper-mediated azide-alkyne cycloaddition (CuAAC) as described below.

In situ treatment of live cells with lysine-reactive electrophiles.

MDA-MB-231 cells were grown to 95% confluence and Ramos cells were grown to 2 × 10^6 cells·mL^{-1} at the time of treatment. Cells were carefully washed with DPBS and replenished with fresh media containing lysine-reactive compounds at indicated concentrations or DMSO vehicle, maintaining total DMSO content below 0.3%. Cells were then harvested in cold DPBS by scraping, centrifuged (1,400 g, 3 minutes, 4 °C), and cell pellets were washed with cold DPBS (2×). Pellets were either directly processed or kept frozen at −80 °C until further use. Cell pellets were next resuspended in DPBS, lysed by sonication (14 pulses, 30% duty cycle, output setting = 4), and fractionated (100,000 g, 45 minutes) to yield soluble and membrane fractions, which were then adjusted to a final protein concentration of 1.8 mg·mL^{-1}. Fractions were treated with the lysine-reactive probe **1** at a final concentration of 100 μM and incubated for 1 hour at 23 °C. Samples were then conjugated by CuAAC as described below.

CuAAC conjugation.

Following *in vitro* or *in situ* fragment treatment and subsequent probe labelling, samples (500 μL) were conjugated to either the *light* (fragment treated) or *heavy* (DMSO treated) isotopically labelled, TEV-cleavable biotin tags (TEV-tags) using CuAAC reaction. CuAAC reagents were pre-mixed prior to their addition to the proteome samples. TEV tags (*light* or *heavy*, 10 μL of 5 mM stocks in DMSO to a final concentration of 100 μM), tris(benzyltriazolylmethyl)amine ligand (TBTA, 30 μL of 1.7 mM stock in DMSO/^{t}BuOH 1:4 to a final concentration of 100 μM), tris(2-carboxyethyl)phosphine hydrochloride (TCEP, 10 μL of freshly prepared 50 mM stock in H_2O to a final concentration of 1 mM), and $CuSO_4$ (10 μL of 50 mM stock in H_2O at a final concentration of 1 mM) were combined in

an Eppendorf tube, vortexed and added to proteomic samples (55 μL per 500 μL sample). The CuAAC reaction mixture containing *heavy* TEV tag was added to DMSO-treated samples and the CuAAC reaction mixture containing *light* TEV tag was added to fragment-treated samples. The reaction was allowed to proceed at 23 °C for 1 hour, *heavy* and *light* samples were combined pairwise in 15 mL conical Falcon tubes on ice containing 4 mL of MeOH (precooled to –80 °C), 1 mL $CHCl_3$ (precooled to 0 °C), and 1 mL H_2O (precooled to 4 °C). Eppendorf tubes from the reaction mixtures were washed with additional cold H_2O (1 mL each) and washes were added to the same Falcon tube to a final ratio of 4:4:1 (H_2O/MeOH/$CHCl_3$). Following centrifugation (5,000 g, 10 minutes, 4 °C), a protein disk formed at the interface of $CHCl_3$ and MeOH/H_2O layers. Top MeOH/H_2O layer was carefully aspirated without perturbing the disk, and additional MeOH (2 mL, precooled to –80 °C) was added and the suspension was mixed by vortexing. The proteins were pelleted (5,000 g, 10 minutes, 4 °C), and the resulting pellets were solubilized in 1.2% SDS in DPBS (1 mL) with sonication (Branson Sonifier probe sonicator, 10 pulses, 40% duty cycle, output setting = 4) and heating (95 °C, 5 min). The insoluble materials were further removed by an additional centrifugation step (5,000 g, 10 min, 23 °C).

Streptavidin enrichment.

The SDS-solubilized protein mixture (1 mL) was diluted with DPBS (4.5 mL) to a final SDS concentration of 0.2%. The streptavidin-agarose beads (Pierce, 20349; 100 μL slurry per sample) were washed with 10 mL of DPBS (3×) and resuspended in DPBS (0.5 mL per sample) prior to addition. The final mixture was rotated for 3 hours at 23 °C. Following this enrichment step, the beads were pelleted by centrifugation (2,000 g, 2 minutes) and washed to remove nonspecifically bound proteins (2 × 10 mL 0.2% SDS in DPBS, 2 × 10 mL DPBS, and 2 × 10 mL H_2O).

Trypsin and TEV digestion.

The beads were transferred to Eppendorf tubes (2 × 500 μL H_2O), pelleted (2,000 g, 2 minutes), and resuspended in 6 M urea in DPBS (500 μL). To this slurry was added DTT (25 μL of a freshly prepared 200 mM stock in H_2O to a final concentration of 10 mM) and samples were incubated at 65 °C for 15 minutes. Then, iodoacetamide (25 μL of a freshly prepared 400 mM stock in H_2O to a final concentration of 20 mM) was added and samples were incubated at 37 °C with shaking for 30 minutes. The bead mixtures were next diluted with 800 μL DPBS, pelleted by centrifugation (2,000 g, 2 minutes), and washed with 2 M urea in DPBS (1 mL). The samples were resuspended in 2 M urea in DPBS (200 μL) and to this slurry was added sequencing grade trypsin (Promega, 2 μg in 4 μL trypsin resuspension buffer containing 1 mM $CaCl_2$). The samples were allowed to digest overnight at 37 °C with shaking. The beads were pelleted (2,000 g, 2 minutes) and the tryptic digest was aspirated. The beads were then washed with DPBS (3 × 1 mL), H_2O (3 × 1 mL), TEV buffer (500 μL, 50 mM Tris, pH 8, 0.5 mM EDTA, 1 mM DTT), and resuspended in TEV buffer (140 μL). TEV protease (4 μL per sample, 80 μM) was then added and the beads were incubated at 30 °C overnight with rotation. Following the TEV digestion, the beads were pelleted by centrifugation (2,000 g, 2 minutes) and the TEV digest was separated from the beads using Micro Bio-Spin columns (Bio-Rad) with centrifugation (800 g, 30 seconds). The beads were washed with H_2O (100 μL, centrifuged at 16,000 g for 1 minute) and the

eluents (300 μL) were acidified by the addition of formic acid (0.1%, 15 μL per sample to a final concentration of 5% *v/v*) and stored at –80 °C prior to analysis.

Liquid-chromatography-mass-spectrometry (LC-MS/MS) analysis.

TEV digested samples were pressure-loaded onto a 250 μm (inner diameter) fused silica capillary columns packed with C18 resin (Aqua 5 μm, Phenomenex) and analyzed by multidimensional liquid chromatography tandem mass spectrometry (MudPIT) using an LTQ-Velos Orbitrap mass spectrometer (Thermo Scientific) coupled to an Agilent 1200-series quaternary pump. The peptides were eluted onto a biphasic column with a 5 μm tip (100 μm fused silica, packed with 10 cm of C18 resin and 4 cm of bulk strong cation exchange resin (SCX, Phenomenex) in a 5-step MudPIT experiment, using 0%, 30%, 60%, 90%, and 100% salt 'bumps' of 500 mM aqueous ammonium acetate and 5%→100% gradient of buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B: 5% water, 95% acetonitrile, 0.1% formic acid) as previously described (Weerapana et al., 2007). Acquired data was collected in a data-dependent acquisition mode with dynamic exclusion enabled (20 s, repeat count of 2). One full MS (MS1) scan (400–1800 m/z) was followed by 30 MS2 scans (ITMS) of the nth most abundant ions.

Peptide identification and quantification.

From each of the five .raw files (one for each salt 'bump') generated by the instrument (Xcalibur software), the MS2 spectra for all fragmented parent ions were extracted from the raw file using RAW Converter (version 1.1.0.22; available at http://fields.scripps.edu/rawconv/). The generated MS2 spectral files (.ms2 files) were uploaded and searched using the ProLuCID algorithm (available at http://fields.scripps.edu/downloads.php) using a reverse concatenated, non-redundant (gene-centric) variant of the Human UniProt database (release-2012_11). Cysteine residues were searched with a static modification for *S*-carboxyamidomethylation (+57.02146). For all competitive and reactivity profiling experiments, lysine residues were searched with up to one differential modification for either the light or heavy TEV tags (+464.24957 or +470.26338, respectively). Peptides were required to have at least one tryptic terminus and to contain the TEV modification. ProLuCID data was filtered through DTASelect (version 2.0) to achieve a peptide false-positive rate below 1%.

R-value calculation and data processing.

The ratios of heavy (DMSO) / light (fragment treated) MS1 peaks (*R*-values) for each unique peptide were quantified with in-house CIMAGE software[7] using default parameters (3 MS1 acquisitions per peak and signal to noise threshold set to 2.5). Site-specific engagement of lysine residues was assessed by blockade of pentynoic acid STP ester probe **1** (Lumiprobe) labeling. A maximal ratio of 20 was assigned for peptides that showed a ≥ 95% reduction in MS1 peak area in the fragment-treated proteome (light TEV tag) compared to the DMSO-treated (control) proteome (heavy TEV tag). Ratios for unique peptide sequences were calculated for each experiment; overlapping peptides with the same modified lysine (e.g., different charge states, chromatographic elution times or tryptic termini) were grouped together and the median ratio was reported as the final ratio (*R*). Additionally, ratios for

peptide sequences containing multiple lysines were grouped together. When aggregating data across experimental replicates, the mean of each experimental median *R* was reported. The peptide ratios reported by CIMAGE were further filtered to ensure the removal or correction of low-quality ratios in each individual dataset. The quality filters applied were the following: (i) removal of peptides with co-elution correlation score R^2 values of ≤ 0.8, (ii) removal of reverse peptide sequences, (iii) removal of half-tryptic peptides, (iv) removal of peptide sequences with tryptic-site modified lysines (e.g., K.K*, R.K*, K*.K, and K*.R), (v) removal of peptides with $R = 20$ and only a single MS2 event triggered during the elution of the parent ion, and (vi) removal of peptides with $R = 20$ and coefficient of variation is ≥ 0.6. For peptide ratios with standard deviations of ≥ 90% from the median, the lowest ratio was taken instead of the median. For each biological replicate, the reported ratio of a given peptide is the median ratio. Across biological replicates for a single fragment: (i) peptides with $R = 20$ are only reported if peptides are quantified and liganded ($R \geq 4 < 20$) in at least one other dataset across all datasets, (ii) peptides with $R \geq 4 < 20$ are reported if peptides are quantified (but not necessarily liganded) in at least one other dataset across all datasets. The remaining peptides with $R = 20$ were manually annotated. Where fragments are aggregated, the reported ratio for a given peptide is the median ratio across the biological replicates. Where chemotypes are aggregated, the reported ratio is the maximum ratio of the constituent fragments.

Recombinant expression of proteins by transient transfection.

HEK293T cells were grown to 60% confluency under standard growth conditions in 10-cm tissue culture dishes. To a 5 μg of DNA diluted in 250 μL of serum-free DMEM was added 15 μL of aqueous polyethyleneimine ‘MAX’ (1 $mg \cdot mL^{-1}$, MW 40,000, PEI, Polysciences, Inc.). ‘Mock’ transfected HEK293T cells were transfected with an empty pRK5 vector. The mixture was incubated at room temperature for 20 minutes and added dropwise to the cells. Cells were grown for 48 h at 37 °C in a humidified 5% CO_2 atmosphere. Cells were then harvested in cold DPBS by scraping, centrifuged (1,400 g, 3 minutes, 4 °C), and cell pellets were washed with cold DPBS (2×). Pellets were either directly processed or kept frozen at −80 °C until further use. Cell pellets were next lysed by sonication (6 pulses, 30% duty cycle, output setting = 4) and fractionated (100,000 g, 45 minutes) to yield soluble and membrane fractions, which were then adjusted to a final protein concentration of 1.0 $mg \cdot mL^{-1}$.

Subcloning and site-directed mutagenesis.

Full-length genes encoding proteins of interest were PCR-amplified from a cDNA library derived from low-passage HEK293T cells using the Ribozol RNA extraction reagent (Amresco) and the iScript Reverse Transcription Supermix kit (Bio-Rad). For the following proteins, cDNA clones were used for PCR-amplification: CPOX (OHu18833, GenScript), SIN3B (OHu28835, GenScript), IFIT3 (OHu10416, GenScript), and RIDA (OHu25061, GenScript). Gene products were subcloned into the pRK5 vector with a C-terminal FLAG tag using SalI (N-terminal) and NotI (C-terminal) restriction sites. DNA was amplified with custom forward and reverse primers (Table S1) using Phusion Polymerase (NEB, M0530S), following the manufacturers’ instructions, digested with the indicated restriction enzyme and ligated into pRK5 vector with the appropriate affinity tag. Lysine mutants

were generated using QuikChange site-directed mutagenesis using Phusion® High-Fidelity DNA Polymerase and custom primers containing the desired mutations and their respective complements (Table S2). All clone sequences were verified (Eton Bioscience).

Western blot analysis.

Cells were collected and lysed in 1% NP-40 lysis buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 10% glycerol, 1% Nonidet P-40) with a complete protease inhibitor cocktail (Roche). Cells were vortexed and sonicated (6 pulses, 30% duty cycle, output setting = 4), and the supernatant was collected after centrifugation (16,000 g, 10 minutes, 4 °C). Protein concentration was determined by detergent compatible assay (5000112, Bio-Rad). Protein lysate was heated at 95 °C for 5 minutes in Laemmli sample buffer (1×). Proteins were resolved by 12% or 14% Novex Tris-Glycine mini gels (Invitrogen) and transferred to 0.45 μm nitrocellulose membrane (GE Healthcare). The membrane was blocked with 5% milk in Tris-buffered saline (20 mM Tris-HCl, pH 7.6, 150 mM NaCl) with Tween (TBST) buffer (0.1% Tween 20, 20 mM Tris-HCl, pH 7.6, 150 mM NaCl) for 1 hour at 23 °C with gentle rocking. The primary antibody (anti-FLAG) was diluted (1:5000) with 5% milk in TBST buffer and incubated with the membrane for 1 hour at 23 °C or overnight at 4 °C with gentle rocking. The membrane was washed with TBST buffer (3×, 5 minutes) and incubated with the secondary antibody (1:5000 dilution in 5% milk in TBST buffer) for 1 hour at 23 °C with gentle rocking. The membrane was washed with TBST buffer (3×, 5 minutes) and western blots were visualized on a LICOR Odyssey scanner. Relative band intensities were quantified using ImageJ software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997–2018).

RIDA deiminase activity assay.

Soluble proteome (100 μL, 1.0 mg·mL^{-1}) from HEK293T cells expressing human RIDA (WT or K117R, K117Q, K117E, K117I mutants) or mock transfected cells (empty vector; negative control) were prepared in 50 mM potassium pyrophosphate (pH 8.5) assay buffer and added into a clear bottom 96-well plate. For compound treatments, 1.0 μL of the lysine-reactive compound (in DMSO) or 1.0 μL of DMSO (positive control) were added and the reactions were incubated for 1 h at 23 °C. A mixture containing 10 μL of semicarbazide·HCl (100 mM in assay buffer, Sigma-Aldrich, S2201), 10 μL of catalase from bovine liver (10 μg in assay buffer, Sigma-Aldrich, C9322), and 10 μL of *L*-amino acid oxidase from *Crotalus adamanteus* (10 μg in assay buffer, Sigma-Aldrich, A9253) was added to each well and the reaction was started by addition of 10 μL of *L*-methionine (2 mM in assay buffer). The absorbance of the semicarbazone formation was measured at 248 nm every minute for 20 minutes at 23 °C.

Ku70/Ku80 heterodimerization assay.

HEK293T cell lysates expressing human FLAG-tagged Ku70 (WT or K351R mutant) were lysed by sonication (5 pulses, 40% duty cycle, output setting = 4) in 1% NP-40 lysis buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 10% glycerol, 1% Nonidet P-40) containing complete protease inhibitor cocktail (Roche). Samples were rotated for 30 minutes at 4 °C to complete lysis, clarified by centrifugation (16,000 g, 10 minutes, 4 °C), and protein concentration was measured using the DC™ Protein Assay (Bio-Rad)

and normalized to 1.0 mg·mL^{-1}. Normalized lysates of cells expressing human HA-tagged Ku70 (WT or K351R mutant) were treated with lysine-reactive compounds or DMSO (control) at indicated concentrations (1 h, 23 °C) and then mixed with lysates expressing WT Ku80 protein (1.0 mg·mL^{-1} in 1% NP-40 buffer) for 1 hour at 23 °C. Samples were then co-immunoprecipitated with ANTI-FLAG® M2 affinity gel (20 μL slurry per sample, Sigma-Aldrich, A2220) by rotation (1 h, 4 °C), washed with 1.0 mL of 0.2% NP-40 washing buffer (4×, 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.2% Nonidet P-40), and heated at 95 °C for 10 minutes in Laemmli sample buffer (2×), followed by western blot analysis with anti-HA immunoblotting.

RNA probe synthesis for IFIT pulldown assay.

3'-Biotinylated 5'-triphosphate (PPP) and 5'-hydroxy (OH) RNA probes were synthesized by *in vitro* transcription using MEGAscript T7 Transcription Kit (Invitrogen, AM1334), according to product guidelines. A 300-nt dsDNA oligo, containing a T7 promoter sequence was used as a template, and biotin-16-UTP (Roche, 11388908910) was incorporated into the reaction mixture at a 1:5 biotin-UTP:UTP ratio. Residual DNA was digested with Turbo DNase and biotinylated probes were subsequently purified using an RNeasy Mini Kit (Qiagen, #74104). 5'-OH-RNA probes were prepared by dephosphorylation of 5'-PPP RNA probes using calf intestinal alkaline phosphatase (CIP) (NEB, M0290) and then purified using an RNeasy Mini Kit). "Mock" dephosphorylated 5'-PPP-RNA probes were prepared alongside 5'-OH-RNA, replacing CIP with water, and were shown to bind comparably to IFITs as untreated 5'-PPP-RNA probes.

IFIT pulldown assay.

Affinity enrichment resins were prepared by coupling biotinylated RNA probes to streptavidin resin (1 μg of RNA per 50 μL of agarose slurry). Streptavidin agarose was initially washed with TAP buffer (3×, 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol, 0.2% Nonidet-P40, 1.5 mM $MgCl_2$) and then incubated with biotinylated probes for 1 hour at 4 °C. Unbound probe was removed by centrifugation, and coupled resin was washed with TAP buffer (1×), prior to dilution for addition to cell lysates. HEK293T cells expressing FLAG-tagged IFIT1/5 (WT or K151R/K150R mutants) were resuspended in TAP buffer, containing complete EDTA-free protease inhibitor tablets (Roche, 04693159001) and allowed to lyse on ice for 15 minutes. Lysates were briefly sonicated (5 pulses, 40% duty cycle, output setting = 4), and then cleared by centrifugation (16,000 g, 5 minutes, 4 °C). The soluble proteome from IFIT5 (WT or K150R mutant) was normalized to 0.25 mg·mL^{-1}, or 1.0 mg·mL for IFIT1 (WT or K151R mutant) and treated with lysine-reactive compounds (5 μL) or DMSO (5 μL, control) at indicated concentrations for 1 hour at 23 °C. Pulldown assays were carried out by rotating samples with 75 μL of IFIT1 (or 50 μL of IFIT5) of RNA-coupled streptavidin for 2 hours at 4 °C. Resins were washed with TAP buffer (4×, 1.0 mL), and bound proteins were eluted with 2× SDS-PAGE sample buffer. Samples were heated (10 minutes, 95 °C) and resolved by gel electrophoresis on Novex 10% Tris-glycine precast gels (Invitrogen), followed by western blot analysis.

LPCAT1 acyltransferase assay.

HEK293T cells expressing human LPCAT1 (WT or K221R) were resuspended in assay buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA, 150 mM NaCl), and lysed by sonication using a probe sonicator (15 pulses, 30% duty cycle, output setting = 3). The lysate was centrifuged (16,000 g, 45 minutes, 4 °C) to collect the membrane fraction. The membrane pellet was then resuspended in assay buffer by sonication (5 pulses, 30% duty cycle, output setting = 3) and diluted to 0.05 mg·mL^{-1}. For acyltransferase assay, 100 μL of 0.05 mg·mL^{-1} lysate was treated with lysine-reactive compounds at indicated concentrations (1 h, 23 °C). After the incubation, 10 μL of 11× substrate cocktail (550 μM 15:0 lyso-PC and 550 μM 10:0 CoA in assay buffer, Avanti Polar Lipids) was added to the sample and incubated for 10 minutes at 23 °C. The reaction was quenched by adding 300 μL of $CHCl_3$/MeOH (2:1, v/v) containing 1 nmol of PC (12:0/12:0, Avanti Polar Lipids) as an internal standard. The suspension was vortexed vigorously and centrifuged (2,000 g, 5 minutes, 4 °C). The bottom layer (150 μL) was collected and mixed with 75 μL of MeOH, and 2.5 μL of the extract was used for MS analysis to measure the production of PC (15:0/10:0). The amount of PC(15:0/10:0) was quantified using an LC/MS-based multiple reaction monitoring (MRM) method in positive mode (Agilent Technologies 6460 Triple Quad). MS analysis was performed using ESI with the following parameters: drying gas temperature, 350 °C; drying gas flow, 9 L/min; nebulizer pressure, 45 Ψ, sheath gas temperature, 375 °C; sheath gas flow, 10 L/min; fragmentor voltage, 100 V; capillary voltage, 3.5 kV. Ammonium acetate (20 mM in dH_2O) and ammonium acetate (20 mM in MeOH) were used as buffer A and B, respectively. The LC gradient was the following after injection: start from 90% B at 0.8 mL/min, increase to 99% B at 0.8 mL/min for 5 min, stay at 99% B at 0.8 mL/min for 1 min, return to 90% B at 0.8 mL/min, then equilibrate for 1.5 min. The MRM transitions for PC (15:0/10:0) and PC (12:0/12:0) were 636.5→184.1 and 622.4→184.1, respectively. The amount of PC (15:0/10:0) was quantified by measuring areas under the curve in comparison to the corresponding PC (12:0/12:0) curve. The hydrolysis activity of LPCAT1 (WT or K221R mutant) was calculated by normalizing to the amount of PC (15:0/10:0) produced against the proteome amount and the incubation time.

LPCAT1 ubiquitination assay.

HA-tagged ubiquitin (2 μg) and FLAG-tagged LPCAT1 (2 μg, WT or K221R mutant) or an empty FLAG-tagged pRK5 vector (2 μg, control) or FLAG-tagged GFP (2 μg, control) were co-expressed in HEK293T cells prior to treatment with or without proteasome inhibitor MG132 (10 μM, SelleckChem) for 2 or 14 hours (37 °C, 5% CO_2). Cells were then collected and lysed by sonication (5 pulses, 40% duty cycle, output setting = 4) in 1% NP-40 lysis buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 10% glycerol, 1% Nonidet P-40) containing complete protease inhibitor cocktail (Roche). Samples were rotated for 30 minutes at 4 °C to complete lysis, clarified by centrifugation (16,000 g, 10 minutes, 4 °C), and protein concentration was measured using the DC™ Protein Assay (Bio-Rad) and normalized to 1.0 mg·mL^{-1}. Samples were then co-immunoprecipitated with ANTI-FLAG® M2 affinity gel (20 μL slurry per sample, Sigma-Aldrich, A2220) by rotation (1 h, 4 °C), washed with 1.0 mL of 0.2% NP-40 washing buffer (4×, 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.2% Nonidet P-40), and heated at 95 °C for 10 minutes in Laemmli sample buffer (2×), followed by western blot analysis with anti-HA immunoblotting. For the endogenous ubiquitination

of LPCAT1, FLAG-tagged LPCAT1 (5 μg, WT or K221R mutant) were overexpressed in HEK293T cells prior to treatment with MG132 (10 μM, 2 hours). Cell lysates were subjected to anti-FLAG immunoprecipitation as described above, and the affinity-enriched precipitates were analyzed by anti-Ubiquitin immunoblotting.

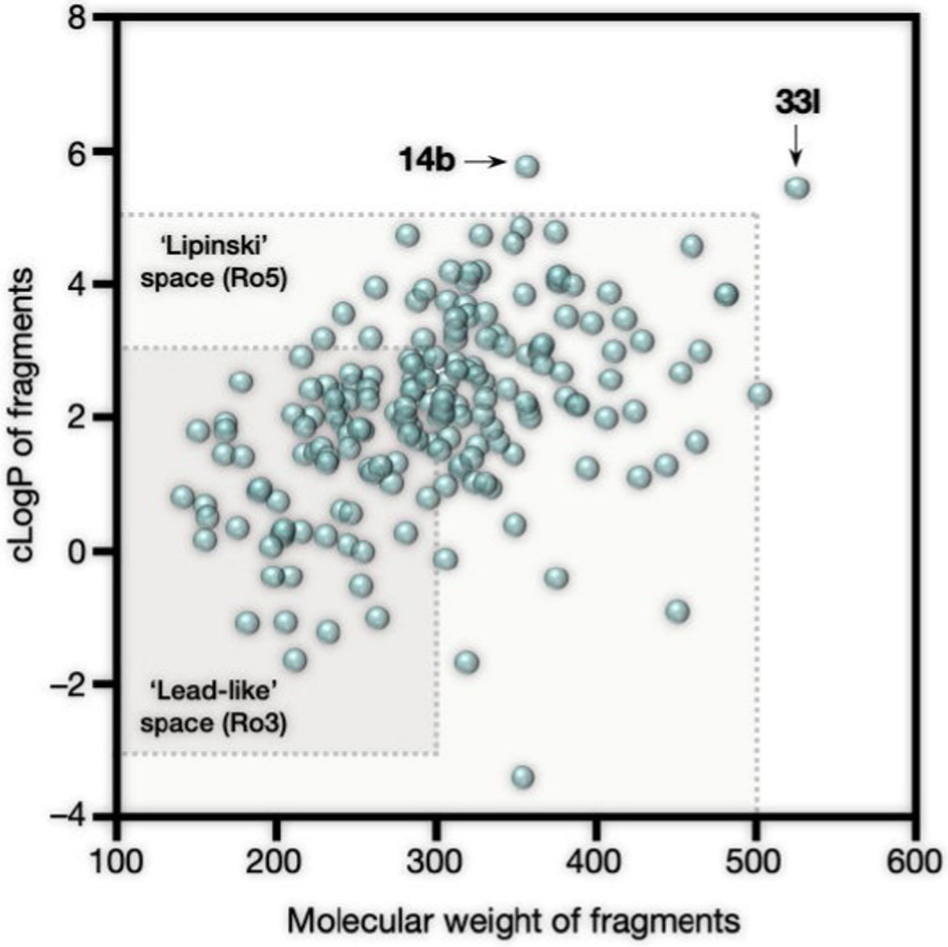
Calculation of relative activity or percent inhibition.

For RIDA, the slope of the linear regression of the linear portion of the absorbance over time was used as the measure of activity. Apparent activity was calculated relative to the WT. Percent inhibition was calculated relative to the positive and negative control and used to calculate IC_{50} values by nonlinear regression analysis from a dose-response curve generated using GraphPad Prism 7. For quantification of inhibition and apparent IC_{50} determination in competitive gel-based ABPP experiments, the percentage of labeling was determined by quantifying the integrated optical intensity of the bands using ImageLab 5.2.1 software (Bio-Rad). Nonlinear regression analysis was used to determine the IC_{50} values from a dose-response curve generated using GraphPad Prism 7.

Statistical analysis

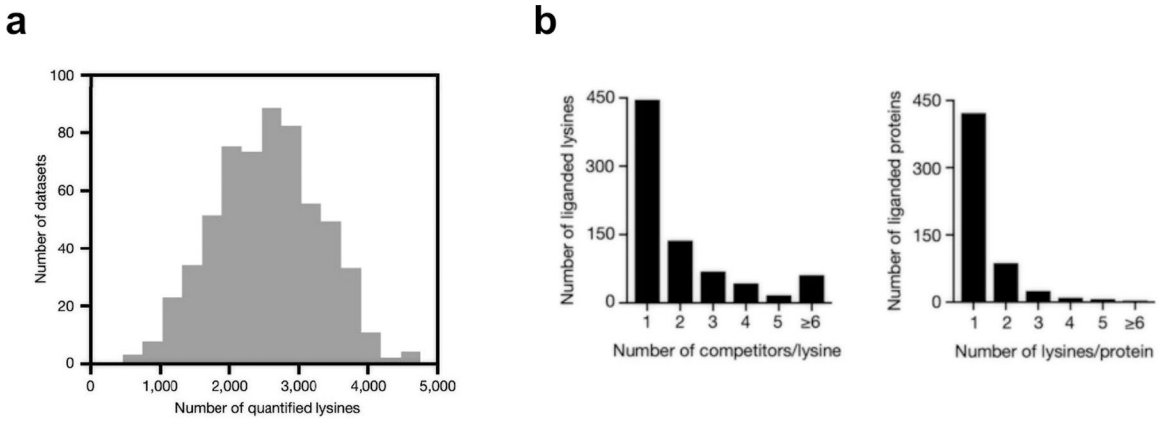
Unless otherwise stated, quantitative data are expressed in bar and line graphs with mean ± s.d. (standard deviation, shown as an error bar) shown. Differences between two groups were examined using an unpaired two-tailed Student’s t-test with equal or unequal variance as noted. Significant P values were indicated (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and ****$P < 0.0001$).

Extended Data



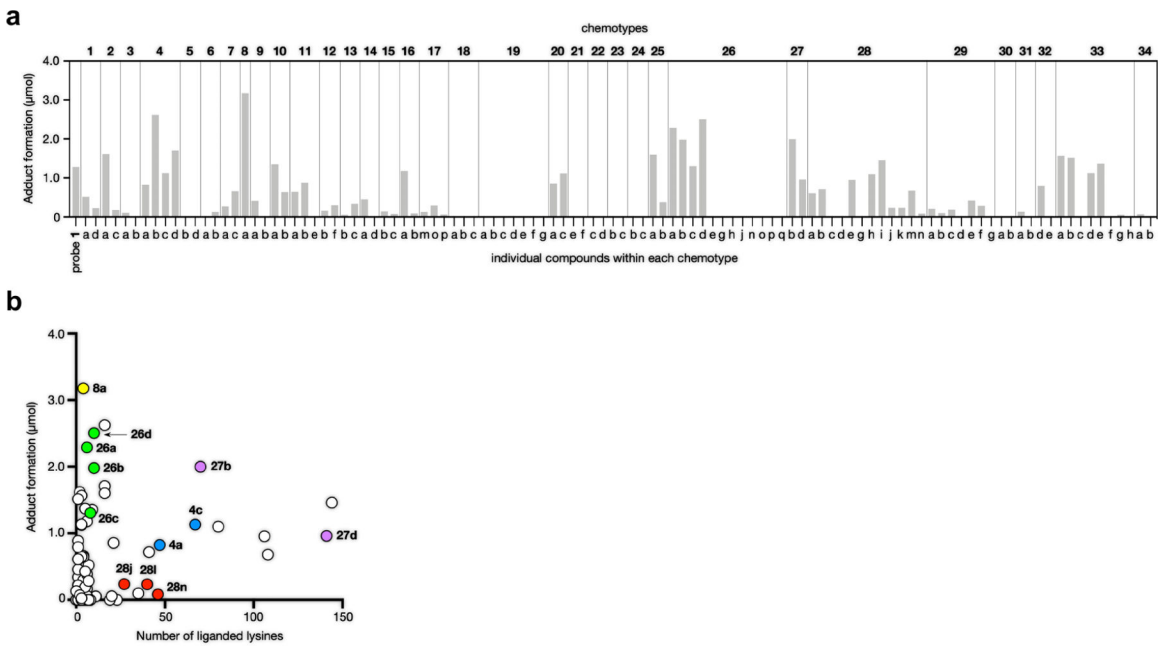
Extended Data Fig. 1. Properties of aminophilic compound library for mapping small molecule-lysine interactions in the proteome.

cLogP versus molecular weight plot showing aminophilic compounds that follow Lipinski's drug-likeness "rule-of-five" (Lipinski space Ro5, light gray box)[1] and Congreve's fragment-based lead-likeness "rule-of-three" (Lead-like space Ro3, dark gray box)[2,3].



Extended Data Fig. 2. Features of aminophilic compound-lysine interaction map in human cancer cell proteomes.

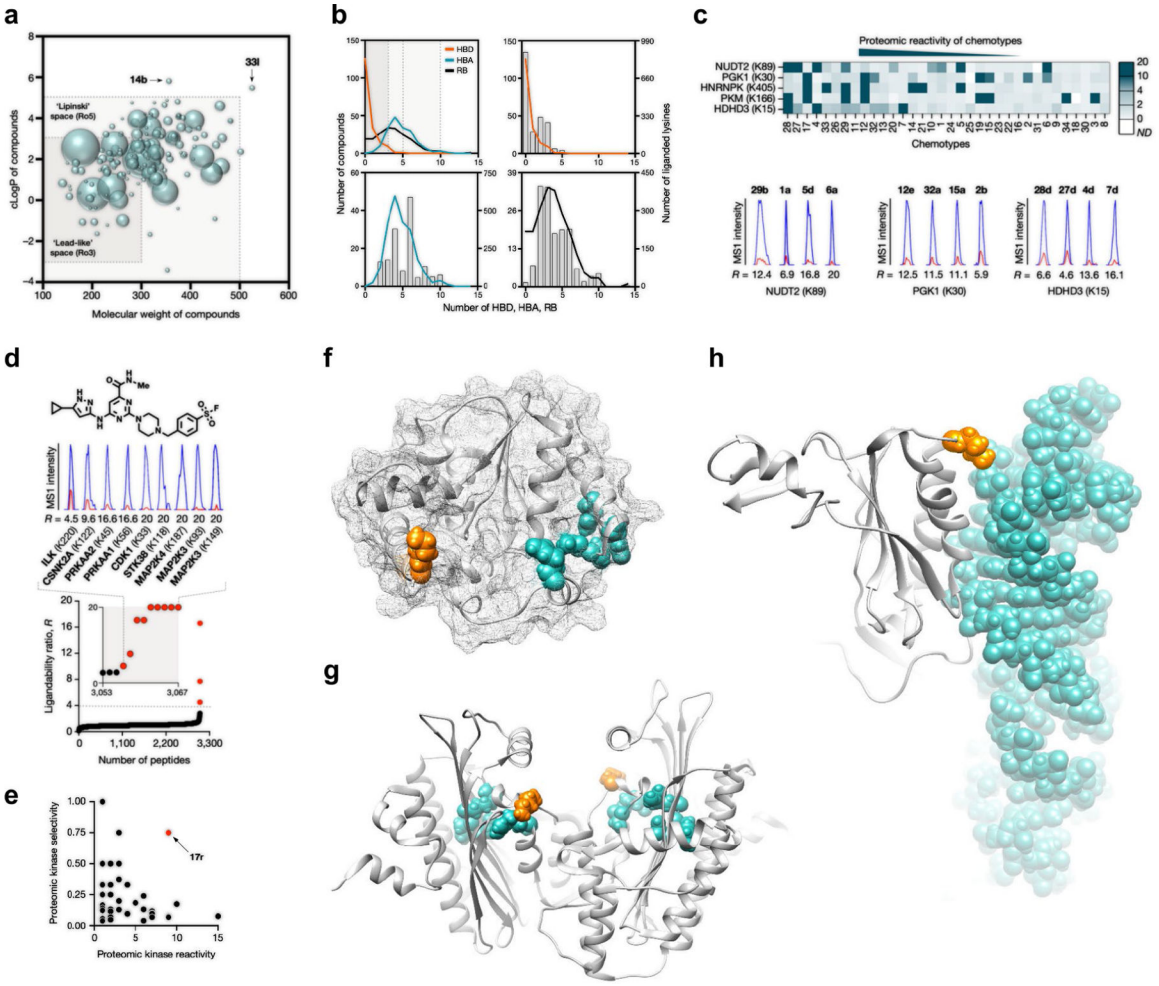
a, Histogram showing number of quantified lysines across all isoTOP-ABPP datasets. **b**, Number of aminophilic compound hits per liganded lysine (left) and the number liganded lysines per protein (right). The results shown are average ratios from three experiments (n = 3 biologically independent experiments).



Extended Data Fig. 3. Reactivity profiles of representative aminophilic compounds with a model amine nucleophile.

a, Aminophilic compounds (125 μM) were incubated at room temperature with the amine nucleophile Nα-acetyl-*L*-lysine-OMe (2 M, 1 h, at pH 10 (0.05 M $NaHCO_3$)). All samples contained 5 μM Nα-acetyl-L-methionine-OH as an internal standard. Samples were neutralized with formic acid and 20 μL of the resulting solution was inject on to an Agilent 6100 series single quadrupole LC/MS system. Samples were run with the following gradient of Buffer A (95/5 Water/MeCN with 0.1% formic acid) and Buffer B (5/95 Water/MeCN with 0.1% formic acid): 100% A from 0–1 min, 100% A → 100% B from 1–11 min,

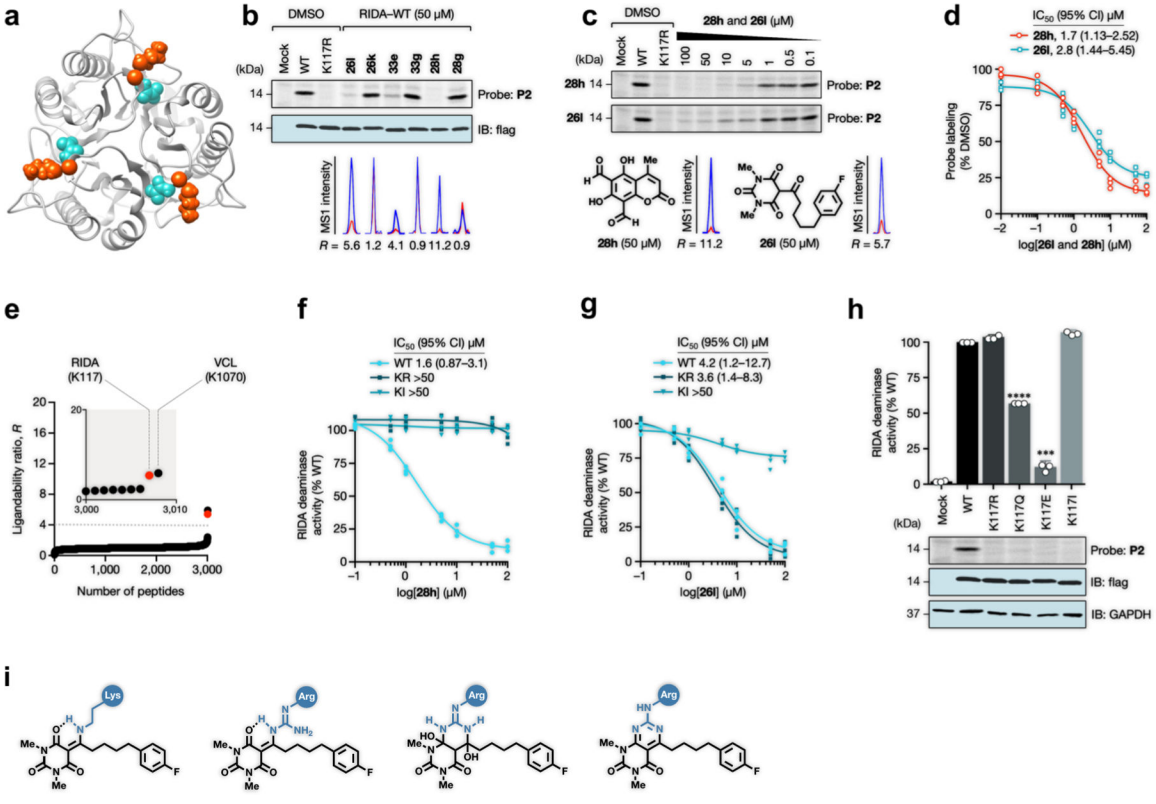
100% B from 11–13 min, and 100% A from 13–15 min. Peaks corresponding to the amine nucleophile adducts were quantified using Agilent Open Lab software. **b**, Correlation plot comparing amine nucleophile adduct formation to liganded lysines for each compound (also see Supplementary Table 1). Representative aminophilic compound chemotypes are color-coded. For **a** and **b**, data represent average values ± SD; n = 2 per group (n = 2 independent experiments).



Extended Data Fig. 4. Relating aminophilic compound-lysine interaction map to compound properties and representative features of liganded lysines.

a, cLogP versus molecular weight plot showing aminophilic compounds that follow Lipinski's "rule of five" (Lipinski space Ro5) and lead-likeness "rule of three" (Lead-like space Ro3). The size of each bubble represents the number of liganded lysines per compound. **b**, Distribution of compounds (top, left) by the number of hydrogen-bond donors (HBDs, orange line), hydrogen-bond acceptors (HBAs, blue line) and rotatable bonds (RBs, black line).[3] Correlation between the compound distribution and the number of liganded lysine interactions (gray bars, right y-axis) as relates to the number of HBDs (top, right), HBAs (bottom, left) and RBs (bottom, right). **c**, Heatmap (top) and extracted MS1 chromatograms (bottom) of representative liganded lysines that show broad reactivity with

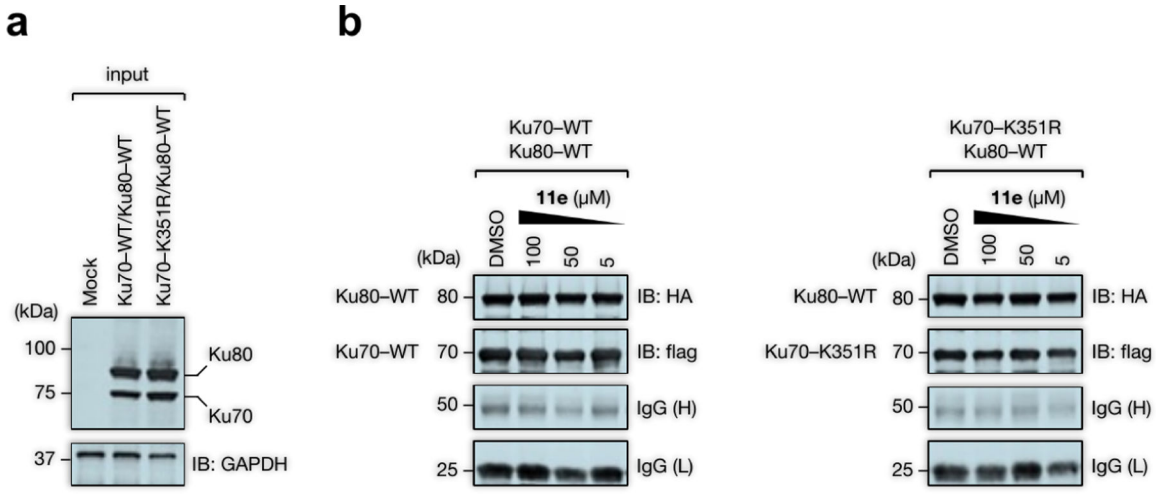
aminophilic compounds (also see Supplementary Dataset 3). **d**, isoTOP-ABPP ratio plot for the sulfonyl fluoride **17r** containing a kinase-directed recognition element. Red points represent liganded active-site lysines in kinases and their corresponding extracted MS1 chromatograms. The dashed line marks the *R* value of 4 used to define a lysine liganding event (also see Supplementary Dataset 3). **e**, Comparison of reactivity of aminophilic compounds toward kinase lysines as a function of selectivity toward kinase lysines across the proteome (right panel). The kinase reactivity of individual compounds was defined by the total number of liganded kinase lysines. The selectivity of individual compounds toward kinase lysines was defined by the fraction of liganded kinase to non-kinase lysines. **f-h**, Location of liganded lysines that are also missense mutated in human disease (orange) in protein crystal structures (gray) of PMVK (K69) (**f**, PDB ID: 3CH4), CPOX (K404) (**g**, PDB ID: 2AEX), and RPL10 (K78) (**h**, PDB ID: 6OLE). Also shown highlighted in blue are active site residues or protein-RNA interaction regions of the proteins where the indicated lysines reside. Note the proximity of K404 in CPOX and K78 in RPL10 to the active site and RNA-interaction region of these proteins, respectively. K69 of PMVK is distant from the active site of the enzyme, but the missense mutation of this lysine causes substantial catalytic defects[4,5], pointing to an allosteric regulatory function.



Extended Data Fig. 5. Functional impact of aminophilic compound-lysine interactions for representative proteins.

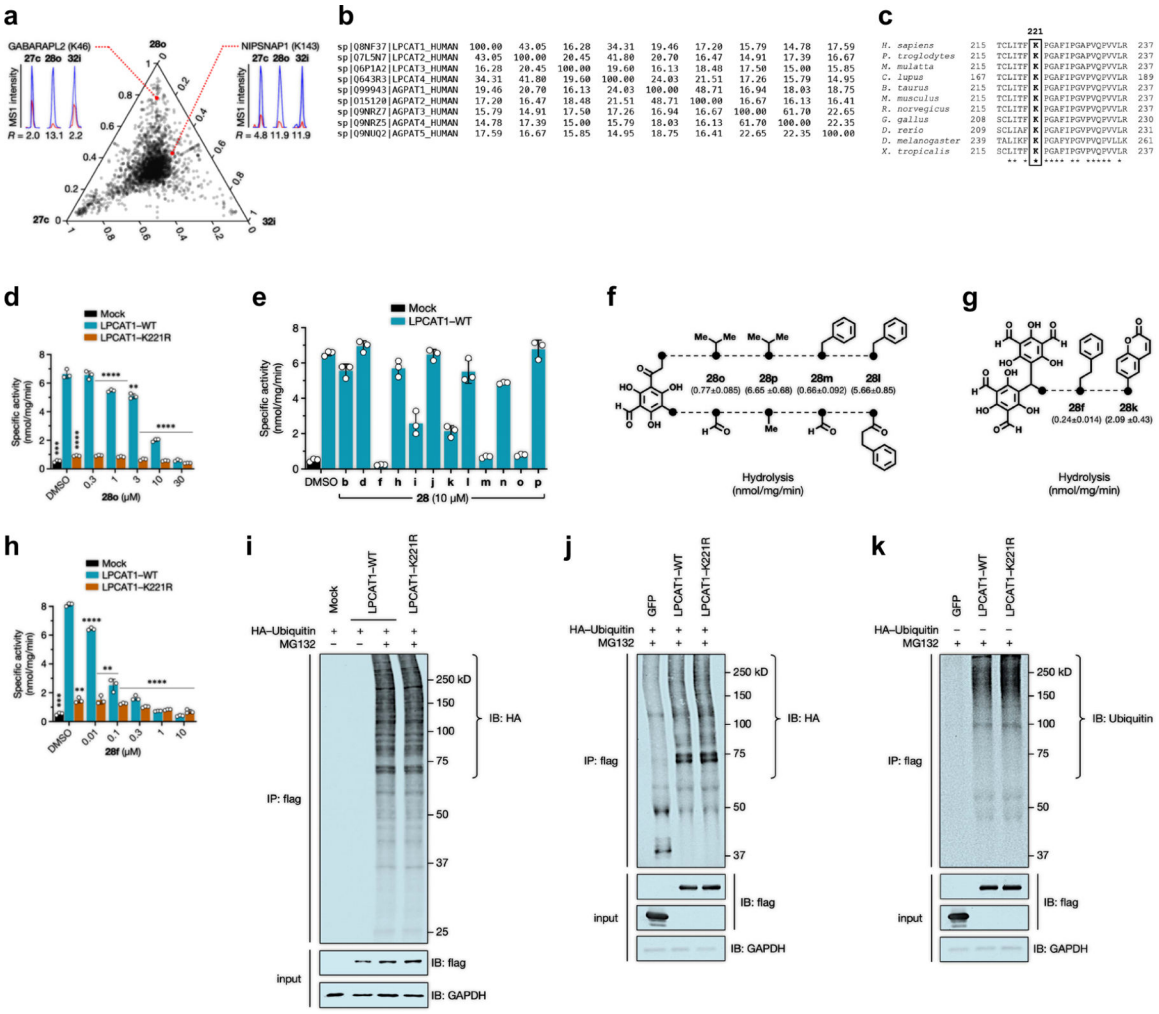
a, The location of liganded lysine K117 (orange) in the RIDA crystal structure (gray, PDB ID: 1ONI). Also shown is bound pyruvate (teal) in each of the three active sites at the interfaces of adjacent monomers. **b,** SAR for aminophilic compound engagement

of K117 in RIDA, as determined by competitive isoTOP-ABPP is recapitulated by gel-ABPP of recombinant protein (also see Supplementary Datasets 3 and 4). Top, HEK293T cells recombinantly expressing WT-RIDA and the corresponding K117R mutant as Flag epitope-tagged proteins were treated with the indicated aminophilic compounds (50 μM, 1 h) followed by treatment with probe **P2** and analyzed by gel-ABPP (top panel) and western blotting (bottom panel). Bottom, Extracted MS1 chromatograms depicting *R* values for the indicated aminophilic compound-RIDA-K117 interactions mapped by competitive isoTOP-ABPP (also see Supplementary Dataset 3). **c**, Top, gel-ABPP data showing concentration-dependent blockade of **P2** labeling of recombinantly expressed WT-RIDA by **28h** and **26l** in HEK293T cell lysates. Bottom, structures of **28h** and **26l** with extracted MS1 chromatograms depicting *R* values for their respective engagement of K117 or RIDA determined by competitive isoTOP-ABPP (also see Supplementary Dataset 3). **d**, Corresponding fitted IC_{50} curves for blockade of probe **2** labeling of WT-RIDA. Data represent average values ± SD; n = 3 per group. CI, confidence interval. **e**, Representative isoTOP-ABPP ratio plot showing proteome-wide lysine reactivity profile for **26l** (50 μM). Among ~3,000 quantified lysines, only two - K117 of RIDA and K1070 of VCL - were liganded. The dashed line marks the *R* value of 4 used to define a liganded lysine event (also see Supplementary Dataset 3). **f, g**, Fitted IC_{50} curves for the concentration-dependent inhibition of the deaminase activity of recombinantly expressed WT- and K117R and K117I mutants of RIDA in HEK293T cell lysates by **28h** (**f**) and **26l** (**g**). Data represent average values ± SD; n = 3 per group. CI, confidence interval. **h**, Catalytic activity (upper panel) and gel-ABPP analysis of **P2** labeling (lower panel) of WT- and indicated K117 mutants. **i**, Presumed reversible-covalent and irreversible adducts formed between **26l** with K117 and R117.[6] Data represent average values ± SD; n = 3 per group. P values were 0.00081 and 0.000066. For western blot and gel-ABPP data in **b**, **c**, and **h**, experiments were conducted three times (n = 3 biologically independent experiments) with similar results. Statistical significance was calculated for changes >25% in magnitude in comparison to DMSO-treated samples with unpaired two-tailed Student's t-tests: *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.



Extended Data Fig. 6. Compound 11e does not block preformed Ku70-Ku80 complex.

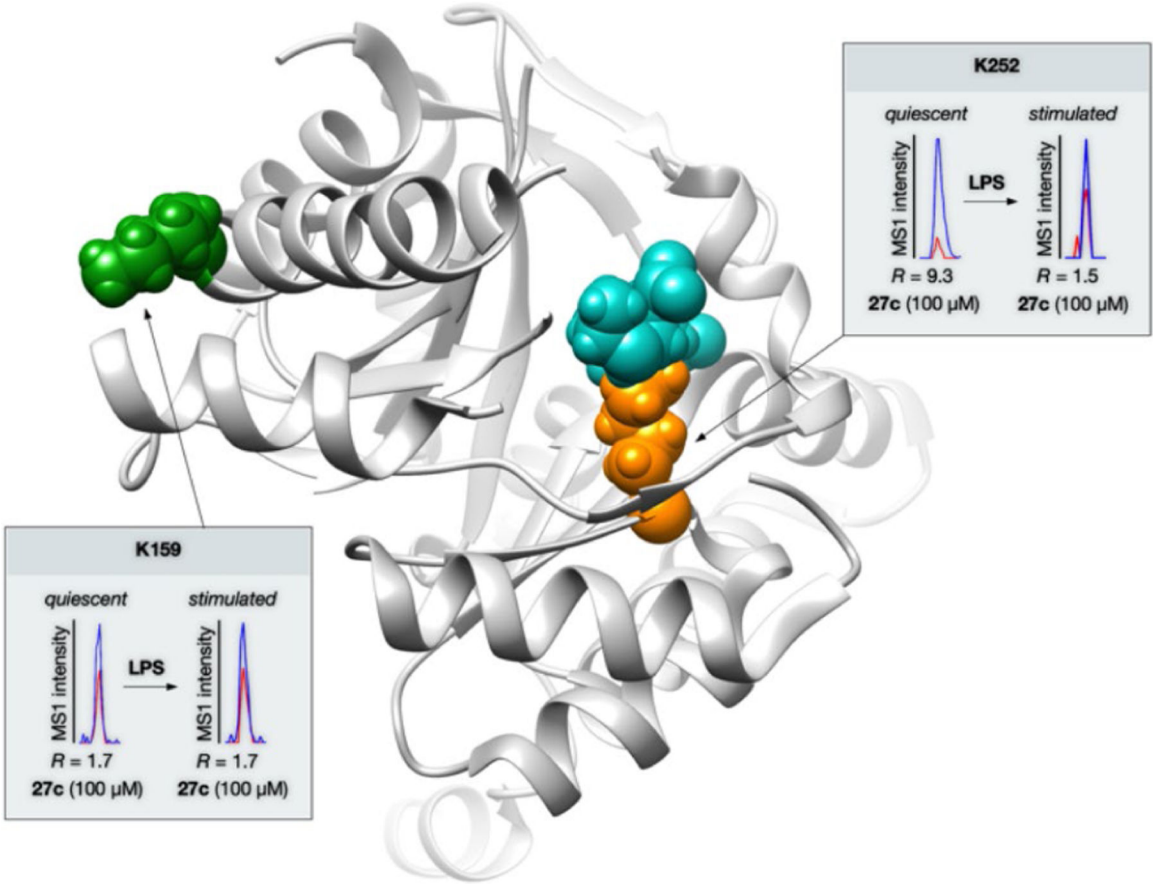
a, Western blot showing recombinantly co-expressed HA-tagged WT Ku80 with Flag-tagged WT and K351R mutant forms of Ku70 in HEK293T cells. **b**, Lysates of HEK293T cells co-expressing WT Ku80 with WT (left panel) and K351R (right panel) mutant forms of Ku70 were co-immunoprecipitated with anti-Flag antibody (1 h, 4 °C), treated with DMSO or **11e** at the indicated concentrations (1 h, 23 °C), washed, and analyzed by Western blotting. Western blots in **a** and **b** are representative of four independent experiments.



Extended Data Fig. 7. Dicarboxaldehyde scout fragments and their functional effects on LPCAT1.

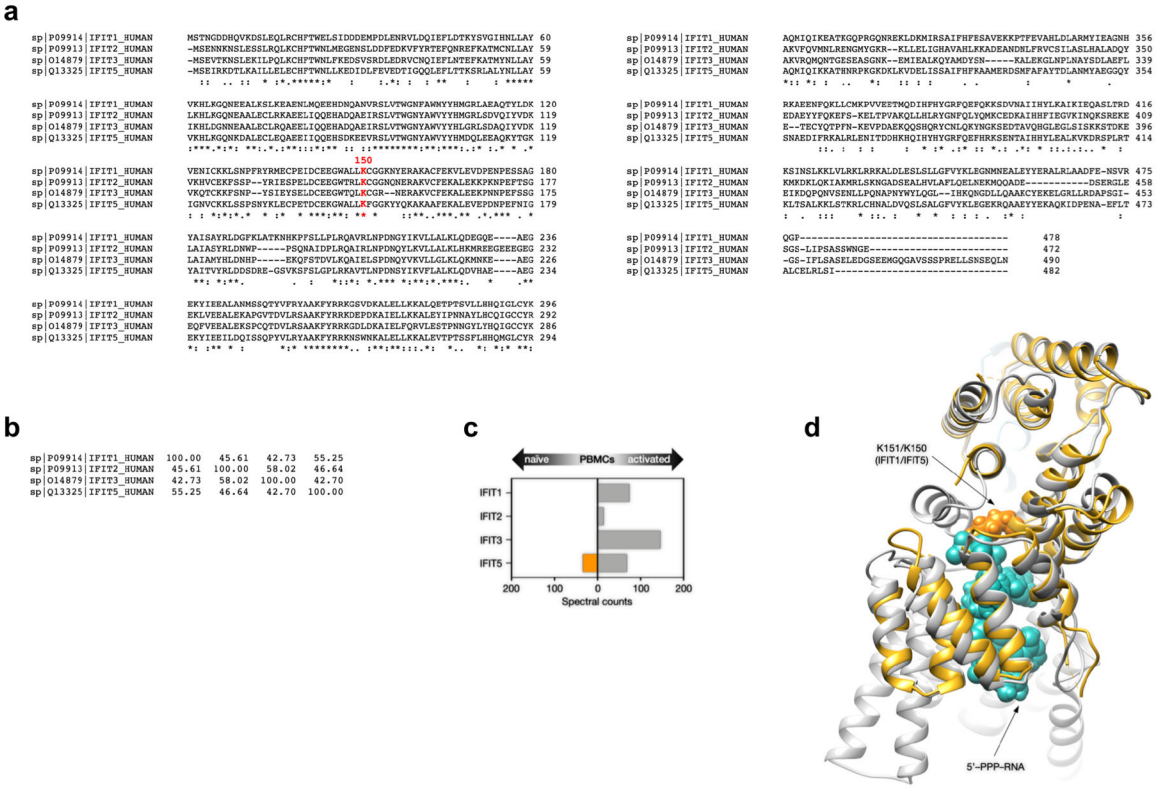
a, Ternary plot showing the proportional lysine reactivity of **27c**, **28o** and **32i** for each lysine. Each point represents a different composition of the three scout fragments based on their individual lysine reactivity ratio (*R*) values, with the maximum proportion (100%) of each fragment in each corner of the triangle and the minimum proportion (0%) at the opposite line. Extracted MS1 chromatograms of representative competed lysines targeted by scout fragments with differential *R* values (also see Supplementary Dataset 3). **b**, Percent identity matrix of human LPCAT1–4 and AGPAT1–4 (https://www.ebi.ac.uk/Tools/msa/clustalo). **c**, Conservation of K221 of LPCAT1 across species (https://www.ncbi.nlm.nih.gov/homologene). **d**, **28o** produces concentration-depended

blockades of WT-LPCAT1 activity. Data represent average values ± SD; n = 3 per group from three biologically independent experiments. P values were 0.00074, 0.000028, 0.000065, 0.0050, and 0.0000062. Statistical significance was calculated for changes >25% in magnitude in comparison to DMSO-treated samples with unpaired two-tailed Student's t-tests: *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001. **e**, SAR for aminophilic compound blockade of lyso-PC hydrolysis activity of recombinantly expressed LPCAT1 in HEK293T cell lysates. Data represent average values ± SD; n = 3 per group. **f**, Compounds **28o** and **28m** produced greater blockade of LPCAT1 enzymatic activity compared to structural analogs **28p** or **28l**. **g**, Compound **28f** produced greater blockade of LPCAT1 enzymatic activity compared to structural analog **28k**. **h**, **28f** produced concentration-depended blockades of WT-LPCAT1 activity. Data represent average values ± SD; n = 3 per group from three biologically independent experiments. Statistical significance was calculated for changes >25% in magnitude in comparison to DMSO-treated samples with unpaired two-tailed Student's t-tests: *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001. **i, j**, HA-tagged ubiquitin and FLAG-tagged WT or K221R LPCAT1 were co-expressed in HEK293T cells in the presence of proteasome inhibitor MG132 (10 μM) for 14 h (**i**) or 2 h (**j**), after which cell lysates were subjected to *anti*-FLAG immunoprecipitation, and the affinity-enriched precipitates analyzed by *anti*-HA immunoblotting. For **i**, mock-transfected cells and LPCAT1-WT cells not treated with MG132 were used as controls. For **j**, FLAG-tagged GFP-transfected cells were used as a control. **k**, FLAG-tagged WT or K221R LPCAT1, or GFP-transfected HEK293T cells were treated with MG132 (10 μM, 2 h), followed by anti-FLAG immunoprecipitation and the affinity-enriched precipitates were analyzed by anti-ubiquitin immunoblotting. Western blots in **i-k** are representative of four independent biological experiments.



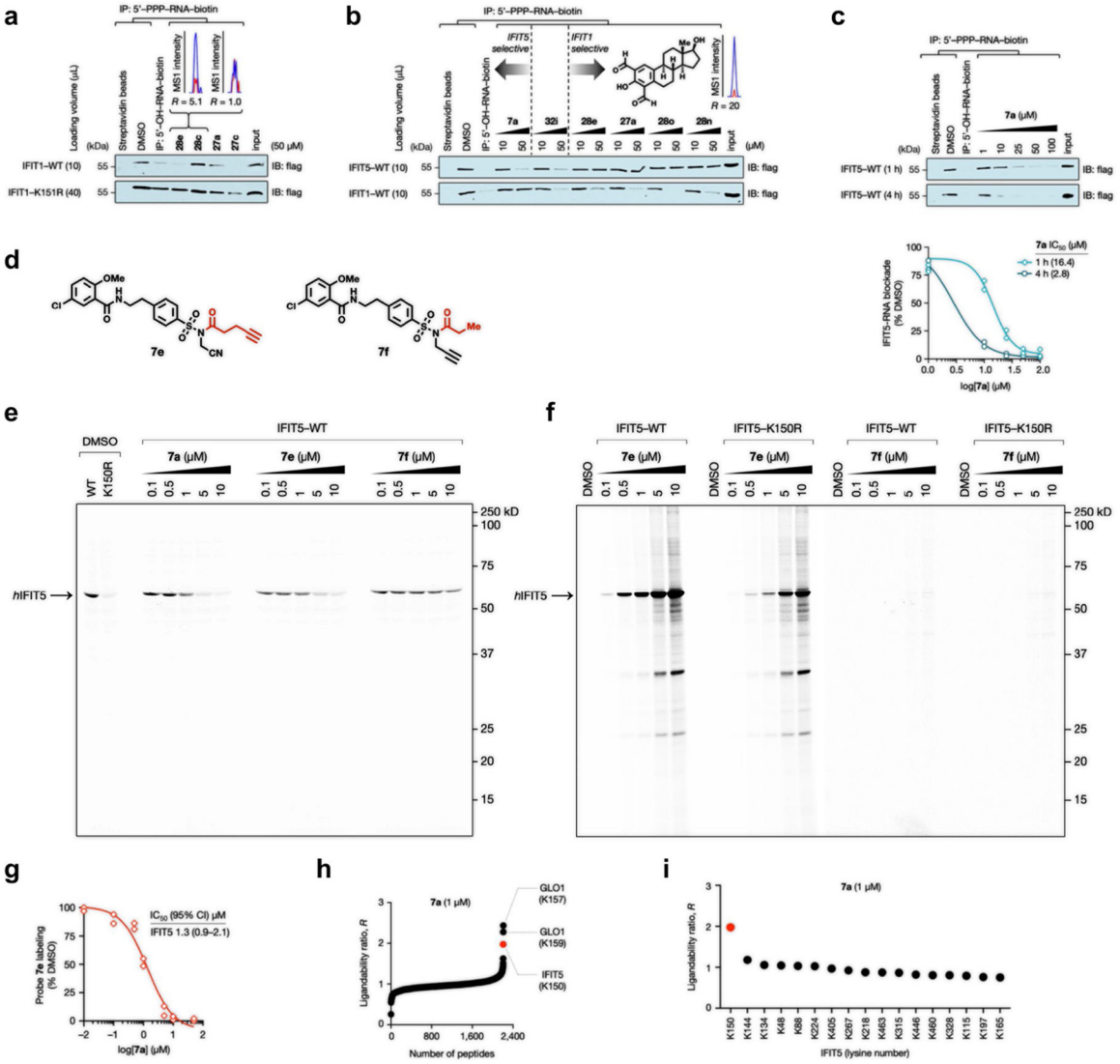
Extended Data Fig. 8. Example of differential lysine ligandability event in stimulated immune cells.

The location of liganded lysine K252 (orange) mapped onto the crystal structure of ALAD (gray, PDB ID: 1PV8). K252 showed substantially weaker interactions with **27c** (100 μM, 23 °C, 1 h) in LPS-stimulated (R = 1.5) vs quiescent PBMCs (R = 9.3), whereas the reactivity of K159 (green) remained largely unchanged by LPS treatment (R = 1.7). K252 is an active-site residue responsible for reversible Schiff-base formation with substrate (blue).



Extended Data Fig. 9. Characterization of aminophilic compounds that selectively inhibit IFIT family of RNA-binding proteins.

a-b, Multiple sequence alignment (**a**) and percent identity matrix (**b**) of human IFIT paralogs (https://www.ebi.ac.uk/Tools/msa/clustalo). The red highlight marks a conserved and liganded lysine. **c**, Aggregate spectral counts for quantified lysine-containing peptides for IFIT proteins in human PBMCs ± LPS treatment. Data represent average values ± SD; n = 3 per group from three biologically independent experiments. **d**, Location of liganded lysines (orange) mapped onto the aligned crystal structures of N-terminal domains in IFIT5 (gray, PDB ID: 4HOT) and IFIT1 (yellow, PDB ID: 4HOU) displaying 5'-PPP-RNA (blue) in the nucleotide binding cleft.



Extended Data Fig. 10. Characterization of aminophilic compounds that inhibit the IFIT family of antiviral RNA-binding proteins.

a, Extracted MS1 chromatograms with corresponding isoTOP-ABPP ratios (top) and Western blot analysis (bottom) from biotinylated RNA pulldown experiments of WT-IFIT1 and the K151R-IFIT1 mutant from HEK293T cell lysates treated with the indicated concentrations of aminophilic compounds. Western blot is representative of three independent experiments). Also see Supplementary Dataset 4. **b**, Western blot analysis from biotinylated RNA pulldown experiments of WT-IFIT1 and IFIT5 from HEK293T cell lysates showing concentration-dependent blockade of RNA binding by indicated aminophilic compounds. Western blot is representative of three independent experiments). Also see Supplementary Dataset 4. **c**, Concentration-dependent blockade (upper panel) and fitted IC_{50} curve (lower panel) of RNA binding of WT-IFIT5 by **7a** after 1 versus 4 h of pre-incubation (n = 2 biologically independent experiments). **d**, Structures of **7e** containing an alkyne moiety on “staying group” and **7f** with an alkyne moiety on “leaving group”. Highlighted in red are “staying groups” in both compounds. **e**, Representative competition

gel showing concentration-dependent blockade of probe **3** labeling by **7a**, **7e**, and **7f** of recombinant WT-IFIT5 in HEK293T cell lysates. **f**, Concentration-dependent labeling of recombinantly expressed WT-IFIT5 and the K150R mutant in HEK293T cell lysates by the clickable probes **7e** and **7f**. gel-ABPP data in **e** and **f** are representative of three independent experiments. **g**, Fitted *in situ* IC_{50} curve for the concentration-dependent blockade of the **7e**-WT-IFIT5 interaction by **7a** in transfected HEK293T cells (n = 4 biologically independent experiments). **h**, Average ratio values for lysines quantified by isoTOP-ABPP in IFIT5-transfected HEK293T cells treated *in situ* with **7a** (1 μM, 2 h) (n = 2 independent experiments; also see Supplementary Dataset 3). **i**, *R* values for quantified lysines in IFIT5 of experiment described in part **h**.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

This work was supported by the NIH (CA231991, Al-126592), a Hewitt Foundation for Medical Research Fellowship (to M.E.A.), a Sir Henry Wellcome Postdoctoral Fellowship, Wellcome Trust, U.K. (to M.E.K.), Pfizer, and Vividion Therapeutics.

Data availability.

All data associated with this study are available in the published article and its supplementary information. All raw proteomics data will be uploaded to the PRIDE repository upon publication and corresponding accession information will be provided in the published article.

References

1. Schreiber SL, Kotz JD, Li M, Aubé J, Austin CP et al. Advancing biological understanding and therapeutics discovery with small-molecule probes. Cell 161, 1252–1265 (2015). [PubMed: 26046436]
2. Hopkins AL & Groom CR The druggable genome. Nat. Rev. Drug Discov 1, 727–730 (2002). [PubMed: 12209152]
3. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA et al. Impact of high-throughput screening in biomedical research. Nat. Rev. Drug Discov 10, 188–195 (2011). [PubMed: 21358738]
4. Scott DE, Coyne AG, Hudson SA & Abell C Fragment-based approaches in drug discovery and chemical biology. Biochemistry 51, 4990–5003 (2012). [PubMed: 22697260]
5. Johnson DS, Weerapana E & Cravatt BF Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. Future Med. Chem 2, 949–964 (2010). [PubMed: 20640225]
6. Backus KM, Correia BE, Lum KM, Forli S, Horning BD et al. Proteome-wide covalent ligand discovery in native biological systems. Nature 534, 570–574 (2016). [PubMed: 27309814]
7. Hacker SM, Backus KM, Lazear MR, Forli S, Correia BE et al. Global profiling of lysine reactivity and ligandability in the human proteome. Nat. Chem 9, 1181–1190 (2017). [PubMed: 29168484]
8. Ward CC, Kleinman JI & Nomura DK NHS-esters as versatile reactivity-based probes for mapping proteome-wide ligandable hotspots. ACS Chem. Biol 12, 1478–1483 (2017). [PubMed: 28445029]
9. Bachovchin DA & Cravatt BF The pharmacological landscape and therapeutic potential of serine hydrolases. Nat. Rev. Drug Discov 11, 52–68 (2012). [PubMed: 22212679]
10. Kato D, Boatright KM, Berger AB, Nazif T, Blum G et al. Activity-based probes that target diverse cysteine protease families. Nat. Chem. Biol 1, 33–38 (2005). [PubMed: 16407991]

11. Chaikuad A, Koch P, Laufer SA & Knapp S The cysteinome of protein kinases as a target in drug development. Angew. Chem. Int. Ed 57, 4372–4385 (2018).
12. Walker CJ, Oaks JJ, Santhanam R, Neviani P, Harb JG et al. Preclinical and clinical efficacy of xpo1/crm1 inhibition by the karyopherin inhibitor kpt-330 in ph+ leukemias. Blood 122, 3034–3044 (2013). [PubMed: 23970380]
13. Ostrem JM, Peters U, Sos ML, Wells JA & Shokat KM K-ras(g12c) inhibitors allosterically control gtp affinity and effector interactions. Nature 503, 548–551 (2013). [PubMed: 24256730]
14. Zhao Q, Ouyang X, Wan X, Gajiwala KS, Kath JC et al. Broad-spectrum kinase profiling in live cells with lysine-targeted sulfonyl fluoride probes. J. Am. Chem. Soc 139, 680–685 (2017). [PubMed: 28051857]
15. Mortenson DE, Brighty GJ, Plate L, Bare G, Chen W et al. "Inverse drug discovery" strategy to identify proteins that are targeted by latent electrophiles as exemplified by aryl fluorosulfates. J. Am. Chem. Soc 140, 200–210 (2018). [PubMed: 29265822]
16. Shannon DA, Banerjee R, Webster ER, Bak DW, Wang C et al. Investigating the proteome reactivity and selectivity of aryl halides. J. Am. Chem. Soc 136, 3330–3333 (2014). [PubMed: 24548313]
17. Choi S, Connelly S, Reixach N, Wilson IA & Kelly JW Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. Nat. Chem. Biol 6, 133–139 (2010). [PubMed: 20081815]
18. Tamura T, Ueda T, Goto T, Tsukidate T, Shapira Y et al. Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed n-acyl-n-alkyl sulfonamide. Nat. Commun 9, 1870 (2018). [PubMed: 29760386]
19. Suh EH, Liu Y, Connelly S, Genereux JC, Wilson IA et al. Stilbene vinyl sulfonamides as fluorogenic sensors of and traceless covalent kinetic stabilizers of transthyretin that prevent amyloidogenesis. J Am Chem Soc 135, 17869–17880 (2013). [PubMed: 24180271]
20. Hunter MJ & Ludwig ML The reaction of imidoesters with proteins and related small molecules. J Am Chem Soc 84, 3491–3504 (1962).
21. Nakamura T, Kawai Y, Kitamoto N, Osawa T & Kato Y Covalent modification of lysine residues by allyl isothiocyanate in physiological conditions: Plausible transformation of isothiocyanate from thiol to amine. Chem. Res. Toxicol 22, 536–542 (2009). [PubMed: 19216492]
22. Metcalf B, Chuang C, Dufu K, Patel MP, Silva-Garcia A et al. Discovery of gbt440, an orally bioavailable r-state stabilizer of sickle cell hemoglobin. ACS Med. Chem. Lett 8, 321–326 (2017). [PubMed: 28337324]
23. Akçay G, Belmonte MA, Aquila B, Chuaqui C, Hird AW et al. Inhibition of mcl-1 through covalent modification of a noncatalytic lysine side chain. Nat. Chem. Biol 12, 931–936 (2016). [PubMed: 27595327]
24. Pettinger J, Le Bihan Y-V, Widya M, van Montfort RLM, Jones K et al. An irreversible inhibitor of hsp72 that unexpectedly targets lysine-56. Angew. Chem. Int. Ed 56, 3536–3540 (2017).
25. Cuesta A & Taunton J Lysine-targeted inhibitors and chemoproteomic probes. Ann. Rev. Biochem 88, 365–381 (2019). [PubMed: 30633551]
26. Wang C, Weerapana E, Blewett MM & Cravatt BF A chemoproteomic platform to quantitatively map targets of lipid-derived electrophiles. Nat. Methods 11, 79–85 (2014). [PubMed: 24292485]
27. Weerapana E, Wang C, Simon GM, Richter F, Khare S et al. Quantitative reactivity profiling predicts functional cysteines in proteomes. Nature 468, 790–795 (2010). [PubMed: 21085121]
28. Ma N, Hu J, Zhang Z-M, Liu W, Huang M et al. 2*H*-azirine-based reagents for chemoselective bioconjugation at carboxyl residues inside live cells. J. Am. Chem. Soc 142, 6051–6059 (2020). [PubMed: 32159959]
29. Bach K, Beerkens BLH, Zanon PRA & Hacker SM Light-activatable, 2,5-disubstituted tetrazoles for the proteome-wide profiling of aspartates and glutamates in living bacteria. ACS Cent. Sci 6, 546–554 (2020). [PubMed: 32342004]
30. Cheng K, Lee J-S, Hao P, Yao SQ et al. Tetrazole-based probes for integrated phenotypic screening, affinity-based proteome profiling, and sensitive detection of a cancer biomarker. Angew. Chem. Int. Ed 56, 15044–15048 (2017).

31. Lin S, Yang X, Jia S, Weeks AM, Hornsby M et al. Redox-based reagents for chemoselective methionine bioconjugation. Science 355, 597–602 (2017). [PubMed: 28183972]
32. Hahm HS, Toroitich EK, Borne AL, Brulet JW, Libby AH et al. Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. Nat. Chem. Biol 16, 150–159 (2020). [PubMed: 31768034]
33. Balthaser BR, Maloney MC, Beeler AB, Porco JA & Snyder JK Remodelling of the natural product fumagillol employing a reaction discovery approach. Nat. Chem 3, 969–973 (2011). [PubMed: 22213919]
34. Lajkiewicz NJ, Cognetta AB, Niphakis MJ, Cravatt BF & Porco JA Remodeling natural products: Chemistry and serine hydrolase activity of a rocaglate-derived β-lactone. J. Am. Chem. Soc 136, 2659–2664 (2014). [PubMed: 24447064]
35. Lipinski CA Lead- and drug-like compounds: The rule-of-five revolution. Drug Discov. Today Technol 1, 337–341 (2004). [PubMed: 24981612]
36. Patricelli MP, Giang DK, Stamp LM & Burbaum JJ Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. Proteomics 1, 1067–1071 (2001). [PubMed: 11990500]
37. Rostovtsev VV, Green LG, Fokin VV & Sharpless KB A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. Angew. Chem. Int. Ed 41, 2596–2599 (2002).
38. Zhang Z, Li C, Wu F, Ma R, Luan J et al. Genomic variations of the mevalonate pathway in porokeratosis. eLife 4, e06322 (2015). [PubMed: 26202976]
39. Brooks SS, Wall AL, Golzio C, Reid DW, Kondyles A et al. A novel ribosomopathy caused by dysfunction of rpl10 disrupts neurodevelopment and causes x-linked microcephaly in humans. Genetics 198, 723–733 (2014). [PubMed: 25316788]
40. Lee D-S, Flachsova E, Bodnarova M, Demeler B, Martasek P et al. Structural basis of hereditary coproporphyria. Proc. Natl. Acad. Sci. U.S.A 102, 14232–14237 (2005). [PubMed: 16176984]
41. Hussey AJ & Hayes JD Characterization of a human class-theta glutathione S-transferase with activity towards 1-menaphthyl sulphate. Biochem. J 286, 929–935 (1992). [PubMed: 1417752]
42. Schmiedeknecht G, Kerkhoff C, Orso E, Stohr J, Aslanidis C et al. Isolation and characterization of a 14.5-kda trichloroacetic-acid-soluble translational inhibitor protein from human monocytes that is upregulated upon cellular differentiation. Eur. J. Biochem 242, 339–351 (1996). [PubMed: 8973653]
43. Katritzky AR & Yousaf TI A c-13 nuclear magnetic resonance study of the pyrimidine synthesis by the reactions of 1,3-dicarbonyl compounds with amidines and ureas. Can. J. Chem 64, 2087–2093 (1986).
44. Kragelund BB, Weterings E, Hartmann-Petersen R & Keijzers G The ku70/80 ring in non-homologous end-joining: Easy to slip on, hard to remove. Front. Biosci 21, 514–527 (2016).
45. Tung CL, Wong CTT, Fung EYM & Li X Traceless and chemoselective amine bioconjugation via phthalimidine formation in native protein modification. Org. Lett 18, 2600–2603 (2016). [PubMed: 27191384]
46. Adhikari S, Ghosh A, Mandal S, Guria S, Banerjee PP et al. Colorimetric and fluorescence probe for the detection of nano-molar lysine in aqueous medium. Org. Biomol. Chem 14, 10688–10694 (2016). [PubMed: 27801458]
47. Bar-Peled L, Kemper EK, Suciu RM, Vinogradova EV, Backus KM et al. Chemical proteomics identifies druggable vulnerabilities in a genetically defined cancer. Cell 171, 696–709.e623 (2017). [PubMed: 28965760]
48. Zhang X, Crowley VM, Wucherpfennig TG, Dix MM & Cravatt BF Electrophilic protacs that degrade nuclear proteins by engaging dcaf16. Nat. Chem. Biol 15, 737–746 (2019). [PubMed: 31209349]
49. Vinogradova EV, Zhang X, Remillard D, Lazar DC, Suciu RM et al. An activity-guided map of electrophile-cysteine interactions in primary human T cells. Cell 182, 1009–1026 (2020). [PubMed: 32730809]
50. Shi C, Qiao S, Wang S, Wu T, Ji G Recent progress of lysophosphatidylcholine acyltransferases in metabolic disease and cancer. Int. J. Clin. Exp. Med 11, 8941–8953 (2018).

51. Zou C, Butler PL, Coon TA, Smith RM, Hammen G et al. Lps impairs phospholipid synthesis by triggering β-transducin repeat-containing protein (β-TRCP)-mediated polyubiquitination and degradation of the surfactant enzyme acyl-coa:Lysophosphatidylcholine acyltransferase 1 (LPCAT1). J. Biol. Chem 286, 2719–2727 (2011). [PubMed: 21068446]
52. Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K et al. Social network architecture of human immune cells unveiled by quantitative proteomics. Nat. Immunol 18, 583–593 (2017). [PubMed: 28263321]
53. Fensterl V & Sen GC Interferon-induced IFIT proteins: Their role in viral pathogenesis. J. Virol 89, 2462–2468 (2015). [PubMed: 25428874]
54. Lo U-G, Bao J, Cen J, Yeh H-C, Luo J et al. Interferon-induced IFIT5 promotes epithelial-to-mesenchymal transition leading to renal cancer invasion. Am. J. Clin. Exp. Urol 7, 31–45 (2019). [PubMed: 30906803]
55. Abbas YM, Pichlmair A, Górna MW, Superti-Furga G & Nagar B Structural basis for viral 5'-PPP-RNA recognition by human ifit proteins. Nature 494, 60–64 (2013). [PubMed: 23334420]
56. Speers AE, Adam GC & Cravatt BF Activity-based protein profiling in vivo using a copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. J. Am. Chem. Soc 125, 4686–4687 (2003). [PubMed: 12696868]
57. Krüger DM, Neubacher S & Grossmann TN Protein-RNA interactions: Structural characteristics and hotspot amino acids. RNA 24, 1457–1465 (2018). [PubMed: 30093489]
58. Zanon Patrick R. A.; Yu Fengchao; Musacchio Patricia; Lewald Lisa; Zollo Michael; Krauskopf Kristina; et al. (2021): Profiling the proteome-wide selectivity of diverse electrophiles. ChemRxiv. Preprint 10.26434/chemrxiv.14186561.v1
59. Congreve M, Carr R, Murray C & Jhoti HA 'rule of three' for fragment-based lead discovery? Drug Discov. Today 8, 876–877 (2003).

Extended Data and Methods References:

1. Lipinski CA Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov. Today Technol 1, 337–341 (2004). [PubMed: 24981612]
2. Congreve M, Carr R, Murray C & Jhoti HA 'rule of three' for fragment-based lead discovery? Drug Discov. Today 8, 876–877 (2003).
3. Sander T, Freyss J, von Korff M & Rufener C DataWarrior: an open-source program for chemistry aware data visualization and analysis. J. Chem. Inf. Model 55, 460–473 (2015). [PubMed: 25558886]
4. Zhang Z, Li C, Wu F, et al. Genomic variations of the mevalonate pathway in porokeratosis. Elife 4, e06322 (2015). [PubMed: 26202976]
5. Herdendorf TJ & Miziorko HM Functional evaluation of conserved basic residues in human phosphomevalonate kinase. Biochemistry 46, 11780–11788 (2007). [PubMed: 17902708]
6. Katritzky AR & Yousaf TI A C-13 nuclear magnetic resonance study of the pyrimidine synthesis by the reactions of 1,3-dicarbonyl compounds with amidines and ureas. Can. J. Chem 64, 2087–2093 (1986).
7. Weerapana E, Wang C, Simon GM, Richter F, Khare S et al. Quantitative reactivity profiling predicts functional cysteines in proteomes. Nature 468, 790–795 (2010). [PubMed: 21085121]

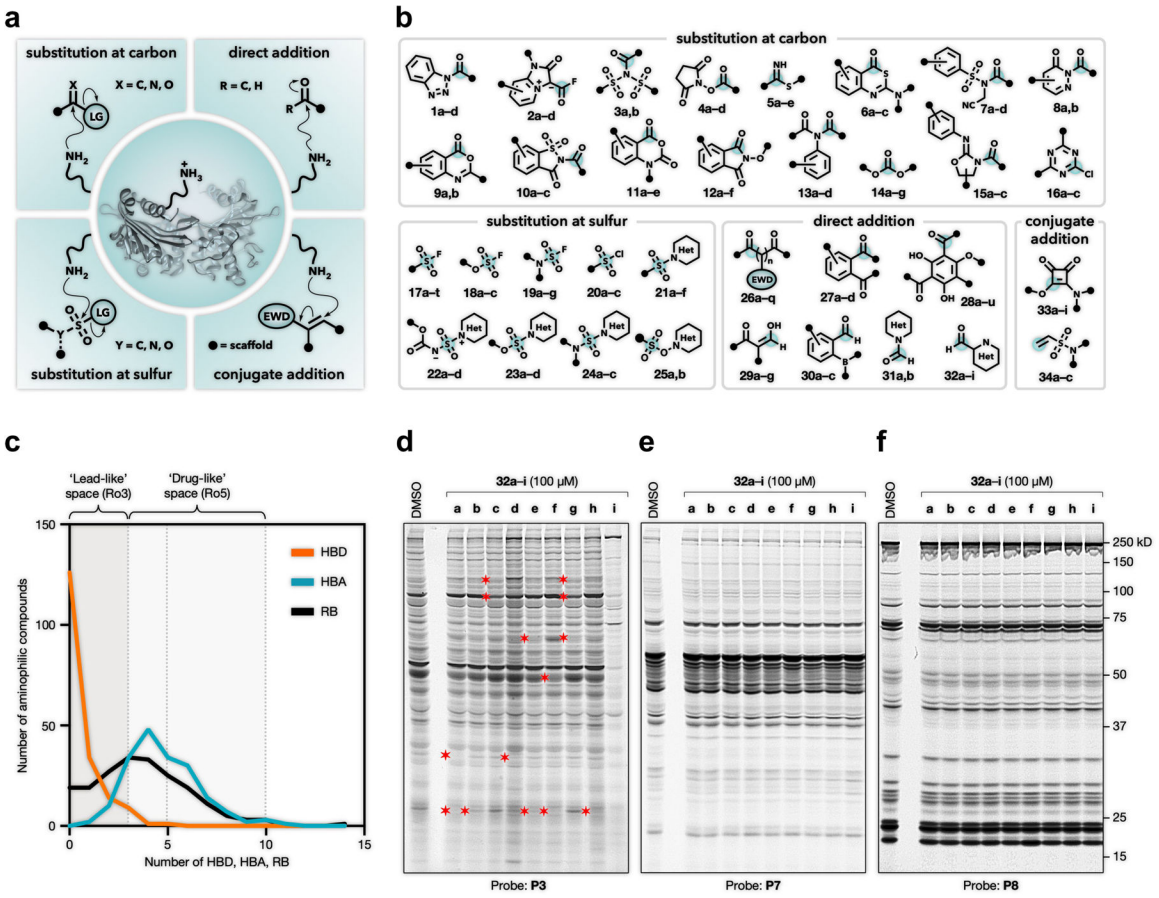


Figure 1. An aminophilic compound library for mapping small molecule-lysine interactions in the proteome.

a, Mechanism-based categorization of aminophilic chemotypes by their predicted modes of reactivity. **B**, Structural composition and diversity of representative aminophilic chemotypes clustered by reactivity modes (presumed electrophilic centers are highlighted by blue circles). See Supplementary Dataset 1 for a complete list of compound structures for each chemotype. **C**, Distribution of aminophilic compounds by the number of hydrogen-bond donors (HBDs, orange line), hydrogen-bond acceptors (HBAs, blue line) and rotatable bonds (RBs, black line). Plot highlights compounds that follow Lipinski's "rule-of-five" (drug-like space, light-grey box)[35] and Congreve's "rule-of-three" (fragment-based lead-like space, dark-grey box)[59]. **D-f**, Qualitative assessment of apparent amino acid reactivity of representative compounds **32a-I** from the heterocyclic aldehyde chemotype, as measured by competitive gel-ABPP with the lysine-directed probe Alexa-Fluor® 488 (**P3**) (**d**)[7], cysteine-directed probe iodoacetamide-rhodamine (**P7**) (**e**)[6] and serine hydrolase-directed probe fluorophosphonate-rhodamine (**P8**) (**f**)[36] in proteomic lysate of the MDA-MB-231 human breast cancer cell line. Competitive profiling experiments were generally performed as follows: soluble proteome from MDA-MB-231 cells was treated with the indicated compounds (100 μM, 1 h, 23 °C), followed by labeling with the indicated fluorogenic probe (2 μM, 1 h, 23 °C) and analysis by SDS-PAGE and in-gel fluorescence scanning. Red asterisks mark representative compound-competed proteins. This experiment was conducted twice (n = 2) with similar results.

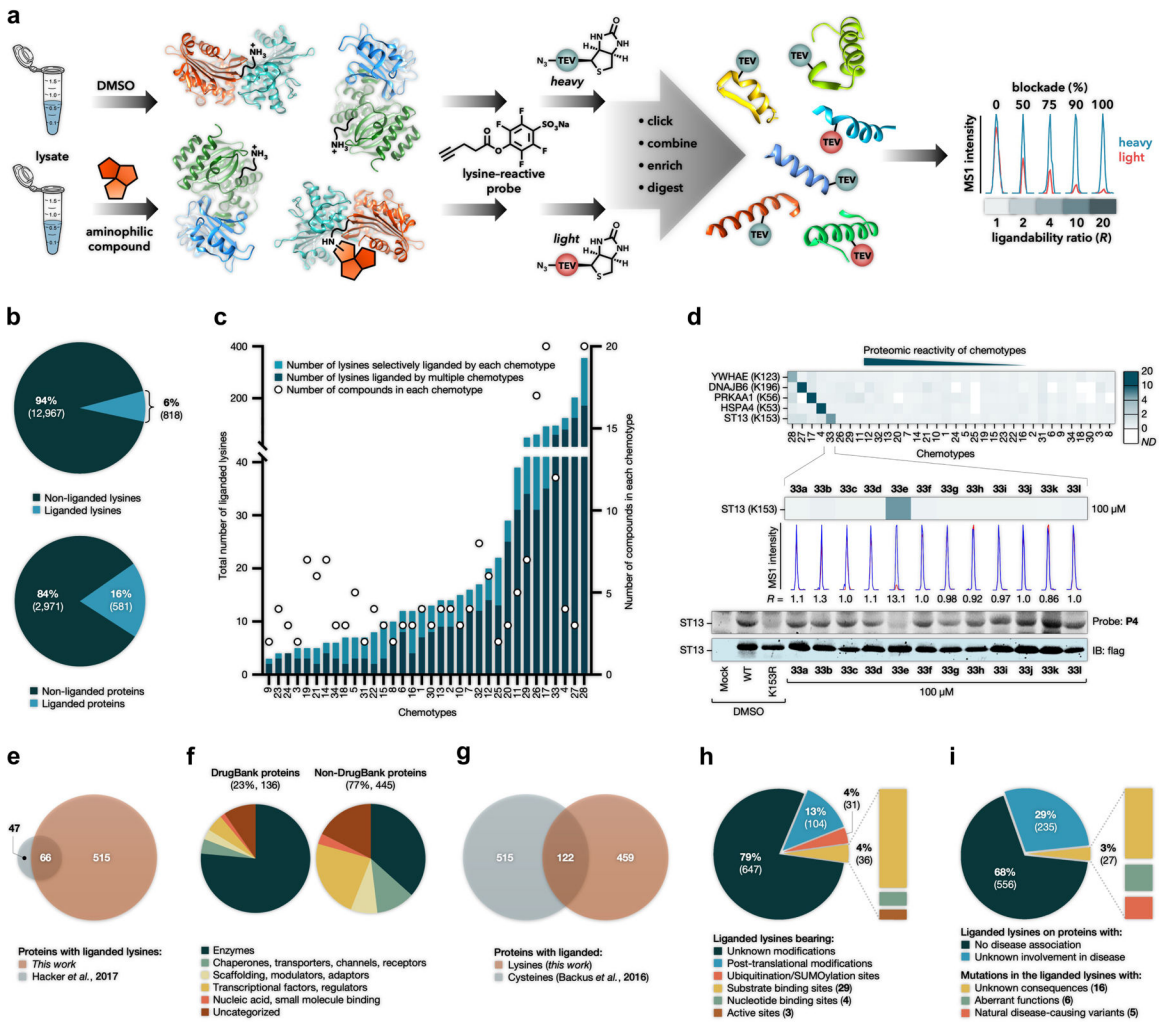


Figure 2. A global map of aminophilic compound-lysine interactions in the human proteome.

a, General schematic for competitive isoTOP-ABPP experiments and experimental workflow to identify lysines liganded by aminophilic compounds. **b**, Fraction of total quantified lysines (left) and proteins (right) liganded by aminophilic compounds. **c**, Plot comparing the number of liganded lysines for each aminophilic chemotype, with blue and black designating lysines that were engaged by a single or multiple chemotypes, respectively. **d**, Top, Heatmap showing *R* values of representative lysines preferentially liganded by a single chemotype. Middle, Extracted MS1 chromatograms with corresponding *R* values for K153 in ST13 showing a highly restricted SAR across individual members of the squarate chemotype (also see Supplementary Dataset 3). Bottom, corresponding recapitulation by gel-based ABPP of the recombinant ST13 (also see Supplementary Dataset 4). **e**, Overlap of proteins with liganded lysines targeted by chemotypes evaluated in this study and by activated esters evaluated in a previous study[7]. **f**, Functional class distribution of liganded DrugBank (left) and non-DrugBank (right) proteins. **g**, Overlap of proteins harboring liganded lysines and liganded cysteines[6] in Ramos and MDA-MB-231 proteomes. **h**, Categorization of liganded lysines based on the indicated functional categories. **i**, Distribution of liganded lysines in proteins that have human disease-relevance (as assessed

by pathogenic mutations that lead to monogenic disorders defined in the OMIM database) and functional consequences of mutations of the liganded lysine residues themselves (in cases, where these mutations are associated with disease). For panels **e-i**, lysines and proteins exclusively liganded by compounds **27c, 28o**, and **32i** were excluded from the analyses, as we revisit these preferred targets of these “scout” compounds in Figure 5.

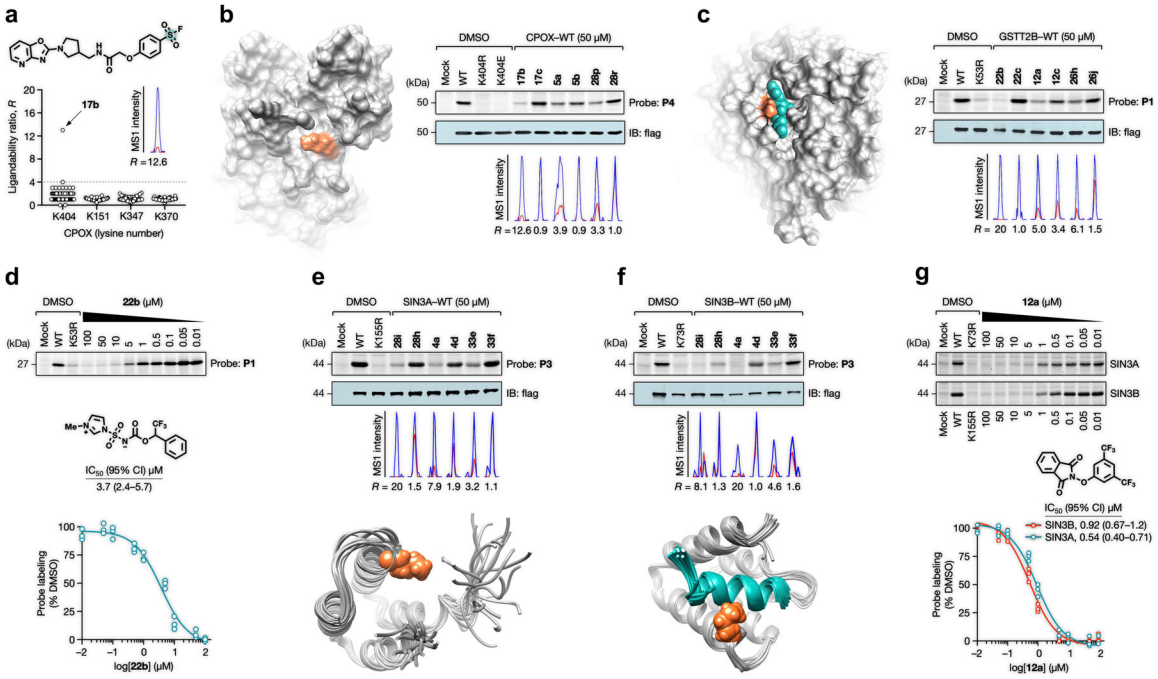


Figure 3. Confirmation and SAR analysis of aminophilic compound-lysine interactions with recombinantly expressed proteins.

a, Structure of sulfonyl fluoride **17b** and *R* values for quantified lysines in CPOX, identifying K404 as a liganded lysine in this protein. Each point represents a distinct aminophilic compound-lysine interaction quantified by isoTOP-ABPP. The dashed line marks the *R* value of 4 used to define a liganding event. The results shown are average ratios from three experiments (n = 3 biologically independent experiments). **b, c**, Site-specific aminophilic compound-lysine interactions are preserved in recombinant proteins. Right panels: Lysates from HEK293T cells recombinantly expressing representative liganded proteins and their corresponding lysine-to-arginine mutants as Flag epitope-tagged proteins were treated with the indicated aminophilic compounds (50 μM, 1 h) followed by treatment with the indicated lysine-reactive probes and analyzed by gel-ABPP (top) and western blotting (middle). Below the blots are shown extracted MS1 chromatograms depicting *R* values for the indicated aminophilic compound-lysine interactions mapped for endogenous proteins by competitive isoTOP-ABPP (also see Supplementary Datasets 3 and 4). Left images: Location of liganded lysines (orange) in protein crystal structures (gray) of CPOX (**b**, PDB ID: 2AEX) and GSTT2B in complex with glutathione (blue) (**c**, PDB ID: 4MPG). **d**, Top, Representative gel-ABPP data showing concentration-dependent blockade of probe **P1** labeling of recombinant GSTT2B in HEK293T cell lysates by ammoniumsulfonyl carbamate **22b**. Middle, Structure of **22b**. Bottom, IC_{50} curve for blockade of **P1** labeling by **22b**. Data represent average values ± SD; n = 3 per group. CI, confidence interval. **e-f**, Structure-activity relationships (SARs) determined for aminophilic compound interactions with the conserved lysine in endogenous SIN3A (**e**) and SIN3B (**f**) by competitive isoTOP-ABPP recapitulated by gel-ABPP of recombinant proteins. Top, HEK293T cells recombinantly expressing representative liganded proteins and their corresponding lysine-to-arginine mutants as Flag epitope-tagged proteins were treated with the indicated aminophilic

compounds (50 μM, 1 h) followed by treatment with the indicated lysine-reactive probes and analyzed by gel-ABPP (top panel) and western blotting (bottom panel). Middle, Extracted MS1 chromatograms depicting *R* values for the indicated aminophilic compound-lysine interactions identified by competitive isoTOP-ABPP (also see Supplementary Datasets 3 and 4). Bottom, Liganded lysines (orange) mapped onto protein crystal structures (gray) of SIN3A (**e**, PDB ID: 2RMR) and SIN3B (**f**, PDB ID: 2CZY) in complex with neural repressor NRSF/REST (blue). **g**, Top, Representative gel-ABPP data showing concentration-dependent blockade of probe **5** labeling by *N*-hydroxyphthalimide **12a** (middle) of recombinant SIN3A and SIN3B in HEK293T cell lysates. Bottom, Corresponding fitted IC_{50} curves. Data represent average values ± SD; n = 3 per group. CI, confidence interval. For gel-ABPP data in **b, c, d,** and **f**, experiments were conducted three times (n = 3 biologically independent experiments) with similar results.

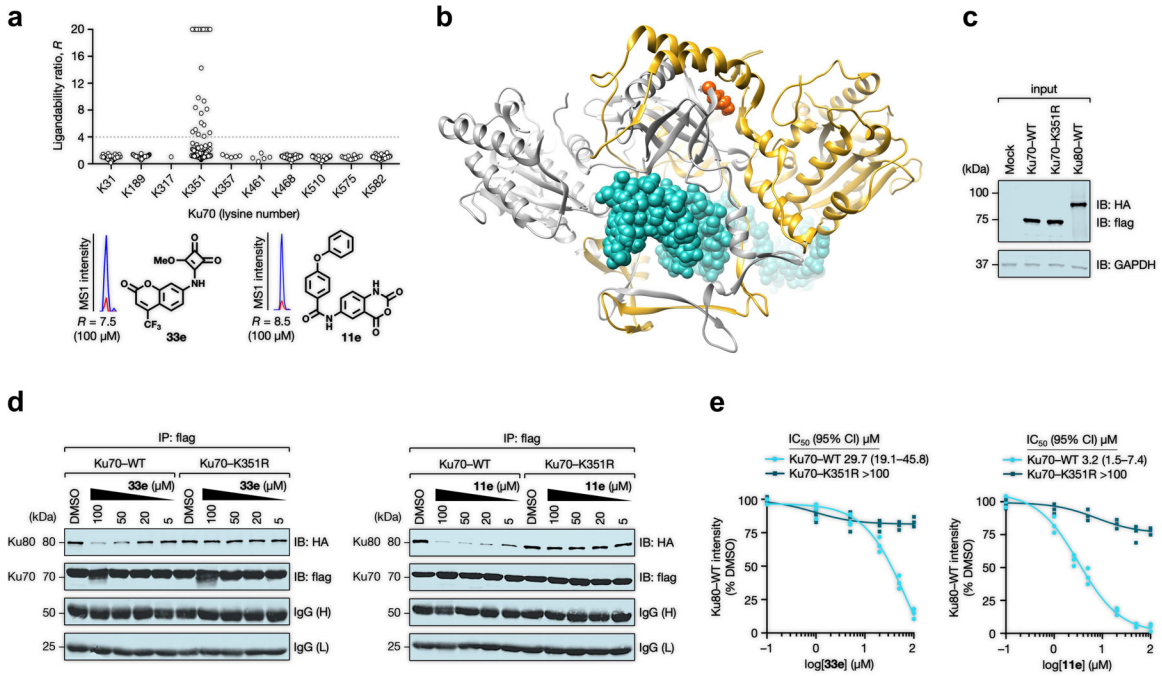


Figure 4. Functional impact of aminophilic compound-lysine interactions for representative proteins.

a, Upper panel, *R* values for quantified lysines in XCRR6 (or Ku70), identifying K351 as the only observed liganded lysine in this protein. Each point represents a distinct aminophilic compound-lysine interaction quantified by isoTOP-ABPP. The dashed line marks the *R* value of 4 used to define a liganded lysine event (also see Supplementary Dataset 3). The results shown are average ratios from three experiments (n = 3 biologically independent experiments). Lower panel, Structures of fragments **33e** and **11e** with extracted MS1 chromatograms depicting *R* values for their respective engagement of K351 in XRCC6 mapped by competitive isoTOP-ABPP. **b**, The location of liganded lysine K351 of Ku70 (orange) mapped onto the crystal structure of the Ku heterodimer (PDB ID: 1JEY) consisting of Ku70 (gray) and Ku80 (yellow) bound to double-stranded DNA (teal). **c-e**, Aminophilic compounds engaging K351 of Ku70 block formation of the Ku70-Ku80 heterodimer. **c**, Western blot analysis showing recombinantly expressed Flag-tagged WT and K351R mutant forms of Ku70, as well as HA-tagged WT-Ku80 in HEK293T cells. **d**, Lysates of cells expressing Ku70 protein variants were treated with DMSO, **33e** (left panel), or **11a** (right panel) at the indicated concentrations (1 h, 23 °C) and then mixed with lysates expressing Ku80 protein (1 h, 23 °C), followed by co-immunoprecipitation with anti-Flag antibody (1 h, 4 °C) and western blot analysis. Also see Supplementary Dataset 4. **e**, Quantification of western blotting data for **33e** and **11e** from three biological replicates. Data represent average values ± SD; n = 3 per group from three biologically independent experiments.

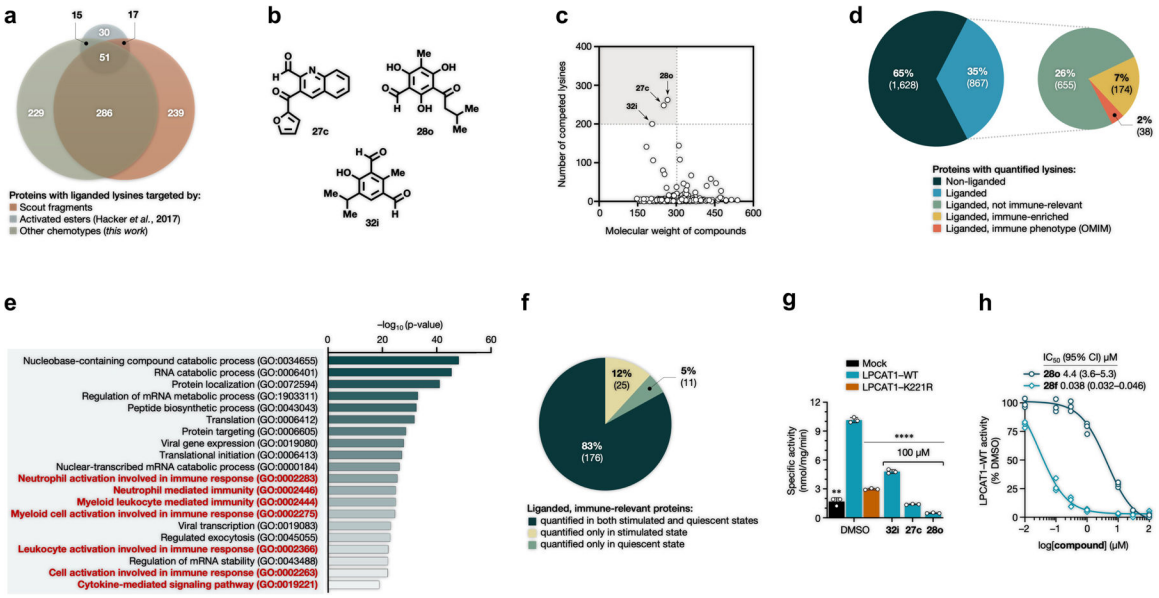


Figure 5. Dicarboxaldehyde scout fragments and their application for profiling lysine ligandability in human immune cells.

a, Overlap of proteins with liganded lysines targeted by scout fragments (**27c**, **28o**, **32i**), other chemotypes from this study, and activated esters from a previous study[7]. **b**, Structures of scout fragments **27c**, **28o**, and **32i**. **c**, Scout fragments **27c**, **28o** and **32i** engage a much larger number of lysines in human cancer cell proteomes compared to other aminophilic compounds. **d**, Fraction of proteins harboring liganded lysines and subset of these liganded proteins that are immune-relevant. **e**, Top-20 enriched clusters of biological processes from gene ontology (GO)-term enrichment analysis of liganded proteins. Red bold font highlights immune-relevant biological processes. **f**, Fraction of liganded, immune-relevant proteins quantified in LPS-stimulated and quiescent PBMCs. **g**, Treatment with scout fragments inhibits the lyso-PC hydrolysis activity of membrane lysates of HEK293T cells recombinantly expressing WT-LPCAT1. Lysates expressing a K221R mutant of LPCAT1 also show impaired lyso-PC hydrolysis activity compared to lysates expressing WT-LPCAT1. The results shown are average ratios from three experiments (n = 3 biologically independent experiments). **h**, Fitted IC_{50} curves for the concentration-dependent inhibition of lyso-PC hydrolysis activity of recombinantly expressed WT LPCAT1 in HEK293T cell lysates by **28o** and **28f**. Data represent average values ± SD; n = 3 per group from three biologically independent experiments. CI, confidence interval. P values were 0.0036 and 0.000027. Statistical significance was calculated for changes >25% in magnitude in comparison to DMSO-treated samples with unpaired two-tailed Student's t-tests: *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

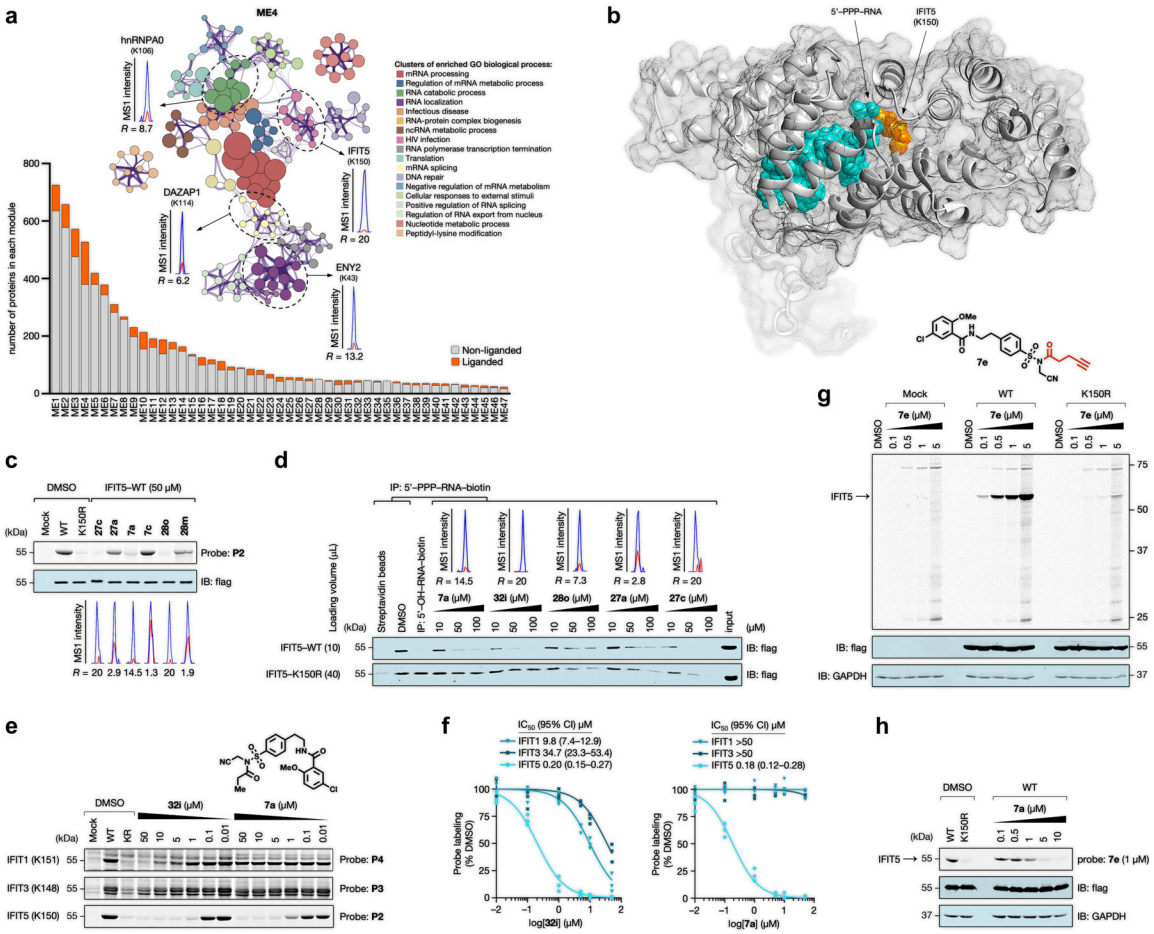


Figure 6. Identification of aminophilic compounds that inhibit the IFIT family of antiviral RNA-binding proteins.

a, Lower panel, Number of proteins with lysines liganded by scout fragments within each functional module of the immune system (modules defined in a previous study[52]). Enrichment network of module 4 (inset, upper panel) showing protein clusters associated with RNA-related functions. Nodes are individual annotation terms, edges represent protein overlap between terms, node size represent annotation enrichment, and fill color represent clusters. Extracted MS1 chromatograms are shown for representative liganded lysines in RNA-binding proteins involved in RNA metabolism, splicing, localization and response to viral infection. **b**, Location of ligandable lysine K150 (orange) mapped onto the crystal structures of IFIT5 (gray, PDB ID: 4HOT) bound to 5’-PPP-RNA (blue) in the nucleotide binding cleft. **c**, SAR for aminophilic compound engagement of K150 of IFIT5, as determined by competitive isoTOP-ABPP is recapitulated by gel-ABPP of recombinant IFIT5. Top, HEK293T cells recombinantly expressing WT-IFIT5 and the corresponding K150R mutant as Flag epitope-tagged proteins were treated with the indicated aminophilic compounds (50 μM, 1 h) followed by treatment with probe **3** and analyzed by gel-ABPP (top panel) and western blotting (bottom panel). Bottom, extracted MS1 chromatograms depicting *R* values for the indicated aminophilic compound-IFIT5 (K150) interactions mapped by competitive isoTOP-ABPP (also see Supplementary Datasets 3 and 4). **d,** Extracted MS1 chromatograms with corresponding isoTOP-ABPP ratios (top) and Western

blot analysis (bottom) from biotinylated RNA pulldown experiments of WT-IFIT5 and the K150R-IFIT5 mutant from HEK293T cell lysates treated with the indicated concentrations of aminophilic compounds. Also see Supplementary Dataset 4. **e, f**, Gel-ABPP data (**e**) and corresponding fitted IC_{50} curves (**f**) for the concentration-dependent blockade of probe labeling of IFIT1, IFIT3, and IFIT5 by **32i** and **7a**. Data represent average values ± SD; n = 3 per group. CI, confidence interval. **g**, Concentration-dependent *in situ* labeling of WT-IFIT5, but not the K150R mutant of IFIT5, by an alkyne probe **7e** in transfected HEK293T cells. Also see Supplementary Dataset 4. **h**, Representative gel-ABPP for the concentration-dependent blockade of the **7e**-WT-IFIT5 interaction by **7a** in transfected HEK293T cells. Data represent average values ± SD; n = 2 per group. CI, confidence interval. For gel-ABPP and Western blotting data in **c**-**e**, **g,** and **h**, experiments were conducted three times (n = 3 biologically independent experiments) with similar results.

Table S1 –

Amplification Primers

Protein	Fwd/ Rev	Sequence	Vector	Restriction Sites	Affinity Tag
CPOX	Fwd	CTGCAGGTCGACGACTTCGCTGGGGAGGCCGGAGG	pRK5	SalI/NotI	N-terminal FLAG
CPOX	Rev	GTTAGCGCGGCCGCTCAACGCACCCAGTCCCTTGGATGGC	pRK5	SalI/NotI	N-terminal FLAG
SIN3A	Fwd	TGGTGCGTCGACGATGAAGCGGCGTTTGGATGA	pRK5	SalI/NotI	N-terminal FLAG
SIN3A	Rev	TACTTAGCGGCCGCTTACACAGAATCAACCTTCTCAGCAGTTG	pRK5	SalI/NotI	N-terminal FLAG
SIN3B	Fwd	CTGCAGGTCGACGATGGCGCACGCTGGCGGTGG	pRK5	SalI/NotI	N-terminal FLAG
SIN3B	Rev	GTTAGCGCGGCCGCTCAGCTGGATCCTATGCGCTTGCAGGATGC	pRK5	SalI/NotI	N-terminal FLAG
GSTT2B	Fwd	CTGCAGGTCGACGATGGGCCTAGAGCTGTTTC	pRK5	SalI/NotI	N-terminal FLAG
GSTT2B	Rev	GTTAGCGCGGCCGCTCAGGGGATCCTGGCGATTC	pRK5	SalI/NotI	N-terminal FLAG
RIDA	Fwd	CTGCAGGTCGACGATGTCGTCCTTGATCAGAAG	pRK5	SalI/NotI	N-terminal FLAG
RIDA	Rev	GTTAGCGCGGCCGCTTATAGTGATGCCGTTGTC	pRK5	SalI/NotI	N-terminal FLAG
Ku70	Fwd	CTGCAGGTCGACGATGTCAGGGTGGGAGTCATAT	pRK5	SalI/NotI	N-terminal FLAG
Ku70	Rev	GTTAGCGCGGCCGCTCAGTCCTGGAAGTGCTTGG	pRK5	SalI/NotI	N-terminal FLAG
Ku80	Fwd	CTGCAGGTCGACGATGGTGCGGTCGGGGAATAAG	pRK5	SalI/NotI	N-terminal HA
Ku80	Rev	GTTAGCGCGGCCGCCTATATCATGTCCAATAAATCGTCCACATCAC	pRK5	SalI/NotI	N-terminal HA
IFIT1	Fwd	CTGCAGGTCGACGATGAGTACAAATGGTGATG	pRK5	SalI/NotI	N-terminal FLAG
IFIT1	Rev	GTTAGCGCGGCCGCCTAAGGACCTTGTCTCACAG	pRK5	SalI/NotI	N-terminal FLAG
IFIT5	Fwd	CTGCAGGTCGACGATGAGTGAAATTCGTAAGGAC	pRK5	SalI/NotI	N-terminal FLAG
IFIT5	Rev	GTTAGCGCGGCCGCTTAAATGGAAAGTCGGAGCTC	pRK5	SalI/NotI	N-terminal FLAG

Table S2 –

Mutagenesis Primers

Protein	Mutation	Fwd/Rev	Sequence
CPOX	K404R	Fwd	GTATGATCGGGGCACAAGGTTTGGCCTCTTC
CPOX	K404R	Rev	CCTTGTGCCCCGATCATACAGCAGATTAAATTC
CPOX	K404E	Fwd	GTATGATCGGGGCACAGAGTTTGGCCTCTTC
CPOX	K404E	Rev	CTCTGTGCCCCGATCATACAGCAGATTAAATTC
SIN3A	K155R	Fwd	CCTTGACATCATGAAGGAATTTCGTTCTCAGAGCATCGACACCCC
SIN3A	K155R	Rev	GGGGTGTCGATGCTCTGAGAACGAAATTCCTTCATGATGTCAAGG
SIN3B	K73R	Fwd	AGATCATGAAGGAGTTCAGGAGCCAGAGCA
SIN3B	K73R	Rev	CCTGAACTCCTTCATGATCTCCAGGAAGCC
GSTT2B	K75R	Fwd	CAACAGCCTGGGGAGGCTGCCGACGCTCAAG
GSTT2B	K75R	Rev	CCTCCCCAGGCTGTTGATCTGCAAGAACTCC
RIDA	K117R	Fwd	CAAGTTGCTGCTTTACCCAGGGGCAGCCGAATTG
RIDA	K117R	Rev	CCTGGGTAAAGCAGCAACTTGGTAAGCAGCTCTA
RIDA	K117E	Fwd	GTTGCTGCTTTACCCGAGGGCAGCCGAATTG
RIDA	K117E	Rev	CTCGGGTAAAGCAGCAACTTGGTAAGCAGC
RIDA	K117Q	Fwd	GTTGCTGCTTTACCCCAGGGCAGCCGAATTG
RIDA	K117Q	Rev	CTGGGGTAAAGCAGCAACTTGGTAAGCAGC
RIDA	K117I	Fwd	GTTGCTGCTTTACCCATCGGCAGCCGAATTG
RIDA	K117I	Rev	GATGGGTAAAGCAGCAACTTGGTAAGCAGC
Ku70	K351R	Fwd	GATGCTCATGGGTTTCAGGCCGTTGGTACTGCTG
Ku70	K351R	Rev	CCTGAAACCCATGAGCATCAAACCTGGATCATC
IFIT1	K151R	Fwd	GGAAGGATGGGCCTTGCTGAGGTGTGGAGGAAAG
IFIT1	K151R	Rev	CCTCAGCAAGGCCCATCCTTCCTCACAGTCTATTTC
IFIT5	K150R	Fwd	GAAAGGCTGGGCACTCTTGAGGTTTGGAGGAAAG
IFIT5	K150R	Rev	CCTCAAGAGTGCCCAGCCTTTCTCACAGTCAGTC