RESEARCH ARTICLE

Efficient Covalent Bond Formation in Gas-Phase Peptide–Peptide Ion Complexes with the Photoleucine Stapler

Christopher J. Shaffer,[1] Prokopis C. Andrikopoulos,[2] Jan Řezáč,[2] Lubomír Rulíšek,[2] František Tureček[1]

[1]Department of Chemistry, University of Washington, Bagley Hall, Box 351700, Seattle, WA 98195-1700, USA
[2]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám 2, 16610, Prague, Czech Republic
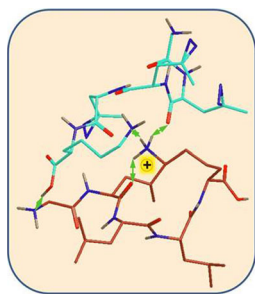
Abstract. Noncovalent complexes of hydrophobic peptides GLLLG and GLLLK with photoleucine (L*) tagged peptides $G(L*_nL_m)K$ ($n = 1,3$, $m = 2,0$) were generated as singly charged ions in the gas phase and probed by photodissociation at 355 nm. Carbene intermediates produced by photodissociative loss of N_2 from the L* diazirine rings underwent insertion into X–H bonds of the target peptide moiety, forming covalent adducts with yields reaching 30%. Gas-phase sequencing of the covalent adducts revealed preferred bond formation at the C-terminal residue of the target peptide. Site-selective carbene insertion was achieved by placing the L* residue in different positions along the photopeptide chain, and the residues in the target peptide undergoing carbene insertion were identified by gas-phase ion sequencing that was aided by specific ^{13}C labeling. Density functional theory calculations indicated that noncovalent binding to GL*L*L*K resulted in substantial changes of the $(GLLLK + H)^+$ ground state conformation. The peptide moieties in $[GL*L*LK + GLLLK + H]^+$ ion complexes were held together by hydrogen bonds, whereas dispersion interactions of the nonpolar groups were only secondary in ground-state 0 K structures. Born-Oppenheimer molecular dynamics for 100 ps trajectories of several different conformers at the 310 K laboratory temperature showed that noncovalent complexes developed multiple, residue-specific contacts between the diazirine carbons and GLLLK residues. The calculations pointed to the substantial fluidity of the nonpolar side chains in the complexes. Diazirine photochemistry in combination with Born-Oppenheimer molecular dynamics is a promising tool for investigations of peptide–peptide ion interactions in the gas phase.

Keywords: Peptide–peptide ion complexes, Laser photodissociation, Diazirine chromophores, Photoleucine, Born-Oppenheimer molecular dynamics

Introduction

Noncovalent protein–protein and protein–ligand interactions are the basis of molecular recognition of critical importance to many areas of biology. In addition to X-ray diffraction and spectroscopic methods of structure elucidation, cross-linking and photoaffinity labeling have been the major

Correspondence to: Lubomír Rulíšek; *e-mail:* lubos@uochb.cas.cz, František Tureček; *e-mail:* turecek@chem.washington.edu

chemical methods to study protein noncovalent interactions, dating back to the pioneering studies of Westheimer et al. [1] and Knowles and coworkers [2] in the 1960s. Detection of noncovalent interactions relies on crosslinking by covalent bond formation between the protein and ligand at their contact sites upon chemical reaction or photolysis. Photochemical crosslinking, in particular, utilizes highly reactive intermediates such as radicals, nitrenes [2], or carbenes [3] produced by photolysis of a chemically stable chromophore that is incorporated into the protein or ligand [4, 5]. With the introduction of photolabile amino acid residues photoleucine and photomethionine, it became possible to achieve photochemical crosslinking in complexes closely resembling native systems,

such as membrane protein [6] and protein–peptide complexes [7]. Photoleucine (L-2-amino-4,4-azi-pentanoic acid, abbreviated as L*) contains a diazirine ring that is a specific chromophore absorbing at 350–370 nm where the native peptide chromophores are transparent and are not photoactivated. Photon absorption causes N_2 elimination, creating a highly reactive singlet carbene that undergoes insertion into proximate X−H bonds (X = C, N, O) to form new covalent H−C−X bonds [8]. Competing with insertion, the carbene can undergo fast intramolecular isomerization by 1,2-hydrogen shift to form a nonreactive olefin in the photoactive component [9–12], which can no longer covalently bind to the complex counterpart. This side reaction provides an internal clock that limits the time scale for carbene-X–H bond interactions to within 10^{-7} s. The photochemical crosslinking ("stitching") can be readily detected by mass spectrometry, and the position of the new covalent link can be located by tandem mass spectrometry sequencing of gas-phase adduct ions [7]. A substantial drawback of solution studies is the very low yields of crosslinked products that must be separated from the unmodified starting material and side products [7, 13–15].

An entirely different approach to studying noncovalent protein–ligand and protein–protein complexes relies on mass spectrometry that has been used to study their stoichiometry [16, 17], stability [18–20], and thermochemistry in the gas phase [21, 22]. Formation of covalent bonds in gas-phase ion complexes has been reported for several systems that relied on collision-induced chemical reactions between an anion and a cation in a complex that mimicked chemical methods used in solution [23, 24]. In addition, covalent bond formation in gas-phase ion complexes with 18-crown-6-ether has been accomplished by collision-induced dissociation of diazomalonate derivatives [25] or photodissociation of a diazirine labeled peptide [26] in which reactive carbenes played the role of reactive intermediates. The singular example whereby photoactivation of a peptide ion led to amide bond formation was severely limited by indiscriminate regioselectivity, use of vacuum ultraviolet light (157 nm), and low yield (<1%) [27]. Herein, we exploit diazirine chromophores that are synthetically incorporated into peptides to generate reactive carbene intermediates for covalent intermolecular bond formation in noncovalent peptide–peptide complexes in the gas phase. This photoactivation scheme provides a unique opportunity for efficient bond formation as a tool for probing the conformational structure of gaseous peptide–peptide complex ions in the absence of surrounding solvent, ions, membranes, or other interacting medium. As the photoactive components, we chose hydrophobic pentapeptides GL*L*L*K, GL*LLK, GLL*LK, and GLLL*K. The first of these provides multiple reactive centers to undergo X−H bond insertion to a peptide counterpart in a nonspecific manner. The other three peptides provide sequence-specific reactive centers produced by photolysis. As counterparts in the noncovalent complexes, we employ hydrophobic peptides GLLLG and GLLLK. The GLLLG sequence motif appears in several transmembrane proteins, e.g., those of

the claudine family, where the Gly residues are flanked by additional hydrophobic residues (Val, Leu, Ile) and are immersed in the lipid membrane [28]. The GLLLK sequence motif appears in over 100 proteins, and the C-terminal Lys is often flanked by another polar residue (Lys, His) [29]. The size of the peptides used for this gas-phase study was deliberately limited to allow us to complement experimental investigations of photo-stitching with a conformational trajectory analysis using Born-Oppenheimer dynamics at an augmented semi-empirical level of quantum theory including dispersion interactions. In this way, both polar and dispersion interactions between the peptide components can be treated at the same level of theory, and their role in affecting the complex' stability and dynamics can be assessed. We wish to illustrate that this new photo-stitching approach achieves high yields of covalent bond formation, provides insight into the structure and dynamics of noncovalent peptide-peptide complexes, and allows us to elucidate the nature and fine features of interactions between the peptide moieties.

Experimental

Materials

All peptides were synthesized on Wang resin (Bachem Americas, Torrance, CA, USA) using commercially available Fmoc peptides and L-photoleucine (Life Technologies, Rockford, IL, USA) according to literature procedures [30, 31].

Methods

Collision induced dissociation (CID) mass spectra were measured on a modified LTQ-XL ETD linear ion trap (LIT) mass spectrometer (ThermoElectron Fisher, San Jose, CA, USA. Peptide solutions (5–10 μM) in 50/50/1 methanol/water/acetic acid were electrosprayed at 2.2–2.3 kV from a pulled fused silica capillary into an open microspray ion source described previously [32, 33]. Ion pair complexes were selected according to their m/z and stored within the linear anion trap. MS^n experiments were carried out by isolating the ions and exposing them to resonant collisional excitation or photoexcitation. The collisional activation times were typically varied between 30 ms for preliminary analysis and 1000 ms to create times similar to photodissociation conditions.

Photodissociation

To accomplish photodissociation of trapped ions in the LIT, the LTQ-XL ETD mass spectrometer was modified as reported previously [26, 34, 35]. Briefly, the chemical ionization (CI) source for the anion production was modified by drilling a 1-mm diameter hole into the insert block to provide a line of sight path to the LIT. The backside vacuum gate to the CI source was replaced by an aluminum plate carrying a quartz window. The irradiating light beam was produced by an EKSPLA NL 301 HT (Altos Photonics, Bozeman, MT, USA) Nd-YAG laser operating at 20 Hz frequency with a 3–6 ns pulse width. The

laser is equipped with a third harmonics frequency generator producing a single 355 nm wavelength at 120 mJ/pulse peak power. The typical light intensity used in the photodissociation experiments was 18 mJ/pulse. The laser beam of 6-mm diameter is aligned by mirrors and focused by a telescopic lens to pass the small aperture drilled in the CI source. The laser beam diameter in the LIT is estimated at 3–4 mm to ensure overlap with the trapped ions. Both the laser system and the LTQ-XL are set on an optical table for optimum alignment. The laser was interfaced to the LTQ by LabView software (National Instruments, Austin, TX, USA) that receives a signal from a TTL pulse on pin14 of the J1 connector on the LTQ console. Laser pulses are triggered internally by the EKSPLA system, but power is controlled for each pulse by commands from the LabView software. The typical experimental set up consists of selecting the ion to be photodissociated and storing it in the LIT for a chosen time period. For example, 400-ms storage time can accommodate up to seven laser pulses spaced by 50 ms. This allows one to vary the number of pulses and determine the photodissociation kinetics. Longer storage times of >3 s, allowing >60 laser pulses, are readily realized.

Computations

Molecular dynamics calculations were run using the AMBER 14 program [36] and the ff14SB force field [37] at the molecular mechanics level. Twenty-one initial structures were assembled that had different orientations of the photocleavable and auxiliary peptide units in which the charging proton was placed on the ε-amine group of one of the lysine residues. For each of these 21 initial structures, an MD run was performed to generate 250 snapshots per initial structure. These 5250 snapshot structures were fully optimized with semi-empirical calculations using PM6 [38] with dispersion corrections (PM6-D3H4) [39] through the MOPAC program [40] that was coupled to the Cuby framework [41], which implements the corrections and provides the molecular dynamics engine. The PM6-D3H4-optimized structures were sorted out by energy and 500 lowest energy structures were reoptimized by density functional theory (DFT) calculations using BLYP [42, 43]/def-SV(P) [44] with dispersion corrections [45]. These DFT calculations were carried out using Turbomole [46]. Twenty-two $[GL*L*K + GLLLK + H]^+$ complexes were selected that had different protonation sites, folding patterns, and orientations of the peptide subunits. The relative energies are summarized in Table S1 (Supplementary Data), the optimized geometries (Cartesian coordinates) can be obtained from the corresponding author upon request. Structures **1a** and **2a** representing the lowest-energy conformers with different orientations of the GL*L*L*K and $(GLLLK + H)^+$ subunits were further reoptimized with a dispersion-corrected hybrid functional ωB97X-D [47, 48] and the 6-31+G(d,p) basis set. These larger calculations produced 310 K enthalpies within 7 kJ mol^{-1} in agreement with the BLYP/def-SV(P) data (Supplementary Table S1). Twenty-five low-energy structures from the DFT optimizations were then used for semi-empirical Born-

Oppenheimer molecular dynamics (BOMD) calculations of trajectories using the augmented PM6-D3H4 method [39]. To perform these simulations, MOPAC program was coupled to the Cuby framework [41], which implements the corrections and provides the molecular dynamics engine. The trajectory calculations were run for 100 ps at 310 K corresponding to the experimental ion trap temperature. In a parallel effort to map the complex conformational space of the charged pentapeptide, structures of the $(GLLLK + H)^+$ ion conformers were generated using the ConformSearch engine described previously [49]. Relative free energies of low-energy conformers were obtained by single-point energy B3LYP [50], M06-2X [51], ωB97XD, and Moller-Plesset (MP2 frozen core) [52] calculations using the 6-311++G(2d,p) basis set. These calculations used the Gaussian 09 suite of programs [53]. The relative free energies are summarized in Supplementary Table S2 and the optimized structures of several low-energy conformers are shown in Supplementary Figure S1. Finally, a set of MD trajectories was obtained for the ester-bound peptide moieties in order to probe the conformations accessible for the remaining two carbenes in a covalent complex.

Results and Discussion

GLLLG + GL*L*L*K

For our initial attempt, we wanted to maximize possible carbene-target efficiencies by encouraging multiple interactions between the probe and target. Thus, we selected GL*L*L*K, with its high photoleucine content, as our first photochemical probe. To minimize structural complexity regarding protonation sites, we selected the non-basic GLLLG peptide as the first target, which lacks the C-terminal lysine, leaving the protonation site at the GL*L*L*K moiety. Selection and collision-induced dissociation (CID) of the $[GLLLG + GL*L*L*K + H]^+$ ion cluster at m/z 1050 to produce $[GL*L*L*K + H]^+$ at m/z 579 revealed that the noncovalent interaction between the two peptides was the weakest associative interaction in the cluster (Figure 1a). This is critically important in comparison to the possible dissociation of N_2 from the L* diazirine rings, which theoretically is another facile dissociation [32], but does not occur under these CID conditions. In comparison, treating the same ion cluster with a single photon pulse (18 mJ) at 355 nm resulted in the loss of a single unit of molecular nitrogen without dissociation of the ion pair (m/z 1022, Supplementary Figure S2). Concomitant with the loss of molecular nitrogen, photodissociation also resulted in a dissociation of the complex, forming the $[GL*L*L*K - N_2 + H]^+$ ion at m/z 551 (Supplementary Figure S2). Irradiation at the same wavelength for 19 pulses allowed for further losses of molecular nitrogen without disruption of the ion pair interaction, producing ions at m/z 994 and 966 from a second and third loss of N_2, respectively (Figure 1b). Overall, photodissociative loss of N_2 without peptide ion pair dissociation accounted for 43% of the total product ion channels from the parent $[GLLLG + GL*L*L*K +$
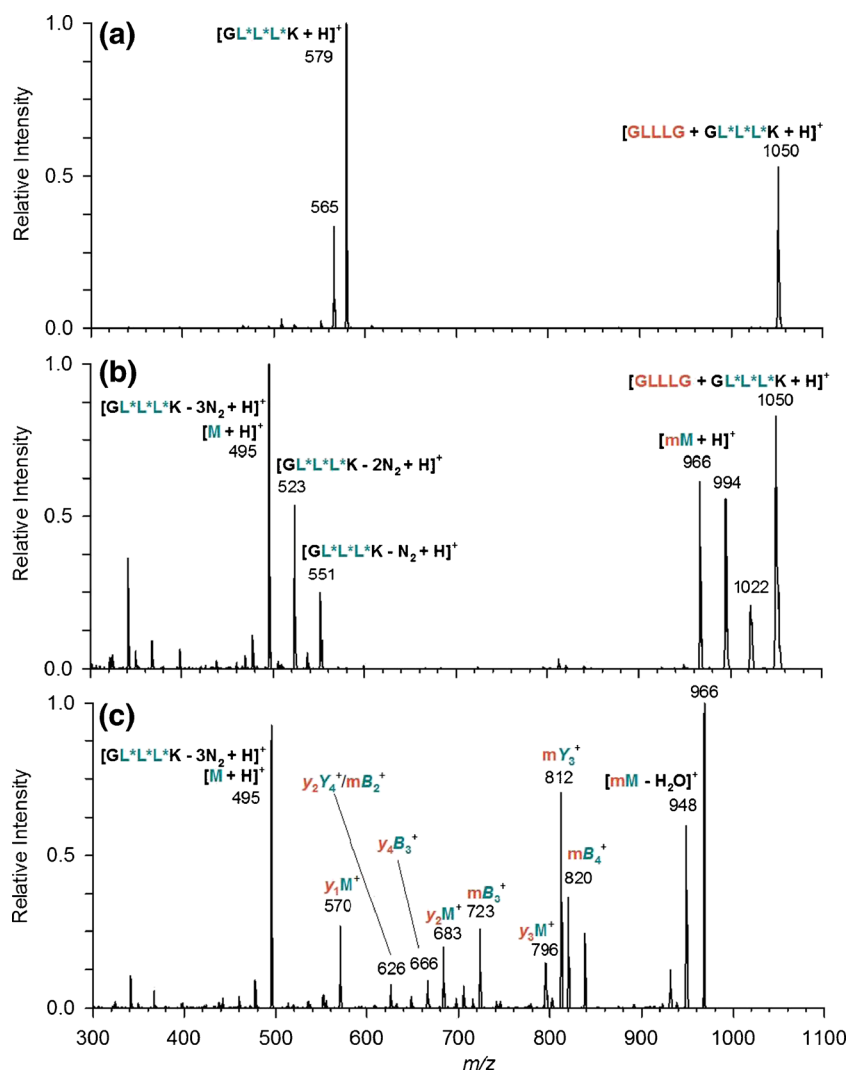
Figure 1. MS^2 spectrum of [GLLLG + GL*L*L*K + H]$^+$ obtained by (**a**) CID at NCE = 19 and (**b**) 355 nm irradiation for 19 pulses (18 mJ), (**c**) UVPD-CID-MS^3 spectrum (19 pulses at 18 mJ, NCE = 28) of the triple-N$_2$ loss product [GLLLG + GL*L*L*K − 3N$_2$ + H]$^+$, labeled as [mM + H]$^+$ here for short, and shown at m/z 966 in spectrum b. Fragments from the target peptide are identified in orange whereas those from the photochemical probe peptide are identified in cyan. See text and Figure 2 for additional explanation of the product ion nomenclature

H]$^+$ complex. This can be compared with amide bond formation via 157 nm irradiation where a similar, but endogenous, peptide dimer lost water without disruption of the noncovalent interaction, accounting for less than 1% of the total ion count [27].

Loss of molecular nitrogen from the ion pair complex does not, however, fully indicate that a bond-forming process has occurred between the two peptides. To this effect, we selected the ion pair complex that resulted from the loss of three N$_2$ units, [GLLLG + GL*L*L*K − 3N$_2$ + H]$^+$, and further subjected it to CID. The greater majority of these UVPD-CID-MS^3 product ions were found between the [GL*L*L*K − 3N$_2$ + H]$^+$ monomer (m/z 495) and the [GLLLG + GL*L*L*K − 3N$_2$ +H]$^+$ parent (m/z 966, Figure 1c), accounting for over 69% of the ion intensities in the spectrum. These ions were most remarkable because in light of the above control experiment they indicated that one or more covalent bonds have been

formed between the two peptide moieties upon diazirine photolysis. Thus, accounting for the yield of survivor complexes from the UVPD step (43%) and these new covalently cross-linked fragments from backbone dissociations on CID (69%), the overall yield for this photo-induced bond creation is 30% from the [GLLLG + GL*L*L*K + H]$^+$ parent.

Each of the major fragment ions from UVPD-CID-MS^3 can be assigned to a b/y-type backbone cleavage [54, 55], as used in the UVPD-CID-MS^3 study at 157 nm [27] that resulted in a single backbone fragmentation. In contrast, our UVPD-CID-MS^3 at 355 nm causes multiple backbone breakages in both chains of the covalently linked products, requiring a modified description. Because the reactive carbenes are in the L* side chains, the expected covalent links are of the L*-side-chain or L*-backbone type, allowing each original peptide to undergo collision-induced backbone dissociation independently. Such fragmentations, when they occur in tandem, do not allow

simple "b/y" nomenclature as in other new methods where crosslinking was carried out in solution and the products underwent a single dissociation [13–15].

Nomenclature for Adduct Fragment Ions

In light of the inadequacy of the established nomenclature, we have started using a new system of fragmentation nomenclature, which is beneficial for distinguishing each of the parent peptide molecules. For this, we label the photochemical probe peptide [GL*L*L*K – 3N$_2$] in capital letters as M and its fragments as B_m and Y_n. This system, as illustrated in Figure 2, then allows the target peptide, GLLLG (or m), and its fragments (b_m and y_n), to be placed immediately alongside the labels for the photochemical probe peptide. For example, the parent peptide ion complex [GLLLG + GL*L*L*K – 3N$_2$ + H]$^+$ can be written as [mM + H]$^+$ in short hand. Referring to the Figure 1c spectrum, the most abundant crosslinked fragment at m/z 812 is written as mY_3^+, indicating retention of the target (m) peptide moiety and backbone fragmentation in the photo probe. The ion at m/z 666, which is the product of backbone dissociations in both peptide moieties, can now easily be labeled as $y_4B_3^+$. Also of note is the ion at m/z 626, which can be assigned either as y_2Y_4, resulting from backbone cleavages in both crosslinked peptide chains, or mB_2 from backbone cleavage in the M chain only. This degeneracy has been removed and the peak was unambiguously assigned to mB_2 by ^{13}C labeling of GLLLG, whereby UVPD-CID-MS^3 of [G(^{13}C$_1$-L)LLG + GL*L*L*K + H]$^+$ gave an ion exclusively at m/z 627, retaining the target peptide Leu-2 label and indicating covalent stitching by the Leu*-2 residue (Supplementary Figure S3).

Analysis of the backbone fragment ions of [GLLLG + GL*L*L*K − 3N$_2$ + H]$^+$ produced by UVPD-CID-MS^3 between m/z 495 and 966 provided information regarding the likely sites of contact that led to photo-crosslinking (Figure 2). Fragment ions resulting from backbone dissociation of the target peptide (e.g., y_1M, y_4B_3, y_2M, and y_3M) point to a specific set of locations where photo-crosslinkage occurred. These specific ions constitute 58% of crosslinked backbone fragments. Out of these, the y_2M ion, which accounts for 4.2% of the total ion abundance, suggests that the same proportion of fragment ions must have a linkage between the photochemical probe peptide and residues 4 and 5 of GLL**LG**. These assignments and percentages for every ion identified in Figure 1c are summarized in Supplementary Table S3 and in the normalized form in Table 1. At 40%, the C-terminal Gly residue, Gly-5, appears the greatest number of times in product ions formed by the target peptide backbone fragmentation. The next highest abundance is for Leu-4, which accounts for 30%. The rest of the residues Gly-1 through Leu-3 range between 6% and 15%. Alone, these numbers suggest that the carbenes created in GL*L*L*K prefer the C-terminal Gly-5 over any of the other residues of GLLLG. Additionally, only very weak b-type ions (total 3%, Supplementary Table S3) are identified as having originated from the target peptide. This suggests that the Gly-5 residue is present in the majority of fragments of the target peptide. If we limited our analysis to these two comparisons and to this particular photo-peptide probe, it would appear that the Gly-5 is the major residue targeted by all three photoleucine residues. This, however, is not the case, and evidence for Leu* site specificity is provided below.

Site-Specific Photocrosslinking to GLLLG

To further locate site-specific interactions within the peptide–peptide ion complex, we systematically replaced a single L* along each leucine position in the photo-peptide probe by using three new peptides, GL*LLK, GLL*LK, and GLLL*K. The pertinent UVPD-CID-MS^3 relative ion intensities for the isomeric complexes are given in Supplementary Tables S4–S6 and the mass spectra are presented in Supplementary Figure S4a–c.

Beginning with the [GLLLG + GL*LLK + H]$^+$ complex, UVPD-CID-MS^3 creates fragments that account for 45% of the final dissociation event and 14% of the total ions in both steps
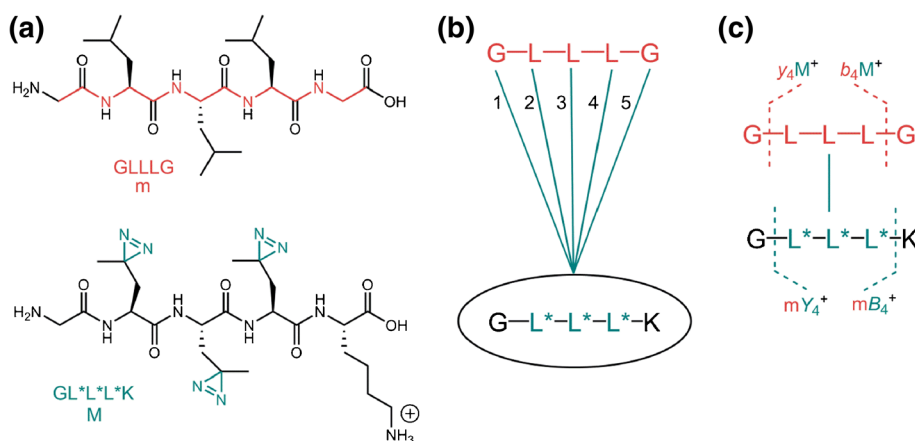
Figure 2. (a) 2D bond-line representation of peptide target GLLLG (orange) and photopeptide probe GL*L*L*K (cyan). (b) Simplified representation of GLLLG + GL*L*L*K ion pair from Table 1, with numbering to emphasize informational positions of peptide target GLLLG. (c) Simplified representation of covalently linked GLLLG/GL*L*L*K ion pair and a few theoretical fragments illustrating the nomenclature introduced in Figure 1c

Table 1. Normalized Intensities of GLLLG Target-Specific Peptide Fragments

GLLLG position	GL*L*L*K	GL*LLK	GLL*LK	GLLL*K
Gly-1	6	16	13	18
Leu-2	8	16	15	17
Leu-3	15	21	20	14
Leu-4	30	24	24	19
Gly-5	40	23	28	32

(Supplementary Figure S4a). One significant difference between these results with GL*LLK and the above ones with GL*L*L*K is the absence of the mY_3^+ ion (the most abundant fragment in Figure 1c), which is entirely consistent with the lack of a photochemical crosslinker in the Y_3^+ ion in GL*LLK. A detailed assignment of the MS^3 fragment ions was achieved with the help of specific ^{13}C labeling of the L residues in GLLLG (Supplementary Figure S5a, b), as described on p. S10 of the Supplementary Data. The fragmentation of the target peptide survivors can be analyzed statistically as it was above for GL*L*L*K. With GL*LLK, the target-specific fragment ions account for 45% of the crosslinked ions. When interacting with GL*LLK, the first three residues of GLLLG, Gly-1, Leu-2, and Leu-3, were more favorable for attachment than they were for GL*L*L*K, now accounting for 16 to 21% of the crosslinked ion count (Table 1). With GL*L*L*K the C-terminal residue Gly-5 was targeted in significant favor of the other four residues. In contrast, when using GL*LLK, the Gly-5 residue participates in only 23% of the crosslinked ion count. This conflict in results is additionally compounded in that $b_m B_n^+$ ions of many types are now observed in significance (Supplementary Table S4), which can only be a result of photo-crosslinkages occurring at positions other than the C-terminus.

Switching the photoleucine into the next position along the photo-peptide probe gives the [GLLLG + GLL*LK + H]$^+$ ion complex. UVPD-CID-MS^3 of this ion (Supplementary Table S5, Supplementary Figure S4b) again produces a spectrum quite similar to that of [GLLLG + GL*L*L*K + H]. The overall UVPD stitching efficiency remains high at 14%, including 43% for the CID of the photolyzed complex. The fragment ions assignable to backbone dissociations in the GLLLG target amount to 51% of all crosslinks. The abundances of each individual target residue amongst fragment ions is quite similar to that when GL*LLK was used, with a range of 13%–20% for the first three residues and 28% for the Gly-5 residue (Table 1). Backbone fragment ion assignment was achieved with the help of specific ^{13}C labeling of Leu residues in GLLLG (Supplementary Figure S6a, b) as described in the Supplementary Data.

Continuing on with GLLL*K, the overall stitching efficiency increased to 20% or 57% for the CID step (Supplementary Table S6, Supplementary Figure S4c), suggesting that the photoleucine at the fourth residue is the one most readily reacting with the target peptide. The sequence fragment ions assignable to the target GLLLG peptide amount to 48% of all crosslinks. Backbone fragment ion analysis was aided by ^{13}C labeling in the GLLLG target peptide (Supplementary

Figure S7), and the results are summarized in Table 1. The first residue, Gly-1, is now represented more frequently (18%) in crosslinked fragment ions, exceeding Leu-2 and Leu-3 at 17% and 14%. The C-terminus still dominates the overall statistics at 32%, but the trend no longer follows a pattern of Gly-5 > Leu-4 > Leu-3 > Leu-2 > Gly-1, indicating that each photoleucine residue is spatially related to the target peptide quite differently. This in effect provides a proof of concept that photoleucine can work as a structural probe in peptide ion clusters.

GLLLK + GL*L*L*K

The above results provided a proof of concept that photo-crosslinking can detail ion complex structure in the gas phase. To further extend the analysis, we aimed for an ion complex structure with a high order of symmetry, while maintaining the high probability for a photo-stitching. This ion complex, [GLLLK + GL*L*L*K + H]$^+$, isolated at m/z 1121 underwent CID-MS^2 and UVPD-MS^2 much in the same way as [GLLLG + GL*L*L*K + H]$^+$ did above. However, instead of producing [GL*L*L*K + H]$^+$, CID-MS^2 gave [GLLLK + H]$^+$ (m/z 543) (Supplementary Figure S8a), whereas the [GL*L*L*K + H]$^+$ ion (m/z 579) was not observed in significant relative abundance. This suggests that GLLLK has a much greater gas-phase basicity than GL*L*L*K, despite the similarities assumed when comparing these two molecules. Photodissociation with a single laser pulse caused loss of molecular nitrogen (m/z 1093, Supplementary Figure S8b). CID of this photofragment ion resulted in a loss of second and third nitrogen molecule (m/z 1065 and 1037) as well as formation of backbone fragments in the m/z 580–980 range (Supplementary Figure S8c). This suggests that one photon pulse is sufficient to induce cross-linking. Similarly, if the ions were treated with a longer train of laser pulses, the second and third N_2 molecules were selectively expelled without additional fragmentation, analogously to the Figure 1b spectrum for [GLLLG + GL*L*L*K + H]$^+$. Here, for [GLLLK + GL*L*L*K + H]$^+$, the product ions resulting from nitrogen loss account for 45% of fragment ion abundance.

In the full UVPD-CID-MS^3 experiment, the [GLLLK + GL*L*L*K – 3N$_2$ + H]$^+$ ion, or [mM + H]$^+$ (m/z 1037), was isolated and activated by CID (Figure 3a), showing a great abundance of backbone fragment ion peaks between the parent ion at m/z 1037 and the dissociated ion pair monomers [GLLLK + H]$^+$ and [GL*L*L*K – 3N$_2$ + H]$^+$ at m/z 543 and 495 respectively. These backbone fragments (Supplementary Table S7) account for 76% of the ions in the spectrum, and 34% of the total ions produced by both UVPD and CID activations. Of main interest in these assignments is the paucity of b-type fragments from the target peptide (Supplementary Table S7), suggesting that the photo-stitching occurs predominantly at its C-terminus, much like it was initially suggested for the GLLLG complex.

Site-Specific Photocrosslinking to GLLLK

We again used the systematic approach where a single photoleucine is moved down the photo-peptide probe's
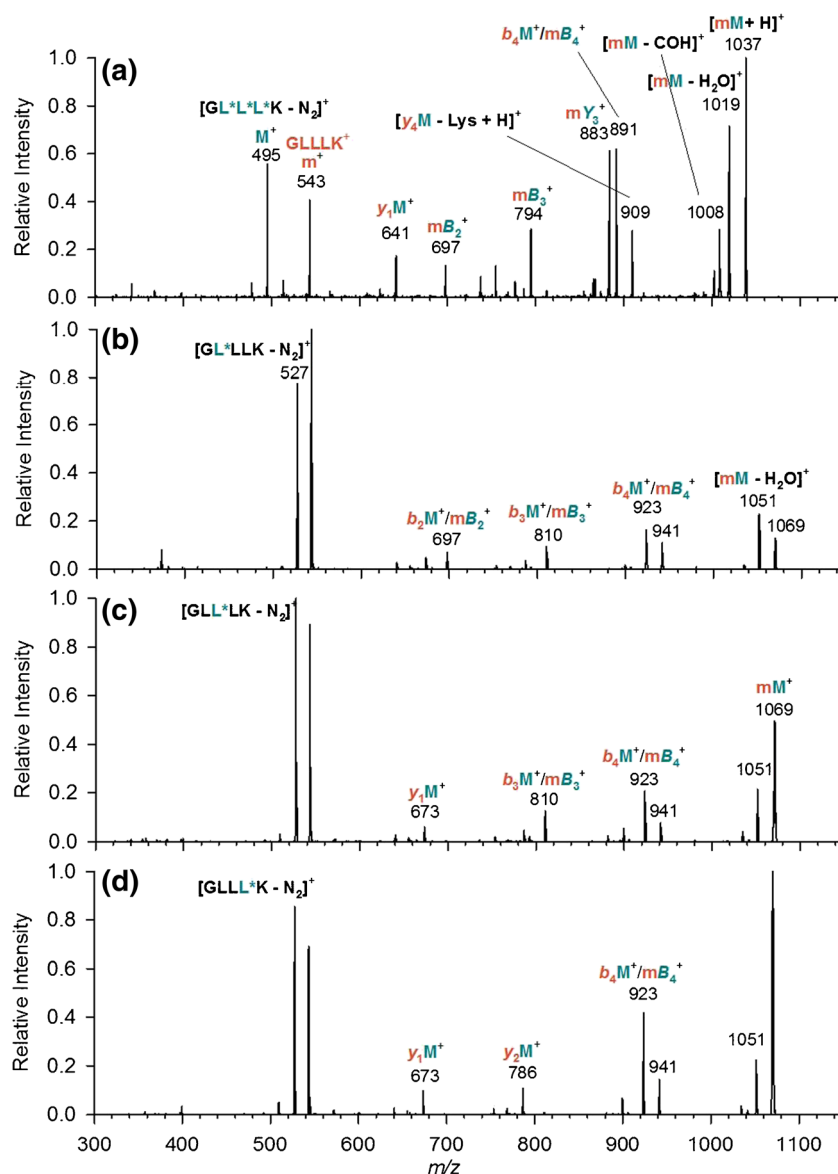
Figure 3. UVPD-CID-MS^3 spectrum (19 pulses at 18 mJ) of [mM + XN$_2$ + H]$^+$, where m represents GLLLK, X is variable from 1 to 3, and (**a**) M represents (GL*L*L*K – 3N$_2$), NCE = 25, (**b**) M represents (GL*LLK – N$_2$), NCE = 22, (**c**) M represents (GLL*LK – N$_2$), NCE = 22, and (**d**) M represents (GLLL*K – N$_2$), NCE = 22. Fragments from the target peptide are identified in orange whereas those from the photochemical probe peptide are identified in cyan

backbone. CID of these complexes again showed preferential protonation of the target GLLLK peptide (Supplementary Figure S9a–c). Starting with [GLLLK + GL*LLK + H]$^+$, fragmentation via UVPD-CID-MS^3 shows that a covalent bond has been created between the two peptide chains (Figure 3b). The assigned backbone fragments (Supplementary Table S8) accounted for 33% of ions in the final spectrum (or 20% of the total ions that survive both UVPD and CID activation events). Although all of the mB-type ions were nominally degenerate with bM-type ions, we eliminated the bM-type as a possibility because of their absence in the GL*L*L*K experiments above where they were nondegenerate. The fragment ions assignable to the target peptide sequence amounted to 21% of all

crosslinks, as most fragmentation occurred in the photo-active peptide moiety.

Moving the diazirine to the next leucine position, [GLLLK + GLL*LK + H]$^+$, created a UVPD-CID-MS^3 spectrum with nearly the same identified peaks (Figure 3c), except for the disappearance of the b_2M$^+$/mB_2^+ degenerate ion at m/z 698 (see Supplementary Table S9 for fragment ion assignment). The fragment ions assignable to the target peptide sequence amounted to 18% of all crosslinks. Notably, photo-stitched products account for 33% or 19% overall between the two UVPD-CID activations.

Finally, when the diazirine is located at the leucine closest to the C-terminus (GLLL*K), the general pattern of the spectrum

is still much the same (Figure 3d). For fragment ion assignment see Supplementary Table S10. Of greatest importance is the increased yield for the photo-stitched ions when measured by CID activation alone, which is now 45% of the ion count, or 26% when summarized in total between both UVPD and CID activations. This significant increase in yield does suggest that the contact between photoleucine and the GLLLK target peptide is most frequent at the Leu-4 position of the photo-peptide probe. The proportion of fragment ions assignable to the target peptide sequence also slightly increased to 25% of all crosslinks.

The above systematic analysis demonstrates a trend where the photo-peptide probe has strong preference for the C-terminus of the target peptide. This is further demonstrated in the statistical analysis of the survival of target peptide fragments in Table 2. Using GL*L*L*K, the photo-peptide probe with the greatest statistical likelihood for probing the entirety of the target peptide space, no fragment ions can be identified to include Gly-1 of the target peptide. This trend is followed throughout the rest of the photo-peptide probe sequences GL*LLK, GLL*LK, and GLLL*K. Target peptide fragments containing the Leu-2 and Leu-3 residues survive to account for 11% of the target fragment ion count with GL*L*L*K, but the probabilities still trend overwhelmingly towards the C-terminus with 28% at Leu-4 and 50% at Lys-5. For GL*LLK and the rest of the series, this trend is even stronger, where no target peptide fragments include Gly-1 or Leu-2, and Leu-3 is present in only 9% of the target peptide fragments. Target peptides with the Leu-4 residue carry over into 33% of fragment ions, but this is overshadowed by Lys-5, which appears in 58%. Fewer and fewer fragments of the target peptide include positions 1, 2, 3, or 4 as the photoleucine moves down the backbone of the photo-peptide probe. By the time GLLL*K is used, only Lys-5 is present in any of the fragments where target peptide fragmentation occurs. These trends can of course result from differences in the propensity of the backbone amide bonds to fragment as a result of the basicity and mobility of the proton within the peptide ion complex. Thus caution should be taken not to over-interpret the data.

Molecular Dynamics Simulations

To model and analyze noncovalent interactions in the peptide complexes, we first selected the [GLLLK + GL*L*L*K + H]$^+$ ion system. This was motivated by the steric similarity of the L* and L residues [6, 32], so that GL*L*L*K can be used as a surrogate for GL*LLK, GLL*LK, and GLLL*K

**Table 2.** Normalized Intensities of GLLLK Target-Specific Peptide Fragments

GLLLK position	GL*L*L*K	GL*LLK	GLL*LK	GLLL*K
Gly-1	0	0	0	0
Leu-2	11	0	0	0
Leu-3	11	9	0	0
Leu-4	28	33	30	0
Lys-5	50	58	70	100

peptides in their respective complexes. At the same time, molecular dynamics trajectory calculations of [GLLLK + GL*L*L*K + H]$^+$ allow one to simultaneously track the behavior of all three L* residues in a comprehensive and economical fashion. We selected 25 lowest-energy complex structures from DFT optimizations and ran trajectory calculations for 100 ps at 310 K corresponding to the experimental ion trap temperature. The 100 ps time frame is compatible with the half-life of a carbene intermediate, which is expected to be in the 0.1–5 ns range [9–12].

The fully optimized static structures of the complexes, corresponding to local energy minima (Supplementary Table S1), already revealed two significant features. The first of these was that regardless of where the ionizing proton was placed in the initially guessed structures, gradient optimization yielded complexes protonated at the GLLLK lysine side chain. This is consistent with the CID spectrum of the [GLLLK + GL*L*L*K + H]$^+$ complex that chiefly produced [GLLLK + H]$^+$ (vide supra), preserving the Lys protonation site at the target peptide in the complex. The other feature follows from the comparison of low-energy conformer structures of [GLLLK + H]$^+$ and those of [GLLLK + GL*L*L*K + H]$^+$ complexes. All low-energy [GLLLK + H]$^+$ conformers, such as **11** and **13** (Figure 4, for other structures see Supplementary Figure S1) show the Lys ε-ammonium group tricoordinated by hydrogen bonds to the backbone amide carbonyls. In stark contrast, coordination of the GLLLK Lys ε-ammonium in the complexes is accomplished by hydrogen bonding to the Lys ε-amine and amide carbonyl of the GL*L*L*K counterpart, whereas only one hydrogen bond is provided by the GLLLK amide carbonyls. In addition, both peptide moieties in the complex are linked by a strong hydrogen bond between the COOH and N-terminal amine groups at 1.58-1.65 Å (structures **2a** and **1a** in Figure 4). This indicates that coordination to GL*L*L*K of the target GLLLK peptide results in substantial changes of the ion conformation. In addition, the nature of the hydrogen bonding interactions between the peptide components in the complexes attests to the low importance of dispersion noncovalent interactions of the hydrophobic side-chain groups for binding of these peptides.

BOMD calculations were run for 100 ps starting from 25 optimized structures of [GLLLK + GL*L*L*K + H]$^+$ complexes. The trajectories were analyzed for close contacts between the diazirine carbon atoms in GL*L*L*K (C-105, C-121, and C-137) and the carbons and heteroatoms in GLLLK to assess the possible stitching sites upon photodissociation. This presumes that the exothermic barrierless insertion of the nascent carbene into a proximate X−H bond is faster than the conformational motion of the side chains and residues in the complex. For atom numbering see Figure 4. The closest contacts between the incipient carbene carbon and C, N, O of the target peptide were estimated as 4-5 Å from atomic van der Waals radii for C, H, N, and O, 1.85, 1.2, 1.4, and 1.54 Å, respectively, and bond geometries in several different orientations. The BOMD analysis revealed a critical effect of intra-complex thermal motion in establishing close contacts between
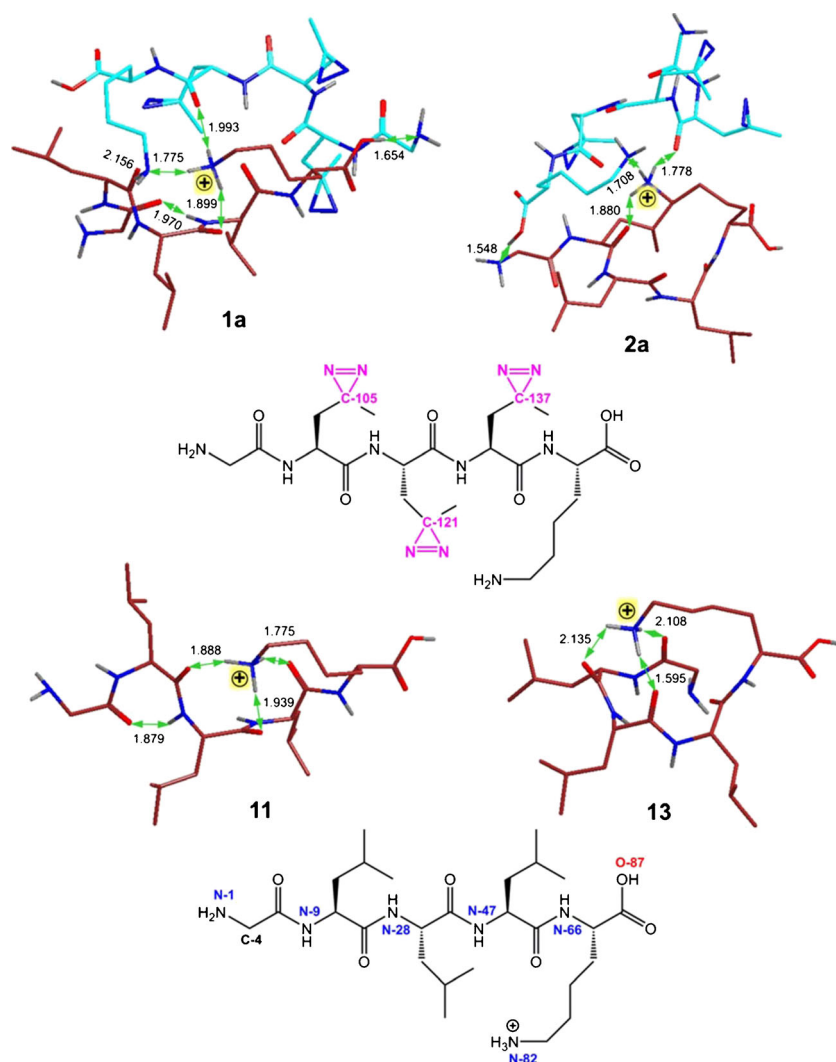
Figure 4. ωB97XD/6-31+G(d,p) optimized structures of low-energy [GLLLK + GL*L*L*K + H]⁺ conformers **1a** and **2a** and (GLLLK + H)⁺ conformers **11** and **13**. Carbon atoms in the GLLLK (bronze) and GL*L*L*K (turquoise) moieties are distinguished by color. Green double-ended arrows show major hydrogen bonds with distances in Ångstrøms. Also shown is the atom numbering in GLLLK and GL*L*L*K

the peptide units. This is illustrated in Figure 5, which shows the percentage of time during the 100 ps BOMD run that the carbon atoms in GL*L*L*K stayed in a 4.5-Å contact with the carbon atoms of the GLLLK residues. The red bars indicate contacts that were also present in the local energy minima corresponding to optimized structures at 0 K and $t = 0$, black bars indicate contacts that developed only as a result of thermal motion at 310 K. For example, in complex **1a** (Figure 5, top panel) there was a Lys-Lys contact in the structure corresponding to a local energy minimum that was retained in 52% of snapshots generated by BOMD at 310 K. In contrast, the most persistent contact (61% of time) between the the GL*L*L*K lysine and Leu-2 of the target peptide developed only as a result of thermal motion. In spite of the thermal breathing of the peptide conformations, the individual complex conformers on average retained their identity throughout the BOMD run. This

is documented by the contact graph of complex **2a** (Figure 5, bottom panel) that displays frequent Leu*-2 …Leu-2 and Leu*-4…Leu-2 contacts that are not present in complex **1a**. Conversely, the most frequent Lys…Leu-2 contact in **1a** is not represented at all in **2a**. Supplementary Figure S10 documents that a level of identity retention was found for each of the other eight complex conformers (**1b-3b**).

To comprise the conformational variability of these peptide complexes, the carbon contacts were summed up for BOMD trajectories of all 25 conformers, as shown in Supplementary Figure S11. Of particular importance were contacts of the photoleucine residues, Leu*-2 through Leu*-4. Out of these, Leu*-2 showed preferential contacts with carbons at Leu-2 and Leu-4 in the target peptide. Leu*-3 showed nonspecific contacts with carbons at all target residues, and Leu*-4 showed preferential contacts with carbons at Gly-1, Leu-2, and Leu-4.
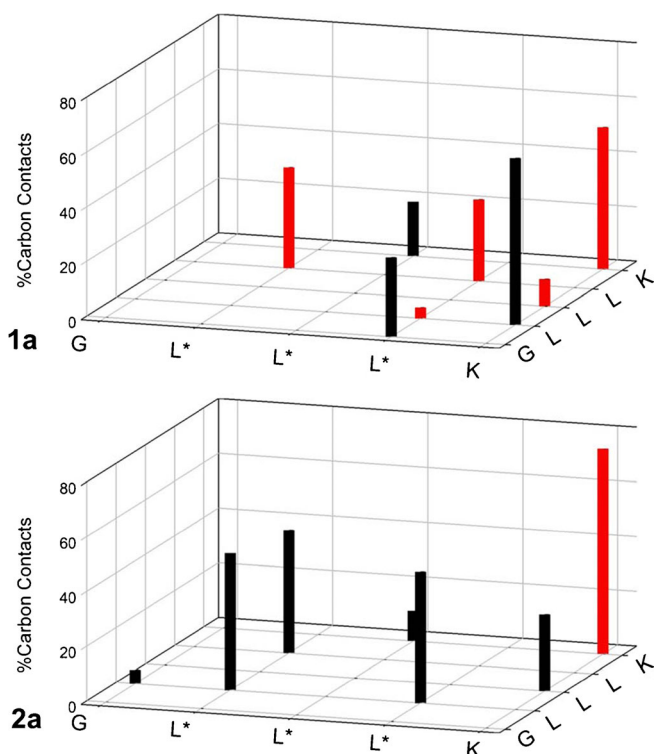
Figure 5. Close contacts within 4.5 Å between the incipient carbenes (C-105, C-121, and C-137) in GL*L*L*K and carbon atoms of GLLLK in conformers **1a** (top panel) and **2a** (bottom panel). Bars show percentage of contacts along the 100 ps BOMD trajectory. Red bars indicate contacts also present in the local energy minimum

In addition to carbon contacts of the incipient carbene reactants, one must also consider the substantial basicity of aliphatic carbenes [56, 57] that can result in preferred insertion into heteroatom−H bonds, such as N−H and O−H (for heteroatom

numbering in GLLLK see Figure 4). Carbene contacts with heteroatoms of the target peptide are displayed in Supplementary Figure S11 bottom and sorted out by their residues. This analysis indicates that Leu*-2 has frequent contacts with the Gly-1, Leu-3, Leu-4, and Lys residues of the target peptide, which are complementary to its carbon contacts. Hence, Leu*-2 is expected to show low residue selectivity in photochemical stitching to the target peptide. In contrast, BOMD analysis leads to frequent contacts of Leu*-3 with the carboxyl group of the target Lys residue, indicating a preferential Leu*-3-Lys cross-linking. Leu*-4 showed similar behavior to Leu*-2 with somewhat higher preference for close contacts with Gly-1.

The relationship of the BOMD contact analysis to the experimentally observed crosslinked fragment ions is visualized for the lowest-energy conformer **1a** (Figure 6). This shows the calculated percentage of contacts between the incipient Leu*-2 through Leu*-4 carbenes and the carbon (black letters), nitrogen (blue letters), and carboxyl oxygen (red letters) atoms in the residues in the target peptide molecule. The predicted fragment ion populations, y_1M (13%), ($mB_2 + mB_3$) (20%), mY_3 (23%), ($mB_4 + b_4M$) (27%), and ($y_4M - $ Lys) (16%), are remarkably close to the experimental fragment ion relative intensities (Figure 3a). Although this result should not be over-interpreted because of the potentially uneven backbone fragmentation of the peptide complex ion, it nevertheless points to the realistic representation of the peptide dynamics in the complex. It is also consistent with the previously postulated extremely low activation energies for carbene insertion [9–12], justifying the assumption that covalent bond formation occurs with high probability upon close contact between the carbene and the X−H bond. The overall conclusion of this theoretical analysis
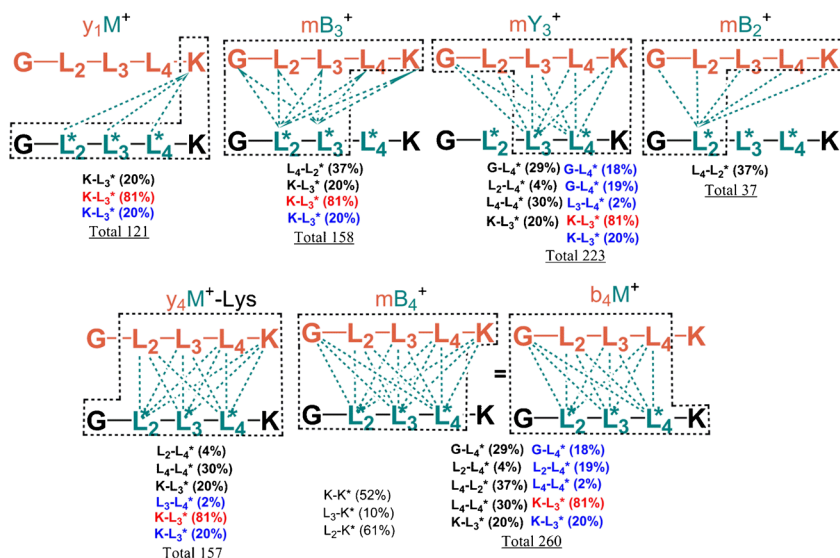
Figure 6. BOMD-calculated contacts in the [GL*L*L*K+GLLLK]+ complex **1a**. The black dotted outlines indicate the residues present in the fragment ions. Incipient carbene 4.5 Å contacts with C–H, N–H, and O–H bonds in the target peptide are indicated by black, blue, and red characters, respectively

is that photolysis of the [GLLLK + GL*L*L*K + H]$^+$ complexes should result in largely nonspecific formation of covalent bonds in all residues of the target GLLLK peptide. This result complements the sequence analysis of the experimental photodissociation mass spectra.

Diazirine ring photodissociation in GL*L*L*K most likely occurs at a random order because the absorbance of the diazirine chromophore is very much constant and does not depend on its position in the complex, as inferred from photodissociation of positionally labeled G[L*,L,L]K peptide complexes. From the pulse dependence of photodissociation (Supplementary Figure S8b, c), it appears that each laser pulse creates a reactive carbene at only one of the L* residues. The laser pulses are spaced by 50 ms, which is substantially longer than the lifetime of the carbene, so that the reactive intermediate formed in the first pulse undergoes complete insertion or quenching by isomerization to an olefin before the arrival of the next laser pulse. The presence of three L* residues in GL*L*L*K raises the question of whether photo-stitching by the first carbene generated from L* can affect the probability of covalent bond formation upon photodissociation of the second and third diazirine ring. In principle, this issue could be addressed by synthesizing adducts in which one L* residue was used to form a covalent bond while the other two L* were retained and to study their photodissociation. However, the variational complexity of such adducts would make this approach a daunting task. Furthermore, double crosslinking would create ring structures that can be expected to be more resistant to backbone fragmentation, thus reducing information obtainable from gas-phase sequencing. We resorted to a more elegant solution using constructs, in which GL*(L*-N$_2$)L*K and G(L*-N$_2$)L*L*K were covalently linked by insertion into the GLLLK carboxyl forming an ester bond in **1c**-ester and **2a**-ester, respectively (Supplementary Figure S12). The constructs were then analyzed by BOMD trajectory calculations. The choice of the ester covalent bond was based on the Table 1 data indicating preferential stitching to the C-terminal residue. The contact analysis disclosed different results for the Leu*-2 and Leu*-3 ester complexes. Crosslinking of Leu*-2 to the GLLLK carboxyl in **2a**-ester (Supplementary Figure S12, right panel) hampers contacts of Leu*-3 with C−H bonds and most heteroatom X−H bonds, while enhancing contact of Leu*-4 with the ester group, which however cannot undergo further X−H insertion. Crosslinking of Leu*-3 to the GLLLK carboxyl in **1c**-ester (Supplementary Figure S12, left panel) enhances contacts of Leu*-2 and Leu*-4 with the Gly-1 and Leu-2 amide nitrogens, possibly resulting in carbene insertion in the pertinent N−H bonds. The CID spectrum of [GLLLK + GL*L*L*K − 3N$_2$ + H]$^+$ shows prominent mB_3 and mB_4 ions. The first of these backbone fragments can be produced when an initial Leu*-3→Lys stitching was followed by Leu*-2→Gly-1 and Leu*-2→Leu-2 bond formation. Conversely, the presence of the y$_n$M fragment ions in the CID spectra indicates the formation of singly stitched adducts in which the other Leu* side chains were converted to nonreactive olefins.

Conclusions

Photodissociation of peptide–peptide ion complexes incorporating diazirine rings in photoleucine residues results in efficient covalent bond formation reaching 30% yields. The regioselectivity of covalent crosslinking by carbene insertion into X−H bonds depends on hydrogen bonding interactions of the charged and neutral polar groups that are the major determinant of the complexes' structure. BOMD trajectory calculations provide evidence of a substantial fluidity of the backbone segments and hydrophobic side chains in these complexes that can engage in multiple contacts caused by thermal motion. In spite of this fluidity, isomers differing in the position of the photoleucine residue and even different conformers of a given complex retain some specificity as to the diazirine-residue contacts and sites of covalent bond formation. This leads to the notion of the complexes consisting of a backbone core, maintained by polar interactions with limited mobility, and fluid hydrophobic side chain groups. The virtual lack of hydrophobic interactions in affecting the equilibrium structure is a salient feature of the gas-phase complexes that provides a dramatic contrast to condensed-phase structures.

Acknowledgments

References

1. Singh, A., Thornton, E.R., Westheimer, F.H.: Photolysis of diazoacetylchymotrypsin. J. Biol. Chem. **237**, PC3006–PC3008 (1962)
2. Fleet, G.W.J., Porter, R.R., Knowles, J.R.: Affinity labeling of antibodies with aryl nitrene as reactive group. Nature (London) **224**, 511–512 (1969)
3. Smith, R.A.G., Knowles, J.R.: Aryldiazirines: potential reagents for photolabeling of biological receptor sites. J. Am. Chem. Soc. **95**, 5072–5073 (1973)
4. Knowles, J.R.: Photogenerated reagents for biological receptor-site labeling. Acc. Chem. Res. **5**, 155–160 (1972)
5. Das, J.: Aliphatic diazirines as photoaffinity probes for proteins: recent developments. Chem. Rev. **111**, 4405–4417 (2011)
6. Suchanek, M., Radzikowska, A., Thiele, C.: Photoleucine and photomethionine allow identification of protein–protein interactions in living cells. Nat. Methods **2**, 261–268 (2005)
7. Koelbel, K., Ihling, C.H., Sinz, A.: Analysis of peptide secondary structures by photoactivatable amino acid analogues. Angew. Chem. Int. Ed. **51**, 12602–12605 (2012)
8. Liu, M.T.H.: Chemistry of Diazirines, Vols. I and II. CRC Press, Boca Raton, FL (1987)
9. Seburg, R.A., McMahon, R.J.: Photochemistry of matrix-isolated diazoethane and methyldiazirine: ethylidene trapping? J. Am. Chem. Soc. **114**, 7183–7189 (1992)
10. Jackson, J.E., Soundararajan, N., Whie, W., Liu, M.T.H., Bonneau, R., Platz, M.S.: Measurement of the absolute rate of 1,2-hydrogen migration in benzylchlorocarbene. J. Am. Chem. Soc. **111**, 6874–6875 (1989)

11. Pezacki, J.P., Couture, P., Dunn, J.A., Warkentin, J., Wood, P.D., Lusztyk, J., Ford, F., Platz, M.S.: Rate constants for 1,2-hydrogen migration in cyclohexylidene and in substituted cyclohexylidenes. J. Org. Chem. **64**, 4456–4464 (1999)

12. Stevens, I.D.R., Liu, M.T.H., Soundararajan, N., Paike, N.: The barrier for 1,2-hydrogen shift in dialkyl carbenes. Tetrahedron Lett. **30**, 481–484 (1989)

13. Mueller, M.Q., Sinz, A.: Chemical cross-linking and high-resolution mass spectrometry to study protein-drug interactions. Methods Mol. Biol. **803**(Chemical Proteomics), 205–218 (2012)

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

15. Clavier, S., Bolbach, G., Sachon, E.: Photocross-linked peptide-protein complexes analysis: a comparative study of CID and ETD fragmentation modes. J. Am. Soc. Mass Spectrom. **26**, 1014–1026 (2015)

16. Heck, A.J.R., van den Heuvel, R.H.H.: Investigation of intact protein complexes by mass spectrometry. Mass Spectrom. Rev. **23**, 368–389 (2004)

17. Andersen, C.B.F., Torvund-Jensen, M., Nielsen, M.J., Pinto de Oliveira, C.L., Hersleth, H.-P., Andersen, N.H., Pedersen, J.S., Andersen, G.R., Moestrup, S.K.: Structure of the haptoglobin–hemoglobin complex. Nature **489**, 456–459 (2012)

18. Liu, L., Bagal, D., Kitova, E.N., Schnier, P.D., Klassen, J.S.: Hydrophobic protein–ligand interactions preserved in the gas phase. J. Am. Chem. Soc. **131**, 15980–15981 (2009)

19. Deng, L., Broom, A., Kitova, E.N., Richards, M.R., Zheng, R.B., Shoemaker, G.K., Meiering, E.M., Klassen, J.S.: Kinetic stability of the streptavidin-biotin interaction enhanced in the gas phase. J. Am. Chem. Soc. **134**, 16586–16596 (2012)

20. Kitova, E.N., Sinelnikov, I., Deng, L., Klassen, J.S.: Thermal dissociation of streptavidin homotetramer in the gas phase: subunit loss versus backbone fragmentation. Int. J. Mass Spectrom. **345/347**, 97–103 (2013)

21. Felitsyn, N., Kitova, E.N., Klassen, J.S.: Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior. Anal. Chem. **73**, 4647–4661 (2001)

22. Liu, L., Baergen, A., Michelsen, K., Kitova, E.N., Schnier, P.D., Klassen, J.S.: Energetics of intermolecular hydrogen bonds in a hydrophobic protein cavity. J. Am. Soc. Mass Spectrom. **25**, 742–750 (2014)

23. Mentinova, M., McLuckey, S.A.: Covalent modification of gaseous peptide ions with N-hydroxysuccinimide ester reagent ions. J. Am. Chem. Soc. **132**, 18248–18257 (2010)

24. McGee, W.M., McLuckey, S.A.: Efficient and directed peptide bond formation in the gas phase via ion/ion reactions. Proc. Natl. Acad. Sci. U. S. A. **111**, 1288–1292 (2014)

25. Julian, R.R., May, J.A., Stoltz, B.M., Beauchamp, J.E.: Biomimetic approaches to gas phase peptide chemistry: combining selective binding motifs with reactive carbene precursors to form molecular mousetraps. Int. J. Mass Spectrom. **228**, 851–864 (2003)

26. Shaffer, C.J., Marek, A., Pepin, R., Slováková, K., Tureček, F.: Combining UV photodissociation with electron transfer for peptide structure analysis. J. Mass Spectrom. **50**, 470–475 (2015)

27. Lee, S., Valentine, S.J., Reilly, J.P., Clemmer, D.E.: Controlled formation of peptide bonds in the gas phase. J. Am. Chem. Soc. **133**, 15834–15837 (2011)

28. Furuse, M., Fujita, K., Hiiragi, T., Fujimoto, K., Tsukita, S.: Claudin-1 and -2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. J. Cell Biol. **141**, 1539–1550 (1998)

29. Basic Local Alignment Search Tool (BLAST) Available at: http://blast.ncbi.nlm.nih.gov. Accessed 20 Dec 2015

30. Janz, J.M., Ren, Y., Looby, R., Kazmi, M.A., Sachdev, P., Grunbeck, A., Haggis, L., Chinnapen, D., Lin, A.Y., Seibert, C., McMurry, T., Carlson, K.E., Muir, T.W., Hunt, S., Sakmar, T.P.: Direct interaction between an allosteric agonist pepducin and the chemokine receptor CXCR4. J. Am. Chem. Soc. **133**, 15878–15881 (2011)

31. Coste, J., LeNguyen, D., Castro, B.: PyBOP: a new peptide coupling reagent devoid of toxic by-product. Tetrahedron Lett. **31**, 205–208 (1990)

32. Marek, A., Tureček, F.: Collision-induced dissociation of diazirine-labeled peptide ions. Evidence for Brønsted-acid assisted elimination of nitrogen. J. Am. Soc. Mass Spectrom. **25**, 779–789 (2014)

33. Marek, A., Pepin, R., Peng, B., Laszlo, K.J., Bush, M.F., Tureček, F.: Electron transfer dissociation of photolabeled peptides. Backbone

34. cleavages compete with diazirine ring rearrangements. J. Am. Soc. Mass Spectrom. **24**, 1641–1653 (2013)

34. Shaffer, C.J., Marek, A., Nguyen, H.T.H., Tureček, F.: Combining near-UV photodissociation with electron transfer. Reduction of the diazirine ring in a photomethionine-labeled peptide ion. J. Am. Soc. Mass Spectrom. **26**, 1367–1381 (2015)

35. Nguyen, H.T.H., Shaffer, C.J., Pepin, R., Tureček, F.: UV action spectroscopy of gas-phase peptide radicals. J. Phys. Chem. Lett. **6**, 4722–4727 (2015)

36. Case, D.A., Berryman, J.T., Betz, R.M., Cerutti, D.S., Cheatham III, T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T.S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K.M., Monard, G., Needham, P., Nguyen, H., Nguyen, H.T., Omelyan, I., Onufriev, A., Roe, D.R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C.L., Smith, W., Swails, J., Walker, R.C., Wang, J., Wolf, R.M., Wu, X., York, D.M., Kollman, P.A.: AMBER 2015. University of California, San Francisco (2015)

37. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., Simmerling, C.: ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput. **11**, 3696–3713 (2015)

38. Stewart, J.J.P.: Optimization of parameters for semi-empirical methods. V. Modification of NDDO approximations and application to 70 elements. J. Mol. Model. **12**, 1173–1213 (2007)

39. Řezáč, J., Fanfrlík, J., Salahub, D., Hobza, P.: Semi-empirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. J. Chem. Theory Comput. **5**, 1749–1760 (2009)

40. MOPAC 2012, Stewart, J.J.P.; Stewart Computational Chemistry: Colorado Springs, CO, USA

41. Řezáč J.: Cuby—computational chemistry framework, ver. 4. Available at: http://cuby4.molecular.cz. Accessed 1 Jul 2015

42. Becke, A.D.: Density-functional exchange-energy approximation with correct asymptotic behavior. Phys. Rev. A **38**, 3098–3100 (1988)

43. Lee, C., Yang, W., Parr, R.G.: Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B **37**, 785–789 (1988)

44. Schäfer, A., Horn, H., Ahlrichs, R.: Fully optimized contracted Gaussian basis sets for atoms lithium to krypton. J. Chem. Phys. **97**, 2571–2577 (1992)

45. Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. **132**, 154104/1–19 (2010)

46. TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2014, TURBOMOLE GmbH, since 2007; available at: http://www.turbomole.com. Accessed 7 May 2015

47. Chai, J.D., Head-Gordon, M.: Systematic optimization of long-range corrected hybrid density functionals. J. Chem. Phys. **128**, 084106/1–084106/15 (2008)

48. Chai, J.D., Head-Gordon, M.: Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. Phys. Chem. Chem. Phys. **10**, 6615–6620 (2008)

49. Moss, C.L., Chamot-Rooke, J., Brown, J., Campuzano, I., Richardson, K., Williams, J., Bush, M., Bythell, B., Paizs, B., Tureček, F.: Assigning structures to gas-phase peptide cations and cation-radicals. An infrared multiphoton dissociation, ion mobility, electron transfer, and computational study of a histidine peptide ion. J. Phys. Chem. B **116**, 3445–3456 (2012)

50. Becke, A.D.: A new mixing of Hartree-Fock and local density-functional theories. J. Chem. Phys. **98**, 1372–1377 (1993)

51. Zhao, Y., Truhlar, D.G.: The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor. Chem. Acc. **120**, 215–241 (2008)

52. Møller, C., Plesset, M.S.: A note on an approximation treatment for many-electron systems. Phys. Rev. **46**, 618–622 (1934)

53. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J.A., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann,

R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox, D.J.: Gaussian 09, Revision A.02. Gaussian, Inc, Wallingford (2009)

54. Roepstorff, P., Fohlman, J.: Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass. Spectrom. **11**, 601 (1984)

55. Biemann, K.: Nomenclature for peptide fragment ions (positive ions). Methods Enzymol. **193**, 886–887 (1990)

56. Dix, E.J., Goodman, J.L.: Protonation of diarylcarbenes by alcohols: the importance of ion pair dynamics. J. Phys. Chem. **98**, 12609–12612 (1994)

57. Mieusset, J.-L., Brinker, U.H.: The nature and extent of pi-stabilization within foiled carbenes. J. Am. Chem. Soc. **128**, 15843–15850 (2006)