RESEARCH ARTICLE

Amino Acid Insertion Frequencies Arising from Photoproducts Generated Using Aliphatic Diazirines

Daniel S. Ziemianowicz,[1] Ryan Bomgarden,[2] Chris Etienne,[2] David C. Schriemer[1]

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB T2N 4N1, Canada
[2]Thermo Fisher Scientific, 3747 N. Meridian Rd., Rockford, IL 61101, USA
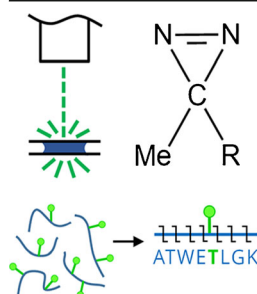
Abstract. Mapping proteins with chemical reagents and mass spectrometry can generate a measure of accessible surface area, which in turn can be used to support the modeling and refinement of protein structures. Photolytically generated carbenes are a promising class of reagent for this purpose. Substituent effects appear to influence surface mapping properties, allowing for a useful measure of design control. However, to use carbene labeling data in a quantitative manner for modeling activities, we require a better understanding of their inherent amino acid reactivity, so that incorporation data can be normalized. The current study presents an analysis of the amino acid insertion frequency of aliphatic carbenes generated by the photolysis of three different diazirines: 3,3'-azibutyl-1-ammonium, 3,3'-azibutan-1-ol, and 4,4'-azipentan-1-oate. Leveraging an improved photolysis system for single-shot labeling of sub-microliter frozen samples, we used EThCD to localize insertion products in a large population of labeled peptides. Counting statistics were drawn from data-dependent LC-MS^2 experiments and used to estimate the frequencies of insertion as a function of amino acid. We observed labeling of all 20 amino acids over a remarkably narrow range of insertion frequencies. However, the nature of the substituent could influence relative insertion frequencies, within a general preference for larger polar amino acids. We confirm a large (6-fold) increase in labeling yield when carbenes were photogenerated in the solid phase (77 K) relative to the liquid phase (293 K), and we suggest that carbene labeling should always be conducted in the frozen state to avoid information loss in surface mapping experiments.

Keywords: Covalent labeling mass spectrometry, Diazirine, Carbene, Footprinting, EThCD

Introduction

Protein function is driven by structure and the conformational states that are sampled by structure. However, to determine exactly how a particular protein functions, we need methods that can generate structural data for the protein in its working state, which always involves interactions with other biomolecules. Proteins are influenced by interactions over a wide range of complexities and timescales, and more methods are needed that can track structural and conformational adaptations in complex states. Simple systems can be studied in reconstitution experiments using a range of spectroscopic, imaging, and crystallographic techniques, but higher-order cellular processes involve many interacting protein networks and complexes, and the regulatory properties of a given protein can only be understood by measuring structure-function in the cell, over a wide timescale.

Mass spectrometry has contributed much towards a coarse mapping of protein interactions and networks [1, 2], and a number of methodological extensions look to add structural resolution (cross-linking, XL-MS) [3–5] and conformational analysis (hydrogen exchange, HX-MS) [6–8] at the same scale. Footprinting methods for higher resolution surface/interface mapping have made less of an impact. Covalent labeling (CL-MS) [9], using reactive probes to identify the chemically-accessible surface areas (CASA) of a protein, can generate a form of topographical map very useful for protein characterization. In the context of intermolecular interactions, this can support the localization of binding surfaces [10–12]. Label data can even constrain 3D protein structure prediction

Correspondence to: David Schriemer; *e-mail:* dschriem@ucalgary.ca

[13]. Pursuing the high-resolution potential of CL-MS has exposed some of the challenges of chemical methods, which are likely at the root of its limited use in the analysis of complex protein systems.

The concept of sampling an interaction footprint makes sense when imagining static protein structures, but as proteins exist as populations of interconverting conformational states that are easily perturbed [14, 15], the validity of detecting a footprint using slow and selective chemistries breaks down. If labeling chemistries are selective and the insertion kinetics are slow relative to the timescales of conformational interconversions, the population being sampled can become progressively skewed in the pursuit of generating enough labeling events to detect by MS. In the extreme, the true equilibrium population can be completely destroyed and the labeling data rendered meaningless.

Nevertheless, bioconjugation strategies are rich and varied, with many options to consider when seeking alternative, higher-performing reagents. Selective reagents provide limited coverage and thus low spatial resolution, requiring surface sampling with multiple compounds for adequate mapping. For example, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) reacts with the carboxylic acid group of glutamate and aspartate and can be combined with organic acid anhydrides or N-hydroxysuccinimide derivatives that react with the amine group of lysine for broader coverage [9, 16]. Less selective labeling methods, such as radical-mediated oxidative labeling, extends the number of side chain chemistries that may be modified, using hydroxyl radicals generated either by radiolysis of H_2O with high-energy X-rays or photolysis of H_2O_2 by a 248 nm laser [17]. Fast photochemical oxidation of proteins (FPOP) is a useful variant that has been used to profile protein folding kinetics on the order of microseconds to hundreds of milliseconds [18, 19]. However, a recent analysis of reaction kinetics shows that radicals can persist for tens of milliseconds after initiation. Attempts to scavenge persistent radicals with glutamine was found to generate metastable secondary radicals that can further extend oxidation reactions [20].

Reaction cascades with end products that are indistinguishable from the primary labeling event are serious complications for radical-based oxidative methods, especially when harsh methods are used for the initiation of reactivity (i.e., high-energy X-rays, H_2O_2, 248 nm). Alternative photolytic processes could offer improvements over oxidative labeling. One such process involves photoconversion of a diazirine to a carbene, a highly-reactive diradical carbon center generated by irradiation at ~310–350 nm [21], a less damaging wavelength for protein states. Carbenes have short lifetimes, on the order of tens of nanoseconds [22], and could permit fast-sampling of an equilibrated protein state. The products of the reaction involve insertions that can generate mass shifts based on the choice of reagent [23]. Carbenes appear relatively nonspecific with respect to substrates: singlet carbenes preferentially insert in O–H, N–H, and S–H bonds but can also insert across C–C and C–H bonds [24]. Historically, carbenes have been incorporated in photoaffinity probes to explore biomolecular interactions such

as receptor-drug binding events [25]. The structure of membranes and associated proteins has also been investigated with diazo- and diazirine-substituted fatty acids [26]. More recently, diazirine precursors have been incorporated into cross-linkers [27, 28] and covalent labeling agents [23, 29, 30]. "Carbene footprinting" has been successfully demonstrated using gaseous diazirine [30], simple aliphatic diazirines [23, 31, 32], and most recently aromatic diazirine derivatives [29]. Carbene labeling appears to offer the hallmarks of an excellent, adaptable labeling tool. We explored the effect of diazirine substituents on protein insertion product distribution and observed differences in the localization of the carbene, arising from the diazirine "scaffold," which steers the solvation properties of the reagent prior to photoconversion [23]. We also observed an improved correlation between insertion yield and CASA, when photolysis was performed on frozen samples.

However, to improve upon our quantitative use of the data for structural purposes, we require a better understanding of amino acid labeling preferences, devoid of higher-order structural influences. Such data would be very useful as a normalization control. Some efforts have been made towards understanding the reactivity of various carbenes towards a variety of substrates [33–36], but the side chain insertion preferences for carbenes have not been explored in sufficient depth [24]. In this study, we characterized the insertion preferences for a set of simple aliphatic diazirine labeling reagents in the solid state, and sought to determine if the nature of the diazirine scaffold has as strong a targeting effect in primary structure as we observed in higher-order structure. We also further explored the intriguing influence of temperature on labeling events.

Methods

Peptide Sample Preparation

Separate aliquots of 200 μM BSA, trypsin, and equine myoglobin (Sigma Aldrich, St. Louis, MO, USA) were digested with porcine pepsin (Sigma Aldrich) at pH 2.5 in 0.1 M glycine HCl for 60 min, 37 °C at an enzyme:substrate ratio of 1:100. Digestion was partially limited to allow for the generation of mid-length peptides for the benefit of efficient ETD fragmentation. Digestion was quenched and pepsin inactivated with 2 M NaOH to pH 7.5. Peptides were extracted with C18 ZipTips (Merck Millipore, Darmstadt, Germany) and eluted using 50% acetonitrile (LC-MS grade; Thermo Scientific, San Jose, CA, USA) with 0.1% trifluoroacetic acid (HPLC grade; Merck Millipore). Eluent was evaporated and peptides resuspended in a labeling buffer consisting of 10 mM NaCl (Biotechnology grade, Amresco, Solon, OH, USA) and 50 mM sodium phosphate (>99.0%; Sigma Aldrich), pH 7.4.

Photolytic Labeling

Diazirine labeling reagents were synthesized using methods previously reported. 4,4′-Azipentan-oate and 3,3′-azibutan-1-ol where prepared using the method of Church et al. [37]. 3,3′-

Azibutyl-1-ammonium was prepared from 3,3′-azibutan-1-ol by conversion to 3-(2-iodo-ethyl)-3-methyl-3H-diazirine using triphenylphosphine, iodine, and imidazole. Subsequent treatment with sodium azide followed by triphenylphosphine gave 3,3′-azibutyl-1-ammonium [38].

The sample for irradiation consisted of 10 µM peptide and 90 mM diazirine in labeling buffer. Photolysis was performed on 0.8 µL aliquots of the equilibrated sample in a windowed 450 µm ID/670 µm o.d. fused silica capillary (Molex, Lisle, IL, USA). Photolysis was achieved with a single 150 mJ (±5%) 355 nm laser pulse of a 10 ns pulse-width, using an Nd:YAG laser (YG 980; Quantel, Les Ulis, France). The laser light was focused using a biconvex lens followed by a plano-concave cylindrical lens into a tight ellipse measuring approximately 1 mm × 7 mm. Solid-phase labeling was performed by snap-freezing the sample at 77 K with liquid nitrogen, on a custom-built apparatus that maintains low temperature during irradiation. The irradiation apparatus is a modified beam-dump design that allows the sample capillary to be held in place and precisely aligned via a three-axis stage.

LC-MS

Following irradiation, samples were discharged from the capillary and diluted to a final concentration of 1 µM. Aliquots were injected into an LC-MS system, conured with an Acclaim PepMap 100 guard column (75 µm × 2 cm C18, 3 µm particles, 100 Å), and separated on a self-packed reverse-phase C18 HPLC column (75 µm × 13 cm, Aeris 3.6 µm particles, Peptide XB C18, 100 Å; Phenomenex, Torrance, CA, USA). Peptides were eluted using a 25 min 5%–40% B gradient at 300 nL/min. Mobile phase A consisted of 0.1% v/v formic acid, 0.1% w/v m-nitrobenzyl alcohol (mNBA; Sigma Aldrich) and 3% acetonitrile. Mobile phase B had the same additives, in 97% acetonitrile. Mobile phases contained mNBA for supercharging to promote higher charge states and more efficient ETD fragmentation [39]. Separation was achieved with an EasyLC 1000 (Thermo Scientific). Peptides were identified and peptide modifications were localized using EThcD fragmentation on an LTQ Orbitrap Fusion Lumos (Thermo Scientific), using a high/high configuration. All spectra were acquired in positive ion mode with a mass range of 300–2000 Th, with the FT analyzer at a resolution of 60,000 for MS and 15,000 for MS/MS scans. Spray voltage was set at 2.2 kV and desolvation temperature at 275 °C. EThcD was performed on the top 10 most intense ions with ≥3+ charge. Unmodified peptides were identified using a set of LC-MS/MS runs and used to build an exclusion list in a two-pass analysis of the modified peptides, to improve the detection of low-intensity modified peptides. The first-pass analysis used a dynamic exclusion time of 35 s, and the second-pass analysis used a dynamic exclusion time of 5 s.

Data Analysis

LC-MS/MS data was processed with the covalent labeling plug-in in Mass Spec Studio for tracking modified peptides [40]. The cross-linking Mass Spec Studio plug-in was used for peptide identification, as it incorporates probabilistic scoring methodology for singly-labeled (so-called "dead-end") peptides [41]. MS and MS/MS mass accuracy windows were set to 20 and 30 ppm, respectively. Ions from the c-, z-, y-, and b-series were searched, with variable modifications, for spectral matching. A 5% false discovery rate was applied for peptide identification, and only unambiguous localizations of the covalent modification were considered for site determination analysis. Insertion frequencies for a given amino acid were quantified by simply taking the ratio of labeled residues positively identified to the total residues in the pool of peptides, expressed as a percent. Quantification of peptide labeling yield was calculated using the tools available within MZmine 2 [42].

Recovery of Ambiguous Carbene Insertion Data

In the subset of data where the site of carbene insertion could not be localized unambiguously, we determined probable sites of carbene insertion constrained by the MS/MS fragmentation data and amino acid labeling probabilities. The likelihood of carbene insertion at any given site x was calculated according to Eqs. 1 and 2,

$$P_x = \frac{w_x}{\sum_{i=1}^{n} w_i} \tag{1}$$

where

$$w_x = \frac{a_{x,r}}{n}. \tag{2}$$

P_x is the probability of labeling a residue at given site x for a peptide with n possible ambiguous insertion sites; w_x is the weighted score of each site x, determined from $a_{x,r}$, which is the preference of reagent r towards residue at site x (see Figure 4).

Results and Discussion

Observations and an Improved Configuration

Our first studies demonstrated that two simple aliphatic diazirines differing only by an amino group diverged in both the amount and the location of labeling [23]. Based on the assumption that carbene insertions should be indiscriminate, we concluded that the "surface-solvating" properties of the diazirine would serve to differentially position the reagent at the surface of the protein, at equilibrium. To extend our understanding of this effect, we collected additional data using whole proteins and illustrated that labeling yields are affected by buffer choice and ionic strength, and diazirine substituents in manner consistent with the reagent's surface activity. In order to better understand the utility of diazirines in footprinting applications, we decided to study the reactivity of these simple diazirines in an experimental configuration that minimizes the complexity associated with higher-order structure, and in a
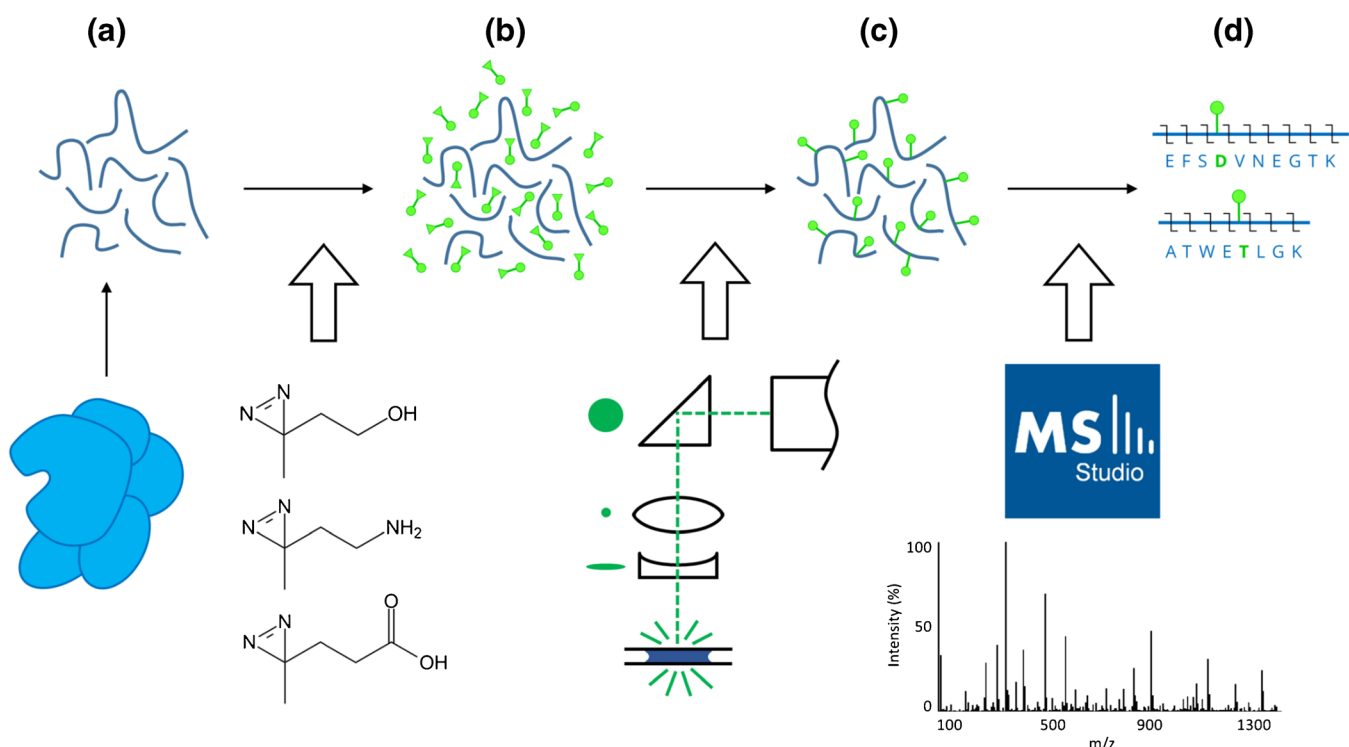
Figure 1. Overview of the photolytic covalent labeling workflow, for peptide labeling. (**a**) Proteolytic digestion of protein generates a population of unlabeled peptides. (**b**) Equilibration of peptides with substituted diazirines in aqueous solution. Diazirine labeling reagents used in this study include 3,3′-azibutan-1-ol, 3,3′-azibutyl-1-ammonium, and 4,4′-azipentan-1-oate. (**c**) Photolysis at λ = 355 nm to generate reactive carbenes that insert into chemically accessible regions of peptide. Photolysis is constrained to a submicroliter volume in a windowed UV-transparent capillary that supports flash-freezing in liquid nitrogen. (**d**) Localization of carbene insertion sites within peptides using high-resolution MS/MS data analyzed in the Mass Spec Studio software package

reaction chamber that promotes efficient single-shot labeling. Our earlier studies suggested that carbenes can diffuse into bulk solvent before reacting [31], even though their lifetimes are short. We demonstrated that freezing improved the correlation between surface accessibility and label incorporation, but to date, all our configurations involved frozen protein samples with long irradiation times. We constructed an experimental configuration that allows for sub-microliter, single-shot irradiation in the frozen state.

We used this new configuration to label pools of peptides equilibrated with relatively high concentrations of diazirine reagent, in order to maximize insertion yields (Figure 1). We evaluated the concentration range that could be accommodated in this constrained system, and found that irradiating diazirine concentrations higher than 90 mM led to occasional fracturing of the capillary, presumably arising from excessive nitrogen liberation from the diazirine. Thus, all of our experiments were limited to 90 mM reagent concentration and a pulse energy of 150 mJ, three times greater than necessary to convert all of the reagent, based on beam and sample dimensions, and a measured quantum yield of 83% (data not shown). We chose not to use multiple laser firings at lower pulse energies. We were concerned with perturbing sample equilibrium with the nitrogen gas, as air–water interfaces can be highly denaturing [43]. Freezing can also influence protein structure [44]. Here, we used small samples in thin-walled capillaries, coupled with

rapid immersion in liquid nitrogen to achieve rates of freezing that are primarily limited by the insulating properties of the silica and the Leidenfrost effect, which creates an insulating layer of nitrogen gas between the cryogenic liquid and the sample [45]. We estimate the freezing rate to be less than 1 s, which is acceptable for our current study, as it focuses on unstructured peptides. Freezing needs to be fast enough to avoid water crystallization, but true vitrification is difficult to achieve with simple plunge-freezing systems. To achieve faster rates of freezing for protein samples, future designs will incorporate spray-freezing and cryofixation protocols.

Peptide Library Composition

We elected to use peptides for insertion site analysis rather than simple amino acids. Given the indiscriminate nature of carbene chemistry, amino acids would present free N and C termini that compete with side-chain reactions. Although capped amino acids could be used for such purposes, the termini would present non-native compositions and similarly bias the yield measurements. A large set of peptides of varying sequence length and composition was produced using protein digests. We identified 777 unique peptides with an average length of 16 residues (Supplementary Figure S1). Our peptide set represents over 12,294 amino acids; thus any possible primary sequence effects on labeling (e.g., adjacent amino acid bias) should be
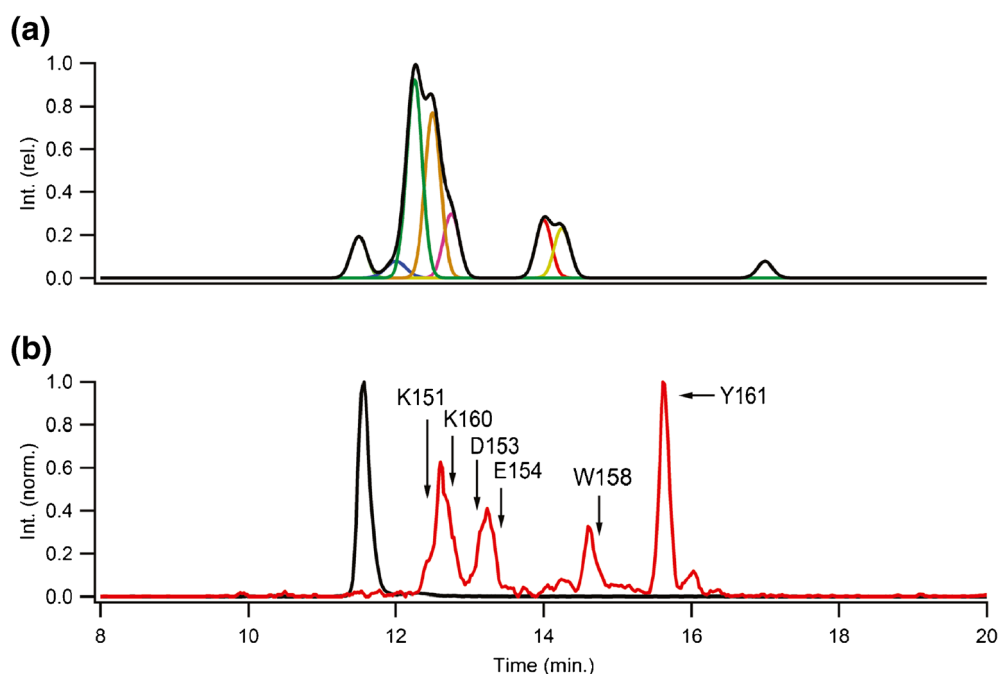
Figure 2. Extracted ion chromatograms (EIC) for peptides and their carbene-labeled counterparts. (**a**) Simulation of a single peptide generating multiple positional isomers with variable retention times. Composite EIC (black trace) and individual positional isomer EICs (colored traces). (**b**) Example EICs for FKADEKKFWGKY (black trace, unlabeled), where photolytic labeling with 4,4'-azipentan-1-oate generated a mixed of resolvable and partially resolvable chromatographic features leading to six site identifications (red trace, labeled)

diluted in the overall results. A data analysis routine was also devised to return reliable amino acid labeling frequencies, which is based on insertion site counting rather than ion chromatogram intensity (see below). Peptide pools were buffered to neutrality at low ion strength. Neutral solution is a relevant labeling condition, and it allows us to examine the effect of a neutral reagent (3,3'-azibutan-1-ol), a negatively charged reagent (4,4'-azipentan-1-oate), and a positively charged reagent (3,3'-azibutyl-1-ammonium) on the site of insertion. Maintaining a low ionic strength during labeling allows us to detect if local electrostatic effects influence the locations of insertion.

Positional Isomers and Effect on MS/MS Sampling

Carbene labeling generated a large population of peptides homogeneous in molecular weight, but heterogeneous in the insertion site. This heterogeneity is particularly apparent in the retention times of the labeled peptides. Retention times for singly labeled peptides, for example, can be influenced in complex ways, sometimes permitting complete chromatographic separation and sometimes generating strongly overlapping chromatographic profiles (Figure 2a). Retention times are difficult to predict, and can spread a signal over several min in our standard gradient (e.g., Figure 2b). As a result, sampling for MS/MS-based site identification was not trivial. Chromatographic signal-splitting coupled with variable insertion yields argue for high-sensitivity product ion scans. However, the unpredictable retention times along with the geometric expansion of the number of chromatographic peaks suggested that MS/MS acquisitions would be difficult to schedule. To

generate a sufficiently comprehensive set of MS/MS spectra, we chose to combine the output of two data-dependent acquisition methods: one with a short dynamic exclusion setting (approximately 1/3 the natural chromatographic peak width) to better sample overlapping isomers, and one with a longer setting (approximately twice the chromatographic peak width) to improve our coverage of the peptide set. The short setting generated 1.6 times as many useable MS^2 scans as the longer setting, yet the latter resulted in 13% more unique identifications.

Site Identification

We previously showed that carbene insertion into carboxylic acids, as found in glutamate and aspartate, could yield labile ester linkages and generate neutral losses in either CID or HCD fragmentation experiments [23]. ETD is the preferred fragmentation mode, but it is ineffective for smaller, lower charge-state peptides, and does not always provide high "sequence reads." Hybrid fragmentation modes improve the localization of post-translation modifications by increasing sequence coverage without necessarily increasing the loss of labile modifications [46, 47]. We used EThcD in this study to improve coverage, but carbene insertions present some unique complications. We cannot make *a priori* assumptions about insert locations, as we do in phosphosite analysis for example, so resolving insertion sites require high-quality MS/MS spectra. Even with efficient chromatography and the data-dependent sampling strategy described above, there is a high incidence rate of overlapping positional isomers, generating chimeric MS/MS spectra.
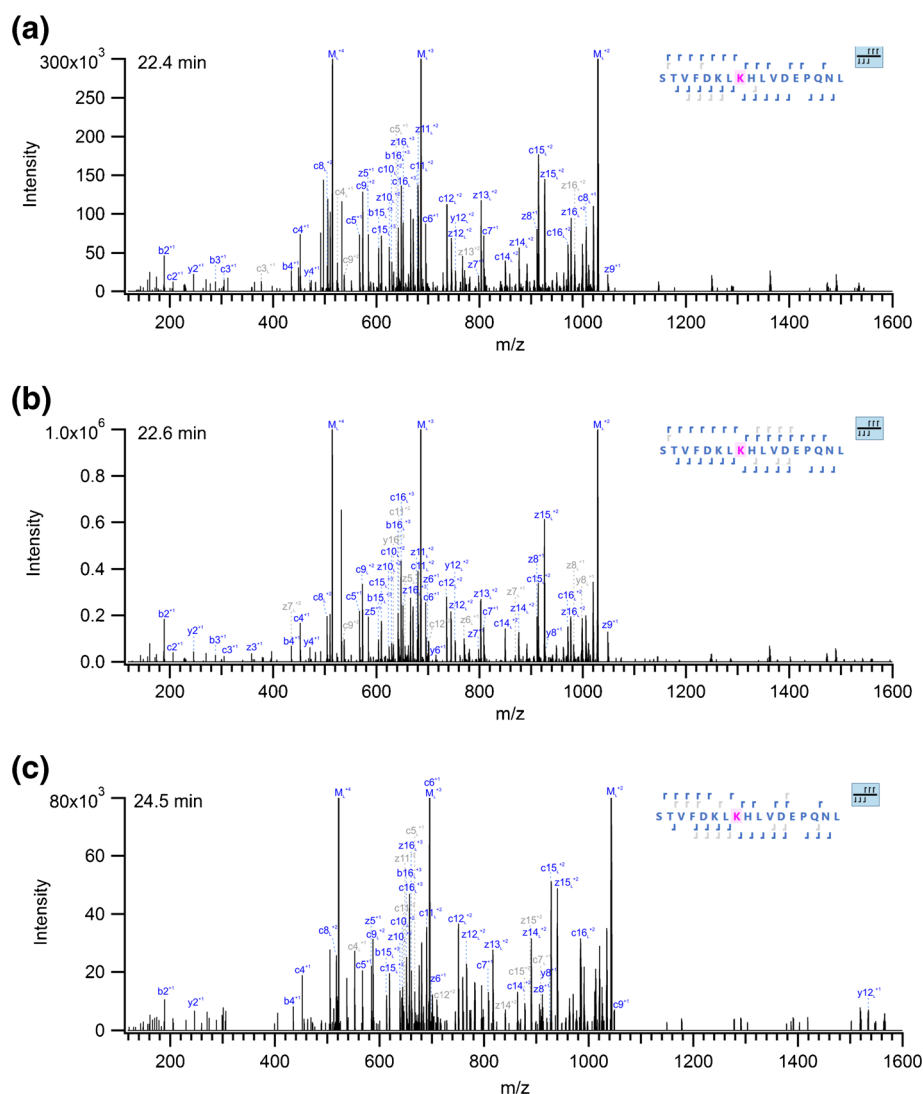
Figure 3. EThcD MS/MS spectra of STVFDKLKHLVDEPQNL in the 4+ charge state, modified at K8 with (**a**) 3,3′-azibutan-1-ol, (**b**) 3,3′-azibutyl-1-ammonium, or (**c**) 4,4′-azipentan-1-oate. Inset sequence shows identified fragment ions with top-scoring site of modification highlighted in pink. Annotations in blue correspond to fragments that support the localization of carbene insertion at K8; annotations in grey support other sites of modification. Subscript 'L' indicates mass modification with respective reagent. Intensity of precursor and charge-reduced precursor ions are truncated for clarity. Retention times of each modified peptide are shown in top left of each spectrum

Further, HCD-based activation of all ETD products returns a measure of neutral loss, as the price for improved sequence coverage. Neutral loss takes the form of expulsion of a stable alkene, having a mass identical to the modification.

For these reasons, we imposed strong criteria for site identification and elected to use a counting method for determining insertion site preference. Only peptides generating a unique, unambiguous site of insertion were considered for site counting. Scoring implemented our cross-linking detection algorithm already adapted for the detection of "dead-end" or single-site insertions [41], and uniqueness was based on score separation from the next highest ranking candidate. For example, the peptide STVFDKLKHLVDEPQNL with a single insertion of a butylammonium group generated a fragment

spectrum attributed to insertion at the second K only, even though credible evidence exists for insertion at E (Figure 3a). In this fashion, we avoid over-counting sites. While this conservative approach likely underestimates the *overall* insertion frequency, we suggest that the large number of peptides used in the study will compensate and generate meaningful *relative* insertion preferences. A counting method based on a large volume of data is preferred over evaluating site preferences using extracted ion chromatogram intensities, since we cannot routinely assign a single site to one chromatographic feature (e.g., Figure 2b). We inspected all MS/MS spectra that generated identifications at common residues, and did not observe a strong impact of the reagent on fragmentation patterns (e.g., Figure 3a–c), except perhaps a slightly higher preference for

Table 1. Carbene Insertion Yields on a Per-Reagent Basis

	3,3-Azibutan-1-ol	3,3-Azibutyl-1-ammonium	4,4-azipentan-1-oate
Total labeled[a]	776	734	553
Unambiguous[b]	354 (45.6%)	333 (45.4%)	239 (43.2%)
Ambiguous[c]	422 (54.4%)	401 (54.6%)	314 (56.8%)

[a] Cumulative values of MS/MS spectra from singly-modified peptide set (where total number of unlabeled peptides in pool is 777)
[b] Number of MS/MS spectra converted into unambiguous site identifications (percentage of total labeled)
[c] Number of MS/MS spectra generating ambiguous site identifications (percentage of total labeled)

multiply charged fragment ions using the butylammonium reagent. A single higher charge state was selected for fragmentation to ensure an adequate population of mobile proton, but we did not observe a strong effect of charge on sequencing effectiveness (Supplementary Figures S2, S3). We anticipate that the greatest variability in site determination arises from different retention times and the corresponding changes in ionization suppression and/or peak overlap. Again, the large number of peptides analyzed should minimize any such effects. Finally, we removed all site identifications attributable to terminal amino acids, to avoid any bias associated with insertions at the free C or N termini, and only mined peptides with one insertion event.

Distribution of Insertion Products

The 777 peptides generated 2063 high quality MS/MS spectra from all three reagents. Table 1 summarizes the site identification statistics. Over 40% of the spectra for each reagent were converted to unambiguous site identifications.

We sorted the data by amino acid for each reagent (Figure 4), where the insertion frequency is expressed as a percentage of a given residue in the entire pool. Aliphatic diazirines of the type used here are able to insert into any amino acid. When reagent type is taken into account, insertion frequencies ranged from over 16% for tyrosine when using 3,3-azibutyl-1-ammonium, to a low of 0.1% for threonine when using 3,3-azibutan-1-ol. When averaged over all reagents, we observe a relatively narrow 20-fold range in insertion frequencies, considerably tighter than hydroxyl radical insertions, for example [17]. Insertion frequencies scale roughly as a function of side-chain polarity and size, but it is noteworthy that even short hydrophobic amino acids sustain measurable insertion rates. The trends in amino acid reactivity show both similarities and differences compared with previous limited studies. For example, we observed that tyrosine and glutamate are indeed common insertion sites, but the bias towards acidic residues is not as dominant as once speculated [24].

We can directly compare insertion rates between the three reagents for individual amino acids, assuming that there is no compound-specific redistribution of reagents during the snap-freezing event. For the most part, labeling trends are similar for the different reagents, reflecting carbene reaction chemistry and its dominant effect on insertion frequency. We do note a number of interesting features in the comparison, however, informed by the measurement precision estimated in Figure 5 (see below). The 3,3'-azibutan-1-ol is relatively biased against F (aromatic) and T (neutral), but favors H strongly (aromatic and neutral at the labeling pH). Both negatively charged 4,4'-azipentan-1-oate and positively charged 3,3'-azibutyl-1-ammonium are remarkably similar in their insertion rates for neutral residues and K, but 4,4'-azipentan-1-oate favors insertion into R, D, and E. All of these differences (and others not highlighted here) likely reflect the underlying interactions with the preactivated reagent. For example, the lower insertion of

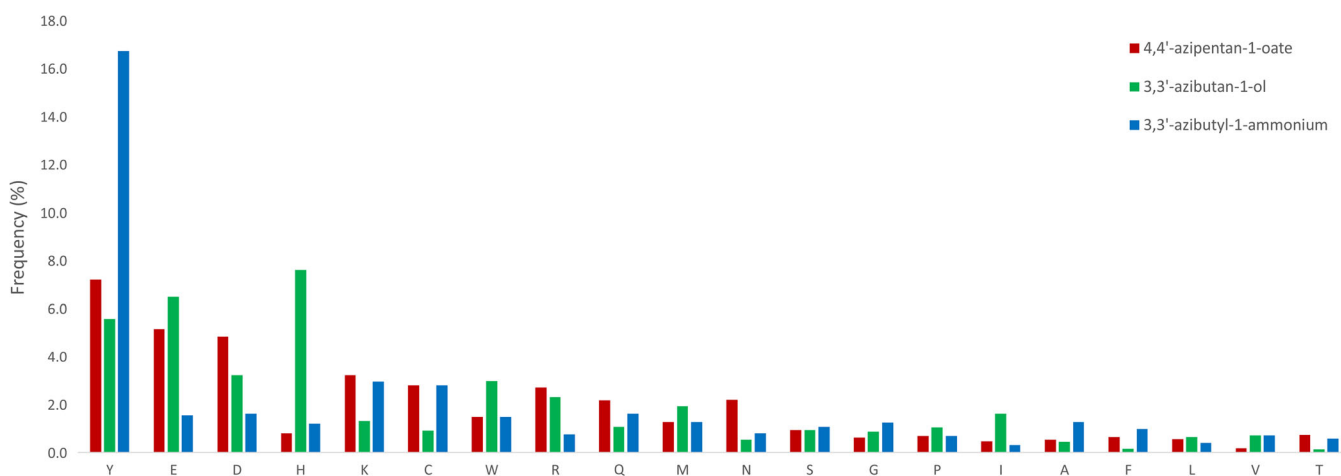**Figure 4.** Average frequency of carbene insertion at each residue generated from the photolysis of either 3,3-azibutyl-1-ammonium, 3,3'-azibutan-1-ol or 4,4'-azipentan-1-oate in the presence of protein digests (777 peptides). Site of label insertion was localized with MS/MS data generated on a Fusion Lumos with EThcD fragmentation and analyzed with the Mass Spec Studio as previously described

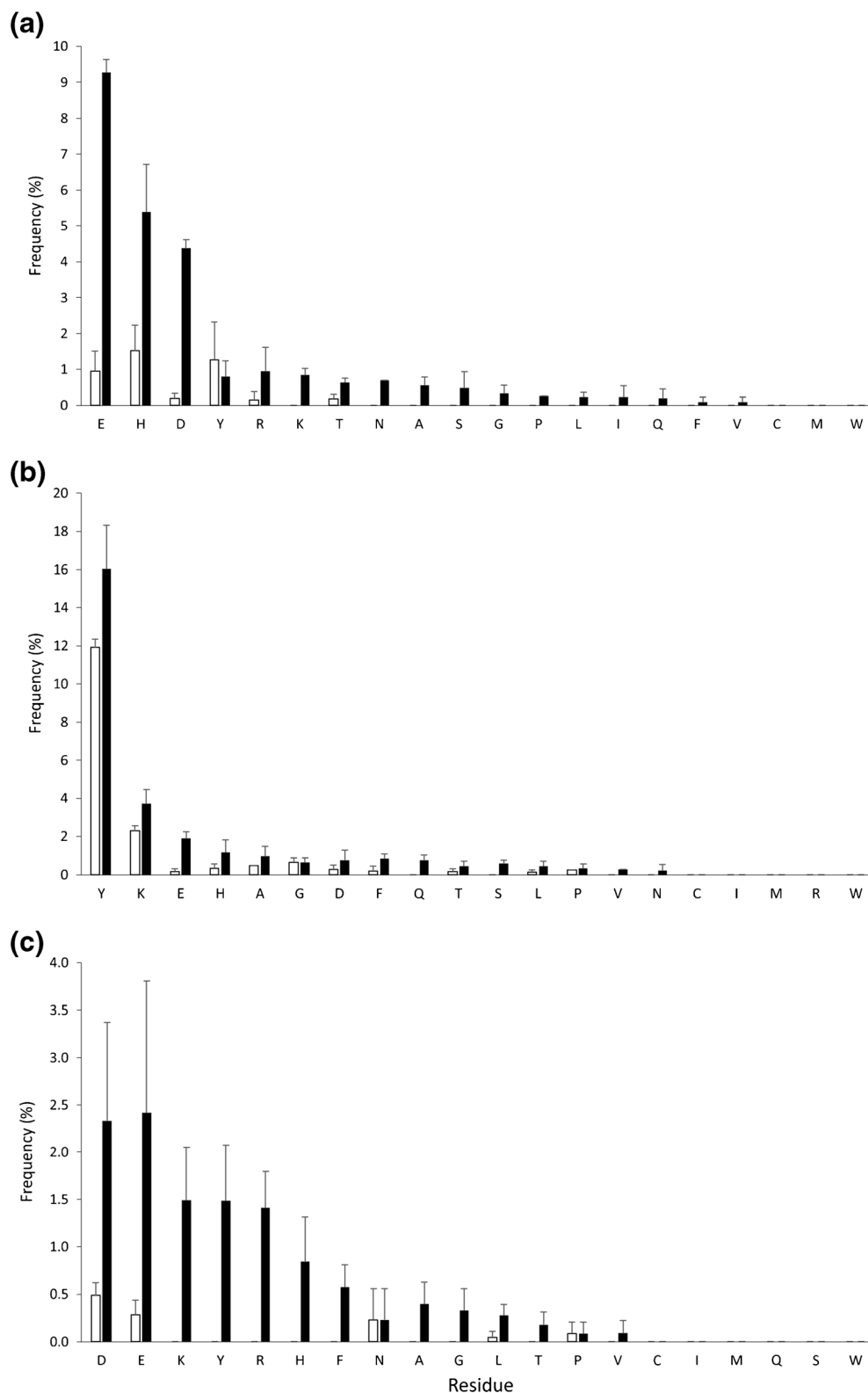**Figure 5.** Comparison of the average frequency of carbene insertion at each residue from diazirine photolysis performed in the liquid (293 K, white) and solid (77 K, black) state. Labeling was performed on peptides generated from the BSA digest only. Reagents used were (**a**) 3,3′-azibutan-1-ol, (**b**) 3,3′-azibutyl-1-ammonium, and (**c**) 4,4′-azipentan-1-oate. Data is the average of three replicates, error bars indicate 1 SD

butylammonium into D and E could arise from ion-pairing orienting the diazirine away from the side chain prior to freezing. It is not immediately obvious why butanol inserts into H so strongly. However, our data confirm that substituent effects are possible at the level of primary structure. Together with our observations of selectivity in higher-order structure, which we observed when labeling intact proteins, reagents that are oriented by the substituents can steer the insertion of the carbene in complex ways, driven by noncovalent molecular interactions. These observations can ultimately be used in the rational design of photoprobes, labeling agents, and cross-linkers, and they suggest a role for molecular simulations to aid in understanding the distribution of labeling products.

Restricting Diffusion

As we noted previously and confirmed recently by Oldham et al. [29], freezing increases the yield of insertion. We observed this to be true for all three reagents. Sampling a subset of common peptides, we estimate a 6-fold greater yield in the frozen solid state (77 K) versus the liquid state (293 K) (Supplementary Figure S4). These values were consistent with protein-level measurements, using MALDI (data not shown). Assuming carbene lifetimes to be <10 ns [48], an active carbene reagent of the molecular weight used in this study can diffuse isotropically ≤130 Å from its site of generation. Such distances are sufficient to increase quenching rates by bulk solvent and thus reduce yield. To test if diffusion could also rebalance insertion frequencies, we compared relative insertion rates for the peptide pool at two temperatures (Figure 5a–c). Generally, labeling in solution at room temperature reduced labeling frequencies across insertion sites (compared with labeling at 77 K), which is expected as frequency determination is a surrogate method for determining yield. However, there are exceptions. For example, we observe very little change in the insertion frequencies for tyrosine using 3,3'-azibutyl-1-ammonium and 3,3'-azibutan-1-ol, which suggests insertion kinetics for these reactions are fast enough to overcome diffusional loss. However, this is not the case for 4,4'-azipentan-1-oate, where no labeling of tyrosine is observed at room temperature, which further suggests that the precise nature of the reagent/amino-acid interaction can influence insertion kinetics. For example, it is possible that 3,3'-azibutyl-1-ammonium and 3,3'-azibutan-1-ol position the carbene

optimally for fast insertion (perhaps at the aryl alcohol) whereas 4,4'-azipentan-1-oate orients the carbene for insertion into a less reactive bond (e.g., the ring itself).

These observations are confirmed by an inspection of ion chromatograms for the two states, which show pattern changes consistent with differential labeling (Supplementary Figure S5). Whatever the precise orientation, the data supports the idea that the differential labeling we observe between states arises from a complex combination of site-specific carbene insertion rates, alternative pre-orientations of the carbene at the amino acid, and/or different reagent-peptide "off-rates." The increased yield and reduction in site labeling bias provide a strong justification for implementing this labeling chemistry in the frozen state.

Photolysis Products – Additional Considerations

Aliphatic diazirines of the type we used can generate a range of intermediates upon photolysis, including singlet and triplet carbene states [33, 49] and diazo isomers that can further decompose to carbocations [24, 50]. These intermediates can undergo insertion to generate the labeled amino acids that we observe [34, 36]. For reactions in the frozen state (77 K), we anticipate based on previous characterizations that amino-acid insertion products could be driven by both short-lived, ground-state carbenes and by diazo-mediated carbocations following carbene rearrangement [50].

However, it is difficult to allocate reaction products to one mechanism or the other, as in most cases the products are indistinguishable with low-energy fragmentation methods of analysis. Reactions may differ only in rates, stereochemistry of the products, or the nature of side reactions. For example, we observed insertion levels for T and S similar to C–H insertions into hydrophobic residues (Figure 4). Insertions could reflect the trapping across the OH bond of a singlet carbene, but they could also involve triplet insertion into the methylene groups of the side chains. The energy difference between the triplet (ground) state and the singlet state is small for aliphatic diazirines (~2 kcal/mol), and inversions in favor of the singlet ground state are possible [51]. The small energy separation may not be great enough to favor one state or the other even at low temperatures; therefore a blend of both states is likely. Finally, the oxidation side products that we observed in intact proteins were confirmed in the peptide pools, but the levels

Table 2. Assigning Insertion Site Probabilities Based on Chemical Labeling Preference

Peptide RT (min)	m/z	Peptide[a]	Start	Stop	Possible residues	Probability (%)		
						Site 1	Site 2	Site 3
18.7	436.5371	FTGHPETLEKFDKFKHLKTEAE**MKA**SEDLKKHGTVVL	34	70	M56, K57, A58	39.8	42.7	17.6
17.0	436.5371	FTGHPETLEKFDKFKHLKTEAE**MK**ASEDLKKHGTVVL	34	70	M56, K57	48.2	51.8	-
7.7	399.7424	KASED**LKK**HGTVVL	57	70	L62, K63, K64	6.2	46.9	46.9
13.1	580.3433	TALGGILKKKGHHEAELKPLAQSHATKHK**IP**IKYLE	71	106	I100, P101	-	27.4	72.6
13.0	580.3433	TALGGILKKKGHHEAELKPLAQSHATKH**KIP**IKYLE	71	106	K99, I100, P101	77.9	6.0	16.0

[a] Singly labeled peptides with ambiguous carbene insertion location narrowed down to the bolded residues. Data obtained from the labeling of myoglobin peptides using 3,3-azibutyl-1-ammonium

were generally low and variable (data not shown). Degassing and/or sparging the samples with helium did little to influence the oxidation levels.

Localizing Carbene Insertion Points – Additional Considerations

Even with improved fragmentation methods and the use of a sensitive detection platform, over 50% of high quality MS/MS spectra do not sustain a precise localization of the reagent. Unfortunately, we cannot fully resolve this site determination ambiguity, as the reagents are capable of inserting into all 20 amino acids. However, some movement in this direction is possible. In the first place, the ambiguity can be simply retained and used as a lower resolution restraint in modeling activities. Additionally, by assigning a chemical labeling probability to each residue based on our profiling data, we can promote or penalize individual residues. An illustration is provided in Table 2. Entries 1 and 2 are chromatographically well-separated and readily support two selections, whereas entries 4 and 5 are complicated by the possibility of chimeric spectra. In both situations, it is appropriate to retain all insertion probabilities. For structure mapping and protein modeling activities, assigning probabilities to topographic profiling can improve robustness in scoring functions, as recently illustrated with ambiguous cross-linking data [52]. A scoring approach of this nature can be extended to include other aspects of the analysis, such as feature abundance, retention-time shifts, and overlapping peptide sequence information.

Conclusions and Perspective

Our study confirms that a class of small carbene precursors based on an aliphatic diazirine backbone can access all 20 amino acids over a relatively narrow range of insertion frequencies, compared with other broadly-active labeling options (e.g., hydroxyl radicals). The comprehensive reactivity is not only the result of carbene functionality but also restricted diffusion in the solid state. Perhaps the greatest challenge in implementing diazirine labeling methods for footprinting involves localizing insertion sites with accuracy. We have shown that EThcD is a viable method for high frequency site determination, and by adopting informatics strategies applied to phosphosite inference [53], we anticipate that otherwise ambiguous data may still have a role in providing additional levels of structural restraints. The best way to use this study is to implement the site determination method in a referencing strategy, where a denatured or digested protein state is labeled and processed. The approach would generate a normalizing set of labeling data for the corresponding structured protein. It may provide a more relevant frequency determination than one built upon random sets of proteins.

There remains room for additional reagent development. Eliminating an abstractable hydrogen in the reagent could prevent neutral loss generating a leaving group with the same mass as the reagent, and improve identification. The concept

that the reagent scaffold is more than just a carrier of the reactive site is a key outcome of this study. While the selectivity of labeling is most dramatic at the protein level, the data in this study show it is also visible at the level of primary structure. Careful scaffold design should permit the targeting of diazirines to distinct structural elements and alleviate some of the difficulties associated with localizing labeling sites.

Acknowledgements

References

1. Dunham, W.H., Mullin, M., Gingras, A.C.: Affinity-purification coupled to mass spectrometry: basic principles and strategies. Proteomics **12**, 1576–1590 (2012)

2. Hesketh, G.G., Youn, J.Y., Samavarchi-Tehrani, P., Raught, B., Gingras, A.C.: Parallel Exploration of interaction space by bioid and affinity purification coupled to mass spectrometry. Methods Mol. Biol. **1550**, 115–136 (2017)

3. Nguyen-Huynh, N.T., Sharov, G., Potel, C., Fichter, P., Trowitzsch, S., Berger, I., Lamour, V., Schultz, P., Potier, N., Leize-Wagner, E.: Chemical cross-linking and mass spectrometry to determine the subunit interaction network in a recombinant human SAGA HAT subcomplex. Protein Sci. **24**, 1232–1246 (2015)

4. Politis, A., Stengel, F., Hall, Z., Hernandez, H., Leitner, A., Walzthoeni, T., Robinson, C.V., Aebersold, R.: A mass spectrometry-based hybrid method for structural modeling of protein complexes. Nat. Methods **11**, 403–406 (2014)

5. Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M.P., Chait, B.T.: Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. Mol. Cell. Proteom. **13**, 2927–2943 (2014)

6. Dai, S.Y., Burris, T.P., Dodge, J.A., Montrose-Rafizadeh, C., Wang, Y., Pascal, B.D., Chalmers, M.J., Griffin, P.R.: Unique ligand binding patterns between estrogen receptor alpha and beta revealed by hydrogen–deuterium exchange. Biochemistry **48**, 9668–9676 (2009)

7. Percy, A.J., Rey, M., Burns, K.M., Schriemer, D.C.: Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry – a review. Anal. Chim. Acta **721**, 7–21 (2012)

8. Wales, T.E., Engen, J.R.: Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrom. Rev. **25**, 158–170 (2006)

9. Mendoza, V.L., Vachet, R.W.: Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. Mass Spectrom. Rev. **28**, 785–815 (2009)

10. Higashimoto, Y., Sugishima, M., Sato, H., Sakamoto, H., Fukuyama, K., Palmer, G., Noguchi, M.: Mass spectrometric identification of lysine residues of heme oxygenase-1 that are involved in its interaction with NADPH-cytochrome P450 reductase. Biochem. Biophys. Res. Commun. **367**, 852–858 (2008)

11. Mendoza, V.L., Antwi, K., Barón-Rodríguez, M.A., Blanco, C., Vachet, R.W.: Structure of the preamyloid dimer of β-2-microglobulin from covalent labeling and mass spectrometry. Biochemistry **49**, 1522–1532 (2010)

12. Nikfarjam, L., Izumi, S., Yamazaki, T., Kominami, S.: The interaction of cytochrome P450 17alpha with NADPH-cytochrome P450 reductase, investigated using chemical modification and MALDI-TOF mass spectrometry. Biochim. Biophys. Acta **1764**, 1126–1131 (2006)

13. Hartlmuller, C., Gobl, C., Madl, T.: Prediction of protein structure using surface accessibility data. Angew. Chem. **55**, 11970–11974 (2016)

14. Boehr, D.D., Nussinov, R., Wright, P.E.: The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. **5**, 789–796 (2009)

15. Liu, J., Nussinov, R.: Allostery: an overview of its history, concepts, methods, and applications. PLoS Comput. Biol. **12**, e1004966 (2016)

16. Means, G.E., Feeney, R.E.: Chemical modification of proteins. Holden-Day, Inc., San Francisco (1971)

17. Xu, G., Chance, M.R.: Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. Chem. Rev. **107**, 3514–3543 (2007)

18. Chen, J., Rempel, D.L., Gau, B.C., Gross, M.L.: Fast photochemical oxidation of proteins and mass spectrometry follow submillisecond protein folding at the amino-acid level. J. Am. Chem. Soc. **134**, 18724–18731 (2012)

19. Vahidi, S., Stocks, B.B., Liaghati-Mobarhan, Y., Konermann, L.: Submillisecond protein folding events monitored by rapid mixing and mass spectrometry-based oxidative labeling. Anal. Chem. **85**, 8618–8625 (2013)

20. Vahidi, S., Konermann, L.: Probing the time scale of FPOP (fast photochemical oxidation of proteins): radical reactions extend over tens of milliseconds. J. Am. Soc. Mass Spectrom. **27**, 1156–1164 (2016)

21. Turro, N.J., Ramamurthy, V., Scaiano, J.C.: Principles of Molecular Photochemistry: an Introduction. University Science Books, Sausalito (2009)

22. Toscano, J.P., Platz, M.S., Nikolaev, V.: Lifetimes of simple ketocarbenes. J. Am. Chem. Soc. **117**, 4712–4713 (1995)

23. Jumper, C.C., Bomgarden, R., Rogers, J., Etienne, C., Schriemer, D.C.: High-resolution mapping of carbene-based protein footprints. Anal. Chem. **84**, 4411–4418 (2012)

24. Das, J.: Aliphatic diazirines as photoaffinity probes for proteins: recent developments. Chem. Rev. **111**, 4405–4417 (2011)

25. Dubinsky, L., Krom, B.P., Meijler, M.M.: Diazirine-based photoaffinity labeling. Bioorg. Med. Chem. **20**, 554–570 (2012)

26. Chowdhry, V., Westheimer, F.H.: Photoaffinity labeling of biological systems. Annu. Rev. Biochem. **48**, 293–325 (1979)

27. Gomes, A.F., Gozzo, F.C.: Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. J. Mass Spectrom. **45**, 892–899 (2010)

28. Suchanek, M., Radzikowska, A., Thiele, C.: Photo-leucine and photo-methionine allow identification of protein–protein interactions in living cells. Nat. Methods **2**, 261–267 (2005)

29. Manzi, L., Barrow, A.S., Scott, D., Layfield, R., Wright, T.G., Moses, J.E., Oldham, N.J.: Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. Nat. Commun. **7**, 13288 (2016)

30. Ureta, D.B., Craig, P.O., Gomez, G.E., Delfino, J.M.: Assessing native and non-native conformational states of a protein by methylene carbene labeling: the case of *Bacillus licheniformis* beta-lactamase. Biochemistry **46**, 14567–14577 (2007)

31. Jumper, C.C., Schriemer, D.C.: Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. Anal. Chem. **83**, 2913–2920 (2011)

32. Zhang, B., Rempel, D.L., Gross, M.L.: Protein footprinting by carbenes on a fast photochemical oxidation of proteins (FPOP) platform. J. Am. Soc. Mass Spectrom. **27**, 552–555 (2016)

33. Irikura, K.K., Goddard, W.A., Beauchamp, J.L.: Singlet-triplet gaps in substituted carbenes CXY (X, Y = H, fluoro, chloro, bromo, iodo, silyl). J. Am. Chem. Soc. **114**, 48–51 (1992)

34. Moss, R.A.: Carbenic reactivity revisited. Acc. Chem. Res. **22**, 15–21 (1989)

35. Moss, R.A., Mallon, C.B., Ho, C.-T.: The correlation of carbenic reactivity. J. Am. Chem. Soc. **99**, 4105–4110 (1977)

36. Moss, R.A., Shen, S., Hadel, L.M., Kmiecik-Lawrynowicz, G., Wlostowska, J., Krogh-Jespersen, K.: Absolute rate and philicity studies of methoxyphenylcarbene. An extended range for carbenic ambiphilicity. J. Am. Chem. Soc. **109**, 4341–4349 (1987)

37. Church, R.F.R., Weiss, M.J.: Diazirines. 2. Synthesis and properties of small functionalized diazirine molecules – some observations on reaction of a diaziridine with iodine–iodide ion system. J. Org. Chem. **35**, 2465 (1970)

38. Shigdel, U.K., Zhang, J.L., He, C.: Diazirine-based DNA photo-cross-linking probes for the study of protein–DNA interactions. Angew. Chem. Int. Edit. **47**, 90–93 (2008)

39. Li, X., Li, Z., Xie, B., Sharp, J.S.: Supercharging by m-NBA improves ETD-based quantification of hydroxyl radical protein footprinting. J. Am. Soc. Mass Spectrom. **26**, 1424–1427 (2015)

40. Rey, M., Sarpe, V., Burns, K.M., Buse, J., Baker, C.A., van Dijk, M., Wordeman, L., Bonvin, A.M., Schriemer, D.C.: Mass spec studio for integrative structural biology. Structure **22**, 1538–1548 (2014)

41. Sarpe, V., Rafiei, A., Hepburn, M., Ostan, N., Schryvers, A.B., Schriemer, D.C.: High sensitivity cross-link detection coupled with integrative structure modeling in the mass spec studio. Mol. Cell. Proteom. **15**, 3071–3080 (2016)

42. Pluskal, T., Castillo, S., Villar-Briones, A., Oresic, M.: MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Ioinformatics **11**, 395 (2010)

43. Graham, D.E., Phillips, M.C.: Proteins at liquid interfaces: I. Kinetics of adsorption and surface denaturation. J. Colloid Interface Sci. **70**, 403–414 (1979)

44. Zhang, A., Qi, W., Singh, S.K., Fernandez, E.J.: A new approach to explore the impact of freeze-thaw cycling on protein structure: hydrogen/deuterium exchange mass spectrometry (HX-MS). Pharm. Res. **28**, 1179–1193 (2011)

45. Dobro, M.J., Melanson, L.A., Jensen, G.J., McDowall, A.W.: Plunge freezing for electron cryomicroscopy. Methods Enzymol **481**, 63–82 (2010)

46. Brunner, A.M., Lossl, P., Liu, F., Huguet, R., Mullen, C., Yamashita, M., Zabrouskov, V., Makarov, A., Altelaar, A.F., Heck, A.J.: Benchmarking multiple fragmentation methods on an Orbitrap fusion for top-down phospho-proteoform characterization. Anal. Chem. **87**, 4152–4158 (2015)

47. Frese, C.K., Zhou, H., Taus, T., Altelaar, A.F., Mechtler, K., Heck, A.J., Mohammed, S.: Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). J. Proteome Res. **12**, 1520–1525 (2013)

48. Pliego, J.R., De Almeida, W.B.: A new mechanism for the reaction of carbenes with OH groups. J. Phys. Chem. A **103**, 3904–3909 (1999)

49. Griesbeck, A.G., El-Idreesy, T.T., Adam, W., Krebs, O.: CRC Handbook of Organic Photochemistry and Photobiology. CRC Press LLC, Boca Raton (2004)

50. Bourissou, D., Guerret, O., Gabbaï, F.P., Bertrand, G.: Stable carbenes. Chem. Rev. **100**, 39–92 (2000)

51. Richards, C.A., Kim, S.-J., Yamaguchi, Y., Schaefer, H.F.: Dimethylcarbene: a singlet ground state? J. Am. Chem. Soc. **117**, 10104–10107 (1995)

52. Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H., Cimermancic, P., Boehringer, D., Sali, A., Aebersold, R., Ban, N.: Molecular architecture of the 40SeIF1eIF3 translation initiation complex. Cell **158**, 1123–1135 (2014)

53. Beausoleil, S.A., Villen, J., Gerber, S.A., Rush, J., Gygi, S.P.: A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat. Biotechnol. **24**, 1285–1292 (2006)