

Supporting Information

Sub-residue Resolution Footprinting of Ligand-Protein Interactions by Carbene Chemistry and Ion Mobility-Mass Spectrometry

Gaoyuan Lu¹, Xiaowei Xu², Gongyu Li³, Huiyong Sun¹, Nian Wang², Yinxue Zhu¹, Ning Wan², Yatao Shi³, Guangji Wang², Lingjun Li^{3,4}, Haiping Hao^{1,2,*}, Hui Ye^{2,*}

¹ School of Pharmacy, China Pharmaceutical University, Tongjiaxiang #24, Nanjing, Jiangsu 210009, China

² Key Laboratory of Drug Metabolism and Pharmacokinetics, State Key Laboratory of Natural Medicines, China Pharmaceutical University, Tongjiaxiang #24, Nanjing, Jiangsu 210009, China

³ School of Pharmacy, ⁴ Department of Chemistry, University of Wisconsin-Madison, 777 Highland Avenue, Madison, Wisconsin 53706, United States

* Correspondence: haipinghao@cpu.edu.cn (H.H.), Tel: 86-25-83271179,

cpuyehui@cpu.edu.cn (H.Y.), Tel: 86-25-83271191.

SUPPORTING METHODS

Evaluation of laser irradiation with HEWL activity assay. The activity of HEWL was detected using a Lysozyme Activity Kit (Sigma Aldrich, St. Louis, MO, USA) as previously reported¹. Briefly, *micrococcus lysodeikticus* cell suspensions were prepared in 20 mM Tris/150 mM NaCl buffer as the substrate of HEWL. A 100 μ L of *micrococcus lysodeikticus* cell suspension was mixed with 5 μ L testing solution, which contained 100 μ M HEWL in 20 mM Tris/150 mM NaCl buffer. The absorbance of the resultant reaction mixture at 450 nm was measured to evaluate the activity of HEWL after laser irradiation.

Residue level analysis using tandem MS. For chromatographically-unresolved peptides, the label ratio of residue/sub-residue was investigated with the method described by Jumper et al². Briefly, carbene labeled and unlabeled fragment ions (y ions in this case) can be obtained after MS/MS fragmentation. The MS/MS scans of the carbene-labeled peptide of interest were combined to deliver a sum spectrum that contains both labeled ions (n_i labeled) and unlabeled ions (n_i unlabeled). The label ratio of n_i residue was calculated using Supplementary Equation 1.

$$\text{Label Ratio}_{n_i} = \frac{I(n_i \text{ labeled})}{I(n_i \text{ labeled}) + I(n_i \text{ unlabeled})} * \text{Label Ratio}_{\text{peptide}} \quad (1)$$

The difference in label ratio between two consecutive fragments will generate the absolute label ratio for the specific residue (Supplementary Equation 2).

$$\text{Label Ratio}_{\text{residue}} = \text{Label Ratio}_{n_i} - \text{Label Ratio}_{n_{i-1}} \quad (2)$$

If the n_i sequence ions cannot be detected, the label ratio of n_i will be grouped together with the following fragment ions.

¹⁶KLVFFA²¹ isomer fractionation and dimethylation. The carbene labeled peptide ¹⁶KLVFFA²¹ was prepared in a relatively large scale (~2 mg) and desalted with Sep-Pack C₁₈

cartridges (Waters, USA). Samples were analyzed by Waters Synapt G2-Si mass spectrometer coupling with a Waters Acquity ultra-performance LC system (Milford, MA, USA). Separation was carried out on a Waters CSH C₁₈ column (150 mm × 2.1 mm, 1.7 μm) with 0.1% formic acid in water as the mobile phase A and 0.1% formic acid in ACN as the mobile phase B. Samples were eluted using a 25 min linear gradient of mobile phase B from 20% to 35% followed by 2 min at 85% B and 3 min of column re-equilibration with the flow rate set at 0.3 μL/min. The parameters of ESI and acquisition setting were described in the Methods section. Three isomers were validated by targeted MS/MS survey and fractionated according to the order of elution. The fractions were lyophilized and then dissolved in 10 μL mobile phase A. 1 μL of 1% CD₂O and 1 μL of 0.6 M NaBH₃CN solution were added to each collected fractions. The resultant peptides were analyzed to examine the mass shift induced by reductive dimethylation.

CCS measurement. The TWIMS was calibrated by the polyalanine solution as described by Bush et al³. Ion mobility data files were acquired by MassLynx 4.1 and analyzed by DriftScope 2.8 (Waters). Selection Tool was used to select the isomers according to the retention time. The MS/MS fragment ions of selected isomers with the corresponding arrival time values were subsequently exported to MassLynx .raw file. The CCS values were further calculated based on measured arrival time as described by Bush et al³ and Ruotolo et al⁴.

The CCS_{He} in TWIMS can be calculated with the following Supplementary Equation 3:

$$\Omega_{He} = \frac{ze}{16} \sqrt{\left[\frac{18\pi}{k_b T} \left(\frac{1}{m} + \frac{1}{M_{He}} \right) \right]} \frac{760}{P} \frac{T}{273.2} \frac{1}{NL} A t^B \quad (3)$$

z is the charge of the ion, e is the elementary charge, k_b is the Boltzmann's constant, m is the mass of ion, M_{He} is the mass of helium gas, P is the pressure, T is the temperature, N is the helium gas number density, L is the length of the drift region and t is the experimental arrival time. A and B are

constants reflecting the non-uniformity of travelling wave electric field, which vary in different instruments. A and B can be determined from standard calibration. For calibration, the following equations can be used.

The corrected arrival time t' was calculated from experimental arrival time t using Supplementary Equation 4.

$$t' = t - \left(\frac{C\sqrt{m/z}}{1000} \right) \quad (4)$$

C is a constant designated as “enhanced duty cycle delay coefficient”. m/z is the mass-to-charge ratio of the ions of interest. The corrected CCS Ω' was corrected with charge and mass (Supplementary Equation 5).

$$\Omega' = \frac{\Omega_{\text{He}}}{z \sqrt{\frac{1}{m} + \frac{1}{M_{\text{He}}}}} \quad (5)$$

A plot of $\ln(t')$ was subsequently plotted against $\ln(\Omega')$ using polyalanine ions. Then, constant A and B can be determined from the intercept and slope of the line. The correlation coefficient (R^2) of the fit should be more than 0.98. Hence, the doubly corrected t'' can be obtained by Supplementary Equation 6.

$$t'' = z(t')^B \sqrt{\frac{1}{m} + \frac{1}{M_{\text{He}}}} \quad (6)$$

Next, the line was replotted with the doubly corrected t'' . The equation of the final line was used to determine CCS based on the measured arrival time of the detected ions.

CCS prediction. DeepCCS software⁵ was downloaded from the website (<https://github.com/plpla/DeepCCS>). The structures of the fragment ions of interest were drawn using ChemDraw (PerkinElmer) and saved in .mol files. A Python Rdkit module (<http://www.rdkit.org/>) was used to convert .mol files to the canonical simplified molecular-input line-entry system (SMILES) annotation. Subsequently, a .csv file includes structure name, SMILES annotation, adduct type ($M+H^+$)

was manually generated. Lastly, the predicted CCS value was obtained by using the “predict” function in DeepCCS software according to the information included in the .csv file. The median relative error of DeepCCS prediction was approximately 2.7%.

FIGURES

Figure S1. The structure of trifluoromethylaryl diazine (TFMAD) and the reaction of carbene labeling.

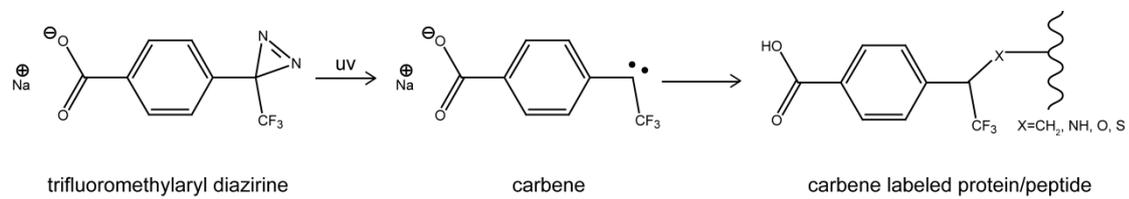


Figure S2. Evaluation of laser irradiation on the enzymatic activity of HEWL. The *micrococcus lysodeikticus* cells were used as the substrate of HEWL (see Methods). After adding HEWL to the cells, the absorbance of the *micrococcus lysodeikticus* cells suspension decreased significantly. The laser irradiation setup used in this study had negligible impact on the enzymatic activity of HEWL using student's t-test.

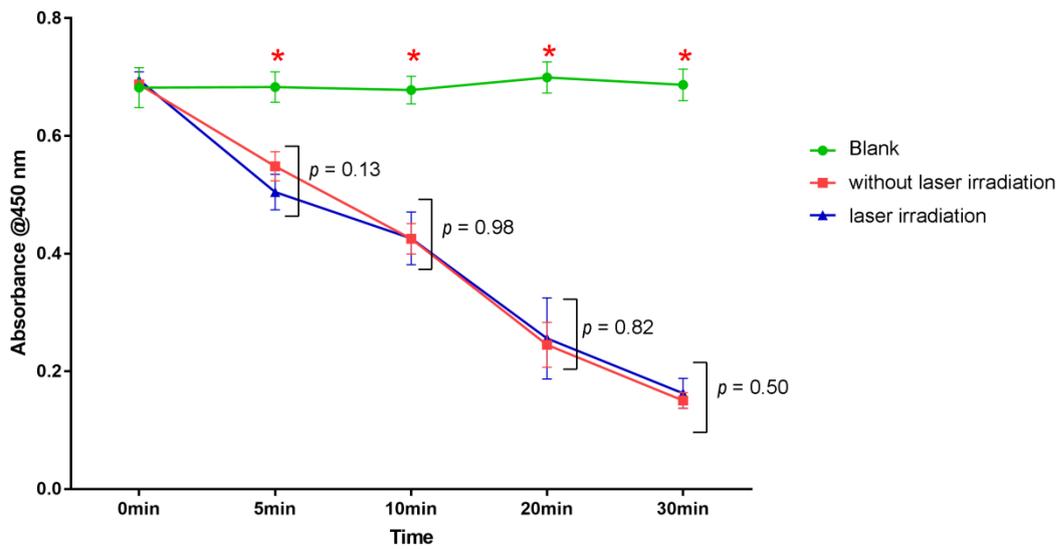


Figure S3. Residue level footprinting of HEWL-NAG4 interaction. (A) Label ratio per amino acid residue for HEWL footprinted with TFMAD in the presence (black bars) and absence (white bars) of NAG4. Error bars are \pm s.d. and significant differences (Student's t-test, $p < 0.05$, $n = 3$) are highlighted with a red or blue dot. Significant difference of Trp62 reactivity is highlighted with a yellow dot. (B) The structure of HEWL footprinted by TFMAD is shown in a surface mode. Color scheme: red = significantly masked by NAG4 binding, blue = increased label ratio in the presence of NAG4, wheat = no difference induced by NAG4 binding, grey = not covered by peptide mapping. Structure is constructed based on the PDB file 1LZC.

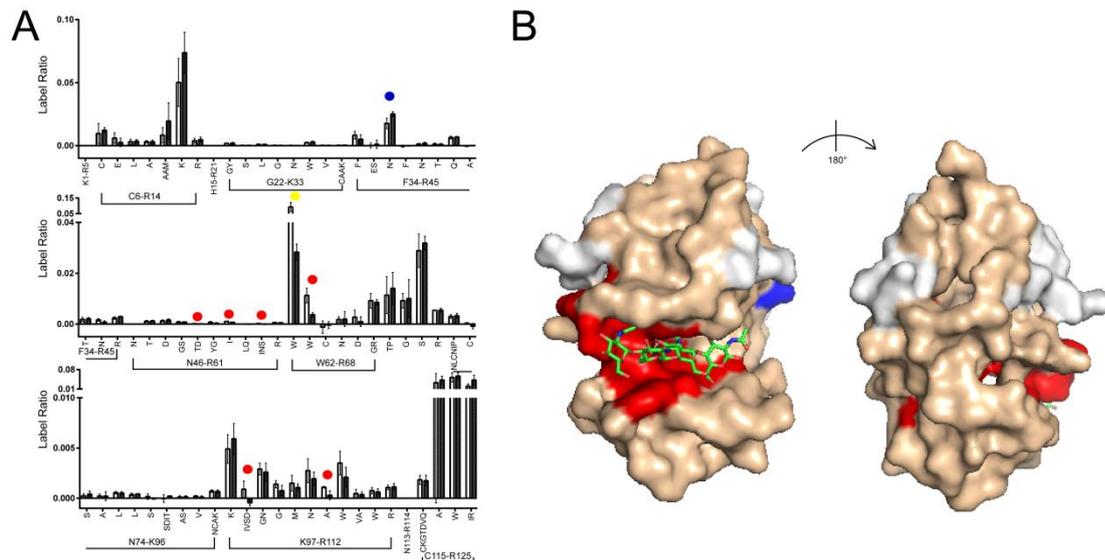


Figure S4. The MS/MS spectra of the three sub-residue isomers that share an identical sequence of $^{62}\text{W}^{\text{carbene}}\text{WCNDGR}^{68}$ yet differ in the X-H bond of carbene insertion at Trp62.

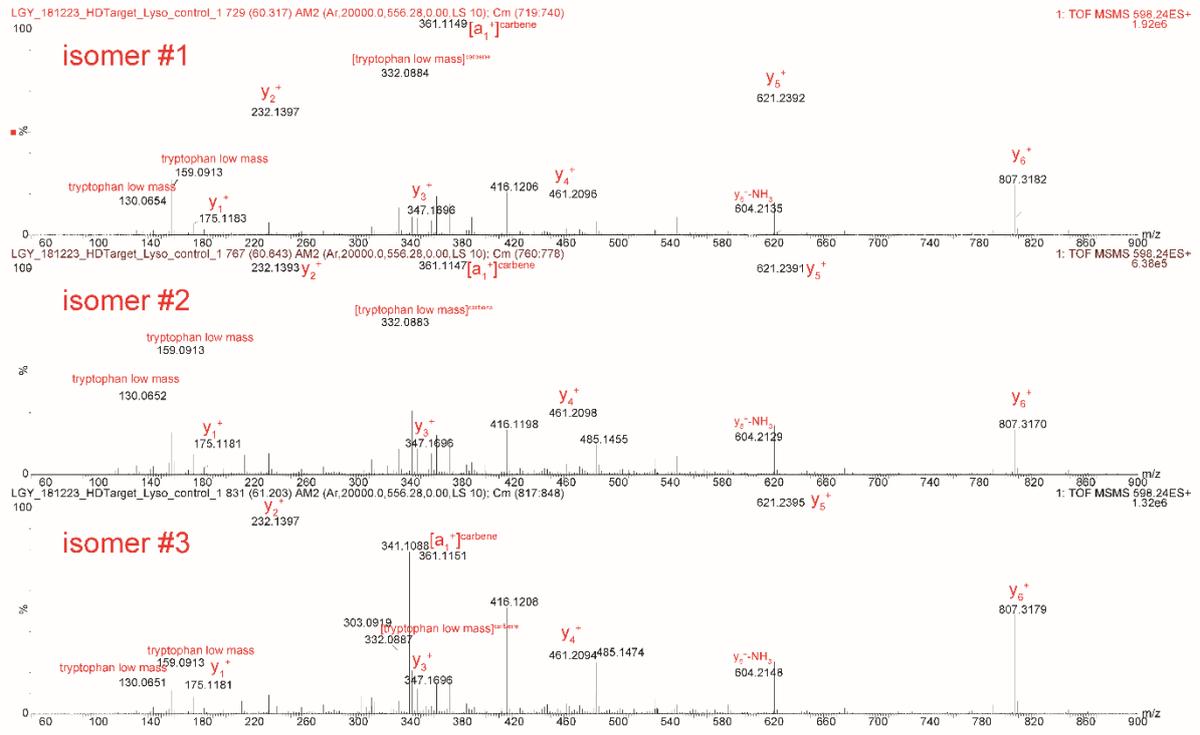


Figure S5. Extracted arrival time distributions of the fragment ions produced from three Trp62

isomers. The measurement was performed in triplicates.

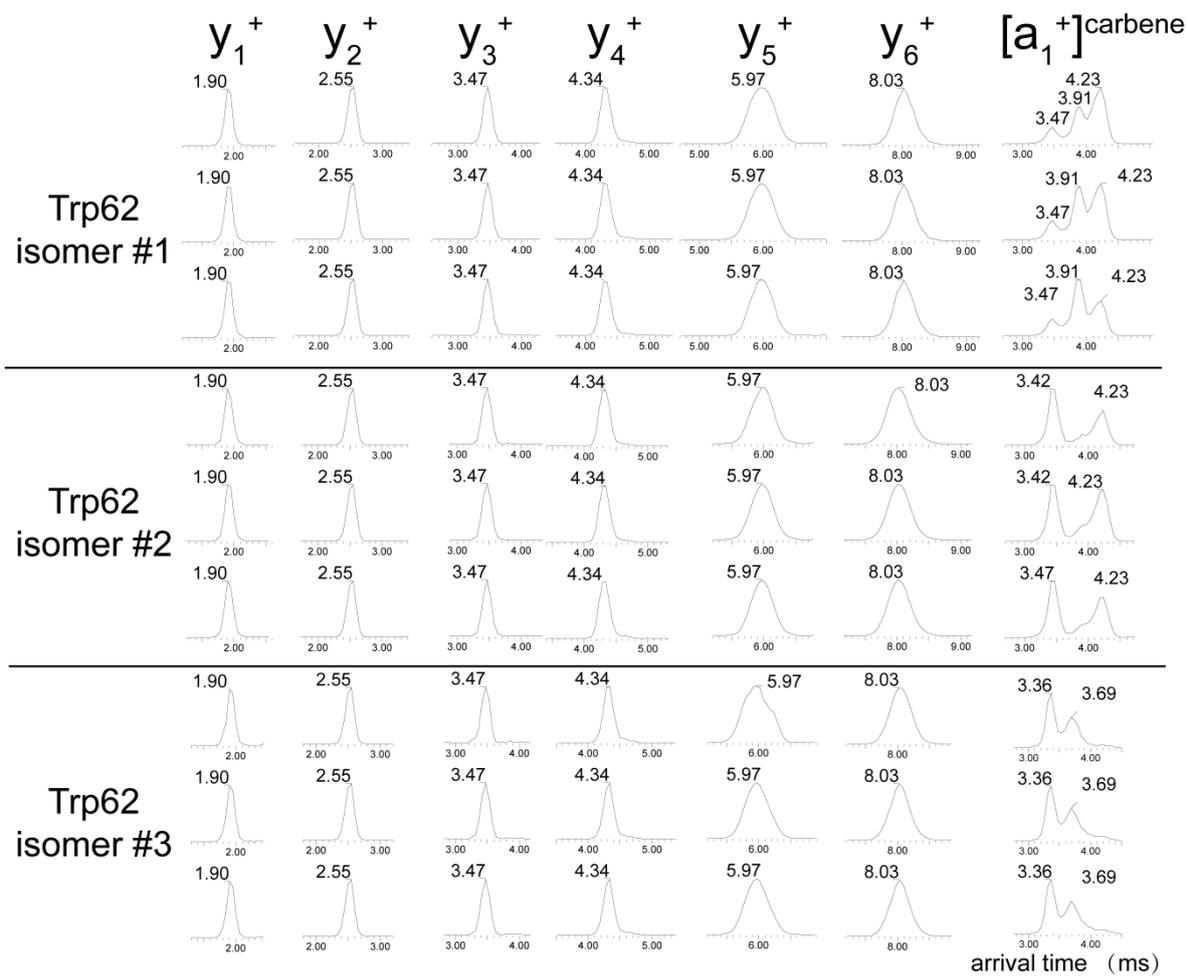


Figure S6. Correlation of the label ratio of tryptophan residue in HEWL with solvent accessible surface area (SASA). The software GETAREA (<http://curie.utmb.edu/getarea.html>) with ligand-free HEWL PDB file 1DPX was used to calculate the SASA value for individual residue.

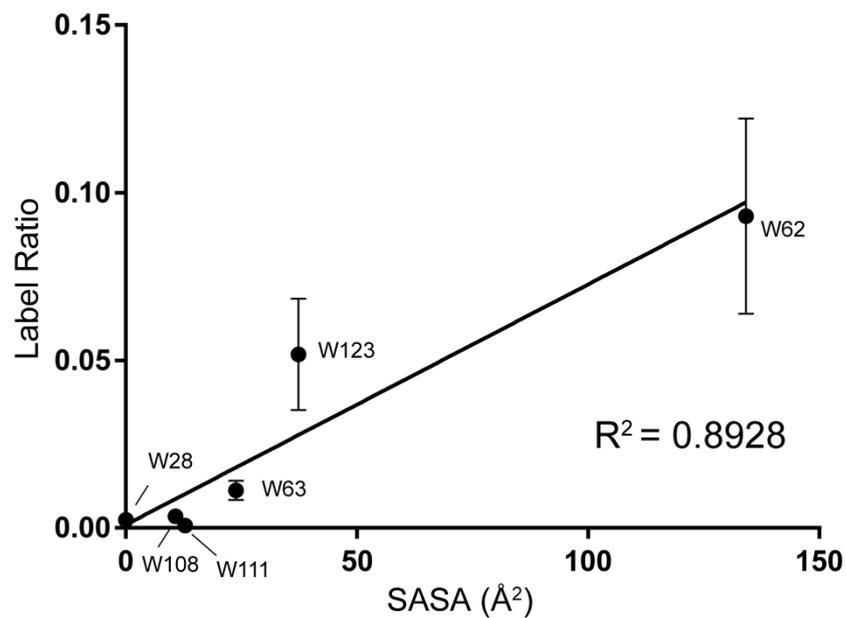


Figure S7. Pathway that leads to the generation of lysine low mass ions. The nitrogen atom in ϵ -NH₂

is kept in the lysine low mass ion at m/z 84.0813.

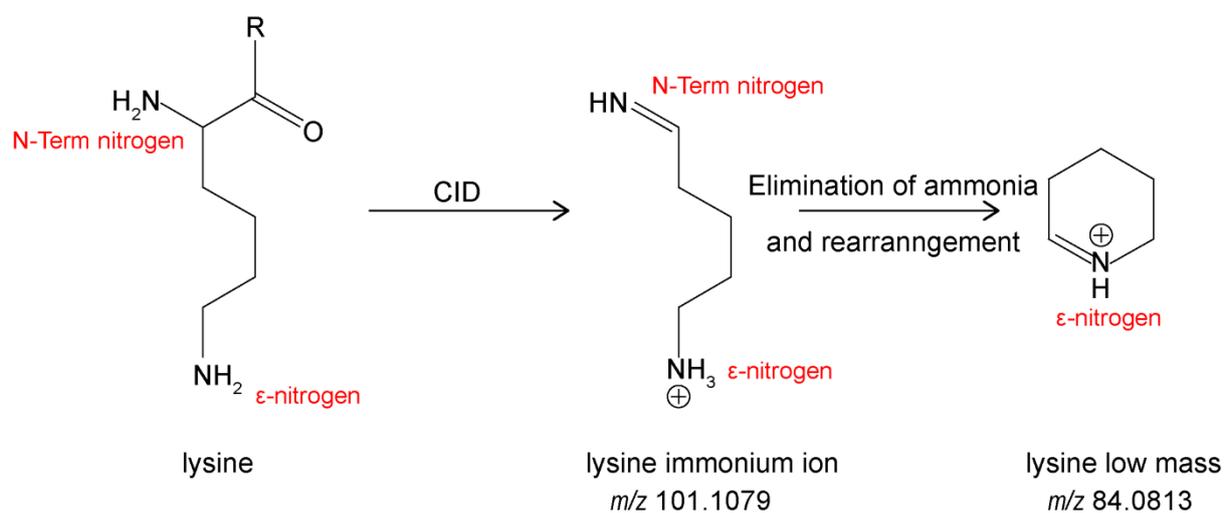
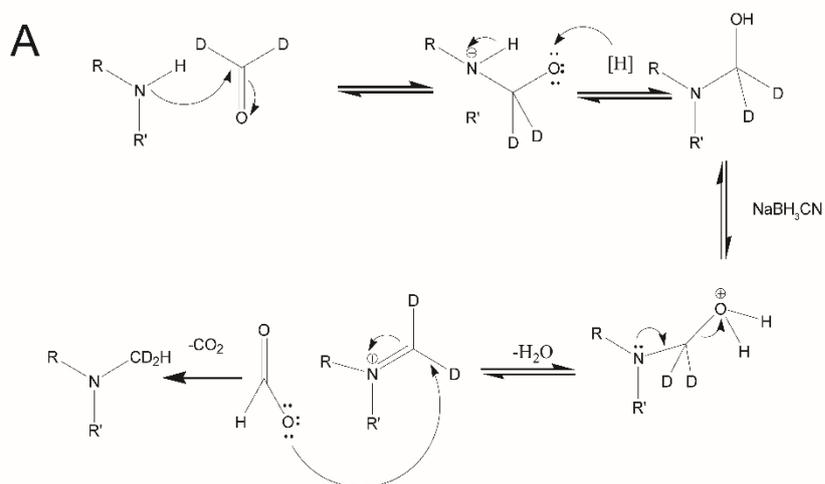


Figure S8. The reaction mechanism of reductive methylation for a primary amine group. (A) The mechanism of Eschweiler-Clarke reaction. The amine group is methylated using excess sodium cyanoborohydride and deuterated formaldehyde. (B) The ranking order of amine nucleophilicity.



B ranking order of amine nucleophilicity

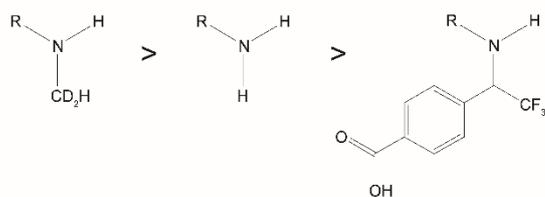


Figure S9. Assigning Lys16 isomers by charge state and the presence of carbene diagnostic ion. (A)

The extracted ion chromatogram (XIC) of carbene-labelled $^{16}\text{KLVFFA}^{21}$ ($z = 1+$). (B) The XIC of

doubly-charged carbene-labeled $^{16}\text{KLVFFA}^{21}$. Due to the strong electrophilic property of the carbene

group, carbene-labeled amine group shows reduced tendency to attract proton, which explains the

absence of doubly charged isomer #2 and isomer #3. (C) The XIC of the dissociated carbene group

during collision induced dissociation (CID).

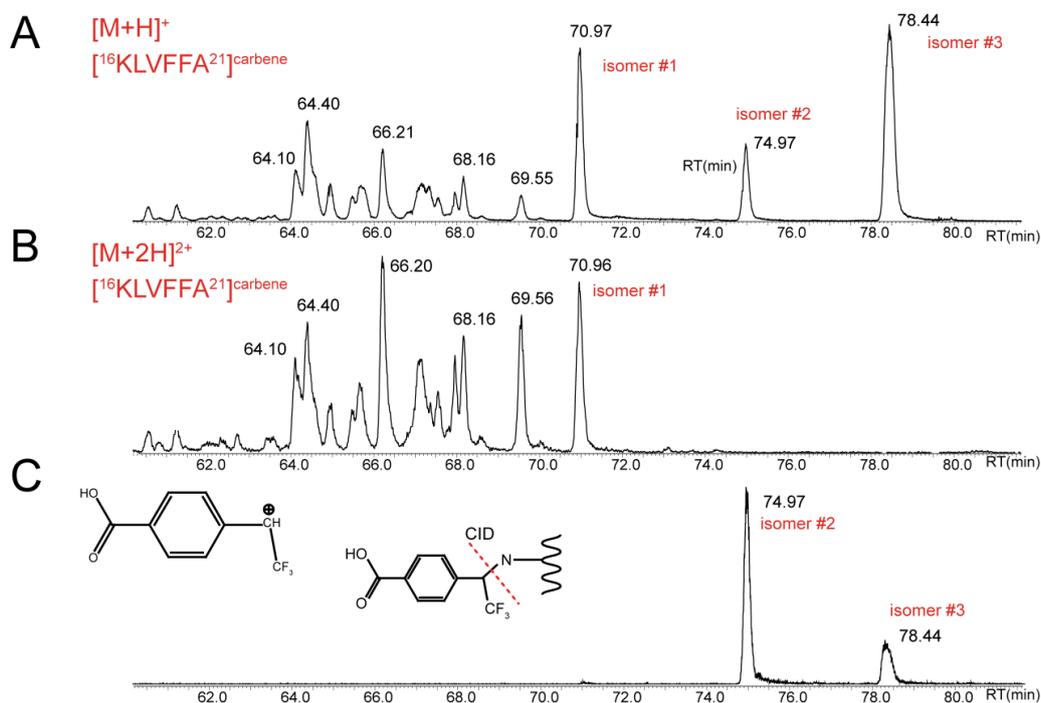


Figure S10. Turbidity of $^{16}\text{KLVFFA}^{21}$ with different incubation duration. The precipitation of $^{16}\text{KLVFFA}^{21}$ contributes to the increased turbidity. In this study, an aggregation model that mimics an early-stage onset of $^{16}\text{KLVFFA}^{21}$ aggregation was used, which avoids extensive precipitation due to a long incubation period.

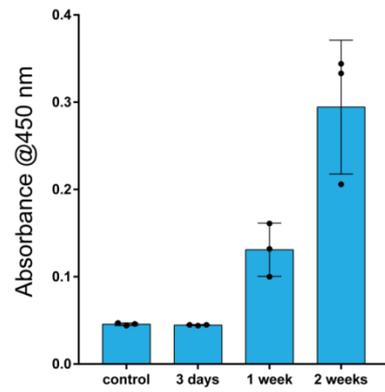


Figure S11. The structure of A β (1-42) fibril. Color scheme: blue = ϵ -NH₂ of Lys16, yellow = ¹⁶KLVFFA²¹ except for ϵ -NH₂ of Lys16, green = the other residues in A β . The structure is constructed based on the PDB file 2MXU.

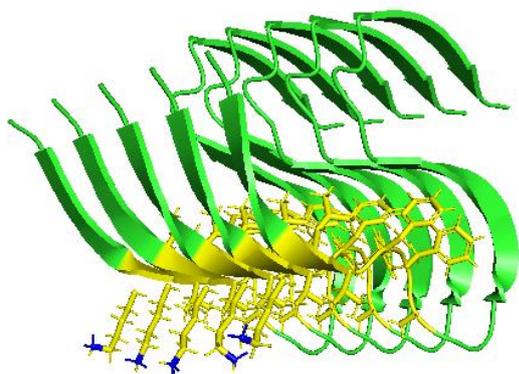


Figure S12. Peptide-level footprinting of ERR α and its three ligands. (A) Label ratio of ERR α in the absence and presence of three ligands. Error bars are \pm s.d. and significant differences (one-way ANOVA, $p < 0.05$, $n = 4$) are highlighted with asterisks. (B) Model of ERR α (based on PDB 2PJL) in a compound 1a-bound state. Color scheme: cyan = compound 1a, red = significant masked by ligands, green = no difference induced by ligand binding, wheat = areas with no peptide coverage or negligible peptide label ratio (less than 0.001).

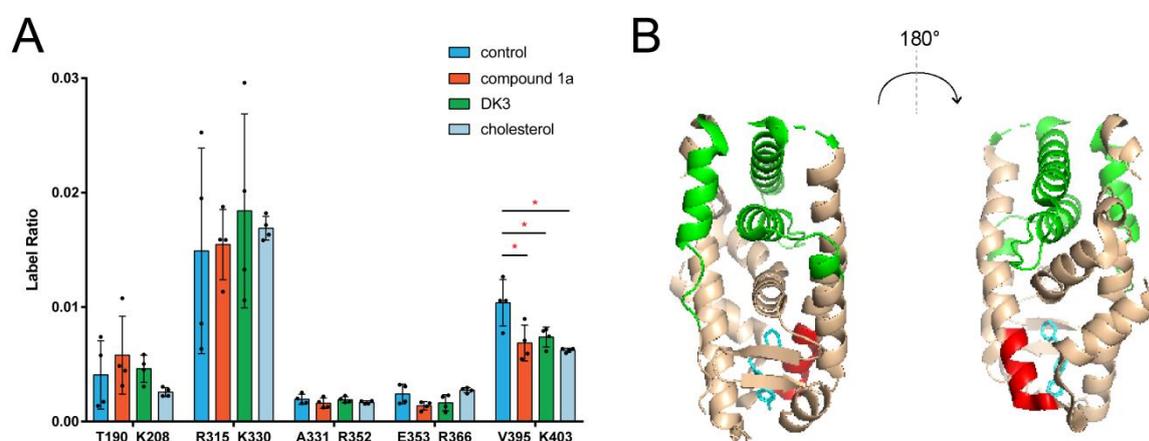


Figure S13. XICs of the fragment ions from $^{395}\text{VLAHFYGVK}^{403}$ reveal the presence of three Phe399 isomers. (A) The XIC of the precursor ion corresponding to the carbene-labeled $^{395}\text{VLAHFYGVK}^{403}$. (B, C) The XIC of y_4^+ and y_5^+ ion of the carbene-labeled $^{395}\text{VLAHFYGVK}^{403}$.

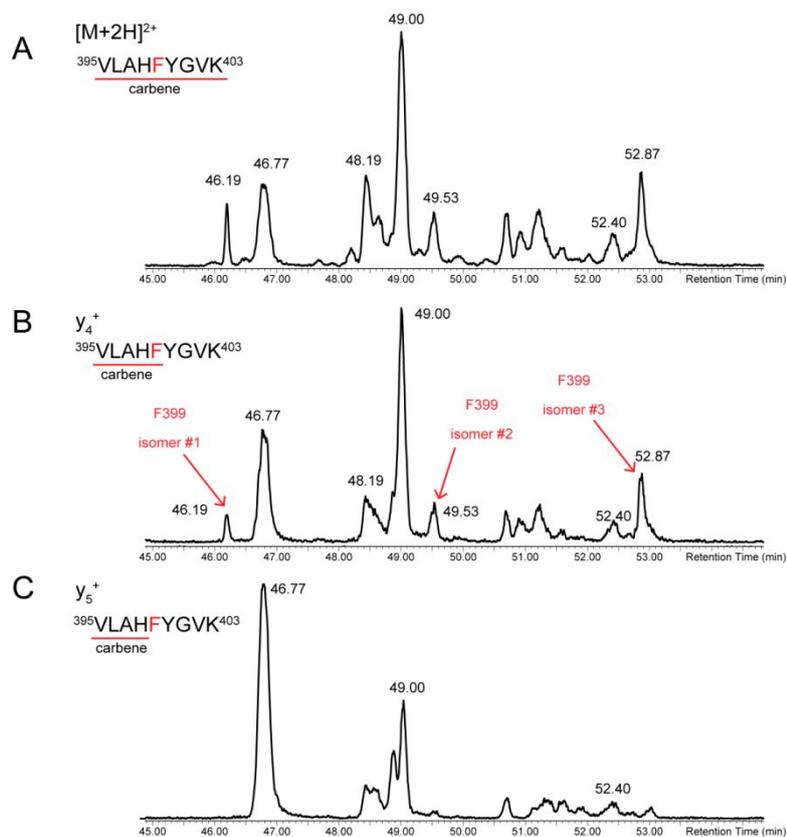


Figure S14. MS/MS spectra of the three Phe399 sub-residue isomers.

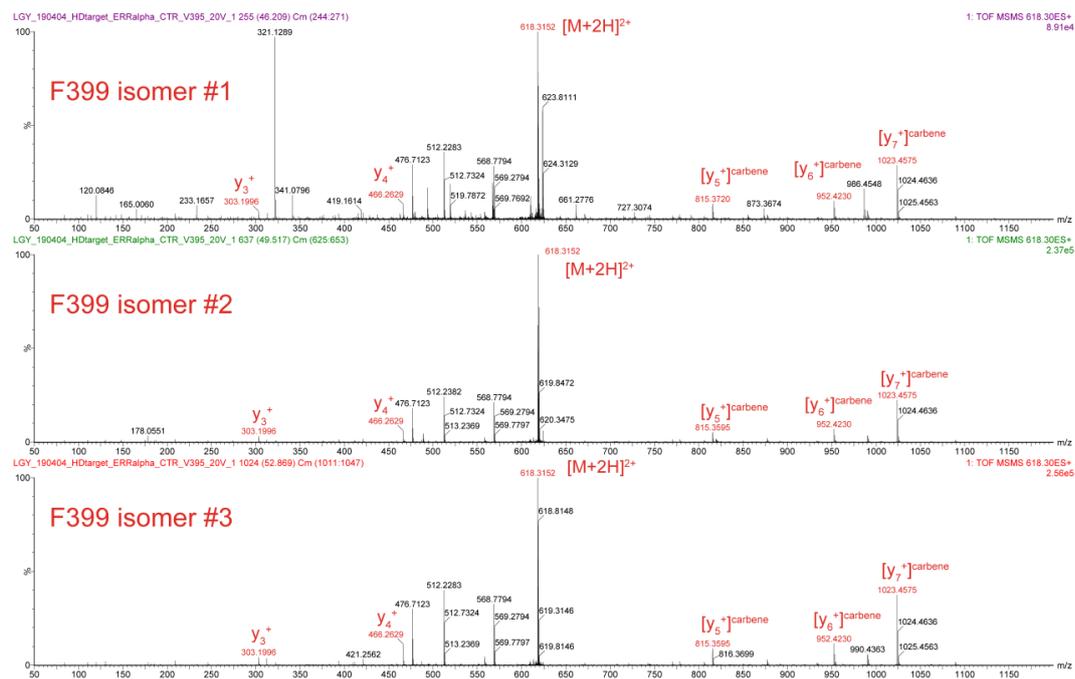


Figure S15. The extracted arrival time distributions of the fragment ions of three Phe399 isomers. The measurement was performed in triplicate.

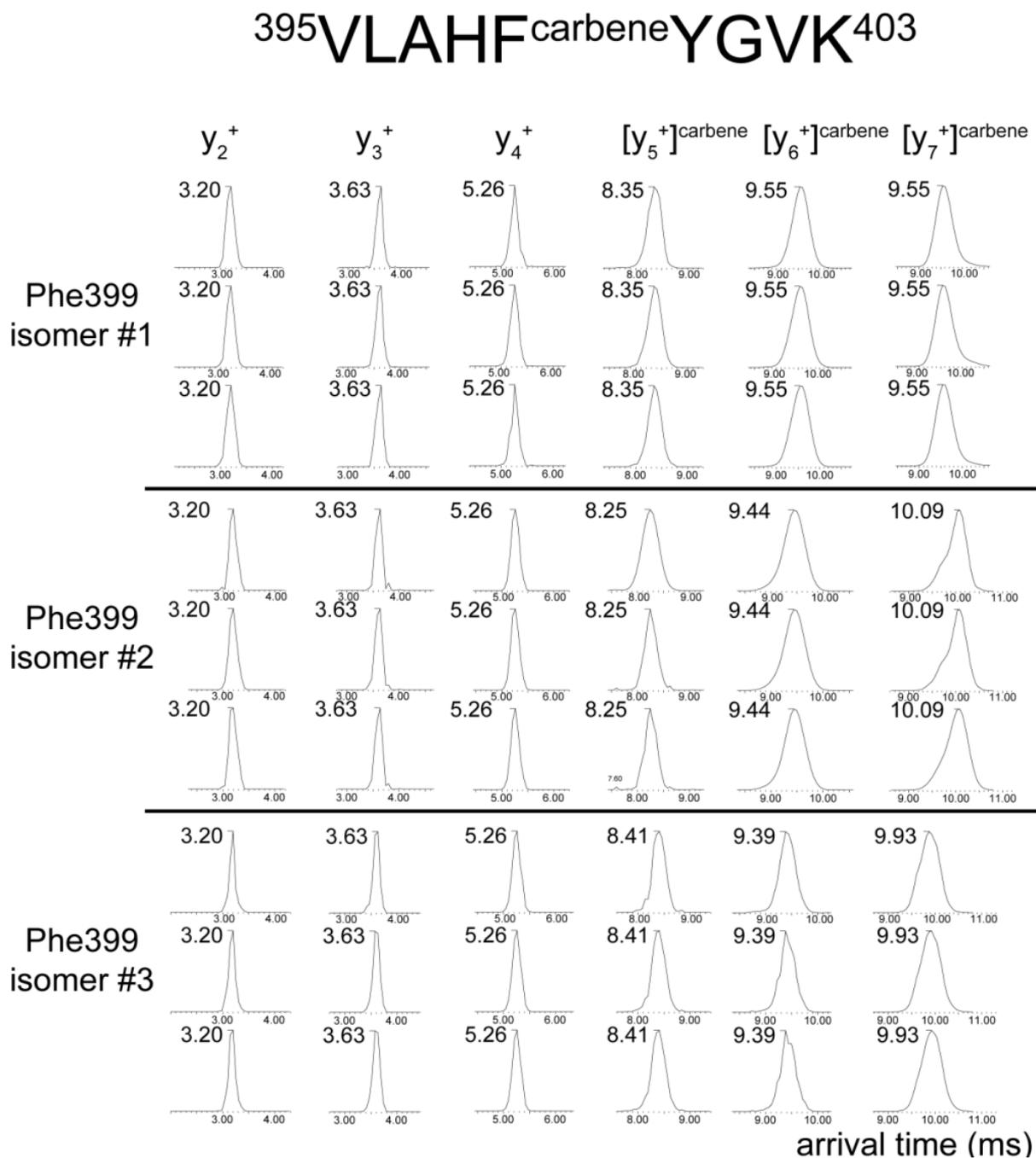


Figure S16. Illustration of the carbene-labeled Phe isomers.

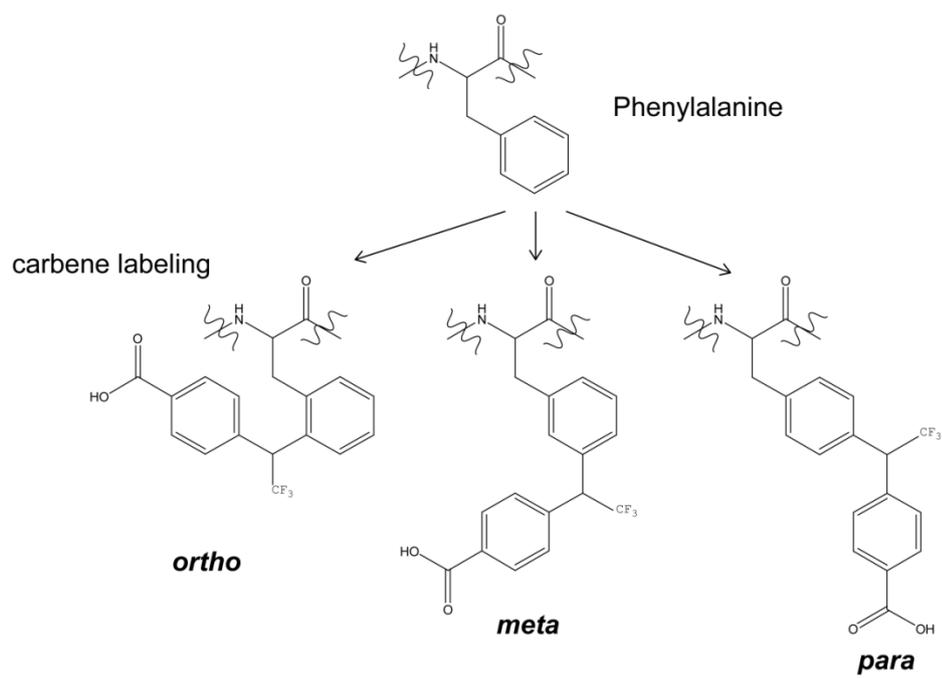


Table S1. Arrival time of the fragment ions of three Trp62-containing isomers (according to **Figure S5**) was compared between isomeric peptides (n=3, arrival time is shown as mean \pm s.d. (standard deviation)). Two-tailed student's t-test was used to determine whether the differences of arrival time of fragment ions produced from different isomers were significant. \checkmark , significant arrival time shift (p value < 0.05); \times , no shift.



ions arrival time (ms) isomers	y_1^+	y_2^+	y_3^+	y_4^+	y_5^+	y_6^+	$[\text{a}_1^+]^{\text{carbene}}$
isomer #1	1.90 \pm 0	2.55 \pm 0	3.47 \pm 0	4.34 \pm 0	5.97 \pm 0	8.03 \pm 0	3.47 \pm 0 3.91 \pm 0 4.23 \pm 0
isomer #2	1.90 \pm 0	2.55 \pm 0	3.47 \pm 0	4.34 \pm 0	5.97 \pm 0	8.03 \pm 0	3.42 \pm 0 4.23 \pm 0
isomer #3	1.90 \pm 0	2.55 \pm 0	3.47 \pm 0	4.34 \pm 0	5.97 \pm 0	8.03 \pm 0	3.36 \pm 0 3.69 \pm 0
significant difference	<i>P</i> value = 1						<i>P</i> value <<0.05
	#1 & #2 \times	#1 & #2 \checkmark					
	#1 & #2 \times	#1 & #2 \checkmark					
	#2 & #3 \times	#2 & #3 \checkmark					

Table S2. The sub-residue label ratios of Trp62 isomers with or without NAG4 incubation. The results were analyzed by two-way ANOVA followed by Fisher's least significant difference (LSD) test.

Two-way ANOVA	alpha = 0.05	
Source of Variation	P value	Significant?
Interaction	0.001	Yes
with or without NAG4	<0.0001	Yes
Sub-residue isomers	0.0002	Yes
Fisher's least significant difference (LSD) test	alpha = 0.05	
	Individual P Value	Significant?
control - NAG4		
isomer #1	0.0022	Yes
isomer #2	0.0174	Yes
isomer #3	<0.0001	Yes

Table S3. Arrival time of the fragment ions of three Phe399-containing peptide isomers (according to **Figure S16**) was compared (n=3, arrival time is shown as mean \pm s.d.). Two-tailed student's t-test was used to determine whether the differences of arrival time of fragment ions produced from different isomers were significant. \checkmark , significant arrival time shift (p value < 0.05); \times , no shift.



ions arrival time (ms) isomers	y_2^+	y_3^+	y_4^+	$[y_5^+]^{\text{carbene}}$	$[y_5^+]^{\text{carbene}}$	$[y_6^+]^{\text{carbene}}$
isomer #1	3.20 \pm 0	3.63 \pm 0	5.26 \pm 0	8.35 \pm 0	9.55 \pm 0	9.55 \pm 0
isomer #2	3.20 \pm 0	3.63 \pm 0	5.26 \pm 0	8.25 \pm 0	9.44 \pm 0	10.09 \pm 0
isomer #3	3.20 \pm 0	3.63 \pm 0	5.26 \pm 0	8.41 \pm 0	9.39 \pm 0	9.93 \pm 0
significant difference	<i>P</i> value = 1			<i>P</i> value \ll 0.05		
	#1 & #2 \times	#1 & #2 \times	#1 & #2 \times	#1 & #2 \checkmark	#1 & #2 \checkmark	#1 & #2 \checkmark
	#1 & #2 \times	#1 & #2 \times	#1 & #2 \times	#1 & #2 \checkmark	#1 & #2 \checkmark	#1 & #2 \checkmark
	#2 & #3 \times	#2 & #3 \times	#2 & #3 \times	#2 & #3 \checkmark	#2 & #3 \checkmark	#2 & #3 \checkmark

Table S4. Measured CCS values of carbene-labeled fragment ions generated from $^{395}\text{VLAHF}^{\text{carbene}}\text{YGVK}^{403}$ and VLAHFVK ($n=3$, arrival time is shown as mean \pm s.d. (standard deviation)) were matched with the DeepCCS-predicted values. The assignments of meta/ortho/para positional isomers were made based on the fragment ions of the peptides. The only irregularity is the $[\text{y}_7^+]^{\text{carbene}}$ -based assignment (not shown). This can be explained by the relatively large ion size and consequently decreased accuracy, since mostly ions of smaller size ($\Omega_{\text{He}} < 300 \text{ \AA}^2$) are included in the training set of DeepCCS⁵. We performed CCS prediction and matching on another shorter version of this peptide, and found the retention order is the same as the full-length one. Moreover, the relative errors between the predicted and measured CCS values for target ions are kept within 1%.

	Phe399 isomer	Ω_{He} measured (\AA^2)	label site	Ω_{He} predicted (\AA^2)	mean relative error
	---	161.10 ± 0	-----	161.90	0.5%
	---	170.74 ± 0	-----	173.82	1.8%
	---	211.44 ± 0	-----	212.27	0.4%
ERR α peptide $^{395}\text{VLAHF}^{\text{carbene}}\text{YGVK}^{403}$	$[\text{y}_5^+]^{\text{carbene}}$	#1	<i>meta</i>	278.78	0.1%
		#2	<i>ortho</i>	276.30	-0.1%
		#3	<i>para</i>	279.61	-0.1%
	$[\text{y}_6^+]^{\text{carbene}}$	#1	<i>meta</i>	305.66	0%
		#2	<i>ortho</i>	301.65	-0.1%
		#3	<i>para</i>	299.46	-1.0%
	Phe isomer	Ω_{He} measured (\AA^2)	label site	Ω_{He} predicted (\AA^2)	relative error
	---	161.10 ± 0	-----	161.90	0.5%
peptide $\text{VLAHF}^{\text{carbene}}\text{VK}$	$[\text{y}_3^+]^{\text{carbene}}$	#1	<i>meta</i>	231.64	-0.3%
		#2	<i>ortho</i>	230.19	0.2%
		#3	<i>para</i>	235.62	0.8%
	$[\text{y}_4^+]^{\text{carbene}}$	#1	<i>meta</i>	257.83	0.3%
		#2	<i>ortho</i>	255.63	-0.1%
		#3	<i>para</i>	261.58	0.2%
	$[\text{y}_5^+]^{\text{carbene}}$	#1	<i>meta</i>	269.54	0.2%
		#2	<i>ortho</i>	267.51	-0.1%
		#3	<i>para</i>	274.08	0.6%

REFERENCES

1. Manzi, L.; Barrow, A. S.; Scott, D.; Layfield, R.; Wright, T. G.; Moses, J. E.; Oldham, N. J., Carbene Footprinting Accurately Maps Binding Sites in Protein–Ligand and Protein–Protein Interactions. *Nat. Commun.* **2016**, *7*, 13288.
2. Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C., High-Resolution Mapping of Carbene-Based Protein Footprints. *Anal. Chem.* **2012**, *84*, 4411.
3. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82*, 9557-9565.
4. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., Ion Mobility–Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3*, 1139.
5. Plante, P.-L.; Francovic-Fontaine, E. I.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F. o.; Marchand, M.; Corbeil, J., Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: Deepccs. *Anal. Chem.* **2019**, *91*, 5191-5199.