

Supplementary Information for

Dissecting Diazirine Photo-reaction Mechanism for Protein Residue-Specific Cross-linking and Distance Mapping

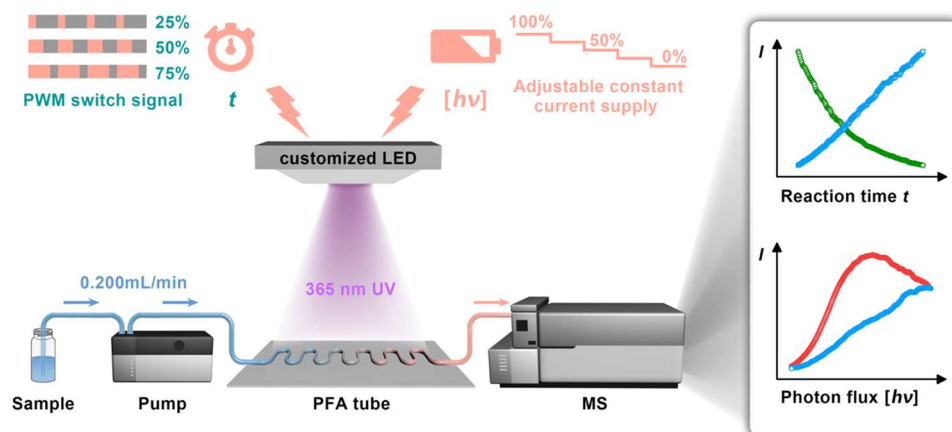
Yida Jiang¹, Xinghe Zhang¹, Honggang Nie^{1,2}, Jianxiong Fan¹, Shuangshuang Di^{1,2}, Hui Fu^{1,2}, Xiu Zhang^{1,2}, Lijuan Wang^{1,2}, Chun Tang^{1,3*}

¹Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. ²Analytical Instrumentation Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. ³Center for Quantitative Biology, PKU-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

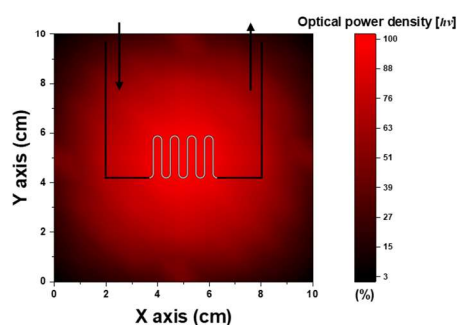
Email: Tang_Chun@pku.edu.cn

Supplementary Figures

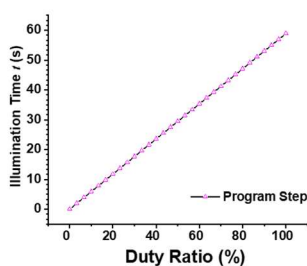
a



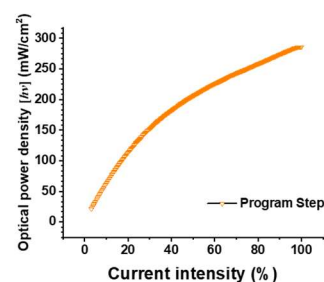
b



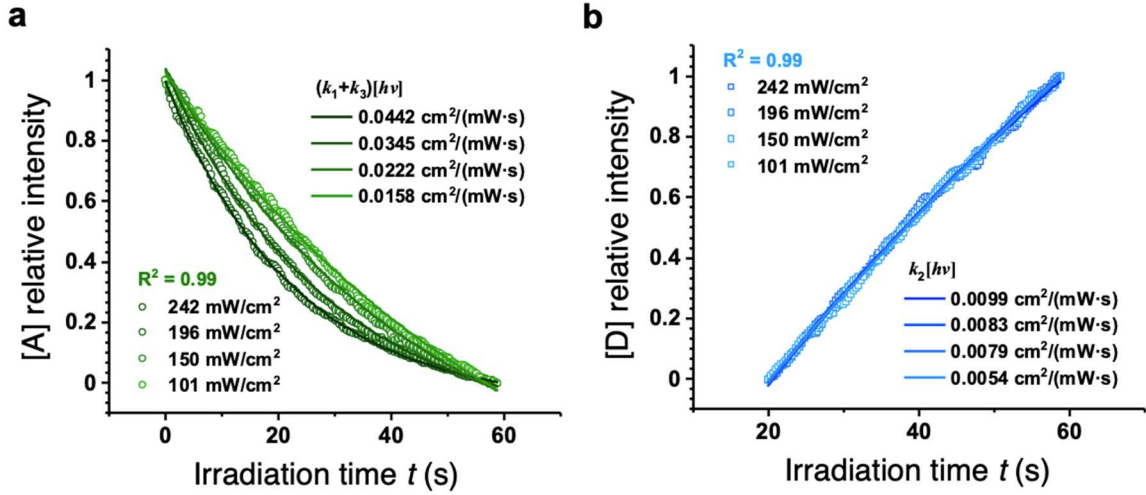
c



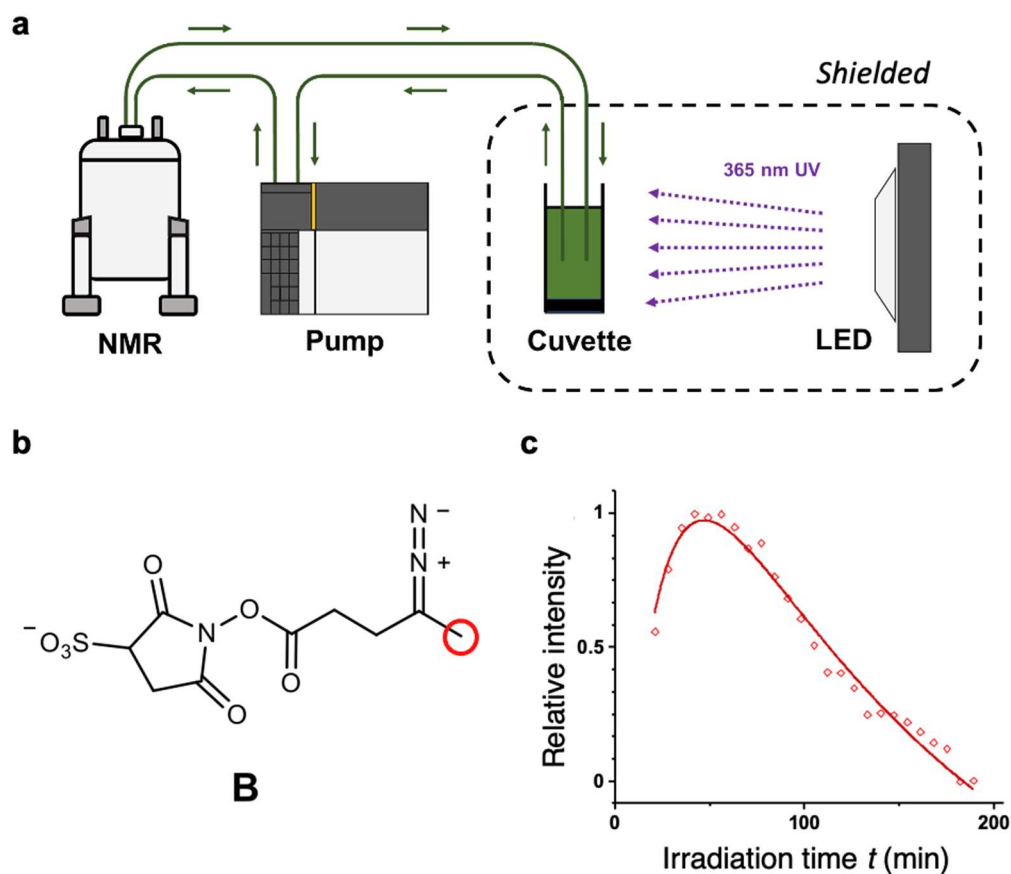
d



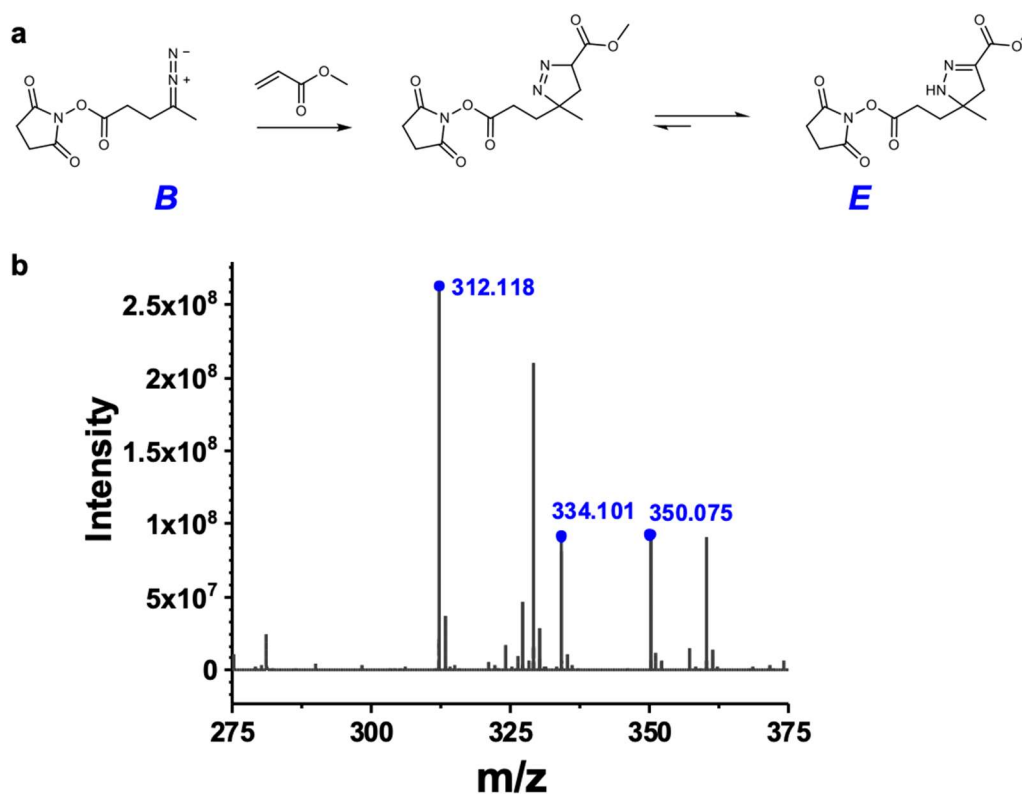
Supplementary Figure 1 | Setting up a power-modulated photo-reaction system with in-line MS monitoring. **a**, Reaction takes place in the PFA tube arranged zigzaggedly under a uniform light field, with the sample injected in a constant flow. Both irradiation time t and optical power density $[h\nu]$ can be modulated to generate a large array of data points, allowing for the fitting of kinetic parameters (insets). Reaction mixture is subsequently injected to MS for real-time Multiple Reaction Monitoring (MRM) analysis. **b**, Calibration of the irradiation field over the PFA sample tube, illustrating homogeneity reached. **c**, Modulation of irradiation time t by using the scheme of pulse-width modulation (PWM), with 31 duty ratios programmed. **d**, Modulation of the optical power density $[h\nu]$ with adjustable constant current supply; 256 current modes, corresponding $[h\nu]$ from 0.1 to 289 mW/cm², can be programmed.



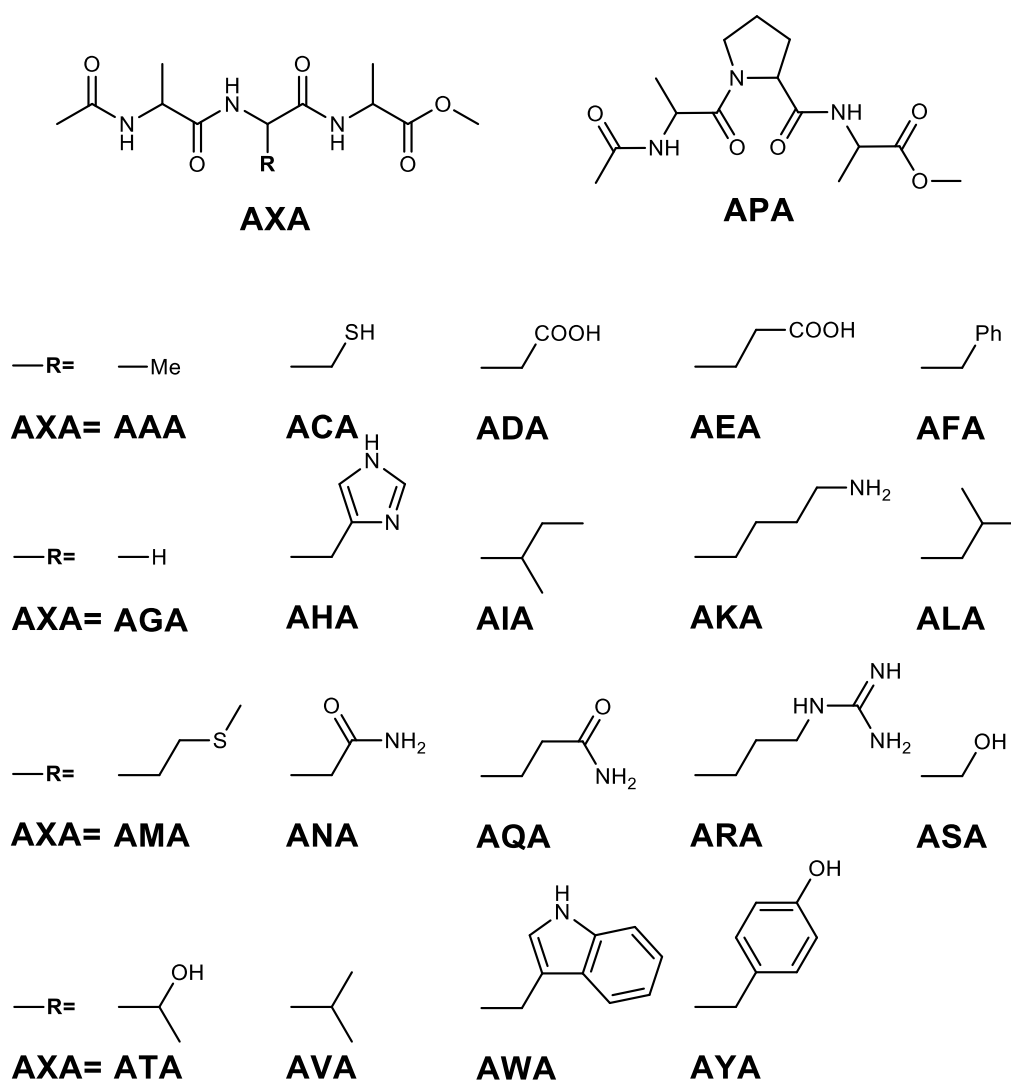
Supplementary Figure 2 | Fitting of the time-dependent relative concentration changes of $[A]$ and $[D]$. **a**, Schematic diagram of single-exponential fitting of A to obtain $(k_1 + k_3)[h\nu]$. At a particular optical power density $[h\nu]$, $[A]$ was fitted using equation of $I_A = [A]_0 \exp(-(k_1 + k_3)[h\nu]t) + I_0$. **b**, Schematic diagram of fitting of $[D]$ to obtain $k_2[h\nu]$, with irradiation time $t > 20$ s. Time-dependent $[D]$ can be described with the equation $I_D = [A]_0 \left(1 - \frac{k_1}{k_1 - k_2 + k_3} \exp(-k_2[h\nu]t) + \frac{k_2 - k_3}{k_1 - k_2 + k_3} \exp(-(k_1 + k_3)[h\nu]t) \right) + I_0$. Four irradiation intensities are plotted, with details of the equations provided in the **Supplementary Note**.



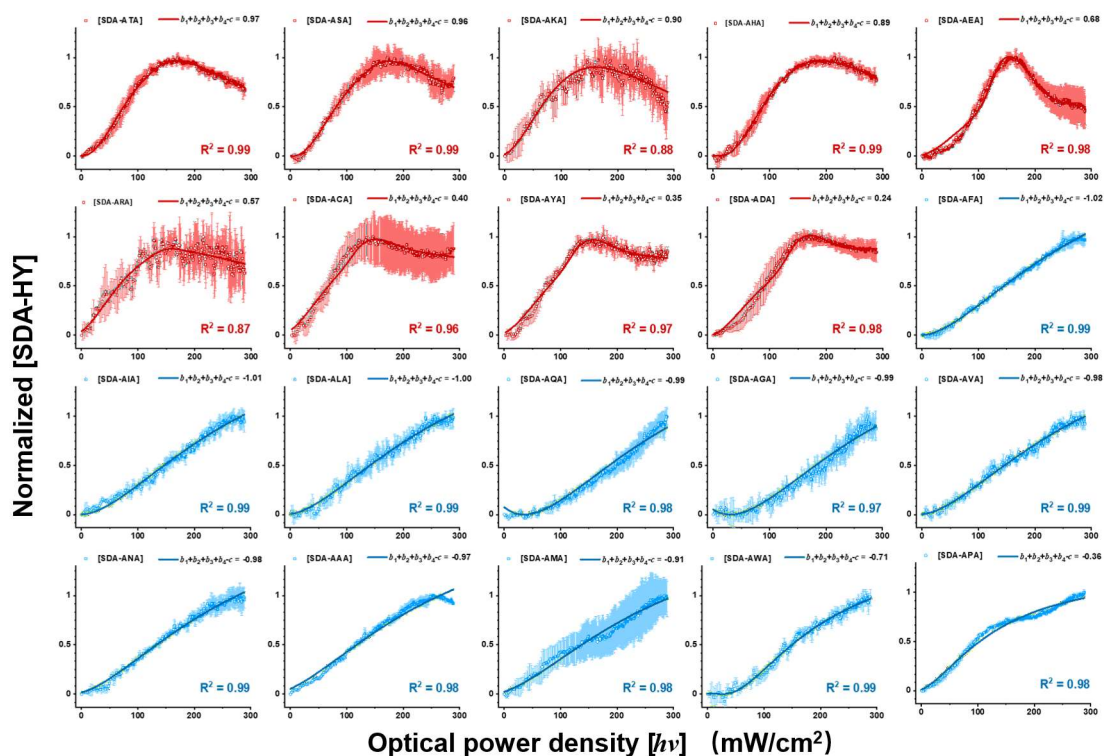
Supplementary Figure 3 | Real-time NMR monitoring of the diazo intermediate *B*. **a**, An in-line NMR spectroscopy setup, which allows for UV irradiation and NMR measurement of the product. **b-c**, The structure of diazo intermediate *B*. The atom denoted, with a characteristic NMR peak at ~ 0.99 ppm, of which the intensity follows a curve that can be described with model **I** or **II**. The peak area (intensity) was integrated and normalized to the highest value.



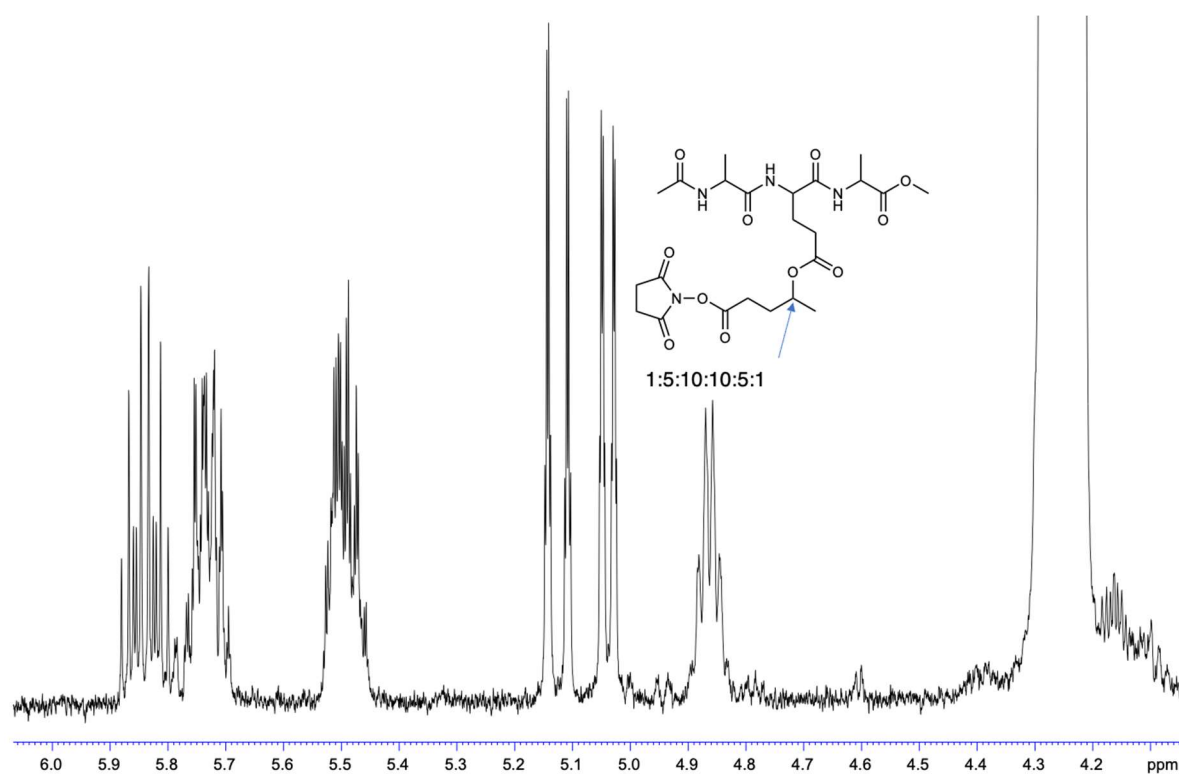
Supplementary Figure 4 | Capturing the diazo intermediate *B* with methyl methacrylate. a, The reaction between diazo intermediate *B* and methyl methacrylate, affording compound *E*. **b,** *E* was confirmed in Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (Solarix XR, Bruker) with the observed *m/z* of 312.118 ($[M+H]^+$), 334.101 ($[M+Na]^+$), and 350.075 ($[M+K]^+$). Expected, 311.112 Da.



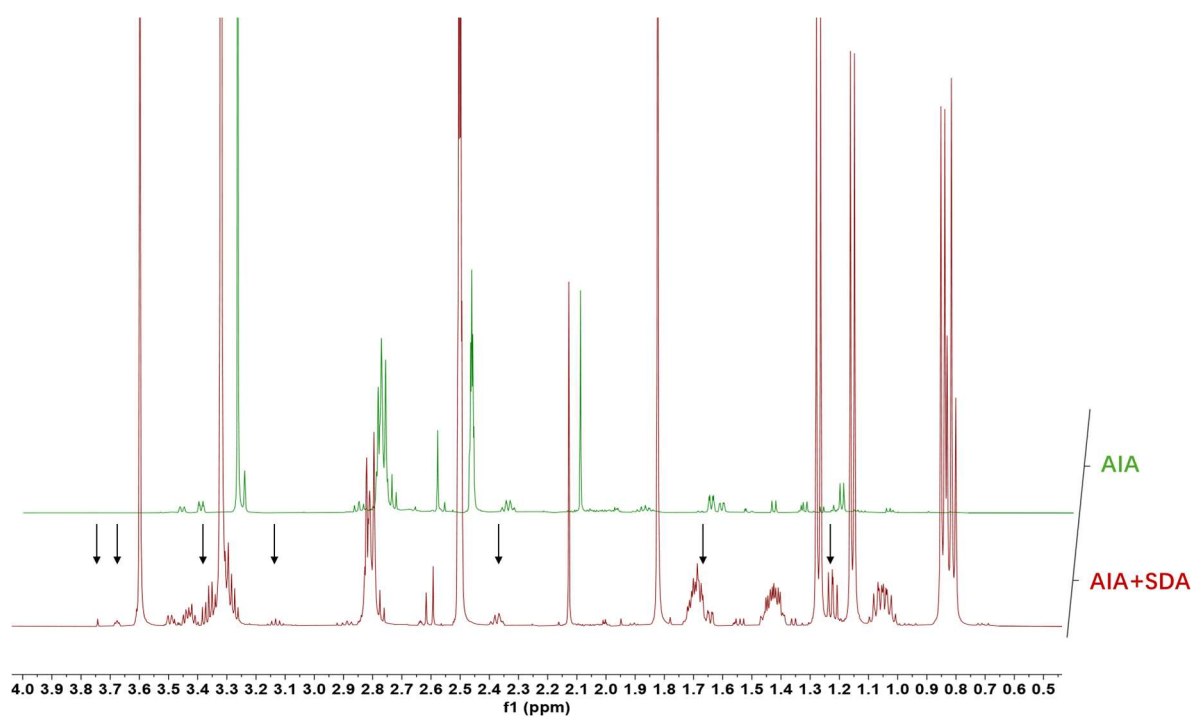
Supplementary Figure 5 | Tripeptides AXA used for assessing photo-reaction kinetics with alkyl diazirine. Here X denotes one of the 20 amino acid residues with various *R* group; the N-terminus of the peptide is acetylated and C-terminus methyl-esterified. Note that though the peptide intends to evaluate the reaction selectivity mimicking a real protein, only enhanced conversion rate may be detected above the intrinsic photo-reaction with AAA.



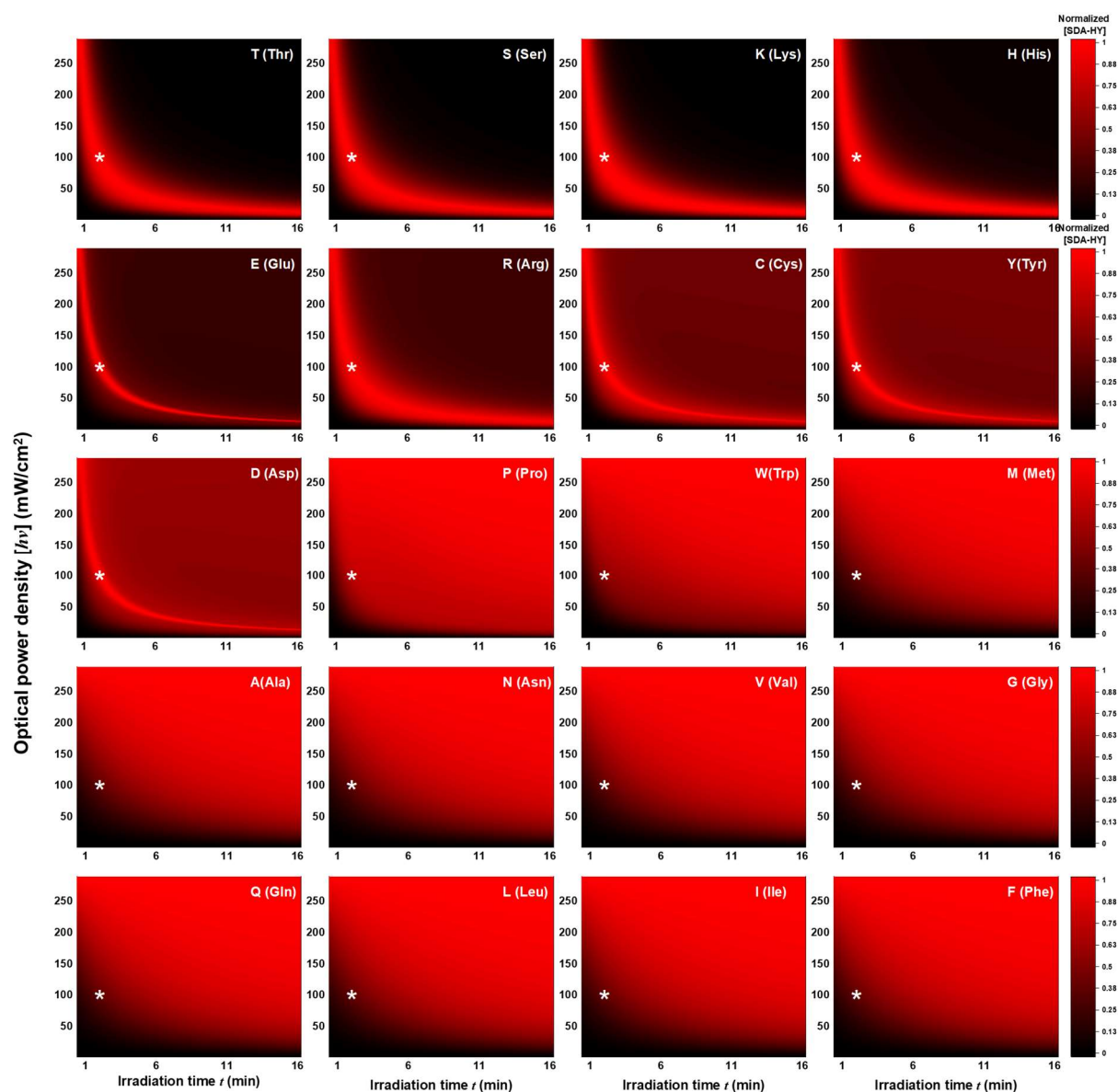
Supplementary Figure 6 | Fitting [SDA-HY] experimental data over optical power density [$h\nu$]. The experiments were performed for 20 amino acid residues in the terminally blocked tripeptide; each measurement was repeated three times. Polar and non-polar residues are colored red and blue, respectively, as judged by the relative contribution from diazo and carbene intermediate; the fitted values are given in **Supplementary Table 1**. The error bars represent 1 S.D. from experimental measurements for $n = 3$; Pearson correlation coefficients R^2 are given.



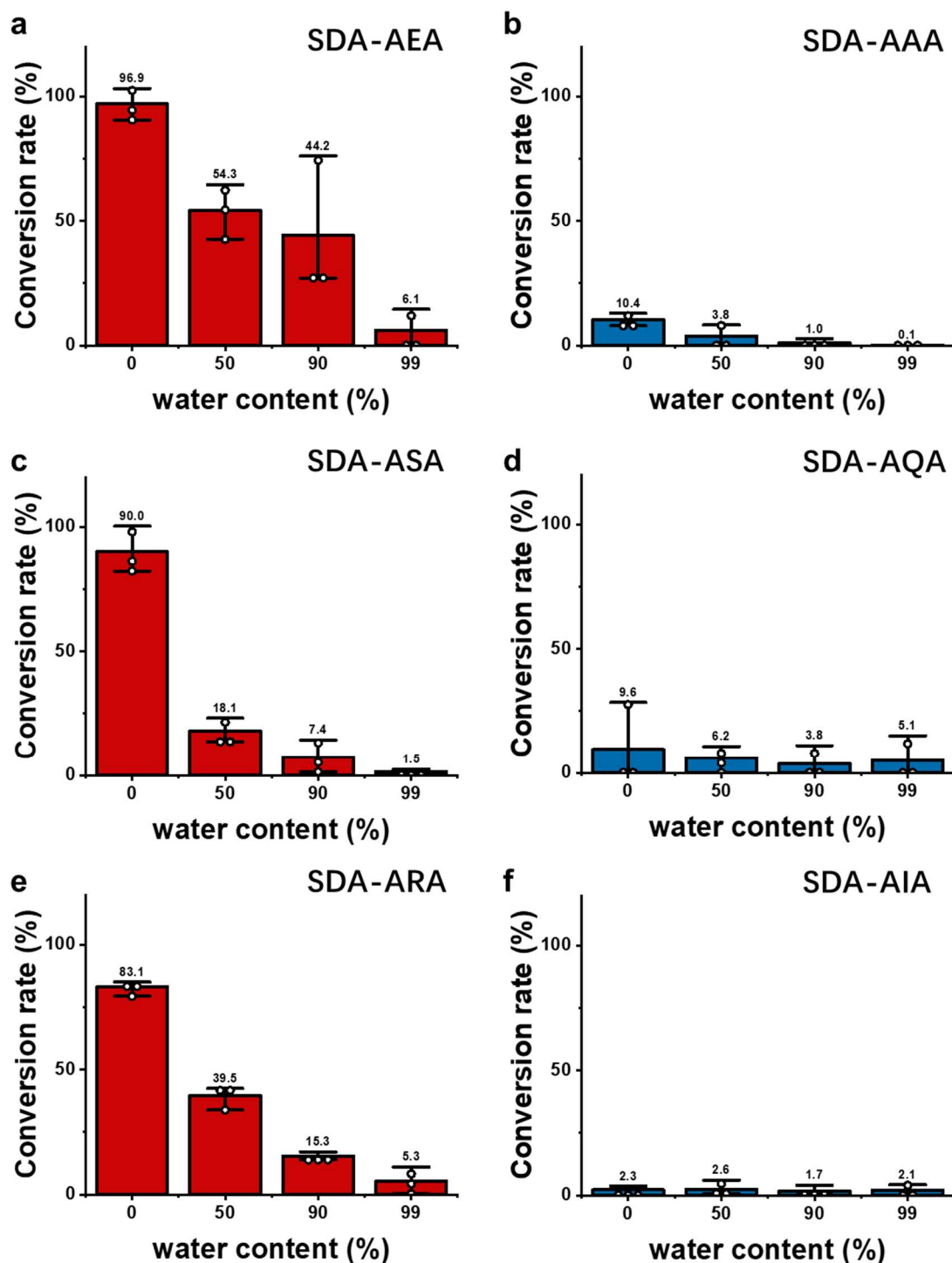
Supplementary Figure 7 | ^1H NMR spectrum for the SDA photo-adduct with tripeptide AEA. With the formation of the new ester bond between with the side-chain carboxylate, a new peak appears at ~4.86 ppm (indicated with the blue arrow), with characteristic multiplet for the three-bond scalar couplings with the adjacent methyl and methylene groups.



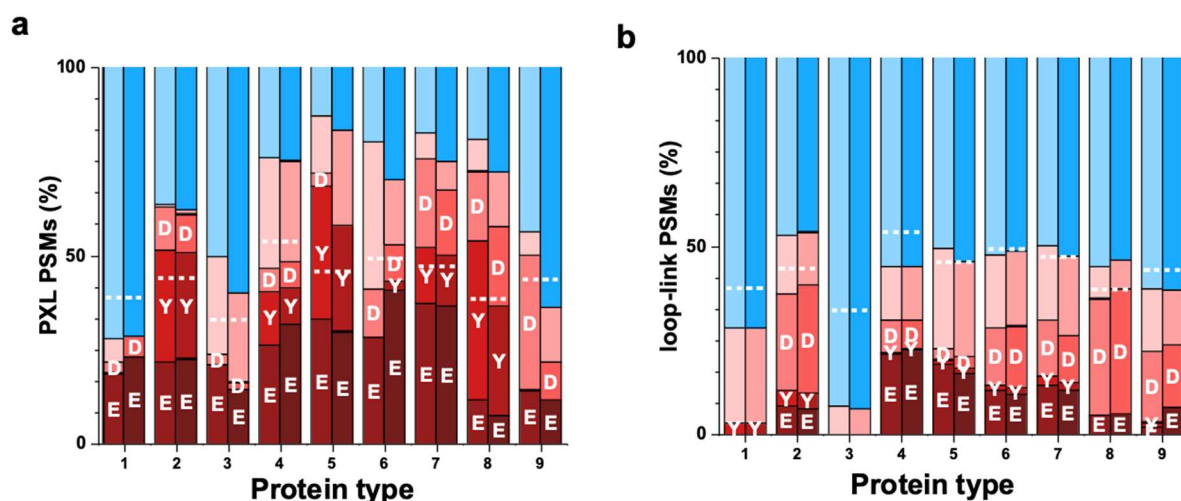
Supplementary Figure 8 | ^1H NMR spectrum for the SDA photo-adduct with tripeptide AIA.
The arrows indicate new peaks upon photo-reaction.



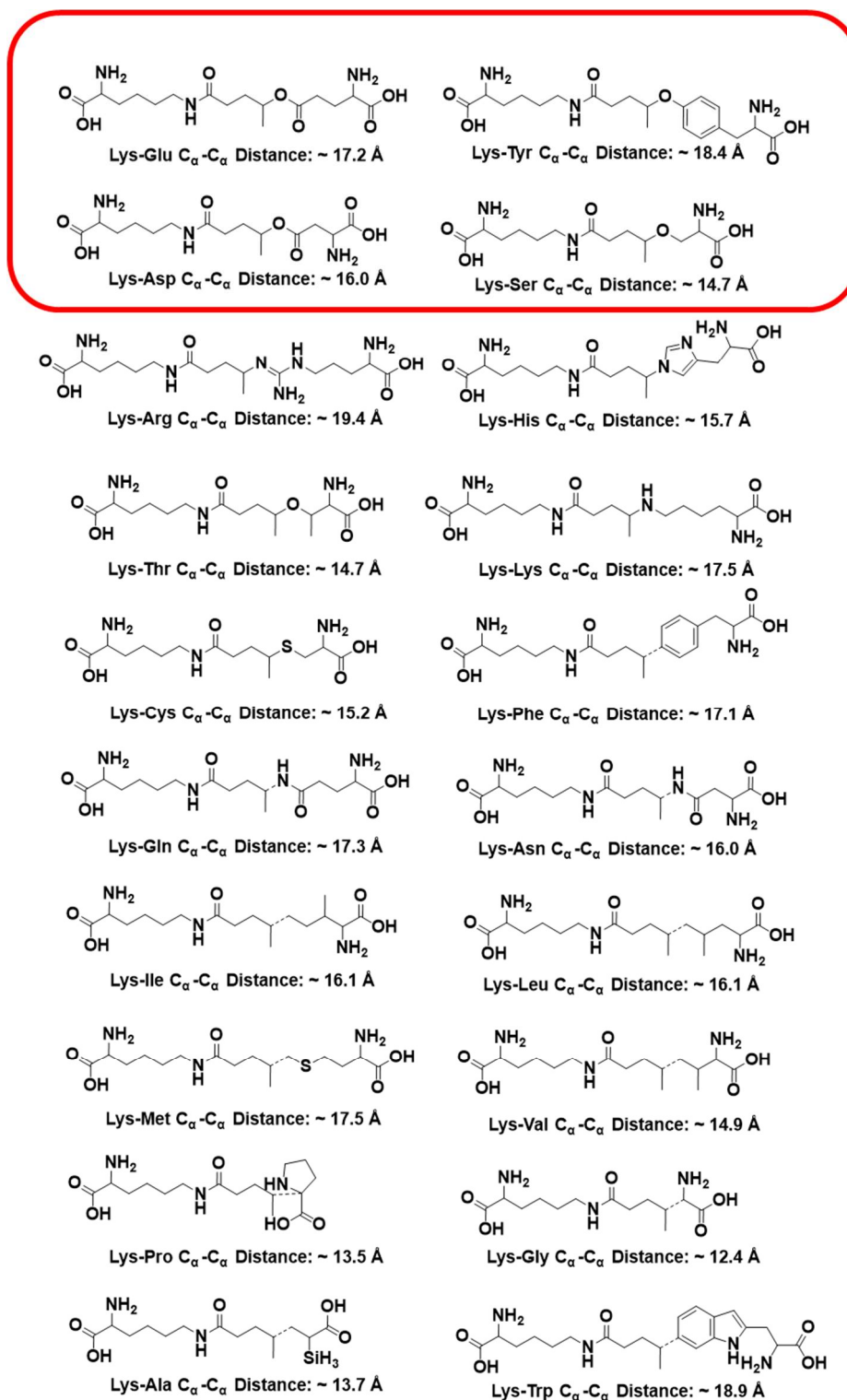
Supplementary Figure 9 | The yield of SDA-HY as a function of irradiation time t and optical power density $[h\nu]$. The heat map scale represents the relative yield of the photo-adduct of SDA-HY, normalized to the highest value of each residue. The asterisks indicate a preferred value of t and $[h\nu]$ for a high yield of SDA-HY for polar residues (100 mW/cm², 2 min).



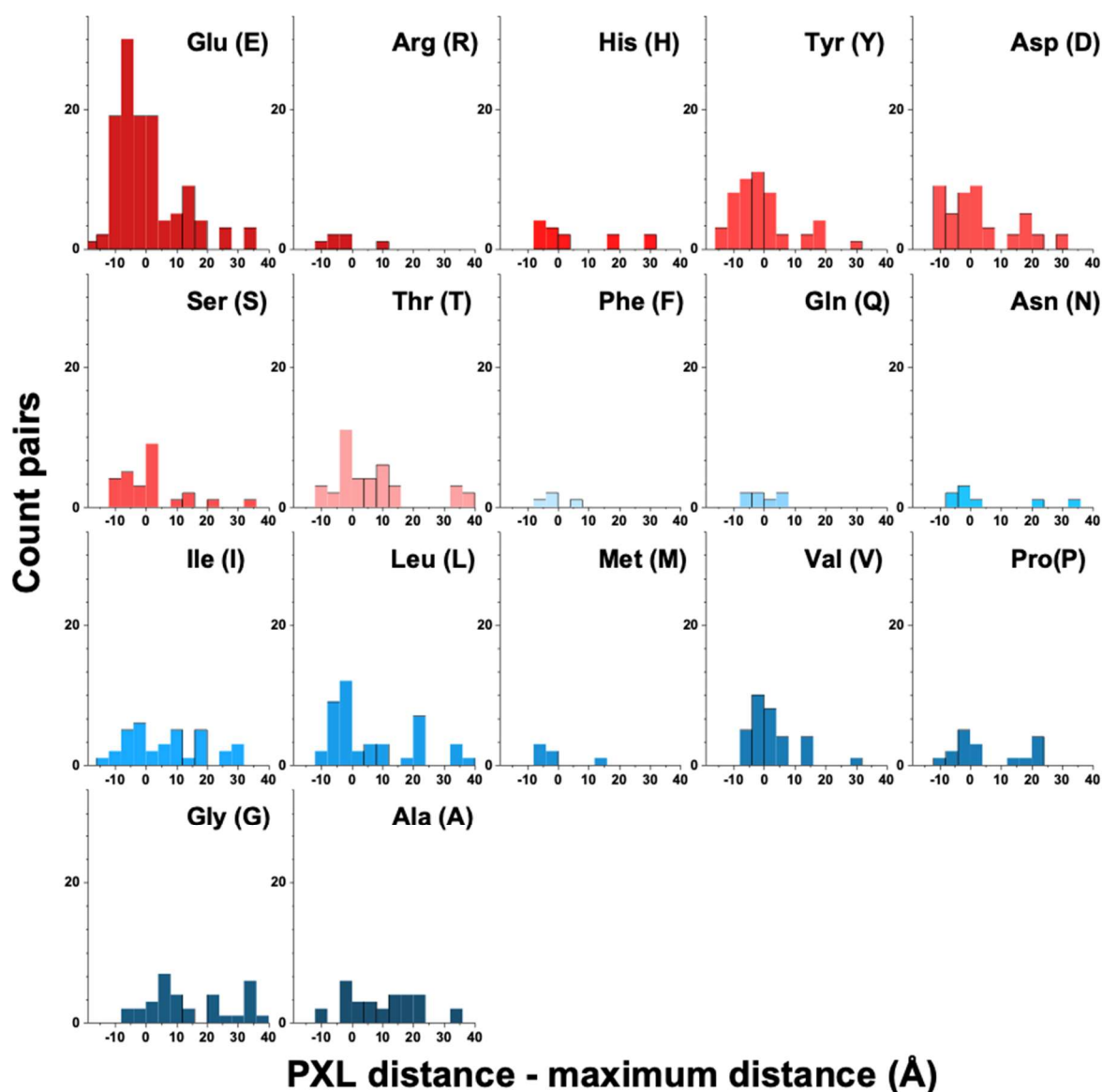
Supplementary Figure 10 | The absolute yield (cf. Fig. 3e) of the SDA photo-adduct decreases with increasing water content (v/v) in the solvent. The conversion rate is computed by the decreasing AXA tripeptide MS signal. Optical power density [$h\nu$] is set at 100 mW/cm² with an irradiation time t for 2 minutes. Error bar represent 1 S. D. from triplicated measurements. The hollow circle represents the result of an independent experiment, and their average value is labelled above the bar. Note that the initial yield for **b**, **d** and **f** in water-free solvent (DMSO) is already much lower than **a**, **c** and **e**. Note that AAA tripeptide can also be considered as a background for the diazirine photo-reaction with AEA and other tripeptides. The results suggest that relatively buried protein residues can be better placed to react with diazirine, either with the diazo or carbene-mediated mechanism.



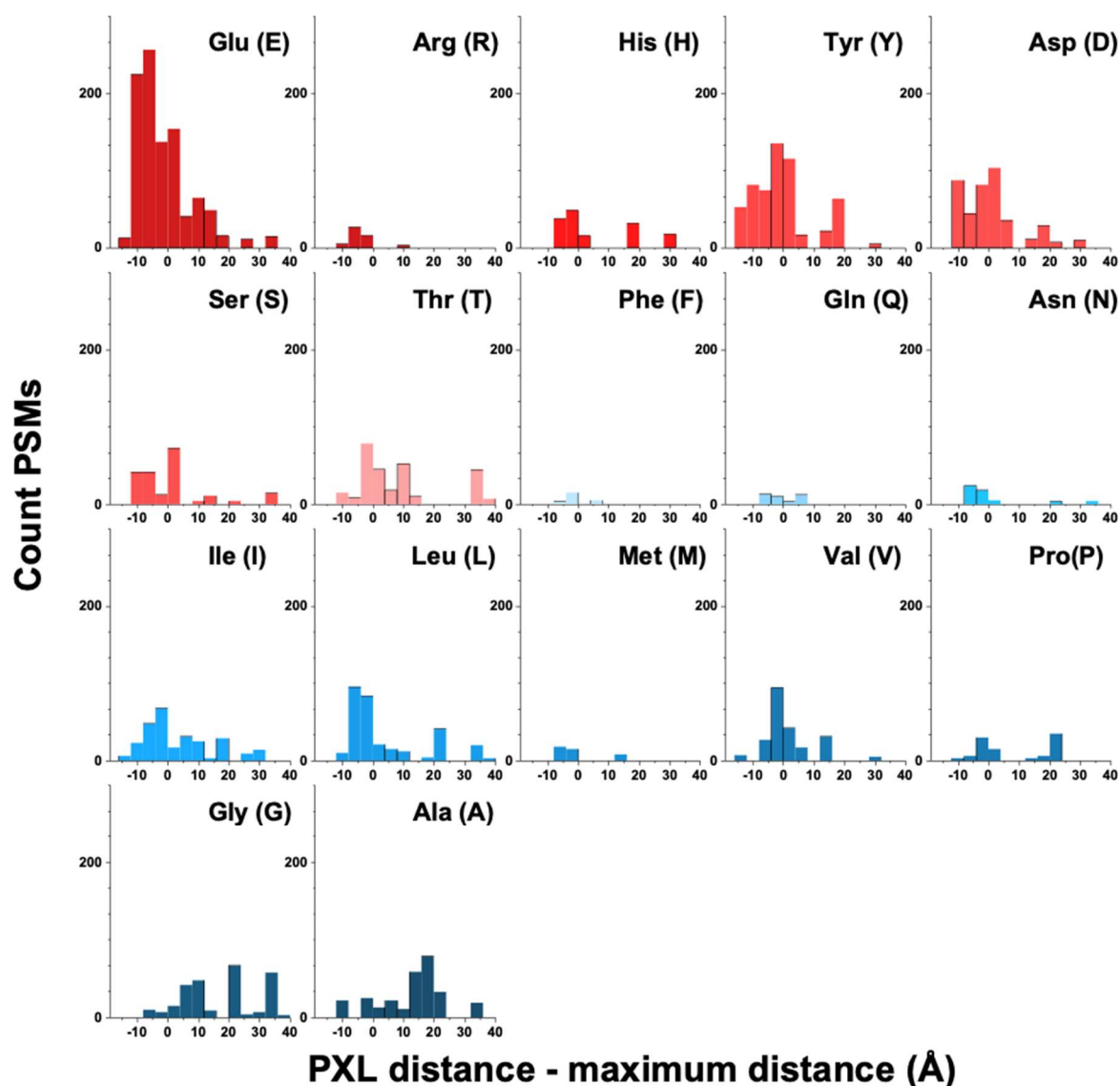
Supplementary Figure 11 | Alkyl diazirine preferentially targets protein polar residues in the test proteins, as evaluated by matched peptide spectra (PSM). **a**, Relative abundance of PSMs for polar residues (red shaded) and non-polar residues (blue shaded), using SDA upon irradiation at an optical power density [$h\nu$] of 35 mW/cm² (left bar, low power) for 10 min or 282 mW/cm² (right bar, low power) for 2 min. The dashed line indicates the natural abundance of the polar residues in each protein. D (Asp), E (Glu) and Y (Tyr) are labeled. **b**, Relative abundance of the loop-link PSMs, for non-polar and polar residues in the model proteins upon low- or high-power irradiation. The details of the 9 model proteins are given in Methods.



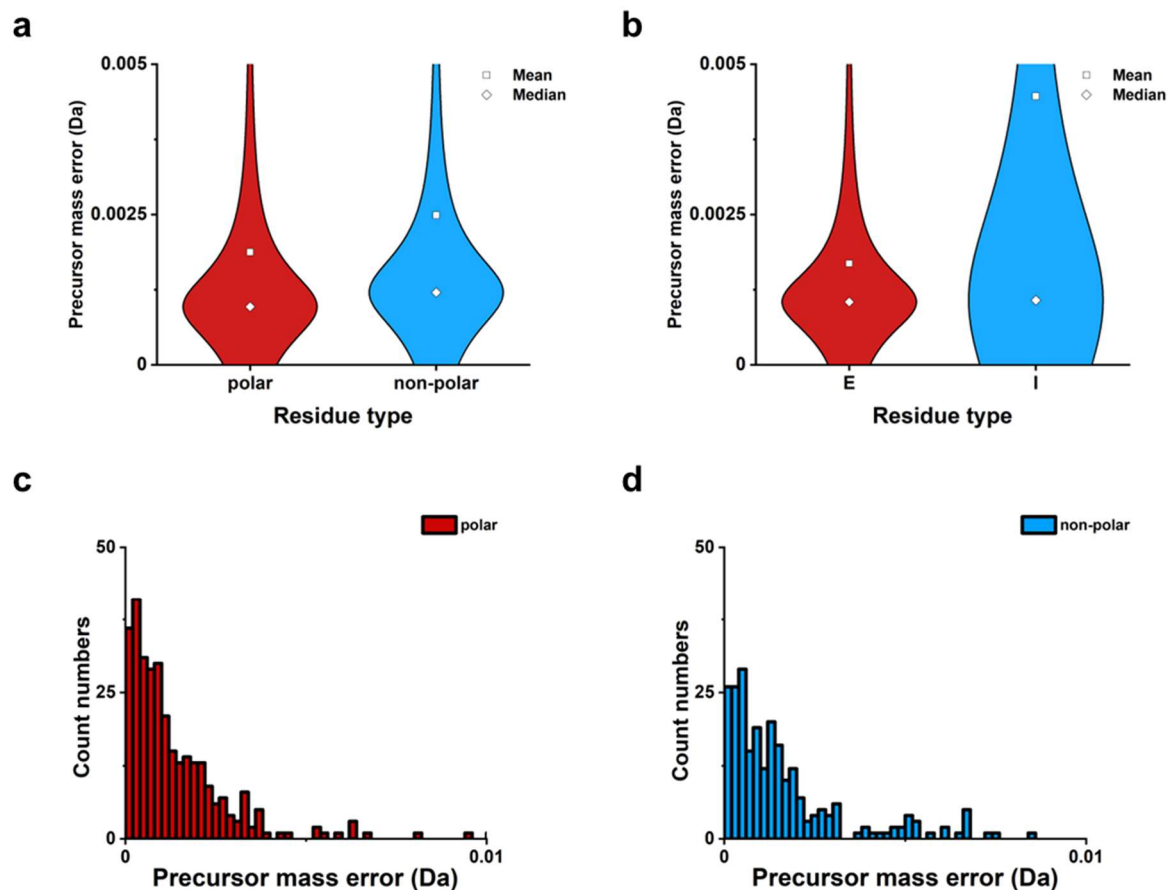
Supplementary Figure 13 | Proposed structures of PXL product involving the side-chain of Lys and another residue. Polar residues largely react with the diazo intermediate, by attacking the tertiary carbon. The four polar residues of which the PXLs can be assigned with highest confidence are boxed. In comparison, the reaction involving non-polar residues yields heterogenous products, with the possible linkages indicated with dashed lines.



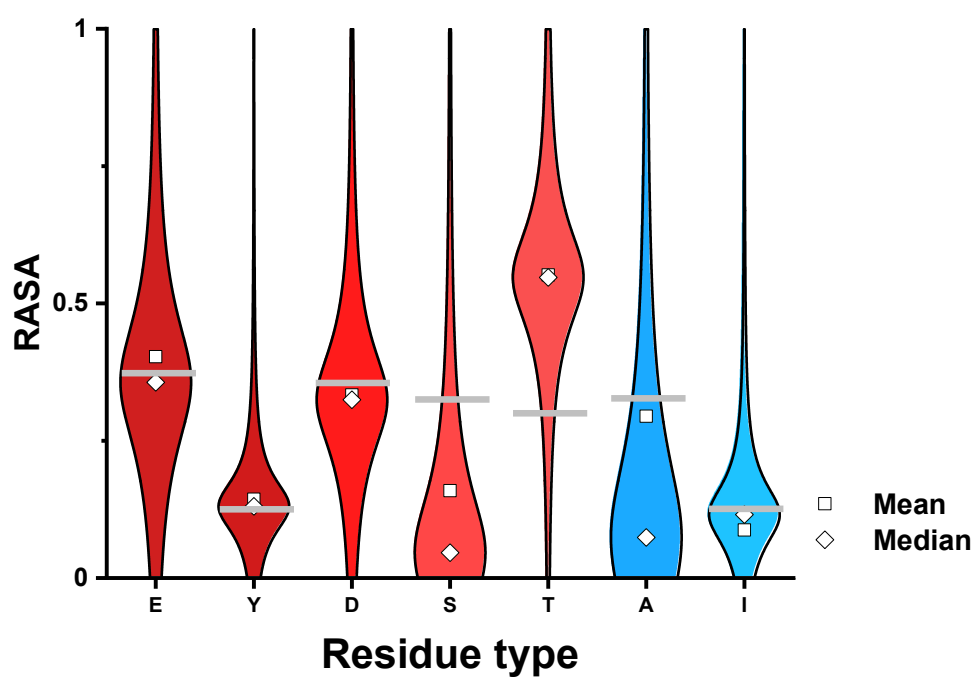
Supplementary Figure 14 | Ca-Ca distance between the PXL residues identified in the model proteins. The distances are computed based on the protein structures, subtracting the maximally allowed theoretical distance (**Supplementary Figure 13**); the relative counts of cross-linked pairs/residues are distributed in 4-Å bins. Note that Lys residues are not included, due to the initial reaction with the NHS moiety, and Cys residues are not included due to alkylation treatment of the cross-linked peptides with 10 mM iodoacetamide. Only few cross-links involving Trp were obtained, which are not included due to lack of statistic meaning. A negative value indicates the consistency between the PXLs and the known protein structure.



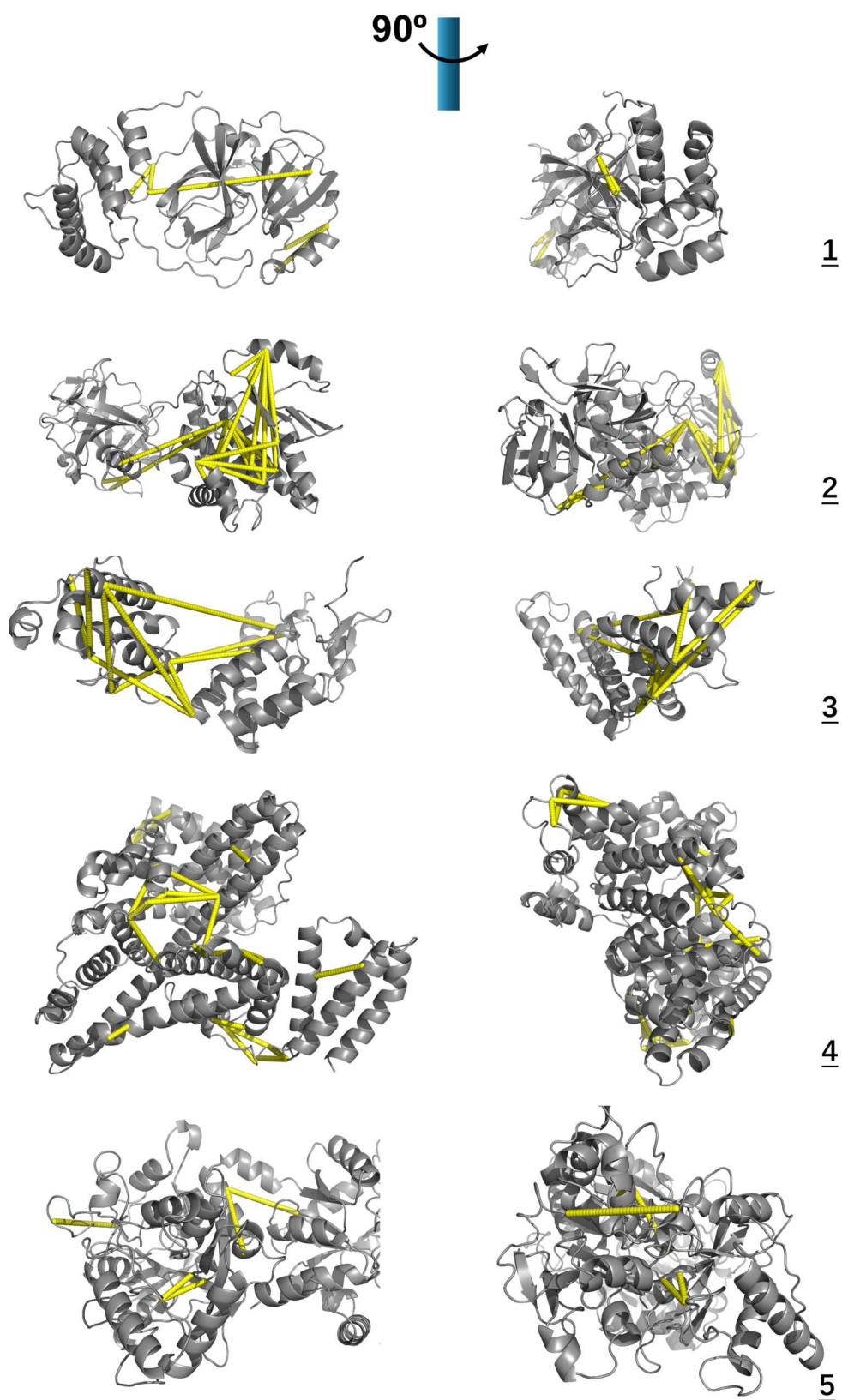
Supplementary Figure 15 | Ca-Ca distance between the PXL residues identified in the model proteins. The distances are computed based on the protein structures, subtracting the maximally allowed theoretical distance (**Supplementary Figure 13**); the relative counts of matched peptide spectra (PSM) are distributed in 4-Å bins. Note that Lys residues are not included, due to the initial reaction with the NHS moiety, and Cys residues are not included due to alkylation treatment of the cross-linked peptides. Few cross-links involving Trp were obtained, which are not included. A negative value indicates the consistency between the PXLs and the known protein structure.

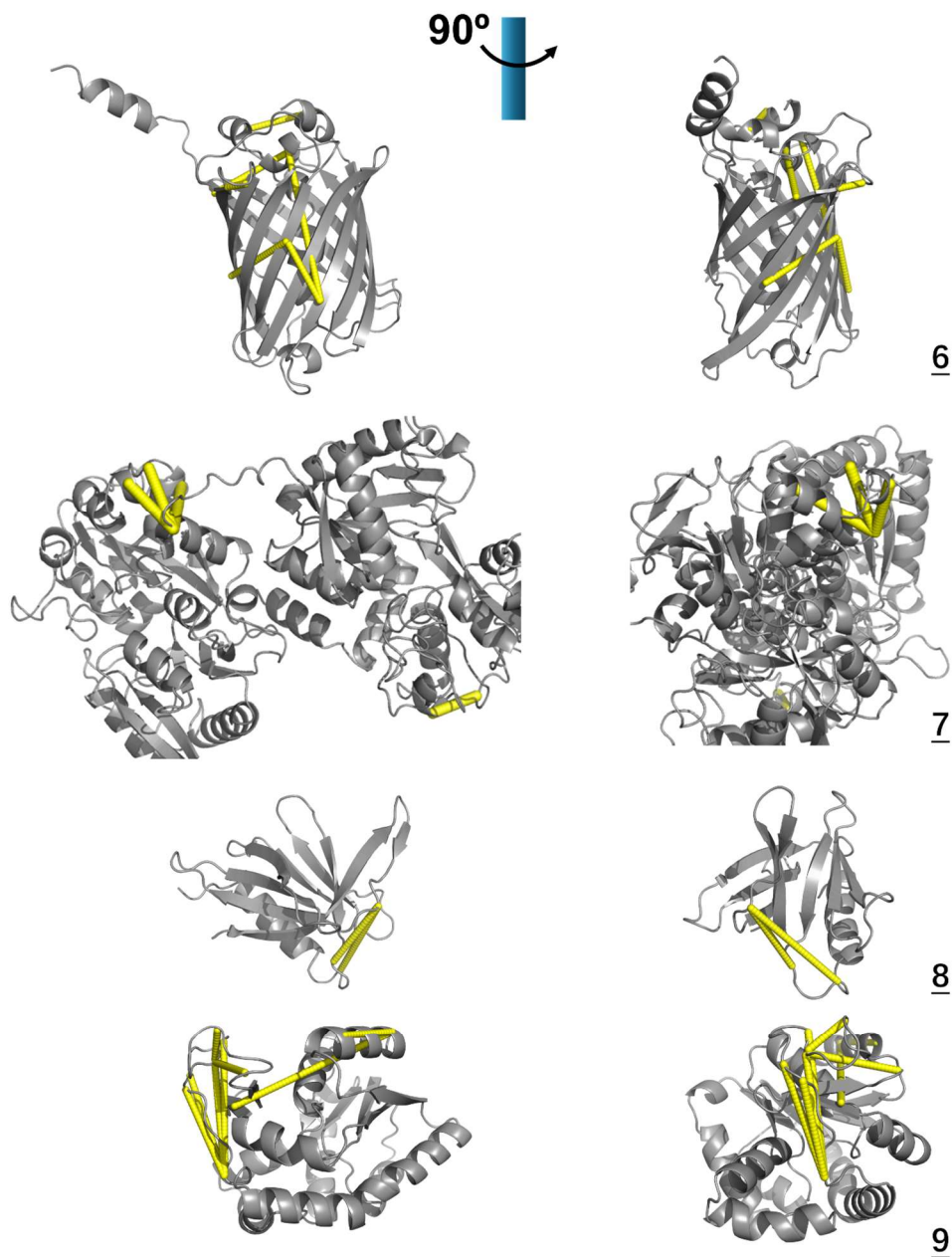


Supplementary Figure 16 | PXLs towards non-polar residues are assigned with higher precursor mass error (PME). In addition, the PME values exhibit larger scattering than those for polar residues. **a, b**, PXLs towards non-polar has higher PME in both mean value and median value than polar residues, significant for Ile to Glu. **c, d**, PME histogram distribution graph of polar and non-polar residues.



Supplementary Figure 17 | Relative solvent-accessible surface area (RASA) of the residues involved in the PXLs. Except for Thr, most cross-linked polar residues have ASA values below the average values of the model proteins, as indicated by the gray lines. The ASA was computed based on the known structures, which was then divided by the total surface area of each type of residue¹.





Supplementary Figure 18 | Distance mapping from the PXLs involving polar residues against protein structures. The details of the nine proteins tested are given in Methods. Cross-links are achieved from experiments under optical power density [$h\nu$] of 35 mW/cm² with the irradiation time of 2 min. The PXLs are provided in the Supplementary Data.

Supplementary Tables

Supplementary Table 1 | Relative contribution of the sub-processes to SDA-AXA photo-reaction[#]

Residue Type (X in AXA)	b_1	b_2	b_3	b_4	c
A	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.97±0.01
C	0.00±0.00	0.59±0.02	0.17±0.01	0.00±0.00	0.36±0.01
D	0.46±0.02	0.00±0.00	0.26±0.01	0.00±0.00	0.48±0.02
E	0.00±0.00	0.38±0.01	0.51±0.01	0.00±0.00	0.21±0.01
F	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	1.02±0.00
G	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.99±0.01
H	0.13±0.20	0.80±0.17	0.01±0.00	0.00±0.00	0.05±0.05
I	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	1.01±0.01
K	0.11±0.86	0.78±0.75	0.01±0.03	0.00±0.00	0.00±0.20
L	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	1.00±0.01
M	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.91±0.00
N	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.98±0.00
P	0.31±0.01	0.00±0.00	0.00±0.00	0.00±0.00	0.67±0.01
Q	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.99±0.00
R	0.20±0.43	0.52±0.38	0.00±0.00	0.04±0.02	0.19±0.09
S	0.17±0.13	0.72±0.11	0.07±0.01	0.00±0.00	0.00±0.03
T	0.00±0.00	0.92±0.01	0.05±0.01	0.00±0.00	0.00±0.01
V	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.98±0.01
W	0.00±0.00	0.18±0.01	0.00±0.00	0.00±0.00	0.89±0.01
Y	0.00±0.00	0.53±0.02	0.21±0.01	0.00±0.00	0.39±0.02

[#] The production of **SDA-AXA** comprises five elementary processes, in which the time t and irradiation power $[h\nu]$ dependent changes of **[SDA-AXA]** can be defined as $b_1 I_{b_1} + b_2 I_{b_2} + b_3 I_{b_3} + b_4 I_{b_4} + c I_c$, in which b_1 , b_2 , b_3 , b_4 , and c are weighting factors, though the absolute yield can vary greatly between residues. I_{b_1} and I_{b_2} represent the direct reaction between protonated **AXA** and diazo intermediate and between deprotonated **AXA** and diazo intermediate, respectively, whereas I_{b_3} and I_{b_4} represent the respective proton-catalyzed processes. I_c represents the carbene-mediated reaction.

Supplementary Table 2 | MRM parameters for the in-line analysis of *A* and *D* upon diazirine photolysis

Interface	Nebulizing Gas Flow (L/min)	Heating Gas Flow (L/min)	Interface Temperature (°C)	DL Temperature (°C)	Heat Block Temperature (°C)	Drying Gas Flow (L/min)
ESI	3	10	300	250	400	10

	Precursor m/z	Product m/z	Dwell Time(ms)	Q1 Pre Bias(V)	CE	Q3 Pre Bias(V)
<i>A</i>	304.00	81.00	100.0	20.0	22.0	28.0
<i>D</i>	292.00	81.00	100.0	19.0	23.0	28.0
	236.04	207.10	100.0	24.0	22.0	19.0
	236.05	115.10	100.0	29.0	35.0	20.0

Supplementary Table 3 | MRM parameters for in-line analysis of diazirine photo-reaction with AXA tripeptide

	Precursor m/z	Product m/z	Dwell Time(ms)	Q1 Pre Bias(V)	CE	Q3 Pre Bias(V)
AAA	288.10	175.20	100.0	-21.0	-10.0	-17.0
	288.10	185.20	100.0	-21.0	-10.0	-19.0
	288.10	104.00	100.0	-22.0	-18.0	-20.0
ACA	320.10	207.20	100.0	-14.0	-12.0	-12.0
	320.10	44.10	100.0	-25.0	-38.0	-14.0
	320.10	76.15	100.0	-26.0	-25.0	-29.0
ADA	354.20	251.15	100.0	-10.0	-19.0	-27.0
	354.20	179.20	100.0	-10.0	-23.0	-11.0
	354.20	224.05	100.0	-17.0	-21.0	-23.0
AEA	368.10	262.25	100.0	-15.0	-21.0	-27.0
	368.10	237.20	100.0	-15.0	-24.0	-23.0
AFA	364.20	120.10	100.0	-29.0	-25.0	-20.0
	364.20	251.20	100.0	-30.0	-13.0	-26.0
	364.20	261.25	100.0	-10.0	-11.0	-12.0
AGA	274.10	161.20	100.0	-23.0	-10.0	-29.0
	274.10	104.15	100.0	-11.0	-22.0	-10.0
	274.10	44.10	100.0	-11.0	-39.0	-16.0
AHA	376.10	245.10	100.0	-15.0	-24.0	-25.0
	376.10	273.10	100.0	-15.0	-21.0	-12.0
	376.10	276.15	100.0	-15.0	-25.0	-17.0
AIA	330.10	86.15	100.0	-13.0	-23.0	-15.0
	330.10	227.20	100.0	-14.0	-10.0	-15.0
	330.10	217.20	100.0	-26.0	-11.0	-22.0
AKA	345.20	84.05	100.0	-14.0	-40.0	-16.0
	345.20	129.20	100.0	-29.0	-19.0	-25.0
	345.20	232.25	100.0	-14.0	-17.0	-10.0
ALA	330.10	86.15	100.0	-13.0	-23.0	-14.0
	330.10	227.15	100.0	-14.0	-10.0	-14.0
	330.10	217.25	100.0	-24.0	-12.0	-29.0
AMA	348.10	104.15	100.0	-14.0	-17.0	-10.0
	348.10	245.15	100.0	-14.0	-10.0	-16.0
	348.10	56.20	100.0	-28.0	-32.0	-24.0
ANA	236.05	195.10	100.0	-16.0	-7.0	-17.0

	236.05	138.25	100.0	-16.0	-12.0	-26.0
	236.05	154.10	100.0	-17.0	-11.0	-15.0
APA	314.10	201.20	100.0	-23.0	-12.0	-22.0
	314.10	70.20	100.0	-13.0	-28.0	-12.0
	314.10	44.10	100.0	-13.0	-48.0	-17.0
AQA	367.15	236.15	100.0	-16.0	-24.0	-15.0
	367.15	264.20	100.0	-16.0	-21.0	-17.0
	367.15	150.20	100.0	-16.0	-21.0	-23.0
ARA	373.20	70.20	100.0	-15.0	-39.0	-28.0
	373.20	243.20	100.0	-15.0	-22.0	-25.0
	373.20	114.10	100.0	-15.0	-23.0	-21.0
ASA	304.10	286.10	100.0	-12.0	-10.0	-19.0
	304.10	191.20	100.0	-24.0	-12.0	-19.0
	304.10	60.05	100.0	-12.0	-24.0	-23.0
ATA	318.20	256.30	100.0	-13.0	-22.0	-17.0
	318.20	258.15	100.0	-25.0	-18.0	-25.0
AVA	236.05	195.10	100.0	-16.0	-7.0	-17.0
	236.05	138.25	100.0	-16.0	-12.0	-26.0
	236.05	154.10	100.0	-17.0	-11.0	-15.0
AWA	425.15	294.20	100.0	-11.0	-26.0	-13.0
	425.15	407.15	100.0	-17.0	-25.0	-19.0
	425.15	312.20	100.0	-17.0	-26.0	-21.0
AYA	380.15	136.25	100.0	-16.0	-22.0	-13.0
	380.15	91.20	100.0	-15.0	-54.0	-17.0
	380.15	267.20	100.0	-30.0	-13.0	-18.0

Supplementary Notes

Mathematic rendering for the analysis of diazirine photolysis mechanism

Upon irradiation, the diazirine compound undergoes photolysis reaction through one of the four possible mechanisms as illustrated in **Fig. 1**. Alkyl diazirine **A** is converted to the diazo intermediate **B** or directly to the carbene intermediate **C**, with the kinetic rates of $k_1[h\nu]$ and $k_3[h\nu]$, respectively. **B** absorbs a second photon and transforms to **C**, with the kinetic rate of $k_2[h\nu]$. **C**, a short-lived carbene intermediate^{2,3}, quickly transforms to an oxidized product **D** or other products with the kinetic rate of k_{ox} or k_{others} ⁴. The **C**-to-**D** conversion ratio m can be defined as $k_{ox}/(k_{ox} + k_{others})$.

Model **I** is a simplified version of model **II**, differing for the lack of a direct **A**-to-**C** process. Model **III** assumes that k_2 process occurs without irradiation, while model **IV** involving no **B**-to-**C** process, which assumption has been held in the previous literature^{5,6}.

We calculated the kinetic equation of **A**, **B** and **D**, which can be monitored with MS or NMR. **A** becomes depleted upon irradiation following a single-exponential process over t , which can be defined as

$$d[A]_t = -k_1[h\nu][A]_t dt - k_3[h\nu][A]_t dt \quad (1)$$

$$[A]_t = [A]_0 \exp(-(k_1 + k_3)[h\nu]t) \quad (2)$$

The curves for the concentration of **B** and **D** include one or two exponential terms (in model **I**, $k_3 = 0$; in model **III**, $k_2' = k_2/[h\nu]$, while $k_2' = k_2$ in all other models; in model **IV**, $k_2 = 0$). Thus, **[B]** and **[D]** can be written as

$$d[B]_t = k_1[h\nu][A]_t dt - k_2'^{[h\nu][B]}_t dt, \text{ with } [B]_0 = 0 \quad (3)$$

$$[B]_t = \frac{k_1[A]_0}{k_1 - k_2' + k_3} \left(\exp(-k_2'[h\nu]t) - \exp(-(k_1 + k_3)[h\nu]t) \right) \quad (4)$$

$$m = k_{ox}/(k_{ox} + k_{othe}) \quad (5)$$

$$d[D]_t = mk_2'^{[h\nu]}_t dt + mk_3[h\nu][A]_t dt, \text{ with } [D]_0 = 0 \quad (6)$$

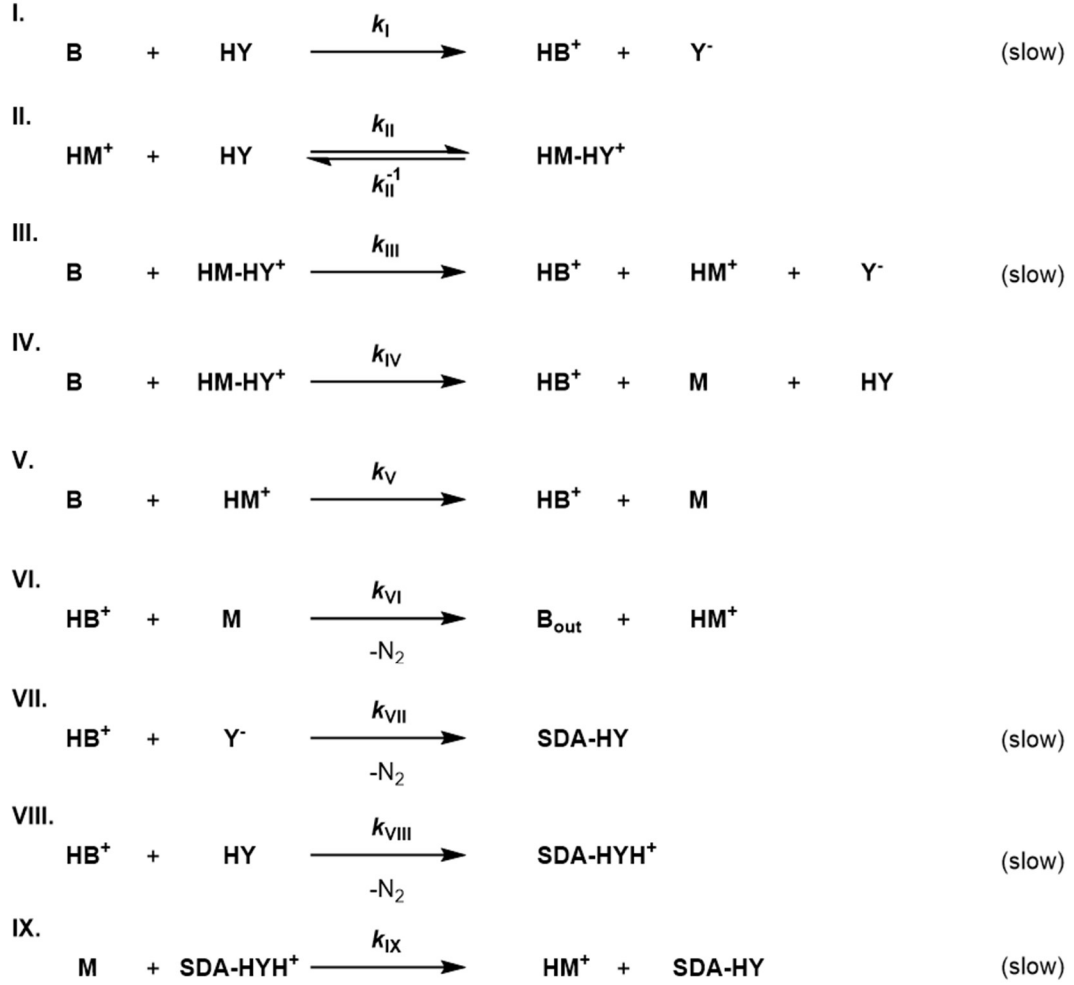
$$[D]_t = m[A]_0 \left(1 - \frac{k_1}{k_1 - k_2' + k_3} \exp(-k_2'[hv]t) + \frac{k_2' - k_3}{k_1 - k_2' + k_3} \exp(-(k_1 + k_3)[hv]t) \right) \quad (7)$$

We can differentiate these models based on MRM data using the power-modulated photo-reaction system, and obtain a best fit for the kinetic parameters. Note that the different appearance of $[D]_t$ is important to model validation, which varies upon the modulation of optical power density $[hv]$.

The exponent $(k_1 + k_3)[hv]$ can be obtained from the curve of A with the fitting to the single-exponential function. The curve of D contains two exponential terms. While it is difficult to fit $k_1[hv]$, $k_2[hv]$ and $k_3[hv]$ directly, $k_2[hv]$ can be fitted from the sectional curve of D directly, when the value of $1/[hv]$ is smaller than the maximum irradiation time t (with the assumption of $k_2[hv] < (k_1 + k_3)[hv]$), which could be validated with good fitting). With $(k_1 + k_3)[hv]$ and $k_2[hv]$ determined, the whole curve of D can be fitted to obtain the individual values of $k_1[hv]$ and $k_3[hv]$. As shown in **Fig. 1** and **Fig. 2**, for model **III**, $k_1 + k_3 = 2.32 \times 10^{-4} \text{ cm}^2/(\text{mW} \cdot \text{s})$, $k_1 = 9k_3$, $k_2 = 73 \times 10^{-4} \text{ s}^{-1}$.

Mathematic treatment for diazirine photo-reaction with tripeptide

Multiple elementary reactions can be involved for the generation SDA-HY (protonated form of the tripeptide AXA), as shown in the following Scheme. The sub-processes include I, in which a proton transfers from HY to **B**; II, III, IV, in which a proton transfers from HY to **B** upon the catalysis of a free proton; V, in which **B** reacts with free proton; VI, in which elimination reaction leads to the generation of by-products **B_{out}**.



Thus, the production of **[SDA-HY]** from diazo intermediate **B** and carbene intermediate **C** can be written as

$$[\text{SDA} - \text{HY}]_{[hv],t} = [\text{SDA} - \text{HY}]_{[hv],t}^{\text{B}} + [\text{SDA} - \text{HY}]_{[hv],t}^{\text{C}} \quad (8)$$

Here $[\text{SDA} - \text{HY}]_{[hv],t}^{\text{B}}$ is a complex function related to $[\mathbf{B}]_{[hv],t}$, $[\mathbf{HY}]$, and $[\mathbf{H}^+]$. For a specific irradiation time t , the production rate of **[SDA-HY]** from **B** can be written as

$$\frac{d[\text{SDA} - \text{HY}]_{[hv],t}^{\text{B}}}{dt} = k_{\text{VII}}[\mathbf{HB}^+]_{[hv],t}[\mathbf{Y}^-]_{[hv],t} + k_{\text{IX}}[\mathbf{SDA} - \mathbf{HYH}^+]_{[hv],t}[\mathbf{M}] \quad (9)$$

Here $[\mathbf{SDA} - \mathbf{HYH}^+]_{[hv],t}$ is a transient intermediate satisfying the steady-state approximation. For simplicity, with a relatively low photo-reaction yield and with **HY** in excess, **[HY]** can be assumed constant as shown, which leads to

$$0 \equiv \frac{d[\text{SDA} - \text{HYH}^+]_{[hv],t}}{dt} = k_{\text{VIII}}[\text{HB}^+]_{[hv],t}[\text{HY}] - k_{\text{IX}}[\text{SDA} - \text{HYH}^+]_{[hv],t}[\text{M}] \quad (10)$$

Equation (10) can then be rearranged as

$$\frac{d[\text{SDA} - \text{HY}]_{[hv],t}^{\text{B}}}{dt} = k_{\text{VII}}[\text{HB}^+]_{[hv],t}[\text{Y}^-]_{[hv],t} + k_{\text{VIII}}[\text{HB}^+]_{[hv],t}[\text{HY}] \quad (11)$$

$[\text{HB}^+]_{[hv],t}$ is also a short-lived intermediate that satisfies steady-state approximation.

Thus, we can obtain the following equation

$$\begin{aligned} 0 &\equiv \frac{d[\text{HB}^+]_{[hv],t}}{dt} \\ &= k_{\text{I}}[\text{B}]_{[hv],t}[\text{HY}] + k_{\text{III}}[\text{B}]_{[hv],t}[\text{HM} - \text{HY}^+]_{[hv],t} \\ &\quad + k_{\text{IV}}[\text{B}]_{[hv],t}[\text{HM} - \text{HY}^+]_{[hv],t} + k_{\text{V}}[\text{B}]_{[hv],t}[\text{HM}^+]_{[hv],t} - k_{\text{VI}}[\text{HB}^+]_{[hv],t}[\text{M}] \\ &\quad - k_{\text{VII}}[\text{HB}^+]_{[hv],t}[\text{Y}^-]_{[hv],t} - k_{\text{VIII}}[\text{HB}^+]_{[hv],t}[\text{HY}] \end{aligned} \quad (12)$$

By omitting the slow processes shown in the Scheme, we can obtain

$$0 \equiv \frac{d[\text{HB}^+]_{[hv],t}}{dt} = k_{\text{IV}}[\text{B}]_{[hv],t}[\text{HM} - \text{HY}^+]_{[hv],t} + k_{\text{V}}[\text{B}]_{[hv],t}[\text{HM}^+]_{[hv],t} - k_{\text{VI}}[\text{HB}^+]_{[hv],t}[\text{M}] \quad (13)$$

Equation (13) can be rearranged to

$$[\text{HB}^+]_{[hv],t} = \frac{k_{\text{IV}}[\text{B}]_{[hv],t}[\text{HM} - \text{HY}^+]_{[hv],t} + k_{\text{V}}[\text{B}]_{[hv],t}[\text{HM}^+]_{[hv],t}}{k_{\text{VI}}[\text{M}]} \quad (14)$$

For the sub-reaction **II** in all possible elementary reactions shown in the Scheme, $[\text{HM} - \text{HY}^+]_{[hv],t}$ and $[\text{HM}^+]_{[hv],t}$ are in rapid equilibrium. Therefore, we can obtain

$$k_{\text{II}}[\text{HM}^+]_{[hv],t}[\text{HY}] = k_{\text{II}}^{-1}[\text{HM} - \text{HY}^+]_{[hv],t} \quad (15)$$

Therefore, $[\text{HB}^+]_{[hv],t}$ has a proportional relationship with $[\text{HM}^+]_{[hv],t}$ (from (14) and (15)). With charge balance, we can obtain that $[\text{HB}^+]_{[hv],t}$ has a proportional relationship with $[\text{Y}^-]_{[hv],t}$.

We can then obtain the following equation by combining Equations (11), (14), and (15)

$$\frac{d[\text{SDA} - \text{HY}]_{[hv],t}^{\text{B}}}{dt} = k_{\text{VII}}[\text{HB}^+]_{[hv],t}[\text{Y}^-]_{[hv],t} + k_{\text{VIII}}[\text{HB}^+]_{[hv],t}[\text{HY}] = k_{\text{b}}[\text{B}]_{[hv],t} \quad (16)$$

As $[\text{HB}^+]_{[hv],t}$ has a proportional relationship with $[\text{Y}^-]_{[hv],t}$, we can obtain the rate constant k_b , which has the following relationship with $[\text{Y}^-]$ (m_1 and m_2 are scaling factors):

$$k_b \propto m_1[\text{Y}^-]_{[hv],t} + m_2[\text{Y}^-]_{[hv],t}^2 \quad (17)$$

If $k_I \gg k_{III}$ (t' is an integration variable):

$$\frac{d[\text{Y}^-]_{[hv],t}}{dt} = k_I[\text{B}]_{[hv],t}[\text{HY}] \quad (18)$$

$$[\text{Y}^-]_{[hv],t}^I \propto \int_0^t [\text{B}]_{[hv],t'} dt' \quad (19)$$

If $k_I \ll k_{III}$:

$$\frac{d[\text{Y}^-]_{[hv],t}}{dt} = k_{III}[\text{B}]_{[hv],t}[\text{HM} - \text{HY}^+]_{[hv],t} \propto [\text{B}]_{[hv],t}[\text{Y}^-]_{[hv],t} \quad (20)$$

$$[\text{Y}^-]_{[hv],t}^{III} \propto \exp\left(\int_0^t [\text{B}]_{[hv],t'} dt'\right) \quad (21)$$

We can thus obtain the following simplified equations, with m_1 , m_2 , n_1 , and n_2 the pre-factors

$$[\text{Y}^-]_{[hv],t} = n_1[\text{Y}^-]_{[hv],t}^I + n_2[\text{Y}^-]_{[hv],t}^{III} \quad (22)$$

$$I_{b_1} = \int_0^t m_1 n_1 [\text{B}]_{[hv],t'} [\text{Y}^-]_{[hv],t'}^I dt' \quad (23)$$

$$I_{b_2} = \int_0^t m_2 n_1 [\text{B}]_{[hv],t'} ([\text{Y}^-]_{[hv],t'}^I)^2 dt' \quad (24)$$

$$I_{b_3} = \int_0^t m_1 n_2 [\text{B}]_{[hv],t'} [\text{Y}^-]_{[hv],t'}^{III} dt' \quad (25)$$

$$I_{b_4} = \int_0^t m_2 n_2 [\text{B}]_{[hv],t'} ([\text{Y}^-]_{[hv],t'}^{III})^2 dt' \quad (26)$$

Thus, the diazo-mediated production of **SDA-HY** can be written as

$$[\text{SDA} - \text{HY}]_{[hv],t}^B = \int_0^t k_b [\text{B}]_{[hv],t'} dt' \approx b_1 I_{b_1} + b_2 I_{b_2} + b_3 I_{b_3} + b_4 I_{b_4} \quad (27)$$

With $[\text{B}]_{[hv],t} = \frac{k_1[A]_0}{k_1 - k_2} (\exp(-k_2[hv]t) - \exp(-(k_1)[hv]t))$, the following equations can be obtained for I_{b_1} and I_{b_2}

$$I_{b_1} = \frac{A_0^2 \left(\frac{e^{-2[hv](k_1+k_2)t_1} (k_1(-e^{[hv]k_1t_1}) + (k_1 - k_2)e^{[hv](k_1+k_2)t_1} + k_2e^{[hv]k_2t_1})^2 + 2[hv]k_1k_2(e^{-[hv]k_2t_1} - e^{-[hv]k_1t_1}))}{\left(-\frac{k_1(e^{-[hv]k_2t_1} - e^{-[hv]k_2t_2})}{[hv]k_2} + \frac{k_2(e^{-[hv]k_1t_1} - e^{-[hv]k_1t_2})}{[hv]k_1} \right) + k_1(-t_1) + k_1t_2 + k_2t_1 - k_2t_2} \right)}{2[hv]^2(k_1 - k_2)^2k_2^2} \quad (28)$$

$$I_{b_2} = \frac{A_0^2 \left(\frac{e^{-2[hv](k_1+k_2)t_1} (k_1(-e^{[hv]k_1t_1}) + (k_1 - k_2)e^{[hv](k_1+k_2)t_1} + k_2e^{[hv]k_2t_1})^2 + 2[hv]k_1k_2(e^{-[hv]k_2t_1} - e^{-[hv]k_1t_1}))}{\left(-\frac{k_1(e^{-[hv]k_2t_1} - e^{-[hv]k_2t_2})}{[hv]k_2} + \frac{k_2(e^{-[hv]k_1t_1} - e^{-[hv]k_1t_2})}{[hv]k_1} \right) + k_1(-t_1) + k_1t_2 + k_2t_1 - k_2t_2} \right)}{2[hv]^2(k_1 - k_2)^2k_2^2} \quad (29)$$

Here, $(t_2 - t_1)$ represents the time for the sample flowing in the PFA tube to MS detection. The parameters can be obtained from experimental observables by using the in-line power-modulated photo-reaction system with these parameters: $k_1 = 0.000232 \text{ cm}^2/(\text{mW} \cdot \text{s})$, $k_2 = 0.000073 \text{ cm}^2/(\text{mW} \cdot \text{s})$, $t_1 = 59 \text{ s}$, $t_2 = 80 \text{ s}$, and $A_0 = 1$.

Though I_{b_3} and I_{b_4} lack analytical mathematical solutions, their numerical results can be fitted, resulting the following equations:

$$I_{b_3} \text{ or } I_{b_4} = A \times \left(1 / (1 + \exp(-(x - x_c + w_1/2)/w_2)) \right) \times \left(1 - 1 / (1 + \exp(-(x - x_c - w_1/2)/w_3)) \right) \quad (30)$$

For I_{b_3} , $x_c = 170.82$, $A = 1.288$, $w_1 = 63.0$, $w_2 = 11.93$, $w_3 = 20.20$. For I_{b_4} , $x_c = 170.60$, $A = 1.286$, $w_1 = 44.7$, $w_2 = 9.17$, $w_3 = 13.32$. Excellent fitting could be obtained ($R^2 = 0.9999$).

For the production of **SDA-HY** from the carbene intermediate **C**, the time-dependent change of **[SDA-HY]** can be written as:

$$[\text{SDA} - \text{HY}]_{[hv],t}^c = [A]_0 \left(1 - \frac{k_1}{k_1 - k_2} \exp(-k_2[hv]t) + \frac{k_2}{k_1 - k_2} \exp(-k_1[hv]t) \right) \times \frac{k_c}{k_c + k_{ox} + k_{others}} = cI_c \quad (31)$$

In brief, according to the k_1 , k_2 given by Model I (a simplified model to Model II) and experiments, we can deduce and calculate the normalized yield of the product **[SDA-HY]** – $[hv]$ curve of **B** or **C** at $t = 1$ min. Assuming that the **HY** is in excess, its concentration can be regarded as a constant; whereas the concentration of **B** and the cumulative yield of **C** are the functions of k_1 , k_2 , $[hv]$, and t . The diazo-mediated reaction can be dissected into elementary reactions involving the transfer of proton, which is thus related to the capability of **HY** to provide proton (b_1 and b_2 processes) and to the proton concentration in the solvent (the enhancement of b_3 over b_1 , and the enhancement of b_4 over b_2). Thus, the acidity of the protein side-chain (pKa value) and the buffer (pH) can have an effect on the kinetic rate of diazo-mediated photo-reaction.

Supplementary References

- S1 Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J. & Wilke, C. O. Maximum allowed solvent accessibilities of residues in proteins. *PLoS ONE* **8**, e80635(2013).
- S2 Admasu, A. *et al.* A laser flash photolysis study of p-tolyl(trifluoromethyl)carbene. *J. Chem. Soc., Perkin Trans. 2*, 1093-1100(1998).
- S3 Toscano, J. P., Platz, M. S. & Nikolaev, V. Lifetimes of simple ketocarbenes. *J. Am. Chem. Soc.* **117**, 4712-4713(1995).
- S4 O'Brien, J. G. K., Jemas, A., Asare-Okai, P. N., Am Ende, C. W. & Fox, J. M. Probing the mechanism of photoaffinity labeling by dialkyldiazirines through bioorthogonal capture of diazoalkanes. *Org. Lett.* **22**, 9415-9420(2020).
- S5 Iacobucci, C. *et al.* Carboxyl-photo-reactive ms-cleavable cross-linkers: Unveiling a hidden aspect of diazirine-based reagents. *Anal. Chem.* **90**, 2805-2809(2018).
- S6 Piersimoni, L., Kastiris, P. L., Arlt, C. & Sinz, A. Cross-linking mass spectrometry for investigating protein conformations and protein–protein interactions—a method for all seasons. *Chem. Rev.* **122**, 7500-7531(2021).