Labeling preferences of diazirines with protein biomolecules

AUTHORS: Alexander V. West, Giovanni Muncipinto, Hung-Yi Wu, Andrew C. Huang, Matthew T. Labenski, Christina M. Woo

ABSTRACT:

Diazirines are widely used in photoaffinity labeling (PAL) to trap non-covalent interactions with biomolecules. However, design and interpretation of PAL experiments is challenging without a molecular understanding of the reactivity of diazirines with protein biomolecules. Here, we report a systematic evaluation of the labeling preferences of alkyl and aryl diazirines with individual amino acids, single proteins, and in the whole cell proteome. We find that aryl-fluorodiazirines react primarily through a carbene intermediate, while alkyl diazirines generate a reactive alkyl diazo intermediate on route to the carbene. The generation of a reactive diazo intermediate leads to preferential labeling of acidic amino acids in a pH-dependent manner. From a survey of 32 alkyl diazirine probes, we use this reactivity profile to rationalize why these probes preferentially enrich highly acidic proteins or those embedded in membranes and why probes with a net positive-charge tend to produce higher labeling yields. These results indicate that alkyl diazirines are an especially effective chemistry for surveying the membrane proteome, and will facilitate probe design and interpretation of biomolecular labeling experiments with diazirines.

Introduction

Since the development of photoaffinity labeling (PAL) in 1962,[1] PAL has emerged as a key approach for measuring biomolecular interactions, protein structure, and the identification of small molecule targets in the cell. The combination of PAL with high-

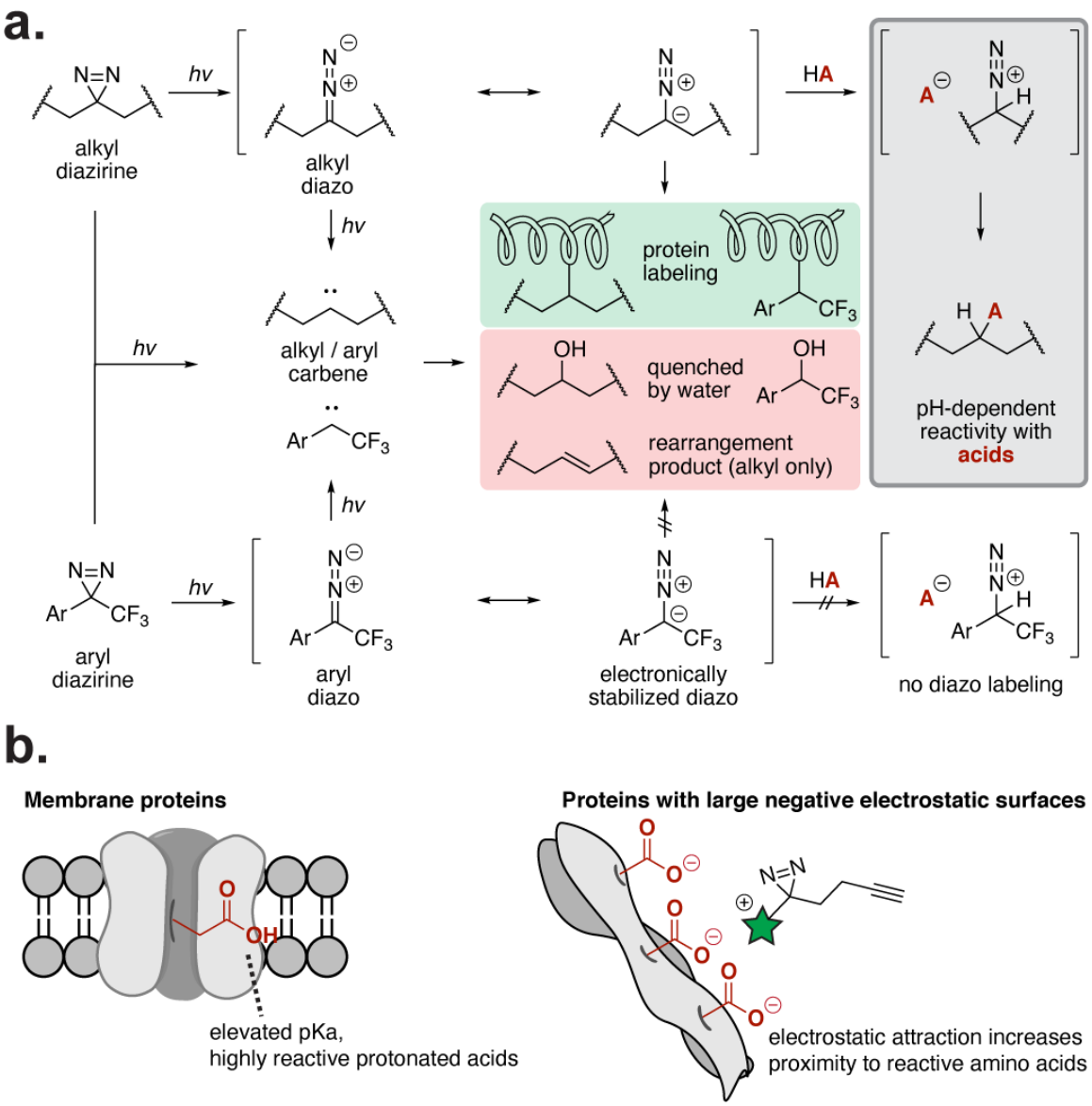

Fig. 1 | Overview of alkyl diazirine reactivity pathways. **(a)** Alkyl and aryl diazirines form carbene and diazo intermediates upon irradiation. Carbenes label nearby proteins, but are rapidly quenched if no protein substrate is nearby. Alkyl diazo intermediates react selectively with acids, while electronically stabilized aryl diazo intermediates do not. **(b)** The acid-selectivity of the alkyl diazo intermediates causes increased labeling of membrane proteins, which have more reactive protonated carboxylic acids, and proteins with large negative electrostatic surfaces. Labeling of these protein surfaces increases when the PAL probe (highlighted in green) is positively charged.

throughput sequencing,[2] mass spectrometry-based proteomics,[3,4] or binding site mapping[5] has significantly increased the depth of information measured from a single experiment. In particular, the design of small tags embedded with the alkyl diazirine as the

PAL functional group and a chemical enrichment handle has accelerated the use of PAL within large-scale biological measurements.[6-8]

However, despite the use of diazirines as probes to measure biomolecular interactions, the incomplete understanding of the labeling preferences of the diazirine limits the design and interpretation of these experiments. To overcome challenges in interpretation, the development of an diazirine "backgroundome"[9,10] or methods to differentiate between selective and non-selective labeling, including the use of enantioprobes,[11] or introduction of a scavenger,[12] have been implemented. During photolysis, the diazirine primarily isomerizes to a diazo intermediate that further converts to the carbene for biomolecular conjugation,[13,14] which will either rapidly form a covalent bond with a nearby biomolecule or be quenched by the aqueous environment (**Figure 1a**).[15] However, the generation of a reactive, electrophilic diazo intermediate during PAL may drive protein labeling through an alternate "pseudo-PAL" mechanism.[15] For example, stabilized diazo compounds exhibit preferential reactivity with organic acids[16] and selectively react with protein carboxylates to generate esters.[17] Evidence for alkyl diazirine reactivity patterns that transition through the diazo isomer to preferentially react with acidic residues has been recently noted using in vitro peptide or protein model systems.[18-20] Despite the investigation of reactivity preferences of other PAL functional groups (e.g., aryl azides,[21,22] benzophenones,[23,24] and aryl tetrazoles[25-27]) the chemical reactivity and resulting labeling preferences of the diazirine with amino acids has only been investigated in part on in vitro systems and peptide digests by mass spectrometry.[28] Systematic establishment of how alkyl diazirines exhibit differential reactivity preferences from aryl diazirines, and the implications this has on biomolecular labeling, would facilitate interpretation of the measurements and conclusions from PAL experiments using diazirines.

Here, we report a systematic study of alkyl and aryl diazirine labeling preferences with individual amino acids and single proteins, and use these data to interpret the protein binding partners visualized by a 32-membered alkyl diazirine library from the whole cell proteome. We demonstrate that the aryl-trifluorodiazirine primarily reacts through a carbene intermediate, while the alkyl diazirine generates a reactive diazo intermediate that underpins preferential reactivity with acidic amino acids in a pH-dependent manner. This chemistry rationalizes the preferential labeling of membrane proteins and proteins with negatively-charged electrostatic surfaces (**Figure 1b**). PAL ligands that are protonated at physiological pH, but not ligands that are neutral or deprotonated at physiological pH, possess affinity for and therefore heightened labeling of these protein surfaces, which is reflected by global protein profiles and binding site maps. This understanding of diazirine labeling preferences improves the ability to design and interpret the results from biological experiments using diazirines.

RESULTS

Alkyl and aryl diazirine reactivity with individual amino acids under neat conditions

We initiated a systematic evaluation of diazirine chemistry by evaluating the photolysis products of the alkyl diazirine[7] in aqueous conditions (20% CD_3CN–D_2O) and recovered a mixture of olefins (50%), a water-insertion product (35%), and a ketone (10–20%), presumably from reaction of the carbene with oxygen (**Figure S1**).[29] We next examined the reactivity of the Fmoc-protected alkyl diazirine **1** with individual amino acids (Ac-AA-OMe) under neat conditions to represent the closest possible binding interaction between the diazirine and a protein. Neat mixtures of the Fmoc-diazirine **1** with each Ac-AA-OMe (4 equiv) were irradiated using a broadband lamp (280–400 nm) and analyzed by liquid chromatography–mass spectrometry (LC-MS) (**Figure 2a**). The Fmoc-diazirine **1** covalently labeled relatively acidic polar amino acids with the highest yields (40–60%

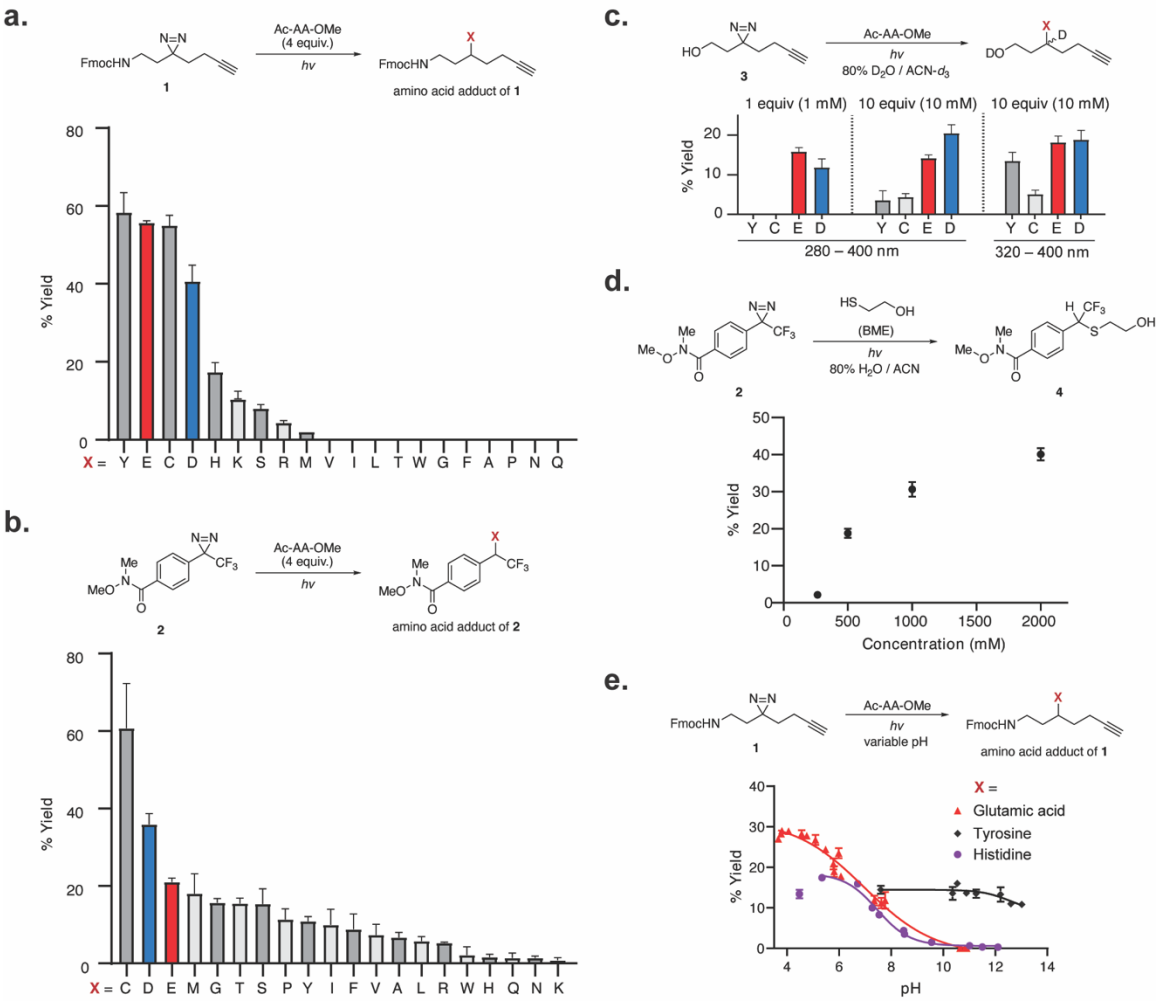

Fig. 2 | Reactivity of alkyl and aryl diazirines with the 20 natural amino acids. **(a)** Reactivity of alkyl diazirines with the 20 N-acetyl, O-Me protected amino acids (4 equiv) in neat conditions. Yields calculated against an internal standard by LC-MS. **(b)** Reactivity of aryl diazirines with the 20 amino acids (4 equiv) in neat conditions. Yields calculated against an internal standard by ^{19}F NMR. **(c)** Reactivity of alkyl diazirines with the 20 amino acids in solution (1 mM **3**, 1 mM or 10 mM amino acid) with 280–400 nm or 320–400 nm excitation. Only reactive amino acids are displayed. Yields calculated against an internal standard by ^{1}H NMR. **(d)** Reaction of aryl diazirine **2** with β-mercapto ethanol at varying concentrations in aqueous conditions. Yields calculated with relative integration by LC-MS. **(e)** Yield of alkyl diazirines with individual amino acids as a function of pH. Yields calculated against an internal standard by LC-MS. Plotted values show the mean with error bars representing the standard deviation with n = 3.

yield), followed by a subset of the additional polar amino acids (<20%), and had no observable labeling of the remaining polar amino acids or aliphatic amino acids (<3%). For comparison, the aryl diazirine **2** was evaluated against individual amino acids (4 equiv) under neat conditions by ^{19}F NMR. The aryl diazirine **2** produced insertion products with all 20 amino acids and afforded the highest yield with cysteine (**Figure 2b, S2**).

Alkyl and aryl diazirine reactivity in aqueous solution

We next evaluated the product distribution in aqueous conditions (80% D_2O–CD_3CN) at equimolar concentrations of the amino acid and the diazirine (1 mM). Under these conditions, we expected that the carbene intermediate generated by each probe would be rapidly quenched by D_2O before reacting with the amino acid residue. Surprisingly, photo-irradiation of the alkyl diazirine **3** at equimolar amounts of the individual amino acids (1 mM) under aqueous conditions revealed selective conjugation with Glu and Asp (10–15% yield, **Figure 2c, S3**). The insertion of the alkyl diazirine **3** to Tyr and Cys was additionally detected in aqueous solution at elevated concentrations (10 mM, 10 equiv). In order to investigate the possibility that the alkyl diazo intermediate **3** was inducing the observed reactivity for acidic residues, analogous to the reactivity of bona fide diazo reagents with acidic amino acid residues,[16] we filtered the broadband excitation wavelength to exclude wavelengths below 320 nm,[30] and thus slow the transition of the diazo to carbene at 300 nm. Indeed, filtration of the light source to a narrower range (320–400 nm) increased reactivity of the diazirine **3** with these acidic amino acid residues (**Figure 2c**).

By contrast, the aryl diazirine **2** formed no observed adducts with any amino acid up to 10 mM (10 equiv) in aqueous solution. To evaluate the limit of reactivity of the aryl diazirine **2** to conjugate with a molecule in aqueous solution, β-mercaptoethanol (BME) was used as a more soluble surrogate for the most-reactive amino acid, cysteine (**Figure 2d, S4**). The aryl diazirine **2** formed conjugated products with BME beginning at 250 mM concentrations (250 equiv). Notably, the aryl diazirine **2** formed comparable yields with BME at 500 mM as the alkyl diazirine **3** with 1 mM Glu. These data suggest that the aryl diazirine **2** primarily reacts through a short-lived carbene and that the aryl diazo

intermediate is not driving alternate reaction pathways under the photolysis conditions used here.

Alkyl diazirine labeling of amino acids is pH-dependent

Reaction of acidic residues with the alkyl diazirine through the diazo intermediate requires a proton transfer from the acid to give a cationic intermediate prior to displacement of dinitrogen by the conjugate base (**Figure 1a**). If the proton transfer is the rate limiting step, the insertion yield will therefore be dependent on the protonation state of the amino acid. The pH-dependence of alkyl diazirine labeling was investigated by titrating aqueous solutions of Glu, His, Tyr, and Lys (10 equiv) with acid or base before addition of the diazirine **1** (1 equiv) and filtered UV irradiation (320–400 nm). Product yields were measured by LC-MS. Glu, His, and Tyr produced pH-dependent insertion products with the alkyl diazirine **1**, while no observable insertion was observed with the fully deprotonated carboxylate and free base His (**Figure 2e**). Protonated Lys was also not reactive under these conditions, suggesting that the pKa of the conjugate acid is too high to react with the diazo intermediate generated by the diazirine **1**. In sum, these data suggest that the alkyl diazirine **1** generates a reactive intermediate that leads to preferential reactivity with protonated acidic residues in aqueous solution.

In the course of these studies, we additionally found that a significant fraction of the photolyzed diazirine underwent rapid internal rearrangement to a mixture of olefins, which may be a major rearrangement product from the carbene intermediate.[20] To minimize this elimination pathway, we drew on the kinetic isotope effect[31] and synthesized a deuterated alkyl diazirine (**Figure S5**). Photolysis of the tetra-deuterated diazirine afforded a greater yield of the ketone (42%) and reduced the yield of the olefin products (34%) as compared

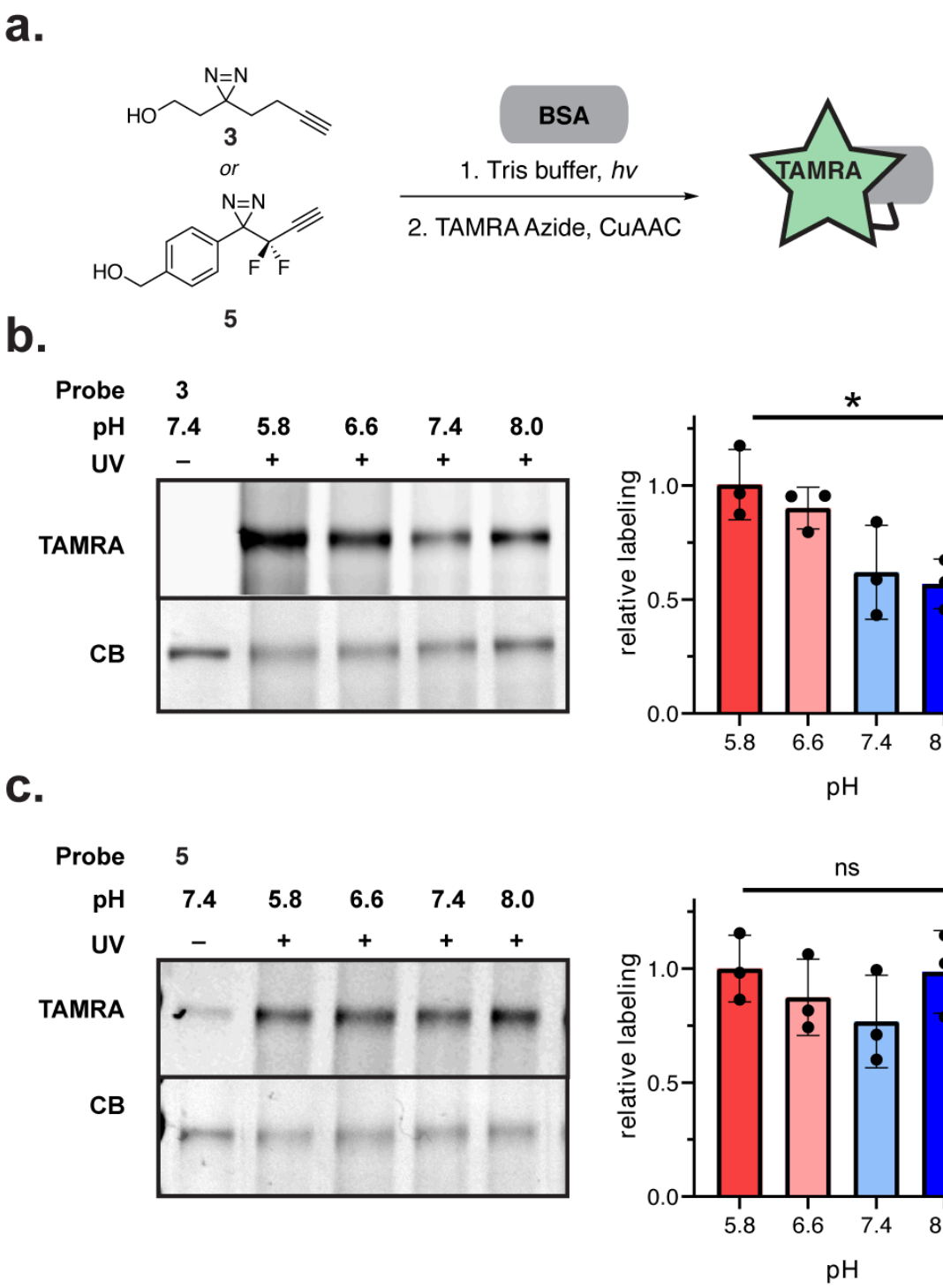

Fig. 3 | Photolabeling of single proteins as a function of pH. (a) Workflow schematic. Alkyl diazirine **3** or aryl-difluorodiazirine **5** were incubated with BSA in tris buffer with varying pH (5.8, 6.6, 7.4, 8.0). The diazirines were crosslinked to the protein by UV irradiation and visualized by in-gel fluorescence following click chemistry with TAMRA-N_3. **(b)** pH-dependent labeling of BSA by alkyl diazirine **3. (c)** pH-independent labeling of BSA by aryl diazirine **5**. Labeling yields of were calculated with TAMRA fluorescence signal normalized to coomassie blue (CB) stain signal. Plotted values show the mean with error bars representing the standard deviation with n = 3. Statistical significance was determined with a t-test between pH 5.8 and 8.0 samples (* = p-value < 0.05).

to the tetra-proton analogue. Thus, deuteration of the alkyl diazirine may improve covalent labeling yields in PAL experiments.

Alkyl diazirine labeling of single proteins is pH-dependent

To translate these observations to proteins, we next examined the labeling efficiency of alkyl and aryl diazirines[32] on an individual protein, bovine serum albumin (BSA), in vitro over a 5.8–8.0 pH range in tris buffer (**Figure 3a**). Analogous to the reactions with single amino acids, labeling from the alkyl diazirine **3** was higher at lower pH, while labeling from the aryldifluorodiazirine **5** was pH-independent over the range examined (**Figure 3b, 3c**). Interestingly, we found that performing a similar labeling experiment of BSA with the alkyl diazirine **3** in acidic phosphate buffer (pH = 5.8) resulted in concentration-dependent labeling of BSA by the alkyl diazirine **3**, presumably via competitive reactivity with the phosphate buffer (**Figure S7**). Indeed, we observed formation of a phosphate adduct with **3** in a pH-dependent manner analogous to the aqueous reactivity of **3** with the protected amino acids, suggesting that acidic phosphate was reacting with the diazo isomer of **3**. These data demonstrate that the differential reactive intermediates from the alkyl diazirine and the aryl diazirine predictably impacts observed labeling patterns on individual protein substrates.

PAL probes with a net positive charge increase alkyl diazirine labeling in the cell

The pH dependence of alkyl diazirine chemistry in vitro implies that charge state will play a significant role in the capture of protein substrates and may explain the inherent labeling efficiency of different alkyl diazirine probes.[11,33] To evaluate the influence of alkyl diazirine reactivity on PAL experiments in cells, we prepared a set of 32 alkyl diazirine probes with a distribution of net positive, negative, or neutral charges (representative structures in **Figure 4a**; for all structures see **Figure S8**). The PAL probes cover a diversity of small

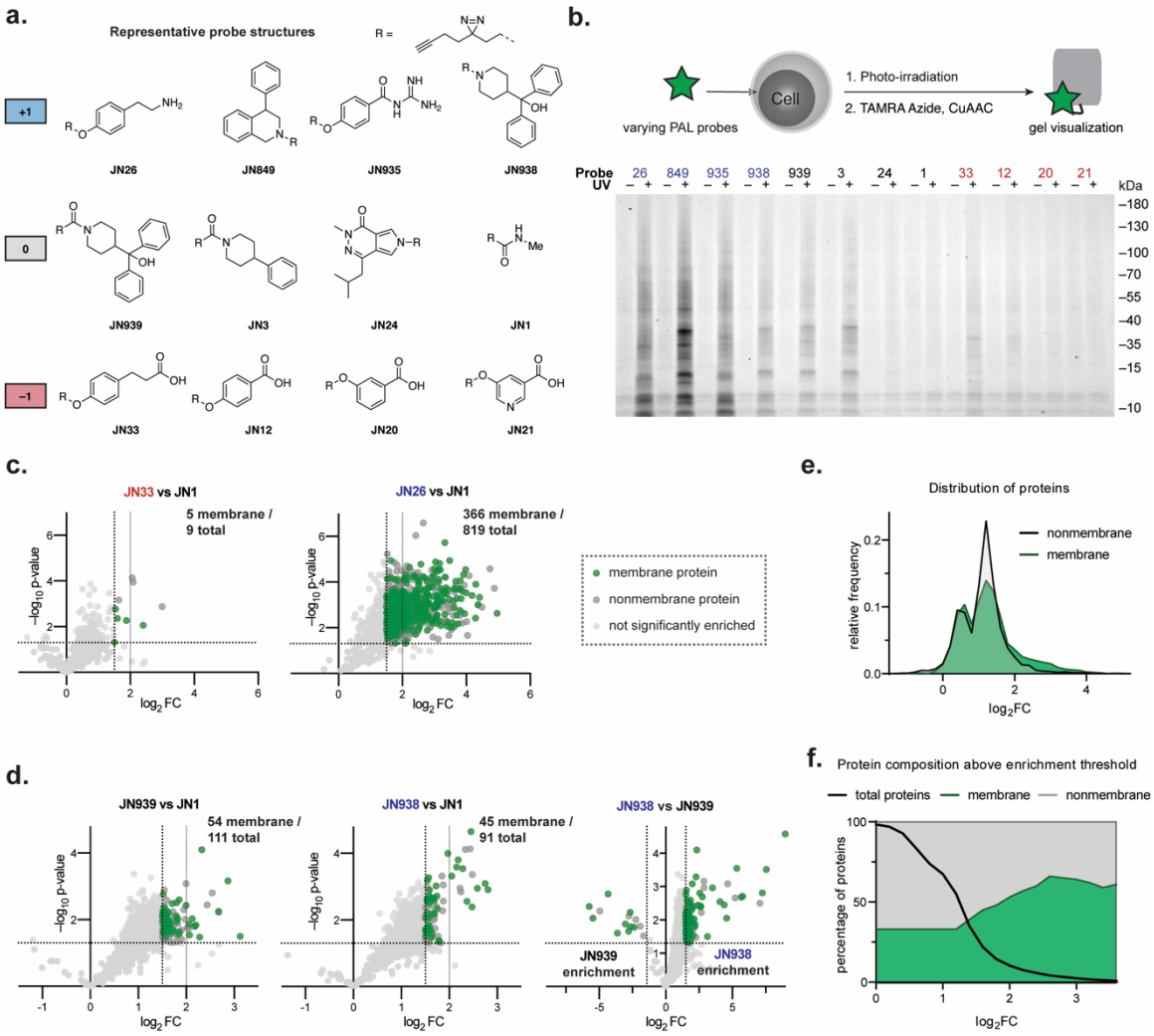

Fig. 4 | Whole proteome photolabeling in cells with PAL probes. **(a)** Representative probe structures (for a full list see **Figure S8**). **(b)** Photolabeling yields of compounds visualized by TAMRA signal. **(c)** Comparison of photolabeling profiles of carboxylic acid probe JN00033 and primary amine JN00026. Significantly enriched proteins are colored as membrane and nonmembrane proteins. **(d)** Photolabeling profiles of neutral amide probe JN939 and positively-charge amine analogue JN938. **(e)** Histogram showing the relative frequency of membrane and nonmembrane protein enrichments for all data in **(c)** and **(d).** **(f)** Percentage of membrane proteins above enrichment thresholds for all data in **(c)** and **(d).**

molecules including aromatic scaffolds (e.g., **JN3**, **JN939**, **JN849**, **JN935**), analogs of nucleotides (e.g., **JN846**, **JN847**, **JN936**), lipids (e.g., **JN38**), and other metabolites (e.g., **JN247**, **JN835**).

From this library, we selected representative probe structures based on their overall charge at physiological pH and evaluated their labeling efficiency by in-gel fluorescence. HEK293T cells were incubated with individual probes at 10 μM and photo-irradiated for 60 sec to crosslink the ligands to their protein binding partners. The cells were harvested, lysed, and visualized by in gel fluorescence following click chemistry with TAMRA-N_3. As expected, probe labeling efficiency generally correlated with the net charge of the probe (**Figure 4b**). The alkyl diazirine probes with a net positive charge exhibited higher labeling than neutral or negatively charged probes, which may reflect increased interaction with negatively-charged protein surfaces composed of acidic residues that react with alkyl diazirines (**Figure 1b**).

To directly examine the contribution of charge to the labeling efficiency in probe design, we compared the enriched proteome of two pairs of alkyl diazirine PAL probes with different net charge by chemical proteomics: first, a primary amine vs. a carboxylic acid (**JN26** vs. **JN33**), and second, a tertiary amine vs. an amide (**JN938** vs. **JN939**). SK-N-SH cells were incubated with individual PAL probes, photo-crosslinked, harvested, and lysed. Lysates were conjugated with biotin-azide using click chemistry and the biotinylated proteins were enriched with streptavidin-agarose. The enriched proteins were digested with trypsin for quantitative proteomics (TMT10plex). The positively-charged **JN26** exhibited a much higher degree of labeling compared to the negatively-charged analogue **JN33** (819 and 9 enriched proteins, respectively, **Figure 4c**). The effect from probe charge state was diminished but still noticeable when comparing neutral versus positively charged probes with larger, more hydrophobic structures. The neutral amide probe **JN939** significantly enriched slightly more proteins than the positively-charged amine **JN938** at a 3-fold threshold (111 and 91 enriched proteins, respectively), but the degree of enrichment was higher for the positive amine **JN938** than the neutral amide **JN939** when the enriched

proteins from each probe are compared directly (**JN938** vs **JN939**, **Figure 4d**). Taken together, these data imply that alkyl diazirine probes with a net positive charge will yield elevated labeling of the proteome, potentially due to increased localization to protein surfaces that are more readily crosslinked by alkyl diazirine chemistry, and that this effect is attenuated with larger hydrophobic probes.

PAL probes disproportionately label membrane proteins in cells

In line with prior reports,[5,10,11,33-35] we observed that alkyl diazirine probes possess an enrichment bias for membrane proteins (**Figure 4e**). While membrane proteins made up 33% of all proteins observed in the dataset, similar to the distribution of these proteins in the global proteome (27%),[36] membrane proteins were over-represented after filtering for protein enrichment: 46% of proteins that had at least a 3-fold enrichment, and 56% of proteins with at least a 4-fold enrichment were membrane proteins (**Figure 4f**). These membrane proteins derived from throughout the cell, including the cell surface, ER/golgi, mitochondrion, and nuclear membranes. The preferential labeling of membrane proteins by alkyl diazirine probes may reflect the elevated pKa of Glu and Asp residues in lipid bilayers.[37-39]

Understanding alkyl diazirine reactivity improves interpretation of binding site mapping results

To further evaluate the labeling trends in cells, the 32-membered alkyl diazirine probe library was divided into groups according to similarities in the probe charge and structure and dosed to SK-N-SH cells for protein and binding site mapping (**Figure 5a**). In brief, SK-N-SH cells were dosed in biological duplicate with the alkyl diazirine probe library in six groups of 5–6 compounds at 10 μM concentrations of each compound for 2 h. The treated cells were photo-irradiated for 60 sec, collected, and the cells were lysed. Labeled

proteins in the cellular lysates were tagged with an acid-cleavable picolyl biotin azide[40] for affinity enrichment and trypsin digestion on streptavidin–agarose beads. The labeled peptides representing binding interactions of the PAL probe with the protein target were subsequently eluted from the bead by acid cleavage and recovered for analysis by LC-MS. The probe-labeled peptide was assigned by database searching against the Swissprot human proteome, followed by filtration based on the embedded isotopic pattern,[41] and manual validation. A total of 632 binding sites from 170 proteins were mapped in SK-N-SH cells across 4,200 peptide spectral matches (PSMs).

Analysis of the mapped binding sites reflected the labeling trends observed on the proteome level. PSMs were used to approximate the relative abundance of labeling events with protein binding partners from each PAL probe (**Figure 5b**). Examination of the number of PSMs per probe revealed a bimodal distribution that clustered as a function of positively charged probes versus neutral or negatively charged probes (**Figure 5b**, inset). Positively charged probes produced on average 418 PSMs corresponding to 50 unique binding sites per probe whereas neutral probes produced an average of 66 PSMs (14 unique binding sites) and negatively-charge probes produced an average of 14 PSMs (5 unique sites). No statistical difference between the neutral or negatively charged PAL probes was observed, although neutral PAL probes with a larger hydrophobic surface area tended to yield a greater number of PSMs (e.g., **JN3**, **JN939**, **JN38**). Similar to observations from the proteome, we found that a majority of the binding sites were from membrane proteins (58% of PSMs) (**Figure 5c**). The wider structural diversity of the 32 PAL probes suggests the increased labeling of membrane proteins is a result of the labeling propensity of alkyl diazirine chemistry rather than the probes' individual structures.

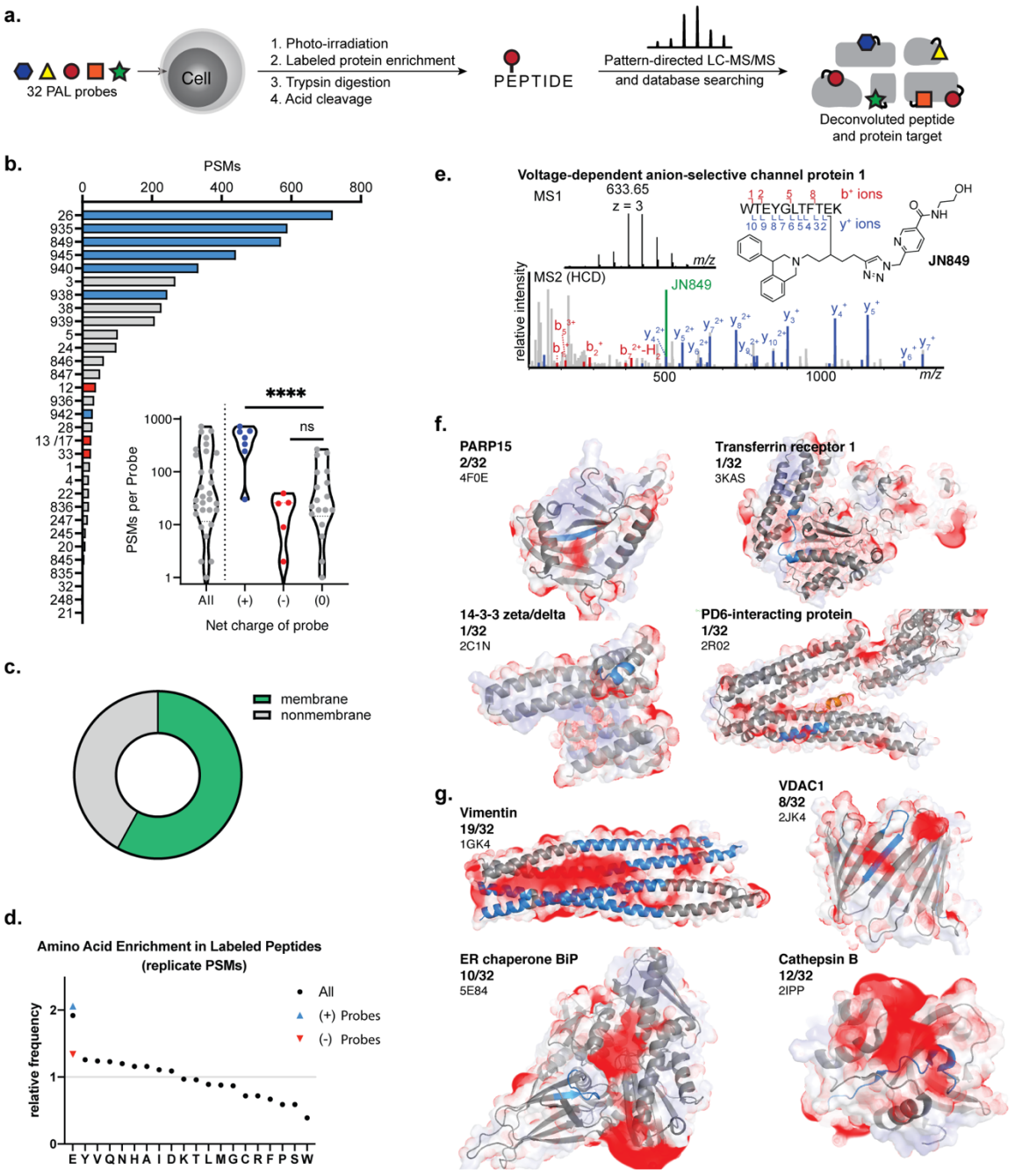

Fig. 5 | Binding site mapping 32 PAL probe library. (a) Cell cultures were treated simultaneously with 5 or 6 PAL probes which were photo-conjugated to protein binding partners. The conjugated proteins were enriched and the conjugated peptide representing the binding site of the PAL probe was isolated for analysis using isotope-targeted MS. **(b)** Count of peptide spectral matches (PSMs) assigned to each PAL probe. PAL probes with a net positive charge highlighted in blue, neutral probes in grey, and negatively-charged probes in red. **(c)** Comparison of binding site PSMs from membrane and nonmembrane proteins. **(d)** Amino acid frequency in the conjugated peptides relative to frequency in human proteome. **(e)** Annotated mass spectra and assignment of a VDAC1 binding site. **(f)** Negative electrostatic maps (red) overlaid on the conjugated peptide (blue) for unique binding interactions. **(g)** Negative electrostatic maps (red) overlaid on conjugated peptide (blue) for frequently conjugated proteins.

Although the positively charged members of the library were six of the top seven probes by PSM counting, the positively-charged probe **JN942** was an outlier at rank 21 of 32 probes. We investigated the cause of this outlier in vitro and observed unaffected labeling efficiencies on individual amino acids, and proteins (**Figure S11**), suggesting that in addition to the net charge, accessibility of the small molecule to biomolecular targets in a cellular context should also be considered in probe design.

As further evidence of labeling preferences between alkyl diazirines and acidic amino acid residues in the whole proteome, we compared the occurrence of amino acids found within labeled peptides to their natural occurrence in the human proteome (**Figure 5d**). Glutamic acid was the most enriched amino acid in labeled peptides, increased specifically in peptides labeled by positively charged probes, and most of this enrichment was eliminated in binding sites measured from negatively-charged probes. The limited enrichment of aspartate residues may be due to steric accessibility.[17] This suggests that for PAL experiments in cells, glutamic acid is labeled more-frequently than other amino acids and that the labeling frequency is influenced by the physical properties of the photoaffinity probe.

Across all the mapped probe-labeled binding sites, we found that a binding site on VDAC1 was frequently observed (representative assignment, **Figure 5e**). This peptide on VDAC1 is in close proximity to two major negatively charged residues on the protein (Y67, E73). VDAC1 E73 is notable for its unusually high pKa (predicted membrane pKa = 7.4) and ability to sensitively mediate dimerization of VDAC1 in a pH-dependent manner.[42] Protonation of VDAC1 E73 would increase the propensity for E73 to react with a diazo intermediate. VDAC1 and VDAC2, but not VDAC3, possess this glutamic acid residue,

which explains why only VDAC1 and VDAC2 are frequently enriched targets in alkyl diazirine studies.[10]

Intrigued by this analysis with VDAC1, we next visualized a subset of the binding sites onto available protein structures (binding site highlighted in dark blue, **Figure 5f, 5g**). The electrostatic potential of each protein surface was calculated using the adaptive Poisson-Boltzmann solver (APBS) and overlaid on the structure of the protein. For clarity, only the negative electrostatic surface map is displayed in red. For binding sites that were observed as labeled by only one or two PAL probes, the mapped peptide has a minor degree of negative electrostatic density nearby (**Figure 5f**). In some cases, the mapped peptide is within 9 Å[18] in proximity to a known binding site for the natural protein ligand. For example, ADP-ribosyltransferase protein PARP15 has a nicotinamide binding cleft that is adjacent to the observed binding site.[43] PD6-interacting protein (PDCD6IP) binds to a natural peptide ligand in close proximity to the observed small molecule binding site (peptide ligand highlighted in orange, **Figure 5f**).[44] The 14-3-3 proteins are receptors for peptides in a cleft within 9 Å of the mapped binding site.[45] Transferrin receptor 1 is a membrane-bound protein; the binding site is in the extracellular region.[46]

In contrast, binding sites on proteins that were mapped to multiple PAL probes exhibited significantly greater negative electrostatic densities around the labeled peptide (highlighted in dark blue, **Figure 5g**). Vimentin is a filamentous protein that is frequently identified in alkyl diazirine labeling experiments, likely due to the significant negative electrostatic density on vimentin (net charge = –27). Unsurprisingly, the 17 unique peptides conjugated across 19 different PAL probes identified from vimentin colocalize with the negative electrostatic density map. Likewise, 8–12 PAL probes conjugated to VDAC1, ER chaperone BiP, and cathepsin B, which each display similar negative

electrostatic densities in close proximity to the mapped peptide binding site. Taken together, these data point to the subset of proteins that are labeled more frequently by alkyl diazirine chemistry typically have negative electrostatic regions that are characterized by a density of acidic residues or those with a relatively high localized pKa, resulting in an increase in the degree of protonation of the amino acid for labeling through the diazo intermediate during PAL experiments.

DISCUSSION

Despite the growing use and application of diazirine chemistry in PAL experiments a systematic understanding of the labeling preferences of the diazirine in cells has yet to emerge. Here, we report a systematic analysis of diazirines and their preferential reactivity pattern with protein biomolecules in vitro and in cells. We find that alkyl diazirines preferentially react with acidic amino acid residues in a pH-dependent manner in vitro, which can be rationalized by the formation of a long-lived diazo intermediate that is intercepted by organic acids prior to formation of an alkyl carbene. By contrast, the aryl-trifluorodiazirine forms insertion products with all 20 amino acids, is less dependent on pH, and the labeling chemistry is readily quenched by water, which are characteristics of a reaction pathway through a carbene intermediate. Although reactions with individual amino acids in neat conditions is only an approximation of the PAL probe reactivity with a target protein in cells, the impact of the alkyl diazirine reactivity pattern translates to a predictable enhancement of cellular labeling with alkyl diazirine probes carrying a net positive charge from a library of 32 alkyl diazirine probes, which may arise from enhanced association of the probe with the matched acidic regions on protein surfaces. In line with this expectation, chemoproteomics and binding site data show that these probes

preferentially enrich proteins with negative electrostatic surfaces or proteins embedded in membranes, which may be a consequence of the predicted elevation of pKa of acidic residues in the hydrophobic membrane environment.[39]

These data assist with the design and interpretation of PAL experiments using alkyl diazirine chemistry. Design of the probe can be tuned to enhance or reduce protein capture simply by altering the net overall charge of the probe molecule. Alkyl diazirine probes with a net neutral or negative charge may selectively visualize stronger binding interactions, while probes with a net positive charge may capture more transient interactions in the proteome. Alkyl diazirine chemistry should also be selected if membrane proteins or those with large acidic patches are desired targets. The preference of alkyl diazirine chemistry for these protein types is reflected herein and on retrospective analysis of ours and other studies.[5,10,11,33-35,40,47] This enhances the utility of diazirine chemistry in probes for visualizing binding over a range of affinities with membrane proteins, which are typically challenging proteins to target.

Although these results point to a significant contribution of alkyl diazirine PAL through the diazo intermediate, dissection of the exact ratio of labeling from the carbene and the diazo intermediate remains to be fully elucidated. Theoretically, products resulting from the carbene are more likely to arise from tight binding interactions with a lower binding constant, since the carbene is a short-lived intermediate that is rapidly quenched in aqueous media.[14,48] Conversely, products resulting from labeling through the diazo are more-likely to represent weaker binding interactions, since the lifetime of the diazo is above the rate of diffusion.[49] There are relatively few chemistries that have been used to selectively label carboxylic acids in whole cells,[50-52] and fewer photoactivatable chemistries,[27] suggesting that this chemistry could be intentionally utilized to profile

carboxylic acids in cells as a method that senses the protonation state of the acids. Additionally, once a carbene intermediate is generated, the contribution from singlet or triplet states is another open question that may affect the amino acid labeling distribution, since the carbene is formed as a singlet, but can relax to the triplet state.[53-57] In addition, overall labeling trends of diazirine probes are also impacted by other factors, such as overall hydrophobicity of the probe or cell permeability, as observed with **JN942**, which must be additionally considered when interpreting data from PAL experiments.

In conclusion, the alkyl diazirine preferentially reacts with acidic amino acid residues due to the major contribution of a reactive diazo intermediate, while the aryl-trifluoro diazirine possesses a reactivity pattern in line with a reactive carbene intermediate. These reactivity patterns are consistent across individual amino acids, single proteins, and the whole proteome, and assist with the design and interpretation of PAL experiments using alkyl diazirine probes. Understanding the reactivity pattern of alkyl diazirines therefore facilitates design and interpretation of PAL experiments. Given these reactivity considerations for the alkyl diazirine, there is opportunity for innovation of new PAL chemistries with different labeling preferences. The design of new chemistries and the systematic examination of their reactivity patterns may provide additional opportunities for generation of an “ideal” PAL functional group or development of methods to selectively map non-covalent interactions with certain regions of the cell or types or proteins.

METHODS

For a full description of the methods used in this study, see the Supporting Information

DATA AVAILABILITY

All proteomics data are available within this paper, in the Supporting Information, or the Supplementary Tables.

AFFILIATIONS

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, 02138

Alexander V. West, Hung-Yi Wu, Christina M. Woo

Jnana Therapeutics, Boston, MA, 02210

Giovanni Muncipinto, Matthew T. Labenski

KAYAK, Cambridge, MA, 02142

Andrew C. Huang

Corresponding author

Correspondence to Christina M. Woo

CONTRIBUTIONS

A.V.W. performed all experiments and synthesized all compounds unless specified otherwise. G.M. synthesized the 32 PAL probes. H.Y.W. performed the binding site proteomics experiments. A.V.W., C.M.W., and M.L. designed chemoproteomic experiments and analyzed the data. A.V.W., C.M.W., and A.C.H. analyzed the binding site data. C.M.W. oversaw the project. A.V.W. wrote the manuscript and all authors revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests

ACKNOWLEDGEMENTS

We would like to thank H. Flaxman, N. Vallavoju, C. Chang, and A. D'Souza for their advice and helpful discussions, and M. Parsons for his assistance. Mass spectrometry data was collected at the Harvard University Mass Spectrometry and Proteomics Resource Laboratory (B. Budnik), and NMR data collected at the Laukin-Purcell Instrumentation Center (D. Cui). Support from NIH NIDA (DP1DA046586), Burroughs Welcome Fund, Sloan Foundation, Camille and Henry Dreyfus Foundation, and Harvard University are gratefully acknowledged.

REFERENCES.

1 Singh, A., Thornton, E. R. & Westheimer, F. H. The Photolysis of Diazoacetylchymotrypsin. *J. Biol. Chem.* **237**, PC3006-PC3008, (1962).

2 Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* **58**, 586-597, (2015).

3 Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C. & Yates, J. R., 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **113**, 2343-2394, (2013).

4 Pappireddi, N., Martin, L. & Wühr, M. A Review on Quantitative Multiplexed Proteomics. *ChemBioChem* **20**, 1210-1224, (2019).

5 Gao, J., Mfuh, A., Amako, Y. & Woo, C. M. Small Molecule Interactome Mapping by Photoaffinity Labeling Reveals Binding Site Hotspots for the NSAIDs. *J. Am. Chem. Soc.* **140**, 4259-4268, (2018).

6 Pan, S. *et al.* A Suite of "Minimalist" Photo-Crosslinkers for Live-Cell Imaging and Chemical Proteomics: Case Study with BRD4 Inhibitors. *Angew. Chem.* **129**, 11978-11983, (2017).

7 Li, Z. *et al.* Design and Synthesis of Minimalist Terminal Alkyne-Containing Diazirine Photo-Crosslinkers and Their Incorporation into Kinase Inhibitors for Cell- and Tissue-Based Proteome Profiling. *Angew. Chem. Int. Ed.* **52**, 8551-8556, (2013).

8 Chang, C.-F., Mfuh, A., Gao, J., Wu, H.-Y. & Woo, C. M. Synthesis of an electronically-tuned minimally interfering alkynyl photo-affinity label to measure small molecule–protein interactions. *Tetrahedron* **74**, 3273–3277, (2018).

9 Bayley, H. *Photogenerated Reagents in Biochemistry and Molecular Biology*. (Elsevier, 1983).

10 Kleiner, P., Heydenreuter, W., Stahl, M., Korotkov, V. S. & Sieber, S. A. A Whole Proteome Inventory of Background Photocrosslinker Binding. *Angew. Chem. Int. Ed.* **56**, 1396-1401, (2017).

11 Wang, Y. *et al.* Expedited mapping of the ligandable proteome using fully functionalized enantiomeric probe pairs. *Nat. Chem.* **11**, 1113-1123, (2019).

12 Ruoho, A. E., Kiefer, H., Roeder, P. E. & Singer, S. J. The Mechanism of Photoaffinity Labeling. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2567-2571, (1973).

13 Platz, M. S. in *J. Org. Chem.* Vol. 79 2341-2353 (American Chemical Society, 2014).

14 Zhang, Y. *et al.* Study of the S1 Excited State of para-Methoxy-3-phenyl-3-methyl Diazirine by Ultrafast Time Resolved UV−Vis and IR Spectroscopies and Theory. *J. Am. Chem. Soc.* **131**, 13784-13790, (2009).

15 Westheimer, F. H. Photoaffinity Labeling-Retrospect and Prospect. *Ann. N. Y. Acad. Sci.* **346**, 134-143, (1980).

16 Roberts, J. D. & Watanabe, W. The Kinetics and Mechanism of the Acid-Catalyzed Reaction of Diphenyldiazomethane with Ethyl Alcohol1,2. *J. Am. Chem. Soc.* **72**, 4869-4879, (1950).

17 Mix, K. A., Lomax, J. E. & Raines, R. T. Cytosolic Delivery of Proteins by Bioreversible Esterification. *J. Am. Chem. Soc.* **139**, 14396-14398, (2017).

18 Flaxman, H. A., Chang, C. F., Wu, H. Y., Nakamoto, C. H. & Woo, C. M. A Binding Site Hotspot Map of the FKBP12-Rapamycin-FRB Ternary Complex by Photoaffinity Labeling and Mass Spectrometry-Based Proteomics. *J. Am. Chem. Soc.* **141**, 11759-11764, (2019).

19 Iacobucci, C. *et al.* Carboxyl-Photo-Reactive MS-Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents. *Anal. Chem.* **90**, 2805-2809, (2018).

20 O'Brien, J. G. K., Jemas, A., Asare-Okai, P. N., am Ende, C. W. & Fox, J. M. Probing the Mechanism of Photoaffinity Labeling by Dialkyldiazirines through Bioorthogonal Capture of Diazoalkanes. *Org. Lett.*, (2020).

21 Smith, R. A. G. & Knowles, J. R. Aryldiazirines. Potential Reagents for Photolabeling of Biological Receptor Sites. *J. Amer. Chem. Soc.* **95**, (1972).

22 Tanaka, Y., Bond, M. R. & Kohler, J. J. Photocrosslinkers illuminate interactions in living cells. *Mol. Biosyst.* **4**, 473-480, (2008).

23 Wittelsberger, A., Thomas, B. E., Mierke, D. F. & Rosenblatt, M. Methionine acts as a "magnet" in photoaffinity crosslinking experiments. *FEBS Lett.* **580**, 1872-1876, (2006).

24 Deseke, E., Nakatani, Y. & Ourisson, G. Intrinsic Reactivities of Amino Acids towards Photoalkylation with Benzophenone − A Study Preliminary to Photolabelling of the Transmembrane Protein Glycophorin A. *Eur. J. Org. Chem.* **1998**, 243-251, (1998).

25 Cheng, K. *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).

26 Tian, Y. *et al.* Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. *J. Am. Chem. Soc.* **139**, 6078-6081, (2017).

27 Bach, K., Beerkens, B. L. H., Zanon, P. R. A. & Hacker, S. M. 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).

28 Ziemianowicz, D. S., Bomgarden, R., Etienne, C. & Schriemer, D. C. Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines. *J. Am. Soc. Mass. Spectrom.* **28**, 2011-2021, (2017).

29 Nielsen, P. E. & Buchardt, O. Aryl Azides as Photoaffinity Labels. A Photochemical Study of some 4-substituted Aryl Azides. *Photochem. Photobiol.* **35**, 317-323, (1982).

30 Pezacki, J. P. *et al.* Laser Flash and Dual Wavelength Photolysis of 3,4-Diaza-2,2-dimethoxy-1-oxa[4.5]spirooct-3-ene. Migration of Hydrogen and Carbon in Cyclobutylidene and in the Excited State of Its Precursor. *J. Am. Chem. Soc.* **119**, 3191-3192, (1997).

31 Albu, T. V. *et al.* Dynamics of 1,2-Hydrogen Migration in Carbenes and Ring Expansion in Cyclopropylcarbenes. *J. Phys. Chem. A* **106**, 5323-5338, (2002).

32 Kumar, N. S. & Young, R. N. Design and synthesis of an all-in-one 3-(1,1-difluoroprop-2-ynyl)-3H-diazirin-3-yl functional group for photo-affinity labeling. *Bioorg. Med. Chem.* **17**, 5388-5395, (2009).

33 Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527-541.e529, (2017).

34 Parker, C. G. *et al.* Chemical Proteomics Identifies SLC25A20 as a Functional Target of the Ingenol Class of Actinic Keratosis Drugs. *ACS Cent. Sci.* **3**, 1276-1285, (2017).

35 Niphakis, Micah J. *et al.* A Global Map of Lipid-Binding Proteins and Their Ligandability in Cells. *Cell* **161**, 1668-1680, (2015).

36 Almén, M. S., Nordström, K. J. V., Fredriksson, R. & Schiöth, H. B. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7**, 50-50, (2009).

37 Panahi, A. & Brooks, C. L., 3rd. Membrane environment modulates the pKa values of transmembrane helices. *J. Phys. Chem.* **119**, 4601-4607, (2015).

38 Marcoline, Frank V., Bethel, N., Guerriero, Christopher J., Brodsky, Jeffrey L. & Grabe, M. Membrane Protein Properties Revealed through Data-Rich Electrostatics Calculations. *Structure* **23**, 1526-1537, (2015).

39 Teixeira, V. H., Vila-Viçosa, D., Reis, P. B. P. S. & Machuqueiro, M. pKa Values of Titrable Amino Acids at the Water/Membrane Interface. *J. Chem. Theory Comput.* **12**, 930-934, (2016).

40 Miyamoto, D. K., Flaxman, H. A., Wu, H.-Y., Gao, J. & Woo, C. M. Discovery of a Celecoxib Binding Site on Prostaglandin E Synthase (PTGES) with a Cleavable Chelation-Assisted Biotin Probe. *ACS Chem. Biol.* **14**, 2527-2532, (2019).

41 Woo, C. M. *et al.* Development of IsoTaG, a Chemical Glycoproteomics Technique for Profiling Intact N- and O-Glycopeptides from Whole Cell Proteomes. *J. Proteome Res.* **16**, 1706-1718, (2017).

42 Bergdoll, L. A. *et al.* Protonation state of glutamate 73 regulates the formation of a specific dimeric association of mVDAC1. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E172-E179, (2018).

43 Andersson, C. D. *et al.* Discovery of Ligands for ADP-Ribosyltransferases via Docking-Based Virtual Screening. *J. Med. Chem.* **55**, 7706-7718, (2012).

44 Zhai, Q. *et al.* Structural and functional studies of ALIX interactions with YPXnL late domains of HIV-1 and EIAV. *Nat. Struct. Mol. Biol.* **15**, 43-49, (2008).

45 Macdonald, N. *et al.* Molecular Basis for the Recognition of Phosphorylated and Phosphoacetylated Histone H3 by 14-3-3. *Mol. Cell* **20**, 199-211, (2005).

46 Abraham, J., Corbett, K. D., Farzan, M., Choe, H. & Harrison, S. C. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. *Nat. Struct. Mol. Biol.* **17**, 438-444, (2010).

47 Gertsik, N. *et al.* Mapping the Binding Site of BMS-708163 on gamma-Secretase with Cleavable Photoprobes. *Cell Chem. Biol.* **24**, 3-8, (2017).

48 Modarelli, D. A., Morgan, S. & Platz, M. S. Carbene formation, hydrogen migration, and fluorescence in the excited states of dialkyldiazirines. *J. Am. Chem. Soc.* **114**, 7034-7041, (1992).

49 Procacci, B., Roy, S. S., Norcott, P., Turner, N. & Duckett, S. B. Unlocking a Diazirine Long-Lived Nuclear Singlet State via Photochemistry: NMR Detection and Lifetime of an Unstabilized Diazo-Compound. *J. Am. Chem. Soc.* **140**, 16855-16864, (2018).

50 Mix, K. A. & Raines, R. T. Optimized Diazo Scaffold for Protein Esterification. *Org. Lett.* **17**, 2358-2361, (2015).

51 Qian, Y., Schürmann, M., Janning, P., Hedberg, C. & Waldmann, H. Activity-Based Proteome Profiling Probes Based on Woodward's Reagent K with Distinct Target Selectivity. *Angew. Chem. Int. Ed.* **55**, 7766-7771, (2016).

52 Martín-Gago, P. *et al.* Covalent Protein Labeling at Glutamic Acids. *Cell Chem. Biol.* **24**, 589-597.e585, (2017).

53 Arenas, J. F., López-Tocón, I., Otero, J. C. & Soto, J. Carbene Formation in Its Lower Singlet State from Photoexcited 3H-Diazirine or Diazomethane. A Combined CASPT2 and ab Initio Direct Dynamics Trajectory Study. *J. Am. Chem. Soc.* **124**, 1728-1735, (2002).

54 Geise, C. M. *et al.* Computational and Experimental Studies of the Effect of Substituents on the Singlet−Triplet Energy Gap in Phenyl(carbomethoxy)carbene. *J. Org. Chem.* **67**, 3079-3088, (2002).

55 Matzinger, S., Bally, T., Patterson, E. V. & McMahon, R. J. The C7H6 Potential Energy Surface Revisited: Relative Energies and IR Assignment. *J. Am. Chem. Soc.* **118**, 1535-1542, (1996).

56 Admasu, A. *et al.* A laser flash photolysis study of p-tolyl(trifluoromethyl)carbene. *J. Chem. Soc., Perkin Trans. 2*, 1093-1100, (1998).

57 Hirai, K., Itoh, T. & Tomioka, H. Persistent Triplet Carbenes. *Chem. Rev.* **109**, 3275-3332, (2009).