Isoforms of Carbohydrates Recognized by Interplay of Non-covalent Bonds with Aromatics.

Erik Saparbaev[1], Vladimir Kopysov[1], Ruslan Yamaletdinov[2], Aleksandr Y. Pereverzev[1], Oleg V. Boyarkin[1]*

[1] Laboratoire de Chimie Physique Moléculaire, École Polytechnique Fédérale de Lausanne, Station-6, 1015 Lausanne, Switzerland.

[2] Nikolaev Institute of Inorganic Chemistry, Novosibirsk, 630090, Russian Federation.

Supporting Information Placeholder

ABSTRACT: Despite the tremendous isomeric diversity of carbohydrates, their non-covalent bindings to proteins in nature are surprisingly selective. We investigated the ability of these intermolecular interactions to communicate structural specificity of binding partners in carbohydrate–aromatic ion complexes, which were formed in solution and then isolated and cooled in the gas phase. Our study revealed that small structural differences between carbohydrate isoforms of any type, including enantiomers, are accurately communicated by these interactions back to aromatic molecules as detectable changes in their vibronic energy levels. Such specific responses of the aromatics to isoforms of carbohydrates are fine-tuned by interplay of the diverse involved non-covalent bonds. These findings enable the identification and relative quantification of any isoforms of carbohydrates in their mixtures.

Non-covalent intermolecular interactions are ubiquitous in nature. Their diversity allows highly selective binding between biomolecules. DNA strands self-assemble precisely to form a double helix, an antibody attaches to a target antigen found among myriads of other molecules, viruses and bacteria invade appropriate cells by, first, binding to specific membrane carbohydrates,[1] *etc*. The selectivity of non-covalent interactions with carbohydrates in particular was explained by NMR[2] and X-ray crystallography[3] studies within a general concept of "lock and key" model. This model implies complementarity of the interacting surfaces that the binding partners should expose to each other for maximum stability of a complex.[4] Little is known about how sensitive the intermolecular non-covalent bonds are in communicating structural details of the partners to each other.[5-7] Will, for instance, a flip of a single bond in a large isomeric oligosaccharide change something observable in a bound partner? Herein we demonstrate how a network of non-covalent bonds between an aromatic host ion and a carbohydrate guest molecule enables specific bilateral responses of the complex to structural differences in isomeric guests. The responses, detected by a combination of gas-phase UV spectroscopy and mass spectrometry reveal key contributions of the non-covalent bonds in communicating structural differences of guests. We explore the use of such networks for the identification and quantification of all types of isoforms of carbohydrates non-covalently bound in solution to protonated aromatic host molecules.

The monomeric structural units of carbohydrates exist in a variety of stereoisomeric forms, such as *D/L*-enantiomers, epimers, and α/β-anomers. For instance, a cyclic aldohexose has 5 stereogenic centers, which implies an existence of 2^5=32 stereoisomers (e.g., α-*D*-glucose, β-*L*-galactose, etc.), although not all of them are essential in nature. In contrast to amino acids and nucleotides, these isomeric units can interconnect through different groups, forming regioisomers, but also at multiple points, assembling into linear and branched structural isomers. Natural modifications (e.g., N-acetylation) of different units further multiply the number of possible isoforms. This enormous isomeric diversity makes carbohydrates efficient at communicating the

specificity of interacting biomolecules, but exceptionally challenging to study glycans experimentally.

In proteins, aromatic residues are by far the most common sites that accommodate carbohydrates by non-covalent binding.[8] UV absorption of aromatic amino acids is highly sensitive to non-covalent interactions of their π-electrons,[9] while carbohydrates are, essentially, transparent in this spectral range. These two facts make UV photofragmentation spectroscopy of aromatic–carbohydrate non-covalent complexes a convenient tool for detecting the response of the intermolecular non-covalent interactions to structural differences of isomeric carbohydrates. In solution, where such complexes are natively prepared, their spectroscopic details are masked by non-covalent interactions of carbohydrates with the infinite number of solvent molecules. Isolation of the protonated complexes in the gas phase eliminates the inhomogeneous broadening induced by these interactions and permits a use of mass spectrometry for sensitive detection of photofragments. This also enables cryogenic cooling of the ions, which greatly enhances the resolution in UV fragmentation spectra.[10-11]

Figure 1A shows a two-dimensional color-coded photofragmentation mass spectrum (2D UV-MS;[11] fragment intensity *vs* wavelength and m/z) of N-acetyl-*D*-glucosamine (GlcNAc) bound to protonated tyrosine. To illustrate the difference between the 2D UV-MS identities of the complexes of $TyrH^+$ with different isomeric monosaccharides, Figures 1B and 1C show pairs of photofragment UV and mass spectra generated from the measured 2D UV-MS of GalNAc-$TyrH^+$ and GlcNAc-$TyrH^+$ complexes. The UV spectra of the complexes differ significantly from each other, in particular in the positions of their electronic band origins. This difference alone makes the two epimers readily distinguishable by spectroscopy of their complexes with $TyrH^+$. Figures 1D and 1E reveal the origin of the spectral difference between the complexes with the two isomers. The figures show the calculated most stable gas-phase structures of the complexes and their networks of non-covalent bonds, which are typical for the other computed low-energy conformers. In both isomers, the two bonds between NH_3^+ of $TyrH^+$ and oxygen atoms of a saccharide allow the complex to survive the harsh conditions of electrospray ionization. Similar pattern of NH_3^+–O interactions was revealed for protonated N-terminus of Tyr residue in some peptides of known geometry

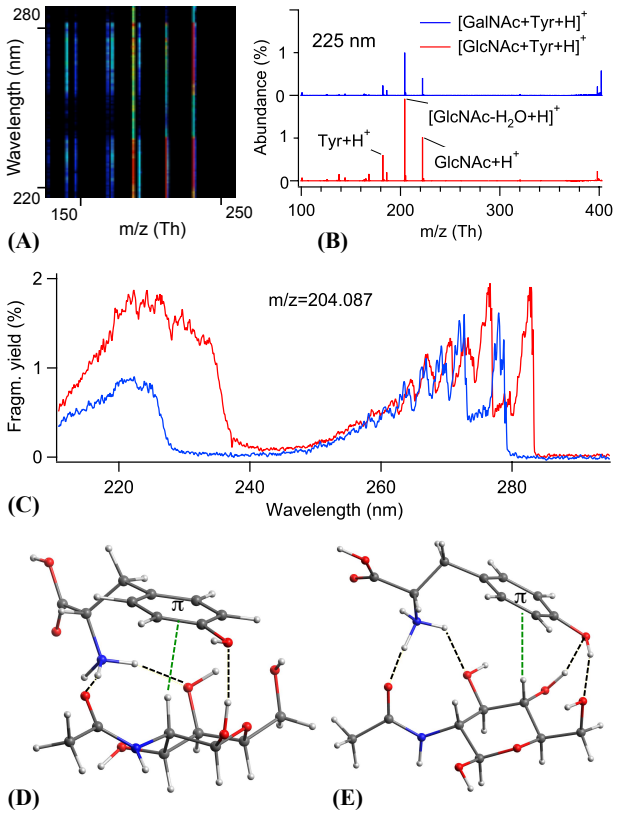

Figure 1. (**A**) Color-coded 2D UV-MS spectrum of GlcNAc–$TyrH^+$ complex cooled to ~10 K [12]. Examples of mass (**B**) UV and photofragment (**C**) spectra, generated by slicing 2D UV-MS fingerprints of GalNAc-$TyrH^+$ (blue trace) and of GlcNAc-$TyrH^+$ (red trace) at m/z=204.087 Th (most intense fragment) and at 225 nm, respectively. The calculated most stable structures of (**D**) GalNAc-$TyrH^+$ and (**E**) GlcNAc-$TyrH^+$ complexes; black and green dashed lines show hydrogen and CH-π bonds, respectively.

(Fig. S1).[13-14] Solvation of the charge by these interactions suppresses the proton–π coupling in the $TyrH^+$ residues, such that their UV band origins (280.3 nm)[15] appear almost at that of neutral Tyr (280.9 nm),[16] whose aromatic ring is free of any non-covalent interactions. Similarly, proton-π interaction should be insignificant in the two complexes and cannot account for the observed large difference in position of their UV band origins.

Structural information is communicated to the aromatics, first, through the hydrogen bonds between the OH group of Tyr side chain and the monosaccharaides. When phenol, which represents the hydroxylphenyl of tyrosine, donates the proton of hydroxyl, its electronic band origin redshifts by as much as ~2.7 nm due to inductive polarization of the π-system.[17-18] Conversely, when the hydroxyl oxygen is the proton acceptor, like in the case of $TyrH^+$–GalNAc, its increased electronegativity near equally shifts the UV absorption of phenol to the blue.[17, 19] Involvement of phenol to both types of H-bonds, like in the case of $TyrH^+$–GlcNAc, redshifts its absorption by ~0.8 nm.[18] Consistently with this picture, the UV band origin in $TyrH^+$–GalNAc complex is shifted from that in $TyrH^+$–GlcNAc to the blue by as much as 4.2 nm. In addition to these two strong H-bonds, both complexes exhibit weaker CH-π interactions. Their strength is governed by the electropositive character of the C−H bonds, which differs between isomeric saccharides.[8] Thus, CH-π bonds may also contribute to the recognition of the isomeric monosaccharaides. To disentangle the strong and weak non-covalent couplings in the complexes, we replaced OH on the Tyr ring with a methyl group (MePhe). This modification removes the H-bonds, but enables coupling of an OH in the saccharides to the aromatic π-system (Fig. S2). CH-π and OH-π bonds now compete in shifting the UV absorption of $MePheH^+$ to the red and to the blue, respectively. Regarding the interatomic distances and bond directions in the calculated GalNAc/GlcNAc–$MePheH^+$ complexes (Fig. S2), OH-π coupling should be particularly prominent in the complexes with GlcNAc, while CH-π is relatively strong in the complex with GalNAc. Consistently, the position of the band origin in the former complex is on the blue side relative to that in the latter complex (Fig. S3). The interplay of OH-π and CH-π couplings thus makes the UV spectroscopic signatures of the two epimers quite different. Similarly, these interactions in the complexes of *L*-$PheH^+$ with *D*- and *L*- enantiomers of glucose modify the UV absorption of this chiral aromatic differently (Figure S4), allowing spectroscopic recognition of the two optically isomeric monosaccharaides.

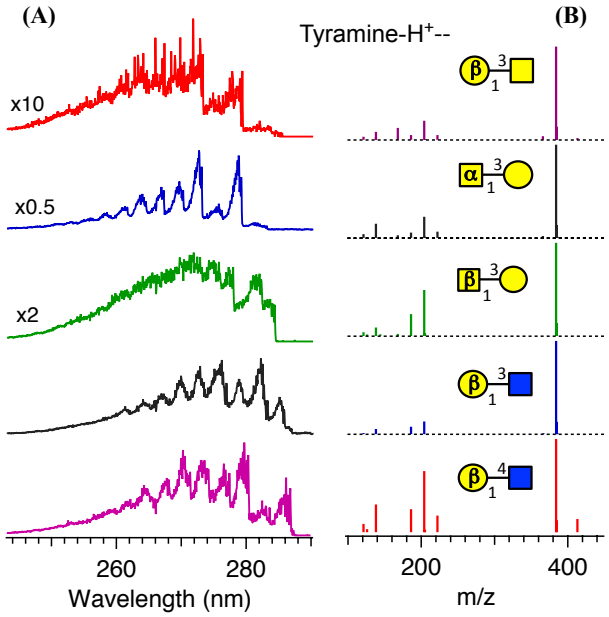

Figure 2. Photofragment (A) UV and (B) mass spectra, generated by integrating over m/z and wavelength, respectively, the 2D UV-MS fingerprints of five isomeric (m=383.16 Da) N-acetylated disaccharides bound in solution with protonated tyramine. The spectra are labeled by the standard symbolic representation of oligosaccharides.[20]

Overall, the interplay of up to five non-covalent interactions between the aromatics and the monosaccharaides determines the net position and structure of the UV absorption of the complexes. Structural alterations of a saccharide change the subtle balance of these interactions, making UV absorption of the complexes exquisitely sensitive to isoforms of carbohydrates (see also Figures 2A, 3A, S4 and S5). In all the calculated structures (e.g., Figs. 1D, 1E and S2), C-terminus of Tyr and MePhe is not involved in H-bonding and therefore may not contribute to this sensitivity. Figure 2 shows fragmentation UV and mass spectra of five N-acetylated isomeric disaccharides bound in solution with tyramine – an aromatic molecule that is similar in chemical structure to Tyr, but has no C-terminus. The disaccharides differ in anomeric and epimeric forms of the units, but also in the connectivity points and in the acetylation sites. The spectra appear visibly different for all the isomers in complexes with tyramine, which confirms that non-covalent bonds with other aromatics besides four amino acids can also be highly sensitive in the recognition of carbohydrates.

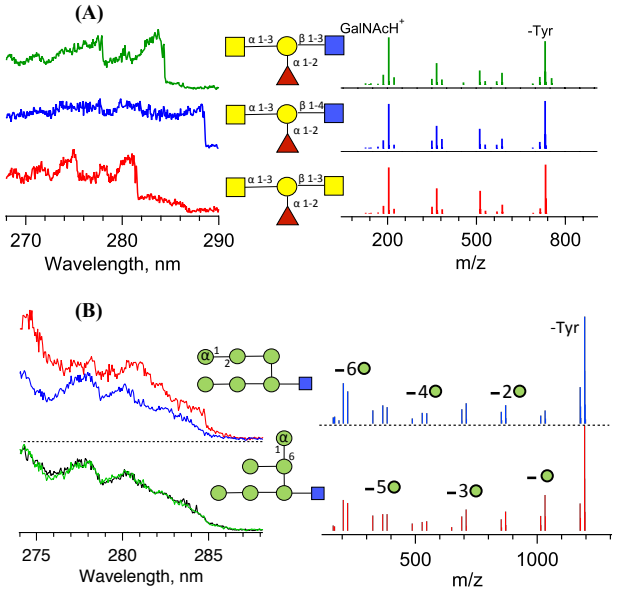

Figure 3. Photofragment UV and mass spectra of (**A**) three isomeric tetrasaccharides, and (**B**) two heptasaccharides ($Man_6GlcNAc$ (II) and $Man_6GlcNAc$ (I), red and blue traces, respectively), bound to $TyrH^+$. Green trace on the left in (**B**) shows the UV spectrum of the 1:2 mixture of the two heptasaccharides, and the overlapped black trace in the same panel is the sum of the red and blue traces taken with the calculated coefficients of 0.29 and 0.71, respectively. The pairs of m/z peaks in (**B**) are labeled by the symbol and number of the residues lost by the heptasaccharides.

Although aromatics, but not carbohydrates, are efficiently protonated in solution, the highly abundant protonated saccharides and their fragments appear in the photofragment MS of the complexes (Figures 1B, 2B and 3B). These observations suggest a proton and energy transfer from aromatics to saccharides in the electronic ground[21] and/or excited state[22] of the complexes. Protonated aromatics are the most abundant photofragments observed for mono- (e.g., *D*-/*L*- glucose, Fig. S4) and disaccharaides (Figure S5) that do not contain NAc groups. This suggests a crucial role for the NAc group in accommodating the proton that is shared with aromatics in the complexes. The rich networks of intermolecular non-covalent bonds in the complexes facilitate the host to guest proton transfer[23], thus allowing a synergy of mass spectrometry and spectroscopy in recognition of carbohydrates.

Aromatic molecules, whose size is comparable with that of a monosaccharide, can sense the structure of a large oligosaccharide only locally near their binding sites. Nevertheless, the 2D UV-MS fingerprints of the complexes with all isomeric carbohydrates studied herein appear visibly distinct. Figures 3A and 3B compare the UV and MS identities of three isomeric tetrasaccharides and two heptasaccharides bound to $TyrH^+$. The oligosaccharides differ in the connectivity points, as well as in the isoforms of the units and in the branching for the tetra- and hepta-saccharides, respectively. We explain the observed difference in the UV-MS fingerprints by the presence of many different binding sites for an aromatic on a large carbohydrate. Although each single aromatic host can sense a structure only locally, different conformers of a complex collectively may reflect the structural details of the whole guest molecule.

The observed differences in the UV-MS identities of isomeric carbohydrates complexed with aromatics allowed for relative quantification of the isomers in their solution mixtures using the methods of 2D data array decompositions.[11, 24] For instance, for a 1:1:2 solution mixture of three isomeric tetrasaccharides (Fig. 3A), the 27:24:49% relative concentrations were determined; the average relative concentrations of 29:71% were calculated for a 1:2 solution mixture of the two heptasaccharides shown in Figure 3B. Most common methods of isomeric identifications of carbohydrates, such as chromatography and ion mobility,[25-26] rely on rates of hindered travel of molecules and ions, respectively. These rates are not fundamental values however and are sensitive to the experimental conditions. Spectroscopy allows identification of molecules on a quantum level, giving fundamental values that can be measured and compared between laboratories with high accuracy. Vibrational spectroscopy (e.g., IRMPD[27] or IR fragmentation of tagged saccharides[28-30]) was able to distinguish without quantification some isoforms of small oligosaccharides. Structural sensitivity of intermolecular non-covalent bonds in aromatic–carbohydrate complexes enables UV-MS recognition of *all* types of isoforms of mono- to heptasaccharides. This sensitivity arises from interplay of the diverse involved non-covalent bonds. Whether in nature the sensitivity of electronic levels of aromatics in proteins to isoforms of the bound carbohydrates contributes to recognition of glycans and to signal transduction in proteins remains an open question.

ASSOCIATED CONTENT

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website at DOI

Materials and method, computational details, additional figures and references (PDF).

AUTHOR INFORMATION

Corresponding Author

*oleg.boiarkin@epfl.ch

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank V. Aladinskaia and E. Solov'eva for assisting in some measurements. We thank Novartis AG (FreeNovation program), Swiss National Science Foundation (grants 206021_164101, 200020_172522 /1) for their support of this work.

REFERENCES

(1) Sharon, N.; Lis, H., Lectins as cell recognition molecules. *Science* **1989,** *246,* 227-234.

(2) Jiménez-Barbero, J.; Peters, T., TR-NOE Experiments to Study Carbohydrate-Protein Interactions. In *NMR Spectroscopy of Glycoconjugates*, Wiley: Weinheim, **2003**, 289-310.

(3) Heightman, T. D.; Vasella, A. T., Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases. *Angew. Chem. Int. Ed.* **1999,** *38*, 750-770.

(4) Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894,** *27*, 2985-2993.

(5) Cocinero, E. J.; Carcabal, P.; Vaden, T. D.; Simons, J. P.; Davis, B. G., Sensing the anomeric effect in a solvent-free environment. *Nature* **2011,** *469*, 76-9.

(6) Garand, E.; Kamrath, M. Z.; Jordan, P. A.; Wolk, A. B.; Leavitt, C. M.; McCoy, A. B.; Miller, S. J.; Johnson, M. A., Determination of Noncovalent Docking by Infrared Spectroscopy of Cold Gas-Phase Complexes. *Science* **2012,** *335*, 694-698.

(7) Sekiguchi, T.; Tamura, M.; Oba, H.; Carcarbal, P.; Lozada-Garcia, R. R.; Zehnacker-Rentien, A.; Gregoire, G.; Ishiuchi, S. I.; Fujii, M., Molecular Recognition by a Short Partial Peptide of the Adrenergic Receptor: A Bottom-Up Approach. *Angew. Chem. Int. Edit.* **2018,** *57*, 5626-5629.

(8) Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N., Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015,** *137*, 15152-60.

(9) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V., Interplay of Intra- and Intermolecular H-Bonding in a Progressively Solvated Macrocyclic Peptide. *Science* **2012,** *336*, 320-323.

(10) Boyarkin, O. V.; Mercier, S. R.; Kamariotis, A.; Rizzo, T. R., Electronic spectroscopy of cold, protonated tryptophan and tyrosine. *J. Am. Chem. Soc.* **2006,** *128*, 2816-2817.

(11) Kopysov, V.; Makarov, A.; Boyarkin, O. V., Colors for molecular masses: fusion of spectroscopy and mass spectrometry for identification of biomolecules. *Anal. Chem.* **2015,** *87*, 4607-11.

(12) Boyarkin, O. V.; Kopysov, V., Cryogenically cooled octupole ion trap for spectroscopy of biomolecular ions. *Rev. Sci. Instrum.* **2014,** *85*, 033105.

(13) Burke, N. L.; Redwine, J. G.; Dean, J. C.; McLuckey, S. A.; Zwier, T. S., UV and IR spectroscopy of cold protonated leucine enkephalin. *Int. J. Mass Spectrom.* **2015,** *378*, 196-205.

(14) Roy, T. K.; Kopysov, V.; Pereverzev, A.; Šebek, J.; Gerber, R. B.; Boyarkin, O. V., Intrinsic structure of pentapeptide Leu-enkephalin: geometry optimization and validation by comparison of VSCF-PT2 calculations with cold ion spectroscopy. *Phys. Chem. Chem. Phys.* **2018,** *20*, 24894-24901.

(15) Kopysov, V.; Boyarkin, O. V., Resonance Energy Transfer Relates the Gas-Phase Structure and Pharmacological Activity of Opioid Peptides. *Angew. Chem. Int. Ed.* **2016,** *55*, 689-692.

(16) Abo-Riziq, A.; Grace, L.; Crews, B.; Callahan, M. P.; van Mourik, T.; Vries, M. S. d., Conformational Structure of Tyrosine, Tyrosyl-glycine, and Tyrosyl-glycyl-glycine by Double Resonance Spectroscopy. *J. Phys. Chem. A* **2011,** *115*, 6077-6087.

(17) Fuke, K.; Kaya, K., Electronic Absorption-Spectra of Phenol-$(H2O)_N$ and $(Phenol)_N$ as Studied by the Ms Mpi Method. *Chem. Phys. Lett.* **1983,** *94*, 97-101.

(18) Ebata, T.; Fujii, A.; Mikami, N., Vibrational spectroscopy of small-sized hydrogen-bonded clusters and their ions. *Int. Rev. Phys. Chem.* **1998,** *17*, 331-361.

(19) Schmitt, M.; Bohm, M.; Ratzer, C.; Krugler, D.; Kleinermanns, K.; Kalkman, I.; Berden, G.; Meerts, W. L., Determining the intermolecular structure in the S0 and S1 states of the phenol dimer by rotationally resolved electronic spectroscopy. *Chemphyschem* **2006,** *7*, 1241-9.

(20) Varki, A. *et al.*, Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **2015,** *25*, 1323-1324.

(21) Vrkic, A. K.; O'Hair, R. A. J., Using non-covalent complexes to direct the fragmentation of glycosidic bonds in the gas phase† ††Gas Phase Ion Chemistry of Biomolecules, Part 39. *J. Am. Soc. Mass. Spectr.* **2004,** *15*, 715-724.

(22) Agmon, N., Elementary Steps in Excited-State Proton Transfer. *J. Phys. Chem. A* **2005,** *109*, 13-35.

(23) Horke, D. A.; Watts, H. M.; Smith, A. D.; Jager, E.; Springate, E.; Alexander, O.; Cacho, C.; Chapman, R. T.; Minns, R. S., Hydrogen Bonds in Excited State Proton Transfer. *Phys. Rev. Lett.* **2016,** *117*, 163002.

(24) Kopysov, V.; Makarov, A.; Boyarkin, O. V., Nonstatistical UV Fragmentation of Gas-Phase Peptides

Reveals Conformers and Their Structural Features. *J. Phys. Chem. Lett.* **2016,** *7*, 1067-1071.

(25) Lee, D.-S.; Wu, C.; Hill, H. H., Detection of carbohydrates by electrospray ionization-ion mobility spectrometry following microbore high-performance liquid chromatography. *J. Chromatogr. A* **1998,** *822*, 1-9.

(26) Gray, C. J.; Thomas, B.; Upton, R.; Migas, L. G.; Eyers, C. E.; Barran, P. E.; Flitsch, S. L., Applications of ion mobility mass spectrometry for high throughput, high resolution glycan analysis. *BBA-Gen. Subjects* **2016,** *1860*, 1688-1709.

(27) Polfer, N. C.; Valle, J. J.; Moore, D. T.; Oomens, J.; Eyler, J. R.; Bendiak, B., Differentiation of isomers by wavelength-tunable infrared multiple-photon dissociation-mass spectrometry: Application to glucose-containing disaccharides. *Anal. Chem.* **2006,** *78*, 670-679.

(28) Stanca-Kaposta, E. C.; Carcabal, P.; Cocinero, E. J.; Hurtado, P.; Simons, J. P., Carbohydrate-aromatic interactions: vibrational spectroscopy and structural assignment of isolated monosaccharide complexes with p-hydroxy toluene and N-acetyl l-tyrosine methylamide. *J. Phys. Chem. B* **2013,** *117*, 8135-42.

(29) Mucha, E.; Florez, A. I. G.; Marianski, M.; Thomas, D. A.; Hoffmann, W.; Struwe, W. B.; Hahm, H. S.; Gewinner, S.; Schollkopf, W.; Seeberger, P. H.; von Helden, G.; Pagel, K., Glycan Fingerprinting via Cold-Ion Infrared Spectroscopy. *Angew. Chem. Int. Edit.* **2017,** *56*, 11248-11251.

(30) Khanal, N.; Masellis, C.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R., Glycosaminoglycan Analysis by Cryogenic Messenger-Tagging IR Spectroscopy Combined with IMS-MS. *Anal. Chem.* **2017,** *89*, 7601-7606.