Studying the Key Intermediate of RNA Autohydrolysis by Cryogenic Gas-Phase Infrared Spectroscopy

Kim Greis, Carla Kirschbaum, Martín I. Taccone, Michael Götze, Sandy Gewinner, Wieland Schöllkopf, Gerard Meijer, Gert von Helden, and Kevin Pagel*
Abstract: Over the course of the COVID-19 pandemic, mRNA-based vaccines have gained tremendous importance. The development and analysis of modified RNA molecules benefit from advanced mass spectrometry and require sufficient understanding of fragmentation processes. Analogous to the degradation of RNA in solution by autohydrolysis, backbone cleavage of RNA strands was equally observed in the gas phase; however, the fragmentation mechanism remained elusive. In this work, autohydrolysis-like intermediates were generated from isolated RNA dinucleotides in the gas phase and investigated using cryogenic infrared spectroscopy in helium nanodroplets. Data from both experiment and density functional theory provide evidence for the formation of a five-membered cyclic phosphate intermediate and rule out linear or six-membered structures. Furthermore, the experiments show that another prominent condensed-phase reaction of RNA nucleotides can be induced in the gas phase: the tautomerization of cytosine. Both observed reactions are therefore highly universal and intrinsic properties of the investigated molecules.

The development of mRNA vaccines has recently experienced an unparalleled boost in the course of the worldwide COVID-19 pandemic. Their sudden breakthrough largely rests on the fact that, contrary to conventional vaccines, mRNA vaccines can be rapidly developed, produced on a large scale and adapted to different pathogens. The development of RNA-based pharmaceuticals requires advanced analytical methods including mass spectrometry (MS) and a profound understanding of fragmentation mechanisms for reliable identification of artificial RNA structures.

DNA is significantly more stable than RNA. This stability difference is mainly attributed to the different sugars incorporated in the backbone. Ribose in RNA contains a hydroxyl group at the C2' position, which destabilizes the phosphodiester bonds. This 2'-OH group can intramolecularly attack the phosphate group at C5' of the same nucleotide, leading to autohydrolysis of RNA even in the absence of degrading enzymes. The deoxyribose in DNA on the other hand, does not contain this OH group and therefore does not undergo autohydrolysis. As a result, DNA fragments may remain stable for hundreds, sometimes even thousands of years.

The mechanism of in vitro RNA-autohydrolysis has been studied extensively. The reaction is initiated by a nucleophilic attack of the 2'-OH group on the 3'-phosphate, proceeding via a phosphorane to yield a cyclic 2',3'-phosphate—the key intermediate of RNA hydrolysis (Figure 1a). The intermediate is subsequently hydrolyzed in aqueous media to yield a mixture of 2'- and 3'-phosphates. RNA autohydrolysis is accelerated up to a million fold in basic or acidic media compared to spontaneous hydrolysis at neutral pH.

In accordance with its reactivity in solution, the 2'-OH group also induces a substantially different behavior of RNA in the gas phase. Tandem mass spectrometry (MS/MS) probes the fragmentation of ions in the gas phase and can be used to deduce structural information of nucleic acids. The major difference in fragmentation of DNA and RNA is the facile formation of c-ions from RNA precursors, which are not observed in DNA fragmentation. c-Ions arise from cleavage between the 5'-O and the phosphorous atom (Figure S4) and their formation relies on the 2'-OH group providing a mobile proton. c-Ions generated from RNA dinucleotides are of identical m/z as the key intermediate of base-catalyzed RNA autohydrolysis in solution. However, there is disagreement about the fragmentation mechanism and exact structure of c-ions. In the literature, mechanisms proceeding via a linear structure or direct formation of 2',3'-cyclic structures as in the condensed phase have been proposed. In the case of dinucleotides, isomeric 3',5'-cyclic nucleoside monophosphates are equally conceivable fragment structures (Figure 1b).

Gas-phase infrared (IR) ion spectroscopy is a powerful tool for the direct structural analysis of intact molecular ions or short-lived fragments. Infrared multiple photon dissociation (IRMPD) spectroscopy has been successfully used to study RNA building blocks and nucleoside triphosphate ions as the key intermediate of base-catalyzed RNA autohydrolysis in solution. More recently, intact DNA dinucleotide anions were analysed by cryogenic IR spectroscopy in helium droplets, a technique offering an increased spectral resolution due to the suppression of ion heating. However, to date no tandem MS fragments of nucleotides have been spectroscopically studied with either method.

Here we investigate c-ions obtained by dissociation of deprotonated RNA dinucleotides ApA, GpG, UpG, and CpG (Figure 1c) using cryogenic IR spectroscopy in helium droplets. The utilized experimental setup has been described previously (see Supporting Information). Briefly, the ions of interest are generated by nano-electrospray ionization in negative ion mode and can be fragmented in the ion source by collisions with residual gas molecules in a process identical to collision-induced dissociation (CID). The ions are mass-to-charge selected in a quadrupole mass filter and accumulated in a hexapole ion trap (90 K). The thermalized, trapped ions are subsequently picked up by superfluid helium nanodroplets, generated by an Even–Lavie valve.
Inside the helium droplets, ions are rapidly cooled to 0.37 K and guided to a detection region where the helium droplet beam overlaps with an IR beam of the Fritz Haber Institute free-electron laser (FHI FEL).[18] Photon absorption excites vibrational modes, leading to evaporation of the helium shell and subsequent release of the ions. The ion yield is measured in a time-of-flight mass spectrometer and plotted against the photon wavenumber to obtain an IR spectrum. The experimentally obtained IR spectra are compared to computed vibrational frequencies. For this, the conformational space of $2',3'$- and $3',5'$-cyclic nucleoside monophosphates was explored for each nucleobase using CREST[19] with the semiempirical method GFN2-xTB.[20] Unique conformations were optimized and harmonic frequencies computed at the PBE0 + D3/def2-TZVP[21] level of theory in Gaussian 16[22] and scaled by an empirical factor of 0.965. Anharmonic frequencies were calculated using the GVPT2[23] method at the PBE0 + D3/def2-TZVP level of theory. Previously reported linear c-fragments were ruled out from consideration because they are not minima on the potential energy surface of the ions at the employed level of theory, and their optimization leads to $2',3'$-cyclic nucleoside monophosphates.

In the condensed phase, autohydrolysis can occur in the absence of enzymes or external reaction partners (Figure 1). In the gas phase, dissociation of deprotonated RNA dinucleotides was induced by in-source fragmentation (Figure S2), leading to c-fragments with identical $m/z$ as the intermediate of RNA autohydrolysis observed in solution. Once transferred to the high vacuum region of the instrument, no further ion activation occurs. The parent ion and its fragments were detected using time-of-flight MS (Figures 2a and S3). Subsequently, cryogenic IR spectra of the generated c-fragments were recorded for each nucleobase to determine their structure (Figure 2b). All IR signatures are similar in the fingerprint region (ca. 1000–1400 cm$^{-1}$) but exhibit significant differences in the functional group region (1400–1800 cm$^{-1}$). The vibrations in the fingerprint region are mostly originating from the ribose and phosphate moieties in the backbone. The two prominent bands at 1082 and 1326 cm$^{-1}$ can be assigned to symmetric and antisymmetric stretches within the phosphate group. The nucleobases mainly absorb in the functional group region. Furthermore, the IR signature of the c-fragment containing cytosine (C) is more prominent than its counterparts. For dinucleotides, two different bicyclic structures are conceivable for c-fragments, as the phosphate backbone can be either attacked by the 2'- or the 5'-OH group leading to a five- or six-membered phosphodiester ring, respectively (Figure 2c). The experimental spectra were compared to

---

**Figure 1.** Autohydrolysis reduces the stability of RNA. a) Schematic autohydrolysis of an RNA strand with a proposed 2',3'-cyclic phosphate intermediate. An intramolecular nucleophilic attack of the 2' OH group is crucial to initiate cleavage of the phosphodiester. In DNA, the absence of a 2' OH group prevents cleavage. b) Structures of cyclic nucleoside monophosphate isomers. c) Nomenclature and structures of RNA nucleobases (N).
computed energetics and frequencies of representative structures for both types of intermediates. First, from a thermodynamic point of view, the formation of a five-membered cyclic phosphate is favored over the formation of a six-membered phosphate by 5–27 kJ mol$^{-1}$ in free energy ($\Delta F_{90K}$), depending on the nucleobase (Tables S1–S4). Moreover, the computed harmonic frequencies of the 2'-3'-cNMPs agree well with the experimental spectra (Figure S6). The frequencies of the 3'-5'-cNMPs on the other hand match less well, especially because of additional absorptions between 1020–1060 cm$^{-1}$ in the computed spectrum, which cannot be observed in the experiment. In addition, it is noteworthy that some absorptions in the experimental spectrum are significantly broader than predicted by the harmonic frequency calculation. Anharmonic frequency calculations show that the band broadening observed for the c-fragments of ApA, GpG and UpG predominantly arises from anharmonicities (Figures 2d and S8). For c-fragments of CpG, on the other hand, the spectrum is considerably more complex and cannot be explained by anharmonicities alone.

The above results suggest an intermediate containing a five-membered phosphate moiety. To further substantiate this finding, the IR signatures of commercially available 2'-3'-cAMP and 3'-5'-cAMP were compared to the IR spectrum of the c-fragment of deprotonated ApA (gray trace) and compared to the computed anharmonic vibrational spectra of the lowest-energy structures of 2',3'-cAMP (red) and 3',5'-cAMP (blue). Free energies of the computed structures at 90 K are indicated. The experimental spectrum of the ApA c-fragment compared to those of deprotonated 2',3'- and 3',5'-cAMP. Key features of the spectra are highlighted with red and blue lines. The absorption bands are clearly diagnostic to distinguish between pentacyclic and hexacyclic phosphates.
shifted and contain an additional band at 1038 cm⁻¹, which was observed previously using IRMPD spectroscopy.⁴

Due to the relatively straightforward assignment of c-fragments from ApA based on cyclic standards, we applied the same methodology to obtain further information on the crowded IR spectrum of CpG fragments. The spectra of 2',3'- and 3',5'-cyclic cytidine monophosphate anions generated under non-activating conditions suggest that the c-fragment of CpG is also a five-membered cyclic phosphate. The absorption bands of 2',3'-cCMP anions match the spectrum of the c-fragment of CpG (Figures 3a and S11). However, some features in the fragment ion spectrum are much more intense than for 2',3'-cCMP and the spectra generally do not match as well as in the case of adenosine. To elucidate the origin of the unusual complexity in the fragment ion spectrum, we subjected the cyclic cytidine anions to activating source conditions and recorded their infrared signatures (Figure 3b). Interestingly, the spectrum of activated 2',3'-cCMP corresponds much better to the c-fragment ion spectrum than its non-activated counterpart. Certain bands are significantly more pronounced and the fragment ion spectrum than its non-activated counterpart.

In vivo, this tautomerization leads to errors in base pairing (CG → C*G) which in turn can cause mutations in the next replication cycle.²⁴ For protonated nucleobases in the gas phase, tautomerization has been observed previously using IRMPD and differential mobility spectrometry.²⁵ For the physiologically more relevant deprotonated ions, however, this process has not been studied to date.

Density functional theory (DFT) calculations reveal that the C* tautomer is destabilized by a mere 3 kJ mol⁻¹ in case of deprotonated 2',3'- and 3',5'-cyclic cytidine monophosphate (Figure 3c). However, for the 2',3'-analog, a substantial activation energy of +156 kJ mol⁻¹ has been computed. This barrier can be lowered to +47 kJ mol⁻¹ when catalyzed by a water molecule, which explains the relatively low activation energy required to induce tautomerization in the presence of water in the source region of the instrument (see Figure S12 and Table S5). The harmonic and anharmonic frequencies of tautomerized cyclic cytidine monophosphate further complement those of the non-tautomered form (Figures 3d, S13 and S14), which explains the wealth of spectral features observed for the c-fragment ions and the activated cCMPs.

Figure 3. Tautomerization of cytosine in the gas phase. The experimental infrared spectrum of the CpG c-fragment, formed under activating source conditions compared to those of deprotonated 2',3'- and 3',5'-cCMP generated at a) non-activating and b) activating source conditions. Key absorption features of 2',3'-cCMP and 3',5'-cCMP spectra are highlighted in red and blue, respectively. The absorption bands are diagnostic and reveal that the fragment ion possesses a pentacyclic structure. Some absorption bands, especially above 1400 cm⁻¹, are more complex and can only be reproduced under activating conditions which induce c) the tautomerization of the cytosine moiety. According to calculations, the free energy (90 K) difference between both tautomers is only 3 kJ mol⁻¹ for both the 2',3' and the 3',5' isomer. d) Anharmonic frequencies confirm that the coexistence of tautomerized and non-tautomerized cCMPs in the ion trap leads to significant broadening of the absorptions in the 1600–1800 cm⁻¹ region.
In summary, we show here that c-fragments in RNA tandem mass spectrometry adopt a structure, which is identical to that of the reactive intermediate occurring during RNA autohydrolysis. Comparison between the infrared signatures of c-fragments and those of synthetic cyclic nucleotides in combination with DFT calculations unambiguously identified the characteristic 2',3'-cyclic structure of the intermediate. This implies that the intramolecular cyclisation reaction of RNA nucleotides is a strongly favored process with a high intrinsic driving force. The results support the previously proposed fragmentation mechanism involving a nucleophilic attack of the 2'-OH group at the phosphate rather than alternative mechanisms proceeding via linear intermediates. Further, our experiments show that tautomerization of cytidine can be followed in the gas phase as well. This tautomerization spontaneously occurs in the source region under activating conditions and leads to more complex infrared signatures. In a broader context, our results highlight that both processes, the intramolecular cyclisation of RNA nucleotides as well as the tautomerization of cytidine can occur in full isolation. This implies that both reactions are intrinsic to the individual molecules and independent of their environment.

Acknowledgements

K.G. thanks the Fonds National de la Recherche, Luxembourg, for funding the project GlycoCat (13549747). C.K. is grateful for financial support by the Fonds der Chemischen Industrie. M.I.T. gratefully acknowledges the support of the Alexander von Humboldt Foundation. M.G. and K.P. acknowledge generous funding by the European Research Council, ERC-2019-CoG-863934-GlycoSpec. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: Autohydrolysis · Fragmentation · Infrared Spectroscopy · Mass Spectrometry · RNA

[1] J. W. Park, P. N. P. Lagnoton, Y. Liu, R. H. Xu, Int. J. Biol. Sci. 2021, 17, 1446–1460.
[2] a) N. Pardi, M. J. Hogan, F. W. Porter, D. Weissman, Nat. Rev. Drug Discovery 2018, 17, 261–279; b) C. Zhang, G. Maruggi, H. Shan, J. Li, Front. Immunol. 2019, 10, 594; c) N. A. C. Jackson, K. E. Kester, D. Casimiro, S. Gurunathan, F. DeRosa, NPJ Vaccines 2020, 5, 11.
[3] a) M. Taucher, K. Breuker, Angew. Chem. Int. Ed. 2012, 51, 11289–11292; Angew. Chem. 2012, 124, 11451–11454; b) C. Wetzel, P. A. Limbach, Analyst 2016, 141, 16–23; c) K. Thüring, K. Schmid, P. Keller, M. Helm, Methods 2016, 107, 48–56; d) R. Lauman, B. A. Garcia, Mol. Biol. Syst. 2020, 16, 305–315.
[4] L. Piabò, H. Poinar, D. Serre, V. Jaucnicke-Despres, J. Hebler, N. Rohland, M. Kueh, J. Krause, L. Vigilant, M. Hofreiter, Annu. Rev. Genet. 2004, 38, 645–679.
[5] H. Lönnberg, Org. Biomol. Chem. 2011, 9, 1687–1703.
[6] G. M. Emilsson, S. Nakamura, A. Roth, R. R. Breakey, DNA 2003, 9, 907–918.
[7] S. Schürch, Mass Spectrom. Rev. 2016, 35, 483–523.
[8] T. Y. Huang, A. Kharlamova, J. Liu, S. A. McLuckey, J. Am. Soc. Mass Spectrom. 2008, 19, 1832–1840.
[9] S. A. McLuckey, G. J. Berkel, G. L. Glish, J. Am. Soc. Mass Spectrom. 1992, 3, 60–70.
[10] a) S. Schürch, E. Bernal-Méndez, C. J. Leumann, J. Am. Soc. Mass Spectrom. 2002, 13, 936–945; b) J. M. Tromp, S. Schürch, J. Am. Soc. Mass Spectrom. 2005, 16, 1262–1268.
[11] a) T. E. Anderssen, F. Kirpekar, K. F. Haselmann, J. Am. Soc. Mass Spectrom. 2006, 17, 1353–1368; b) C. Riml, H. Glasner, M. T. Rodgers, R. Micura, K. Breuker, Nucleic Acids Res. 2015, 43, 5171–5181; c) E. Fuchs, C. Falschung, R. Micura, K. Breuker, Nucleic Acids Res. 2019, 47, 7223–7234.
[12] a) Y. W. Ni, K. T. Crampton, G. Berden, J. Oomens, M. T. Rodgers, J. Phys. Chem. A 2013, 117, 10634–10649; b) R. R. Wu, B. Yang, G. Berden, J. Oomens, M. T. Rodgers, J. Phys. Chem. B 2015, 119, 2795–2805.
[13] R. E. van Oosterj, J. Martens, G. Berden, J. D. Steill, J. Oomens, A. M. Rijs, Phys. Chem. Chem. Phys. 2018, 20, 28319–28330.
[14] B. Chiavarino, M. E. Crestoni, S. Fornarini, F. Lanucara, J. Lemaire, P. Maltaire, D. Scuderi, Int. J. Mass Spectrom. 2008, 270, 111–117.
[15] D. A. Thomas, R. Chang, E. Mucha, M. Lettow, K. Greis, S. Gewinner, W. Schöllkopf, G. Meijer, G. von Helden, Phys. Chem. Chem. Phys. 2020, 22, 18400–18413.
[16] a) E. Mucha, A. I. González Flórez, M. Marianski, D. A. Thomas, W. Hoffmann, W. B. Struve, H. S. Hahn, S. Gewinner, W. Schöllkopf, P. H. Seeberger, G. von Helden, K. Pagel, Angew. Chem. Int. Ed. 2017, 56, 11248–11251; Angew. Chem. 2017, 129, 11400–11404; b) K. Greis, C. Kirschbaum, S. Leichnitz, S. Gewinner, W. Schöllkopf, G. von Helden, G. Meijer, P. H. Seeberger, K. Pagel, Org. Lett. 2020, 22, 8916–8919; c) C. Kirschbaum, K. Greis, L. Polewski, S. Gewinner, W. Schöllkopf, G. Meijer, G. von Helden, K. Pagel, J. Am. Chem. Soc. 2021, 143, 14827–14834.
[17] U. Even, EPJ Tech. Instrum. 2015, 2, 17.
[18] W. Schöllkopf, S. Gewinner, H. Junkes, A. Paarmann, G. von Helden, H. P. Blume, A. M. M. Todd, Proc. SPIE-Int. Soc. Opt. Eng. 2015, 9512, 95121 L.
[19] P. Pracht, F. Bohle, S. Grimmer, Phys. Chem. Chem. Phys. 2020, 22, 7169–7192.
[20] C. Bannworth, S. Ehlerst, S. Grimmer, J. Chem. Theory Comput. 2019, 15, 1652–1671.
[21] a) J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 1996, 77, 3865–3868; b) C. Adamo, V. Barone, J. Chem. Phys. 1999, 110, 6158–6170; c) S. Grimmer, J. Antony, S. Ehrlich, H. Krieg, J. Chem. Phys. 2010, 132, 154104; d) F. Weigend, R. Ahlrichs, Phys. Chem. Chem. Phys. 2005, 7, 3297–3305.
[22] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. J. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Yang, F. Ding, F. Lippertini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G.
