PNAS

Supporting Information for

Aqueous Microdroplets Enable Abiotic Synthesis and Chain Extension of Unique Peptide Isomers from Free Amino Acids

Dylan T. Holden, Nicolás M. Morato, and R. Graham Cooks*

Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA
*Corresponding author: R. Graham Cooks
Email: cooks@purdue.edu

This PDF file includes:

Figures S1 to S14
Tables S1 to S4
Equation S1

Supporting Figures

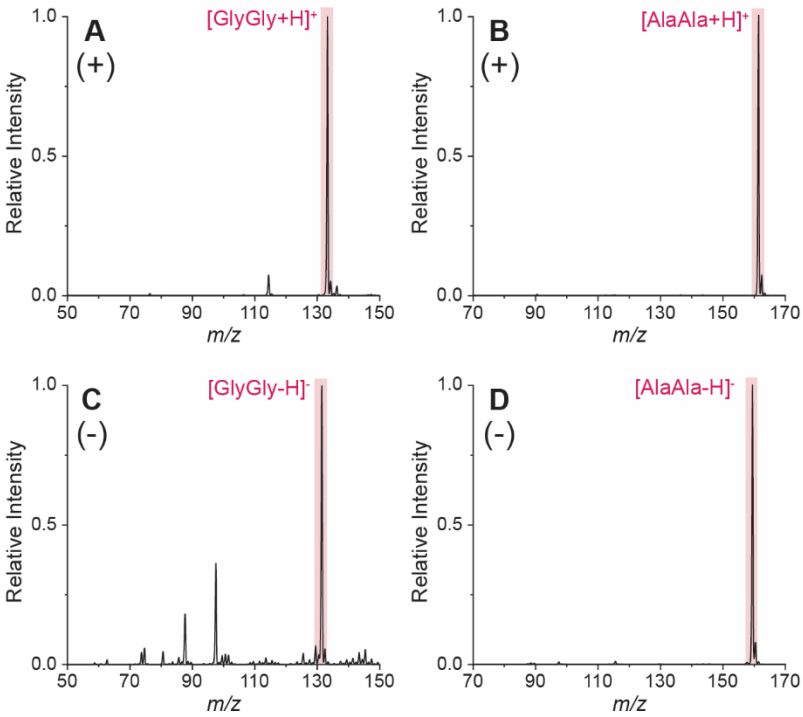


Figure S1. Full scan nESI-MS spectra of aqueous solutions (5 mM) of glycylglycine (GlyGly, **A** and **C**) and L-alanyl-L-alanine (AlaAla, **B** and **D**). Spectra were acquired in both positive (*top row*, **A, B**) and negative (*bottom row*, **C, D**) ion mode. The observed dipeptide peaks are highlighted in all cases.

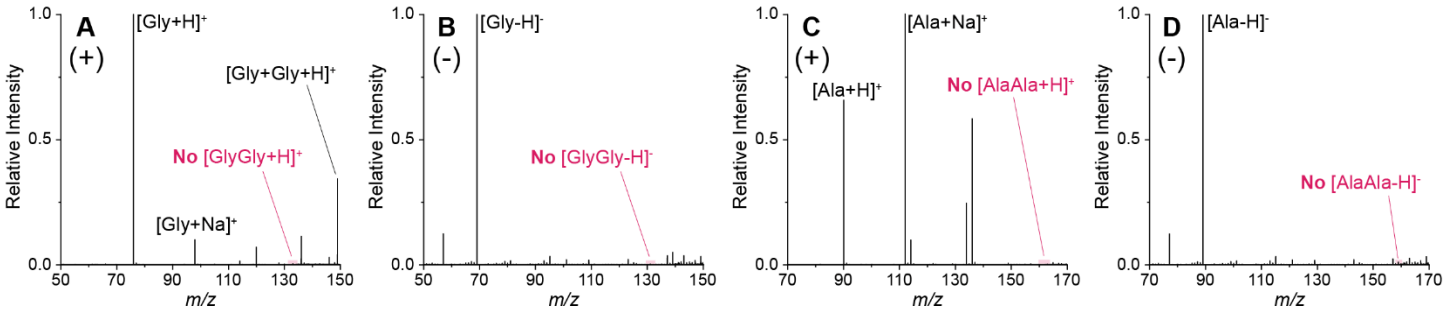


Figure S2. Full scan ESI-MS spectra obtained for aqueous solutions of Gly (**A** and **B**) and Ala (**C** and **D**) after incubation at room temperature for 2 hours. Results in both the positive (**A** and **C**) and negative (**B** and **D**) modes are shown for both amino acids. No dipeptide product is observed in any of the spectra.

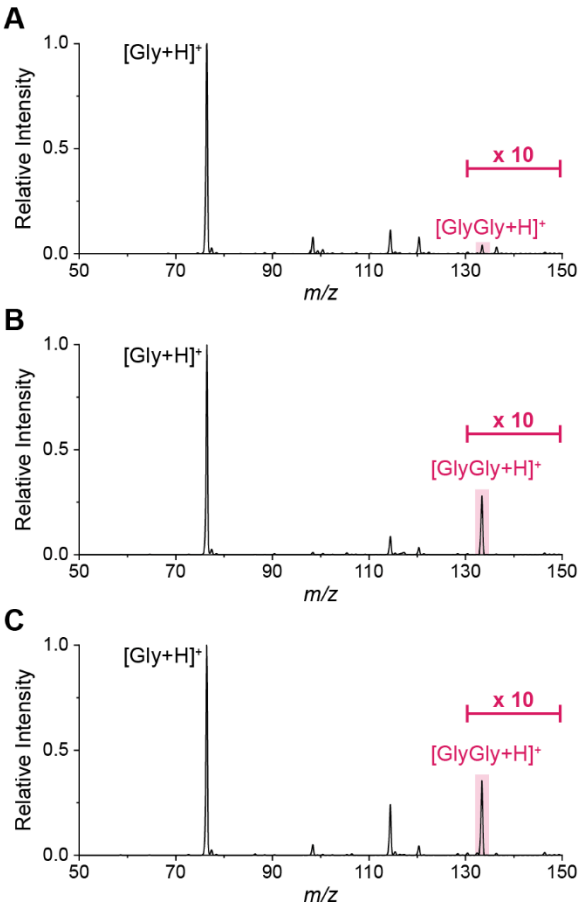


Figure S3. Full scan nESI-MS spectra obtained by spraying an aqueous solution of Gly while varying the spray distance (i.e. distance between the nESI emitter and the inlet of the mass spectrometer). Shown are representative results for spray distances of 1 cm (**A**), 10 cm (**B**), and 20 cm (**C**). The generated dipeptide species is highlighted in all cases.

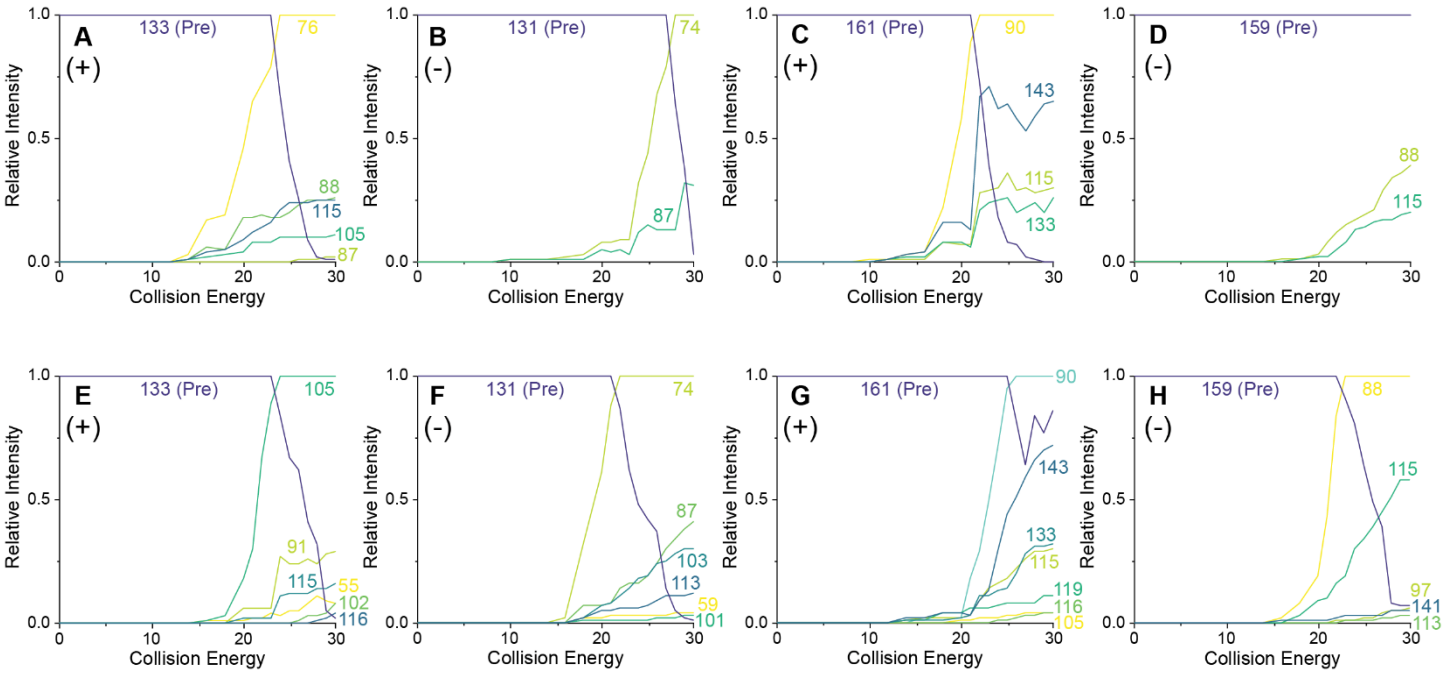


Figure S4. Breakdown curves for the standard (*top row*, **A – D**) and isomeric (*bottom row*, **E – H**) dipeptides. Results are shown for GlyGly in both protonated (**A** and **E**; precursor *m/z* 133) and deprotonated (**B** and **F**; precursor *m/z* 131) forms, as well as for AlaAla in both protonated (**C** and **G**; precursor *m/z* 161) and deprotonated (**D** and **H**; precursor *m/z* 159) forms. Note that the isomer spectra (*bottom row*) are obtained by spraying aqueous solutions of pure amino acids, whereas standard spectra (*top row*) are acquired by spraying aqueous solutions of authentic dipeptide standards.

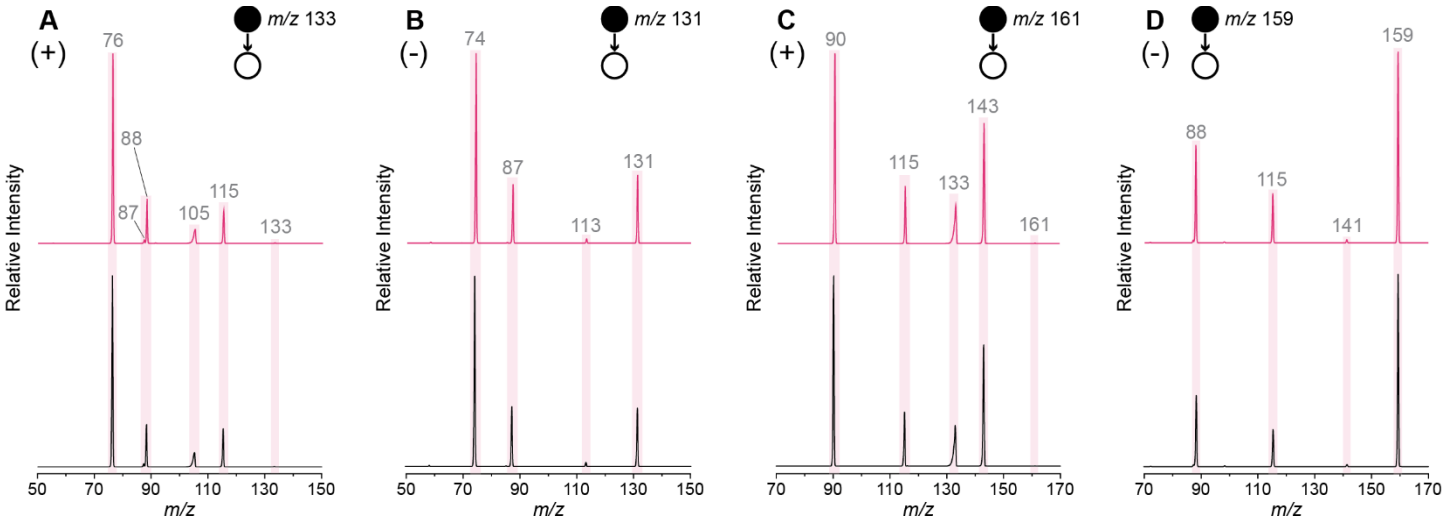


Figure S5. Comparison of the MS/MS spectra of the standard dipeptides before (*bottom*) and after (*top*) collisional heating (30 seconds; CE below fragmentation threshold). Results for both protonated (**A**) and deprotonated (**B**) GlyGly, as well as protonated (**C**) and deprotonated (**D**) AlaAla, are shown. Note that no changes are observed in any of the cases. This result should be contrasted with that for the isomeric peptides shown in Fig. 3 in the main text.

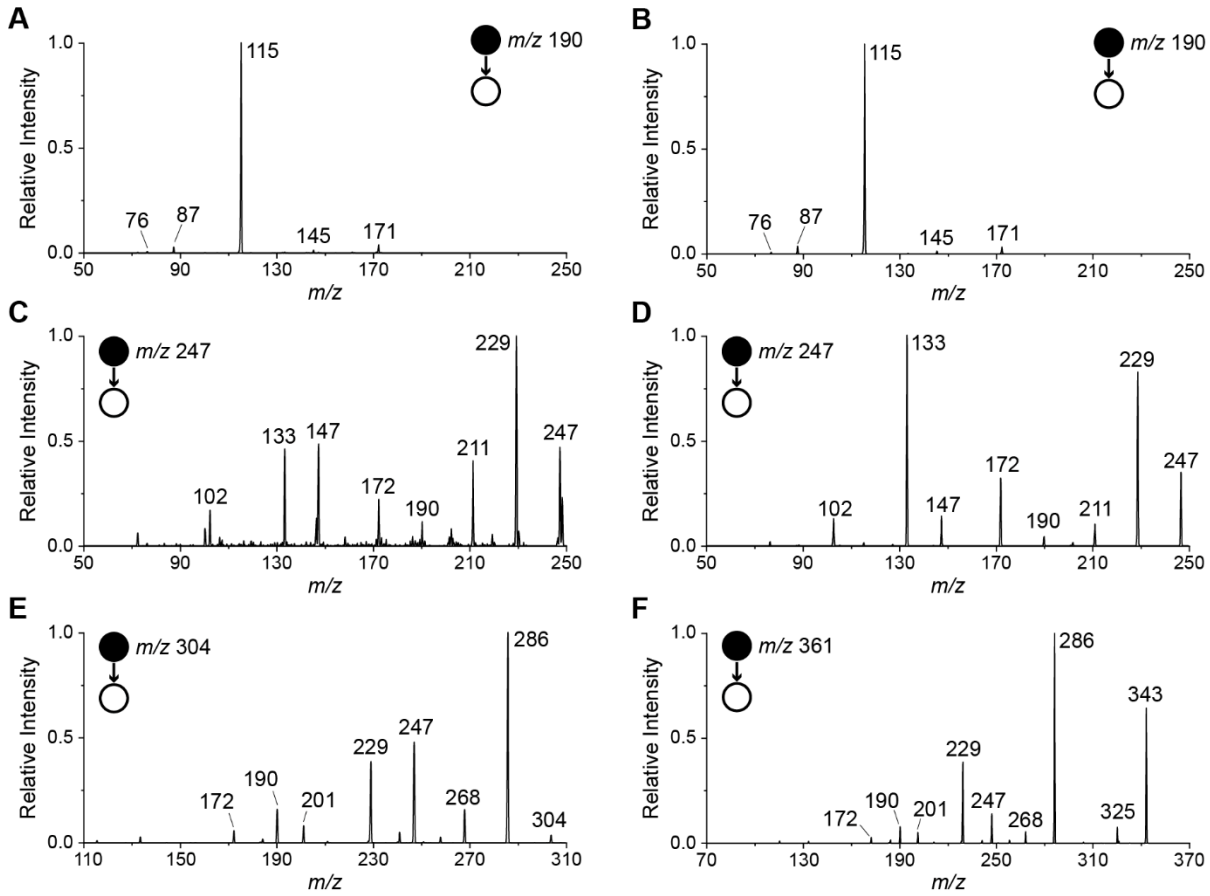


Figure S6. MS/MS analysis of the higher order Gly peptides generated through droplet-fusion experiments. The MS/MS spectrum of the generated Gly_3 species (**A**) is compared with that of the standard Gly tripeptide (**B**), and no significant differences are identified. Similarly, the MS/MS spectra of the synthesized Gly_4 (**C**) and its standard (**D**) gave identical spectra. MS/MS data for the microdroplet-generated Gly_5 (**E**) and Gly_6 (**F**) are also included. All data are for positively charged ions.

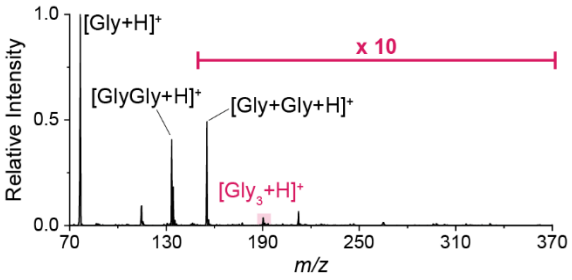


Figure S7. Full scan nESI-MS spectrum of a mixture of Gly and GlyGly in the positive ion mode This single-emitter experiment leads to insignificant amounts of higher order peptides, a result that contrasts with the droplet-fusion experiments where oligomers up to hexapeptides are observed.

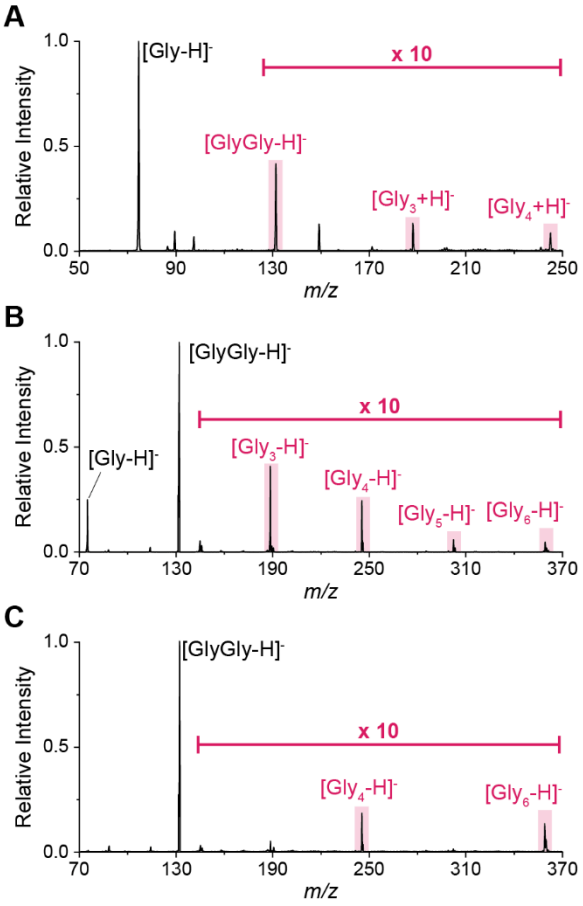


Figure S8. Droplet-fusion experimental results in the negative ion mode. When aqueous solutions of Gly are sprayed from two nESI emitters (**A**), di, tri and tetrapeptides are observed. Similarly, when spraying Gly from one nESI emitter and GlyGly from another (**B**), oligomers up to hexapeptide species are detected. Finally, if both emitters are used to spray GlyGly (**C**) only tetra and hexapeptides are obtained, as would be expected. The generated peptide species are highlighted in all cases. Note the indicated zoomed-in region in each spectrum.

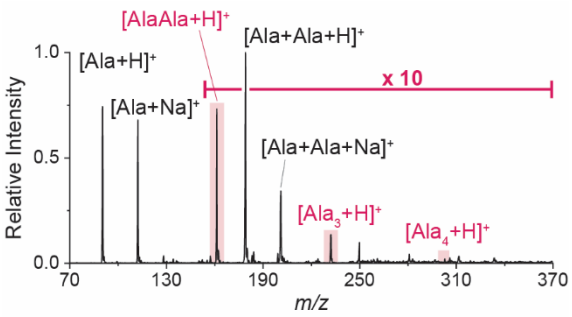


Figure S9. Droplet-fusion experimental results for Ala in the positive ion mode. When aqueous solutions of Ala are sprayed from two nESI emitters, di, tri and tetrapeptides are observed. The generated peptide species are highlighted. Note that the indicated zoomed-in region on the spectrum does not cover the base peak.

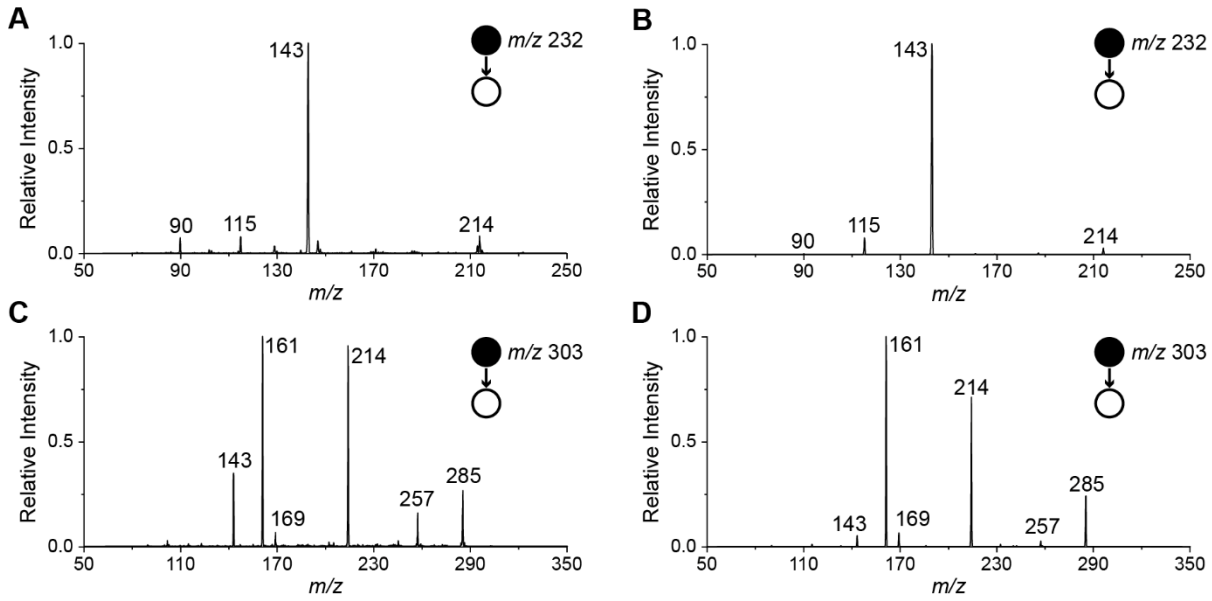


Figure S10. MS/MS analysis of the higher order Ala peptides generated through droplet-fusion experiments and recorded in the positive ion mode. MS/MS spectrum of the generated Ala_3 species (**A**) shows no significant differences compared with that of the standard Ala tripeptide (**B**). MS/MS spectrum of the synthesized Ala_4 (**C**), and its corresponding standard (**D**) are also shown.

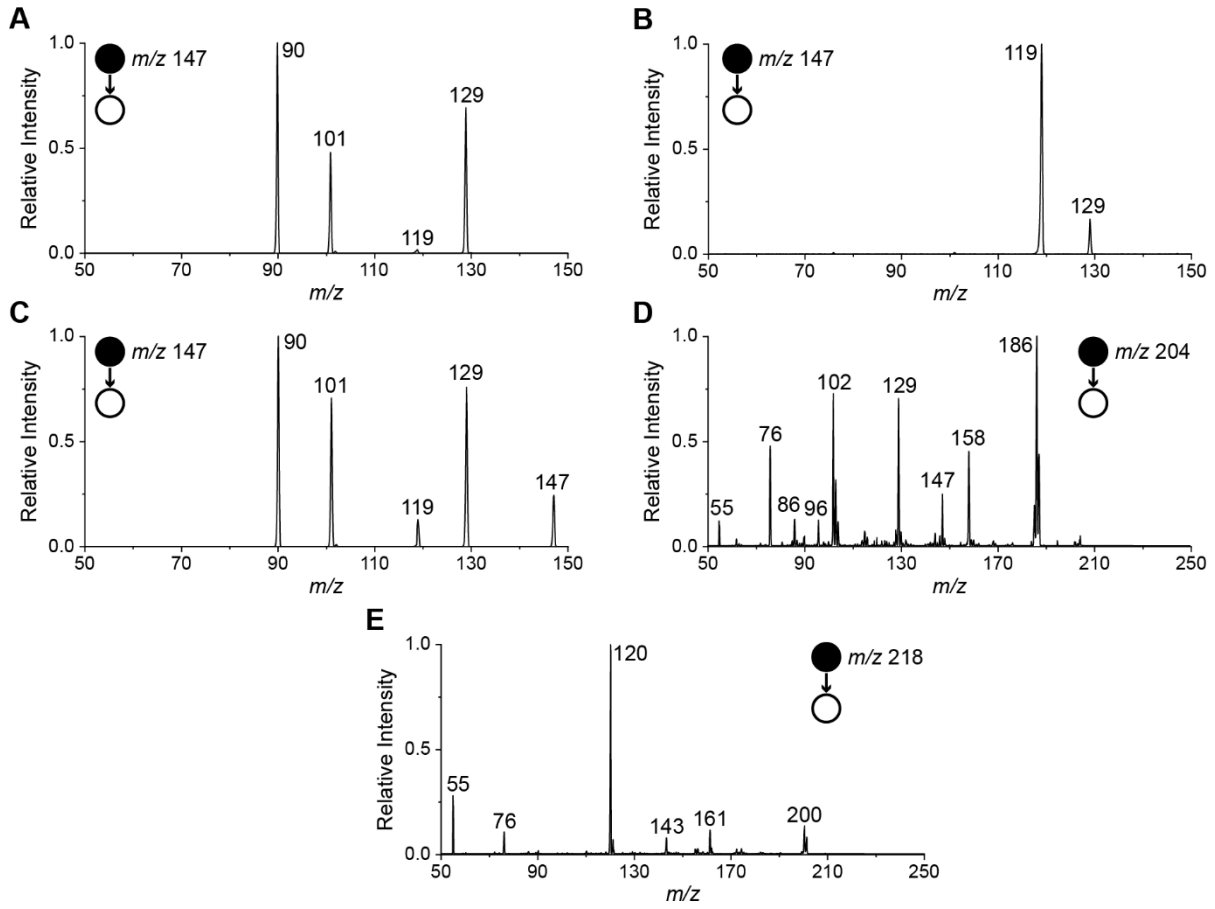


Figure S11. MS/MS analysis in the positive ion mode of the higher order heteropeptides generated through droplet-fusion experiments. MS/MS spectrum of the generated hetero dipeptide (**A**) is compared with that of authentic AlaGly (**B**) and authentic GlyAla (**C**). Clearly, the synthesized heterodipeptide is comprised largely of the GlyAla sequence isomer. MS/MS spectra of the synthesized tripeptide containing 2 Gly and 1 Ala (**D**) as well as that with 1 Gly and 2 Ala (**E**) are also shown.

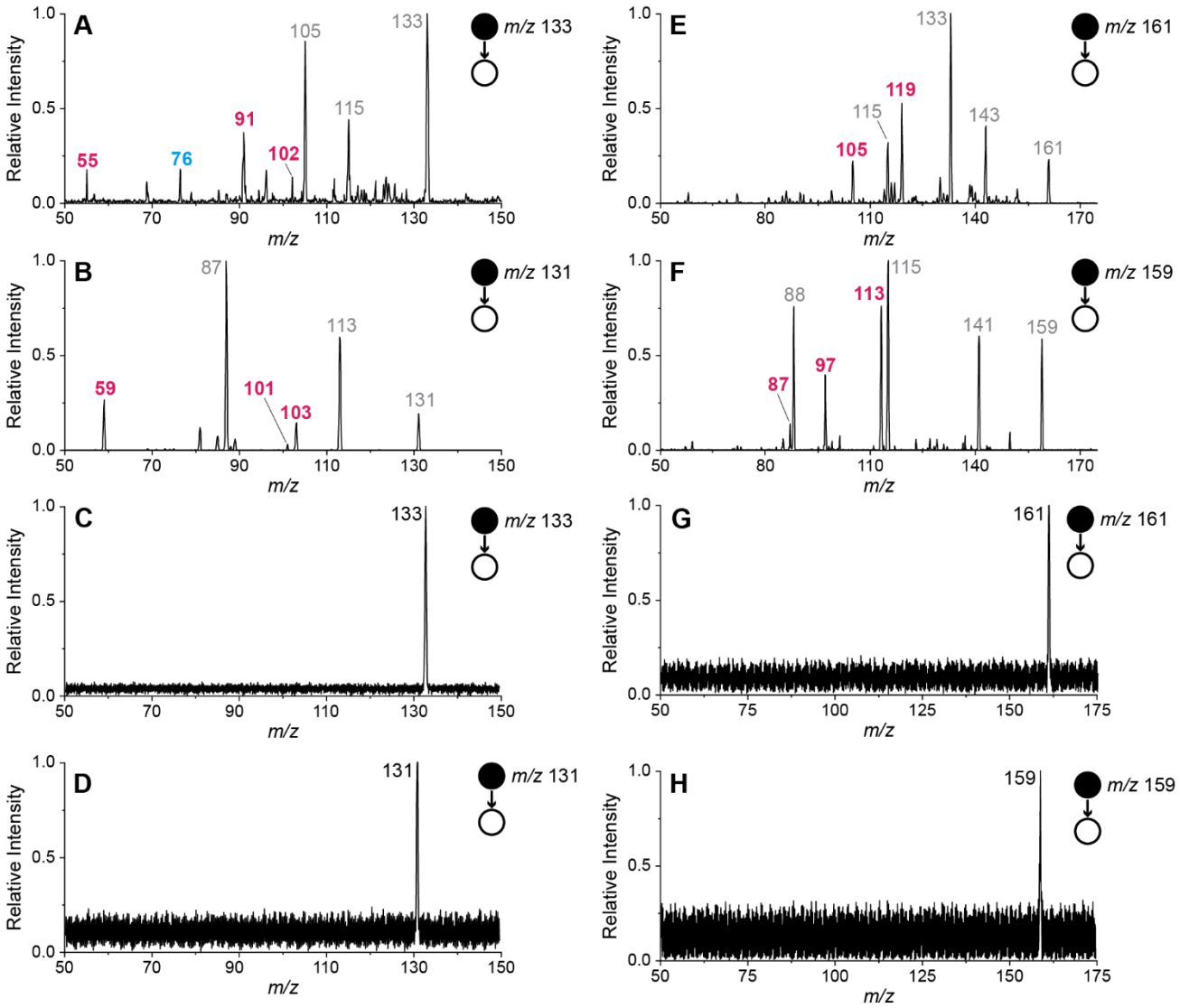


Figure S12. MS/MS spectra of standard GlyGly (**A-B**) and AlaAla (**E-F**) in the positive (**A** and **E**) and negative (**B** and **F**) ion modes acquired using 10 pM aqueous solutions of the reference compounds. In all cases, the unique peaks from standard GlyGly are highlighted in blue, while the fragments of the unique isomer (shown to involve the oxazolidinone) are highlighted in pink. Blank MS/MS analysis of background ions at all the expected *m/z* values for protonated (**C** and **G**) and deprotonated (**D** and **H**) dipeptides was carried out using pure water to assess for the presence of possible chemical interferences. The absolute intensities of the base peaks of these spectra (**C**, **D**, **G**, **H**) are ca. 100 times lower than those of spectra **A, B, E** or **F**.

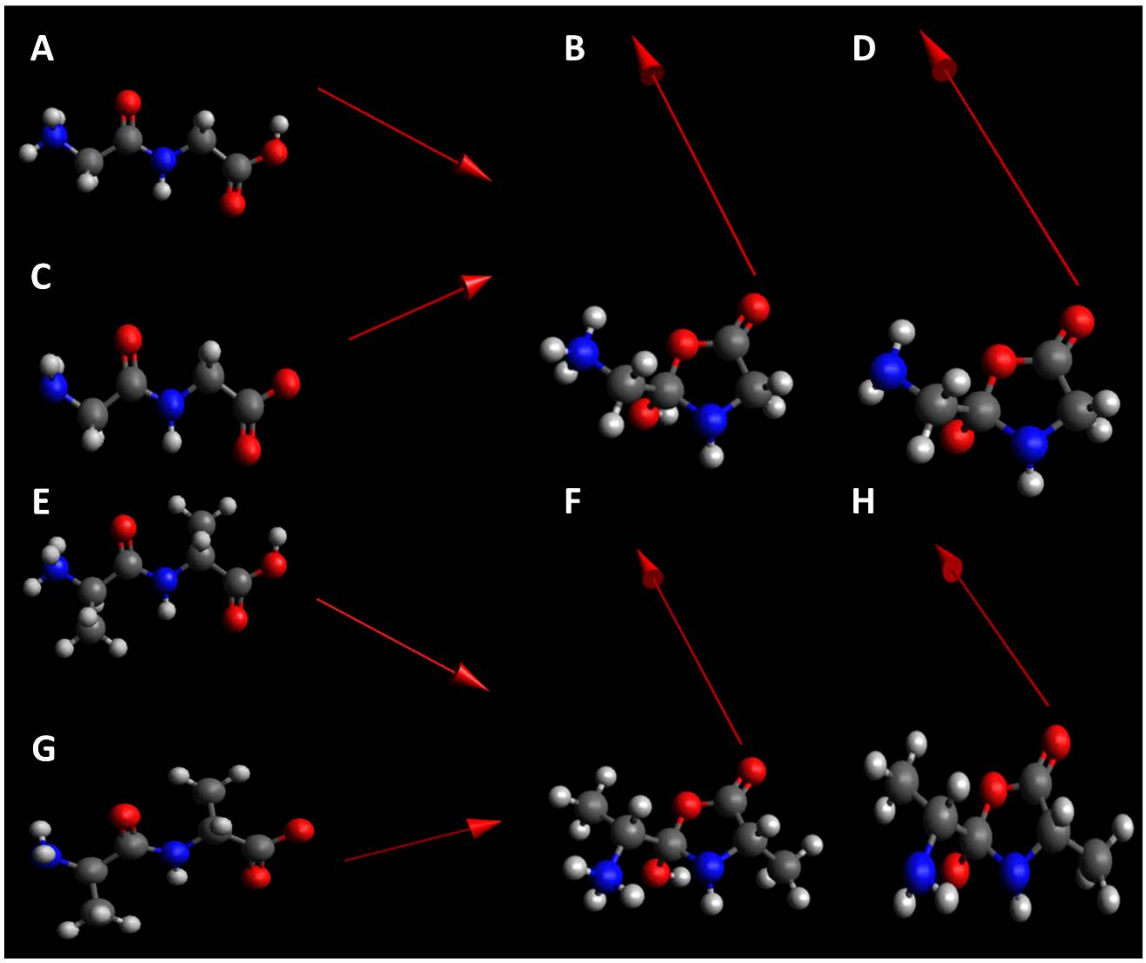


Figure S13. Dipole moment vectors for the authentic dipeptides (**A,C,E,G**) and proposed oxazolidinone isomers (**B,D,F,H**) in both protonated (**A,B,E,F**) and deprotonated (**C,D,G,H**) forms. Geometry optimizations were performed with a UFF forcefield using a Steepest Descent algorithm until minima were located. Dipole moments were calculated using Gasteiger partial charges. All calculations were carried out in Avogadro (Seattle, WA, USA).

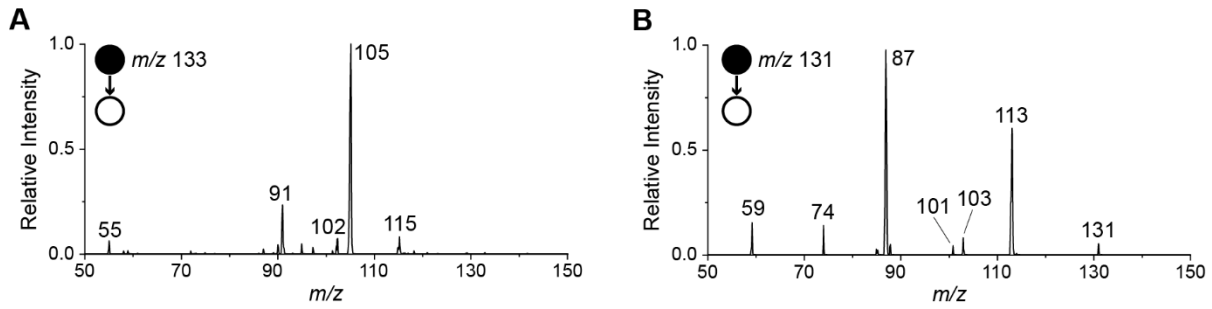


Figure S14. MS/MS analysis of 10 mM Gly in a 1% (m/v) NaCl aqueous solution sprayed via nESI in both the positive (**A**) and negative (**B**) ion modes. In both cases the fragmentation profile associated with isomeric GlyGly was observed. Experimental parameters were identical as those described for nESI experiments using pure water solutions.

Supporting Tables

Table S1. Summary of the product ions identified for all the dipeptide species (standards and synthesized isomers). The measured exact masses and the derived chemical composition (all within 5 ppm mass error) are included. Unique fragments associated with standard compounds are highlighted in pink, while those fragment ions characteristic of the isomeric species are labeled with blue.

Precursor Species	Precursor *m/z*	Product *m/z*	Elemental Composition
[GlyGly + H]$^+$	133.06038	55.0184	$C_2HNO^{+\cdot}$
		76.0388	$C_2H_6O_2N^+$
		87.0550	$C_3H_7ON_2^+$
		88.0386	$C_3H_6O_2N^+$
		91.0506	$C_2H_7O_2N_2^+$
		102.0194	$C_3H_4O_3N^+$
		105.0656	$C_3H_9O_2N_2^+$
		115.0501	$C_4H_7O_2N_2^+$
		116.0341	$C_4H_6O_3N^+$
[GlyGly - H]$^-$	131.0488	59.0140	$C_2H_3O_2^-$
		74.0249	$C_2H_4O_2N^-$
		87.0452	$C_3H_7N_2O^-$
		101.0305	$C_3H_5O_2N_2^-$
		103.0530	$C_3H_7O_2N_2^-$
		113.0344	$C_4H_5O_2N_2^-$
[AlaAla + H]$^+$	161.09141	90.0547	$C_3H_8O_2N^+$
		105.0696	$C_3H_9O_2N_2^+$
		115.0865	$C_5H_{11}ON_2^+$
		119.0816	$C_4H_{11}O_2N_2^+$
		133.0968	$C_5H_{13}O_2N_2^+$
		143.0814	$C_6H_{11}O_2N_2^+$
[AlaAla - H]$^-$	159.077	87.0563	$C_3H_7ON_2^-$
		88.0404	$C_3H_6O_2N^-$
		97.1157	$C_5H_7ON^-$
		113.0718	$C_5H_9ON_2^-$
		115.0876	$C_5H_{11}ON_2^-$
		141.0556	$C_6H_9O_2N_2^-$

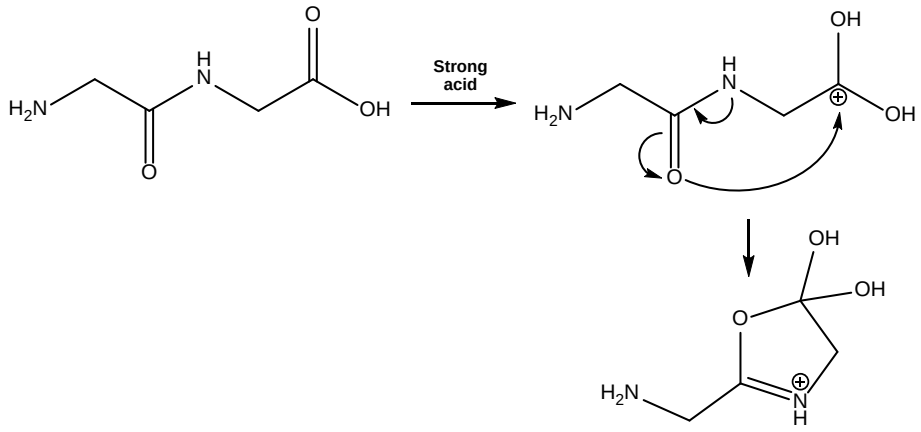
Table S2. Conditions explored in the attempted ambient collection of the microdroplet-generated dipeptide isomers. Note that in all cases the collected material was characterized through Raman spectroscopy, ^{1}H, ^{13}C and ^{15}N NMR and by mass spectrometry performed under non-accelerating conditions (spray distance < 0.5 cm). The dipeptide isomers could not be detected in the collected materials regardless of the conditions utilized for collection. Note that in the methods where solvent was allowed to fully evaporate, the remaining material was reconstituted in water (for MS and paper-SERS analysis), deuterated water (NMR analysis) or used directly (Raman analysis).

Method	Description
ESSI[1] spray collection	Aqueous amino acid solutions were sprayed by ESSI for 1 hour and the spray plume was collected in a dry RBF[2]
ESSI spray collection on water	Aqueous amino acid solutions were sprayed by ESSI for 1 hour and the spray plume was collected in an RBF containing ~ 2 mL of H_2O
ESSI spray collection on methanol	Aqueous amino acid solutions were sprayed by ESSI for 1 hour and the spray plume was collected in an RBF containing ~ 2 mL of MeOH
ESSI spray collection with heated tube	Aqueous amino acid solutions were sprayed by ESSI for 1 hour through a grounded stainless-steel tube (3.2 mm o.d., 1.6 mm i. d., length 22 cm) heated to 300 °C and the spray plume was collected in a dry RBF. Material deposited on the interior of the heated tube was also analyzed.
ESSI spray collection on cooled surface	Aqueous amino acid solutions were sprayed by ESSI for 1 hour and the spray plume was collected in a dry RBF cooled by surrounding dry ice
ESSI spray collection with heated tube and cooled surface	Aqueous amino acid solutions were sprayed by ESSI for 1 hour through a grounded stainless-steel tube (3.2 mm o.d., 1.6 mm i. d., length 22 cm) heated to 300 °C and the spray plume was collected in a dry RBF surrounded by dry ice
Solvent removal by rotary evaporation	Aqueous amino acid solutions were placed in a rotovap for 10 min to remove solvent and increase the solution surface area
Solvent removal by lyophilization	Frozen aqueous amino acid solutions and collected ESSI spray were placed in a freeze dryer and cooled to -50 °C under vacuum to remove the solvent
Thin film reactions (pH 7)	1 mL of aqueous amino acid solutions were dropcast onto glass slides, spread evenly across the surface, and allowed to fully evaporate
Thin film reactions (pH 10)	1 mL of pH-adjusted (NaOH) aqueous amino acid solutions were dropcast onto glass slides, spread evenly across the surface, and allowed to fully evaporate
Thin film reactions (pH 4)	1 mL of pH-adjusted (HCl) aqueous amino acid solutions were dropcast onto glass slides, spread evenly across the surface, and allowed to fully evaporate
Thin film reaction with nonvolatile components	1 mL of aqueous amino acid solutions combined with 1 uL of DMAPA[3] were dropcast onto glass slides, spread evenly across the surface, and allowed to fully evaporate

[1] ESSI: Electrosonic spray ionization; [2] RBF: Round bottom flask; [3] DMAPA: 3-dimethylamino-1-propylamine

Note: The proposed isomeric dipeptide structure, in the fully covalent or ion-neutral forms of the ion, is likely to rapidly decompose once isolated from the partially solvated environment of the microdroplet air-water interface or the gas phase, and then readily neutralize. Collection of these products is hindered also by expected rapid hydrolysis into constituent amino acids or decomposition into volatile species (i.e. isocyanic acid, ethane-1,2-diimine, N-methylene methanimidamide, and acetic acid)

Table S3. Structures considered for isomer of dipeptide. Several structures have been considered for the isomeric compound generated at the micro-droplet interface in the nESI spray experiments. These structures are based on consideration of the IMS and MS/MS data for the $(M+H)^+$ and $(M-H)^-$ ions generated from this isomer in comparison with the IMS and MS/MS data for the authentic dipeptides. For simplicity, only the GlyGly case is considered in detail, but the arguments extend to AlaAla.

Compound	Main Features
Authentic dipeptide (***S1***)	For comparison with the isomers that will be discussed below, the main features of the positive and negative ion MS/MS spectra of the authentic dipeptide GlyGly (***S1***) are summarized first. The ions considered are $[M + H]^+$, principally the *N*-protonated form, and $[M – H]^-$, principally considered to be in the carboxylate form. ***Positive ions*** Principal and most diagnostic ions are (cf. **Figure 2A** and **2E**): *m/z* 76 $H_3N^+–CH_2–CO_2H$ *y-cleavage ion* *m/z* 87 $H_3N^+–CH_2–CO–N=CH_2$ *a-cleavage ion* *m/z* 88 $(M + H – CO – NH_3)^+$ This agrees with the reference database MS/MS spectrum https://hmdb.ca/spectra/ms_ms/446206 ***Negative ions*** Principal ions are: *m/z* 74 $H_2N–CH_2–CO_2^-$ *y- ion* *m/z* 87 $(M – H – CO_2)^-$ This agrees with the reference database MS/MS spectrum https://hmdb.ca/spectra/ms_ms/439069. Note that these species are not exclusively diagnostic of the authentic dipeptide as they are also observed (though in different intensities) for the isomeric $[GlyGly-H]^-$ (cf. **Figure 2A** and **2E**).
Geminal diol oxazolidine (***S2***)	Interconversion between the gem diol species (***S2***) and the authentic dipeptide is readily accounted for in a highly acidic environment. The carboxylic acid terminus acts as an electrophile causing cyclization. This chemistry is analogous to known microdroplet chemistry, e.g. formic acid reaction with *o*-diamines to form benzimidazoles (Basuri et al. *Chemical Science* **2020**, 11, 47, 12686-12694).  We expect fragmentation to be dominated by loss of H_2O. This product ion is observed, though not with the intensity that would be expected from the fragmentation of a geminal diol. In fact, water loss occurs less readily than in authentic dipeptides (**Figure 2**, *m/z* 115/133 ratio in authentic peptide >> 100, while for the isomer this ratio is ~2). This information suggests that the isomer does not have structure (**2**).

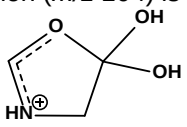
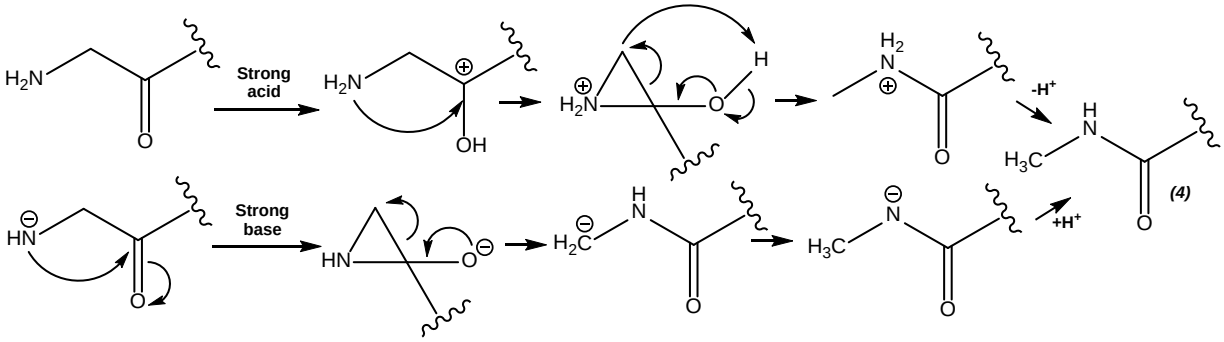
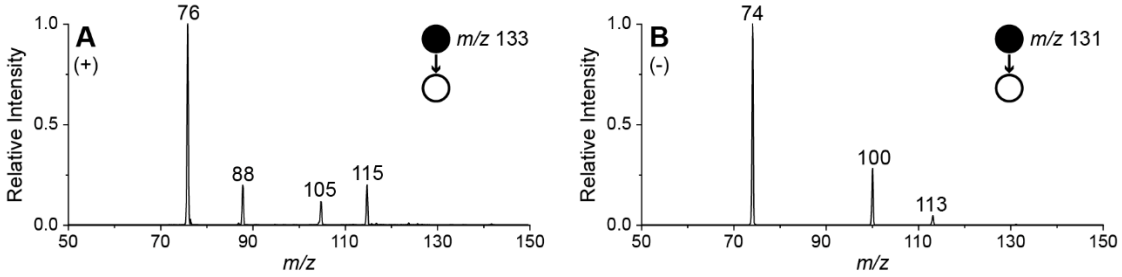
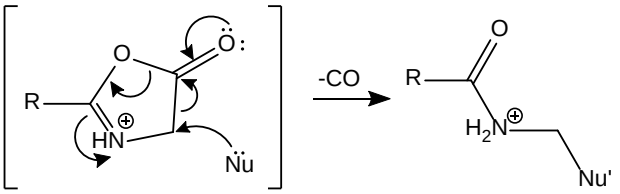
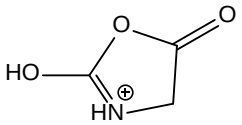
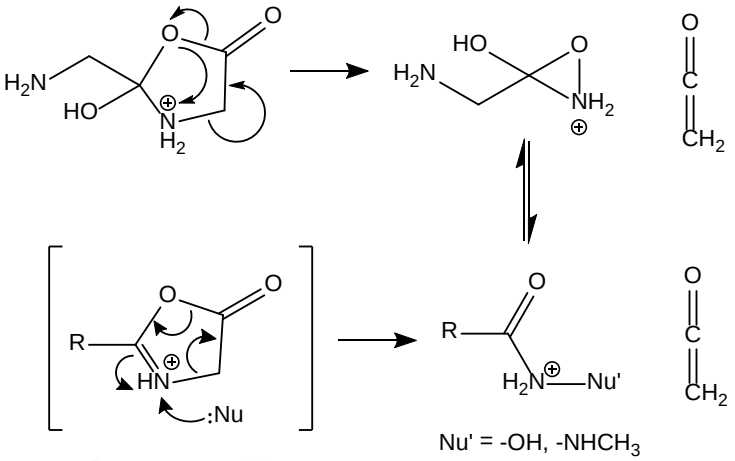
	Loss of $NHCH_2$ to give the stable oxazolium ion (*m/z* 104) is expected for this structure but **not** observed.  The cyclic structure fits the IMS data, but the fragmentation data rule out this structure.
Urea (***S3***)	The urea (***S3***) is a possible peptide isomer. It could be formed via a tricyclic intermediate in the presence of strong acid (or base) at the droplet interface. As before, the chemistry can be intramolecular as illustrated, or intermolecular before formation of the dipeptide, which is the favored interpretation.  However, the IMS data does not fit the extended chain structure of the urea ***S3***. Moreover, the MS/MS spectra of the authentic urea (obtained from a commercial standard of 2-[(methylcarbamoyl)amino] acetic acid purchased from Enamine and shown below in positive ***A*** and negative ***B*** ion modes) does not resemble that of the microdroplet-synthesized dipeptide isomer. This fact rules out the urea structure for the isomer. 

Table S4. Evidence for the assignment of the proposed structures (Scheme 1, Main text) of the dipeptide isomer. Arguments refer directly to **ion** structures as those are the species on which information was obtained.

Criteria for Isomeric Structures	Oxazolidinone Ion and Associated Ion-Neutral Features
Reasonable isomerization pathways for interconversion to/from dipeptide and isomer	Yes, see **Scheme 1**
More compact structure than authentic dipeptide	Yes, all species contain a 5-membered cyclic moiety whereas authentic dipeptides are linear
Reasonable structures for isomer-specific product ions	*Diagnostic ions of isomeric [GlyGly+H]$^+$* ***m/z* 105** $[M+H-CO]^+$ (Not diagnostic but significantly increased in intensity) Loss of carbonyl readily facilitated in the ion-neutral complexes  R = -OH, $-CH_2NH_2$ Nu = H_2O, NH_2CH_3 Nu' = -OH, $-NHCH_3$ **m/z 102** $[M+H-CH_3NH_2]^+$ Proposed species has a facile NH_2CH_3 to generate uniquely stabilized 2-hydroxyoxazolidinonium ion (extended delocalization), either from covalent oxazolidinones or either of the methylamine ion-neutral complexes  ***m/z* 91** $[M+H-C_2H_2O]^+$ The oxazolidinone ion (in covalent or ion-neutral complex forms) are capable of ketene loss via intramolecular cyclization to form an oxaziridine or nucleophilic attack of the neutral (i.e. methylamine or water), respectively. Both product ions can also interconvert.  R = -OH, $-CH_2NH_2$ Nu = H_2O, NH_2CH_3 Nu' = -OH, $-NHCH_3$

	***m/z* 55:** $[M+H-C_2H_8NO_2]^{+\cdot}$ Ion is likely $[HN=C=C=O]^{+\cdot}$ unknown in normal peptide but accessible from isomeric structures through extensive rearrangements *Authentic peptide gives two characteristic ions **not seen** for the isomer:* ***m/z* 76** a *y*-type peptide fragment likely $^{+}H_3N\text{-}CH_2\text{-}CO_2H$ ***m/z* 88**, a *b*-type peptide fragment, likely $^{+}H_3N\text{-}CH_2\text{-}CO\text{-}NH{=}CH2$
Mechanism for chain extension reaction	Yes, see **Fig. 5B** which shows dehydration from the 2-hydroxyoxazol-4-one. Ring-opening and dehydration is even simpler considering the ion-neutral complex of the dipeptide isomer

Table S5. Single-point energy and dipole moment for the authentic and proposed oxazolidinone isomers in both protonated and deprotonated forms. The ratio of the dipole moments between each isomer pair (either protonated or deprotonated) is also shown. In all cases the dipole moment of the oxazolidinone isomers is larger than that of the authentic dipeptides. Geometry optimizations were performed with a UFF forcefield using a Steepest Descent algorithm until minima were located. Dipole moments were calculated using Gasteiger partial charges. All calculations were carried out in Avogadro (Seattle, WA, USA). Note that calculations were not performed for the ion-neutral complexes which likely have even larger dipole moments.

Structure	Calculated Energy (kJ/mol)	Dipole Moment (D)	Direction of Moment Vector	Dipole Moment Ratio
$[GlyGly+H]^+$	44.0806	2.213	Figure S13A	1.486
$[Gly\ Oxazolidinone+H]^+$	143.006	3.289	Figure S13B	
$[GlyGly-H]^-$	50.671	1.631	Figure S13C	2.356
$[Gly\ Oxazolidinone-H]^-$	132.917	3.843	Figure S13D	
$[AlaAla+H]^+$	50.671	2.144	Figure S13E	1.767
$[Ala\ Oxazolidinone+H]^+$	161.868	3.782	Figure S13F	
$[AlaAla-H]^-$	74.279	1.575	Figure S13G	2.117
$[Ala\ Oxazolidinone-H]^-$	156.539	3.334	Figure S13H	

Supporting Equations

$$Conversion\ Ratio\ (\%) = \frac{Product\ Intensity}{Product\ Intensity + Reactant\ Intensity}$$

Equation S1. Conversion ratio calculation.