Supplementary Information (SI) for New Journal of Chemistry.

ELECTRONIC SUPPLEMENTARY MATERIAL

The reaction of glyoxylic acid with lysine chemically protected on the α-amine group studied by mass spectrometry

Katarzyna Wrobel[1], Alma Rosa Corrales Escobosa[1], Francisco Javier Acevedo-Aguilar[1], Israel Enciso Donis[1], Kazimierz Wrobel[1]*

[a] Chemistry Department, University of Guanajuato, L. de Retana 5, Guanajuato, 36000, Mexico.

* Corresponding author. E-mail: kazimier@ugto.mx

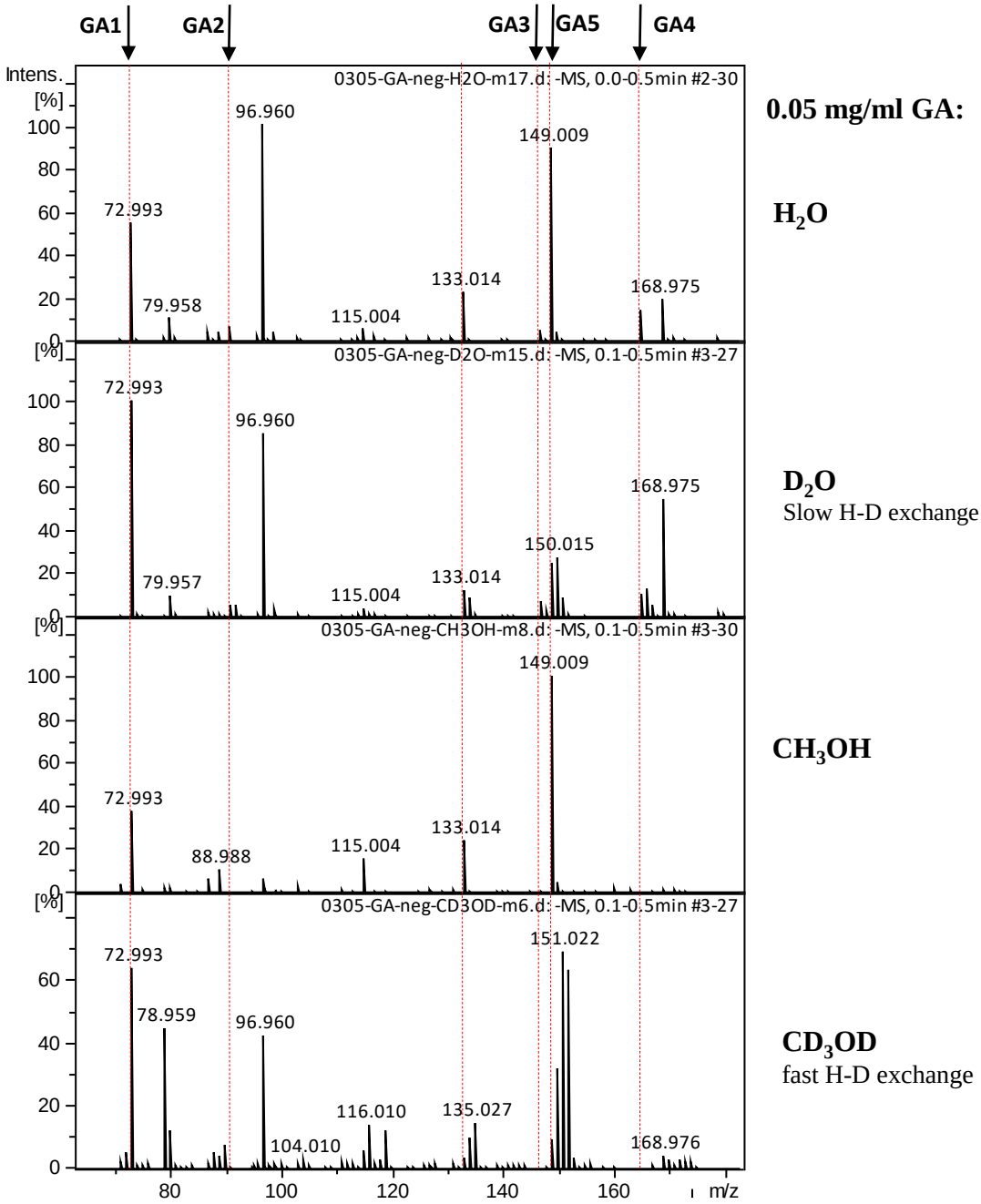
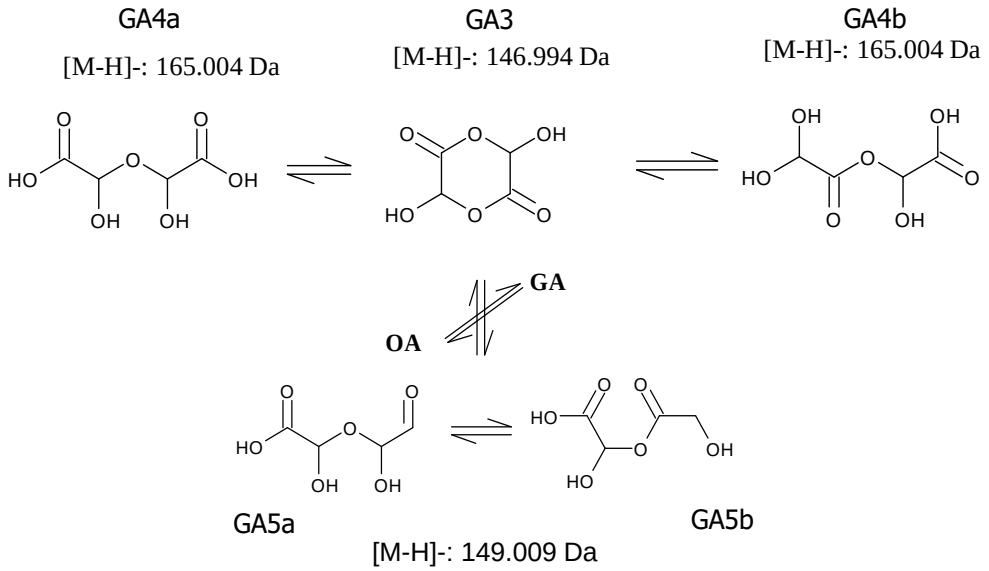


Fig. 1S. ESI(-)-QTOF-MS spectra of GA (0.05 mg/mL) prepared in H_2O, D_2O, CH_3OH, and CD_3OD, respectively. The *m/z* values corresponding to protonated GA1-GA5 are marked with dotted lines; *m/z* 168.976 was assigned as sodium adduct of GA3, *m/z* 133.014 is a GA5 fragment after oxygen loss and *m/z* 115.004 corresponds to further loss of water molecule.



$$3\ C_2H_2O_3 + H_2O \rightarrow C_4H_6O_6 + C_2H_2O_4$$

Fig. 2S. Proposed scheme of reversible conversion of cyclic GA dimer (GA3) into two plausible forms of GA4 (GA4a and GA4b) and their subsequent reduction to GA5a and GA5b, respectively. Both processes occur with simultaneous oxidation of GA to oxalic acid. The equation shows the overall redox process, starting from GA and producing GA5 and oxalic acid.

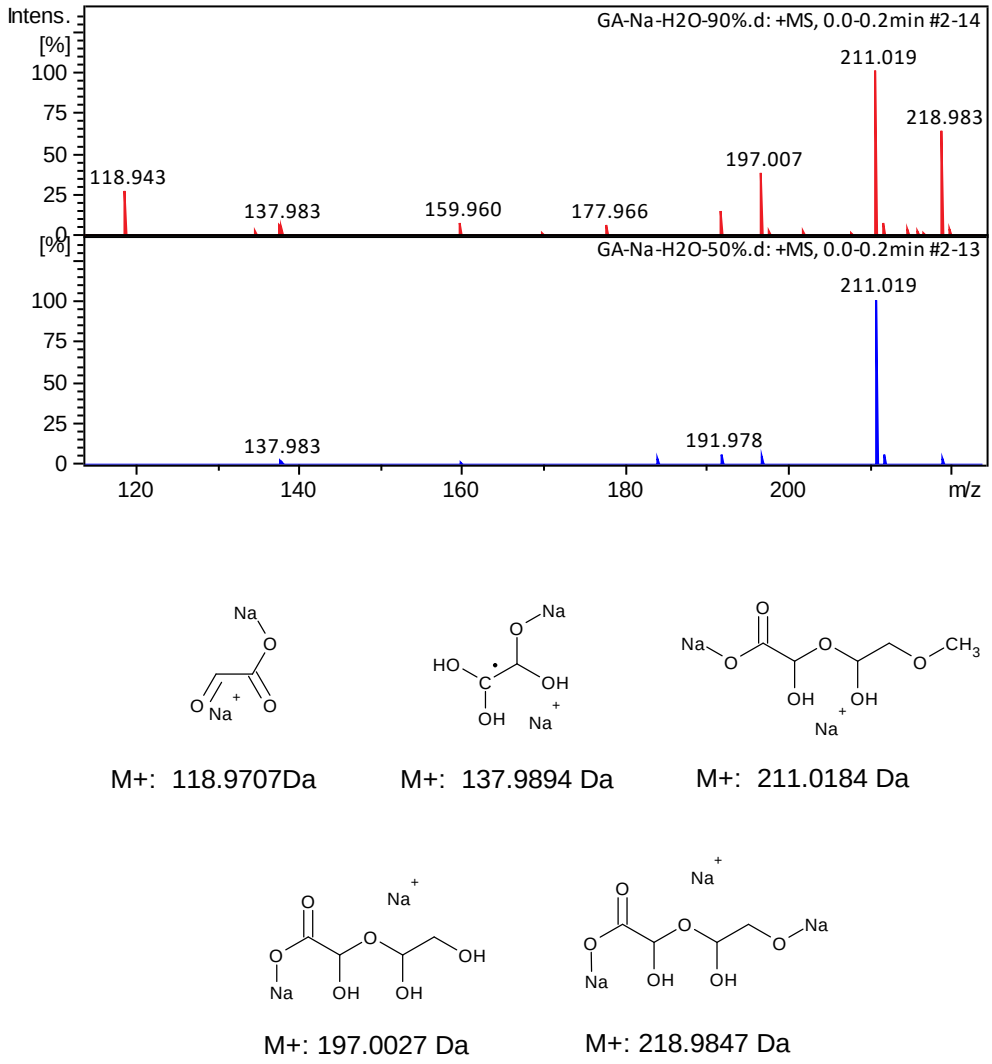


Fig. 3S. ESI(+)-QTOF-MS spectra for glyoxalic acid in water-methanol mixtures corresponding to 90% water (upper) and 50% water (lower). Below, structures assigned to *m/z* values are shown. It is noteworthy that *m/z* 137.989 corresponds to a radical ion.

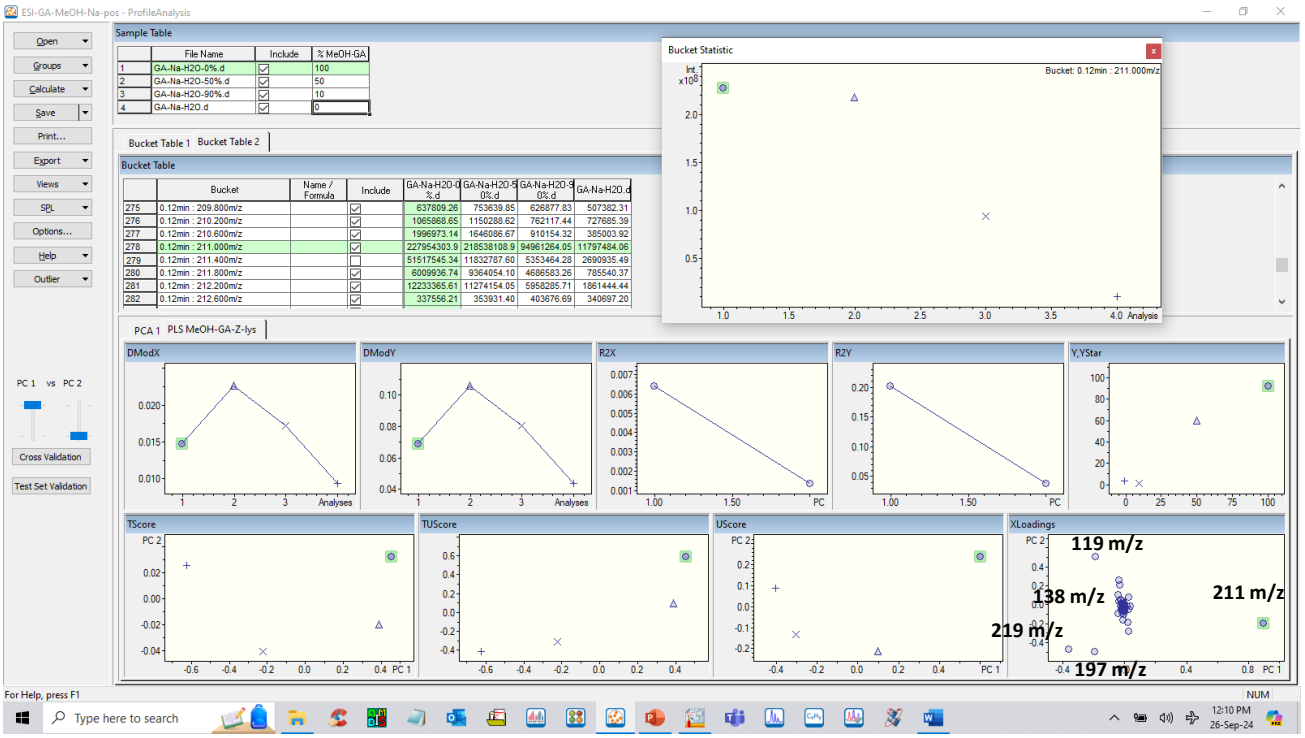


Fig. 4S. Partial least square (PLS) analysis. Based on four ESI(+)-QTOF-MS datasets acquired for GA solutions in 0, 50, 90 and 100% water.

PLS analysis was used to detect explanatory variables (*m/z* values) contributing to spectral changes between different GA solutions due to the water to methanol ratio used in each of them. Additionally, PLS showed that the spectral changes were significant enough to allow prediction of the solvent composition based on the MS spectrum. This was an important finding reinforcing that GA6 and GA5 were preferentially formed in methanol. Noteworthy, GA5 was better detected in ESI(-) and GA6 in ESI(+), but both in the same solutions of GA in water - methanol.

A software ProfileAnalysis from Bruker was used and a PLS print screen from this program is presented. In the upper part, a fragment of the buckets table is located showing the four cases (X- four solutions of GA in different proportions of water-methanol) and their respective spectral data in columns (Y- spectral buckets). A plot on the right shows that the intensity of bucket containing GA6 (*m/z* 211) was high in pure methanol and decreased when the percentage of water in the solution was increasing.

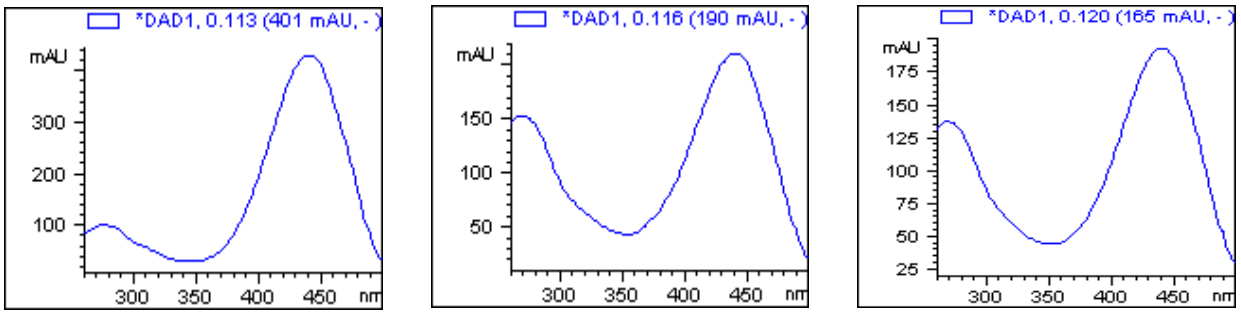
The obtained PLS model is projected below with nine windows displaying the following plots: DModX, DModY, R2X, R2Y, Y,Ystar, TScore, TUScore, UScore and XLoadings. The first four of them evaluate the quality of PLS model; DModX and DModY show the distance of X data (spectral buckets) and Y data (cases or solutions), respectively, from the model indicating potential outilers; no outliers are observed. Then, R2X and R2Y present the residual variance in the X and Y space, respectively, for increasing number of latent variables or principal components used to create the model. As can be observed, the first two components (PC1 and PC2) accounted practically for 100% of total data variability in both spaces. TUScore represents correlation between X (spectral buckets) and Y (cases); the observed deviation from linear plot was ascribed to processing errors. The UScore plot indicates if there are deviations among Y observations when replicates of cases are included; no deviations are observed in this plot because we used only one replicate.

Most importantly, the TScore presents the distribution of cases in the coordinates of PC1 and PC2. In the XLoadings plot, the distribution of variables (buckets) is projected in the same space of PC1, PC2 coordinates, helping to visualize the association of cases with specific variables. When comparing TScore with XLoadings, the bucket *m/z* 211 (corresponding to GA6) is clearly associated with high methanol content in the solution. In Y,Ystar graph, the measured spectral data are plotted against those predicted by the model for each GA solution. A straight diagonal obtained demonstrates the prediction capability of the model.

GA species	Neutral molecular formula (M)	Ion detected	Exact mass	Δ, ppm*
GA1	$C_2H_2O_3$	$[M-H]^-$	72.9931	1.4
		$[M+2Na-H]^+$	118.9710	23
GA2	$C_2H_4O_4$	$[M-H]^-$	91.0037	5.5
		$[M+2Na]^{\cdot+}$	137.9894	43
GA3	$C_4H_4O_6$	$[M-H]^-$	146.9935	4.2
GA4	$C_4H_6O_7$	$[M-H]^-$	165.0041	1.8
GA5	$C_4H_4O_6$	$[M-H]^-$	149.0092	0.7
GA6	$C_5H_9NaO_6$	$[M+Na]^+$	211.0184	1.9
		$[M-CH_2+Na]^+$	197.0027	19
		$[M-CH_3+Na]^+$	218.9847	5.5

*experimental mass error; since spectra were acquired several times, the highest relative difference between experimental and exact *m/z* value is presented in each case.

Table 1S. The list of GA species annotated based on ESI-QTOF-MS data acquired for the GA solutions prepared in water, methanol and their mixtures.



Control (no GA)
100%

GA-H2O
47.5+/-2.3

GA-MeOH
44.9+/-0.9

Fig. 5S. Spectrophotometric assay examining possible formation of free radicals in GA solutions. Absorbance spectra are presented that were obtained for DNMA in water and for GA solution with DNMA addition in water and in methanol, respectively (procedural details given is section 2.2). Absorbance decrease at 440 nm is due to decolorating of DNMA caused by entrapment of free radicals.

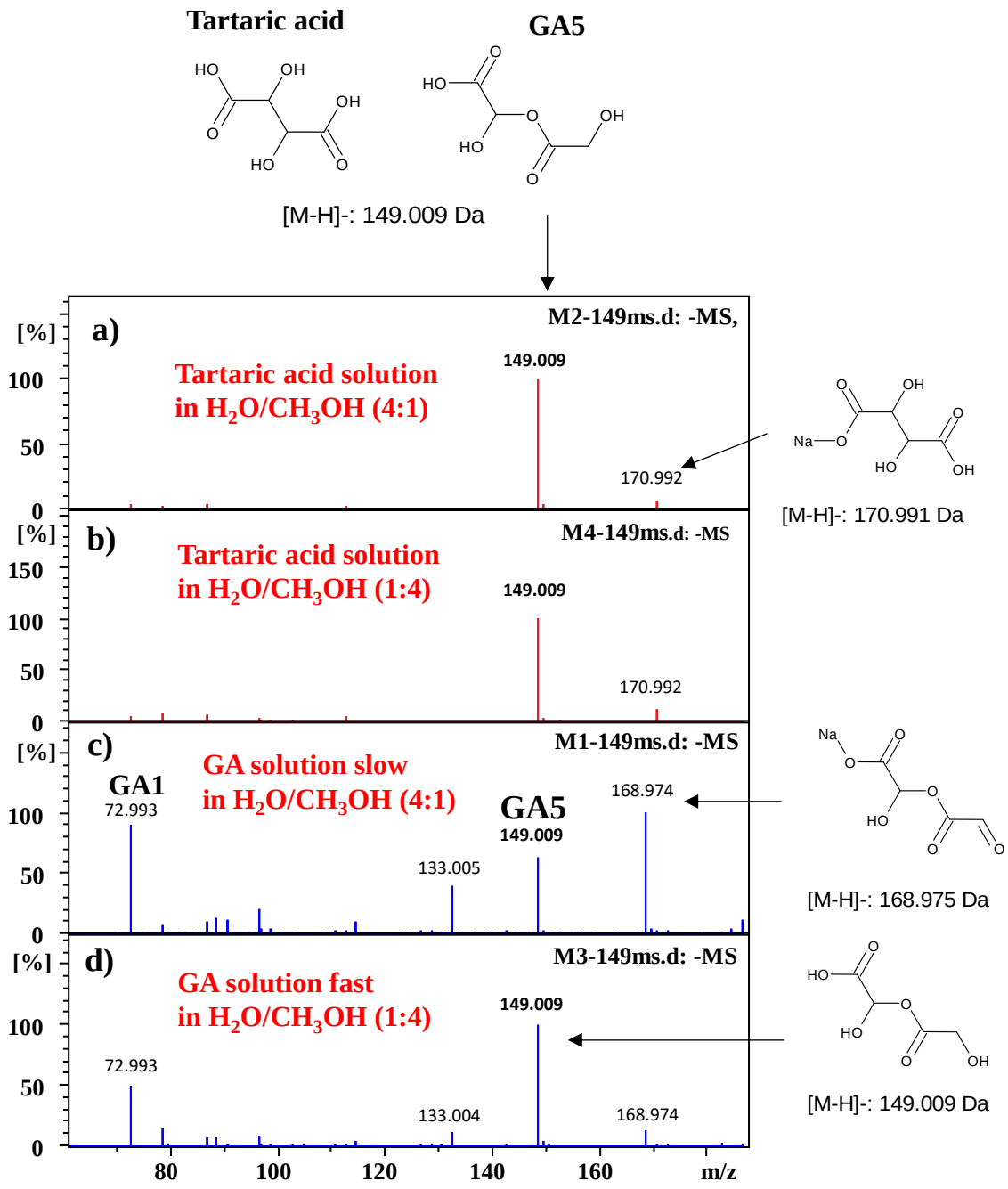


Fig. 6S. ESI(-)-QTOF-MS spectra obtained for tartaric acid standard in a) water - methanol mixtures 4:1 and b) 1:4 ratios and spectra obtained for GA solutions c), d) with the same solvent compositions.

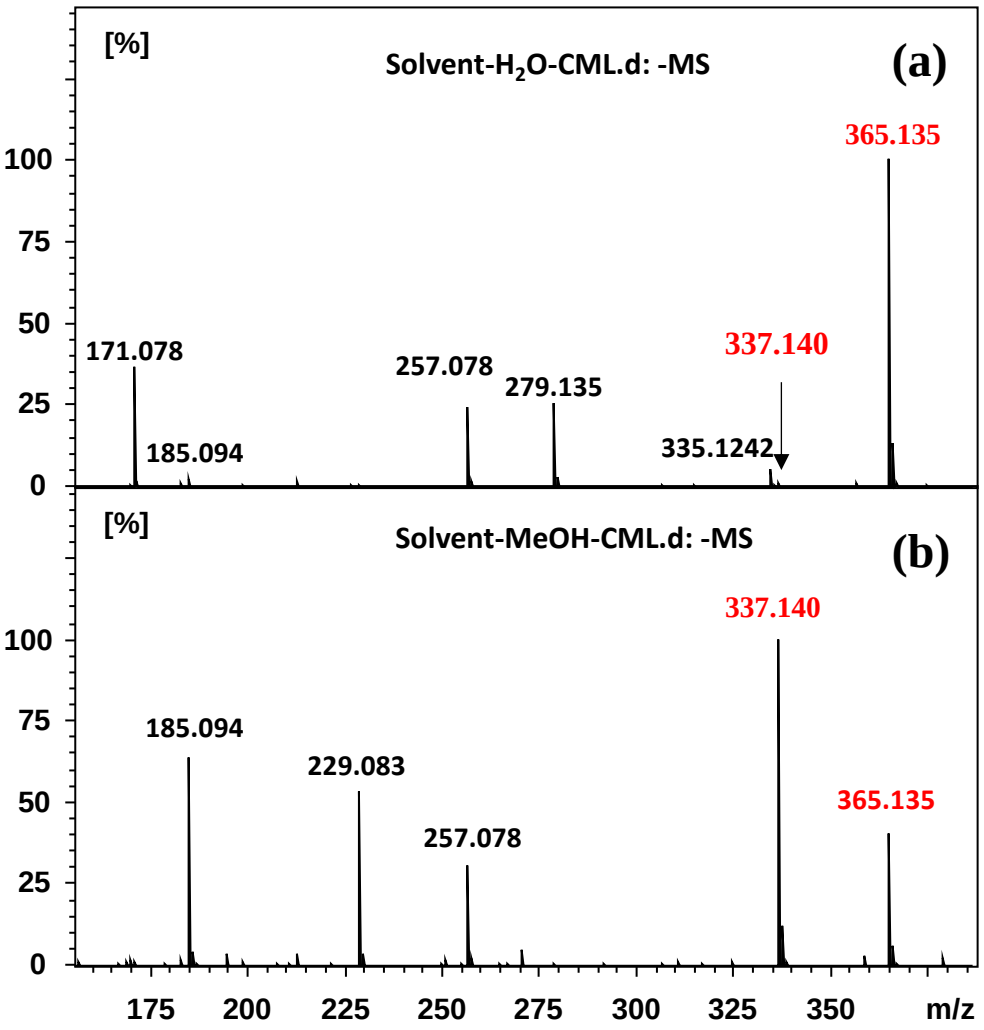


Fig. 7S. ESI(-)-QTOF-MS spectra obtained for GA mixture with Z-Lys in water (a) and in methanol (b). GA to Z-Lys molar ratio 2:1, heating at 60 °C for 72 h. Z-CML ion $[M-H]^-$ observed at *m/z* 337.140 and *m/z* 365.135 is a deprotonated molecule of intermediate **4** (Table 1, Scheme 2 in the main text).

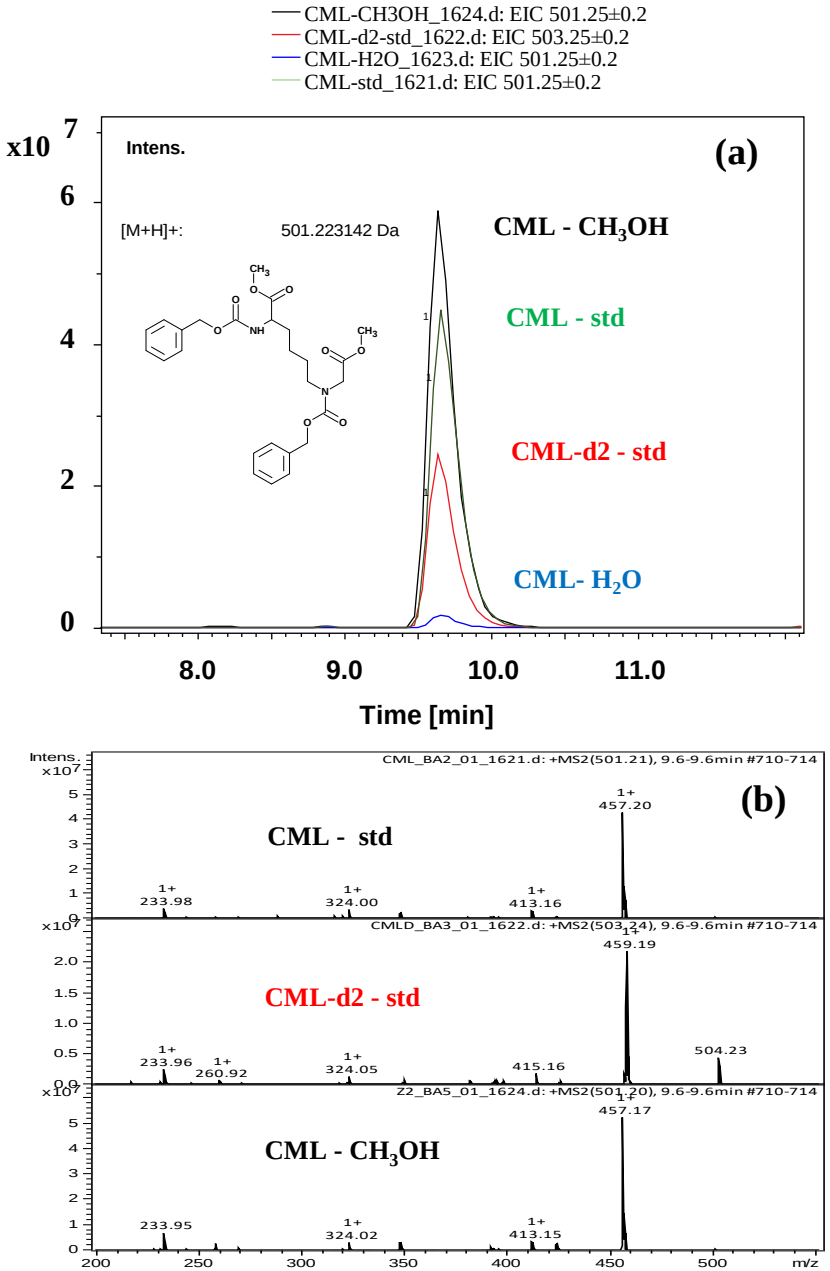


Fig. 8S. **(a)** Extracted ion chromatograms of benzyl chloroformate derivatives of CML standard, deuterated CML standard (CML-d2) and Z-CML synthesized from GA and Z-Lys in water and in methanol (GA to Z-Lys molar ratio 2:1, 60 °C, 72 h, same experiment as in Fig. 6S.).

(b) Fragmentation spectra acquired for the protonated CML derivatives during respective chromatographic runs shown in (a).

As can be observed, the precursor ion *m/z* 503.24 is efficiently fragmented and absent in MS/MS spectra (b). On the other hand, the signal at *m/z* 504.12 is present as a non-fragmented part of isotopic pattern of this same species or trace contamination, or it could be due to traces of CML-d3 in a commercial CML-d2.

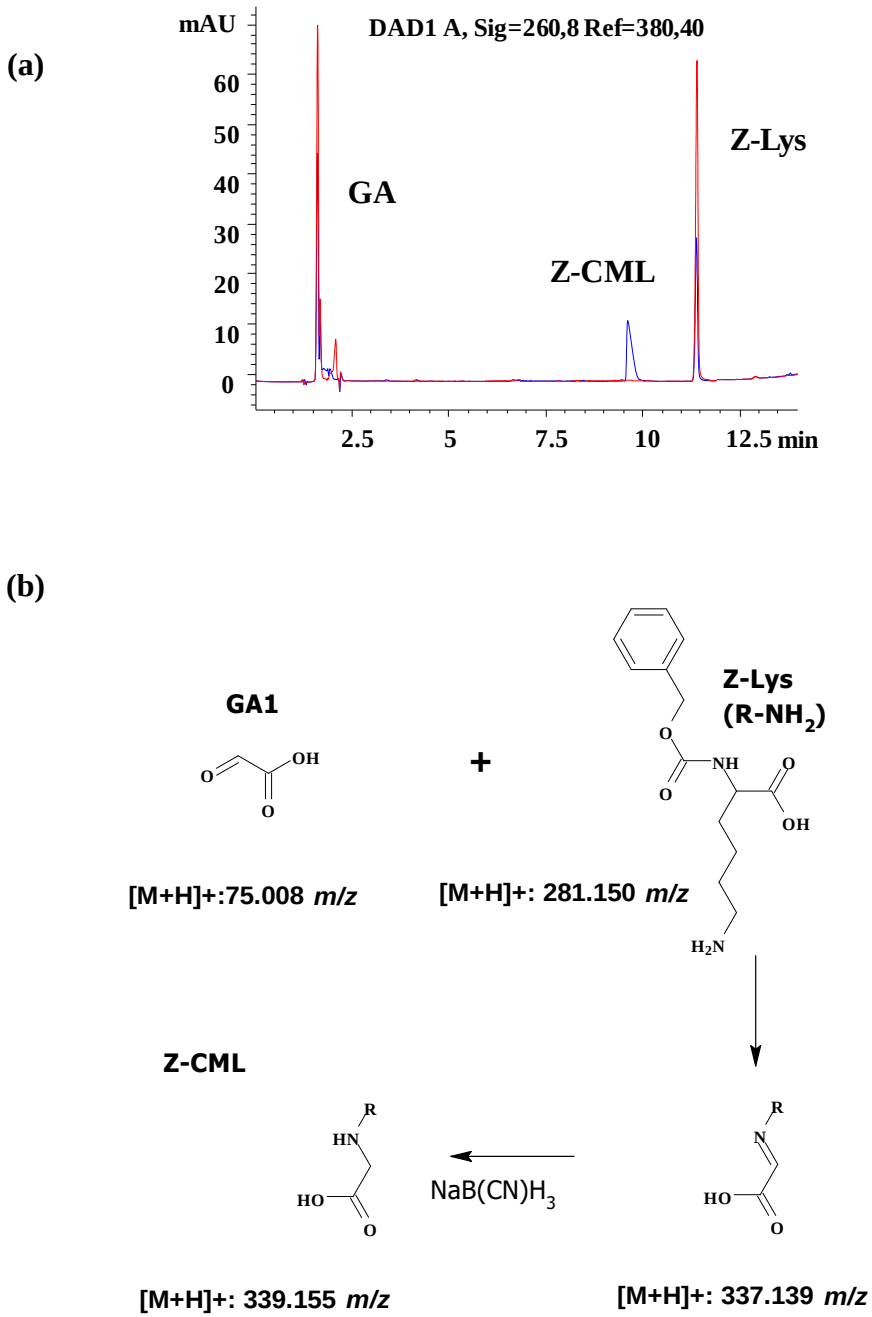


Fig. 9S. **(a)** Typical HPLC-DAD chromatogram for the Z-Lys + GA reaction mix which was incubated at 25 °C for 15 min (red) and a chromatogram obtained for the same mix after addition of sodium cyanoborohydride (blue). The separation/detection conditions are given in section 2.4. **(b)** Mechanism of Z-CML formation from Z-Lys and GA in the presence of a strong reductant.

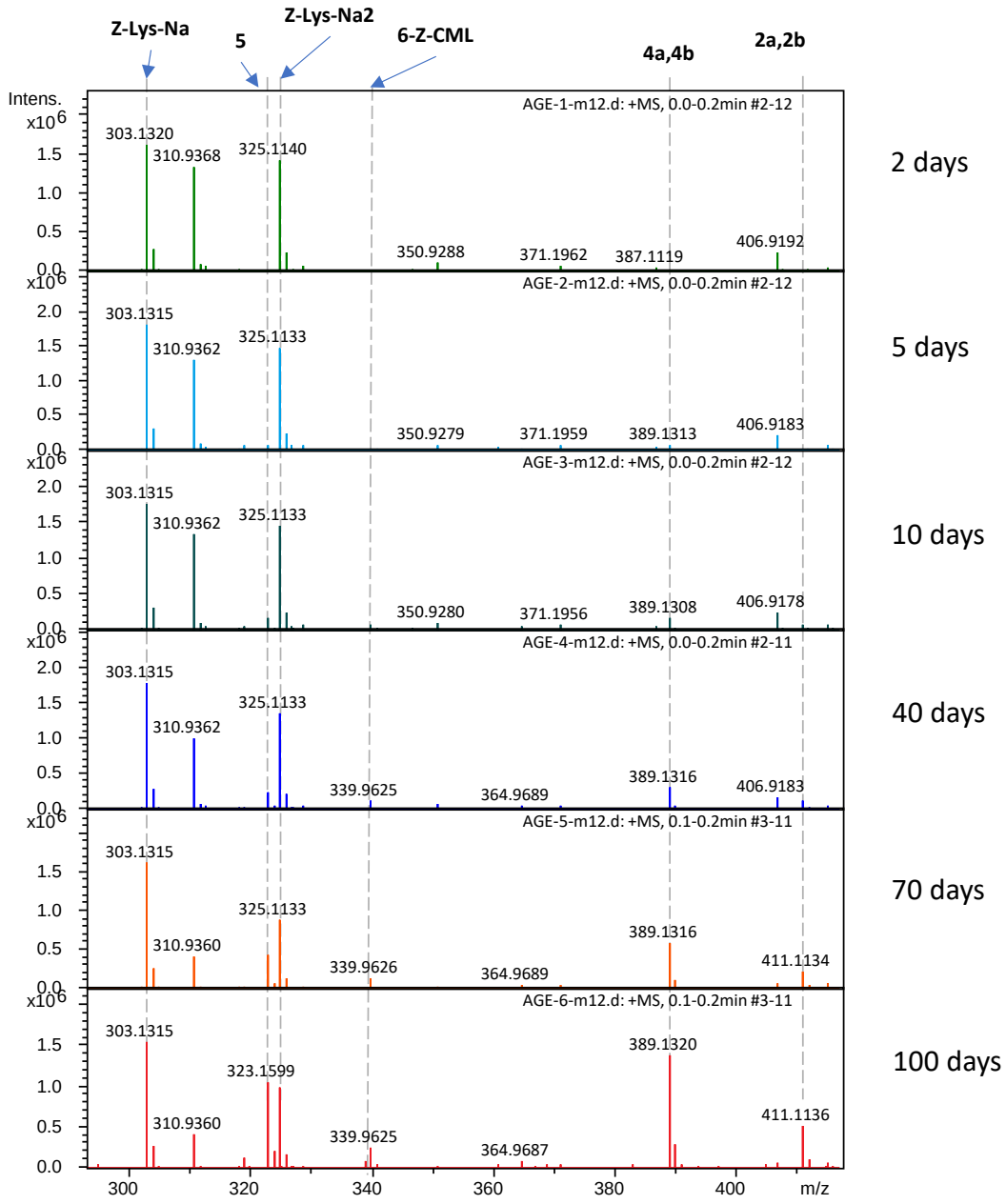
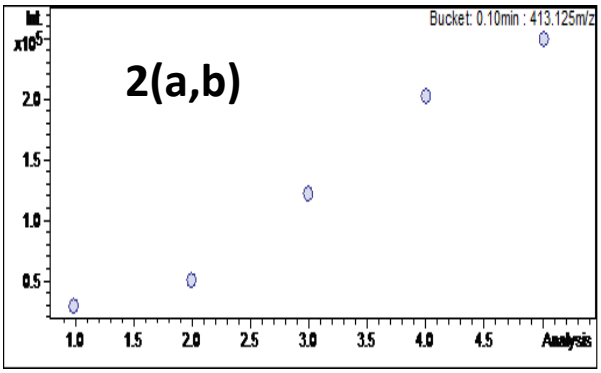
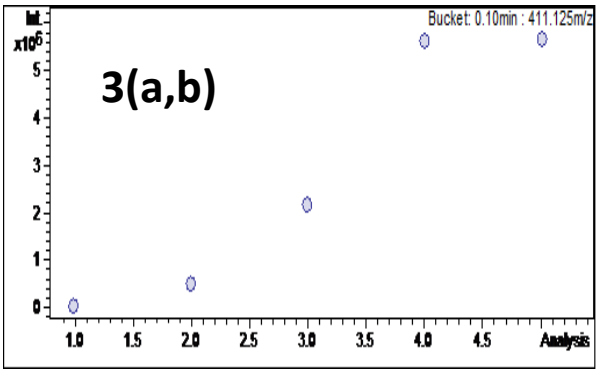
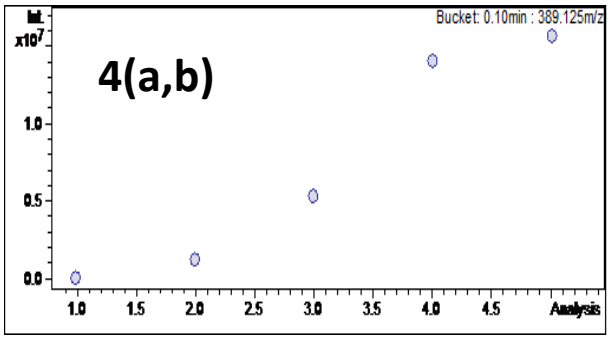
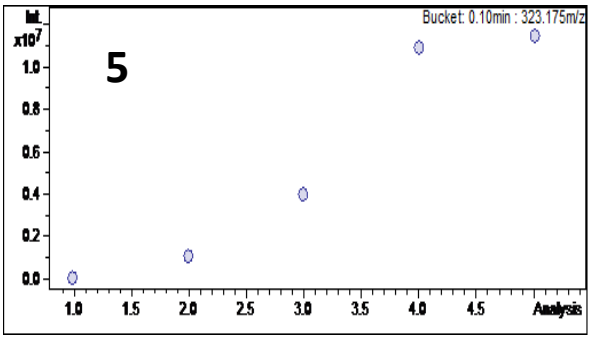


Fig. 10S. ESI(+) - QTOF-MS **s**pectra acquired for the reaction mixture at different times during 100-day experiment. As can be observed, individual ions presented very different abundances; non-reacted Z-Lys produced very strong signals corresponding to different adducts, while signals of intermediates are hardly observed. For clarity of presentation, Fig. 2 in the main text shows the MS spectrum of the reaction mixture acquired after 100 days, reducing the number of cationized adducts by salt elimination via SPE. Changes of the abundance over the reaction time for the substrate, intermediates and Z-CML are presented in Figure 11S. The abundances obtained for different ions at different times of reaction were used to construct PCA model (Fig. 1 in the main text).









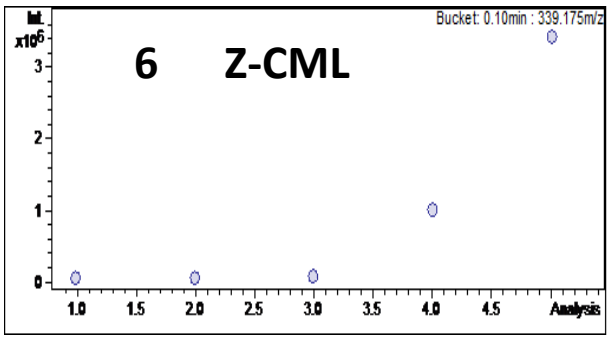
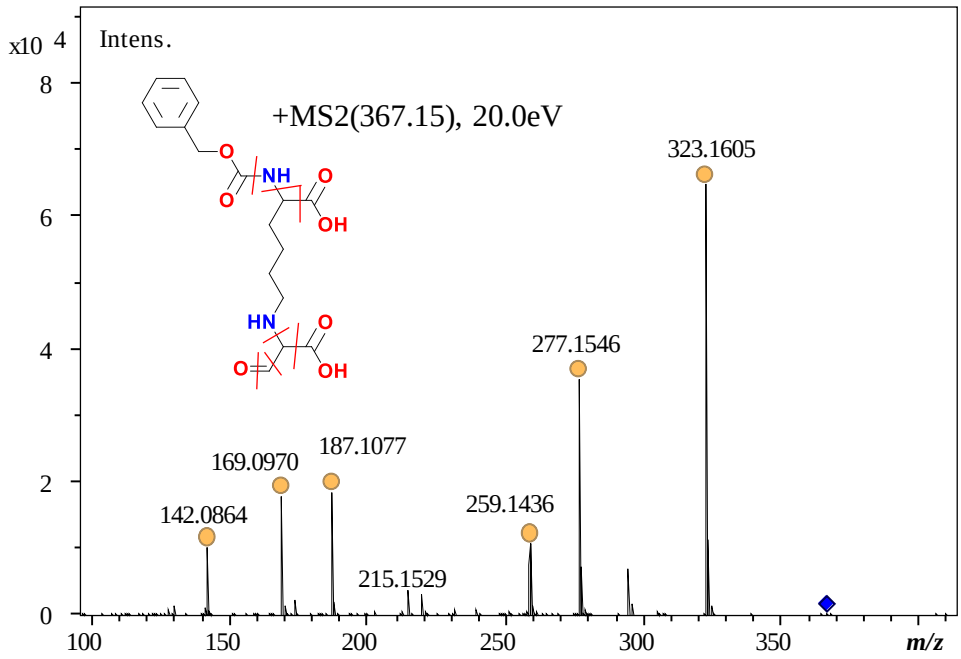
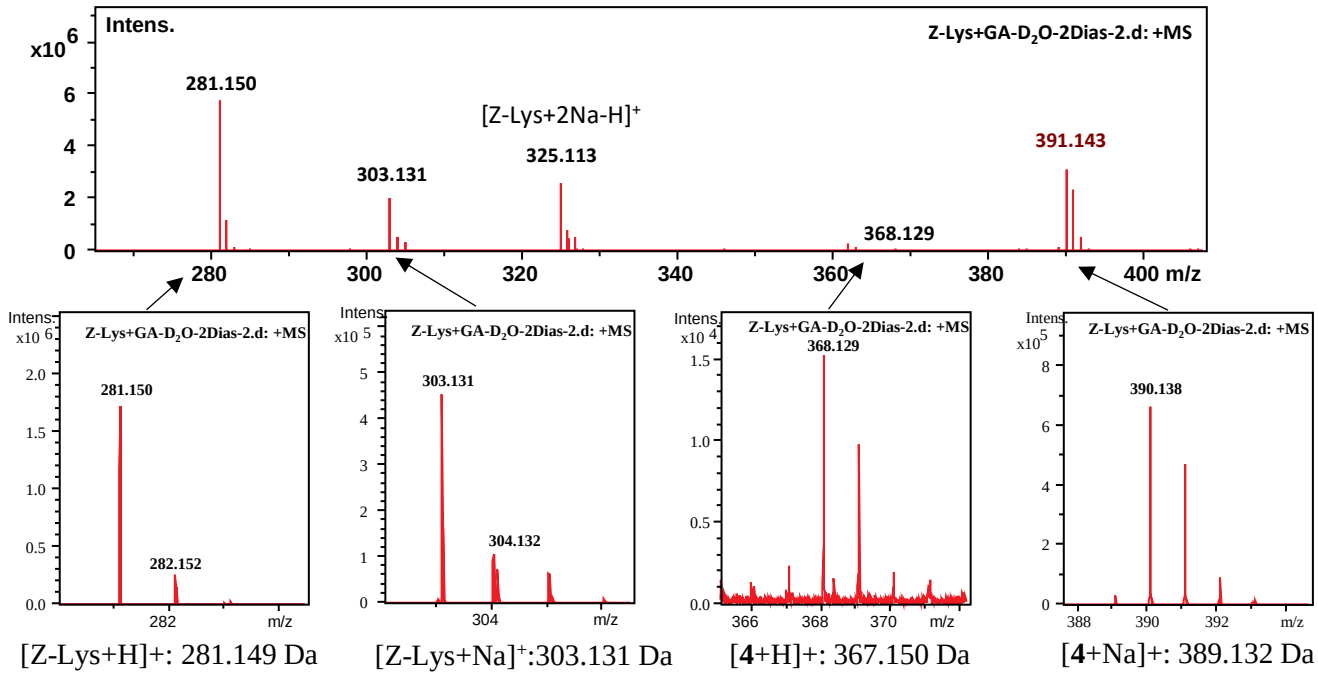


Fig. 11S. Changes of the abundance of individual ions during the reaction. The annotation of ions given in Table 1, in PCA model (Fig. 1) and in Scheme 2 in the main text. The X-axis represents the timeline (5, 10, 40, 70 and 100 days, respectively). Intensities were obtained after bucketing the mass spectra in Profile Analysis 2.3 with an *m/z* window of 0.01(Bruker Daltonics).



Meas. m/z	Ion Formula	m/z	err [ppm]	mSigma
142.0864	C7H12NO2	142.0863	-1.3	10.2
169.0970	C8H13N2O2	169.0972	1.2	8.3
187.1077	C8H15N2O3	187.1077	-0.1	13.0
259.1436	C15H19N2O2	259.1441	1.9	28.1
277.1546	C15H21N2O3	277.1547	0.3	21.1
323.1605	C16H23N2O5	323.1601	-1.0	4.5

Fig. 12S. Tandem mass spectrum obtained for protonated molecule of intermediate **4** in the cleaned-up reaction mix by SPE after 100-day incubation (GA to Z-Lys molar ratio 5:1, HEPES, pH 7.2. The ions identified as the precursor fragments are marked with yellow circles, their formulas, mass errors and mSigma values are listed below. Fragmentation sites are marked in the intermediate structure.



The above m/z values are given for formulas containing H (without considering H-D exchange)

H-D exchange in Z-Lys (*m/z* 281.149) = **0.18 +/- 0.09**
H-D exchange in **4** (*m/z* 367.150) = **1.21 +/- 0.08**
H-D exchange in **4** excluding Z-Lys moiety = **1.03**

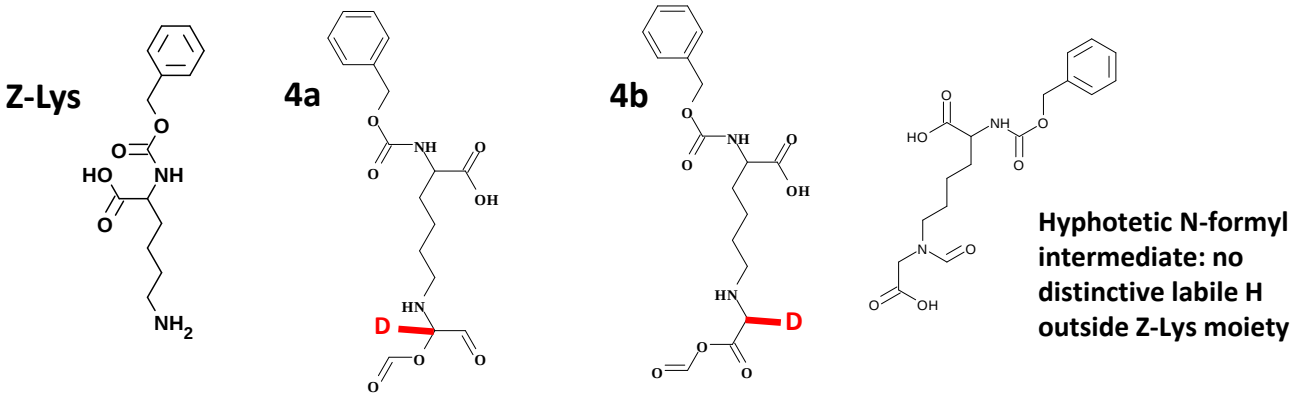


Fig. 13S. ESI(+)-QTOF-MS spectrum acquired for the reaction mixture in D_2O (GA to Z-Lys molar ratio 2:1, 60 °C, 72 h; then 100 dilution with H_2O). For intermediate **4** (protonated molecule and sodium adduct) and in respective ions of Z-Lys, the rate of D incorporation was assessed using multivariate calibration and prediction. PLS model was obtained using Unscrambler software, based on the experimental isotopic pattern acquired for natural isotopes distribution and the theoretical patterns created for different number of hydrogen atoms exchanged to deuterium. As indicated in molecular structures above, H-D exchange can take place only at carbonyl α-hydrogen in the proposed **4a** or **4b** structures but not in N-formyl compound that has no exchangable hydrogen available outside Z-Lys moiety.

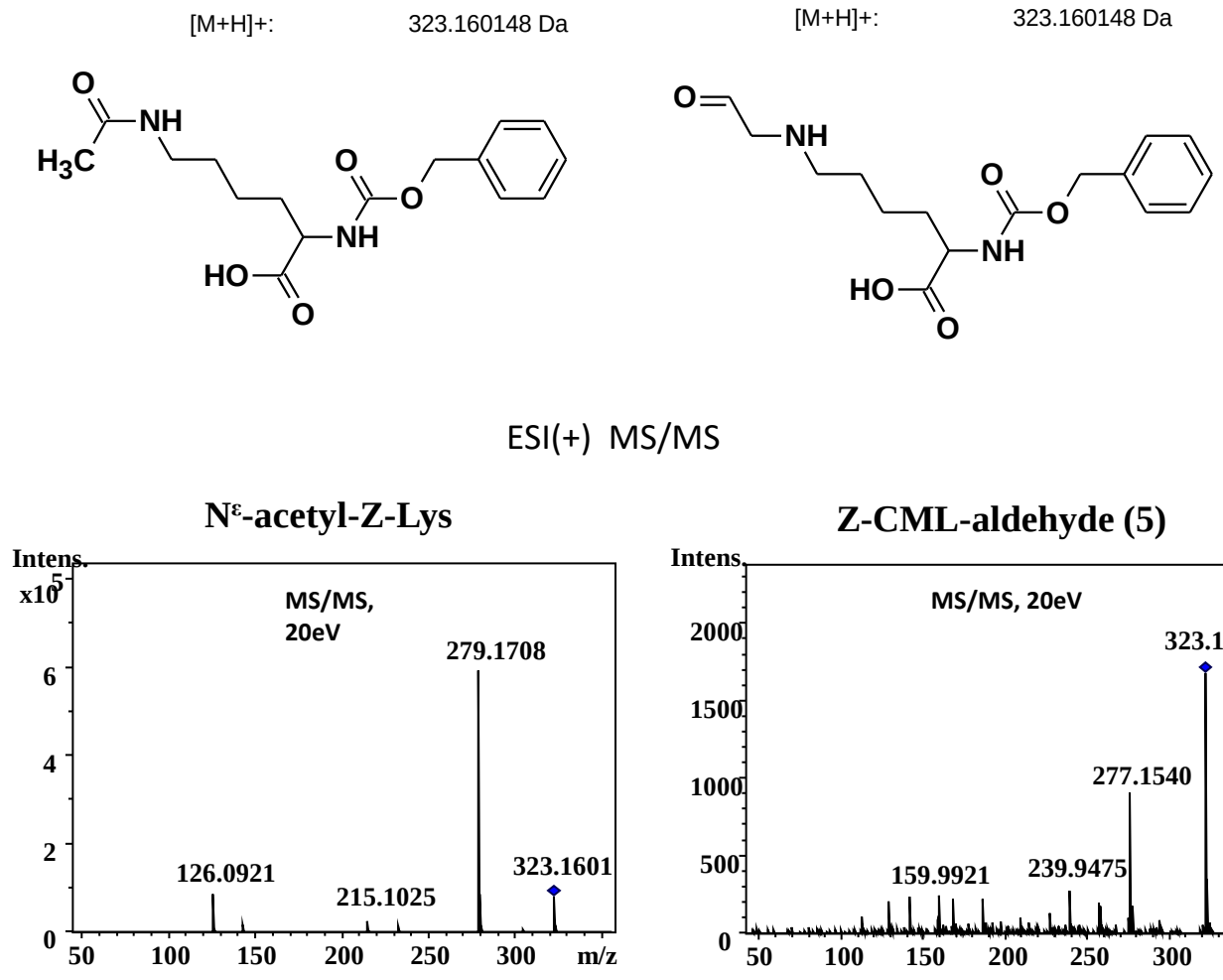


Fig. 14s. ESI(+)-QTOF-MS/MS spectrum acquired for the synthesized Nε-acetyl-Z-Lys and for the precursor *m/z* 323.160 corresponding to intermediate **5** (reaction mixture containing GA to Z-Lys molar ratio 2:1, 60 °C, 72 h). Molecular structures of both compounds are given above.

Nε-acetyl-Z-Lys was obtained by mixing Z-Lys with acetyl chloride in acetonitrile (molar ratio 1:1).