

Structural and analytical identifications of biomolecules by cold ion spectroscopy of non-covalent complexes

Présentée le 12 octobre 2021

Faculté des sciences de base
Laboratoire de chimie physique moléculaire
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Erik SAPARBAYEV

Acceptée sur proposition du jury

Prof. R. Beck, président du jury
Dr O. Boyarkine, directeur de thèse
Prof. J. Roithova, rapporteuse
Dr E. Concinero, rapporteur
Dr S. Sekatskii, rapporteur

ABSTRACT

Non-covalent interactions are ubiquitous in nature. The diversity and functionality of these interactions are employed by nature as a universal tool for building the molecular mechanism of life. Although weak, these interactions play a key role in both formation of 3D structures of biomolecules through intra- and intermolecular binding and highly selective coupling between biomolecules. In this thesis, we explore these interactions in two directions: (i) for solving native 3D structures of biomolecules and (ii) for analytical identifications of isomeric carbohydrates. Both studies use the technique of cold ion spectroscopy (CIS), which often enables vibrational resolution in infrared and electronic spectra of biomolecules isolated in the gas. Although the resolved spectra can be used in solving the intrinsic molecular 3D structures of biomolecules, life science research desperately needs their native structures. It is challenging however to get vibrational resolution in solution phase, where these structures originally reside. As a compromise, we do not fully dehydrate the biomolecules during their gentle electrospray ionization (ESI) and leave a few water molecules attached to the ion. The structure of such microhydrated molecule may resemble its native one in solution while working in the gas phase enables the desired vibrational spectral resolution.

The second direction of the study aims to explore the sensitivity of intermolecular noncovalent interactions in biomolecular complexes to fine structural details of the partners. The work demonstrates how this sensitivity can be used for analytical identifications of the biomolecules that are highly rich in isomers.

In the first part of this work, we report on the elucidation of native structures of small protonated biomolecules glycine and triglycine. We apply cold ion IR spectroscopy to the gas-phase complexes of these biomolecules, in which they are microhydrated by a controlled number of

water molecules. The complexes were produced directly from an aqueous solution, using a gentle ESI source. Our studies suggest that retaining molecules in the complexes generated from the solution allows for preserving the main structural features of the native structures.

In the second part of the thesis, we report on the investigation of the sensitivity of non-covalent interaction between aromatic molecules and carbohydrates to structural details of binding partners. We calculated structures of non-covalent complexes of aromatic molecules with isomeric monosaccharides and validated them by IR spectroscopy. It appears, that the change of an analyzing isomeric carbohydrate molecule significantly influences the UV absorption of the aromatic sensor molecule. The spectral difference between the complexes is explained by the interplay of different intermolecular non-covalent bonds, which may not be all identical for different isomers. Even structural changes in the groups that are not directly involved in non-covalent binding with chromophore exhibit a significant impact on UV spectra.

In the last part of the thesis, we explore the use of non-covalent carbohydrate-aromatic complexes for the identification of isomeric glycans using a unique method of 2D UV-MS CIS. We demonstrated the applicability of this method for accurate identification and quantification of all types of isomeric carbohydrates and tested the performance of the technically simpler version of the method, 1D UV fragmentation spectroscopy.

Keywords: noncovalent interactions, cold ion spectroscopy, high-resolution mass spectrometry, ultraviolet spectroscopy, infrared spectroscopy, microsolvation, isomers, anomers, glycans, carbohydrates.

RÉSUMÉ

Les interactions non covalentes (INC) sont omniprésentes dans la nature. Bien que faibles, ces interactions jouent un rôle clé à la fois dans la formation de structures 3D de biomolécules par liaison intra- et intermoléculaire. Dans cette thèse, nous explorons ces interactions dans deux directions: (i) pour résoudre des structures 3D natives de biomolécules et (ii) pour des identifications analytiques de glucides isomériques. Les deux études utilisent la technique de spectroscopie d'ions froids (CIS), qui permet souvent une résolution vibrationnelle dans les spectres IR et UV des biomolécules isolées dans le gaz. Bien que les spectres résolus puissent être utilisés pour résoudre les structures moléculaires 3D intrinsèques des biomolécules, la recherche en sciences de la vie a besoin de leurs structures natives. Cependant, il est difficile d'obtenir une résolution vibrationnelle dans la phase de solution d'origine, où résident ces structures. En guise de compromis, nous ne déshydraterons pas complètement les biomolécules lors de leur ionisation douce par électrospray (ESI) et laissons quelques molécules d'eau attachées à l'ion. La structure d'une telle molécule microhydratée peut ressembler à sa molécule native en solution.

La deuxième direction de l'étude vise à explorer la sensibilité des INC intermoléculaires dans les complexes biomoléculaires aux détails structuraux fins des partenaires. Les travaux montrent comment cette sensibilité peut être utilisée pour des identifications de biomolécules isomériques.

Dans la première partie de ce travail, nous rapportons l'élucidation des structures natives de petites biomolécules protonées glycine et triglycine. Nous appliquons la spectroscopie IR à ions froids aux complexes en phase gazeuse de ces biomolécules, dans lesquels elles sont microhydratées par un nombre contrôlé de molécules. Les complexes ont été produits directement à partir d'une solution aqueuse, en utilisant une

source ESI douce. Nos études suggèrent que la rétention de molécules dans les complexes générés à partir de la solution permet de préserver les principales caractéristiques structurales des structures natives.

Dans la deuxième partie de la thèse, nous rapportons l'étude de la sensibilité de l'INC entre les molécules aromatiques et les glucides aux détails structuraux des partenaires de liaison. Nous avons calculé les structures de complexes non covalents de molécules aromatiques avec des monosaccharides isomériques et les avons validées par spectroscopie IR. Il semble que le changement d'une molécule de glucide isomérique invite l'influence de manière significative l'absorption UV du capteur hôte. La différence spectrale entre les complexes s'explique par l'interaction de différentes liaisons intermoléculaires non covalentes, qui peuvent ne pas être toutes identiques pour différents isomères. Même les changements structuraux dans les groupes qui ne sont pas directement impliqués dans la liaison non covalente avec le chromophore présentent un impact significatif sur les spectres UV.

Dans la dernière partie de la thèse, nous explorons l'utilisation de complexes glucides-aromatiques non covalents pour l'identification de glycanes isomériques en utilisant une méthode unique de CIS UV-MS 2D. Nous avons démontré l'applicabilité de cette méthode pour une identification de tous les types des glucides isomériques et testé les performances de la version techniquement plus simple de la méthode, la spectroscopie de fragmentation UV 1D.

Mots clés: interactions non covalentes, spectroscopie d'ions froids, spectrométrie de masse haute résolution, spectroscopie ultraviolette, spectroscopie infrarouge, microsolvation, isomères, anomères, glycanes, glucides.

KURZFASSUNG

Nicht-kovalente Wechselwirkungen (NKW) sind in der Natur allgegenwärtig. Obwohl schwach, spielen diese Wechselwirkungen eine Schlüsselrolle bei der Bildung von 3D-Strukturen von Biomolekülen durch intra- und intermolekulare Bindung. In der vorliegenden Dissertation untersuchen wir diese NKW mit zwei Zielen: (i) die Auflösung native 3D-Strukturen von Biomolekülen und (ii) die analytische Identifizierung isomerer Glykane. Beide Studien verwenden die Technik der Kaltionenspektroskopie (KIS), die oft eine Schwingungsauflösung in IR- und UV-spektren von im Gas isolierten Biomolekülen ermöglicht. Obwohl die aufgelösten Spektren zur Aufklärung der intrinsischen molekularen 3D-Strukturen von Biomolekülen verwendet werden können, benötigt die Life-Science-Forschung ihre nativen Strukturen. Es ist jedoch eine Herausforderung, eine Schwingungsauflösung in der Lösungsphase zu erhalten, in der sich diese Strukturen ursprünglich befinden. Als Kompromiss dehydrieren wir die Biomoleküle während ihrer sanften Elektrospray-Ionisation (ESI) nicht vollständig, so dass einige Wassermoleküle am Ion haften bleiben. Die Struktur eines solchen mikrohydratisierten Moleküls kann seine nativen Struktur in Lösung ähneln. Die zweite Richtung der Studie besteht darin, die Empfindlichkeit intermolekularer NKW in biomolekularen Komplexen gegenüber feinen Strukturdetails der Partner zu untersuchen. Die Arbeit demonstriert die Nutzung dieser Sensitivität für analytische Identifizierungen der Biomoleküle, die sehr reich an Isomeren sind.

Im ersten Teil dieser Arbeit berichten wir über die Aufklärung nativer Strukturen der kleinen protonierten Biomoleküle Glycin und Triglycin. Wir wenden Kaltionen-IR-Spektroskopie auf die Gasphasenkomplexe dieser Biomoleküle an, in denen sie durch eine kontrollierte Anzahl von Wassermolekülen mikrohydratisiert werden. Die Komplexe wurden direkt aus einer wässrigen Lösung unter Verwendung einer sanften ESI-Quelle

hergestellt. Unsere Studien legen nahe, dass das Zurückhalten von mikrohydratisierten Komplexen aus der Lösung die Erhaltung der wichtigsten Strukturmerkmale der nativen Strukturen ermöglicht.

Im zweiten Teil der Dissertation berichten wir über die Untersuchung der Sensitivität der NKW zwischen aromatischen Molekülen und Glykane auf strukturelle Merkmale von Bindungspartnern. Wir berechneten und validierten Strukturen von nicht-kovalenten Komplexen aromatischer Moleküle mit isomeren Monosacchariden durch IR-Spektroskopie. Anhand der Strukturen zeigten wir, dass eine Veränderung eines isomeren Glykanemoleküls eines Gastes die UV-Absorption des aromatischen Rings signifikant beeinflusst. Der spektrale Unterschied zwischen den Komplexen wird durch das Zusammenspiel verschiedener intermolekularer KNW erklärt, die für die verschiedenen Isomere möglicherweise nicht alle identisch sind. Sogar strukturelle Veränderungen in den Gruppen, die nicht direkt an der NKW mit dem Chromophor beteiligt sind, zeigen einen signifikanten Einfluss auf die UV-Spektren.

Im letzten Teil der Dissertation untersuchen wir die Verwendung nicht-kovalenter Kohlenhydrat-aromatischer Komplexe zur Identifizierung isomerer Glykane mit einer einzigartigen 2D UV-MS KIS Methode. Wir demonstrierten die Anwendbarkeit dieser Methode zur genauen Identifizierung und Quantifizierung aller Arten isomerer Glykane und testeten die Leistung der technisch einfacheren Version der 1D-UV-Fragmentierungsspektroskopie Methode.

Stichwörter: nichtkovalente Wechselwirkungen, Kaltionenspektroskopie, hochauflösende Massenspektrometrie, Ultraviolettsspektroskopie, Infrarotspektroskopie, Mikrosolvation, Isomere, Anomere, Glykane, Kohlenhydrate

LIST OF ABBREVIATIONS AND ACRONYMS

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
B3LYP	Becke-3-Lee-Yang-Parr hybrid functional
BBO	beta barium borate
DC	direct current
DFT	density functional theory
DNA	deoxyribonucleic acid
ESI	electrospray ionization
FAIMS	field asymmetric ion mobility spectrometry
Gal	galactose
GalNAc	N-acetylgalactosamine
GC	gas chromatography
Glc	glucose
GlcNAc	N-acetylglucosamine
Gly	glycine
H-bond	hydrogen bond
HPLC	high-performance liquid chromatography
IMS	ion mobility spectrometry
IR	infrared
IRMPD	infrared multiple photon dissociation
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionization
Man	mannose
MD	molecular dynamics
α/β -MeGalNAc	α/β -methyl-N-acetylgalactosamine
MePhe	4-methyl-phenylalanine
MS	mass spectrometry
NAc	N-acetyl

YAG	yttrium aluminum garnet
NMR	nuclear magnetic resonance
OPO	optical parametric oscillator
Phe	phenylalanine
QMS	quadrupole mass spectrometer
RF	radio frequency
RMSD	root-mean-square deviation
RNA	ribonucleic acid
Trm	tyramine
Tyr	tyrosine
UV	ultraviolet
UVPD	ultraviolet photodissociation

TABLE OF CONTENT

ABSTRACT	1
RÉSUMÉ	3
KURZFASSUNG	5
LIST OF ABBREVIATIONS AND ACRONYMS	7
TABLE OF CONTENT	9
CHAPTER 1. Introduction	11
<i>1.1. Introduction</i>	11
<i>References</i>	22
CHAPTER 2. Experimental Setup	30
<i>2.1. Instrument Overview</i>	30
<i>2.2. RF ion funnel and skimmers</i>	33
<i>2.3. Timing of events</i>	35
<i>2.4. Optical setup</i>	37
<i>References</i>	39
CHAPTER 3. Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy	40
<i>3.1. Introduction</i>	41
<i>3.2. Results and discussion</i>	42
<i>3.3. Conclusion</i>	48
<i>References</i>	50
<i>Appendix</i>	53
CHAPTER 4. Interplay of H-Bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates	57
<i>4.1. Introduction</i>	58
<i>4.2. Results and discussion</i>	60
<i>4.3. Conclusion</i>	64
<i>References</i>	68
<i>Appendix</i>	71

CHAPTER 5. Revealing Single-bond Anomeric Selectivity in Carbohydrate-Protein Interactions	85
5.1. <i>Introduction</i>	86
5.2. <i>Results and discussion</i>	87
5.3. <i>Conclusion</i>	94
<i>References</i>	95
<i>Appendix</i>	99
CHAPTER 6. Identification and Quantification of Any Isoforms of Carbohydrates by 2D UV-MS Fingerprinting of Cold Ions	104
6.1. <i>Introduction</i>	106
6.2. <i>Experimental approach</i>	109
6.3. <i>Results and discussion</i>	111
6.4. <i>Conclusion</i>	128
<i>References</i>	130
<i>Appendix</i>	135
CHAPTER 7. SUMMARY AND OUTLOOK	139
7.1. <i>Summary</i>	139
7.2. <i>Outlook</i>	140
CURRICULUM VITAE	142

CHAPTER 1.

Introduction

Non-covalent interactions are ubiquitous in nature. Their importance cannot be overstated. The balance of these relatively weak interactions defines the native three-dimensional (3D) structure of the most important biomolecules: proteins, nucleic acids (DNA, RNA), etc. In addition, their diversity allows highly selective binding between biomolecules. For instance, the processes of molecular recognition of an analyte (e.g. antigen, enzyme substrate) by bioreceptors (e.g. antibodies, enzymes) are based entirely on non-covalent interactions between specific binding sites.¹⁻² Unlike the covalent bonds, these interactions do not involve sharing of an electron pair, but rather rely on a variety of electromagnetic interactions, which include electrostatic (e.g. hydrogen bonding), π effects (e.g. stacking, $X-H \cdots \pi$, $charge \cdots \pi$), hydrophobic effect, Van der Waals forces (e.g. dipole - dipole, dipole - induced dipole), with typical energies on the order of 0.1 - 5 kcal/mol.³ The overwhelming majority of biochemical processes occur in water and its solutes, which crucially affect the 3D structure, properties, and functionality of biomolecules through the non-covalent interactions.⁴⁻⁵ The main objective of this research is to demonstrate how one can study native conformations of biomolecules in the gas phase. In addition, we employ the fact that intermolecular non-covalent interactions are capable of transferring structural information to binding partners to develop an approach for identifications of isomeric molecules, for which the use of standard techniques is complicated.

Determination of the three-dimensional structures of biomolecules in their native environment remains an important problem to solve for understanding the molecular mechanisms of life. The native environment for biomolecules is an aqueous solution, where they reside in different conformations while carrying out their functions in living organisms. These native conformations are stabilized by interactions with the

surrounding water and its solutes. However, the inhomogeneity of the interactions in the liquid phase, the dynamic nature of H-bonds together with an infinite number of molecular degrees of freedom make the identification of native structures challenging. One can get rid of the interactions by isolating the molecules in the gas phase. There are several different techniques, nowadays, that enable a gentle transfer of non-volatile molecules from solution to the gas phase. The most widespread soft ionization techniques, like MALDI and ESI, are capable of producing gas phase protonated/deprotonated species as large as protein complexes and even viruses. Isolation in the gas phase suppresses inhomogeneous spectral broadening, which is generally present in solution phase. In combination with cryogenic cooling of the isolated ions, this allows attaining vibrational resolution in their IR and UV spectra, which subsequently can be used in structural investigations.⁶⁻⁷

Optical spectroscopy has emerged as a powerful tool for structural determinations and analytical identifications of biological molecules. Although not directly, the spectroscopic techniques provide specific information about the 3D structure. For example, measurements of IR spectra in combination with quantum chemistry calculation enable a stringent validation of calculated 3D structures of biomolecules. The optical spectra are extremely sensitive to the spatial arrangement of atoms, immediately responding to even small structural alterations by changing shapes and positions of absorption peaks. Compared to other widely-used techniques (chromatography, ion mobility), the distinct advantage of optical spectroscopy is that it neither separates species nor assigns them by respective time tags. The optical spectra reflect differences in the 3D structures of molecules on a quantum mechanical level by mapping up the energies of electronic/vibrational/rotational molecular transitions. These fundamental characteristics of molecules are well reproducible over time and between different instrumentations, making the spectra convenient and very individual molecular tags.

Historically, the first electronic and vibrational spectra of bare gas-phase biomolecules were measured at room temperature.⁸⁻⁹ IRMPD is one

of the first methods widely used for vibrational spectroscopy of biomolecules in the gas phase. Although many IR stretch transitions could be resolved at room temperature, the observed spectral bands usually are quite broad. In addition, IRMPD does not offer any conformational selectivity: a single spectrum reflects IR transitions in all thermally well-populated species. This makes ambiguous the comparison of the IRMPD spectra with the conformer-specific computational data, which drastically lowers the confidence in assigning the computed structures. Despite all these limitations, the simplicity and convenience of IRMPD in some cases can become the decisive factor for using the technique in structural identifications and analytical applications.¹⁰⁻¹³

Room temperature UV spectra of the biomolecules that contain an aromatic ring are typically very broad and unstructured. Such spectra are suitable neither for obtaining conformer selective spectra nor for analytical identification.¹⁴ It worth mentioning that even if the UV spectra were resolved, the use of UV spectra for structural validations would be fairly difficult due to the fact, that modern methods of quantum chemistry used for calculating UV spectra are still insufficiently accurate for reproducing the experiment.

Stringent validation of the gas-phase structures requires some number of vibrationally resolved transitions in their spectra. These transitions then serve for an accurate comparison with theoretical predictions. One of the techniques that greatly improves vibrational resolution in spectra of large (e.g., biological) molecules is the cooling of molecules from room to cryogenic temperatures. Such cooling reduces the population of excited vibrational states, therefore, suppressing the thermal congestion arising mainly due to vibrational “hot” bands. Once the molecules are cooled down, their spectra may become structured.^{7, 15} This greatly increases the confidence in comparison of experimental and theoretical IR spectra. In addition, the resolved transitions in UV spectra of cold ions can be used for conformer-selective vibrational spectroscopy (e.g., IR-UV double-resonance and IR-IR-UV triple-resonance spectroscopic techniques).¹⁶⁻²¹ Such techniques enable an accurate conformational

assignment of the gas-phase ions, because they provide conformer-selective experimental spectra, which, one-by-one, can be compared with calculated vibrational frequencies of each particular conformer.

Another advance in the conformational assignment is the use of ion mobility spectrometry for separating molecular conformers and/or isomers prior their spectroscopic analysis.²²⁻²³ For instance, the combination of field-asymmetric ion mobility spectrometry (FAIMS) with cold ion spectroscopy allowed for separation of the conformational families of nonapeptide bradykinin.²⁴ Vibrational cooling is not a panacea, however, because for some molecules, for example, tryptophan, an intrinsic lifetime broadening may become the main source of spectral congestion and it cannot be suppressed. Interestingly, that in the case of protonated tryptophan this lifetime broadening can be greatly reduced by solvation of the charge by a very few water molecules non-covalently attached to the bare ion. The charge solvation breaks the interactions between the charged N-terminus and the aromatic ring of the ion. This removal of the electronic perturbation of the ring drastically lengthens the lifetime of the excited electronic state, thus sharpening its electronic spectra.²⁵ It worth noting that in all our studies herein, because of the low concentration of ions in the cold trap ($\sim 10^5 - 10^6$ ions/cm³), we do not detect a direct absorption by the gas-phase ions. Instead, we employ photofragmentation “action” spectroscopy, where appearance of fragments induced by absorption of laser light is monitored as a function of its wavelength. The recorded action spectra are often similar to the absorption spectra, but occasionally can be substantially offset by the fragmentation yield, which may not be a flat function of the wavelength.

Overall, high vibrational resolution achievable with cryogenic cooling enables an unambiguous validation of 3D structures of molecules. Sensitivity of well-resolved IR and UV spectra to structural details of molecules allows for distinguishing structurally similar molecules, such as isomers. This opens two main directions in developing cold ion spectroscopy: (i) solving intrinsic and native-like 3D structures of ions and (ii) analytical identification of biomolecules.

Over the last two decades, a significant amount of data has been produced by cold ion spectroscopy and used for validation of structures of small and mid-size molecules, including peptides, carbohydrates, lipids, metabolites, drugs, intermediates, etc.²⁶⁻⁴¹ So far the largest molecule for which the structures were validated by cold-ion spectroscopy remains decapeptide gramicidin S.⁴² The intrinsic limitation for sensing larger molecules is the lifetime broadening of vibrational transitions, which occurs due to the anharmonic coupling of vibrational states⁷. Moreover, the number of IR transitions rapidly increases with the molecular weight of molecules. This increase suppresses the vibrational resolution in the spectra, because the increased number of characteristic for biomolecules IR transitions remains located within the two limited spectroscopic regions around 3 μm and 6 μm . But, even if the vibrational transitions were well-resolved, the accuracy of quantum chemistry calculations for such large molecules would be insufficient for assigning the numerous closely spaced IR transitions.

Cryogenic cooling of ionic species allows for their non-covalent tagging by different small neutral molecules (e.g. He, N₂, H₂O, etc.) that may condense onto large cold ions in a cryogenic ion trap. The low energies of the non-covalent interactions between the ion of interest and tag molecules enable an efficient dissociation of the complexes after absorbing a single UV or IR photon. The spectra of tagged ions are measured by monitoring the depletion of the number of the complexes as a function of wavelength. Compared with bare ions, the interaction with tag molecules often causes shifts of some transitions in spectra of the latter.⁴³⁻⁴⁴ Depending on the nature of the tag molecule, the extent of the shifts can be different. The IR spectra of ions tagged by, for example, weakly bound noble gases are still might be suitable for certain structural validations.^{28, 45}

Despite all the advantages of cold ion spectroscopy, the value of the gas-phase intrinsic structures for life science is fairly limited. Water plays a key role in building the native 3D structure of biological molecules by breaking intra- and forming inter-molecular bonds. Although the main

structural motif of large biomolecules (e.g. proteins) may survive the desolvation, the pattern of non-covalent bonds in small-size biomolecules (e.g. metabolites, drugs, peptides, etc.) are likely to undergo a rearrangement, altering the 3D structures. As was mentioned above, the vibrational resolution decreases upon progressing solvation, greatly complicating the assignment of calculated structures. A compromise between the gas phase, where spectral resolution can be high but structures are intrinsic and the solution phase, where the structures are native but high resolution is difficult to achieve, one can interrogate in the gas-phase microhydrated biomolecular ions. Binding with a limited number of water molecules may retain the main structural motives of a native structure while keeping the high spectral resolution. Apparently, the calculations of structures and theoretical spectra are more challenging for microsolvated biomolecules than for the same bare molecules due to the flexibility of the intermolecular non-covalent bonds and the increased size of the molecular system. Nevertheless, the computational methods are becoming more accurate and less costly.

There are two principally different methods for isolating microsolvated ions in the gas phase. The most straightforward method is to gently transfer them directly from solution using electrospray ionisation. The mechanism of formation of bare ions from solution by ESI includes the formation of nanodroplets, where analyte ions are surrounded by water molecules. The conditions of ESI can be optimised such that to protect the solvated ions from complete drying. This approach can be called “top-down”, because it resembles the mechanism of formation of small peptides from a large protein. However, harsh conditions of ESI usually do not allow for retaining water molecules on an ion and evaporate them. There is only a narrow interval of ESI conditions, where such non-covalent complexes can survive while keeping the high ionization efficiency. This method of production of hydrated ions was implemented for the native structure elucidation of many microsolvated organic⁴⁶⁻⁴⁸ and inorganic ions⁴⁹⁻⁵¹. Alternatively, the microsolvated molecules can be built in the gas phase by condensing water molecules onto the bare ions that

are produced by ESI and cooled in a cryogenic trap. Opposite to the “top-down” approach, where water molecules undergo evaporation from nanodroplets forming smaller microsolvated ions, this method of building aqueous shells on bare ions can be called “bottom-up”. In this case, the steps of the ionization and the formation of the clusters are separated. This allows the conditions of ESI to be optimized for desolvated ions, while tuning temperature and buffer gas pressure in subsequent cryogenic ion trap can be optimized for forming a high concentration of the hydrated complexes with the desired number of waters (up to 50 molecules).⁵²⁻⁵³

In chapter 3, we experimentally answer an important question: are “bottom-up” (condensation) and “top-down” (dehydration) processes produce the same structures? We studied the microhydration of a small amino acid GlyH⁺ and a model tripeptide Gly₃H⁺, natively produced from solution by the top-down approach. We applied cold ion IR spectroscopy to the gas-phase complexes, in which these biomolecules are microhydrated by a controlled number of water molecules. The protonated complexes were produced directly from an aqueous solution, using a gentle electrospray ionization source. We compared our IR spectra with the available in literature spectra of the same complexes but produced by cryogenic condensation of water molecules onto the gas-phase ions.⁵⁴⁻⁵⁶ The comparison revealed a difference in the conformational distribution of the embedded biomolecules for the two methods of hydration. Already one water molecule retained on GlyH⁺ keeps the main motive of the native structure for this amino acid: the lack of an intramolecular hydrogen bond. This structure remains almost unchanged in all larger complexes that we studied (with up to 6 waters), which allows us to suggest that it is, indeed, a native-like structure of GlyH⁺. In contrast, the method of hydration by condensation produces only a fraction of the complexes with this native-like structure of the embedded glycine, while in the rest of them the amino acid remains kinetically trapped in its intrinsic state. This implies that the identification of the native-like structures with the complexes produced by the condensation can be ambiguous. Only upon hydration by 4-5 water molecules, the two methods converged in producing the same conformers.

Similarly, a fraction of Gly₃H⁺ tripeptides remains kinetically trapped in the complexes with one water molecule.

Another developing field for the application of cold ion spectroscopy is the analytical identifications of isomeric molecules. Isomers, which are molecules that have identical chemical composition but distinct arrangements of atoms, vary not only in chemical and biological properties but also have very different quantum energy levels. This fact can be employed for analytical identification of isomers. When combined with cryogenic cooling of ions, optical spectroscopy enables the mapping of ionic vibrational energy levels that are characteristic to the 3D structure of the species at a fundamental level. Instead of assigning time respective tags (retention or drift time measured by LC and IMS), which are known to be sensitive to experimental conditions, one can distinguish isomeric molecules by means of their inherent properties. Several techniques based on IR or UV spectroscopy of cryogenically cooled ions have been developed. The methods imply a library-based approach, in which a preliminarily recorded library of either IR or UV spectra of pure molecules are compared with a spectrum of an unknown sample. As it was mentioned above, the use of IR spectroscopy for structural validations is limited to relatively small molecules, such as peptides with 10-15 residues, because for larger molecules the IR spectra become blurred. The same applies for analytical applications of IR spectroscopy, with the exception that in some cases shapes of IR spectra alone can be used for distinguishing isomers.⁵⁷

⁵⁸ Dissimilar to IR, UV spectroscopy involves electronic excitation of particular chromophores, for which the width of peaks is not directly influenced by an increase in the size of a molecule. The drawback of this approach is that a chromophore may not sense the remote changes in the structure of the molecule.

In 2014 our group invented a technique, named 2D UV-MS fingerprinting, which combines UV spectroscopy of cold ions (UV) with high-resolution mass spectrometry (MS). UV electronic spectra measured with vibrational resolution, achievable by cryogenic cooling, contain a great number of spectral features, constituting unique fundamental

fingerprints of ions. The mechanism of UV photofragmentation can be different from IRMPD or thermal dissociation. This often allows for observing the photofragments that are specific to isomeric structures.⁵⁹ The relative abundance of all photofragments is detected at once by a broadband high-resolution mass spectrometer (e.g., Orbitrap-based). This additional MS dimension reduces the dispersion of RMSD values, enabling identification and quantification of challenging for other methods cases of isomerism. The identification by this approach implies a preliminary measurement of 2D UV-MS fingerprints of the species to be identified, constructing a library of the corresponding 2D matrices, and a decomposition of the recorded 2D spectra of the samples to be analysed into a linear combination of the library matrices. This procedure is very similar to the expressing a k -dimensional vector as a sum of k basis set vectors multiplied by the coefficients that reflect cosines of the angles between the sample and the basis set vectors.⁶⁰

An additional advantage of this approach is that it can be used online with liquid chromatography, which requires the timescale of spectroscopic measurement to be on a scale shorter than a minute. Although measurement of one 2D spectrum for an isomeric library compound may take a few dozens of minutes, the measurements of the spectra for an unknown analyte that belongs to this library can be done just at a few critical wavelengths, where spectral differences between the library fingerprints appeared to be the most prominent. This allows for shortening the acquisition time to a few tens of seconds with only an insignificant loss of the method accuracy.⁶¹⁻⁶²

Although relatively new, this approach has been already applied prior the studies described in this thesis for the identification of different isomeric molecules, including drugs, and peptides.⁶³⁻⁶⁵ All the molecules that were studied in the demonstrations contain at least one aromatic ring (e.g. phenyl, hydroxyphenyl aromatic groups) that strongly absorbs in UV. But, what if the analyte molecules do not contain a UV chromophore, as it is in the case of carbohydrates and lipids? The identification of these classes of molecules is extremely challenging, because they contain an

incredible number of diverse isomers. In this work, we extended application of the 2D UV-MS technique for identification of isomeric carbohydrates.

Carbohydrates are the most abundant organic molecules present in living organisms. Existing in myriads of structural forms carbohydrates decorate all cells and are used by nature to perform myriads of biological functions, including immune response, cell-cell communication, assisting protein folding. Living organisms exploit actively carbohydrates during the course of different diseases, changing their structures⁶⁶. The appearance of some glycans is associated with pathogenesis.⁶⁷ The early recognition of the glycomarkers enables more efficient treatment of a corresponding disease. Establishing connections between structures and functions of carbohydrates requires first their structural identification. However, the tremendous isomeric diversity of carbohydrates makes their identification and study extremely challenging.

The most widespread methods for structural identifications of carbohydrates are based on a combination of controlled enzymatic cleavage of large glycans, liquid chromatography (LC) to separate the produced structural units, and single/tandem mass spectrometry (LC/GC-MS) to characterise them. Despite its wide applicability, the method has certain limitations, such as the use of compound-specific conditions for separation, the need for derivatizations (by reductive amination, permethylation, etc.), and, often, is not capable alone of distinguishing carbohydrates in all their numerous isoforms.⁶⁸

In chapter 4 we demonstrate the use of non-covalent carbohydrate-aromatic complexes for the identification of isomeric glycans by developed in our group the unique method of 2D UV-MS cold ion spectroscopy. To make the method applicable to glycans, which have no functional groups active in near UV, we make use of non-covalent complexes of carbohydrates with aromatic molecules. Such complexes can be readily formed in solution, brought to the gas phase using ESI and then isolated and cooled in a cryogenic trap. The experimental and theoretical investigations show that an interplay of different types of inter- and intra-

molecular hydrogen bonds (CH- π , OH- π , cation- π , etc.) in such complexes makes UV absorption of the aromatic ion highly sensitive to the small structural differences between isomeric carbohydrates. These findings enable the identification and relative quantification of isoforms of carbohydrates in their mixtures.

In chapter 5, we report the atomic level studies of isolated in the gas phase non-covalent complexes of monosaccharide α/β -GalNAc with protonated aromatic molecule tyramine. We experimentally observed that hydrogen bonds in carbohydrate-protein interactions are amazingly sensitive to the anomeric form of carbohydrates. Although not directly involved in hydrogen bonding, these anomeric groups influence the length of a remote intermolecular hydrogen bond, thus contributing to the anomeric selectivity of the binding. Our estimations of the difference in binding energy for a single bond is 1 kcal/mol, which cannot be neglected for the total budget of the glycan-protein affinity. The finding demonstrates an incredible sensitivity of the aromatic molecule that can feel the changes in analyte molecules even if the pattern of H-bond in the complexes is the same.

In chapter 6 we experimentally demonstrate that all classes of isomeric carbohydrates can be distinguished and their relative solution concentrations can be quantified using the method of 2D UV-MS fingerprinting of cold ions. We also test the performance of a simplified, 1D UV-MS, approach for the identification of isomeric carbohydrates. It turned out that UV spectroscopy alone is capable of accurate quantification of carbohydrates, however, for the more complex mixtures, the 2D UV-MS approach is required. This finding allows for using an inexpensive lower resolution quadrupole mass spectrometer, instead of the costly Orbitrap-based MS without a significant loss of accuracy of the quantifications. The reduced cost of the required hardware may lead to a more widespread application of the method for identifications of isomeric species.

In the final chapter, we summarise the main results presented in this thesis and give an outlook for future research.

REFERENCES

1. Gellman, S. H., Introduction: Molecular Recognition. *Chemical Reviews* **1997**, 97 (5), 1231-1232.
2. Sela-Culang, I.; Kunik, V.; Ofra, Y., The structural basis of antibody-antigen recognition. *Front Immunol* **2013**, 4, 302-302.
3. Hobza, P., Calculations on Noncovalent Interactions and Databases of Benchmark Interaction Energies. *Accounts Chem Res* **2012**, 45 (4), 663-672.
4. Laage, D.; Elsaesser, T.; Hynes, J. T., Water Dynamics in the Hydration Shells of Biomolecules. *Chemical Reviews* **2017**, 117 (16), 10694-10725.
5. Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E., Water Determines the Structure and Dynamics of Proteins. *Chemical Reviews* **2016**, 116 (13), 7673-7697.
6. Rijs, A. M.; Oomens, J., IR Spectroscopic Techniques to Study Isolated Biomolecules. In *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Rijs, A. M.; Oomens, J., Eds. Springer International Publishing: Cham, 2015; pp 1-42.
7. Boyarkin, O. V., Cold ion spectroscopy for structural identifications of biomolecules. *International Reviews in Physical Chemistry* **2018**, 37 (3-4), 559-606.
8. Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W., Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy. *Proceedings of the National Academy of Sciences* **2002**, 99 (25), 15863.
9. Andersen, L. H.; Lapierre, A.; Nielsen, S. B.; Nielsen, I. B.; Pedersen, S. U.; Pedersen, U. V.; Tomita, S., Chromophores of the green fluorescent protein studied in the gas phase. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics* **2002**, 20 (3), 597-600.
10. Martens, J.; Koppen, V.; Berden, G.; Cuyckens, F.; Oomens, J., Combined Liquid Chromatography-Infrared Ion Spectroscopy for

Identification of Regioisomeric Drug Metabolites. *Anal. Chem.* **2017**, *89* (8), 4359-4362.

11. 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* (3), 670-679.

12. Barnes, L.; Schindler, B.; Chambert, S.; Allouche, A.-R.; Compagnon, I., Conformational preferences of protonated N-acetylated hexosamines probed by InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and ab initio calculations. *Int. J. Mass Spectrom.* **2017**, *421*, 116-123.

13. Schindler, B.; Barnes, L.; Gray, C. J.; Chambert, S.; Flitsch, S. L.; Oomens, J.; Daniel, R.; Allouche, A. R.; Compagnon, I., IRMPD Spectroscopy Sheds New (Infrared) Light on the Sulfate Pattern of Carbohydrates. *J. Phys. Chem A* **2017**, *121* (10), 2114-2120.

14. 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* (9), 2816-2817.

15. Rizzo, T. R.; Boyarkin, O. V., Cryogenic Methods for the Spectroscopy of Large, Biomolecular Ions. *Gas-Phase Ir Spectroscopy and Structure of Biological Molecules* **2015**, *364*, 43-97.

16. Hünig, I.; Seefeld, K. A.; Kleineremanns, K., REMPI and UV-UV double resonance spectroscopy of tryptophan ethylester and the dipeptides tryptophan-serine, glycine-tryptophan and proline-tryptophan. *Chem Phys Lett* **2003**, *369* (1), 173-179.

17. Gerardi, H. K.; Breen, K. J.; Guasco, T. L.; Weddle, G. H.; Gardenier, G. H.; Laaser, J. E.; Johnson, M. A., Survey of Ar-Tagged Predissociation and Vibrationally Mediated Photodetachment Spectroscopies of the Vinylidene Anion, $C_2H_2^-$. *J. Phys. Chem A* **2010**, *114* (3), 1592-1601.

18. Florio, G. M.; Gruenloh, C. J.; Quimpo, R. C.; Zwier, T. S., The infrared spectroscopy of hydrogen-bonded bridges: 2-pyridone-(water)_n and 2-hydroxypyridine-(water)_n clusters, n=1,2. *The Journal of Chemical Physics* **2000**, *113* (24), 11143-11153.

19. Carney, J. R.; Zwier, T. S., The infrared and ultraviolet spectra of individual conformational isomers of biomolecules: Tryptamine. *J. Phys. Chem. A* **2000**, *104* (38), 8677-8688.
20. Zwier, T. S., Laser Spectroscopy of Jet-Cooled Biomolecules and Their Water-Containing Clusters: Water Bridges and Molecular Conformation. *J. Phys. Chem A* **2001**, *105* (39), 8827-8839.
21. Shubert, V. A.; Zwier, T. S., IR–IR–UV Hole-Burning: Conformation Specific IR Spectra in the Face of UV Spectral Overlap. *J. Phys. Chem A* **2007**, *111* (51), 13283-13286.
22. Warnke, S.; Ben Faleh, A.; Scutelnic, V.; Rizzo, T. R., Separation and Identification of Glycan Anomers Using Ultrahigh-Resolution Ion-Mobility Spectrometry and Cryogenic Ion Spectroscopy. *J. Am. Soc. Mass. Spectr.* **2019**, *30* (11), 2204-2211.
23. Ben Faleh, A.; Warnke, S.; Rizzo, T. R., Combining Ultrahigh-Resolution Ion-Mobility Spectrometry with Cryogenic Infrared Spectroscopy for the Analysis of Glycan Mixtures. *Anal. Chem.* **2019**, *91* (7), 4876-4882.
24. Voronina, L.; Rizzo, T. R., Spectroscopic studies of kinetically trapped conformations in the gas phase: the case of triply protonated bradykinin. *Phys Chem Chem Phys* **2015**, *17* (39), 25828-25836.
25. Mercier, S. R.; Boyarkin, O. V.; Kamariotis, A.; Guglielmi, M.; Tavernelli, I.; Cascella, M.; Rothlisberger, U.; Rizzo, T. R., Microsolvation Effects on the Excited-State Dynamics of Protonated Tryptophan. *J. Am. Chem. Soc.* **2006**, *128* (51), 16938-16943.
26. 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.
27. DeBlase, A. F.; Harrilal, C. P.; Lawler, J. T.; Burke, N. L.; McLuckey, S. A.; Zwier, T. S., Conformation-Specific Infrared and Ultraviolet Spectroscopy of Cold [YAPAA+H]⁺ and [YGPAA+H]⁺ Ions: A Stereochemical “Twist” on the β -Hairpin Turn. *J. Am. Chem. Soc.* **2017**, *139* (15), 5481-5493.

28. Roithová, J.; Gray, A.; Andris, E.; Jašík, J.; Gerlich, D., Helium Tagging Infrared Photodissociation Spectroscopy of Reactive Ions. *Accounts Chem Res* **2016**, *49* (2), 223-230.
29. Andris, E.; Navrátil, R.; Jašík, J.; Terencio, T.; Srnec, M.; Costas, M.; Roithová, J., Chasing the Evasive Fe=O Stretch and the Spin State of the Iron(IV)-Oxo Complexes by Photodissociation Spectroscopy. *J. Am. Chem. Soc.* **2017**, *139* (7), 2757-2765.
30. Gerlich, D.; Jašík, J.; Strelnikov, D. V.; Roithová, J., IR Spectroscopy of Fullerene Ions in a Cryogenic Quadrupole Trap. *The Astrophysical Journal* **2018**, *864* (1), 62.
31. Schwarz, H.; Asmis, K. R., Identification of Active Sites and Structural Characterization of Reactive Ionic Intermediates by Cryogenic Ion Trap Vibrational Spectroscopy. *Chemistry – A European Journal* **2019**, *25* (9), 2112-2126.
32. Kirschbaum, C.; Greis, K.; Mucha, E.; Kain, L.; Deng, S.; Zappe, A.; Gewinner, S.; Schöllkopf, W.; von Helden, G.; Meijer, G.; Savage, P. B.; Marianski, M.; Teyton, L.; Pagel, K., Unravelling the structural complexity of glycolipids with cryogenic infrared spectroscopy. *Nat Commun* **2021**, *12* (1), 1201.
33. Greis, K.; Kirschbaum, C.; Lechnitz, S.; Gewinner, S.; Schöllkopf, W.; von Helden, G.; Meijer, G.; Seeberger, P. H.; Pagel, K., Direct Experimental Characterization of the Ferrier Glycosyl Cation in the Gas Phase. *Organic Letters* **2020**, *22* (22), 8916-8919.
34. Marianski, M.; Mucha, E.; Greis, K.; Moon, S.; Pardo, A.; Kirschbaum, C.; Thomas, D. A.; Meijer, G.; von Helden, G.; Gilmore, K.; Seeberger, P. H.; Pagel, K., Remote Participation during Glycosylation Reactions of Galactose Building Blocks: Direct Evidence from Cryogenic Vibrational Spectroscopy. *Angew. Chem. Int. Ed.* **2020**, *59* (15), 6166-6171.
35. Gorlova, O.; Colvin, S. M.; Brathwaite, A.; Menges, F. S.; Craig, S. M.; Miller, S. J.; Johnson, M. A., Identification and Partial Structural Characterization of Mass Isolated Valsartan and Its Metabolite with Messenger Tagging Vibrational Spectroscopy. *J. Am. Soc. Mass. Spectr.* **2017**, *28* (11), 2414-2422.

36. Pereverzev, A. Y.; Szabó, I.; Kopysov, V. N.; Rosta, E.; Boyarkin, O. V., Gas-phase structures reflect the pain-relief potency of enkephalin peptides. *Phys Chem Chem Phys* **2019**, *21* (41), 22700-22703.
37. Kirschbaum, C.; Saied, E. M.; Greis, K.; Mucha, E.; Gewinner, S.; Schöllkopf, W.; Meijer, G.; von Helden, G.; Poad, B. L. J.; Blanksby, S. J.; Arenz, C.; Pagel, K., Resolving Sphingolipid Isomers Using Cryogenic Infrared Spectroscopy. *Angew. Chem. Int. Ed.* **2020**, *59* (32), 13638-13642.
38. Scutelnic, V.; Rizzo, T. R., Cryogenic Ion Spectroscopy for Identification of Monosaccharide Anomers. *J. Phys. Chem A* **2019**, *123* (13), 2815-2819.
39. Cismesia, A. P.; Bell, M. R.; Tesler, L. F.; Alves, M.; Polfer, N. C., Infrared ion spectroscopy: an analytical tool for the study of metabolites. *Analyst* **2018**, *143* (7), 1615-1623.
40. Scutelnic, V.; Perez, M. A. S.; Marianski, M.; Warnke, S.; Gregor, A.; Rothlisberger, U.; Bowers, M. T.; Baldauf, C.; von Helden, G.; Rizzo, T. R.; Seo, J., The Structure of the Protonated Serine Octamer. *J. Am. Chem. Soc.* **2018**, *140* (24), 7554-7560.
41. Wako, H.; Ishiuchi, S.-i.; Kato, D.; Féraud, G.; Dedonder-Lardeux, C.; Jouvet, C.; Fujii, M., A conformational study of protonated noradrenaline by UV-UV and IR dip double resonance laser spectroscopy combined with an electrospray and a cold ion trap method. *Phys Chem Chem Phys* **2017**, *19* (17), 10777-10785.
42. Roy, T. K.; Kopysov, V.; Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V.; Gerber, R. B., Conformational Structures of a Decapeptide Validated by First Principles Calculations and Cold Ion Spectroscopy. *Chemphyschem* **2015**, *16* (7), 1374-1378.
43. Johnson, C. J.; Wolk, A. B.; Fournier, J. A.; Sullivan, E. N.; Weddle, G. H.; Johnson, M. A., Communication: He-tagged vibrational spectra of the SarGlyH⁺ and H⁺(H₂O)_{2,3} ions: Quantifying tag effects in cryogenic ion vibrational predissociation (CIVP) spectroscopy. *The Journal of Chemical Physics* **2014**, *140* (22), 221101.
44. Masson, A.; Williams, E. R.; Rizzo, T. R., Molecular hydrogen messengers can lead to structural infidelity: A cautionary tale of

protonated glycine. *The Journal of Chemical Physics* **2015**, *143* (10), 104313.

45. Jašík, J.; Žabka, J.; Roithová, J.; Gerlich, D., Infrared spectroscopy of trapped molecular dications below 4K. *Int. J. Mass Spectrom.* **2013**, *354-355*, 204-210.

46. Chang, T. M.; Chakrabarty, S.; Williams, E. R., Hydration of Gaseous m-Aminobenzoic Acid: Ionic vs Neutral Hydrogen Bonding and Water Bridges. *J. Am. Chem. Soc.* **2014**, *136* (29), 10440-10449.

47. 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* (6079), 320.

48. Spieler, S.; Duong, C. H.; Kaiser, A.; Duensing, F.; Geistlinger, K.; Fischer, M.; Yang, N.; Kumar, S. S.; Johnson, M. A.; Wester, R., Vibrational Predissociation Spectroscopy of Cold Protonated Tryptophan with Different Messenger Tags. *J. Phys. Chem A* **2018**, *122* (40), 8037-8046.

49. Asmis, K. R.; Neumark, D. M., Vibrational Spectroscopy of Microhydrated Conjugate Base Anions. *Accounts Chem Res* **2012**, *45* (1), 43-52.

50. Mitra, S.; Duong, C. H.; McCaslin, L. M.; Gerber, R. B.; Johnson, M. A., Isomer-specific cryogenic ion vibrational spectroscopy of the D2 tagged $\text{Cs}+(\text{HNO}_3)(\text{H}_2\text{O})_{n=0-2}$ complexes: ion-driven enhancement of the acidic H-bond to water. *Phys Chem Chem Phys* **2020**, *22* (8), 4501-4507.

51. Sun, S.-T.; Jiang, L.; Liu, J. W.; Heine, N.; Yacovitch, T. I.; Wende, T.; Asmis, K. R.; Neumark, D. M.; Liu, Z.-F., Microhydrated dihydrogen phosphate clusters probed by gas phase vibrational spectroscopy and first principles calculations. *Phys Chem Chem Phys* **2015**, *17* (39), 25714-25724.

52. Marsh, B. M.; Voss, J. M.; Garand, E., A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *The Journal of Chemical Physics* **2015**, *143* (20), 204201.

53. Garand, E., Spectroscopy of Reactive Complexes and Solvated Clusters: A Bottom-Up Approach Using Cryogenic Ion Traps. *J. Phys. Chem A* **2018**, *122* (32), 6479-6490.

54. Fischer, K. C.; Sherman, S. L.; Garand, E., Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and β -Alanine. *J. Phys. Chem A* **2020**, *124* (8), 1593-1602.
55. Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E., Microsolvation Structures of Protonated Glycine and l-Alanine. *J. Phys. Chem A* **2019**, *123* (15), 3355-3366.
56. Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E., Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H⁺·(H₂O). *J. Phys. Chem A* **2018**, *122* (41), 8213-8221.
57. Khanal, N.; Masellis, C.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R., Cryogenic IR spectroscopy combined with ion mobility spectrometry for the analysis of human milk oligosaccharides. *Analyst* **2018**, *143* (8), 1846-1852.
58. Masellis, C.; Khanal, N.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R., Cryogenic Vibrational Spectroscopy Provides Unique Fingerprints for Glycan Identification. *J. Am. Soc. Mass. Spectr.* **2017**, *28* (10), 2217-2222.
59. Reilly, J. P., Ultraviolet photofragmentation of biomolecular ions. *Mass spectrometry reviews* **2009**, *28* (3), 425-447.
60. 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* (9), 4607-4611.
61. Kopysov, V.; Gorshkov, M. V.; Boyarkin, O. V., Identification of isoforms of aspartic acid residues in peptides by 2D UV-MS fingerprinting of cold ions. *Analyst* **2018**, *143* (4), 833-836.
62. Lobas, A. A.; Solovyeva, E. M.; Saparbaev, E.; Gorshkov, M. V.; Boyarkin, O. V., Accelerating photofragmentation UV Spectroscopy-Mass spectrometry fingerprinting for quantification of isomeric peptides. *Talanta* **2021**, *232*, 122412.
63. Kopysov, V.; Makarov, A.; Boyarkin, O. V., Identification of Isomeric Ephedrines by Cold Ion UV Spectroscopy: Toward Practical Implementation. *Anal. Chem.* **2017**, *89* (1), 544-547.

64. Kopysov, V.; Nagornova, N. S.; Boyarkin, O. V., Identification of Tyrosine-Phosphorylated Peptides Using Cold Ion Spectroscopy. *J. Am. Chem. Soc.* **2014**, *136* (26), 9288-9291.
65. Solovyeva, E. M.; Kopysov, V. N.; Pereverzev, A. Y.; Lobas, A. A.; Moshkovskii, S. A.; Gorshkov, M. V.; Boyarkin, O. V., Method for Identification of Threonine Isoforms in Peptides by Ultraviolet Photofragmentation of Cold Ions. *Anal. Chem.* **2019**, *91* (10), 6709-6715.
66. Varki, A., Biological roles of glycans. *Glycobiology* **2017**, *27* (1), 3-49.
67. Pierce, J. M., Glycomarkers. In *Glycoscience: Biology and Medicine*, Endo, T.; Seeberger, P. H.; Hart, G. W.; Wong, C.-H.; Taniguchi, N., Eds. Springer Japan: Tokyo, 2021; pp 1-6.
68. Gray, C. J.; Migas, L. G.; Barran, P. E.; Pagel, K.; Seeberger, P. H.; Evers, C. E.; Boons, G.-J.; Pohl, N. L. B.; Compagnon, I.; Widmalm, G.; Flitsch, S. L., Advancing Solutions to the Carbohydrate Sequencing Challenge. *J. Am. Chem. Soc.* **2019**, *141* (37), 14463-14479.

CHAPTER 2.

Experimental Setup

2.1. INSTRUMENT OVERVIEW

Our cold ion spectrometer, built a decade ago and upgraded on several occasions, has been described elsewhere.¹⁻⁴ The most recent developments of the instrument relate to the ion source. Currently, the instrument can be used with two different electrospray ion sources that target generation (1) of high quantity of biomolecular ions and (2) of their microhydrated complexes. The highly efficient but relatively harsh ion source uses a double electrodynamic ion funnel, and the ultra-soft but moderately efficient ion source, designed in our group, is based on the use of three consecutive skimmers. Schematic views of the instrument are shown in Figures 2.1 and 2.2. In the high productive configuration (Figure 2.1), the ions, generated from acidic solution by a nano-electrospray ionization (n-ESI) source, enter an electrodynamic ion funnel (IF) orthogonally to its axis through a 10 cm long stainless-steel capillary of 0.7 mm internal diameter. The ions are then turned by 90° to travel along the funnel axis, which guides them into a hexapole ion trap (HEX) where they are accumulated and thermalized for approximately 50 or 100 ms, depending on the repetition rate of experiment.

In the soft configuration (Figure 2.2), the ions, injected in parallel to the axis, pass through the three consecutive skimmers (~2 mm in diameter of the orifice), which allow for a gradual reduction of pressure from 0.5 mbar to 10⁻⁴ mbar, and enter an octupole ion guide. Guided by the octupole, they are accumulated in an octupole ion trap, where they are thermalized.

After being released from hexapole or octupole ion traps, ions pass through a quadrupole mass filter (Q1), which is set to select the ions of interest. Mass-selected ions are then turned by 90° using an electrostatic bender, focused by a stack of three electrostatic lenses and moved through an RF octupole guide into a cold octupole trap, which is kept at 6 K. The

trap is driven by two 1 MHz sinus waveforms with peak-to-peak amplitudes of 50-100 V. The stored in the trap ions are cooled down to ~10 K in collisions with He buffer gas, which is pulsed into the trap shortly before the arrival of the ions. The cold ions are irradiated by UV, IR or IR-UV laser pulses, depending on the type of experiments, and undergo photo fragmentation. The fragment and parent ions were 90° turned by the second electrostatic quadrupole bender either toward a highly sensitive quadrupole mass spectrometer (QMS) or toward high-resolution broadband Orbitrap-based MS (Exactive, Thermo Fisher). The spectra and fingerprints are measured by continuously recording the yield of the photo fragments using the QMS and the Orbitrap-based MS, while scanning UV or IR wavelength.

The repetition rate of the cooling/fragmentation cycle was determined by the 10 Hz repetition rate of the OPO, such that the parent ions experienced only one OPO shot in each cycle. At each UV wavelength the yield for a single fragment (with QMS) or the entire fragment mass spectrum was measured in 10 cycles and averaged to give a data point in UVPD spectrum or a fragment MS in 2D UV-MS fingerprint, respectively.

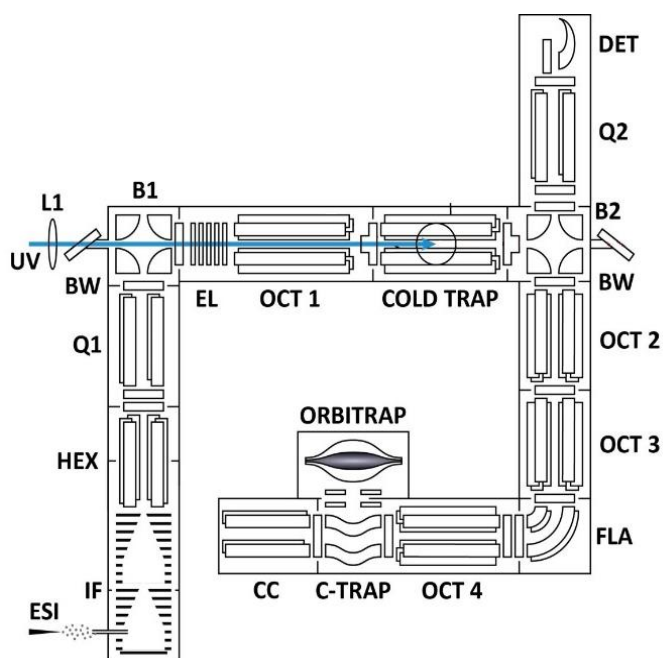


Figure 2.1. Schematic view of the highly efficient modification of the instrument for spectroscopy of cryogenically cooled ions.

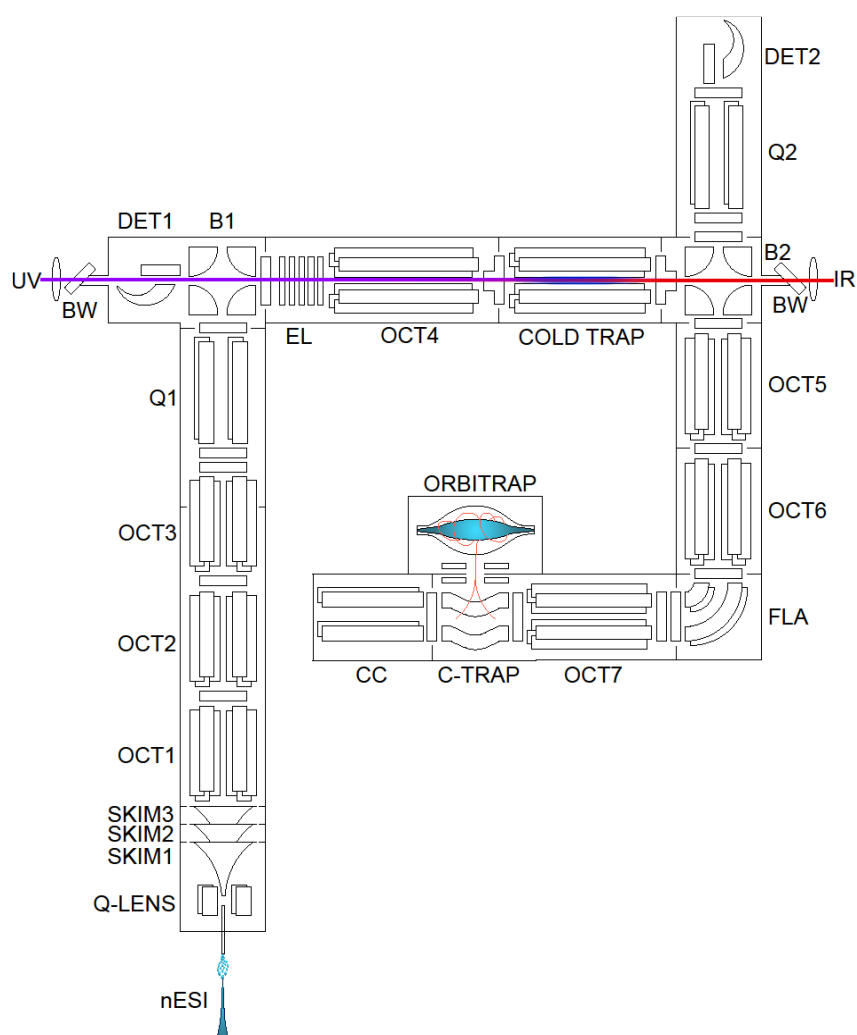


Figure 2.2. Schematic view of the ultra-soft modification of the instrument for spectroscopy of cryogenically cooled microsolvated ions.

In measurements with Orbitrap-MS, the whole fragment mass spectrum was normalized to the total ion signal detected in the cycle. In QMS measurements we employ 20 Hz cooling cycle and detect fragment or parent ions in 10 Hz alternative cycles with and without UV pulse, respectively.

2.2. RF ION FUNNEL AND SKIMMERS

Ion cloud produced by ESI travels through a metal capillary into the first vacuum chamber maintained at the pressure 2-6 mbar with a raw pump. The injected ions are further guided into the lower pressure sections using two different designs, depending on the type of experiment. The first modification was designed for efficient guidance of ions and based on an **electrodynamic ion funnel**, while the second design was developed for a soft transition of microsolvated ions and based on the system of three consecutive **skimmers**.

RF ION FUNNEL

The design of ion funnels allows for efficient focusing of ions, increasing, therefore, the throughput of the ions to the next vacuum sections. An ion funnel consists of a stack of cylindrical electrodes with gradually decreasing inner diameter and a final plate that serves as a conductance limiting orifice. The funnel needs to keep the ion beam radially focused, while moving them axially toward the orifice. The radial confinement is provided via radiofrequency oscillation voltage (RF). When the ions are far from the electrodes, they do not feel radial force or feel only a little. However, as the ions start movements outward toward the electrodes they experience the field that returns them to the central axis. Although the RF field effectively keeps the ion beam far from the electrodes it does not push the ions in the axial direction. To provide the axial movement, in addition to the RF, DC potential gradient is applied to the electrodes. When using the ion funnel the orthogonal injection of the ions is exploited. The ions, orthogonally injected into the funnel, are first bent perpendicularly by a repeller (with relatively high voltage applied), then are caught by RF, and pushed by the DC gradient towards the next stages. The orthogonal injection enables removing not charged molecules, which are pumped away by the grease pump.

SKIMMERS

A skimmer samples a portion of a gas from a flow, enabling to achieve a drastic decrease of pressure in the further vacuum stages. Our system, which will be described elsewhere,⁵ consists of three consecutive skimmers with the apertures of 1-3 mm spaced apart by 7 to 16 mm. The produced by ESI ions and solvent molecules are transferred toward the 1st skimmer through the heated capillary. They are further transported in the expanding gas jet through the skimmers. In addition, the ions are pulled in and focused to the axis by an electrical field, created between the capillary and the 1st skimmer and between the skimmers by the applied to these elements DC voltages. The focusing allows the neutral molecules to be pumped away, as the DC voltages do not influence their pathway. Although consecutive skimmers substantially limit the ion transmittance, such design enables a soft transition of non-covalently bound complexes, including microsolvated ions. The number of transmitted ions can be increased by increasing the diameters of the opening in the skimmers, but that would result in the build-up of the pressure in further sections. The source based on the ion funnel is more efficient, but the applied RF field and the relatively high pressure in the funnel heat the ions up in the volume close to the conductance limit. This vibrational heating often causes the breakage of weak non-covalent bonds.

2.3. TIMING OF EVENTS

Different hardware subunits working in our experiments must operate synchronously. Because the repetition rate of our lasers is 10 Hz, our measurements by quadrupole mass spectrometer are done in 20 Hz repetitive cycles. The scheme allows for measuring alternatively one value with a laser pulse and one without it. Typically, we use the cycles with laser on for measuring appearances of fragments, while the cycles with laser off are used for measuring parent ions. The signal of parent ions is used for normalisation to account for its fluctuations.

Figure 2.3 shows a diagram of the events in our experiments. The experimental cycle begins with the injection of He gas into the cold trap, just shortly before the accumulated in hexapole/octupole ions, are released by lowering (for duration of 3ms) the trapping potential of the exit lens. After preselection by the first mass filter, the parent ions enter the octupole cold trap, where they are trapped and cooled down in collisions with the injected He. When ions are cold, the photofragmentation by a laser is induced. In the 50 ms cycles, a laser pulse typically arrives 21 ms from the start, allowing ~20 ms for dissociation of photo excited ions. The exit lens of the cold trap opens for 5 ms by pulling down the trapping potential in about 10 ms and keeping the trap opened until the end of the cycle. The appeared photofragments and the remained parent ions are released from the cold trap and travel towards the quadrupole mass analyser. In order to efficiently empty the cold trap before the next cycle, we switch off RF waveform applied to from its rods 5 ms before the end of the cycle for 3 ms. The appearance of selected ions is recorded by one of the two channels of a photon counter. In the next cycle, the ions are accumulated in the same manner, but no laser pulse is applied. The quadrupole mass analyser is typically set to detect parent ions, whereas the signal is measured by the second channel of the photon counter. Typically, we use 10 laser shots to record one data point for UV spectra and 20 laser shots for IR spectra.

Similar workflow is used when measuring the data by Orbitrap, but because Orbitrap allows for simultaneous detection of fragment and

parent ions, the laser fires in every cooling/trapping cycle, and the ions are guided to an Orbitrap mass analyser.

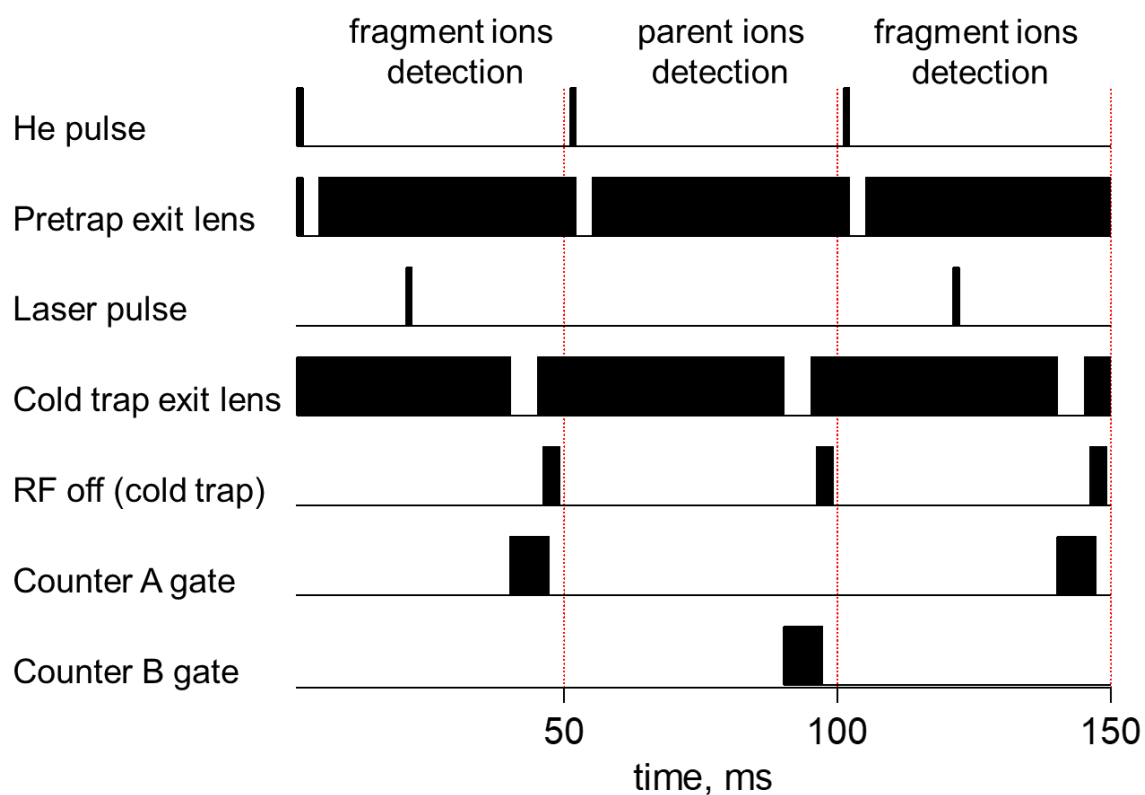


Figure 2.2. Typical time-diagram of the laser and He gas pulses and electrical potentials applied to different electrodes of the instrument during a photodissociation experiment at the repetition rate of 20 Hz.

2.4. OPTICAL SETUP

Previously, our group exploited dye lasers for analytical experiments.⁶⁻⁹ The linewidth of the dye lasers is significantly narrower ($\sim 0.2\text{ cm}^{-1}$) than of the UV OPO light ($\sim 5\text{ cm}^{-1}$). The high spectral resolution of dye lasers allows for very accurate identification and quantification of components, but modern standard analytical techniques require also speed, compactness, and simplicity of use. UV OPO systems bring these qualities without any significant compromise in the accuracy of analytical identifications.

UV light for analytical experiments is produced by 10 Hz UV OPO (EKSPLA NT 340C)¹⁰. This is a widely wavelength tunable system, which covers 192 – 2600 nm spectral range. Our experimental data were mainly recorded using 210 – 310 nm spectral range, which covers the absorption of small aromatic molecules such as tyrosine, phenylalanine, tryptophan. The pulse energy of the OPO in this UV range can achieve 10 mJ (Figure 2.3). We tuned the OPO to produce a lower ($\sim 2\text{ mJ}$) but uniform energy through the entire operating spectral range of 210 – 310 nm. The angular positions of non-linear crystals in the OPO doubling/mixing stages were detuned from their optimal phasematching angles, such that the output energy at each wavelength was close to the minimum energy attainable within the desired spectral range. This minimizes the potential contributions to UVPD from non-linear absorption and, therefore, increases reproducibility of the fingerprints. In addition, the measured ion signals were normalized to OPO pulse energy, which was measured by a broadband pyroelectric detector. The laser beam from the UV OPO enters the instruments through a BaF_2 window, after being loosely focused by a $F=1\text{m}$ fused silica lens $\sim 2\text{ cm}$ from the centre of the cold trap.

For structural investigations that require high spectral resolution, the UV light is produced by frequency doubling the visible output of a dye laser (HD-500, Lumonics) in a BBO (beta barium borate) crystal of an Autotracker II (Inrad, IS) device. The dye laser is pumped by the third harmonic (355 nm) of a Nd:YAG laser (Spectra Physics GCR-250). In our experiments we used 1-2 mJ of energy per pulse at 10 Hz repetition rate.

To generate IR laser light we use two tunable IR OPO systems (Laser Vision), pumped by 1064 nm of two different Nd:YAG lasers (Surelite IIIEx, Continuum; SpitLight 600, Innolas). The systems can operate at 5 or 10 Hz repetition rate with 1 to 10 mJ energy per pulse. The spectral resolutions of the IR OPOs are about 1 cm^{-1} . The laser beam from IR OPO enters the instruments through another a BaF_2 window placed at Brewster's angle. To measure conformer selective IR and UV spectra, IR-UV double resonance techniques were employed.^{2, 4}

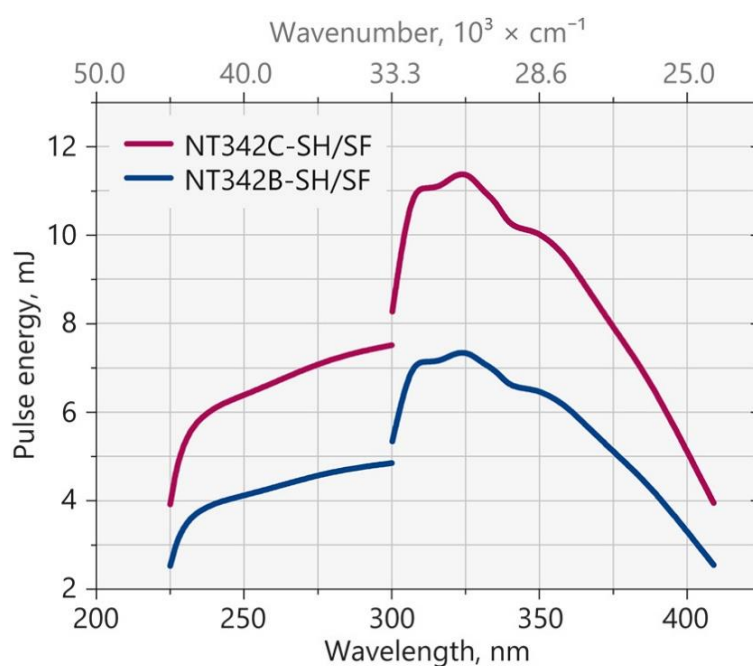


Figure 2.3. Typical output energy in used spectral range before tuning¹⁰

REFERENCES

1. 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* (9), 4607-4611.
2. Boyarkin, O. V., Cold ion spectroscopy for structural identifications of biomolecules. *International Reviews in Physical Chemistry* **2018**, *37* (3-4), 559-606.
3. Solovyeva, E. M.; Kopysov, V. N.; Pereverzev, A. Y.; Lobas, A. A.; Moshkovskii, S. A.; Gorshkov, M. V.; Boyarkin, O. V., Method for Identification of Threonine Isoforms in Peptides by Ultraviolet Photofragmentation of Cold Ions. *Anal. Chem.* **2019**, *91* (10), 6709-6715.
4. Pereverzev, A. Multilaser Conformer-Selective Spectroscopy of Cold Biomolecular Ions in the Gas Phase. EPFL, PhD Thesis, 2017.
5. Zviagin, A. to be published. EPFL, PhD Thesis, 2023.
6. 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* (6079), 320.
7. Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R., Conformation-specific infrared and ultraviolet spectroscopy of tyrosine-based protonated dipeptides. *J Chem Phys* **2007**, *127* (15), 154322-7.
8. Kopysov, V.; Makarov, A.; Boyarkin, O. V., Identification of Isomeric Ephedrines by Cold Ion UV Spectroscopy: Toward Practical Implementation. *Anal. Chem.* **2017**, *89* (1), 544-547.
9. Kopysov, V.; Nagomova, N. S.; Boyarkin, O. V., Identification of Tyrosine-Phosphorylated Peptides Using Cold Ion Spectroscopy. *J. Am. Chem. Soc.* **2014**, *136* (26), 9288-9291.
10. EKSPLA High Energy Broadly Tunable Lasers, NT340 series. <https://ekspla.com/product/nt340-series-tunable-lasers/#tab6>.

CHAPTER 3.

Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy

OVERVIEW

The native-like structures of protonated glycine and peptide Gly₃H⁺ were elucidated using cold ion IR spectroscopy of these biomolecules hydrated by a controlled number of water molecules. The non-covalent complexes of Gly₃H⁺ with H₂O were generated directly from an aqueous solution using gentle electrospray ionization. Already with a single retained water molecule, GlyH⁺ exhibits the native-like structure characterized by a lack of intramolecular NH...O=C hydrogen bonds. We use our spectra to calibrate the available data for the same complexes, but produced by cryogenic condensation of water onto the gas-phase glycine. Only in some conformers of these complexes GlyH⁺ adopts the native-like structure, while in the others remains “kinetically” trapped in the intrinsic state. Only upon condensation of 4-5 water molecules the embedded amino acid fully adopts its native-like structure. Similar, condensation of one water molecule onto the tripeptide is insufficient to fully eliminate its kinetically trapped intrinsic states.

The results presented in this chapter were published in:

E. Saparbaev, V. Aladinskaia, A. Zviagin, and O. Boyarkin. Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy. *Journal of Physical Chemistry Letters* 2021, 12, 10644-10648, DOI: <https://doi.org/10.1021/acs.jpcllett.0c03678>

3.1. INTRODUCTION

Three-dimensional structures of biological molecules isolated in the gas phase may differ drastically from the geometries that they adopt in aqueous solutions. These native conformations, which are stabilized by hydrogen bonds with surrounding water, define the functionality of the biological molecules in living organisms. Many biological processes are mediated by water, including molecular recognition, protein folding, proton transfer, etc. All this makes the determination of the 3D structures of biomolecules in their native environment a pivot objective for many fields of life science.

IR spectroscopy of cryogenically cold ions isolated in the gas phase is a recent approach that allows for solving the intrinsic structures of small to midsize biological molecules with high accuracy and conformational resolution. The cooling sharpens vibrational transitions, thus providing detailed spectroscopic signatures that can be used for stringent validation of the computed molecular structures. While, potentially, high in the gas phase, the spectroscopic resolution degrades drastically in aqueous solutions, where biomolecules naturally adopt their native structures. The inhomogeneity and dynamical nature of non-covalent interactions with surrounding water make IR spectra unacceptable for validating structural calculations. As a compromise between the solution phase, where the structures are native but high resolution is difficult to achieve, and the gas phase, where spectral resolution can be high but structures are intrinsic, one can interrogate in the gas phase microhydrated biomolecular ions. Already a few water molecules may lock the main features of native structures of the ions in such complexes, while their limited size yet enables the use of cold ion spectroscopy.^{1,2} One of the challenges of this approach is in producing a high number of hydrated complexes, where water molecules are only weakly coupled to a parent ion via hydrogen bonds. The most straightforward method to produce hydrated ions is to transfer them directly from a solution to the gas phase using gentle conditions in an electrospray ionization (ESI) source and subsequently cool the ions down in a cold trap.^{1,3-5} There is a narrow interval of ESI conditions,

where such complexes can survive while keeping the ionization efficiency high. The use of soft ESI has been demonstrated on many occasions for generating microsolvated complexes of organic^{1,5,6} and inorganic⁷⁻¹¹ ions, as well as for transferring to the gas phase intact protein-ligand complexes.¹²

An alternative and, perhaps, a more universal way of microhydration uses a gas-phase condensation of water molecules onto cryogenically pre-cooled ions.¹³⁻¹⁵ Recently, this approach was employed by the group of Garand for spectroscopic study of microhydrated protonated amino acid glycine¹⁶ and a small peptide triglycine.¹⁷ Here we use IR spectroscopy to investigate the same hydrated molecules, but extracted directly from aqueous solution to the gas phase by soft ESI and subsequently cooled to cryogenic temperature. We address the questions of whether the structure of the ions depends on the method of their hydration.

3.2. RESULTS AND DISCUSSION

Apart from the modified ion source, our experimental setup has been described elsewhere¹⁸ (see appendix chapter 3 for details). Briefly, the hydrated ions are produced from solution by a nano-ESI source and transferred through a metal capillary and three consecutive inline molecular skimmers to a room temperature octupole ion trap for accumulation and thermalization. The thermalized ionic complexes are mass-selected by a quadrupole mass-filter and then guided into a cold ($T=6\text{K}$) octupole trap,¹⁹ where they undergo photofragmentation by an IR laser pulse. The appearing fragments that lost one or two waters are detected by the second quadrupole MS.

Figure 3.1 (red traces) shows the IR spectra of GlyH^+ hydrated by the well-defined numbers (1 to 6) of water molecules that were retained on the amino acid. For comparison, the blue traces in the same figure reproduce the IR spectra earlier measured by the group of Garand for the same complexes ($n=1$ to 4), but produced in the gas phase by the cryogenic condensation.¹⁶ Note, that the relative intensities of the transitions in each of the two “action” spectra may differ due to the difference in the used methods of the measurement. The assignment of the peaks in these

spectra was based on the comparison with the computed spectra of the low-energy ionic structures that were calculated for the complexes by the same group.¹⁶ For $n=1$ the spectra of the complexes extracted from the solution and the complexes prepared by the condensation exhibit all the same transitions, except one extra peak at 3336 cm^{-1} in the spectrum of the latter.

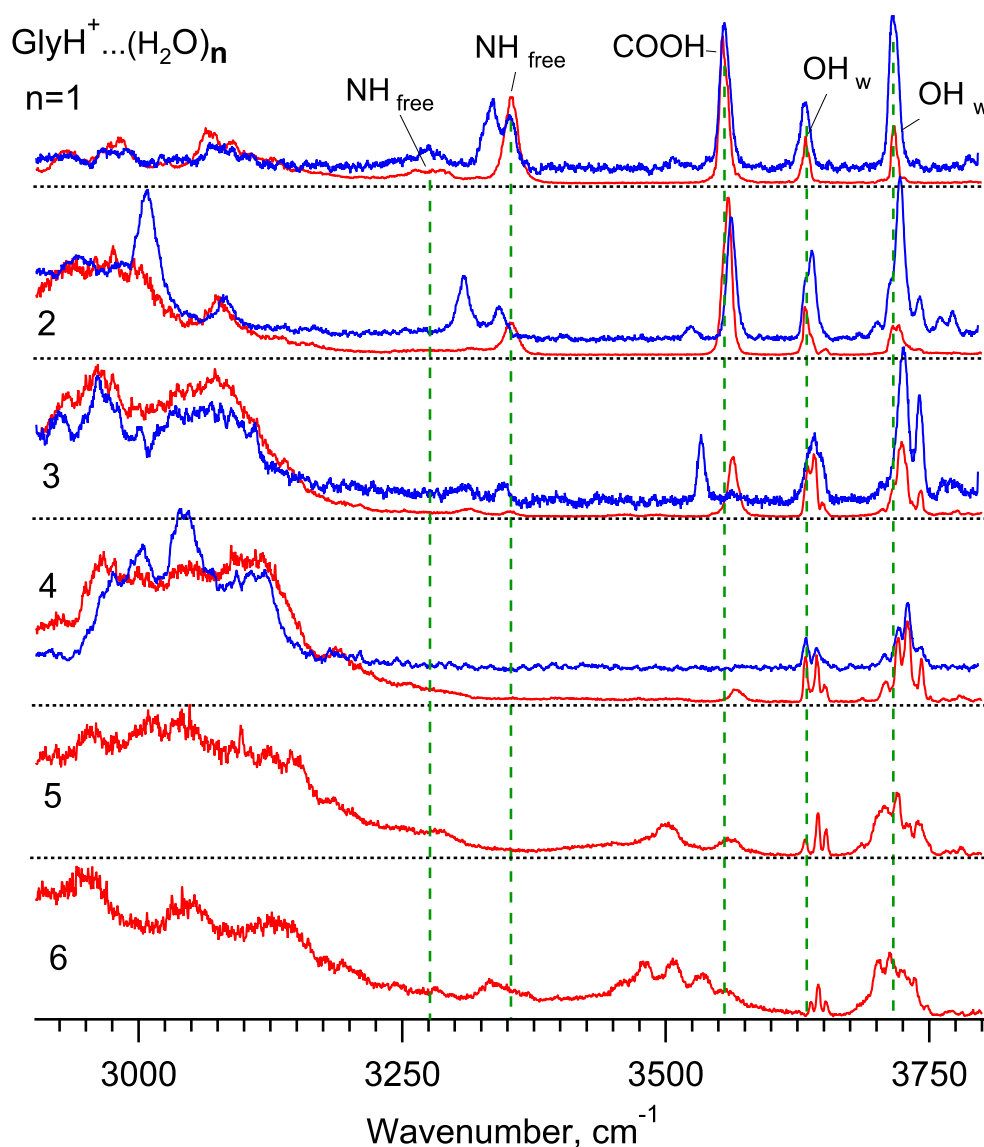


Figure 3.1. IR spectra of cold $\text{GlyH}^+\cdots(\text{H}_2\text{O})_n$ ($1 \leq n \leq 6$), produced directly from aqueous solution (red trace; this work) and by cryogenic condensation (blue trace, ref.¹⁶). The reproduced original blue traces were shifted by -4 cm^{-1} to match to the position of free OH-stretches of waters around 3700 cm^{-1} measured herein (red traces) with the accuracy of $\pm 0.2\text{ cm}^{-1}$. The blue traces are offset up for graphical clarity.

The structure associated with this peak was assigned to the most stable conformer of the singly hydrated GlyH⁺. This conformer is characterized by an intramolecular H-bond between the NH₃⁺ and C=O groups. The same strong H-bond is present in the most stable gas-phase structure of GlyH⁺, which was the only conformer observed experimentally.¹⁷ Consistently, the gas-phase hydration by condensation conserves its main structural features. In an aqueous solution, glycine does not have this intramolecular H-bond, because the surrounding water molecules fully solvate the charge and the hydrophilic functional groups of the ion.²⁰ We thus are to conclude that, the structure assigned to the peak at 3336 cm⁻¹ in the spectrum of GlyH⁺...H₂O complex produced in the gas phase is relevant not to the native solution phase, but rather to the intrinsic gas-phase structure of the amino acid. The lack of this conformer for the complexes produced from the solution provides direct evidence of that. The peak at 3355 cm⁻¹, which appears in the spectra of the complexes produced by both methods of hydration, was earlier assigned to the conformer that has no intramolecular H-bond between the NH₃⁺ and C=O groups.¹⁶ Competing with this bond, the first condensed or the last retained water molecule non-covalently binds to the protonated N-terminus, partially solvating its charge. Consistently, this is the only structure detected for the complexes extracted from the solution in our experiments (Figure 3.1, n=1, red trace). We may speculate, that at the mild cryogenic temperature of ~80 K used for condensation,¹⁵ a fraction of complexes residing in the gas-phase structures overcomes the barrier toward the solution-like structure of GlyH⁺. The rest of the complexes remain “kinetically” trapped in the gas-phase conformational state. This is a mirroring of the process, where a molecule isolated in the gas phase from solution remains “kinetically” trapped in a solution-like state.^{6,11,21,22} In opposite, retaining a single water molecule on GlyH⁺ conserves the main features of the solution-phase structure of the ion in the gas phase. This single lowest-energy structure remains stable despite the thermalization at room temperature for ~45 ms prior to the cryogenic cooling.

Retaining one more water molecule on GlyH⁺...H₂O complex almost does not change the resolved intense peaks assigned to weakly bound NH- and OH-stretches of glycine (compare the red traces for n=1 and 2 in Figure 1). A close inspection of the spectra reveals two spectroscopic manifestations of binding the second water to the N-terminus. The broad weak peak at ~ 3280 cm⁻¹, which for n=1 was earlier assigned to one of the two “free” NH stretches,¹⁶ disappears (for n=2). It, likely, shifts to the red due to the H-bond between this NH group and the second water molecule. In opposite, the solvation of the charge weakens the long-distance non-covalent interaction of the two termini. This becomes evident from a slight, but well detected 5.6 cm⁻¹ blue shift of the OH-stretch transition of the C-terminus. The spectroscopic signature of the two additional OH-stretches of the second water molecule in the spectrum for n=2 (Figure 3.1, n=2, red trace) is the slightly split broaden peak at 3722 cm⁻¹ and the small peak at 3652 cm⁻¹. In comparison with this, the spectrum of the same complex, GlyH⁺...(H₂O)₂, but produced in the gas phase by condensation¹⁶ (blue trace in Fig. 1, n=2), exhibits 4-5 additional peaks in the region of the absorption by free OH-stretches of water (3750-3780 cm⁻¹), as well as 3 peaks earlier assigned to the NH-stretches of GlyH⁺. This larger number of transitions reflects the existence of more than one conformer of the GlyH⁺...(H₂O)₂. The calculations and IR-IR double resonance spectroscopy, indeed, revealed two main conformers, which do not have the intramolecular H-bond between the two termini of GlyH⁺, as well as, at least, one minor conformer with this bond.¹⁶ The most abundant conformer of GlyH⁺...(H₂O)₂ produced in the gas-phase is associated with the two characteristic intense NH-stretch spectral transitions at 3007 and 3308 cm⁻¹ (Figure 3.1, n=2, blue trace).¹⁷ These peaks are not present, however, in the spectrum of the same complex, but produced from solution (red trace in Figure 3.1, n=2). This observation clearly demonstrates once more that, the gas-phase hydration “kinetically” traps the embedded molecules in their intrinsic structures.

The IR transitions of the second main conformer¹⁶ of the gas-phase produced GlyH⁺...(H₂O)₂ look similar to the spectrum of this complex

prepared from solution, except the 14 cm^{-1} redshift of the characteristic NH-stretch peak at 3353 cm^{-1} . We may suggest that the structure of this conformer is similar to the structure of the only conformer (conformational family) of the complexes prepared from solution, provided the shift can be explained by the influence of the D_2 tag, which was attached to the complexes (produced by the condensation) for their detection.^{23,24}

The influence of the D_2 tag was also invoked by the authors¹⁶ to explain the 30 cm^{-1} redshift of the (CO)O-H stretch transition in the spectrum of $\text{GlyH}^+\cdots(\text{H}_2\text{O})_3$ relative to $\text{GlyH}^+\cdots(\text{H}_2\text{O})_2$ (blue trace in Figure 3.1, $n=2, 3$). With this reservation, the spectra of GlyH^+ triply hydrated by the two methods look very similar. The calculations suggested that the third water molecule energetically almost equally binds either to the C-terminus, inducing a large calculated redshift of the (CO)O-H stretch transition, or to the protonated N-terminus to fully solvate the charge.¹⁶ The former binding results in a reduction of the intensity of the characteristic peak of the (CO)O-H stretch in both spectra for $n=3$. The continuing charge solvation becomes evident from the frequency of this transition, which shifts to blue upon retaining one more water on $\text{GlyH}^+\cdots(\text{H}_2\text{O})_2$ (red traces in Figure 3.1, $n=2$ and 3). We, thus, may suggest that the structures of $\text{GlyH}^+\cdots(\text{H}_2\text{O})_3$ complexes prepared by the two methods become similar and are both native-like.

Retaining one more water molecule to generate $\text{GlyH}^+\cdots(\text{H}_2\text{O})_4$ from solution further decreases the intensity of the peak, which was assigned to the free (CO)O-H stretch,¹⁶ but doesn't change anymore its frequency. The implication of this, confirmed by the calculations, is that the charge has been already fully solvated by the first three waters, such that the fourth molecule may now attach to the H atom of the C-terminus. The last three retained water molecules, therefore, already form the first solvation shell of the protonated N-terminus. The remaining intensity of the (CO)O-H stretch peak suggests the presence of another conformer, earlier predicted theoretically,^{25,26} in which the fourth water molecule does not bind to the C-terminus, but to the first solvation shell. This conformer is

absent, however, for the same $\text{GlyH}^+\cdots(\text{H}_2\text{O})_4$ complex prepared by the condensation in the gas phase.

The spectra of GlyH^+ hydrated by 5 and 6 waters become similar for both methods of their production (Figure 3.S2, $n=5$ and 6). In all the calculated conformers, the fifth and the sixth water molecules belong to the second solvation shell. We thus may propose that the hydration of GlyH^+ by 5-6 water molecules using any of the two methods results in the native structures of this small biomolecule. The identification of such structures, which lack an intramolecular H-bond, becomes evident already for the singly-hydrated GlyH^+ released from solution, but remains ambiguous in the case of the gas-phase prepared complexes until as many as four to five water molecules were condensed onto the amino acid.

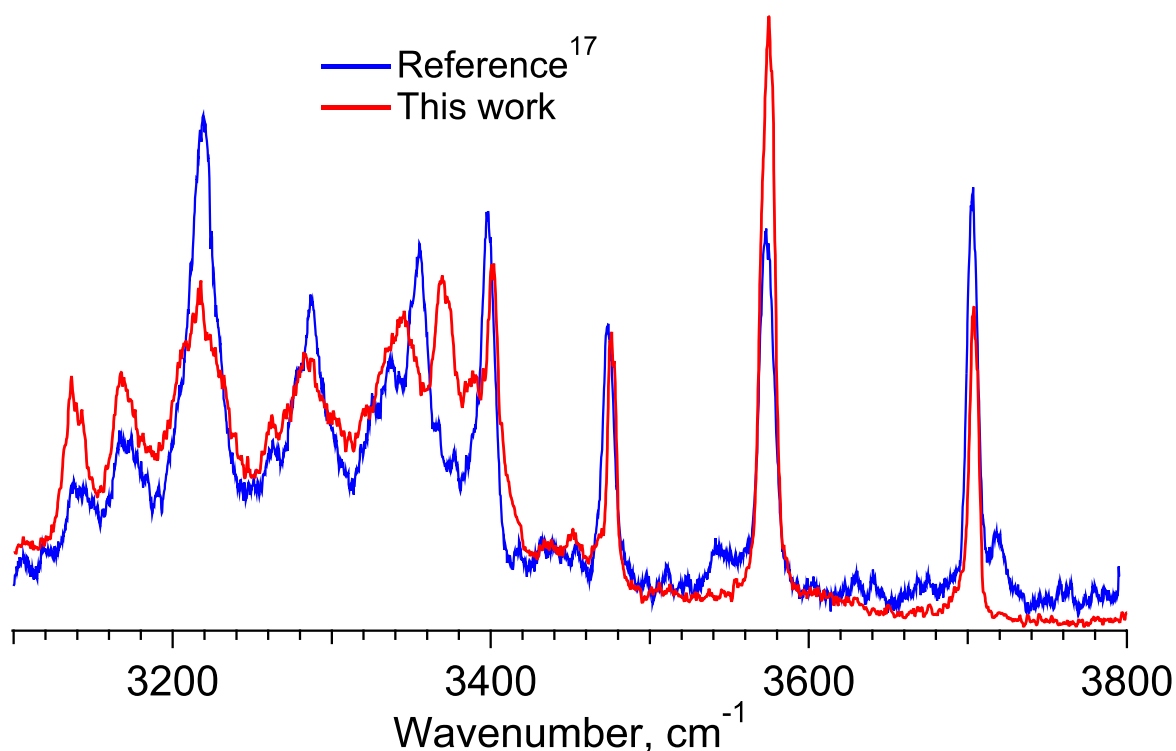


Figure 3.2. IR vibrational spectra of $\text{Gly}_3\text{H}^+\cdots(\text{H}_2\text{O})$. The red and blue traces correspond to the complexes produced from an aqueous solution and in the gas phase, respectively. The red trace was recorded by detecting Gly_3H^+ as a photofragment. The blue spectrum was measured by the group of Garand.¹⁷

The *cis* configuration of the amide group is also atypical for peptides, except those that contain proline residues. Condensation of a single water molecule to the gas phase conformers, indeed, made the *trans* conformer energetically favorable and reduced the population of the *cis* conformer to 10%. This conformer is naturally not present at all for the same $\text{Gly}_3\text{H}^+\cdots\text{H}_2\text{O}$ complex extracted from the solution. The only detected conformer with one retained water molecule exhibits all the same characteristic IR transitions (Figure 3.2) that were assigned to the main *trans* conformer for the complexes prepared by the condensation. The characteristic IR transitions at 3718 cm^{-1} and 3541 cm^{-1} (Figure 3.2, blue trace), which were earlier assigned to the *cis* conformer prepared by the condensation, are clearly absent in our IR spectrum (Figure 3.2, red trace). Apart from this, the only difference in the position of the peaks in the spectra of the differently prepared complexes is a distinct peak at 3369.8 cm^{-1} that appears only in the case of solution-prepared $\text{Gly}_3\text{H}^+\cdots\text{H}_2\text{O}$. We may speculate that this feature can be attributed either to the certain structural difference of the differently prepared complexes or to the influence of tagging in the case of the gas-phase hydration.

3.3. CONCLUSION

In conclusion, retaining a single water molecule on GlyH^+ already allows for the determination of its native-like structure, which changes only little upon retaining more waters (Table 3.S1). In comparison, the cryogenic condensation may produce some additional conformers of the amino acid, where the intrinsic structures of the biomolecule are “kinetically” trapped. This makes the identification of the native structures among all the available conformers ambiguous. Upon increasing the number of the attached water molecules, the method of condensation produces, essentially, the same conformers as the structures extracted directly from the solution. Because of the large number of vibrational degrees of freedom and increasing spectral congestion, the structures of these larger complexes are more difficult to solve and to validate, however.

Finally, for larger gas-phase protonated peptides, the increased number of hydrophilic groups will increase the probability for a charge to be self-solvated;¹³ more H-bonds between these groups will be deeply buried into the molecule. These strong intramolecular interactions may compete with a subsequent microhydration in the gas phase, protecting peptides in their “kinetically” trapped intrinsic states. We may speculate that the difference between the structures of microsolvated peptides prepared from solution and formed in the gas will increase for larger molecules.

REFERENCES

1. 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.
2. Roy, T. K.; Nagornova, N. S.; Boyarkin, O. V.; Gerber, R. B. A Decapeptide Hydrated by Two Waters: Conformers Determined by Theory and Validated by Cold Ion Spectroscopy. *J. Phys. Chem. A* **2017**, *121*, 9401-9408.
3. Mercier, S. R.; Boyarkin, O. V.; Kamariotis, A.; Guglielmi, M.; Tavernelli, I.; Cascella, M.; Rothlisberger, U.; Rizzo, T. R. Microsolvation Effects on the Excited-State Dynamics of Protonated Tryptophan. *J. Am. Chem. Soc.* **2006**, *128*, 16938-16943.
4. Spieler, S.; Duong, C. H.; Kaiser, A.; Duensing, F.; Geistlinger, K.; Fischer, M.; Yang, N.; Kumar, S. S.; Johnson, M. A.; Wester, R. Vibrational Predissociation Spectroscopy of Cold Protonated Tryptophan with Different Messenger Tags. *J. Phys. Chem. A* **2018**, *122*, 8037-8046.
5. Chang, T. M.; Chakrabarty, S.; Williams, E. R. Hydration of Gaseous m-Aminobenzoic Acid: Ionic vs Neutral Hydrogen Bonding and Water Bridges. *J. Am. Chem. Soc.* **2014**, *136*, 10440-10449.
6. Silveira, J. A.; Fort, K. L.; Kim, D.; Servage, K. A.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H. From Solution to the Gas Phase: Stepwise Dehydration and Kinetic Trapping of Substance P Reveals the Origin of Peptide Conformations. *J. Am. Chem. Soc.* **2013**, *135*, 19147-19153.
7. Asmis, K. R.; Neumark, D. M. Vibrational Spectroscopy of Microhydrated Conjugate Base Anions. *Accounts Chem. Res.* **2012**, *45*, 43-52.
8. Inokuchi, Y.; Ebata, T.; Rizzo, T. R.; Boyarkin, O. V. Microhydration Effects on the Encapsulation of Potassium Ion by Dibenzo-18-Crown-6. *J. Am. Chem. Soc.* **2014**, *136*, 1815-1824.
9. Sun, S. T.; Jiang, L.; Liu, J. W.; Heine, N.; Yacovitch, T. I.; Wende, T.; Asmis, K. R.; Neumark, D. M.; Liu, Z. F. Microhydrated Dihydrogen Phosphate Clusters Probed by Gas Phase Vibrational Spectroscopy and First Principles Calculations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25714-25724.

10. Mitra, S.; Duong, C. H.; McCaslin, L. M.; Gerber, R. B.; Johnson, M. A. Isomer-Specific Cryogenic Ion Vibrational Spectroscopy of the D₂ Tagged Cs⁺(HNO₃)(H₂O)_{n=0-2} Complexes: Ion-Driven Enhancement of the Acidic H-bond to Water. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4501-4507.
11. Hebert, M. J.; Russell, D. H. Tracking the Structural Evolution of 4-Aminobenzoic Acid in the Transition from Solution to the Gas Phase. *J. Phys. Chem. B* **2020**, *124*, 2081-2087.
12. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85*, 2-9.
13. Liu, D.; Wyttenbach, T.; Barran, P. E.; Bowers, M. T. Sequential Hydration of Small Protonated Peptides. *J. Am. Chem. Soc.* **2003**, *125*, 8458-8464.
14. Xu, S.; Smith, J. E. T.; Weber, J. M. Hydration of a Binding Site with Restricted Solvent Access: Solvatochromic Shift of the Electronic Spectrum of a Ruthenium Polypyridine Complex, One Molecule at a Time. *J. Phys. Chem A* **2016**, *120*, 7650-7658.
15. Voss, J. M.; Fischer, K. C.; Garand, E. Accessing the Vibrational Signatures of Amino Acid Ions Embedded in Water Clusters. *J. Phys. Chem. Lett.* **2018**, *9*, 2246-2250.
16. Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J. Phys. Chem. A* **2019**, *123*, 3355-3366.
17. Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly₃H⁺·(H₂O). *J. Phys. Chem A* **2018**, *122*, 8213-8221.
18. Boyarkin, O. V. Cold Ion Spectroscopy for Structural Identifications of Biomolecules. *Int. Rev. Phys. Chem.* **2018**, *37*, 559-606.
19. Boyarkin, O. V.; Kopysov, V. Cryogenically Cooled Octupole Ion Trap for Spectroscopy of Biomolecular Ions. *Rev. Sci. Instrum.* **2014**, *85*, 033105.
20. Kayi, H.; Kaiser, R. I.; Head, J. D. A Theoretical Investigation of the Relative Stability of Hydrated Glycine and Methylcarbamic Acid—from Water Clusters to Interstellar Ices. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4942-4958.

21. Patrick, A. L.; Cismesia, A. P.; Tesler, L. F.; Polfer, N. C. Effects of ESI Conditions on Kinetic Trapping of the Solution-Phase Protonation Isomer of p-Aminobenzoic Acid in the Gas Phase. *Int. J. Mass Spectrom.* **2017**, *418*, 148-155.
22. Zhang, H.; Cao, W.; Yuan, Q.; Zhou, X.; Valiev, M.; Kass, S. R.; Wang, X.-B., Cryogenic “Iodide-Tagging” Photoelectron Spectroscopy: A Sensitive Probe for Specific Binding Sites of Amino Acids. *J. Phys. Chem. Lett.* **2020**, *11*, 4346-4352.
23. Kamrath, M. Z.; Garand, E.; Jordan, P. A.; Leavitt, C. M.; Wolk, A. B.; Van Stipdonk, M. J.; Miller, S. J.; Johnson, M. A. Vibrational Characterization of Simple Peptides Using Cryogenic Infrared Photodissociation of H₂-Tagged, Mass-Selected Ions. *J. Am. Chem. Soc.* **2011**, *133*, 6440-6448.
24. Masson, A.; Williams, E. R.; Rizzo, T. R. Molecular Hydrogen Messengers Can Lead to Structural Infidelity: A Cautionary Tale of Protonated Glycine. *J. Chem. Phys.* **2015**, *143*, 104313.
25. Wei, Z.; Chen, D.; Zhao, H.; Li, Y.; Zhu, J.; Liu, B. Ab Initio Investigation of the First Hydration Shell of Protonated Glycine. *J. Chem. Phys.* **2014**, *140*, 085103.
26. Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. Microhydration of Protonated Glycine: An ab initio Family Tree. *J. Phys. Chem. B* **2008**, *112*, 2430-2438.
27. Pereverzev, A. Y.; Szabó, I.; Kopysov, V. N.; Rosta, E.; Boyarkin, O. V. Gas-Phase Structures Reflect the Pain-Relief Potency of Enkephalin Peptides. *Phys. Chem. Chem. Phys.* **2019**, *21*, 22700-22703.

APPENDIX CHAPTER 3

Experimental and chemicals

The hydrated ions are produced in the gas phase from solution by a nano-electrospray ionization (n-ESI) source and transferred through a metal capillary and three consecutive inline molecular skimmers to a room temperature octupole ion trap for accumulation and thermalization. The thermalized ionic complexes are mass-selected by a quadrupole mass-filter and guided into a cold octupole trap,²⁴ which is kept at 6 K. The stored in the trap complexes are cooled down to ~10 K in collisions with He buffer gas²⁴ and then irradiated by a pulse of IR light (6 ± 2 mJ/pulse, 1 cm^{-1} spectral linewidth), produced by an optical parametric oscillator (OPO, Laser Vision). The absorption of the light results in the loss by a complex of a few water molecules. The remaining parent and the reduced fragment complexes are released from the trap and detected by a quadrupole mass spectrometer, which is tuned to transmit alternatively either one of the fragment or parent complexes. We average 10 measurements at each IR wavelength at a repetition rate of 10 Hz. Each spectrum was recorded 2-3 times to ensure its reproducibility. A wavelength meter measures the wavenumber of the Nd:YAG laser and of the signal wave of the IR OPO, pumped by this laser, thus providing the wavenumber of the generated by the difference frequency mixing IR light with $\pm 0.2\text{ cm}^{-1}$ accuracy.

The spectra of $\text{GlyH}^+(\text{H}_2\text{O})_n$ complexes have been measured by detecting the fragments $\text{GlyH}^+(\text{H}_2\text{O})_m$ ($m = n - 2$ for $n \geq 2$ and $m=0$ for $n=1$); and the spectrum of $\text{Gly}_3\text{H}^+(\text{H}_2\text{O})$ has been measured by detecting the bare peptide.

Glycine were purchased from ACROS Organics (> 99% purity) and used without further purification. All the solvents of HPLC grade and acetic acid of >99% purity are from Sigma-Aldrich.

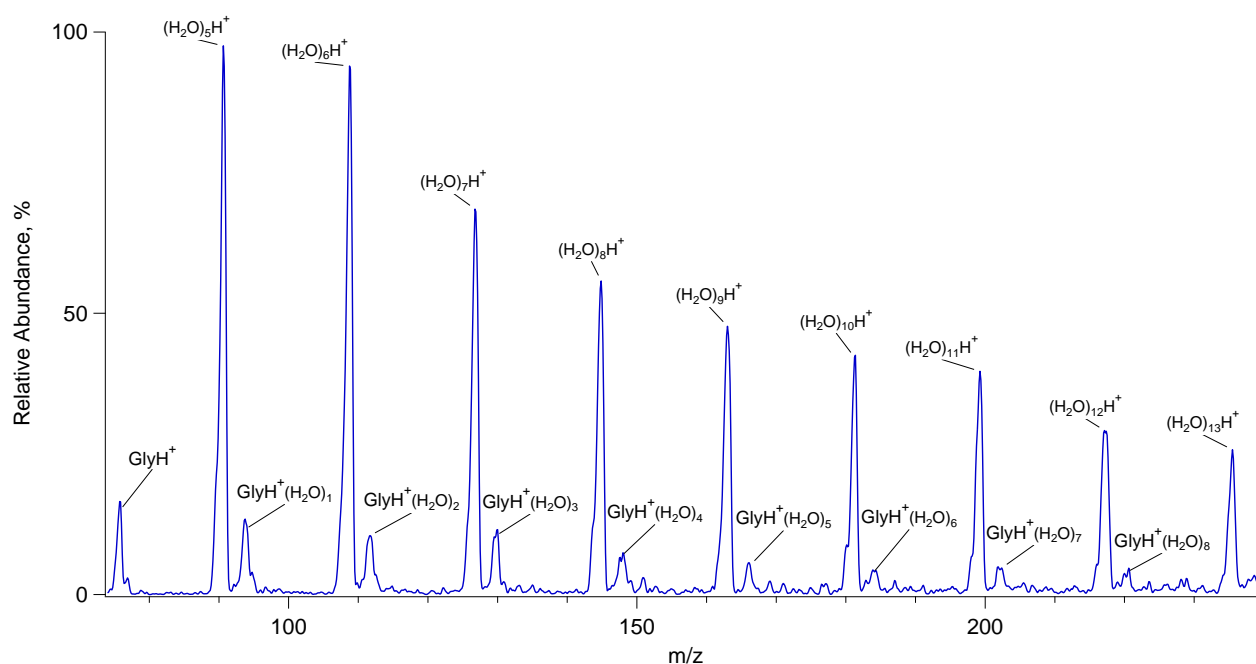


Figure 3.S1. Mass-spectrum, showing distribution of GlyH⁺(H₂O)_n complexes, produced by a nano-electrospray under the conditions, not optimized for a particular number of water molecules n in the complexes. Adjustments of voltages allow for some control of the complex size distribution.

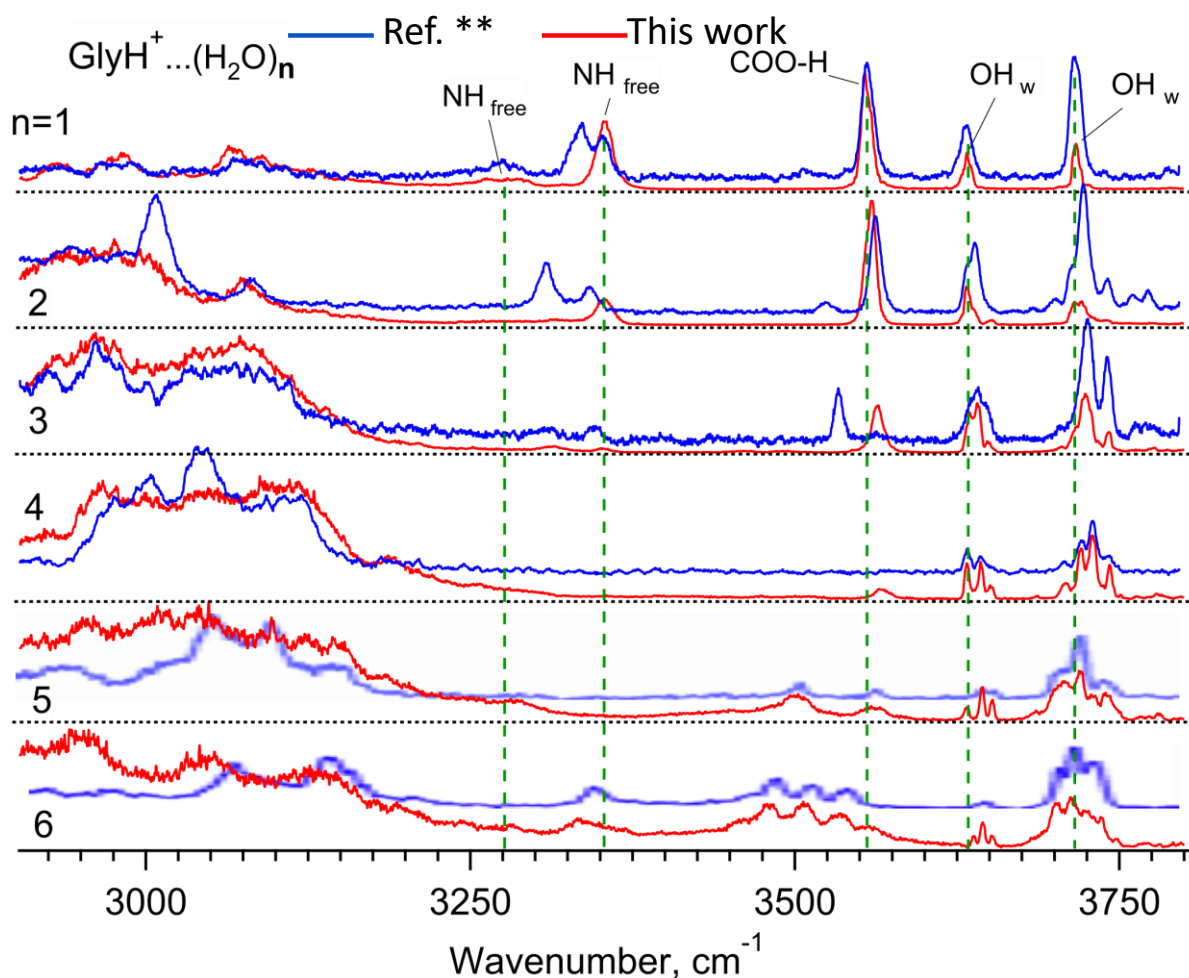
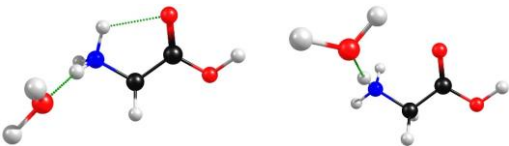
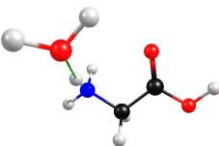
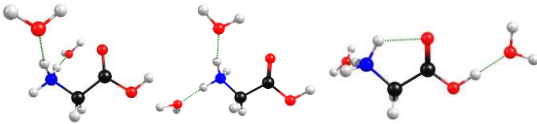
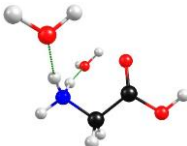
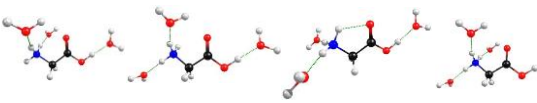
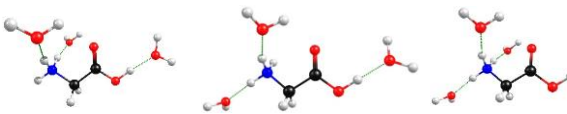
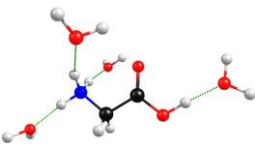
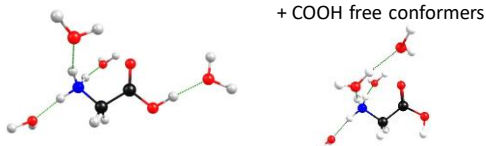
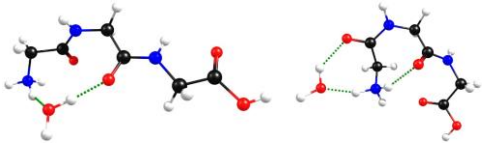
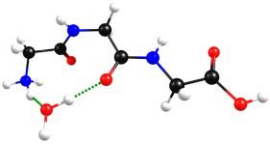


Figure 3.S2. IR spectra of cold $\text{GlyH}^+(\text{H}_2\text{O})_n$ ($1 \leq n \leq 6$), produced directly from aqueous solution (red trace; this work) and by cryogenic condensation (blue trace, ref.²¹). Note, that the reproduced original blue traces were shifted by -4 cm^{-1} to match to the manifold of the high-frequency transitions (free OH-stretches of waters around 3700 cm^{-1}) measured herein (red traces) with the accuracy of $\pm 0.2 \text{ cm}^{-1}$. The blue traces are offset up for graphical clarity.

Table 3.S1. Comparison of structures for the complexes prepared by retaining from solution and by cryogenic condensation of water molecules. All the structures were calculated by the group of Garand (ref.^{16, 17}).

	Condensed clusters	Retained clusters
n	GlyH ⁺ (H ₂ O) _n	
1		
2		
3		
4		
	Gly ₃ H ⁺ (H ₂ O) _n	
1		

CHAPTER 4.

Interplay of H-Bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates

OVERVIEW

The tremendous isomeric diversity of carbohydrates enables a wide range of their biological functions, but makes identification and study of these molecules difficult. We investigated the ability of intermolecular interactions to communicate structural specificity of carbohydrates to protonated aromatic molecules in non-covalent complexes, 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 electronic excitation spectra. Such specific response of the aromatic molecules to isoforms of carbohydrate is fine-tuned by interplay of the diverse involved non-covalent bonds. This finding enables gas-phase identification and relative quantification of any isoforms of oligosaccharides in their solution mixtures using 2D UV-MS fingerprinting technique.

The results presented in this chapter were published in:

E. Saparbaev, V. Kopysov, R. Yamaletdinov, A. Pereverzev, O. V. Boyarkin. Interplay of H-Bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates. *Angew. Chem. Int. Ed.* 2019, 58, 7346-7350, DOI: <https://doi.org/10.1002/anie.201902377>

4.1. INTRODUCTION

Carbohydrates are by far the most abundant organic molecules present in living organisms.¹ Existing in myriads of structural forms, many of which are isomeric, carbohydrates may function as messenger and recognition molecules in signaling the type and the state of living cells in complex chains of biological processes.² Based on this, cancer cells can be identified, for instance, by detecting their specific membrane glycoproteins,^[3] which may differ from those of the healthy cells by the isoform of the glycan only; viruses and bacteria adhere to appropriate cells for invasion by selective binding to specific membrane glycans,⁴ distinguishing them from many other structurally similar (e.g. isomeric) saccharides; etc. These and many other examples illustrate the importance of comprehensive structural studies of carbohydrates, including the identification of their isoforms.⁵

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 aldohexopyranose 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 OH 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. Human milk alone, for instance, contains, at least, about 200 different oligosaccharides, many of which are essential not only for supplying energy to newborns, but also for their antibacterial defense.⁶

The enormous isomeric diversity of carbohydrates makes their identification and study exceptionally challenging. Most common analytical methods of isomeric identifications of carbohydrates, such as chromatography⁷ and, more recently, ion mobility⁸, demonstrate high capability in separating small isomeric saccharides, although become inefficient in separating complex mixtures of oligosaccharides without

their chemical derivatizations.⁹ The methods rely on rates of hindered travel of molecules and ions, which are not fundamental values and are known to be sensitive to the experimental conditions and instrumentation. 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 in the gas phase (e.g., IRMPD¹⁰ or IR-induced fragmentation of tagged saccharides¹¹) was able to distinguish, yet without quantification, many stereoisomers of small oligosaccharides. UV photofragmentation spectroscopy of cryogenically cooled ions in conjunction with mass spectrometric detection of the charged fragments recently demonstrated highly accurate identifications and quantification of protonated isomeric peptides and drug molecules.^[12] A straightforward extension of this approach to carbohydrates is, however, challenging due to the fact that they do not absorb in visible and UV. In addition, carbohydrates are not prone to protonation. Both problems, potentially, can be solved by sensing non-covalent complexes of carbohydrates with protonated aromatic molecules,¹³ relying on the fact that UV absorption of many aromatic molecules, including aromatic amino acids, is sensitive to intra- and intermolecular hydrogen bonds.¹⁴ But, how sensitive are the intermolecular non-covalent bonds in communicating structural details between molecules? Will, for instance, a flip of a single OH group in a large isomeric oligosaccharide noticeably change UV spectrum of a weakly bound aromatic? Herein we demonstrate how a network of non-covalent bonds between an aromatic ion and a carbohydrate analyte molecule enables specific bilateral responses of the complex to structural differences in isomeric analytes. The responses, detected by a combination of the gas-phase UV spectroscopy and mass spectrometry reveal key functions of the non-covalent bonds in communicating structural differences of the analytes. We explore the use of such networks for identification and quantification of *all* types of isoforms of carbohydrates non-covalently bound in solution to protonated aromatic molecules.

4.2. RESULTS AND DISCUSSION

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.^{12a, 15}

Figure 4.1A shows a two-dimensional color-coded photofragmentation mass spectrum (2D UV-MS;^{12a} 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 4.1B and 4.1C 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 4.1D and 4.1E 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. The measured and the calculated (for the most stable structure) IR spectra of GlcNAc-TyrH⁺ complex exhibit very good match (Figure 4.S1), validating the structure in Fig. 4.1E as the only highly abundant conformer of the complex present in the cold trap at our experimental conditions. We are thus confident in the appropriate level of accuracy of our computations and accept the calculated lowest energy structure of the GalNAc-TyrH⁺ as a reasonable model for analysis of non-covalent interactions in this complex. In both isomers, the two strong bonds between NH₃⁺ of TyrH⁺ and oxygen atoms of a saccharide allow the complex to survive the harsh conditions of electrospray ionization. Similar

pattern of $\text{NH}_3^+\text{-O}$ interaction was revealed for protonated N-terminus of Tyr residue in some peptides of known geometry (Fig. 4.S2).¹⁷ 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)^[18] appear almost at that of neutral Tyr (280.9 nm),¹⁹ 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.

An important change in UV absorption of the aromatic is induced by 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.^{14a, 14b} Conversely, when the hydroxyl oxygen is the proton acceptor, like in the case of $\text{TyrH}^+\text{-GalNAc}$, its increased electronegativity near equally shifts the UV absorption of phenol to the blue.^[14a, 20] Involvement of phenol to both types of H-bonds, like in the case of $\text{TyrH}^+\text{-GlcNAc}$, redshifts its absorption by ~ 0.8 nm.^{14b} Consistently with this picture, the UV band origin in $\text{TyrH}^+\text{-GalNAc}$ complex is shifted from that in $\text{TyrH}^+\text{-GlcNAc}$ to the blue by as much as 4.2 nm. In addition to these two H-bonds, both complexes exhibit weaker $\text{CH-}\pi$ interactions. Their relative strength is governed by the electropositive character of the C-H bonds, which differs between isomeric saccharides.²¹ Thus, $\text{CH-}\pi$ bonds may also contribute to the distinguishing 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; Table 4.S1). This modification removes the H-bonds, but enables coupling of an OH in the saccharides to the aromatic π -system (Fig. 4.S3). $\text{CH-}\pi$ and $\text{OH-}\pi$ 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 $\text{GalNAc/GlcNAc-MePheH}^+$ complexes (Fig. 4.S3), $\text{OH-}\pi$ 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. 4.S4). The interplay of OH- π and CH- π couplings thus makes the UV spectroscopic signatures of the two epimers quite different. The calculated networks of non-covalent interactions in figures 4.1D and 4.1E are thus fully consistent with the observed difference in the onsets of UV absorption for the respective complexes. Similarly, these interactions in the complexes of *D*- and *L*-enantiomers of glucose with *L*-PheH⁺ modify the UV absorption of this chiral aromatic differently (Figure 4.S5), allowing spectroscopic identification of the two enantiomeric monosaccharaides.

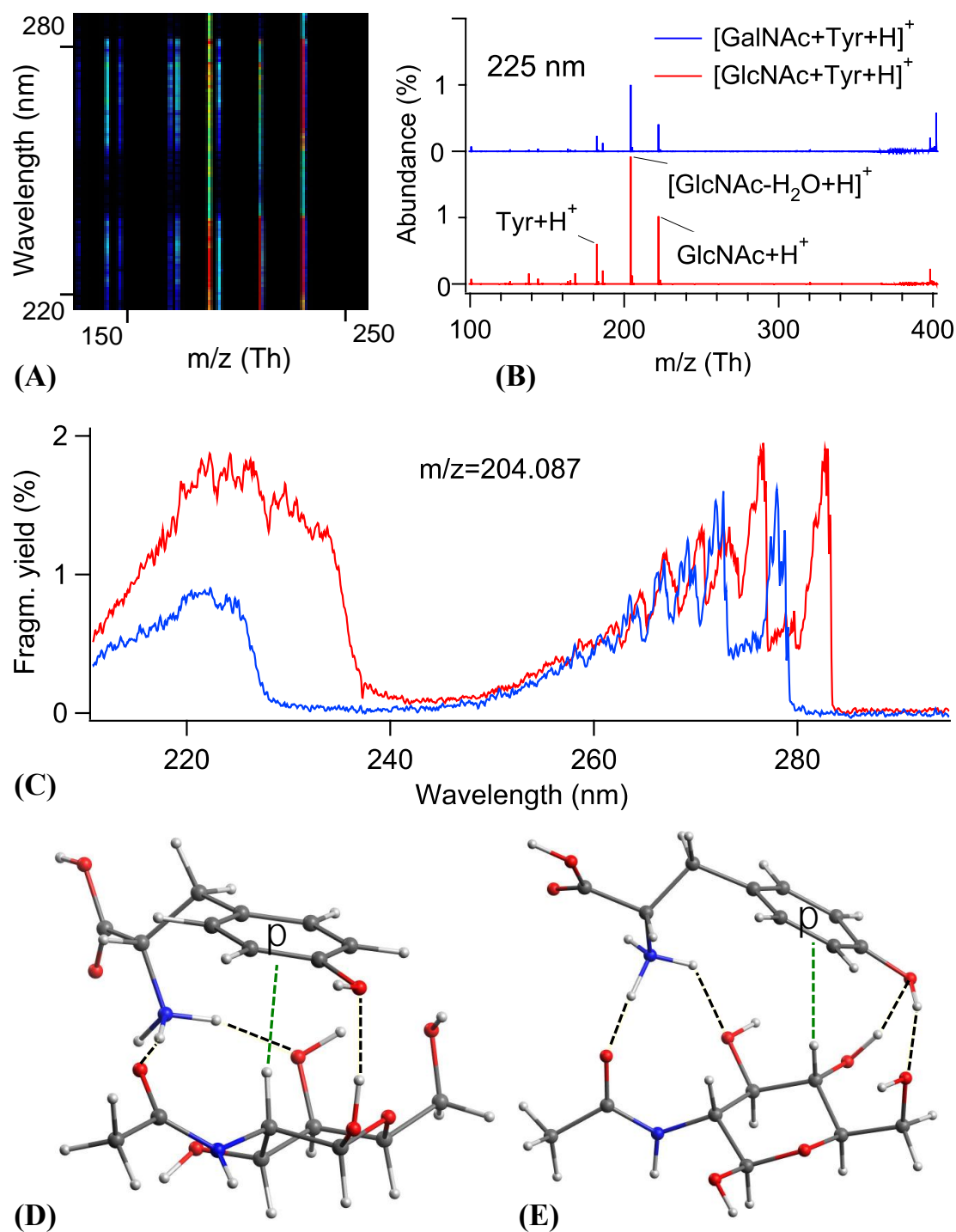


Figure 4.1. (A) Color-coded 2D UV-MS spectrum of GlcNAc-TyrH⁺ complex cooled to ~10 K¹⁶. Examples of mass (B) UV and photofragment (C) spectra, generated by slicing 2D UV-MS fingerprints of GalNAc-TyrH⁺ (blue trace) and of GlcNAc-TyrH⁺ (red trace) at m/z=204.087 Th (most intense fragment) and at 225 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.

4.3. CONCLUSION

Overall, the interplay of up to five non-covalent interactions between the aromatic molecules and the monosaccharides 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 4.2A, 4.3A, 4.S5 and 4.S6). In all the calculated structures (e.g., Figs. 4.1D, 4.1E and 4.S3), C-terminus of Tyr and MePhe is not involved in H-bonding and therefore may not contribute to this sensitivity. Figure 4.2 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 (Table 4.S1). 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 aromatic molecules besides four amino acids can also be highly sensitive in the identification of carbohydrates.

Although aromatic amines and amino acids, 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 4.1B, 4.2B and 4.3B). These observations suggest a proton and energy transfer from aromatic molecules to saccharides in the electronic ground^{13a} and/or, upon UV excitation, excited state^[22] of the complexes. Protonated aromatic molecules are the most abundant photofragments observed for mono- (e.g., *D-/L-* glucose, Fig. 4.S5) and disaccharides (Figure 4.S6) that do not contain NAc groups. This suggests a crucial role for the NAc group in accommodating the proton that is shared with aromatic molecules in the complexes. The rich networks of intermolecular non-covalent bonds in the complexes facilitate the sensor to analyte proton transfer²³, thus allowing a synergy of mass spectrometry and spectroscopy in identification of carbohydrates. Regarding all these observations, we don't rule out that in carbohydrate-protein non-covalent

complexes a proton transfer from a reporter aromatic residue to the analyzing carbohydrate and the changes in energetics of electronic levels of the sensor residue might contribute to finer recognition of carbohydrates by proteins in complex signaling processes.

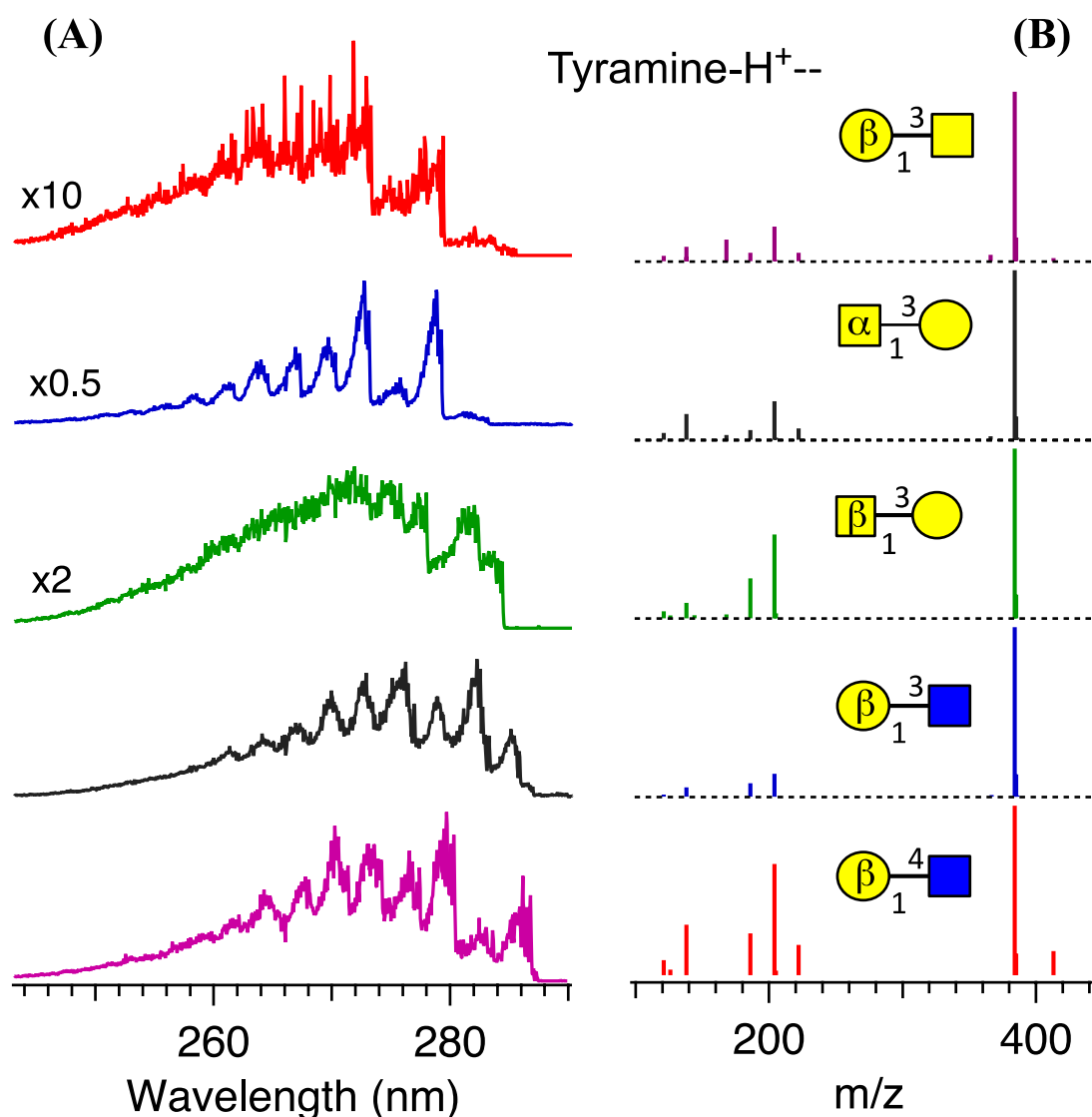


Figure 4.2. Photofragment (A) UV and (B) mass spectra, generated by integrating over m/z and wavelength, respectively, the 2D UV-MS fingerprints of five isomeric (m=383.16 Da) N-acetylated disaccharides bound in solution with protonated tyramine. The spectra are labeled by the standard symbolic representation of oligosaccharides (see Table 4.S2).

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 4.3A and 4.3B 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 heptasaccharides, 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 sensor can sense a structure only locally, different conformers of a complex collectively may reflect the structural details of the whole analyte molecule.

The differences in the UV-MS identities of isomeric carbohydrates in complexes with aromatic molecules allowed for relative quantification of the isomers in their solution mixtures using the method of the 2D UV-MS data array decomposition.^{12a, 24} It is worth mentioning that this library-based method intrinsically provides the relative concentrations of isomers in solution, but not in the gas phase, and is insensitive to (likely) structural differences between the same molecules in the two phases. As a test, here we applied the method to the studied tetra- and heptasaccharides. For instance, for a 1:1:2 solution mixture of three isomeric tetrasaccharides (Fig. 4.3A), 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 4.3B. These examples of accurate quantifications illustrate the high structural sensitivity of intermolecular non-covalent bonds in aromatic-carbohydrate complexes. The sensitivity arises from interplay of the diverse involved non-covalent bonds and enables identification and quantification of *all* types of isoforms of mono- to oligosaccharides, which is challenging for most common separation techniques. This approach can

be used for assignment of isomeric carbohydrates separated by chromatography or ion mobility, or as an alternative of these techniques, when a reliable separation cannot be achieved. Isomers of other types of biomolecules (e.g. lipids), which are difficult to separate by these techniques and which do not exhibit resolved UV spectra, potentially, can be also identified by the 2D UV-MS fingerprinting of their non-covalent complexes with suitable aromatic molecules using the same universal UV-MS instrumentation.

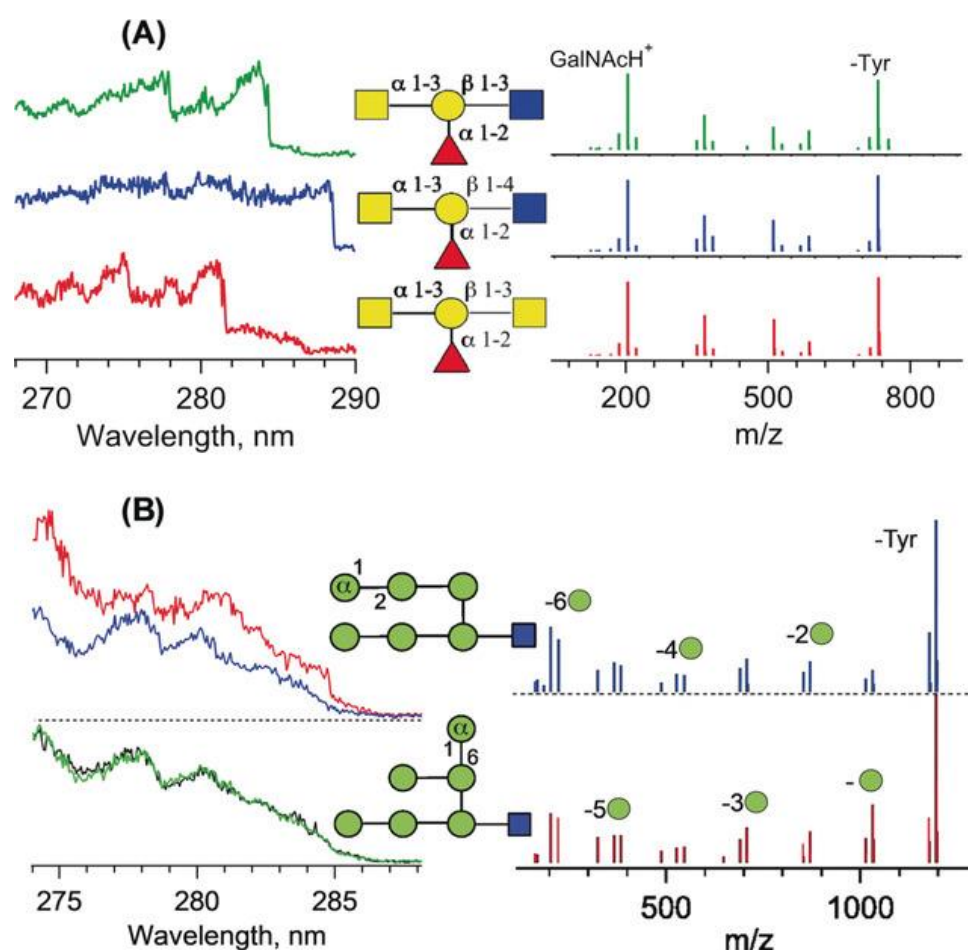


Figure 4.3. Photofragment UV and mass spectra of (A) three isomeric tetrasaccharides, and (B) two heptasaccharides (Man₆GlcNAc (II) and Man₆GlcNAc (I), red and blue traces, respectively), bound to TyrH⁺. Green trace on the left in (B) shows the UV spectrum of the 1:2 mixture of the two heptasaccharides, 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.

REFERENCES

1. L. G. Wade, *Organic chemistry*, Prentice Hall, Upper Saddle River, NJ, **1999**.
2. a) R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683-720; b) N. K. Broeker, D. Andres, Y. Kang, U. Gohlke, A. Schmidt, S. Kunstmann, M. Santer, S. Barbirz, *Perspect. Sci.* **2017**, *11*, 45-52; c) J. L. Asensio, A. Ardá, F. J. Cañada, J. Jiménez-Barbero, *Accounts Chem. Re.s* **2013**, *46*, 946-954.
3. H. Ghazarian, B. Idoni, S. B. Oppenheimer, *Acta Histochem.* **2011**, *113*, 236-247.
4. N. Sharon, H. Lis, *Science* **1989**, *246*, 227-234.
5. A. Lakshminarayanan, M. Richard, B. G. Davis, *Nat. Rev. Chem.* **2018**, *2*, 148-159.
6. a) L. Bode, *Glycobiology* **2012**, *22*, 1147-1162; b) T. Hennet, A. Weiss, L. Borsig, *Swiss Med Wkly* **2014**, *144*, w13927.
7. a) M. J. Kailemia, L. R. Ruhaak, C. B. Lebrilla, I. J. Amster, *Anal. Chem.* **2014**, *86*, 196-212; b) A. A. Ghfar, S. M. Wabaidur, A. Y. B. H. Ahmed, Z. A. Alothman, M. R. Khan, N. H. Al-Shaalan, *Food Chem.* **2015**, *176*, 487-492.
8. a) D.-S. Lee, C. Wu, H. H. Hill, *J. Chromatogr. A* **1998**, *822*, 1-9; b) C. J. Gray, B. Thomas, R. Upton, L. G. Migas, C. E. Eyers, P. E. Barran, S. L. Flitsch, *Biochim. Biophys. Acta (BBA) - Gen. Subj.* **2016**, *1860*, 1688-1709; c) M. M. Gaye, G. Nagy, D. E. Clemmer, N. L. Pohl, *Anal. Chem.* **2016**, *88*, 2335-2344.
9. J. Schenk, G. Nagy, N. L. B. Pohl, A. Leghissa, J. Smuts, K. A. Schug, *J. Chromatogr. A* **2017**, *1513*, 210-221.
- [10] a) N. C. Polfer, J. J. Valle, D. T. Moore, J. Oomens, J. R. Eyler, B. Bendiak, *Anal. Chem.* **2006**, *78*, 670-679; b) B. Schindler, L. Barnes, G. Renois, C. Gray, S. Chambert, S. Fort, S. Flitsch, C. Loison, A. R. Allouche, I. Compagnon, *Nat. Commun.* **2017**, *8*, 973.
11. a) E. J. Cocinero, P. Carcabal, T. D. Vaden, J. P. Simons, B. G. Davis, *Nature* **2011**, *469*, 76-79; b) E. Mucha, A. I. G. Florez, M. Marianski, D. A. Thomas, W. Hoffmann, W. B. Struwe, H. S. Hahm, S. Gewinner, W. Schollkopf, P. H. Seeberger, G. von Helden, K. Pagel, *Angew. Chem. Int. Ed.* **2017**, *56*, 11248-11251; c) N. Khanal, C. Masellis, M. Z. Kamrath, D. E.

- Clemmer, T. R. Rizzo, *Anal. Chem.* **2017**, *89*, 7601-7606; d) J. M. Voss, S. J. Kregel, K. C. Fischer, E. Garand, *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 42-50.
12. a) V. Kopysov, A. Makarov, O. V. Boyarkin, *Anal. Chem.* **2015**, *87*, 4607-4611; b) V. Kopysov, A. Makarov, O. V. Boyarkin, *Anal. Chem.* **2017**, *89*, 544-547; c) V. Kopysov, M. V. Gorshkov, O. V. Boyarkin, *Analyst* **2018**, *143*, 833-836.
13. a) A. K. Vrkic, R. A. J. O'Hair, *J. Am. Soc. Mass. Spectr.* **2004**, *15*, 715-724; b) A. Ozdemir, J. L. Lin, K. J. Gillig, M. Gulfen, C. H. Chen, *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1113-1121.
14. a) K. Fuke, K. Kaya, *Chem. Phys. Lett.* **1983**, *94*, 97-101; b) T. Ebata, A. Fujii, N. Mikami, *In. Rev. Phys. Chem.* **1998**, *17*, 331-361; c) J. Screen, E. C. Stanca-Kaposta, D. P. Gamblin, B. Liu, N. A. Macleod, L. C. Snoek, B. G. Davis, J. P. Simons, *Angew. Chem. Int. Ed.* **2007**, *46*, 3644-3648; d) N. S. Nagornova, T. R. Rizzo, O. V. Boyarkin, *Science* **2012**, *336*, 320-323; (e) T. Sekiguchi, M. Tamura, H. Oba, P. Carcarbal, R. R. Lozada-Garcia, A. Zehnacker-Rentien, G. Gregoire, S. I. Ishiuchi, M. Fujii, *Angew. Chem. Int. Ed.* **2018**, *57*, 5626-5629.
15. O. V. Boyarkin, S. R. Mercier, A. Kamariotis, T. R. Rizzo, *J. Am. Chem. Soc.* **2006**, *128*, 2816-2817.
16. O. V. Boyarkin, V. Kopysov, *Rev. Sci. Instrum.* **2014**, *85*, 033105.
17. a) N. L. Burke, J. G. Redwine, J. C. Dean, S. A. McLuckey, T. S. Zwier, *Int. J. Mass Spectrom.* **2015**, *378*, 196-205; b) T. K. Roy, V. Kopysov, A. Pereverzev, J. Šebek, R. B. Gerber, O. V. Boyarkin, *Phys. Chem. Chem. Phys.* **2018**, *20*, 24894-24901.
18. V. Kopysov, O. V. Boyarkin, *Angew. Chem. Int. Ed.* **2016**, *55*, 689-692.
19. A. Abo-Riziq, L. Grace, B. Crews, M. P. Callahan, T. van Mourik, M. S. d. Vries, *J. Phys. Chem. A* **2011**, *115*, 6077-6087.
20. M. Schmitt, M. Bohm, C. Ratzer, D. Krugler, K. Kleinermanns, I. Kalkman, G. Berden, W. L. Meerts, *Chemphyschem* **2006**, *7*, 1241-1249.
21. K. L. Hudson, G. J. Bartlett, R. C. Diehl, J. Agirre, T. Gallagher, L. L. Kiessling, D. N. Woolfson, *J. Am. Chem. Soc.* **2015**, *137*, 15152-15160.
22. N. Agmon, *J. Phys. Chem. A* **2005**, *109*, 13-35.

23. D. A. Horke, H. M. Watts, A. D. Smith, E. Jager, E. Springate, O. Alexander, C. Cacho, R. T. Chapman, R. S. Minns, *Phys. Rev. Lett.* **2016**, *117*, 163002.
24. V. Kopysov, A. Makarov, O. V. Boyarkin, *J. Phys. Chem. Lett.* **2016**, *7*, 1067-1071.

APPENDIX CHAPTER 4

I. Materials and Methods

a) Experimental method

Protonated gas-phase carbohydrate-aromatic complexes were produced from the solutions containing $5 \cdot 10^{-5}$ M of a carbohydrate and $5 \cdot 10^{-5}$ M of an aromatic in the 1:1 water-methanol mixture with 1% of acetic acid using a nano-electrospray ionization source. As estimated from the relative abundances of protonated aromatic molecules and the complexes in MS spectra, in the experiments with N-acetyl carbohydrates, up to 30% of the molecules form complexes with TyrH⁺ and survive during the transfer to the gas phase. This number drops to a few percent for the monosaccharides that do not contain the NAc group.

Protonated species pass through a quadrupole mass filter, which selects parent ions of a particular mass-to-charge ratio (m/z). The selected ions are transferred into an octupole ion trap, which is cooled to 6 K by a closed-cycle refrigerator (SRDK-408, Sumitomo). Ions in the trap are cooled by collisions with helium atoms, which are introduced into the trap before the arrival of the ion packet. Approximately 40 ms later, when ions are thermalized and He has been pumped out, a pulse of UV light induces fragmentation of the stored parent ions. The photofragmentation is performed by the 2-3 mJ output of a widely-tunable UV OPO (EKSPLA, NT 3542C). The parent and fragment ions are released from the trap 1 ms after the excitation and directed into an Orbitrap-based mass spectrometer (Thermo Scientific (Bremen), Exactive-II), where the abundance and m/z of the parent and all fragment ions are measured simultaneously. Alternatively, the fragments were detected one-by-one in a more sensitive than Orbitrap quadrupole mass spectrometer (Extrel). We perform 10-20 measurements at each UV wavelength at a repetition rate of 10 Hz. The linewidth of the UV OPO light is $6\text{-}7\text{ cm}^{-1}$. Each spectrum was measured 3 times to ensure its reproducibility. The details of the setup can be found elsewhere.^{11, 24, 25, 26}

The recorded with the Orbitrap-based mass spectrometer two-dimensional data array (ion abundance vs m/z and wavelength) was, first,

pre-treated with the Peak-to-Peak software package (Spectroswiss) for peak detection and a baseline correction. The rectified data array then was stored as the 2D matrix identity of an isomer or a mixture of them.

b) Chemicals

All the carbohydrates were purchased from Carbosynth (>95% purity) and ROTH (>98% purity) and used without further purification. Aromatic molecules of >98% purity were purchased from Sigma-Aldrich and TCI. All the solvents of HPLC grade and acetic acid of >99% purity are from Sigma-Aldrich.

c) Library-based identification of isomers in their mixtures

Briefly, the library-based identification of isomers in their mixtures implies a preliminary measurement of 2D UV-MS data arrays of the species to be identified, constructing a library of the corresponding 2D matrices, and a decomposition of the recorded 2D spectra of the mixtures to be analyzed into a linear combination of the library matrices, using Matlab software package. The coefficients of this decomposition give the relative concentrations of the mixed in solution isomers. The detailed description of the procedure has been published elsewhere.^{11a, 24}

d) Computational details

The lowest energy conformers of the protonated carbohydrate-aromatic complexes were calculated in a two-step procedure: 1) Molecular Dynamics (MD) annealing and 2) Density Functional Theory (DFT) rectifications.

The MD simulations were performed with NAMD2 package,^[27] referenced as “NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign”. To describe interatomic interactions, the CHARMM36 force-field optimized for carbohydrates²⁸ and for proteins²⁹ were used. Calculations were performed in a spherical cell with the radius of 3 nm and with 1 fs integration time step. In each simulation, the system was disordered for 20 ps at temperature 1000 K and subsequently cooled at a rate of 5 K/ps to 300 K. After this annealing, the system was equilibrated for 125 ps, and, finally,

the energy was minimized in 1000 time-steps. The above procedure was repeated 500 times for every carbohydrate-aromatic pair. The obtained conformations were clustered on the basis of their geometry: (i) the two angles of relative position of the carbohydrate and aromatic rings; (ii) the distance between rings; (iii) the angle between the C-NH₃⁺ and C-NH (-NAc) groups (or C-CH₂OH for the non-modified molecules). The energies of the 20 lowest-energies conformers were further optimized by DFT calculations.

The DFT optimizations of the structures and the calculations of harmonic IR spectra for the final lowest energy geometries were performed with NWChem package. The geometries of the 20 lowest-energies conformers were optimized with a low accuracy (energy precision ~0.6 kcal/mol) using the B3LYP functional with the Grimme's DFT-D3 empirical dispersion correction and the 6-31+G(d,p) basis set. The geometries of the unresolved lowest-energies conformers were further optimized with more accurate convergent criteria (energy precision ~0.03 kcal/mol).

Table 4.S1. Structural formulas of the used aromatic sensor molecules

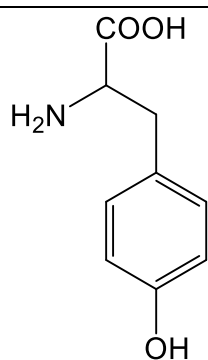
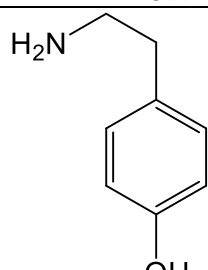
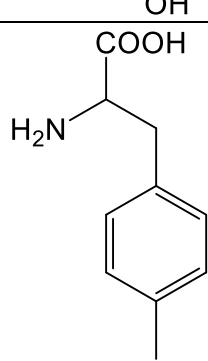
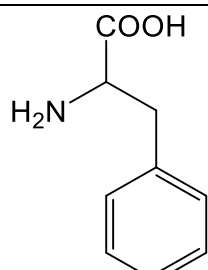





Name and abbreviation	Structural formula
Tyrosine (Tyr)	 <chem>NC(Cc1ccc(O)cc1)C(=O)O</chem>
Tyramine (Trm)	 <chem>NCCc1ccc(O)cc1</chem>
4-methylphenylalanine (MePhe)	 <chem>NC(Cc1ccc(C)cc1)C(=O)O</chem>
Phenylalanine (Phe)	 <chem>NC(Cc1ccccc1)C(=O)O</chem>

Table 4.S2. Symbolic representation of some monosaccharaides.

Symbolic representation	Name
	D-galactose
	D-mannose
	N-acetyl-D-glucosamine
	N-acetyl-D-galactosamine
	L-fucose

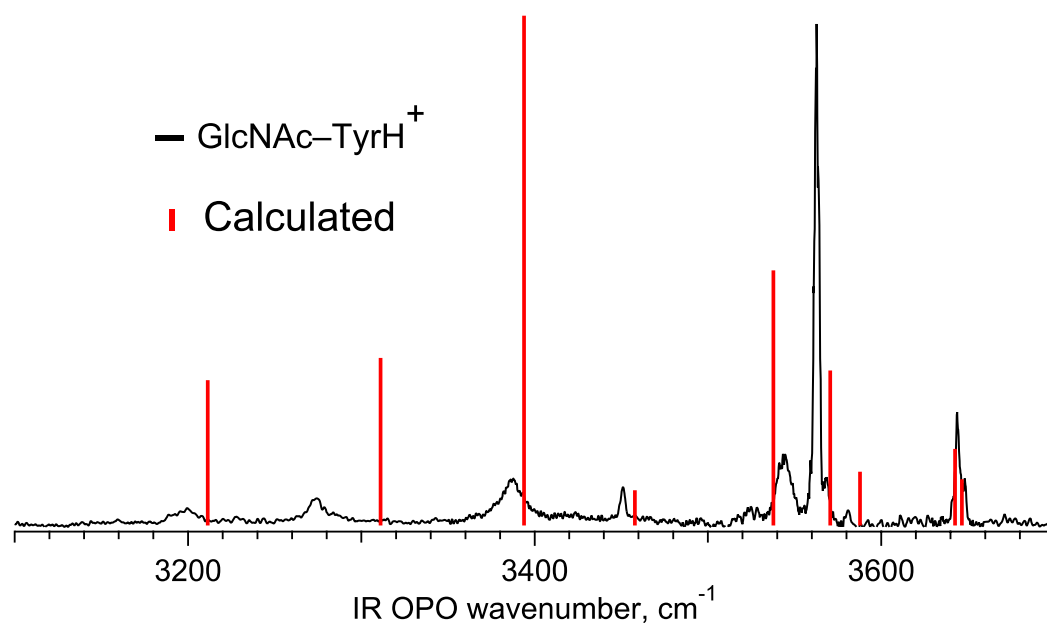


Figure 4.S1. IR-UV vibrational gain spectrum of GlcNAc-TyrH⁺ complex (black) and its calculated spectrum (harmonic approximation; scaling factor of 0.956) of the lowest energy computed β -anomeric structure (Figure 4.1E). The spectrum reflects vibrational transitions in *all* highly abundant conformers of the complex. The spectrum was measured by fixing UV OPO wavelength at 283.42 nm while scanning the wavenumber of the preceding pulse of IR OPO.^{25, 26} The good experiment-calculation match (except the typically overestimated in harmonic approximation frequencies of strongly red-shifted vibrations) validates the only highly abundant in our experiments conformer (conformational family) of the complex.

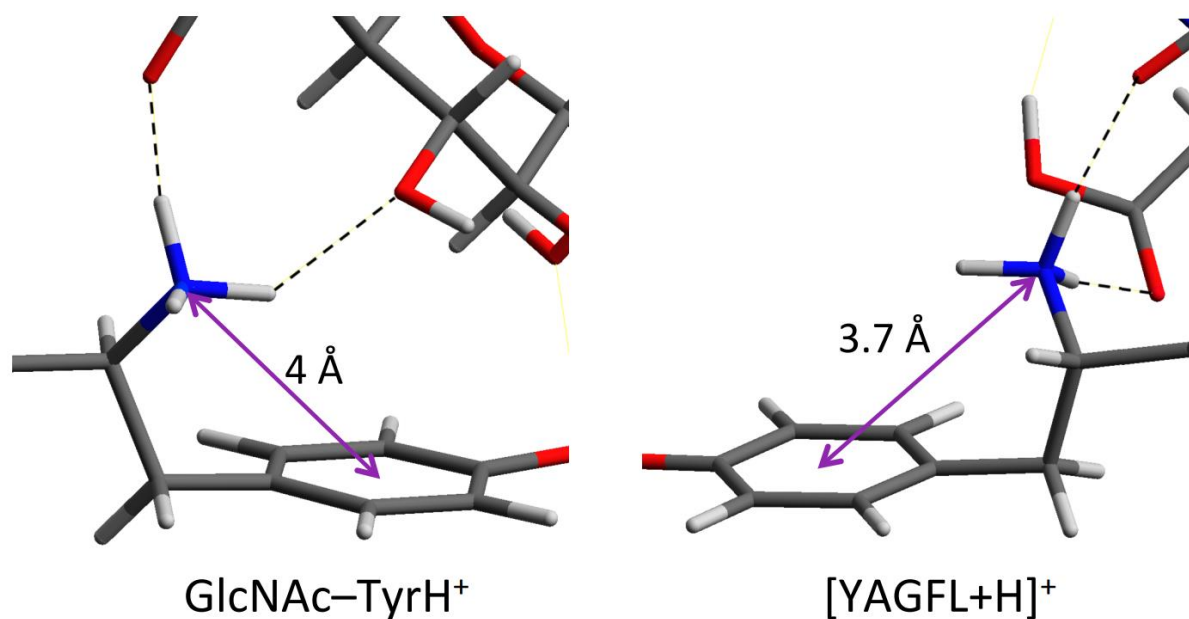
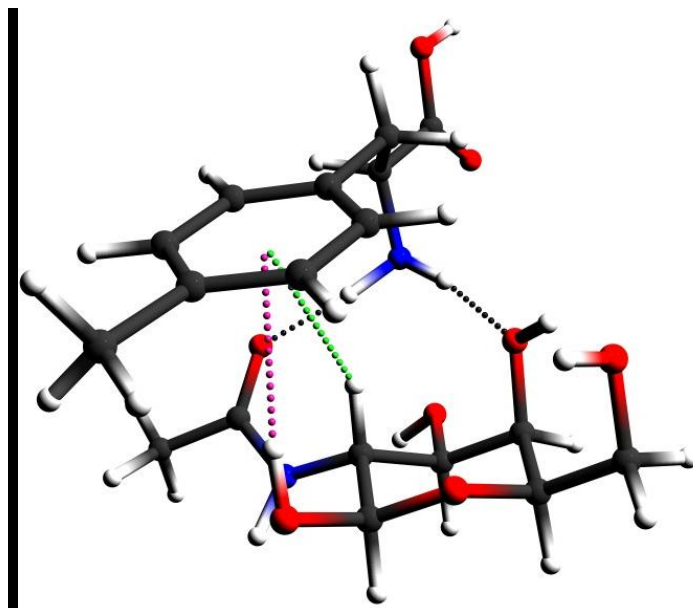


Figure 4.S2. Comparison of the computed the lowest energy structures of GlcNAc-TyrH^+ complex (this work) and $[\text{YAGFL+H}]^+$ peptides around their N-termini, illustrating the similarity of their H-bonding pattern. The dashed lines denote H-bonds; the double arrows show the distance from N atom to the center of the aromatic ring.

(A)



(B)

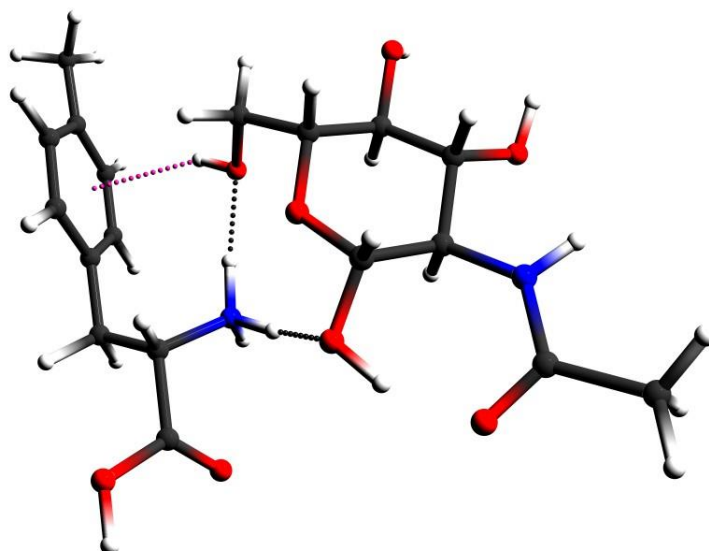


Figure 4.S3. The calculated lowest-energy structure of (A) GalNAc—MePheH⁺ and (B) GlcNAc—MePheH⁺ complexes; black, green and purple dashed lines indicate the O-H, CH- π and OH- π intermolecular non-covalent bonds, respectively.

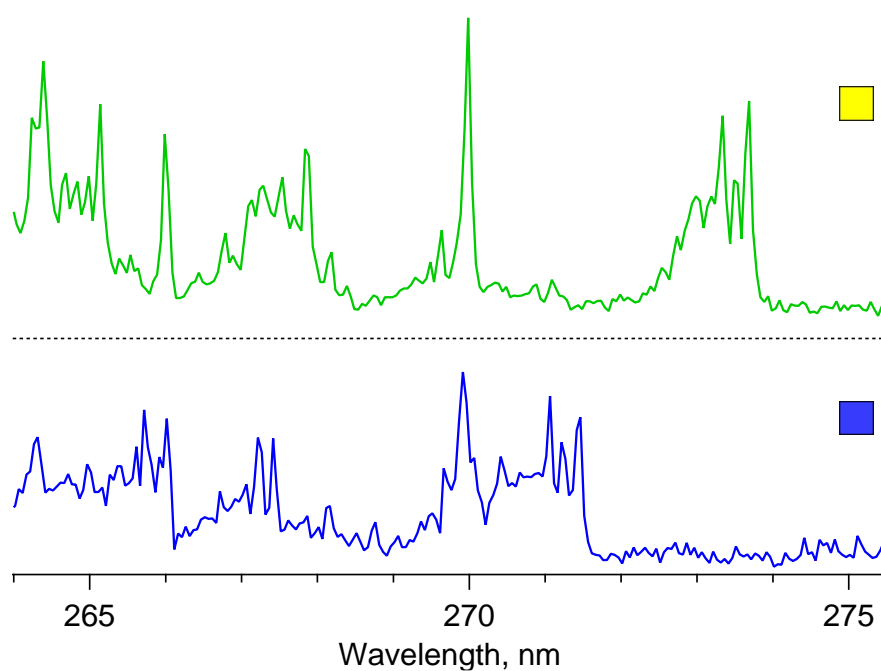


Figure 4.S4. Photofragmentation UV spectra of protonated 4-methylphenylalanine (red trace) and its complexes: GlcNAc—MePheH⁺ (blue trace) and GalNAc—MePheH⁺ (green trace). The UVPD spectra are normalized to UV OPO pulse energy. The spectra are labeled by the symbolic representation of monosaccharide building blocks.

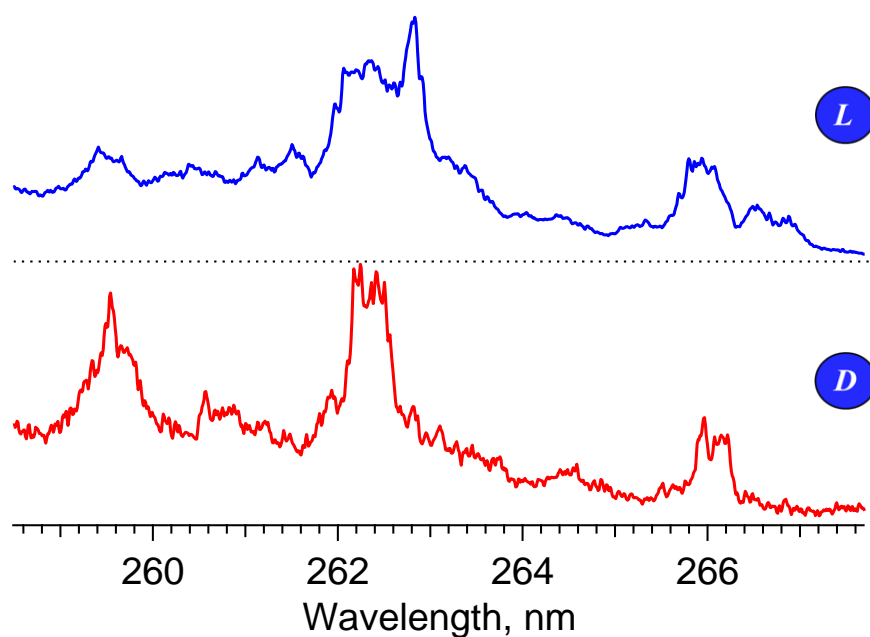


Figure 4.S5 Photofragmentation UV spectra of *D*-Glc—*L*-PheH⁺ (red trace) and *L*-Glc—*L*-PheH⁺ (blue trace) complexes at T= 6 K. The spectra of these fragile complexes were recorded by monitoring the intensity of the fragment with $m/z = 121$ Th. (the loss of neutral glucose and ammonia from *D/L*-Glc—*L*-PheH⁺) using a quadrupole mass spectrometer. The spectra are labeled by the symbolic representation of the monosaccharide building blocks.

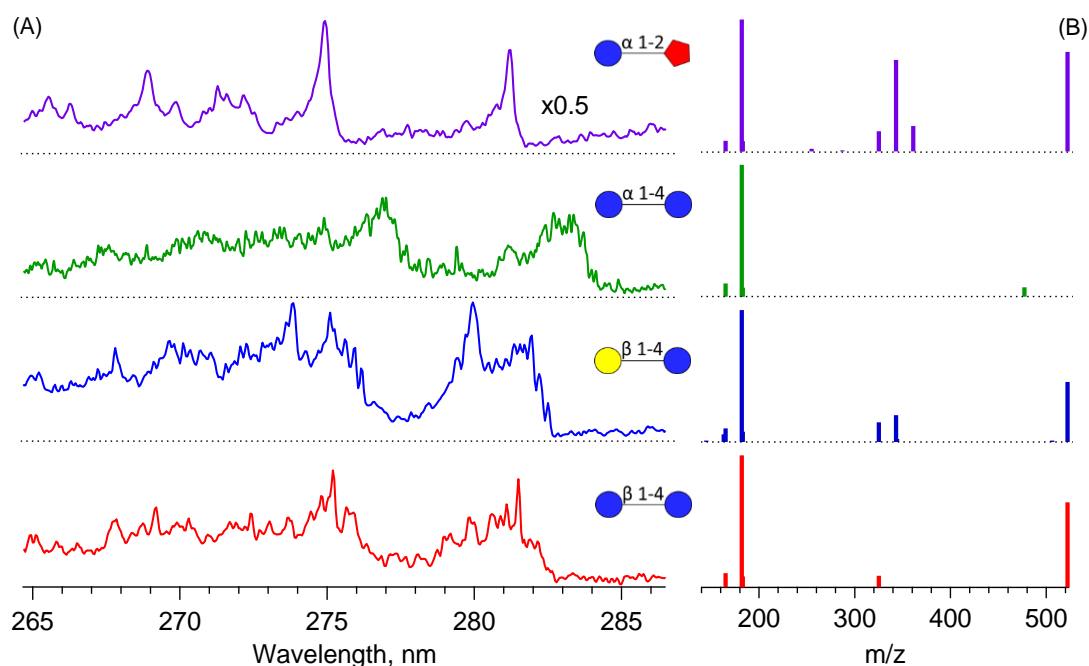


Figure 4.S6. Photofragmentation (A) UV and (B) mass spectra of the complexes of tyrosine with non-acetylated disaccharides (from bottom to top): β -Glc-1,4-Glc-TyrH⁺, β -Gal-1,4-Glc-TyrH⁺, α -Glc-1,4-Glc-TyrH⁺, α -Glc-1,2-Fru-TyrH⁺. The spectra were generated by integrating the respective 2D UV-MS data arrays over wavelength. The 2D UV-MS arrays are normalized to the UV OPO pulse energy; each mass spectrum is normalized to the total ion signal, including the parent ion and scaled to the most intense peak. The spectra are labelled by the symbolic representation of the monosaccharide building blocks.

Coordinates of the lowest energy calculated GalNAc-TyrH⁺ non-covalent complex

C	-1.92543	-0.68940	12.79178
H	-2.53136	-1.21227	13.55101
O	-2.46558	0.57795	12.70021
H	-1.89542	1.18551	12.19311
C	0.03532	-1.88282	13.51432
H	-0.50027	-2.37827	14.34102
O	-0.57339	-0.61582	13.23323
C	-1.94655	-1.49426	11.46758
H	-1.25307	-1.01972	10.76773
N	-3.28212	-1.47937	10.88225
H	-4.03595	-1.14460	11.46924
C	-3.50850	-1.51806	9.54369
O	-2.60864	-1.75326	8.71940
C	-4.92318	-1.26177	9.08003
H	-5.64201	-1.16090	9.89642
H	-5.23237	-2.07860	8.42315
H	-4.93004	-0.34456	8.48313
C	-1.45461	-2.91986	11.79935
H	-2.07065	-3.32517	12.61651
O	-1.49920	-3.81260	10.69205
H	-2.41804	-4.06029	10.52454
C	-0.01196	-2.83274	12.29822
H	0.31462	-3.83024	12.62134
O	0.81979	-2.39530	11.21513

Coordinates of the lowest energy calculated GalNAc- TyrH⁺ non-covalent complex

C	-1.91325	6.45039	2.23724
H	-1.59683	5.90780	1.32389
O	-1.46322	5.81950	3.38749
H	-0.49837	5.74810	3.35891
C	-1.77083	8.51015	1.02937
H	-1.52064	7.96583	0.10404
O	-1.36065	7.75900	2.18409
C	-3.45475	6.49994	2.31376
H	-3.71906	6.93957	3.27911
N	-3.98289	5.14170	2.28894
H	-3.55809	4.49016	1.64206
C	-4.85536	4.63949	3.18868
O	-5.41406	5.35440	4.04499
C	-5.12719	3.15694	3.13629
H	-4.74614	2.70985	4.05990
H	-4.66654	2.65449	2.28232
H	-6.20780	2.99602	3.11516
C	-4.01069	7.36920	1.18181
H	-3.87424	6.85273	0.21883
O	-5.41021	7.58381	1.40040
H	-5.68394	8.22219	0.72123
C	-3.28784	8.71293	1.12367
H	-3.49752	9.24140	2.05907
O	-3.81832	9.41625	0.00980
H	-3.79433	10.36895	0.20581

Additional References

25. A. Y. Pereverzev, V. Kopysov, O. V. Boyarkin, High Susceptibility of Histidine to Charge Solvation Revealed by Cold Ion Spectroscopy. *Angew. Chem. Int. Edit.* 2017, 56, 15639-15643
26. A. Y. Pereverzev, V. N. Kopysov, O. V. Boyarkin, Peptide Bond Ultraviolet Absorption Enables Vibrational Cold-Ion Spectroscopy of Nonaromatic Peptides. *J. Phys. Chem. Lett.* 2018, 9, 5262-5266
27. J. C. Phillips et al., Scalable molecular dynamics with NAMD. *J. Comp. Chem.* 2005, 26, 1781-1802
28. O. Guvench et al., Additive empirical force field for hexopyranose monosaccharides. *J. Comp. Chem.* 2008, 29, 2543-2564
29. A. D. MacKerell, M. Feig, C. L. Brooks, Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* 2004, 126, 698-699

CHAPTER 5.

Revealing Single-bond Anomeric Selectivity in Carbohydrate-Protein Interactions

OVERVIEW

Non-covalent binding of proteins to glycans is amazingly selective to the isoforms of carbohydrates, including α/β anomers that co-exist in solution. We isolate in the gas phase and study at the atomic level the simplest model system: non-covalent complexes of monosaccharide α/β -GalNAc and protonated aromatic molecule tyramine. IR/UV cold ion spectroscopy and quantum chemistry calculations jointly solve the structures of the two complexes. Although the onsets of the measured UV absorptions of the complexes differ significantly, the networks of H-bonds in both complexes appear identical and do not include the anomeric hydroxyl. The detailed analysis reveals that, through inductive polarization, the α - to β - re-orientation of this group nevertheless reduces the length of one remote short intermolecular H-bond by 0.03 Å. Although small, this change substantially strengthens the bond, thus contributing to the anomeric selectivity of the binding. The finding demonstrates an amazing sensitivity of non-covalent bonds to structural changes in a partnering molecule.

The results presented in this chapter were published in:

E. Saparbaev, V. Aladinskaia, R. Yamaletdinov, A. Pereverzev, and O. Boyarkin. Revealing Single-Bond Anomeric Selectivity in Carbohydrate-Protein Interactions. *Journal of Physical Chemistry Letters* 2020, 11, 3327-3331, DOI: <https://doi.org/10.1021/acs.jpclett.0c00871>

5.1. INTRODUCTION

Protein-carbohydrate non-covalent interactions are ubiquitous in nature. The flexibility and the diversity of these weak interactions make proteins sensitive to finer structural details of the binding partners, greatly increasing their biological functionality. Despite the tremendous isomeric diversity of carbohydrates, the glycan binding proteins (GBPs), for instance, lectins, are capable of finding specific membrane carbohydrates to adhere viruses and bacteria to appropriate cells for invasion.¹

Hydrogen bonds between the hydroxyl groups of carbohydrates and the polar groups of amino acid residues play a crucial role in binding to lectins and other GBPs. Difference in the relative orientation of the hydroxyl groups in glycans has a pivotal impact on the ability of GBPs to distinguish different isoforms of carbohydrates, in particular their epimers (e.g., monosaccharides Gal, Glc, Man, etc.)²⁻³ Regardless of its epimeric isoform, a glycan in solution exists as two interconverting isomers: α - and β -anomers, which differ only by the orientation of the first OH group (1-OH) in the reducing-end monosaccharide. In contrast to epimers, which are usually recognized through binding to the 3-OH and 4-OH groups, the anomeric 1-OH often is not involved in the binding to GBPs.^{2, 4} Nevertheless, similar to epimers, the two anomers often exhibit very different affinities to, for instance, lectins and to some artificial receptors, which allows them a selective recognition of anomeric carbohydrates.³⁻⁴ Despite many excellent studies in this field, the detailed mechanism of the anomeric recognition remains obscure, in particular, due to the large size of the interacting partners and the presence of solvent molecules.⁵⁻⁸ Here we use cold-ion UV/IR spectroscopy and structural quantum chemistry computations to investigate at the atomic level the simplest anomeric system that models a local protein-glycan interaction: monosaccharide anomers, α/β -GalNAc, non-covalently bound to protonated tyramine (TrmH^+) and isolated from solvent in the gas phase. The key questions that we address herein are, whether the spectroscopic and structural signatures of the anomeric effect can be detected for such

a small systems, and, if yes, what is their origin and the implication for the recognition of anomers.

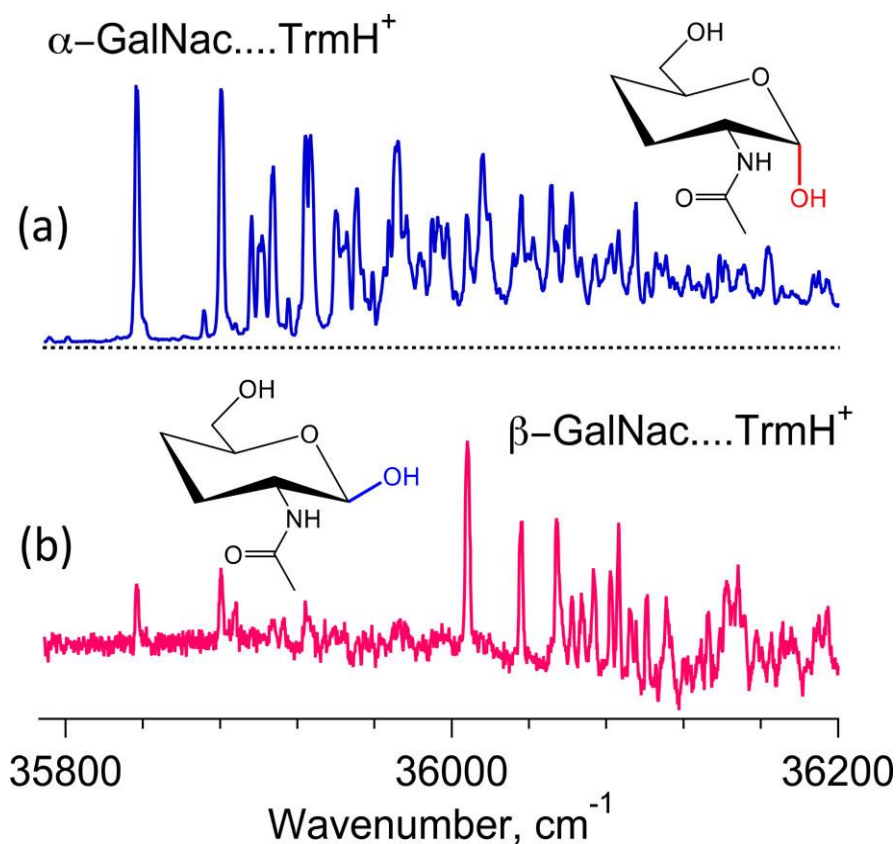


Figure 5.1. (a) UV photo fragmentation spectra of α/β -GalNac-TrmH⁺ cooled to T~10 K, and (b) IR-UV hole-burning spectrum of the complex measured with IR excitation of the α -anomer at 3534.3 cm⁻¹. The inserts show the structures of the respective glycans with the anomeric hydroxyl group colored.

5.2. RESULTS AND DISCUSSION

GalNac is one of the most frequent carbohydrate units involved in binding with proteins;⁴ tyramine is structurally similar to tyrosine but smaller than this amino acid due to the lack of C-terminus. It was recently shown that in the GalNac-TyrH⁺ complex this carboxylic group is not involved in H-bonding.⁹ GalNac-TrmH⁺ complex therefore is a reasonable small system to model local protein-glycan interactions. In solution, where the complexes are natively prepared, their spectroscopic details are masked by non-covalent interactions with solvent molecules. Gas-phase isolation and cryogenic cooling of such small complexes enables their

vibrationally resolved photofragment IR/UV spectroscopy and the use of mass spectrometry for sensitive detection of the charged fragments.¹⁰⁻¹¹ Our tandem cold-ion mass spectrometer¹²⁻¹⁴ (see appendix chapter 5 for details) allows for measuring of four types of photofragmentation spectra of the complexes: electronic *UV* and *IR-UV* hole burning,¹⁵ vibrational *IR-UV* gain^{14, 16} and *IR-UV* depletion¹² (here and below the *italic* font designates the scanned laser). Structures of the complexes and their “harmonic” IR spectra were calculated using standard approaches of quantum chemistry (see SI for details).

The UV spectra of the GalNAc-TrmH⁺ complex (Figure 5.1a), as well as its methylated derivatives (Fig. 5.S1), exhibit well-resolved electronic band origins. This observation enables measuring of the *IR-UV* depletion and background-free gain spectra of the complexes. Figure 5.2a shows a conformer non-selective IR gain spectrum of GalNAc-TrmH⁺. The spectrum reflects vibrational transitions in both anomers of the monosaccharide. For comparison, Figures 5.2b and 5.2c show the spectra of the α -MeGalNAc-TrmH⁺ and β -MeGalNAc-TrmH⁺ complexes, in which anomeric forms of the glycan are fixed by its methylation at 1-OH. The spectra with the methylated glycans clearly differ in the pairs of peaks (Fig 5.2a) 1-2, 5-6 and 9-10. Except peak 3, every peak of the spectrum in figure 5.2a is mutually reproduced (within ± 0.8 cm⁻¹) in the spectra 5.2b and/or 5.2c. This observation implies that the anomeric 1-OH, indeed, is not involved in any strong H-bonds in both complexes and allows for an unambiguous assignment of peaks 1, 4 5, 7, 8 and 10 to α -GalNAc-TrmH⁺ and peaks 2, 6, 7, 9 and 11 (unresolved) to the β -GalNAc-TrmH⁺ anomer. The remaining peak (3) has been assigned to the 1-OH stretch, which is absent in the methylated glycans. The gain spectra of the complexes, in which both partners are methylated, α/β -MeGalNAc-MeTrmH⁺, allow for a tentative assignment of the disappearing peaks (Figures 5.2d and 5.2e) to the OH stretches of Trm in the α and β anomeric GalNAc-TrmH⁺ (peak 4 and shoulder 4 in Fig. 5.2a). The detected essential difference of the frequencies in the pairs of identical vibrational transitions 5-6 and 10-11 is the first sign of the anomeric specificity of the complexes.

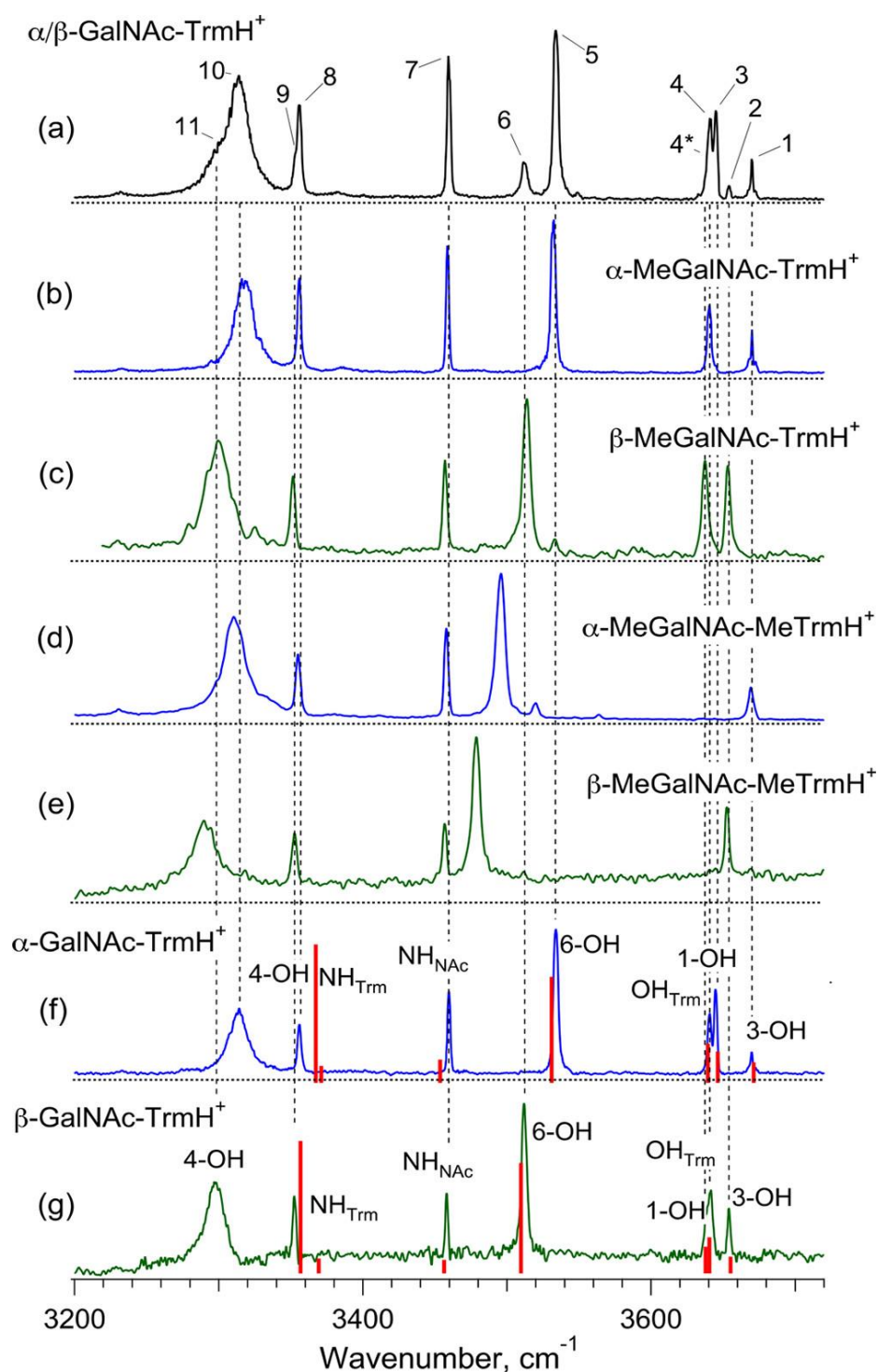


Figure 5.2. (a) IR-UV gain spectrum of GalNAc-TrmH⁺ complex and (b–e) of its methylated derivatives (as labelled in the figure). IR-UV depletion spectra of GalNAc-TrmH⁺ with UV OPO tuned to the electronic band origin in (f) α -GalNAc-TrmH⁺ (35836 cm⁻¹) and (g) β -GalNAc-TrmH⁺ (36008 cm⁻¹). Red sticks indicate the scaled (0.956) vibrational frequencies calculated for the respective lowest-energy GalNAc-TrmH⁺ complexes (Figs. 5.3a and 5.3b).

Using the intense isolated peak 5 of the α -anomer, we have performed IR-UV hole-burning spectroscopy of this complex, which give us the anomeric assignment of the UV transitions (Fig. 5.1b). The electronic band origins of the two anomers appear to be shifted by as much as 171 cm^{-1} . This large shift is the second and the most striking spectroscopic manifestation of the anomeric effect observed for the complex. To reveal the origin of the effect we, first, solved the structures of the anomers. Figures 5.2f and 5.2g show the anomer-selective IR-UV depletion spectra of GalNAc-TrmH⁺. Every peak in the gain spectrum 5.2a of the complex appears in one of its two depletion spectra, which implies that, under our experimental conditions, only a single conformer is populated for each anomeric complex; the next nearest in the energy conformer should be higher by, at least, 2 kcal/mol.¹⁷ Our computations reveal that the most stable structures for both anomeric complexes are, indeed, below the closest in energy conformers by more than 4 kcal/mol. The comparison of the measured and computed IR spectra validates these lowest energy structures (Figures 5.3a, 5.3b). All the peak assignments experimentally derived above are reproduced by the computations (Figs. 5.2f and 5.2g); frequencies of the vibrations with the atoms that are not involved in strong H-bonds match very well to the measured peaks. The frequencies of the coupled vibrations are slightly overestimated, which is typical for computations in a harmonic approximation. None of the tested high-energy structures exhibit the computed IR spectra that are similar to the measured ones. The validated most stable geometries of the α/β -GalNAc-TrmH⁺ complexes look very similar with exactly the same network of H-bonds that does not involve the anomeric OH. But then, where are those structural differences that make the UV spectra anomer-specific? In order to reveal these differences we next perform a finer analysis of the non-covalent bonds and their influence on UV absorption of the aromatic moiety.

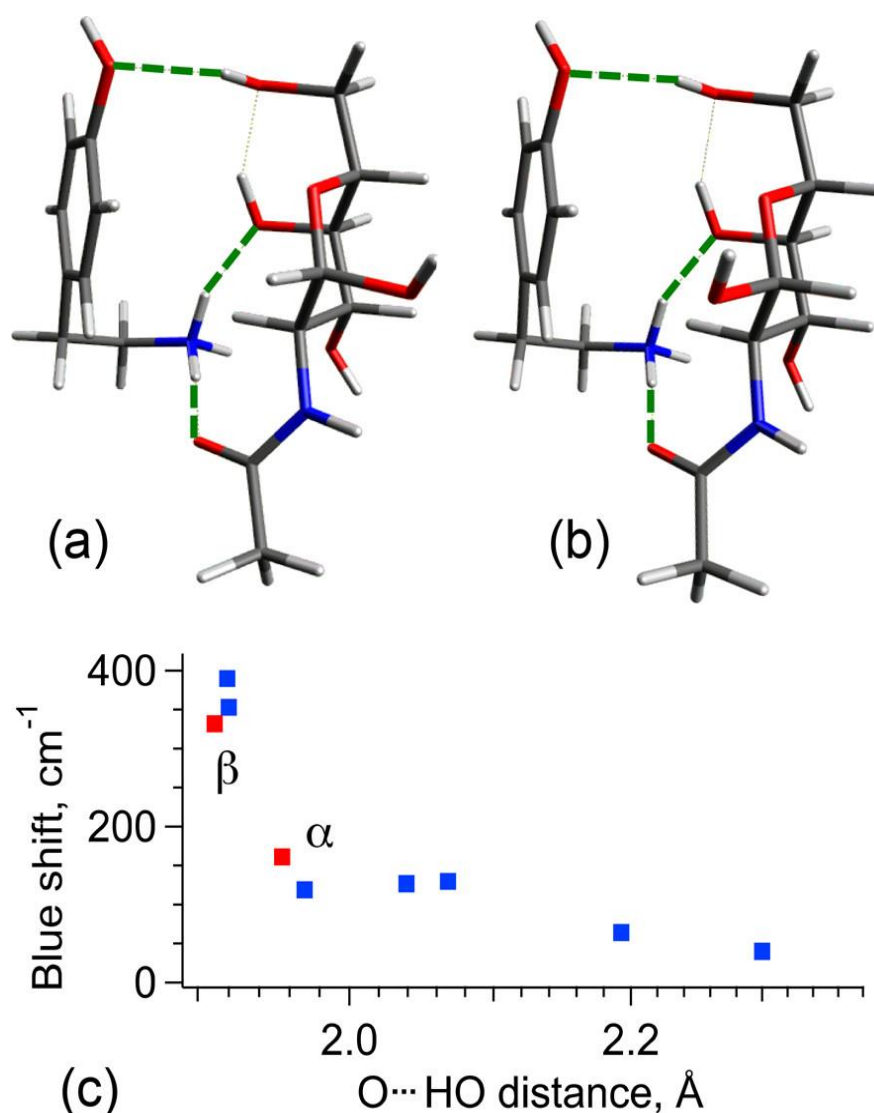


Figure 5.3. The most stable calculated structures of (a) α -GalNAc-TrmH⁺ and (b) β -GalNAc-TrmH⁺ gas-phase complexes. (c) Blue shifts of UV band origins in aromatic molecules induced by (aromatic)-O...HO bonding as a function of the bond length: collected from literature (blue dots; Table 5.S1) and measured here (red dots).

The two strong bonds formed by the NH₃⁺ group of TrmH⁺ with 4-OH and C=O of anomeric sugars appear in both complexes, allowing them to survive under the harsh conditions of ESI. The length of the bond to 4-OH is the same in both complexes (1.821 Å); the bond to CO differs negligibly between the two complexes (1.710 Å and 1.715 Å for the α and β anomers, respectively), such that this bond cannot contribute significantly to the anomeric selectivity of the binding. The two bonds also cannot account for

the large blue shifts of the UV absorption in the two complexes relative to the electronic band origin in neutral Trm,¹⁸ since the charge of the NH_3^+ group solvated by intermolecular H-bonds influences only little the absorption of aromatic ring.¹⁹⁻²⁰ Although the protonated N-terminus of a protein might be remote from the domain of the contact with glycan and not be involved to the binding, herein such interaction with the charge mimics the *in vivo* anchoring of terminal GalNAc group of glycans to lectins *via* a Ca^{2+} cation, which also imposes strong steric constraints on geometries of the complex.²

A much weaker $\text{CH}\cdots\pi$ interaction is known to exhibit anomeric selectivity in the aromatic-glycan complexes that have no strong steric constraints.^{3, 21-22} Statistically, glycans prefer binding to aromatic systems by their α and β faces for β -GalNAc and α -GalNAc, respectively. This facial anomeric selectivity is suppressed in our complexes, which are constrained (by the $\text{NH}_3^+\cdots\text{O}$ bonds), such that TrmH⁺ interacts with the β -face.²¹ The quite small relative difference between the lengths of the (shortest) $\text{CH}\cdots\pi$ bonds (3.447 Å and 3.462 Å for the α and β anomers, respectively) indicates their nearly identical binding energies in the two complexes. $\text{CH}\cdots\pi$ interactions commonly shift UV absorption of aromatic systems to the red by a few tens of cm^{-1} at most.²³ This binding therefore cannot explain the blue shifts of the electronic band origins observed in the UV spectra of both complexes. We thus are left with the H-bond that couples the oxygen atom of the tyramine side-chain and the hydrogen of 6-OH in the glycan as the only strong intermolecular bond to account for the anomeric specificity of the blue shifts. This H-bond is short and different in the two complexes (1,985 Å for and 1,955 Å in the α - and β -anomers, respectively). Hydrogen binding of hydroxyl oxygen in aromatic molecules (e.g., phenol, tyrosine) always induces a blue shift of their electronic band origins, which is attributed to inductive polarization of the aromatic π -system.²⁴ The blue shifts reported earlier for different non-covalent complexes of aromatic molecules clearly correlate with the H-bond length and increase steeply upon shortening of the bond (Fig.

5.3b).²⁵⁻³² Our data are consistent with this correlation. They fall onto the dependence in Figure 5.3b at near the shortest length of the bond. This explains the high sensitivity of the observed blue shifts of the electronic band origin in the anomeric complexes to the small change of the H-bond length. The shifts must reflect the strength of the 6-OH...OH(Trm) bond. We therefore may expect this bond to be noticeably stronger in the complex of the β -anomer. Consistently, the frequency of the 6-OH stretch vibration in the IR spectrum of this anomer is substantially red-shifted with respect to that in the α -anomer (Figs. 5.2f and 5.2g).

The difference of binding energies was estimated based on such energies calculated for (i) fucose-phenol non-covalent complexes with HO...H bonds of different lengths³³ and for (ii) 6-HO...HN bonds in MeGal-aromatic complexes⁵ as ~0.75 kcal/mol and ~1 kcal/mol, respectively. The calculated energies were linearly interpolated and extrapolated for the first and the second case, respectively, to the length of the HO...H bond in GalNAc-TrmH⁺. The estimated difference is substantial and, being accumulated together with a few similar H-bonds (and/or H- π couplings) in a protein-glycan complex may influence the overall binding affinity of the partners. Although demonstrated here with a Tyr-like aromatic molecule, one may expect similar anomeric effect for the non-aromatic residues that make H-bonds with glycans.

Anomeric effects were also observed for isolated non-covalent complexes of a neutral aromatic molecule bound to β/α faces of α -MeGal and β -MeGal, respectively.⁵ Methylation was required for fixing the anomeric identity of the glycans, which otherwise could not be distinguished. The difference between the geometries of the α - and β - complexes was explained by the difference in the inductive polarization of the glycan atoms involved in the multiple intermolecular H-bonds. The large size of the methyl group that stuck into the area of the contact in the complexes left an uncertainty in whether the detected anomeric effects are of the steric origin or, indeed, solely due to the difference in the inductive polarization. Another uncertainty aroused from

the conformational heterogeneity of the complexes, which reduced confidence in the assignment of the computed structures. Our study removes both uncertainties. The experimental approach used herein enables anomeric- and conformer-specific sensing of the complexes with the unmodified α/β -monosaccharaides, naturally coexisting in solution. The rigorously assigned intrinsic structures of the observed single conformer of each complex are almost identical. The structures reveal that the unbound anomeric group sticks away from the contact area and there is only one H-bond, 6-OH...OH(Trm), whose length differ substantially between the two anomeric complexes. The steric effect is therefore ruled out and the anomeric specificity of the binding is exclusively due to the difference in polarization of this bond in the two complexes, which differ only by the flip of the anomeric 1-OH in the glycan.

5.3. CONCLUSION

Overall, our study demonstrates that, even in the case when binding of a glycan to a protein is sterically constrained by strong interactions with a charge (Ca^{2+} or NH_3^+), the binding may remain anomer-selective due to the difference in the strength of inductive polarization of intermolecular H-bonds for two anomers. The difference in binding energy for such single bond can be as large as ~ 1 kcal/mol, which is not negligible for the total budget of the glycan-protein affinity. The induced anomer-specific polarization can subsequently weaken intramolecular non-covalent interactions in the protein, which might be a mechanism of signaling to the protein the anomeric identity of the attached carbohydrate.

REFERENCES

1. Sharon, N.; Lis, H., Lectins as cell recognition molecules. *Science* **1989**, *246*, 227-234.
2. Weis, W. I.; Drickamer, K., Structural Basis Of Lectin-Carbohydrate Recognition. *Ann. Rev. Biochem.* **1996**, *65*, 441-473.
3. Davis, A. P.; Wareham, R. S., Carbohydrate Recognition through Noncovalent Interactions: A Challenge for Biomimetic and Supramolecular Chemistry. *Angew. Chem. Int. Ed.* **1999**, *38*, 2978-2996.
4. Sharon, N.; Lis, H., Specificity and Affinity. In *Lectins*, 2 ed.; Springer: The Netherlands, 2007; pp 63-104.
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. Contreras, C. S.; Polfer, N. C.; Oomens, J.; Steill, J. D.; Bendiak, B.; Eyler, J. R., On The Path To Glycan Conformer Identification: Gas-Phase Study Of The Anomers Of Methyl Glycosides Of N-Acetyl-D-Glucosamine And N-Acetyl-D-Galactosamine. *Int. J. Mass Spectrom.* **2012**, *330*, 285-294.
7. 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.
8. Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K., Identification Of Carbohydrate Anomers Using Ion Mobility-Mass Spectrometry. *Nature* **2015**, *526*, 241-4.
9. Saparbaev, E.; Kopysov, V.; Yamaletdinov, R.; Pereverzev, A.; Boyarkin, O. V., Interplay of H-bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates. *Angew. Chem. Int. Ed.* **2019**, *58*, 7346-7350.
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. Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R., Conformation-Specific Infrared And Ultraviolet Spectroscopy Of Tyrosine-Based Protonated Dipeptides. *J. Chem. Phys.* **2007**, *127*, 154322-7.
13. Pereverzev, A. Y.; Cheng, X.; Nagornova, N. S.; Reese, D. L.; Steele, R. P.; Boyarkin, O. V., Vibrational Signatures of Conformer-Specific Intramolecular Interactions in Protonated Tryptophan. *J. Phys. Chem A* **2016**, *120*, 5598-5608.
14. Pereverzev, A. Y.; Boyarkin, O. V., Exploring The Relevance Of Gas-Phase Structures To Biology: Cold Ion Spectroscopy Of The Decapeptide Neurokinin A. *Phys. Chem. Chem. Phys.* **2017**, *19*, 3468-3472.
15. Shubert, V. A.; Zwier, T. S., IR-IR-UV Hole-Burning: Conformation Specific IR Spectra In The Face Of UV Spectral Overlap. *J. Phys. Chem A* **2007**, *111*, 13283-13286.
16. Inokuchi, Y.; Boyarkin, O. V.; Kusaka, R.; Haino, T.; Ebata, T.; Rizzo, T. R., UV and IR Spectroscopic Studies of Cold Alkali Metal Ion-Crown Ether Complexes in the Gas Phase. *J. Am. Chem. Soc.* **2011**, *133*, 12256-12263.
17. Pereverzev, A. Y.; Kopysov, V.; Boyarkin, O. V., High Susceptibility of Histidine to Charge Solvation Revealed by Cold Ion Spectroscopy. *Angew. Chem. Int. Ed.* **2017**, *56*, 15639-15643.
18. Yoon, I.; Seo, K.; Lee, S.; Lee, Y.; Kim, B., Conformational Study of Tyramine and Its Water Clusters by Laser Spectroscopy. *J. Phys. Chem A* **2007**, *111*, 1800-1807.
19. 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.
20. 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.

21. Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J., Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946-954.
22. 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.
23. Tsuzuki, S., CH/ π interactions. *Ann. Rep. Sec. C (Phys. Chem.)* **2012**, *108*, 69.
24. Hsu, C. H.; Park, S.; Mortenson, D. E.; Foley, B. L.; Wang, X.; Woods, R. J.; Case, D. A.; Powers, E. T.; Wong, C. H.; Dyson, H. J.; Kelly, J. W., The Dependence of Carbohydrate-Aromatic Interaction Strengths on the Structure of the Carbohydrate. *J. Am. Chem. Soc.* **2016**, *138*, 7636-48.
25. Biswas, N.; Chakraborty, S.; Wategaonkar, S., Gas Phase Spectroscopic Studies of Hydroquinone Dimer. *J. Phys. Chem A* **2004**, *108*, 9074-9081.
26. Weichert, A.; Riehn, C.; Brutschy, B., High-Resolution Rotational Coherence Spectroscopy of the Phenol Dimer. *J. Phys. Chem A* **2001**, *105*, 5679-5691.
27. Seifert, N. A.; Steber, A. L.; Neill, J. L.; Pérez, C.; Zaleski, D. P.; Pate, B. H.; Lesarri, A., The Interplay Of Hydrogen Bonding And Dispersion In Phenol Dimer And Trimer: Structures From Broadband Rotational Spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11468-11477.
28. Bernhard, D.; Dietrich, F.; Fatima, M.; Perez, C.; Gottschalk, H. C.; Wuttke, A.; Mata, R. A.; Suhm, M. A.; Schnell, M.; Gerhards, M., The Phenyl Vinyl Ether-Methanol Complex: A Model System For Quantum Chemistry Benchmarking. *Beilstein J. Org. Chem.* **2018**, *14*, 1642-1654.
29. Yi, J. T.; Ribblett, J. W.; Pratt, D. W., Rotationally Resolved Electronic Spectra of 1,2-Dimethoxybenzene and the 1,2-Dimethoxybenzene–Water Complex. *J. Phys. Chem A* **2005**, *109*, 9456-9464.
30. Reimann, B.; Buchhold, K.; Barth, H. D.; Brutschy, B.; Tarakeshwar, P.; Kim, K. S., Anisole-(H₂O)_N (N=1-3) Complexes: An Experimental And Theoretical Investigation Of The Modulation Of Optimal Structures, Binding Energies, And Vibrational Spectra In Both The Ground And First Excited States. *J. Chem. Phys.* **2002**, *117*, 8805-8822.

31. Ribblett, J. W.; Sinclair, W. E.; Borst, D. R.; Yi, J. T.; Pratt, D. W., High Resolution Electronic Spectra of Anisole and Anisole–Water in the Gas Phase: Hydrogen Bond Switching in the S1 State. *J. Phys. Chem A* **2006**, *110*, 1478-1483.
32. Macleod, N. A.; Simons, J. P., Beta-Blocker Conformations In The Gas Phase: 2-Phenoxy Ethylamine, Its Hydrated Clusters And 3-Phenoxy Propanolamine. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2878-2884.
33. Tsuzuki, S.; Uchimaru, T.; Mikami, M., Magnitude And Nature Of Carbohydrate-Aromatic Interactions In Fucose-Phenol And Fucose-Indole Complexes: CCSD(T) Level Interaction Energy Calculations. *J. Phys. Chem A* **2011**, *115*, 11256-62.

APPENDIX CHAPTER 5

II. Materials and Methods

a) Experimental method

Protonated gas-phase carbohydrate-aromatic complexes were produced from the solutions containing $5 \cdot 10^{-5}$ M of a carbohydrate and $5 \cdot 10^{-5}$ M of an aromatic in the 1:1 water-methanol mixture with 1% of acetic acid using a nano-electrospray ionization source. As estimated from the relative abundances of protonated aromatic molecules and the complexes in MS spectra, in the experiments with N-acetyl carbohydrates, up to 30% of the molecules form complexes with TrmH^+ and survive during the transfer to the gas phase.

Charged species pass through a quadrupole mass filter, which selects parent ions of a particular mass-to-charge ratio (m/z). The selected ions are transferred into an octupole ion trap, which is cooled to 6 K by a closed-cycle refrigerator (SRDK-408, Sumitomo). Ions in the trap are cooled by collisions with helium atoms, which are introduced into the trap before the arrival of the ion packet. Approximately 40 ms later, when ions are thermalized and He has been pumped out, a pulse of UV light induces fragmentation of the stored parent ions. The UV photofragmentation of methylated derivatives of GalNAc- TrmH^+ complexes were performed by the 2-3 mJ output of a widely-tunable UV OPO (EKSPLA, NT 3542C). The linewidth of the UV OPO light is 6-7 cm^{-1} . The UV light for the experiments with non-modified GalNAc- TrmH^+ complexes is produced by frequency doubling the output of a dye laser (HD-500, Lumonics) in a KDP crystal. The dye laser is pumped by 150 mJ of the 3rd harmonic of a Nd:YAG laser (GCR-210, Spectra-Physics). IR beams are generated by two tunable optical parametric oscillator (OPO) (Laser Vision), pumped by two different Nd:YAG lasers (Surelite IIIEx, Continuum; SpitLight 600, Innolas). Spectral resolution of the IR OPO is about 1 cm^{-1} .

The parent and fragment ions are released from the trap 1 ms after the excitation and directed into an Orbitrap-based mass spectrometer (Thermo Scientific (Bremen), Exactive-II), where the abundance and m/z of the parent and all fragment ions are measured simultaneously.

Alternatively, the fragments were detected one-by-one in a more sensitive than Orbitrap quadrupole mass spectrometer (Extrel). We perform 10-20 measurements at each UV wavelength at a repetition rate of 10 Hz. Each spectrum was measured 3 times to ensure its reproducibility. The details of the setup can be found elsewhere.^[10]

The recorded with the Orbitrap-based mass spectrometer two-dimensional data array (ion abundance vs m/z and wavelength) was, first, pre-treated with the Peak-to-Peak software package (Spectroswiss) for peak detection and a baseline correction. The rectified data array then was stored as the 2D matrix identity of an isomer.

b) Chemicals

All the carbohydrates were purchased from Carbosynth (>95% purity) and ROTH (>98% purity) and used without further purification. Aromatic molecules of >98% purity were purchased from Sigma-Aldrich and TCI. All the solvents of HPLC grade and acetic acid of >99% purity are from Sigma-Aldrich.

c) Computational details

The lowest energy conformers of the protonated carbohydrate-aromatic complexes were calculated in a two-step procedure: 1) Molecular Dynamics (MD) annealing and 2) Density Functional Theory (DFT) rectifications.

The MD simulations were performed with NAMD2 package,^[18] referenced as “NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign”. To describe interatomic interactions, the CHARMM36 force-field optimized for carbohydrates^[19] and for proteins^[20] were used. Calculations were performed in a spherical cell with the radius of 3 nm and with 1 fs integration time step. In each simulation, the system was disordered for 20 ps at temperature 1000 K and subsequently cooled at a rate of 5 K/ps to 300 K. After this annealing, the system was equilibrated for 125 ps, and, finally, the energy was minimized in 1000 time-steps. The above procedure was repeated 500 times for every carbohydrate-aromatic pair.

The obtained conformations were clustered on the basis of their geometry: (i) the two angles of relative position of the carbohydrate and aromatic rings; (ii) the distance between rings; (iii) the angle between the C-NH₃⁺ and C-NH (-NAc) groups (or C-CH₂OH for the non-modified molecules). The energies of the 20 lowest-energies conformers were further optimized by DFT calculations.

The DFT optimizations of the structures and the calculations of harmonic IR spectra for the final lowest energy geometries were performed with NWChem package. The geometries of the 20 lowest-energies conformers were optimized with a low accuracy (energy precision ~0.6 kcal/mol) using the B3LYP functional with the Grimme's DFT-D3 empirical dispersion correction and the 6-31+G(d,p) basis set. The geometries of the unresolved lowest-energies conformers were further optimized with more accurate convergent criteria (energy precision ~0.03 kcal/mol).

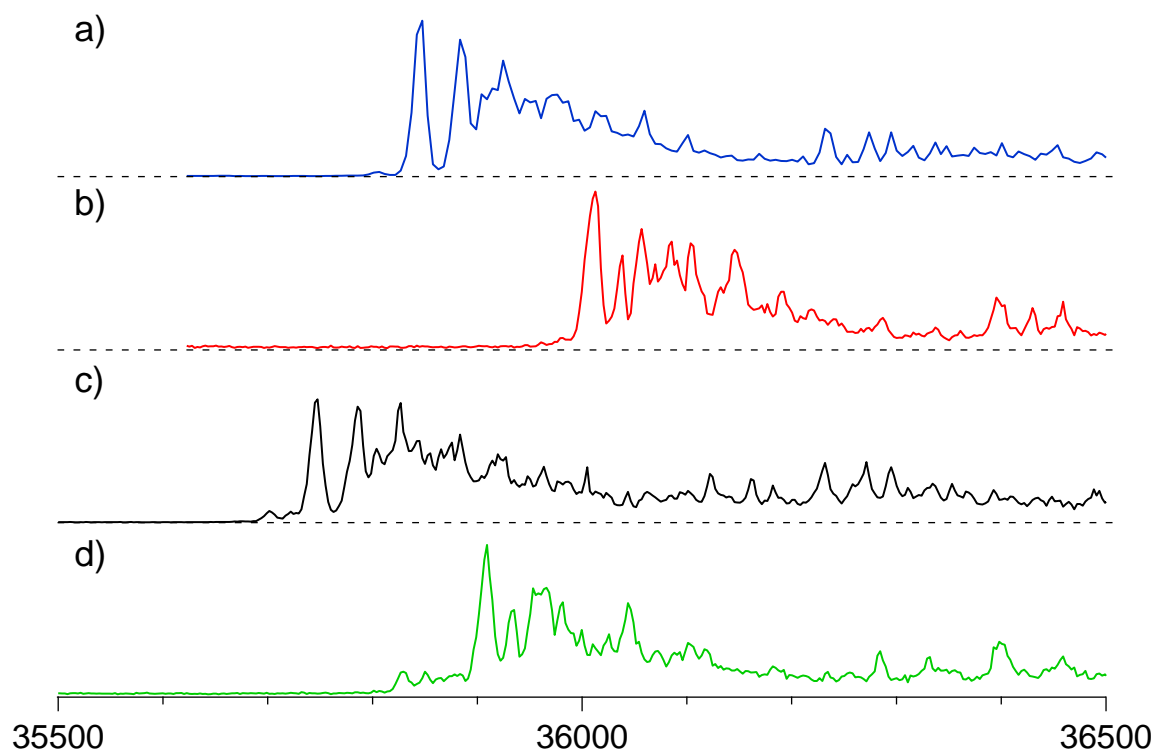


Figure 5.S1. UV gain spectra of the methylated derivatives of GalNAc-TrmH⁺ non-covalent complexes **a)** α -MeGalNAc-TrmH⁺ **b)** β -MeGalNAc-TrmH⁺ **c)** α -MeGalNAc-MeTrmH⁺ and **d)** β -GalNAc-MeTrmH⁺.

Table 5.S1. Blue shifts of UV band origins in aromatic as a function of the bond length.

Complex	Shift, cm ⁻¹	H-bond length, Å	Ref.
(hydroquinone) ₂	390	1.92	[16a]
(phenol) ₂	353	1.92	[16b], [16c]
phenyl vinyl ether-CH ₃ OH	130	2.07	[16d]
1,2-dimethoxybenzene-H ₂ O	127	2.04	[16e]
anisole-H ₂ O	119	1.97	[16f], [16g]
anisole-(H ₂ O) ₂	64	2.19	[16f]
3-phenoxy propanolamine-H ₂ O	40	2.30	[16h]

Additional References

34. J. C. Phillipset al., Scalable molecular dynamics with NAMD. *J.Comp. Chem.* 2005, 26, 1781-1802.
35. O. Guvenchet al., Additive empirical force field for hexopyranose monosaccharides. *J.Comp. Chem.* 2008, 29, 2543-2564.
36. A. D. MacKerell, M. Feig, C. L. Brooks, Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* 2004, 126, 698-699.

CHAPTER 6.

Identification and Quantification of Any Isoforms of Carbohydrates by 2D UV-MS Fingerprinting of Cold Ions

OVERVIEW

Biological functionality of isomeric carbohydrates may differ drastically, making their identifications indispensable in many applications of life science. Due to the large number of isoforms, structural assignment of saccharides is challenging and often requires a use of different orthogonal analytical techniques. We demonstrate that isomeric carbohydrates of any isoforms can be distinguished and quantified using solely the library-based method of 2D UV-MS photofragmentation of cold ions. The two-dimensional “fingerprint” identities of UV transparent saccharides were revealed by photofragmentation of their non-covalent complexes with aromatic molecules. We assess the accuracy of the method by comparing the known relative concentrations of isomeric carbohydrates mixed in solution with the concentrations that were mathematically determined from the measured in the gas phase fingerprints of the complexes. For the tested sets with up to five isomers of di- to heptasaccharides, the root-mean-square deviation of 3-5% was typically achieved. This indicates the expected level of accuracy in analysis of unknown mixtures for isomeric carbohydrates of similar complexity.

The results presented in this chapter were published in:

E. Saparbaev, V. Kopysov, V. Aladinskaia, V. Ferriere, L. Legentil, O. Boyarkin. Identification and Quantification of Any Isoforms of Carbohydrates by 2D UV-MS Fingerprinting of Cold Ions. ***Analytical Chemistry*** 2020, 92, 21, 14624-14632, DOI: <https://doi.org/10.1021/acs.analchem.0c03122>

6.1. INTRODUCTION

Carbohydrates are ubiquitous in nature. Their various biological functions range from being a fuel for our brain and signalling the state of living cells to antimicrobial defence of infants¹ and shielding viruses from recognition by human immune system.² The tremendous isomeric diversity of carbohydrates allows nature to assign glycans of specific structures to each function of this variety, but it also makes identification of these isoforms challenging.³ The composition of glycans of the same mass may differ by the monosaccharide units this polymer is composed of, by their order, linkage (1→3, 1→4, etc. and α/β) and branching. The structural difference between the isomeric monosaccharide units includes orientation of hydroxyl groups (epimers and α/β anomers), the ring size (e.g., pyranose and furanose) and the absolute configuration (*D* or *L*). Natural modifications, such as N-acetylation, phosphorylation, etc., further multiply the isomeric diversity of carbohydrates *in vivo*.

There is no single analytical approach capable of distinguishing carbohydrates in all their numerous isoforms.⁴ The time-proven methods of x-ray diffraction, NMR⁵ and, since recently, cryo-electron microscopy^{6,7} are capable to provide direct structural information for carbohydrates, although each of these techniques has its own limitations and practical drawbacks. The most common method of structural analysis of isomeric carbohydrates involves combination of enzymatic digestion for a controlled cleavage of large oligosaccharides, liquid chromatography (LC) to separate the produced isomers and single/tandem mass spectrometry (MS) to identify them.⁸ Despite its wide time-proven applicability, the method has certain limitations, such as the use of compound-specific conditions for separation, the need for chemical derivatization (e.g., by reductive amination and/or permethylation) and, often, an insufficient resolution in separating multicomponent mixtures of oligosaccharides.^{9,10} Ion mobility spectrometry (IMS) is another technique that is capable of separating isomeric ions that exhibit different collision cross sections.¹¹ Structural difference between isomeric glycans is often insufficient however for their separation in complex mixtures,¹² although the technique

is demonstrating a rapidly increasing selectivity.¹³ More generally, the values measured by chromatography and ion mobility (retention and arrival times, respectively) are not fundamental to molecules and dependent on specific experimental conditions and their stability (e.g., type of chromatographic column, pressure, temperature, etc.),¹⁴⁻¹⁶ This may limit reproducibility in high-resolution measurements, making a use of internal calibrants indispensable.

2D UV-MS fingerprinting is a recently developed approach that is capable of highly accurate identification of isomeric biomolecules.¹⁷⁻²² This analytical technique integrates UV photofragmentation (UVPD) spectroscopy of cryogenically cooled ions with broadband high-resolution (e.g. Orbitrap-based) mass spectrometry. It measures abundances of all UV-induced photo fragments at once in function of UV wavelength, thus providing a two-dimensional data array, named 2D UV-MS fingerprint, as a characteristic of an ion. Cryogenic cooling of mid-size biomolecular ions often enables vibrational resolution in their electronic spectra,²³ which makes them highly distinguishable. The UVPD mass spectra may also exhibit certain isomeric specificity. The synergy of the two techniques, spectroscopy and mass spectrometry, makes 2D UV-MS fingerprints highly individual to structure of ions, including isomers and conformers. Different from LC and IMS, optical spectroscopy reflects molecular transitions between quantum states, which are fundamental to ions. This makes 2D UV-MS fingerprints the highly reproducible standards that can be shared across different laboratories. The high performance of 2D UV-MS cryogenic fingerprinting was earlier demonstrated in library-based identification and accurate quantification of conformers and isoforms of different peptides in their multicomponent mixtures,^{17,18,20,24} as well as of small isomeric drugs on a concentration scale of a few ng/mL.¹⁹ Here we demonstrate the use of 2D UV-MS cryogenic fingerprinting for identification and quantification of natural carbohydrates for all types of their numerous isoforms. The tested sets of isomers range from mono- to hepta-saccharides and from two to five compounds in a set. The accuracy of relative library-based quantifications for some of the sets was

determined by fingerprinting several multi-compound solution mixtures of isomeric molecules. Finally, we demonstrate that in certain cases the 2D UV-MS method can be reduced to the technically less demanding one-dimensional UV fragment spectroscopy, yet without a significant loss of accuracy in the quantification.

A straightforward application of 2D UV-MS fingerprinting to glycans is challenging, because they do not absorb in UV. Instead, we make use of non-covalent complexes of glycans with protonated aromatic molecules, such as phenylalanine, tyrosine and its decarboxylated analog tyramine.^{22,25-27} These complexes can be readily prepared by dissolving a glycan and an aromatic molecule in a standard for MS methanol/water solution. The complexes appear to be strong enough to survive electrospray ionization that brings them to the gas phase. Our estimate suggests that, under mild conditions of our electrospray, up to half of the dissolved N-acetylated glycans bind to the aromatic molecules and appear in the gas phase as protonated complexes. The isomeric structure of an analyzing carbohydrate is communicated to the reporter aromatic molecule by non-covalent interactions, which induce changes in the UV absorption of the latter. In particular, an interplay of different types of H-bonds in isomeric complexes makes the position of their UV absorption onset very sensitive to the isoform of an analyzing carbohydrate.²² We also found that the number of low-energy conformers is much lower for non-covalent complexes of glycans with aromatic molecules than for glycan-alkaline ion adducts,^{27,28} which are often used for identifications of isomeric carbohydrates by IR spectroscopy.²⁹ This fortunate circumstance slows spectral congestion in UV spectra of cold carbohydrates upon increase of their size. The conserved vibronic structure enables high spectral isomeric specificity of 2D UV-MS fingerprints. The MS isomeric specificity arises from the phenomena of UV-induced proton transfer from sensor ion to analyzing carbohydrate molecule in the complexes.³⁰ The transfer results in fragmentation of glycans, which often yields isomer-specific pattern of fragment-MS.

6.2. EXPERIMENTAL APPROACH

Chemicals. All the carbohydrates, except for furanoside Gal/NAc, were purchased from Carbosynth (>95% purity), ROTH (>98% purity) and Acros Organics (>98% purity), and used without any further purification. The Gal/NAc glycan is not commercially available and was synthesized according to the procedure described elsewhere³¹ and purified to >95%. Aromatic molecules of >98% purity were purchased from Sigma-Aldrich and TCI. All the solvents were of HPLC grade and acetic acid was of >99% purity. Protonated gas-phase carbohydrate-aromatic complexes were produced from $5 \cdot 10^{-5}$ M equimolar solutions of a carbohydrate and an aromatic molecule in the 1:1 water-methanol mixture with 0.5 % of acetic acid.

Apparatus. Apart from a few minor changes, the details of our experimental setup has been describes elsewhere.³² Briefly, the ions, produced in the gas phase from solution by a nano-electrospray ionization (n-ESI) source, enter an electrodynamic ion funnel (IF) orthogonally to its axis through a 100 mm long stainless steel capillary of 0.7 mm ID. The ions are then turned by 90° to travel along the IF axis at pressure of 6 mbar and exit the funnel through the conductance limit of 2.5 mm diameter. They subsequently pass through two skimmers (1.5 and 2 mm in diameter), which allow for a gradual reduction of pressure from 0.5 mbar to 10^{-4} mbar. The ions are then accumulated and thermalized in an octupole ion trap for 97 ms, before they are released and pass through a quadrupole mass filter (Q1), which is set to select the ions of interest. Mass-selected ions are then turned by 90° using an electrostatic bender, focused by a stack of three electrostatic lenses and moved through an RF octupole guide into a cold octupole trap,³³ which is kept at 6 K. The trap is driven by two 1 MHz sinus waveforms with peak-to-peak amplitudes of 50-100 V. The stored in the trap non-covalent complexes are cooled down to ~10 K in collisions with He buffer gas,³³ which is pulsed into the trap shortly before arrival of the ions. Approximately 85 ms later the ions are irradiated by a pulse of UV light (1.8 ± 0.7 mJ/pulse, 5 ns duration, ~ 6 cm⁻¹ spectral linewidth),

produced by a UV optical parametric oscillator (OPO; NT 342C, EKSPLA) and undergo photo fragmentation. The fragment and parent ions were 90° turned by the second electrostatic quadrupole bender either toward a highly sensitive quadrupole mass spectrometer (QMS) or toward high-resolution broadband Orbitrap-based MS (Exactive, Thermo Fisher). UVPD spectra and 2D UV-MS fingerprints are measured by continuously recording the yield of the photofragments using the QMS and the Orbitrap-based MS, respectively, while scanning UV wavelength. The repetition rate of the cooling/fragmentation cycle was determined by the 10 Hz repetition rate of the OPO, such that the parent ions experienced only one OPO shot in each cycle. At each UV wavelength the yield for a single fragment (with QMS) or the entire fragment mass spectrum was measured in 10 cycles and averaged to give a data point in UVPD spectrum or a fragment MS in 2D UV-MS fingerprint, respectively.

Special measures have been taken to narrow the variation of pulse energy of the OPO over the wide tuning range (210-350 nm). The angular positions of non-linear crystals in the OPO doubling/mixing stages were detuned from their optimal phasematching angles, such that the output energy at each wavelength was close to the minimum energy attainable within the desired spectral range. This minimizes the potential contributions to UVPD from non-linear absorption and therefore increases reproducibility of the fingerprints. In addition, the measured ion signals were normalized to OPO pulse energy, which was measured by a broadband pyroelectric detector.

In measurements with Orbitrap-MS, the whole fragment mass spectrum was normalized to the total ion signal detected in the cycle. In QMS measurements we employ 20 Hz cooling cycle and detect fragment or parent ions in 10 Hz alternative cycles with and without UV pulse, respectively.

Data processing. The measured 2D UV-MS data arrays (intensity vs UV wavelength and m/z) were, first, pre-processed for detection of mass-peaks using the PeakbyPeak software package (Spectraswiss). The reduced 2D data arrays of the known isomeric compounds were stored as a library

tagged by the exact mass. 2D UV-MS fingerprints of solution mixtures of (unknown) isomeric molecules were pre-processed the same way and then mathematically decomposed in the basis set of the 2D matrices from the library tagged by the mass of the isomers. The details of the matrix decomposition procedure are described elsewhere.^{17,19} The same software is used for decomposition of UVPD spectra measured with QMS.

6.3. RESULTS AND DISCUSSION

Monosaccharide isomers. The first but essential step toward structural identification of isomeric oligosaccharides is to distinguish their structural units – cyclic monosaccharides. These building blocks are classified by five types of isomerism as epimers, α/β -anomers, *D/L*-enantiomers, aldose/ketose and ring-size isomers.

Aldose-ketose. Glucose and galactose example the most common mammalian aldose; ketoses (e.g. fructose) occurs widely in plants. Aldoses have a terminal formyl COH group, while ketoses contain a carbonyl group bound to two carbons. Different chemical properties of these isomers allow their reliable identification by a variety of techniques.³⁴ Nevertheless, for the sake of completeness we tested the UVPD spectroscopy approach for distinguishing isomers of this type. Figures 6.1a and 6.1b show UVPD optical spectra of protonated aromatic molecule tyramine (Trm) in non-covalent complexes with *D*-glucose (*D*-Glc) and *D*-fructose (*D*-Fru), respectively. Despite the very close positions of the onsets of UV absorption, the very different shapes, widths and positions of the measured vibronic transitions make the spectra readily distinguishable. Regarding the relative simplicity and similarity of the UV fragment MS for both isomeric complexes, a use of a more complex method of 2D UV-MS fingerprinting seems to be unnecessary in this case.

Epimers. Different from aldose-ketose isomers, epimers are stereoisomers that differ in absolute configuration about single of their several asymmetric carbons. For example, N-acetyl-D-glucopyranose (GlcNAc) and N-acetyl-D-mannopyranose (ManNAc) differ only in the configuration at the C2 position, whereas N-acetyl-D-galactopyranose

(GalNAc) is an epimer of N-acetyl-D-glucopyranose at the C4 position. Figure 6.1 (c-e) compares UVPD optical spectra of these three isoforms. The revealed differences in the spectral shapes and/or the onsets of UV absorption make the spectra unambiguously specific to these isomers. In opposite, the UVPD MS of the isomers do not exhibit any clear specificity.

When complexes of all isomers of a library exhibit sharp and well separated (compared with the width of the rising edge of the absorptions) onsets of their UV absorptions, the positions of the onsets alone can be used as reliable spectroscopic tags of these isomers. The onsets of unknown library compounds that are mixed together in solution can be located by measuring the UVPD yields at a few but “critical” pre-selected wavelengths only, instead of measuring the whole continuous spectrum at, typically, hundreds of wavelengths.^{17,19,20} Such approach may shorten the measurements to a few seconds, which would allow for performing isomeric identifications online with HPLC.

Anomers. Anomerism is a particular case of epimerism, where structural variations are about the hemiacetal/hemiketal carbon in a carbohydrate ring (e.g. about C1 for aldoses). Cyclisation of an open-chain carbohydrate results in one of the two possible anomeric forms, notated as α - and β -. The anomers coexist at equilibrium in aqueous solutions and spontaneously interconvert upon reopening-closing the ring. Methylation of the anomeric 1-OH group inhibits this conversion. Noticeable differences between IR spectra of α/β anomers were earlier revealed for neutral and protonated complexes of aromatic molecules with MeGalNAc and with GalNAc.^{26,27} Figures 6.1f-6.1g illustrate how much UVPD spectra of complexes of TrmH⁺ with methylated glycans α -MeGalNAc and β -MeGalNAc differ in the onsets of UV absorption, but also in many details of highly structured absorption bands. These differences ensure an unambiguous spectral identification of the two anomers. An earlier detailed study revealed that in complexes with GalNAc the anomeric 1-OH group, which position is the only structural difference between the α/β isomers, is not involved to any non-covalent bonds with the sensor aromatic and therefore cannot directly influence the UV absorption of Trm. Instead, the isomeric

information is communicated by the analyte to the reporter through an inductive polarization of the H-bond, formed between the hydrogen of 6-OH in the glycan and the oxygen atom of the tyramine side chain.²⁷ A stronger polarization of the bond results in a larger blue shift of UV absorption by the aromatic, whose ring is conjugated with the hydroxyl.

Ring-size isomers. Upon cyclization of open-chain glycan molecules, monosaccharides can form rings with different number of carbon centers. For example, N-acetyl-galactosamine may form five- or six-membered rings, named furanose (Gal*f*NAc) and pyranose (Gal*p*NAc), respectively. Although, forming pyranoses is energetically more favorable, the furanose isomers are also observed in nature as units of polysaccharides.³⁵ Figures 6.1g and 6.1h compare UVPD spectra of protonated tyramine in complexes with methylated N-acetyl-galactosamine in pyranose (β -MeGal*p*NAc) and furanose (β -MeGal*f*NAc) forms. The methylation was required to lock the carbohydrates in their β -anomeric form. Overall, the significant, ~5 nm, spectral shift of the electronic band origins and the very different vibronic structures of the two spectra make them readily distinguishable. The spectra thus can be used for a reliable visual identification of the isomers.

Enantiomers. When a carbohydrate and its mirror image cannot be superimposed, such two isomers are called *D* and *L* enantiomers of this chiral molecule. Except for circular dichroism,^{36,37} enantiomers of glycans have exactly the same physical and optical properties and therefore identical spectroscopy. Complexes of *D/L* glycans with tyramine, which is an achiral molecule, must adopt exactly the same geometry, such that their UV spectra is to be indistinguishable too. Figures 6.2a and 6.2b illustrate this fundamental property of enantiomers, comparing UVPD spectra of Trm in complexes with *D/L*-glucose. Apart from some small deviations that characterize the reproducibility of the experiment, the two spectra are, indeed, identical. The *D/L* isoforms may drastically differ however in interactions with chiral molecules. This structural property allows, for instance, performing separation of enantiomers by chiral HPLC or IMS.³⁸ Similarly, non-covalent complexes of *D/L* glycans with a chiral aromatic

should have different geometries and therefore distinguishable spectroscopy.

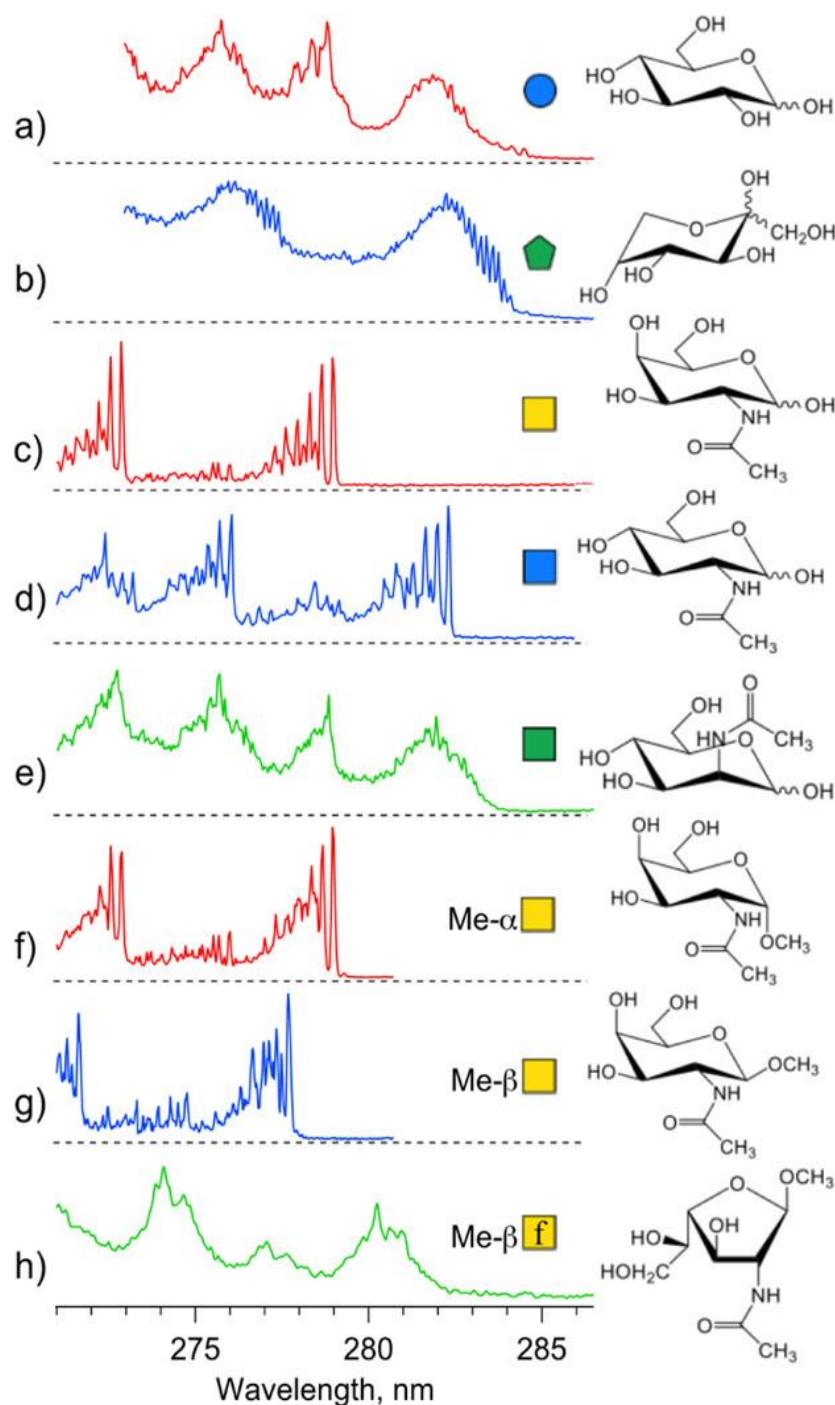


Figure 6.1. UVPD spectra of cold non-covalent complexes of protonated tyramine (TrmH^+) with monosaccharides: a) *D*-Glc, b) *D*-Fru, c) *D*-GalNAc, d) *D*-GlcNAc, e) *D*-ManNAc, f) α -MeGalNAc, g) β -MeGalNAc, h) β -MeGal/NAc. The spectra (a) and (b) were measured by detecting TrmH^+ photofragment with QMS; the spectra (c) to (h) were generated from the 2D UV-MS fingerprints measured with the Orbitrap-based MS by integrating them

over m/z dimension. The standard colour/shape-coded symbols of the saccharides and their structures are shown on the right.

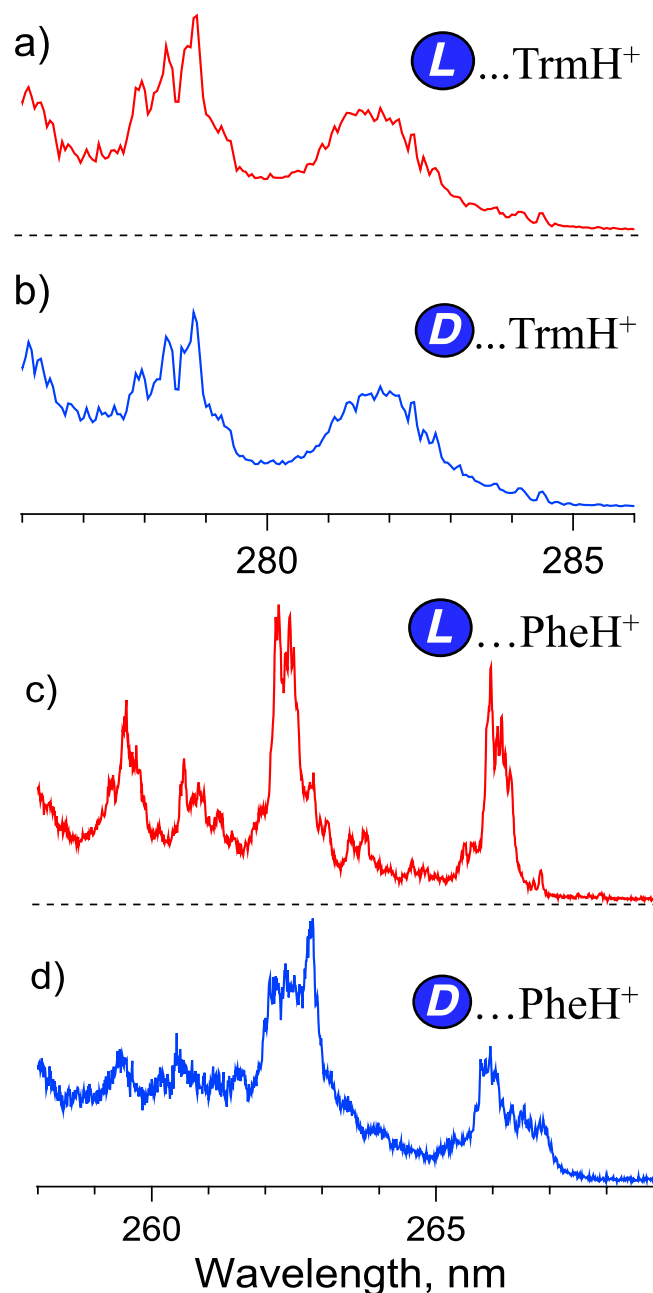


Figure 6.2. Photofragmentation UV spectra of the complexes of enantiomeric *L/D* glucose with non-chiral (Trm) and chiral (*L*-Phe) chromophores.

Figures 6.2c and 6.2d show UVPD spectra of *L*-Phe, which is a chiral molecule, in complexes with *D/L*-Glc. Different from Trm, the side chain of Phe does not have an OH group, whose H-bonding to glycans induces the most significant differences in UV spectra (e.g. shifts of the electronic band

origins) of the complexes.²² Nevertheless, the change of the reporter molecule makes the spectra of *D/L* -Glc...PheH⁺ complexes clearly different, which is likely due to interplay of the relatively weak OH- π and CH- π interactions.

Fundamentally, the spectra of complexes of a chiral aromatic with a carbohydrate must remain the same upon a simultaneous change of chirality of both partners: $D \leftrightarrow L$. Figure 6.3 compares UVPD spectra of Gal-PheH⁺ complexes for all four combinations of chirality. Within the accuracy of the experiment, the spectra of *L-D* and *D-L* complexes look, indeed, identical (Figures 6.3a and 6.3b) Similarly, the complexes with *D-D* and *L-L* combinations of enantiomers are also indistinguishable by spectroscopy (Figures 6.3c and 6.3d), while their spectroscopic signature is clearly different from that of the complexes with the molecules of the mixed chirality. Apart from curiosity, the analytical implication of this comparison is in the possibility to detect appearance of a previously unobserved enantiomer (e.g. in asymmetric catalysis) using UV spectroscopy of its existing mirror image in complex with a chiral reporter molecule. The complementary enantiomer then can be identified by spectroscopy of the reporter of the opposite chirality in complex with the suspected isomer.

Overall, the presented above data demonstrate that monosaccharides of any type of isoformerism can be distinguished by cold-ion UVPD spectroscopy of the easily prepared in solution glycan-aromatic non-covalent complexes.

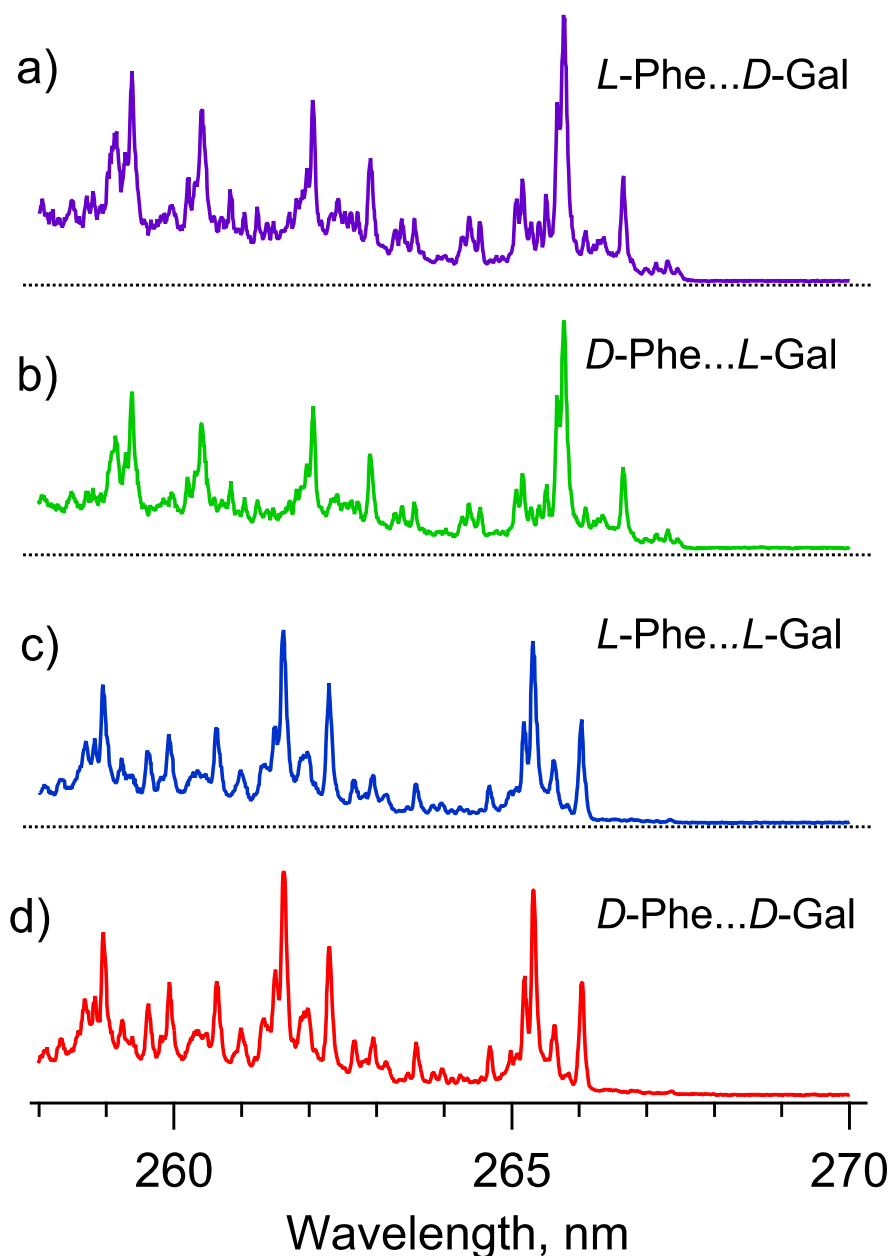


Figure 6.3. UVPD spectra of protonated *L/D*-Phe with *L*- and *D*-enantiomers of Gal.

Disaccharide isomers. In di-, oligo- and polysaccharides the glycosidic linkage may occur (i) between the same or different monosaccharaides and in different order (*compositional* isomers), (ii) at different points of attachments (*connectivity* isomers) and (iii) with different orientations of glycosidic bonds (*configurational* isomers). In addition, large glycans may form not only linear, but also branched

polymeric structures, which further add to the diversity of the connectivity isomers. Furthermore, glycans can naturally undergo different modifications, such as, for instance, N-acetylation, phosphorylation and sulfation. The modifications can be multiple and occur on different units of the polymers (*regioisomers*), thus increasing the number of possible isomers. All these additional features multiply the already high isomeric diversity of monosaccharaide building blocks.

Compositional. A disaccharide with monoisotopic mass of 342.12 Da can be composed, for instance, of either glucose or galactose or mannose, linked in different orders. UVPD fragment mass spectra of the complexes of these disaccharides with protonated Trm or Tyr are still low informative and contain, mainly, the protonated aromatic molecules and their fragments (Figure 4). This is similar to the UVPD MS of the monosaccharaides that have no NAc group. The presence of this group, which has high proton affinity, enables proton transfer from the reporter aromatic and, subsequently, fragmentation of the analyzing glycan. The isomer-specific content of the 2D UV-MS fingerprints for the non-acetylated disaccharides remains, largely, in the spectroscopic domain. Figures 6.4a-6.4c illustrate the difference between the UV spectra of compositional isomers of disaccharides. All three glycans exhibit clearly distinct spectroscopic signatures, which can be used for identification of the isomers.

Configurational. Maltose and cellobiose (α/β -Glc-1,4-Glc) differ only by the orientation of glycosidic linkage between a pair of monosaccharaide units (figures 6.4c and 6.4d, respectively). Nevertheless, their spectroscopic signatures in complexes with TyrH⁺ differ substantially. The pair of isomers of mannobiose (α/β -Man-1,4-Man) in figures 4b and 4f also exhibits well distinguishable UV spectra that can be used for identification of these disaccharides.

Connectivity. In these isomers the two monosaccharaides are linked by a glycosidic bond between carbon atoms at different positions (e.g. 1→2, 1→3, 1→4 etc.). In the case of the disaccharides containing, for instance, only mannoses as building blocks and just for the

α -configuration of the glycosidic bond, four connectivity isomers are naturally abundant: α -Man-1,2-Man, α -Man-1,3-Man, α -Man-1,4-Man, α -Man-1,6-Man. 2D UV-MS fingerprinting method is capable of distinguishing all these isoforms. Figures 6.4e-6.4h show the UVPD spectra of these isomers in complexes with *L*-tyrosine. The spectra differ significantly from each other in the positions of the onset of UV adsorption and/or in the shapes of the vibronic bands. Overall, each of the spectra in Figure 6.4 is unique and can be used as a spectroscopic tag for identification of the isomers.

Regioisomers. Figures 6.5a and 6.5b compare UVPD optical and mass spectra of TyrH⁺ in complexes with disaccharide regioisomers β -GalNAc-1,3-Gal and β -Gal-1,3-GalNAc, which differ only by the location of N-acetylation. Figure 6.5 also compares UVPD identities of the disaccharides that differ by combinations of two and even three types of isomerism: regio- and configurational (6.5a vs 6.5c), regio- and compositional (6.5b vs 6.5d) and regio-, compositional and configurational (6.5c vs 6.5e). In addition to the apparent isomeric specificity of the optical spectra, the UVPD fragment mass spectra of different isomers also exhibit a clear difference in presence and abundance of many photo fragments. The presence of NAc group, which has high proton affinity, enables UV-driven proton transfer from the reporter aromatic molecule to an analyzing glycan with subsequent fragmentation of the latter (Fig. 6.5). The transfer greatly enriches the fragment MS, which appear to be isomer-specific. The fact, that the 2D UV-MS fingerprints of the amino glycans become isomer-specific in both the wavelength and the m/z dimensions, greatly improves the reliability of the method in identifications of single isomeric carbohydrates, but also increases its accuracy in quantification of their multicomponent mixtures (see below).

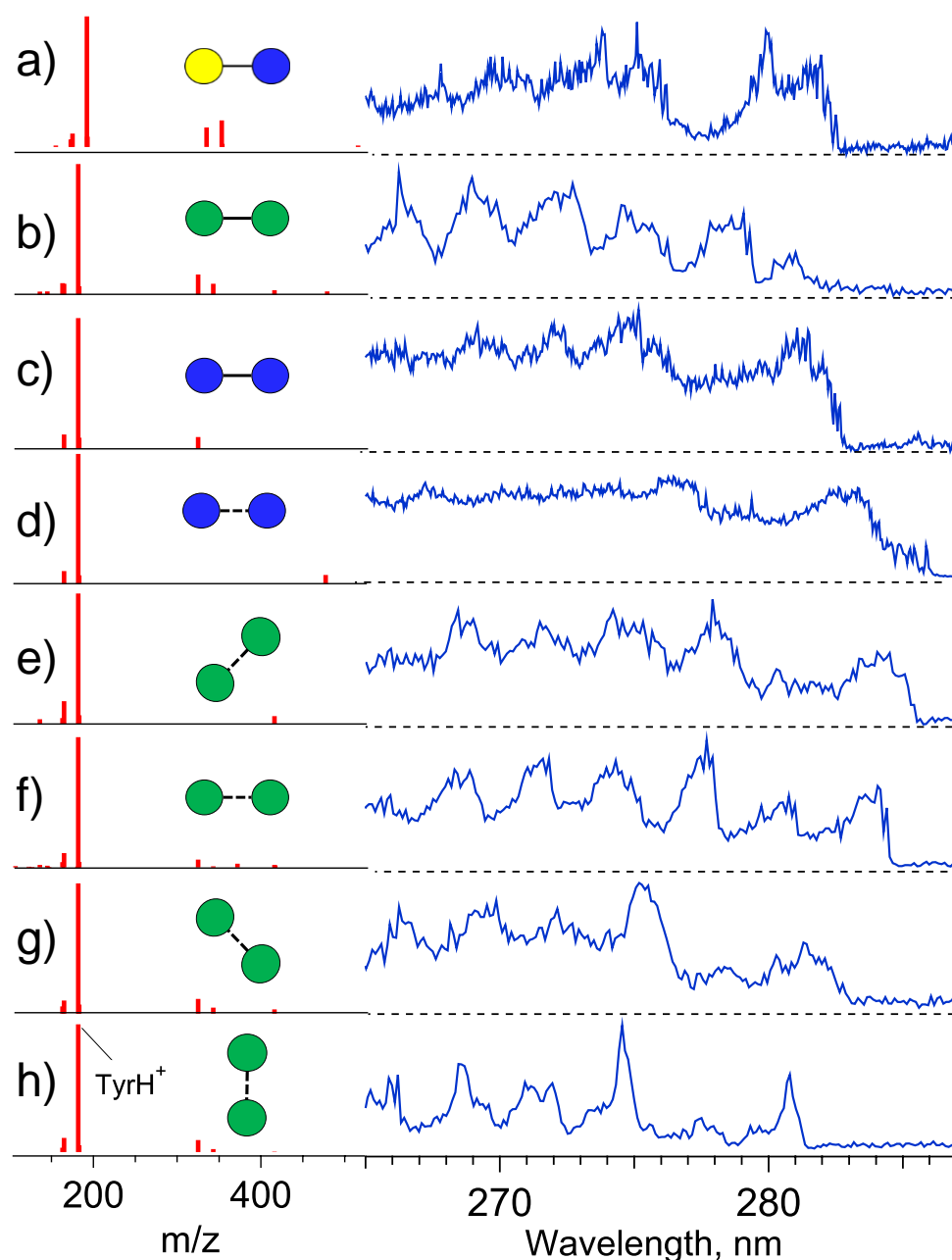


Figure 6.4. UVPD optical and mass spectra, generated by integrating over m/z and the wavelength, respectively, the 2D UV-MS fingerprints of complexes of tyrosine (*L*-Tyr) with isomeric disaccharides ($m/z=524.31\text{Th}$): a) β -Gal-1,4-Glc, b) β -Man-1,4-Man, c) β -Glc-1,4-Glc, d) α -Glc-1,4-Glc, e) α -Man-1,3-Man, f) α -Man-1,4-Man, g) α -Man-1,6-Man, h) α -Man-1,2-Man. Conventionally, dashed and solid lines used in the symbolic representation indicate α and β orientations of glycosidic bond, respectively; the orientation of the lines indicates the connectivity points.³⁹

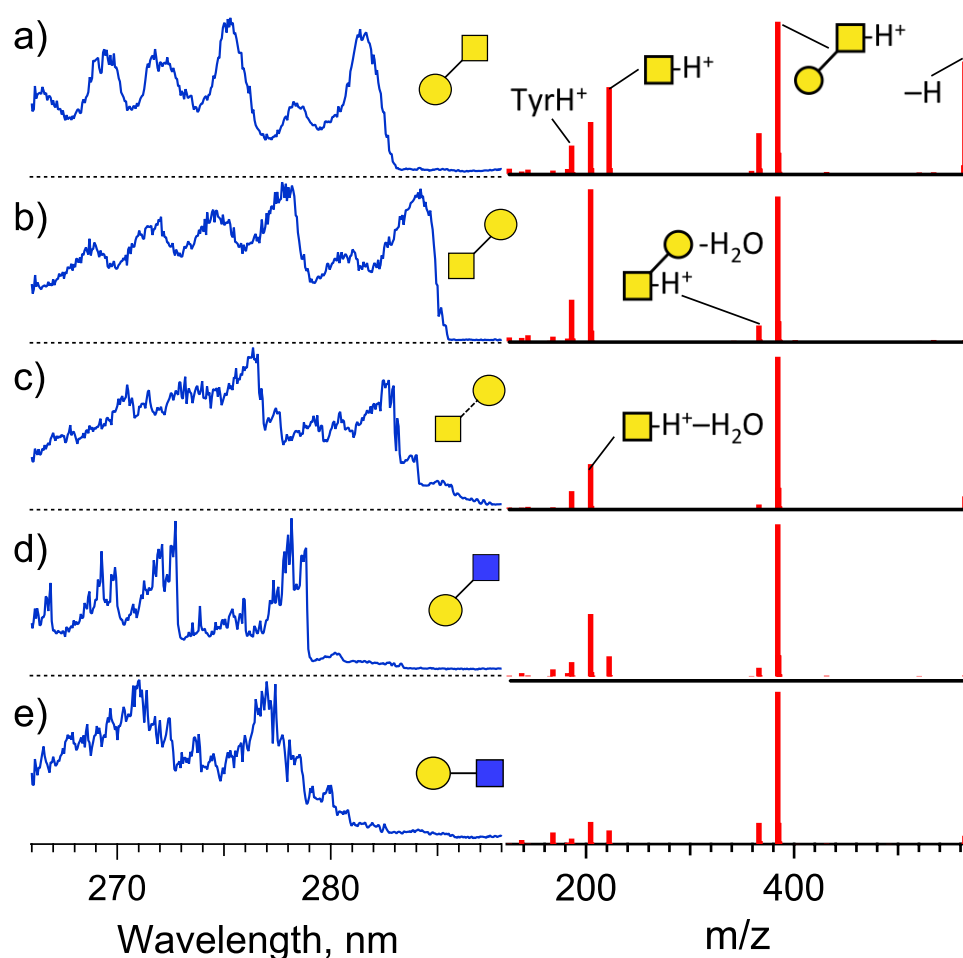


Figure 6.5. UVPD optical and mass spectra generated by integrating over m/z and wavelength, respectively, the 2D UV-MS fingerprints of non-covalent complexes of L -TyrH⁺ with a) β -Gal-1,3-GalNAc, b) β -GalNAc-1,3-Gal, c) α -GalNAc-1,3-Gal, d) β -Gal-1,3-GlcNAc, e) β -Gal-1,4-GlcNAc isomeric N-acetylated disaccharides. Some intense fragment MS peaks are labelled using standard colour/shape presentation of carbohydrates and chemical symbols.

Oligosaccharides. The isomeric diversity of glycans grows rapidly with their size due to the increasing number of combinations for composition and connectivity of monosaccharaides. This implies that identification of oligosaccharides, as compared with mono-/disaccharides, requires more selective analytical methods. Two circumstances may,

potentially, inhibit a use of 2D UV-MS fingerprinting of oligosaccharide-aromatic complexes for identification of large glycans. Single aromatic, which size is comparable with size of a monosaccharide, may not be capable to sense isoforms of the oligosaccharide units that are remote from the docking site of the aromatic molecule in the complexes. In addition, for larger glycans one may expect a larger number of conformers. The overlap of their slightly different UV absorptions would increase the inhomogeneous broadening of transitions in (non-conformer specific) UVPD optical spectra of the complexes. The broadening may wash out any details in the spectra, potentially, making them low isomer-specific. Figure 6.6 compares UVPD optical and mass spectra derived from 2D UV-MS fingerprints of trisaccharides. The optical spectra still remain sufficiently different to distinguish the three isomers easily. Compared with the disaccharides in Figure 6.5, UVPD of these N-acetylated trisaccharides results in rich fragment MS, which contain several highly abundant fragments. Although most of them are the same for all three isomers, the overall patterns of the MS become well distinguishable too. The glycans as large as tetra- and even hepta-saccharides also exhibit quite distinct 2D UV-MS fingerprints²² (Figures 6.S1 and 6.S2). We explain the ability of a small aromatic to sense isoforms of a large glycan by non-covalent binding of the former to different sites of the latter. For each particular configuration of the complex its UVPD would reflect only the local isoform of a glycan close to the binding site. The contributions from different available configurations collectively may reflect all the isomeric features of a glycan in its 2D UV-MS fingerprint.

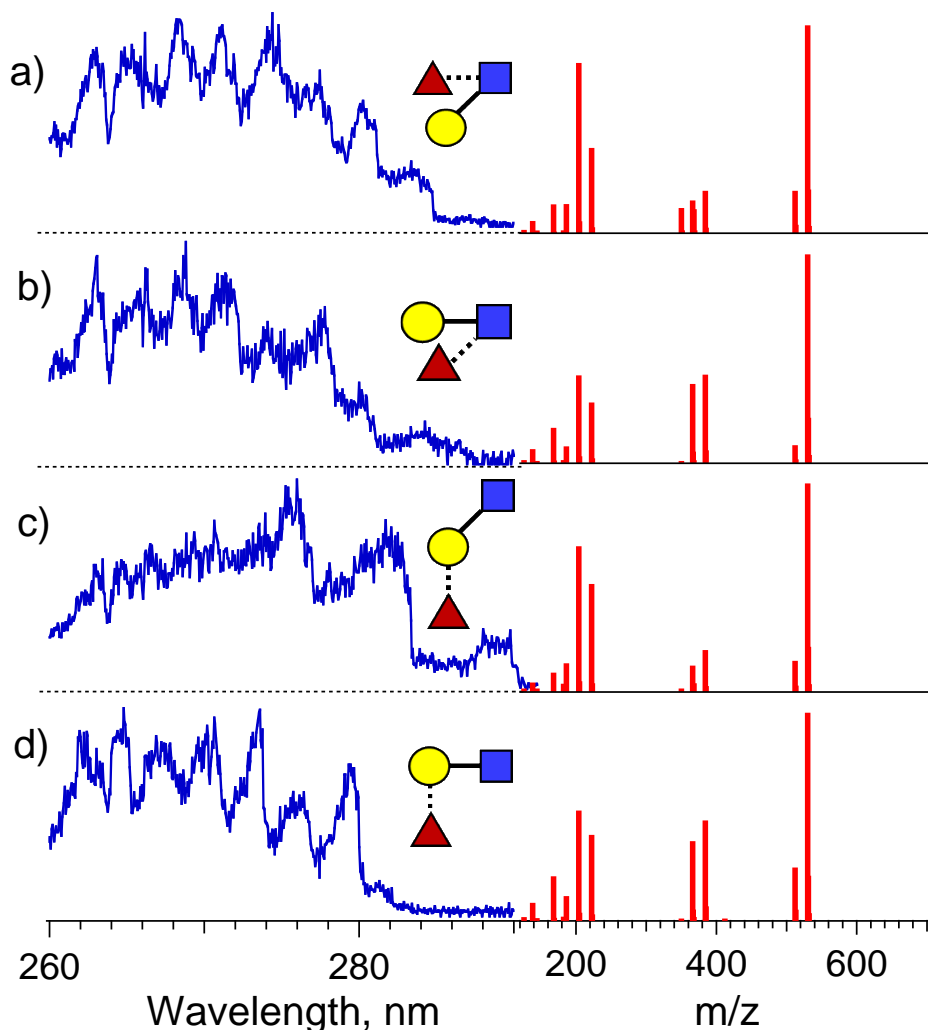


Figure 6.6. Photofragmentation UVPD optical and mass spectra, generated by integrating over m/z and wavelength, respectively, the 2D UV-MS fingerprints of non-covalent complexes of $L\text{-TyrH}^+$ with trisaccharides a) Lewis A, b) Lewis X, c) Blood Group H1, d) Blood Group H2.

Along with 2D UV-MS fingerprinting, over last few years different types of gas-phase IR spectroscopy also have been tested for recognition of single isomeric carbohydrates^{29,40-42} Visually distinct IRPD and IRMPD spectra were demonstrated for some isoforms of glycans, making a promising ground for further developments of these approaches as analytical methods. The latter requires however not only some visually distinct identities of isomers, but a way to quantify them in solution mixtures and an assessment of accuracy for such quantifications. While

this is likely still in agenda of IR spectroscopy-based approaches, the library-based 2D UV-MS fingerprinting technique has already demonstrated accurate quantifications of relative concentration of isomeric peptides and drug molecules in their multicomponent solution mixtures.^{17-20,32} Herein we use the same instrumentation and workflow for quantification of isomeric carbohydrates mixed in solution, but perform fingerprinting of their non-covalent complexes with aromatic molecules.

Quantification of isomers. First, 2D UV-MS fingerprints of the complexes of an aromatic molecule with each isomer of interest were measured and processed to create a library of 2D data arrays labelled by the mass of these isomers. Next, a fingerprint of the isomeric complex produced from a solution mixture of unknown number of the isomeric glycans of unknown relative concentrations was measured too. For each solution mixture its 2D data array was decomposed in the basis set of the isomers of the respective library using mathematical methods of non-negative matrix decomposition. The fit coefficients then give the relative concentrations of the isomers in the solution mixture.

We applied this workflow to 12 solution mixtures of one to three (out of five) isomeric disaccharides and an aromatic molecule tyramine (Trm), which was added to the solution for forming complexes with the carbohydrates. The use of Trm instead of Tyr as a reporter aromatic was not critical, although the lack of the -COOH group in tyramine makes UVPD optical spectra of the complexes more structured and therefore more isomer-specific. Figures 6.7a and 6.7b illustrate the quality of matrix decompositions for a 50:50 % solution mixture of β -Gal-1,3-GlcNAc and α -Gal-1,4-GlcNAc disaccharides. The figures compare the optical and mass spectra visualized by integrating over m/z and wavelength dimensions, respectively, the measured and the calculated 2D UV-MS matrices. The five fit coefficients of the decomposition imply 0%, 1.2%, 0%, 47.5% and 51.3% relative concentrations for the five isomeric disaccharides listed in Fig 6.7c (from top to bottom, respectively). The identity of the two mixed isomers was determined correctly; their calculated concentrations differ from the prepared solution ones by 2.5% at most. The analysis shows that, in

average over the prepared pool of 12 mixtures (Table 6.S1), relative concentrations of these five isomers have been determined with 5.3% accuracy (Fig. 6.7c). The error is substantially higher (8% and 15%) for the two mixtures that include β -Gal-1,3-GalNAc glycan, for which an exceptionally low yield of UVPD (Figure 6.S3) has been measured. For most of the mixtures the error is below 3%, however. It is exactly the synergy of spectroscopy and mass spectrometry that allows for such accurate quantification of isomeric carbohydrates in their mixtures.

As we noted above (e.g. Fig. 6.5), UVPD of complexes of protonated aromatic molecules with carbohydrates that have no NAc groups often yields trivial mass spectra, dominated by the aromatic and its fragments. Such spectra exhibit low isomeric specificity, which largely comes from optical spectra. This observation suggests that, perhaps, measuring a UVPD spectrum alone might be a sufficiently good but simpler alternative to the technically more demanding 2D UV-MS fingerprinting. In order to simulate this approach, we compared the accuracy of 2D UV-MS decomposition with the accuracy of one-dimensional fits of optical spectra. The latter have been generated from the 2D UV-MS fingerprints of the library isomers and of their solution mixture by integrating the 2D data arrays over m/z dimension. We did such a simulation for two heptasaccharides, Man₆GlcNAc-I and Man₆GlcNAc-II, and their 2:1 solution mixture (Fig. 6.S2). Both the 2D UV-MS and 1D UVPD-only deconvolutions earlier determined the relative concentrations as 71:29 and 68:32, respectively.²² This example demonstrates that even in the case of large oligosaccharides, which exhibit structureless and broad UVPD optical spectra, the small differences between them, collectively, enable high accuracy in quantification of the mixed isomers.

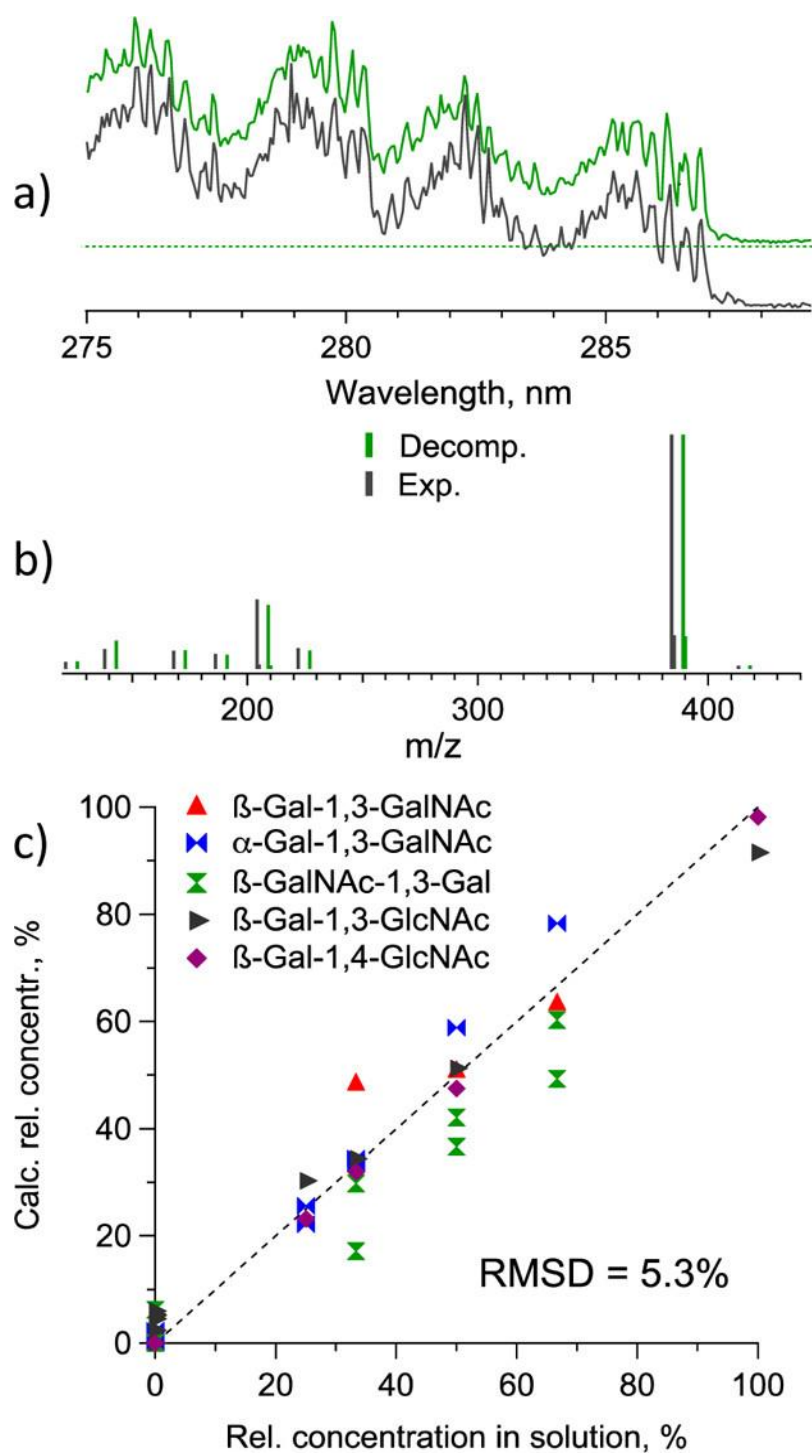


Figure 6.7. Comparison of the experimental (black) and decomposition (green) UVPD (a) optical and (b) mass spectra for the equimolar solution mixture of β -Gal-1,3-GlcNAc and β -Gal-1,4-GlcNAc disaccharides and of protonated Trm aromatic reporter molecule. For graphical clarity the decomposition spectra are offset up and by +4 Th in a) and b), respectively. (c) Calculated relative concentrations for 12 mixtures of one to three

isomers from the 5-component library versus their real solution concentrations (Table 6.S1).

In order to demonstrate this technical simplification, we used a quadrupole mass filter instead of the Orbitrap-based MS for detecting UVPD fragments. The filter was tuned to $m/z=182$ Th for maximum transmission of TyrH^+ in a low-resolution mode just sufficient to discriminate the protonated parent complexes. Figure 6.8ab compares UVPD spectra of anomeric disaccharides maltose and cellobiose in their complexes with the aromatic. Despite a lack of a resolved vibronic structure, the difference in many reproducible details makes the spectra suitable for identification and accurate quantification of this pair of isomers. Figure 6.8c illustrates the quality of the decomposition by comparing the spectra calculated and measured for the 1:1 solution mixtures of the two anomers. The fit coefficients suggest 47:53% for the relative concentration of the two glycans in this mixture. The decompositions of the spectra for 2:1 and 1:2 solution mixtures of the glycans give their relative concentrations as 64:35 and 27:73, respectively. Together, the three decompositions allow for an estimate of RMSD, which characterizes the expected accuracy in quantification of unknown sample mixtures, to 4.4%.

In comparison with 2D UV-MS fingerprinting, the evaluated above high accuracy of 1D UV deconvolution drops quickly however upon increasing the number of isomers in the library. Our simulations using 2D UV-MS fingerprints show that, for example, for a 1:1:2 mixture of three tetrasaccharides (blood group A1, A2 and A3/4; Figure 6.S1) the 2D decomposition gives the relative concentrations of 27:24:49, while 1D UV measurements and fit would result in a less accurate relative concentrations of 30:28:42. These data are consistent with the earlier reported simulations, that considered 2D UV-MS, 1D UV-only and 1D MS-only decompositions of 2D data arrays measured for isomeric peptides. Overall, for multicomponent libraries 2D UV-MS decompositions exhibit better stability in wide wavelength ranges, while 1D fits may give good accuracy only within specific narrow wavelength or m/z windows that

have to be predetermined in advance. It is worth stressing that for the spectroscopically large systems studied herein, the high isomeric specificity of UV spectra is based on cooling ions to a cryogenic temperature, which ensures no/little population of vibrationally excited levels.

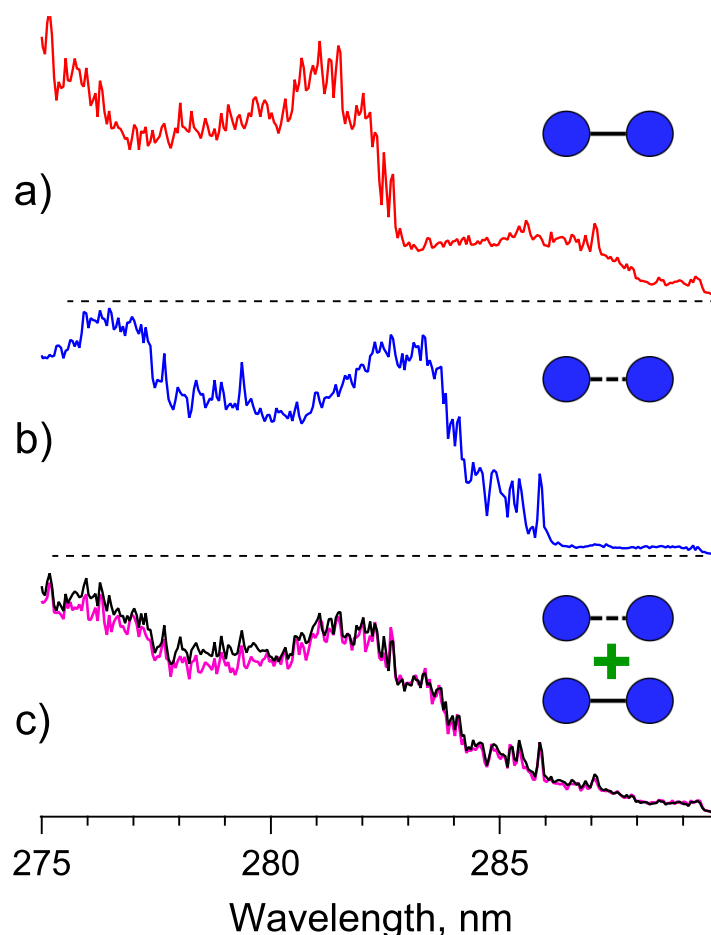


Figure 6.8. Measured with QMS UVPD spectra of non-covalent complexes of TyrH⁺ with isomeric disaccharides a) α -Glc-1,4-Glc (cellobiose), b) β -Glc-1,4-Glc (maltose) and (c) with both isomers in 1:1 mixture (pink trace). The black trace in (c) is the best numerical 1D fit that adds the spectra in (a) and (b) with coefficients of 0.465 and 0.535, respectively.

6.4. CONCLUSIONS

In this study, we have demonstrated how isomeric carbohydrates in any of their numerous isoforms can be distinguished and relative solution concentrations of saccharides can be quantified using our recently

developed approach of 2D UV-MS fingerprinting of cold ions. Mixing carbohydrates and suitable aromatic molecules in solution efficiently produces their protonated non-covalent complexes, a good fraction of which can be brought intact to the gas phase using soft ESI. Hydrogen bonds to aromatic molecule make glycans “visible” in UV. This enables recording the yield of photodissociation of the complexes as a 2D function of UV wavelength and of m/z of the appearing fragments. When the complexes are cooled to cryogenic temperature, such 2D fingerprints become quite specific to isomers of glycans. The sets of up to five isomers, and the carbohydrates as large as heptasaccharides, were tested. For most of the sets the expected accuracy in determination relative concentrations of isomers mixed in solution is better than 3-5%. For sets of a few isomers the 1D UVPD cold ion spectroscopy can be a sufficiently accurate while technically less demanding alternative to the 2D UV-MS sensing. Our test, in which a quadrupole mass filter replaced a broadband high-resolution Orbitrap-based MS, gives 4.4% accuracy 1D quantification of mixtures of two isomeric disaccharides. 2D UV-MS fingerprinting remains indispensable however for accurate quantifications with larger libraries and bigger glycans.

2D UV-MS fingerprinting of carbohydrates has potential for online coupling to liquid chromatography. To shorten the time of measurement to the scale of LC, the aromatic-glycan complexes can be sensed not continuously, but at a few “critical” wavelengths only, similar to what was earlier shown for peptides.^{17,19,20} In such a case, yet to be demonstrated, the premixing of aromatic molecules can be performed in between an LC column and an ESI capillary.

Overall, adding carbohydrates to the sample box of 2D UV-MS fingerprinting expands its capabilities in identification of isomeric biomolecules. It is practically important that the same hardware and software can be used for very different types of biomolecules, making the method quite flexible in analytical identifications of isomers.

REFERENCES

1. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **2012**, *22*, 1147-1162.
2. Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chemical Reviews* **1996**, *96*, 683-720.
3. Sharon, N.; Lis, H. Lectins as cell recognition molecules. *Science* **1989**, *246*, 227-234.
4. Gray, C. J.; Migas, L. G.; Barran, P. E.; Pagel, K.; Seeberger, P. H.; Evers, C. E.; Boons, G. J.; Pohl, N. L. B.; Compagnon, I.; Widmalm, G.; Flitsch, S. L. Advancing Solutions to the Carbohydrate Sequencing Challenge. *J Am. Chem. Soc.* **2019**, *141*, 14463-14479.
5. Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. Conformational Studies of Oligosaccharides and Glycopeptides: Complementarity of NMR, X-ray Crystallography, and Molecular Modelling. *Chemical Reviews* **2002**, *102*, 371-386.
6. Adrian, M.; Dubochet, J.; Lepault, J.; McDowell, A. W. Cryo-electron microscopy of viruses. *Nature* **1984**, *308*, 32.
7. Wu, X.; Delbianco, M.; Anggara, K.; Michnowicz, T.; Pardo-Vargas, A.; Bharate, P.; Sen, S.; Pristl, M.; Rauschenbach, S.; Schlickum, U.; Abb, S.; Seeberger, P. H.; Kern, K. Imaging single glycans. *Nature* **2020**, *582*, 375-378.
8. Han, L.; Costello, C. E. Mass spectrometry of glycans. *Biochemistry (Mosc)* **2013**, *78*, 710-720.
9. Schenk, J.; Nagy, G.; Pohl, N. L. B.; Leghissa, A.; Smuts, J.; Schug, K. A. Identification and deconvolution of carbohydrates with gas chromatography-vacuum ultraviolet spectroscopy. *Journal of Chromatography A* **2017**, *1513*, 210-221.
10. Tang, Y.; Wei, J.; Costello, C. E.; Lin, C. Characterization of Isomeric Glycans by Reversed Phase Liquid Chromatography-Electronic Excitation Dissociation Tandem Mass Spectrometry. *J. Am. Soc. Mass. Spectr.* **2018**, *29*, 1295-1307.
11. May, J. C.; McLean, J. A. Ion mobility-mass spectrometry: time-dispersive instrumentation. *Anal. Chem.* **2015**, *87*, 1422-1436.

12. Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* **2015**, *526*, 241-244.
13. Kirk, A. T.; Bohnhorst, A.; Raddatz, C. R.; Allers, M.; Zimmermann, S. Ultra-high-resolution ion mobility spectrometry-current instrumentation, limitations, and future developments. *Anal. Bioanal. Chem.* **2019**, *411*, 6229-6246.
14. Norbeck, A. D.; Monroe, M. E.; Adkins, J. N.; Anderson, K. K.; Daly, D. S.; Smith, R. D. The utility of accurate mass and LC elution time information in the analysis of complex proteomes. *J. Am. Soc. Mass. Spectr.* **2005**, *16*, 1239-1249.
15. Riddle, L. A.; Guiochon, G. Influence of mobile phase gradients on the retention and separation of peptides from a cytochrome-c digest by reversed-phase liquid chromatography. *Chromatographia* **2006**, *64*, 121-127.
16. Tarasova, I. A.; Perlova, T. Y.; Pridatchenko, M. L.; Goloborod'ko, A. A.; Levitsky, L. I.; Evreinov, V. V.; Guryca, V.; Masselon, C. D.; Gorshkov, A. V.; Gorshkov, M. V. Inversion of chromatographic elution orders of peptides and its importance for proteomics. *J Anal Chem+* **2012**, *67*, 1014-1025.
17. 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*, 4607-4611.
18. 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.
19. Kopysov, V.; Makarov, A.; Boyarkin, O. V. Identification of Isomeric Ephedrines by Cold Ion UV Spectroscopy: Toward Practical Implementation. *Anal. Chem.* **2017**, *89*, 544-547.
20. Kopysov, V.; Gorshkov, M. V.; Boyarkin, O. V. Identification of Isoforms of Aspartic Acid Residue in Peptides by 2D UV-MS Fingerprinting of Cold Ions. *Analyst* **2018**, *143*, 833-836.
21. USA, US 10,283,336 B2, 2019.

22. Saparbaev, E.; Kopysov, V.; Yamaletdinov, R.; Pereverzev, A.; Boyarkin, O. V. Interplay of H-bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates. *Angew Chem Int Ed Engl* **2019**, *58*, 7346-7350.
23. 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.
24. Solovyeva, E. M.; Kopysov, V. N.; Pereverzev, A. Y.; Lobas, A. A.; Moshkovskii, S. A.; Gorshkov, M. V.; Boyarkin, O. V. Method for Identification of Threonine Isoforms in Peptides by Ultraviolet Photofragmentation of Cold Ions. *Anal Chem* **2019**, *91*, 6709-6715.
- (25) 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.
26. 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*, 76-79.
27. Saparbaev, E.; Aladinskaia, V.; Yamaletdinov, R.; Pereverzev, A. Y.; Boyarkin, O. V. Revealing Single-Bond Anomeric Selectivity in Carbohydrate-Protein Interactions. *The Journal of Physical Chemistry Letters* **2020**, *11*, 3327-3331.
28. Voss, J. M.; Kregel, S. J.; Fischer, K. C.; Garand, E. IR-IR Conformation Specific Spectroscopy of Na⁽⁺⁾(Glucose) Adducts. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 42-50.
29. 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.
30. Agmon, N. Elementary Steps in Excited-State Proton Transfer. *J. Phys. Chem A* **2005**, *109*, 13-35.
31. Schindler, B.; Legentil, L.; Allouche, A. R.; Ferrieres, V.; Compagnon, I. Spectroscopic diagnostic for the ring-size of carbohydrates in the gas

phase: furanose and pyranose forms of GalNAc. *Phys Chem Chem Phys* **2019**, *21*, 12460-12467.

32. Boyarkin, O. V. Cold ion spectroscopy for structural identifications of biomolecules. *International Reviews in Physical Chemistry* **2018**, *37*, 559-606.

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

34. Ma, C.; Sun, Z.; Chen, C.; Zhang, L.; Zhu, S. Simultaneous separation and determination of fructose, sorbitol, glucose and sucrose in fruits by HPLC-ELSD. *Food chemistry* **2014**, *145*, 784-788.

35. Lowary, T. L. Twenty Years of Mycobacterial Glycans: Furanosides and Beyond. *Accounts Chem Res* **2016**, *49*, 1379-1388.

36. Hong, A.; Choi, C. M.; Eun, H. J.; Jeong, C.; Heo, J.; Kim, N. J. Conformation-Specific Circular Dichroism Spectroscopy of Cold, Isolated Chiral Molecules. *Angew. Chem. Int. Ed.* **2014**, *53*, 7805-7808.

37. Daly, S.; Rosu, F.; Gabelica, V. Mass-resolved electronic circular dichroism ion spectroscopy. *Science* **2020**, *368*, 1465-1468.

38. Dwivedi, P.; Wu, C.; Matz, L. M.; Clowers, B. H.; Siems, W. F.; Hill, H. H. Gas-Phase Chiral Separations by Ion Mobility Spectrometry. *Anal. Chem.* **2006**, *78*, 8200-8206.

39. Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; Varki, A.; Woods, R. J.; The, S. D. G. Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **2019**, *29*, 620-624.

40. Schindler, B.; Laloy-Borgna, G.; Barnes, L.; Allouche, A. R.; Bouju, E.; Dugas, V.; Demesmay, C.; Compagnon, I. Online Separation and Identification of Isomers Using Infrared Multiple Photon Dissociation Ion Spectroscopy Coupled to Liquid Chromatography: Application to the Analysis of Disaccharides Regio-Isomers and Monosaccharide Anomers. *Anal Chem* **2018**, *90*, 11741-11745.

41. Ben Faleh, A.; Warnke, S.; Rizzo, T. R. Combining Ultrahigh-Resolution Ion-Mobility Spectrometry with Cryogenic Infrared

Spectroscopy for the Analysis of Glycan Mixtures. *Anal. Chem.* **2019**, *91*, 4876-4882.

42. Mucha, E.; Stuckmann, A.; Marianski, M.; Struwe, W. B.; Meijer, G.; Pagel, K. In-depth structural analysis of glycans in the gas phase. *Chem Sci* **2019**, *10*, 1272-1284.

APPENDIX CHAPTER 6

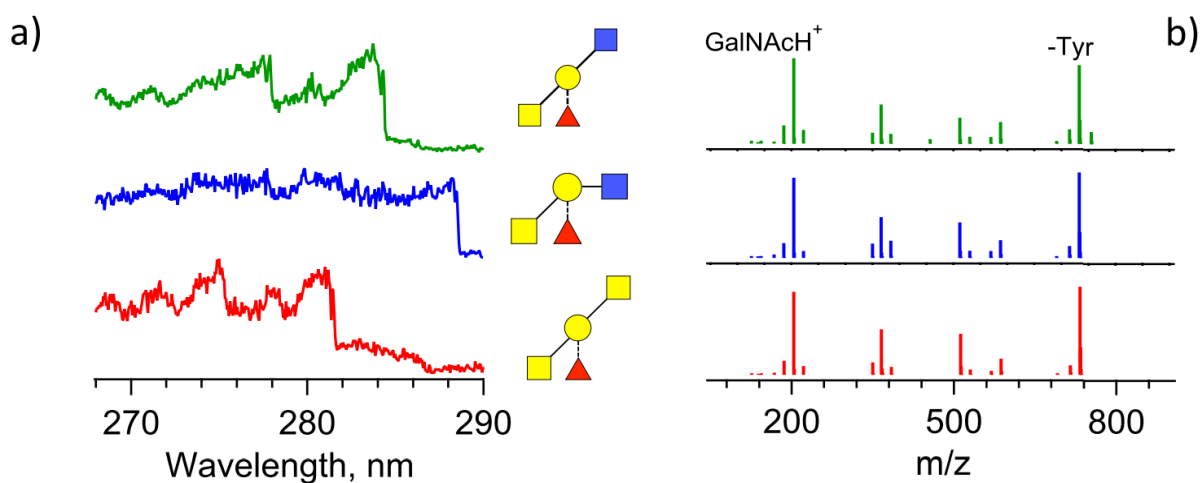


Figure 6.S1. UVPD (a) optical and (b) mass spectra of TyrH⁺ in complexes (from top to bottom) with Blood Type A1, Blood Type A2 and Blood Type A13/4 tetrasaccharides. The spectra are obtained by integrating the respective 2D UV-MS fingerprints over m/z and wavelength dimensions for optical and mass spectra, respectively.

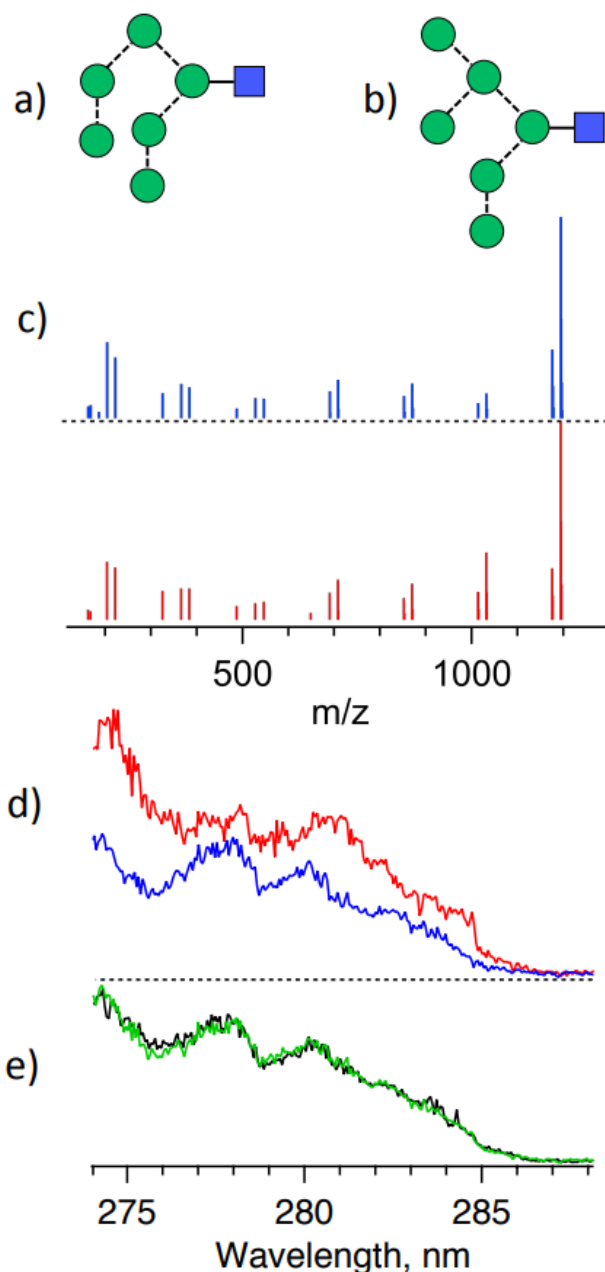


Figure 6.S2. Symbolic structures of (a) Man₆GlcNAc-II and (b) Man₆GlcNAc-I hepta saccharides. (c) UVPD mass spectra (in red and blue, respectively) and (d) UVPD optical spectra (the same colour coding) of the two heptasaccharides in complexes with TyrH⁺; the spectra are derived by integrating the respective measured 2D UV-MS fingerprints. (e) UVPD optical spectrum (green trace) derived from the 2D UV-MS fingerprint measured for 1:2 mixture of the glycans and the best fit of these spectrum (red trace) obtained from matrix decomposition procedure. The calculated relative concentrations, 29% and 71%, differ from the prepared solution ones by 4%.

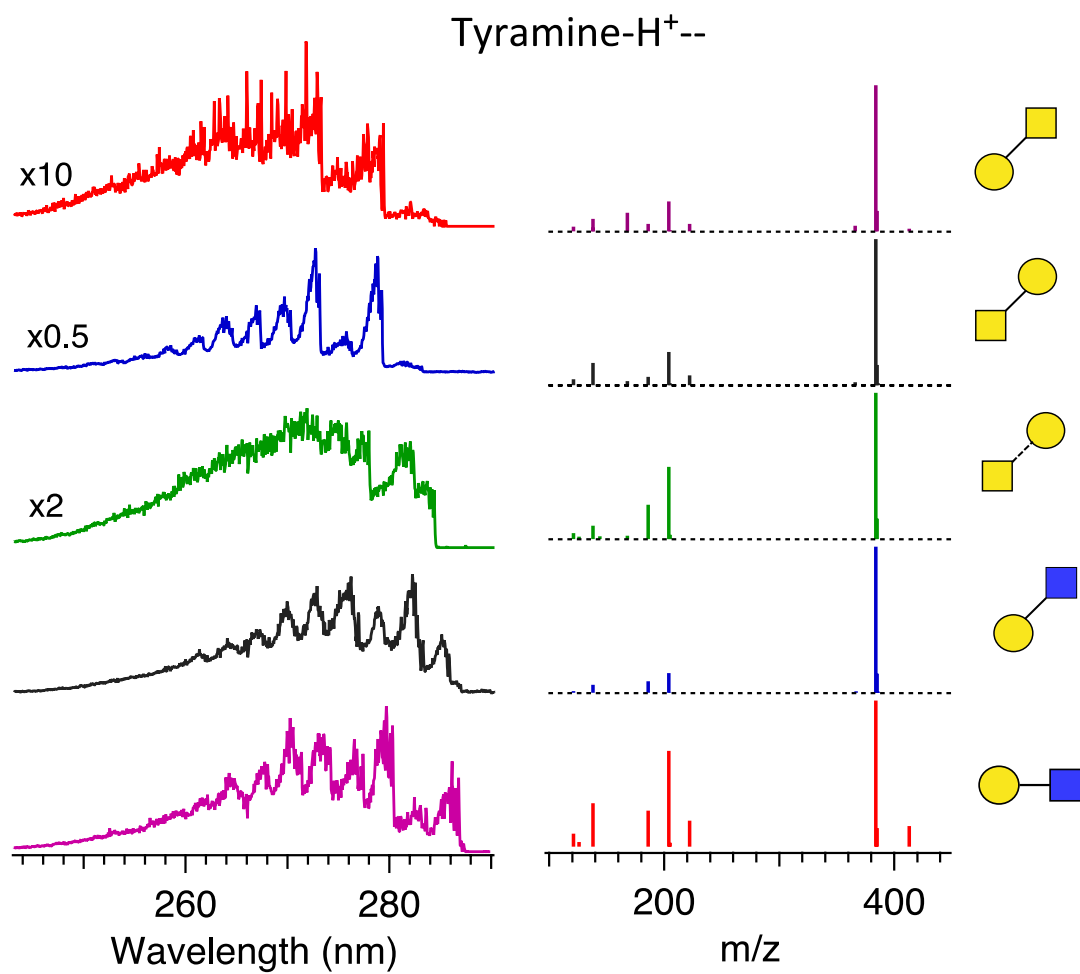


Figure 6.S3. UVPD optical and MS identities of five isomeric disaccharides in non-covalent complexes with TrmH⁺ aromatic molecule.

Calculated relative concentrations

Saccharide Mixture	A	B	C	D	E
B2:C1	0.00	78.30	17.15	4.55	0.00
B1:C1	0.00	58.85	36.65	4.50	0.00
B1:C2	0.00	34.40	60.25	5.35	0.00
A2:C1	63.17	1.93	29.80	5.10	0.00
A1:C1	50.60	1.30	42.10	5.99	0.00
A1:C2	48.24	0.05	49.30	2.41	0.00
D1:E1	0.00	1.22	0.00	51.26	47.52
B1:D1:E1	0.00	33.57	0.12	34.38	31.93
B1:D2:E1	0.00	25.60	0.00	51.25	23.14
B1:D1:E2	0.00	22.18	0.00	30.30	47.52
E	0.00	1.85	0.00	0.00	98.15
D	0.00	2.29	6.19	91.52	0.00

Table 6.S1. Calculated relative concentrations for 12 mixtures of one to three isomeric disaccharides from the 5-component library of individual disaccharides.

The identities of the isomers are encoded by A-E letters, where **A** - β -Gal-1,3-GalNAc, **B** - α -GalNAc-1,3-Gal, **C** - β -GalNAc-1,3-Gal, **D** - β -Gal-1,3-GlcNAc, **E** - β -Gal-1,4-GlcNAc; the relative solution concentrations of the mixed isomers are shown next to the letters. The leftmost column lists the tested mixtures of disaccharides and the other five columns show the relative concentrations of the five isomers calculated from matrix decompositions.

CHAPTER 7.

Summary and Outlook

7.1. SUMMARY

In this work, we demonstrated the use of non-covalent interactions between solvent molecules and biomolecules for mimicking the native environment of the latter in the gas phase. We also employed these relatively weak interactions for interrogating structural information from isomers – the molecules that have absolutely the same mass but different arrangement of the atoms.

Firstly, we investigated structures of microsolvated small molecules GlyH⁺ and GlyGlyGlyH⁺ produced by retaining a controlled number of water molecules onto the generated by ESI hydrated biomolecules. Results of our studies suggest that solvent molecules, retained from the aqueous solution, may preserve the main structural features of the native structures. The same microsolvated complexes can be alternatively produced in the gas phase by cryogenic condensation of water molecules onto bare ions. Comparison of our results with earlier published experimental results brings clear evidence that cryogenic condensation does not unambiguously reflect the solution-like structure of the hydrated biomolecules, because some of the gas phase conformers may remain kinetically trapped after hydration. Although cryogenic condensation is, likely, a more universal method of hydration, its use for elucidation of native structures of biomolecules may lead to false identifications. Our experimental results may also help many groups that work on the investigation of structures of biomolecules in the gas phase by mass spectrometry and/or gas-phase spectroscopy. Harsh conditions of ESI may influence spatial geometries of biomolecules, making the structures far from being native. The presence of the relatively weak complexes in mass spectra of biomolecules guarantees soft conditions in electrospray, increasing the probability of appearance of native structures.

Secondly, thanks to non-covalent interactions, we extended our fast and reproducible 2D UV-MS approach for identification and quantification of the most diverse and, therefore, challenging class of isomeric biomolecules: glycans. This approach cannot be directly applied to this class of biomolecules, because due to the lack of UV chromophore they are “invisible” in UV. The key finding that allows us to make them visible and distinguishable in UV is a noncovalent attachment of aromatic molecules. The differences in structures of analyte molecules can be transferred through intermolecular non-covalent interactions to an aromatic reader-molecule. This aromatic molecule «reads» even small structural alterations in a partnering glycan through non-covalent interactions and responds to the alterations by shifting the UV spectra and/or changing its shape. Several types of non-covalent interactions are involved in the sensing of the analysing molecules, including OH ... OH, CH ... π , OH ... π interactions. These relatively weak bonds are incredibly sensitive towards the structural changes in analysing molecules, which was demonstrated using as an example complexes between two anomers of GalNAc and tyramine. The pattern of hydrogen bonds in these complexes is the same and the only significant difference is the lengths of one remote bond between oxygen of tyramine and hydrogen of glycans. The anomeric OH groups, which orientations are different in two complexes, are not even involved in any type of hydrogen bonds, but spectra of two anomers are significantly shifted.

7.2. OUTLOOK

The demonstrated sensitivity of aromatic reporter molecules to the structural changes in analyzing carbohydrate molecules together with an almost infinite choice of reporter molecules makes us believe that this approach can be easily extended for any other classes of biomolecules, where the problem of identification of isomers exists. Using this approach, we already showed that any classes of isomeric glycans can be identified and relative concentrations in their mixtures can be determined.

A logical continuation of this project would be demonstration of the applicability of this approach for the identification and quantification of isomeric lipids, which are, differently from glycans, do not contain many polar functional groups to form noncovalent bonds with aromatic rings. Our preliminary results (to be published) suggest that instead of strong H-bonds (e.g., OH...OH) such isomeric lipids can be recognised by much weaker CH- π interactions with aromatic rings of, for instance, small peptides that include phenylalanine residues. It implies that isomeric lipids can be identified and quantified using the same hardware and software as can be used for the identification of isomeric peptides, glycans, and small drugs.

The performance of this approach can be further explored for identifications of other classes of biomolecules. Future efforts should target the development of the methods for *de novo* sequencing of glycans, which requires a library of 2D UV-MS spectra of small carbohydrates. This library can be used for sequencing large glycans with preceding enzymatic digestion or hydrolysis. Over the last decade, many groups have already demonstrated the applicability of IR and UV cold ion spectroscopy for the identifications of different isomeric biomolecules, including carbohydrates, lipids, peptides, metabolites, and drugs, but the application of these methods for solving practical problems, in reality, is yet to be shown. We believe that universality, rapidity, and robustness of the cold ion spectroscopy-based methods can be employed for industrial needs, as many companies are interested in accelerating speed and in improving the accuracy of standard analyses. Over the last few years our group has made several significant steps towards practical applications of our method: (1) we significantly decreased the time of the analysis by reducing the number of required experimental points (2) we also demonstrated the performance of a 1D UVPD approach, which drastically lowers the cost of the required instrumentation.

CURRICULUM VITAE

EDUCATION

Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland 08/2017 – 09/2021
PhD candidate in Chemical Engineering
Novosibirsk State University, Russia 09/2015 – 06/2017
MSc in Chemistry
Novosibirsk State University, Russia 09/2011 – 06/2015
BSc in Chemistry

CORE EXPERIENCE

Laboratory of Molecular Physical Chemistry, EPFL 08/2017 – present
Structural and Analytical Identification of Biomolecules by Cold Ion Spectroscopy of Non-Covalent Complexes
Laboratory of Catalytic Processes in Fuel Cells, BIC SB RAS 09/2015 – 06/2017
Bimetallic Nanoalloy Catalysts for Selective Oxidation of Carbon Monoxide in the Presence of Hydrogen as a Probe Reaction and for Sensors Applications
Laboratory of Cluster and Supramolecular Compounds, NIIC SB RAS 02/2012 – 08/2015
Synthesis, Structure and Properties of New Metal-Organic Frameworks Based on Heteroaromatic Dicarboxylic Acids

ADDITIONAL EXPERIENCE

INTERNSHIPS:

Laboratory of the Mechanisms of Catalytic Reactions, BIC SB RAS 11/2015 – 12/2015
Catalyst systems based on ruthenium complexes for stereoselective oxidation of olefins and sulfides
Laboratory of Cluster and Supramolecular Compounds, NIIC SB RAS 04/2014 – 05/2014
Determination of the diffusion coefficient of nitrogen into the metal-organic coordination polymer $[\text{Zn}_4(\text{dmf})(\text{ur})_2(\text{ndc})_4]$
Department of Heterogeneous Catalysis, BIC SB RAS 11/2013 – 12/2013
Butanal production by reaction of ethylene oxide with ethylene on the Zeolite FeZSM-5 surface
Laboratory of Multispin Coordination Compounds, ITC SB RAS 04/2013 – 05/2013
Synthesis of 3,3,4-trimethyl-1-(1-methyl-1H-pyrazol-4-yl)-4-nitropentane-1-one by Stille reaction

TEACHING:

Swiss Federal Institute of Technology in Lausanne (EPFL) 02/2018 – 05/2021
Organic Chemistry, General Physics: Electromagnetism, Lasers and Applications in Chemistry
Novosibirsk State University (NSU) 09/2015 – 06/2017
General and Inorganic Chemistry
The Specialized Educational Scientific Center of NSU 09/2012 – 05/2017
Advanced Inorganic Chemistry, General Chemistry, Physical Chemistry

LANGUAGES

English (Fluent C1), German (Upper-Intermediate B2), French (Elementary A2), Russian (Native)

PUBLICATIONS

1. **E. Saparbaev**, V. Aladinskaia, A. Zviagin, and O. Boyarkin. Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy. *Journal of Physical Chemistry Letters* 2021, 12, 10644-10648
2. A. Lobas, Anna, E. Solovyeva, **E. Saparbaev**, M. Gorshkov and O. Boyarkin. Accelerating Photofragmentation UV Spectroscopy - Mass Spectrometry Fingerprinting for Quantification of Isomeric Peptides. *Talanta* 2021, 232, 122412
3. **E. Saparbaev**, V. Kopysov, V. Aladinskaia, V. Ferriere, L. Legentil, O. Boyarkin. Identification and Quantification of Any Isoforms of Carbohydrates by 2D UV-MS Fingerprinting of Cold Ions. *Analytical Chemistry* 2020, 92, 21, 14624–14632
4. **E. Saparbaev**, V. Aladinskaia, R. Yamaletdinov, A. Pereverzev, and O. Boyarkin. Revealing Single-Bond Anomeric Selectivity in Carbohydrate–Protein Interactions. *Journal of Physical Chemistry Letters* 2020, 11, 3327-3331
5. A. Pereverzev, Z. Koczor-Benda, **E. Saparbaev**, V. Kopysov, E. Rosta, and O. Boyarkin. Spectroscopic Evidence for Peptide-Bond-Selective Ultraviolet Photodissociation. *Journal of Physical Chemistry Letters* 2020, 11, 206–209
6. **E. Saparbaev**, V. Kopysov, R. Yamaletdinov, A. Pereverzev, O. V. Boyarkin. Interplay of H-Bonds with Aromatics in Isolated Complexes Identifies Isomeric Carbohydrates. *Angew. Chem. Int. Ed.* 2019, 58, 7346-7350
7. D. Potemkin, **E. Saparbaev**, A. Zadesenets, E. Filatov, P. Snytnikov, V. Sobyenin. Preferential CO Oxidation on Bimetallic Pt_{0.5}M_{0.5} Catalysts (M = Fe, Co, Ni) Prepared from Double Complex Salts. *Catalysis in Industry* 2018, 10, 62-67
8. **E. Saparbaev**, S. Sapchenko, D. Samsonenko, D. Dybtsev, V. Fedin. Synthesis, structure, and luminescent properties of layered coordination polymer based on cadmium(II) 2,5-furandicarboxylate. *Russian chemical bulletin* 2015, 3, 0613-0617.
9. S. Sapchenko, **E. Saparbaev**, D. Samsonenko, D. Dybtsev, V. Fedin. Synthesis, structure, and properties of a new layered coordination polymer based on zinc(II) carboxylate. *Russian Journal of Coordination Chemistry* 2013, 39, 8, 549–552