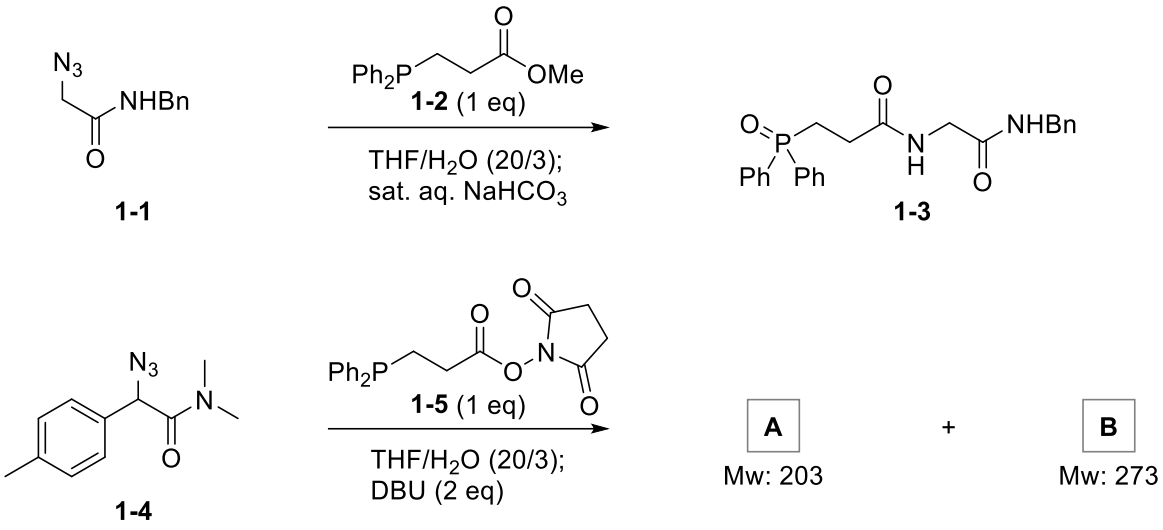
Problem Session

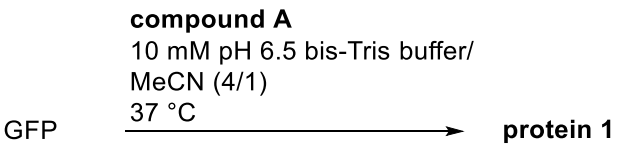
Oct. 1, 2022

Hiroaki Itoh

1-1. Please explain the reaction mechanisms and provide the structures of **compounds A** and **B**.

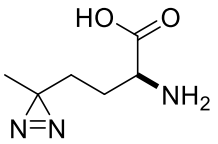


1-2. Green fluorescent protein (GFP) or **protein 1** (GFP treated with **compound A**, see below) was added to mammalian cells. After the incubation at 37 °C, GFP-derived intracellular fluorescence was observed only from the cells incubated with **protein 1**. Please explain the reason based on the possible reactions of **compound A** with GFP.

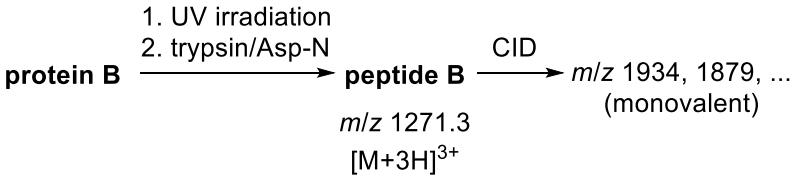


2. **Protein 2** possesses L-photo-methionine (**2-1**) residues instead of several L-methionine residues. After UV irradiation (320-400 nm), the protein was digested by trypsin/Asp-N. The resultant peptides were subjected to LC-MS/MS analysis.

Peptide B (*m*/*z* 1271.274 $[M+3H]^{3+}$) was detected in the LC-MS, which indicated the crosslinking between **peptide B_1** (VPDNEEATQYVEAxFR, "x" indicates photo-methionine) and **peptide B_2** (MLQMDLNPSSWISQQR). In the MS/MS analysis of **peptide B** by collision-induced dissociation (CID), product ions *m*/*z* 1934 and *m*/*z* 1879 (monovalent) were observed. Please predict the residue that involves with the reaction in **peptide B_2**, and explain the possible mechanisms of the reaction and the fragmentation by CID.



L-photo-methionine (**2-1**)



Diazo/diazirine compounds for protein modification

1. Background

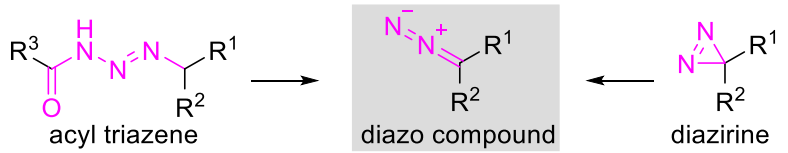


Figure 1. Generation of diazo compounds from acyl triazene and diazirine.

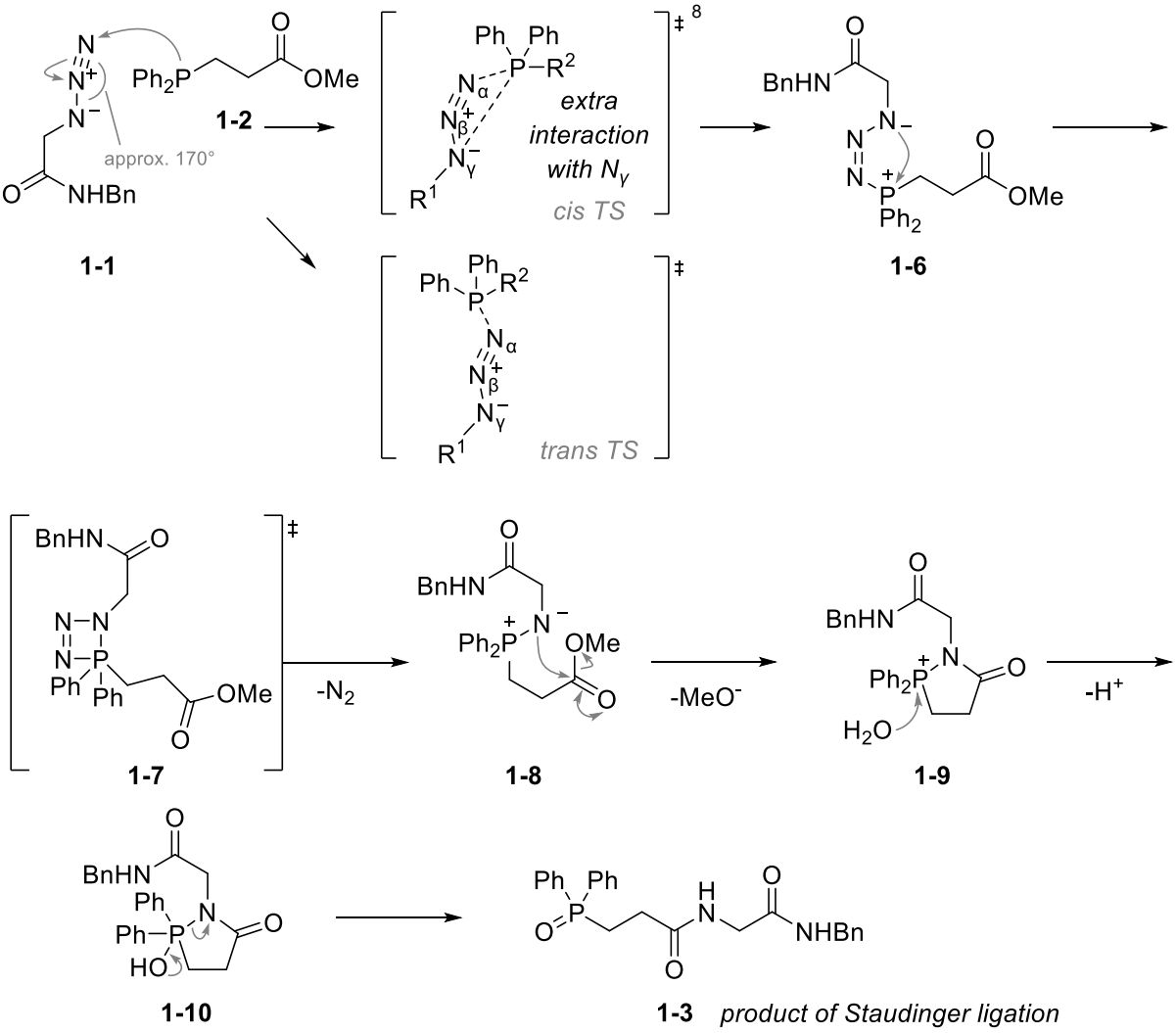
・Generation of diazo compounds from acyl triazene species and its application for modifying the membrane permeability of proteins have been reported by Raines group.[1,2,3,4]

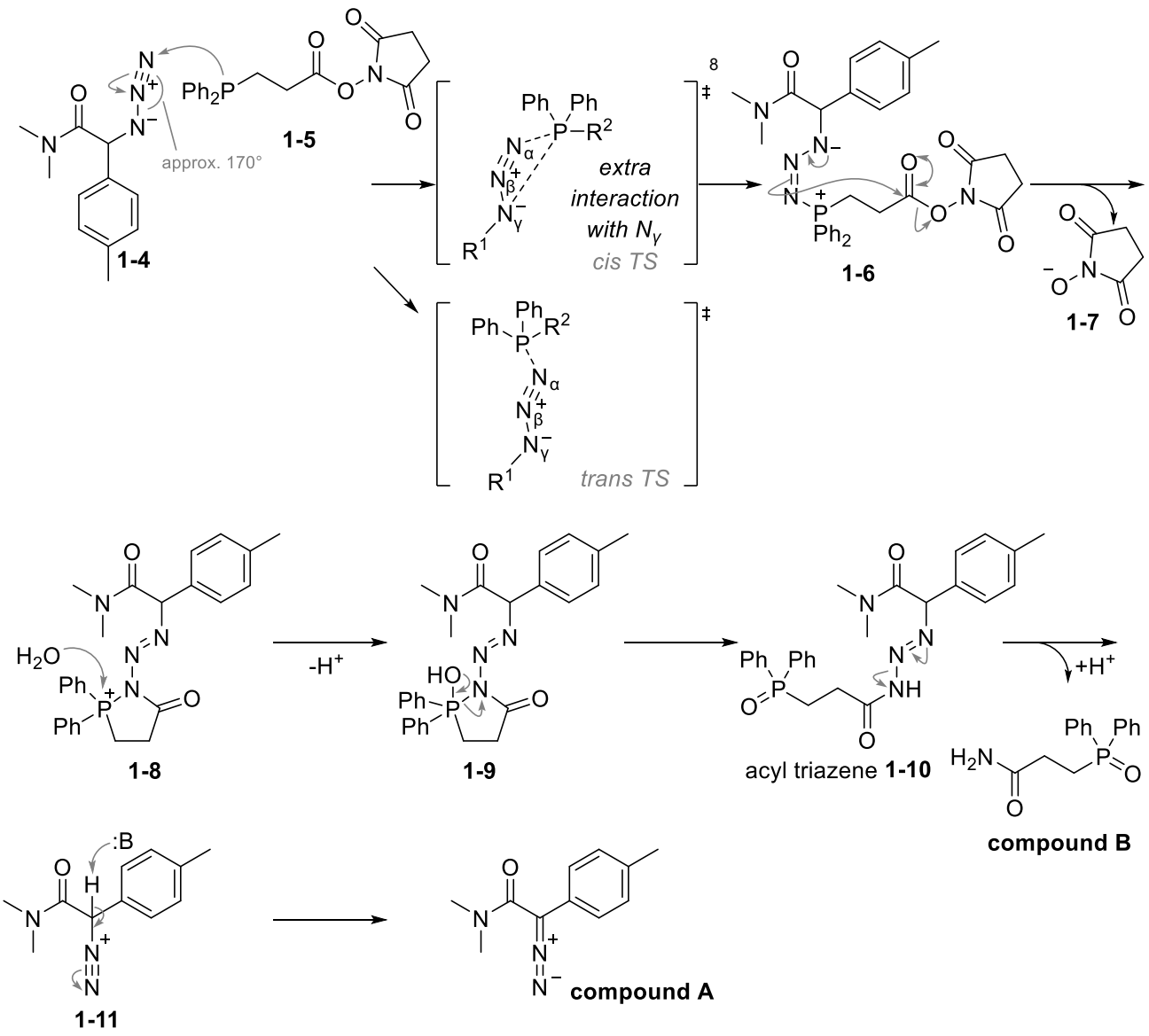
・In the context of photoaffinity labeling (PAL), generation of diazo compounds from diazirines had been recognized as undesired pathway because of the reactivity in the dark and longer solution life-time (larger labeling radius) than carbene species. However, revealing the contribution of these species has become important to understand the labeling preference of the alkyl diazirine PAL tag.[5]

2. Diazo compound from acyl triazene

2-1. Answer of problem 1

・Acyl triazene species is possibly generated as a side product of Staudinger ligation.[6,7]





・Key points of answer to problem 1-2

1. Side chains of D (Asp) and E (Glu), and C-terminal carboxylic acid could be esterified by **compound A**.
2. The negative charge is reduced or eliminated by esterification.
3. The cationic and neutral molecules are delivered more readily into cells than anionic molecules.
4. Referring to the possibility of the hydrolysis of the ester in the cells.

2-2. Determination of acyl triazene structure[6]

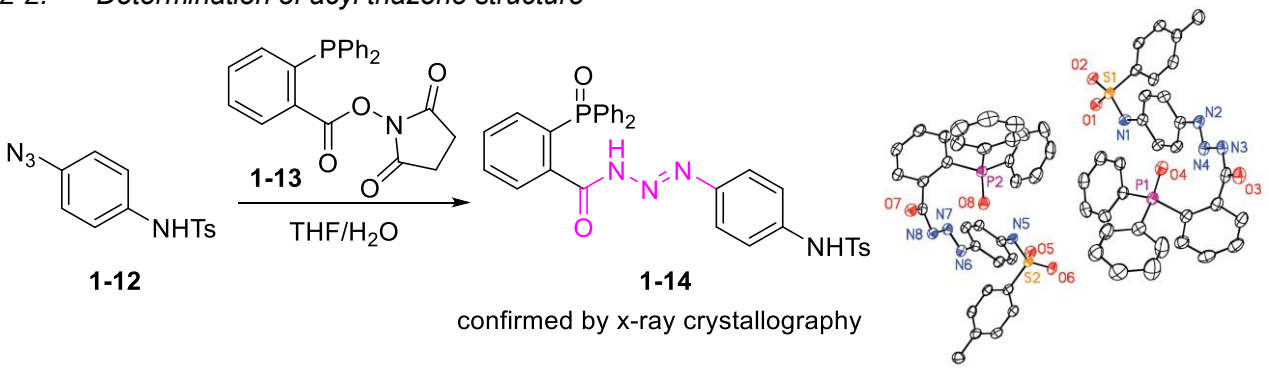
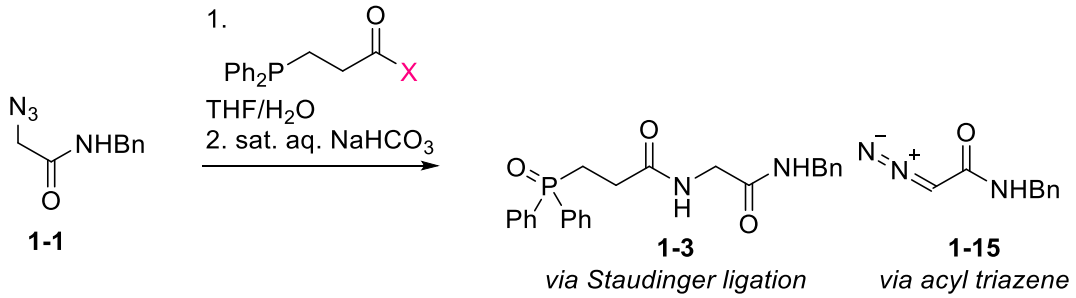


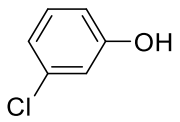
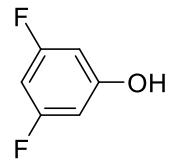
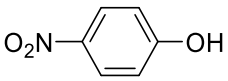
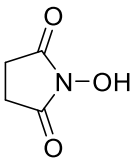
Figure 2. Generation of thermally stable acyl triazene **14** from aryl azide **12**.

2-3. Ratio of ligation and diazo product[9]

・Product ratio depends on the leaving group (pK_a, acyl migration vs. N_2 expulsion).

Table 1. Reactivity of phosphinoester



XH	pK_a	ratio of **1-3**:**1-15**
MeOH	15.5	100:0
	9.0	67:33
	8.7	27:63
	7.1	3:97
	6.0	2:98

2-4. Cell membrane permeability

・Cationic and neutral molecules are more readily delivered into cells than anionic molecules because of

a. Exclusion of hydrophilic charged molecules from the hydrophobic environment of the lipid bilayer

b. Electrostatic repulsion from the anionic glycocalyx (glycoprotein and glycolipid covering the membrane of cells)[10]

c. Plasma membrane potential [c.f. poly R,[11] $DiBAC_4(3)$ assay]

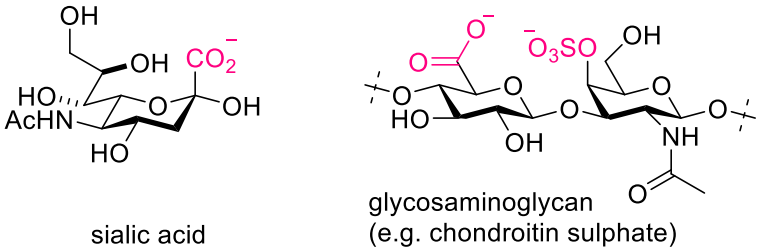


Figure 3. Representative anionic units in glycocalyx.

・Esterification of carboxylic acids [D (Asp), E (Glu), and C-terminus, indicated in red in the below sequence] of proteins can reduce the charge repulsion and increase hydrophobicity, which could facilitate the membrane permeability.

superfolder GFP (sfGFP, the N-terminus was modified by His6-spacer-TEV protease recognition sequence)
261 amino acids (exact mass 29,343 Da)
sequence:

MHHHHHHSSG VDLGTENLYF QGMVSKGEEL FTGVVPILVE LDGDVNGHKF SVRGEGEGDA TIGKLTLKFI
CTTGKLPVPW PTLVTTLTYG VQCFSRYPDH MKQHDFFKSA MPEGYVQERT ISFKDDGKYK TRAVVKFEGD
TLVNRIELKG TDFKEDGNIL GHKLEYNFNS HNVYITADKQ KNGIKANFTV RHNVEDGSVQ LADHYQQNTP
IGDGPVLLPD NHYLSTQTVL SKDPNEKRDH MVLHEYVNAA GITLGMDELY K(C-terminus)

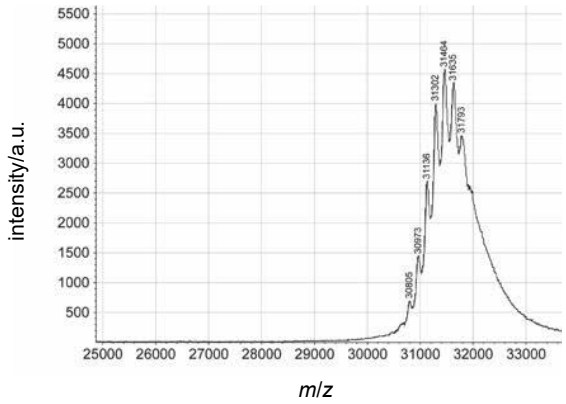


Figure 4. MALDI-TOF MS spectrum of GFP esterified with **compound A** (100 eq, 3 eq. per carboxyl group). *m/z* + 175 u per ester.

・The esterified protein permeates the cell membrane even at 4 °C, indicating that endocytotic pathway does not involve with the cellular uptake of the protein (Figure 5).

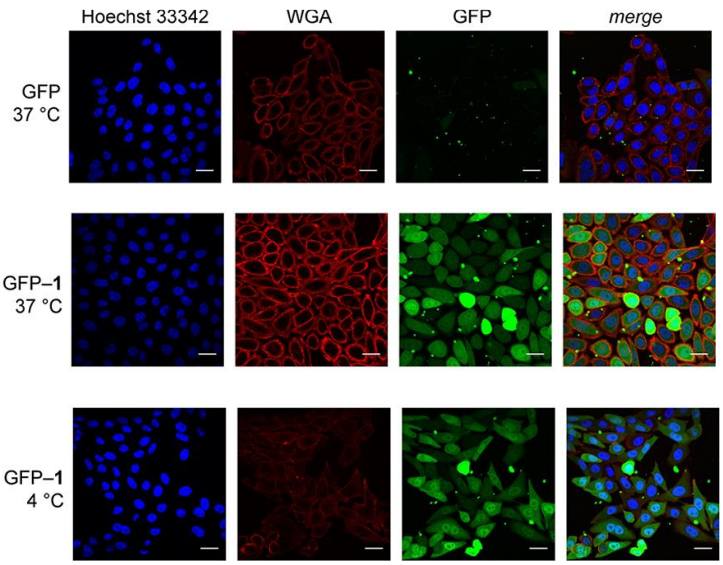


Figure 5. Fluorescence imaging of GFP/**protein 1**-treated CHO (Chinese hamster ovary)-K1 cells. The cells were incubated with protein (15 μM) for 2 h at 37 °C. WGA = fluorescence-labeled wheat germ agglutinin (plasma membrane indicator).

・The accumulation of GFP indicates that the esterified protein could be hydrolyzed by cytosolic esterase (similar to acetoxymethyl ester of fluorescent dyes such as calcein-AM):

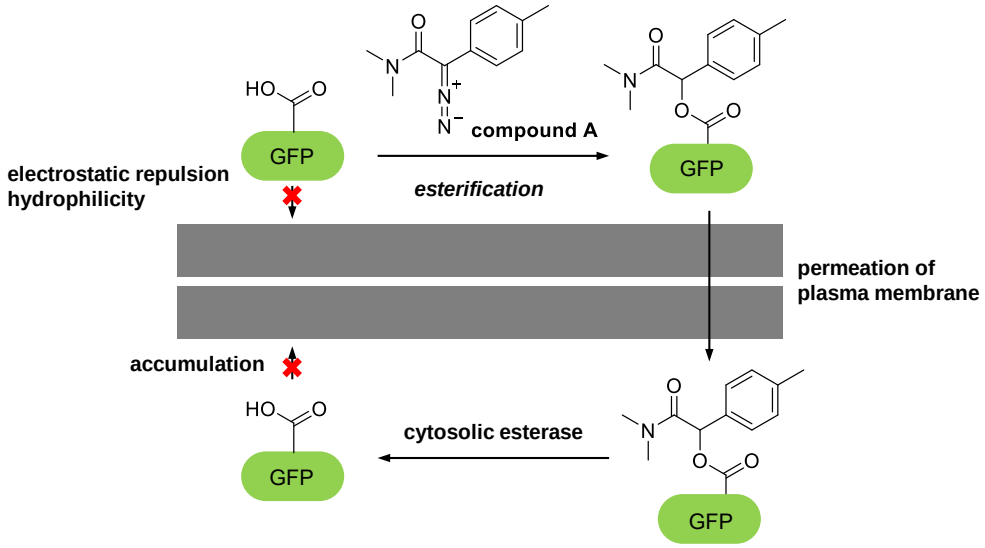


Figure 6. Schematic diagram of cellular accumulation of the GFP.

・Supporting data for involvement of the cytosolic esterase:

a. Unmodified protein signal (FLAG-angiogenin, *m*/*z* 15,270 + 175 u per ester) was observed in MALDI-MS analysis after treatment of the esterified protein with CHO cell extract.

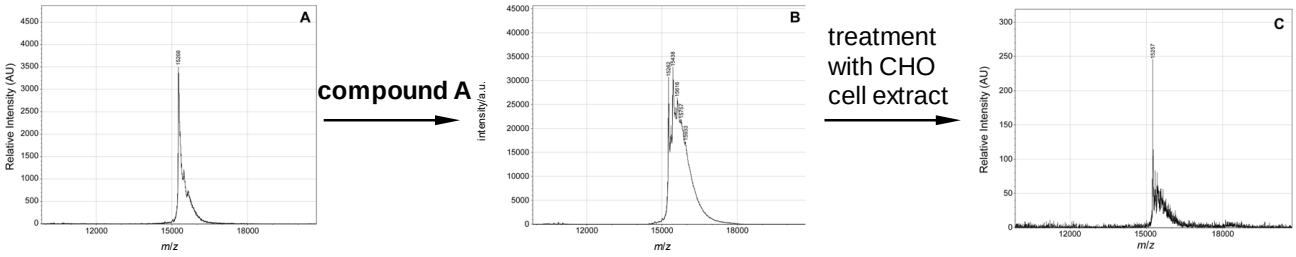


Figure 7. Change of MALDI-TOF MS spectra of FLAG-angiotensin.

b. Fluorescence intensity increased depending on the incubation time: intracellular trapping **protein 1** occurs by hydrolysis of the esters.

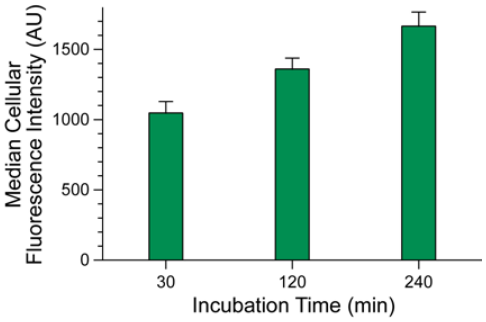


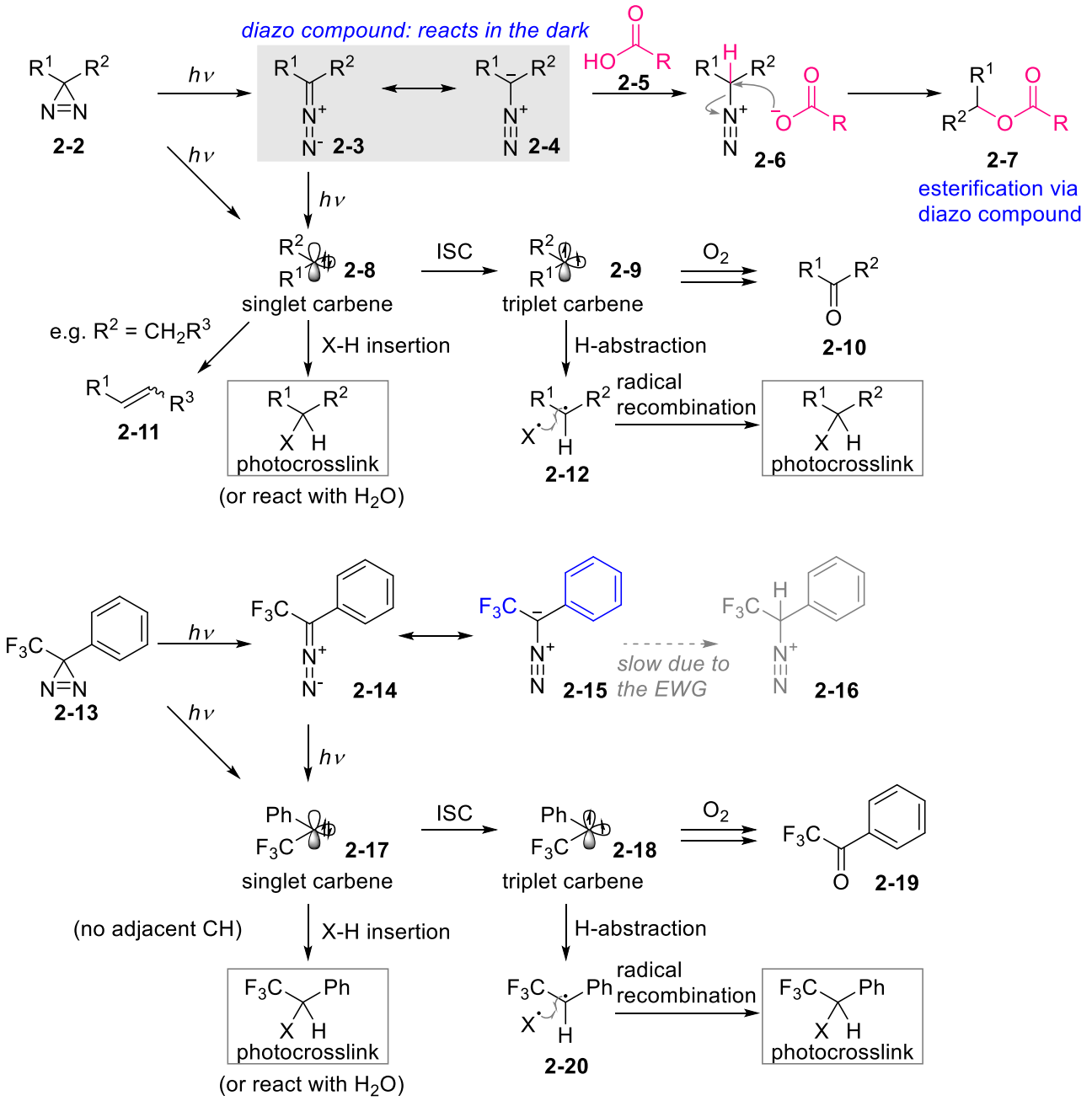
Figure 8. Time-course for the cellular internalization of **protein 1**. CHO-K1 cells were incubated with **protein 1** (4 μM) at 37 °C. The cellular fluorescence intensity was quantified by flow cytometry.

3. Diazo compound from diazirine

3-1. Diazirine species for PAL

・As a tag for PAL, 3-trifluoromethyl-3-phenyldiazirine (a.k.a. TPD, see also below) is now widely used.[12] In many reports, alkyl diazirines also gave successful results, while it is important to consider the difference in their reactivity for interpreting the outcomes.

3-2. Trifluoromethylphenyl diazirine vs alkyl diazirine



・In the case of TPD, triplet carbene is ground state,[13,14] although stable spin state could be affected by hydrogen bonding of solvent molecules.[15]

・note: solution half-life:[16] singlet carbene $t_{1/2}$ < 1 ns; triplet carbene $t_{1/2}$ = 2 μs; diazonium ion ~ 0.4 s

・As aforementioned, D (Asp), E (Glu), and C-terminal carboxylic acid are reactive against diazo compounds generated from the alkyl diazirines in the protein modification.[17,18]

3-3. Photo-methionine/photo-leucine for photo-crosslinking[19]

・The structures and properties of photo-methionine and photo-leucine show high similarity to the proteinogenic methionine (Met, M) and leucine (Leu, L), which allows these photo-activatable amino acids to be incorporated into proteins by the mammalian translation machinery. It is important that both amino acids behave as alkyl diazirines.

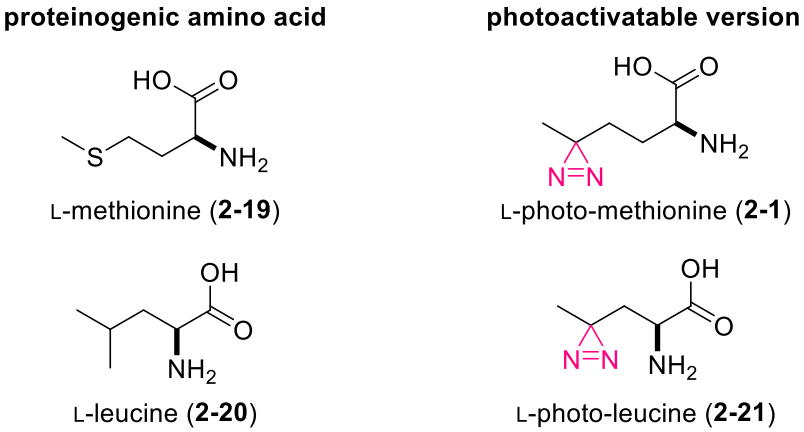


Figure 9. Structures of photo-methionine **2-1** and photo-leucine **2-21**.

3-4. Sequence of bovine GCAP2

Protein 1 is photomethionine-modified guanylyl cyclase-activating protein 2 (GCAP2):

guanylyl cyclase-activating protein 2 (GCAP2, *Bos taurus*)
204 amino acids (exact mass 23,728 Da)
sequence:
MGQQFSWEEA EENGAVGAAD AAQLQEWYKK FLEECPSGTL FMHEFKRFFK VPDNEEATQY
VEAMFRAFDT NGDNTIDFLE YVAALNLVLR GTLEHKLKWT FKIYDKDRNG CIDRQELLDI
VESIYKLKKA CSVEVEAEQQ GKLLTPEEVV DRIFLLVDEN GDGQLSLNEF VEGARRDKWV
MKMLQMDLNP SSWISQQRRK SAMF

・Several methionine (M) residues are substituted with photo-methionine (x) in **protein 1**. At least, Met64 (highlighted in red in the above sequence) is substituted with x.

peptide B_1 (residues 51-66: VPDNEEATQYVEAxFR) is highlighted in yellow.
exact mass: 1905.8
peptide B_2 (residues 183-198: MLQMDLNPSSWISQQR) is highlighted in cyan.
exact mass: 1932.9

・note: digestion of the protein for LC-MS/MS

a. Trypsin cleaves amide bonds at the C-terminal side of K (lysine) and R (arginine) residues

b. Asp-N cleaves amide bonds at the N-terminal side of D (aspartic acid) and C (cysteine) residues

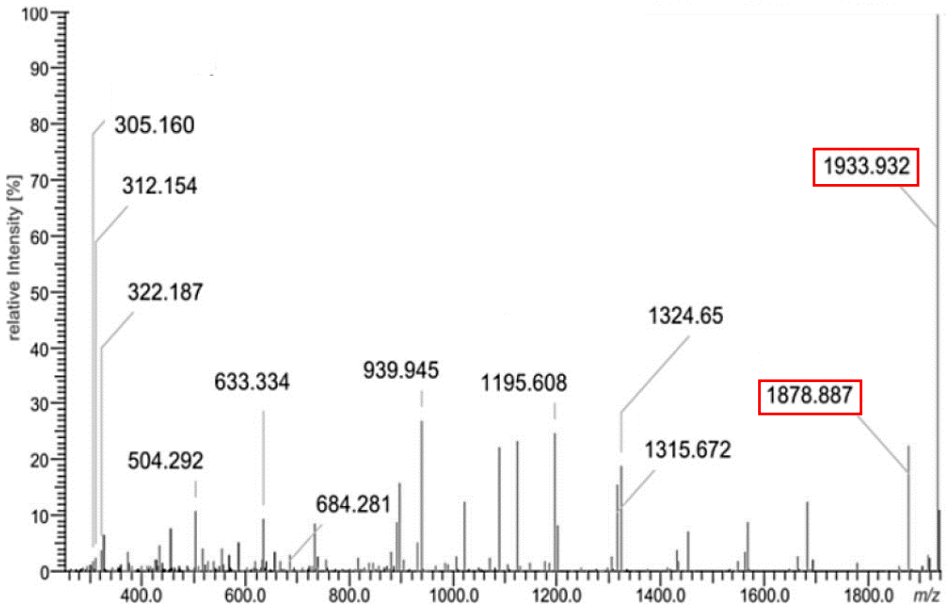
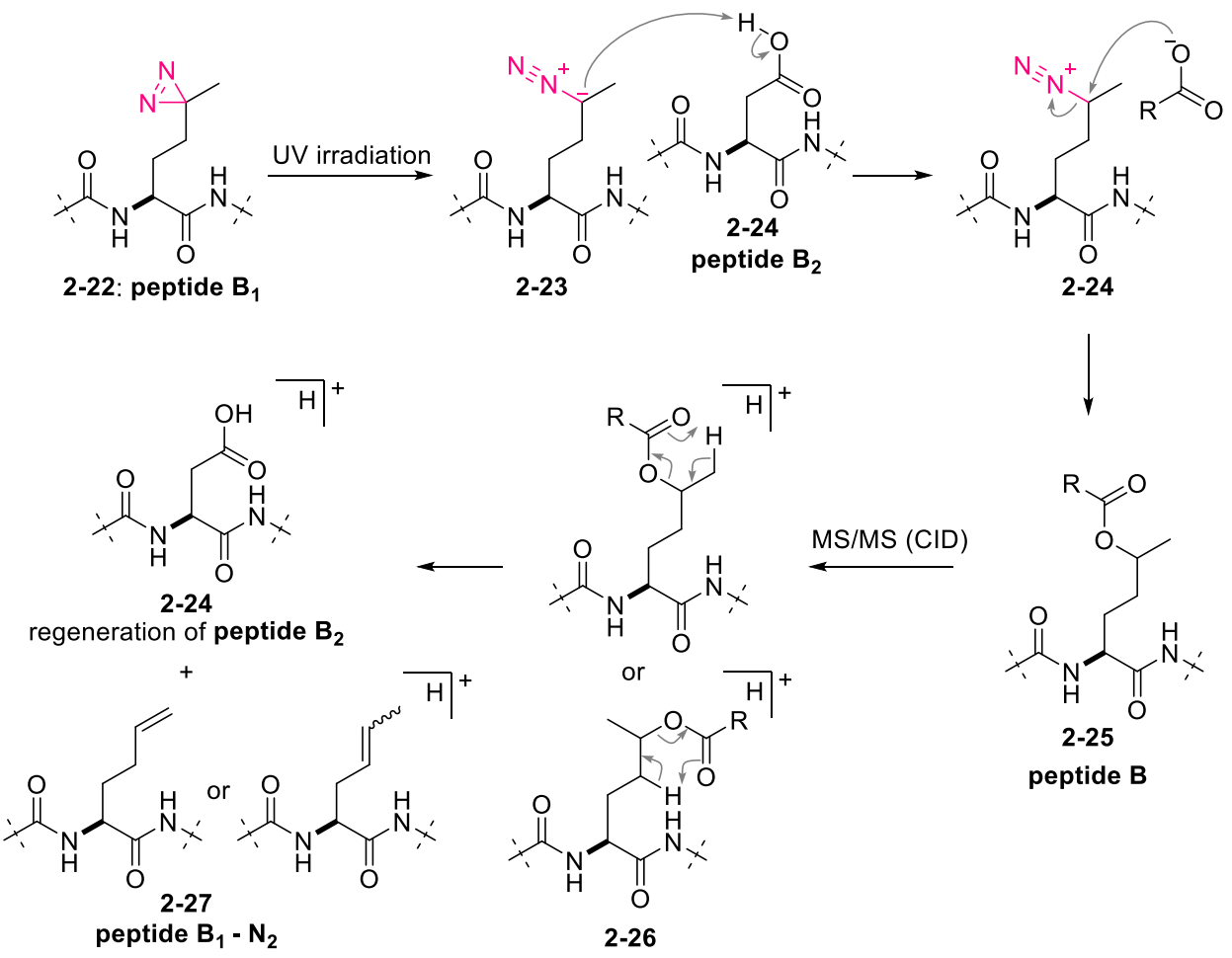


Figure 10. MS/MS spectrum of **peptide B**. The analysis was conducted on an Orbitrap Fusion Tribrid or an Orbitrap Q-Exactive Plus mass spectrometer.

・Based on the product ions in the MS/MS analysis, the carbene-mediated reaction is unlikely (signals 1934 and 1879 correspond to **peptide B_2** and **peptide B_1**-28 u $[M+H]^+$)

3-5. Answer of problem 2

・Based on the reactivity and fragmentation in the MS/MS, the esterification via diazo compound likely occurred at the side-chain carboxylic acid of D (Asp). In the three-dimensional structure, two amino-acid residues have spatial proximity (Figure 11).



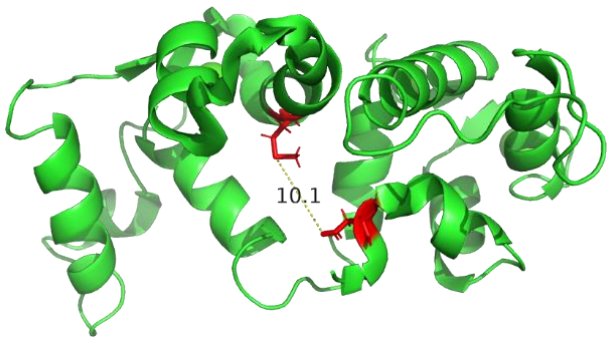


Figure 11. Three-dimensional structure of bovine GCAP2 (solution NMR-based structure, PDB ID: 1JBA). Met-64 and Asp-187 are indicated in red. The distance of the S atom of Met and the oxygen atom of the carboxylic acid of Asp is also displayed.

References

1. Mix, K. A.; Lomax, J. E.; Raines, R. T. Cytosolic delivery of proteins by bioreversible esterification. *J. Am. Chem. Soc.* **2017**, *139*, 14396-14398.
2. Mix, K. A.; Raines, R. T. Optimized diazo scaffold for protein esterification. *Org. Lett.* **2015**, *17*, 2358-2361.
3. McGrath, N. A.; Andersen, K. A.; Davis, A. K. F.; Lomax, J. E.; Raines, R. T. Diazo compounds for the bioreversible esterification of proteins. *Chem. Sci.* **2015**, *6*, 752-755.
4. Mix, K. A.; Aronoff, M. R.; Raines, R. T. Diazo compounds: Versatile tools for chemical biology. *ACS Chem. Biol.* **2016**, *11*, 3233-3244.
5. Lin, Y. LS211203
6. Myers, E. L.; Raines, R. T. A phosphine-mediated conversion of azides into diazo compounds. *Angew. Chem., Int. Ed.* **2009**, *48*, 2359-2363.
7. For other method for generation of diazo compounds from acyl triazenes, see: Lu, G.-H.; Huang, T.-C.; Hsueh, H.-C.; Yang, S.-C.; Cho, T.-W.; Chou, H.-H. Novel *N*-transfer reagent for converting α-amino acid derivatives to α-diazo compounds. *Chem. Commun.* **2021**, *57*, 4839-4842.
8. Tian, W. Q.; Wang, Y. A. Mechanisms of Staudinger reactions within density functional theory. *J. Org. Chem.* **2004**, *69*, 4299-4308.
9. Chou, H.-H.; Raines, R. T. Conversion of azides into diazo compounds in water. *J. Am. Chem. Soc*. **2013**, *135*, 14936-14939.
10. Palte, M. J.; Raines, R. T. Interaction of nucleic acids with the glycocalyx. *J. Am. Chem. Soc.* **2012**, *134*, 6218-6223.
11. Rothbard, J. B.; Jessop, T. C.; Lewis, R. S.; Murray, B. A.; Wender, P. A. Role of membrane potential and hydrogen bonding in the mechanism of translocation of guanidinium-rich peptides into cells. *J. Am. Chem. Soc.* **2004**, *126*, 9506-9507.
12. Brunner, J.; Senn, H.; Richards, F. M. 3-Trifluoromethyl-3-phenyldiazrine. *J. Biol. Chem.* **1980**, *255*, 3313-3318.
13. Musolino, S. F.; Pei, Z.; Bi, L.; DiLabio, G. A.; Wulff, J. E. Structure-function relationships in aryl diazirines reveal optimal design features to maximize C-H insertion. *Chem. Sci.* **2021**, *12*, 12138-12148.
14. Earley, D. F.; Guillou, A.; Klingler, S.; Fay, R.; Gut, M.; d'Orchymont, F.; Behmaneshfar, S.; Reichert, L.; Holland, J. P. *JACS Au* **2022**, *2*, 646-664.
15. Costa, P.; Sander, W. Hydrogen bonding switches the spin state of diphenylcarbene from triplet to singlet. *Angew. Chem., Int. Ed.* **2014**, *53*, 5122-5125.
16. Seath, C. P.; Trowbridge, A. D.; Muir, T. M.; MacMillan, D. W. C. Reactive intermediates for interactome mapping. *Chem. Soc. Rev.* **2021**, *50*, 2911-2926.
17. Iacobucci, C.; Götze, M.; Piotrowski, C.; Arlt, C.; Rehkamp, A.; Ihling, C.; Hage, C.; Sinz, A. Carboxyl-photo-reactive MS-cleavable cross-linkers: Unveiling a hidden aspect of diazirine-based reagents. *Anal. Chem.* **2018**, *90*, 2805-2809.
18. West, A. V.; Muncipinto, G.; Wu, H.-Y.; Huang, A. C.; Lebenski, M. T.; Jones, L. H.; Woo, C. M. Labeling preferences of diazirines with protein biomolecules. *J. Am. Chem. Soc.* **2021**, *143*, 6691-6700.
19. Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2005**, *2*, 261-267.