Revealing Single-bond Anomeric Selectivity in Carbohydrate– Protein Interactions.

Cite This: J.Phys.Chem.Lett.2020, 11, 3327–3331

<https://pubs.acs.org/doi/10.1021/acs.jpcllett.0c00871>

Erik Saparbaev,^a Viktoriia Aladinskaia,^a Ruslan Yamaletdinov,^b Aleksandr Y. Pereverzev,^a Oleg V. Boyarkin^{*a}

^a Laboratoire de Chimie Physique Moléculaire, École Polytechnique Fédérale de Lausanne, Station-6, 1015 Lausanne, Switzerland.

^b Nikolaev Institute of Inorganic Chemistry, Novosibirsk, 630090, Russian Federation

ABSTRACT

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.

TOC GRAPHICS

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, Glu, 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.

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 SI 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 1a), as well as its methylated derivatives (Fig. 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 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 2b and 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 2a) 1-2, 5-6 and 9-10. Except peak 3, every peak of the spectrum in figure 2a is mutually reproduced (within $\pm 0.8 \text{ cm}^{-1}$) in the spectra 2b and/or 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 2d and 2e) to the OH stretches of Trm in the α and β anomeric GalNAc-TrmH⁺ (peak 4 and shoulder 4* in Fig. 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.

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. 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 2f and 2g show the anomer-selective IR-UV depletion spectra of GalNAc-TrmH⁺. Every peak in the gain spectrum 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 3a, 3b). All the peak assignments experimentally derived above are reproduced by the computations (Figs. 2f and 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.

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 NH₃⁺ 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 aromatics 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 aromatics 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 α - and 1.955 Å in the 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 aromatics clearly correlate with the H-bond length and increase steeply upon shortening of the bond (Fig. 3b).²⁵⁻³² Our data are consistent with this correlation. They fall onto the dependence in Figure 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. 2f and 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 $\cdots$ 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.

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.

ASSOCIATED CONTENT

Supporting Information. Computational and experimental details; additional UV spectra.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We thank Prof. A. Zehnacker-Rentien for useful discussions, Novartis AG (FreeNovation program; Dr. Hans Widmer) and Swiss National Science Foundation (grant 200020_172522) for financial support of this work.

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.; Uchamaru, 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.

Figure 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.

Figure 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. 3a and 3b).

Figure 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 S1) and measured here (red dots).

Supporting Information for

**Revealing Single-bond Anomeric Selectivity in
Carbohydrate–Protein Interactions.**

Erik Saparbaev, Viktoriia Aladinskaia, Ruslan Yamaletdinov, Aleksandr Y. Pereverzev, Oleg
V. Boyarkin*
Correspondence to: oleg.boiarkin@epfl.ch

Table of Contents

Materials and Experimental method	S2
Computational details	S3
Figure S1. UV spectra of methylated modifications of α/β -GalNAc–TrmH ⁺ complex	S4
Table S1. Details of the data used in Figure 3b	S5
Additional References.....	S5

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 $6\text{-}7\text{ cm}^{-1}$. 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.¹⁷

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 host 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.

b) 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,³⁴ 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 carbohydrates³⁵ and for proteins³⁶ 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 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 1. 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. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781.
- [35] O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor and A. D. Mackerell, *Journal of Computational Chemistry*, 2008, **29**, 2543.
- [36] A. D. MacKerell, M. Feig and C. L. Brooks, *J. Am. Chem. Soc.*, 2004, **126**, 698.