eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

FACULTY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF THE BASQUE COUNTRY

Rotational spectroscopy as a Chemical tool for the study of sugars and molecular interactions

Ph.D. Thesis

Aran Insausti Beiro

Supervised by:

Dr. Emilio J. Cocinero Pérez
Dra. Camilla Calabrese

Department of Physical Chemistry
Leioa, July 2022

Aran Insausti obtained a predoctoral fellowship from the Basque Government. Thanks also to the SGIKer (UPV/EHU) Laser and Computational resources, Biofisika Insitute (EHU-CSIC) and Centro de Supercomputación de Galicia (CESGA). and Compute Canada Computation Services.

I want to thank my thesis directors Emilio J. Cocinero Pérez and Camilla Calabrese for giving me the opportunity to do this work. I also want to thank the members (past and present) of the Spectroscopy Group at the University of the Basque Country (UPV/EHU).

An immense thank to my friend and workmate Elena R. Alonso for all his help and time spent together in the laboratory. Really, a great deal of work presented here has been carry out with her.

I also want to thank, Jose Luis Alonso, Santiago Mata and Cristobal Perez Cuadrado whose knowlegment to assemble the new arrangement of chirp instrument has been essential.

I would like to thank all the people of the Yunjie Xu and Wolfgang Jäeger group at University of Alberta. Yunjie, thanks to give me the opportunity to join your group. Yunjie, Wolfgang, Jiarui, Arsh, Mo, Gojie, Alex, Fan, Qian and all other members, I cannot stress enough my gratitude to all of you.

I want to thank Centro de SuperComputacion de Galicia (CESGA) for their service, especially to Pablo Rey for his help with all problems with calculations. I also want to thank also Scientific Computation Service of SGIKER.

Me gustaría también agradecer a Ana Rosa Plaza Urbina, del departamento de Educación de Gobierno Vasco por toda su buena voluntad y ayuda con las gestiones burocráticas.

Azkenik ezkerrak Iñaki Porres irakasleari orain dela bederatzi urte kimika ikastera animatzeagatik.

Familia, amigos y a los más cercanos.

Eskerrik asko!!

Index

The appendices on this Thesis are available in electronic format. The reader can access it online as a pdf file in the following website link:

https://drive.google.com/drive/folders/1eCMEAxz0oucAeHFGeUdkCoJbOy95utdT?usp=sharing

Acronyms and symbols

Here we collect some of the acronyms and symbols using in this Thesis.

Acronym	Meaning
cav-FTMW	Cavity Fourier Transform Microwave (refers to a Flygare-Balle's spectrometer)
CP-FTMW	Chirped Pulse Fourier Transform Microwave
FID	Free Induction Decay
DFT	Density Functional Theory
PES	Potential energy surface
QC	Quantum chemistry
MM	Molecular mechanics
H-bond	Hydrogen bond
NCI	Non covalent interactions
ISM	Interstellar medium

Chapter I.

Introduction

I.1 Introduction

Nowadays, structure determination of organic and inorganic compounds is part of the scientific background of chemists, physicist, as well as of biochemists, engineering and pharmacists among other disciplines. Concretely, the structure determination can be defined as the *procedure by which the three-dimensional atomic coordinates of a molecule or biomolecule are solved.*

According to the atomic disposition and electronic distribution of the systems, the molecule or particles that are bound can only take on certain discrete values of energy called energy levels. The information encoded in these energy levels may be spectroscopically decoded to obtain information about the structure and properties of chemical and biological systems. The structural determination of the molecules or bonded molecular structures, has several goals. First, the structure of the molecules is related to their physical,[1–3] chemical[4,5] and biological properties,[6,7] so the knowledge of their conformational behaviour is essential to understand the properties of each molecular system. Second, the experimental molecular structure obtained with a very high spatial resolution can be used as benchmark of the theoretical chemistry.[8–10] Although great progresses in theoretical chemistry have been done during the last years, the computational results do not present enough accuracy to reproduce all the experimental results. Thus, the theoretical results have to be always corroborate experimentally.[11,12] Third, the atomic disposition of the molecule and the energy levels abovementioned, are intrinsic for each molecule and represent their own fingerprint. Not just that, the spectroscopy values are essential to identify unequivocally the molecules in the unknown sample or material, as well as in the interstellar medium (using the radio-telescopes data).

On this point, there are many different experimental techniques that allow us to know the molecular structures with high resolution. Leaving aside widely knowledge classical structural determination procedures (based on nuclear magnetic resonance, X-ray spectroscopy or infrared spectroscopy) nowadays, there is a new variety of techniques to determine the 3D atomic shape of molecules with an high spatial resolution. Maybe, the most current spectroscopic techniques developed to determine the molecular conformation are atomic force microscopy and cryo-electron microscopy.[13–15] They are very extended techniques and they present good results for several molecules. However, the spatial resolution of most of these methods (> 1.5 Å) does not allow showing the finest molecular details.[13,16] Furthermore, these experiments are carried out in condensed environment, where the intrinsic molecular conformations are perturbed by the intermolecular interactions. There are also other non spectroscopic techniques able to detect the molecular conformations, for example some techniques based on mass spectrometry and others based on coulomb

explosion fragments analysis.[17–19] However, these methods do not have a high enough spatial resolution to determine accurate atomic positions.

As aforementioned, many spectroscopic techniques are available, but this Thesis is focused on rotational or microwave spectroscopy in a free-jet environment. Rotational spectroscopy is a technique that catches changes in rotational energy within a molecule (or small molecular aggregates) to obtain the physical constants that are related to the concrete atomic distribution in the molecule. The correct treat of rotational data can be used to determine the molecular structure in gas phase with unrivaled resolution. This is the principal feature that outstand rotational spectroscopy compared with the other spectroscopy techniques, and that allows the unequivocal assignment of conformers,[20] tautomers[21,22] and isotopomers.[23,24] Further, new remarkable rotational spectroscopy experiments have demonstrated its ability to distinguish chiral molecules.[25,26] The rotational spectra associated with different conformational geometries can be used to measure the energy-resolved unimolecular isomerization rate in free environment.[27] Microwave spectroscopy provides one of the highest spectral resolution in the molecular spectroscopy techniques. As a result, the technique has no trouble resolving the spectra from molecules and either for all cluster geometries present in the sample.

Although the rotational spectroscopy begins with the rotational spectrum of ammonia in 1934,[28] the study of molecules started to become a viable technique after the II World War. The first spectrometers were based on the absorption signals of the molecules, they used the continuum microwave source and the modulation of the spectra were done using Stark-modulation or Zeeman-modulation. They studied gas samples introduced in airtight cells, commonly at room temperature. This means that the molecular population was distributed over many rotational and vibrational states, which provided less intense and more complex spectra. This problem was overcome when the supersonic jet spectrometers were developed, in which the molecules were jet-cooled to a rotational temperature of few Kelvin.[29] This environment changes the Boltzmann distribution of the molecules energy levels to the centimeter wavelength of the spectra.

The next large advance for microwave spectroscopy came when the first Fourier transform microwave (FTMW) spectrometer was developed. The FTMW spectrometers were developed thanks to the advances in the speed of microwave switches that allow the polarisation of the molecules with a short, fast and pulsed microwave radiation. Flygare's group designed the spectrometer combining the Fabry-Pérot cavity resonator with the supersonic jet expansion.[30,31] After that, several different spectrometers have been developed.[32–34] The latest breakthrough in the rotational spectrometers was done by Brooks H. Pate's group in 2008, when they published the novel rotational spectrometer set-up based on

chirped pulses.[35] The reported instrument excites molecules across the bandwidth of the *chirped* pulse (8 GHz or more). This experiment allows acquiring a spectral range of several GHz in each experiment, and the good signal-to-noise ratio is obtained accumulating the signal of thousand or million free induction decay (FID). The operation is similar to that of Nuclear Magnetic Resonance experiments.

One of the unusual features of this Thesis is that the experimental apparatuses are not commercially available, so we have spent long time maintaining, building, optimising and developing the spectrometers. Concretely, we used the most common spectrometers to work in the supersonic jet expansion: Fabry-Pérot cavity and *chirped* pulse spectrometer available at the University of the Basque Country (all details in the Chapter II).[36–38]

This thesis consists of two conceptually connected parts. The first part is devoted to the characterization of monosaccharides with rotational spectroscopy. Concretely, in this part we used Fabry-Pérot Fourier transform rotational spectrometer and due to the thermolability of the sugars, all these molecules were vaporised using laser vaporization technique.

In relation with the first part (Chapters III-VI), monosaccharides are the smaller building block of glycobiology that aims to identify the structure and function of the complete set of glycans.[39] Their molecular diversity encodes structural information of many physiological processes, such as production of energy, immune response and cell recognition.[39] Furthermore, the conformational panorama of carbohydrates is directly related to their biological activity. For this reason, understanding the rules that governs the structural behaviour of carbohydrates can help to know the language that the biological molecules and cells use to communicate between them. Learn this language is fundamental for the rational design of pharmacologic agents based on carbohydrates. Additionally, the carbohydrates are the essential building blocks of life, so their chemistry is closely related to the origins of life. For this reason, one of the principal issues in the astrochemistry consists on study the carbohydrate related molecules in the interstellar medium.[40,41]

Regarding to their chemical structure, the monosaccharides are polyhydroxy aldehydes or ketones, with two or more hydroxyl groups and also a carbonyl group either at the terminal carbon atom (aldose) or at the second carbon atom (ketose). Regularly, in solid, liquid and gas phase, carbonyl group combines with one hydroxyl group to form a cyclic compound. The monosaccharides are classified according to the number of carbon atoms they contain: triose (C = 3), tetrose (C = 4), pentose (C = 5), hexose (C = 6), etc. In addition, monosaccharides could suffer different modifications, such as acetylation, fluorination, methylation, oxidation, lactonization and sulfonation among others. These modifications could drastically change the properties of each sugar. For these reasons, it is evident

that there are a large number of sugars and their complexity and variability are very high.

The sugars' studies in gas phase were initialised by Professor John P. Simons at the University of Oxford in the beginning of the 21st century using ion depletion laser spectroscopy techniques.[42] The ability to study molecules in supersonic jet in combination with the high resolution that microwave spectrometers provide, makes rotational spectroscopy an attractive alternative to laser spectroscopy.[43,44] After the detection of the first monosaccharides, other few biologically relevant sugars have recently been studied by high resolution rotational microwave spectroscopy.[45–48] Gas-phase rotational spectroscopy not only has the potential to determine accurate structures of complex molecular systems under controlled conditions, also give information about the population of each conformer,[49] atomic dynamics[50] and subtler structural parameters.[51] Thus, in the first part of this Thesis, the structure, population and properties of several monosaccharides will be presented, and this information will be use for different goals.

In the second part (Chapters VII-VIII) of this Thesis, several studies carried out with the *chirped* pulse spectrometer will be presented. In particular, studies performed at the University of the Basque Country and at the University of Alberta will be presented. For what concerns the University of the Basque Country, the chirped pulsed spectrometer was updated and modified, and several studies are still in progress, despite this the results obtained experimentally will be reported. Concretely, we will focus on water micro-solvation of methyl benzoate and 4-oxobutanonitrile astrochemical molecule whose studies are not concluded. Moreover, studies on proton transfer reactions performed during the abroad stay at the University of Alberta will be described.

I.2 Aims and objectives of this Thesis

The spectroscopy Group at the University of the Basque Country in the Microwave Region has large experience in the determination of the structures of organic molecules in gas phase. Since its formation in 2009, the group has been pioneer in the study of several biological molecules like sugars. The group has built two spectrometers and during this thesis we have updated and optimised them with new electronic components and configurations. This thesis has two general goals. First, continue with the studies of the biological molecules commonly studied in this group. Second, continue to develop and build spectrometers for future works.

So, the objectives of this thesis are:

- → Conformational elucidation of several biological relevant molecules using the combination of rotational spectroscopy and quantum chemistry calculations. The results will be used for different purposes: know the factors that govern the conformation of the molecules, astronomical searches and as a benchmark for computational chemistry, among others.
- → Move and updated the previously built *chirped* pulse spectrometer to the new laboratory. Extend the spectral range in which the spectrometer initially works (from 6 to 18 GHz) to the lower frequency range (from 2 to 6 GHz) and modify the electronic configuration to measure more efficiently in the 7 to 14 GHz range.

In **Chapter II**, the basic concepts about rotational spectroscopy will be reported, in order to understand the studies presented in this dissertation. Here, the experimental setups will be described in detail together with the new *chirped pulsed* experiment installed during this Thesis. Further, other basic concepts (like computational chemistry calculations) necessary to carry out the studies will be slightly explained.

Chapter III focuses on a erythrulose sugar. In this chapter, we combine the laboratory results obtained by rotational spectroscopy with the results of the radio telescopes, in order to try to detect the molecule in the Inter Stellar Medium (ISM). The rotational constants obtained in our laboratory were used as fingerprints of the molecule radio-telescopes research. Here, we combine a new method to prepare solid sample, the rotational spectroscopy data, astronomical searches and high accurate molecular structure determination to describe the erythrulose molecule as never done before.

In **Chapter IV,** we present a project about fluoro/deoxy derivative aldohexoses. The importance of this mutation in the molecules are extremely high because the biological activity of these sugars drastically differs from their analogues. In this chapter, we pay attention to the conformation and the population of the hydroxymethyl groups in both liquid and gas phase. The study

was carry out through the collaboration between different research groups in order to combine synthesis, Nuclear Magnetic Resonance studies, Quantum Chemistry calculations, Molecular Dynamics and rotational spectroscopy data, to understand the arrangement of the cooperative H-bonds and hydroxymethyl groups in several glucose, mannose and galactose derivatives.

In **Chapter V**, the case of gluconolactone and glucoronolactone, the most important glucose derivative lactones, will be presented. Here, we take advance of rotational spectroscopy to see the difference between these two glucose derivatives. The high resolution of rotational spectroscopy allows the determination of the ring configuration and the ring puckering of these molecules in gas phase, and the study shows the large difference between the glucose and its lactone derivatives.

In **Chapter VI**, we present the study of the acid derivatives of the most abundant monosaccharides: glucose and galactose. We combine the solid state Nuclear Magnetic Resonance, recrystallizations, rotational spectroscopy and quantum chemical calculations to decode the H-bond network in both anomers (alpha and beta) of the glucuronic acid and galacturonic acid. The extremely importance of the anomeric hydroxyl group in the H-bond network will be showed.

In the **Chapter VII,** the results using the *chirp* pulsed spectrometer will be introduced. This chapter reflects the work done at the University of Alberta. Here, the conformational panorama of the monomer of furoic acid and the proton transfer reaction in the furoic acid dimers are presented. The proton transfer reaction is complex and subtly includes processes that have great interest in chemistry and in biology. Here, we combine the rotational spectroscopy data and new computational chemistry calculations to describe the proton transfer reaction.

In **Chapter VIII,** other in progress studies, generally obtained with the *chirped* pulse experiment, will be summarised. The long time spent developing the experimental setups meant that these studies are still ongoing. In particular, the micro-solvation of methyl benzoate will be presented and also the rotational spectrum of 4-oxobutanonitrile molecule, for which a future astronomical search in the ISM will be performed. These researches will be published during the next few years.

Finally, the **Chapter IX** summarizes the general conclusions of this Thesis and the future outlooks of the presented works.

I.3 Bibliography

[1] S. G. Davey, *Nat. Rev. Chem.* **2020**, *4*, 507.
[2] C. Wang, C. C. Sun, *CrystEngComm* **2020**, *22*, 1149–1153.
[3] M. Schleeger, C. C. Vandenakker, T. Deckert-Gaudig, V. Deckert, K. P. Velikov, G. Koenderink, M. Bonn, *Polymer (Guildf).* **2013**, *54*, 2473–2488.
[4] A. Zunger, *Nat. Rev. Chem.* **2018**, *2*, 1–16.
[5] P. G. Seybold, M. May, U. A. Bagal, **1987**, *64*, 575–581.
[6] S. Lüdtke, P. Neumann, K. M. Erixon, F. Leeper, R. Kluger, R. Ficner, K. Tittmann, *Nat. Chem.* **2013**, *5*, 762–767.
[7] D. D. Boehr, P. E. Wright, *Science (80-.).* **2008**, *320*, 1429–1430.
[8] K. L. Bak, J. Gauss, P. Jørgensen, J. Olsen, T. Helgaker, J. F. Stanton, *J. Chem. Phys.* **2001**, *114*, 6548–6556.
[9] K. L. K. Lee, M. McCarthy, *J. Phys. Chem. A* **2020**, *124*, 898–910.
[10] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.
[11] I. Uriarte, A. Insausti, E. J. Cocinero, A. Jabri, I. Kleiner, H. Mouhib, I. Alkorta, *J. Phys. Chem. Lett.* **2018**, *9*, 5906–5914.
[12] E. Gloaguen, B. De Courcy, J. P. Piquemal, J. Pilmé, O. Parisel, R. Pollet, H. S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier, M. Mons, *J. Am. Chem. Soc.* **2010**, *132*, 11860–11863.
[13] G. R. Heath, E. Kots, J. L. Robertson, S. Lansky, G. Khelashvili, H. Weinstein, S. Scheuring, *Nature* **2021**, *594*, 385–390.
[14] E. DM, S. EK, *Nature* **1990**, *344*, 524.
[15] M. Herzik, M. A. H. Jr, *Encycl. Ref. Genomics Proteomics Mol. Med.* **2006**, 361–361.
[16] T. Kato, F. Makino, T. Nakane, N. Terahara, T. Kaneko, Y. Shimizu, S. Motoki, I. Ishikawa, K. Yonekura, K. Namba, *Microsc. Microanal.* **2019**, *25*, 998–999.
[17] H. Wu, R. Zhang, W. Zhang, J. Hong, Y. Xiang, W. Xu, *Chem. Sci.* **2020**, *11*, 4758–4765.
[18] T. M. Schwarz, C. A. Dietrich, J. Ott, E. M. Weikum, R. Lawitzki, H. Solodenko, E. Hadjixenophontos, B. Gault, J. Kästner, G. Schmitz, P. Stender, *Sci. Rep.* **2021**, *11*, 1–19.
[19] W. Zhou, L. Ge, G. A. Cooper, S. W. Crane, M. H. Evans, M. N. R. Ashfold, C. Vallance, *J. Chem. Phys.* **2020**, *153*, 184201.
[20] S. R. Domingos, C. Pérez, C. Medcraft, P. Pinacho, M. Schnell, *Phys. Chem. Chem. Phys.* **2016**, *18*, 16682–16689.
[21] J. L. Alonso, V. Vaquero, I. Peña, J. C. Lõpez, S. Mata, W. Caminati, *Angew. Chemie - Int. Ed.* **2013**, *52*, 2331–2334.
[22] L. B. Favero, I. Uriarte, L. Spada, P. Écija, C. Calabrese, W. Caminati, E. J. Cocinero, *J. Phys. Chem. Lett.* **2016**, *7*, 1187–1191.
[23] Z. Kisiel, B. A. Pietrewicz, P. W. Fowler, A. C. Legon, E. Steiner, *J. Phys. Chem. A* **2000**, *104*, 6970–6978.
[24] G. Klapper, L. Surin, F. Lewen, H. S. P. Muller, I. Pak, G. Winnewisser, *Astrophys. J.* **2003**, *582*, 262–268.
[25] Z. Su, N. Borho, Y. Xu, *J. Am. Chem. Soc.* **2006**, *128*, 17126–17131.
[26] D. Patterson, M. Schnell, J. M. Doyle, *Nature* **2013**, *497*, 475–477.
[27] Z. Kisiel, O. Dorosh, A. Maeda, I. R. Medvedev, F. C. De Lucia, E. Herbst, B. J. Drouin, J. C. Pearson, S. T. Shipman, *Phys. Chem. Chem. Phys.* **2010**, *12*, 8329–8339.
[28] C. E. Cleeton, N. H. Williams, *Phys. Rev.* **1934**, *45*, 234–237.
[29] K. G. Garrison, W. Gordy, **1957**, 899–901.
[30] T. J. Balle, E. J. Campbell, M. R. Keenan, W. H. Flygare, *J. Chem. Phys.* **1980**, *72*, 922–932.
[31] T. J. Balle, W. H. Flygare, *Rev. Sci. Instrum.* **1981**, *52*, 33–45.

[32] W. Stahl, G. Bestmann, H. Dreizler, U. Andresen, R. Schwarz, *Rev. Sci. Instrum.* **1985**, *56*, 1759–1762.
[33] J. U. Grabow, W. Stahl, H. Dreizler, *Rev. Sci. Instrum.* **1996**, *67*, 4072–4084.
[34] U. Andresen, H. Dreizler, J. U. Grabow, W. Stahl, *Rev. Sci. Instrum.* **1990**, *61*, 3694–3699.
[35] G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman, B. H. Pate, *Rev. Sci. Instrum.* **2008**, *79*, 053103.
[36] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[37] I. Uriarte, C. Pérez, E. Caballero-Mancebo, F. J. Basterretxea, A. Lesarri, J. A. Fernández, E. J. Cocinero, *Chem. - A Eur. J.* **2017**, *23*, 7238–7244.
[38] E. J. Cocinero, A. Lesarri, P. Ecija, F. J. Basterretxea, J. U. Grabow, J. A. Fernandez, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[39] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, *Essentials of Glycobiology, Third Edition*, **2017**.
[40] G. Cooper, N. Kimmich, W. Belisle, J. Sarinana, K. Brabham, L. Garrel, *Nature* **2001**, *414*, 879–883.
[41] M. Nuevo, G. Cooper, S. A. Sandford, *Nat. Commun.* **2018**, *9*, 1–10.
[42] F. O. Talbot, J. P. Simons, *Phys. Chem. Chem. Phys.* **2002**, *4*, 3562–3565.
[43] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[44] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.
[45] E. R. Alonso, I. Peña, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2016**, *7*, 845–850.
[46] I. Peña, L. Kolesniková, C. Cabezas, C. Bermúdez, M. Berdakin, A. Simão, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2014**, *16*, 23244–23250.
[47] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.
[48] I. Peña, C. Cabezas, J. L. Alonso, *Chem. Commun.* **2015**, *51*, 10115–10118.
[49] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[50] G. Feng, L. B. Favero, A. Maris, A. Vigorito, W. Caminati, R. Meyer, *J. Am. Chem. Soc.* **2012**, *134*, 19281–19286.
[51] E. R. Alonso, I. Peña, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2016**, *7*, 845–850.

Chapter II

Scientific Method

This chapter has been described in such a way that any reader, regardless of his or her field of knowledge, can understand the general concepts of our studies. The following is a brief description of the methodology used.

As in any research, the first fundamental step concerns the topic to be studied or, at least, knowing the question(s) to be asked in our research ("*Ask a question*" in Figure II.1). There are different ways of approaching this. In general, we use background knowledge and information of other researches (available in the specialise bibliography) to find the missing pieces of knowledge and the specific issue to be explored. This step is crucial, because all subsequent steps will focus on answering the initial question. Another key factor to be considered is the feasibility of the project. This point is very essential because it is related to the feasibility of the project.

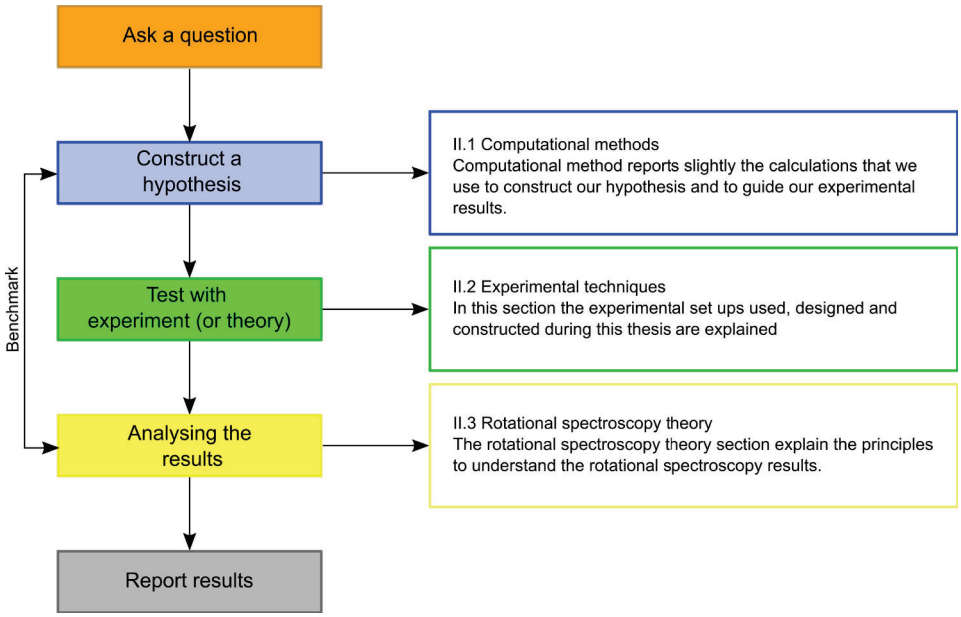


Figure II.1 Block diagram of the methodology used in this doctoral Thesis.

Once the research topic has been identified, the next step is to formulate a (or several) hypothesis that can solve the initial question(s). In our research, the hypothesis are commonly supported and reinforced by several computational chemistry calculations methods ("*Construct the hypothesis*" in Figure II.1). For this reason, a specific section on **Computational methods** is described below, in order to explain the general concepts of the computational calculations that were used in this Thesis.

Once the question is established and the different hypothetical answers have been outlined, it is time to prove which of the hypotheses may be the correct ("*Test with experiment (or theory)*" in Figure II.1). Throughout this Thesis, different experimental set-ups have been used, built and developed. The **Experimental techniques** section within this chapter is dedicated to describe the instruments used and developed during the Thesis.

Experimental results require a carefully validation based on scientific certainties ("*Analysing the results*" in Figure II.1). In our case, the experimental results are obtained by rotational spectroscopy and supported with high-level computational chemistry calculations. The general aspects of rotational spectroscopy are explained in the **Rotational spectroscopy theory** section.

Finally, all the information have to be explained in detail and shared with the scientific community. For this purpose, there are different specialised journals in which our studies can be published, and several scientific congresses for dissemination activity ("*Report results*" in the Figure II.1).

II.1. Computational Methods

Previously, rotational spectroscopists had to build molecular models manually, based on chemical intuition and experimental parameters from other studies and techniques (e. g. crystal structures, …). Nowadays, high accessibility to a different computational tools, as well as the enormous progress in the computational chemistry and High-Performance Computing (HPC) quantum chemistry computations have revolutionised the rotational spectroscopy researches. This has allowed us to study very complex molecular systems whose structures cannot be predicted without computation assistance, i.e. low interaction energy molecular aggregates,[1] flexible molecules[2] and very big molecules.[3]

In many cases, as shown in Figure II.1, computational calculations can be used to raise the initial hypothesis. Although nowadays computational methods increase their accuracy and robustness, as has been shown in many papers. It is also true that there are cases where computational methods do not give consistent results and the theory fails. So, it is highly desirable that the results could be verified by experiment(s).[4–6]

In this thesis, quantum-chemical calculations are used as a chemical tool. For this reason, in the following section we focus on the main purposes for which we use theoretical calculations, without going into extensive details. In our case, theoretical calculations are used for several goals. The first is to know the spatial arrangement of the atoms in a molecule that can be interconverted by rotations around single bonds. In this way, we can find out the possible conformations that the molecules can adopt (*conformational landscape*). At the same time, theoretical calculations allow us to estimate the internal energy differences

between each conformer. These relative energies give us an idea of the population of each of these conformations and hence a probability of being detected in the gas phase. Once the most stable structures have been identified, different molecular properties can be predicted with different theoretical models. These molecular properties will first serve as a guide and then help us to identify each of the possible conformers, as they are a fingerprint of the conformer and molecule. Finally, the experimental results will allow us to test the accuracy, calibrate and validate the different theoretical models. The demands on computational time in this thesis work have been high, so we have relied on computational facilities and supercomputers (IZO-SGI Scientic Computing Service (SGIker) of the University of the Basque Country (UPV/EHU) and *Centro de Super Computación* in Galicia (CESGA).

Conformational search

The molecular system under study can be very simple and its molecular and conformational structure prediction could be evaluated with chemical intuition.

Figure II.2 Summary of the theoretical methodologies applied for the rotational spectroscopy studies described in this Thesis.

However, the molecular systems interrogated nowadays (and of course the systems showed in this Thesis) present too many degrees of freedom doing impossible predict their conformers and/or interactions manually. Thanks to the great advances in computational calculations, there are tools that allow us to obtain a fairly accurate conformational scenario of complex molecular systems. In this context, the *conformational search* is the basic approach to study the conformational distribution of the molecules (or molecular aggregates), that evaluates using classical mechanics the structural, thermodynamic and kinetic properties of the. In our case, we used a two-step protocol, and this protocol has been successfully used in several studies[7,8] and is similar to that used in other research groups.[9,10]

First, we use molecular mechanics (MM), which allows us to scan the full potential energy surface (PES) in a very fast way, thanks to sophisticated computational algorithms. This speed is achieved by compromising on the accuracy of molecular properties such as the energy. After a first MM conformational search, the resulted structures are used as input structures to run another conformational search. With this procedure, we ensure that all the possible stable conformers according to the employed force field will be find during

the calculation. In the next step, the more stable structures (and therefore more likely to be detected in our experiment) are re-calculated with more robust quantum-chemical (QC) calculations using Density Functional Theory (DFT) and *ab initio* theory. In the following, we describe our two-step procedure.

The MM is a quick and easy way to evaluate the PES. Energy values of each conformer are calculated on basic principles of classical physics and energy gradient algorithms, where a force field is used for the energetic description. In this thesis, different force fields specifically designed to organic/bio-organic molecules have been used. The most employed was *Merck Molecular Force Field* (MMFFs),[11] while other methods such as *Assisted Model Building with Energy Refinement* (AMBER)[12] and *Optimized Potential for Liquid Simulations* (OPLS)[13] were used to a lesser degree. All these force fields are addressed to describe different molecular systems, some of them are built for big molecules like proteins or nucleic acids chains, and others for molecules in solution. The principal goal of the conformational search is to find all the possible low energy conformers that will be optimised with QC calculations in the next step. Now, we have to be sure that the local minima of MM conformational search will converge in the QC calculations to all low energy stable conformers, so we commonly also repeat this step using different force fields, because they may lead different stable conformers. In our experience, they provide a good starting point for our calculations.

However, the MM gradient optimization method only allow us to find a local minimum structure from each *starting conformer*. In order to scan the entire PES, different search engines are used. Specifically, the algorithm used to sample the conformational landscape is a combination of the Monte Carlo and large-scale low-scale sampling (LLMOD).[14] Finally, the energy window used in the conformational search varies depending on the system, typically 10-25 $kJmol^{-1}$.

We only use the structures from the MM conformational search as input coordinates for high-quality QC calculations. In our studies, we are interested in finding all the most stable conformers that could be detected experimentally (depending on the spectral S/N ratio, but usually conformers in the QC optimizations have Relative Energy < 10 $kJmol^{-1}$ or Relative Gibbs Free Energy < 10 $kJmol^{-1}$). The MM treatment for most of the presented systems was carried out using *MacroModel* software implemented in Schrödinger/Maestro Release 2017-1, LLC, New York.[15] However, the *Conformer–Rotamer Ensemble Sampling Tool* (CREST) software, which applies other conformational search methods and mathematical functions, has also been used in some specific studies (more details in the Results section).[16]

Structural optimization

The MM is a very powerful and user-friendly tool to fully explore the PES. However, the initial geometries from the conformational searches are tentative in both structure and energy values. So, they were re-optimised with QC calculations. In addition, high-resolution rotational spectroscopy is very sensitive to the atomic position in the molecule, so it must be described as accurately as possible, reducing and simplifying spectral analysis time. Different QC methods and basis sets are routinely used in computational modelling for rotational spectroscopy analysis. Hence, several methods and basis sets were used in this Thesis. All presented systems have been interrogated using both *ab initio* and several DFT methods.

On the one hand, an approach to solve the Schrödinger equation using the *ab initio procedure* is the Hartree-Fock method, in which the electron-electron interaction is replaced by an average interaction. However, this method is not very accurate for chemical phenomena because it does not consider the *electron correlation energy*. The Møller-Plesset second-order perturbation theory (MP2) is one of the most widely used *ab initio* methods to estimate molecular properties (structures, energies…) with electron correlation descriptions.[17] It is used routinely in high-resolution rotational spectroscopy and this method has been employed in all the systems presented in this thesis. MP2 is renowned for its accuracy but also for its high computational cost. On the other hand, *Density Functional Theory* (DFT) methods have also been used in this work. In DFT methods, the ground state electronic energy is fully determined by the electron density. The main issue in DFT methods is their deficiency in describing some important features, such as van der Waals interactions, dispersive effects, …. One of the most used DFT calculation in this Thesis has been B3LYP, that it does not explicitly include the dispersion correction. Such deficiency can be semi-empirically corrected with excellent results and low computational requirements, at least compared to MP2 methods. In our case, we have employed the dispersive Grimme's D3BJ correction, including Becke-Johnson damping.[18]

The basis set most commonly used in this Thesis is the standard Pople's basis set with polarization and diffuse functions (6-311++G(d,p)),[19,20] as they have been shown to be very robust. Alternatively, other basis sets have also been found to show very good agreement with experimental data, sometimes better than Pople's basis sets - for example Weigend's def2-TZVPP[21,22] or Dunning's different basis-set like cc-pVXZ[23,24] which has been employed more occasionally.

For sake of clarity, the specific QC methods and the basis sets used for each system are discussed in each section of the results obtained and the choice depended in many cases on the systems to be studied. Two different versions of

the commercial software Gaussian (Gaussian 09 and Gaussian 16)[25] and free ORCA software (Version 4.2.1)[26] have been used during this Thesis.

Molecular Properties

Once the molecular structure has been described by QC methods, other interesting molecular properties can be estimated to help and guide us in the analysis of rotational spectra. For example, in some cases, the molecules present some large-amplitude motion perturbation in the spectra due to their ability to cross some barriers through the tunneling effect. Tunneling across the barrier causes a doubling of rotational states with opposite parities, and in some cases transitions can be observed between the corresponding symmetric and antisymmetric rovibrational states. For this reason, this situation cannot be treated as a common perturbation of the rotation. A classic example arises from the inversion tunneling of ammonia caused for the inversion of the amino group. However, there are different motions that may generate the tunneling effects, such as internal rotation of methyl group, ring puckering of rings, pseudorotation and/or internal motion of light asymmetric groups and bending vibrations. In chapter VII of this Thesis, we will show the concrete case of the proton transfer tunneling which allow for the estimatation of the proton transfer reaction energy.

One of the principal abilities of of rotational spectroscopy is determine the molecular structures in gas phase. However, many molecules and aggregates present different conformations or structures that the high resolution of this technique allows discriminating unambiguously. Furthermore, rotational transition intensities could be use to determine the relative conformational population of the conformers. In this way, there are two different pathways to explain the conformational preference of the molecules. On the one hand, we could assume the thermodynamic equilibrium, where the population of each conformer corresponds to a Boltzmann distribution at prior expansion conditions (pressure and temperature). In this case, we can assume the equilibrium conditions and the lack of interconversion between the conformers to obtain relative energy difference between conformers. On the other hand, it is the kinetically controlled molecular conditions, and in this case the conformational population of the conformers come from the different contribution of other factors (monomer unit population, tautomeric population...).[27,28] Related to this conformational distribution, in some cases, due to the molecular interconversion processes, some low energy conformers cannot be detected in our experimental conditions. This effect is called “missing conformer” and is dependent on the weight of the carrier gas atoms.[29,30] Controlled conformational interconversion through different carrier gases could be used to find the lower energy conformers and also to know the lower energy interconversion pathways and dynamics of the molecules.[31]

In addition, other computational tools have also been used in this Thesis in order to help us interpreting the results obtained. For example, *Non Covalent*

Interaction (*NCI*) *method* is employed to visualise non-covalent interactions by making use of the distribution of both the electron density and its gradient;[32] *Natural Bond Orbital (NBO) analysis* provides an idea of the interaction energies between orbitals;[33] *Symmetry-Adapted Perturbation Theory* (*SAPT) calculation* yields the energy decomposition analysis between different molecules or molecular systems;[34] and *Quantum Theory of Atoms in Molecules* (QTAIM) gives the topology of the electron density distribution.[32] More detailed information can be found in their references.

II.2. Experimental Techniques

Microwave spectroscopy has long been considered a powerful technique for accurately determining the structure of molecules and their aggregates in gas-phase, being restricted to molecules of small size (typically 30 atoms). Moreover, the devices were not commercially available, so it was necessary to design and build the equipment/s yourself. Nowadays, their applicability has been extended with the introduction of the broadband spectrometers and the first commercial equipment. This significantly increases the size of the systems studied. In the *Grupo de Espectroscopía* of the University of the Basque Country (UPV/EHU) and in particular in this Thesis, the potential of the technique has been pushed to the limit in many cases. Due to the high-resolution power of the rotational spectroscopy, it is possible to distinguish unambiguously between different conformational structures, tautomers,[35,36] isotopologues[37] and enantiomers.[38,39] The only requirement is that the molecule or cluster must have a non-zero electric dipole moment.

Rotational spectroscopy could be separated according to the frequency region of the spectrum. Rotational transitions of the molecules appear commonly in the centimetric, millimetric and sub-millimetric wavelength region of the spectra. This Thesis is focused on the centimetric wavelength and we have employed supersonic jet expansion conditions to cool down the temperature of the molecules (~1K rotational). This experimental conditions allow us to increase the population of the lower rotational energy levels of the systems (all details in the Molecular-beam supersonic expansion environment).

The following sections present the microwave spectrometers used to carry out the experiments. Two different home-made rotational spectrometers are available at the University of the Basque Country (UPV/EHU) and both of them acquire the molecular free induction emission in the time domain, then transformed with Fourier transform to obtain the frequency domain spectra. The experiments are carried out in supersonic molecular-beam expansion conditions (all details in Molecular-beam supersonic expansion environment section of this chapter). The older spectrometer is the molecular beam Fabry-Perot resonator cavity microwave (cav-FTMW) spectrometer,[40] that works in the 4-18 GHz

spectral region and it is equipped with a laser vaporization system to study thermally unstable molecules (the Laser Ablation (LA) process will be discussed later).[41] The second is a chirped-pulse broadband Fourier transform microwave (CP-FTMW) spectrometer, based on the B. H. Pate's original design (University of Virginia)[42,43], that can measure in the 2-18 GHz spectral range using two different arrangements. The devices have many similarities to Fourier transform nuclear magnetic resonance (NMR) spectroscopy, although the relaxation time of magnetic resonance is 10^6-10^7 times longer than that of rotational resonance relaxation. Thanks to the development of electronic components, we were able to update, modify and optimise the CP-FTMW experiments during the developing of this Thesis. Although the pros and cons of the two set-ups could be extensively discussed, it is clear that a key advantage of CP-FTMW over cav-FTMW is the ability to cover a much larger frequency range in a single event and consequently it provides spectra with higher sensitivity through the accumulation of millions of spectra. In terms of accuracy, the measurements in the cav-FTMW spectrometer have an accuracy of ±5 kHz, while the CP-FTMW spectrometer has an uncertainty of ±15 kHz. Hence, the cavity allows us to better separate congested signals in a narrower frequency range. This feature can be of interest in cases where the spectrum exhibits fine and hyperfine structures, such as internal rotation or quadrupole tensor effects. As part of the internationalization of this Thesis, we will explain some experimental set-ups used during the stay at the University of Alberta.

Basically, all FTMW spectroscopic experiments have three main stages. The first is the excitation stage, in which a microwave pulse is generated and amplified to the appropriate power and frequency range. The second is the interaction stage, in which the radiation pulse polarises the rotational resonances of the system in the jet. And the third is the detection stage in which the coherent molecular emission, also known as free induction decay (FID), is amplified and/or heterodyned for digitisation in the time domain, and then Fourier transformed to the frequency domain using an oscilloscope or a computer.

Cav-FTMW

In 2010, the cav-FTMW spectrometer was built in the University of the Basque Country (UPV/EHU).[40] Later, in 2012, the laser vaporization system was installed, making possible the study of thermally unstable biomolecules. [41]

The spectrometer is based on the original design of the Fabry-Perot MW cavity, whose resolution was improved by locating the molecular beam along the resonator axis (called coaxially oriented beam-resonator arrangement or COBRA).[44–46] The mirror diameter is 33 cm. The electronic scheme of the cavity spectrometer is shown in the Figure II.3.

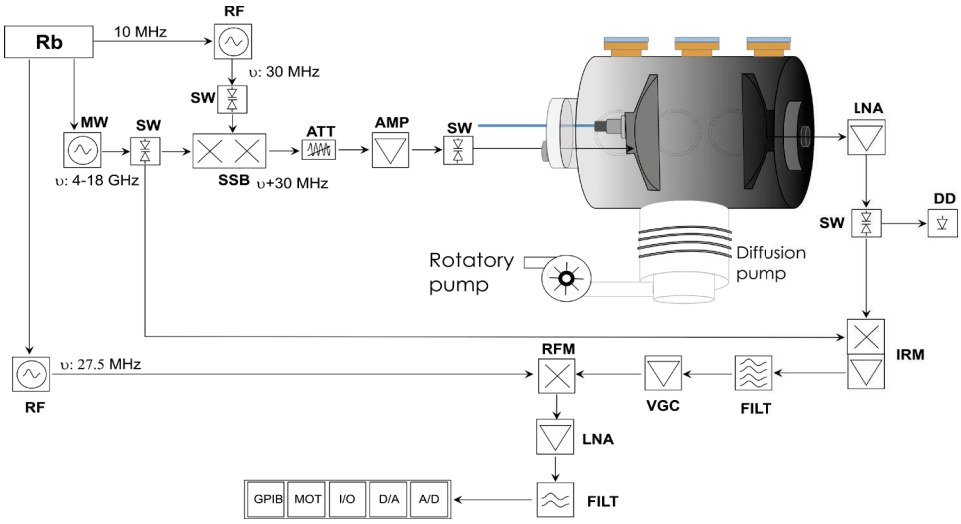


Figure II.3 cav-FTMW spectrometer at the UPV-EHU.

The cav-FTMW electronics work as follows: referring to Figure II.3. A primary microwave continuum (CW) radiation at frequency υ generated by the synthesiser is shifted to $\upsilon + 30$ MHz. Each molecule or conformer requires a different polarization power, so the radiation power is optimized through the attenuation/amplification sequence. The excitation of a molecule requires fast amplitude modulated MW pulses (~1 μs) which are produced with fast pin-diode switches (SW). Approximately, 0.5-1.5 μs MW excitation pulse is broadcasted in the resonator through a L-shape antenna located on an aluminium mirror and then excites the molecular rotational resonance transitions by aligning the molecular dipole moments. After the dissipation of the excitation pulse, the molecular emission is recorded by the other antenna in the time-domain and amplified with a low noise amplifier (LNA, NF = 1.6). Finally, the FID is down-converted to 2.5 MHz in two mixing steps and acquired with a National Instrument PXI Oscilloscope target. The cavity-FTMW set-up is fully controlled by the software implemented in the National Instruments PXI System which manages the pulse generation, the digital and analogue input/output, the transient acquisition and the mirror position.

The cav-FTMW apparatus is very robust as demonstrated in different publications,[37,41] but the greatest disadvantage is the extremely narrow bandwidth for each microwave pulse (~1 MHz). This is because the distance between the cavity mirrors should be integer multiple of the polarization/acquisition wavelength, so for each frequency, the mirrors should be mechanically tuning leading to a tedious spectral acquisition process. Cavity-FTMW has been used to record the spectra of all sugars presented in this Thesis (Chapters III-VI).

CP-FTMW (EHU/UPV)

Due to the advances of digital electronics, in 2006 B. H. Pate's group designed the first broadband microwave spectrometer [42,43] with the application of chirped-pulse techniques. The instrument is able to generate phase coherent chirped-pulse in the form of a broadband linear frequency sweep and is equipped with broadband frequency receivers. In 2014, 6-18 GHz CP-FTMW spectrometer was built in the University of the Basque Country (UPV/EHU).[47] The apparatus was then reconditioned and updated to obtain a more efficient 7-14 GHz polarization and the frequency extension to 2-6 GHz. The molecular jet is generated using homemade nozzle, pulsed with *Parker Serie 9* valve powered by *Iota One* driver. The CP-FTMW works using 2 different arrangements in order to collect broadband data in 2-6 GHz and 7-14 GHz bands. Although the principles of the experiment are very similar, the polarization stage is different for each arrangement.

For the lower 2-6 GHz band, the chirped signal is directly generated by the Arbitrary Waveform Generator (AWG). We usually use 2-4 μs chirp duration, but generally its duration is one of the parameters that we need to optimize depending on the different molecular systems. The MW pulse is amplified by the Solid State Amplifier (200W - SSA) to an adequate input power and then the overtones are removed by filter (FILT). (Figure II.4).

For the upper 7-14 GHz band, the electronic scheme is more complicated (Figure II.5). The reason is that the AWG cannot generate waveforms with frequencies higher than 10.5 GHz. Therefore, to reach the 7-14 GHz frequency

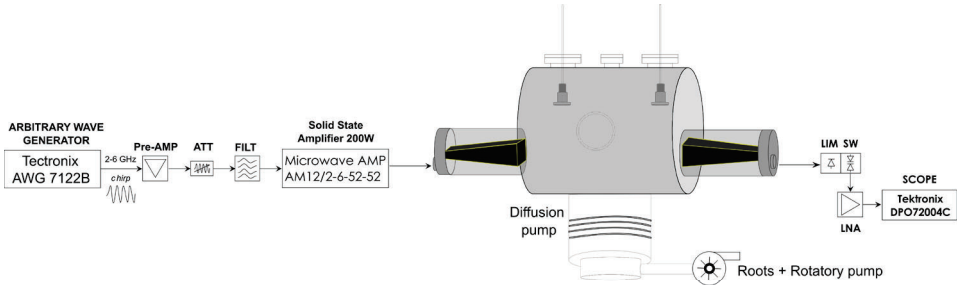


Figure II.4 Scheme of the 2–6 GHz CP-FTMW spectrometer.

range, we initially generate a chirp in the fundamental band 3.5-7 GHz with the AWG. The chirp is amplified (Pre-AMP) and then attenuated to enter with the correct power in the doubler (Dobl) in order to obtain the 7-14 GHz frequency range. Since the fundamental band and the second overtone still exist in the signal when the fundamental chirp is doubled, the pulse is cleaned with a high pass filter (FILT 1) and then a low pass filter (FILT 2). The obtained 7-14 GHz chirped-pulse is further amplified by the 250W Traveling Wave Tube (TWT).

For both 2-6 GHz and 7-14 GHz arrangements, the chirped-pulse is broadcasted by the same Q-Par horn antenna in the high-vacuum chamber. These antennas are designed to operate from 2 to 18 GHz. Thus, it is unnecessary to change the components inside the vacuum chamber when the working frequency band is changed. In the radiation-sample interaction region, there are three nozzles installed. It has been showed that the signal amplitude using two nozzles is twice than that with only one nozzle (more details in the Molecular-beam supersonic expansion environment section in this Chapter).

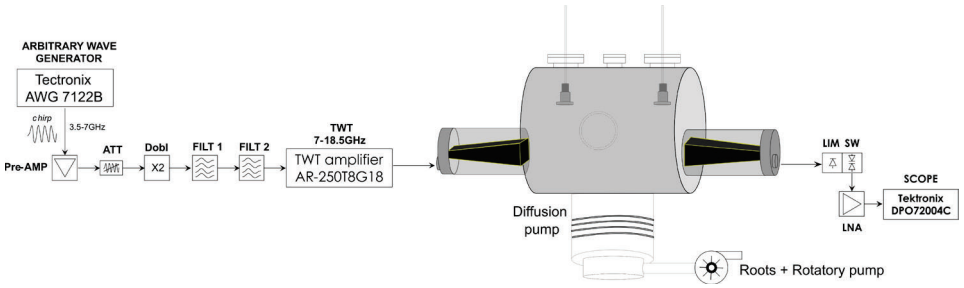


Figure II.5 Scheme of the 7–14 GHz CP-FTMW spectrometer.

The detection line is similar for both spectral ranges. The only variation regards a Low Noise Amplifier (LNA), being specific for each frequency range.

The set-ups should be optimised for each molecular system aimed to be measured. Usually, we use between 6-15 excitation/detection cycles for each molecular pulse. Furthermore, different valve opening times, backing pressures and carrier gas have been used depending on the type of molecular target and experiment. Also, the FID time of the molecular signal depends on the molecular system and the electronic components and the resolution that we want to obtain in our experiment. Depending on the system and the spectral band, the range between10-40μs FID times have been used to record spectra.

CP-FTMW (University of Alberta)

In the University of Alberta, three MW spectrometers are available. However, during the period abroad of this Thesis, only the CP-FTMW was utilized, so in the following paragraphs it is the only that is described.

The 2-6 GHz CP-FTMW spectrometer in the University of Alberta is similar to the Bilbao's 2-6 GHz spectrometer explained before. Actually, both set-ups are based on the B. H. Pate's original design (University of Virginia).[43] At the University of Alberta, a 12 Gs/s arbitrary waveform generator (AWG) generated a 2-6 GHz 1-2 μs chirped-pulse which is directly amplified by a 400 W, 2.5-7.5 GHz traveling wave tube (TWT) amplifier.[43] Two horn antennas located in the vacuum chamber are used to broadcast the amplified MW pulse and to receive the molecular emission free induction decay (FID) signal. The molecular jet is generated using homemade nozzle, pulsed with *Parker Serie 9* valve powered by *Iota One* driver. Limiter and switch elements protect the LNA and the scope when molecules are polarized and TWT is working. Once the polarization wave dissipated, the protection switch is opened and the molecular signal is recorded and averaged in the Tektronix 25 Gs/s oscilloscope. The frequency precision measured by the CP-FTMW spectrometer is approximately 10 kHz.

Through slight changes of electronic components, the 2-6 GHz chirp spectrometer of the University of Alberta can operate also in the 8-12 GHz band. In this case, 0.5 GHz bandwidth and 4 μs duration 4-8 GHz chirped signal has to be generated by the AWG. This signal is then doubled to 8-12 GHz frequency range while the doubled bandwidth is 1 GHz. The power is optimised to satisfy the SSA amplifier input power specifications and the fundamental frequencies and overtones are removed using low pass and high pass filters. Then the microwave signal is amplified by a 20W SSA to gain enough power to polarize the transitions at rotational energy levels of molecules in 1 GHz bandwidth. Lastly, the acquisition is carried out in the same way as the 2-6 GHz set-up, but special electronic components are required for the detection in the 8-12 GHz frequency range.

Molecular Vaporization Methods

The rotational spectroscopy techniques used in this Thesis were carried out in gas phase. One critical process is to bring intact molecules into the gas phase. An appropriate vaporization method is selected depending on the characteristics of the sample (thermal stability, vapor pressure, …). In this Thesis, we used the conventional heating method for liquid and some solid samples and laser ablation for thermally unstable molecules such as sugars.

Conventional heating methods

Heating is the most typical method to vaporize the molecules, however some molecules cannot provide enough vapor pressure to be detected by heating. We usually use this method for liquids and solids that are thermally stable and that can be heated below 230ºC. For samples that need to be heated more that 230ºC, we use laser ablation because the solenoid valve does not work properly with T>230ºC. Actually, the experiment that does not involve laser ablation is easier and more stable. Our heating system is formed by three main components.

The first one is the nozzle. Both spectrometers of the University of the Basque Country use Parker *Serie 9* commercial valve powered by *Iota One* driver to generate the molecular beam (more details are in the following sections). Unfortunately, commercial nozzles do not have enough space to locate the sample, so the nozzles have to be redesigned. The customised home-made nozzles have a crown shape which allow the introduction of the sample below vacuum chamber. The other two components are the heating wire and the thermocouple. The resistance heating wire is located inside housing. Its shape is specifically designed to be adapted to the customised nozzles. The thermocouple measures the temperature of the nozzle. Finally, there is the temperature controller. There are several commercial temperature controllers available. We use *Fuji Electric PXR4* programmable temperature controller and 12V AC power supply for the heating wire. The programmable temperature controllers switch on/off the current that the heating wire received, guided by digitally adjustable parameters which results in more controlled heating ramps. The controller monitors the nozzle temperature in real time using the thermocouple.

Laser vaporization

Figure II.6 shows a fragment of the spectrum of 2-deoxy-D-ribose heated in the conventional way, in which the most intense signals come from thermal decomposition products, that usually belong to water and acetone in sugar samples. This example illustrates the importance of not using conventional heating method in sugar measurements.

During these years, thermally unstable molecules and molecules whose vapor pressure are extremely low were undetectable with conventional vaporization methods in rotational spectroscopy. Although some refractory small molecules have been generated using discharges, there are only a few samples that can be directly generated using this method. Further, thermally labile molecules cannot be usually generated using discharge techniques. In order to vaporise this kind of samples, the combination of jet expansion and laser-induced molecular plasma ejection systems were designed. Different laser vaporization methods have been developed. To our knowledge, the first laser vaporization system coupled to MW spectrometer was presented in 1989 by R. D. Suenran *et al*, when they measured the 1_{01}-0_{00} transition of SiC_2 (Melting Point > 2000 ºC).[48] After that, K. A. Walker and M. C. L. Gerry developed other two laser-ablation devices. The first laser ablation (LA) set-ups were focused on studying systems with very high boiling point and very low vapor pressure like metal oxides,[49] halides[50] and sulfides.[51] After that, J. L. Alonso's group developed LA system for thermally unstable molecules, concretely organic molecules.[52] In 2012, the cavity spectrometer of the University of the Basque Country was equipped with a pico-second laser ablation system. This strategy enabled the study of different

biomolecules like monosaccharides, hormones, nucleosides and small peptides.[3,41,53,54]

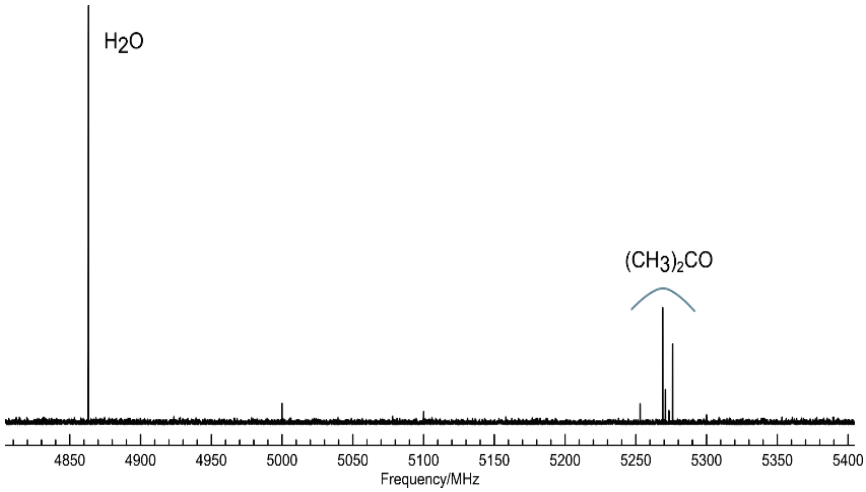


Figure II.6 Section of the spectra of 2-deoxy-D-Ribose sample using conventional heating method.

Briefly, the sample is prepared by grinding and mixing with a commercial binder and then, the mixture is pressed to get a cylinder shaped solid rod (approximately 1 cm length and 0.6 cm of diameter). Once the sample target rod is sufficiently dry, it will be loaded close to the Parker valve. The vaporization process goes as follows. The sample is irradiated by the laser pulses coming from a Nd-YAG laser in the third harmonic (355 nm). The vaporized material is entrained in an inert carrier gas stream (Ne or Ar) which cools rotationally the molecules in the beam to 0.5-5K. Once the molecules reach the center of the Fabry-Perot mirrors, a microwave radiation is applied to the cavity.

Molecular-beam Supersonic Expansion Environment

The supersonic expansion molecular-beam source is an important tool in the high-resolution spectroscopy. Experimentally, high particle densities and low internal temperatures have facilitated studies of electronic, vibrational, and rotational molecular (and small clusters) structures. The spectroscopy experiments presented in this Thesis were carried out in supersonic expansion molecular-beam conditions, which provide a virtual environment free of collision and interactions.

Experimentally, we obtain the supersonic expansion conditions expanding a gas from the high-pressure region (p_0: 0.5-15 bar) into a low-pressure region (10^{-7}mbar in standby conditions) through a small hole (usually 0.25-0.5 mm of

radius) (see Figure II.7). Pulsed gas beam is generated with a fast solenoid magnetic valve (*Parker* Serie 9 commercial valve powered by *Iota One* driver). This valve has a conical Teflon poppet tip which seals directly against the output hole of the valve. The duration of the gas pulse is approximately 300-1000 μs. It expands molecules into a high vacuum chamber, which is pumped by a sequence of diffusion, roots and rotatory pumps (roots pump are only installed in the CP-FTMW of the University of the Basque Country). This system allows us to keep the pressure in the vacuum chamber around 10^{-5}-10^{-4} mbar in the working conditions.

In particular, the systems presented in this Thesis were ejected using pure noble gases (He, Ne or Ar). It is necessary to emphasize that with the laser ablation, the molecular dilution and expansion occurs at the same time inside the vacuum chamber. However, in conventional experiments the mixture is done before expansion. As consequence of the molecular-beam expansion, the rotational temperature of the molecules (or clusters) decreases to ~ 0.5-5 K, increasing significantly the population of the rotational and vibrational ground states. During the jet expansion, the molecules which travel in the 'zone of silence' (Figure II.7), are in a virtual condition free of collisions and interactions.

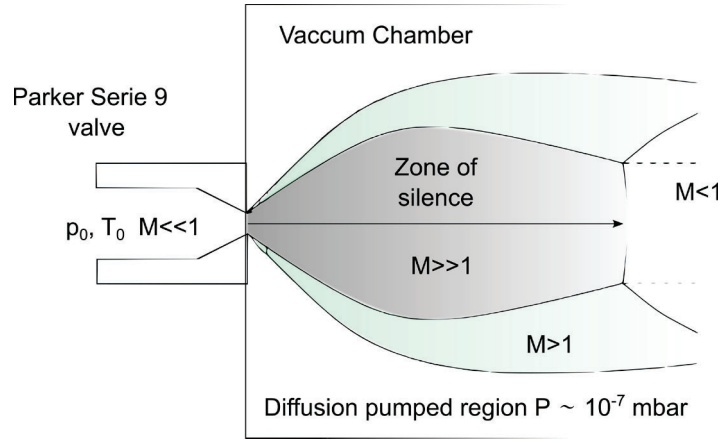


Figure II.7 Scheme of a supersonic gas jet.

Experimental measurements in supersonic expansion molecular-beam conditions offer the following advantages:

1. The linewidth that comes from Doppler broadening and collision broadening is greatly reduced. It allows exceptionally high-resolution and high-precision measurements.
2. At room temperature the population of the molecules is divided in several rotational and vibrational excited states. In the jet conditions only the

lowest vibrational and rotational states are populated due to the low internal temperature and the lack of intermolecular collisions. Experimentally, in the molecular emission spectroscopy, we can in some way control the temperature of the experiment in order to increase the probability of exciting the systems at the spectral ranges in which the population is higher (in our experiments 2-18 GHz concretely).

3. Supersonic jet environment allows the formation and, therefore, conformational lock of weakly bound complexes, such as, Van der Waals, weak H-bond or π-π complexes. For this reason, it is the most popular environment to study small molecular clusters using high resolution spectroscopy techniques.
4. Last but not least, supersonic jet conditions make possible vaporizing the molecules using laser ablation vaporization system. Low vapour pressure molecules can be carried to the gas phase using laser source.

II.3. Rotational Spectroscopy Theory:

Rotational spectroscopy is directly encoded by the geometric distribution of the atoms in the molecule. Pure rotational transitions can be detected in the microwave or millimeter-wave region, although for some small molecules, they can be detected in the far IR region. Like many other spectroscopy techniques, quantum mechanics is employed to understand and interpret the results obtained from rotational spectroscopy. Indeed, pure rotational quantized energy of its states can be calculated with the time-independent Schrödinger equation. In the following sections, we present general theoretical principles for rotational spectroscopy research. Further details can be found in the specialized literature.[55,56]

Rigid Rotor

Derived from the rotational Hamiltonian that describes the rotation of the molecule, the constants of interest are presented as eigenvalues, which refers to the energy levels. The Hamiltonian operator is usually expressed in terms of angular momentum operators.

$$\hat{\mathcal{H}} = \frac{1}{2}\left(\frac{\mathrm{P}_x^2}{\hat{I}_x} + \frac{\mathrm{P}_y^2}{\hat{I}_y} + \frac{\mathrm{P}_z^2}{\hat{I}_z}\right)$$

Considering the approximation of the rigid rotor and the mass center of atoms, for a molecule having *i* atoms whose nuclei mass is m_i and nuclear coordinates are (x_i, y_i, z_i), the elements of the inertial tensor *I* can be calculated as:

$$I_{ij} = \sum_i m_i(b_i^2 + c_i^2), \;\; i = j \quad ; \quad I_{ij} = -\sum_i m_i a_i b_i, \;\; i \neq j$$

In order to obtain the diagonal inertial tensor which can be easily treated mathematically, the initial inertial tensor is transformed by applying the appropriate transformation (inertial matrix *R*), forming a new diagonal inertial tensor. This diagonal tensor represents a transformation of the original coordinate system to one 3x3 diagonal matrix, each lying along the axes of the new system. This coordinate system is called the principal axis system (PAS) where the axes are typically represented as (a, b, c). The diagonal elements of this matrix correspond to the principal moments of inertia of the molecule. For geometrical convenience, the axis of inertia is defined as I_c being always the largest moment of inertia, and I_a as the smallest one. Depending on the value of the inertial moment, molecules can be distinguished into five groups which are related with their symmetry: linear tops ($I_a = 0$, $I_b = I_c$), spherical tops ($I_a = I_b = I_c$), symmetric oblate tops ($I_a = I_b < I_c$), symmetric prolate tops ($I_a < I_b = I_c$) or asymmetric tops ($I_a < I_b < I_c$). All the systems presented in this Thesis are asymmetric tops, and the systems that nowadays are mostly studied using rotational spectroscopy are asymmetric tops systems.

One of the problems with these types of molecules is that the Schrödinger equation for the asymmetric top molecules does not have an analytical solution, so it should be solved numerically with the help of a computer. Luckily, their solution can be derived from oblate and prolate symmetric top molecules whose Schrödinger equation can be solved analytically. For this reason, the asymmetric top Hamiltonian is solved by using the symmetric top basis set. The classical energy-level expression for the rigid symmetric top is defined as:

$$\mathcal{H}_{rig} = A\mathbf{P}_a^2 + B\mathbf{P}_b^2 + C\mathbf{P}_c^2$$

Resolving the symmetric top energy Schrödinger equation generated with the symmetric top Hamiltonian, (omitted in this work) we can obtain the following energy terms for oblate and prolate tops,

$$E = BJ(J+1) + (A-B)K^2 \qquad \text{(oblate top)}$$

$$E = BJ(J+1) + (C-B)K^2 \qquad \text{(prolate top)}$$

where A, B y C are the rotational constants which corresponds to $A = h^2/8\pi^2 I_a$, $B = h^2/8\pi^2 I_b$, $C = h^2/8\pi^2 I_c$, in which *J* and *K* are the rotational quantum numbers.

Now, to extend above results to an asymmetric molecule, it is necessary to estimate “how far the asymmetric tops deviate from the symmetry one” or “the level of asymmetry” of the asymmetric tops aimed to be studied. For this purpose, Ray’s constant is defined based on the rotational constants:

$$\kappa = \frac{2B - A - C}{A - C}$$

The value of Ray's constant is between -1 and 1, and from 0 to 1 corresponds to an oblate type system, while from -1 to 0 corresponds to a prolate one. All the values between -1 and +1 describe the asymmetry degree of the molecule. Moreover, by using the value of Ray's constant, we can obtain the rotational energy levels of one concrete system, through the simply extrapolation of the results from the symmetric top molecules, using a linear approximation between the limited oblate and prolate cases (see Figure II.8). Now, the quantum numbers who define the rotational transitions of the symmetric molecules are not valid for the asymmetric systems. Therefore, the rotational transitions of asymmetric tops are defined by introducing King-Hainer-Cross notation ($J_{K-1\,K1}$) where K_{-1} and K_1 are the pseudo quantum numbers, meaning that they are not pure quantum numbers, or rather extrapolated from the K values within the prolate and oblate limitation.

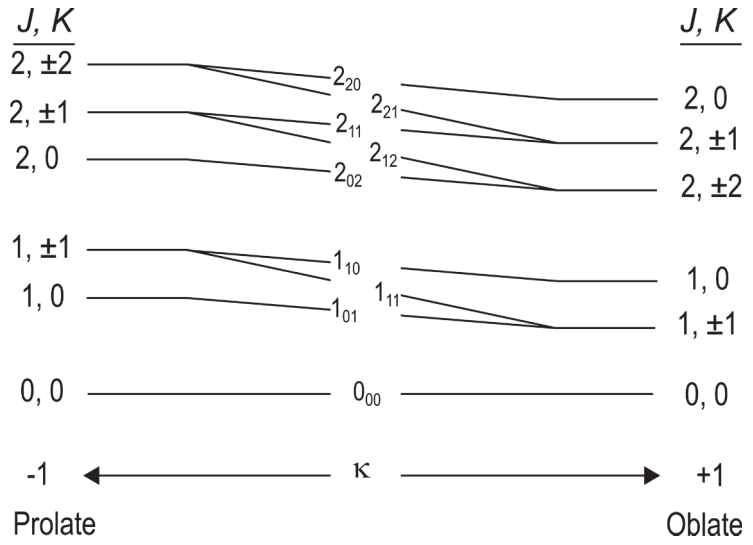


Figure II.8 Relation of the asymmetric rotor energy levels and the limiting prolate (left) and oblate (right) cases.

In general, one of the most important point required for the interpretation of the rotational spectra is to explain which types of transitions have more probability compared to others to be observed in the spectra, and in some cases, why certain transitions cannot be observed at all. Therefore, the selection rules establish which transitions are allowed and which are forbidden, and typically, those transitions who are allowed are the signals that we observe and use to fit the rotational spectra and to find different conformers. The principal rule for observing a rotational transition is that the molecule should have a non-zero permanent dipole moment, which can be expressed in Dirac notation:

$$< J, K_{-1},\ K_1|\mu|J', K'_{-1}, K'_1 >\neq 0$$

For *J* quantum numbers, the allowed transitions between two states should have the difference $\Delta J = 0, \pm 1$. It is defined that $\Delta J = 0$ transitions represent the *Q*-branch, $\Delta J = -1$ for *P*-branch and $\Delta J = 1$ for *R*-branch. Furthermore, for the asymmetric molecules, the allowed pseudo-quantum number changes in K_{-1} and K_1 are summarised in Table I.1.

Table I.1 Selection rules for the different transitions of the asymmetric molecules.

Type of transition	ΔK_{-1}	ΔK_{-1}
a (if $\mu_a \neq 0$)	$0, (\pm even)$	$\pm 1 (\pm odd)$
b (if $\mu_b \neq 0$)	$\pm 1, (\pm odd)$	$\pm 1 (\pm odd)$
c (if $\mu_c \neq 0$)	$\pm 1, (\pm odd)$	$0\ (\pm even)$

Hence, for the asymmetric molecules, μ_a-, μ_b- and μ_c-type lines can be identified. On the other hand, the signal intensity depends on different factors, such as the dipole moment components, the rotational temperature, the population of the conformer in the experiment and the type of the transition or the polarization energy.

Rigid rotor perturbations

In the previous section, the molecular rotation has been treated with the approximation of the rigid rotor model. This model, in which the molecule is considered that is formed by atoms without volume, connected by invisible and massless rigid bonds, is useful for the molecules whose atomic disposition is very stable or immobile. This approximation is ideally assumed and it cannot be applied in much more complicated real cases. In reality, different kinds of perturbation exist, for example, the chemical bonds are somehow like strings, which means that with the increasing of the rotational energy, the centrifugal force takes over trying to pull the atoms apart. For this reason, the introduction of perturbation corrections to the rigid rotor approximation is required. In this way, the Hamiltonian presented for the rigid rotors should be modified by adding a correction term as:

$$\mathcal{H}_R = \mathcal{H}_{rig} + \mathcal{H}_{correc.}$$

One of the most common corrections regards the **centrifugal distortion** constants. The distortion constants can be defined as distortions derived from the stretching of bonds that the molecule suffers when it is rotating at high rotational energy levels, as described before. Generally, the value of these constants is very small (only few kHz) with respect to the rotational constants. To introduce centrifugal distortion constants into the Hamiltonian H_R, it is convenient to apply

Watson's reduced Hamiltonian expressed by the unitary transformation of H_R in the following term:[57,58]

$$\mathcal{H}_R^{Red} = U^{-1}\mathcal{H}_R U$$

where the unitary operator ($U^{-1} = U^{\dagger}$) is defined as $U = e^{iS_3}e^{iS_5}$. S_3 and S_5 are transformation coefficients. Without going into details, Watson's reduction provides two kinds of reduced Hamiltonian, asymmetric (A) and symmetric (S). A-reduced distortion Hamiltonian is obtained including all the sextic terms expressed as:

$$\mathcal{H}_R^{(A)} = \mathcal{H}_r + \mathcal{H}_d^{(4)} + \mathcal{H}_d^{(6)}$$

$$\mathcal{H}_r = B_x^{(A)}P_x^2 + B_y^{(S)}P_y^2 + B_z^{(S)}P_z^2$$

$$\mathcal{H}_d^{(4)} = -\Delta_J P^4 - \Delta_{JK}P^2P_z^2 - \Delta_K P^2P_z^2 + 2\delta_j P^2\left(P_x^2 + P_y^2\right) + \delta_k\left[P_z^2\left(P_x^2 + P_y^2\right) + \left(P_x^2 + P_y^2\right)P_z^2\right]$$

$$\mathcal{H}_d^{(6)} = -\Phi_J P^6 - \Phi_{JK}P^4P_z^2 + \Phi_{KJ}P^2P_z^4 + \Phi_K P_z^6 + \phi_j P^4\left(P_x^2 + P_y^2\right) + \phi_{jk}P^2\left[P_z^2\left(P_x^2 + P_y^2\right) + \left(P_x^2 + P_y^2\right)P_z^2\right] + \phi_k\left[P_z^4\left(P_x^2 + P_y^2\right) + \left(P_x^2 + P_y^2\right)P_z^4\right]$$

where B_x, B_y and B_c are the rotational constants, Δ_J, Δ_{JK}, Δ_K, δ_j , δ_k are the quadratic distortion coefficients and Φ_J, Φ_{JK}, Φ_K, ϕ_j, ϕ_k , ϕ_{jk} are the sextics. For the same propose, Watson defined the S-type Hamiltonian as:

$$\mathcal{H}_R^{(S)} = \mathcal{H}_r + \mathcal{H}_d^{(4)} + \mathcal{H}_d^{(6)}$$

$$\mathcal{H}_r = B_x^{(S)}P_x^2 + B_y^{(S)}P_y^2 + B_z^{(S)}P_z^2$$

$$\mathcal{H}_d^{(4)} = -D_J P^4 - D_{JK}P^2P_z^2 - D_K P^2P_z^2 + d_1P^2(P_+^2 + P_-^2) + d_2(P_+^4 + P_-^4)$$

$$\mathcal{H}_d^{(6)} = -H_J P^6 - H_{JK}P^4P_z^2 + H_{KJ}P^2P_z^4 + H_K P_z^6 + h_1P^4(P_+^2 + P_-^2) + h_2P^2(P_+^4 + P_-^4) + h_3(P_+^6 + P_-^6)$$

where B_x, B_y and B_c are the rotational constants, D_J, D_K, D_{JK}, d_1 , d_2 are the quadratic distortion coefficients and H_J, H_K, H_{JK}, H_{KJ}, h_1, h_2 , h_3 are the sextics.

In the spectra recorded during this Thesis, only the quadratic distortion coefficients of Watson's reductions were used. Generally, the sextic distortion coefficients are requiredfor the interpretation of the rotational spectra recorded at higher rotational energy levels, generally, in millimetre spectroscopy, where the molecules are normally studied at room temperature.

So far, the rigid rotor model has been modified with the consideration of vibrational contribution and other distortion contributions. From the spectrum, we can easily observe how the molecular motions affect the rotational transitions.

Usually, one can observe line splitting caused by large amplitude motions (LAMs) in some molecules possessing certain groups or unique structure. In these cases, additional parameters are required in order to fit the spectrum.

Structure determination

One of the applications of rotational spectroscopy is the accurate determination of the molecular structure. The rotational constants are related to the atomic disposition in the molecule. This information is directly related with the bond lengths and the bond angles. The extremely high accuracy of the determination of the rotational constants and the high dependency of the rotational constants with the atomic arrangement, makes rotational spectroscopy one of the most powerful techniques to determine the atomic disposition in the molecules.

Commonly, in rotational spectroscopy studies, the experimental rotational constants are compared with the ones obtained from high level computational calculations, in order to identify the different conformers. However, the bond angles and distances have high dependency with the computational method employed. On the other hand, experimentally, there are several factors that influence the bond lengths and angles. Although several uncertainties could affect the values, like the imprecision in Planck's constant and/or in the atomic masses, one major error in the structures can be attributed to their own vibrations even if the molecule is in its ground vibrational state. For this reason, different sophisticated mathematical methods have been developed to obtain the reliable structural parameters correcting the vibrational effects. Concretely, different types of bond lengths are defined, for example r_e, r_0, r_s, $\langle r \rangle$ and r_m. In the following sections, the methods used in this Thesis will be described.

Equilibrium structure (r_e)

Although the r_e structure is undetectable by rotational spectroscopy due to a zero-point vibrational energy effect, to know its physical definition is essential to understand the general concepts of our research. The equilibrium structure r_e corresponds to an atomic disposition in which the potential energy surface is a minimum without taking into account the vibrational effects (Figure II.9).

The relation between the equilibrium structure and the different vibrational states is applicable to the three rotational constants in an asymmetric-top molecule using the following expression:

$$B_e = B_v + \sum \alpha_i (\vartheta_i + \frac{d_i}{2})$$

Where the α_i is the rotation-vibration interaction constant and d_i is the vibrational degeneracy. In this Thesis, due to the supersonic expansion conditions, mainly the *v*=0 vibrational state is observable. However, it is important to know that the structural optimization using quantum chemistry calculations provides the r_e structure, so the comparison of the rotational constant (in a specific *v*=n state) with the theoretical values should be done carefully.

Effective structure (r_0)

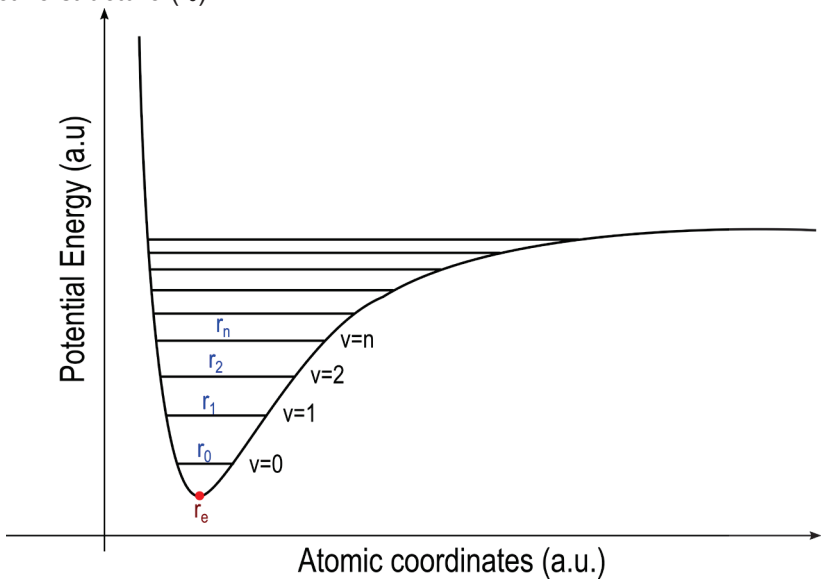


Figure II.9 Potential energy function with vibrational states.

The effective structure r_0 represents the real and the experimentally observable structure in a given vibrational state. When this structure corresponds to the structure of the ground vibrational state, we refer to it as r_0 structure. The experimental rotational constants and the geometry derived by theoretical calculations can be used to made a least square fit, to obtain the atomic disposition which reflects the experimentally rotational constants. This procedure may be done using or not the isotopic substituted rotational constants, more experimental constants are used more structural parameters can be adjusted. Despite of this, the results may not be reliable if the molecule contains large amplitude motions such as internal rotations or out-of-plane bending motions.

Substitution structure (r_s)

When the isotopic substituted species (isotopologues) are determined experimentally, one method commonly used to obtain the accurate position of the atoms was developed by Kraitchman.[59] In the Kraitchman's method, it is assumed that the molecule is rigid and the bond distances angles are the same

in the parent and isotopologues, which means that the changes of the rotational constants are only due to the changes of the mass of one atom. For the atoms in which this change is very low (C, O, N…), this approximation operates well, although for example for the H, this approximation could not be completely accurate due to the great mass change.

Therefore, to determine the r_s structure of asymmetric top molecules with a single isotopic substitution using Kraitchman's equations, we define firstly the tensor ***P*** whose diagonal elements are:

$$P_x = P_{xx} = \sum m_i x_i^2 \qquad P_y = P_{yy} = \sum m_i y_i^2 \qquad P_z = P_{zz} = \sum m_i z_i^2$$

In the principal axis system, the ***P*** matrix is diagonal, so off-diagonal matrix elements have been omitted. Briefly, the coordinates of the substituted atom can be calculated using the following expression:

$$|x| = \left[\frac{\Delta P_x}{\mu}\left(1 + \frac{\Delta P_y}{I_x - I_y}\right)\left(1 + \frac{\Delta P_z}{I_x - I_z}\right)\right]^{1/2}$$

$$|y| = \left[\frac{\Delta P_y}{\mu}\left(1 + \frac{\Delta P_z}{I_y - I_z}\right)\left(1 + \frac{\Delta P_x}{I_y - I_x}\right)\right]^{1/2}$$

$$|z| = \left[\frac{\Delta P_z}{\mu}\left(1 + \frac{\Delta P_x}{I_z - I_x}\right)\left(1 + \frac{\Delta P_y}{I_z - I_y}\right)\right]^{1/2}$$

where changes in the ***P*** tensor elements can be defined as:

$$\Delta P_x = \frac{1}{2}\left(-\Delta I_x + \Delta I_y + \Delta I_z\right) \qquad \Delta P_y = \frac{1}{2}\left(-\Delta I_y + \Delta I_x + \Delta I_z\right)$$

$$\Delta P_z = \frac{1}{2}\left(-\Delta I_z + \Delta I_y + \Delta I_x\right)$$

and the reduced mass for the isotopic substitution is:

$$\mu = \frac{M\Delta m}{M + \Delta m}$$

In this formula, *M* represents the total mass of the molecule and Δm is the variation in mass of the isotopic substitution. If one atom is very close to an inertial axis, this coordinate value is very small and this method could yield imaginary values. Anyway, this method gives in general very accurate values of the atomic disposition.

II.4. References

[1] C. Calabrese, B. Temelso, I. Usabiaga, N. A. Seifert, F. J. Basterretxea, G. Prampolini, G. C. Shields, B. H. Pate, L. Evangelisti, E. J. Cocinero, *Angew. Chemie* **2021**, *60*, 16894–16899.

[2] E. Burevschi, M. Eugenia Sanz, *Molecules* **2021**, *26*, 5162.

[3] I. León, E. R. Alonso, S. Mata, J. L. Alonso, *J. Phys. Chem. Lett.* **2021**, *12*, 6983–6987.

[4] F. Xie, M. Fusè, A. S. Hazrah, W. Jäger, V. Barone, Y. Xu, *Angew. Chemie* **2020**, *132*, 22613–22616.

[5] I. Uriarte, A. Insausti, E. J. Cocinero, A. Jabri, I. Kleiner, H. Mouhib, I. Alkorta, *J. Phys. Chem. Lett.* **2018**, *9*, 5906–5914.

[6] E. Gloaguen, B. De Courcy, J. P. Piquemal, J. Pilmé, O. Parisel, R. Pollet, H. S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier, M. Mons, *J. Am. Chem. Soc.* **2010**, *132*, 11860–11863.

[7] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.

[8] I. Peña, E. J. Cocinero, C. Cabezas, A. Lesarri, S. Mata, P. Écija, A. M. Daly, Á. Cimas, C. Bermúdez, F. J. Basterretxea, S. Blanco, J. A. Fernández, J. C. Lõpez, F. Castaño, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2013**, *52*, 11840–11845.

[9] A. D. Becke, *J. Chem. Phys.* **1993**, DOI 10.1063/1.464913.

[10] F. Xie, N. A. Seifert, M. Heger, J. Thomas, W. Jäger, Y. Xu, *Phys. Chem. Chem. Phys.* **2019**, *21*, 15408–15416.

[11] T. A. Halgren, *J. Comput. Chem.* **1999**, *20*, 720–729.

[12] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, J. R., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J Am Chem SocJ. Am. Chem. Sc* **1995**, *117*, 5179–5197.

[13] G. A. Kaminski, R. A. Friesner, J. Tirado-rives, W. L. Jorgensen, *Quantum* **2001**, *2*, 6474–6487.

[14] I. Kolossváry, G. M. Keserü, **2001**, *22*, 21–30.

[15] L. L. C. Schrodinger, *New York* **2013**.

[16] P. Pracht, F. Bohle, S. Grimme, *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

[17] P. Møller, *Eur. J. Cardio-thoracic Surg.* **1934**, *53*, 1237–1243.

[18] A. Allouche, *J. Comput. Chem.* **2012**, *32*, 174–182.

[19] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650–654.

[20] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265–3269.

[21] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571–2577.

[22] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[23] R. A. Kendall, T. H. Dunning, R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[24] D. E. Woon, T. H. Dunning, *J. Chem. Phys.* **1993**, *98*, 1358–1371.

[25] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 16, Rev. C. 01* **2016**.

[26] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, 4–9.
[27] W. C. Keene, J. N. Galloway, *Atmos. Environ.* **1984**, *18*, 2491–2497.
[28] F. Xie, N. A. Seifert, W. Jäger, Y. Xu, *Angew. Chemie - Int. Ed.* **2020**, *59*, 15703–15710.
[29] P. D. Godfrey, R. D. Brown, F. M. Rodgers, *J. Mol. Struct.* **1996**, *376*, 65–81.
[30] P. Felder, **1982**, 9–25.
[31] S. R. Domingos, C. Pérez, C. Medcraft, P. Pinacho, M. Schnell, *Phys. Chem. Chem. Phys.* **2016**, *18*, 16682–16689.
[32] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
[33] F. Weinhold, C. R. Landis, *Valency and Bonding: A Natural Bond Orbital Donor–Acceptor Perspective*, **2005**.
[34] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, C. D. Sherrill, *J. Chem. Phys.* **2014**, *140*, 094106.
[35] L. B. Favero, I. Uriarte, L. Spada, P. Écija, C. Calabrese, W. Caminati, E. J. Cocinero, *J. Phys. Chem. Lett.* **2016**, *7*, 1187–1191.
[36] J. L. Alonso, V. Vaquero, I. Peña, J. C. Lõpez, S. Mata, W. Caminati, *Angew. Chemie - Int. Ed.* **2013**, *52*, 2331–2334.
[37] E. J. Cocinero, A. Lesarri, P. Écija, Á. Cimas, B. G. Davis, F. J. Basterretxea, J. A. Fernández, F. Castaño, *J. Am. Chem. Soc.* **2013**, *135*, 2845–2852.
[38] D. Patterson, M. Schnell, J. M. Doyle, *Nature* **2013**, *497*, 475–477.
[39] Z. Su, N. Borho, Y. Xu, *J. Am. Chem. Soc.* **2006**, *128*, 17126–17131.
[40] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[41] E. J. Cocinero, A. Lesarri, P. Ecija, F. J. Basterretxea, J. U. Grabow, J. A. Fernandez, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[42] G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman, B. H. Pate, *Rev. Sci. Instrum.* **2008**, *79*, 053103.
[43] C. Pérez, S. Lobsiger, N. A. Seifert, D. P. Zaleski, B. Temelso, G. C. Shields, Z. Kisiel, B. H. Pate, *Chem. Phys. Lett.* **2013**, *571*, 1–15.
[44] J. U. Grabow, W. Stahl, H. Dreizler, *Rev. Sci. Instrum.* **1996**, *67*, 4072–4084.
[45] T. J. Balle, W. H. Flygare, *Rev. Sci. Instrum.* **1981**, *52*, 33–45.
[46] U. Andresen, H. Dreizler, J. U. Grabow, W. Stahl, *Rev. Sci. Instrum.* **1990**, *61*, 3694–3699.
[47] I. Uriarte, C. Pérez, E. Caballero-Mancebo, F. J. Basterretxea, A. Lesarri, J. A. Fernández, E. J. Cocinero, *Chem. - A Eur. J.* **2017**, *23*, 7238–7244.
[48] R. D. Suenram, F. J. Lovas, K. Matsumura, *Astrophys. J.* **1989**, *342*, L103.
[49] R. D. Suenram, F. J. Lovas, G. T. Fraser, K. Matsumura, *J. Chem. Phys.* **1990**, *92*, 4724–4733.
[50] K. D. Hensel, C. Styger, W. Jäger, A. J. Merer, M. C. L. Gerry, *J. Chem. Phys.* **1993**, *99*, 3320–3328.
[51] K. A. Walker, M. C. L. Gerry, *J. Mol. Spectrosc.* **1997**, *182*, 178–183.
[52] A. Lesarri, S. Mata, J. C. López, J. L. Alonso, *Rev. Sci. Instrum.* **2003**, *74*, 4799–4804.
[53] C. Cabezas, M. Varela, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2017**, *56*, 6420–6425.
[54] I. Peña, C. Cabezas, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2015**, *54*, 2991–2994.
[55] P. F. Bernath, *Spectra of Atoms and Molecules*, Oxford University Press, New York, Oxford, **1995**.
[56] W. Gordy, R. L. Cook, *Microwave Molecular Spectra. Von*, John Wiley & Son, New York, Chichester, Brisbane, Toronto, Singapore, **1984**.
[57] J. K. G. Watson, *J. Chem. Phys.* **1968**, *48*, 4517–4524.
[58] J. K. G. Watson, *J. Chem. Phys.* **1966**, *45*, 1360–1361.
[59] J. Kraitchman, *Am. J. Phys.* **1953**, *21*, 17–24.

Chapter III.

Laboratory observation and Astrochemical Search of Erythrulose in the Interstellar Medium

This study has been published and adapted to this chapter. Aran Insausti, Elena R. Alonso, Belen Tercero, José I. Santos, Camilla Calabrese, Natalja Vogt, Francisco Corzana, Jean Demaison, Jose Cernicharo, and Emilio J. Cocinero. Laboratory Observation of, Astrochemical Search for, and Structure of Elusive Erythrulose in the Interstellar Medium. *J. Phys. Chem. Lett. 2021, 12, 4, 1352–1359*

III.1 Introduction

Rotational spectroscopy provides the most powerful means of identifying molecules of biological interest in the interstellar medium (ISM), but despite their importance, the detection of carbohydrates has remained rather elusive. In this chapter, I present a comprehensive Fourier transform rotational spectroscopic study of elusive erythrulose, a sugar building-block likely to be present in the ISM, employing a novel method of transferring the hygroscopic oily carbohydrate into the gas phase. Searches employing the new experimental data for erythrulose, have been undertaken in different ISM regions, so far including the cold areas B1-b, TMC-1 and the envelope of Sgr B2.

Carbohydrates presence in the interstellar medium (ISM) could be connected to the appearance of life on Earth. The question of the origin of life has intrigued human beings for centuries, and nowadays is one of the popular research topics for astrochemists and astrobiologists.[1,2] The detection of over 200 molecules in the ISM suggests that prebiotic molecules are present in space. Almost half of the detected molecules contain five or more atoms - labelled by astrophysicists as 'Complex Organic Molecules' (COMs). Glycolaldehyde, a two-carbon (2C) molecule (Figure III.1) and the simplest sugar-related building block, was first detected in 2000 in the ISM[3] and with the recently improved sensitivity in the observational facilities, is now recurrently identified in ISM surveys.[4] Glyceraldehyde, a (3C) compound, and other sugar related molecules like lactaldehyde, characterized through rotational spectroscopy in the laboratory have also been sought in the ISM.[5,6] The corresponding (3C) keto-sugar, 1,3-dihydroxyacetone (1,3-DHA, Figure 1), has been spectroscopically characterised and detected in the Murchison and Murray meteorites[7] together with (3C) to (6C) sugar alcohols, sugar acids, dicarboxylic sugar acids or deoxy sugar acids:[7] their abundances were comparable to those of amino acids.

The importance of detecting a sugar (or other building blocks, such as amino acids, nucleobases or fatty acids) vital to all known lifeforms in the ISM is helpful to model the reactions and formation pathways that could occur in the different regions of the ISM that lead to the formation of more complex COMs. Detection of molecules in the ISM is not trivial since their rotational spectra are extremely congested or contaminated with "astrophysical weeds" associated with a multiplicity of compounds. Laboratory rotational spectroscopy plays a crucial role in providing the reference data needed to establish molecular identification in the ISM.[8]

Unfortunately, despite their small size and apparent simplicity there are almost no rotational spectroscopic data for sugars or their smaller building-blocks, which prevents their possible subsequent astrophysical detection. The main reason for this is the difficulty of bringing unaltered sugars into the gas phase because they decompose into caramel when heated but in 2012, the problem was

resolved by coupling an ultrafast UV laser source to the microwave spectrometer which allowed vaporization of intact sugars.[9] A similar approach had previously been successful with other thermo-labile biomolecules, such as amino acids.[10]

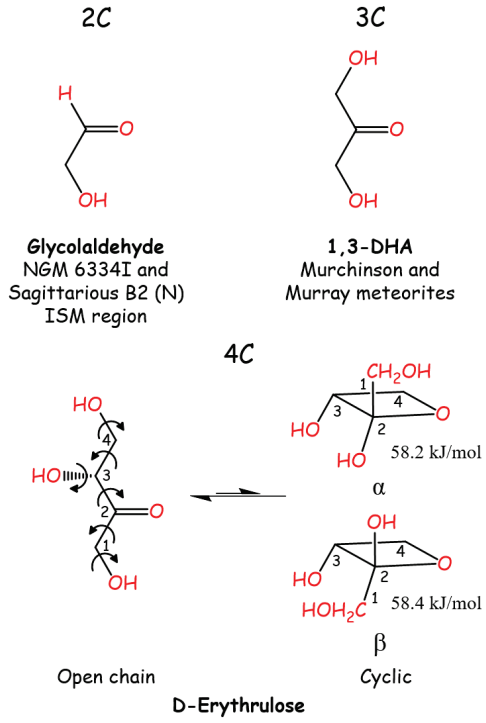


Figure III.1 Sequence of building block sugars with astro(bio)chemical interest in increasing order of carbon units. Glycolaldehyde (up left), 1,3-dihydroxyacetone (1,3-DHA, up right) and erythrulose (down). Glycolaldehyde and 1,3-DHA have already been detected in ISM and meteorites, respectively. Erythrulose is also shown as one of the following possible targets to be identified in the ISM. Erythrulose only allows the formation of four-membered rings which are very high in energy compared to linear chains. Hence, structural polymorphism is reduced to linear forms exclusively.

To begin to remedy the lack of rotational data for sugars, we present herein a comprehensive study of erythrulose, from its observation and characterization in the laboratory to its search in different regions of the ISM. Our strategy exploits a new procedure for sample preparation, rotational spectroscopic data, accurate structural determination and radio-astronomical observations in the ISM. Exploiting the high resolution achieved in the experiment, the isolated experimental equilibrium structure of erythrulose has been determined with high precision.

Erythrulose (1,3,4-trihydroxybutan-2-one, $C_4H_8O_4$, Figure III.1) has an oily appearance. So, our first attempts were to study it by heating in the CP-FTMW.

However, it did not reveal any lines attributable to erythrulose, presumably because the sample had decomposed. To solve this problem, we decided to use the ultrafast laser vaporization technique as in the earlier studies of (5C) and (6C) sugars.[9,11–16] However, the production of the solid rods, required for this approach, was not straightforward since the compound was a viscous syrup. Neither were the alternative methods described in the literature, (such as mixing the sample with NaCl (in (4C)-erythrose)[17] or glycine (in scopine)[18] and then compacting and drying it to create solid rods that would be subsequently ablated by laser radiation) successful. After testing various substances as solidifying agents (*i.e.* NaCl, glycine, glue (loctite), talcum powder, anhydrous magnesium sulphate and silica gel), using different proportions (about 25, 50 and 75%) and different drying times, the best results were obtained with a mix of talcum powder and erythrulose in a 50:50 (w/w) ratio with drying times of 2-3 weeks in the vacuum desiccator. Although the rod was not completely solid, it had enough consistency and quality to be able to perform the rotational spectroscopy experiments. As well as these difficulties, solid-state NMR measurements of the sample used in the study showed a ratio keto:hydrated form of ~92:8, which complicated both the detection and the analysis.

III.2 Methods

Experimental details

The rotational spectrum of erythrulose was recorded by means of homemade Molecular-Beam Fourier-transform microwave-spectrometer (cav-FTMW) at the Universidad del País Vasco (all details in Chapter 2).[19] The purchased sample of erythrulose presents an oleaginous appearance, hindering the process of making a solid rod for laser vaporization. Hence, we designed a new protocol to prepare the sample (see above). These samples were intact transferred to the gas phase by laser beam at 7-8 mJ pulse^{-1} from the third harmonic (355 nm) of a picosecond Nd:YAG laser. The vaporized sample was mixed with carrier gas (Ar, 6-8 bar), and the molecules presented in the resulting flow were supersonically expanded into the vacuum resonator chamber through a small orifice (1 mm) to achieve low rotational temperatures (T<5 K). After expansion, molecules in the jet absorb low power (0.1-100 mW) microwave pulses of 0.2-0.5 μs duration, which cause a spontaneous emission of rotational levels that are detected in the time domain and Fourier transformed in the frequency domain. The accuracy of the frequency measurements is better than 3 kHz, and spectral lines separated by less than about 10 kHz can be resolved.

Semiexperimental equilibrium geometry

Briefly, we combined rotational data with results of high-level *ab initio* computations (see Chapter 2 of this thesis for more details) whereby equilibrium rotational constants are derived from experimental ground state rotational

constants and rovibrational corrections (see Table AIII.1 in the appendix) resulting from a quantum chemical MP2/cc-pVTZ cubic force field.[20] As the number of experimental rotational constants is not large enough to determine a complete structure, the semiexperimental predicate-regression mixed-estimation method was used where no parameters are fixed.[21,22] This approach is relatively easy because quantum chemical calculations at a medium level of theory permit determining the internal coordinates of many light atoms with a reasonable accuracy.

Computational details

Computational simulations were used as a tool to guide the searches of rotational spectrum of erythrulose. Hence, the conformational panorama of linear and cycled forms was explored. A first energy screening was done using fast molecular mechanic methods (force fields: MMFFs, OPLS and AMBER) as a previous step to the geometry optimization[23] by DFT (B3LYP-D3BJ)[24–26] and *ab initio* (MP2)[27] quantum mechanics calculations, both with the Pople split-valence triple-zeta basis set augmented with diffuse and polarization functions on all atoms (6-311++G(d,p) basis set).[28] The rotational parameters determined by computations methods are given in Tables AIII.6-AIII.7 in the appendix.

III.3 Results and discussion

The (5*C*) and (6*C*) sugar families, key components of living systems, have been the focus of several spectroscopic studies[9,11–16] although as was mentioned earlier they were delayed by the experimental difficulty in bringing intact samples of thermo-labile molecules into the gas phase (first study in 2012),[9] for example, the amino acid family (glycine was first reported in 1978[29]) and regularly studied by Alonso's group from 2002.[10,30] To date, only one (4*C*) sugar, erythrose, has been reported in the literature despite its smaller size and apparent greater chemical simplicity.[17] The main reason for the absence of information on 4*C* sugars is the high hygroscopicity compared to the (5*C*) and (6*C*) sugars: this feature, together with the fact that the sugars tend to decompose when heated,[31] makes their study much more challenging.

Erythrulose (1,3,4-trihydroxybutan-2-one, $C_4H_8O_4$, Figure III.1) a keto-tetrose, is a (4*C*) monosaccharide with a carbonyl group in the C2 position. It has important industrial applications, particularly in the field of self-tanning cosmetics:[32] recently, it has been identified as one of the main degradation products of Maillard degradation pathways of ascorbic acid (vitamin C).[33] More importantly, in the present context, the monosaccharide is an excellent candidate, and a logical next step, for potential identification in the ISM, following the detection of, and searches for, (2*C*) and (3*C*) sugars (Figure 1). Initial attempts to study the (oily) liquid sample by heating in our home-built chirped-pulse Fourier-transform microwave apparatus (cav-FTMW) at monitored temperatures

(20-140 ºC) were unsuccessful -presumably because the sample had decomposed. To avoid this problem, samples were vaporized with an ultrafast laser as in the earlier studies of (5*C*) and (6*C*) sugars.[9,11–16] The production of the solid rods, required for this approach, was not an easy task since the compound was a viscous syrup and the standard sample preparation procedure had to be redesigned. All the details about the new methodology of sample preparation can be found in the Methods Section. As well as these difficulties, solid-state NMR measurements of the sample used in the study showed a ratio keto:hydrated form of ~92:8, which complicated both the detection and the analysis.

Experimental observation of isolated erythrulose by means of rotational spectroscopy.

Once the problem of sample preparation was solved, an exhaustive and updated[34] analysis of the conformational panorama of erythrulose was performed theoretically in a two-step (1st MM + 2nd QM) strategy to guide the experimental rotational work (see Computational Method in Chapter 2 of this Thesis). The cyclic forms were also considered in our computational study, although the calculations clearly indicated that the acyclic conformer(s) would be the only ones present in erythrulose, due to the energy difference with the cyclic forms of 60 kJ mol^{-1} (Figure III.1).

The rotational spectrum of unbiased erythrulose in the gas phase was recorded in a home-built Molecular Beam Fourier-transform microwave-spectrometer (cav-FTMW)[19] equipped with an UV ultrafast laser vaporization system (the set-up details are in the Experimental method section of Chapter 2).[9] The *UV* (355 nm), short (~40 ps) and highly energetic (7-8 mJ pulse^{-1}) laser radiation vaporizes the *intact* sample, avoiding the decomposition that attends conventional heating methods.[31] Through a process of several scans in the spectral regions where the most intense transitions were predicted, a set of rotational transitions (Table AIII.1 in the appendix, Figure III.2) could be identified: they were fitted to a semirigid rotor Hamiltonian (Watson's *S*-reduction in the *I^r*-representation[35,36]), yielding the rotational parameters shown in Table 1 - a fingerprint of the molecule providing its molecular structure. Comparing the experimental and theoretical values (Table III.1), it was unequivocally concluded that the observed structure was ***I*-1**, the main conformer of the open chain form predicted as the global minimum energy of the molecule. No more transitions corresponding to a second conformer were found in the spectrum.

Table III.1 Experimental and Theoretical Rotational Parameters of the Observed Conformer of Erythrulose and its monosubstituted ^{13}C isotopologues.

	***I*-1**[d]	**Parent**	**$^{13}C1$-$^{12}C_3H_8O_4$**	**$^{13}C2$-$^{12}C_3H_8O_4$**	**$^{13}C3$-$^{12}C_3H_8O_4$**	**$^{13}C4$-$^{12}C_3H_8O_4$**
A[a]	3135	3133.10679(27)	3119.677(10)	3130.2228(82)	3121.348(17)	3110.608(16)
B	1410	1422.65810(10)	1413.06870(26)	1421.56732(22)	1417.83894(45)	1412.17541(39)
C	1143	1157.320960(77)	1150.01309(19)	1156.25737(15)	1155.08211(31)	1148.45454(29)
D_J		0.2699(11)	[0.2699][e]	[0.2699]	[0.2699]	[0.2699]
D_{JK}		0.2245(47)	[0.2245]	[0.2245]	[0.2245]	[0.2245]
D_K		1.225(27)	[1.225]	[1.225]	[1.225]	[1.225]
d_1		-69.30(81)	[-69.30]	[-69.30]	[-69.30]	[-69.30]
d_2		-14.89(29)	[-14.89]	[-14.89]	[-14.89]	[-14.89]
N_{lines}[b]		45	9	9	9	9
σ_{fit}[c]		1.1	1.6	1.2	2.5	2.4

[a] Experimental rotational constants (in MHz) and centrifugal distortion constant (D_J, D_{JK}, and D_K in kHz and d_1 and d_2 in Hz).
[b] Number of fitted transitions.
[c] MW *root-mean-square* deviation of the fit in kHz.
[d] Equilibrium values predicted at B3LYP-D3BJ / 6-311++G(d,p) at level of theory.
[e] Centrifugal distortion constants were fixed to the values of the parent species.

Although the sample contained < 40% of erythrulose, a high degree of sensitivity was achieved, allowing the observation of ^{13}C isotopologues in natural abundance (~1.1%) and the rotational spectra of all the monosubstituted $^{13}C^{12}C_3H_8O_4$ species could be detected for ***I*-1**. The same fitting procedure used for the parent species was followed for the four additional isotopologues. The rotational parameters are given in Table III.1. The measured experimental frequencies for all the species are shown in Tables AIII.1-AIII.5. Due to the hygroscopic character of erythrulose, further efforts were performed in order to try to observe its clusters with a water molecule, the first step in the solvation but no lines were found in our spectra records.

Erythrulose adopts an open chain with up to six possible internal torsions, leading presumably to a rich conformational landscape, However, only the most stable conformer ***I*-1** could be observed. This conformer is stabilized with three intramolecular hydrogen bonds (2x O-H···O=C and 1x O-H···O-H). According to the predicted conformational landscape at B3LYP-D3BJ / 6-311++G(d,p), the second most stable conformer ***I*-2** lies at 5.2 kJmol^{-1} above the one observed. The main structural difference between them is the disposition of the hydroxyl group -O4H respect to the -O3H, with a g- disposition in ***I*-1** and g+ in ***I*-2** (∠O4-C4-C3-O3 of ~-65º (***I*-1**) and ~+55º (***I*-2**)). The interconversion barrier from ***I*-2** to ***I*-1** is predicted to ~20 kJmol^{-1}, not being possible any relaxation process, but the combination of the higher instability and the low values of the dipole moment components of ***I*-2**, explain their non-observation. The next conformers (***I*-3**, ***I*-4** and ***I*-5**) are predicted higher in energy (>6 kJmol^{-1}) due to presence of only two hydrogen bonds (HBs hereafter) against the three HBs that stabilize the observed

conformer. Figures and spectroscopic parameters of ***l*-2** to ***l*-5** are shown in the appendix.

Determination of the *unbiased* equilibrium structure of Erythrulose with high accuracy.

The high sensitivity of our cav-FTMW spectrometer which allows observation of the ^{13}C-isotopologues species in natural abundance (1.1%) also enabled the determination, for a 4*C*-sugar, of the precise equilibrium structure of erythrulose. In contrast to monosaccharides bearing four or more carbons, which can adopt favoured 5- or 6-membered rings, erythrulose adopts only a linear form.[37]. In this case, the formation of a 4-membered ring would provoke a large angular tension (Figure III.1). This work represents the example in which a sugar is isolated and observed in its linear form.

Most of the current structural knowledge of carbohydrates is based on condensed phases, using techniques such as X-ray crystallography or NMR. However, it is important to note that all these structures can be biased by the solvent, matrix, crystal or packing effects. Theoretical chemistry is also contributing significantly to obtaining sugar structures *in silico*[38–42] but despite the enormous advances in computational chemistry, the need of experimental data to test the validity of the theoretical results is still required. Gas phase experiments based on infra-red laser techniques provided early experimental data on isolated carbohydrates,[43–48] subsequently followed by other groups using alternative methodologies, to obtain valuable conformational information.[49–54] They all required interpretation through theoretical calculations and, consequently, none of them were able to determine accurate experimental structures.

In this context, rotational spectroscopy emerges as a unique tool because of its unrivalled resolution (~kHz), its ability to discriminate unambiguously between conformers, tautomers and isotopologues but its greatest advantage arises when isotopologues data are also available allowing the sets of rotational parameters to provide a pathway to the accurate, isolated molecular structure. In this work, we used the semi-experimental equilibrium geometry (r_e^{SE}), which is probably the most sophisticated method of determining structures at present. It combines information from quantum chemical calculations with experimental data, whereby equilibrium rotational constants are derived from experimental ground state rotational constants and rovibrational corrections, resulting from a quantum chemical cubic force field. However, its application has been very restricted for molecules of this size and there are very few cases reported in the literature to date.[55–57] There are three main reasons for this. Firstly, the evaluation of these molecules is not trivial; secondly, experimental isotopologue data are required, which is not straightforward for a molecule of this size. Thirdly, the high-level of calculations needed are very time-consuming. A brief description of the methodology used in this work is given in the Methods section and the full

procedure can be found in the appendix. Figure III.3 shows the most relevant structural parameters and full structure is given in the Table AIII.8 of appendix. The results of the fit are given in Table AIII.10 and the derived Cartesian coordinates in Table AIII.11 of the appendix.

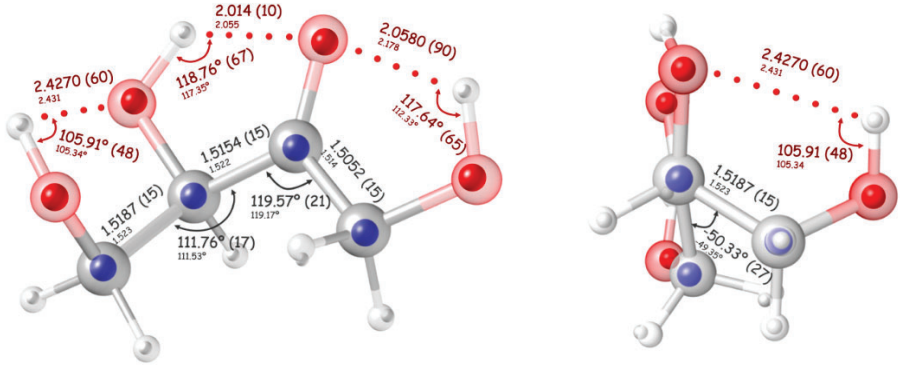


Figure III.2 Two views of observed experimental structure of D-erythrulose. Superposition of the semiexperimental equilibrium geometry (r_e^{SE}, small dark spheres) and the MP2/6-311++G(d,p) structure (larger transparent atoms) of ***I*-1** erythrulose. Some relevant structural parameters such as hydrogen bonds (in Å), bond lengths (in Å) and angles (in degrees) are also indicated for both experimental and theoretical equilibrium geometries.

The structure ***I*-1** erythrulose is stabilized by three moderate HBs, two O-H···O=C [bond lengths: 2.014(10) and 2.058(9) Å; angles: 118.76(67) and 117.64(65)º) and an O-H···O-H (bond length: 2.427(6) Å; angle: 105.91(48)º]. Similar behaviour to that observed in glyceraldehyde[58] and dihydroxyacetone[59] where two moderate HBs O-H···O=C [2.019/2.346 and 2x2.060(14) Å; angles: 119.8/108.7 and 2x115.3(9)º] were observed in each system, respectively. The side chain adopts a twisted (gauche) conformation in contrast to the extended hydrocarbon chain to encourage the formation of HBs (∠C1-C2-C3-C4 = -50.33(27)º) with eclipsed conformers.

Astronomical search of Erythrulose in cold regions of the ISM: B1-b, TMC-1 and Sgr B2.

Once the spectral frequencies of erythrulose were characterized in the laboratory, a search was begun for the molecule in space, focusing on cold regions of the ISM to prevent the high values of the partition function associated with a molecule of this size, which would result in a significant reduction in line intensities. Particular targets included the cold dark cloud Barnard 1 (B1-b) in the first stages of low-mass star formation, the pre-stellar core TMC-1 and the cold envelope towards the active high-mass star forming region Sagittarius B2 (Sgr B2). The searches were limited to low energy lines (E_{upp} < 40 K) of *a*-type or *c*-

type transitions where the frequency uncertainties were below 1 MHz, exploring the IRAM 30 m data of B1-b and TMC-1 at 3 mm[60,61] and the public data of the PRebiotic Interstellar MOlecule Survey[1] (PRIMOS)[62] towards Sgr B2 collected with the 100 m Green Bank Telescope (GBT). The MADEX code[63] was used to exploit the spectroscopic parameters presented in this work and to derive the synthetic spectrum of the molecule (assuming local thermodynamic equilibrium) according to the typical physical parameters of these sources listed in Table III.2.[61,62] The column density was the only free parameter for these models.

Table III.2. Parameters used in the searches of Erythrulose in the ISM.

Source	HPBW[a] [´´]	Frequency[b] [GHz]	V_{LSR}[c] [km s^{-1}]	$\Delta\nu_{FWHM}$[d] [km s^{-1}]	D_{sou}[e] [´´]	T_{rot}[f] [K]	N × 10^{13}[g] [cm^{-2}]
B1-b	30-21	82.5-115.5	6.7	0.7	60	12	≤ (1.0 ± 0.3)
TMC-1	30-21	82.5-115.5	6.0	0.7	60	10	≤ (1.0 ± 0.3)
Sgr B2	80-15	7-50	64	25	60	10	≤ (50 ± 20)

[a] Half power beam width (HPBW) of the telescope at the considered frequencies.
[b] Range of frequencies considered in the analysis.
[c] Radial velocity of the source with respect to the Local Standard of Rest (LSR).
[d] Line width at full width at half maximum (FWHM).
[e] Adopted source size.
[f] Rotational (excitation) temperature.
[g] Derived upper limit to the erythrulose column density.

Figure III.2 shows selected frequencies of these data together with the model derived by MADEX. It demonstrates the absence of erythrulose signals above the noise level of these data. The lines depicted in Figure III.3 are those which present the highest intensity in our models of the different sources for the frequency range available in the dataset. The upper limits to the erythrulose column density in these sources are shown in Table III.2.

[1] Access to the entire PRIMOS data set, specifics on the observing strategy, and overall frequency coverage information is available at http://www.cv.nrao.edu/~aremijan/PRIMOS/

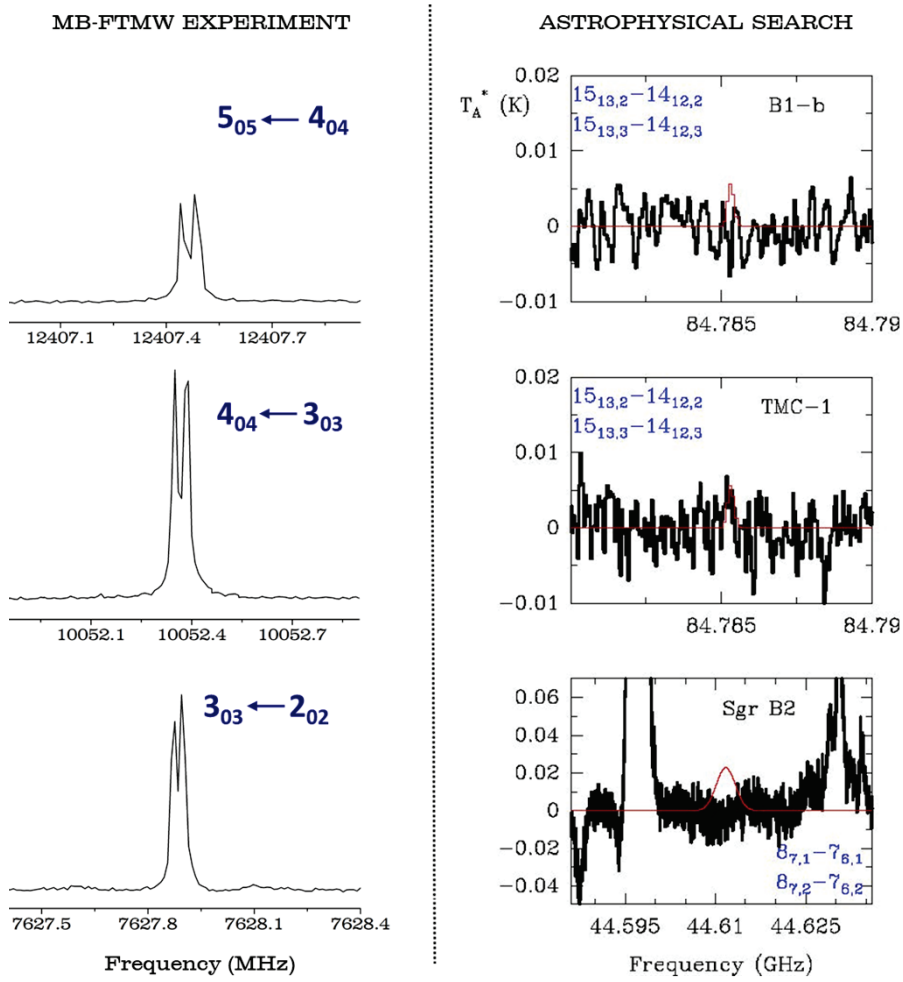


Figure III.3 (Left side) *a*-type progression ($J{+}1_{0,J+1} \leftarrow J_{0,J}$ with J = 2, 3, 4) registered in the laboratory with a homemade cav-FTMW spectrometer equipped with an UV ultrafast vaporization system. All rotational transitions are split by the instrumental Doppler effect. (Right side) Astronomical observations of B1-b, TMC-1 and Sgr B2 taken with different instruments (histogram black spectra) at selected frequencies together with the synthetic spectra obtained using the column densities given as upper limits in Table 2 (red profile). A V_{LSR} of +6.7, +6.0, and +64 km/s is assumed for B1-b, TMC-1 and Sgr B2, respectively.

III.4 Conclusions

Given the essential role that sugars play in life on Earth, any appearance in an environment unrelated to our planet is very significant. In the last decade, with the improvements in the astrophysical observatories, it has been possible to detect relevant prebiotic molecules, including amino acids and sugars precursors in the ISM, mainly located in the regions of stellar dust between the stars.[64] In this context, spectroscopic characterization of erythrulose in the laboratory and the accurate determination of its rotational parameters would provide a key, opening the door to the possible detection of this (4*C*) sugar in different regions of the interstellar medium.

The rotational spectrum of elusive erythrulose, recorded using a home-built Fourier-transform microwave-spectrometer, was facilitated by a novel sample preparation technique and the provision of an ultra-fast laser vaporization source, which will enable the future study of other oily-like molecules. Its sensitivity has allowed the detection of mono-substituted ^{13}C isotopologues in natural abundance (~1.1%) and the accurate determination of the equilibrium structure of isolated erythrulose (through its semi-experimental equilibrium geometry parameters, r_e^{SE}). The molecule adopts an open-chain form, observed for the first time in a sugar, stabilised by three HBs.

The new laboratory data have been implemented in the MADEX code to obtain simulated spectra in the spectral range appropriate to the observational data obtained for three ISM sources: the cold dark cloud Barnard 1 (B1-b), the pre-stellar core (TMC-1) and the cold envelope towards the active high-mass star forming region Sagittarius B2 (Sgr B2). Although the comparison of the observational surveys with the simulated rotational spectra did not lead to a positive result, the new experiment and reference data should encourage the search for, and future identification of, this and other sugar molecules present in interstellar molecular clouds.

III.5 References

[1] E. Herbst, E. F. van Dishoeck, *Annu. Rev. Astron. Astrophys.* **2009**, *47*, 427–480.

[2] P. Ehrenfreund, J. Cami, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002097.

[3] J. M. Hollis, F. J. Lovas, P. R. Jewell, *Astrophys. J.* **2000**, *540*, L107–L110.

[4] B. A. McGuire, C. L. Brogan, T. R. Hunter, A. J. Remijan, G. A. Blake, A. M. Burkhardt, P. B. Carroll, E. F. van Dishoeck, R. T. Garrod, H. Linnartz, C. N. Shingledecker, E. R. Willis, *Astrophys. J.* **2018**, *863*, L35.

[5] E. R. Alonso, B. A. McGuire, L. Kolesniková, P. B. Carroll, I. León, C. L. Brogan, T. R. Hunter, J.-C. Guillemin, J. L. Alonso, *Astrophys. J.* **2019**, *883*, 18.

[6] J. M. Hollis, P. R. Jewell, F. J. Lovas, A. Remijan, H. Møllendal, *Astrophys. J.* **2004**, *610*, L21–L24.

[7] G. Cooper, N. Kimmich, W. Belisle, J. Sarinana, K. Brabham, L. Garrel, *Nature* **2001**, *414*, 879–883.

[8] E. R. Alonso, L. Kolesniková, E. Białkowska-Jaworska, Z. Kisiel, I. León, J.-C. Guillemin, J. L. Alonso, *Astrophys. J.* **2018**, *861*, 70.

[9] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.

[10] A. Lesarri, S. Mata, E. J. Cocinero, S. Blanco, J. C. López, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2002**, *41*, 4673–4676.

[11] E. R. Alonso, I. Peña, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2016**, *7*, 845–850.

[12] I. Peña, E. J. Cocinero, C. Cabezas, A. Lesarri, S. Mata, P. Écija, A. M. Daly, Á. Cimas, C. Bermúdez, F. J. Basterretxea, S. Blanco, J. A. Fernández, J. C. Lõpez, F. Castaño, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2013**, *52*, 11840–11845.

[13] E. J. Cocinero, A. Lesarri, P. Écija, Á. Cimas, B. G. Davis, F. J. Basterretxea, J. A. Fernández, F. Castaño, *J. Am. Chem. Soc.* **2013**, *135*, 2845–2852.

[14] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.

[15] C. Calabrese, I. Uriarte, A. Insausti, M. Vallejo-López, F. J. Basterretxea, S. A. Cochrane, B. G. Davis, F. Corzana, E. J. Cocinero, *ACS Cent. Sci.* **2020**, *6*, 293–303.

[16] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.

[17] C. Cabezas, I. Peña, A. M. Daly, J. L. Alonso, *Chem. Commun.* **2013**, *49*, 10826–10828.

[18] P. Écija, M. Vallejo-López, I. Uriarte, F. J. Basterretxea, A. Lesarri, J. A. Fernández, E. J. Cocinero, *ChemPhysChem* **2016**, *17*, 3030–3034.

[19] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.

[20] J. Vázquez, J. Stanton, in *Equilib. Mol. Struct.* (Eds.: J. Demaison, J.E. Boggs, A.G. Csaszar), CRC Press, **2010**, pp. 53–87.

[21] L. S. Bartell, D. J. Romanesko, T. C. Wong, *Chemical Society Specialist Periodical Report No. 20, Vol. 3*, The Chemical Society, London, **1975**.

[22] S. J. MILLER, in *Probab. Lifesaver* (Eds.: J. Demaison, J.E. Boggs, A.G. Császár), CRC Press, **2019**, pp. 625–635.

[23] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R.

Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09, Revis. D.01, Gaussian, Inc., Wallingford CT* **2009**.

[24] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[25] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[26] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[27] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.

[28] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265–3269.

[29] R. D. Brown, P. D. Godfrey, J. W. V. Storey, M. P. Bassez, *J. Chem. Soc. Chem. Commun.* **1978**, 547–548.

[30] J. L. Alonso, J. C. López, in *Top. Curr. Chem.*, Springer Verlag, **2015**, pp. 335–402.

[31] R. A. Motiyenko, E. A. Alekseev, S. F. Dyubko, F. J. Lovas, *J. Mol. Spectrosc.* **2006**, *240*, 93–101.

[32] A. N. Simonov, L. G. Matvienko, O. P. Pestunova, V. N. Parmon, N. A. Komandrova, V. A. Denisenko, V. E. Vas'kovskii, in *Kinet. Catal.*, Springer, **2007**, pp. 550–555.

[33] M. Smuda, M. A. Glomb, *Angew. Chemie - Int. Ed.* **2013**, *52*, 4887–4891.

[34] J. R. Aviles-Moreno, T. R. Huet, *J. Mol. Struct. THEOCHEM* **2008**, *858*, 113–119.

[35] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.

[36] R. K. Thomas, *Phys. Bull.* **1975**, *26*, 501–501.

[37] W. Zhang, A. S. Serianni, *J. Am. Chem. Soc.* **2012**, *134*, 11511–11524.

[38] R. J. Woods, *Chem. Rev.* **2018**, *118*, 8005–8024.

[39] S. Pérez, T. Tubiana, A. Imberty, M. Baaden, *Glycobiology* **2015**, *25*, 483–491.

[40] I. Cumpstey, *Org. Biomol. Chem.* **2012**, *10*, 2503–2508.

[41] I. Alkorta, P. L. A. Popelier, *Carbohydr. Res.* **2011**, *18*, 2933–2939.

[42] B. Brauer, M. Pincu, V. Buch, I. Bar, J. P. Simons, R. B. Gerber, *J. Phys. Chem. A* **2011**, 115, 5859–5872.

[43] R. A. Jockusch, R. T. Kroemer, F. O. Talbot, J. P. Simons, *J. Phys. Chem. A* **2003**, *107*, 10725–10732.

[44] R. A. Jockusch, F. O. Talbot, J. P. Simons, *Phys. Chem. Chem. Phys.* **2003**, *5*, 1502–1507.

[45] C. S. Barry, E. J. Cocinero, P. Çarçabal, D. P. Gamblin, E. C. Stanca-Kaposta, S. M. Remmert, M. C. Fernández-Alonso, S. Rudić, J. P. Simons, B. G. Davis, *J. Am. Chem. Soc.* **2013**, *135*, 16895–16903.

[46] E. J. Cocinero, D. P. Gamblin, B. G. Davis, J. P. Simons, *J. Am. Chem. Soc.* **2009**, *131*, 11117–11123.

[47] E. J. Cocinero, P. Carcabal, T. D. Vaden, J. P. Simons, B. G. Davis, *Nature* **2011**,

469, 76–80.
[48] P. Çarçabal, E. J. Cocinero, J. P. Simons, *Chem. Sci.* **2013**, *4*, 1830–1836.
[49] A. M. Rijs, J. Oomens, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, **2015**.
[50] E. Mucha, M. Marianski, F. F. Xu, D. A. Thomas, G. Meijer, G. von Helden, P. H. Seeberger, K. Pagel, *Nat. Commun.* **2018**, *9*, 4174–4178.
[51] C. J. Gray, L. G. Migas, P. E. Barran, K. Pagel, P. H. Seeberger, C. E. Eyers, G. J. Boons, N. L. B. Pohl, I. Compagnon, G. Widmalm, S. L. Flitsch, *J. Am. Chem. Soc.* **2019**, *141*, 14463–14479.
[52] E. Saparbaev, V. Aladinskaia, R. Yamaletdinov, A. Y. Pereverzev, O. V. Boyarkin, *J. Phys. Chem. Lett.* **2020**, *0*, 3327–3331.
[53] E. J. Cocinero, P. Çarçabal, in *Gas-Phase IR Spectrosc. Struct. Biol. Mol.* (Eds.: A.M. Rijs, J. Oomens), Springer International Publishing, Cham, **2015**, pp. 299–333.
[54] R. Lozada-Garcia, D. Mu, M. Plazanet, P. Çarçabal, *Phys. Chem. Chem. Phys.* **2016**, *18*, 22100–22107.
[55] N. Vogt, J. Demaison, E. J. Cocinero, P. Écija, A. Lesarri, H. D. Rudolph, J. Vogt, *Phys. Chem. Chem. Phys.* **2016**, *18*, 15555–15563.
[56] N. Vogt, J. Demaison, S. V. Krasnoshchekov, N. F. Stepanov, H. D. Rudolph, *Mol. Phys.* **2017**, 942–951.
[57] M. Vallejo-López, P. Écija, N. Vogt, J. Demaison, A. Lesarri, F. J. Basterretxea, E. J. Cocinero, *Chem. - A Eur. J.* **2017**, *23*, 16412.
[58] N. Vogt, E. G. Atavin, A. N. Rykov, E. V. Popov, L. V. Vilkov, *J. Mol. Struct.* **2009**, *936*, 125–131.
[59] O. V. Dorofeeva, N. Vogt, J. Vogt, M. V. Popik, A. N. Rykov, L. V. Vilkov, *J. Phys. Chem. A* **2007**, *111*, 6434–6442.
[60] N. Marcelino, J. Cernicharo, M. Agúndez, E. Roueff, M. Gerin, J. Martín-Pintado, R. Mauersberger, C. Thum, *Astrophys. J.* **2007**, *665*, L127–L130.
[61] J. Cernicharo, N. Marcelino, E. Roueff, M. Gerin, A. Jiménez-Escobar, G. M. Muñoz Caro, *Astrophys. J. Lett.* **2012**, *759*, L43.
[62] J. L. Neill, M. T. Muckle, D. P. Zaleski, A. L. Steber, B. H. Pate, V. Lattanzi, S. Spezzano, M. C. McCarthy, A. J. Remijan, *Astrophys. J.* **2012**, *755*, 153.
[63] J. Cernicharo, in *EAS Publ. Ser.*, In ECLA-2011: Proceedings Of The European Conference On Laboratory Astrophysics, Laboratory Astrophysics And Astrochemistry In The Herschel/ALMA Era, **2012**, pp. 251–261.
[64] B. A. Mcguire, *Astrophys. J. Suppl. Ser.* **2018**, *239*, 17.

Chapter IV.

Conformational preferences of fluoro and deoxypyranoses

Manuscript for publication in preparation

IV.1 Introduction

In this Chapter, a combination of spectroscopic (Fourier Transform Microwave and Nuclear Magnetic Resonance) experiments and theoretical (quantum mechanics and molecular dynamics) calculations has been used to assess the conformational preferences of the hydroxymethyl group of several pyranose sugars in both the gas phase and in aqueous solution.

From a conformational perspective, the glycosidic linkages with two torsional degrees of freedom (Φ/Ψ) display a moderate flexibility, due to the *exo*-anomeric and steric effects.[1] However, disaccharides containing the (1→6)-*O*-glycosidic bond, contain an extra torsion angle (ω = O6–C6–C5–O5, Figure IV.1a), which significantly enhances their flexibility. Obviously, the increased dynamics has consequences in molecular recognition events, especially through the associated conformational entropy factor. Rotation around the C5–C6 bond gives rise to three possible conformations, *gg*, *gt*, and *tg* (Figure IV.1b). D-Glc and D-Man derivatives, regardless of the substitution at O6, present near equal *gg* and *gt* populations and a near complete absence of the *tg* conformation, both in solution[2–7] and in the solid state.[8,9] The propensity for the ω-torsion angle to adopt *gg/gt* conformations in D-Glc/Man has been attributed to the *gauche* effect[10] and to unfavorable 1,3-diaxial O4/O6 interactions operating in the *tg* rotamer. D-Gal moieties, in contrast, show a markedly different conformational distribution around the C5–C6 bond, where *gg* contributions are almost negligible, probably due to unfavorable 1,3-diaxial O4/O6 interactions.[4] Based on quantum mechanical (QM) methods and molecular dynamics (MD) simulations in explicit water, Woods and coworkers reported that water plays a key role in determining the conformation around ω,[11] by disrupting the intramolecular HB networks within Gal monosaccharides that stabilize the *gg* conformers. However, strikingly different results have been reported for **α-Gal**[12] and **β-Gal**[13] in the gas phase. The α-anomer[12] exhibits a major *gg* orientation (87%) with a clockwise (*c*) arrangement of the intra-molecular cooperative HB network while a dominant *gt* arrangement with a counterclockwise (*cc*) orientation of the cooperative HB network has been described for the **β-Gal** derivative.[13] Thus, the observations in solution for **Gal** are similar to that observed for **β-Gal**, but strikingly different to the behavior of **α-Gal** in the gas phase. Therefore, a systematic experimental study of representative monosaccharides and their derivatives by comparing their relative rotamer (*gg*:*gt*:*tg*) populations in the gas phase and in water is missing. Notwithstanding the enormous advances in the application of QM calculations to systems displaying many torsional degrees of freedom,[14–16] one must be aware of their limitations. The ideal scenario to study these systems is to interrogate the isolated sugars in the gas phase using experimental methods, free of any interference, and then to compare these results to those obtained in solution. In this scenario, Fourier-transform microwave

(FTMW) spectroscopy provides a unique tool to obtain an accurate view of molecular structure in the gas phase, given its unrivalled resolution (~3 kHz ≈ 10^{-7} cm^{-1}). Indeed, it has already provided valuable data for several monosaccharides,[12,14,17–22] including a nucleoside,[23] and sugar derivatives.[24] The observed structures and their conformational populations are independent of the quality of QM models; rather they serve to benchmark their quality and even to calibrate the theoretical methods.

In the present work, we have designed an experimental strategy to evaluate the role of the different factors that influence the conformational behavior around the hydroxymethyl group of pyranoses. The effect of the intrinsic intramolecular HB network within a given monosaccharide and that of the environment (gas phase *versus* solution) has been assessed. It has been inspired by the use of bio-isosteric OH-to-F substitutions in molecular recognition studies,[25,26] including the development of novel (glyco)drugs,[27–32] selective tracers for non-invasive imaging modalities as ^{19}F-MR/MRI[33,34] or positron emission tomography (^{18}F-PET).[35–38] Fluorine and hydroxyl groups are isoelectronic and possess similar polarity and steric impact.[39–41] The ability of the hydroxyl groups to act as hydrogen bond (H-bond) acceptors and donors can be chemically altered/disrupted by OH-to-F (or OH-to-H) substitution at position C2 of the ring to provide the corresponding deoxyfluoro (or deoxy) monosaccharides. Thus, several compounds derived from the representative hexopyranosyl monosaccharides D-glucose (**Glc**), D-mannose (**Man**), and D-galactose (**Gal**) have been synthesized. In the case of F-sugars, fluorine can act as a moderate

H-bond acceptor and can be involved in hydrophobic contacts[42,43] but (obviously) it cannot influence the intra-molecular HB network by acting as an H-bond donor! On the other hand, OH-to-H substitution completely removes the ability to donate or accept H-bonds.

The multidisciplinary strategy to determine their conformational preferences in the gas phase and in solution and to identify the factors that control them has involved the multi-gram synthesis of fluorinated monosaccharide mimetics (omitted in this Chapter); rotational spectroscopy in the gas phase; solid-state and solution NMR experiments; QM calculations and extensive molecular dynamics (MD) simulations.

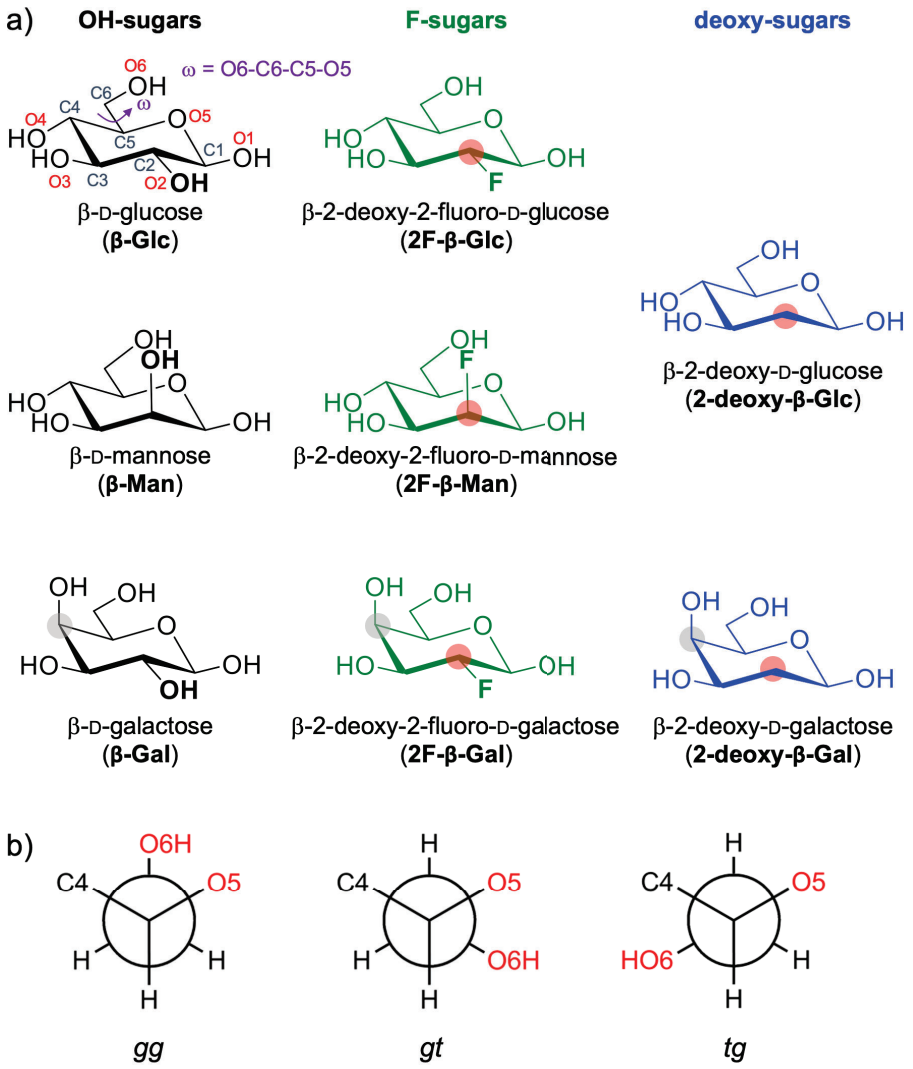


Figure IV.1 Monosaccharides and nomenclature. a) 2-hydroxy, 2-fluorodeoxy, and 2-deoxy hexopyranoses studied in this work, along with the definition of the torsion angle ω. b) Notation used for the more stable dispositions of ω, labeled as *gg* (*gauche-gauche*), *gt* (*gauche-trans*), and *tg* (*trans-gauche*). The first letter indicates the torsion angle between the endocyclic oxygen O5 and the exocyclic oxygen O6, and the second letter refers to the dihedral angle between C4 and O6.

IV.2 Methods

Computational methods

The conformational landscape of both α and β-anomers was successfully explored using an already established strategy based on fast molecular mechanics methods with a sophisticated algorithm with different force fields (MMFFs, OPLS, AMBER). The structures thus obtained were geometry-optimized using *ab initio* (MP2) and DFT (B3LYP functional) methods, including the Grimme D3 dispersion interactions with Becke−Johnson damping, respectively, as provided by the Gaussian suite of programs. In all cases, the Pople split-valence triple-ζ basis set augmented with diffuse and polarization functions on all atoms (6-311++G (d,p)) was used (Details and References in the Chapter II of this Thesis).

Rotational spectroscopy in gas phase

Besides considering the previously published data for **Glc**,[20] **Man**,[13,44] and **Gal**[12,13] analogues, new data for **2F-β-Glc**, **2F-β-Man**, **2F-β-Gal**, **2-deoxy-β-Glc**, and **2-deoxy-β-Gal** have now been acquired in our home-built cav-FTMW spectrometer.[17,45] It is important to stress that recrystallized samples of the synthesized and commercially available derivatives only contain the β-anomer (Figure IV.2a and appendix, Figures AIV.11–AIV.18). In all cases, the solid samples were vaporized with UV radiation (355 nm) from a pulsed Nd:YAG *picosecond* laser (~40 ps). This way of vaporization has been proven to be very successful to transfer "*intact"* sugars to the gas phase.[17] This step is critical due to the high hygroscopic and thermo-labile character of the molecules.[46] (see the appendix and Chapter II for a detailed explanation of the gas phase experiments).

Conformational Analysis of β-D-Hexopyranoses in the Solution

To investigate the effect of hydration in the structure of these sugars, their conformational analysis in aqueous solution was carried out by ^{1}H NMR spectroscopy measurements (homonuclear coupling constants, $^{3}J_{H,H}$) assisted by MD simulations.[47–50] For J-coupling analysis (2-deoxy- and 2-deoxy-2-fluor derivatives), NMR spectra (1H NMR, 1D-selective 1H NMR TOCSY, band selective 1H-13C HSQC) in D2O solution were acquired using 3 mM samples on an 800 MHz BRUKER AVANCE III spectrometer, equipped with a TCI cryo-probe with z-gradient coil. TopSpin 3.2.7 (BRUKER) software was employed for data acquisition and processing. Calculations were performed using AMBER 20,[51] implemented with the latest version of the GAFF force field[52] to study the conformational behavior of fluorinated and 2-deoxy variants.

IV.3 Results and Discussion

Conformational Analysis of β-D-Hexopyranoses in the Gas Phase

Figure IV.1 shows the carbohydrates synthesized and studied in this work. Interestingly, the solid samples of such fluorinated derivatives contained exclusively the β-anomer, as determined by ^{19}F and ^{13}C CP/MAS solid-state NMR (Figure IV.2a and apendix, Figures AIV.11–AIV.16.[53–58] Similarly, only the β-anomer was also found by ^{13}C CP/MAS solid-state NMR in 2-deoxy-D-glucose (**2-deoxy-Glc**) and 2-deoxy-D-galactose (**2-deoxy-Gal**), which were acquired from commercial suppliers in the required amounts (appendix, Figures AIV.17, AIV-18).

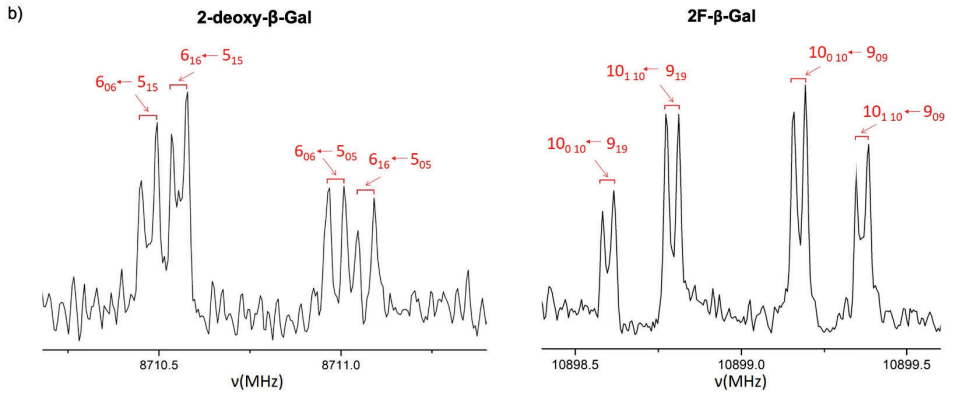


Figure IV.2 Experimental data. Typical rotational transitions of **2-deoxy-β-Gal** (left) and **2F-β-Gal** (right) hexopyranoses were observed using Fourier-transform microwave (cav-*FTMW*) spectroscopy equipped with a UV ultrafast laser vaporization system. Rotational transitions are split due to the instrumental Doppler effect.

Previous studies in the gas phase using rotational spectroscopy have been performed with **Glc**, **Man**, and **Gal**. While both anomers were observed in **Glc**,[20] only α-anomers could be detected for **Man**[59] and **Gal**.[12] It is important to note that mutarotation (α↔β interconversion) is impeded in the gas.[14] Hence, the gas phase populations of the α- or β-anomers reflect the values of the solid sample. Therefore, and due to the lack of rotational experimental data for **β-Man** and **β-Gal**, their conformational/population data were obtained from previous laser spectroscopy experiments (*ion-dip* spectroscopy) by Simons and coworkers on **β-Man-OPh**[44] and **β-Gal-OPh**[13] also in the gas phase. The obtained results for these molecules were consistent with the computationally predicted data (Table IV.2).

Table IV.1 Most relevant experimental rotational parameters of the observed conformers for the studied species. The full data can be found in the appendix (Tables AIV.15–AIV.21).

		A[a]	B	C
2F-β-Glc	**1** (*gt*/*cc*)	1174.97088(22) [b]	804.19900(29)	528.30649(47)
	2 (*gg*/*cc*)	1171.00499(44)	779.36429(19)	489.58154(18)
2F-β-Gal	**2** (*gt*/*cc*)	1269.18963(19)	755.20420(13)	524.177553(88)
	3 (*gg*/*c*)	1289.06503(18)	796.27051(12)	563.133262(85)
2F-β-Man	**1** (*gt*/*cc*)	1080.6644(11)	898.87960(43)	593.39487(22)
	2 (*gg*/*cc*)	1117.2134(99)	848.8623(69)	531.96736(55)
2-deoxy-β-Glc	**1** (gg/c)	1189.644727(86)	1074.782499(81)	632.855274(49)
	4 (*gt*/*cc*)	1239.26352(26)	1005.92933(15)	580.78082(13)
	6 (*tg*/*c*)	1362.98879(17)	950.247495(79)	586.510948(66)
2-deoxy-β-Gal	**1** (*gg*/*c*)	1293.38327(15)	1065.40506(11)	685.853728(73)
	3 (*gt*/*c*)	1381.23062(48)	935.04220(16)	633.71769(10)

[a] Experimental rotational constants A, B and C in MHz.
[b] Standard error (SE) in round brackets in units of the last digit.

As a summary of the antecedents in the gas phase, and independently of the experimental technique (see Table IV.2 for the full set of gas phase population data), the clockwise orientation of the HB network is the major one for α-**D**-anomers and is mostly correlated with a predominance of the *gg* rotamer for the hydroxymethyl group, while the counterclockwise orientation is mostly observed for β-**D** derivatives and associated to the presence of *gt* rotamers. The *tg* conformers are present in minor extent. This data for **Glc** and **Man** derivates are fully consistent with the values found both in solution[2,3,7] and in the solid state.[9] It should be noted that in solution, and due to mutarotation, the populations of α- and β-anomers cannot be separated. In contrast, as mentioned in the Introduction, the hydroxymethyl conformational behavior of **Gal** in solution (*gt/tg* equilibrium) resembles that observed for its **β-Gal-OPh**[13] derivative (*gt/tg* equilibrium with *cc*-HB network), but is strikingly different to that of α-**Gal**[12] in the gas phase (major *gg* around ω and clockwise HB network).

Several frequency scans were guided by computational predictions (appendix, Tables AIV.1–AIV.14 and Figures AIV.19–AIV.15) for each sample. The rotational spectra identified a total of eleven structures among all investigated compounds. Specifically, three structures (conformations) were detected for **2-deoxy-β-Glc** and two for each of the remaining sugars (**2F-β-Glc**, **2F-β-Man**, **2F-β-Gal**, and **2-deoxy-β-Gal**). For each conformation, the set of experimental rotational transitions (Figure IV.2b) was fitted to Watson's *S*-reduced semi-rigid rotor Hamiltonian in I^r representation yielding to rotational parameters (Table IV.1 and appendix, Tables AIV.15– AIV.21). These parameters are considered as fingerprints of the 3D-molecular structures. The fitting of the experimental data to the theoretically predicted values produced the unequivocal characterization of

the existing conformers. In particular, the species that were unambiguously identified are gathered in Figure IV.3a, labelled with consecutive numbers according to the calculated energy ranking: **2F-β-Glc-1**, **2F-β-Glc-2**; **2F-β-Man-1**, **2F-β-Man-2**; **2F-β-Gal-2**, **2F-β-Gal-3**; **2-deoxy-β-Glc-1**, **2-deoxy-β-Glc-4**, **2-deoxy-β-Glc-6** and **2-deoxy-β-Gal-1**, **2-deoxy-β-Gal-3**.

Moreover, given the high sensitivity of our FTMW spectrometer, ^{13}C-isotopomers in natural abundance (~1.1%) were also detected. Hence, the rotational spectra of nine additional mono-substituted isotopologue species ($^{13}CC_5H_{12}O_5$) were observed for the most abundant conformers of **2-deoxy-β-Gal** (6 monosubstituted ^{13}C species, appendix, Tables AIV.20, AIV34–AIV39) and **2-deoxy-β-Glc** (3 monosubstituted ^{13}C species, appendix, Tables AIV.16, AIV.25–AIV.27). These rotational data allowed determining the r_0 and r_S structures (Details in Chapter II, parameters in appendix Figure AIV.30). Importantly, as mentioned above, the resulting structures are independent of the quality or processes of the associated models. It is noteworthy to note that, in some cases, the QM results are method-dependent and therefore, inconclusive.[14,60] In contrast, FTMW spectroscopy has been proved to be a benchmarking tool for QM methods.[60]

The experimental population ratios shown in Figure 3a for the different observed conformers of **2F-β-Glc**, **2F-β-Man**, **2F-β-Gal**, **2-deoxy-β-Glc**, and **2-deoxy-β-Gal** were deduced from the analysis of the relative intensities of rotational transitions.[61] The available data for the native sugars (**β-Glc**, **β-Man**, and **β-Gal**) have been added to facilitate comparison. Interestingly, the three possible staggered conformations for the hydroxymethyl group (*gg*, *gt*, and *tg*) were observed, although with rather distinct populations. Notably, in all cases, the structures are stabilized by *cooperative* intra-molecular HB networks that can be arranged in either clockwise (*c*) or counterclockwise (*cc*) orientations. The experimentally derived population ratios are presented in Figure IV.3 and Table IV.2.

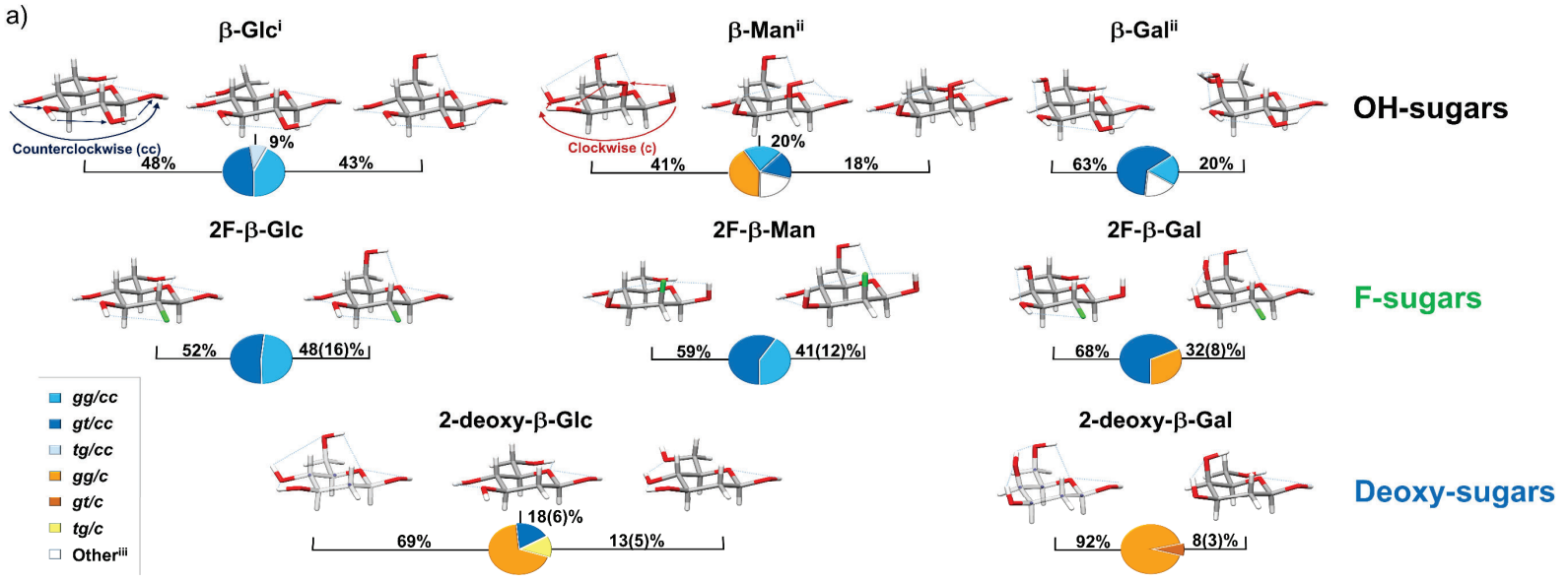


Figure IV.3 Populations and observed structures of monosaccharides in the gas phase. a) All the structures and populations of the OH-sugars, F-sugars, and deoxy-sugars. All the monosaccharides are stabilized by H-bonding (HB) cooperative networks. A structure with a counterclockwise (*cc*) HB network orientation is highlighted for **β-Glc** (blue arrow), while the alternative clockwise (*c*) orientation in **β-Man** (red arrow). *Notes:* [i] Experimental values from reference.[20] [ii] Structures and theoretical populations at MP2/6-311++G(d,p) level. These theoretical data are consistent with the experimental values observed by *ion-dip* spectroscopy for the **β-Man-OPh**[44] and **β-Gal-OPh**[13] derivatives (Table IV.2). [iii] Conformations with theoretical populations of less than 10% have been omitted from the figure for clarity, but all values can be found in Table 2. Their values have been grouped as '*other*', which would be the total contribution of the remaining conformers.

Table IV.2 Experimentally and theoretically derived percentage population ratios for the hexopyranoses in gas phase. Standard deviation (SD) is shown in round brackets.

Molecule \ Conformer	*gg/cc*	*gt/cc*	*tg/cc*	*gg/c*	*gt/c*	*tg/c*
α-Glc[20]	36	32	18	14	-	-
β-Glc[20]	43	48	9	-	-	-
β-Glc-OPh[62], [a]	25	68	7	-	-	-
2F-β-Glc This work	48(17)	52	-	-	-	-
2-deoxy-β-Glc[b], This work	-	18(6)	-	69	-	13(5)
α-Man[c], This work	11	6	2	66	4	10
α-Man-OPh[13], [a]	10	5	-	85	-	-
β-Man[c], This work	20	18	8	41[d]	9	4
β-Man-OPh[44], [a]	30	70	-	-[d]	-	-
2F-β-Man This work	41	59(8)	-	-	-	-
2-deoxy-β-Man[b], This work	-	18(6)	-	69	-	13(5)
α-Gal[12]	2	9	4	85	-	-
α-Gal-OPh[63], [a]	-	63	17	20[d]	-	-
β-Gal[c], This work	9	63	20	7	<1	<1
β-Gal-OPh[13]	-	75	20	5	-	-
2F-β-Gal This work	-	68	-	32(8)	-	-
2-deoxy-β-Gal This work	-	-	-	92	-	8(3)

[a] Experimental values for OPh derivatives based on laser spectroscopy. The estimation of the population has been determined using the signal intensity of the R2PI spectra. Although the precision of this data is not very high, the values reveal the general trend of the populations of the observed conformers.
[b] **2-deoxy-β-Glc** and **2-deoxy-β-Man** are the same molecule.
[c] Theoretical populations at MP2/6-311++G(d,p) level.
[d] The non-observation of the *gg/c* conformer for **β-Man-OPh** or the significant decline in the population in **α-Gal-OPh** analogue by laser spectroscopy is due to the lack of hydrogen atom at the O1 position. The O1H···O2 interaction is essential to stabilize the *c*-HB networks. For this reason, in native OH-sugar **β-Man** and **α-Gal**, these are the main conformers with populations 41% and 85%, respectively.

The case of native sugars

The reported experimental and theoretical data compiled for native β-sugars indicate that there are three different and well-defined behaviors; while a dominant species (*gt*/*cc*) is observed/predicted for **β-Gal**, several conformations coexist for **β-Glc** and **β-Man**. Interestingly, while all conformers in **β-Glc** exhibited a *cc*-HB pattern, the *c*-HB network was present for the major (41%) conformation (*gg/c*) of **β-Man**, although the *cc*-HB pathway is also highly populated (46%).

It is important to note, and surprising to some extent, that the population of the *tg* rotamer in the gas phase is very small for **β-Glc**[20] and basically negligible for **β-Man**, despite being stabilized by a stronger hydrogen bond than that present in the *gg* and *gt* arrangements (O6H···O4 *vs*. O6H···O5). The higher strength of

O6H···O4 HB was verified by NBO calculations (appendix, Tables AIV.42–AIV.44). Moreover, contrary to the results reported for **α-Gal**, the population of the *gg* conformation is almost negligible for **β-Gal**, where the *gt* alternative is dominant, exhibiting the O6H···O5 HB. Therefore, to explain this important finding, a systematic energy scan around ω torsion angle was carried out (Figures IV.4 and appendix, Figures AIV.32–AIV.38), by allowing or excluding the stabilizing hydrogen bonds between OH6 and O5/O4. Remarkably, the three rotamers (*gg*, *gt,* and *tg*) are far from being energetically equivalent. The *tg* conformation of **β-Glc** exhibits a strong "repulsive" energy penalty (~15 kJ mol^{-1}), which cannot be compensated by the strong hydrogen bond (O6H···O4) that can be formed in this geometry and is destabilized respect to the more stable *gg* and *gt* rotamers. Moreover, if this hydrogen bond is impeded, the *tg* conformer is not even a local minimum (Figure IV.4). The same scenario is observed for **β-Man** in the *cc*-orientation. Similarly, the *gg* conformer in the gas phase of **β-Gal** is highly destabilized, with an energy penalty (>20 kJ mol^{-1}) respect to the *tg* and *gt* rotamer when the O6H···O4 is impeded (*cc*-HB pattern, Figures 4 and SI, Figures S42–S48). Obviously, the O4H···O6 HB is much stronger in **Gal** than in **Glc**/**Man** due to the axial arrangement of the O4H.

The case of deoxyfluorinated analogues

The analysis of the experimental data for the F-sugars was then carried out. Only the *gg* (32-48%) and *gt* (52-68%) conformers with significant populations coexist, with the *cc*-HB pattern being largely predominant. The presence of the *tg* rotamer is always negligible. Fluorination destabilizes both cooperative HB networks, but especially the clockwise one. The fluorine atom cannot act as an H-donor, yet it may perform as a partial H-acceptor, but only for the counter-clockwise orientation. This is especially evident for **2F-β-Man** and **2F-β-Glc** where there is not any trace of clockwise oriented structures. The axial (**Man**) or equatorial (**Glc**) orientation of the fluorine atoms at C2 seems to be irrelevant for the outcome. Only for **2F-β-Gal,** there is certain population (32%) of the *c*-HB pathway (associated to the *gg* rotamer), but the *cc* alternative is more abundant, since the fluorine atom at C2 can act as weak HB acceptor. Thus, the *gt/cc* geometry is largely preferred (68%).

.

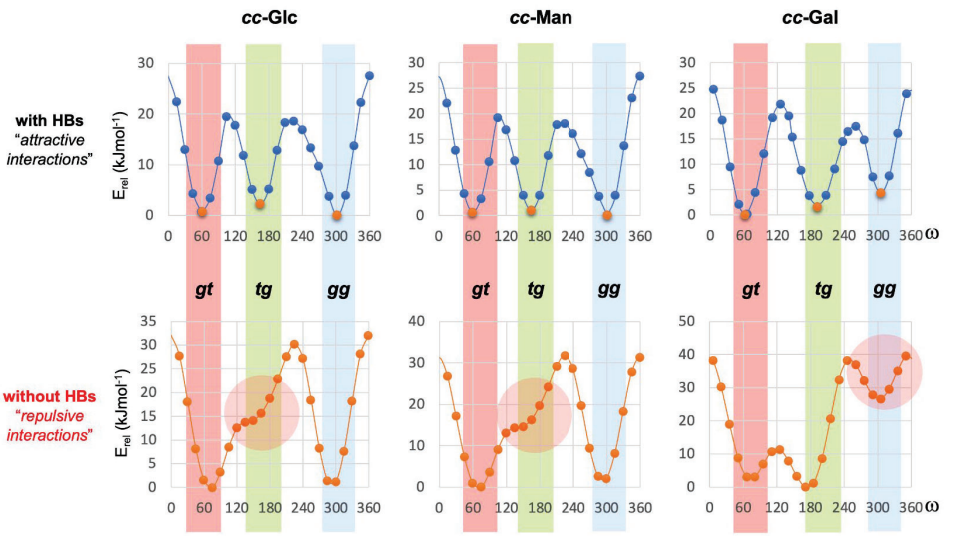


Figure IV.4 The interconversion barrier of ω angle at MP2/6-311++G(d,p) level of ***cc*-β-Glc**, ***cc*-β-Man**, and ***cc*-β-Gal**. The attractive scan has been carried out by rotating ω, while optimizing the rest of the parameters. In the case of the "repulsive" scans the H–O(6)–C(6)–C(5) has been fixed to 180º to avoid that this group is involved as hydrogen bond donor with the hydroxyl group at C(6) and the ring oxygen or the hydroxyl group of C(4)

The case of deoxy analogues

Deoxygenation eliminates any possibility of acting as an H-donor/acceptor at C2. For both **2-deoxy-Glc/Man** (69%) and **2-deoxy-Gal** (92%) analogues, the preferred choices for both cases are *gg* rotamers with *c*-HB networks.

For **2-deoxy-Glc/Man**, the HB network is rather inefficient, and now a small population (13%) of the *tg* rotamer appears. Although the *tg* rotamer of the hydroxymethyl group can be stabilized by either a *c*- or *cc*-HB network, the population of this rotamer is associated to the *c*-pathway. Thus, the O3eqH···O4eq/O4eqH···O6 HB *c*-type network is stronger than the *cc* counterpart (O6H···O4eq/O4eqH···O3eq), but not strong enough to completely overcome the O4/O6 destabilizing interaction. This fact could be explained by the unfavorable 1,3-diaxial interactions in this geometry. There is a predominance of *gg* (69%) and *c* (82%) orientations. These *c*-arrangements are favored by being able to create longer HB networks by involving endocyclic oxygen (O5) as well.

The **β-Gal** case deserves special attention. For **Gal** derivatives, the presence of the *gg* conformer is associated to the presence of an intra-molecular *c*-HB network. In fact, as mentioned above, the *gg* conformer is the most populated one in the gas phase for **α-Gal**,[12] and the *c*-type HB network stabilizes

this hydroxymethyl conformation. Obviously, the removal of the hydroxyl group at C2 destabilizes both HB networks. For **2-deoxy-β-Gal**, the O3eqH···O4ax/O4axH···O6 HB *c*-network is now preferred (*ca*. 100%) over the *cc* alternative (O6H···O4ax/O4axH···O3eq) and the *gg* rotamer increases up to 92%.

In summary, the comparison between the fluorodeoxy and deoxy analogues permits to assess the key role of the fluorine atom as a weak HB acceptor. The *cc* cooperative HB network is largely stabilized for the fluorodeoxy derivatives in comparison to the deoxy analogues and this is due to the presence of the fluorine atom.

The importance of the substitution at C1 for the existence of a predominant HB network can also be highlighted. A cooperative *c*-HB network starts with OH1 as HB donor. Indeed, **β-Glc-OPh**, **β-Man-OPh**, and **β-Gal-OPh**,[13] which lack this OH1, displayed very major or exclusive *cc*-pathways. The *c*-HB network is also negligible for native unsubstituted **β-Glc**[20] and **β-Gal**, which also adopt the *cc*-pathway. It is tempting to point out that the participation of HO1 as HB donor would require the unfavorable non-exo-anomeric orientation around the C1–O1 linkage to provide efficient donation to O2. Therefore, only the *cc*-HB network takes place. Only for **β-Man** there is certain population of the *c*-HB network. In this case, the axial orientation of O2 does require that the C1–O1 torsion adopts the non-exo-anomeric conformation. In contrast, for **α-Gal** or **α-Glc**, the axial orientation of O1 allows O1H to act as HB donor in an exo-anomeric orientation (*anti*). As experimentally demonstrated for **α-Gal**, the clockwise network from O1H is stable and the *gg* rotamer is dominant (85%).

Conformational Panorama in Aqueous Media

As expected, the analysis showed that all derivatives adopt the $^4C_1$ chair conformation in water (appendix, Figure AIV.39).[64] The analysis of the experimental $^3J_{H5,H6}$ coupling constants[64,65] showed that the hydroxymethyl group is flexible and preferentially adopts the staggered rotamers *gt* and *gg* for **β-Glc** and **β-Man** derivatives, regardless of the presence of –CHOH–, –CHF– or –CH_2– groups at C2 (Figure IV.5a and IV.5b). Indeed, the experimental $^3J_{H5,H6}$ values only change marginally upon deoxygenation or fluorination (Table IV.3).

Table IV.3 Coupling constants $^3J_{H5,H6}$ determined in aqueous solution by NMR for the different derivatives studied in this work (see the SI for experimental details) .

Hexopyranose	anomer	$J_{H5,H6R}$ (Hz)	$J_{H5,H6S}$ (Hz)
2-deoxy-2-fluoro-D-Glc*p*	α	5.3[a]	2.3[a]
(**2F-Glc**)	β	6.1[a]	2.3[a]
2-deoxy-2-fluoro-D-Man*p*	α	5.4[b], 5.7[a]	2.2[b], 2.1[a]
(**2F-Man**)	β	6.2[b], 6.7[a]	2.3[b], 2.3[a]
2-deoxy-2-fluoro-D-Gal*p*	α	8.3[b]	4.1[b]
(**2F-Gal**)	β	7.8[b]	4.5[b]
2-deoxy-D-Glc*p*	α	5.1[b]	2.3[b]
(**2-deoxy-Glc/Man**)	β	6.1[b]	2.3[b]
2-deoxy-D-Gal*p*	α	6.2[b]	6.2[b]
(**2-deoxy-Gal**)	β	7.8[b], 7.3[c,d]	4.6[b], 5.5[c,d]
D-Glc*p* (**Glc**)	α	5.8[c,e,f]	2.0,[c] 1.0,[e] 1.9[f]
	β	5.9,[c] 6.0[e,f]	2.0,[c] 2.1,[e] 1.9[f]
D-Man*p* (**Man**)	α	5.8,[g] 5.6[h]	2.0[g,h]
	β	6.4,[g] 5.7[h]	2.3[g,h]
D-Gal*p* (**Gal**)	α	6.0,[b] 6.8,[c] 7.9[f]	6.0,[c] 4.6[f]
	β	7.8[f]	4.6[f]

[a] Ref. [66]; [b] This work. Measured at 800 MHz (see the SI for experimental details). The found ABX, AX2 Spin systems were solved at: (http://anorganik.unituebingen.de/klaus/nmr/spinsystems/index.php?p=abx_ab_an) [c] Ref. [50]. [d] Apparent coupling constant. [e] Ref. [3]. [f] Ref. [67]. [g] Ref. [7]. [h] Ref.[68].

According to the employed Karplus-like equations,[64] the experimental $^3J_{H5,H6}$ coupling constants for **β-Glc**, **2F-β-Glc**, and **2-deoxy-β-Glc** correspond to a rotamer distribution for the hydroxymethyl group (*gg*:*gt*:*tg)* of ≈ 55:45:0. These values are also very similar to those observed for **β-Man** and **2F-β-Man**. The same trend was obtained for the α-anomers (appendix, Figure AIV.41) and the experimental data agrees with the results estimated by MD simulations (Figures IV.5 and appendix, Figures AIV41–AIV.43). Therefore, in aqueous solution, independently of the configuration, deoxygenation, or fluorination at C2, all **Man** and **Glc** derivatives display *ca.* ≈ 50:50 *gg/gt* conformational equilibria. In all

cases, the experimental data indicate that the *tg* conformer is basically negligible in solution, as also described above for the gas phase. Now, the analysis of all MD trajectories showed that there are no significant intra-molecular hydrogen bonds in water, with populations <10% in all cases, as expected given the presence of competitive surrounding water molecules.

For the **β-Gal** variants (Figure IV.6a and Table IV.3), the $^{3}J_{H5,H6}$ values are consistent with the occurrence of *gt* and *tg* conformers in solution[11,64,69] with a smaller *gg* population, in agreement with those obtained in the gas phase for **β-Gal**, but opposed to those reported for **α-Gal**,[12] where the *gg* was the most populated rotamer. Notably, the solvated MD simulations with the GAFF force field overestimate the population of the *gg* conformer, particularly for **2F-Gal** and **2-deoxy-Gal** derivatives, regardless the α- or β-anomer. However, the MD simulations carried out using the latest version of GLYCAM[70] for the native derivatives **α-Gal** and **β-Gal** lead to values close to the experimental ones (appendix, Figures AIV.40–AIV.43). As for **Glc** and **Man**, no relevant intramolecular HBs were detected in the simulations. In contrast, all hydroxyl groups, including O6, participate in multiple hydrogen bonds with the water molecules of the first hydration shell (Figure IV.6b). Therefore, in aqueous solution, the role of the intramolecular HB networks to rule the conformation around C5–C6 is negligible.

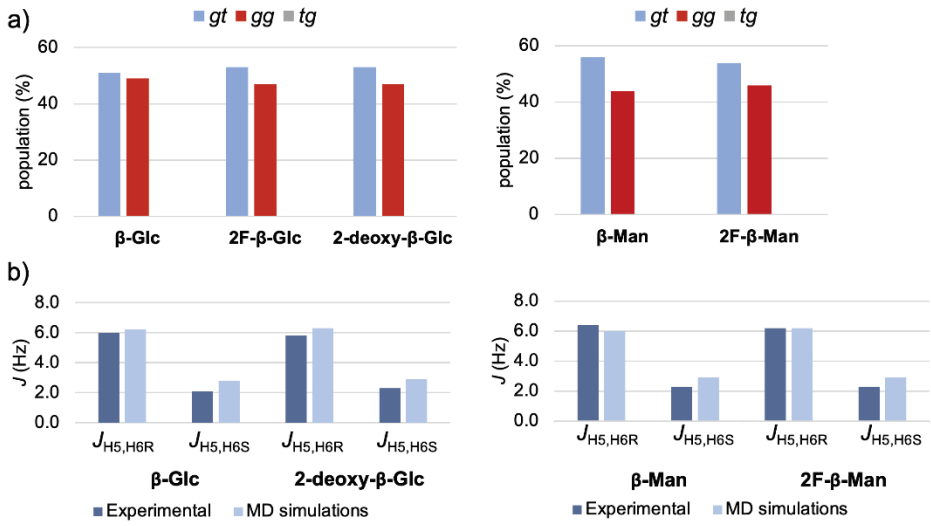


Figure IV.5 Conformational study of β-Glc and β-Man derivatives in water. a) Distribution of ω-torsion angle (O5-C5-C6-O6) for **β-Glc** and **β-Man** variants derived from NMR data (Table IV.3). b) Experimental (by ^{1}H NMR) and predicted $^{3}J_{H5,H6}$ constants for **β-Glc** and **β-Man** derivatives.

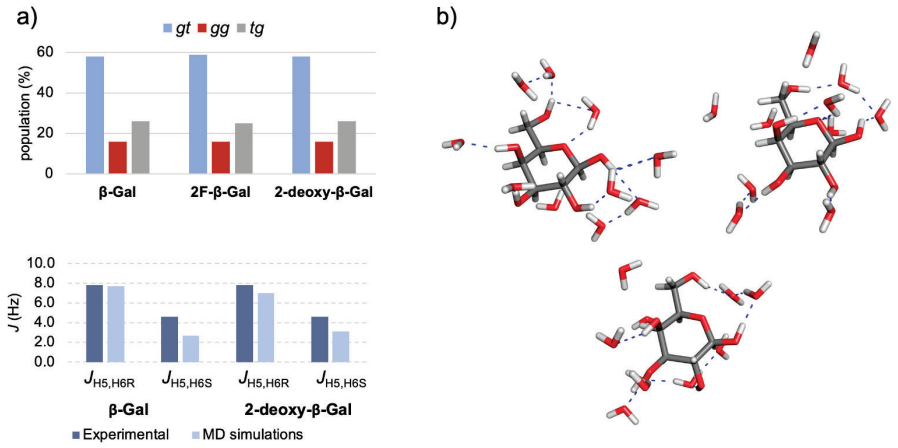


Figure IV.6 Conformational study of β-Gal derivatives in water. a) Distribution of ω-torsion angle (O5-C5-C6-O6) for **β-Gal** variants derived from experimental data (Table IV.3), together with the experimental (by ^{1}H NMR) and predicted $^{3}J_{H5,H6}$ constants for these compounds (bottom). b) Representative frames for **β-Gal**, exhibiting a *gt* conformation for ω-torsion angle, obtained from the simulations.

IV.4 Conclusions

In this work we demonstrate that in the gas-phase and aqueous solution the **β-Glc** and **β-Man** and their fluorinated and deoxy variants at position C2 show that the *gg* and *gt* rotamers dominate, while the *tg* geometry is rarely populated, suggesting that water molecules display a minor influence in the conformational preferences of the corresponding hydroxymethyl groups. A different scenario was observed for the **Gal** derivatives. While the population of the *gg* rotamer is negligible for natural **β-Gal**, both in the gas phase and in aqueous solution, the *gg* rotamers of the 2-fluoro and 2-deoxy analogues are significantly populated in the gas phase. For the gg rotamer of **Gal**, the O4H···O6 HB is significantly stronger than the O6H···O4 HB alternative. This is not the case for the *tg* rotamers in the **Glc**/**Man** analogues. This O4H···O6 HB is notably favored for the axial arrangements of O4 in **Gal** derivatives. A similar tendency is deduced for the fluorodeoxy and deoxy analogues.

Nevertheless, this *gg* rotamer disappears in the presence of water, which disrupts the clockwise intramolecular hydrogen bond network for facilitating the *gg* geometry. Fittingly, in the gas phase, *gg* also dominates (85%) in **α-Gal**, also showing the clockwise orientation for the cooperative intra-molecular HB network. This situation is also reversed in aqueous solution. Under these solvated experimental conditions, the conformation around C5–C6 of **Gal** derivatives can be described by a *gt/tg* conformational equilibrium with a very minor participation of the *gg* geometry.

Therefore, the conformational behavior around the hydroxymethyl group for a given monosaccharide in the gas phase depends on different factors: the configuration at its stereogenic centers, which provide the orientation of the hydroxyl groups to participate in cooperative HB networks. This is especially important at the anomeric carbon. The orientation of the hydroxyl group at C4 generates stabilizing (O4H···O6 HB) and destabilizing (O4/O6 steric) interactions. The predominance of the stabilizing or destabilizing forces depend on the associated intramolecular cooperative HB network, *clockwise* or *counterclockwise*, which can be modulated depending on the chemical nature and orientation of the substituent at C2. Water disrupts the HB network and the conformational behavior around C5–C6 mostly depends on steric interactions.

Thus, the work presented here demonstrates how OH-to-F/H replacements may modify the presentation, cooperative HB pattern, hydration, and dynamic properties of sugars. Hence, these simple chemical modifications are not innocuous in sugars and may strongly alter the tailored presentation to the appropriate biological targets, with a corresponding effect in their molecular recognition features and eventually in their biological activity.

IV.5 References

[1] A. García-Herrero, E. Montero, J. L. Muñoz, J. F. Espinosa, A. Vian, J. L. García, J. L. Asensio, F. J. Cañada, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **2002**, *124*, 4804–4810.

[2] H. Ohrui, Y. Nishida, M. Watanabe, H. Hori, H. Meguro, *Tetrahedron Lett.* **1985**, *26*, 3251–3254.

[3] Y. Nishida, H. Ohrui, H. Meguro, *Tetrahedron Lett.* **1984**, *25*, 1575–1578.

[4] K. Bock, J. Duus, *J. Carbohydr. Chem.* **1994**, *13*, 513–543.

[5] C. Nóbrega, J. T. Vázquez, *Tetrahedron Asymmetry* **2003**, *14*, 2793–2801.

[6] G. D. Rockwell, T. B. Grindley, *J. Am. Chem. Soc.* **1998**, *120*, 10953–10963.

[7] H. Hori, Y. Nishida, H. Ohrui, H. Meguro, *J. Carbohydr. Chem.* **1990**, *9*, 601–618.

[8] G. A. Jeffrey, R. K. McMullan, S. Takagi, *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **1977**, *33*, 728–737.

[9] R. H. Marchessault, S. Perez, *Biopolymers* **1979**, *18*, 2369–2374.

[10] S. Wolfe, *Acc. Chem. Res.* **1972**, *5*, 102–111.

[11] K. N. Kirschner, R. J. Woods, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10541–10545.

[12] I. Peña, C. Cabezas, J. L. Alonso, *Chem. Commun.* **2015**, *51*, 10115–10118.

[13] P. Çarçabal, R. A. Jockusc, I. Hünig, L. C. Snoek, R. T. Kroemer, B. G. Davis, D. P. Gamblin, I. Compagnon, J. Oomens, J. P. Simons, *J. Am. Chem. Soc.* **2005**, *127*, 11414–11425.

[14] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.

[15] M. Szczepaniak, J. Moc, *Carbohydr. Res.* **2014**, *384*, 20–36.

[16] I. Alkorta, P. L. A. Popelier, *Carbohydr. Res.* **2011**, *18*, 2933–2939.

[17] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.

[18] C. Calabrese, I. Uriarte, A. Insausti, M. Vallejo-López, F. J. Basterretxea, S. A. Cochrane, B. G. Davis, F. Corzana, E. J. Cocinero, *ACS Cent. Sci.* **2020**, *6*, 293–303.

[19] I. Peña, E. J. Cocinero, C. Cabezas, A. Lesarri, S. Mata, P. Écija, A. M. Daly, Á. Cimas, C. Bermúdez, F. J. Basterretxea, S. Blanco, J. A. Fernández, J. C. Lõpez, F. Castaño, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2013**, *52*, 11840–11845.

[20] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.

[21] I. Peña, S. Mata, A. Martín, C. Cabezas, A. M. Daly, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2013**, *15*, 18243–18248.

[22] P. Écija, I. Uriarte, L. Spada, B. G. Davis, W. Caminati, F. J. Basterretxea, A. Lesarri, E. J. Cocinero, *Chem. Commun.* **2016**, *52*, 6241–6244.

[23] I. Peña, C. Cabezas, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2015**, *54*, 2991–2994.

[24] I. Uriarte, P. Écija, R. Lozada-Garcia, P. Çarçabal, E. J. Cocinero, *ChemPhysChem* **2018**, *19*, 766–773.

[25] J. D. Martínez, A. I. Manzano, E. Calviño, A. de Diego, B. Rodriguez De Francisco, C. Romanò, S. Oscarson, O. Millet, H. J. Gabius, J. Jiménez-Barbero, F. J. Cañada, *J. Org. Chem.* **2020**, *85*, 16072–16081.

[26] J. D. Martínez, A. S. Infantino, P. Valverde, T. Diercks, S. Delgado, N. C. Reichardt, A. Ardá, F. J. Cañada, S. Oscarson, J. Jiménez-Barbero, *Pharmaceuticals* **2020**, *13*, 1–18.

[27] Y. Zhou, J. Wang, Z. Gu, S. Wang, W. Zhu, J. L. Acenã, V. A. Soloshonok, K. Izawa, H. Liu, *Chem. Rev.* **2016**, *116*, 422–518.

[28] E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly, N. A. Meanwell, *J. Med. Chem.* **2015**, *58*, 8315–8359.

[29] J. Wang, M. Sánchez-Roselló, J. L. Aceña, C. Del Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok, H. Liu, *Chem. Rev.* **2014**, *114*, 2432–2506.
[30] T. Fujiwara, D. O'Hagan, *J. Fluor. Chem.* **2014**, *167*, 16–29.
[31] S. Purser, P. R. Moore, S. Swallow, V. Gouverneur, *Chem. Soc. Rev.* **2008**, *37*, 320–330.
[32] J. Jiménez-Barbero, B. Linclau, A. Ardá, N. C. Reichardt, M. Sollogoub, L. Unione, S. P. Vincent, *Chem. Soc. Rev.* **2020**, *49*, 3863–3888.
[33] J. Wahsner, E. M. Gale, A. Rodríguez-Rodríguez, P. Caravan, *Chem. Rev.* **2019**, *119*, 957–1057.
[34] I. Tirotta, V. Dichiarante, C. Pigliacelli, G. Cavallo, G. Terraneo, F. B. Bombelli, P. Metrangolo, G. Resnati, *Chem. Rev.* **2015**, *115*, 1106–1129.
[35] X. Deng, J. Rong, L. Wang, N. Vasdev, L. Zhang, L. Josephson, S. H. Liang, *Angew. Chemie - Int. Ed.* **2019**, *58*, 2580–2605.
[36] S. M. Ametamey, M. Honer, P. A. Schubiger, *Chem. Rev.* **2008**, *108*, 1501–1516.
[37] P. W. Miller, N. J. Long, R. Vilar, A. D. Gee, *Angew. Chemie - Int. Ed.* **2008**, *47*, 8998–9033.
[38] S. Maschauer, O. Prante, *Biomed Res. Int.* **2014**, *2014*, 871609.
[39] D. L. Orsi, R. A. Altman, *Chem. Commun.* **2017**, *53*, 7168–7181.
[40] C. Ni, J. Hu, *Chem. Soc. Rev.* **2016**, *45*, 5441–5454.
[41] L. Hunter, *Beilstein J. Org. Chem.* **2010**, *6*, 38.
[42] V. H. Dalvi, P. J. Rossky, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13603–13607.
[43] J. C. Biffinger, H. W. Kim, S. G. DiMagno, *ChemBioChem* **2004**, *5*, 622–627.
[44] E. J. Cocinero, E. C. Stanca-Kaposta, E. M. Scanlan, D. P. Gamblin, B. G. Davis, J. P. Simons, *Chem. - A Eur. J.* **2008**, *14*, 8947–8955.
[45] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[46] R. A. Motiyenko, E. A. Alekseev, S. F. Dyubko, F. J. Lovas, *J. Mol. Spectrosc.* **2006**, *240*, 93–101.
[47] W. Xu, H. Yang, Y. Liu, Y. Hua, B. He, X. Ning, Z. Qin, H. M. Liu, F. W. Liu, *Synth.* **2017**, *49*, 3686–3691.
[48] C. Falk, P. E. Jansson, M. Rinaudo, A. Heyraud, G. Widmalm, P. Hebbar, *Carbohydr. Res.* **1996**, *285*, 69–79.
[49] S. J. Angyal, *Carbohydr. Res.* **1994**, *263*, 1–11.
[50] A. De Bruyn, M. Anteunis, *Bull. des Sociétés Chim. Belges* **1975**, *84*, 1201–1209.
[51] A. Avenoza, C. Cativiela, F. Corzana, J. M. Peregrina, M. M. Zurbano, *Synthesis (Stuttg).* **1997**, 1146–1150.
[52] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
[53] D. Šišak, L. B. McCusker, G. Zandomeneghi, B. H. Meier, D. Bläser, R. Boese, W. Bernd Schweizer, R. Gilmour, J. D. Dunitz, *Angew. Chemie - Int. Ed.* **2010**, *49*, 4503–4505.
[54] M. Kibalchenko, D. Lee, L. Shao, M. C. Payne, J. J. Titman, J. R. Yates, *Chem. Phys. Lett.* **2010**, *498*, 270–276.
[55] Y. Y. Chen, S. Y. Luo, S. C. Hung, S. I. Chan, D. L. M. Tzou, *Carbohydr. Res.* **2005**, *340*, 723–729.
[56] A. N. Appleyard, R. B. Herbert, P. J. F. Henderson, A. Watts, P. J. R. Spooner, *Biochim. Biophys. Acta - Biomembr.* **2000**, *1509*, 55–64.
[57] P. E. Pfeffer, K. B. Hicks, M. H. Frey, S. J. Opella, W. L. Earl, *J. Carbohydr. Chem.* **1984**, *3*, 197–217.
[58] P. De Andrade, J. C. Muñoz-garcía, G. Pergolizzi, V. Gabrielli, R. Harniman, S. J. Eichhorn, J. Angulo, Y. Z. Khimyak, P. de Andrade, J. . Munoz-garcia, G. Pergolizzi, V. Gabrielli, S. A. Nepogodiev, D. Iuga, L. Fabian, R. Nigmatullin, M. A. Johns, R. Harniman, S. J. Eichhorn, J. Angulo, Y. Z. Khimyak, R. A. Field, *ChemRxiv* **2020**, 1–5.

[59] I. Peña, A. M. Daly, C. Cabezas, S. Mata, J. L. Alonso, in *68th OSU Int. Symp. Mol. Spectrosc. Ohio State Univ. Columbus*, **2013**.
[60] I. Uriarte, A. Insausti, E. J. Cocinero, A. Jabri, I. Kleiner, H. Mouhib, I. Alkorta, *J. Phys. Chem. Lett.* **2018**, *9*, 5906–5914.
[61] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[62] F. O. Talbot, J. P. Simons, *Phys. Chem. Chem. Phys.* **2002**, *4*, 3562–3565.
[63] N. Mayorkas, S. Rudić, E. J. Cocinero, B. G. Davis, J. P. Simons, *Phys. Chem. Chem. Phys.* **2011**, *13*, 18671.
[64] C. A. G. Haasnoot, F. A. A. M. de Leeuw, C. Altona, *Tetrahedron* **1980**, *36*, 2783–2792.
[65] J. Gonzalez-Outeiriño, K. N. Kirschner, S. Thobhani, R. J. Woods, *Can. J. Chem.* **2006**, *84*, 569–579.
[66] M. Ziemniak, A. Zawadzka-Kazimierczuk, S. Pawlędzio, M. Malinska, M. Sołtyka, D. Trzybiński, W. Koźmiński, S. Skora, R. Zieliński, I. Fokt, W. Priebe, K. Woźniak, B. Pająk, *Int. J. Mol. Sci.* **2021**, *22*, 3720.
[67] Y. Nishida, H. Hori, H. Ohrui, H. Meguro, *J. Carbohydr. Chem.* **1988**, *7*, 239–250.
[68] A. De Bruyn, M. Anteunis, *Carbohydr. Res.* **1976**, *47*, 311–314.
[69] A. Lonardi, P. Oborský, P. H. Hünenberger, *Helv. Chim. Acta* **2017**, *100*, e1600158.
[70] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

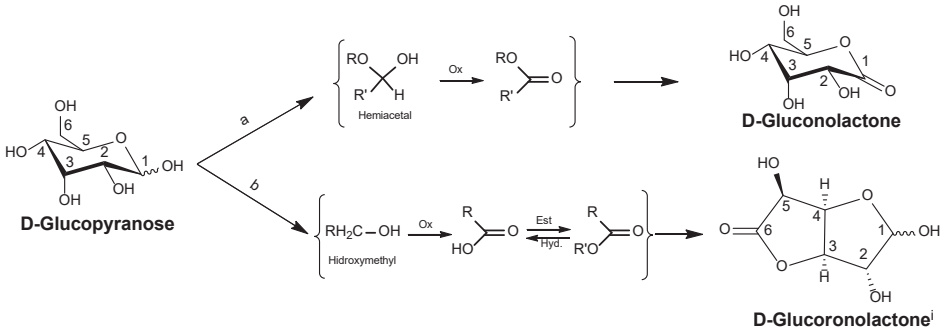
Chapter V.

Sugar Lactones

V.1 Introduction

Oxidation is a very important biological reaction in sugar chemistry because it enables the synthesis of a great variety of monosaccharides and their derivatives from simple monosaccharides. During the oxidation processes, one of the recurrent group of molecules that appears are lactones.[1] In this chapter, we will show our study about the two most and important D-glucose derivative lactones in the nature: D-gluconolactone and D-Glucoronolactone (see Scheme V.1).

They are two different pathways to obtain lactones through sugar oxidation. First, the selective oxidation of hemiacetal anomeric hydroxyl group and second, oxidation of any other secondary hydroxyl group followed by the intramolecular esterification (Scheme V.1). Despite their small apparent structural change during oxidation/esterification processes, conformational, chemical and biological properties of the sugars could be completely altered.[2]



Scheme V.1 Two glucose lactone formation mechanisms and its structures. The hemiacetal oxidation (a) form the gluconolactone molecule, and the hydroxymethyl oxidation and intramolecular esterification (b) forms the glucoronolactone. Atom numbering used in this work has been added. Note: [i]Depicted furanose D-glucoronolactone isomer is the same as appears in the crystal structure (see text for more details).

Glucose molecule is a prototype monosaccharide that has been studied extensively for its ubiquitous presence in structural role of complex biomolecules.[3] Glucose is also the most abundant monosaccharide and consequently, its lactones are also very common in nature. Several glucose oxidation/lactonization products are known, and although much of them are very common in several cosmetic and food products their structure, metabolism and biological effects are not completely clear.

One of the most consumed glucose derived lactone is D-gluconolactone (**GlcL**) and it comes from the oxidation of hydroxylation at the anomeric carbon of glucose (Scheme V.1a). **GlcL** is ingested as common commercial food additive (stabilizing agent E575) and it has also been incorporated into several cosmetic products. These include functions such as UV protection,[4] humectant,[5]

stabilizing agent[6] and antioxidant.[5] In the biological media, the molecule is generated oxidizing glucose by glucose oxidase enzyme but depending on the temperature, it is spontaneously converted to gluconic acid.[7] Curiously, the **GlcL** can also be generate by the thermal degradation process of glucose without the presence of enzymes.[8]

The second glucose derivative lactone is D-glucuronolactone (**GlcLA**) that is formed by the second mechanism of Scheme V.1b, i.e. through the oxidation of the hydroxymethyl group of glucose and then by intramolecular esterification in C4. Esterification and hydrolysis reactions make that **GlcLA** and glucuronic acid (the study of the glucuronic acid molecule is also reported in Chapter VI of this Thesis) form an equilibrium mixture in aqueous solution and their ratio is dependent on the temperature (Scheme V.1b).[9,10] **GlcLA** is found in nature as an important structural component of nearly all connective tissues. It has found application in medicinal uses, including improvement of joint health, anti-inflammatory effects for the skin, or in lowering cholesterol or triglyceride levels. Specifically, it has been reported to relieve hepatotoxicity,[11] to have protective effects against cell apoptosis during brain development[12] or to attenuate para-xylene-induced defects in neuronal development and plasticity.[13] It is also frequently used in energy drinks to increase energy levels and improve alertness, together with other substances as caffeine, taurine, guaranine, and B vitamins.[14]

GlcL and **GlcLA**, the molecules of interest, structurally present noticeable differences. For example, **GlcL** and **GlcLA** have different lactone ring members. In the case of **GlcL**, it is formed by 5-member lactone ring while **GlcLA** is formed by 6-member lactone ring. This is evidenced in the crystal structures of glucuronolactone[15] and gluconolactone,[16] reported using X-ray diffraction. In crystal, the gluconolactone is very similar to glucose, with the hydroxymethyl group in a perpendicular position with respect to the ring (*gg* conformer) but with the pyranose ring puckering distorted. Further, the lack of anomeric hydroxyl group in **GlcL** inhibits the ring-opening reaction. For this reason, there is only one possible isomer for this molecule. Nevertheless, **GlcLA** presents a more complicate molecular isomerism. The non-oxidation of anomeric oxygen presumes the possibility to see five separate isomers so far: straight chain, and both alpha and beta anomers of pyranose and furanose form. Interestedly, the crystal structure of **GlcLA** have two 5-member rings, the γ-lactone formed by C3-C4-C5-C6-O_{ester} and the second one forms the furanose ring through C1-C2-C3-C4-O5 atoms (See Scheme V.1). To our knowledge, no other structural studies on any of these molecules have been reported so far.

The aim of this work is to study conformational possibilities in the gas phase of **GlcL** and **GlcLA**, two molecules formed by the oxidation of glucose, that play significant roles in many biological processes. Herein, we present a detailed analysis of the rotational spectra of **GlcL** and **GlcLA** in the 6-18 GHz frequency

range. For this purpose, we used a combination of computational chemistry calculations, Fourier Transform rotational spectrometer and a laser ablation system to vaporize the samples.

V.2 Methodology

Rotational spectra were recorded in a 6-18 GHz Fourier transform microwave spectrometer (cav-FTMW) based on the Balle-Flygare design[17] constructed at the UPV/EHU and described in detail elsewhere. [18,19] Commercial samples of D-(+)-glucuronolactone and D-(+)-Gluconic acid δ-lactone (gluconolactone) were used without further purification. The solid samples were transferred intact to the gas phase by vaporizing a solid rod of mechanically pressed sample with an ultrafast 7−8 mJ laser pulse (35 ps) from the third harmonic (355 nm) of a Nd:YAG laser. The vaporized sample was mixed with Ar as carrier gas at 8 bar.

Experimental work was aided by computational procedures: the conformational landscape was initially scanned by Merck Molecular Mechanics Ford Field (MMFFs) molecular mechanics (MM) calculations,[20] that combine advanced Monte Carlo and large-scale low-mode algorithms.[21] MM calculated energy minima structures were further refined by ab initio Møller-Plesset (MP2)[22] and density functional theory (DFT) quantum chemistry methods, implemented in the Gaussian 16 package. For DFT, Becke's B3LYP [23] hybrid functional with GD3BJ empirical dispersion was used. The Pople's 6-311++G(d,p) basis set was used in all calculations (all details in next section and in the Appendix V, as well as in Chapter II).

V.3 Results and disscusion

Glucoronolactone

The initial exploration of glucurolactone began by predicting its possible structures and their rotational spectra by density functional theory (B3LYP(GD3BJ)) and MP2 theory using 6-311++G(d,p) basis functions (all details in Methodology section). The molecule can exist as, apart from linear isomer, α and/or β anomers in the pyranose and/or furanose forms (Figure V.1). So the initial conformational search has been focused on all the possible isomers that **GlcLA** can adopt and separate calculations were carried out for each of the anomers/isomers. One of the advantage of gas phase spectroscopy is that α ↔ β or furanose ↔ pyranose interconversion is prevented.[24,25] Therefore, in the gas phase we can only detected the same isomer that is present in the solid sample. In this study, we cannot predict the isomer that forms the sample, but the unique crystal structure that has been detected is for beta furanose isomer, so initially we suggested that probably the solid sample could be formed only by β-*f* isomer. The predicted rotational constants and dipole moment components of **α-** and **β-*f*-GlcLA** conformers are presented in Table V.1 and in the Appendix (Tables AV.1-

AV.2 and Figures AV1-AV2) and the results of **α-** and **β-*p*-GlcLA** and linear isomers are in the Appendix (Tables AV.3-AV.5 and Figures AV3-AV5). As can be seen from the predictions, rotational constants of the conformers of the **β-*f*-GlcLA** isomer differ drastically from those of **α-*f*-GlcLA, α/β-*p*-GlcLA** and ***l*-GlcLA**, so they could be in principle easily distinguished using rotational spectroscopy data.

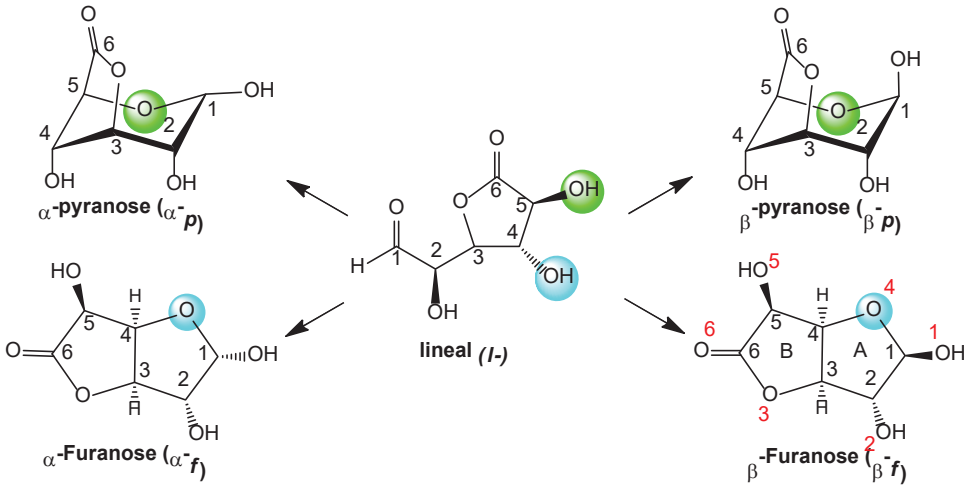


Figure V.1 Possible isomers of D-glucoronolactone. The numbers represent the labeling of the carbons (in black) and of the oxygens (in red). In the β-*f*, ring A (hemiacetal ring) and B (lactone ring) are indicated.

The search for conformers in the supersonic expansion began in the spectral regions where the most stable structures of all isomers were predicted theoretically (linear, α/β-*p* and α/β-*f*). After several scans, two distinct series of rotational transitions were observed, originating from different species. The rotational transitions of each species were fitted by using a Watson *A*-type semi-rigid Hamiltonian, including centrifugal distortion constants. The obtained rotational constants are shown in Table V.1. and all the fitted transitions are reported in Tables AV8-AV9 of the Appendix. Comparing the fitted values to the predicted ones, the observed structures were assigned to **β-*f*-GlcLA 1** and **β-*f*-GlcLA 2** (the two lowest energy conformers of β-*f*). Observed and calculated rotational constants of **β-*f*-GlcLA 1** are very similar, although conformers number 3 and 4 (2.8 and 4.3 kJ/mol higher than conformer 1, respectively) show less deviation from experimental values. However, these conformers can rotationally relax to **β-*f*-GlcLA 1** (the explication is given later). Furthermore, the most intense observed lines were μ_c-type transitions, so conformer 3 and 4 should be excluded due to a lower predicted value for this dipole moment component. For **β-*f*-GlcLA 2** (2.7 kJ/mol higher than **β-f-GlcLA 1**), the deviation between observed and

calculated rotational constants is smaller than for **β-*f*-GlcLA 1**. Other alternative assignments could be conformer 6 (5 kJ/mol higher than conformer 1, respectively), although with a slightly larger deviation than in **β-*f*-GlcLA 2** and the transition intensities and dipole components are not consistent. So, we can conclude that the detected conformers were unambiguously the lowest energy conformers **β-*f*-GlcLA 1** and **β-*f*-GlcLA 2**.

Although the search for higher energy conformers was extensive, no other signals could be detected in the supersonic expansion. In order to rationalize the absence of more conformers and to address the possible relaxation among species in the jet, β-*f*-GlcLA conformer interconversion pathways were carefully examined. A complete scheme of relaxation paths between conformers is proposed, based on the lower energy pathways (see Figure V.2.). It should be noted that the interconversion between conformer is predicted with a very low energy gap and this can explain the lack of more conformers in the experimental data. Particularly relevant are the following pathways: 5 → 7 → 4 → 3 → 1 (this one would explain the absence of 5, 7, 4 and 3 species in the jet) and 6 → 2 (explains the absence of 6 conformer). Finally, 9 → 1, and 8 → 2 relaxation pathways explain the absence of these energetically higher species in the jet. On the other hand, the lack of other isomers (linear, *α-f* and *α*/β-*p*) could be attributed to the absence of these isomers in the solid commercial sample used in the experiment.

Table V.1 Experimentally fitted rotational parameters and theoretical predictions using MP2/6-311++G(d,p) method of **β-*f*-GlcLA**. Energy and Gibbs energy are in cm^{-1}.

	β-*f*-GlcLA 1		**β-*f*-GlcLA 2**	
	Theo.	Exp.	Theo.	Exp
$A^{[a]}$	1670	1656.56600(17)[d]	1583	1578.36796 (26)
B	885	874.851218(84)	811	806.99379 (12)
C	787	782.182652(82)	775	777.96259 (15)
D_J		0.04755(67)		0.05715 (79)
D_{JK}		0.1285(46)		0.1790 (53)
d_1		-0.00627(35)		-0.00810 (69)
$\mu_a^{[b]}$	0.8	w.	4.3	vs.
μ_b	1.8	s.	1.2	w.
μ_c	-2.3	vs.	-1.9	s.
$\Delta E^{[c]}$	0.0		229	
ΔE_{ZPE}	0.0		60	
ΔG^{298K}	95		0.0	
$N^{[d]}$		99		59
$\sigma^{[e]}$		2.2		2.8

[a] The rotational constants are in MHz and the distortion constantans are in kHz.
[b] Dipole moments are in Debye units. In the experimental column, w., s, and vs. refers to weak, strong and very strong respectively, the experimental intensity of the lines.
[c] The energies and Gibbs energy are in cm^{-1}
[d] *N* represents the fitted number of lines and σ is the root-mean-square deviation of the fit expressed in kHz.
[e] Numbers in parentheses represent the error propagated from the errors in the rotational constants.

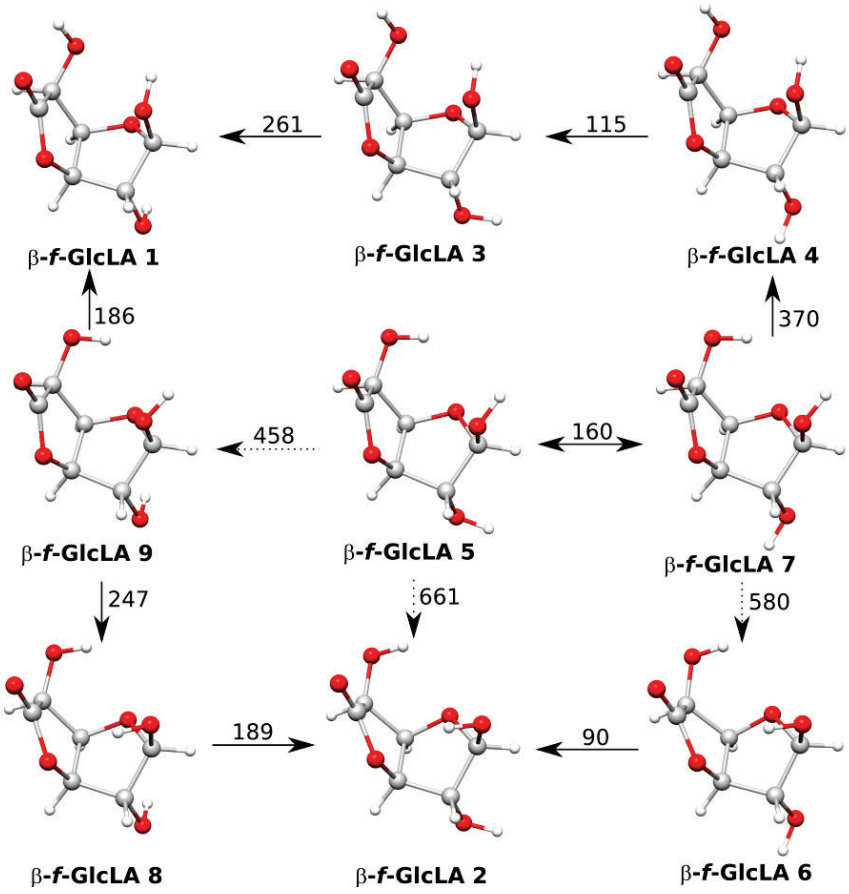


Figure V.2 Possible interconversion pathways for relaxation among conformers of β-GlcLA calculated at B3LYP (GD3BJ)/6-311++G(d,p) level of theory. The arrows with dots represent higher energetic pathways and the numbers represent the energy barrier reported in cm^{-1}.

For the conformational point of view, **β-*f*-GlcLA** isomer (Figure V.2) presents the O1H and O2H hydroxyl groups on the opposite side of the furanose A ring, and this disposition prevents the formation of cyclic cooperative intramolecular H-bonds, characteristic of monosaccharides. Conformer **β-f-GlcLA 1** (Figure V.2) has three non-cooperative weak intramolecular H-bonds; $O1H\cdots O4_{hemiacetal}$ $O2H\cdots O4_{hemiacetal}$ and $O5H\cdots O{=}C$, all of them with $r_{O\cdots HO}$ > 2.4 Å. Meanwhile, in **β-f-GlcLA 2** two stronger intramolecular H-bonds are present: $O1H\cdots O3_{ester}$ and $O5H\cdots O4_{hemiacetal}$ with $r_{O\cdots HO}$ < 2.4 Å. In the case of **β-*f*-GlcLA 1,** it present three intramolecular H-bonds (2x $OH\cdots$ $O4_{hemiacetal}$ and 1 x

OH···O=C). **β-*f*-GlcLA 2** has one H-bond less than **β-*f*-GlcLA 1**, because the O2H of **β-*f*-GlcLA 2** is not pointed to $O4_{hemiacetal}$. The lack of stabilization of $O2H\cdots O4_{hemiacetal}$ interaction in the **β-*f*-GlcLA 2** (like in the **β-f-GlcLA 1**) could be attributed to the ring puckering differences in these two conformers.

A study of ring torsional states has been carried out for the assigned structures of **β-*f*-GlcLA**, through the use of Cremer-Pople diagrams, for both ring A (hemiacetal ring) and B (lactone ring), (see Figure VA.5 of Appendix V).[26] For comparison purposes, we have also included the crystal structure of β-*f*-GlcLA.[15] For **β-*f*-GlcLA 1**, ring A shows a ^2E envelope conformation, whereas ring B exhibits a $^4T_5$ twist state. Curiously**, β-*f*-GlcLA** solid samples (the unique anomer reported in the solid state), ring A state is quite the same (in both amplitude and phase angle) of ring A of gaseous **β-*f*-GlcLA 1**, and ring B state is practically coincident with ring B of **β-*f*-GlcLA-1**. For **β-*f*-GlcLA-2**, puckering and amplitude of ring A and ring B are very similar and they show a $^4T_3$ twist conformation. This ring configuration keeps away the O2H from hemiacetal oxygen (O4) and for this reason the interaction $O2H\cdots O4_{hemiacetal}$ in the **β-*f*-GlcLA-2** is not favoured**.**

Gluconolactone

For gluconolactone (**GlcL**), the methodology proceeded in the same way as before. The predicted most stable structures using MM methods were recalculated at MP2/6-311++G(d,p) and B3LYP(GD3BJ)/6-311++G(d,p) theory levels, both methods gave very similar results for the three lowest-lying structures (see Table V.2 and Table AV.6 and Figure AV.8 of Appendix V). T Conformer energies span the 0-600 cm^{-1} range, and their main structural differences lie in the different orientations of the CH_2OH hydroxymethyl group with respect to the cycle frame, the orientation of the OH group and the ring torsional states. The conformers have been labeled according to their hydroxymethyl group dispositions, where *gt*, *gg* or *tg*, describe the torsion angle O5–C5–C6–O6 corresponding to approximately 60º, -60º or 180º, respectively.

Two conformers of **GlcL** were detected in the rotational spectrum and a rigid rotor model was used to fitted rotational transitions, see Table V.2 (all transitions and errors are in Tables AV.10-AV11 of the Appendix). They were assigned to **GlcL 1** and **GlcL 2** predicted structures. The concordance of *A*, *B* and *C* experimental values with the predicted ones is very good, so the assignment can be made confidently. These conformers are the two most stable structures, and their energy difference is 223 cm^{-1}, according to MP2 calculations. For **GlcL 1**, μ_b-type transitions have been observed, whereas for **GlcL 2** μ_a - and μ_b -types have been measured. These results are in accordance with the theoretical predictions, that indicate these dipole moment components the most intense ones.

On the matter of conformation, **GlcL** molecule displays clear similarities with the **Glc** molecule (see Figure V.3).[25] The most stable structures of both, **Glc** and **GlcL**, has preference of a *cc* H-bond network (O4H → O3H → O2H → O1) with cooperative H-bond extended through all the hidroxil groups of the molecule. However, the conformational population are considerably different. In the case of glucose (α and β anomers) the population of *gg* (*G-g*+) and *gt* (G+g-) conformers are very similar and the third conformer *tg* (Tg+) has enough stability to be detected experimentally (**β-Glc *tg*** lies 279 cm^{-1} above the global minimum according to the MP2/6-311++G(d,p)). Nevertheless, **GlcL *tg* 4** lies at 591 cm^{-1} above the most stable conformer, making its detection impossible in our experimental conditions. We suggest that the lack of O1H···O4 interaction in the GlcL could contribute to the preference of O6H···O4 interaction to the detriment of the stability of **GlcL *tg* 4**.

Table V.2 Experimentally fitted rotational parameters and theoretical predictions using MP2/6-311++G(d,p) method of **GlcL**.

	GlcL *gg* 1		**GlcL *gt* 2**		**GlcL *tg* 4**
	Theo.	Exp.	Theo.	Exp.	Theo.
***A*[a]**	1205	1202.555(21)	1230	1232.346(23)	1369
B	848	844.597(10)	803	798.740(10)	753
C	552	552.80540(36)	503	502.24230(51)	504
μ_a[b]	-0.6	s	-2.1	s	-1.3
μ_b	-0.7	-	-1.4	s	-2.6
μ_c	0.5	-	-1.0	-	-0.9
ΔE[c]	0.0		232		468
ΔE_{ZPE}	0.0		223		523
ΔG^{298K}	0.0		243		591
N[d]		8		8	
σ[e]		2.0		1.3	

[a] The rotational constants are in MHz and the distortion constantans are in kHz.
[b] Dipole moments are in Debye units. In the experimental column, w., s, and vs. refers to weak, strong and very strong respectively, the experimental intensity of the lines.
[c] The energies and Gibbs energy are in cm^{-1}.
[d] *N* represents the fitted number of lines and σ is the root-mean-square deviation of the fit expressed in kHz.
[e] Numbers in parentheses represent the error propagated from the errors in the rotational constants.

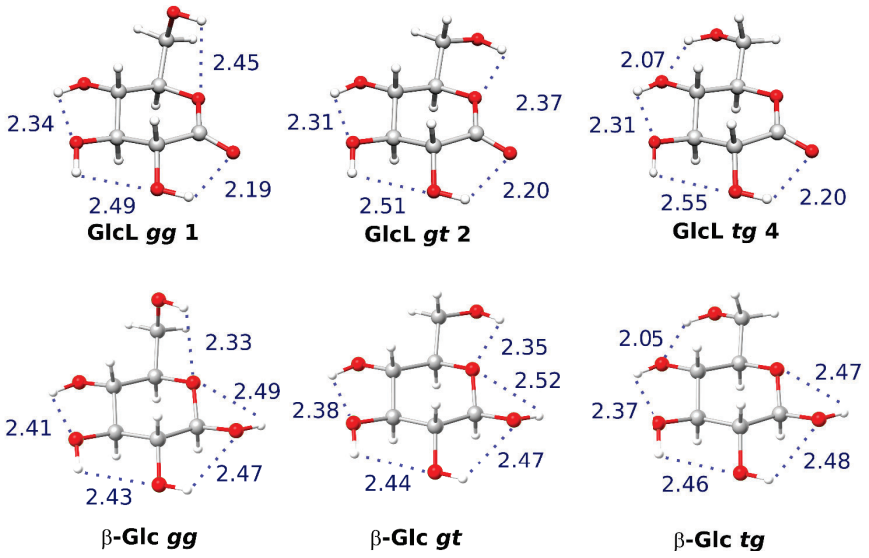


Figure V.3 The most stable structures of **GlcL** calculated at MP2/6-311++G(d,p) level of theory and their β-Glc analogues detected in a previous rotational study. [25]

Finally, predicted structural parameters of **GlcL** have been compared with those of its precursor β-D-glucopyranose (both structures calculated at MP2/6-311++G(d,p)). H-bond distance comparison between the conformers whose hydroxymethyl group in the same arrangement are depicted in Figure V.3. A careful analysis of the data reveals that the H-bond distances near the carbonyl group in **GlcL** are considerably shorter than in β-D-glucopyranose (O2H···O1H with $r_{O2H\cdots O1H}$ = 2.47 - 2.48 Å vs. $r_{O2H\cdots O=C}$ =2.19 - 2.20 Å). Further, the O6H···O distance between *gg* conformers of **Glc** and **GlcL** are quite different. This could be attributed to the different ring puckering of **Glc** and **GlcL**. In this contest, it is known that Glc presents pure $^4C_1$ ring puckering, but the lactone group of **GlcL**, forced in planarity by the sp^2 hybridization of C1, acquires a $^4C_1$ distorted ring puckering (all values are reported in Table AV.7 of Appendix V).

V.4 Conclusions

In this study, we present the first experimental data of glucose derivative sugar lactones in the gas phase, which has led to compare their intrinsic conformational landscape with its precursor glucose. For each molecule, we experimentally assigned two conformers. **GlcLA** presented β-furanose isomer in gas phase and **GlcL** does not have anomeric hydroxyl groups, so we observed its unique pyranose isomeric form. In general, the structure of the **β-*f*-GlcLA** drastically differs for the other aldohexose sugars. The three hydroxyl groups of **β-*f*-GlcLA** do not present the OH···O H-bond cooperativity, commonly presents in other aldohexoses, and for this reason, there are several conformers in a very low energy window. Unfortunately, the low interconversion barriers generated by the lack of the abovementioned intramolecular H-bond network, make that only two of them were shown experimentally. On the other hand, **GlcL** and **Glc** molecules have clear similarities in the gas phase. Their hydroxymethyl groups have similar arrangement and their H-bond networks are practically identical. However, we detected small ring puckering differences and high preference in the **GlcL** of O6H···O5 interaction that would be related also with the lack of **GlcL *gg*** conformer in our experimental data.

V.5 References

[1] A. P. Rauter, S. Vogel, Y. Queneau, *Carbohydrates in Sustainable Development II*, **2010**.
[2] M. Miljković, *Carbohydrates: Synthesis, Mechanisms, and Stereoelectronic Effects*, **2009**.
[3] *Carbohydrate Chemistry and Biochemistry*, **2013**.
[4] E. F. Bernstein, D. B. Brown, M. D. Schwartz, K. Kaidbey, S. M. Ksenzenko, Z. D. Draelos, *Dermatologic Surg.* **2004**, *30*, 189–196.
[5] E. Berardesca, F. Distante, G. P. Vignoli, C. Oresajo, B. Green, *Br. J. Dermatol.* **1997**, *137*, 934–938.
[6] X. Wei, J. Zhu, H. Gong, J. Xu, Y. Xu, *Contrast Media Mol. Imaging* **2011**, *6*, 465–473.
[7] R. Wilson, A. P. F. Turner, *Biosens. Bioelectron.* **1992**, *7*, 165–185.
[8] I. S. Cherepanov, O. O. Botalova, *AIP Conf. Proc.* **2020**, *2280*, 050016–1.
[9] S. Suzuki, S. Hayase, M. Nakano, Y. Oda, K. Kakehi, *J. Chromatogr. Sci.* **1998**, *36*, 357–360.
[10] R. Wang, T. L. Neoh, T. Kobayashi, Y. Miyake, A. Hosoda, H. Taniguchi, S. Adachi, *Biosci. Biotechnol. Biochem.* **2010**, *74*, 601–605.
[11] Z. Yu, F. Wu, J. Tian, X. Guo, R. An, *Mol. Med. Rep.* **2018**, *18*, 2551–2560.
[12] J. Gao, H. Ruan, X. Qi, X. Guo, J. Zheng, C. Liu, Y. Fang, M. Huang, M. Xu, W. Shen, *Neuroscience* **2016**, *331*, 177–185.
[13] Y. Liao, Y. Luo, N. Ding, J. Gao, X. Wang, W. Shen, *Toxicology* **2020**, *430*, 152341.
[14] J. P. Higgins, T. D. Tuttle, C. L. Higgins, *Mayo Clin. Proc.* **2010**, *85*, 1033–1041.
[15] S. H. Kim, G. A. Jeffrey, R. D. Rosenstein, P. W. Corfield, *Acta Crystallogr.* **1967**, *22*, 733–743.
[16] M. L. Hackert, R. A. Jacobson, *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **1971**, *27*, 203–209.
[17] T. J. Balle, W. H. Flygare, *Rev. Sci. Instrum.* **1981**, *52*, 33–45.
[18] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[19] E. J. Cocinero, A. Lesarri, P. Écija, Á. Cimas, B. G. Davis, F. J. Basterretxea, J. A. Fernández, F. Castaño, *J. Am. Chem. Soc.* **2013**, *135*, 2845–2852.
[20] T. A. Halgren, *J. Comput. Chem.* **1999**, *20*, 720–729.
[21] I. Kolossváry, G. M. Keserü, **2001**, *22*, 21–30.
[22] P. Møller, *Eur. J. Cardio-thoracic Surg.* **1934**, *53*, 1237–1243.
[23] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[24] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.
[25] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.
[26] D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

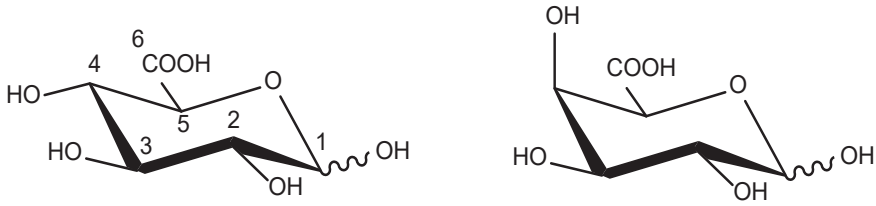
Chapter VI.

Uronic acids

Manuscript for publication in preparation

VI.1 Introduction

In this chapter we will present the study of the most abundant hexuronic acids in nature: glucuronic acid and galacturonic acid. Monosaccharides suffer modifications, including acetylation, methylation, oxidation, and sulfonation, creating a great chemical diversity from the simplest carbohydrate building blocks. One of the most current modification is the oxidation of the hydroxymethyl group of hexoses into carboxylic acid group. These new molecules are called uronic acid or hexuronic acids. This is the case of D-Glucuronic acid (**GlcA**) and D-Galacturonic acids (**GalA**), which are derived from the oxidation of the primary alcohol on C6 of D-Glucose (**Glc**) and D-Galactose (**Gal**), respectively, to a carboxyl group (see Scheme VI.1).



D-Glucuronic acid **D-Galacturonic acid**

Scheme VI.1 $^4C_1$ chair conformations for D-glucuronic acid and D-galacturonic acid in a generic anomeric configuration.

The hexuronic acids are very common monosaccharides, mainly as constituents of polysaccharides. The presence of **GlcA** in nature is very high and it was firstly isolated by M. Jaffé in 1874[1] in the urine of animals and chemically synthetized in 1891 by E. Fischer and O. Piloty.[2] In the biological media, it is one of the most current monosaccharides in glycosaminoglycan polysaccharides like heparin,[3] dermantan sulfates,[4] hyaluronic acid[5] and it also appears in several plant polysaccharides like hemicelluloses[6] and in several bacterial polysaccharides. There are many types of these polysaccharides that contain **GlcA** monomer as principal building block and, obviously, they are involved in a broad range of biological processes. For example, one of the most studied glycosaminoglycan, heparin, has been shown to possess several activities that extend beyond anticoagulation, including anti-inflammatory and anticancer properties. Numerous scientific studies have identified a role for heparin oligosaccharide structures in mediating several diseases, such as Alzheimer's disease,[7] viral infection[8] and cancer.[9–11]

Regarding to the **GalA**, it is commonly present in the plant kingdom, concretely in pectins.[12] Pectin is a polysaccharide with a backbone consisting mostly of α-1,4-linked D-galacturonic acid, as well as secondary ramifications composed by a large number of other sugars. Nowadays, pectin represents an

important part of the research of natural medicines and health products not only due to its wide availability in nature, also because it has been effective preventing several diseases like cancer.[13–15] In the pharmaceutical industry, it is used to reduce blood cholesterol levels and gastrointestinal.[16,17]

Since the last century, polysaccharides have found their place in the biomedical field but the complexity of sugar chemistry makes impossible to predict their biological functions without biomedical and clinical trials. For this reason, one of the principal topics in glycobiology consist on the prediction of the biological function of each polysaccharide knowing only their structure and composition. In this context, it is well known that the function of a molecule is strongly connected to its three dimensional shape,[18,19] and of course, the conformational behavior of polysaccharides is directly related to their biological activity.[20–22] Thus, the essential first step to predict the functionality of each sugar is to know perfectly their conformation. However, although the conformational panorama of polysaccharides could be elucidated using several experimental techniques, such as nuclear magnetic resonance,[23] atomic force microscopy,[24] crystallography[25] or/and ion-mobility/mass-spectrometry,[26] commonly these results have to be accompanied and supported by computational chemistry calculation. This raises two issues: first, we need to synthesize or obtain the sugar before knowing its structure; second, once the compound has been obtained, it is not always possible to elucidate the structure, or the result obtained is not what expected.

The solution would be to computationally predict the sugar conformation and use these data to evaluate their biological activity before synthesizing or acquiring the sample of interest. However, an accurate *in-silicon* prediction of the structures is not accessible with current computational methods. In this context, the logic way to understand the laws that govern the conformational shape of these molecules should start understanding perfectly the conformational panorama of each monosaccharide in an isolated environment.

For that purpose, in the present study, the conformational behaviors of **GlcA** and **GalA**, successfully transferred in the gas phase by laser ablation, are reported for the first time using microwave spectroscopy. We combine the solid state Nuclear Magnetic Resonance (ss-NMR), recrystallization, rotational spectroscopy and high level quantum chemistry calculations to elucidate the conformational panorama of the α and β anomers of **GlcA** and **GalA**.

VI.2 Methodology

Commercial solid samples of D-glucuronic acid and D-galacturonic acid hydrate were used to obtain rotational spectra. Besides, a portion of D-galacturonic acid hydrate was dissolved and recrystallized in acetic acid following a previous report in order to obtain the β-anomer (all details in Point AVI.2 of the Appendix VI).[27] Then, the same procedure mentioned in detail in Chapter II of

this thesis was followed to create a solid rod. The molecules were measured using the Fabry-Perot Molecular Beam Fourier Transform microwave spectrometer (cav-FTMW) already explained in Chapter II. We used picosecond laser ablation (355nm) to vaporize the samples and Ar (~ 55 kPa) as carrier gas.

The conformational search was done using molecular mechanics calculations. Concretely, we used the *Merck Molecular Force Field* (MMFFs) implemented in Maestro2016 and we selected all the conformers with relative energy lower that 15 kJ/mol for α and β pyranose forms of **GalA** and **GlcA**. Then, the Møller–Plesset second order method (MP2) and some Density Functional Theory (DFT) methods together with the 6-311++G(d,p) basis set were used to optimize the geometries of all the structures and to calculate the relevant spectroscopic properties. The obtained results are summarized in Tables VI.1-VI.2 and in Tables AVI.1-AVI.4 and in Figures AVI.1-AVI.4 of the Appendix VI (all details are in the Computational Methods of Appendix VI). We rejected the possibility to detect the furan forms for their bigger ring torsion.

VI.3 Results and discussion

Once the conformational landscape was computationally predicted for α and β-anomers of **GlcA** and **GalA** (see Methodology, as well as Tables AVI.1-AVI.4 and Figure AVI.1-AVI.4 in the Appendix VI) two independent experiments were performed, one for **GlcA** and another for **GalA**. Frequency scans were recorded by means of the home-made cav-FTMW[28] spectrometer equipped with a ultrafast laser vaporization system at the Universidad del País Vasco[29,30], where the most intense rotational transitions were predicted. First, **GlcA** and **GalA** commercial samples were used in the experiments three different sets of μ_a-*R*-type rotational transitions could be identified, belonging to two and one independent structure(s) of **GlcA** and **GalA**, respectively. Each set was fitted with Pickett's SPFIT program[31] using Watson's symmetrically reduced semirigid rotor Hamiltonian and I^r representation[32] to obtain the rotational parameters of Table VI.1-VI.2. In particular, the rotational constants are very useful since are a *unique fingerprint* of the molecular structure. The comparison of the experimental and theoretical rotational constants (Table VI.1-VI.2, Table AVI.1-AVI.4 in the Appendix VI) allowed us to straightforwardly and unambiguously identify the **α-GlcA 1** and **α-GalA 1** confomers. In fact, both structures are the corresponding global minima according to the theory. A second conformer was identified for **GlcA, β-GlcA 1**, although it had to be further supported by the dipole moment components with the availability or absence of their corresponding spectra. This also resulted in an unambiguous assignment (see Table AVI.2 in the Appendix VI). No more lines corresponding to any possible conformers were identified. The experimental sets of rotational parameters and the measured transition frequencies can be found in Table VI.1- VI.2 and Tables AVI.8-AVI.10 in the Appendix VI, respectively.

Table VI.1 The experimental and computational results of both α and β anomers of **GlcA**.

	α-GlcA 1		β-GlcA 1	
	Theo.[a]	**Exp.**	**Theo.**	**Exp.**
A [MHz][b]	1239	1235.96890(50)[f]	1137	1135.72639(45)
B [MHz]	645	642.711001(73)	661	658.97886(10)
C [MHz]	468	466.342985(42)	436	434.973356(63)
D_J [kHz]		7.03(12)		0.00797(21)
d_1 [kHz]		--		-0.00197(15)
$\mu_a/\mu_b/\mu_c$ (D)	2.9/0.5/0.5	Yes/No/No[g]	2.6/0.6/0.7	Yes/No/No
σ [kHz][c]		1.1		2.0
N[d]		37		85
Com. gas/sol. [%][e]		24(8):10		76:90

[a] *Ab initio* calculations (MP2/6-311++G(d,p)).
[b] Rotational constants (A, B, C); μ_a, μ_b and μ_c are the absolute values of the electric dipole moment components.
[c] The *rms* deviation of the fit.
[d] Number of fitted rotational transitions.
[e] Experimental anomeric population in the commercial sample (Com.) in gas and solid phase.
[f] Standard error in parenthesis in the units of the last digit.
[g] Experimental observation of *a*-, *b*-, and *c*- *μ*-type transitions for each structure.

Table VI.2 The experimental and computational results of both, α and β anomers of **GalA**.

	α-GalA 1		β-GalA 1	
	Theo.[a]	**Exp.**	**Theo.**	**Exp.**
A [MHz][b]	1310	1303.8852(32)[f]	1242	1234.2059(36)
B [MHz]	644	642.79144(12)	633	633.08463(27)
C [MHz]	486	483.60745(13)	474	470.39353(11)
D_J [kHz]		0.00839(39)		
D_{JK} [kHz]		0.0451(54)		0.102(25)
D_K [kHz]		-0.19(14)		
$\mu_a/\mu_b/\mu_c$ (D)	2.5/0.0/0.5	Yes/No/No[g]	5.1/2.4/1	Yes/No/No
σ [kHz][c]		2.8		3.2
N[d]		52		20
Com. gas/sol. [%][e]		100:>95		0/5
Recr. gas/sol. [%]		17(5):50		83:50

[a] *Ab initio* calculations (MP2/6-311++G(d,p)).
[b] Rotational constants (A, B, C); μ_a, μ_b and μ_c are the absolute values of the electric dipole moment components.
[c] The *rms* deviation of the fit.
[d] Number of fitted rotational transitions.
[e] Experimental anomeric population in the commercial sample (Com.) and in the recrystallized sample (Recr.) in gas and solid phase.
[f] Standard error in parenthesis in the units of the last digit.
[g] Experimental observation of *a*-, *b*-, and *c*- *μ*-type transitions for each structure.

It is worth noting the lack of any β-conformer of **GalA,** although this could be due to the α/β population present in the commercial sample. It is quite well-known that α↔β and furanose↔pyranose interconversions (*mutarotation*) are hindered in the gas phase [33], so the lack of this anomer in the solid sample makes unviable its detection in gas phase. To address this point, the anomeric population of the solid sample was investigated thanks to ss-NMR (see Figure VI.1). The analysis of ss-NMR spectra of this commercial sample confirmed our suspicions on the residual proportion of β-anomer (α:β, 95:<5). On the contrary, a higher proportion of the β-anomer in **GlcA** was revealed (α:β, 10:90) . In order to try to observe the **β-GalA** conformers, a portion of the commercial sample of **GalA** (approximately 1 g) was recrystallized in acetic acid following the bibliography[34] and again the population ratio was followed by ss-NMR (Figure VI.1). The recrystallization was successful and we achieved a population ratio (α:β, 50:50) (see Figure VI.1). At this point, we were once again ready to search for this β-anomer in the gas phase. Following the above procedure, **β-GalA 1** was detected in the rotational spectrum (Table VI.2 and Table AVI.11), being predicted the global minimum (Table AVI.4 of Appendix VI). No more conformers were detected in the recorded spectrum, so the discussion below will only be focused on the detected structures.

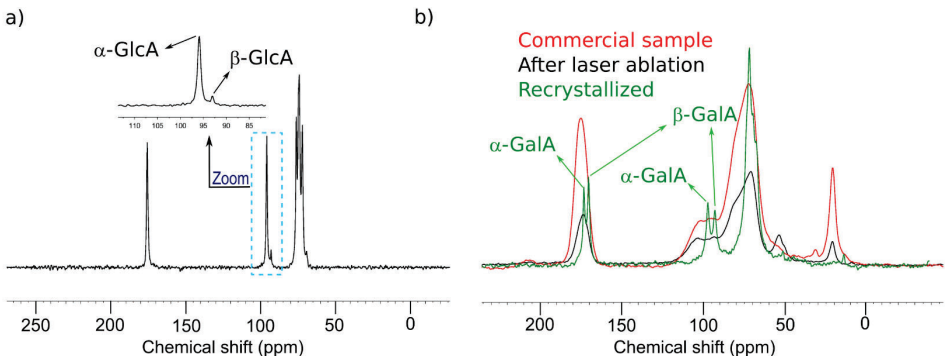


Figure VI.1 a) ss-NMR spectrum of the commercial sample of **GlcA**. b) ss-NMR spectrum of the commercial sample (red line) of **GalA**, the sample after laser ablation process (black line) and the sample recrystallized with acetic acid (green line). The peaks of each anomer are labelled.

Now, we were intrigued to know if the anomeric population in gas phase and in the ss-NMR were similar so we compared them using the experimental rotational spectroscopy signals intensity and theoretical dipole moments for the gas phase results,[35] and the intensity of ss-NMR for the solid state population. The obtained results surprised us because we expected a similar trend for both molecules. However, for **GlcA** the solid and the gas phase populations were very similar, but in the case of **GalA** the two populations differ drastically. We do not

know exactly the reason of this result, but we hypothesize that the population differs because of the different vaporization/degradation efficiency of each anomer during laser vaporization and the subsequent jet cold down process in the rotational spectrometer.

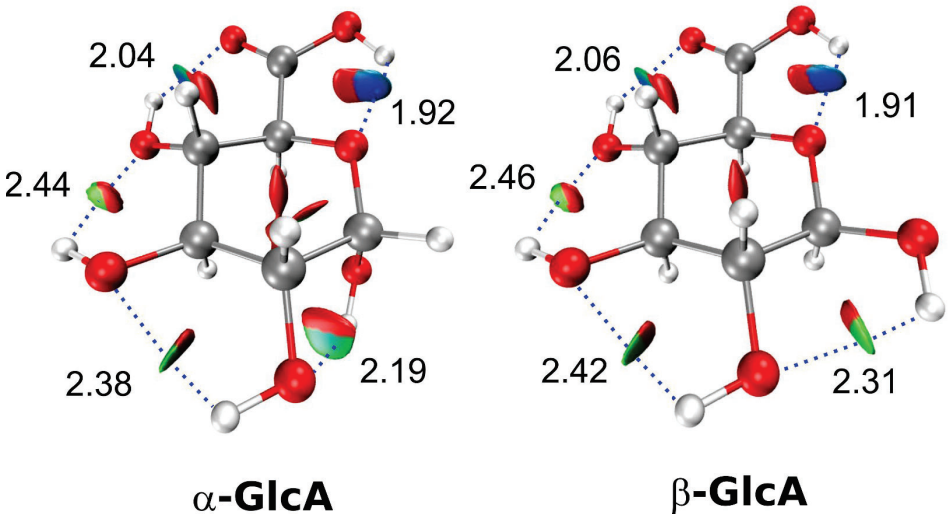


Figure VI.2 Observed conformers and intramolecular interaction (obtained from NCI analysis) of **GlcA** and **GalA**. Dashed lines indicate the hydrogen bond interaction. In the spatial distribution of the non-covalent interactions, the gradient isosurfaces (s = 0.5 a.u.) are coloured on a blue-green-red scale according to the values of the electron density multiplied by the sign of the second Hessian eigenvalue (sign(λ_2)ρ), ranging from -0.03 to 0.01 a.u.. The strongest O-H···O hydrogen bonds (blue) are accompanied by weak dispersive interactions (green shades) and repulsive zones (red).

In light of these experimental data, the conformations that compose the structural landscape of these two molecules can be rationalized in terms of the different interactions that stabilize them, concretely intramolecular hydrogen bonding networks. The spatial distribution of the non-covalent interactions (Figure VI.2 and Figure VI.3) has been mapped according to NCI-Plot analysis.[36] In the **GlcA** molecule, for both α and β anomers assigned in the spectra, a *clockwise* hydrogen bonding network is spread throughout the atoms of the molecule (O1H(ax/eq) → O2H(eq) → O3H(eq) → O4H(eq) → O=COH → O5) and it is responsible for the stabilization of these structures over the rest, with two specially strong hydrogen bonds –O4H(eq) → O=C (MP2/6-311++G(d,p): r_e=2.04-2.06 Å) and O=COH → O5 (MP2/6-311++G(d,p): r_e=1.91-1.92 Å) (Figure VI.2). These results are in contrast with the structure of **α/β-Glc** and α-D-glucosamine (**α-GlcN**), where the hydrogen bond network in the most stable conformers adopts a *counter-clockwise* direction with the hydroxymethyl group stabilised by O6H → O5 interaction (structures of **α/β-Glc** in Figure AVI.8-AVI.9 of Appendix VI).[37,38]

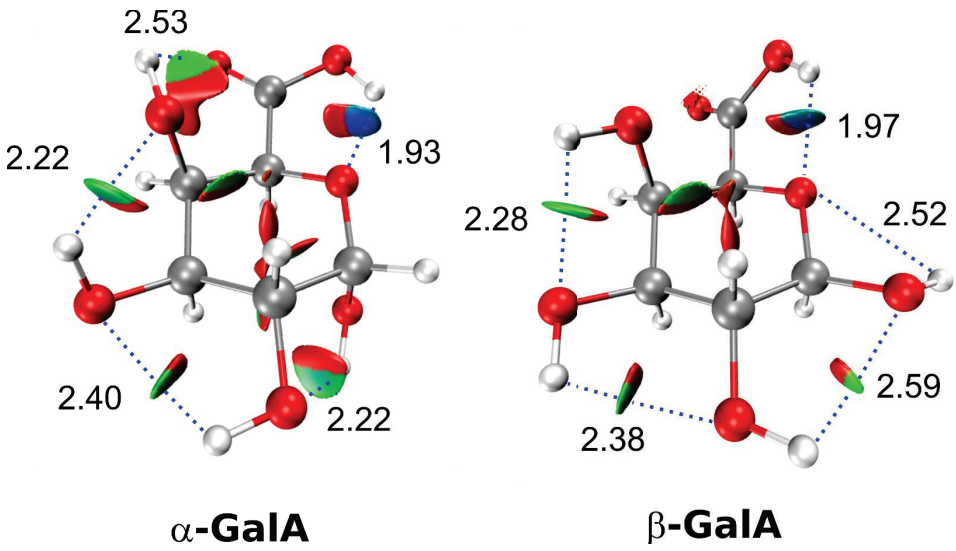


Figure VI.3 Observed conformers and intramolecular interaction (obtained from NCI analysis) of **GalA**. Dashed lines indicate the hydrogen bond interaction. In the spatial distribution of the non-covalent interactions, the gradient isosurfaces (s = 0.5 a.u.) are coloured on a blue-green-red scale according to the values of the electron density multiplied by the sign of the second Hessian eigenvalue (sign(λ_2)ρ), ranging from -0.03 to 0.01 a.u.. The strongest O-H···O hydrogen bonds (blue) are accompanied by weak dispersive interactions (green shades) and repulsive zones (red).

Now, a particular attention should be paid to **GalA** molecule (see Figure VI.3). In the case of the α anomer, it also adopts a *clockwise* intramolecular cooperative hydrogen bonding network, although the O4H(ax) → O=C interaction is much weaker than the analogue **α-GlcA** (MP2/6-311++G(d,p): r_e=2.53 Å). Actually, the hydrogen bond network direction is the same as in the most stable structure of **α/β-Gal**[39] and **β-2-deoxy-Gal** (Chapter IV of this thesis) but in the no oxidized sugars, O4H(ax) → O6 interaction is stronger (MP2/6-311++G(d,p): r_e=1.94-1.96 Å). These experimental data suggest that the attribution of the *clockwise* hydrogen bond direction in the galactose sugar derivatives to a strong O4H → O6 H-bond, should be revised in the future. Interestingly, the most stable structure of **β-GalA** adopts a *counter-clockwise* hydrogen bonding network disposition of (O4H(Ax) → O3H(eq) → O2H(eq) → O1H(eq) and a non-cooperative O=COH → O5 hydrogen bond, breaking the complete network that takes place in the rest of the identified conformers. The *counter-clockwise* hydrogen bond disposition was observed in the most stable structures of non-galactose and non-deoxidized aldohexose derivatives, such as **Glc**, **GlcN**, **2F-Glc** and **2F-Man** (see Chapter IV of this thesis).[37,38]

VI.4 Conclusion

In summary, conventional aldohexoses have exocyclic hydroxymethyl (-CH_2OH) flexible group that lead to three possible configurations.[38–40] In the cases of **GlcA** and **GalA**, instead of the hydroxymethyl group, there is a carboxylic group (–COOH), which leads to a unique disposition governed by a strong hydrogen bond and forming a five member ring interaction between the hydroxyl of the carboxylic group and the oxygen of the ring: O=COH $\rightarrow$ O_{Ring} (MP2/6-311++G(d,p): r_e=1.91-1.97 Å). Structurally, we can conclude that the carboxylic group does not assume the same role as the hydroxymethyl group at C5 present in natural sugars, reducing the conformational panorama in **GlcA** and **GalA** molecules. Further, contrary to the pure **Glc**, in which in both anomers the most table structures adopt a *counter-clockwise* hydrogen bond network, the acid group of the glucuronic acid forces all the structures to adopt a *clockwise* hydrogen bond direction. In the case of **α-GalA**, the reported results suggest the intrinsic preference to the *clockwise* H-bond network and the change of H-bond network direction in **α-GalA** and **β-GalA** anomers have demonstrated experimentally how changes in the anomeric form could affect in all molecular structure.

VI.5 References

[1] M. Jaffé, *Berichte der Dtsch. Chem. Gesellschaft* **1874**, *7*, 1673.
[2] E. Fischer, O. Piloty, *Chem. Ber.* **1891**, *24*, 4214–4225.
[3] R. D. Rosenberg, L. Lam, *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 1218–1222.
[4] L. C. Rosenberg, H. U. Choi, L. H. Tang, T. L. Johnson, S. Pal, C. Webber, A. Reiner, A. R. Poole, *J. Biol. Chem.* **1985**, *260*, 6304–6313.
[5] L.-A. Franson, in *Polysaccharides, Vol. 3* (Ed.: G.O. Aspinall), Academic Press, New York, **1985**, pp. 338–415.
[6] T. E. Timell, *Adv. Carbohydr. Chem.* **1964**, *19*, 247–302.
[7] Z. Scholefield, E. A. Yates, G. Wayne, A. Amour, W. McDowell, J. E. Turnbull, *J. Cell Biol.* **2003**, *163*, 97–107.
[8] V. Tiwari, C. Clement, M. B. Duncan, J. Chen, J. Liu, D. Shukla, *J. Gen. Virol.* **2004**, *85*, 805–809.
[9] B. Casu, I. Vlodavsky, R. D. Sanderson, *Pathophysiol. Haemost. Thromb.* **2008**, *36*, 195–203.
[10] L. Borsig, R. Wong, J. Feramisco, D. R. Nadeau, N. M. Varki, A. Varki, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3352–3357.
[11] S. N. Ma, Z. X. Mao, Y. Wu, M. X. Liang, D. D. Wang, X. Chen, P. an Chang, W. Zhang, J. H. Tang, *Cell Adhes. Migr.* **2020**, *14*, 118–128.
[12] L. Flutto, in *Encycl. Food Sci. Nutr.*, Elsevier, **2003**, pp. 4440–4449.
[13] H. Ohkami, K. Tazawa, I. Yamashita, T. Shimizu, K. Murai, K. Kobashi, M. Fujimaki, **1995**, *86*, 523–529.
[14] D. W. Heitman, W. E. Hardman, I. L. Cameron, *Carcinogenesis* **1992**, *13*, 815–818.
[15] L. Leclere, P. Van Cutsem, C. Michiels, *Front. Pharmacol.* **2013**, *4 OCT*, 1–8.
[16] S. T. Minzanova, V. F. Mironov, D. M. Arkhipova, A. V. Khabibullina, L. G. Mironova, Y. M. Zakirova, V. A. Milyukov, *Polymers (Basel).* **2018**, *10*, 1–31.
[17] F. Brouns, E. Theuwissen, A. Adam, M. Bell, A. Berger, R. P. Mensink, *Eur. J. Clin. Nutr.* **2012**, *66*, 591–599.
[18] D. D. Boehr, R. Nussinov, P. E. Wright, *Nat. Chem. Biol.* **2009**, *5*, 789–796.
[19] D. D. Boehr, P. E. Wright, *Science (80-.).* **2008**, *320*, 1429–1430.
[20] Y. Yu, M. Delbianco, *Chem. - A Eur. J.* **2020**, *26*, 9814–9825.
[21] C. R. Bertozzi, L. L. Kiessling, *Science (80-.).* **2001**, *291*, 2357–2364.
[22] M. Hasan, H. Khakzad, L. Happonen, A. Sundin, J. Unge, U. Mueller, J. Malmström, G. Westergren-Thorsson, L. Malmström, U. Ellervik, A. Malmström, E. Tykesson, *Chem. Sci.* **2021**, *12*, 1869–1885.
[23] H. Y. Y. Yao, J. Q. Wang, J. Y. Yin, S. P. Nie, M. Y. Xie, *Food Res. Int.* **2021**, *143*, 110290.
[24] Y. Meng, X. Shi, L. Cai, S. Zhang, K. Ding, S. Nie, C. Luo, X. Xu, L. Zhang, *Macromolecules* **2018**, *51*, 10150–10159.
[25] M. Plevin, *Encyclopedia of Biophysics*, **2013**.
[26] J. Sastre Toraño, O. Aizpurua-Olaizola, N. Wei, T. Li, L. Unione, G. Jiménez-Osés, F. Corzana, V. J. Somovilla, J. M. Falcon-Perez, G. J. Boons, *Chem. - A Eur. J.* **2021**, *27*, 2149–2154.
[27] W. G. Ferrier, *Acta Crystallogr.* **1963**, *16*, 1023–1031.
[28] T. J. Balle, W. H. Flygare, *Rev. Sci. Instrum.* **1981**, *52*, 33–45.
[29] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[30] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[31] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.
[32] R. K. Thomas, *Phys. Bull.* **1975**, *26*, 501–501.
[33] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J.

Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.
[34] H. S. Isbell, H. L. Frush, *J. Res. Natl. Bur. Stand. (1934).* **1943**, *31*, 33.
[35] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[36] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
[37] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.
[38] I. Peña, L. Kolesniková, C. Cabezas, C. Bermúdez, M. Berdakin, A. Simão, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2014**, *16*, 23244–23250.
[39] I. Peña, C. Cabezas, J. L. Alonso, *Chem. Commun.* **2015**, *51*, 10115–10118.
[40] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.

Chapter VII.

Rotational Spectroscopy of 2-Furoic Acid and Its Dimer

This work was carry out at the University of Alberta. I would like to thank all the members of Yunjie Xu and Wolfgang Jäger groups.

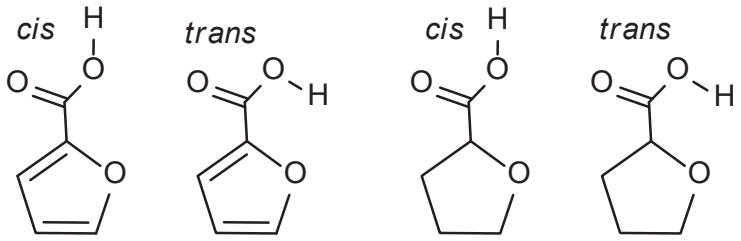
This study has been published and adapted to this chapter. Aran Insausti, Jiarui Ma, Qian Yang, Fan Xie and Yunjie Xu. Rotational Spectroscopy of 2-Furoic Acid and Its Dimer: Conformational Distribution and Double Proton Tunneling. *ChemPhysChem*, **2022**, e202200176

VII.1 Introduction

In this chapter we present the work carry out at the University of Alberta in stay the during 2021-2022 (*Egonlabur* EP_2021_1_0032 grant). Organic carboxylic acids occur widely in nature and may be involved in a variety of biological and environmental processes, such as cellular recognition[1] and the formation of acid rain.[2] Often, the associated COOH functional group may take on the *cis* or *trans*-COOH configuration with the C=O and the OH groups on the same or opposite side, respectively, and may serve as a hydrogen (H)-bond donor and/or acceptor, leading to rich non-covalent binding topologies. This label is shown in Scheme VII.1. With the ability of forming stable binary and even larger aggregates via H-bonding interaction, many carboxylic acids present somehow unconventional features, such as higher boiling and melting points than other organic hydrocarbons with similar molecular weights.

2-furoic acid (FA), the molecule of interest, is an aromatic carboxylic acid that composes a 5-membered aromatic ring and a carboxylic acid group. It is widely applied in food industry as a preservative, a flavoring agent, as well as a food sterilizer. In the early X-ray crystallographic work, one type of centrosymmetric homodimer which consists of two *cis*-COOH FA subunits forming a cyclic double H-bonded ring, was identified.[3] In 2009, two different dimeric structures consisting of two *cis*-COOH subunits were identified in a polarized IR study of FA crystals where the early X-ray report was also re-examined and re-assigned.[4] Interestingly, a combined theoretical and experimental IR and Raman study of FA crystals in 2014 stated that *only* the most stable *cis*-COOH dimer existed in solid,[5] echoing the conclusion of the earlier X-ray work.[3] More recently, an IR matrix isolation investigation revealed two different *cis*-COOH FA conformers with nearly 1 : 1 abundance ratio in Ar or Ne matrixes.[6] Upon near-IR excitation at the $\upsilon_{O\text{-}H}$ overtone of the *cis*-COOH FA monomer,[6] a photoproduced conformer with the *trans*-COOH (see Scheme VII.1) was generated and observed. In addition, a closely related system, the hydrogenated FA, i.e. tetrahydro-2-furoic acid (THFA), was recently studied using rotational spectroscopy and matrix isolation IR and vibrational circular dichroism spectroscopy.[7,8] The THFA monomer was shown to strongly favor the *trans*-COOH conformation, in contrast to FA. When THFA interacts with itself, water, or another chiral molecule such as propylene oxide, the conformational distributions were shifted by different degrees towards the *cis*-COOH conformation,[7,9,10] highlighting the importance of environmental perturbation. Therefore, with the exquisite conformational sensitivity available with rotational spectroscopy, it would be of substantial interest to apply it to examine the conformational structures and relative conformational distribution of the FA monomer and dimer, and how such distribution alters moving from the monomer to dimer.

A further point of significant interest is the possible double proton tunneling event in which the two protons from OH groups transfer to the aldehyde groups of the other binding partner. Such proton transfer events also play an important role in the acid-base characteristics of the nucleobases in Watson-Crick base pairs and may even lead to the formation of some rare tautomers, impacting the stability of DNA.[11] Not surprisingly, researchers have applied several different types of spectroscopies, for example, laser induced fluorescence and IR spectroscopy,[12,13] as well as rotational spectroscopy, to examine double proton tunneling. Closely related to the current study are the reports of rotational spectroscopy of hetero- and homodimers of several carboxylic acids, such as the dimers of formic acid with itself,[14] with propiolic acid,[15,16] acetic acid,[17] and difluroacetic acid,[18] o- and m-anisic acid,[19,20] 3,3,3-trifluoro-2-(trifluoromethyl)propanoic acid[21] and perfuorobutyric acid,[22] benzoic acid,[23] as well as the acrylic acid homodimer.[24] It is interesting to note that tunneling splittings were detected only in some and not in others including the THFA homodimer. Generally, the rotational spectroscopic data of tunnelling splittings are high quality data for benchmarking different approaches, such as the instanton theory[25] and the simplified one-dimensional approach,[26] for quantifying such splittings, a research topic of significant current interest.



Scheme VII.1 The *cis*- and *trans*-COOH labels of FA and THFA.

In the present study, we applied chirped pulse Fourier transform microwave (CP-FTMW) spectroscopy to address whether the *trans*-COOH conformer of the FA monomer can be observed directly without IR pumping. In terms of the FA dimer, we focused on two aspects: 1) to identify the associated geometries theoretically and experimentally using CP-FTMW spectroscopy; and 2) to analyze the possible double proton tunneling event in the dimer and apply a recently developed, simplistic computational procedure[26] to evaluate the tunneling splitting caused by proton transfer.

VII.2 Methodology

Rotational Spectroscopy

This work was carry out in the Proffesor's Xu group in the University of Alberta during the *Egonlabur* stay. So Alberta's University 2-6 GHz CP-FTMW spectrometer was utilized for the rotational measurements.[27,28] The spectrometer was built based on a similar instrument by Pate and co-workers and basic experimental details has been explained in Chapter 2 of this Thesis.[29,30] Briefly, a 12 Gs s^{-1} arbitrary waveform generator (AWG) generated a 2-6 GHz, 1 μs chirped pulse which was then amplified by a 400 W, 2.5-7.5 GHz traveling wave tube (TWT) amplifier. Two horn antennas located in the vacuum chamber were used to broadcast the amplified MW pulse and to receive the free induction decay (FID) signal. The FID signal was then recorded and digitized using a 25 Gs s^{-1} oscilloscope. For the 6-8 GHz measurements, a Fabry-Perot based resonant cavity spectrometer[31] was used (The experimental set-up is very similar as cav-FTMW presented in the Chapter II of this Thesis). For the 8-12GHz measurements, a different CP-FTMW set-up was used. The chirp pulse generated in the AWG in the 4-6 GHz range was doubled to the 8-12 GHz range. A 4 μs chirped pulse with 1GHz bandwidth was amplified with a 20 W solid state amplifier.

In the 2-6 (8-12) GHz CP-FTMW experiments, for each gas pulse, the signals of 15 (6) excitation/detection cycles were measured and a total of about 700k (200k) FIDs were averaged and then Fourier transform was performed to get a frequency spectrum. For the cavity measurements, the frequency uncertainty is approximately ~1 kHz and the corresponding values for the CP-FTMW data are ~10 kHz.

FA (98% purity) purchased from Millipore-Sigma was used without further purification. For the deuterated species, the FA solid was dissolved in D_2O and the resulting solution was sonicated for 2 hours at approximately 60 to 70 ºC and then dried. For the CP-FTMW measurements, the FA solid was placed directly inside a modified General Valve nozzle cap and heated to ~ 110 ºC and the helium backing pressure used was about 50 psi. For the cavity measurements, the FA solid was placed inside a stainless steel sample compartment, right behind the nozzle, and heated to about ~ 110 ºC with a neon backing pressure of 25 psi.

Computational Methods

The systematic conformational searches were carried out using the CREST (conformer-rotamer ensemble sampling tool) program developed based on GFN-xTB, a semiempirical tight-binding (TB) quantum chemistry code.[32,33] All subsequent geometry optimization and harmonic frequency calculations were done with the Gaussian 16 suite of programs[34] at the B3LYP-D3(BJ)/def2-TZVP level of theory with the dispersion correction (D3) and Becke-Johnson (BJ)

damping function,[35] and combined with the def2-TZVP basis set.[36,37] Finally, the proton transfer tunnelling splittings were computed using Orca 4 and the related python software.[26,38]

VII.3 Results and Discussion

The FA Monomer

For the FA monomer, 14 CREST[32,33] candidates initially identified (see the section on 'Methodology') were re-optimized at the B3LYP-D3(BJ)/def2-TZVP level of theory, leading to the four stable structures shown in Figure VII.1. The related spectroscopic constants are summarized in Table VII.1. Each structure is labeled with two identifiers: the first two letters represent the position of carboxyl with respect to the furan ring, with ap or sp indicating that the dihedral angle $\angle_{O\text{ring-}C\text{-}C=O}$ is ~180º or 0º, respectively, the second is associated with the cis- or trans-COOH configuration as already discussed before. For simplicity, these four conformers are also simply named as **I**, **II**, **III**, and **IV**, in decreasing stability, with **I** being the most stable one.

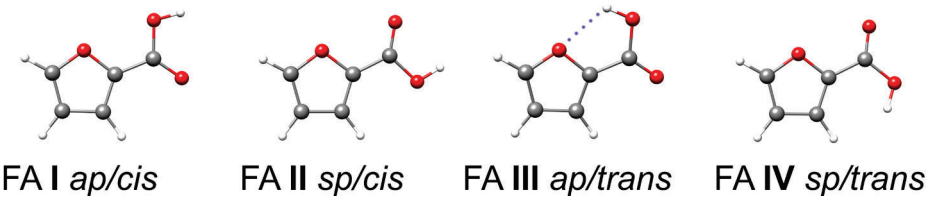


Figure VII.1 Optimized geometries of the four most stable conformers of FA. The close contact between OH and the heteroatom of the ring is indicated with a dashed line.

As can be seen in Table VII.1, FA **I** and **II** have strong *a*-type transitions, and their patterns were straightforwardly recognized in the 2-6 GHz spectrum and the 8-12 GHz spectrum. Subsequentially, the *b*-type transitions of **I** and **II** were also assigned, whereas no c-type transitions were observed. FA **III**, which is about 8.9 kJ mol^{-1} less stable than the global minimum, was predicted to have a percentage abundance of ~ 3% based on the Boltzmann distribution at the source temperature of 383 K. Thanks to its very large *a*-dipole component, the *a*-type transitions of **III** were also assigned. Some example sections of the spectra are depicted in Figure VII.2. FA **IV**, which is over 20 kJ mol^{-1} less stable than **I**, was not observed experimentally because of its very low abundance of ~0.1%. We used Watson's *S*-reduction semi-rigid rotor Hamiltonian and the *Pickett* program[39] to fit the three sets of measured transitions of **I**, **II** and **III**. The resulting spectroscopic parameters are also collected in the second half of Table VII.1.

Table VII.1 Relative raw (ΔE_e) and anharmonic *ZPE* corrected (ΔE_0) energies, Gibbs energy at 383 K ($\triangle G_0$) (in kJ mol^{-1}), rotational constants (in MHz), and electric dipole moment components (in Debye) of the four FA conformers and the corresponding experimental spectroscopic constants in the lower half.

	I	**II**	**III**	**IV**
$\triangle E_e$	0	0.6	9.1	25.2
$\triangle E_0$	0	0.6	8.8	23.8
$\triangle G_0$	0.0	0.5	8.9	21.2
A	5270	5271	5195	5213
B	1639	1634	1653	1633
C	1250	1247	1254	1246
μ_a	2.0	1.9	4.9	4.4
μ_b	0.8	2.2	1.7	3.4
μ_c	0.0	0.0	0.0	0.3
		Experimental[a]		
A	5239.81504(73)	5238.51317(86)	5167.20252(91)	-
B	1638.51689(23)	1633.13797(27)	1650.88707(27)	-
C	1248.75945(23)	1245.55338(26)	1251.65659(25)	-
μ[b]	$\mu_a > \mu_b$ no μ_c	$\mu_a \approx \mu_b$ no μ_c	$\mu_a > \mu_b$ no μ_c	-
N[c]	35	38	28	-
σ[c]	5.6	7.1	5.0	-
Ab%[c]	51	47	2	-

[a] The fits were done without centrifugal distortion constants since the system is quite rigid.

[b] The estimated relative magnitudes of the transition dipole moment components.

[c] N is the number of lines in the fit; σ is the root-mean-square deviation of the fit and Ab% is the estimated percentage conformational abundance.

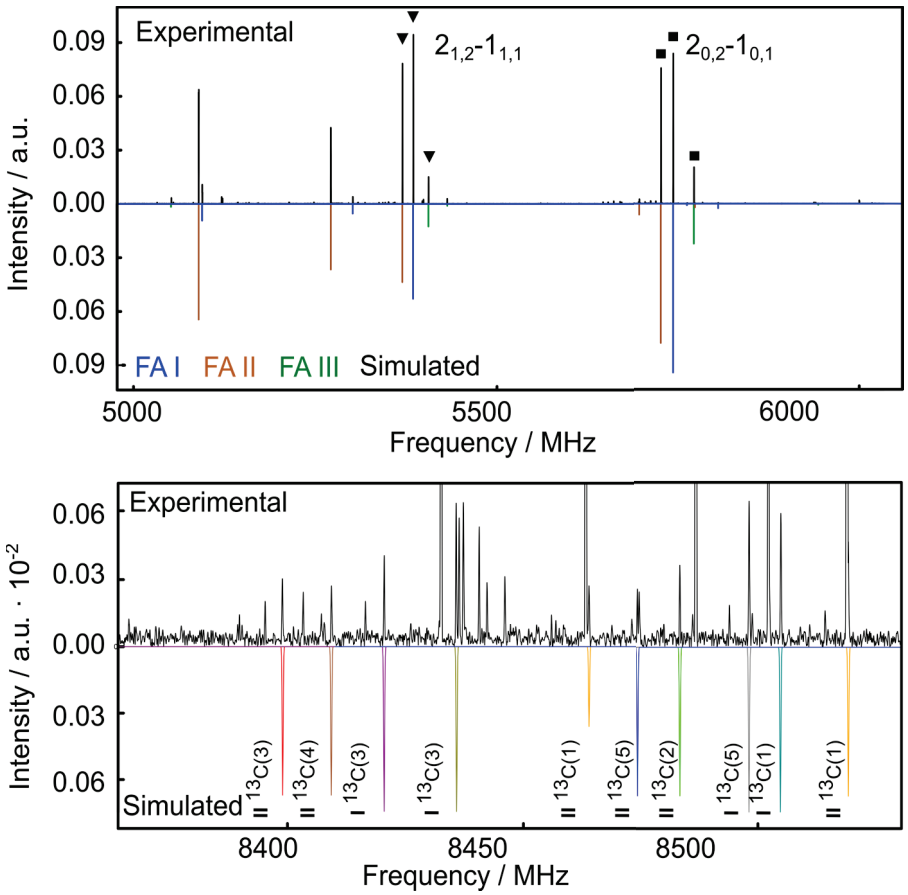
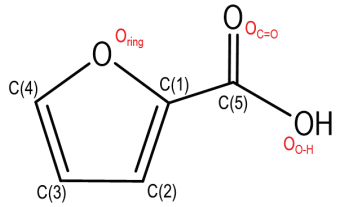


Figure VII.2 A section of the experimental broadband spectrum of FA and the simulated spectra of the conformers assigned by using the experimental spectroscopic constants, an estimated rotational temperature of 1 K, and the theoretical dipole moment components, as well as the relative experimental abundances of the conformers. See the main text for detail.

The high signal-to-noise ratio achieved in the experimental spectra allowed one to assign almost all ^{13}C-isotopologues in its natural abundance (1.1%) for **I** and **II**. The only exception is the ^{13}C(1) isotopologues of **I** whose rotational transitions are essentially on top of the parent transitions because C(1) is located very close to the centre of mass, rendering it impossible the detection of rotational transitions of **I** ^{13}C(1). In addition, the acid ^{2}H isotopologues of **I**, **II** and **III** were also assigned using the enrich deuterated sample. The same spectroscopic fitting procedure used for the parent species was followed for all isotopic species. The resulting spectroscopic constants of all the rarer isotopologues are summarized in Table VII.2, while the measured experimental frequencies of all monomeric species are listed in Table AVII.1-AVII.15, appendix.

Table VII.1 Experimental rotational constants of **I** and **II** ^{13}C and **I**, **II**, and **III** D isotopologues.[a]



	I $^{13}C(1)$	**I** $^{13}C(2)$	**I** $^{13}C(3)$	**I** $^{13}C(4)$	**I** $^{13}C(5)$
A	N/A	5162.0485(84)	5218.4353(62)	5218.124(15)	5239.4251(95)
B	N/A	1633.9486(15)	1612.2135(16)	1615.8896(28)	1628.4613(17)
C	N/A	1241.7767(14)	1232.2341(12)	1234.0719(22)	1242.8880(12)
N	N/A	7	9	6	9
σ	N/A	5.9	7.0	10.9	7.4
	II $^{13}C(1)$	**II** $^{13}C(2)$	**II** $^{13}C(3)$	**II** $^{13}C(4)$	**II** $^{13}C(5)$
A	5238.3374(43)[d]	5164.2378(20)	5213.2807(28)	5215.2948(25)	5238.3619(27)
B	1633.2103(11)	1629.14021(47)	1607.67689(60)	1609.88263(56)	1623.24533(57)
C	1245.5941(15)	1238.99845(38)	1229.29041(68)	1230.69401(52)	1239.78259(56)
N	8	15	13	14	15
σ	5.3	4.6	6.2	5.7	6.3
	I D at OH	**II** D at OH	**III** D at OH		
A	5185.3218(16)	5195.36495(89)	5006.21657(98)		
B	1596.18967(42)	1588.32847(28)	1642.66320(30)		
C	1221.04512(37)	1216.98884(26)	1237.34286(28)		
X_{aa}	0.2641(33)	0.2638(30)	0.0753(60)		
X_{bb}	[-0.1164][b]	[-0.1097]	0.0925(61)		
X_{cc}	[-0.1476]	[-0.1531]	-0.1679(61)		
N	57	84	76		
σ	6.1	6.7	6.0		

[a] *A*, *B* and *C* are rotational constants in MHz. *N* is the number of transitions included in the fit. σ is the standard deviation of the fit. All centrifugal distortion constants are fixed at the same as the corresponding parent species.
[b] Fixed for theoretical values.

The isotopic rotational constants allow one to determine the carbon backbone substitution (r_s) and effective (r_0) of **II** using the *Kraitchman's* equations and mass-dependence molecular structures theory, respectively. In addition, the theoretical anharmonic vibrational contributions of the rotational constants were used to obtain semi-experimental equilibrium rotational constants[40–43] which in turn provided the semi-experimental equilibrium (r_e^{SE}) structures of **II**. The coordinates and structural parameters obtained are summarized in Table AVII.16-AVII.19, appendix. The experimental ground-state inertial defect is Δ_0 = -0.1799 uÅ^2 for the parent species; after the rot-vibrational correction, based on the B3LYP-D3(BJ)/def2-TZVP anharmonic calculations, it becomes −0.0201 uÅ^2 in the r_e^{SE} structural analysis, consistent with the expectation, i.e. 0 uÅ^2, for a planar structure of **II** at equilibrium.

Using the predicted electric dipole moment components and an estimated rotational temperature of 1 K, the experimental conformational percentage abundances of **I**, **II** and **III** were estimated to be 51%, 47% and 2%,[27,44] in

reasonably agreement with the Boltzmann percentages of 52%, 45% and 3% calculated based on the FA source temperature of 383 K and their relative free energies. The slight decrease in the percentages of **II** and **III** suggests perhaps some minor cooling effect taking place in the jet expansion, although the experimental uncertainty in the percentage values may also provide an adequate explanation.

Clearly, the above results demonstrate that *cis*-COOH conformers of FA are by far more preferred than the *trans*-COOH conformer, in stark contrast to THFA where the opposite trend was observed experimentally.[7] The current results are common for most carboxylic acids, especially where there are no strong intramolecular H-bonds or close contacts to stabilize the *trans*-COOH conformers. The *cis*-COOH conformational preference has been called the Z effect, although its origin is still being debated.[45,46] To map out the different non-covalent intramolecular interactions in FA and THFA, noncovalent interaction (NCI) analyses[47] were performed and the results are provided in Figure VII.3. Clearly, the much stronger intramolecular OH···O_{ring} interaction in THFA **I** and **II** than in FA **III** is responsible for the strong preference of *trans*-COOH configuration in THFA.

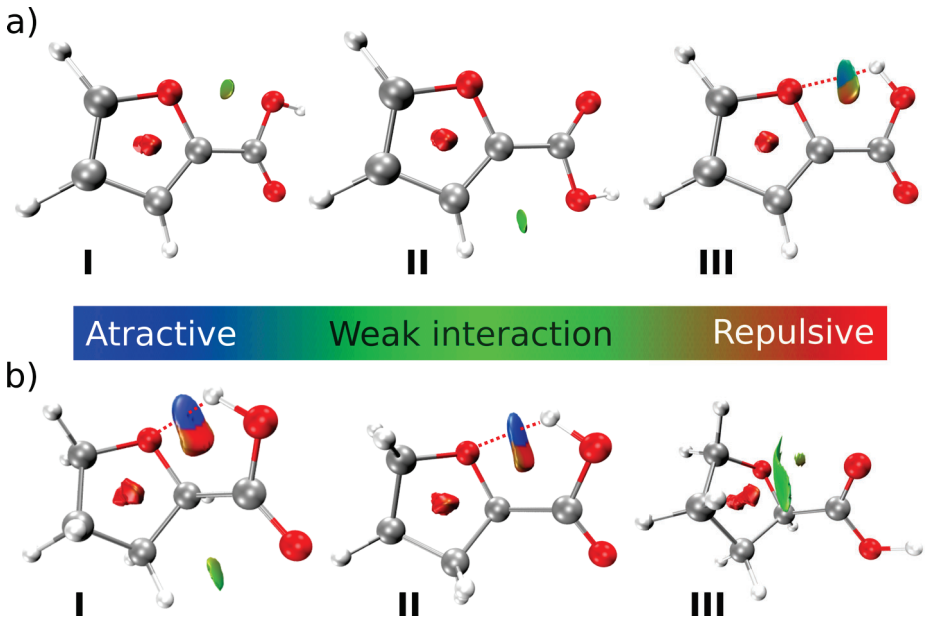


Figure VII.3 NCI analyses of the non-covalent interactions present in a) the three observed FA conformers, **I**, **II**, and **III**; and b) the three observed THFA conformers **I**, **II**, and **III.** The NCI iso-surfaces (s = 0.60) were calculated at the B3LYP-D3(BJ)/def2-TZVP level of theory.

To quantify the energies associated with the O-H···O_{ring} contacts, we also utilized the NBO analyses where the energy of a H-bond (or contact) can be obtained by considering the orbitals directly engaged in such interaction, i.e. n_O and σ^*_{OH} for the O-H···O_{ring} contacts.[48] The results (Table AVII.20, appendix) show that in THFA **I** and **II**, OH···O_{ring} H-bond energy is about 19 to 21 kJ mol^{-1}, whereas this value drops significantly to about 2.4 kJ Mol^{-1} in FA **III.** This trend was also captured by using the Intrinsic Bond Strength Index (IBSI) results[49] where *trans*-THFA **I** and **II** have the IBSI value of 0.047-0.049, higher than 0.025 of FA **III**.

Overall, three conformers of the FA monomer were detected experimentally in the current study, consistent with the result reported previously with the matrix isolation IR study.[6] The difference is that one could directly observe *trans*-COOH FA **III** without the need of laser pumping utilized previously,[6] thanks to the high sensitivity of the rotational spectroscopic instrument. The isotopic data and the semi-experimental equilibrium structural analysis show that FA **II** is planar at equilibrium. FA and its hydrogenated counterpart, TFHA, exhibit very different OH···O_{ring} interactions, leading to vastly different conformation preference.

The FA Dimer

After removing all known transitions of FA **I**, **II** and **III** and their rarer isotopologues, some weaker spectral patterns emerged, suggesting the possible presence of the FA dimer. It was known in the previous THFA dimer study,[10] some barriers, for example those associated with the *cis*- and *trans*-COOH conversion, are difficult to overcome during a CREST search. Therefore, different FA conformational starting geometries were used to ensure a complete survey of the binary FA landscape. We name the resulting binary geometries based on its monomeric subunits. Table VII.3 shows the calculated relative energies and the spectroscopic constants of the five low energy binary conformers. Interestingly, it appears that the non-covalent interaction of FA with itself favors **II** over **I**, making **II-II** more stable than **I-I**. The three most stable binary FA geometries, which contain a double H-bonded ring, are depicted in Figure VII.4.

Thermodynamically, **I-I**, **I-II**, and **II-II** are similar in their stability and are strongly favored over **I-III** and **II-III**, whereas **III-III** are the least stable one. The geometries of **I-III** and **II-III** are provided in Figure AVII.2, appendix. For **I-I** and **II-II** which are formed by the same monomeric subunits, the associated geometries are of C_{2h} symmetry with a σ_h symmetry plane and an inversion center. Consequently, it would not be possible to detect their rotational spectra because of a zero dipole moment. The double proton transfer event would convert one binary conformer to another, as depicted in Figure VII.3. For **I-II** which belongs to the C_s group with a non-zero *b*-type dipole component, the double proton transfer event would convert **I-II** to its equivalent geometry as shown also in Figure VII.3. One would expect to observe only **I-II** based on a thermodynamical model.

On the other hand, the previous jet-cooled rotational spectroscopic study of the THFA dimer demonstrated that both thermodynamic and kinetic factors play important roles in the final abundances of different binary conformers.[10] The thermodynamically favored binary conformers with two *cis*-COOH THFA subunits were detected, so were the kinetically favored and thermodynamically moderate binary conformers with one *trans*- and one *cis*-COOH THFA subunits. Interestingly, the most strongly kinetically favored binary conformers with two *trans*-COOH subunits were not observed in a jet expansion, but were detected in the matrix isolation IR and VCD spectra.[8] Assuming a kinetically controlled process, one would expect the abundances of **I-I**, **I-II**, **I-III**, **II-II**, **II-III**, and **III-III** to be 26%, 48%, 2%, 22%, 2% and 0%, respectively, based on the experimental conformational abundances of **I** (51%), **II** (47%) and **III** (2%). Therefore, the situation with the FA dimer is greatly simplified because there is no competition between the kinetic and thermodynamic processes, i.e. both processes favor binary geometries containing only *cis*-COOH conformers, **I-I**, **I-II**, and **II-II**.

Table VII.3 Calculated relative raw (ΔE_e) and *ZPE* corrected (ΔE_0) energies and raw (D_e) and *ZPE/BSSE* corrected (D_0) binding energies (in kJ mol^{-1}), rotational constants (in MHz), and electric dipole moment components (in Debye) of the three binary FA conformer. [a]

$(FA)_2$	**II-II**	**I-I**	**I-II**	**I-III**	**II-III**
A	2614	2630	2612	1318	1298
B	167	167	167	198	199
C	157	157	157	172	173
$\mu_a/\mu_b/\mu_c$	0/0/0	0/0/0	0/1.2/0	2.1/3.2/0	2.0/2.1/0
ΔE_e	0	0.4	0.3	40.9	40.1
ΔE_0	0	2.1	1.2	40.4	39.6
D_e[a]	79.6	78.0	78.7	47.3	48.8
D_0[a]	74.4	75.2	74.7	25.6	27.1

[a] The equations used to obtain D_e and D_0 are given in Point AVI.1, appendix.

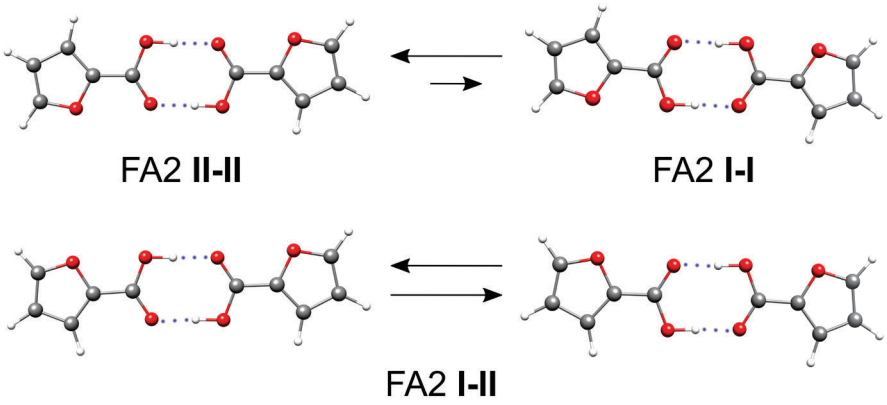


Figure VII.4 Optimized geometries of the three most stable conformers of the FA dimer at the B3LYP-D3(BJ)/def2-TZVP level of theory. The arrows show the interconversion pathways in the double proton transfer event. The two H atoms involved in double proton transfer are labelled as H_Z and H_E.

With the predicted *b*-type transitions of **I-II**, it was immediately obvious that many rotational transitions split into doublets. The spin statistical weights for the double proton transfer systems were established as 9/7 in favor of even K_a+K_c (lower level) transitions in the v=0 tunneling state and in favor of odd K_a+K_c (lower level) transitions in the v =1 tunneling state.[24] The cross-tunneling transitions, on the other hand, are driven by the *a*-type dipole component which is zero in the current case, rendering their non-observation. The 9/7 intensity character was very helpful in differentiating the two tunneling components, since an excited state component can be higher or lower in frequency relative to the related ground state component. Overall, a set of 180 transitions in the 2-12 GHz region were measured and fitted simultaneously using the *Pickett's* spectroscopic fitting program.[39] The coupled Hamiltonian used has the following expression:

$$H = \sum_i H_i^R + \Delta E_{01} + F_{ab} \text{ x } (P_a P_b + P_b P_a) \quad (1)$$

H_i^Rrepresents the rotational Hamiltonian for the state *i*, ΔE_{01} is the energy difference between the v = 0 and v = 1 state and F_{ab} is the rotation-vibration coupling parameter between the two states. For the current system, it appears that no centrifugal distortional constants are needed to obtain a good fit, indicating the rigidness of the dimer. The experimental spectroscopic parameters obtained are given in Table VII.4 and the experimental transition frequencies are given in Table AVII.21, appendix.

Table VII.4 Experimental spectroscopic constants (in MHz) of all isotopologues of **I-II** measured.[a]

	HH	D_ZH_E	H_ZD_E	DD
A_0	2592.40067(24)	2577.28179(85)	2572.28060(77)	2557.34150(95)
B_0	166.70081(12)	166.46913(17)	66.475770(86)	166.22106(14)
C_0	156.691493(79)	156.43200(29)	156.418791(83)	156.17785(12)
A_1	2592.38604(17)	2577.27917(50)	2572.27583(72)	2557.32015(96)
B_1	166.69857(13)	166.46936(10)	66.475882(86)	166.21898(14)
C_1	156.689428(45)	156.431260(90)	156.41879(11)	156.17624(12)
ΔE_{01}	1056.0(10)	116.5(60)	141.2(28)	-
F_{ab}	-10.347(15)	[-10.347]	[-10.347]	-
N^a	180	89	81	104
σ	6.9	5.6	65	6.5

[a] The fits were done with out the centrifugal distortion constants. A, B and C are rotational constants in MHz where the subscripts of 0 and 1 correspond to the ground and the tunneling excited state, respectively. $\Delta E_{01}$is tunneling splitting in MHz. F_{ab} is the Coriolis coupling constant in MHz. *N* is the number of transitions included in the fit. σ is the standard deviation of the fit in kHz

In addition, the deuterated binary FA species were also studied where different tunnelling H atom(s) were substituted. The four possible isotopologues are labeled as HH (the parent species), D_ZH_E, H_ZD_E and DD where the Z and E positions are indicated in Figure VII.4. A portion of the broadband spectrum is depicted in Figure VII.5 where some transitions of the mono and di-deuterated binary FA species are present.

The same fitting procedure was applied to the deuterated species, and two issues were noticed. First, in the case of the monodeuterated species, since the ΔE_{01} values are considerably smaller than that of the parent (HH), the simultaneous fitting of ΔE_{01} and F_{ab} were unsuccessful. For this reason, the F_{ab} value was fixed at that of the parent species in the fitting procedure for the monodeuterated species. Second, for the DD species, the double proton tunneling splittings were only partially resolved in some transitions. As a result, we could not obtain the ΔE_{01} and F_{ab} parameters of the DD species. The nuclear quadruple coupling splittings were negligible, based on the predicted nuclear quadrupole coupling constants of the D species. Therefore, only the rotational constants in the ground and the excited tunneling states were included in the fitting procedure of the DD species. All experimental constants obtained are also summarized in Table AVII.4, while the corresponding transition frequencies are listed in Table AVII.22-AVII.25**,** appendix.

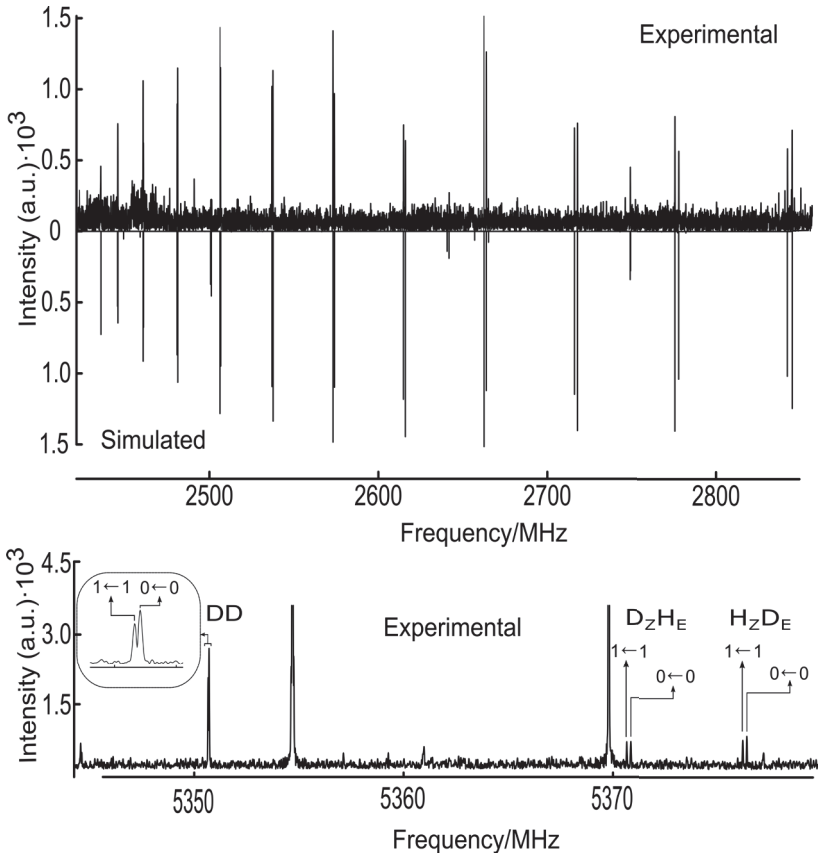


Figure VII.5 A section of the experimental broadband spectrum of FA and the simulated spectra of the FA dimers assigned by using the experimental spectroscopic constants, an estimated rotational temperature of 1 K, and the theoretical dipole moment.

In Table VII.5, we compare the ΔE_{01} values of a range of binary carboxylic acid complexes reported previously with that of the FA dimer. The tunnelling band origin value of the FA dimer is among the biggest, very similar to that of the benzoic acid homodimer[50] and somewhat bigger than that of the acrylic acid homodimer.[24] Interestingly, it appears that the ΔE_{01} value tends to be bigger in the systems whose carboxylic acid side chain is conjugated with the carbon backbone, a subject which is worthy of further theoretical investigations.

Table VII.5 The tunneling splitting, ΔE_{01} (in MHz), of several carboxylic acid dimer specie.

Molecule	ΔE_{01} HH	Ref.
$(C6H5COOH)_2$	1114.0(10)	[50]
$(HCOOH)_2$	474(12)	[12]
$(HCOOHD)_2$	331.6(5)	[14]
2x(HCOOH) – F-benzene	267.6080(13)	[51]
HCOOH − HC2COOH	291.428(5)	[16]
HCOOH − CH3COOH	250.4442(1)	[17]
acrylic acid dimer	880.6(6)	[24]
BzOH − HCOOH	548.72(6)	[23]
$(FA)_2$	1056.0(10)	This work

While it is known that the experimental ΔE_{01} value of the double proton transfer can be related to the tunnelling barrier, additional information is needed to extract an experimentally estimated barrier in the FA dimer. To achieved that, we utilized a recently developed simplistic computational procedure[26] for computing tunnelling splittings for large amplitude motions such as the double proton transfer event. The code provides a solution of an effective one-dimensional Schrödinger equation with an effective mass and an effective PES composed of electronic and harmonic zero-point vibrational energies of small amplitude motions in the molecule. For the FA dimer **I-II**, a double proton transfer relaxed PES was generated by simultaneously changing the O-H distance of each monomer, following the formic acid dimer example,[26] using the *Orca* software.[38] The proton transfer dynamics was then calculated using the publicly available Python code.[26] This procedure was used recently for a complex consists of a formic acid dimer and a fluorobenzene molecule.[51] For the reaction coordinate, the following expression was used:

$$\xi = \big(r(O-H) + r(O-H) - r(O\cdots HO) - r(O\cdots HO)\big)/2\sqrt{2} \quad (2)$$

The obtained values of ΔE_{01} (tunneling splitting), E_{HB} (raw barrier), and E_{HB}^{Eff} (ZPE corrected barrier) at several different DFT calculations are reported in the Table VII.6. At first slight, the E_{HB}, and E_{HB}^{Eff} and the resulting ΔE_{01} values obtained depend strongly on the functionals used, and the precision of the theoretical results appear to be too low to reproduce the experimental data. Similar variations in the theoretical results of proton transfer were reported for the dimers of formic acid and maloic acid, formic acid dimer with fluorobenzene, and also in the large amplitude motion of thiophenol dimers.[14,51,52] Generally, while some usual trend could be predicted, the extremely accurate results provided by rotational spectroscopy are very difficult to reproduce computationally.

Table VII.6 The theoretical ΔE_{01} (tunneling splitting), E_{HB} (raw barrier), and E_{HB}^{Eff} (ZPE corrected barrier), and E_{HB}^{Eff} (scaled) calculated with a range of functional and the def2-TZVP basis set.[a]

Functional	ΔE_{01}	E_{HB}	E_{HB}^{Eff}	E_{HB}^{Eff} (scaled)[b]
b1lyp	$4.80 \cdot 10^{-1}$	2289	1108	360
b3lyp	$1.89 \cdot 10$	2043	776	394
b3lyp-gd3bj	$1.90 \cdot 10^{2}$	2016	629	456
b97-3c	$3.25 \cdot 10^{3}$	1554	384	492
cam-b3lyp	$7.77 \cdot 10$	1882	632	406
cam-b3lyp-gd3bj	$2.76 \cdot 10^{2}$	1882	563	442
mPW1LYP	1.99	2220	1008	391
x3lyp	$7.79 \cdot 10$	2018	709	447
wB97	$1.00 \cdot 10^{-2}$	2936	1486	333

[a] ΔE_{01} is in MHz and other energy terms are in cm^{-1}.
[b] E_{HB}^{Eff} (scaled) is obtained with a scaled PES to reproduce the experimental tunneling splitting. See the main text for details.

Recently, Medel summarized a range of simple models for tunneling splitting, showing that the most crucial parameter for tunneling splitting is the barrier height.[53] Therefore, we scaled the effective PESs obtained by different scaling factors and solved the Schrödinger equation in order to reproduce the experimental tunneling splitting. Somewhat surprisingly, the final E_{HB}^{Eff} barriers obtained from all the levels of theory are fairly consistent, in the range of 333~492 cm^{-1}. The results are also summarized in Table VII.5. The closest agreement with the experiment before scaling is provided with the cam-b3lyp-gd3bj functional, and the scaled result with the same functional produced an E_{HB}^{Eff} value of approximately 442 cm^{-1}, which is regarded as the best estimate for the double proton transfer barrier of **I-II** in the current work.

VII.4 Conclusions

The structural and double proton transfer properties the FA monomer and dimer were interrogated using jet-cooled rotational spectroscopy and quantum chemistry calculations. Three monomeric conformers were identified experimentally with a strong preference for the *cis*-COOH configuration. The high energy *trans*-COOH FA **III** conformer was also observed, in contrast to the previous matrix isolation IR experiment where **III** was only detected with the assistant of laser excitation. The rotational transitions of the heterodimer, **I-II**, exhibit characteristic proton transfer tunnelling splittings, which were analysed, leading to a fitted ΔE_{01} value of 1056.0 (12) MHz. By using a recently developed one-dimensional tunnelling model, it was possible to estimate the zero point energy corrected, double proton transfer barrier, E_{HB}^{Eff} to be ~ 442 cm^{-1} using the experimental ΔE_{01} value. In general, different levels of theory produce very different tunnelling splittings, suggesting that more theoretical developments would be needed in order to reproduce the accurate experimental tunnelling data of the double proton transfer event.

VII.5 References

[1] K. Ahmed, *Front. Endocrinol. (Lausanne)*. **2011**, *2*, 1–12.
[2] W. C. Keene, J. N. Galloway, *Atmos. Environ.* **1984**, *18*, 2491–2497.
[3] C. J. Gilmore, P. R. Mallinson, J. C. Speakman, *Acta Cryst.* **1983**, *1983*, 1111–1113.
[4] H. T. Flakus, M. Jabłońska, J. Kusz, *Vib. Spectrosc.* **2009**, *49*, 174–182.
[5] H. Ghalla, N. Issaoui, M. V. Castillo, S. A. Brandán, H. T. Flakus, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2014**, *121*, 623–631.
[6] A. Halasa, L. Lapinski, I. Reva, H. Rostkowska, R. Fausto, M. J. Nowak, *J. Phys. Chem. A* **2015**, *119*, 1037–1047.
[7] F. Xie, X. Ng, N. A. Seifert, J. Thomas, W. Jäger, Y. Xu, *J. Chem. Phys.* **2018**, *149*, 224306.
[8] Y. Yang, J. Cheramy, Y. Xu, *ChemPhysChem* **2021**, *22*, 1336–1343.
[9] F. Xie, S. Mahendiran, N. A. Seifert, Y. Xu, *Phys. Chem. Chem. Phys.* **2021**, *23*, 3820–3825.
[10] F. Xie, N. A. Seifert, W. Jäger, Y. Xu, *Angew. Chemie - Int. Ed.* **2020**, *59*, 15703–15710.
[11] R. Srivastava, *Front. Chem.* **2019**, *7*, 1–17.
[12] M. Ortlieb, M. Havenith, *J. Phys. Chem. A* **2007**, *111*, 7355–7363.
[13] K. G. Goroya, Y. Zhu, P. Sun, C. Duan, *J. Chem. Phys.* **2014**, *140*, 164311.
[14] W. Li, L. Evangelisti, Q. Gou, W. Caminati, R. Meyer, *Angew. Chemie - Int. Ed.* **2019**, *58*, 859–865.
[15] A. M. Daly, P. R. Bunker, S. G. Kukolich, *J. Chem. Phys.* **2010**, *132*, 201101.
[16] A. M. Daly, K. O. Douglass, L. C. Sarkozy, J. L. Neill, M. T. Muckle, D. P. Zaleski, B. H. Pate, S. G. Kukolich, *J. Chem. Phys.* **2011**, *135*, 154304.
[17] M. C. D. Tayler, B. Ouyang, B. J. Howard, *J. Chem. Phys.* **2011**, *134*, 054316.
[18] Q. Gou, G. Feng, L. Evangelisti, W. Caminati, *Chem. Phys. Lett.* **2014**, *591*, 301–305.
[19] A. Macario, S. Blanco, J. Thomas, Y. Xu, J. C. López, *J. Phys. Chem. A* **2019**, *123*, 6772–6780.
[20] A. Macario, S. Blanco, J. Thomas, Y. Xu, J. C. López, *Chem. - A Eur. J.* **2019**, *25*, 12325–12331.
[21] J. Thomas, M. J. Carrillo, A. Serrato, F. Xie, W. Jäger, Y. Xu, W. Lin, *Mol. Phys.* **2019**, *117*, 1193–1199.
[22] J. Thomas, M. J. Carrillo, A. Serrato, W. Lin, W. Jäger, Y. Xu, *J. Mol. Spectrosc.* **2017**, *335*, 88–92.
[23] L. Evangelisti, P. Écija, E. J. Cocinero, F. Castaño, A. Lesarri, W. Caminati, R. Meyer, *J. Phys. Chem. Lett.* **2012**, *3*, 3770–3775.
[24] G. Feng, L. B. Favero, A. Maris, A. Vigorito, W. Caminati, R. Meyer, *J. Am. Chem. Soc.* **2012**, *134*, 19281–19286.
[25] Y. Litman, J. O. Richardson, T. Kumagai, M. Rossi, *J. Am. Chem. Soc.* **2019**, *141*, 2526–2534.
[26] D. S. Tikhonov, *Struct. Chem.* **2021**.
[27] F. Xie, N. A. Seifert, M. Heger, J. Thomas, W. Jäger, Y. Xu, *Phys. Chem. Chem. Phys.* **2019**, *21*, 15408–15416.
[28] N. A. Seifert, J. Thomas, W. Jäger, Y. Xu, *Phys. Chem. Chem. Phys.* **2018**, *20*, 27630–27637.
[29] C. Pérez, S. Lobsiger, N. A. Seifert, D. P. Zaleski, B. Temelso, G. C. Shields, Z. Kisiel, B. H. Pate, *Chem. Phys. Lett.* **2013**, *571*, 1–15.
[30] G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman, B. H. Pate, *Rev. Sci. Instrum.* **2008**, *79*, 053103-1-053103–13.
[31] Y. Xu, W. Jäger, *J. Chem. Phys.* **1997**, *106*, 7968–7980.
[32] S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

[33] S. Grimme, C. Bannwarth, P. Shushkov, *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
[34] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 16, Rev. C. 01* **2016**.
[35] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *123*, 0–9.
[36] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
[37] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571–2577.
[38] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, *J. Chem. Phys.* **2020**, *152*, 224108.
[39] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.
[40] J. K. G. Watson, A. Roytburg, W. Ulrich, **1999**, *119*, 102–119.
[41] J. Kraitchman, *Am. J. Phys.* **1953**, *21*, 17–24.
[42] Z. Kisiel, *J. Mol. Spectrosc.* **2003**, *218*, 58–67.
[43] P. Groner, R. D. Warren, *J. Mol. Struct.* **2001**, *599*, 323–335.
[44] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[45] D. Ferro-Costas, R. A. Mosquera, *Phys. Chem. Chem. Phys.* **2015**, *17*, 26946–26954.
[46] C. H. Görbitz, M. C. Etter, *J. Am. Chem. Soc.* **1992**, *114*, 627–631.
[47] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
[48] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899–926.
[49] J. Klein, H. Khartabil, J. C. Boisson, J. Contreras-Garciá, J. P. Piquemal, E. Hénon, *J. Phys. Chem. A* **2020**, *124*, 1850–1860.
[50] I. Kalkman, C. Vu, M. Schmitt, W. L. Meerts, *ChemPhysChem* **2008**, *9*, 1788–1797.
[51] W. Li, D. S. Tikhonov, M. Schnell, *Angew. Chemie - Int. Ed.* **2021**, *60*, 25674–25679.
[52] R. T. Saragi, M. Juanes, C. Pérez, P. Pinacho, D. S. Tikhonov, W. Caminati, M. Schnell, A. Lesarri, *J. Phys. Chem. Lett.* **2021**, *12*, 1367–1373.
[53] R. Medel, *Phys. Chem. Chem. Phys.* **2021**, *23*, 17591–17605.

Chapter VIII.

Other studies

In this chapter, we summarize other two ongoing studies experimentally carried out during the installation of the new bands in the CP-FTMW spectrometer. The new experimental arrangement has demonstrated extremely good sensitivity obtaining the spectra of penta-hydrated clusters of Methyl Benzoate (MeBnz). In the second section of this chapter, an interesting astrochemical molecule, 4-oxobutanonitrile, will be presented.

VIII.1 Micro-solvation of Methyl Benzoato

Understanding the interactions between water and organic molecules is a fundamental issue in chemistry and biochemistry.[1] The ideal case for a systematic study of solvation processes would be through the generation and detection of bare solutes, studying each solute-water aggregate with an increasing number of attached water molecules. One of the most sophisticated method to carry out these studies in order to show the effects caused by the introduction of water molecules in the solute, is by means of rotational spectroscopy.[2–5]

In this study, the rotational spectra of MeBnz–W_n (n=1–5) clusters are analyzed by CP-FTMW spectroscopy and quantum chemistry calculations. The MeBnz is a rigid molecule, so it is not affected by structural changes upon micro-solvation effects. However, this molecule is composed by different polar areas. On the one hand, the molecule presents a carbonyl group, whose affinity for the water through a $O_wH\cdots O{=}C$ H-bond has been already proved in several studies.[5–7] On the other hand, MeBnz presents an aromatic ring, and although in solution water does not present a special affinity for aromatic compounds, in gas phase water electron donor $H_2O_W\cdots\pi$ and water electron acceptor $O_WH\cdots\pi$ complexes have been identified.[8,9] Due to the possibility of multiple binding sites in between water and MeBnz, here we examine the interaction between the water and MeBnz organic molecule on the formation of its first solvation shell.

In the present work, we record the 2 - 6 GHz and 7 - 14 GHz spectra of MeBnz with water, using CP-FTMW spectrometer. Concretely, we used the 2 - 6 GHz and 7 – 14 GHz band built during this thesis (all the details are in the Chapter II). The sample has a very high vapor pressure, so we did not need to heat it to obtain enough concentration in gas phase to register a good spectrum. The water was added using a simple bubbler and it was dragged with the carrier gas (~6 psi of Ne).

The spectral assignment was supported by quantum chemistry calculations. The equilibrium structures of the clusters were calculated in two steps. First, an extensive conformational search was done using *Merk Molecular Force Field (MMFFs)* molecular mechanics algorithm.[10] Second, the obtained structures were fully optimized using DFT quantum mechanical methods. For this system, we used B3LYP and Pople's 6-311++G(d,p) functions. The quantum chemistry calculations were carried out using Gaussian 09 and Gaussian 16 packages (all details in the Chapter II of this Thesis).[11,12] The theoretical structures, spectroscopic parameters, and relative energies of the most stable geometries are reported in the appendix (Table VIII.A1-5 and Figure VIII.A1-5).

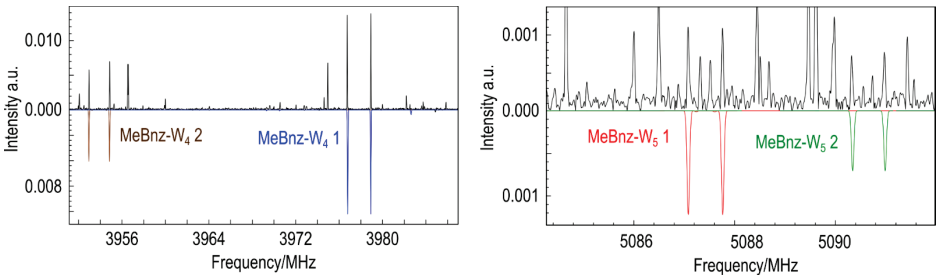


Figure VIII.1: The CP-FTMW pulsed-jet spectrum of MeBnz with water after a 1.2 million acquisitions in the time domain. In positive the experimental spectrum has been depicted and in negative the simulated spectrum of fitted conformers. On the left side, the 6_{16}-5_{15} and 6_{06}-5_{05} transitions of the two detected conformers of tetra-hydrated clusters (MeBnz-W_4 1 and MeBnz-W_4 2), and on the right the 4_{14}-3_{13} and 4_{04}-3_{03} doublets transitions of the two detected conformers penta-hydrated clusters (MeBnz-W_5 1 and MeBnz-W_5 2) are shown.

For each water added to the MeBnz, two set of transitions has been identified, which were fitted by Pickett's SPFIT program, with Watson's semirigid Hamiltonian (*S* reduction, I^r representation).[13,14] Although for some conformers the splitting for the methyl rotor was detected, in this chapter it will be omitted. The determined rotational constants of all ten complexes are summarized in Table VIII.1. The data of the final fittings with the first-order centrifugal distortion constants and the line list have been summarised in the appendix (Table VIII.A.6-15). Despite that the calculations treatment are not completed yet, the comparison of the rotational constants with the theoretically calculated rotational constants provide an identification of the clusters. Initially, we only suggest the possibility to see the thermodynamic conformers although in some cases the kinetically controlled conformers could also appears in the spectra. However in this case, the calculation are still in progress to try to search the kinetically controlled conformers in the recorded spectra.

Table VIII.1: Most relevant experimental rotational parameters of the observed conformers of the studied species. The full data can be found in the Appendix, Tables VIII.A.6-15.

	A[a]	*B*	*C*	N[b]	σ[c]
MeBnz W_1 1	2468.90074(59) [d]	530.56282(14)	438.55917(12)	143	8.0
MeBnz W_1 2	1246.73721(57)	806.50710(37)	491.79012(22)	102	9.5
MeBnz W_2 1	2196.82502(72)	346.90332(11)	300.56801(10)	192	5.7
MeBnz W_2 2	801.9746(11)	684.26762(50)	371.08993(17)	82	6.1
MeBnz W_3 1	789.3731(13)	571.7201(22)	475.80124(85)	16	4.7
MeBnz W_3 4	1482.533(95)	262.03998(61)	225.77101(66)	32	9.2
MeBnz W_4 1	620.2820(10)	522.21855(86)	379.18729(90)	25	9.0
MeBnz W_4 2	619.5287(15)	523.1454(24)	376.48325(55)	14	9.6
MeBnz W_5 3	545.25096(53)	397.38944(91)	308.91633(59)	21	7.4
MeBnz W_5 4	544.69171(51)	398.10871(86)	309.10556(65)	25	8.7

[a] Experimental rotational constants *A*, *B* and *C* in MHz.
[b] Number of fitted transitions.
[c] MW *root-mean-square* deviation of the fit in kHz.
[d] Standard error (SE) in round brackets in units of the last digit.

Similarly as was showed in previous studies, the most favorable water-solute interaction is an $O_wH\cdots O{=}C$ H-bond,[5,7,15] which appears in all detected clusters. In the monohydrated cluster, as computational results predicted, the water lies in the symmetry plane of MeBnz, interacting with the carboxylic oxygen as proton donor through a strong $O_wH\cdots O{=}C$ H-bond. Further, in the detected clusters, the water also have a second interaction site with low hidrogen donor capacity protons of MeBnz through C_{Me}-H$\cdots O_W$ and C_{arom}-H$\cdots O_W$ interaction in MeBnz W_1 1 and MeBnz-W_1 2, respectively.

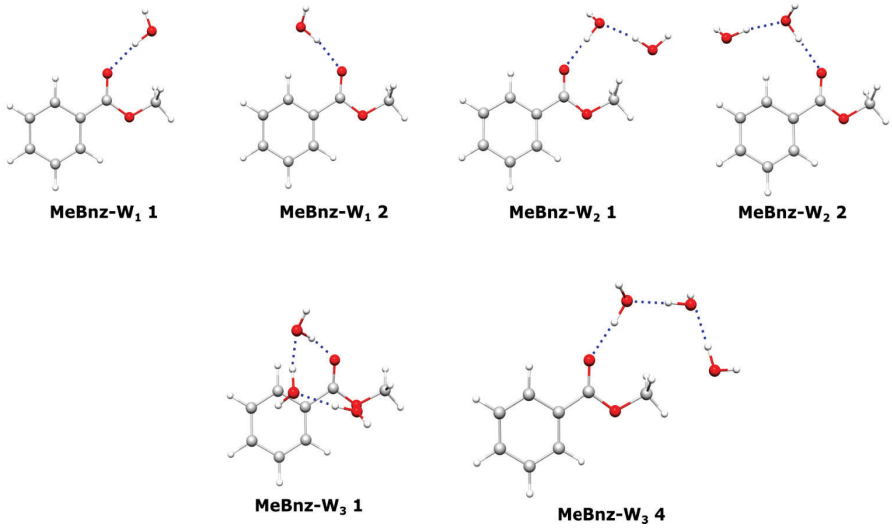


Figure VIII.2 Theoretical structures calculated at the B3LYP-GD3BJ/6-311++G(d,p) level of detected conformers.

MeBnz-W_2 displays clear similarities with the MeBnz-W_1 clusters and, as for monohydrated conformers, the computation results reproduce well the experimentally obtained data. Conformationally, the addition of the second water molecule does not modify the position of first water molecule and the O_w-H···O=C H-bond created by the water in the MeBnz-W_1 but due to the H-bond cooperativity, the calculation predicts the O_w-H···O=C H-bond distance shorter in the MeBnz-W_2 clusters (~ -0.05 Å). , In the detected two conformers of dehydrated MeBnz, the second water also present some C-H···O_w interactions with the MeBnz, but as for the one water cluster the interaction is very weak.

For MeBnz-W_3 cluster, the B3LYP-GD3BJ/6-311++G(d,p) calculations predicted three conformers with a very similar population (Table VIII.A.3 of Appendix VIII). However, although extensive searches of the three conformers were done, experimentally we only observed two of them, MeBnz-W_3 1 and MeBnz-W_3 4. In the case of MeBnz-W_3 4, the water molecule interacts in the similar fashion that in the case of MeBnz-W_1 1 and MeBnz-W_2 1, with all the oxygens of water in the symmetry plane of the MeBnz molecule and interaction slightly with methyl group. Curiously, this tendency changed completely in the MeBnz-W_3 1 where the water molecules leave the plane symmetry disposition presented in all the previously shown structures. The new water arrangement in the MeBnz-W_3 1 keep the O_wH···O=C interaction, but in this case, the new water arrangement at the top of the aromatic π cloud makes possible an O_wH···π interaction. This interaction was detected in other aromatic-water clusters, such as in the benzene-water complex.[8]

Now, in the tetra-hydrated MeBnz-W_4 and penta-hydrated MeBnz-W_5 clusters, the auto-aggregation of water gains predominance and in the most stable conformers water molecules, and the water molecules form 4 and 5-member ring structures respectively (See Tables VIII.A.4-5 and Figures VIII.A.4-5 of Appendix). Interestingly, the water molecules in the MeBnz-W_4 have the same structure that the pure water clusters, where the water presents an up-down-up-down orientation in the free hydrogen atoms, exactly the same as in the minimum energy structure of the water tetramer (See Figure VIII.2).[16] The two detected tetra-hydrated conformers differ in the cyclic water-water H-bond network direction. The structures are also stabilised by an $O_wH\cdots\pi$ interaction and, although the H-bond network direction changes in each of the detected conformer, the oxigens of the water molecules are positioned in the same position in the MeBnz-W_4 1 and MeBnz-W_4 2 clusters.

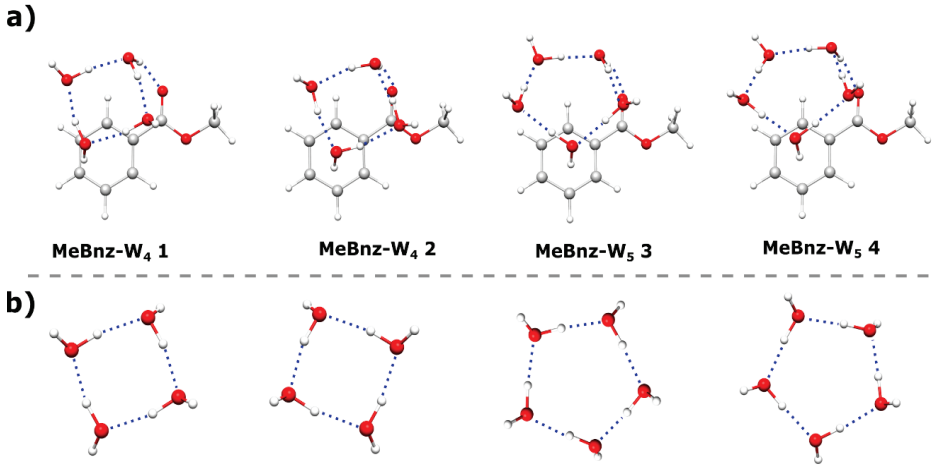


Figure VIII.3: a) Theoretical structures calculated at the B3LYP-GD3BJ/6-311++G(d,p) level of detected conformers. b) The most stable structures with clockwise and counter-clockwise H-bond network of water calculated at B3LYP-GD3BJ/6-311++G(d,p) level of theory.

Finally, on the MeBnz-W_5 cluster, the quantum chemistry predicted several similar stable structures (Figure VIII.A.5 and Table VIII.A.5 of Appendix VIII). The assignment of the observed pentahydrate conformers to each respective geometry identified as minimum was not trivial. The most stable conformers only differ in the position of the water with respect to MeBnz. water that acts as proton donor with π cloud electrons and in the directionality of the hydrogen bond network (Figure VIII.A.5 of Appendix). We needed to compare the rotational constants and the dipole moments of the detected transitions to assign the conformers to MeBnz-W_5 3 and MeBnz-W_5 4, in which water adopts a quasi-planar

geometry, similar to the most stable structure of the pure water pentamer (Figure VIII.3).[17] As for the MeBnz-W_4, the detected structures of the MeBnz-W_5 only differ in the H-bond network direction. Regarding the structures, the detected conformers are stabilised by two water-solute attractive interactions: the $O_wH\cdots O{=}C$ H bond and through additional interaction between the water proton and MeBnz π electron cloud, $HO_W\cdots\pi$.

In summary, we have assigned the rotational spectra of ten structures of MeBnz-W_n (n=1-5). The most important results obtained up to now are: (i) the combination of rotational spectroscopic and quantum chemistry methods gives a deeper insight in the structure of MeBnz-W_n clusters with up to five water molecules; (ii) the self-aggregation of water is of crucial importance in the solvation of MeBnz and prevails over the formation of other H-bonds with the solvent molecules; (iii) when the cluster size grows from n=1,2 to n=3, the bonding waters moves from the plane of symmetry of the phenyl group to the top of the aromatic cloud and this new waters' disposition is maintained for the bigger clusters (n = 4-5). The analysis of these systems is expected to shed light on the influence of water in the solvation of the MeBnz.

VIII.2 4-oxobutanonitrile

The analyzed molecule, 4-Oxobutanenitrile ($HCOCH_2CH_2CN$), has a high prebiotic relevance in the context of the RNA-world hypothesis for the origin of Life, since it has been invoked in prebiotic experiments as a precursor of the glutamic acid (Glu),[18] an amino acid that is used by almost all living beings in the biosynthesis of proteins, through the route:

$$C_2H_3CHO + HCN \rightarrow HCOCH_2CH_2CN \rightarrow Glu$$

The first two species, propenal (C_2H_3CHO) and hydrogen cyanide (HCN) are detected in the quiescent giant molecular cloud G + 0.693-0.027,[19,20] so 4-Oxobutanenitrile is a good candidate to be detected. The improved sensitivity of current single dish telescopes, such as Yebes 40m, IRAM 30m and GBT 100m, and interferometers such as ALMA, is allowing us to detect a plethora of molecules with similar weight of 4-oxobutanonitrile (MW: 83.05 Da), such as, ethanolamine ($NH_2CH_2CH_2OH$; MW: 61.06 Da)[21] ethyl isocyanate (C_2H_5NCHO, MW: 72.05 Da)[22] hydrocarbon cycles aromatic rings such as o-benzyne, (o–C_6H_4 MW: 76.03 Da)[23], cyclopentadiene (c–C_5H_6, MW: 66.05 Da)[24] and cyano derivatives of hydrocarbon cycles like benzonitrile (c–C_6H_5CN, MW:103.05)[25], cyanocyclopentadiene (c- C_5H_5CN, MW: 91.04)[26], and also double aromatic rings such as indene (c–C_9H_8, MW:116.06)[24,27], and 1− and 2−cyanonaphthalene (c-$C_{10}H_7CN$, MW: 153.06 Da)[28], molecules that are related with the chemical feedstock RNA world.

The radio-telescopes are not looking for organisms living in ISM today. However, the astronomical search collects data of the molecules that could be used to identify biosignatures related with the life and with the origin of life. Detection of molecules in the ISM is not trivial since their rotational spectra are extremely congested or contaminated with “astrophysical weeds” associated with a multiplicity of compounds. For this reason, the detection of the molecules in the laboratory is in many times essential to detect the molecules in the ISM. In this context, we carried out the rotational characterization of 4-oxobutanonitrile in the laboratory for future searches in the ISM. We combined rotational spectroscopy with quantum chemical calculations to obtain the necessary experimental rotational constants for the comparison of the experimental results with the radio-telescopes data.

Because there was no previous information about 4-oxobutanenitrile, it was necessary to carry out computational calculations to support the analysis of the spectrum. The conformational space of 4-oxobutanenitrile was explored using the molecular mechanics method, obtaining five possible stable conformations. These structures were then optimized by DFT method, concretely B3LYP density functional,[29] including the Grimme D3 dispersion interactions with Becke–Johnson damping,[30,31] and Pople’s 6-311++G(d,p) basis set, available in

Gaussian 16 software (all details are in the Chapter II of this Thesis).[12] The most stable structures are displayed in Figure VIII.4 and their spectroscopic parameters and energy values are in Table VIII.2. The information of all the most stable conformers are in the appendix of this chapter.

The nomenclature employed to name them is based on two dihedral angles, C1-C2-C3-C4 and C2-C3-C4=O, defined by two letters respectively, according to IUPAC recommendations. The first letter defines the carbon backbone disposition, *trans* (T) for approximately 180° and *gauche* dispositions, G+ and G- for +60° or -60°, respectively. The second letter defines the C2-C3-C4=O dihedral angle, where besides T, G+ and G-, it was necessary define the *anticlinal* positions A^+ or A^- when the dihedral angle has +120° or -120° values, and *synperiplanar* or *cis* (C) when this dihedral angle assumes values around 0°.

The rotational spectrum of the molecule in the 2-18 GHz frequency range was obtained using the CP-FTMW spectrometer.[32] The commercial sample was heated at 45° C, and the vaporized products were seeded in Neon at backing pressures of 2-4 bar, to then be supersonically expanded through a nozzle into the vacuum chamber (10^{-6} mbar) of the spectrometer. In total, three independent experiments were carried out in the 2-6 GHz, 7-14 GHz and 12-18 GHz frequency ranges to cover the complete 2-18 GHz range, each one with up to 100k individual free induction decays, which were averaged in the time domain and Fourier transformed to obtain the broadband frequency domain spectrum shown in Figure VIII.5.

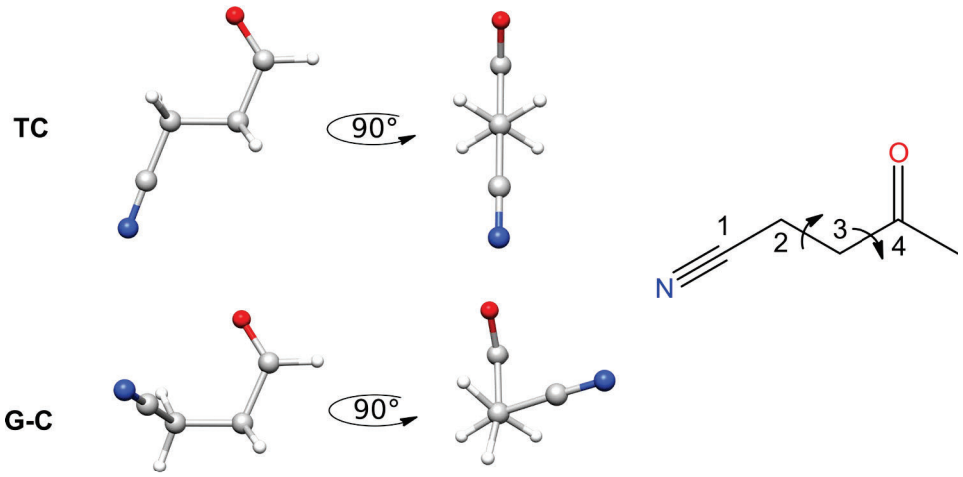


Figure VIII.4. The most stable conformers of of 4-oxobutanenitrile calculated at B3LYP-GD3BJ/6-311++G(d,p) level.

Table VIII.2. Theoretically (B3LYP-GD3BJ/6-311++G(d,p)) and experimentally obtained parameters of 4-oxobutanonitrile..

	TC		G-C
	Theo	**Exp**	**Theo**
A[a]	16182	16000.7597(21)[f]	6637
B	1611	1622.40667(57)	2447
C	1492	1500.80611(64)	2051
D_J[b]		0.216(16)	
X_{aa}[c]	-4.18	-3.8989(49)	-0.99
X_{bb}	1.98	1.8536(57)	-0.43
X_{cc}	2.20	2.0453(57)	1.42
ΔE^{ZPE}	0.0		370
ΔG^{298K}	0.0		381
N[d]		91	
σ[e]		9.9	

[a] Experimental rotational constants *A*, *B* and *C* in MHz.
[b] The distortion constant has been fixed from the parent specie (kHz).
[c] X_{aa}, X_{bb} and X_{cc} correspond to the quadrupole coupling constants derived for the presence of nitrogen in the molecule (MHz).
[d] Number of fitted transitions.
[e] MW *root-mean-square* deviation of the fit in kHz.
[f] Standard error (SE) in round brackets in units of the last digit.

In the recorded CP-FTMW spectra, one rotamer was easily identified due to a very intense pattern of $1_{0,1} \leftarrow 0_{0,0}$, $2_{1,2} \leftarrow 1_{1,1}$, $2_{0,2} \leftarrow 1_{0,1}$ and $2_{1,1} \leftarrow 1_{1,0}$ μ_a-type *R*-branch transitions. We measured μ_a-type R-branch transitions up to $J' = 5$ and $K'_a = 2$. Also, μ_b-type transitions of the *R*, *Q* and *P* branches were identified. The observed rotational transitions of 4-oxobutanenitrile are split into a complicated hyperfine structure. These hyperfine patterns were accounted for in terms of nuclear quadrupole coupling interaction effects due to the presence of one ^{14}N nuclei with nonzero electric quadrupole moment and spin (I = 1), interacting with the electric field gradient created by the rest of the molecule. The ^{14}N nuclear quadrupole coupling splits the rotational energy levels, which leads to a decrease of the overall intensity of each rotational transition and produce a complex hyperfine structure. In our case, the lines were split enough to be resolved by our CP-FTMW spectrometer (see Figure VIIII.5(b) and VIIII.5 (c)). A first semi-rigid rotor analysis was done using Watson's *A* type reduced Hamiltonian[13,14] and a first set of experimental rotational constants was obtained, which enables by comparing them with the predicted spectroscopic constants, the unambiguous identification of the observed rotamer as conformer **TC** (Table VIII.2). Although the conformer **G-C** could be a good candidate to be detected in the spectrum, no signals of this conformer were detected in the recorded spectra, maybe for the low energy interconversion barrier (≈ 3.5 kJ/mol using B3LYP-GD3BJ/6-311++G(d,p) method) between the two conformers.

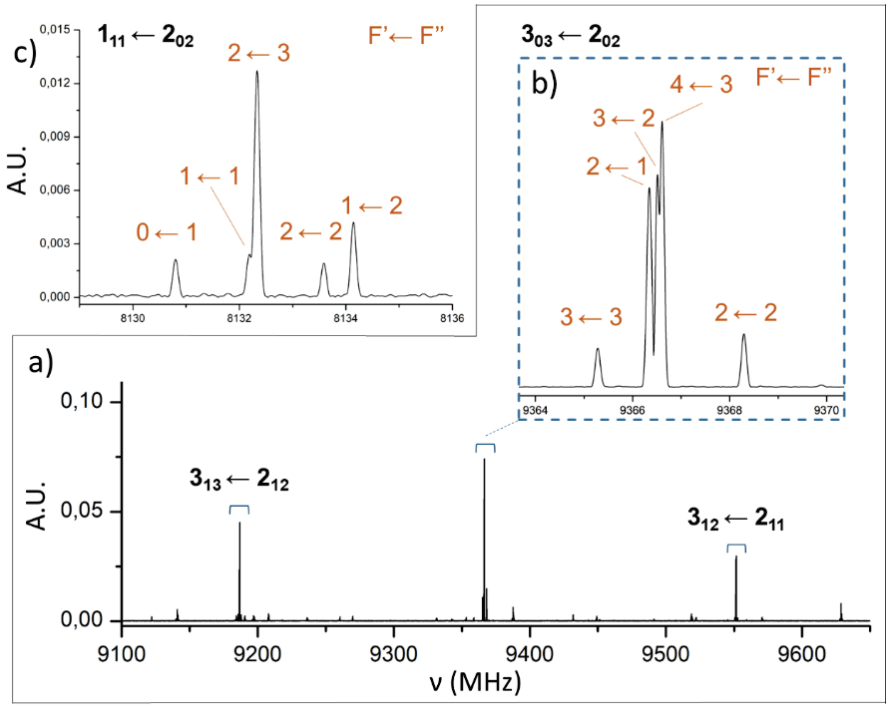


Figure VIII.5 (a) Section of the broadband CP-FTMW spectrum of 4-oxobutanenitrile from 9100 to 9600 MHz, with the typical pattern of μ_a-type *R*-branch transitions *J*''=2. (b) Section of the spectrum showing the completely resolved nuclear hyperfine structure of the $3_{03} \leftarrow 2_{02}$ transition. Each hyperfine component is labelled with the corresponding values of F', F'' quantum numbers. (c) Section of the spectrum showing the completely resolved nuclear hyperfine structure of the μ_b-type *P*-branch $1_{11} \leftarrow 2_{02}$ transition.

The extremely high sensitivity of the chirp spectrometer allows us to detect the rotational spectra of several monosustituted isotopologues in natural abundance (1.1% ^{13}C, 0.4% ^{15}N and 0.2% ^{18}O). The transitions were fitted as described for the parent species. Measured transitions and rotational constants of all isotopic species obtained from a semi-rigid rotor analysis are given in Table VIII.3 and in the appendix of this chapter.

Table VIII.3. Experimentally obtained parameter of 4-oxobutanonitrile isotopologues in natural abundance.

	^{13}C1	^{13}C2	^{13}C3	^{13}C4	^{15}N	^{18}O
A[a]	15981.6594(24)	15877.7461(29)	15699.5659(32)	15882.5401(28)	15998.1167(47)	15636.5335(45)
B	1605.1063(10)	1621.5958(13)	1621.5308(12)	1603.9020(12)	1578.4909(18)	1566.5687(19)
C	1485.8252(11)	1499.0409(10)	1497.3477(12)	1483.9303(11)	1463.1476(20)	1450.7792(14)
X_{aa}[b]	-3.8853(41)	-3.8933(45)	-3.9093(62)	-3.9093(59)	-	-3.9427(29)
X_{bb}	1.8373(70)	1.842(12)	1.820(11)	1.820(10)	-	[1.8536][f]
X_{cc}	2.0481(70)	2.051(45)	2.089(11)	2.089(10)	-	[2.0453]
D_J[c]	[0.216]	[0.216]	[0.216]	[0.216]	[0.216]	[0.216]
N[d]	24	25	28	30	8	8
σ[e]	6.1	7.0	7.8	7.4	7.4	3.9

[a] *A*, *B* and *C* are the rotational constants fitted using *A*-type Watson reduction in MHz.
[b] X_{aa}, X_{bb} and X_{cc} correspond to the quadrupole coupling constants derived for the presence of nitrogen in the molecule (MHz).
[c] The distortion constant has been fixed from the parent specie (kHz).
[d] *N* are the number of lines fitted.
[e] σ is the room-mean-square error of the fit in kHz.
[f] For ^{18}O, the X_{aa}, X_{bb} and X_{cc} could not be determined, so the values of the parent specie were used.

The availability of the rotational constants of the parent species and of all the singly-substituted isotopologues allowed for the determination of the exact position of the heavy atoms in the molecule. In this work, we have determined the substitution structure (r_s) using Kraitchman's equations.[33] A comparison of the results obtained with theoretical method and with the Kraitchman's equation's in this work is provided below in the Figure VIII.6.

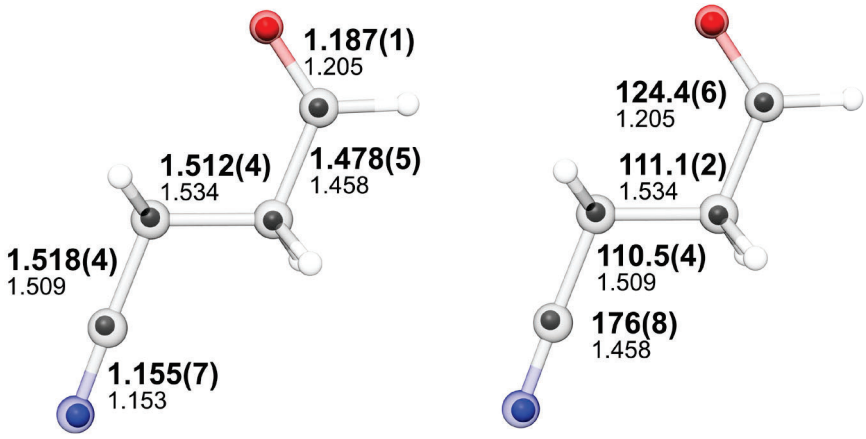


Figure VIII.6 Comparison between the theoretical structure (r_e) and the substitution structure (r_s) of the observed conformer TC of 4-oxobutanonitrile. The full body of the molecule represents the DFT structure calculated at the B3LYP-GD3BJ/6-311++G(d,p) level of theory. The solid spheres represent the position of the N, C and O atoms calculated using Kraitchman's equations. The errors of r_s structure are reported in round brackets in terms of the last digit.

We only perform a laboratory characterization of the molecule in the 2-18 GHz range, so the next step of this work will be the extension of the recorded spectrum to the upper region (30-60 GHz) using the Free-Jet spectrometer of the University of Bologna. The principal goal of this extension is to obtain experimental signals in the region in which the radio-telescopes are more sensitive, increasing the probability to detect the target molecule in the ISM. When the data will be collected and treated, the obtained lines will be used as fingerprints to try to detect this molecule in several region of the Inter Stellar Medium.

VIII.1 Bibliography

[1] P. Ball, *Chem. Rev.* **2008**, *108*, 74–108.
[2] C. Calabrese, W. Li, G. Prampolini, L. Evangelisti, I. Uriarte, I. Cacelli, S. Melandri, E. J. Cocinero, *Angew. Chemie* **2019**, 8525–8530.
[3] J. L. Alonso, E. J. Cocinero, A. Lesarri, M. E. Sanz, J. C. López, *Angew. Chemie Int. Ed.* **2006**, *45*, 3471–3474.
[4] C. Pérez, A. L. Steber, B. Temelso, Z. Kisiel, M. Schnell, *Angew. Chemie - Int. Ed.* **2020**, *59*, 8401–8405.
[5] C. Pérez, J. L. Neil, M. T. Muckle, D. P. Zaleski, I. Peña, J. C. Lopez, J. L. Alonso, B. H. Pate, *Angew. Chemie - Int. Ed.* **2015**, *54*, 979–982.
[6] W. Li, M. M. Quesada-Moreno, P. Pinacho, M. Schnell, *Angew. Chemie* **2021**, *133*, 5383–5390.
[7] J. Thomas, O. Sukhorukov, W. Jäger, Y. Xu, *Angew. Chemie - Int. Ed.* **2014**, *53*, 1156–1159.
[8] S. Suzuki, P. G. Green, R. E. Bumgarner, **1992**, *257*, 942–944.
[9] L. Evangelisti, K. Brendel, H. Mäder, W. Caminati, S. Melandri, *Angew. Chemie - Int. Ed.* **2017**, *56*, 13699–13703.
[10] T. a Halgren, *J. Comput. Chem.* **1996**, *17*, 490–519.
[11] M. J. et al Frisch, *Gaussian 09, Revis. A.02* **2009**.
[12] G. A. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, *Gaussian, Inc.* **2016**.
[13] R. K. Thomas, *Phys. Bull.* **1975**, *26*, 501–501.
[14] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.
[15] C. Pérez, A. Krin, A. L. Steber, J. C. López, Z. Kisiel, M. Schnell, *J. Phys. Chem. Lett.* **2016**, *7*, 154–160.
[16] J. D. Cruzan, L. B. Braly, K. Liu, M. G. Brown, J. G. Loeser, R. J. Saykally, *Science (80-.).* **1996**, *271*, 59–62.
[17] K. Liu, M. G. Brown, J. D. Cruzan, R. J. Saykally, *Science (80-.).* **1996**, *271*, 62–64.
[18] N. Kitadai, S. Maruyama, *Geosci. Front.* **2018**, *9*, 1117–1153.
[19] V. M. Rivilla, J. Martín-Pintado, I. Jiménez-Serra, S. Martín, L. F. Rodríguez-Almeida, M. A. Requena-Torres, F. Rico-Villas, S. Zeng, C. Briones, *Astrophys. J.* **2020**, *899*, L28.
[20] I. Jiménez-Serra, J. Martín-Pintado, V. M. Rivilla, L. Rodríguez-Almeida, E. R. Alonso Alonso, S. Zeng, E. J. Cocinero, S. Martín, M. Requena-Torres, R. Martín-Domenech, L. Testi, *Astrobiology* **2020**, *20*, 1048–1066.
[21] V. M. Rivilla, I. Jiménez-Serra, J. Martín-Pintado, C. Briones, L. F. Rodríguez-Almeida, F. Rico-Villas, B. Tercero, S. Zeng, L. Colzi, P. De Vicente, S. Martín, M. A. Requena-Torres, *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, 9pp.
[22] L. F. Rodríguez-Almeida, I. Jiménez-Serra, V. M. Rivilla, J. Martín-Pintado, S. Zeng, B. Tercero, P. de Vicente, L. Colzi, F. Rico-Villas, S. Martín, M. A. Requena-Torres, *Astrophys. J. Lett.* **2021**, *912*, L11.
[23] J. Cernicharo, M. Agúndez, R. I. Kaiser, C. Cabezas, B. Tercero, N. Marcelino, J. R. Pardo, P. De Vicente, *Astron. Astrophys.* **2021**, *652*, 1–7.
[24] J. Cernicharo, M. Agúndez, C. Cabezas, B. Tercero, N. Marcelino, J. R. Pardo, P. De Vicente, *Astron. Astrophys.* **2021**, *649*, 1–12.
[25] B. A. Mcguire, *Astrophys. J. Suppl. Ser.* **2018**, *239*, 17.
[26] M. C. McCarthy, K. L. K. Lee, R. A. Loomis, A. M. Burkhardt, C. N. Shingledecker, S. B. Charnley, M. A. Cordiner, E. Herbst, S. Kalenskii, E. R. Willis, C. Xue, A. J. Remijan, B. A. McGuire, *Nat. Astron.* **2021**, *5*, 176–180.
[27] A. M. Burkhardt, K. Long Kelvin Lee, P. Bryan Changala, C. N. Shingledecker, I. R. Cooke, R. A. Loomis, H. Wei, S. B. Charnley, E. Herbst, M. C. McCarthy, B. A. McGuire, *Astrophys. J. Lett.* **2021**, *913*, L18.

[28] B. A. McGuire, R. A. Loomis, A. M. Burkhardt, K. L. K. Lee, C. N. Shingledecker, S. B. Charnley, I. R. Cooke, M. A. Cordiner, E. Herbst, S. Kalenskii, M. A. Siebert, E. R. Willis, C. Xue, A. J. Remijan, M. C. McCarthy, *Science (80-.).* **2021**, *371*, 1265–1269.
[29] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[30] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[31] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
[32] I. Uriarte, C. Pérez, E. Caballero-Mancebo, F. J. Basterretxea, A. Lesarri, J. A. Fernández, E. J. Cocinero, *Chem. - A Eur. J.* **2017**, *23*, 7238–7244.
[33] J. Kraitchman, *Am. J. Phys.* **1953**, *21*, 17–24.

Chapter IX.

Conclusions

IX.1 Conclusions

In this dissertation some of the works done during my PhD period have been presented. The research study was focus on the conformational characterization of several molecules of biological interest, starting from the monosaccharides and continuing with other molecules that are involved in several biological functions. Concretely, the structural determination of several molecules and some molecular aggregates were done using the combination of rotational spectroscopy and quantum chemistry calculations. My work has been accomplished by using, assembling, maintaining and developing the *chirped* pulse Fourier transform microwave spectrometer and the Fabry-Perot cavity based Fourier transform spectrometer. As aforementioned, the thesis is divided into two connected sections. The first part is focus on the study of several monosaccharides and due to the thermolability of the sugars, all the experiments of this part were carried out using the Fabry-Pérot cavity (cav-FTMW) with laser ablation vaporization system. Although the monosaccharides are very similar molecules, the transformation of their OH groups change completely their function.

In **chapter III**, the study or erythrulose molecule has been presented. Erythrulose is a four carbon ketone sugar, which presents in gas, solid and liquid phase a linear structure. Following the detection of several sugar related molecules in Murchison meteorites and in several regions of the ISM, we developed a new sample preparation method to obtain the rotational spectrum of erythrulose in our rotational spectrometer. With the obtained experimental results, erythrulose has been searched in several regions of the ISM. Although the detection was unsuccessful, the obtained data will be essential for its future searches in the ISM.

From **chapters IV** to **chapter VI**, several glucose, galactose and mannose derivatives studies have been presented. The study of glucose, galactose and mannose using the combination of the rotational spectroscopy and quantum chemistry calculation were carried out some years ago. However, the conformational behaviour of the analogues presented in this dissertation have not been studied before.

In **chapter IV** the conformational effect due to the bio-isosteric OH to F substitutions and dihydroxylation of glucose, galactose and mannose have been presented. In this work, we studied eight molecules using the combination of the theoretical results and several spectroscopic methods, although the chapter is specially focalised in the rotational spectroscopy analysis. In general, we showed that bio-isosteric OH to F substitutions do not have special effects in the hydroxymethyl and H-bond arrangements of glucose and galactose. However, the replacement of the second hydroxyl group for fluorine atom in the mannose, changes the direction of the H-bond network between this Mannose and fluorinated mannose derivative. On the other hand, the dihydroxylation in the

second carbon of the glucose and galactose, stabilise the *clockwise* H-bond network reversed with respect to the hydroxylated and fluorinated analogues. The results show the different conformational arrangements of the no natural sugars, and these results are very useful to understand how the small changes of the natural sugars could induce extremely high conformational changes in the mutant sugars.

Chapter V reports the spectroscopic analysis of glucose derivative sugar lactones. We observed that the gluconorolactone and the glucose have a completely different 3D atomic shape. In the case of the glucoronolactone, the molecule presents in gas phase two furanose rings and the hidroxyl group is not connected through an intramolecular H-bond network, characteristic of many sugars. On the other hand, the gluconolatone is quite similar to glucose, with the characteristic cooperative H-bond network and three stable hydroxymethyl conformations. However, the ring puckering and the population of the conformers are different, due to the different hybridization ($sp^3 \rightarrow sp^2$) of the anomeric carbons.

Other current glucose and galactose derivatives comes from the oxidation of the hydroxymethyl group to the carboxylic acid. The oxidation of the hydroxymethyl group of the glucose and galactose forms glucuronic and galacturonic acids. The study of these molecules are reported in **chapter VI**. In this study, we combine the recrystallization and solid state nuclear magnetic resonance to obtain and verify the presence of both anomers of glucuronic and galacturonic acids. In the case of glucuronic acid, only one conformer of each anomer was detected in gas phase with the same H-bond arrangenment. Curiously, in the galacturonic acid the H-bond network is the opposite for each anomer. In this chapter, we demonstrate that the transformation of the hydroxymethyl group to a carboxylic acid changes completely the structure of the sugars comparing with their analogues.

The second part of this thesis, **chapters VII** and **VIII**, reports several experimental studies measured by the *chirped* pulse Fourier transform spectrometer.

The first chapter of this section, **chapter VII**, presents experiments done at the Xu's laboratory at the University of Alberta. In this chapter, we have described the study of the monomer and dimer of furoic acid. We detected in gas phase three monomers and one dimer, where the degenerate proton transfer was experimentally detected. We used computational chemistry calculations to estimate the proton transfer barrier and we compared the result with the values reported for other dimers. Curiously, we observed a very high energy for the double proton transfer, and we suggested that probably the tunneling energy is bigger when the carbonyl groups are connected with the aromatic ring.

Finally, in **chapter VIII** we summarized two studies that are currently in progress. The experiments were carried out in the *chirped* pulse Fourier transform

spectrometer at the University of the Basque Country. Concretely, we focused on the micro-solvation of the methyl benzoate where we observed the importance of the auto-aggregation of water in the clusters formed by four or more water molecules. On the other hand, we also showed the first steps of the study of 4-oxobutanonitrile, molecule related with the presences of amino-acids in the interstellar medium.

As summary, we have demonstrated that the rotational spectroscopy is a versatile technique to study different properties of monosaccharides in gas phase. Although the first monosaccharides were measured using this technique some years ago, the results in gas phase results still provide essential information to understand the rules that govern their conformational distribution. On the other hand, the successful construction of the new arrangement of *chirped* pulse rotational spectrometer has been demonstrated with the detection of the methyl benzoate five water molecule cluster, one of the bigger water cluster detected using the rotational spectroscopy. Finally, fitting the proton transfer large amplitude motion of the furoic acid dimers, I demonstrate the ability to work with complex spectra.

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
FACULTY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF THE BASQUE COUNTRY

Espektroskopia errotazionala tresna gisa azukreak eta interakzio molekularrak aztertzeko

Doktorego Tesia

Aran Insausti Beiro

Zuzendariak:

Dr. Emilio J. Cocinero Pérez
Dra. Camilla Calabrese

Kimika fisiko departamendua
Leioa, Ekaina 2022

Aurkibidea

Tesi honen eranskinakk format elektronikoan aurki daitezke. Irakurleak hurrengo web gunean aurki ditzazke pdf formatuan:

https://drive.google.com/drive/folders/1eCMEAxz0oucAeHFGeUdkCoJbOy95utdT?usp=sharing

Akronimoak eta simboloak

Tesi honertan erabilitako zenbait akronimo eta sinbolo bildu dira hurrengo taulan.

Akronimoa	Esanahia
cav-FTMW	Barrumbeko Fourier Transformatuko Mikrouhina (Flygare-Balle-ren espektrometroarentzat erabilia)
CP-FTMW	Chirped Pulse Fourier Transform Microwave
FID	Molekulen indukzio askeko emisioa
DFT	Dentsitate Funtsional Teoria
PES	Energia potentzial gainazala
QC	Kimika kuantikoa
MM	Mekanika molekularra
H-lotura	Hidrogeno lotura
NCI	Interakzio ez kobalenteak
IEG	Izararteko eremu gunea

I. kapitulua

Sarrera

I.1 Sarrera

Konposatu organiko eta ez-organikoen egituren determinazioa, beste diziplina batzuetaz gain, kimika, fisika, biokimika, ingeniaritza eta farmazia ezagutza zientifikoaren zati garrantzitsua da. Laburbilduz, egitura determinazioa *molekula edo biomolekula baten hiru dimentsioko koordenatu atomikoak ebazteko prozedura* da.

Sistema edo molekula baten atomo eta elektroien kokapen eta banaketaren arabera, konektatuta dauden molekula edo partikula errealek, energia maila balio diskretu jakin batzuk bakarrik har ditzakete. Energia maila horien informazioa, espektroskopikoki deskodifikatu daiteke sistema kimikoen eta biologikoen egitura eta ezaugarriei buruzko informazioa lortzeko. Molekulen edo konektatutako molekula sistemen egituraren ebazpenak hainbat helburu nagusi ditu. Lehenik eta behin, molekulen egitura bere propietate fisiko,[1–3] kimiko[4,5] eta biologikoekin erlazionatuta dago,[6,7] eta beraz, euren ezaugarri konformaziomalak ezagutzea funtsezkoa da sistema molekularren propietateak ulertzeko. Bigarrenik, erresoluzio espazial oso altuko egitura molekularren zehazpenak kimika teorikoaren erreferentzia moduan erabili daitezke.[8–10] Azken urteetan, kimika teorikoan aurrerapen ikaragarriak egin diren arren, konputazionalki lortzen diren emaitzek ez dute esperimentalki lortu daitezkeen datu guztiak erreproduzitzeko adina zehaztasunik ematen. Hori dela eta, emaitza teorikoek esperimentalki berretsi behar dira beti.[11,12] Hirugarrena, molekulen atomoen antolaketa espaziala eta aurretik aipatutako energia mailak, molekula bakoitzarentzat intrintsekoak dira eta beraz, beren zalantza gabeko hatz marka eratzen dituzte. Ez hori bakarrik, balio espektroskopikoak funtsezkoak dira lagin edo material ezezagun, eta baita zeruko espazio gunean dauden, molekulak zalantza gabe identifikatzeko (erradio teleskopioen datuak erabiliz adibidez).

Puntu honetan, teknika esperimental ezberdin askok erresoluzio altuko egitura molekularrak ezagutzeko aukera ematen dute. Egituren determinaziorako klasikoak diren prozedura esperimentalak alde batera utzita (erresonantzia magnetiko nuklearra, X-izpiko espektroskopia edo espektroskopia infragorria, adibidez), 3 dimentsioko egitura atomikoa determinatzeko erresoluzio espazial altua duten teknika berri ugari garatu dira azken urteetan. Agian, egitura eta konformazio molekularra zehazteko garatu diren teknikarik egungoenak, indar atomikoko mikroskopia eta krio-elektroi mikroskopia dira.[13–15] Teknika horiek oso hedatuak daude eta emaitza onak ematen dituzte molekula mota askoren kasuan. Hala ere, esperimentalki lortzen diren datuen ziurgabetasun espazialak (> 1.5 Å) ez du ahalbidetzen molekulen xehetasun txikienak detektatzea.[13,16] Gainera, esperimentu hauek ingurune kondentsatuan burutzen dira, non molekulen konformazio intrintsekoak, interakzio intermolekularren bidez maskaratuak izan daitezkeen. Halaber, konformazio molekularra determinatzea ahalbidetzen duten beste teknika ez espektroskopiko batzuk ere existitzen dira, adibidez, badaude

masa espektrometrian oinarrituak diren teknika batzuk, eta baita beste batzuk Coulomb-en eztandan sortzen diren pusketen analisian oinarritutakoak ere.[17–19] Hala ere, metodo hauek ez dute atomoen kokapen zehatza finkatzeko erresoluzio nahikorik.

Aurretik aipatutako moduan, teknika espektroskopiko desberdin ugari dauden arren, Tesi hau interakziorik gabeko ingurunean egindako errotazio edo mikrouhin espektroskopioan zentratuta dago. Errotazio espektroskopia, molekula baten (edo molekula agregatu txiki baten) energia errotazionalaren maila energetikoetatik, gas fasean dagoen molekularen atomoen antolamenduak zehaztea ahalbidetzen duten konstante fisikoak lortzeko erabiltzen den teknika da. Datu errotazionalen tratamendu egokiak, gas fasean dagoen egitura molekularraren paregabeko erresoluzio espazialeko egitura zehazteko erabil daitezke. Are nabarmenagoa da espektroskopia-teknika honek, konformero,[20] tautomero[21,22] eta isotopomeroak[23,24] zalantza barik desberdindu ditzakeen teknika bakarra dela. Horretaz gain, errotazio espektroskopia esperimentu iraultzaile berrien bidez molekula kiralak ere desberdintzeko gaitasuna frogatu da.[25,26] Konformero desberdinetatik lortutako errotazio espektroak erabilita, energian ebatzitako isomerizazio unimolekularren energia ebazteko erabili daiteke.[27] Mikrouhinen espektroskopiako espektroek erresoluzio altuenetako emaitzak eskaintzen dituen teknika da. Ondorioz, teknikak ez du arazorik lagin batean dauden molekulen edo molekula agregatu txikien geometriak desberdindu eta ebazteko.

Lehenengo errotazio espektroa 1934an neurtu zen eta amoniakoaren espektroa izan zen hain zuzen ere.[28] Hala ere, teknika hau molekulak ikertzeko bideragarri bihurtzen hasi zen Bigarren Mundu Gerraren ostean. Lehen espektrometroak, molekulen erradiazio xurgaketan zeuden oinarrituta, mikrouhin iturri jarraitua erabiltzen zuten eta espektroaren modulazioa *Stark*-modulazioa edo *Zeeman*-modulazioaren bidez egiten zen. Ikerketa horietan, gas laginak aztertu zituzten, gelaxka isolatuetan eta normalean ingurugiro tenperaturan. Baldintza horien ondorioz, molekulen populazioa oinarrizko egoeran egoteaz gain, egoera bibrazional eta errotazional kitzikatu askotan banatuta dago, eta honek espektroaren seinaleen intentsitatea jaitsi eta espektroaren konplexutasuna asko areagotzen du. Arazo hori, jet supersoniko espektrometroak garatu zirenean gainditu zen, non molekulak Kelvin bakar batzuetako tenperaturatara hozten diren.[29] Ingurune berri horretan, molekulen Boltzmann-en energia mailen populazioa aldatzen da eta molekularen energia mailen banaketa espektroaren uhin zentrimetrikoetan kontzentratzen da.

Mikrouhinen espektroskopiaren hurrengo aurrerapen handia, Fourier transformatuan oinarritutako lehen mikrouhin (FTMW) espektrometroa garatu zenean eman zen. Fourierren transfomatuan oinarritutako mikrouhin espektrometroak, erradiazioa ireki eta ixteko etengailuen abiaduraren

areagotzeari esker garatu ziren, izan ere, horrek molekulak mikrouhin erradiazio labur, azkar eta pultsatu batekin populatzeko aukera eman baitzuen. Flygare-ren taldeak, Fabry-Pérot barrunbean jet supersonikoaren hedapena konbinatzen zuen espektrometroa diseinatu zuen.[30,31] Ondoren, hainbat espektrometro desberdin garatu dira.[32–34] Baina hurrengo aurrerapen instrumental iraultzailea B. H. Pate-en taldeak egin zuen 2008an pultsu multi-frekuentzian (*chirped pulse*) oinarritutako espektrometro errotazionala argitaratu zutenean.[35] Aurkeztutako aparatu berriak, molekularen espektroaren banda zabal bat (8 GHz edo gehiago) esperimentu bakoitzeko kitzika dezake. Esperimentu honek gertaera bakoitzean espektroko hainbat GHz lortzeko aukera ematen du, eta seinale-zarataren ratio ona lortzeko milaka edo milioika espektro denboraren domeinuan metatuz eta batezbestekoa eginez lortzen da.

Tesi honen ezaugarri bereizgarrienetako bat da mikrouhin-espektrometro hauek ez direla komertzialak; horren ondorioz, denbora luzea eman dut espektrometroak mantentzen, eraikitzen, optimizatzen eta garatzen. Laburbilduz, gaur egun jet-supersonikoaren hedapen ingurunean lan egiteko espektrometrorik erabilienaz baliatu naiz tesia egiteko: Fabry-Pérot barrunbe espektrometroa eta pultsu multifrekuentzian oinarritutako errotazio espektrometroa, biak Euskal Herriko Unibertsitatean eskuragarri daudenak (xehetasun guztiak II. kapituluan).[36–38] Tesi hau erlazionatuta dauden bi zatitan bereiztuta dago. Lehenengo zatia zenbait monosakaridoren karakterizazio espektroskopikoan oinarrituta dago. Zehazki, zati honetan Fabry-Pérot barrunbean oinarritutako Fourier espektrometro errotazionala erabili nuen eta azukreen termoegonkortasun falta dela eta, molekula hauek guztiak pultsu ultraazkarreko laser izpien bidez lurrundu ziren.

Tesi honen lehen zatiarekin erlazionatuta (III-VI. kapituluak), monosakaridoak osagai glikobiologiko txikienak dira. Zehazki, glikobiologia azukre eta deribatuen funtzioa eta egituraren identifikatzea helburu duen zientzia da.[39] Azukreen aniztasunak, prozesu fisiologiko askoren informazio estrukturalak kodifikatzen du, hala nola energiaren ekoizpena, erantzun immunologikoa eta zelulen errekonozimendua.[39] Gainera, karbonatoen egitura konformazionala zuzenean erlazionatuta dago euren aktibitate biologikoarekin. Arrazoi honengatik, karbohidratoen egiturak zehazten dituen arau estrukturalak ulertuta, molekula biologikoek eta zelulek euren artean komunikatzeko erabiltzen duten hizkuntza ezagutzen lagun dezakete. Hizkuntza hau ikastea oinarrizkoa da karbohidratoetan oinarrituta dauden agente farmakologikoen diseinu arrazionala garatzeko. Horretaz gain, karbohidratoak bizitzaren funtsezko eraikuntza blokeak dira, beraz, haien kimikak eta bizitzaren jatorria zuzenki erlazionatu daitezke. Arrazoi horregatik, astrofisikaren gai nagusietako bat, izar arteko gunean dauden karbohidrato eta deribatuak aztertzean datza.[40,41]

Haien egitura kimikoari dagokionez, monosakaridoak polihidroxi aldehido edo zetonak dira, hau da, bi edo hidroxilo talde gehiago izateaz gain, karbonilo talde bat amaierako karbono atomoan (aldosa) edo bigarren karbono atomoan (ketosa) dute. Erregulartasunez, solido, likido eta gas fasean, karbonilo taldea hidroxilozko talde batekin konbinatzen da konposatu ziklikoa osatzeko. Monosakaridoak, karbono atomo kopuruaren arabera sailkatzen dira: triosa (C = 3), tetrosa (C = 4), pentosa (C = 5), hexosa (C = 6), etab. Horretaz gain, monosakaridoek eraldaketa kimiko ezberdin asko jasan ditzakete, hala nola, azetilazioa, fluorazioa, metilazioa, oxidazioa, laktonizazio eta sulfonazioa, besteak beste. Eraldaketa hauek azukre bakoitzaren ezaugarriak guztiz alda ditzakete. Hori dela eta, argi dago azukre asko daudela eta horien konplexutasuna eta aldakortasuna oso handia dela.

Azukreen ikerketak gas fasean Oxford Unibertsitateko John P. Simons irakasleak hasi zituen XXI. mendearen hasieran, ioi murrizketa laser espektroskopia teknikak erabiliz.[42] Hedapen supersonikoan eta mikrouhin espektroskopiaren erresoluzio altuaren konbinaketak, teknika hau laser espektroskopiaren alternatiba erakargarri bihurtzen du.[43,44] Lehen monosakaridoak detektatu ondoren, azken urteetan errotazio espektroskopiaren bidez, biologikoki esanguratsuenak diren beste azukre gutxi batzuk aztertu dira [45–48] Gas faseko errotazio espektroskopia, tesi honetan nagusiki erabilitako teknika esperimentala, sistema molekular konplexuen egitura zehatzak baldintza kontrolatuetan zehazteko ahalmena izateaz gain, konformero bakoitzaren populazioa,[49] dinamika atomikoak[50] eta egituraren parametro finen informazioa ere ematen du.[51] Horrela, tesi honen lehen zatian, monosakaridoen egitura, populazioa eta propietateak agertuko dira, eta lortutako informazioa helburu desberdinetarako erabiliko da.

Tesi honen bigarren zatian (VII-VIII kapituluak), pultsu multzifrekuentzian oinarritutako espektrometroarekin egindako hainbat lan aurkeztuko dira. Bereziki, Euskal Herriko Unibertsitatean eta Albertako Unibertsitatean egindako ikerketak aurkeztuko dira. Euskal Herriko Unibertsitateari dagokionez, espektrometro multifrekuentzia pultsatua eguneratu eta bere konfigurazioa aldatu egin zen, eta oraindik datu asko prozesatzen ari dira; hala ere, tesi honetan zehar esperimentalki lortutako emaitzen berri emango da. Zehazki, metil bentzoato uretako mikro solbatazioan eta 4-oxobutanonitrilo molekula astrokimikoan jarriko dugu arreta. Gainera, Albertako Unibertsitatean atzerrian egindako egonaldian protoi transferentzia bikoitza gertaerari buruz egindako azterlanak deskribatuko da.

I.2 Tesi honen helburu eta xedeak

Euskal Herriko Unibertsitateko Espektroskopia Taldeak esperientzia handia du gas fasean dauden molekula organikoen egiturak zehazten. 2009tik aurrera, taldea aitzindaria izan da hainbat molekula biologiko aztertzen, hala nola azukreak. Taldeak bi espektrometro eraiki ditu eta tesi honetan osagai elektroniko eta konfigurazio berriekin eguneratu eta optimizatu ditugu. Tesi honek bi helburu orokor ditu. Lehenik eta behin, talde honetan ikasi ohi diren molekula biologikoen azterketekin jarraitzea. Bigarrenik, etorkizuneko lanetarako espektrometroa garatzen eta eraikitzen jarraitzea.

Horrela, tesi honen helburuak honako hauek dira:

→ Hainbat molekula biologiko garrantzitsuren konformazioen determinazioa, espektroskopia errotazionalaren konbinazioa eta kimika kuantikoko kalkuluak erabilita. Emaitzak hainbat helburutarako erabiliko dira: molekulen osaera gidatzen duten faktoreak ezagutzea, bilaketa astronomikorako eta kimika konputazionalerako erreferentzia gisa, besteak beste.

→ Lan tartea edatu eta eguneratu laborategi berriaren aurretik eraikitako chirp pultsuko espektrometroa. Espektrometroaren hasierako espektro-tartea (6 eta 18 GHz artekoa) maiztasun-tarte txikienera zabaltzea (2 eta 6 GHz artekoa), eta elektronika aldatzea, 7 eta 14 GHz arteko tartean modu eraginkorragoan neurtzeko.

II. kapituluan, tesi honetan erabilitako espektroskopia errotazionalaren teknikari buruzko oinarrizko kontzeptuak laburbilduko ditut. Oinarrizko kontzeptuek tesi honetan aurkeztutako azterketa guztiak ulertzeko aukera ematen dute. Hemen ere, muntaia esperimentalak xehetasunez deskribatuko dira eta tesi honetan garatutako multifrekuentzia polarizazionan oinarritatako esperimentu berria azalduko da. Gainera, gure ikasketak egiteko behar ditugun oinarrizko beste kontzeptu batzuk laburki azalduko dira (hala nola kimika konputazionaleko kalkuluak).

III. kapituluak lehen azukrea den eritrulosa du ardatz. Kapitulu honetan, espektroskopia errotazionaleko laborategiko emaitzak eta irrati-teleskopioen emaitzak konbinatuko ditugu, molekula izarren arteko espazio gunean detektatzeko. Gure laborategian lortutako errotazio-konstanteak molekularen hatz-marka gisa erabili ziren irrati-teleskopioen datuak erabiliz bilatzeko. Hemen metodo berri bat konbinatzen dugu lagin solidoak prestatzeko, espektroskopia errotazionalaren datuak, bilaketa astronomikoak eta zehaztasun espazial handiko egitura molekularra zehaztea, eritrulosa molekula inoiz egin ez den bezala deskribatzeko.

IV. kapituluan fluoro/desoxi aldohexosatik eratorritako zenbait azukreri buruzko proiektu bat aurkezten dugu. Mutazio horrek molekuletan duen garrantzia oso handia da, azukre horien jarduera biologikoa eta antzekoak zeharo

desberdinak baitira. Kapitulu honetan, arreta jartzen dut hidroximetilo taldearen konformazio eta populazioan, bai fase likidoan, bai gaseosoan. Ikerketa hainbat ikerketa talderen lankidetzarekin egin zen, eta, jatorrian, sintesi, erresonantzia magnetiko nuklear, kimika kuantiko, dinamika molekular eta espektroskopia errotazionaleko datuak konbinatu ziren hidrogeno lotura kooperatiboaren eta hidroximetilo taldearen kokapena ulertzeko hainbat glukosa, manosa eta galaktosatan deribatuetan.

V. kapituluan glukonolaktona eta glukoronolaktonaren kasuak aurkeztuko dira, glukosatik eratorritako laktona garrantzitsu eta ugarienak. Hemen espektroskopia errotazionalaz baliatzen gara glukosaren bi deribatu horien arteko aldea ikusteko. Espektroskopia errotazionalaren erresoluzio handiak aukera ematen du gas-fasean dauden molekula horien eraztunaren konfigurazioa eta eraztunaren paketatzea zehazteko, eta azterketak erakusten du alde handia dagoela glukosaren eta haren deribatu laktonikoen artean.

VI. kapituluan monosakarido ugarienen deribatu azidoen ikerketa aurkezten dut: glukosa eta galaktosa. Egoera solidoko Erresonantzia Magnetiko Nuklearra, berkristalizazioak, espektroskopia errotazionala eta kalkulu kimiko kuantikoak konbinatzen ditugu azido glukuronikoaren eta azido galakturonikoaren bi anomeroetan (alfa eta beta) hdrogeno lotura sarea ikertzeko. Hidroxilo talde anomerikoak hidrogeno lotura sarean duen garrantzi handia erakutsiko da.

VII. kapituluan Flygare-Balle espektrometroan egindako azukreen azterketak alde batera utziko ditugu, eta espektrometro multifrekuentziaren emaitzetan zentratuko naiz. Kapitulu hau, Albertako Unibertsitatean egin zen (Egonlabur bekaren egonaldian). Hemen azido furoikoaren monomeroaren konformazio panorama eta azido furoiko dimeroetan protoi-transferentzia bikoitza aurkezten dira. Protoiak transferitzeko erreakzioa prozesu konplexua eta sotila da, eta interes handia du kimikan eta biologian. Hemen konbinatzen ditugu espektroskopia errotazionalaren datuak eta kimika konputazionalaren kalkulu berriak protoien transferentziaren erreakzioa deskribatzeko.

VIII. kapituluan jardunean dauden beste ikerketa batzuk laburbildu ditugu eta pultsu multifrekuentzia esperimentuan lortutako datuak dira. Konfigurazio esperimentalak garatzen emandako denbora luzearen ondorioz, ikasketa horiek guztiak oraindik amaitu gabe daude. Zehazki, metil bentzoatoaren mikrosolbatazioa eta 4-oxobutanonitriloarena aurkeztuko ditut. 4-oxobutanonitrilo molekula, izarren arteko eremu gunean bilatzen saiatuko gara etorkizunean. Ikerketa horiek datozen urteetan argitaratuko dira.

Azkenik, IX. kapituluak aurkeztutako lanaren ondorio orokorrak eta etorkizuneko joerak laburbiltzen ditu. Kapitulu bakoitzaren ondorioa bakoitzaren amaieran azalduko den arren, hemen tesi osoaren ondorio orokorrak azalduko ditut.

I.3 Erreferentziak

[1] S. G. Davey, *Nat. Rev. Chem.* **2020**, *4*, 507.
[2] C. Wang, C. C. Sun, *CrystEngComm* **2020**, *22*, 1149–1153.
[3] M. Schleeger, C. C. Vandenakker, T. Deckert-Gaudig, V. Deckert, K. P. Velikov, G. Koenderink, M. Bonn, *Polymer (Guildf).* **2013**, *54*, 2473–2488.
[4] A. Zunger, *Nat. Rev. Chem.* **2018**, *2*, DOI 10.1038/s41570-018-0121.
[5] P. G. Seybold, M. May, U. A. Bagal, **1987**, *64*, 575–581.
[6] S. Lüdtke, P. Neumann, K. M. Erixon, F. Leeper, R. Kluger, R. Ficner, K. Tittmann, *Nat. Chem.* **2013**, *5*, 762–767.
[7] D. D. Boehr, P. E. Wright, *Science (80-.).* **2008**, *320*, 1429–1430.
[8] K. L. Bak, J. Gauss, P. Jørgensen, J. Olsen, T. Helgaker, J. F. Stanton, *J. Chem. Phys.* **2001**, *114*, 6548–6556.
[9] K. L. K. Lee, M. McCarthy, *J. Phys. Chem. A* **2020**, *124*, 898–910.
[10] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.
[11] I. Uriarte, A. Insausti, E. J. Cocinero, A. Jabri, I. Kleiner, H. Mouhib, I. Alkorta, *J. Phys. Chem. Lett.* **2018**, DOI 10.1021/acs.jpclett.8b02339.
[12] E. Gloaguen, B. De Courcy, J. P. Piquemal, J. Pilmé, O. Parisel, R. Pollet, H. S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier, M. Mons, *J. Am. Chem. Soc.* **2010**, *132*, 11860–11863.
[13] G. R. Heath, E. Kots, J. L. Robertson, S. Lansky, G. Khelashvili, H. Weinstein, S. Scheuring, *Nature* **2021**, *594*, 385–390.
[14] E. DM, S. EK, *Nature* **1990**, *344*, 524.
[15] M. Herzik, M. A. H. Jr, *Encycl. Ref. Genomics Proteomics Mol. Med.* **2006**, 361–361.
[16] T. Kato, F. Makino, T. Nakane, N. Terahara, T. Kaneko, Y. Shimizu, S. Motoki, I. Ishikawa, K. Yonekura, K. Namba, *Microsc. Microanal.* **2019**, *25*, 998–999.
[17] H. Wu, R. Zhang, W. Zhang, J. Hong, Y. Xiang, W. Xu, *Chem. Sci.* **2020**, *11*, 4758–4765.
[18] T. M. Schwarz, C. A. Dietrich, J. Ott, E. M. Weikum, R. Lawitzki, H. Solodenko, E. Hadjixenophontos, B. Gault, J. Kästner, G. Schmitz, P. Stender, *Sci. Rep.* **2021**, *11*, 1–19.
[19] W. Zhou, L. Ge, G. A. Cooper, S. W. Crane, M. H. Evans, M. N. R. Ashfold, C. Vallance, *J. Chem. Phys.* **2020**, *153*, 184201.
[20] S. R. Domingos, C. Pérez, C. Medcraft, P. Pinacho, M. Schnell, *Phys. Chem. Chem. Phys.* **2016**, *18*, 16682–16689.
[21] J. L. Alonso, V. Vaquero, I. Peña, J. C. Lõpez, S. Mata, W. Caminati, *Angew. Chemie - Int. Ed.* **2013**, *52*, 2331–2334.
[22] L. B. Favero, I. Uriarte, L. Spada, P. Écija, C. Calabrese, W. Caminati, E. J. Cocinero, *J. Phys. Chem. Lett.* **2016**, *7*, 1187–1191.
[23] Z. Kisiel, B. A. Pietrewicz, P. W. Fowler, A. C. Legon, E. Steiner, *J. Phys. Chem. A* **2000**, *104*, 6970–6978.
[24] G. Klapper, L. Surin, F. Lewen, H. S. P. Muller, I. Pak, G. Winnewisser, *Astrophys. J.* **2003**, *582*, 262–268.
[25] Z. Su, N. Borho, Y. Xu, *J. Am. Chem. Soc.* **2006**, *128*, 17126–17131.
[26] D. Patterson, M. Schnell, J. M. Doyle, *Nature* **2013**, *497*, 475–477.
[27] Z. Kisiel, O. Dorosh, A. Maeda, I. R. Medvedev, F. C. De Lucia, E. Herbst, B. J. Drouin, J. C. Pearson, S. T. Shipman, *Phys. Chem. Chem. Phys.* **2010**, *12*, 8329–8339.
[28] C. E. Cleeton, N. H. Williams, *Phys. Rev.* **1934**, *45*, 234–237.
[29] K. G. Garrison, W. Gordy, **1957**, 899–901.
[30] T. J. Balle, E. J. Campbell, M. R. Keenan, W. H. Flygare, *J. Chem. Phys.* **1980**, *72*, 922–932.
[31] T. J. Balle, W. H. Flygare, *Rev. Sci. Instrum.* **1981**, *52*, 33–45.

[32] W. Stahl, G. Bestmann, H. Dreizler, U. Andresen, R. Schwarz, *Rev. Sci. Instrum.* **1985**, *56*, 1759–1762.
[33] J. U. Grabow, W. Stahl, H. Dreizler, *Rev. Sci. Instrum.* **1996**, *67*, 4072–4084.
[34] U. Andresen, H. Dreizler, J. U. Grabow, W. Stahl, *Rev. Sci. Instrum.* **1990**, *61*, 3694–3699.
[35] G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman, B. H. Pate, *Rev. Sci. Instrum.* **2008**, *79*, 053103.
[36] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[37] I. Uriarte, C. Pérez, E. Caballero-Mancebo, F. J. Basterretxea, A. Lesarri, J. A. Fernández, E. J. Cocinero, *Chem. - A Eur. J.* **2017**, *23*, 7238–7244.
[38] E. J. Cocinero, A. Lesarri, P. Ecija, F. J. Basterretxea, J. U. Grabow, J. A. Fernandez, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[39] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, *Essentials of Glycobiology, Third Edition*, **2017**.
[40] G. Cooper, N. Kimmich, W. Belisle, J. Sarinana, K. Brabham, L. Garrel, *Nature* **2001**, *414*, 879–883.
[41] M. Nuevo, G. Cooper, S. A. Sandford, *Nat. Commun.* **2018**, *9*, 1–10.
[42] F. O. Talbot, J. P. Simons, *Phys. Chem. Chem. Phys.* **2002**, *4*, 3562–3565.
[43] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[44] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.
[45] E. R. Alonso, I. Peña, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2016**, *7*, 845–850.
[46] I. Peña, L. Kolesniková, C. Cabezas, C. Bermúdez, M. Berdakin, A. Simão, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2014**, *16*, 23244–23250.
[47] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.
[48] I. Peña, C. Cabezas, J. L. Alonso, *Chem. Commun.* **2015**, *51*, 10115–10118.
[49] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[50] G. Feng, L. B. Favero, A. Maris, A. Vigorito, W. Caminati, R. Meyer, *J. Am. Chem. Soc.* **2012**, *134*, 19281–19286.
[51] E. R. Alonso, I. Peña, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2016**, *7*, 845–850.

III. kapitulua

Eritrulosaren laborategiko behaketa eta bilaketa astromikimikoa IEG-an

Kapitulu hau aurretik argirtaratua, itzulia eta egokitua izan da. Aran Insausti, Elena R. Alonso, Belen Tercero, José I. Santos, Camilla Calabrese, Natalja Vogt, Francisco Corzana, Jean Demaison, Jose Cernicharo, and Emilio J. Cocinero. Laboratory Observation of, Astrochemical Search for, and Structure of Elusive Erythrulose in the Interstellar Medium. *J. Phys. Chem. Lett. 2021, 12, 4, 1352–1359*

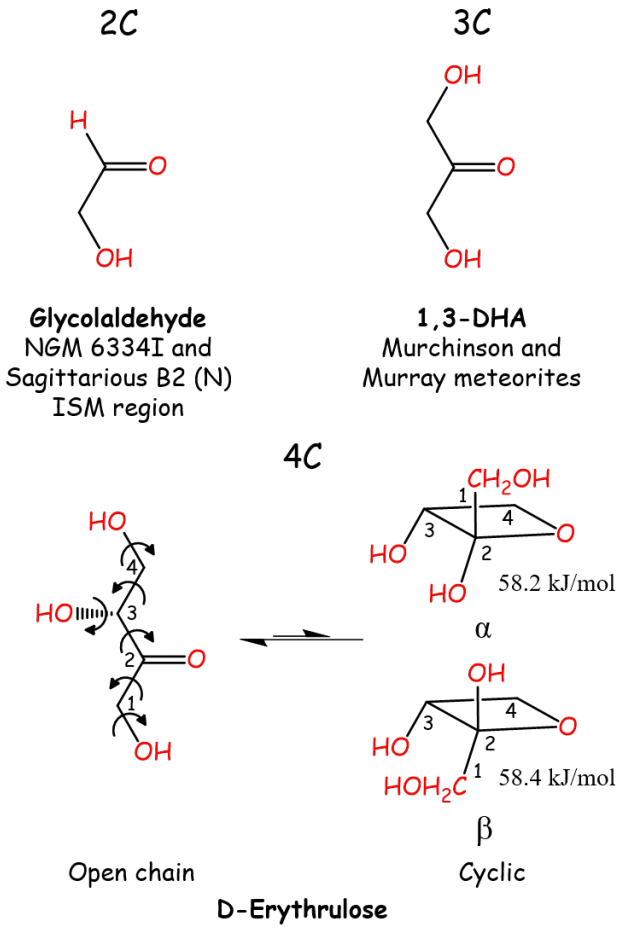
III.1 Sarrera

Espektroskopia errotazionala izarrarteko espazioaren gunean (IEG) dauden interes biologikodun molekulak identifikatzeko teknikarik erabiliena eta baliotsuena da. Hala ere, gaur egun ezin izan da karbohidrato bat ere ez identifikatu lurretik kanpo. Kapitulu honetan, IEG egon litekeen eritrulosa azukrearen Fourierren transformatuan oinarritutako errotazio espektroaren identifikazioa eta interpretazioa egin da. Horretarako, eritrulosa lagin oliotsua lurruntzeko lagin-prestaketa metodo berria diseinatu behar izan da. Lortutako errotazio espektroen datuak, eritrulosa molekula B1-b, TMC-1 eta Sgr B2 IEG-ko eremu desberdinetan bilatzeko erabili dira.

Karbohidratoen presentzia IEG-n, Lurreko bizitzaren eratorriarekin erlazionatu liteke. Bizitzaren agerpena, gizakiaren jakin-mina piztu du historia osoan zehar, eta gaur egungo astrokimika eta astrobiologia ikerketen gai ezaguratsuenetakoak dira. IEG detektaturiko 200 molekulek, molekula prebiotikoak ere IEG-n ageri direla iradokitzen du.[1,2] Detektatutako molekuletatik, erdiak baino gehiagok, bost atomo baino gehiago dituzte. Molekula horiei astrofisikariek, 'Molekula Organiko Konplexu' (MOK) deritzo. Glikolaldehido, bi karbonodun (2*C*) molekula (III.1 irudia) eta azukreekin erlazionatutako unitate kimiko txikiena, 2000. urtean detektatu zen IEG-n.[3] Ondoren, irrati-teleskopioen hobekuntza teknologikoak, bere noizbehinkako detekzioa zeruko gune desberdinetan ahalbidetu du.[4] Bestalde, glizeraldehido, hiru karbonodun azukrea, eta baita ere azukreekin erlazionatutako beste molekula ugari, adibidez laktaldehidoa, laborategietan lortutako errotazio espektroen bidez IEG-ko gune desberdinetan bilatu dira.[5,6] Hiru karbonodun keto-azukrea den 1,3-dihidroxiazetona (1,3-DHA, III.1 irudia) Murchison eta Murray meteoritoetan bestelako hirutik (3*C*) seira (6*C*) karbono atomoz osatutako bestelako azukre alkohol, azukre azido, azido dikarboxiliko eta desoxiazukre azidoekin batera detektatu eta karakterizatu da.[7] Gainera, euren ugaritasuna, amino azidoen ugaritasunen parekoak ziren.

Bizitzarako ezinbestekoak diren azukreen detekzioa IEG-an (bestelako molekula biologikoen detekzioen antzera) oso erabilgarria da IEG-ko gune desberdinetan ematen diren erreakzio molekularrak eta euren mekanismoak ezagutzeko, eta modu horretan, MOK molekula konplexuagoen eratorri kimikoa ezagutzeko. Halaber, molekulen detekzioa irrati-teleskopioen bidez ez da bat ere prozesu tribiala, izan ere behaketetan lortutako espektroak IEG-n dauden konposatu ugarien emisioak direla eta "belar astrofisiko" (*astrophysical weeds*) deritzon kutsadura eta seinaleez josita daude. Hori dela eta, laborategiko errotazio espektroen identifikazioa ezinbestekoak dira IEG-ko espektroetan seinaleen zalantza gabeko identifikazioan, izan ere, laborategiko lortutako seinaleak zuzenean erreferentzia moduan erabiltzen baitira.[8]

Tamalez, monosakarido eta antzeko molekulen itxurazko tamaina txikia eta sinplea duten arren, detekzio astrofisikoa ahalbidetu dezakeen errotazio espektroen datuen falta nabaria da. Arazo nagusia, azukreak gas fasera transferitzeko aurkitzen da, izan ere, lurruntze konbentzionaleko berotze prozesuan, lurruntzeko tenperaturara iritsi aurretik azukreak normalean deskonposatzen direlako. 2012. urtean, arazo hori konpondu zen pultsu ultra-azkarren bidezko laser ablazio baporizazio prozesuaren bidez.[9] Antzeko prozedura esperimentala aurretiaz bestelako molekula termolabilak ikertzeko erabili izan da, adibidez amino azidoak. [10]



III.1. irudia Karbono-unitate orden gorakorrean interes astro(bio)kimikoa duten azukre-blokeen sekuentzia. Glikolaldehidoa (goian ezkerrean), 1,3 dihidroxiazetona (1,3-DHA, goian eskuinean) eta eritrulosa (behean). Glikolaldehidoa eta 1,3-DHA detektatu dira dagoeneko IEG-an eta meteoritoetan, hurrenez hurren. Eritrulosa, kapitulu honetako intereseko molekula da eta IEG-an identifikatzeko helburu posiblea da. Eritrulosak lau muturreko eraztunak sortzeko aukera du, baina kate linealekin alderatuta, energia oso handia du. Beraz, polimorfismo estrukturala, forma linealetara murrizten da soilik.

Gaur egun, azukre askoren errotazio espektroak neurtu eta karakterizatu gabe daude. Datu gabezia horrek bultzatuta, kapitulu honetan eritrulosa azukrearen azterketa aurkezten dugu, alde batetik bere errotazio espektroaren identifikazio eta karakterizazioa laborategian egin da eta beste alde batetik, EIG-ko hainbat gunetan bilaketak eta detekzio ahaleginak egin dira. Gure ikerketan, laginak prestatzeko prozedura berria, errotazio espektroaren erregistroa, eritrulosa molekularen determinazio estrukturala eta irrati-teleskopioen behaketak aurkezten dira. Lortutako paregabeko sentikortasun eta erresoluzio esperimentala erabili da eritrulosaren orekako egitura zehaztasun handiz determinatzeko.

Eritrulosa (1,3,4-trihidroxibutan-2-ona, $C_4H_8O_4$, III.1 irudia) itxura oliotsua duen lagina da. Hori ikusita, errotazio espektroa erregistratzeko lehenengo saiakerak beroketa baporizazio bidez CP-FTMW esperimentuan burutu ziren. Hala ere, lortutako espektroetan ez zen eritrulosa molekularen seinalerik ikusi, zihurrenik lagina berotzean lurrundu aurretik deskonposatu zelako. Arazo hori konpontzeko, laser baporizazio sistema erabiltzea erabaki genuen, izan ere, bestelako (5*C*) eta (6*C*) azukreak teknika horren bidez aurretiaz ikertuak izan direlako.[9,11–16] Kasu honetan, eritrulosa molekula sirope antzeko konposatua da eta hori dela eta, laser sistemaren lagin euskarrian beharrezkoak diren hagaxka moduko lagin solidoak prestatzeko arazoak aurkitu ziren. Aitzitik, antzeko laginekin erabilitako teknikek (adibidez lagina NaCl-rekin nahaztu eritrosan[17] edo glizinarekin nahastu eskopinan,[18] eta gero konpaktatu eta lehortu hagaxka solidoa lortzeko) ez zuten esperotako emaitza onak eman eritrulosarekin. Gure kasuan, zenbait substantzia solidotzaile frogatu ziren (NaCl, glicina, *Loctite* pegamentua*,* talko ahutza, magnesio sulfatoa eta silika gela) proportzio desberdinekin (gutxi gorabehera 25, 50 eta %75) eta lehortze denbora desberdinak erabilita. Laginak banan-banan frogatu ondoren, esperimentua egiteko lagin aproposena, talko-hautsa eta eritrulosa 50:50 (m/m) eta hutseko lehorgailuan 2-3 astez lehortu ondoren aurkitu zen. Lortutako hagatxoa guztiz solidoa ez zen arren, kontsistentzia egokia zuen energia handiko laser pultsuak pairatzeko eta hortaz errotazio espektroa erregistratzeko. Traba esperimental horretaz gain, solidoan egindako erresonantzia magnetiko nuklearrak, lagina keto eta hidratatu forman ~92:8 proportzioan zegoela ikusi genuen, bere detekzio eta analisia are gehiago eragotziz.

III.2 Metodologia

Xehetasun esperimentalak

Eritrulosaren errotazio-espektroa, Euskal Herriko Unibertsitatean bertan eraikitako Fourierren transformatuan eta izpi-molekularrean oinarritutako mikrouhin espektrometroaren bidez (cav-FTMW) neurtu zen (xehetasun guztiak 2. kapituluan).[19] Eritrulosa, lagin oliotsua denez laser baporizazioak behar duen hagaxka egiteko arazoak egon ziren. Hori dela eta lagin prestaketa prozesu berria diseinatu behar izan genuen (begiratu sarrera ataletan). Lagina gas fasera deskonposatu gabe lurrundu zen 7-8 mJ pultsuko ultramore laser izpien bidez (355nm-ko uhin luzeera eta 30 ps pultsu denbora). Lurrundutako lagina gas garraiatzailearekin nahastu (Ar, 6-8 bar), eta molekula/gas garraiatzaile nahasketa izpi supersonikoaren bidez barrumbe ganbara barruan hedatzen da zirrikitu txiki baten bidez (1mm), lortutako errotazio tenperatura oso baxua delarik (T < 5K). Hedapenaren ostean, molekulak potentzia baxuko (0.1-100mW) eta 0.2-0.5 μs iraupeneko mikrouhin pultsuaz irradiatzen dira. Erradiazioaren frekuentzia errotazio maila batekin bat egiten duenean, molekulek erradiazioa absorbatu eta denbora domeinuan erregistratzen den emisio espontaneoa ematen da. Denbora domeinuan erregistratutako espektroa frekuentzia domeinura Fourierren transformatuaren bidez bihurtzen da. Neurketaren zehaztasuna 3 kHz ingurukoa da eta 10 kHz baino bananduago dauden seinaleak desberdindu daitezke.

Orekako geometria erdi-esperimentala

Laburki, errotazio espektroskopian lortutako datuen eta maila altuko *ab initio* konputazio datuen konbinazioren bidez orekako geometria erdi-esperimentala kalkulatu genuen (irakurri 2. kapitulua xehetasun gehiago jakiteko). Oreka geometriako errotazio konstanteak kalkulatzeko, esperimentalki lortutako oinarrizko egoerako errotazio konstanteen zuzenketa MP2/cc-pVTZ indar-eremu kubikoaren bidez egin zen.[20] Lortutako errotazio konstante kopurua ez denez nahikoa molekularen egitura osoa determinatzeko erregresio metodo mixtoa erabili zen, non parametrorik ez diren finkatu.[21,22] Metodo hau, nahiko erraza da, izan ere maila ertaineko kalkulu mekano kuantikoek atomo gehienen posizio zehatza ezagutzea ahalbidetzen du erresoluzio nahiko zehatzerekin.

Xehetasun konputazionalak

Simulazio konputazionalak eritrulosaren errotazio-espektroan seinaleen bilaketan eta identifikazioan laguntzeko erabili ziren. Hortaz, hasiera batean, isomero lineal zein ziklikoen panorama konformazionala aztertu zen. Lehenengo energia baheketa, mekanika molekularrean oinarritutako metodo azkarren desberdinen bidez egin zen (erabilitako indar eremuk: MMFFs, OPLS eta AMBER). Ondoren, lortutako egituren optimizazio geometrikoan,[23] DFT (B3LYP-GD3BJ) [24–26] eta *ab initio* (MP2)[27] metodo mekano kuantiko aurreratuak erabili ziren. Kalkulu mekano-kuantiko guztietan *Pople*-ren balentzia banatutako eta

difusio eta polarizazio funtzioen bidez handitutako "*triple-ζ*" deritzon basea (6-311++G(d,p)) erabili zen.[28] Metodo konputazionalen bidez lorturiko errotazio parametroak eranskinetako AIII.6-AIII.7 tauletan daude.

III.3 Emaitzak eta eztabaida

Bizitzaren oinarri diren (5*C*) eta (6*C*) familiako azukreak, hainbat ikerketa espektroskopikoren ardatz izan dira,[9,11–16] baina, aurretiaz adierazi den moduan, euren datu esperimentalak nolabaiteko atzerapen kronologikoa izan zuten, batez ere gas fasera deskonposatu gabe pasatzeko arazoak zirela eta (lehengo ikerkuntza 2012. urtean).[9] Horren adibide da, Alonsoren taldeak 2002-tik aurrera erregularki ikertutako amino-azidoak,[10,30] izan ere lehenengo ikerketa 1978. urtekoa baita.[29] Gaur egun arte, naiz eta (4*C*) azukreak tamaina txikikoak eta kimikoki oso sinpleak diren, bibliografian, errotazio espektroskopiaren bidez bakarra ikertu dela ikusi daiteke, eritrosa hain zuzen ere.[17] Emaitza horien faltaren arrazoi nagusia, 4*C* azukreak, 5*C* eta 6*C* azukreekin konparatuta, lagin higroskopikoagoak eta oliotsuagoak direla da. Ezaugarri horrek eta azukreen deskonposatzeko erraztasunak erronka esperimental latza bilakatzen ditu. [31]

Eritrulosa (1,2,4-trihidroxibutan-2-ona, $C_4H_8O_4$, III.1 irudia), zeto-treosa, C2 posizioan karbonilo taldea duen 4*C* motako monosakaridoa da. Industrian zenbait garrantzizko aplikazio ditu, batez ere autobeltzatzeko kosmetikoen alorrean.[32] Gainera, orain dela gutxi, azido askorbikoaren (C-bitamina) Maillard-en degradazio bidearen produktu nagusietako dela ere ikusi da.[33] Baina gure ikerkuntzarako are garrantzitsuagoa dena, eritrulosa monosakaridoa hautagai bikaina, eta hurrengo urrats logikoa da IEG bilatzeko, izan ere, 2*C* eta 3*C* azukreak detektatu eta bilatu dira zeruko zenbait guneetan (III.1 irudia). Lagina esperimentalki ikertzeko lehenengo ahalegin porrota CP-FTMW errotazio espektrometroan egin zen, laginaren tenperatura apurka-apurka igoz (20-140 ºC), arazoa ziurrenik laginaren deskonposatzea berotze prozesua izan zen. Arazo hori ekiditeko, lagina aurretiaz bestelako 5*C* eta 6*C* azukreak lurruntzeko erabilitako laser ultra-azkarren ablazio metodoa erabili zen. [9,11–16] Baporizazio teknika honek behar duen lagin hagatxoa egitea ez zen lan erreza izan, izan ere, eritrulosa olio antzeko lagin likatsua da eta oiko lagin prestaketa prozesua berriz diseinatu behar izan zen. Lagina prestatzeko metodologia berriaren informazioa kapitulu honen sarreran aurki dezakezue. Gainera, arazo horretaz gain, fase solidoko erresonantzia magnetiko nuklearrak, keto:hidratatu ~92:8 proportzioa ikusten zen, bere detekzio eta analisia are gehiago eragotziz.

Eritrulosa isolatuaren identifikazioa errotazio espektroskopiaren bidez.

Behin lagin prestaketaren arazoak konponduta zeudela, eritrulosaren panorama konformazional zehatza eta eguneratua[34] metodo konputazionalen bidez ikertu zen bi urratseko (1 MM + 2 QM) estrategia erabilita. Datuak, errotazio esperimentua giatzeko erabili ziren (begiratu Metodologia konputazionala tesi

honen 2. kapituluan). Hasiera batean, isomero ziklikoak ere kontuan hartu ziren gure abiapuntu teorikoan, baina kalkuluen arabera, forma aziklikoa bakarrik aurkituko zen gure laginean, izan ere energia diferentzia isomero ziklikoarekiko 60 kJ/mol ingurukoa baita (III.1 irudia).

Eritrulosaren errotazio espektroa gas fasean, unibertsitatean bertan eraikitako eta ultra-more laser baporizazio sistemaz hornitutako izpi molekular eta Fourierren transformatuan oinarritutako mikrouhin espektrometroan (cav-FTMW) [19] neurtu zen (esperimentuaren zehaztasun guztiak tesi honen 2. kapituluan aurki daitezke).[9] Laserraren ezaugarriak, ultra-morea erradiazioa (355 nm-ko uhin luzerakoa), pultsu ultra-laburra (40ps inguru) eta energia altua (7-8 mJ/pultsuko), lagina deskonposizioa saihesten du, beroketa prozesuaren deskonposatze arazoak konponduz.[31] Konputazionalki trantsizio intentsuenak iragartzen ziren frekuentzia inguruetan, mikrouhin espektrometroarekin frekuentzia desberdinetan ekorketa eginez, errotazio trantsizio seinale multzoak osatzen zuen patroia identifikatu zen: seinaleak Hamiltondar erdi-zurrunaren bidez doitu ziren (Watson-en S-erredukzioa eta I^r-errepresentazioaz[35,36]), eta lortutako balioak III.1 taulan bildu dira. Errotazio konstanteak molekularen hatz-marka ukiezina da. Lortutako balio esperimentak eta teorikoak konparatuta, (III.1 taula), zalantzarik gabe, detektatuko konformeroa ***I*-1** dela ondorioztatu zen. ***I*-1**, energia baxueneko konformero lineala da. Beste konformeroen detekzioa saiatu zen arren ez zen beste konformerorik identifikatu errotazio espektroan.

III.1. **taula** Eritrulosaren molekula eta ^{13}C isotopologoen parametro errotazionala teoriko eta esperimentalak.

	***I*-1**[d]	**Parent**	**^{13}C1-$^{12}C_3H_8O_4$**	**^{13}C2-$^{12}C_3H_8O_4$**	**^{13}C3-$^{12}C_3H_8O_4$**	**^{13}C4-$^{12}C_3H_8O_4$**
***A*[a]**	3135	3133.10679(27)	3119.677(10)	3130.2228(82)	3121.348(17)	3110.608(16)
B	1410	1422.65810(10)	1413.06870(26)	1421.56732(22)	1417.83894(45)	1412.17541(39)
C	1143	1157.320960(77)	1150.01309(19)	1156.25737(15)	1155.08211(31)	1148.45454(29)
D_J		0.2699(11)	[0.2699][e]	[0.2699]	[0.2699]	[0.2699]
D_{JK}		0.2245(47)	[0.2245]	[0.2245]	[0.2245]	[0.2245]
D_K		1.225(27)	[1.225]	[1.225]	[1.225]	[1.225]
d_1		-69.30(81)	[-69.30]	[-69.30]	[-69.30]	[-69.30]
d_2		-14.89(29)	[-14.89]	[-14.89]	[-14.89]	[-14.89]
***N*[b]**		45	9	9	9	9
σ[c]		1.1	1.6	1.2	2.5	2.4

[a] Errotazio konstante esperimentalak (MHz-tan) eta distortsio zentrifugoko konstanteak (D_J, D_{JK}, and D_K in kHz and d_1 and d_2 kHz-tan)
[b] Doitutako trantsizio kopurua.
[c] Mikrouhin doiketaren errore batezbesteko koadratikoa kHz-tan
[d] Oreka egituraren balioak B3LYP-D3BJ / 6-311++G(d,p) maila teorikoan.
[e] Distortsio zentrifugoko konstanteak jatorrizko espeziearen balioetara finkatuak

Laginaren purutasuna < %40 den arren, esperimentalki lortutako sensibilitate handiak, ***I*-1** konformeroaren ^{13}C isotopologo mono-ordezkatuak

$^{13}C^{12}C_3H_8O_4$ abundantzia naturalean (%1.1) detektatzea ahalbidetu zuen. Errotazio konstanteen doiketak jatorrizko espezian erabilitako Hamiltondar berdinarekin egin ziren. Lortutako balioak III.1 taulan adierazi dira. Esperimentalki neurtutako espezie guztien frekuentziak kapitulu honen eranskinetako AIII.1-AIII.5 tauletan daude. Eritrulosaren izaera higroskopikoa ikusita, eritrulosa eta uraren arteko klusterra ikusten saiatzera animatu ginen, baina ezin izan genuen detektatu esperimentalki.

Erithrulosak, kate irekiko molekula da eta hori dela eta, ustez konformazio egonkor ugari izan dezake bere sei barne-tortsioen errotazioaren bidez (III.1 irudia). Hala ere, bakarrik konformerorik egonkorrena ***l*-1** esperimentalki detektatu zen. ***l*-1**, hiru H-lotura intramolekularren bidez (2x O-H···O=C eta 1x O-H···O-H) egonkortuta dago. Iragarritako ikuspegi konformazionalaren arabera B3LYP-D3BJ maila teorikoan, bigarren konformero egonkorrena, ***l*-2**, detektaturiko konformeroren gainetik 5.2 kJ/mol ezegonkorragoa da. Bi konformero horien arteko egitura desberdintasun nabariena, -O4H-k -O3H-ren arabera duen disposizioan dago, izan ere, ***l*-1**-en *g-* eta ***l*-2** *g+* diedroak dituzte (∠O4-C4-C3-O3 of ∼-65º (***l*-1**) eta ∼+55º (***l*-2**)). ***l*-1**-tik ***l*-2**-rainoko bihurtze energia 20 kJ/mol ingurukoa da, eta beraz, erlaxazio prozesua jet espantsioan ez da printzipioz posible. Hala ere, ***l*-2** konformeroaren ezegonkortasuna eta dipolo momentu baxua, bere detekzioa eza azal dezake. Hurrengo konformeroak (***l*-3**, ***l*-4** eta ***l*-5**), energian ezegonkorragoak dira (> 6 kJ/mol), bi hidrogeno lotura bakarrik dituztelako, eta detektatutako konformeroak hiru hidrogeno lotura intramolekular ditu. ***l*-2**-tik ***l*-5**-erainoko konformeroen egiturak eta parametro espektroskopiko teorikoak tesi honen eranskinetan aurki daitezke.

Eritrulosaren zehaztasun handiko orekarko egitura

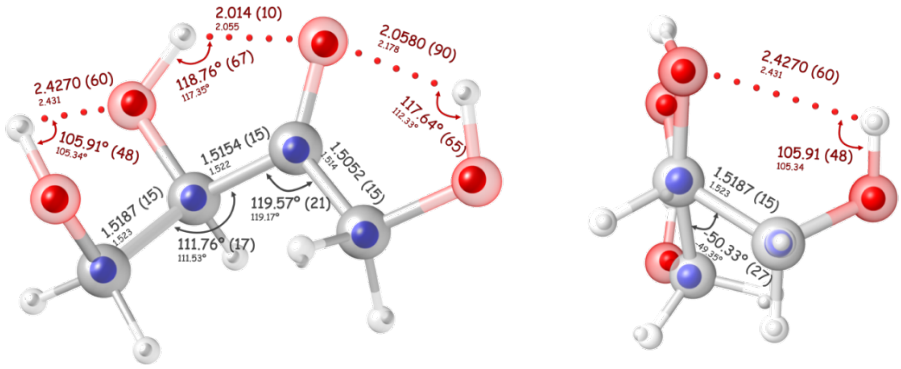
Gure cav-FTMW espektrometroaren sentikortasun altuaren bidez ^{13}C isotopologo mono-ordezkatuak abundantzia naturalean (%1.1) detektatu ziren eta, horrek, 4*C* eritrulosa azukrearen, oreka parametro egitura zehatza ezagutzea ahalbidetu zuen. Eritrulosaren egitura lau edo karbono gehiago duten bestelako monosakaridoen egiturarekin konparatuta, bestelako azukreak 5- edo 6-muturreko forma ziklikoa duten arren eritrulosak kate lineala baino ez du hartzen.[37] Erithrulosaren kasuan, eratu daitekeen ziklo bakarrak lau muturreko eraztuna da eta horrek tentsio handia sortzen du eraztunean (III.1 irudia). Lan honek, azukreak fase isolatuan kate lineala osatzen duen adibidea aurkezten du.

Gaur egun ezagutzen den karbohidratoen egitura estrukturala fase kondentsatuko egituretan oinarritzen da, batez ere, X-izpien edo erresonantzia magnetiko nuklearra moduko tekniken bidez lortuta. Hala ere, kontuan izan behar da, egitura horiek disolbatzailea, matrizea edo kristalen paketatzearen eragina izaten dezakela. Horretaz gain, kimika teorikoak ere azukreen egitura ezagutzera lagundu du,[38–42] baina kimika konputazionalean izugarrizko aurrerakuntzak egin diren arren, lortutako emaitzen zehazkabetasuna dela eta, lortutako datu

konputazionalak oraindik esperimentalki balioztatuak izan behar dira. Laser infragorri espektroskopian oinarritutako gas faseko esperimentuek hainbat karbohidratoen datuak helaratu zituen[43–48] eta horren ondoren, bestelako hainbat taldek metodologia alternatiboak erabili zituzten zenbait azukreen informazio estruktural eta konformazionala lortzeko.[49–54] Metodo guzti horien emaitzek, halabeharrez kalkulu teorikoak behar dituzte emaitzen interpretazio egokia lortzeko eta gainera, batek ere ez du ahalbidetzen egitura esperimental zehatza ezagutzea.

Testuinguru horretan, errotazio espektroskopia tresna paregabe moduan nabarmendu da, izan ere bere paregabeko erresoluzioaz baliatuz (~kHz), tautomero, konformero eta isotopologoak diskriminatzeko aukera ematen baitu. Ez hori bakarrik, isotopologoen errotazion konstante esperimentalak, molekula isolatuen hiru dimentsioko egitura zehatza determinatzeko bidea irekitzen du. Lan honetan, orekako geometria semi-esperimental (r_e^{SE}) deritzon medoa erabili dugu, ziurrenik metodorik sofistikatuena eta zehatzena baita molekulen egiturak ezagutzeko. Metodo honek, informazio kimiko kuantiko teorikoa eta esperimentala konbinatzen ditu, non orekako errotazio konstanteak, esperimentalki lortutako oinarrizko egoerako errotazio konstante eta erroto-bibrazio zuzenketen bidez lortzen diren, indar eremu kubikoak erabili direlarik. Dena den, bere aplikazioa sistema txikietara mugaturik dago eta gaur egun kasu gutxi batzuetan bakarrik aplikatua izan da.[55–57] Hiru arrazoi nagusi daude horretarako. Lehenengo, datuen tratamendua ez da tribiala; bigarrena, isotopologoen errotazio konstante esperimentalak behar dira, eta normalean molekularen txikietara daude mugatuta. Hirugarrena, beharrezkoak diren maila-altuko kalkuluak egiteko beharrezkoa den denbora oso luzea da. Lan honetan erabilitako metodologiaren laburpena kapitulu honen Metodologia atalean aurki daiteke eta informazio zehatzagoa kapitulu honen eranskinetan dago. III.2 irudian egituraren parametro garrantzitsuenak daude adierazita eta egituraren parametro guztiak eranskinetako AII.8 taulan ageri dira. Egituraren doiketaren emaitzak, eranskinetako AIII.10 taulan daude eta atomoen koordenatu kartesiarrak eranskinetako AIII.11 taulan ageri dira.

***I*-1** konformeroaren egiturak hiru interakzio moderatuko H-lotura bidez egonkortuta dago, bi OH···O=C [lotura distantzia: 2.014(10) eta 2.058(9) Å; angeluak: 118.76(65) eta 117.64(65)º)] eta O-H···O-H bat [lotura distantzia: 2.427(6) Å; angelua: 105.91(48) º]. Antzeko egiturak ikusi ziren glizeraldehido[58] eta dihidroxiacetona[59] molekuletan, non bi H-lotura moderatu O-H···O=C [2.019/2.346 eta 2x2.060(14) Å; angeluak: 119.8/108.7 eta 2x 115.3(9)º] ere bai ikusi ziren. Beraz, molekulan bi diedro desberdin ditugu, albo kateak osatzen duen txandakatu konformazioa eta karbono kateak duen eklipsatu motako konformazioa (∠C1-C2-C3-C4 = -50.33(27)º).



III.2. irudia Identifikatutako D-eritrulosaren egitura esperimentalaren bi ikuspuntu. *I*-1 eritrulosaren orekako geometria erdi-esperimentalaren (r_e^{SE}, esfera ilun txikiak) eta MP2/6-311++G(d,p) egitura (atomo garden handiagoak) gainjartzea. Egitura-parametro garrantzitsu batzuk, hala nola hidrogeno-loturak (Å-tan), lotura-luzerak (Å-tan) eta angeluak (gradutan) ere adierazten dira oreka geometria esperimental zein teorikoetarako.

Eritrulosaren bilaketa astronomikoak IEG-ko gune hotzetan: B1-b, TMC-1 eta Sgr B2.

Eritrulosaren karakterizazioa laborategian burutu eta gero, molekularen bilaketa espazioan hasi zen, ISM-ko zenbait gune hotzetan fokalizatuz. Gune hotz horiek, errotazio partiketa funtzioaren balio altuak saihesteko aukeratu ziren, izan ere balio altuek seinaleen intentsitatea txikitu dezakete. Bilaketa guneak, masa txikiko eta izarraren sorreran lehenen etapan dagoen Barnard 1 (B1-b) hodei ilun hotza, aurre-izar zuloa TMC-1 eta masa handiko izarra eratzeko eskualde aktiboa den Sagittariun B2 (Sgr B2) izan ziren. Detekzioen saiakuntzak energia baxuko a-motako eta c-motako seinaleetara mugatu ziren non frekuentziaren jakingabetasuna 1 MHz baino txikiagoak ziren eta laborategiko datuak irrati-teleskopio datuekin konparatzeko estrapolatu ziren. Irrati-teleskopio datuak 30 m IRAM irrati-teleskopioa 3mm-ko B1-b eta TMC-1 espazio guneko datuak erabili ziren.[60,61] Sgr B2-ko datuak “PRebiotic Interstellar MOlecule Survey[1]” (PRIMOS)-ko[62] datu publikoak 100 m-ko Green Bank teleskopioko (GBT) datuak erabili ziren. MADEX kodearen[63] erabili zen laborategian lortutako datu espektroskopikoak ustiatzeko eta molekularen espektro sintetikoa lortzeko III.2 taulan adierazitako, eta normalean erabilitako, parametroak erabili ziren.[61,62] Dentsitate zutabea izan zen libre utzitako parametro bakarra erabilitako ereduan.

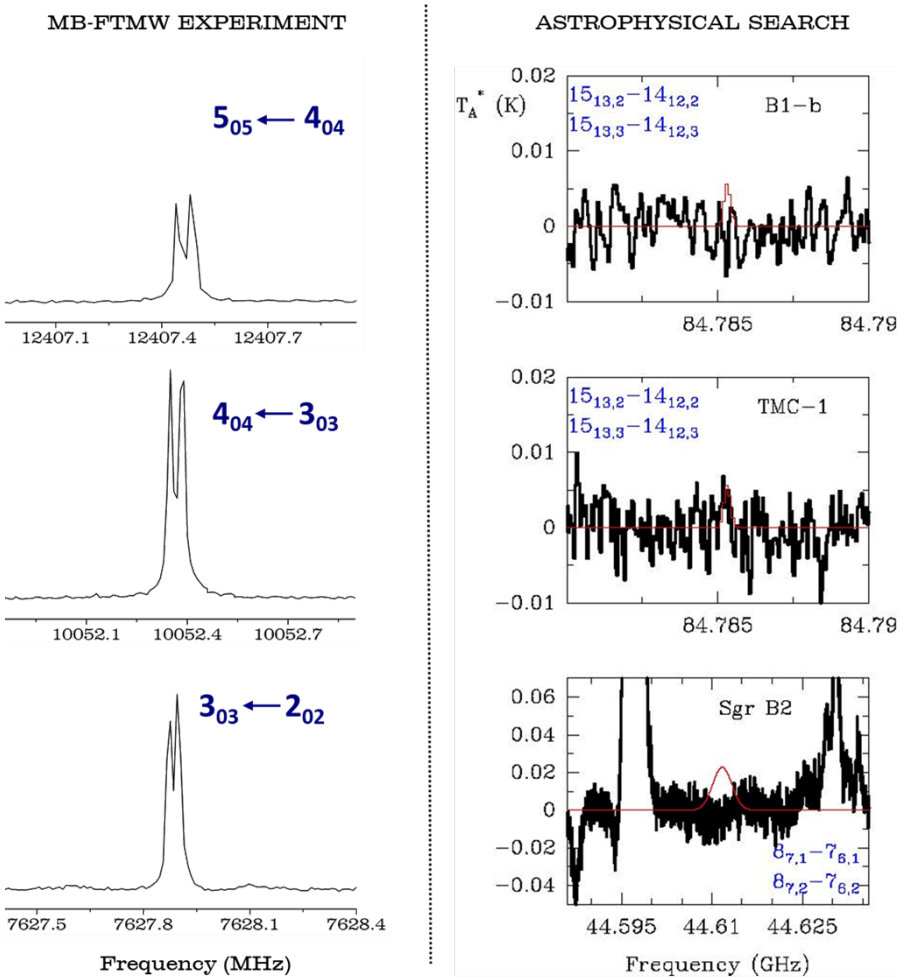
[1] http://www.cv.nrao.edu/~aremijan/PRIMOS/ helbidean eskuragarri dago PRIMOS datu sorta osoa, behaketa-estrategiaren xehetasunak eta maiztasun-estaldura orokorraren informazioa

III.2. **taula** IEG eritrulosaren bileketak egiteko erabilitako parametroak.

Iturria	HPBW[a] [´´]	Frekuentzia[b] [GHz]	V_{LSR}[c] [km s^{-1}]	$\Delta\nu_{FWHM}$[d] [km s^{-1}]	D_{sou}[e] [´´]	T_{rot}[f] [K]	N × 10^{13}[g] [cm^{-2}]
B1-b	30-21	82.5-115.5	6.7	0.7	60	12	≤ (1.0 ± 0.3)
TMC-1	30-21	82.5-115.5	6.0	0.7	60	10	≤ (1.0 ± 0.3)
Sgr B2	80-15	7-50	64	25	60	10	≤ (50 ± 20)

[a] Half power beam width (HPBW) of the telescope at the considered frequencies.
[b] Analisian kontuan hartutako maiztasun tartea.
[c] Iturriaren abiadura erradiala Atseden Sistema Lokalarekiko (LSR).
[d] Gailurraren altueraren erdialdeko zabalera (FWHM).
[e] Adopted source size.
[f] Errotazio tenperatura (kitzikapena).
[g] Dentsitate zutabearengoiko muga maximo deribatua.

III.3. irudian esperimentalki lortutako datuen artean aukeratutako frekuentzien datuak eta MADEX-en bidez lortutako datuak erakusten dira. Datu horiek zarata-mailaren gainetik eritrulosaren seinalerik ez dagoela erakusten dute. III.2 irudian adierazitako trantsizioak intentsitate handieneko trantsizioak dira erabilitako modeloaren arabera. Dentsitate zutabeko limite maximoa III.2 taulan adierazita daude.



III.3. irudia (Ezkerreko aldean) a motako progresioa ($J+1_{0,J+1} \leftarrow J_{0,J}$, J = 2, 3, 4 balioekin) laborategian erregistratuta, UV ultraazkar lurruntze-sistemaz hornitutako cav-FTMW espektrometroa erabilita. Trantsizio guztiak Doppler efektu instrumentalaren bidez banatzen dira. (Eskuineko aldean) B1-b, TMC-1 eta Sgr B2-ren behaketa astronomikoak tresna ezberdinekin (histograma-espektro beltzak) aukeratutako maiztasunetan egindako espektro sintetikoekin batera, 2. taulan (profil gorria)). A V_{LSR} of +6.7, +6.0, eta +64 km/s B1-b, TMC-1 eta Sgr B2, suposatu da, hurrenez hurren.

III.4 Ondorioak

Lurreko bizitzarako ezinbestekoa den azukreen garrantzia ikusita, edozein motako detekzioa gure planetatik at oso informazioa esanguratsua da. Azken hamarkadetan, behatoki astrofisikoen aurrerakuntzez baliatuz, zenbait molekula prebiotikoren detekzioa lortu da; adibidez amino azido eta azukre aitzindariak, zeintzuk normalean IAE-n izar arteko hauts guneetan lokaliza daitezke.[64] Kontestu horretan, eritrulosaren laborategiko karakterizazio espektroskopikoa eta bere errotazio parametroen determinazioak, bere etorkizuneko detekzioa IAE desberdinentan ahalbidetu dezake. Lortutako datuek eritrulosa 4*C* azukrearen detekzioa izar-arteko detekzioaren ateak irekitzen ditu.

Eritrulosaren errotazio espektroa, unibertsitatean bertan eraikitako Fourierren transformatuan oinarritutako mikro-uhin espektrometroan neurtu zen. Horretarako, etorkizunean lagin oliotsuak prestatzeko erabili daitekeen lagin prestaketa prozesu berria garatu da, non lagina laser ultra-azkarraren pultsuetara ezarri daiteke. Lortutako sentikortasun esperimentalak ^{13}C espezie isotopologiko mono-ordezkatuak abundantzia naturalean detektatzea ahalbidetu du eta datu horien bidez eritrulosa isolatuaren orekako egitura zehatza lortzea ahalbidetu du (horretarako orekako geometria erdi-esperimentalaren metodoa erabili da, r_e^{SE}). Molekulak kate irekiko forma dauka gas egoeran, azukre batean lehen aldiz ikusita eta hiru hidrogeno loturen bidez dago egonkortuta.

Lortutako datu espektroskopiko berriak MADEX kodean sartu dira eta irrati-teleskopioetan datuekin konparatzeko espektroaren tarte egokietan espektro berria simulatu da. Lortutako datuak IAE-ko zenbait guneetako datuekin alderatu dira molekula detektatu nahian: Barnard 1 hodei ilun hotza (B1-b), izarraurreko nukleoa (TMC- 1) eta Sagittarius B2 (Sgr B2) izar sorrera aktiboko eskualde hotzeko datuekin hain zuzen ere. Behaketa astrofisiko eta simulaturiko errotazio espektroen konparaketak emaitza positiborik eman ez duen arren, esperimentalki lortutako datuak etorkizuneko molekula honen eta bestelakoen bilaketa eta identifikazioa sustatu dezake.

III.5 Erreferentziak

[1] E. Herbst, E. F. van Dishoeck, *Annu. Rev. Astron. Astrophys.* **2009**, *47*, 427–480.

[2] P. Ehrenfreund, J. Cami, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002097.

[3] J. M. Hollis, F. J. Lovas, P. R. Jewell, *Astrophys. J.* **2000**, *540*, L107–L110.

[4] B. A. McGuire, C. L. Brogan, T. R. Hunter, A. J. Remijan, G. A. Blake, A. M. Burkhardt, P. B. Carroll, E. F. van Dishoeck, R. T. Garrod, H. Linnartz, C. N. Shingledecker, E. R. Willis, *Astrophys. J.* **2018**, *863*, L35.

[5] E. R. Alonso, B. A. McGuire, L. Kolesniková, P. B. Carroll, I. León, C. L. Brogan, T. R. Hunter, J.-C. Guillemin, J. L. Alonso, *Astrophys. J.* **2019**, *883*, 18.

[6] J. M. Hollis, P. R. Jewell, F. J. Lovas, A. Remijan, H. Møllendal, *Astrophys. J.* **2004**, *610*, L21–L24.

[7] G. Cooper, N. Kimmich, W. Belisle, J. Sarinana, K. Brabham, L. Garrel, *Nature* **2001**, *414*, 879–883.

[8] E. R. Alonso, L. Kolesniková, E. Białkowska-Jaworska, Z. Kisiel, I. León, J.-C. Guillemin, J. L. Alonso, *Astrophys. J.* **2018**, *861*, 70.

[9] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.

[10] A. Lesarri, S. Mata, E. J. Cocinero, S. Blanco, J. C. López, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2002**, *41*, 4673–4676.

[11] E. R. Alonso, I. Peña, C. Cabezas, J. L. Alonso, *J. Phys. Chem. Lett.* **2016**, *7*, 845–850.

[12] I. Peña, E. J. Cocinero, C. Cabezas, A. Lesarri, S. Mata, P. Écija, A. M. Daly, Á. Cimas, C. Bermúdez, F. J. Basterretxea, S. Blanco, J. A. Fernández, J. C. Lõpez, F. Castaño, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2013**, *52*, 11840–11845.

[13] E. J. Cocinero, A. Lesarri, P. Écija, Á. Cimas, B. G. Davis, F. J. Basterretxea, J. A. Fernández, F. Castaño, *J. Am. Chem. Soc.* **2013**, *135*, 2845–2852.

[14] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.

[15] C. Calabrese, I. Uriarte, A. Insausti, M. Vallejo-López, F. J. Basterretxea, S. A. Cochrane, B. G. Davis, F. Corzana, E. J. Cocinero, *ACS Cent. Sci.* **2020**, *6*, 293–303.

[16] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.

[17] C. Cabezas, I. Peña, A. M. Daly, J. L. Alonso, *Chem. Commun.* **2013**, *49*, 10826–10828.

[18] P. Écija, M. Vallejo-López, I. Uriarte, F. J. Basterretxea, A. Lesarri, J. A. Fernández, E. J. Cocinero, *ChemPhysChem* **2016**, *17*, 3030–3034.

[19] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.

[20] J. Vázquez, J. Stanton, in *Equilib. Mol. Struct.* (Eds.: J. Demaison, J.E. Boggs, A.G. Csaszar), CRC Press, **2010**, pp. 53–87.

[21] L. S. Bartell, D. J. Romanesko, T. C. Wong, *Chemical Society Specialist Periodical Report No. 20, Vol. 3*, The Chemical Society, London, **1975**.

[22] S. J. MILLER, in *Probab. Lifesaver* (Eds.: J. Demaison, J.E. Boggs, A.G. Császár), CRC Press, **2019**, pp. 625–635.

[23] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R.

Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09, Revis. D.01, Gaussian, Inc., Wallingford CT* **2009**.

[24] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[25] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[26] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[27] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.

[28] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265–3269.

[29] R. D. Brown, P. D. Godfrey, J. W. V. Storey, M. P. Bassez, *J. Chem. Soc. Chem. Commun.* **1978**, 547–548.

[30] J. L. Alonso, J. C. López, in *Top. Curr. Chem.*, Springer Verlag, **2015**, pp. 335–402.

[31] R. A. Motiyenko, E. A. Alekseev, S. F. Dyubko, F. J. Lovas, *J. Mol. Spectrosc.* **2006**, *240*, 93–101.

[32] A. N. Simonov, L. G. Matvienko, O. P. Pestunova, V. N. Parmon, N. A. Komandrova, V. A. Denisenko, V. E. Vas'kovskii, in *Kinet. Catal.*, Springer, **2007**, pp. 550–555.

[33] M. Smuda, M. A. Glomb, *Angew. Chemie - Int. Ed.* **2013**, *52*, 4887–4891.

[34] J. R. Aviles-Moreno, T. R. Huet, *J. Mol. Struct. THEOCHEM* **2008**, *858*, 113–119.

[35] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.

[36] R. K. Thomas, *Phys. Bull.* **1975**, *26*, 501–501.

[37] W. Zhang, A. S. Serianni, *J. Am. Chem. Soc.* **2012**, *134*, 11511–11524.

[38] R. J. Woods, *Chem. Rev.* **2018**, *118*, 8005–8024.

[39] S. Pérez, T. Tubiana, A. Imberty, M. Baaden, *Glycobiology* **2015**, *25*, 483–491.

[40] I. Cumpstey, *Org. Biomol. Chem.* **2012**, *10*, 2503–2508.

[41] I. Alkorta, P. L. A. Popelier, *Carbohydr. Res.* **2011**, *18*, 2933-2939.

[42] B. Brauer, M. Pincu, V. Buch, I. Bar, J. P. Simons, R. B. Gerber, *J. Phys. Chem. A* **2011**, 115, 5859–5872.

[43] R. A. Jockusch, R. T. Kroemer, F. O. Talbot, J. P. Simons, *J. Phys. Chem. A* **2003**, *107*, 10725–10732.

[44] R. A. Jockusch, F. O. Talbot, J. P. Simons, *Phys. Chem. Chem. Phys.* **2003**, *5*, 1502–1507.

[45] C. S. Barry, E. J. Cocinero, P. Çarçabal, D. P. Gamblin, E. C. Stanca-Kaposta, S. M. Remmert, M. C. Fernández-Alonso, S. Rudić, J. P. Simons, B. G. Davis, *J. Am. Chem. Soc.* **2013**, *135*, 16895–16903.

[46] E. J. Cocinero, D. P. Gamblin, B. G. Davis, J. P. Simons, *J. Am. Chem. Soc.* **2009**, *131*, 11117–11123.

[47] E. J. Cocinero, P. Carcabal, T. D. Vaden, J. P. Simons, B. G. Davis, *Nature* **2011**,

469, 76–80.
[48] P. Çarçabal, E. J. Cocinero, J. P. Simons, *Chem. Sci.* **2013**, *4*, 1830–1836.
[49] A. M. Rijs, J. Oomens, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, **2015**.
[50] E. Mucha, M. Marianski, F. F. Xu, D. A. Thomas, G. Meijer, G. von Helden, P. H. Seeberger, K. Pagel, *Nat. Commun.* **2018**, *9*, 4174–4178.
[51] C. J. Gray, L. G. Migas, P. E. Barran, K. Pagel, P. H. Seeberger, C. E. Eyers, G. J. Boons, N. L. B. Pohl, I. Compagnon, G. Widmalm, S. L. Flitsch, *J. Am. Chem. Soc.* **2019**, *141*, 14463–14479.
[52] E. Saparbaev, V. Aladinskaia, R. Yamaletdinov, A. Y. Pereverzev, O. V. Boyarkin, *J. Phys. Chem. Lett.* **2020**, *0*, 3327–3331.
[53] E. J. Cocinero, P. Çarçabal, in *Gas-Phase IR Spectrosc. Struct. Biol. Mol.* (Eds.: A.M. Rijs, J. Oomens), Springer International Publishing, Cham, **2015**, pp. 299–333.
[54] R. Lozada-Garcia, D. Mu, M. Plazanet, P. Çarçabal, *Phys. Chem. Chem. Phys.* **2016**, *18*, 22100–22107.
[55] N. Vogt, J. Demaison, E. J. Cocinero, P. Écija, A. Lesarri, H. D. Rudolph, J. Vogt, *Phys. Chem. Chem. Phys.* **2016**, *18*, 15555–15563.
[56] N. Vogt, J. Demaison, S. V. Krasnoshchekov, N. F. Stepanov, H. D. Rudolph, *Mol. Phys.* **2017**, 942–951.
[57] M. Vallejo-López, P. Écija, N. Vogt, J. Demaison, A. Lesarri, F. J. Basterretxea, E. J. Cocinero, *Chem. - A Eur. J.* **2017**, *23*, 16412.
[58] N. Vogt, E. G. Atavin, A. N. Rykov, E. V. Popov, L. V. Vilkov, *J. Mol. Struct.* **2009**, *936*, 125–131.
[59] O. V. Dorofeeva, N. Vogt, J. Vogt, M. V. Popik, A. N. Rykov, L. V. Vilkov, *J. Phys. Chem. A* **2007**, *111*, 6434–6442.
[60] N. Marcelino, J. Cernicharo, M. Agúndez, E. Roueff, M. Gerin, J. Martín-Pintado, R. Mauersberger, C. Thum, *Astrophys. J.* **2007**, *665*, L127–L130.
[61] J. Cernicharo, N. Marcelino, E. Roueff, M. Gerin, A. Jiménez-Escobar, G. M. Muñoz Caro, *Astrophys. J. Lett.* **2012**, *759*, L43.
[62] J. L. Neill, M. T. Muckle, D. P. Zaleski, A. L. Steber, B. H. Pate, V. Lattanzi, S. Spezzano, M. C. McCarthy, A. J. Remijan, *Astrophys. J.* **2012**, *755*, 153.
[63] J. Cernicharo, in *EAS Publ. Ser.*, In ECLA-2011: Proceedings Of The European Conference On Laboratory Astrophysics, Laboratory Astrophysics And Astrochemistry In The Herschel/ALMA Era, **2012**, pp. 251–261.
[64] B. A. Mcguire, *Astrophys. J. Suppl. Ser.* **2018**, *239*, 17.

IV. kapitulua

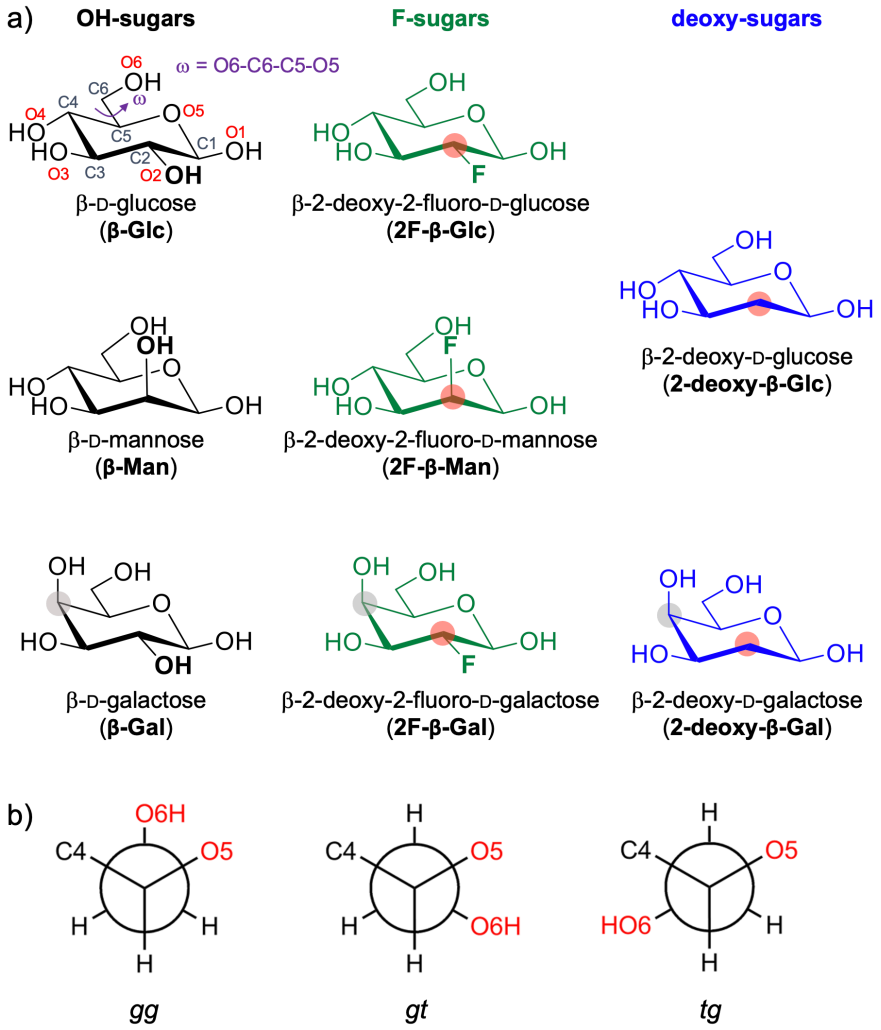
Fluoro eta desoxihexopiranosen konformazio-lehentasunak

IV.1 Sarrera

Kapitulu honetan, esperimentu espektroskopiko (Fourierren transformatuko mikrouhinak eta Erresonantzia Magnetiko Nuklearra (EMN)) eta kalkulu teorikoen (kimika kuantikoan (KK) eta dinamika molekularrean (DM) oinarritutakoak) konbinaketaren bidez, hainbat piranosa monosakaridoren hidroximetilo taldearen konformazio lehentasuna gas eta disoluzio urtsuan egindako ikerketa aurkezten da.

Konformazionalki, efektu exo-anomerikoa eta esterikoaren ondorioz, lotura glikosidikoak neurrizko malgutasuna duten bi tortsio askatasun gradu ditu (Φ/Ψ).[1] Hala ere, (1→6)-*O*-glikosido lotura duten disakaridoek, tortsio angelu gehigarri bat dute (ω = O6–C6–C5–O5, IV. irudia, 1a atala) euren malgutasuna nabarmenki handituz. Noski, malgutasun horrek zelulen arteko interakzioetan zenbait eragin ditu, batez ere efektu entropikoen ondorioz ematen direnak. C5-C6 loturaren biraketak hiru konformazio egonkor posible sortzen ditu, *gg*, *gt* eta *tg* (IV.irudia, 1b atala). D-Glc eta D-Man deribatuetan, O6-a edozein talderekin ordezkatzen bada ere, disoluzio[2–7] eta solido egoeran,[8,9] *gg* eta *gt* hidroximetilo konfigurazioen populazioak ia berdin mantentzen dira eta *tg* konformazioa aldiz ia ez da igertzen. Glc/Man molekula eta euren deribatuetan *gg/gt* konformazioen lehentasuna *gauche* efektuari[10] eta 1,3-diaxial O4/O6 interakzioei egotzi zaie. D-Gal molekula eta deribatuetan aldiz, C5-C6 loturaren konformazio diferentzia nahiko desberdina da, izan ere *gg* konformeroa da ugariena ziurrenik 1,3-diaxial O4/O6 interakzio ez faboratua dela eta.[4] Ur molekulak esplizituki gehitutako kimika kuantikoan eta dinamika molekularretan lortutako emaitzetatik, Wood eta lankideek, urak ω angeluren konfigurazioan efektu zuzena duela determinatu zuten,[11] izan ere Gal monosakaridoan *gg* egituran egonkortzen duen hidrogeno lotura (H-lotura) intramolekular sarearekin interakzionatzen baitu. Hala ere, emaitza oso desberdinak lortu izan dira gas fasean **α-Gal**[12] eta **β-Gal**[13] molekulen artean. α-anomeroan,[12] *gg* hidroximetilo konformazioa (%87) eta erloju-orratzen noranzkoan (*c*) dauden H-loturak nagusi diren bitartean, β-anomeroan, erlojuen norantzaren kontrakoan (*cc*) eta *gt* hidroximetilo konformazioa da ugariena. Horretaz gain, disoluzioan ikusten den **Gal**-ren konformazio lehenetsiak **β-Gal**-ren antz xamarra duen arren, **α-Gal**-k gas fasean duenaren oso desberdina da.[13] Hala eta guztiz ere, monosakarido esanguratsu eta bere deribatuen hidroximetilo errotameroen populazioa (*gg*:*gt*:*tg*) gas fasean eta disoluzio urtsuan aztertzen duen ikerkuntza sistematikorik ez da aurretiaz egin. Kimika kuantikoan tortsio askatasun gradu ugari dituzten sistemetarako aurrerakuntza ikaragarriak egin diren arren,[14–16] bere mugak oraindik nahiko handiak dira. Sistema hauek ikertzeko aukerarik egokiena, azukreak gas fasean analizatzen duen metodo esperimentala erabiltzea da, non inguruneko interferentziarik ez den ematen, eta ondoren emaitza horiek disoluzioan lortutakoekin konparatu. Horretarako, Fourierren transformatuan oinarritutako

mikrouhin espektroskopia, bere paregabeko erresoluzioa dela eta (~3 kHz ≈ 10^{-7} cm^{-1}), molekulen egitura zehatza gas fasean ikusteko aukera ematen duen teknikarik apartenetakoa da. Zehazki, teknika honen bidez, aurretiaz zenbait monosakarido,[12,14,17–22] nukleosido[23] eta azukre deribatuen[24] datu esperimentalak lortu ahal dira. Lortutako egiturak eta konformazioen populazioak, kimika konputazionalaren metodoen independenteak dira; aitzitik, datu esperimentalak kalkulu konputazionalen kalitatea bermatzeko eta kalkuluak kalibratzeko erabili daitezke.



IV.1. irudia Monosakaridoak eta euren nomenklatura. a) lan honetan ikertutako 2-hidroxi, 2-fluorodesoxi, and 2-desoxi hexopiranosak, ω tortsio angeluaren identifikazioarekin. b) ω tortsio angeluaren nomenklatura, *gg* (*gauche-gauche*), *gt* (*gauche-trans*), and *tg* (*trans-gauche*) moduan izendatua. Lehenengo letrak oxigeno endozikliko O5 eta O6 arteko posizioa deskribatzen du eta bigarrenak aldiz C4 eta O6 artekoa.

Lan honetan, zenbait piranosaren hidroximetiloaren konformazioan eragina duten faktoreak ebaluatu eta aztertzeko estrategia esperimental bat diseinatu dugu. Monosakarido jakin batean ematen den H-lotura sare intrintsekoa eta baita azukrearen hidroximetiloan eman daitezkeen alterazioen eraginak (gas fasea vs. disoluzio) ebaluatu dira. Lana, OH eta F taldeen ordezkapen bio-isosterikoan oinarritutako ezagupen molekularreko ikerkuntzetan inspiratua izan da,[25,26] eta baita (gliko)botiken garapenean,[27–32] trazatzaile ez inbaditzaile selektiboetan ^{19}F MR/MRI[33,34] edo positroi emisio tomografian erabilitako trazatzaileetan (^{18}F PET) ere bai.[35–38] Fluor eta hidroxilo taldeak euren artean isoelektronikoak dira eta antzeko polaritatea eta efektu esterikoa dute.[39–41] Hidroxilo taldek duten hidrogeno emaile zein hartzaile izaera kimikoki eraldatua izan daiteke C2 karbonoan OH taldea F edo H-rekin (fluor edo hidrogenoarekin) elkar trukatuz, dagokion desoxifluoro (edo desoxi) monosakaridoa lortzen delarik. Horrela, hexopiranosil monosakarido adierazgarrietatik (D-glukosa (Glc), D-manosa (Man) eta D-galaktosa (Gal)) eratorritako zenbait konposatu sintetizatu dira. F azukrearen kasuan, fluorrak H-lotura ahulak era ditzake eta kontaktu hidrofobikoetan parte har dezake[42,43] baina (bistan da) H-lotura sare intramolekularrean ezin du hidrogeno emaile gisa jokatu! Bestalde, H ordezkapenak erabat deuseztatzen du hidrogenoa emateko edo eskuratzeko gaitasuna H-loturan.

Gas-fasean eta soluzio-fasean dituzten konformazio lehentasunak zehazteko eta horiek kontrolatzen dituzten faktoreak identifikatzeko diziplina anitzeko estrategia zabala diseinatu zen. Erabilitako estrategian, monosakariko fluoratuen gramo eskalako sintesia (kapitulu honetan aipatu gabe), espektroskopia errotazionala gas-fasean, egoera solidoan eta disoluzioan egindako EMN esperimentuak, QM kalkuluak eta dinamika molekular (DM) simulazio estentsiboen konbinaketa erabili ziren.

IV.2 Metodoak

Metodo konputazionalak

α eta β-anomeroen egitura konformazionala, aurretiaz finkatutako estrategia bat erabiliz arakatu zen, mekanika molekularrean (MM) oinarritutako indar-eremu ezberdineko algoritmo sofistikatuetaz baliatuta (MMFFs, OPLS, AMBER). MM-ren bidez lortutako egiturak, ab initio (MP2) eta DFT (B3LYP) metodoak erabiliz optimizatu ziren, Grimme-ren D3 dispertsio-interakzioak eta Becke-Johnson trukaketa potentziala erabiliz, Gaussian 16 programan. Kasu guztietarako, Pople-ren zatitutako balentzia hirukoitza deritzon metodoa erabili zen, difusio eta polarizazio handitze funtzioak atomo guztiei aplikatuta (6-311++G (d,p)). (xehetasunak eta erreferentzia guztiak tesi honen II kapituluan eta eranskinetan daude).

Errotazio espektroskopia gas fasean

Aurretik argitaratutako Glc,[20] Man,[13,44] eta Gal[12,13] molekulen datuak kontuan hartzeaz gain, 2F-β-Glc, 2F-β-Man, 2F-β-Gal, 2-desoxi-β-Glc eta 2-desoxi-β-Gal molekulen datu esperimental berriak unibertsitatean bertan eraikitako cav-FTMW espektrometroan eskuratuak izan ziren.[17,45] Garrantzitsua da azpimarratzea sintetizatutako eta komertzialki eskuratutako laginetan, bakarrik β anomeroa zegoela (IV.2a irudia eta eranskinetako AIV.11–AIV.18 irudiak). Kasu guztietarako, laginak UV erradiazioarekin (355 nm) lurrundu ziren Nd:YAG pikosegundoko laser pultsu ultra-azkarrak erabiliz (~40 ps). Lurruntze-modu hori arrakastatsua izan da azukre desberdinak deskonposatu gabe gas fasera transferitzeko.[17] Lagin hauetarako, lurrunketa prozesu hori, urrats kritikoa da molekulen izaera higroskopiko eta termo ezegonkorra dela eta.[46] (ikus eranskina eta II. kapitulua gas faseko esperimentuen azalpen zehatza lortzeko).

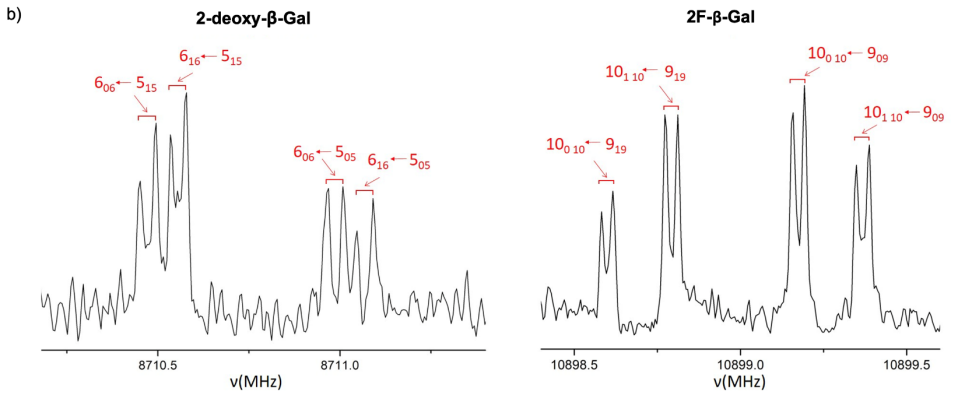
β-D-Hexopiranosen konformazioa disoluzio urtsuan

Hidratazioak azukreen egituran duen efektua ikusteko, euren konformazio analisia ur disoluzioan dinamika molekular simulazioez lagundutako ^{1}H EMN espektroskopia esperimentuen bidez egin zen (akoplamendu homonuklearren konstanteak, $^{3}J_{H,H}$).[47–50] J-akoplamenduaren analisirako (2-desoxi- eta 2-desoxi-2-fluor deribatuetan), EMN espektroak (^{1}H EMN, 1D-selektiboa ^{1}H EMN TOCSY, banda selektiboa ^{1}H-13C HSQC) D_2O disoluzioan 3 mM kontzentrazioan neurtu ziren 800 MHz-ko BRUKER AVANCE III z-gradienteko bobina eta TCI krio-zuntza batez hornitua dagoen espektrometroaren bidez. TopSpin 3.2.7 (BRUKER) softwarea erabili zen datuak eskuratu eta prozesatzeko. Kalkuluak AMBER 20 metodoa erabiliz egin ziren,[51] fluoratuen eta 2-desoxiaren portaera konformazionala aztertzeko, GAFF indar-eremuaren bertsio eguneratua erabili zen.[52]

IV.3 Emaitzak eta eztabaida

β-D-Hexopiranosen konformazio analisia gas fasean

IV.1 irudiak lan honetan sintetizatu eta ikertutako karbohidratoak erakusten ditu. Aipagarria da, fluoro deribatuen lagin solidoak, ^{19}F eta ^{13}C CP/MAS egoera solidoko EMN-n emaitzen arabera, nagusiki β anomeroz osatuta zeudela (IV.2a eta eranskinetako AIV.11–AIV.16 irudiak).[53–58] Era berean, komertzialki eskuratutako desoxi-D-glukosa (**2-desoxi-Glc**) eta 2-desoxi-D-galaktosari egindako (**2-desoxi-Gal**) ^{13}C CP/MAS EMN esperimentuetan, β anomeroa bakarrik aurkitu zen, (eranskinetako AIV.17 eta AIV-18. irudiak).



IV.2. irudia Datu esperimentalak. **2-desoxi-β-Gal** (ezkerra) and **2F-β-Gal** (eskuina) hexopiranosen errotazio trantsizio tipikoak, Fourier transformatuan oinarritutako mikrouhin espektrometroan (cav-FTMW) eskuratutakoak. Errotazio trantsizioak banatuta daude Doppler efektuaren ondorioz.

Glc, **Man** eta **Gal**-ren errotazio espektroak gas-fasean aurretiaz eginda daude. **Glc**-ren kasuan bi anomeroak detektatu ziren bitartean,[20] **Man**[59] eta **Gal**-ren[12] kasuan, α anomeroa soilik detektatu ahal izan zen. Garrantzitsua da aipatzea, gas fasean errotamutazio prozesua (α↔β anomeroen arteko bihurketa) eragotzita dagoela.[14] Beraz, α edo β anomeroen populazioak gas fasean, lagin solidoan dagoen populazio bera islatzen du. Hortaz, **β-Man** eta **β-Gal**-ren errotazio datu esperimentalen faltaren ondorioz, lan honetan euren konformazio/populazioen datuak *Simons* eta lankideek **β-Man-OPh**[44] eta **β-Gal-OPh**[13] molekuletan gas fasean egindako "*ion-dip*" deritzon laser espektroskopia esperimentuen emaitzetatik lortu ziren. Molekula hauetarako lortutako emaitzak konputazionalki lortutako datuekin bat datoz (IV.2 taula).

IV.1. taula. Aztertutako molekulen konformeroen errotazio parametro esperimental garrantzitsuenak. Datu guztiak eranskinetan aurki daitezke (AIV.15 - AIV.21 tauletan).

		A[a]	B	C
2F-β-Glc	**1** (*gt*/*cc*)	1174,97088(22) [b]	804,19900(29)	528,30649(47)
	2 (*gg*/*cc*)	1171,00499(44)	779,36429(19)	489,58154(18)
2F-β-Gal	**2** (*gt*/*cc*)	1269,18963(19)	755,20420(13)	524,177553(88)
	3 (*gg*/*c*)	1289,06503(18)	796,27051(12)	563,133262(85)
2F-β-Man	**1** (*gt*/*cc*)	1080,6644(11)	898,87960(43)	593,39487(22)
	2 (*gg*/*cc*)	1117,2134(99)	848,8623(69)	531,96736(55)
2-desoxi-β-Glc	**1** (gg/c)	1189,644727(86)	1074,782499(81)	632,855274(49)
	4 (*gt*/*cc*)	1239,26352(26)	1005,92933(15)	580,78082(13)
	6 (*tg*/*c*)	1362,98879(17)	950,247495(79)	586,510948(66)
2-desoxi-β-Gal	**1** (*gg*/*c*)	1293,38327(15)	1065,40506(11)	685,853728(73)
	3 (*gt*/*c*)	1381,23062(48)	935,04220(16)	633,71769(10)

[a] Errotazio konstante esperimentalak A, B eta C MHz-tan.
[b] Desbideratze estandarra (DE) parentesi artean adierazita dago, azken digituaren unitatetan.

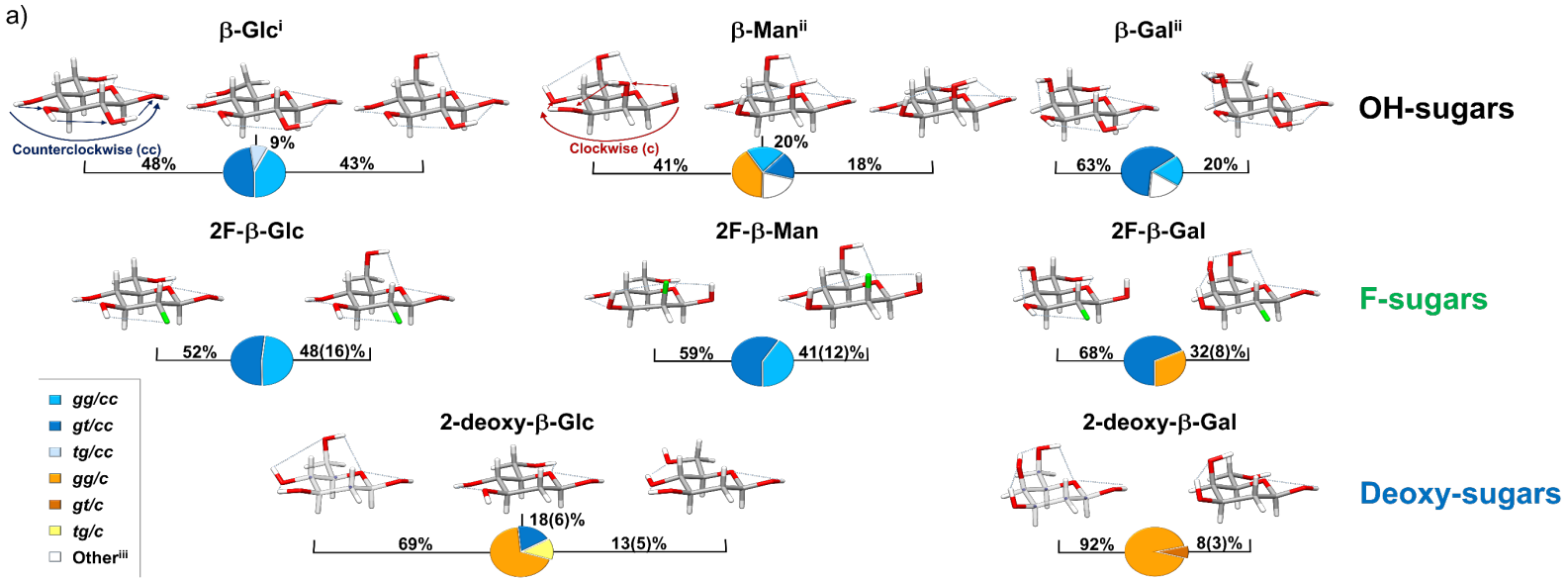
Gas faseko aurreko lanen laburpen gisa, eta teknika esperimentalen menpekotasuna alde batera utzita (ikus IV.2 taula gas-faseko populazioaren datuen multzo osoa ikusteko), α-anomeroentzat erloju-orratzen noranzko H-lotura sareak du lehentasuna eta gehienetan hidroximetil albo katean *gg* errotameroaren nagusitasunarekin erlazionatuta dago. Aldiz β-deribatuetan, erloju-orratzen kontrako noranzko H-lotura sarea ikusten da eta *gt* errotameroaren presentziari lotuta dago. Kasu guztietarako, *tg* konformazioaren populazioa baztergarria da. **Glc** eta **Man** deribatuen datu horiek, disoluzioan[2,3,7] zein solido egoeran[9] lortutako balioekin bat datoz. Hala ere, kontuan izan behar da disoluzioan ahalbidetua dagoen errotamutazioaren ondorioz, α eta β anomeroen populazioak disoluzioan ezin direla bereizi. Aitzitik sarreran esan den moduan, **Gal**-ren hidroximetil albo-katearen portaerak disoluzioan (*gt*/*tg* oreka) bere **β-Gal-OPh**[13] deribatuan (*gt*/*tg* oreka cc H-lotura sarearekin) ikusitakoaren antza du, baina oso desberdina da gas-fasean dagoen **α-Gal**-rekin konparatuta,[12] non *gg* konformeroa den nagusi eta H-lotura sarea *c* norantzan dagoen.

Lagin bakoitzeko, kalkulu konputazionalen emaitzak lagunduta, hainbat frekuentzia ekorketa egin ziren errotazio espektrometroarekin (eranskinetako AIV.1–AIV.14 taulak eta AIV.19–AIV.15 irudiak). Ikertutako konposatu guztien artean, guztira hamaika egitura (konformero) identifikatu ziren. Zehazki, hiru konformero detektatu ziren **2-desoxi-β-Glc**-rako eta bi konformero gainontzeko azukre bakoitzeko (**2F-β-Glc**, **2F-β-Man**, **2F-β-Gal** eta **2-desoxi-β-Gal**). Konformero bakoitzerako detektatutako errotazio trantsizio esperimental multzoa Watson-en *S* Hamiltondar erdi-zurrun murriztuaren bidez doitu zen I^r deritzon adierazpidean (edo errepresentazioan), errotazio konstanteak lortzeko. (IV.1 taula eta eranskinetako AIV.15– AIV.21 taulak). Errotazio konstante horiek, molekularen hiru dimentsioko egituraren hatz-markak dira. Datu esperimentalen

eta teorikoen arteko konparaketak, zalantza gabeko konformeroen esleitzea ahalbidetu zuen. Identifikatu ziren konformero guztiak IV.3a irudian irudikatu dira, kalkulatutako energiaren arabera zenbakien bidez izendatuak: **2F-β-Glc-1**, **2F-β-Glc-2**; **2F-β-Man-1**, **2F-β-Man-2**; **2F-β-Gal-2**, **2F-β-Gal-3**; **2-desoxi-β-Glc-1**, **2-desoxi-β-Glc-4**, **2-desoxi-β-Glc-6** eta **2-desoxi-β-Gal-1**, **2-desoxi-β-Gal-3**.

Gainera, gure cav-FTMW espektrometroaren sentikortasun handiak ahalbidetuta, ugaritasun naturalean zeuden zenbait ^{13}C isotopomero ere detektatu ziren (~1.1%). Zehazki, **2-desoxi-β-Gal** eta **2-desoxi-β-Glc** konformero ugarienen ($^{13}CC_5H_{12}O_5$) bederatzi espezie isotopologo mono-ordezkatu gehigarrien errotazio espektroak doitu ziren (**2-desoxi-β-Gal**-ren 6 ^{13}C espezie mono-ordezkatu, eranskinetan AIV.20. eta AIV34–AIV39. taulak; eta 2-desoxi-β-Glc-ren 3 ^{13}C espezie, eranskinetan AIV.16. eta AIV.25–AIV.27. taulak). Isotopologo horien errotazio parametroek, r_0 eta r_S egiturak zehaztea ahalbidetu zuten (xehetasunak II. kapituluan, parametroak eranskinetako AIV.30 taulan). Aurretik aipatu den bezala, metodo horien bidez lortutako egiturak eredu konputazionalekiko independenteak dira. Kasu batzuetan, kimika kuantikoaren bidez lortutako emaitzak, erabilitako metodoaren menpekoak dira eta, beraz, ez dira guztiz erabakigarriak.[14,60] Aitzitik, mikrouhin espektroskopiako datuak kimika konputazionalean lortutako emaitzetarako erreferentzia tresna direla frogatu da.[60]

2F-β-Glc, **2F-β-Man**, **2F-β-Gal**, **2-desoxi-β-Glc** eta **2-desoxi-β-Gal**-ren trantsizioen intentsitate erlatiboen azterketatik, 3a irudian agertzen diren populazio esperimentalak kalkulatu ziren.[61] Azukre naturalei dagozkien datuak (**β-Glc**, **β-Man** eta **β-Gal**) gehitu dira konparazioa errazteko. Nahiz eta nahiko populazio desberdina eduki, hidroximetilo taldearen hiru konformazio txandakatu posibleak (*gg*, *gt* eta *tg*) behatu ziren ikertutako azukreetan. Kasu guztietan, egitura molekularrak H-lotura sare kooperatiboen bidez egonkortzen dira, erlojuaren orratzen (*c*) edo erlojuaren orratzen kontrako (*cc*) norabidean antolatu daitezkeenak. Populazio esperimentalak IV.3 irudian eta IV.2 taulan aurkezten dira.



IV.3. irudia Gas fasean behatutako monosakaridoen populazioak eta egiturak. a) OH, F eta desoxi azukreen egitura eta populazio guztiak. Monosakarido guztiak H-lotura sare kooperatiboek egonkortzen dituzte. H-lotura erloju orratzen kontrako noranzkoa (*cc*) duen Glc egitura bat (gezi urdina), eta kontrako noranzkoa, erloju orratzen orientazio berdina (*c*) duen H-lotura sarea aldiz Man molekulan irudikatu da (gezi gorria). Oharrak: [i] Aurreko laneko balio esperimentalak.[20] [ii] Lortutako egiturak eta populazioak MP2/6-311+G(d,p) maila teorikoan. Datu teoriko horiek koherenteak dira *"ion-dip"* deritzon espektroskopian lortutako balio esperimentalekin **β-Man-OPh**[44] eta **β-Gal-OPh**[13] deribatuentzat (IV.2 taula). [iii] Konformazioak, % 10 baino gutxiagoko populazioa teorikoa duten datu konformeroen egiturak, ez dira irudikatu baina balioak IIV.2 taulan ikusi daitezke. Haien balioak *'other'* bezala taldekatu dira eta gainontzeko konformero guztien ekarpena batzen du.

IV.2. taula Esperimentalki eta teorikoki, gas fasean hexopiranosei dagokien konformero bakoitzaren populazio ratioa. Desbideratze estandarra (DE) parentesi artean azken dezimalerako erakusten da.

Molekula \ Konformeroa	*gg/cc*	*gt/cc*	*tg/cc*	*gg/c*	*gt/c*	*tg/c*
α-Glc[20]	36	32	18	14	-	-
β-Glc[20]	43	48	9	-	-	-
β-Glc-OPh[62], [a]	25	68	7	-	-	-
2F-β-Glc Lan hau	48(17)	52	-	-	-	-
2-desoxi-β-Glc[b], Lan hau	-	18(6)	-	69	-	13(5)
α-Man[c], Lan hau	11	6	2	66	4	10
α-Man-OPh[13], [a]	10	5	-	85	-	-
β-Man[c], Lan hau	20	18	8	41[d]	9	4
β-Man-OPh[44], [a]	30	70	-	-[d]	-	-
2F-β-Man Lan hau	41	59(8)	-	-	-	-
2-desoxi-β-Man[b], Lan hau	-	18(6)	-	69	-	13(5)
α-Gal[12]	2	9	4	85	-	-
α-Gal-OPh[63], [a]	-	63	17	20[d]	-	-
β-Gal[c], Lan hau	9	63	20	7	<1	<1
β-Gal-OPh[13]	-	75	20	5	-	-
2F-β-Gal Lan hau	-	68	-	32(8)	-	-
2-desoxi-β-Gal Lan hau	-	-	-	92	-	8(3)

[a] OPh deribatuen balio esperimentalak laser espektroskopia datuetatik eratorriak. Populazioaren erizpena R2PI seinaleen intentsitateetatik lortu da. Balio hauen prezisio nahiko txikia den arren, balio hauek konformeroen populazioen balio orokorrak erakusten dituzte.
[b] **2-desoxi-β-Glc** eta **2-desoxi-β-Man** molekula berdina dira.
[c] MP2/6-311++G(d,p) mailan lortutako balio teorikoak.
[d] *gg/c* konformeroaren falta **β-Man-OPh** eta bere ugaritasun txikia **α-Gal-OPh** analogoan, O1 posizioaren hidrogeno faltaren ondorioa da. O1H···O2 interakzioa ezinbestekoa da *cc* H-lotura egonkortzeko. Horren ondorioz **β-Man** and **α-Gal** azukreetan *c* motako H-lotura duten konformeroak dira nagusi, %41 eta %85eko populazioarekin hurrenez hurren.

Jatorrizko azukreen kasua

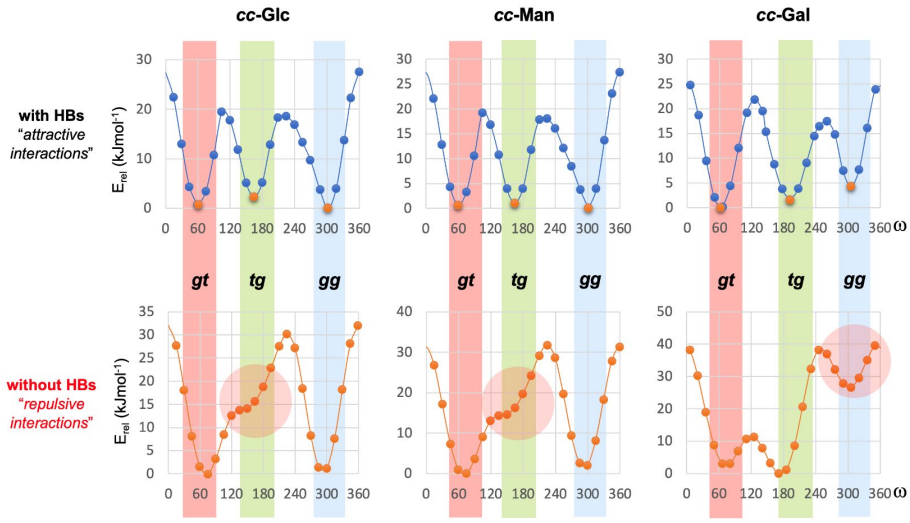
Jatorrizko azukreen β-anomeroen datu esperimental eta teorikoen multzoak, hiru jokabide desberdin eta ondo zehaztuta daudela adierazten du; **β-Gal**-rentzat espezie nagusi bat (*gt/cc*) ikusten/iragartzen den bitartean, **β-Glc** eta **β-Man**-rentzat hainbat konformero batera ikusi/iragarri dira. **β-Glc**-ko konformero guztiek, *cc* H-lotura erakusten zuten bitartean, *c* motako H-lotura sarea **β-Man**-ren konformazio nagusia zen (*gg/c*: %41), nahiz eta *cc* H-lotura duten konformeroen populazioa nahiko handia den ere (%46).

Hein batean harritzekoa da, **β-Glc**-ren *tg* rotameroaren populazioa gas fasean oso txikia dela eta,[20] funtsean, arbuiagarria **β-Man**-ren kasuan, nahiz eta

gg eta *gt* konformeroekin konparatuta, *tg* egitura hidrogeno lotura sendoago batez egonkortua dagoen (O6H···O4 vs. O6H···O5). O6H···O4 H-loturaren indar handiagoa NBO kalkuluen bidez egiaztatu zen (AIV.42–AIV.44 taulak eranskinetan). Gainera, **α-Gal**-rako aurretik esperimentalki lortutako emaitzetan ez bezala, **β-Gal**-rentzat *gg* konformeroaren populazioa ia arbuiagarria da, eta horren ordez O6H···O5 H-lotura duen *gt* konformeroa nagusi da. Aurkikuntza baliotsu hau azaltzeko, ω tortsio angeluaren inguruko energia ekorketa sistematikoa egin zen (IV.4 irudia eta eranskina, AIV.32–AIV.38 irudiak), OH6 eta O5/O4 arteko H-lotura egonkortzailea ahalbidetuz eta galaraziz. Energia ekorketetan, hiru errotameroak (*gg*, *gt* eta *tg*) energetikoki baliokideak izatetik urrun daudela ikusten da. **β-Glc**-ren kasuan, *tg* konformeroak energia penalizazio "errepultsibo" handia erakusten du (~15 kJ·mol^{-1}) eta nahiz eta geometria honetan hidrogeno lotura sendoa sor daitekeen (O6H···O4), lotura hori ez da nahiko energetikoa penalizazioa pairatzeko eta *tg* konformazioa egonkortzeko *gg* eta *gt*-ren aurrean. Gainera, H-lotura hori eragozten bada, *tg* konformeroa ez da ezta minimo lokala ere (IV.4. irudia). Jokabide bera ikusten da **β-Man**-rako *cc* orientazioan. Era berean, **β-Gal**-en *gg* konformeroa gas fasean oso desegonkortuta dago, eta O6H···O4 eragozten denean (*cc* H-lotura eredua) *tg* eta *gt* errotameroekin konparatuta penalizazio energetikoa oso handia da (>20 kJ·mol^{-1}) (IV.4. irudiak eta eranskinak, AIV.42-AIV.48 irudiak). Jakina, O4H-ren antolamendu axiala dela eta, O4H···O6 H-lotura **Gal**-an **Glc**/**Man**-an baino askoz ere indartsuagoa da.

Azukre desoxifluoratuen kasua

Ondoren, F-azukreen datu esperimentalen analisia egin zen. Populazio esperimentalen arabera, bakarrik *gg* (% 32-48) eta *gt* (% 52-68) hidroximetiloaren antolamendua duten konformeroak ageri dira, *cc* H-lotura norabidea nagusi izanik. Are gehiago, *tg* egituraren presentzia arbuiagarria da kasu guztietan. Fluorazioaren ondorioz, H-lotura sare kooperatiboak ezegonkortzen dira, batez ere *c* H-loturaren noranzkoan. Izan ere, fluor atomoak ezin du hidrogeno emaile gisa jokatu, baina bai hidrogeno hartzaile gisa, zein erlojuaren orratzen kontrako orientaziorako bakarrik eman daitekeen. Hau bereziki nabaria da **2F-β-Man** eta **2F-β-Glc**-ren kasuetan, non erloju-orratzen norantzan bideratutako egituren arrastorik ez dagoen. C2-ko fluor atomoaren orientazio axialak (**Man**) edo ekuatorialak (**Glc**) badirudi ez duela garrantzirik lortutako emaitzetan. **2F-β-Gal**-rako bakarrik *c* H-loturaren populazio apur bat (%32) dago (*gg* errotameroari dagokiona), baina C2-ko fluor atomoak bakarrik H-lotura hartzaile ahul gisa joka dezakenez gero, *cc* alternatiba ugariagoa da. Hori dela eta, *gt/cc* geometria neurri handi batean egonkortzen da (%68).



IV.4. irudia *cc*-β-Glc, ***cc*-β-Man**, eta ***cc*-β-Gal** molekulen ω angeluaren interkonbertsio energia langa MP2/6-311++G(d,p) maila teorikoan. ω angeluan H-lotura ahalbidetuz egindako ekorketa ("*attractive interactions*"). H-lotura galaraziz egindako ω angeluaren ekorketa energetikoan ("*repulsive interactions*"), H–O(6)–C(6)–C(5) angelua 180º-tan finkatu zen, horrela hidrogeno emaile bezala jokatzeko aukera galaraziz

Desoxi analogoen kasua

Desoxigenazioak, C2-ko H emaile/hartzaile izaera deuseztatzen du. Horren ondorioz, **2-desoxi-Glc/Man** (%69) eta **2-desoxi-Gal** (%92) azukreetan, *c* H-lotura sareak dituzten *gg* errotameroak dira nagusi.

2-desoxi-Glc/Man-an, H-lotura sarearen eraginkortasuna nahiko kaskarra da, eta kasu horretan, *tg* errotameroaren populazio txiki bat (%13) agertzen da. Hidroximetil albokateko *tg* konformeroa *c* edo *cc* H-lotura sarearen bidez egonkortu daitekeen arren, *tg*-ren populazioa *c* H-lotura noranzkoarekin erlazionatuago dago. Horrela, O3eqH···O4eq/O4eqH···O6 H-lotura *c* motako H-lotura sarean *cc* H-loturako (O6H···O4eq/O4eqH···O3eq) sarean baino sendoagoa da, baina hala ere ez da nahikoa O4/O6 elkarrekintza ezegonkortzeko. Jokabide hori, geometria horretarako 1,3-interakzio-diaxial kaltegarriaren bidez azal liteke. Beraz, *gg* (%69) eta *c* (%82) orientazioak nagusitzen dira. Aipatutako *c* H-lotura sareak, oxigeno endoziklikoa (O5) inplikatzen duen H-lotura sare luzeagoak sortzeko gai izateak hauen alde egiten du.

β-Gal deribatuen kasuak arreta berezia merezi du. **Gal** deribatu guztietarako, *gg* konformeroaren agerpena *c* motako H-lotura sarearekin erlazionatuta dago. Izan ere, arestian esan bezala, **α-Gal**-an *gg* konformero ugariena da eta konformero hori *c* noranzkoa duten H-lotura sareak egonkortzen

du.[12] Jakina, C2-ko hidroxilo taldea kentzeak bi H-lotura norabideak desegonkortzen ditu. **2-desoxi-β-Gal**-rentzat, O3eqH···O4ax/O4axH···O6 H-lotura *c* noranzkoan egonkortzen da (% 100 inguru), cc alternatibarekin konparatuta (O6H···O4ax/O4axH···O3eq) eta *gg* rotameroaren populazioa % 92ra igotzen da.

Laburbilduz, fluor eta desoxi analogoen datuen arteko konparaketak, fluor atomoak H-lotura hartzaile ahul gisa duen funtsezko eginkizuna ebaluatzeko aukera ematen du. *cc* H-lotura sarea, fluor atomoaren presentzia dela eta, neurri handi batean egonkortuta dago fluoro deribatuetarako desoxi analogoekin alderatuta.

C1-eko ordezkapenak H-lotura sarean duen garrantzia ere nabarmengarria da. *c* noranzko H-lotura sare kooperatiboan, OH1 taldeak hidrogeno emaile gisa jarduten du. Izan ere, OH1 hori ez duten **β-Glc-OPh**, **β-Man-OPh** eta **β-Gal-OPh**[13] molekuletan, *cc* H-lotura duten konformeroak bakarrik identifikatu ziren. **β-Glc**[20] eta **β-Gal** azukreentzat aldiz, *cc* H-lotura kooperatiboa duten konformeroak dira nagusi eta *c* H-lotura dutenen populazioa aldiz arbuiagarria da. Azpimarra daiteke, HO1···O2 H-loturan C1-O1 loturaren orientazioak, efektu exo-anomeriko ez faboratua eratzen duela. Beraz, *cc* H-lotura sarean baino ez da ematen. **β-Man** bakarrik *c* motako konformeroen populazio txiki bat dago. Kasu horretan, O2-ren orientazio axiala dela eta, C1-O1 loturak ez du efektu exo-anomeriko ezegonkortze posizioa hartzen. Aitzitik, **α-Gal** edo **α-Glc**-rentzat, O1-ren orientazio axialak O1H-k O2-rekin H emaile gisa jardutea ahalbidetzen du orientazio exo-anomerikoan (anti). **α-Gal**-rako esperimentalki frogatu denez, *gg* hidroximetilo orientazio eta erlojuaren orratzen noranzkoan zuzendutako sarea da egonkorrena (%85eko populazioarekin).

Konformazioa disoluzio urtsuan

Espero zen moduan, datuen analisiak erakutsi zuen deribatu guztiek disoluzio urtsuan $^4C_1$ aulki konformazioa dutela (eranskinak, AIV.39. irudia).[64] $^3J_{H5,H6}$ akoplamendu konstante esperimentalen analisiaren arabera,[64,65] hidroximetiloa albo kate malgua den arren *gt* eta *gg* errotameroak lehentasuna dute **β-Glc** eta **β-Man** deribatuetarako C2-n dagoen taldea zeinahi ere den, –CHOH–, –CHF- edo CH2–. (IV.5a eta IV.5b irudia). Izan ere, $^3J_{H5,H6}$ balio esperimentalen desberdintasuna azukre desoxigenatu edo fluoratuetan oso txikia da (IV.3 taula).

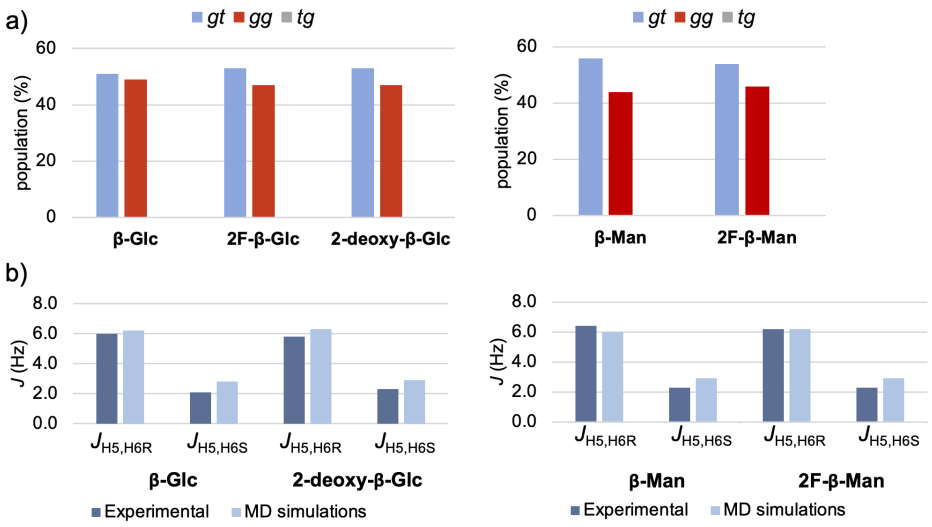
IV.3. taula $^3J_{H5,H6}$ akoplamendu konstanteak disolusio urtsuan EMN teknikaren bidez determinatuak lan honetan (xehetasun experimentalak eranskinetan) .

Hexopiranosa	anomeroa	$J_{H5,H6R}$ (Hz)	$J_{H5,H6S}$ (Hz)
2-desoxi-2-fluoro-D-Glc*p*	α	5,3[a]	2,3[a]
(**2F-Glc**)	β	6,1[a]	2,3[a]
2-desoxi-2-fluoro-D-Man*p*	α	5,4;[b] 5,7[a]	2,2;[b] 2,1[a]
(**2F-Man**)	β	6,2;[b] 6,7[a]	2,3;[b] 2,3[a]
2-desoxi-2-fluoro-D-Gal*p*	α	8,3[b]	4,1[b]
(**2F-Gal**)	β	7,8[b]	4,5[b]
2-desoxi-D-Glc*p*	α	5,1[b]	2,3[b]
(**2-desoxi-Glc/Man**)	β	6,1[b]	2,3[b]
2-desoxi-D-Gal*p*	α	6,2[b]	6,2[b]
(**2-desoxi-Gal**)	β	7,8;[b] 7,3[c,d]	4,6;[b] 5,5[c,d]
D-Glc*p* (**Glc**)	α	5,8[c,e,f]	2,0;[c] 1,0;[e] 1,9[f]
	β	5,9;[c] 6,0[e,f]	2,0;[c] 2,1;[e] 1,9[f]
D-Man*p* (**Man**)	α	5,8;[g] 5,6[h]	2,0[g,h]
	β	6,4;[g] 5,7[h]	2,3[g,h]
D-Gal*p* (**Gal**)	α	6,0;[b] 6,8;[c] 7,9[f]	6,0;[c] 4,6[f]
	β	7,8[f]	4,6[f]

[a] Err. [66]; [b] Lan hau. 800 MHz neurtuta (Eranskinetan xehetasuna). ABX, AX2 espin sistemak hurrengo gunean doitu ziren: (http://anorganik.uniuebingen.de/klaus/nmr/spinsystems/index.php?p=abx_ab_an) [c] Err. [50]. [d] Itxurazko akoplamendu konstanteak. [e] Err. [3]. [f] Err. [67]. [g] Err. [7]. [h] Err.[68].

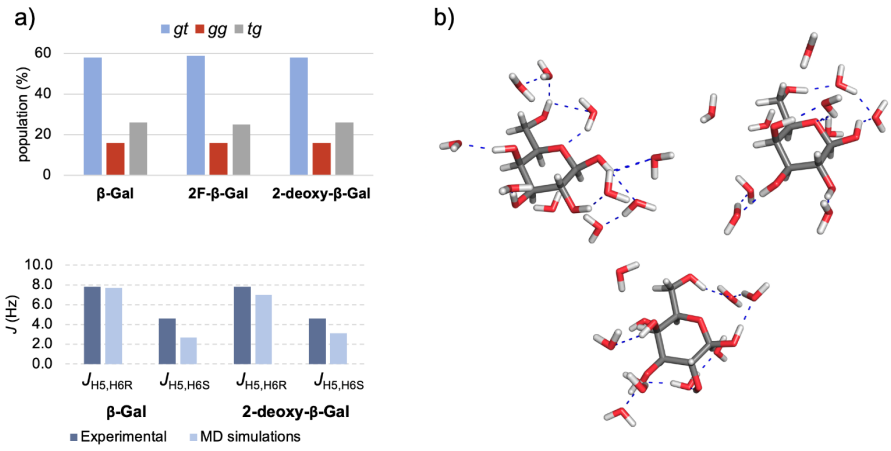
Erabilitako Karplus motako ekuazioaren arabera,[64] **β-Glc**, **2F-β-Glc**, eta **2-desoxi-β-Glc**-ren $^3J_{H5,H6}$ akoplamendu konstante esperimentalak, hidroximetilo taldearentzako (*gg*:*gt*:*tg*) errotameroaren ≈ 55:45:0 banaketari dagozkio. Balio hauek **β-Man** eta **2F-β-Man** molekulen datuen antzekoak dira. α-anomeroetan joera bera ikusi da (eranskinak, AIV.41 irudiak) eta datu esperimentalak bat datoz DM simulazioen emaitzekin (IV.5. irudia eta eranskinak, AIV41–AIV.43. irudiak). Beraz, disoluzio urtsuan, ikertutako C2-ko ordezkatzaile guztietarako, **Man** eta **Glc** deribatuek 50:50 *gg/gt* konformazio banaketa erakusten dute. Kasu guztietan, datu esperimentalek adierazten dute *tg* konformeroaren populazioa, gas fasean gertatzen den moduan, ez dela adierazgarria disoluzioan. DM-ko datu guztien

analisiek erakutsi zuten uretan ez dela monosakaridoetan hidrogeno lotura intramolekular sendorik eratzen, populazioak <%10 dira kasu guztietarako, inguruko disolbatzaile ur molekulen eragina dela eta.



IV.5. irudia β-Glc and β-Man deribatuen azterketa konformazionala uretan. a) EMN datuetatik deribatutako ω-tortsio angeluaren (O5-C5-C6-O6) banaketa **β-Glc** and **β-Man** deribatuetarako. (IV.3. taula). b) $^{3}J_{H5,H6}$ akoplamendu konstante esperimentalak (^{1}H EMN-ren bidez lortuta) eta predikzioak **β-Glc** and **β-Man** deribatuentzat.

β-Gal deribatuetarako, $^{3}J_{H5,H6}$ balioak bat datoz *gt* eta *tg* konformeroen presentziarekin[11,64,69] eta *gg* konformeroaren populazio txikiagoarekin. Balio horiek bat datoz gasaren fasean **β-Gal** deribatuetarako lortutako balioekin, baina **α-Gal**-rako lortutako balioen kontrakoak dira.[12] Batez ere, disoluzio fasean egindako GAFF indar-eremuarekin lotutako DM simulazioek, *gg* konformeroaren populazioa handiegia iradokitzen dute, bereziki **2F-Gal** eta **2-desoxi-Gal** deribatuetarako, nahiz eta α- edo β-anomeroa izan. Hala ere, GLYCAM-en azken bertsioa[70] erabiliz egin ziren DM simulazioak **α-Gal** and **β-Gal** molekulentzako, esperimentalki lortutako balioen hurbil dauden emaitzak eman zituzten (eranskinak, AIV.40 – AIV.43 irudiak). **Glc** eta **Man** simulazioetan berriz, ez zen aparteko H-lotura intramolekularrik antzeman. Beraz, hidroxilozko talde guztiek, lehenengo hidratazio geruzan dauden ur molekulekin H-loturak eratzen dituzte (IV.6b irudia). Ondorioz, disoluzio urtsuan, C5-C6 loturan H-lotura intramolekularraren sarean eragina oso txikia da.



IV.6. irudia β-Gal deribatuen azterketa konformazionala ur disoluzioan. a) ω-tortsio angeluaren (O5-C5-C6-O6) banaketa **β-Gal** deribatuetarako, datu esperimentaletatik lortutako emaitzak (IV.3 taula), datu esperimentalekin batera (^{1}H EMN-ren bidez lortutakoak), eta aztertutako konposatuen $^3J_{H5,H6}$ akoplamendu konstanteak (behean). b) Simulazioetatik lortutako **β-Gal**-ren egitura adierazgarrienek ω-tortsio angeluan *gt* konformazioa dute.

IV.4 Ondorioak

Kapitulu honetan lortutako emaitzen bidez gas fasean eta soluzio urtsuan **β-Glc** eta **β-Man** eta euren C2-ko deribatu fluoratu eta desoxidatuetan *gg* eta *gt* errotameroak nagusi direla frogatzen dugu. Aldiz, *tg* konformeroaren populazioa oso txikia da. Horren arabera iradoki daiteke, ur molekulek hidroxilo taldearen konformazioan eragin txikia dutela. **Gal** deribatuetan aldiz, beste izaera bat antzeman zen. **β-Gal**-erako *gg* errotameroen populazioa gas fasean zein disoluzio urtsuan oso txikia den bitartean, *gg* rotameroaren populazioa 2-fluoro eta 2-desoxi galaktosa analogoetan nahiko handia da. **Gal**-ren *gg* konformeroentzat, O4H···O6 H-lotura O6H···4 H-lotura baina sendoagoa da, izan ere O4H···6 lotura O4 posizio axialaren bidez faboraturik baitago. Antzeko tendentzia ondorioztatu daiteke fluoro eta deoxi deribatuetan.

Hala ere, *gg* errotamero hau desagertu egiten da ur disoluzioan eta, aldi berean, *gg* geometria faboratzen duen *c* H-lotura intramolekularra desagertzen da. Gas fasean **α-Gal**-an *gg* konformeroa ere da nagusi (%85), non *c* norabidean H-lotura sare intra-molekular kooperatiboa dagoen. Egoera hau, ur disoluzioan trukatu egiten da. Baldintza solbatatuaren eraginaren ondorioz, C5–C6 loturaren konformazioa **Gal** deribatuetan, *gt/tg* konformazio oreka gisa deskriba daiteke, *gg* geometriaren presentzia oso txikiarekin.

Beraz, hidroximetiloaren konfomazioa ikertutako monosakaridoetan zenbait faktoreren menpe dago: zentro estereogenikoen konfigurazioa, hidroximetilo taldea H-lotura kooperatiboa eratzeko duen gaitasunarekin erlazionatuta dago. Hau bereziki garrantzitsua da karbono anomerikoan. C4-ko hidroxil taldearen kokapenak egonkortze (O4H···O6 H-lotura) eta ezegonkortze (O4/O6 esterikoa) interakzioak eragiten ditu. Egonkortzeko edo desegonkortzeko indarren nagusitasuna H-loturaren noranzkoaren (*c* edo *cc*) araberakoa da, zein C2-ko ordezkatzailearen izaera kimikoaren eta orientazioaren arabera moldatu daitekeen. Urak H-lotura sarea eragozten du eta C5 – C6 elkarreragin esterikoen araberakoa da gehienbat.

Horrela, aurkeztutako lanak frogatzen du nola OH eta F/H-ren arteko ordezkapenak azukreen izaera, H-loturen kooperazioa, hidratazioa eta propietate dinamikoak aldatu ditzakeen. Horregatik, eraldaketa kimiko sinple horiek eragina dute azukreen propietatetan eta noski helburu biologiko konkretuetara begira ere, beren propietate molekularretan eta azken finean beren aktibitate biologikoan dagokion efektua dutelarik.

IV.5 Erreferentziak

[1] A. García-Herrero, E. Montero, J. L. Muñoz, J. F. Espinosa, A. Vian, J. L. García, J. L. Asensio, F. J. Cañada, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **2002**, *124*, 4804–4810.

[2] H. Ohrui, Y. Nishida, M. Watanabe, H. Hori, H. Meguro, *Tetrahedron Lett.* **1985**, *26*, 3251–3254.

[3] Y. Nishida, H. Ohrui, H. Meguro, *Tetrahedron Lett.* **1984**, *25*, 1575–1578.

[4] K. Bock, J. Duus, *J. Carbohydr. Chem.* **1994**, *13*, 513–543.

[5] C. Nóbrega, J. T. Vázquez, *Tetrahedron Asymmetry* **2003**, *14*, 2793–2801.

[6] G. D. Rockwell, T. B. Grindley, *J. Am. Chem. Soc.* **1998**, *120*, 10953–10963.

[7] H. Hori, Y. Nishida, H. Ohrui, H. Meguro, *J. Carbohydr. Chem.* **1990**, *9*, 601–618.

[8] G. A. Jeffrey, R. K. McMullan, S. Takagi, *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **1977**, *33*, 728–737.

[9] R. H. Marchessault, S. Perez, *Biopolymers* **1979**, *18*, 2369–2374.

[10] S. Wolfe, *Acc. Chem. Res.* **1972**, *5*, 102–111.

[11] K. N. Kirschner, R. J. Woods, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10541–10545.

[12] I. Peña, C. Cabezas, J. L. Alonso, *Chem. Commun.* **2015**, *51*, 10115–10118.

[13] P. Çarçabal, R. A. Jockusc, I. Hünig, L. C. Snoek, R. T. Kroemer, B. G. Davis, D. P. Gamblin, I. Compagnon, J. Oomens, J. P. Simons, *J. Am. Chem. Soc.* **2005**, *127*, 11414–11425.

[14] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J. Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.

[15] M. Szczepaniak, J. Moc, *Carbohydr. Res.* **2014**, *384*, 20–36.

[16] I. Alkorta, P. L. A. Popelier, *Carbohydr. Res.* **2011**, DOI 10.1016/j.carres.2011.10.013.

[17] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.

[18] C. Calabrese, I. Uriarte, A. Insausti, M. Vallejo-López, F. J. Basterretxea, S. A. Cochrane, B. G. Davis, F. Corzana, E. J. Cocinero, *ACS Cent. Sci.* **2020**, *6*, 293–303.

[19] I. Peña, E. J. Cocinero, C. Cabezas, A. Lesarri, S. Mata, P. Écija, A. M. Daly, Á. Cimas, C. Bermúdez, F. J. Basterretxea, S. Blanco, J. A. Fernández, J. C. Lõpez, F. Castaño, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2013**, *52*, 11840–11845.

[20] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.

[21] I. Peña, S. Mata, A. Martín, C. Cabezas, A. M. Daly, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2013**, *15*, 18243–18248.

[22] P. Écija, I. Uriarte, L. Spada, B. G. Davis, W. Caminati, F. J. Basterretxea, A. Lesarri, E. J. Cocinero, *Chem. Commun.* **2016**, *52*, 6241–6244.

[23] I. Peña, C. Cabezas, J. L. Alonso, *Angew. Chemie - Int. Ed.* **2015**, *54*, 2991–2994.

[24] I. Uriarte, P. Écija, R. Lozada-Garcia, P. Çarçabal, E. J. Cocinero, *ChemPhysChem* **2018**, *19*, 766–773.

[25] J. D. Martínez, A. I. Manzano, E. Calviño, A. de Diego, B. Rodriguez De Francisco, C. Romanò, S. Oscarson, O. Millet, H. J. Gabius, J. Jiménez-Barbero, F. J. Cañada, *J. Org. Chem.* **2020**, *85*, 16072–16081.

[26] J. D. Martínez, A. S. Infantino, P. Valverde, T. Diercks, S. Delgado, N. C. Reichardt, A. Ardá, F. J. Cañada, S. Oscarson, J. Jiménez-Barbero, *Pharmaceuticals* **2020**, *13*, 1–18.

[27] Y. Zhou, J. Wang, Z. Gu, S. Wang, W. Zhu, J. L. Acenã, V. A. Soloshonok, K. Izawa, H. Liu, *Chem. Rev.* **2016**, *116*, 422–518.

[28] E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly, N. A. Meanwell, *J. Med. Chem.*

2015, *58*, 8315–8359.
[29] J. Wang, M. Sánchez-Roselló, J. L. Aceña, C. Del Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok, H. Liu, *Chem. Rev.* **2014**, *114*, 2432–2506.
[30] T. Fujiwara, D. O'Hagan, *J. Fluor. Chem.* **2014**, *167*, 16–29.
[31] S. Purser, P. R. Moore, S. Swallow, V. Gouverneur, *Chem. Soc. Rev.* **2008**, *37*, 320–330.
[32] J. Jiménez-Barbero, B. Linclau, A. Ardá, N. C. Reichardt, M. Sollogoub, L. Unione, S. P. Vincent, *Chem. Soc. Rev.* **2020**, *49*, 3863–3888.
[33] J. Wahsner, E. M. Gale, A. Rodríguez-Rodríguez, P. Caravan, *Chem. Rev.* **2019**, *119*, 957–1057.
[34] I. Tirotta, V. Dichiarante, C. Pigliacelli, G. Cavallo, G. Terraneo, F. B. Bombelli, P. Metrangolo, G. Resnati, *Chem. Rev.* **2015**, *115*, 1106–1129.
[35] X. Deng, J. Rong, L. Wang, N. Vasdev, L. Zhang, L. Josephson, S. H. Liang, *Angew. Chemie - Int. Ed.* **2019**, *58*, 2580–2605.
[36] S. M. Ametamey, M. Honer, P. A. Schubiger, *Chem. Rev.* **2008**, *108*, 1501–1516.
[37] P. W. Miller, N. J. Long, R. Vilar, A. D. Gee, *Angew. Chemie - Int. Ed.* **2008**, *47*, 8998–9033.
[38] S. Maschauer, O. Prante, *Biomed Res. Int.* **2014**, *2014*, DOI 10.1155/2014/214748.
[39] D. L. Orsi, R. A. Altman, *Chem. Commun.* **2017**, *53*, 7168–7181.
[40] C. Ni, J. Hu, *Chem. Soc. Rev.* **2016**, *45*, 5441–5454.
[41] L. Hunter, *Beilstein J. Org. Chem.* **2010**, *6*, 38.
[42] V. H. Dalvi, P. J. Rossky, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13603–13607.
[43] J. C. Biffinger, H. W. Kim, S. G. DiMagno, *ChemBioChem* **2004**, *5*, 622–627.
[44] E. J. Cocinero, E. C. Stanca-Kaposta, E. M. Scanlan, D. P. Gamblin, B. G. Davis, J. P. Simons, *Chem. - A Eur. J.* **2008**, *14*, 8947–8955.
[45] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[46] R. A. Motiyenko, E. A. Alekseev, S. F. Dyubko, F. J. Lovas, *J. Mol. Spectrosc.* **2006**, *240*, 93–101.
[47] W. Xu, H. Yang, Y. Liu, Y. Hua, B. He, X. Ning, Z. Qin, H. M. Liu, F. W. Liu, *Synth.* **2017**, *49*, 3686–3691.
[48] C. Falk, P. E. Jansson, M. Rinaudo, A. Heyraud, G. Widmalm, P. Hebbar, *Carbohydr. Res.* **1996**, *285*, 69–79.
[49] S. J. Angyal, *Carbohydr. Res.* **1994**, *263*, 1–11.
[50] A. De Bruyn, M. Anteunis, *Bull. des Sociétés Chim. Belges* **1975**, *84*, 1201–1209.
[51] A. Avenoza, C. Cativiela, F. Corzana, J. M. Peregrina, M. M. Zurbano, *Synthesis (Stuttg).* **1997**, 1146–1150.
[52] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.
[53] D. Šišak, L. B. McCusker, G. Zandomeneghi, B. H. Meier, D. Bläser, R. Boese, W. Bernd Schweizer, R. Gilmour, J. D. Dunitz, *Angew. Chemie - Int. Ed.* **2010**, *49*, 4503–4505.
[54] M. Kibalchenko, D. Lee, L. Shao, M. C. Payne, J. J. Titman, J. R. Yates, *Chem. Phys. Lett.* **2010**, *498*, 270–276.
[55] Y. Y. Chen, S. Y. Luo, S. C. Hung, S. I. Chan, D. L. M. Tzou, *Carbohydr. Res.* **2005**, *340*, 723–729.
[56] A. N. Appleyard, R. B. Herbert, P. J. F. Henderson, A. Watts, P. J. R. Spooner, *Biochim. Biophys. Acta - Biomembr.* **2000**, *1509*, 55–64.
[57] P. E. Pfeffer, K. B. Hicks, M. H. Frey, S. J. Opella, W. L. Earl, *J. Carbohydr. Chem.* **1984**, *3*, 197–217.
[58] P. De Andrade, J. C. Muñoz-garcía, G. Pergolizzi, V. Gabrielli, R. Harniman, S. J. Eichhorn, J. Angulo, Y. Z. Khimyak, P. de Andrade, J. . Munoz-garcia, G. Pergolizzi, V. Gabrielli, S. A. Nepogodiev, D. Iuga, L. Fabian, R. Nigmatullin, M. A.

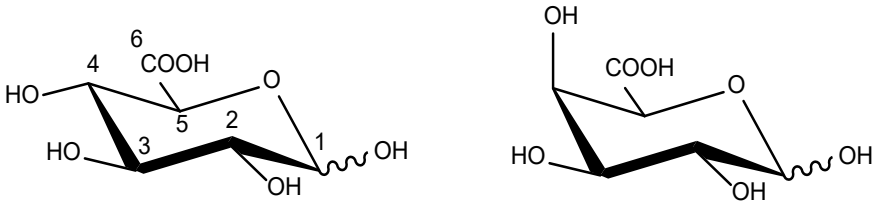
Johns, R. Harniman, S. J. Eichhorn, J. Angulo, Y. Z. Khimyak, R. A. Field, *ChemRxiv* **2020**, 1–5.
[59] I. Peña, A. M. Daly, C. Cabezas, S. Mata, J. L. Alonso, in *68th OSU Int. Symp. Mol. Spectrosc. Ohio State Univ. Columbus*, **2013**.
[60] I. Uriarte, A. Insausti, E. J. Cocinero, A. Jabri, I. Kleiner, H. Mouhib, I. Alkorta, *J. Phys. Chem. Lett.* **2018**, DOI 10.1021/acs.jpclett.8b02339.
[61] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[62] F. O. Talbot, J. P. Simons, *Phys. Chem. Chem. Phys.* **2002**, *4*, 3562–3565.
[63] N. Mayorkas, S. Rudić, E. J. Cocinero, B. G. Davis, J. P. Simons, *Phys. Chem. Chem. Phys.* **2011**, DOI 10.1039/c1cp22348h.
[64] C. A. G. Haasnoot, F. A. A. M. de Leeuw, C. Altona, *Tetrahedron* **1980**, *36*, 2783–2792.
[65] J. Gonzalez-Outeiriño, K. N. Kirschner, S. Thobhani, R. J. Woods, *Can. J. Chem.* **2006**, *84*, 569–579.
[66] M. Ziemniak, A. Zawadzka-Kazimierczuk, S. Pawlędzio, M. Malinska, M. Sołtyka, D. Trzybiński, W. Koźmiński, S. Skora, R. Zieliński, I. Fokt, W. Priebe, K. Woźniak, B. Pająk, *Int. J. Mol. Sci.* **2021**, *22*, 3720.
[67] Y. Nishida, H. Hori, H. Ohrui, H. Meguro, *J. Carbohydr. Chem.* **1988**, *7*, 239–250.
[68] A. De Bruyn, M. Anteunis, *Carbohydr. Res.* **1976**, *47*, 311–314.
[69] A. Lonardi, P. Oborský, P. H. Hünenberger, *Helv. Chim. Acta* **2017**, *100*, e1600158.
[70] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

VI. kapitulua

Azido uronikoak

VI.1 Sarrera

Kapitulu honetan, naturan ugarienak diren azido hexuronikoen azterketa aurkeztuko dugu: azido glukuronikoa eta azido galakturonikoa ain zuzen ere. Monosakaridoek eraldaketa kimikoak izaten dituzte, adibidez, azetilazioa, metilazioa, oxidazioa eta sulfonazioa eta horrek karbohidrato sinpleenen oinarritik abiatuta dibertsitate kimiko handiak sortzen ditu. Ohiko eraldaketatko bat, hexosen hidroximetiloa taldearen oxidazioa da, zein azido karboxilikoa ematen duen. Molekula berri horiei azido uroniko edo azido hexuroniko deitzen zaie. Hori da azido glukuroniko (**GlcA**) eta azido galakturonikoaren (**GalA**) kasua, non D-Glukosaren (Glc) eta D-Galaktosaren (Gal) seigarren karbonoko alkohol primarioaren oxidaziotik eratortzen diren molekulak dira hurrenez hurren, hidroxilo taldea talde karboxilikoa bihurtzen delarik (ikusi VI.1 eskema).



D-Glucuronic acid **D-Galacturonic acid**

VI.1 eskema azido D-glukuroniko eta acido D-galacturonikoaren $^4C_1$ aulki egitura egitura anomeriko generikoan

Azido hexuronikoak monosakarido oso arruntak dira, batez ere polisakaridoen osagai gisa. Naturan, **GlcA**-ren presentzia oso handia da, eta M. Jafféḱ isolatu zuen lehen aldiz 1874. urtean animalien gernutik,[1] eta E. Fischer eta O. Pilotyk kimikoki sintetizatu zuten 1891. urtean.[2] Ingurune biologikoetan, glikosaminoglikano polisakaridoetan hala nola heparina,[3] dermantan sulfatoak,[4] azido hialuronikoa,[5] gehien agertzen den monosakaridoetako bat da, eta baita hainbat polisakarido begetaletan ere agertzen da, adibidez hemizelulosan[6] eta hainbat bakterioen polisakaridoetan. **GlcA** monomeroa eraikuntza bloke nagusi gisa duten polisakarido mota asko daude, eta, jakina, prozesu biologiko ugaritan parte hartzen dute. Adibidez, gehien aztertu den glikosaminoglikano batek, heparinak, antikoagulaziotik haratago zabaltzen diren hainbat efektu dituela frogatu da, hanturaren aurkako eta ankerigenen aurkako propietateak barne. Hainbat ikerketa zientifikoak heparina-oligosakaridoen egituren funtzio bat identifikatu dute hainbat gaixotasunen bitartekotzan, hala nola Alzheimer gaixotasuna,[7] infekzio birala[8] eta minbizia.[9–11]

GalA-ri dagokionez, eskuarki landare erreinuan dago, zehazki pektinetan.[12] Pektinak, α-1,4 lotutako azido D-galakturonikoa eskeleto bat duen polisakarido bat da, eta baita beste azukre batzuetaz osatutako bigarren mailako adardurak ditu ere. Gaur egun, pektina sendagai naturalen eta osasunerako

produktuen ikerketaren zati garrantzitsua da, ez bakarrik naturan oso ugaria delako, baita minbizia bezalako hainbat gaixotasunen prebentzioan eraginkorra izan delako ere.[13–15] Industria farmazeutikoan, odoleko eta gastroeste kolesterol mailak murrizteko erabiltzen da.[16,17]

Joan den mendetik, polisakaridoek ohikoak izan dira biomedikuntza munduan, baina azukreen kimikaren konplexutasunak ezinezko egiten du haien funtzio biologikoak aurreikustea saiakuntza biomediko eta klinikorik egin gabe. Horregatik, glikobiologiaren gai nagusietako bat polisakarido bakoitzaren funtzio biologikoa iragartzea da, haren egitura eta konformazioa bakarrik ezagututa. Testuinguru horretan, jakina da molekula baten funtzioa oso lotuta dagoela bere hiru dimentsioko formarekin,[18,19] eta, jakina, polisakaridoen konformazio izaera zuzenean lotuta dagoela bere jarduera biologikoarekin.[20–22] Horrela, azukre bakoitzaren funtzionaltasuna aurreikusteko ezinbesteko lehen urratsa azukrearen egitura zehatza ondo ezagutzea da. Hala ere, polisakaridoen konformazio-panorama zenbait teknika esperimental erabiliz argitu daitekeen arren, hala nola erresonantzia magnetiko nuklearra, [23] indar atomikoko mikroskopia,[24] kristalografia[25] edo/eta mugikortasun ionikoa/masa-espektrometria,[26] orokorrean, emaitza horien interpretazioa kimika konputazionaleko kalkuluez lagunduta egin behar dira. Honek bi galdera planteatzen ditu: lehenik, azukrea sintetizatu edo lortu behar dugu bere egitura ezagutu aurretik; bigarrenik, konposatua lortu ondoren, ezin da beti egitura argitu, edo lortutako emaitza espero ez dena izan daiteke.

Irtenbidea, azukrearen konformazioa konputazionalki aurreikustea litzateke, eta datu horiek erabiltzea eragin biologikoa ebaluatzeko, intereseko lagina sintetizatu edo eskuratu aurretik. Hala ere, egungo metodo konputazionalekin ezin da egiturak zehatza iragarri. Testuinguru horretan, molekula horien forma konformazionala gobernatzen duten legeak modu logikoak ulertzeko aurretiaz monosakarido bakoitzak ingurune isolatu batean duen konformazio panorama ondo baino hobeto ulertu behar da.

Horretarako, azterlan honetan, mikrouhinen espektroskopiaren bidez, **GlcA** eta **GalA**-ren konformazio portaeren berri ematen da lehen aldiz, laser ablazioaren bidez gas fasean arrakastaz transferitu direnalarik. Egoera solidoko erresonantzia magnetiko nuklearra (ss-NMR), birkristalizazioa, espektroskopia errotazionala eta maila altuko kimika kuantikoko kalkuluak konbinatu ditugu, **GlcA** and **GalA**-ren α eta β anomeroak ikertzeko.

VI.2 Metodologia

Lan honetan, azido D-glukuroniko eta azido D-galakturoniko hidratoaren lagin solido komertzialak erabili ziren espektro errotazionalak lortzeko. Gainera, aldez aurretik argitaratutako lan bati jarraituz, azido D-galakturoniko hidratoaren zati bat azido azetikoan disolbatu eta birkristalizatu zen, β-anomeroa lortzeko

(xehetasun guztiak VI. eranskineko AVI.2 puntuan).[27] Gero, tesi honetako II. kapituluan zehatz aipatutako prozedura bera jarraitu zen hagatxo solido bat sortzeko. Molekulak Fabry-Perot barrumbea eta Fourierren transformatuan oinarritutako mikrouhin espektrometroa (cav-FTMW) erabiliz neurtu ziren, II. kapituluan azaldu den bezala, pikosegundu pultsu denborako laser (355 nm) bidezko ablazioa erabilti dugu laginak lurruntzeko, eta Ar (~ 55 kPa) gas garraiatzaile gisa erabili da.

Bilaketa konformazionala mekanika molekularreko kalkuluen bidez egin zen. Zehazki, Maestro2016 ezarrita dagoen *Merck Molecular Force Field* (MMFFs) indar eremua erabili da, eta **GalA** and **GlcA**-ren α eta β piranosa egiturako 15 kJ/mol baino energia erlatibo txikiagoa duten konformero guztiak hartu dira. Ondoren, egitura guztien geometriak optimizatzeko eta propietate espektroskopiko garrantzitsuak kalkulatzeko, Møller-Plesset-en bigarren mailako metodoa (MP2) eta Dentsitatearen Teoria Funtzionalaren (DFT) metodo batzuk erabili ziren 6-311++G (d,p) oinarri teorikoarekin batera. Lortutako emaitzak VI.1-VI.2 tauletan, AVI.1 AVI.4 tauletan eta VI. eranskineko AVI.1 AVI.4 irudietan laburbiltzen dira (xehetasun guztiak VI. eranskineko metodo konputazionaletan daude). Furano eraztunak detektatzeko aukerari uko egiten genion, bihurdura tentsio handiagoa dutelako.

VI.3 Emaitzak eta eztabaida

Konputazionalki, **GlcA** eta **GalA** α eta β-anomeroen konformazio panorama aurreikusi ondoren, (ikus Metodologia, AVI.1 AVI.4 taulak eta AVI.1 AVI.4 irudia VI. eranskinean), bi esperimentu independente egin ziren, bata **GlcA**-rentzat eta bestea **GalA**-rentzat. Mikro uhin espektroaren ekorketak, teorikoki iragarritako errotazio trantsizio intentsuenen frekuentzietan egin ziren cav-FTMW espektrometroaren bidez.[28] Espektrometro hori, unibertsitatean bertan eraiki da eta laser ultra azkarreko lurrutze sistema batez hornituta dago.[29,30] Lehenik eta behin, **GlcA** eta **GalA** lagin komertzialak erabili ziren esperimentuetan. Esperimentalki μ_a-*R*-motako trantsizio errotazionalen hiru multzo identifikatu ahal izan ziren, **GlcA** bi egitura eta **GalA**-ren egitura bati dagozkionak. Multzo bakoitza *Pickett*-en SPFIT programarekin doitu zen,[31] *Watson*-en Hamiltondar simetriko erdi-zurrun murriztua eta I^r irudikapena erabiliz[32] VI.1-VI.2. taulako errotazio parametroak lortzeko. Bereziki, errotazio konstanteak oso erabilgarriak dira, izan ere egitura molekularraren aztarna ukigabea baitira. Errotazio konstante esperimental eta teorikoen konparazioari esker (VI.1-VI.2 taula, VI. gehigarriko AVI.1-AVI.4 taula) **α-GlcA 1** eta **α-GalA 1** konomeroak zuzenean eta zalantzarik gabe identifikatzeko aukera izan genuen. Gainera, bi egitura horiek teoriaren arabera dagozkien gutxieneko globalak dira. **GlcA**-ren bigarren konformeroa identifikatzeko momentu dipolarreko osagaiek erabili behar izan genituen, dagozkien espektroen seinaleen eskuragarritasunarekin edo gabeziarekin erlazionatuta daudenak. Horrekin, zalantzarik gabeko esleipena lortu genuen

(ikusi AVI.2 taula AVI eranskinean). Ez ziren bestelako konformeroen seinalerik identifikatu egindako lehenengo espektroetan. Esperimentalki lortutako errotazio parametro multzoa eta neurtutako trantsizioen frekuentziak VI.1-VI.2 taulan eta VI. ranskinean, AVI.8-AVI.10 tauletan aurki daitezke, hurrenez hurren.

VI.1 taula Esperimentalki eta teorikoki lortutako emaitzak **GlcA**-ren α and β anomeroetzat.

	α-GlcA 1		**β-GlcA 1**	
	Theo.[a]	**Exp.**	**Theo.**	**Exp.**
A [MHz][b]	1239	1235.96890(50)[f]	1137	1135.72639(45)
B [MHz]	645	642.711001(73)	661	658.97886(10)
C [MHz]	468	466.342985(42)	436	434.973356(63)
D_J [kHz]		7.03(12)		0.00797(21)
d_1 [kHz]		--		-0.00197(15)
$\mu_a/\mu_b/\mu_c$ (D)	2.9/0.5/0.5	Yes/No/No[g]	2.6/0.6/0.7	Yes/No/No
σ [kHz][c]		1.1		2.0
N[d]		37		85
Com. gas/sol. [%][e]		24(8):10		76:90

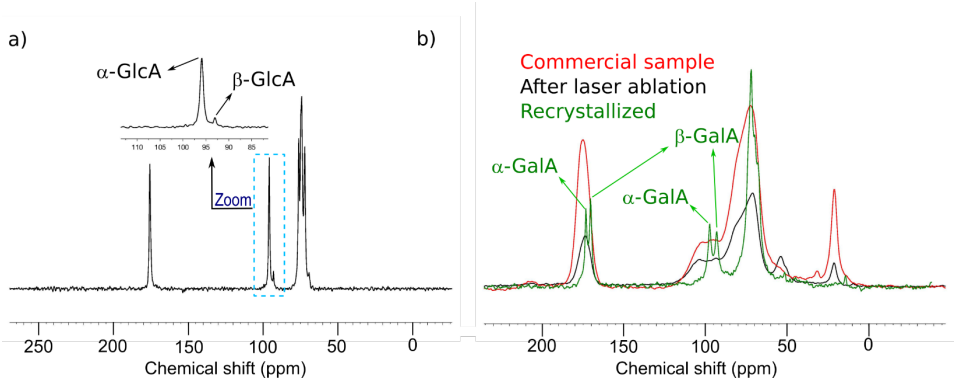
[a] *Ab initio* kalkulual (MP2/6-311++G(d,p)).
[b] Errotazio konstanteak (A, B, C); μ_a, μ_b and μ_c momentu dipolarren balioak dira Debye-tan.
[c] Doiketaren bataz besteko desbiazioa.
[d] Doitutako trantsizio kopurua.
[e] Anomeroen populazio esperimentala lagin komertzialean (Com.) and eta birkristaldutako lagina (Recr.) gas eta solido egoeran.
[f] Asken digituen errore estandarra parentesi artean dago..
[g] Egitura bakoitzarentzat esperimentalki ikusitako *a*-, *b*-, eta *c*- μ- motako trantsizioak.

VI.2 taula Esperimentalki eta teorikoki lortutako emaitzak **GalA**-ren α and β anomeroetzat

	α-GalA 1		**β-GalA 1**	
	Theo.[a]	**Exp.**	**Theo.**	**Exp.**
A [MHz][b]	1310	1303.8852(32)[f]	1242	1234.2059(36)
B [MHz]	644	642.79144(12)	633	633.08463(27)
C [MHz]	486	483.60745(13)	474	470.39353(11)
D_J [kHz]		0.00839(39)		
D_{JK} [kHz]		0.0451(54)		0.102(25)
D_K [kHz]		-0.19(14)		
$\mu_a/\mu_b/\mu_c$ (D)	2.5/0.0/0.5	Yes/No/No[g]	5.1/2.4/1	Yes/No/No
σ [kHz][c]		2.8		3.2
N[d]		52		20
Com. gas/sol. [%][e]		100:>95		0/5
Recr. gas/sol. [%]		17(5):50		83:50

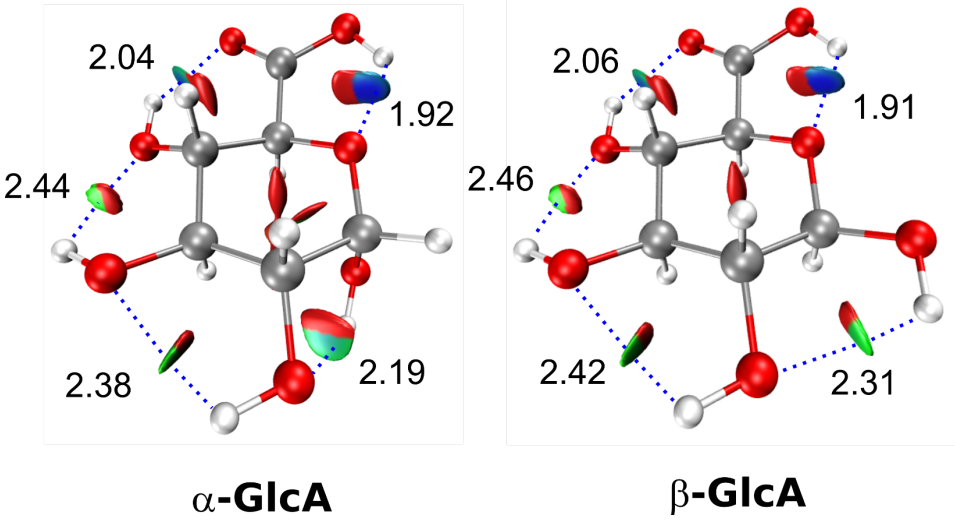
[a] *Ab initio* kalkulual (MP2/6-311++G(d,p)).
[b] Errotazio konstanteak (A, B, C); μ_a, μ_b and μ_c momentu dipolarren balioak dira Debye-tan.
[c] Doiketaren bataz besteko desbiazioa.
[d] Doitutako trantsizio kopurua.
[e] Anomeroen populazio esperimentala lagin komertzialean (Com.) and eta birkristaldutako lagina (Recr.) gas eta solido egoeran.
[f] Asken digituen errore estandarra parentesi artean dago..
[g] Egitura bakoitzarentzat esperimentalki ikusitako *a*-, *b*-, eta *c*- μ- motako trantsizioak.

Aipatzekoa da, hasiera batean **GalA** molekularen β-anomeroaren konfermerorik ez zeal ikusi lortutako emaitza esperimentaleta, nahiz eta hori komertzialki eskuratutako laginaren α/β populazioaren ondorio izan litekeen. Badakigu α↔β eta furanosa↔piranosa interkonbertsioa (errotamutazioa) gas-fasean eragotzita dagoela,[33] eta, beraz, lagin solidoan ez dagoen anomeroa, ezin da gas fasean detektatu. Hori aztertzeko, lagin solidoaren populazio anomerikoa ikertu zen ss-NMR-a erabilita (ikus VI.1 irudia). Lagin komertzialearen ss-NMR espektroen analisiak baieztatu egin zuen aurretik susmatutako α-anomeroaren hondar proportzioa lagin komertzialean (α:β, 95:<5). Aitzitik, **GlcA**-n β-anomeroaren proportzio handiagoa agertu zen (α:β, 10:90). **β-GalA**-ren konformatzaileak behatzeko, **GalA**-ren lagin komertzial zati bat (1 g inguru) azido azetikoan birkristalizatu zen bibliografiari jarraituz,[34] eta berriro anomeroen populazio analisia ss-NMR bidez egin zen (VI. irudia). Birkristalizazioa arrakastatsua izan zen eta α:β populazio proportzio egokia lortu genuen errotazio esperimentua egiteko (α:β, 50:50) (ikus VI.1 irudia). Puntu honetan, berriz ere prest geunden anomero hau bilatzeko gas fasean. Aurreko prozedurari jarraituz, **β-GalA 1**-ren errotazio espektroa detektatu eta doitu zen (VI.2 taula eta AVI.11 taula), eta lortutako egitura energia minimodun konformeroa zen (VI. eranskineko AVI.4 taula). Erregistratutako espektroan ez zen konformero gehiago detektatu, eta, beraz, hurrengo eztabaida detektatutako egituretan soilik zentratuko da.



VI.1 irudia a) **GlcA** lagin komentzialaren ss-NMR espektroa. b) **GalA** lagin komertzial (marra gorria) ablazio prozesu osteko (marra beltza) eta azido azetikoan berkristaldutako laginaren (marra berdea) ss-NMR espektroa. Anomero bakoitzaren gailurrak adierazi dira irudian.

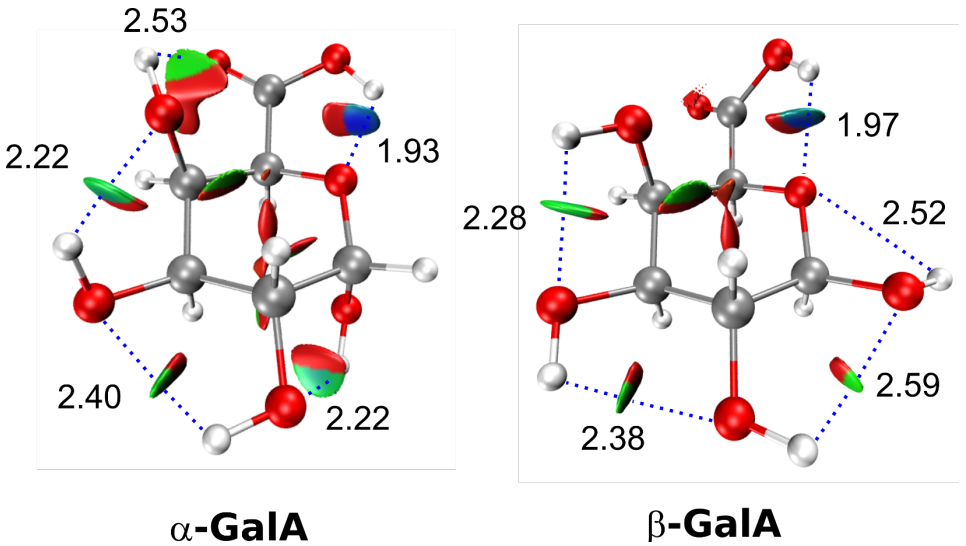
Orain, jakinminez geunden ea gas-faseko eta solido faseko biztanleria anomerikoa antzekoak ziren, eta, beraz, fase bakoitzeko populazioak ezagutzeko gas-faseko espektroskopia errotazional esperimentaleko seinaleen intentsitatea eta momentu dipolar teorikoak erabiliz,[35] eta egoera solidoko ss-NMR intentsitatea erabiliz euren populazioak lortu ziren. Lortutako emaitzek harritu egin gintuzten, bi molekuletarako antzeko joera espero genuelako. Hala ere, **GlcA**-ren kasuan gas eta solido fasean zeuden populazioak oso antzekoak ziren arren, baina **GalA**-ren kasuan bi anomeroen populazioak zeharo ezberdinak ziren. Ez dakigu zehazki zein den emaitza horren arrazoia, baina honako hipotesi hau planteatzen dugu: anomero biztanleria desberdina bi tekniketan errotazio espektroskopian laser bidezko lurruntze eta ondoren espektrometro errotazionalean hedapeneko hozteko prozesuaren eraginkortasuna eta degradazioren desberdinaren ondorioa da.



VI.2 irudia GlcA laginean behatutako konformeroak eta interakzio intramolekularrak (NCI analisitik lortuak). Lerro etenek hidrogeno loturaren interakzioa adierazten dituzte. Elkarrekintza ez-kobalenteen banaketa espazialean, gradient gainazala (s = 0,5 u.a.) dentsitate elektronikoaren balioen arabera eskala urdine berde eta gorrian koloreztatu dira, bigarren Hesiar balio propioaen arabera (sign(λ_2)ρ), -0.03 eta 0.01 bitarteko tartean. O-H···O hidrogeno lotura zendoenak (urdine) interakzio ahoulagoekin (tonu berdeak) eta aldarapen guneekin (gorri) batera adierazi dira.

Datu esperimental horien arabera, bi molekula horien egitura arrazionalizatu egin daitezke euren energia egonkortzen dituzten elkarrekintza desberdinen arabera, zehazki, hidrogeno lotura intramolekular sareen arabera. Elkarrekintza ez-kobalenteen banaketa espaziala, NCI-plot analisiaren arabera irudikatu da (VI.2 irudia eta VI.3 irudia).[36] **GlcA** molekularen bi anomeroetan, erlojuaren orratzen noranzkoan dagoen eta molekula osoan zehar hedatzen den H-lotura sareak (O1H(ax/eq) → O2H(eq) → O3H(eq) → O4H(eq) → O=COH →

O5) konformero egonkorrenen egonkortasun bereziaren erantzulea da. Gainera detektatutako GlcA-ren egiturak bi h lotura bereziki sendoak ditu –O4H(eq) → O=C (MP2/6-311++G(d,p): r_e=2.04-2.06 Å) eta O=COH → O5 (MP2/6-311++G(d,p): r_e=1.91-1.92 Å) (VI.2 irudia) Emaitza horiek, **α/β-Glc** eta α-D-glukosamine (**α-GlcN**) egiturarekin konparatuta, nahiko desberdinak dira, izan ere, molekula horien konformero egonkorrenetan hidrogeno lotura sareak erlojuaren orratzen aurkako noranzkoa hartzen dute eta egiturak O6H → O5 interakzioaren bidez egonkortuta daude (**α/β-Glc**-ren egiturak eranskinetako AVI.8-AVI.9 irudietan daude).[37,38]



VI.3 irudia GalA laginean behatutako konformeroak eta interakzio intramolekularrak (NCI analisitik lortuak). Lerro etenek hidrogeno loturaren interakzioa adierazten dituzte. Elkarrekintza ez-kobalenteen banaketa espazialean, gradient gainazala (s = 0,5 u.a.) dentsitate elektronikoaren balioen arabera eskala urdine berde eta gorrian koloreztatu dira, bigarren Hesiar balio propioaen arabera (sign(λ_2)ρ), -0.03 eta 0.01 bitarteko tartean. O-H···O hidrogeno lotura zendoenak (urdine) interakzio ahoulagoekin (tonu berdeak) eta aldarapen guneekin (gorri) batera adierazi dira.

Orain, arreta berezia jarri behar zaio **GalA** molekulari (ikusi VI.3 irudia). α-anomeroaren kasuan, hidrogeno lotura sare kooperatibo intramolekularra bat ere erlojuaren orratzen noranzkoa hartzen du, nahiz eta O4H(ax) → O=C interakzioa α α-GlcA analogoa baino askoz ahulagoa den (MP2/6-311++G(d,p): r_e=2.53 Å). Egia esan, hidrogeno-loturen sarearen norabidea **α/β-Gal**[39] and **β-2-deoxy-Gal**-ren (tesi honen IV. kapitulua) egitura egonkorrenaren berdina da, baina oxidatu gabeko azukreetan, O4H(ax) → O6 interakzioa sendoagoa da (MP2/6-311++G(d,p): r_e=1.94-1.96 Å). Datu esperimental horiek iradokitzen dute H-loturaren norabidea erlojuaren orratzen noranzkoaren erantzual agian O4H → O6 interakzio sendoari esleitzea berrikusi behar dela. Bitxia bada ere, **β-GalA**-ren

egiturarik egonkorrenak hidrogenozko lotura sarearen ordutegiaren aurkako norabidean dagoela, lotura kooperatibo bat O4H(Ax) → O3H(eq) → O2H(eq) → O1H(eq) eta ez kooperatiboa den beste atal bat O=COH → O5 H-lotura du, beraz ez da molekula osoan zehar edatzen. Hidrogenozko loturaren kokapena erlojuaren orratzen kontrako norabidean ikusi zen deribatu galaktosa ez deribatuetan eta desoxidatu gabeko aldohexosen egitura egonkorrenetan, hala nola **Glc**, **GlcN**, **2F-Glc** and **2F-Man** (ikus tesi honen IV. kapitulua).[37,38]

VI.4 Ondorioak

Laburbilduz, aldohexosa konbentzionalek hidroximetilo talde exozikliko (-CH_2OH) malgu bat dute, hiru konfigurazio posible sortzen dituena. [38–40] **GlcA** eta **GalA**-ren kasuan, hidroximetilo taldearen ordez, azido karboxiliko talde bat dago (– COOH), eta karboxilo taldea disposizio bakarrean finkatzen da, izan ere molekulak bost kideko eraztuna eratzen du azidoaren hidroxilo taldea eta eraztuneko oxigenoaren H-lotura lotura sendoren bidez: O=COH → O_{Ring} (MP2/6-311++G(d,p): r_e=1.91-1.97 Å). Estrukturalki ondoriozta dezakegu talde karboxilikoak ez duela bere gain hartzen bestelako azukreen hidroximetilo taldearen rol bera, eta horrek **GlcA** eta **GalA** molekulen konformazio egonkor kopurua murrizten du. Gainera, **Glc** puruarekin konparatuta, non bi anomeroetan egitura gehienek hidrogenozko lotura sare bat hartzen duten erlojuen orratzen aurkako norabidean, azido glukuronikoaren talde azidoak hidrogenozko lotura sarearen norabidea bakarrik noranzko bat hartzera behartzen ditu egitura guztietan, erlojuaren orratzen noranzkoan ain zuzen ere. **α-GalA**-ren kasuan, elarazitako emaitzek iradokitzen dute erlojuaren orratzen noranzkoan H-lotura sarearen norabide lehentasun intrintsekoak eta **α-GalA** eta **β-GalA** anomeroetako, forma anomerikoaren aldaketek, H-lotura sarearen norabide aldaketak molekula osoari eragiten diola esperimentalki frogatu da.

VI.5 Erreferentziak

[1] M. Jaffé, *Berichte der Dtsch. Chem. Gesellschaft* **1874**, *7*, 1673.
[2] E. Fischer, O. Piloty, *Chem. Ber.* **1891**, *24*, 4214–4225.
[3] R. D. Rosenberg, L. Lam, *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 1218–1222.
[4] L. C. Rosenberg, H. U. Choi, L. H. Tang, T. L. Johnson, S. Pal, C. Webber, A. Reiner, A. R. Poole, *J. Biol. Chem.* **1985**, *260*, 6304–6313.
[5] L.-A. Franson, in *Polysaccharides, Vol. 3* (Ed.: G.O. Aspinall), Academic Press, New York, **1985**, pp. 338–415.
[6] T. E. Timell, *Adv. Carbohydr. Chem.* **1964**, *19*, 247–302.
[7] Z. Scholefield, E. A. Yates, G. Wayne, A. Amour, W. McDowell, J. E. Turnbull, *J. Cell Biol.* **2003**, *163*, 97–107.
[8] V. Tiwari, C. Clement, M. B. Duncan, J. Chen, J. Liu, D. Shukla, *J. Gen. Virol.* **2004**, *85*, 805–809.
[9] B. Casu, I. Vlodavsky, R. D. Sanderson, *Pathophysiol. Haemost. Thromb.* **2008**, *36*, 195–203.
[10] L. Borsig, R. Wong, J. Feramisco, D. R. Nadeau, N. M. Varki, A. Varki, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3352–3357.
[11] S. N. Ma, Z. X. Mao, Y. Wu, M. X. Liang, D. D. Wang, X. Chen, P. an Chang, W. Zhang, J. H. Tang, *Cell Adhes. Migr.* **2020**, *14*, 118–128.
[12] L. Flutto, in *Encycl. Food Sci. Nutr.*, Elsevier, **2003**, pp. 4440–4449.
[13] H. Ohkami, K. Tazawa, I. Yamashita, T. Shimizu, K. Murai, K. Kobashi, M. Fujimaki, **1995**, *86*, 523–529.
[14] D. W. Heitman, W. E. Hardman, I. L. Cameron, *Carcinogenesis* **1992**, *13*, 815–818.
[15] L. Leclere, P. Van Cutsem, C. Michiels, *Front. Pharmacol.* **2013**, *4 OCT*, 1–8.
[16] S. T. Minzanova, V. F. Mironov, D. M. Arkhipova, A. V. Khabibullina, L. G. Mironova, Y. M. Zakirova, V. A. Milyukov, *Polymers (Basel).* **2018**, *10*, 1–31.
[17] F. Brouns, E. Theuwissen, A. Adam, M. Bell, A. Berger, R. P. Mensink, *Eur. J. Clin. Nutr.* **2012**, *66*, 591–599.
[18] D. D. Boehr, R. Nussinov, P. E. Wright, *Nat. Chem. Biol.* **2009**, *5*, 789–796.
[19] D. D. Boehr, P. E. Wright, *Science (80-.).* **2008**, *320*, 1429–1430.
[20] Y. Yu, M. Delbianco, *Chem. - A Eur. J.* **2020**, *26*, 9814–9825.
[21] C. R. Bertozzi, L. L. Kiessling, *Science (80-.).* **2001**, *291*, 2357–2364.
[22] M. Hasan, H. Khakzad, L. Happonen, A. Sundin, J. Unge, U. Mueller, J. Malmström, G. Westergren-Thorsson, L. Malmström, U. Ellervik, A. Malmström, E. Tykesson, *Chem. Sci.* **2021**, *12*, 1869–1885.
[23] H. Y. Y. Yao, J. Q. Wang, J. Y. Yin, S. P. Nie, M. Y. Xie, *Food Res. Int.* **2021**, *143*, 110290.
[24] Y. Meng, X. Shi, L. Cai, S. Zhang, K. Ding, S. Nie, C. Luo, X. Xu, L. Zhang, *Macromolecules* **2018**, *51*, 10150–10159.
[25] M. Plevin, *Encyclopedia of Biophysics*, **2013**.
[26] J. Sastre Toraño, O. Aizpurua-Olaizola, N. Wei, T. Li, L. Unione, G. Jiménez-Osés, F. Corzana, V. J. Somovilla, J. M. Falcon-Perez, G. J. Boons, *Chem. - A Eur. J.* **2021**, *27*, 2149–2154.
[27] W. G. Ferrier, *Acta Crystallogr.* **1963**, *16*, 1023–1031.
[28] T. J. Balle, W. H. Flygare, *Rev. Sci. Instrum.* **1981**, *52*, 33–45.
[29] E. J. Cocinero, A. Lesarri, P. Écija, J. U. Grabow, J. A. Fernández, F. Castaño, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486–12493.
[30] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J. U. Grabow, J. A. Fernández, F. Castaño, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3119–3124.
[31] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.
[32] R. K. Thomas, *Phys. Bull.* **1975**, *26*, 501–501.
[33] C. Calabrese, P. Écija, I. Compañón, M. Vallejo-López, Á. Cimas, M. Parra, F. J.

Basterretxea, J. I. Santos, J. Jiménez-Barbero, A. Lesarri, F. Corzana, E. J. Cocinero, *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.
[34] H. S. Isbell, H. L. Frush, *J. Res. Natl. Bur. Stand. (1934).* **1943**, *31*, 33.
[35] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[36] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
[37] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.
[38] I. Peña, L. Kolesniková, C. Cabezas, C. Bermúdez, M. Berdakin, A. Simão, J. L. Alonso, *Phys. Chem. Chem. Phys.* **2014**, *16*, 23244–23250.
[39] I. Peña, C. Cabezas, J. L. Alonso, *Chem. Commun.* **2015**, *51*, 10115–10118.
[40] J. L. Alonso, M. A. Lozoya, I. Peña, J. C. López, C. Cabezas, S. Mata, S. Blanco, *Chem. Sci.* **2014**, *5*, 515–522.

VII. kapitulua

Azido 2-furoiko: monomero eta dimeroaren errotazio espektroskopia ikerketa

Lan hau Albertako Unibertsitatean egin zen. Yunjie Xu and Wolfgang Jäger taldeko kide guztiei ezkerrak eman nahi dizkiet.

Kapitulu hau aurretik argirtaratua, itzulia eta egokitua izan da. Aran Insausti, Jiarui Ma, Qian Yang, Fan Xie and Yunjie Xu. Rotational Spectroscopy of 2-Furoic Acid and Its Dimer: Conformational Distribution and Double Proton Tunneling. *ChemPhysChem*, **2022**, e202200176

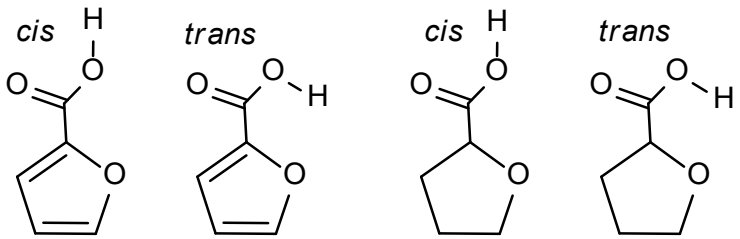
VII.1 Sarrera

Kapitulu honetan, 2021-2022 ikasturtean Albertako Unibertsitateko egonaldian (*Egonlabur* EP_2021_1_0032 beka) egindako lan bat aurkezten da. Konposatu organiko karbonilikoak oso ugariak dira eta ingurumen eta prozesu biologiko ugarietan parte hartzen dute, hala nola errekonozimendu zelularrean[1] eta euri azidoaren sorreran.[2] Orokorrean, COOH talde funtzionalak *cis* edo *trans*-COOH konfigurazioak har ditzazke, non C=O eta OH taldeak alde (gune) berean edo kontrakoan kokatuta dauden, hurrenez hurren. Horren ondorioz, karbonilo taldeak hidrogeno emaile edo hartzaile izaera izan dezake, lotura ez-kobalente mota desberdin ugari sortuz. Nomenklatura hori VII.1. eskeman irudikatu da. Azido karboxilikoak agregatu bitar eta handiagoak eratzeko gaitasunak nolabaiteko propietate bereziak izatea ahalbidetzen du, adibidez pisu molekular antzeko beste hidrokarburo organikoek baino irakite eta urtze puntu altuagoak dituzten molekulak dira.

Intereseko molekula den azido-2-furoikoa (furano-2-azido karboxilikoa, FA), azido karboxiliko aromatikoa da, 5 muturreko eraztun aromatiko batez eta azido karboniko talde batez osatuta dagoena. Elikagaien industrian oso erabilia da, kontserbatzaile, zapore emaile eta elikagai esterilizatzaile gisa batez ere erabiltzen dena. X-izpien bidez aurretiaz egindako lan kristalografikoan, homodimero zentro simetriko bat identifikatu zen, bi *cis*-COOH FA monomeroz osatuta, euren artean H-lotura bikoitzaren bidez komunikatuta dagoen eraztun bat osatuz.[3] 2009. urtean, aurretik X-izpien bidez ikertutako kristal mota berdinak berriz infragorri espektroen bidez aztertu eta doitu ziren, eta *cis*-COOH monomero bikotez eratutako bi dimero berri identifikatu ziren.[4] Gainera, 2014. urteko FA kristalen emaitza teoriko eta, infragorri (IR) eta Raman espektroskopiaren datu esperimentalen konbinaketaren bidez, solido egoeran bakarrik *cis*-COOH konformeroa zegoela ondorioztatzen zen,[5] X-izpien aurreko emaitzak berretsiz.[3] Duela gutxi, matrizean isolatuko molekulen infragorri ikerketa batean bi *cis*-COOH FA konformero desberdin detektatu ziren, Ar eta Ne matrize solidoan ia 1:1 populazio erlatiboarekin.[6] FA monomeroaren *cis*-COOH konformeroaren infragorri hurbileko $\upsilon_{O\text{-}H}$ gaintonuaren ponpaketaren bidez, *trans*-COOH konformeroa fotoien kitzikapenaren bidez sortu eta detektatu zen (VII. eskema begiratu).[6] Horretaz gain, guztiz erlazionatuta dagoen FA hidrogenatua, azido tetrahidro-2-furoikoa (THFA), orain dela gutxi espektroskopia errotazional, matrizean isolatutako IR espektroskopia eta dikroismo zirkular bibrazionalaren (VCD) bidez ikertua izan da.[7,8] THFA monomeroan, FA-ren kasuan ez bezala, *trans*-COOH konformeroa oso faboratua dagoela ikusi zen. THFA bere buruarekin, urarekin edo beste molekula kiral batekin (adibidez propileno oxidoa) interakzionatzean, konformeroen distribuzioa aldatu egiten da eta *cis*-COOH konformazioaren populazioa faboratzen da,[7,9,10] ingurumenaren garrantzia nabarmenduz. Hortaz, espektroskopia errotazionalak duen sentikortasun

konformazional bikaina dela eta, bere erabilpena FA monomeroaren eta dimeroen egitura konformazionalak eta populazio erlatiboa aztertzeko teknika oso interesgarria da, eta gainera, konformeroen banaketa monomeroetatik dimeroetara nola aldatzen den ikertzea ahalbidetzen du.

Beste intereseko puntu bat aldibereko protoi transferentzia bikoitza da, non OH taldeetako protoiak beste bazkideko karbonilo taldera transferitzen diren une berean. Protoi transferentzia gertaerak paper garrantzitsua du *Watson-Crick* base bikoteen ezaugarri azido-baseetan, eta era berean, produktu tautomeriko arraro batzuk ere era daitezke, DNA-ren egonkortasunari eraginez.[11] Ez da harrigarria, ikertzaileek espektroskopia mota ezberdinak aplikatu izatea, adibidez laser fluoreszentzia eta IR espektroskopia[12,13] eta noski errotazio espektroskopia ere bai, protoi transferentzia bikoitza ikertzeko. Lan honekin erlazionatuta dauden aurreko zenbait argitalpenetan, karbonilo taldea duten hetero- eta homodimeroen errotazio espektroskopia aztertu da, adibidez azido formikoa bere buruarekin,[14] azido propiolikoarekin,[15,16] azido azetikoarekin,[17] azido difluoroazetikoarekin,[18] o- eta m-anisikoarekin,[19,20] azido 3,3,3-trifluoro-2-(trifluorometil) propanoikoarekin,[21] azido perfluorobutirikoarekin[22] eta azido bentzoikoarekin.[23] Ez hori bakarrik, baita ere azido akrilikoaren homodimeroak aurretik ikusiak eta aztertuak izan dira.[24] Aipagarria da *tunel* bikoizketa bakarrik sistema batzuetan antzeman zela, eta ez beste batzuetan, adibidez TFA-ren homodimeroan ez zen igarri. Orokorrean, *tunel* bikoizketari buruzko datu espektroskopiko errotazionalak kalitate oso handikoak dira eta bikoizketa horren energia kuantifikatzen saiatzen diren zenbait hurbilketetarako erreferentziarako balio apartak dira, hala nola *Instanton* teoria[25] eta dimentsio bakarreko hurbilketa sinplifikatua.[26] *Tunel* bikoizketa azaltzen eta kuantifikatzen duten teoriak gaur egungo interes esanguratsua duen ikerketa topikoa dira.



VII.1. eskema FA eta THFA-ren *cis* eta *trans*-COOH konformeroen nomenklatura.

Kapitulu honetan pultsu multifrekuentzian (*chirped pulse*) eta Fourierren transformatuan oinarritutako mikrouhin (CP-FTMW) espektroskopia teknika aplikatu genuen FA monomeroaren *trans*-COOH konformeroa infragorri ponpaketarik gabe ikus litekeen aztertzeko. FA dimeroaren kasuan, bi alderdi

desberdinetan zentratu ginen: 1) emaitza teoriko eta esperimentalak konparatuz, CP-FTMW espektroskopia erabiliz lotutako geometrien identifikazioan; eta 2) FA dimeroan gerta litekeen protoi transferentzia *tunel* bikoizketa gertaera aztertu eta horren ebaluazioa orain dela gutxi garatutako prozedura konputazional sinple baten bidez aztertzea.[26]

VII.2 Metodologia

Errotazio espektroskopia

Aurkeztutako lana, Albertako Unibertsitateko Xu irakaslearen taldean burutu zen *Egonlabur* internazionalizazio egonaldian zehar. Hortaz, Albertako Unibertsitateko 2-6 GHz CP-FTMW espektrometroa erabili zen errotazio espektroa eskuratzeko.[27,28] Espektrometroa B. H. Pate eta bere lankideen instrumentuan oinarrituta dago eta oinarrizko xehetasun esperimentalak tesi honen 2. kapituluan daude azalduta.[29,30] Laburki, 12 $Gs \cdot s^{-1}$-ko uhin sorgailu (*Arbitrary Waveform Generator*, AWG) batek 2-6 GHz eta 1 μs-ko mikrouhin pultsua sortu, eta uhin hori ondoren 400 W-ko 2,5-7,5 GHz uhin-hodi biderkatzaile (*Traveling Wave Tube*, TWT) batekin anplifikatu zen. Huts ganbaran aurkitzen diren bi adar motako antenak erabili ziren mikrouhin pultsu anplifikatua ganbaran zabaltzeko eta molekulen indukzio askeko emisioa (*Free Induction Decay,* FID) jasotzeko. FID seinalea gorde eta digitalizatzeko 25 $Gs \cdot s^{-1}$ osziloskopioa erabili zen. 6-8 GHz-ko neurketetarako, Fabry-Perot barrunbean oinarritutako espektrometroa[31] erabili zen (esperimentua tesi honen II. kapituluan aurkeztutakoaren oso antzekoa da eta xehetasun guztiak bibliografian aurki daitezke). 8-12 GHz neurketarako, CP-FTMW-ren konfigurazio ezberdin bat erabili zen: 4-6 GHz-ko pultsua AWG-an sortu eta bikoiztu zen 8-12 GHz bandan lan egiteko. 1 GHz banda-zabalera eta 4μs-ko denbora tarteko pultsua egoera solidoko 20 W-ko anplifikagailuaz potentziatu zen.

2-6 (8-12) GHz CP-FTMW esperimentuetan, gas pultsu bakoitzeko, 15 (6) kitzikapen/detekzio zikloen seinaleak neurtu ziren eta guztira 700k (200k) FID-en batezbestekoa egin zen eta ondoren Fourier transformatuaren bidez frekuentzia espektroa lortu zen. Barrunbe esperimentuaren neurketarako, frekuentzien ziurgabetasuna ~1 kHz da gutxi gorabehera, eta CP-FTMW datuei dagozkien ~10 kHz.

FA (%98ko purutasuna), *Millipore-Sigma*-n erosi zen eta zuzenean purifikatu gabe erabili genuen. Espezie deuteratuak lortzeko, FA solidoa D_2O-n disolbatu zen, ondoren disoluzioa 2 orduz ultra-soinu bainuan 60-70 ºC inguruan berotu eta azkenik lehortu egin zen. CP-FTMW neurketetarako, FA solidoa *General Valve* pita eraldatuaren barruan jarri zen zuzenean eta ~ 110 ºC-tan berotu zen. Gas garraiatzaile gisa helioa erabili zen 50 psi-ko presioarekin. Barrunbeko neurketetarako, FA solidoa altzairu herdoilgaitz konpartimentu baten

barruan jarri zen, pita aurretik, eta ~ 110 ºC inguru berotu zen. Ne gas garraiatzailea erabili zen 25 psi-ko presioarekin.

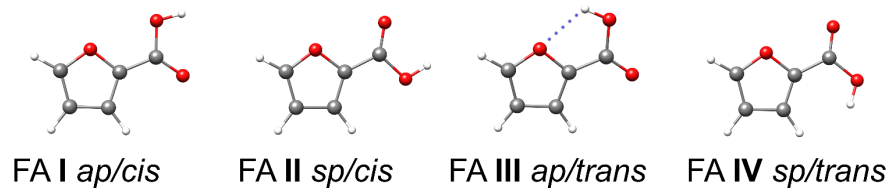
Metodologia konputazionala

Bilaketa konformazional sistematikoa CREST (*conformer-rotamer ensemble sampling tool*) programaren bidez egin zen, GFN-xTBn kimika kuantikoko metodo semi-empirikoan oinarrituta dagoen programa erabiliz.[32,33] Ondorengo urratseko geometria optimizazio eta frekuentzien kalkulu harmoniko guztiak Gaussian 16 programarekin[34] egin ziren, B3LYP-D3(BJ)/def2-TZVP teoriaren maila, dispertsio zuzenketarekin (D3) eta Becke-Johnson-en *damping* deritzon zuzenketekin (BJ),[35] guztiak noski def2-TZVP oinarri matematikoan.[36,37] Protoi transferentzia bikoitzeko *tunnelling* bikoizketa *Orca 4* eta *Python* softwarea erabiliz kalkulatuak izan ziren.[26,38]

VII.3 Emaitzak eta eztabaida

FA monomeroa

FA monomeroarentzat, 14 CREST[32,33] hautagai identifikatu ziren hasiera batean (ikusi aurreko Metodologia atala xehetasun gehiago ezagutzeko), eta ondoren B3LYP-D3 (BJ)/def2-TZVP teoriaren mailan bir-optimizatu ziren,VII.1 irudian agertzen diren lau egitura egonkorrak lortuz. Egitura bakoitzaren konstante espektroskopikoak VII.1.taulan laburbildu dira. Egiturak izendatzeko bi etiketa erabili dira: lehenengo bi hizkiek karboxiloak furano eraztunarekiko duen posizioa adierazten dute, *ap* edo *sp*, zeintzuk $\angle_{\text{Oring-C-C=O}}$ ~ 180º edo 0º hurrenez hurren adierazten duten; bigarren zatia lehen aipatutako *cis*- edo *trans*-COOH konfigurazioarekin erlazionatuta dago. Sinplifikatzearren, lau konformero horiek **I**, **II**, **III** eta **IV** bezala izendatu dira, egonkortasun erlatiboaren arabera ordenatuak daudenak, **I** izanda konformerorik egonkorrena.



VII.1 Irudia FA lau egitura egonkorrenen geometria optimizatua FA. OH eta eraztuneko heteroatomoaren kontaktu hurbilena puntuekin irudikatu da.

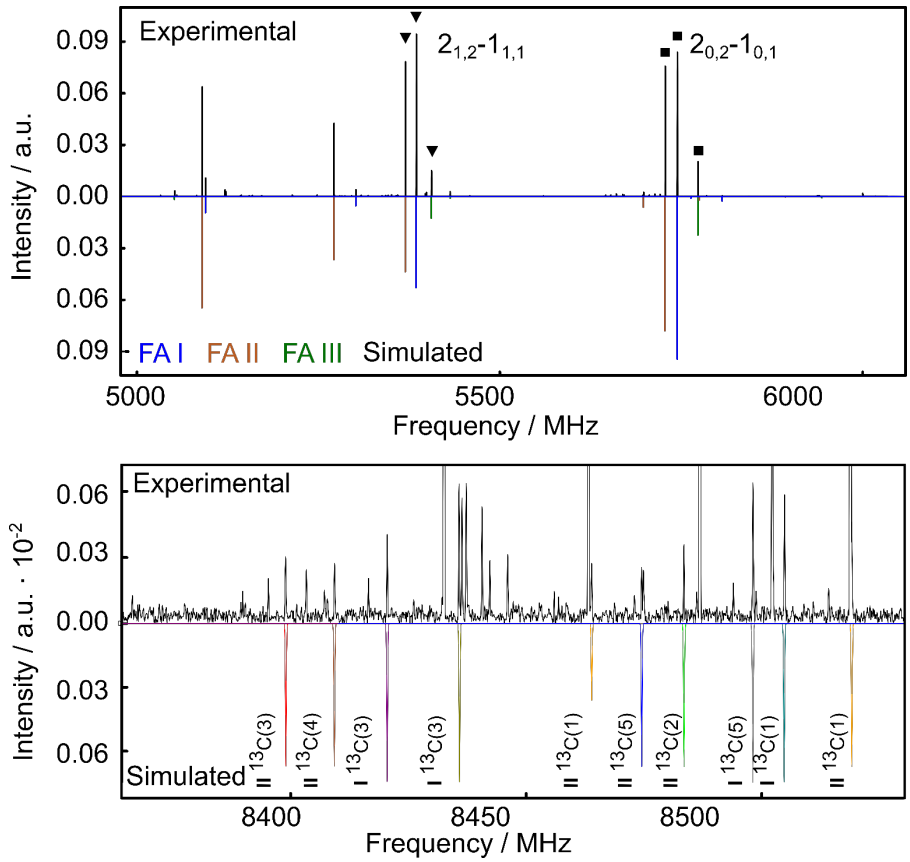
VII.1 taulan ikus daitekeen moduan, FA **I** eta **II** egiturek *a*-motako trantsizio sendoak dituzte, eta euren patroiak erraz identifikatu ziren 2-6 GHz eta 8-12 GHz tarteko espektroetan. Ondoren, **I** eta **II** konformeroen *b*-motako trantsizioak ere esleitu ziren, baina ez zen *c*-motako seinale esperimentalik ikusi. Konformero egonkorrena baino gutxi gorabehera 8,9 kJ·mol^{-1} ezegonkorragoa den FA **III**

konformeroaren populazioa, Boltzmann banaketaren arabera, 383 K-eko tenperaturan % 3-koa zela iradoki zen. Bere *a*-motako dipolo momentu handiari esker, **III** konformeroaren *a*-trantsizioak ere esleitu ziren. Espektroaren adibide batzuk VII.2. irudian ikus daitezke. **I** konformeroa baino 20 $kJ \cdot mol^{-1}$ ezegonkorragoa den FA IV konformeroa, ez zen esperimentalki detektatu, ziurrenik bere ugaritasun eskasagatik, ~ % 0,1. *Watson*-en *S* hurbilketaren bidezko Hamiltondar murriztu zurruna erabiliz, **I**, **II** eta **III** konformeroen errotazio espektroak doitu ziren *Pickett* programa erabilita.[39] Lortutako parametro espektroskopikoak ere VII.1 taulako bigarren atalean daude adierazita

VII.1 taula Lau FA konformeroentzako energia erlatiboak (ΔE_e) *ZPE* zuzenketaren bidez lortutako energiak (ΔE_0), *Gibbs*-en energia 383 K-tan ($\triangle G_0$) ($kJ \cdot mol^{-1}$ unitateak erabilita kasu guztietarako), errotazio konstanteak (MHz-tan), eta dipolo momentu elektrikoak (Debye-etan). Balio esperimentalak taulako beheko atalean daude adierazita.

	I	**II**	**III**	**IV**
$\triangle E_e$	0	0,6	9,1	25,2
$\triangle E_0$	0	0,6	8,8	23,8
$\triangle G_0$	0,0	0,5	8,9	21,2
A	5270	5271	5195	5213
B	1639	1634	1653	1633
C	1250	1247	1254	1246
μ_a	2,0	1,9	4,9	4,4
μ_b	0,8	2,2	1,7	3,4
μ_c	0,0	0,0	0,0	0,3
		Experimental[a]		
A	5239,81504(73)	5238,51317(86)	5167,20252(91)	-
B	1638,51689(23)	1633,13797(27)	1650,88707(27)	-
C	1248,75945(23)	1245,55338(26)	1251,65659(25)	-
μ[b]	$\mu_a > \mu_b$ no μ_c	$\mu_a \approx \mu_b$ no μ_c	$\mu_a > \mu_b$ no μ_c	-
N[c]	35	38	28	-
σ[c]	5,6	7,1	5,0	-
Ab%[c]	51	47	2	-

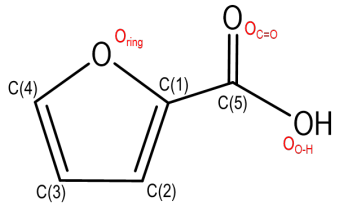
[a] Doiketak distortsio zentrifugorik gabe egin ziren sistema nahiko zurruna delako. [b] Estimatutako trantsizio dipolo momentuak. [c] N doitutako lerro kopurua; σ doiketaren desbideratze estandarra da eta Ab% estimatutako konformero bakoitzaren populazioaren ehunekoa.



VII.2. irudia FA molekularen espektroaren zatia eta doitutako konformeroen espektro simulatua. Simulazioa egiteko errotazio tenperatura 1 K dela suposatu da; dipolo momentuetarako balio teorikoak erabili dira eta azkenik ugaritasun esperimentalen balioak ere erabili dira. Xehetasun guztiak testuan ikus daitezke.

Espektro esperimentalean lortutako seinale/zarata ratio altuak, **I** eta **II** konformeroen ia ^{13}C-isotopologo mono-ordezkatu guztiak bere ugaritasun naturalean (% 1,1) esleitzeko aukera eman zuen. Salbuespen bakarra **I** konformeroaren ^{13}C(1) isotopologoa izan zen, zeinaren trantsizio errotazionalak, funtsean, jatorrizko konformeroaren trantsizioez gainjarrita dauden, izan ere karbono hori masaren zentroko erdigunetik oso hurbil baitago, eta ezinezkoa da **I** konformeroaren ^{13}C(1)-ren errotazio trantsizioak detektatzea. Horrez gain, **I**, **II** eta **III** konformeroen protoi azidoaren ^{2}H isotopologoak ere esleitu ziren, lagin deuteratua erabiliz. Jatorrizko espezieak doitzeko erabili zen prozedura espektroskopiko bera jarraitu zen espezie isotopiko guztiak doitzeko. Isotopologo guztien konstante espektroskopikoak VII.2 taulan laburbiltzen dira eta espezie bakoitzarentzat neurtutako trantsizio esperimentalak eta ziurgabetasunak eranskinetako AVII.1-AVII.15 tauletan daude.

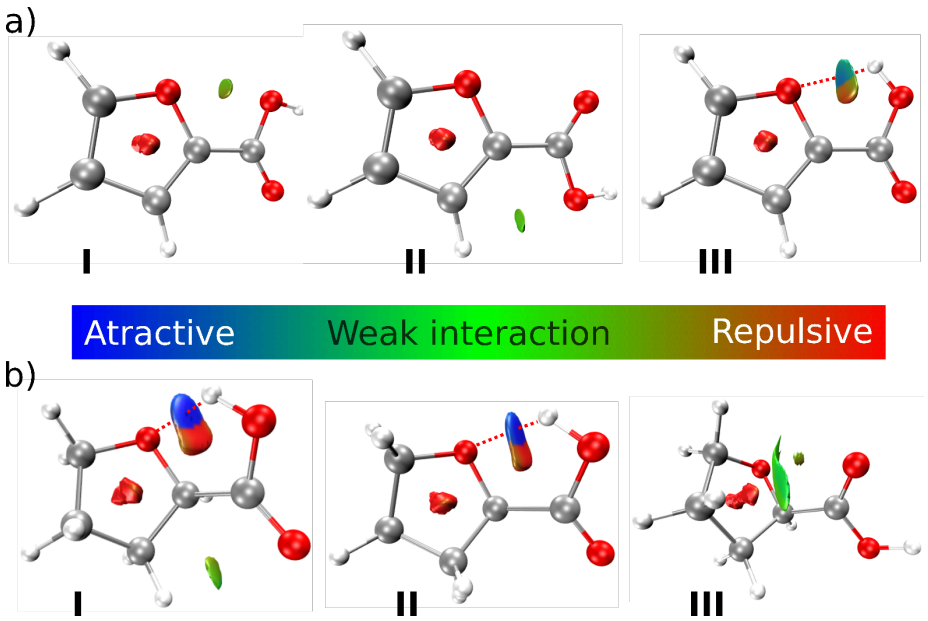
VII.2. taula I eta **II** ^{13}C eta **I**, **II**, eta **III** konformeroen D isotopologoen errotazio konstante esperimentalak.[a]



	I ^{13}C(1)	**I** ^{13}C(2)	**I** ^{13}C(3)	**I** ^{13}C(4)	**I** ^{13}C(5)
A	N/A	5162,0485(84)	5218,4353(62)	5218,124(15)	5239,4251(95)
B	N/A	1633,9486(15)	1612,2135(16)	1615,8896(28)	1628,4613(17)
C	N/A	1241,7767(14)	1232,2341(12)	1234,0719(22)	1242,8880(12)
N	N/A	7	9	6	9
σ	N/A	5,9	7,0	10,9	7,4
	II ^{13}C(1)	**II** ^{13}C(2)	**II** ^{13}C(3)	**II** ^{13}C(4)	**II** ^{13}C(5)
A	5238,3374(43)[d]	5164,2378(20)	5213,2807(28)	5215,2948(25)	5238,3619(27)
B	1633,2103(11)	1629,14021(47)	1607,67689(60)	1609,88263(56)	1623,24533(57)
C	1245,5941(15)	1238,99845(38)	1229,29041(68)	1230,69401(52)	1239,78259(56)
N	8	15	13	14	15
σ	5,3	4,6	6,2	5,7	6,3
	I D at OH	**II** D at OH	**III** D at OH		
A	5185,3218(16)	5195,36495(89)	5006,21657(98)		
B	1596,18967(42)	1588,32847(28)	1642,66320(30)		
C	1221,04512(37)	1216,98884(26)	1237,34286(28)		
X_{aa}	0,2641(33)	0,2638(30)	0,0753(60)		
X_{bb}	[-0,1164][b]	[-0,1097]	0,0925(61)		
X_{cc}	[-0,1476]	[-0,1531]	-0,1679(61)		
N	57	84	76		
σ	6,1	6,7	6,0		

[a] *A*, *B* eta *C* errotazio konstanteak MHz-tan adierazten dira. *N* doitutako trantsizio kopurua da. *σ* doiketaren desbideratze estandarra da. Balio zentrifugoak isotopologo ugarienaren balio esperimentaletan finkatuak izan dira. [b] Balio teorikoan finkatua.

Isotopologoen errotazio konstanteen doiketak, FA **II** konformeroaren karbono katearen ordezkapen egitura (r_s) eta egitura eraginkorra (r_0) deritzon molekularen egitura zehaztera ahalbidetu dute. Horretarako, *Kraitchman*-en ekuazioak eta masaren mendekotasun egitura molekularren teoria hurrenez hurren, erabili dira. Horretaz gain, **II** konformeroaren orekako egitura semi-empirikoa (r_e^{SE}),[40–43] errotazio konstanteetan bibrazio kontribuzio kalkulu teoriko anhamonikoen bidez kalkulatu zen. Lortutako koordenatuak eta egituren parametro estrukturalak AVII.16-AVII.19 tauletan laburbildu dira. Oinarrizko egoerako inertzia defektu esperimentala jatorrizko espeziarentzat Δ_0 = -0,1799 uÅ^2 da, B3LYP-D3(BJ)/def2-TZVP kalkulu anharmonikoen errotazio-bibrazio zuzenketaren ondoren balio hori -0,0201 uÅ^2 bihurtzen da r_e^{SE} egituraren analisian, molekula lauentzat esperotako 0 uÅ^2 baliotik oso hurbil dagoena.



VII.3. irudia Interakzio ez kobalenteen NCI analisia non a) detektaturiko hiru FA konformeroak irudikatu diren, **I**, **II**, eta **III**; eta b) aurreko lan batean doitutako hiru THFA konformeroak **I**, **II**, and **III** adierazi diren. NCI iso-gainazala (s = 0.60) B3LYP-D3(BJ)/def2-TZVP maila teorikoan kalkulatu da.

Teorikoki lortutako dipolo momentuko osagaiak eta 1 K-eko errotazio tenperatura zenbatetsiz, **I**, **II** eta **III**-ren populazio esperimentala % 51, % 47 eta % 2 zela lortu zen,[27,44] zeina bat dator FA jatorrizko 383 K-ko tenperaturaren eta haien energia libre erlatiboaren arabera kalkulatutako Boltzmannen distribuzio portzentaiarekin, %52, %45 eta %3. **II** eta **III**-ren populazio ehuneko balio esperimental txikiagoak, agian, hozte efektu txikiren bat iradokitzen du, nahiz eta ziurgabetasun esperimentalak ehuneko balio horien azalpen egokia eman dezakeen.

Helarazitako emaitzek, FA-ren *cis*-COOH konformeroa *trans*-COOH konformeroa baino askoz ere faboratuagoa dagoela frogatzen dute, zeinek THFA-ren alderantzizko joera duen.[7] Erakutsitako emaitzak azido karboxiliko gehienetan ohikoak dira, batez ere COOH taldea egonkortzeko H-lotura intramolekular indartsurik edo kontaktu esturik ematen ez denean. *Cis*-COOH konformeroaren nagusitasunari *Z*-efektua deritzo, nahiz eta bere jatorria oraindik eztabaidatzen ari den.[45,46] FA eta THFA-ren elkarrekintza intramolekularrak irudikatzeko, elkarrekintza ez-kobalente (*non-covalent interaction,* NCI) analisia egin zen eta emaitzak VII. irudian jasotzen dira.[47] Argi dago, OH···O_{ring} elkarrekintza askoz ere sendoagoa dela THFA **I** eta **II** koformeroetan FA **III**-n baino, THFA **I** eta **II-ren** *trans*-COOH konformazioaren lehentasunaren erantzulea dela.

O-H···O_{ring} interakzioarekin erlazionatutako energia kuantifikatzeko, NBO analisiak ere erabili genituen non H-lotura baten (edo kontaktu baten) energia lor daitekeen elkarrekintza horrekin zuzenean erlazionatuta dauden orbitalen energia ezagutuz, hau da, O-H···O_{ring} orbitalen n_O eta σ^*_{OH}, erabiliz.[48] Emaitzen arabera (eranskinetako AVII.20. taula), THFA **I** eta **II** konformeroetan, OH···O_{ring} H-lotura energia 19 eta 21 kJ·mol^{-1} bitartekoa da, eta aldiz balio hau FA **III**-ren kasuan, 2,4 kJ·mol^{-1}-ra jaisten da. *Lotura Sendotasun Intrintsekoaren Indizea* delakoak (*Intrinsic Bond Strength Index, IBSI*) ere joera hori erakusten zuten, izan ere *trans*-COOH THFA **I** eta **II** emaitzek 0,047-0,049 IBSI balioa duten, eta aldiz FA **III**-ren kasuan, balio hori 0,025 da.

Orokorrean, FA monomeroaren hiru konformero esperimentalki detektatu ziren egindako esperimentuan, aurretik IR matrize isolatuan egindako azterketarekin bat datorren emaitza lortuz.[6] Desberdintasuna, errotazio espektrometroaren sentikortasun handiari esker, *trans*-COOH duen FA **III** konformeroa zuzenean IR laser ponpaketarik gabe detektatu daitekeela da.[6] Isotopologoen datuen bidez lortutako orekako egitura erdi esperimentalak, FA **II** konformeroa oreka egoeran laua dela erakusten du. FA eta bere homologo hidrogenatuak, THFA, OH···O_{ring} elkarrekintza oso ezberdinak erakusten dituzte eta horren ondorioz konformazio lehentasun oso desberdina sortzen da.

FA Dimeroa

FA **I**, **II** eta **III**-ren trantsizio ezagun guztiak eta detektatutako monomeroen isotopologo mono-ordezkatu guztien seinaleak kendu ondoren, espektro garbian FA dimeroaren presentzia iradokitzen zuen patroi ahulago bat agerian zegoen. Aurreko THFA dimeroaren ikerketan,[10] konformeroen arteko interkonbertsio energia langa batzuk, adibidez *cis* eta *trans*-COOH interkonbertsioarekin erlazionatutakoak daudenak, zailak dira gainditzen CREST bilaketa konformazionalean zehar. Horregatik, hasierako FA geometria konformazional ezberdinak erabili ziren FA dimeroaren bilaketa konformazionala egiteko. Dimeroak izendatzeko, monomero azpiunitateen konformazioan oinarritu ginen. VII.3. taulak kalkulatutako bost dimero egonkorrenen energia erlatibo eta konstante espektroskopikoak erakusten ditu. Interesgarria da ikustea nola FA interakzio intramolekularrak **II** konformeroaren alde faboratzen direla, **II-II** konformeroa **I-I** baino egonkorragoa bihurtuz. Hiru FA dimeroen geometria egonkorrenak, H-lotura bikoitza dutenak, VII.4 irudian irudikatu dira.

Termodinamikoki, **I-I**, **I-II** eta **II-II** dimeroek egonkortasun oso antzekoak dituzte, eta **I-III** eta **II-III** dimeroekin alderatura, energetikoki oso faboratuta daude, azkenik **III-III** konformeroa ezegonkorrena izanda. **I-III** eta **II-III**-ren egiturak, eranskinetako AVII.2 irudian jaso dira. **I-I** eta **II-II** dimeroen kasuan, azpiunitate monomeriko berdinak eratutako dimeroak dira eta C_{2h} simetria aurkezte dute, σ_h plano simetriko eta inbertsio zentro batekin. Egitura horiek, ez dute dipolo momenturik eta horren ondorioz, ez da posible haien espektro errotazionala

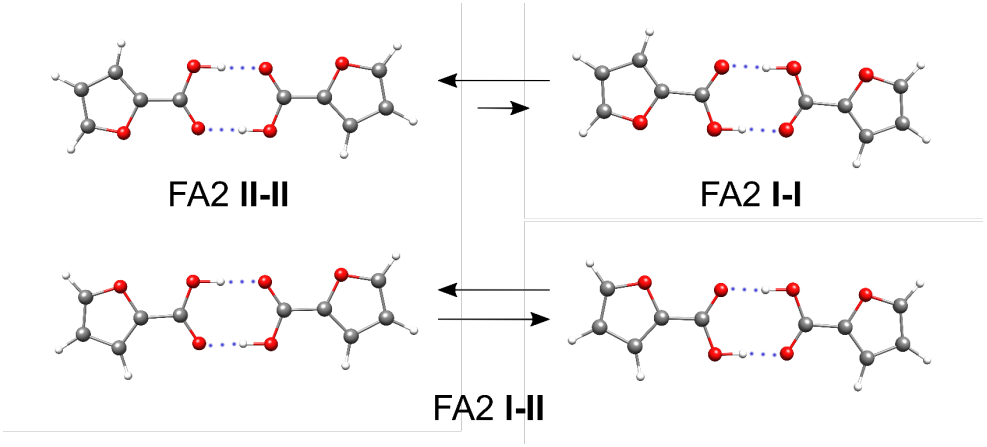
detektatzea. Protoi transferentzia bikoitza gertaerak, konformero bat beste batean bihurtuko luke, VII.3 irudiak adierazten den moduan. **I-II**, C_s motako simetria erakusten duen konformeroa da eta horren ondorioz zero ez den *b*-motako dipolo momentua dauka. Gainera protoi transferentzia bikoitz gertaerak, konformero baliokidea ematen du, VII.3 irudian ikus daitekeen moduan. Eredu termodinamikoaren arabera bakarrik **I-II** konformeroa ikustea espero izango litzateke.

Bestalde, THFA dimeroaren aurreko azterketa espektroskopikoek frogatu zuen nola faktore termodinamiko eta zinetikoek eragin garrantzitsuak dituztela esperimentalki agertzen diren dimeroetan eta noski euren ugaritasunean ere bai. [10] Zehazki, THFA-ren kasuan, termodinamikoki faboratutako dimeroak esperimentalki detektatu ziren, bi *cis*-COOH THFA monomero azpiunitaterekin eratuta zeudenak, eta aldi berean, zinetikoki faboratuak eta termodinamikoki ez faboratuak ziren *trans*-COOH eta *cis*-COOH monomero unitatez eratutako dimeroak ere detektatu ziren. Bitxiki, zinetikoki gehien faboratuta zegoen *trans*-COOH bi azpiunitatez osatutako dimeroak ez ziren *jet* hedapen ingurunean detektatu, baina konformero horiek bai detektatu zirela matrize isolatuan egindako IR eta VCD espektroetan.[8] Gure esperimentuan, prozesua guztiz zinetikoki kontrolatua izanez gero, **I-I**, **I-II**, **I-III**, **II-II**, **II-III** eta **III-III** taldeen ugaritasunak % 26, % 48, % 2, % 22, % 2 eta % 0 izatea espero zen, hurrenez hurren, **I** (% 51), **II** (% 47) eta **III.** (% 2) monomeroen ugaritasun esperimentala kontuan hartuz. Hortaz, FA dimeroaren kasuan egoera asko sinplifikatzen da, prozesu zinetiko eta termodinamikoen artean lehiarik ez baitago, hau da, zinetikak zein termodinamikak bi *cis*-COOH monomero azpiunitatez eratutako dimeroen alde daudelako, **I-I**, **II-I** eta **II-II**

VII.3. taula Bost FA dimeroen energia erlatiboak (ΔE_e), *ZPE* zuzenketaren bidez lortutako energia erlatiboak (ΔE_0), eta lotura energia erlatiboa (D_e) eta *ZPE/BSSE* zuzendutako lotura energia erlatiboak (D_0) (kJ mol^{-1} unitateak erabilita kasu guztietarako), errotazio konstanteak (MHz-tan), eta dipolo momentu elektrikoak (Debye-etan). [a]

$(FA)_2$	**II-II**	**I-I**	**I-II**	**I-III**	**II-III**
A	2614	2630	2612	1318	1298
B	167	167	167	198	199
C	157	157	157	172	173
$\mu_a/\mu_b/\mu_c$	0/0/0	0/0/0	0/1,2/0	2,1/3,2/0	2,0/2,1/0
$\triangle E_e$	0	0,4	0,3	40,9	40,1
$\triangle E_0$	0	2,1	1,2	40,4	39,6
D_e[a]	79,6	78,0	78,7	47,3	48,8
D_0[a]	74,4	75,2	74,7	25,6	27,1

[a] D_e eta D_0 kalkulatzeko erabilitako ekuazioak eranskinetako AVI.1 puntuan adierazi dira



VII.4. irudia Optimizatutako FA hiru dimero egonkorrenak B3LYP-D3(BJ)/def2-TZVP maila teorikoan. Geziek protoi transferentzia bikoitzeko bidea adierazten dute. Protoi transferentzia bikoitzean parte hartzen duten H-ak H_Z eta H_E moduan izendatu dira.

I-II konformeroaren *b*-motako errotazio trantsizioen iragarpenarekin soilik, berehala ikusi zen errotazio trantsizio asko bikotetan bikoizten zirela. Protoi transferentzia gertaeraren *spinen* pisu estatistikoa, 9/7 ezarri ziren k_a+k_c bikoitien alde v = 0 *tunel*-egoeran, eta k_a+k_c bakoitien alde v = 1 *tunel*-egoeran.[24] *Tunneling* gurutzatuko trantsizioak, beste alde batetik, ezin dira ikusi, izan ere *a*-motako dipolo trantsizioetan emango litzateke zeintzuk zero momentu dipolarra duten. 9/7 intentsitate erlatiboa oso lagungarria izan zen *tunneling* bi osagaiak desberdintzeko, izan ere egoera kitzikatuko trantsizioak jatorrizko egoerakoak baina maiztasun altuago edo baxuagoan ager baitaitezke. Orokorrean, 2-12 GHz maiztasun tartean 180 transferentzia-multzo neurtu eta doitu ziren bi egoerak batera doituz, *Pickett* programa espektroskopikoa erabiliz.[39] Horretarako, hurrengo Hamiltondar bikoitza erabili zen:

$$H = \sum_i H_i^R + \Delta E_{01} + F_{ab} \text{ x } (P_a P_b + P_b P_a) \quad (1)$$

H_i^R terminoak *i* egoerako errotazio Hamiltondarra adierazten du, ΔE_{01} v = 0 eta v = 1 egoeren arteko energia diferentzia da eta F_{ab} bi egoeren arteko errotazio eta bibrazio akoplamendu parametroa da. Doitutako sistemarako, ez dirudi distortsio konstanteen erabilera beharrezkoa denik doiketa ona lortzeko, beraz horren arabera sistema nahiko zurruna da. Lortutako balio esperimentalak VII.4 taulan adierazi dira eta trantsizio esperimentalen frekuentziak eta erroreak eranskinetako AVII.21 taulan ikus daitezke.

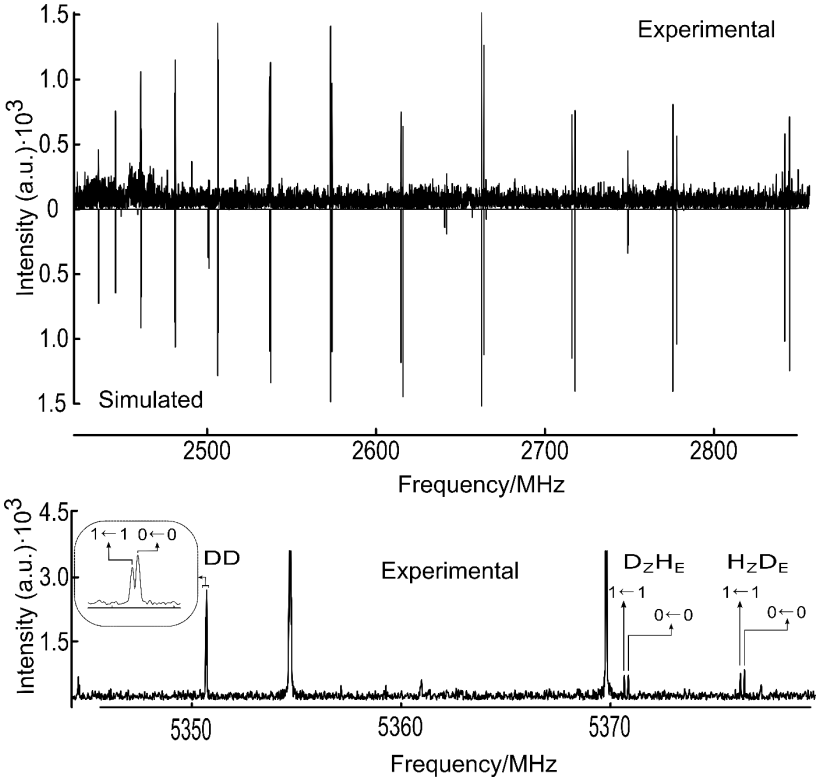
Table VII.4 I-II konformero eta isotopologo guztien errotazio konstante esperimentalak (MHz-tan).[a]

	HH	**D_ZH_E**	**H_ZD_E**	**DD**
A_0	2592,40067(24)	2577,28179(85)	2572,28060(77)	2557,34150(95)
B_0	166,70081(12)	166,46913(17)	66,475770(86)	166,22106(14)
C_0	156,691493(79)	156,43200(29)	156,418791(83)	156,17785(12)
A_1	2592,38604(17)	2577,27917(50)	2572,27583(72)	2557,32015(96)
B_1	166,69857(13)	166,46936(10)	66,475882(86)	166,21898(14)
C_1	156,689428(45)	156,431260(90)	156,41879(11)	156,17624(12)
ΔE_{01}	1056,0(10)	116,5(60)	141,2(28)	-
F_{ab}	-10,347(15)	[-10,347]	[-10,347]	-
N^a	180	89	81	104
σ	6,9	5,6	65	6,5

[a] Doiketak distortsio zentrifugo konstanterik gabe egin ziren. A, B eta C errotazio konstanteak dira, MHz unitateetan adierazita non 0 eta 1 azpi-indizeak oinarrizko egoera eta *tunneling* egoera kitzikatuko balioak adierazten dituzte hurrenez urren. ΔE_{01} *tunneling* bikoizketa balioa da MHz-tan. F_{ab} Coriolis akoplamendu konstantea da MHz-tan ere bai. N doitutako trantsizio kopurua da. σ doiketaren desbideratze estandarra kHz-tan adierazten da.

Horretaz gain, FA dimeroen espezie deuteratuak ere ikertu ziren, non *tunneling* gertaeran parte hartzen duten H atomoak ordezkatu diren. Lau isotopologo posibleak HH (jatorrizko espezierako), D_ZH_E, H_ZD_E eta DD nomenklaturaren bidez izendatu dira. Z eta E hidrogeno atomoen kokapenak VII.4 irudian adierazi dira. Banda zabaleko espektroaren zati bat eranskinetako AVII1. irudian adierazi da. Bertan, FA dimero mono eta di-deuteratuaren zenbait trantsizio ikus daitezke.

Doiketa prozedura bera aplikatu zitzaien deuteratutako espezieei, halaber, doiketetan zehar bi bitxikeria hauteman ziren. Lehenik, espezie monodeuteratuen kasuan, ΔE_{01} balioak jatorrizko espeziearenak (HH) baino askoz txikiagoak direnez, aldi bereko ΔE_{01} and F_{ab}-ren doiketak ez ziren arrakastatsuak izan. Hori dela eta, jatorrizko espeziearenean F_{ab}-ren balio finkoa espezie monodeuteratuak doitzeko erabili zen. Bigarrena, DD espeziearentzat, *tunel* seinale bikoizketak zenbait lerrotan bakarrik igertzen ziren espektro esperimentalean. Ondorioz, ezin izan genituen DD espeziearen ΔE_{01} eta F_{ab} parametro esperimentalak doitu. D espezieen kalkuluen arabera, momentu kuadrupolar nuklearra dimero deuteratu guztietarako deuseztagarria zen. Hori dela eta, DD espeziearen doiketan bakarrik oinarrizko egoerako eta tunneling egoerako konstante errotazionalak doitu ziren. Lortutako konstante esperimental guztiak AVII.4 taulan laburbildu dira eta trantsizio bakoitzaren frekuentzia eta errorea eranskinetako AVII.22-AVII.25 tauletan adierazi dira



VII.5. irudia FA dimeroaren espektro doitutako isotopomeroen espektro simulatua. Simulazioa egiteko errotazio tenperatura 1 K dela suposatu da; dipolo momentuetarako balio teorikoak erabili dira .

VII.5 taulan, aurretik ikertutako zenbait azido karboxilikoren konplexu binarioen ΔE_{01} balioak konparatu dira. *Tunneling* bandaren jatorria FA dimeroan, bestelako ikerkuntzetan lortutako baliorik altuenetakoa da, azido bentzoiko homodimeroaren[50] oso antzekoa da eta azido akriliko homodimeroaren balioa baino dezente handiagoa da.[24] Nolabait harritzekoa da, geroz eta albokatearen konjokazioa handiagoa izan, ΔE_{01} balioak ere handiagoa izateko joera duela eta ziurrenik etorkizuneko azterketa sakonago bat beharko lukeen gaia da.

Table VII.5. Zenbait azido karboxilikoren koplexu dimerikoen *tunneling* ΔE_{01} balioa (MHz).

Molekula	ΔE_{01} HH	Ref.
$(C_6H_5COOH)_2$	1114,0(10)	[50]
$(HCOOH)_2$	474(12)	[12]
$(HCOOHD)_2$	331,6(5)	[14]
2x(HCOOH) – F-bentzeno	267,6080(13)	[51]
HCOOH − HC_2COOH	291,428(5)	[16]
HCOOH − CH_3COOH	250,4442(1)	[17]
Azido akriliko dimeroa	880,6(6)	[24]
BzOH − HCOOH	548,72(6)	[23]
$(FA)_2$	1056,0(10)	Lan hau

Ondo dakigun moduan, protoi transferentzia bikoitzaren ΔE_{01} balioa, *tunneling* prozesuaren langa energetikoarekin erlaziona daitekeen parametroa da, baina tamalez, FA dimeroetan gainontzeko informazio gehiago behar da prozesuaren langa energetikoa lortzeko. Datu horiek eskuratzeko, protoi transferentzia bikoitzeko anplitude handiko mugimenduak aztertzeko garatutako prozedura konputazional sinplea erabili dugu.[26] Kodeak, masa efektibo eta anplitude handiko mugimenduaren energia potentzialaren ganazalaz sortutako dimentsio bakarreko *Schrodinger* ekuazioa ebazten du. FA **I-II** dimeroarentzat, protoi transferentzia bikoitzeko energia potentzial gainazala ekorketa erlaxatuaren bidez egin zen, aldi berean erreakzioan parte hartzen duten bi O-H···O distantzia aldatuz. Ekorketa egiteko, *Orca 4* softwarea erabili zen.[38] Metodo hori aurretik azido formikoan protoi transferentzia bikoitzeko *tunneling* bikoizketa ikertzeko erabili da.[26] Bigarren urrats batean, protoi transferentzia dinamika kalkulatu zen aurretiaz argitaratutako *Python* kode publiko bat erabiliz.[26] Prozedura hau orain dela gutxi argitaratutako azido formiko eta fluorobentzeno molekulen arteko klusterrean erabili da.[51] Erreakzio koordenatua hurrengo moduan adierazi da:

$$\xi = \big(r(O-H) + r(O-H) - r(O\cdots HO) - r(O\cdots HO)\big)/2\sqrt{2} \quad (2)$$

DFT metodo desberdinen bidez lortutako ΔE_{01} balioak (*tunneling*-aren bikoizketa energia balioak), E_{HB} (energia langa), and E_{HB}^{Eff} (ZPE zuzendutako energia langa) VII.6. taulan adierazi dira. Lehenengo begirada batean, lortutako E_{HB}, eta E_{HB}^{Eff} eta ondorioz lortzen diren ΔE_{01} balioak, erabilitako metodoaren dependentzia handia duela ikus daiteke, eta lortutako balioen zehaztasuna nahiko kaskarra da balio esperimentalak interpretatzeko. Teoria eta esperimentuen arteko antzeko ziurgabetasuna ere detektatu da beste azido batzuen protoi transferentzia bikoizketetan, hala nola azido formiko dimeroetan eta azido malonikoan, azido formiko dimeroak fluorobentzenoarekin eta baita ere anplitude handiko tiofenol dimeroen mugimenduen doiketan ere bai.[14,51,52] Orokorrean, kalkulu hauek erreakzio edo mekanismoen joerak nahiko ondo iragar ditzakeen arren, espektroskopiak eskaintzen dituen emaitza oso zehatzak oso zailak dira konputazionalki erreproduzitzen.

VII.6. taula ΔE_{01} (tunneling bikoizketa) balio teorikoa, E_{HB} (energia langa), and E_{HB}^{Eff} (ZPE zuzendutako energia langa), and E_{HB}^{Eff} (eskalatua) kalkulatutata zenbait funtzional desberdinekin, guztietan def2-TZVP base matematikoa erabilita.[a]

Funtzionala	ΔE_{01}	E_{HB}	E_{HB}^{Eff}	E_{HB}^{Eff} (eskalaktua)[b]
b1lyp	$4{,}80\cdot10^{-1}$	2289	1108	360
b3lyp	$1{,}89\cdot10$	2043	776	394
b3lyp-gd3bj	$1{,}90\cdot10^{2}$	2016	629	456
b97-3c	$3{,}25\cdot10^{3}$	1554	384	492
cam-b3lyp	$7{,}77\cdot10$	1882	632	406
cam-b3lyp-gd3bj	$2{,}76\cdot10^{2}$	1882	563	442
mPW1LYP	1,99	2220	1008	391
x3lyp	$7{,}79\cdot10$	2018	709	447
wB97	$1{,}00\cdot10^{-2}$	2936	1486	333

[a] ΔE_{01} MHz-tan eta beste energia balioak cm^{-1} unitateetan.
[b] E_{HB}^{Eff} (eskalatua) energia potentzial gainazal eskalatuarekin lortutako balioa tunneling bikoizketa balio esperimentala lortzeko. Xehetasun guztiak testuan azalduta

Oran dela gutxi, Medel-ek zenbait *tunneling* bikoizketa azaltzeko eredu sinple desberdinak laburbildu zituen, eta demostratu zuen nola *tunneling* bikoizketak azaltzeko parametrorik garrantzitsuena energia langa dela.[53] Hori ikusita, gure kasuan, teorikoki protoi transferentziarako lortutako energia potentzial gainazala zenbait faktorerekin eskalatu eta dimentsio bakarreko *Schrodinger* ekuazioa ebatzi genuen esperimentalki lortutako *tunneling* bikoizketa energiaren balioa lortzeko. Nolabait harrigarria iruditu zitzaigun maila teoriko diferenteak erabiliz lortutako E_{HB}^{Eff} balioak nahiko konsistenteak zirela euren artean eta guztiak 333-492 cm^{-1} bitartean zeudela. Lortutako emaitzak VII.5 taulan adierazi dira. Eskalatu aurretik balio hurbilena cam-b3lyp-gd3b funtzionalarekin lortu zen eta eskalatu ondoren lortutako E_{HB}^{Eff} balioan metodo horrekin gutxi gorabehera 442 cm^{-1}, zein azken finean aurkeztutako protoi transferentzia bikoitzaren erreakzio langa hoberen aurresaten duen baliotzat har daitekeen.

VII.4 Ondorioak

FA monomeroaren eta dimeroaren egitura eta protoi transferentzia bikoitzeko gertaeraren propietateak, jet hedapeneko espektroskopia eta kimika kuantikoaren kalkuluak erabiliz ikertu ziren. Monomeroaren hiru konformero esperimentalki identifikatu ziren, *cis*-COOH konfigurazioaren zaletasuna ikusiz. Energia altua duen *trans*-COOH FA **III** konformeroa ere detektatu eta identifikatu zen, aurreko IR esperimentu matrizearen aurrean, non **III** laser kitzikapenaren laguntzarekin bakarrik detektatu zen. **I-II** heterodimeroaren errotazio trantsizioek protoi transferentzia bikoizketa erakusten zuten, eta bere analisiak ΔE_{01} balio esperimentala 1056,0 (12) MHz-tan doitu zen. Orain dela gutxi garatutako dimentsio bakarreko eredua *tunnelling* bikoizketak ebazteko, protoi transferentzia bikoitzeko langa energetikoaren E_{HB}^{Eff} balioa 442 cm^{-1}-koa dela kalkulatu zen, horretarako ΔE_{01} balio esperimentala erabili zelarik. Oro har, teoriaren maila ezberdinek *tunnelling* bikoizketa nahiko desberdinak igortzen dituzte, hau da, garapen teoriko gehiago beharko lirateke protoi transferentzia bikoitzaren datu esperimental zehatzak lortzeko.

VII.5 Erreferentziak

[1] K. Ahmed, *Front. Endocrinol. (Lausanne)*. **2011**, *2*, 1–12.
[2] W. C. Keene, J. N. Galloway, *Atmos. Environ.* **1984**, *18*, 2491–2497.
[3] C. J. Gilmore, P. R. Mallinson, J. C. Speakman, *Acta Cryst.* **1983**, *1983*, 1111–1113.
[4] H. T. Flakus, M. Jabłońska, J. Kusz, *Vib. Spectrosc.* **2009**, *49*, 174–182.
[5] H. Ghalla, N. Issaoui, M. V. Castillo, S. A. Brandán, H. T. Flakus, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2014**, *121*, 623–631.
[6] A. Halasa, L. Lapinski, I. Reva, H. Rostkowska, R. Fausto, M. J. Nowak, *J. Phys. Chem. A* **2015**, *119*, 1037–1047.
[7] F. Xie, X. Ng, N. A. Seifert, J. Thomas, W. Jäger, Y. Xu, *J. Chem. Phys.* **2018**, *149*, 224306.
[8] Y. Yang, J. Cheramy, Y. Xu, *ChemPhysChem* **2021**, *22*, 1336–1343.
[9] F. Xie, S. Mahendiran, N. A. Seifert, Y. Xu, *Phys. Chem. Chem. Phys.* **2021**, *23*, 3820–3825.
[10] F. Xie, N. A. Seifert, W. Jäger, Y. Xu, *Angew. Chemie - Int. Ed.* **2020**, *59*, 15703–15710.
[11] R. Srivastava, *Front. Chem.* **2019**, *7*, 1–17.
[12] M. Ortlieb, M. Havenith, *J. Phys. Chem. A* **2007**, *111*, 7355–7363.
[13] K. G. Goroya, Y. Zhu, P. Sun, C. Duan, *J. Chem. Phys.* **2014**, *140*, 164311.
[14] W. Li, L. Evangelisti, Q. Gou, W. Caminati, R. Meyer, *Angew. Chemie - Int. Ed.* **2019**, *58*, 859–865.
[15] A. M. Daly, P. R. Bunker, S. G. Kukolich, *J. Chem. Phys.* **2010**, *132*, 201101.
[16] A. M. Daly, K. O. Douglass, L. C. Sarkozy, J. L. Neill, M. T. Muckle, D. P. Zaleski, B. H. Pate, S. G. Kukolich, *J. Chem. Phys.* **2011**, *135*, 154304.
[17] M. C. D. Tayler, B. Ouyang, B. J. Howard, *J. Chem. Phys.* **2011**, *134*, 054316.
[18] Q. Gou, G. Feng, L. Evangelisti, W. Caminati, *Chem. Phys. Lett.* **2014**, *591*, 301–305.
[19] A. Macario, S. Blanco, J. Thomas, Y. Xu, J. C. López, *J. Phys. Chem. A* **2019**, *123*, 6772–6780.
[20] A. Macario, S. Blanco, J. Thomas, Y. Xu, J. C. López, *Chem. - A Eur. J.* **2019**, *25*, 12325–12331.
[21] J. Thomas, M. J. Carrillo, A. Serrato, F. Xie, W. Jäger, Y. Xu, W. Lin, *Mol. Phys.* **2019**, *117*, 1193–1199.
[22] J. Thomas, M. J. Carrillo, A. Serrato, W. Lin, W. Jäger, Y. Xu, *J. Mol. Spectrosc.* **2017**, *335*, 88–92.
[23] L. Evangelisti, P. Écija, E. J. Cocinero, F. Castaño, A. Lesarri, W. Caminati, R. Meyer, *J. Phys. Chem. Lett.* **2012**, *3*, 3770–3775.
[24] G. Feng, L. B. Favero, A. Maris, A. Vigorito, W. Caminati, R. Meyer, *J. Am. Chem. Soc.* **2012**, *134*, 19281–19286.
[25] Y. Litman, J. O. Richardson, T. Kumagai, M. Rossi, *J. Am. Chem. Soc.* **2019**, *141*, 2526–2534.
[26] D. S. Tikhonov, *Struct. Chem.* **2021**.
[27] F. Xie, N. A. Seifert, M. Heger, J. Thomas, W. Jäger, Y. Xu, *Phys. Chem. Chem. Phys.* **2019**, *21*, 15408–15416.
[28] N. A. Seifert, J. Thomas, W. Jäger, Y. Xu, *Phys. Chem. Chem. Phys.* **2018**, *20*, 27630–27637.
[29] C. Pérez, S. Lobsiger, N. A. Seifert, D. P. Zaleski, B. Temelso, G. C. Shields, Z. Kisiel, B. H. Pate, *Chem. Phys. Lett.* **2013**, *571*, 1–15.
[30] G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman, B. H. Pate, *Rev. Sci. Instrum.* **2008**, *79*, 053103-1-053103–13.
[31] Y. Xu, W. Jäger, *J. Chem. Phys.* **1997**, *106*, 7968–7980.
[32] S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

[33] S. Grimme, C. Bannwarth, P. Shushkov, *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
[34] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 16, Rev. C. 01* **2016**.
[35] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *123*, 0–9.
[36] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
[37] A. Schäfer, H. Horn, R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571–2577.
[38] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, *J. Chem. Phys.* **2020**, *152*, 224108.
[39] H. M. Pickett, *J. Mol. Spectrosc.* **1991**, *148*, 371–377.
[40] J. K. G. Watson, A. Roytburg, W. Ulrich, **1999**, *119*, 102–119.
[41] J. Kraitchman, *Am. J. Phys.* **1953**, *21*, 17–24.
[42] Z. Kisiel, *J. Mol. Spectrosc.* **2003**, *218*, 58–67.
[43] P. Groner, R. D. Warren, *J. Mol. Struct.* **2001**, *599*, 323–335.
[44] G. T. Fraser, R. D. Suenram, C. L. Lugez, *J. Phys. Chem. A* **2000**, *104*, 1141–1146.
[45] D. Ferro-Costas, R. A. Mosquera, *Phys. Chem. Chem. Phys.* **2015**, *17*, 26946–26954.
[46] C. H. Görbitz, M. C. Etter, *J. Am. Chem. Soc.* **1992**, *114*, 627–631.
[47] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
[48] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899–926.
[49] J. Klein, H. Khartabil, J. C. Boisson, J. Contreras-Garciá, J. P. Piquemal, E. Hénon, *J. Phys. Chem. A* **2020**, *124*, 1850–1860.
[50] I. Kalkman, C. Vu, M. Schmitt, W. L. Meerts, *ChemPhysChem* **2008**, *9*, 1788–1797.
[51] W. Li, D. S. Tikhonov, M. Schnell, *Angew. Chemie - Int. Ed.* **2021**, *60*, 25674–25679.
[52] R. T. Saragi, M. Juanes, C. Pérez, P. Pinacho, D. S. Tikhonov, W. Caminati, M. Schnell, A. Lesarri, *J. Phys. Chem. Lett.* **2021**, *12*, 1367–1373.
[53] R. Medel, *Phys. Chem. Chem. Phys.* **2021**, *23*, 17591–17605.

IX. kapitulua

Ondorioak

IX.1 Ondorioak

Lan honetan, nire doktoretzan egindako lan batzuk aurkeztu dira. Ikerketan, interes biologikodun hainbat molekularen konformazioren karakterizazioan oinarritu zen, monosakaridoetatik hasita eta hainbat funtzio biologikotan inplikatuta dauden beste molekula batzuekin jarraituz. Zehazki, molekula eta agregatu molekular batzuen egituraren zehaztapen espektroskopikoa, errotazio espektroskopia eta kimika kuantikoaren kalkuluen konbinazioaren egin zen. Nire lana mikro uhinen espektrometroa pultsu multifrekuentzia Fourierren transformatuarekin eta Fabry-Pérot-en barrunbean eta Fourierren transformatuan oinarritutako multifrekuentzia espektrometroa, erabili, mantenduz eta garatuz egin da. Lehen aipatu den bezala, tesia bi sekziotan banatzen da. Lehenengo zatian monosakarido batzuk aztertzen dira, eta azukreen termolabilitatea dela eta, zati honetako esperimentu guztiak laser ablazio bidezko lurrun sistemarekin hornituta dagoen Fabry-Pérot-en (cav-FTMW) barrunbea erabiliz egin ziren. Monosakaridoak oso antzeko molekulak diren arren, beren OH taldeen eraldaketak beren funtzioak erabat aldatzen ditu.

III. kapituluan, eritrulosa molekularen azterketa aurkeztu da. Eritrulosa, lau karbono dituen zetona azukrea da, eta egitura lineala aurkezten du gas fase, solido eta disoluzioan. Murchison-go meteoritoetan eta IEG-ko hainbat eskualdetan azukrearekin lotutako hainbat molekula detektatu ondoren, laginak prestatzeko metodo berri bat garatu genuen, gure mikro uhin espektrometoroan eritrulosaren espektro errotazionala lortu ahal izateko. Lortutako emaitza esperimentalekin, eritrulosa IEG-ko hainbat guneetan bilatu da. Detekzioak arrakastarik izan ez bazuen ere, lortutako datuak funtsezkoak izango dira etorkizunean IEG-n bilaketak egiteko.

IV-VI. kapituloetan, glukosa, galaktosa eta manosaren zenbait deribaturen ikerketak egin dira. Glukosa, galaktosa eta manosaren azterketa, orain dela urte batzuk egin zen espektroskopia errotazionala eta kimika kuantikoaren konbinaketaren bidez. Hala ere, tesi honetan aurkeztutako deribatuak ez dira aurretiaz ikertu erabilitako teknikarekin.

IV. kapituluan, glukosa, galaktosa eta manosan OH eta F-ren arteko ordezkapen bio-isosterikoen eta dihidroxilazioaren ondoriozko ematen den konformazio eraldaketak aurkeztu da. Nahiz eta kapitulu hau bereziki espektroskopia errotazionalaren bidezko analisian oinarrituta dagoen, lan honetan aurkeztutako zortzi molekulak emaitza teorikoen eta hainbat metodo espektroskopiko desberdinen konbinazioaren bidez ikertu dira. Oro har, glukosa eta galaktosaren kasuan, OH eta F arteko ordezkapen bioisoterikoek ez dutela efektu berezirik hidrogeno loturen orientazioan eta hidroximetilo konponketetan frogatzen dugu. Hala ere, manosaren kasuan bigarren karbonoko hidroxilo taldea fluor atomoaz ordezkatzen badugu, jatorrizko manosa eta manosa fluoratuaren artean H lotura-sarearen norabidea guztiz aldatzen da. Bestalde, glukosa eta galaktosaren bigarren karbonoan hidroxilo taldea hidrogenoaren ordezkapekak,

analogo hidroxilatu eta fluoratuekin konparatuz, H lotura sarea erlojuaren orratzen noranzkoan egonkortzen dute. Emaitzek azukre ez-naturalen konformazio eraldakak erakusten dituzte, eta emaitza horiek oso baliagarriak dira azukre naturalen aldaketa txikiek konformazio aldaketa oso altuak eragin ditzaketela ulertzeko.

V. kapituluak glukosatik eratorritako laktona azukreen analisi espektroskopikoa erakusten du. Emaitzetan, glukonorolaktonak eta glukosak hiru dimentsioko forma espazial erabat desberdina dutela ikusi dugu. Glukoronolaktonaren molekulak, bi furano eraztunez osatuta dago gas fasean, eta hidroxilo taldeak ez daude euren artean H lotura intramolekularren bidez konektatuta, azukreetan ohikoa den moduan. Bestalde, glukonolatonak glukosaren antz zamarra du, H lotura sare kooperatibo bereizgarriarekin eta hidroximetiloaren hiru konformazio egonkorrekin. Hala ere, eraztunaren paketatzea eta konformeroen populazioa desberdinak dira, kabono anomerikoaen hibridazio desberdinagatik ($sp^3 \rightarrow sp^2$).

Glukosaren eta galaktosaren beste deribatu ohiko batzuk talde hidroximetiloaren oxidazioaren ondorioz eratzen diren azido karboxiliko azukreak dira. Glukosaren eta galaktosaren hidroximetiloaren oxidazioak, azido glukuroniko eta galakturonikoak eratzen ditu. Molekula horien azterketa, **VI. kapituluan** jasotzen da. Ikerketa honetan, berkristalizazioa, egoera solidoko erresonantzia magnetiko nuklearra konbinatzen ditugu azido glukuroniko eta galakturonikoen bi anomeroen presentzia lortu eta egiaztatzeko. Azido glukuronikoaren kasuan, gas fasean anomero bakoitzaren konformazio bakarra detektatu zen, hidrogeno lotura antolamendu berbera duena. Aitzitik, azido galakturonikoan H lotura sarearen orientazioa anomero bakoitzarentzat kontrakoa da. Kapitulu honetan frogatzen dugu talde hidroximetiloa azido karboxiliko bihurtzeak erabat aldatzen duela azukreen egitura, haien eratorriekin alderatuta.

Tesi honen bigarren zatian, **VII. eta VIII. kapituluetan**, pultsu multifrekuentzia eta Fourierren transformatuan oinarritutako espektrometroan egindako hainbat ikerketa erakusten dira.

Atal honetako lehen kapitulua, **VII. kapitulua**, Albertako Unibertsitateko Xu irakaslearen laborategian egindako ikerketa aurkezten du. Kapitulu honetan, azido furoiko monomer eta dimeroaren kasua erakusten dugu. Gas-fasean, hiru monomero eta dimero bat detektatu genituen, non protoien transferentziaren ondorioz ematen den degenerazioa esperimentalki detektatu zen. Protoi transferentzia prozesuaren energia estimatzeko, kimika konputazionaleko kalkuluak erabili ditugu eta lortutako emaitzak aurretik beste dimeroetan lortutako balioekin alderatu ditugu. Bitxia bada ere, energia oso altua ikusten dugu protoi bikoitzaren transferentziarako, eta ziurrenik tunel efektuko energia handiagoa dela iradokitzen dugu karbonilo taldeak eraztun aromatikoarekin konektatuta daudenean.

Azkenik, **VIII. kapituluan**, oraindik egiten ari diren bi ikerketa laburbiltzen ditugu. Lan hauetako parte esperimentala, Euskal Herriko Unibertsitateko pultsu multifrekuentzian eta Fourierren transformatuan oinarritutako espektrometroan egin ziren. Zehazki, metil bentzoatoaren mikro-solbatazioan zentratzen gara, eta bertan uraren autoagregazioak lau molekula edo gehiagoz osatutako klusterretan duen garrantzia ikusten dugu. Bestalde, 4-oxobutanonitriloaren azterketaren lehen urratsak ere erakutsi ditugu, IEG-n egon liteken eta amino azidoen presentziarekin lotuta dagoen molekula.

Laburbilduz, espektroskopia errotazionala teknika moldakorra dela monosakaridoek gas fasean dituzten propietateak aztertzeko frogatu dugu. Nahiz eta lehen monosakaridoak orain dela urte batzuk teknika horrekin neurtu ziren, gas faseko egiturek funtsezko informazioa ematen dute oraindik, konformazio banaketa gobernatzen duten arauak ulertzeko. Bestalde, pultsu multifrekuentzia espektrometro konfigurazio berriaren eraikuntza arrakastatsua metil bentzoato molekula bost ur molekulez osatutako klusterrarekin detektatzean frogatu da, espektroskopia errotazionalaren bidez detektatutako ur konplexurik handienetakoa baita. Azkenik, azido furoikoko dimeroen protoien transferentzia zabaltasun handiko mugimendua doituz, espektro konplexuekin lan egiteko gaitasuna erakusten dut.

Chapter X

Appendix of other concluded studies

Tn antigens

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Cite This: J. Am. Chem. Soc. 2018, 140, 9952−9960
Article
pubs.acs.org/JACS

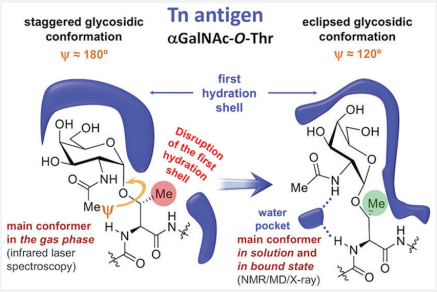
Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition

Iris A. Bermejo,[†,□] Imanol Usabiaga,[‡,□] Ismael Compañón,[†,□] Jorge Castro-López,[§] Aran Insausti,[‡,¶] José A. Fernández,[‡] Alberto Avenoza,[†] Jesús H. Busto,[†] Jesús Jiménez-Barbero,[∥,⊗,▽] Juan L. Asensio,[⊥] Jesús M. Peregrina,[†] Gonzalo Jiménez-Osés,[†] Ramón Hurtado-Guerrero,[*,§,#] Emilio J. Cocinero,[*,‡,¶] and Francisco Corzana[*,†]

[†]Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain
[‡]Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV-EHU), 48080 Bilbao, Spain
[§]Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), 50018 Zaragoza, Spain
[∥]CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain
[⊗]Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain
[▽]Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain
[⊥]Instituto de Química Orgánica General, IQOG-CSIC. 28006 Madrid, Spain
[#]Fundación ARAID, 50018 Zaragoza, Spain
[¶]Biofisika Institute (CSIC, UPV/EHU), University of the Basque Country (UPV/EHU), 48940 Leioa, Spain

Supporting Information

ABSTRACT: The tumor-associated carbohydrate Tn antigens include two variants, αGalNAc-*O*-Thr and αGalNAc-*O*-Ser. In solution, they exhibit dissimilar shapes and dynamics and bind differently to the same protein receptor. Here, we demonstrate experimentally and theoretically that their conformational preferences in the gas phase are highly similar, revealing the essential role of water. We propose that water molecules prompt the rotation around the glycosidic linkage in the threonine derivative, shielding its hydrophobic methyl group and allowing an optimal solvation of the polar region of the antigen. The unusual arrangement of αGalNAc-*O*-Thr features a water molecule bound into a "pocket" between the sugar and the threonine. This mechanism is supported by trapping, for the first time, such localized water in the crystal structures of an antibody bound to two glycopeptides that comprise fluorinated Tn antigens in their structure. According to several reported X-ray structures, installing oxygenated amino acids in specific regions of the receptor capable of displacing the bridging water molecule to the bulk-solvent may facilitate the molecular recognition of the Tn antigen with threonine. Overall, our data also explain how water fine-tunes the 3D structure features of similar molecules, which in turn are behind their distinct biological activities.



INTRODUCTION

The Tn antigens are among the most specific human tumor-associated carbohydrate antigens (TACAs), present in approximately 90% of tumors.[1,2] In general, the aggressiveness of the carcinoma and the occurrence of these antigens has a clear direct correlation,[3] promoting their use as biomarkers and potential therapeutic targets against cancer.[4] Vaccines based on peptide fragments carrying this determinant are able to induce antibodies in mice that discriminate between normal and cancer cells, reducing in some cases, the size of the tumor and increasing the survival of the animals.[5]

In general, the Tn antigen is referred to as GalNAc α-*O*-linked to a serine or a threonine residue (**Tn-Ser** and **Tn-Thr**, respectively), without specifying which of the two amino acids the GalNAc is linked to (Figure 1a). However, despite their structural simplicity and similarity, differing only in a methyl group, they display totally different conformations in solution (Figure 1b), leading to significant biological consequences.[6−12] For instance, while anti-MUC1 antibodies recognize glycopeptides bearing a **Tn-Thr** moiety, they show very low affinity toward derivatives with the **Tn-Ser** residue.[6] On the other hand, several anti-Tn antibodies show a clear preference for glycopeptides containing the **Tn-Ser** antigen.[13] A recent study

Received: May 7, 2018
Published: July 13, 2018

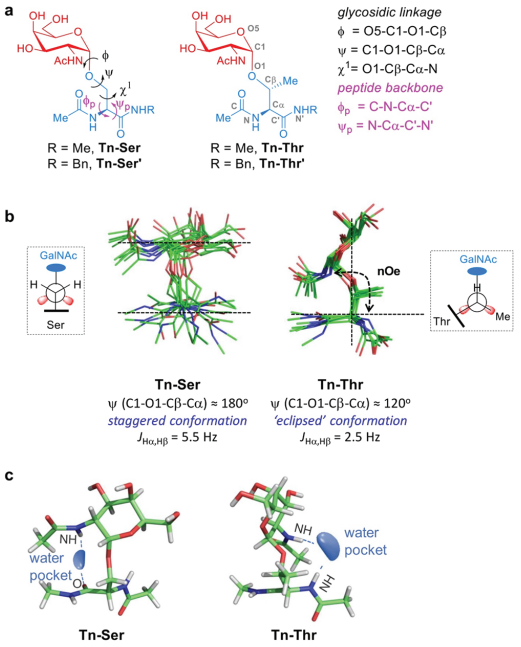


Figure 1. Conformational behavior of the antigens **Tn-Ser** and **Tn-Thr** in water. (a) The two Tn-antigens studied in this work, together with the definition of the most relevant torsional angles and atom labels. (b) Major conformations in solution for the Tn antigen with either a serine (**Tn-Ser**) or a threonine (**Tn-Thr**) derived from experiment-guided molecular dynamics (MD) simulations.[7,8] The values of the ψ torsion angle of the glycosidic linkage and of the associated coupling constant $J_{H\alpha,H\beta}$ are shown. The Newman projections for the $C\beta-O1$ bonds are also given, showing the staggered (**Tn-Ser**) and eclipsed (**Tn-Thr**) conformations. (c) Water pockets derived from experiment-guided MD simulations between the peptide fragment and the GalNAc.[7,8] Antigens **Tn-Ser** and **Tn-Thr** accommodate different water pockets owing to their distinct conformational behavior in solution.

conducted by our own group concluded that while some lectins select the **Tn-Thr** determinant, others showed a higher affinity to glycopeptides carrying the **Tn-Ser** epitope.[9] In the context of antifreeze glycoproteins, which consist of the tandem repeating $(Thr\text{-}Ala\text{-}Ala)_n$ polypeptide glycosylated with Galβ-(1,3)-GalNAc, the replacement of a threonine by a serine residue eliminates the antifreeze activity.[14] It is important to note that these Tn antigens are also present in the structure of other significant TACAs, such as T, STn, or ST antigens.[4]

Our conformational analysis based on NMR data combined with experiment-guided Molecular Dynamics (MD) simulations,[15] showed that **Tn-Thr** (Figure 1b) is rather rigid in solution,[8] with its *O*-glycosidic linkage in the so-called "eclipsed" conformation ($\phi \approx 80°$, $\psi \approx 120°$), and its side chain fixed at a χ^1 torsional angle value around 60° (Figure 1a). Conversely, **Tn-Ser** (Figure 1b) displays the typical *exo*-anomeric/*syn* conformation for the glycosidic linkage with ϕ and ψ values ~80° and ~180°, respectively. This latter system is more flexible and exhibits the three possible staggered conformers for the side chain in solution.[7] In **Tn-Thr**, the carbohydrate lies almost perpendicular to the peptide but in **Tn-Ser** it adopts a parallel orientation. These conformational differences can be attributed to intrinsic steric interactions between the endocyclic oxygen and the methyl group of the threonine, which force the GalNAc unit to be located distant to the threonine, leading to an eclipsed conformation of the ψ torsion angle.[8] Concomitant with these differences, water molecules around both antigens are predicted to be distributed in a different way (Figure 1c). **Tn-Thr** displays a potential water-binding pocket localized between the NH groups of the GalNAc and threonine moieties but in the **Tn-Ser** a bridging water molecule would engage the carbonyl group of the amino acid. Their different locations could play an additional role in stabilizing the observed conformations. Unfortunately, despite various attempts, using ^{15}N- and ^{13}C-labeled Tn variants, no evidence of the presence of the proposed water pockets could be directly deduced from NMR experiments to support the computer predictions.[8] Here, by combining synthesis, crystallographic, spectroscopic and computational studies conducted in the solid and gas phases, and comparing these results with our previous and updated NMR and MD simulations data in solution, we elucidate the molecular basis behind the distinct conformations of **Tn-Thr** and **Tn-Ser** in an aqueous environment.

RESULTS AND DISCUSSION

Conformational Analysis of the Tn Antigens in the Gas Phase. First, the factors governing the preferred conformations of the Tn antigens in the gas phase, free of any interference of solvent, were investigated using mass- and conformer-selected infrared laser spectroscopy conducted under molecular beam conditions and coupled with quantum chemical computations. This strategy has been successfully employed to deduce the gas phase conformational preferences of many biomolecules.[16−18] To facilitate their detection through mass-selected ultraviolet (UV) photoionization, the Tn antigens were synthesized as benzylamide derivatives (**Tn-Ser′** and **Tn-Thr′** derivatives in Figure 1a). Of note, these variants displayed a comparable behavior in solution than that observed for the methylated variants, confirmed by 2D-ROESY spectra and experiment-guided MD simulations[15] (see Supporting Information for methods, Table S1, Schemes S1 and S2, and Figures S1, S2, and S11−S42).

The infrared ion-dip (IRID) spectra in the gas phase of **Tn-Thr′** and **Tn-Ser′** are shown in Figure 2 (see also Supporting Information, Figures S3−S7). Remarkably, they are nearly identical, particularly in the regions of the N−H and O−H stretching modes, a region highly sensitive to the presence of specific hydrogen-bonding, which suggests that these entities display the same pattern of hydrogen bonds in the gas phase. Indeed, the lowest free energy conformers calculated for both derivatives using different quantum mechanical methods (Supporting Information, Tables S2−S4) are very similar and show an excellent agreement between experimental and calculated spectra (Figure 2), confirming the observed conformation. In these preferred conformations the amino acid backbone adopts an inverse γ-turn stabilized by a strong hydrogen bond between the amino acid *C*-terminal amide and the *N*-terminal acetamide carbonyl group (band at ~3370 cm^{-1} for NH). Synergistically with the amino acid conformation, the *N*-acetyl group of the carbohydrate—essential for biological activity—is engaged in two strong hydrogen bonds that constitute the main driving force for the special architecture of the glycosidic bond in the Tn antigens in the gas phase: the acetamide carbonyl acts as a H-bond acceptor to

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

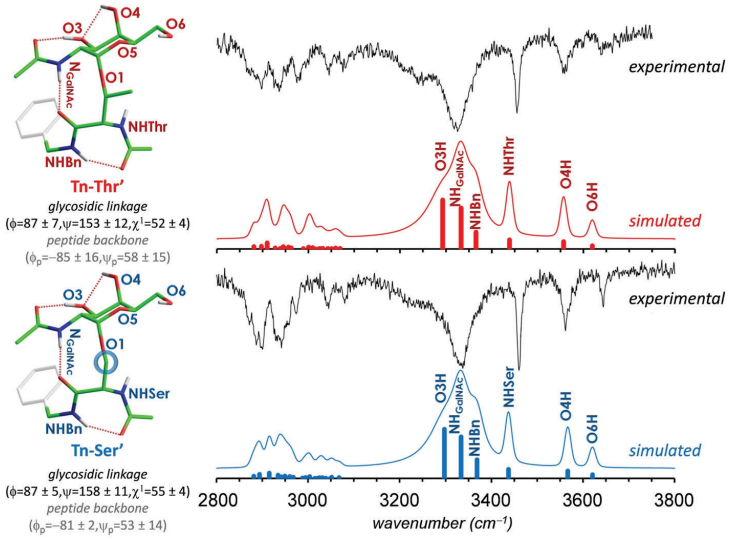


Figure 2. Infrared ion-dip (IRID) spectra of antigens **Tn-Ser′** and **Tn-Thr′** in the gas phase. Experimental and simulated infrared ion-dip (IRID) spectra of derivatives **Tn-Thr′** (upper panel) and **Tn-Ser′** (lower panel), together with representative minimum free energy conformers calculated for both compounds at the M06-2X/6-31+G(d) level.[20] These conformers are virtually identical for both derivatives in the gas phase and show the typical staggered conformation for the glycosidic linkage found in solution for the serine derivative. The values for the glycosidic linkages and the peptide backbone dihedrals are an average of the different computational methods used to predict the IRID spectra (Supporting Information, Tables S2−S4).

O3H hydroxyl group and the acetamide NH acts as a H-bond donor to the underlying amino acid *C*-terminal carbonyl group. As a result, the ψ torsion angle of the glycosidic linkage is locked to the value 158 ± 11° for the **Tn-Ser′** antigen and 153 ± 12° for the **Tn-Thr′** analogue, depending on the computational method used (Tables S2−S4). The occurrence of these common stabilizing interactions is in good agreement with the bands observed within the ∼3200−3350 cm^{-1} region for the O3H and the NH of the carbohydrate. In general, the hydroxyl groups that are involved in strong-moderate hydrogen bonds show red-shifted and broad bands. In this particular case, O3—H (strong O3—H···O═C) is displaced about 300 cm^{-1} and it is broadened around 200 cm^{-1}. Although NH groups display similar behavior, they display narrower bands and smaller displacements than OH groups.[16b,c] In this case, the acetamide NH of GalNAc (strong NH_{GalNAc}···O═C) is displaced around 100 cm^{-1}, and it is broadened about 100 cm^{-1}. The remaining hydroxyl groups of the sugar are engaged in weak hydrogen bonds, characterized by bands within the 3500−3700 cm^{-1} region: O4H interacts with O3 while O6H is in *gg* conformation and engaged in a hydrogen bond with the endocyclic oxygen O5.

Notably, the theoretical IR spectra derived from the low-energy structures are in excellent agreement with the IRID data, demonstrating the correct prediction of the common hydrogen bond network occurring in both Tn antigen derivatives in the gas phase. These experimentally validated structures also coincide with the one proposed by Csonka and co-workers[19] for the **Tn-Ser** antigen in the gas phase, based on *ab initio* calculations (HF/6-31G(d)). The new spectroscopic evidence conclusively proves that the different behavior observed for the glycosidic linkages in the Tn antigens in *solution* does not *solely* reflect the influence of steric repulsions between the carbohydrate moiety and the β-methyl group of the threonine derivative.[8] Thus, the dominant population of the eclipsed conformation of **Tn-Thr** found experimentally in solution suggests an important role for differently organized water around this substrate with respect to its **Tn-Ser** analogue.

Gradual Solvation of the Tn Antigens. Taking into account that all attempts to experimentally characterize the **Tn-Ser′** and **Tn-Thr′** antigens solvated with a discrete number of water molecules in the gas phase were unsuccessful, a comprehensive theoretical conformational analysis of the Tn antigens (**Tn-Ser** and **Tn-Thr**) with 1 and 20 water molecules was conducted to fulfill the conditions for the first hydration shell of both molecules (Figure 3a and Supporting Information, Tables S5 and S6). This analysis involved, as in the case of the isolated structures, an exhaustive conformational search to find the lowest energy structures for each system and the subsequent minimization through quantum mechanics (Supporting Information and Figure 3a). Although both monohydrated Tn antigens share the staggered conformation around the glycosidic linkage (with $\psi \approx 165°$), the addition of a single water molecule to the **Tn-Ser** derivative disrupts the hydrogen bond between the sugar and the amino acid and promotes a backbone conformational transition from the inverse γ-turn form observed *in vacuo* toward the extended arrangement populated in solution. In contrast, the **Tn-Thr···H_2O** complex retained the folded arrangement for the amino acid fragment, requiring up to 20 discrete water molecules to complete the same conformational shift. However, these water molecules were still not enough to force the **Tn-Thr** antigen to adopt the $\psi \approx 120°$ geometry. Accordingly, when experiment-guided MD simulations on these antigens were conducted in explicit water (Figure 3a, right panel, and Supporting Information, Figure S8), the **Tn-Thr** derivative adopted the "eclipsed" conformation for the

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

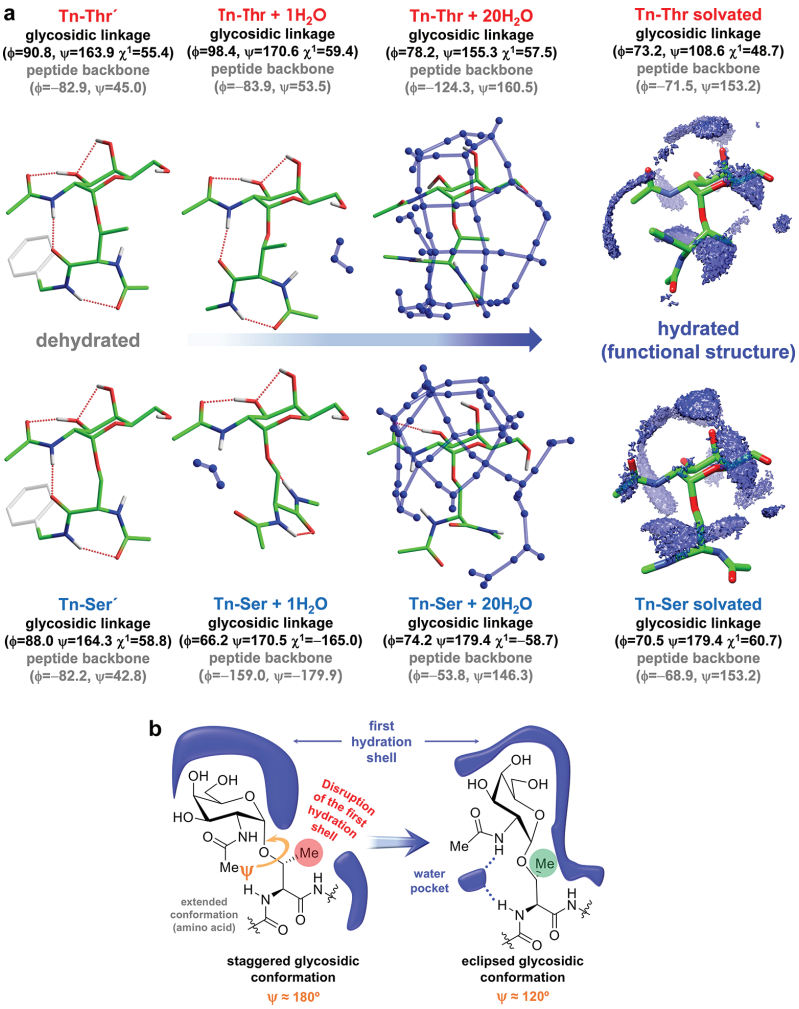


Figure 3. Gradual solvation and first hydration shell of the Tn antigens. (a) Lowest energy conformers calculated at the M06-2X/6-31+G(d) level[20] for **Tn-Thr** (upper panel) and **Tn-Ser** (lower panel) with discrete water molecules (1 and 20), together with the averaged first hydration shell derived from the experiment-guided 1 μs MD simulations. In the **Tn-Ser**, the first hydration shell is not obstructed when the staggered conformer is displayed. On the contrary, in the **Tn-Thr** a conformational shift toward the eclipsed conformer occurs, promoting an efficient solvation of the entire molecule. The geometry of the glycosidic linkage (in black) and the conformation of the peptide backbone (in gray) are also shown. (b) Schematic representation of the proposed role of the water molecules for determining the 3D structure of the **Tn-Thr** antigen.

glycosidic linkage and an extended conformation for the amino acid. Thus, the change in the underlying amino acid backbone from the inverse γ-turn (in the gas phase) to extended PPII conformations (in water) and complete solvation of the molecule are both crucial to achieve the "eclipsed" glycosidic linkage in **Tn-Thr**.

Full water solvation impairs the key hydrogen bonds involving the *N*-acetyl group of GalNAc described above, disconnecting the sugar and amino acid moieties and exacerbating both the steric and hydrophobic influence of the β-methyl group of threonine and ultimately unveiling the differences between serine and threonine Tn antigens in solution. In **Tn-Ser**, the contacts between the GalNAc and the serine moieties through water molecules take place without any interference and the more stable staggered form observed in vacuum is retained. Conversely, for **Tn-Thr**, the first solvation shell clashes with the β-methyl group of the threonine residue (Figure S8), forcing the ψ glycosidic torsion to rotate around 60° to accommodate the complete solvation shell of the antigen. The resulting conformer shows an alternative water pocket between the *N*-acetyl group of the GalNAc and the amino group of the Thr residue. Nevertheless, it is worth mentioning that the entire first hydration shell, and not only the bridging water molecule, causes this particular orientation of the glycosidic linkage in **Tn-Thr** antigen. The proposed mechanism is schematically represented in Figure 3b.

Analysis of the Crystal Structures of Fluorinated Glycopeptides Bound to an Anti-MUC1 Antibody. As previously

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952–9960

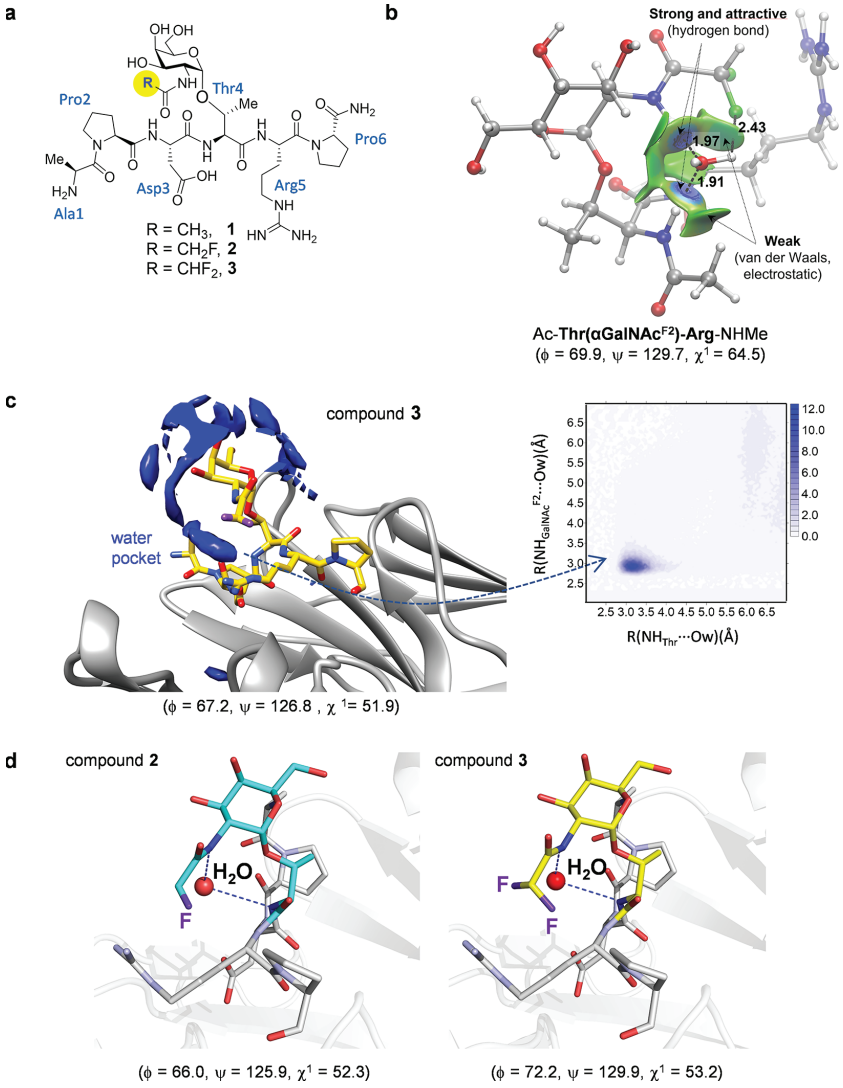


Figure 4. Use of fluorinated derivatives to trap bridging water molecules in the solid state. (a) Natural and fluorinated glycopeptides synthesized in this work, comprising the peptide epitope recognized by anti-MUC1 antibodies.[22] (b) Non-covalent interactions for the water molecule bound to the hydrophilic pocket of model glycopeptide **Ac-Thr(*α*GalNAc^{F2})-Arg-NHMe** deduced by QM calculations (M06-2X/6-311G(d,p)). Weak attractive polar interactions between the fluorine atoms and water hydrogens stabilize the water into this pocket. The geometries were fully optimized with the PCM(water)/M06-2X/6-311G(d,p) method.[20,23] (c) Representation of the first hydration shell around the fluorinated Tn antigen derived from 200 ns MD simulations performed on glycopeptide **3** in the SM3-bound state. The 2D radial distribution function[24] calculated for the nitrogen atoms involved in the bridging water molecule is also shown. (d) Views of the binding sites of the complexes between glycopeptides **2** and **3** and the scFv-SM3 antibody (PDB IDs: 6FZR and 6FZQ, respectively), showing the key water molecule between the *N*-fluoroacetyl groups of the sugar and the amino group of the threonine residue. The geometry of the glycosidic linkage is shown in parentheses in (b), (c), and (d).

reported,[6a] the structure of the complex between the Thr-containing glycopeptide **1** (Figure 4a) when bound to an anti-MUC1 antibody (SM3)[21] did not show any bridging water molecule between the sugar and the threonine residue. Probably, the high water exchange rate precluded the experimental detection of the water molecules at this site.[8] In an attempt to detect this relevant water-mediated carbohydrate/amino acid interactions, we hypothesized that a more hydrophilic pocket would be able to bind water molecules more efficiently. Thus, the hydrogen-bonding donor character of the sugar *N*-acetyl fragment could be enhanced by replacing its constituent methyl group by fluoromethyl groups[6b] (compounds **2** and **3** in Figure 4a; see

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

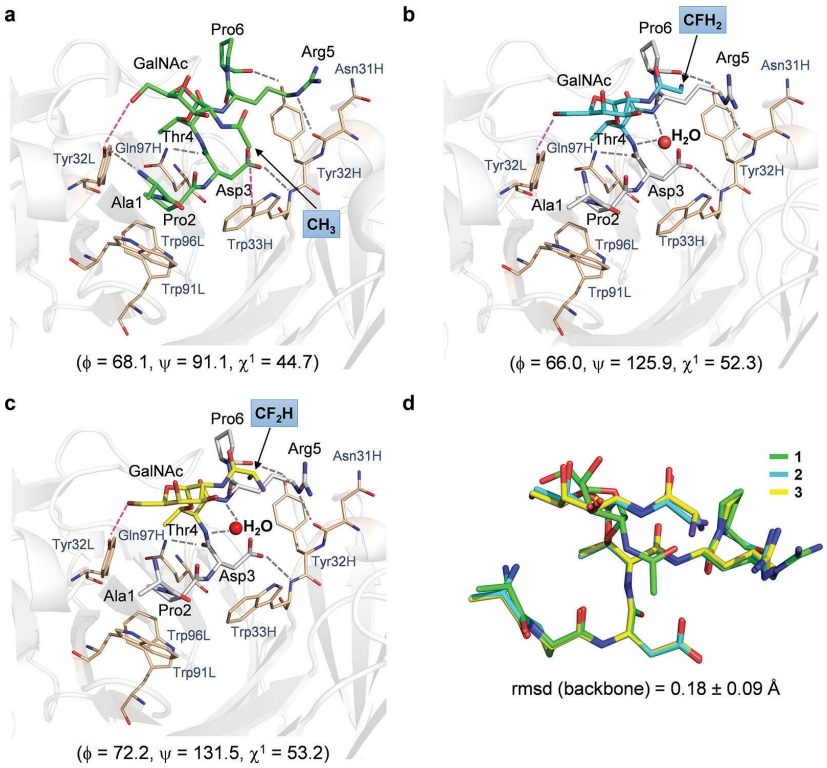


Figure 5. Analysis of the X-ray structures of glycopeptides **1** (ref 6), **2**, and **3** in complex with scFv-SM3. Key binding interactions of glycopeptides **1** (a), **2** (b), and **3** (c) with the antibody, as observed in the X-ray crystal structures (PDB IDs: 5A2K, 6FZR, and 6FZQ, respectively). Pink dashed lines indicate hydrophobic and hydrogen bond interactions between GalNAc and SM3 surface, and gray dashed lines indicate hydrogen bonds between peptide backbones and SM3 antibody. (d) Superposition of the peptide backbone of glycopeptides **1**–**3** in complex with SM3.

also Supporting Information, Scheme S3 and Figures S43–S56). In line with this idea, quantum mechanical calculations performed on a reduced model of **3** indicate that the water molecule is stabilized not only by two hydrogen bonds with the NH groups of the sugar and Thr residues, but also through an O—H···F contact, providing a negative water binding free energy compared to the positive value calculated for the non-fluorinated analogue (Figure 4b and Supporting Information, Figure S9).

Moreover, 200 ns MD simulations performed on glycopeptide **3** in the SM3-bound state[6,22] revealed an unusually high water density located between the sugar and the peptide units (Figure 4c), strongly suggesting that the fluorinated *N*-acetyl group indeed configures a more hydrophilic water pocket between the carbohydrate–peptide interface. In light of these theoretical predictions and to provide certain experimental evidence for the proposed solvent-mediated conformational transition in **Tn-Thr** antigen, we determined the X-ray structures of derivatives **2** and **3** bound to the antibody scFv-SM3[22] at high resolution (<2.0 Å, Figures 4a,d and 5; see also Supporting Information, Table S7 and Figure S10). To our delight, these data allowed us, for the first time, to visualize interfacial water molecules. Certainly, the crystal structures reveal a structural water molecule located between the amino group of the GalNAc and the NH group of the glycosylated Thr residue, as we have proposed earlier for the natural **Tn-Thr** antigen in solution.[8,25] To the best of our knowledge, this is the first confirmation of a commonly accepted hypothesis in the field of *O*-glycopeptides.

Crystallographic analysis also revealed that the surface groove of the antibody fits all the peptide residues in the three studied complexes (Figure 5a–c), irrespective of the presence of the natural or fluorinated GalNAc. The overall conformation of the peptide fragment of glycopeptides **1**–**3** is nearly identical, except for the side chain of the arginine reside in the fluorinated variants, and is similar to that found in the crystal structure reported for the naked peptide[21] (Figure 5d). The stabilizing contacts in these complexes between the antigen and the antibody involve several hydrogen bonds, some of them mediated by water molecules, as well as several stacking interactions (Figure 5a–c). For instance, while side chains of Asp3 and Arg5 in all glycopeptides are involved in hydrogen bonds with Trp33H and Asn31H, respectively, the carbonyl group of Thr4 and Pro6 are engaged in a hydrogen bond interaction with Gln97H and Tyr32H. In addition, Pro2 stacks with Trp91L, Trp96L, and Tyr32L, while side chains of Asp3 and Arg5 are engaged in hydrophobic contacts with Trp33H and Tyr32H, respectively. In contrast, while Ala1 in **1** is interacting with Tyr32L through a hydrogen bond, this interaction is not observed in the X-ray structures of the fluorinated glycopeptides. Concerning the glycosidic linkage, it adopts the common "eclipsed" conformer

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952–9960

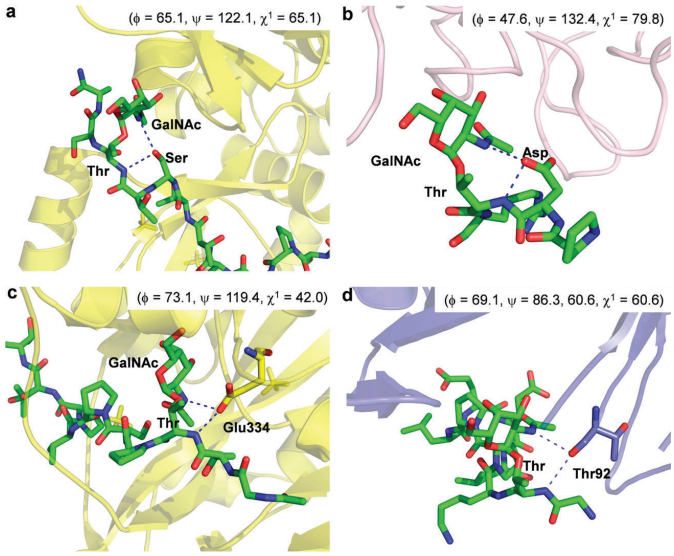


Figure 6. Molecular recognition of glycopeptides bearing the **Tn-Thr** antigen. The bridging water molecule is replaced by an atom of either the glycopeptide or the receptor. (a) Zoom-in of the crystal structure of the active form of GalNAc-Transferase 2 (GalNAc-T2) in complex with UDP and the glycopeptide MUC5AC-13, showing the lectin domain (PDB ID: 5AJP).[27] (b) Zoom-in of the crystal structure of soybean agglutinin from *Glycine max* in complex with the glycopeptide PDT(αGalNAc)R (PDB ID: 4D69).[30] (c) Zoom-in of the crystal structure of the inactive form of GalNAc-T2 in complex with UDP and the glycopeptide MUC5AC-3,13, showing the catalytic domain (PDB ID: 5AJO).[27] (d) Zoom-in of the crystal structure of the antibody 237 in complex with its glycopeptide epitope (PDB ID: 3IET).[28] In all cases, the geometry of the glycosidic linkage is shown in parentheses.

in the solid state for glycopeptides **2** and **3**, with ψ close to 120° (Figure 5b,c), which is stabilized by the occurrence of the bridging water molecule. In the natural glycopeptide, however, the torsional angle ψ takes a value close to 90°. This orientation favors a CH-π interaction between the methyl group of GalNAc and Trp33H. In this regard, it is important to note that MD simulations performed on the complex of glycopeptide **1** and scFv-SM3 suggested that the glycosidic linkage of the antigen adopts an "eclipsed" conformation, with a value for ψ around 120° in solution.[6] Moreover, the glycosidic linkage of **Tn-Thr** antigen in complex with other proteins/enzymes can adopt variable values for $\psi \approx 90°$ to 130° in the solid state (Figure 6). Finally, in the three complexes the hydroxymethyl group of GalNAc is engaged in a hydrogen bond with Tyr32L. It is important to note that the density for the fluorine atoms shown in Figures 4d and 5b–d, and Supporting Information Figure S10 is weak, impeding the accurate location of these atoms in the X-ray structures. This is likely due to the rotational mobility of the CH_2F and CHF_2 groups. As a consequence, the crystal structures do not confirm the proposed additional O–H···F contact aforementioned and shown in Figure 4b.

Implications for the Molecular Recognition of the Tn Antigens. The occurrence of this persistent water pocket in the **Tn-Thr** antigen may have important implications in the stabilization of the eclipsed conformation and, in turn, in the binding of this entity to the corresponding receptors. Indeed, according to the data presented in this work and the crystal structures of receptors complexed with glycopeptides bearing this Tn antigen (Figure 6), two different scenarios are likely:

(1) The water pocket can be retained upon binding. This is the most plausible situation for MUC1-like glycopeptides bound to anti-MUC1 antibodies. In these cases, the bridging water molecules help the antigen to display the bioactive conformation in solution, therefore assisting the binding process.

(2) The "bridging" water molecule can be replaced upon binding to the biological target by an oxygen atom of either the ligand or the receptor. As a result, the eclipsed conformation observed in solution is also maintained in the bound state. According to several reported X-ray structures,[26−28] installing oxygenated amino acids in specific regions of the receptor capable of displacing the bridging water molecule to the bulk-solvent may entropically facilitate the molecular recognition[29] of glycopeptides bearing the **Tn-Thr** antigen. This novel strategy, which resembles the well-known water displacement approach in drug design, although in a reverse manner (i.e., water in the ligand is displaced by the protein receptor) could be valuable for designing receptors with an enhanced affinity toward the **Tn-Thr** antigen.

■ CONCLUSIONS

A multidisciplinary approach that includes the experimental and theoretical study of the Tn antigens in the gas, solution, and solid phases has been applied to deduce the key role of water in the modulation of the conformational preferences of these molecules and therefore in their presentations for interacting with protein receptors. In the **Tn-Ser** antigen, water molecules can efficiently solvate the whole molecule in the typical *exo*-anomeric/*syn* conformation also present in the gas phase. However, in the **Tn-Thr** derivative, the methyl group at C_β disturbs the proper solvation of the "native" gas-phase

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952–9960

geometry and ψ rotates around 60° to exhibit an eclipsed conformation. In this geometry, the GalNAc moiety drastically modifies its presentation and displays an almost perpendicular arrangement with respect to the amino acid. Fittingly, this arrangement structure facilitates the efficient accommodation of a water pocket between the NH groups of the sugar and the threonine residues. This mechanism is reinforced by the observation, for the first time, of such structural water in the crystal structures of scFv-SM3 antibody in complex with two fluorinated Tn-Thr antigens. Interestingly, this peculiar arrangement of the **Tn-Thr** antigen is also observed in the bound state of this antigen to different biological receptors, including antibodies,[6,13] enzymes (GalNAc-transferases),[26,27] and lectins.[30] In contrast, for the **Tn-Ser** antigen, different arrangements of the glycosidic linkage may occur in the bound state since the lack of the β-methyl group renders a more flexible architecture.

In addition, proving the importance of the *O*-GlcNAcylation of threonine and serine residues in different biological events,[31−33] the extension of this mechanism to the β-*O*-GlcNAc-Ser and β-*O*-GlcNAc-Thr analogues, is also possible. Overall, our data provide compelling evidence of the molecular basis behind the different conformations of the **Tn-Thr** and **Thr-Ser** antigens in solution and in the enzyme/protein-bound state, which are determinant for their distinct biological functions and outcomes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b04801.

Synthesis and characterization of **Tn-Ser′**, **Tn-Thr′**, and glycopeptides **2** and **3**; experimental data registered for **Tn-Ser′** and **Tn-Thr′** in the gas phase; computational protocols and Cartesian coordinates of the lowest energy DFT calculated structures; conformational analysis of **Tn-Ser′** and **Tn-Thr′** in solution; data collection and refinement statistics for the X-ray structures of complexes **2**/**scFv-1SM3** and **2**/**scFv-1SM3** (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*rhurtado@bifi.es
*emiliojose.cocinero@ehu.eus
*francisco.corzana@unirioja.es

ORCID

Imanol Usabiaga: 0000-0002-1621-8536
José A. Fernández: 0000-0002-7315-2326
Jesús H. Busto: 0000-0003-4403-4790
Jesús Jiménez-Barbero: 0000-0001-5421-8513
Jesús M. Peregrina: 0000-0003-3778-7065
Gonzalo Jiménez-Osés: 0000-0003-0105-4337
Emilio J. Cocinero: 0000-0001-7632-3728
Francisco Corzana: 0000-0001-5597-8127

Author Contributions

□I.A.B., I.U., and I.C. have contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank MINECO (projects CTQ2015-67727-R and UNLR13-4E-1931 to F.C. and J.M.P; CTQ2013-44367-C2-2-P and BFU2016-75633-P to R.H.-G.; CTQ2015-64597-C2-1P to J.J.-B.; CTQ2015-70524-R and RYC-2013-14706 to G.J.O; SEV-2016-0644 to CIC bioGUNE; CTQ2017-89150-R to E.J.C.; and CTQ2015-68148-C2-1-P to J.A.F.). F.C. thanks the EU (Marie-Sklodowska Curie ITN, *ProteinConjugates*). J.A.F. and E.J.C. thank UPV/EHU (UFI11/23 and PPG17/10) for financial support. I.A.B. thanks the *Asociación Española Contra el Cancer* en La Rioja for a grant. I.C. thanks Universidad de La Rioja for the FPI grant. R.H-G. thanks *Agencia Aragonesa para la Investigación y Desarrollo* (ARAID) and the *Diputación General de Aragón* (DGA, group number E34_R17) for financial support. The research leading to these results has also received funding from the FP7 (2007-2013) under BioStruct-X (grant agreement no. 283570 and BIOSTRUCTX_5186). We thank synchrotron radiation source ALBA (Barcelona) and beamline XALOC. Computational resources of CESGA, Universidad de La Rioja (BERONIA) and UPV/EHU (SGIker and I2Basque) were used in this work. We thank Prof. J. P. Simons (Oxford University) for valuable comments and scientific discussions.

■ REFERENCES

(1) Ju, T.; Wang, Y.; Aryal, R. P.; Lehoux, S. D.; Ding, X.; Kudelka, M. R.; Cutler, C.; Zeng, J.; Wang, J.; Sun, X.; Heimburg-Molinaro, J.; Smith, D. F.; Cummings, R. D. *Proteomics: Clin. Appl.* **2013**, *7*, 618−631.

(2) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 1770−1791.

(3) Springer, G. F. *J. Mol. Med.* **1997**, *75*, 594−602.

(4) Martínez-Sáez, N.; Peregrina, J. M.; Corzana, F. *Chem. Soc. Rev.* **2017**, *46*, 7154−7175.

(5) Lakshminarayanan, V.; Thompson, P.; Wolfert, M. A.; Buskas, T.; Bradley, J. M.; Pathangey, L. B.; Madsen, C. S.; Cohen, P. A.; Gendler, S. J.; Boons, G.-J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 261−266.

(6) (a) Martínez-Sáez, N.; Castro-López, J.; Valero-Gónzalez, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Bernardes, G. J. L.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem., Int. Ed.* **2015**, *54*, 9830−9834. (b) Unione, L.; Alcalá, M.; Echeverria, B.; Serna, S.; Ardá, A.; Franconetti, A.; Cañada, F. J.; Diercks, T.; Reichardt, N.; Jiménez-Barbero, J. *Chem. - Eur. J.* **2017**, *23*, 3957−3965.

(7) Corzana, F.; Busto, J. H.; Jiménez-Oses, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2006**, *128*, 14640−14648.

(8) Corzana, F.; Busto, J. H.; Jiménez-Oses, G.; García de Luis, M.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2007**, *129*, 9458−9467.

(9) Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; García-García, L.; Berbis, M. Á.; Valero-Gónzalez, J.; Martín-Santamaría, S.; Hurtado-Guerrero, R.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoza, A.; Busto, J. H.; Corzana, F.; Peregrina, J. M. *Chem. - Eur. J.* **2014**, *20*, 12616−12627.

(10) Mazal, D.; Lo-Man, R.; Bay, S.; Pritsch, O.; Dériaud, E.; Ganneau, C.; Medeiros, A.; Ubillos, L.; Obal, G.; Berois, N.; Bollati-Fogolin, M.; Leclerc, C.; Osinaga, E. *Cancer Immunol. Immunother.* **2013**, *62*, 1107−1122.

(11) Zhang, Y.; Li, Q.; Rodriguez, L. G.; Gildersleeve, J. C. *J. Am. Chem. Soc.* **2010**, *132*, 9653−9662.

(12) Kanekura, T.; Sakuraba, H.; Matsuzawa, F.; Aikawa, S.; Doi, H.; Hirabayashi, Y.; Yoshii, N.; Fukushige, T.; Kanzaki, T. *J. Dermatol. Sci.* **2005**, *37*, 15−20.

(13) Coelho, H.; Matsushita, T.; Artigas, G.; Hinou, H.; Cañada, F. J.; Lo-Man, R.; Leclerc, C.; Cabrita, E. J.; Jiménez-Barbero, J.; Nishimura, S.-I.; Garcia-Martin, F.; Marcelo, F. *J. Am. Chem. Soc.* **2015**, *137*, 12438−12441.

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

(14) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S.-I. *Angew. Chem., Int. Ed.* **2004**, *43*, 856−862.
(15) Corzana, F.; Busto, J. H.; Engelsen, S. B.; Jiménez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Avenoza, A. *Chem. - Eur. J.* **2006**, *12*, 7864−7871.
(16) (a) Cocinero, E. J.; Stanca-Kaposta, E. C.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 1282−1287. (b) Barry, C. S.; Cocinero, E. J.; Carçabal, P.; Gamblin, D. P.; Stanca-Kaposta, E. C.; Remmert, S. M.; Fernández-Alonso, M. C.; Rudić, S.; Simons, J. P.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135*, 16895−16903. (c) Cocinero, E. J.; Stanca-Kaposta, E. C.; Dethlefsen, M.; Liu, B.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. *Chem. - Eur. J.* **2009**, *15*, 13427−13434.
(17) Cocinero, E. J.; Carçabal, P.; Vaden, T. D.; Simons, J. P.; Davis, B. G. *Nature* **2011**, *469*, 76−79.
(18) León, I.; Millán, J.; Cocinero, E. J.; Lesarri, A.; Fernández, J. A. *Angew. Chem., Int. Ed.* **2013**, *52*, 7772−7775.
(19) Csonka, G. I.; Schubert, G. A.; Perczel, A.; Sosa, C. P.; Csizmadia, I. G. *Chem. - Eur. J.* **2002**, *8*, 4718−4733.
(20) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(21) Dokurno, P.; Bates, P. A.; Band, H. A.; Stewart, L. M.; Lally, J. M.; Burchell, J. M.; Taylor-Papadimitriou, J.; Snary, D.; Sternberg, M. J.; Freemont, P. S. *J. Mol. Biol.* **1998**, *284*, 713−728.
(22) Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycobiology* **2004**, *14*, 681−692.
(23) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3094.
(24) Andersson, C.; Engelsen, S. B. *J. Mol. Graphics Modell.* **1999**, *17*, 101−105.
(25) Corzana, F.; Busto, J. H.; García de Luis, M.; Fernández-Tejada, A.; Rodríguez, F.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Eur. J. Org. Chem.* **2010**, *2010*, 3525−3532.
(26) Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañón, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. *Angew. Chem., Int. Ed.* **2014**, *53*, 8206−8210.
(27) Lira-Navarrete, E.; de Las Rivas, M.; Compañón, I.; Pallarés, M. C.; Kong, Y.; Iglesias-Fernández, J.; Bernardes, G. J. L.; Peregrina, J. M.; Rovira, C.; Bernadó, P.; Bruscolini, P.; Clausen, H.; Lostao, A.; Corzana, F.; Hurtado-Guerrero, R. *Nat. Commun.* **2015**, *6*, 6937.
(28) Brooks, C. L.; Schietinger, A.; Borisova, S. N.; Kufer, P.; Okon, M.; Hirama, T.; Mackenzie, C. R.; Wang, L.-X.; Schreiber, H.; Evans, S. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10056−10061.
(29) Persch, E.; Dumele, O.; Diederich, F. *Angew. Chem., Int. Ed.* **2015**, *54*, 3290−3327.
(30) Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-Gónzalez, J.; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ACS Chem. Biol.* **2015**, *10*, 747−756.
(31) Brister, M. A.; Pandey, A. K.; Bielska, A. A.; Zondlo, N. J. *J. Am. Chem. Soc.* **2014**, *136*, 3803−3816.
(32) Rani, L.; Mallajosyula, S. S. *J. Phys. Chem. B* **2017**, *121*, 10618−10638.
(33) Fernández-Tejada, A.; Corzana, F.; Busto, J. H.; Jiménez-Oses, G.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. *Chem. - Eur. J.* **2009**, *15*, 7297−7301.

DOI: 10.1021/jacs.8b04801
J. Am. Chem. Soc. 2018, 140, 9952−9960

Ribosugars

ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Observation of the Unbiased Conformers of Putative DNA-Scaffold Ribosugars

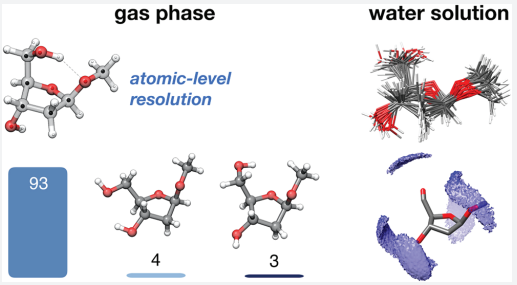
Camilla Calabrese, Iciar Uriarte, Aran Insausti, Montserrat Vallejo-López, Francisco J. Basterretxea, Stephen A. Cochrane, Benjamin G. Davis,* Francisco Corzana,* and Emilio J. Cocinero*

Cite This: *ACS Cent. Sci.* 2020, 6, 293−303

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The constitution, configuration, and flexibility of the core sugars of DNA molecules alter their function in diverse roles. Conformational itineraries of the ribofuranosides (*f*s) have long been known to finely determine rates of processing, yet we also know that, strikingly, semifunctional DNAs containing pyranosides (*p*s) or other configurations can be created, suggesting sufficient but incompletely understood plasticity. The multiple conformers involved in such processes are necessarily influenced by context and environment: solvent, hosts, ligands. Notably, however, to date the *un*biased, "naked" conformers have not been experimentally determined. Here, the inherent conformational biases of DNA scaffold deoxyribosides in unsolvated and solvated forms have now been defined using gas-phase microwave and solution-phase NMR spectroscopies coupled with computational analyses and exploitation of critical differences between natural-abundance isotopologues. Serial determination of precise, individual spectra for conformers of these 25 isotopologues in alpha (*α*-D) and beta (*β*-D); pyrano (*p*) and furano (*f*) methyl 2-deoxy-D-ribosides gave not only unprecedented atomic-level resolution structures of associated conformers but also their quantitative populations. Together these experiments revealed that typical ^{2}E and ^{3}E conformations of the sugar found in complex DNA structures are not inherently populated. Moreover, while both OH-5′ and OH-3′ are constrained by intramolecular hydrogen bonding in the unnatural *αf* scaffold, OH-3′ is "born free" in the "naked" lowest lying energy conformer of natural *βf*. Consequently, upon solvation, unnatural *αf* is strikingly less perturbable (retaining $^{2}T_{1}$ conformation *in vacuo* and water) than natural *βf*. Unnatural *αp* and *βp* ribosides also display low conformational perturbability. These first experimental data on inherent, unbiased conformers therefore suggest that it is the background of conformational flexibility of *βf* that may have led to its emergence out of multiple possibilities as the sugar scaffold for "life's code" and suggest a mechanism by which the resulting freedom of OH-3′ (and hence accessibility as a nucleophile) in *βf* may drive preferential processing and complex structure formation, such as replicative propagation of DNA from 5′-to-3′.



INTRODUCTION

Structural variability and flexibility of ribonucleic acids are not only apparent and immense in scope but also intimately linked to both the existence[1−3] and emergence[4,5] of biological function. Moreover, ever-expanding interest in the design and use of both natural and unnatural ribonucleotides in diagnostic and therapeutic applications continues to highlight a key role for an understanding of the fundamentals that generate associated structural populations.[6−8] For example, while chemical modifications at phosphate[9] or nucleobase[10] can usefully increase *in vivo* stability (reduced *reactivity*), it is also the correct manipulation of the conformations of the core sugar scaffold that has proven key to optimal functional *activity*. In DNA polymerases, an essential factor that prevents improper inclusion and extension of nucleotides appears to be governed by the preferred conformations of the furanose moiety of each incoming nucleotide during both incorporation and extension.[11,12] Interestingly, as the pioneering work of Eschenmoser[13] highlighted, there is also no necessity for ribosidic or even furanosidic structures, and alternative polynucleotides can be constructed based on, for example, L-*threo*-furanosides[14] or even configurationally varied pyranosides.[13] While their functions are typically moderated (e.g., reduced base-pairing strengths), such altered-sugar polynucleotides can still adopt relevant duplex structures via typical (e.g., Watson−Crick) patterns and can even be processed by appropriate variant enzymes, albeit at reduced rates.

Received: December 14, 2019
Published: February 18, 2020

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

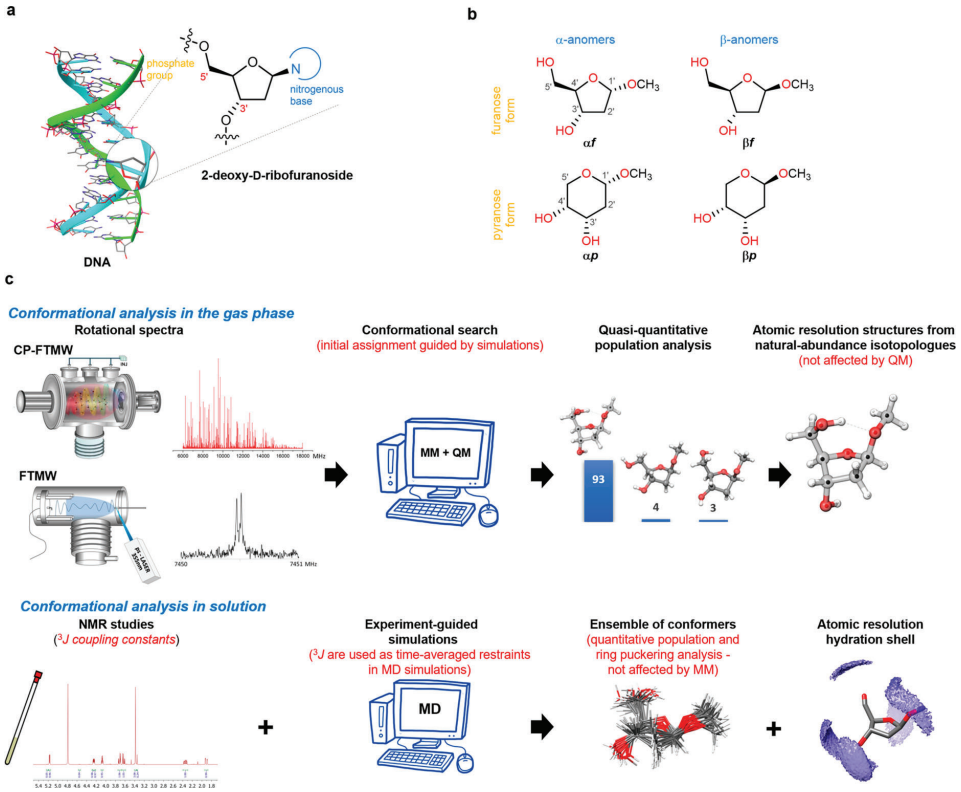


Figure 1. Coherent conformational analyses and comparison of DNA-scaffold ribosugars. (a) Schematic representation of the structure of dominant B-form DNA. (b) Compounds studied in this work: two 2-deoxyribofuranose forms (compounds *αf* and *βf*) and two corresponding pyranose forms (derivatives *αp* and *βp*). (c) Overview of the precise, comparative protocol used in this work to determine D-ribosides in the gas phase (upper panel) and in aqueous solution (lower panel).

Together, such examples illustrate that both the fuller understanding of natural nucleotide function and the development of useful unnatural nucleotides require a deep insight into associated conformational preferences and, in particular, those of the native sugar moiety, 2-deoxy-D-riboside. However, no furanoside structures to date have been determined under fully isolated (and hence artefact free) conditions. Crystallographic data can be biased by both solvent effects and crystal-packing interactions. Spectroscopy in the solution phase is similarly dominated by environmental contacts.[15] Indeed, extensive analyses performed on nucleosides and nucleotides, as well as in DNA, suggest that the sugar cores of such structures often exist in a dynamic equilibria between multiple conformations in solution.[16]

Here we present a strategy for the complete structural analysis of the core sugar scaffold of DNA, 2-deoxy-D-riboside (Figure 1a,b), that exploits custom-made, high-resolution microwave spectrometers combined with complementary vaporization and sampling techniques in the gas phase. Importantly, while associated quantum chemistry methods have been integral to prior analyses/interpretations of most gas phase structures, these microwave methods do not require quantum chemical computations during structural determination. Instead, natural abundance isotopologues act as key structural references to allow atomic-level resolution. The resulting structures are therefore critically independent of the quality or processes of any associated quantum mechanical models. Comparison of these first unbiased conformational analyses with those affected by solution was then achieved by combining NMR experiments with experiment-guided molecular dynamics (MD) simulations (Figure 1c). This allows precise study of the inherent intramolecular interactions responsible for the structural diversity found in the core of DNA without interference from surrounding molecules in condensed media (e.g., water as solvent).

RESULTS

Design of a System for Generating and Analyzing Gaseous, Isolated DNA Scaffolds. Current understanding of DNA structure is based in significant part on X-ray crystallography data,[17−20] providing information concerning global helical structure and the geometry of local features, such as base-pair stacking patterns and backbone conformation driven by that structure. However, such data can not only be biased by both solvent effects and crystal-packing interactions; it also provides only static structures. Conformational analyses of DNA fragments in solution can overcome some of these problems.[21] However, the lack of sequence variety (dominated

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

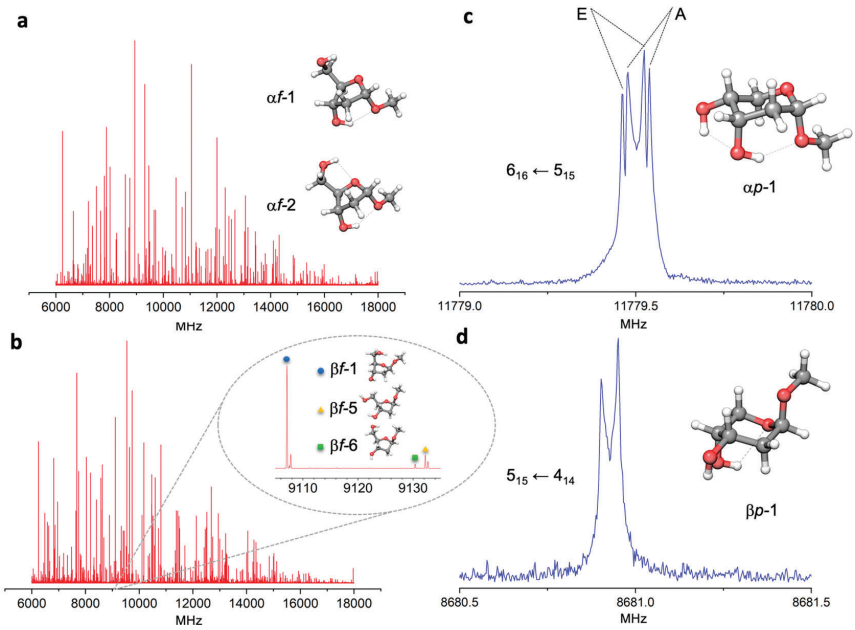


Figure 2. High-resolution rotational spectra of *αf*, *βf*, *αp*, and *βp*. (a and b) Overview of the CP-FTMW rotational spectra of *α* and *β* deoxyriboside anomers *αf*, *βf* in the 6−18 GHz region, respectively. In b, an expanded view highlights rotational transitions of each of the *β*-conformers. (c and d) Typical rotational transitions of *αp* and *βp* were observed using FTMW spectroscopy equipped with a UV ultrafast laser vaporization system. Rotational transitions showed hyperfine splitting due to internal rotation of the methyl group. Both transitions (c and d) are additionally split by the instrumental Doppler effect.

in general by A-tracts and the Drew−Dickerson dodecamer) combined with limitations in refinement protocols (sometimes biased by the quality of calculations used to interpret experimentally derived restraints[22]) has not yet delivered a data set of structures with consistent information on the intrinsic mechanics of DNA in solution.[23] For instance, in some structures, significant deviations are seen from similar NMR data when refined with different backbone restraints,[24,25] and in most studies conformational analyses rely on highly variable two-state models.[26,27] Consequently, to date, no precise (atomic resolution) structures of DNA nor the biases of its scaffold have yet been reported.

In previous pioneering studies, conformational analyses of smaller carbohydrates fragments were accomplished using gas phase vibrational laser spectroscopy.[28−31] However, these studies typically demand the use of molecules tagged with a chromophore that can be a potential cause of undesirable artefacts. While these can be replaced by external probes (e.g., toluene[32,33] or peptides[34]), this alternative methodology does not fully remove the need for chromophore and can induce additional unwanted environmental interference. Moreover, the method is critically dependent on quantum mechanical (QM) models to relate determined spectra to derived structures.

In this context, gas phase rotational spectroscopy emerges as a unique method that can avoid such artifacts. In particular, observation of multiple rotational spectra from isotopologues (e.g., where ^{13}C replaces ^{12}C) allows direct structural determination without dependence on QM models or chromophores. In addition, it possesses a superior inherent resolution due to unsurpassed frequency resolution (∼kHz) and unrivalled chemical (conformers, tautomers, isotopologues, even enantiomers) discrimination.[35,36] However, the ability to measure isotopic species directly requires sufficient sensitivity (or enriched samples), which in turn demands sufficient partial pressure of the analyte in the gas phase. For analytes, such as carbohydrates, that have low volatility and are thermolabile, this has proven challenging using traditional heating methods[37] or even IR nanosecond-pulsed laser vaporization techniques.[38] Recently, we have shown that UV (355 nm) picosecond-pulsed (∼40 ps) laser vaporization can allow observations of certain reducing sugars,[39−42] but their mutarotation[39,40] and/or lack of sensitivity[39,41] prevented either observation of biologically relevant constitutional forms or structures with atomic resolution; consequently, observation of core DNA scaffold 2-deoxy-D-ribosides (Figure 1b) was not previously possible.

We reasoned that the required critical increase in sensitivity might be gained in several ways. This proved successful through the custom design (Supplementary Figures 1 and 2) of microwave apparatus that included a Fourier-transform microwave-spectrometer (FTMW)[43] coupled to a UV ultrafast laser vaporization system[39] and a set-up with chirped-pulses (CP-FTMW) (Figure 1c).[44] Specifically, the latter system allowed: (a) simultaneous three-nozzle gas injection; (b) longer acquisition times (allowing even spectral registration for 1 week continuously averaging ∼40 million of rotational spectra); and (c) more efficient radiation pulse sequences (30 microwave pulses per molecular pulse). These features together provided an increase of an order of magnitude (or

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

Table 1. Experimental Rotational Constants of the Observed Conformers of *αf*, *βf*, *αp*, and *βp*

	αf-1	*αf*-2	*βf*-1	*βf*-5	*βf*-6	*αp*-1	*βp*-1
A/MHz[a]	2056.39188(60)[d]	1882.30839(76)	1890.65747(49)	1999.87647(81)	1481.77004(72)	2153.8853(66)	2358.67704(36)
B/MHz	1019.67311(21)	1024.81006(26)	1145.05589(21)	991.81629(38)	1256.10078(60)	1058.89820(11)	1001.97250(15)
C/MHz	860.52781(23)	804.74073(31)	910.72056(22)	771.18931(31)	784.89436(42)	960.74320(13)	831.19774(12)
N[b]	95	99	140	52	86	55	90
σ[c]/kHz	8.2	9.3	10.5	6.7	7.4	2.3	3.5

[a]Rotational constants (*A*, *B*, *C*). [b]Number of rotational transitions (*N*). [c]Root-mean-square (rms) deviation (*σ*) of the fit. [d]Standard errors in units of the last digit.

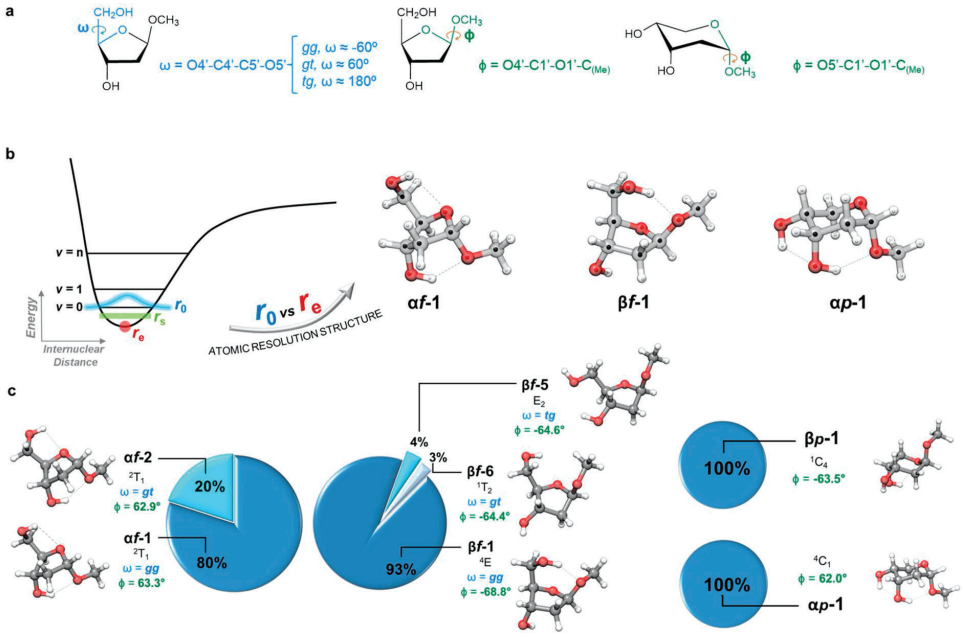


Figure 3. Observed conformers of *αf*, *βf*, *αp*, and *βp* in the gas phase. (a) Definition of relevant torsional angles. (b) Conceptualization of gas phase molecular structure-determination methods. Experimental molecular structures could be determined for *αf*-1, *βf*-1, and *αp*-1 with atomic resolution due to the observation of isotopologues in natural abundance. (c) Conformers detected for furanosides and pyranosides, together with their population and key geometrical parameters.

greater) sensitivity, which in turn allowed lower consumption of the sample (down to 0.7−2.0 g of synthetic samples). In this way, even each of the multiple isotopologues present in samples at only natural abundance ($^{13}C^{12}C_5H_{12}O_4 \approx 1.1\%$) could be detected and characterized. These advances allowed us to determine, for the first time, "nonbiased" atomic resolution structures of core DNA scaffold 2-deoxy-D-ribosides.

Experimental Determination of Unbiased Structures of DNA Scaffold Sugars. First, core putative scaffolds *αf*, *βf*, *αp*, and *βp* (Figure 1b) were readily synthesized by complementary methods on the multigram scales required. Briefly (see Supplementary Methods for further details), in one method 2-deoxy-D-ribose was subjected to classical Fischer glycosylation conditions (1% HCl in methanol); the equilibrium mixture after 18 h (57% *βp*, 15% *αp*, 11% *βf*, 17% *αf*) yielded *αp* by column chromatography and *βp* by further crystallization. To generate *αf* and *βf* on scale, we preferentially employed brief (15 min) treatment, giving an ∼1:1 mixture of essentially only furanosides; subsequent peracetylation/deacetylation allowed clean isolation via column chromatography (33% *βf*, 36% *αf*).

Next, gas-phase DNA scaffold generation was tested under various modes (Figures 1c and 2). Of these, CP-FTMW spectroscopy allowed collection of rotational spectra of *αf* and *βf*. (Figure 2, panels a and b, respectively). An FTMW spectrometer coupled to UV ultrafast laser vaporization proved most successful for *αp* and *βp*. Together, these experiments allowed long acquisition times from gram-scale samples thereby yielding rotational spectra of the "nonbiased" structures of *αf*, *βf*, *αp*, and *βp* with an unprecedented signal-to-noise ratio (S/N for most intense observed transitions: >500/1 and >1800/1 for *αf* and *βf*, respectively). In turn, such S/N orders allowed determination of structures with atomic resolution. In this way, despite heavily congested spectra—*αf* (Figure 2a) and *βf* (Figure 2b) showed ∼600 and ∼1800 rotational lines, respectively—sets of rotational transitions were identified belonging to independent structures.

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

Each transition set was fitted to a semirigid rotor Hamiltonian based on Watson's symmetric reduction and I^r representation to obtain *A*, *B*, and *C* rotational constants and critical centrifugal distortion constants.[45,46] Conformational assignments were guided and supported by (but not dependent on, *vide infra*) theoretical calculations, using a two-step MM then QM (DFT and *ab initio*) strategy[39,42] (see Supplementary Methods). Rotational constants can be correlated with moments of inertia, which are a fingerprint of 3D molecular structure (Table 1). Direct comparison of simulated and experimental rotational constants (Table 1 and Supplementary Tables 1, 3, 5, and 6) therefore allowed an unequivocal assignment.

Two αf conformers (named αf-1 and αf-2, numbered by calculated energy ranking, Figure 2a) and three different βf conformers (βf-1, βf-5, and βf-6, Figure 2b) were identified, whereas for αp and βp only single (dominant) conformers (Figure 2c,d) were detected (Table 1 and Supplementary Tables 1, 3, 5, and 6). Other low energy conformers as βf-2 or βf-3 were not detected, likely due to relaxation to the most stable βf-1 through interconversion pathways (see Supplementary Figures 3 and 4). Notably, in both αp and βp, even hyperfine-splitting due to internal rotation of the methyl group could be observed and analyzed (Figure 2c,d). The resulting experimental spectroscopic parameters and experimental frequencies provide an extensive "structural map" of all possible 2-deoxy-riboside sugar scaffolds (Supporting Information).

Notably, such was the sensitivity that it allowed detection of ^{13}C isotopes in natural abundance (~1.1%). Strikingly, we were able to observe the rotational spectra for all (18 additional $^{13}C^{12}C_5H_{12}O_4$ species) of the monosubstituted ^{13}C isotopologues distributed within this small abundance for the most populated conformers αf-1, βf-1, and αp-1. As noted, correlation of moments of inertia with rotational constants renders them a key source for structural information. Moments of inertia depend critically on distribution of atomic masses, and so molecular geometry, and are affected by vibrational energy even in the ground state, since molecules are not rigid systems. Except for simplified cases, the explicit correction of the moments of inertia for vibrational contributions has not been possible because of related experimental difficulties. Therefore, different procedures and/or evaluations of molecular structures were tested, all exploiting isotopologue spectra (Figure 3b). *Equilibrium structures* (r_e, Figure 3b) represent the hypothetical vibration-less state that corresponds to the minimum of the potential energy surface. This structure, essentially inaccessible to molecules, is the one obtained by typical computational methods. *Effective structures* (r_0, Figure 3b)[47] reproduce rotational constants in the ground state and were derived here starting from the calculated geometry and through iterative least-squares adjustments of experimental rotational constants of each deoxyriboside (21 = 3 for parent +6 × 3 for isotopologues). *Substitution structures* (r_S, Figure 3b),[48,49] obtained from the analysis of the changes of the moments of inertia resulting from a single isotopic substitution, allow determination of the atomic coordinates of the substituted atom. This "r_S method" has the advantage that it provides the position of the substituted atom free from other assumptions about molecular structure and so does not depend on computational calculations. Such r_S structures are generally assumed to be intermediate between r_0 and r_e.[45] For these reasons, such "real" experimental geometries (r_S and r_0) are not directly comparable with those obtained by typical computational methods (r_e). In this way, the accuracy of rotational spectroscopy yielded riboside structures with unprecedented resolution (Figure 3b and Supplementary Tables 8−13).

Analysis of Gas-Phase Rotational Transitions of Core DNA Scaffolds Reveals Conformers and their Inherent Populations. Rotational spectroscopy allowed not only determination of unbiased, experimentally accurate structures but also an estimation of the population ratios between these different observed conformers by analyzing the relative intensities of associated rotational transitions.[50] αf adopts preferably two conformers in the gas phase, αf-1 ($^2T_1$ *gg*, Figure 3) and αf-2 ($^2T_1$ *gt*), in a ratio αf-1/αf-2 of 80(20):20, whereas for βf, three different conformers were identified (Figure 3c): βf-1 (^4E *gg*), βf-5 (E_2 *tg*), and βf-6 ($^1T_2$ *gt*) in a 93(10):4(1):3 population ratio. Only one conformer was detected for either αp ($^4C_1$) and βp ($^1C_4$).

For αf, both observed conformers (αf-1 and αf-2) share the same ring shape $^2T_1$, stabilized by two intramolecular hydrogen bonds: O3′−H3′···O1′ and O5′−H5′···O4′ across both α and β faces. On the other side, βf-1 (*gg*) adopts a partly distorted ^4E conformation. This puckering is stabilized by a strong hydrogen bond O5′−H5′···O1′, which is viable only when the C5′-hydroxymethyl group adopts a *gg* arrangement. Notably, a "switched" loss of this stabilizing interaction in βf-5 (to *tg*) and βf-6 (to *gt*) causes them to adopt somewhat distorted puckered ring E_2 and $^1T_2$ conformations, respectively, similar to those found in methyl β-D-ribofuranoside.[41]

The pyranosides αp and βp were much more rigid in the gas phase, exhibiting single dominant conformers (Figure 3c). αp adopts a $^4C_1$ chair, while βp adopts the inverted $^1C_4$. In both, reinforcing hydrogen bond networks on the α face, akin to reinforcing hydrogen bonding networks seen previously using IR spectroscopy in hexopyranosides,[51] further drive this rigidity (for αp-1: O4′−H4′···O3′−H3′···O-1′ and for βp-1: O3′−H3′···O4′−H4′···O5′) beyond the known greater barriers to conformational interconversion found in pyranosides.

In all scaffolds, the orientation of the methyl group was in accordance with the *exo*-anomeric effect,[34,52] with values of ϕ close to either 60° or −60° for the α- (αf, αp) and β- (βp, βf) anomers, respectively. This serves to place the methyl group in a site of essentially low influence with regard to conformation. Strikingly, and in clear contrast to α-anomer αf and both pyranosides αp and βp, only in βf was the hydroxyl O3′-H3′ not engaged in hydrogen bonding—the lowest lying conformer of natural scaffold βf therefore uniquely frees OH-3′, and hence the α-face remains consequentially more accessible (and potentially reactive, *vide infra*).

Comparison of Inherent Conformers of DNA scaffolds with Solvated Structures Reveals Selectively Driven Conformational Change for 2-deoxy-β-D-ribofuranoside βf. DNA functions in both aqueous and partly hydrated (e.g., enzyme active site) environments. To study the effect of the hydration state on the structure of these core scaffolds derivatives, αf, αp, βp, and βf were subjected to conformational analysis in aqueous solution by combining NMR spectroscopy measurements ($^3J_{H,H}$ coupling constants as constraints) with molecular dynamics (MD) simulations. Calculations were performed in AMBER18,[53] implemented with the GLYCAM06 force field.[54] Notably, extended (5 μs) MD simulations *alone* performed on derivatives αf and βf grossly failed to reproduce experimental data (Supplementary

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

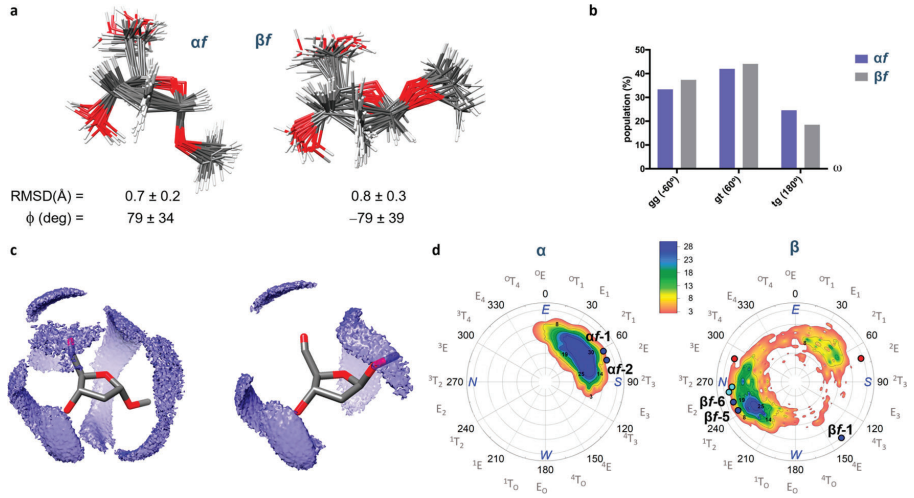


Figure 4. Conformational analysis of *αf* and *βf* in aqueous solution. (a) Structural ensembles derived from 0.2 *μs* MD-tar simulations, together with the root-mean-square deviation (RMSD) values for the heavy atoms in both anomers and the average value of ϕ torsional angle through the entire MD-tar trajectory. (b) Distribution of torsional angle ω for *αf* and *βf* derived from MD-tar simulations in aqueous solution. (c) Water oxygen density for the first hydration shell derived from experiment-guided MD simulations for *αf* and *βf*. The average structure of the carbohydrate is represented. (d) Cremer−Pople diagrams for *αf* and *βf* derived from experiment-guided MD simulations. N, S, E, and W stand for "North", "South", "East", and "West" forms. The dark blue circles represent the conformations found in the gas phase. The red circles denote the conformation found in DNA. Finally, the light blue circles show the conformations found in the gas phase for methyl *β*-D-ribofuranoside.[41] The contour coloring indicates the population (in arbitrary units) of the different conformers obtained from experiment-guided MD simulations.

Table 15), predicting a rather flexible five-membered ring with almost free pseudorotation (Supplementary Figures 5 and 6). However, MD simulations (0.2 *μs*) using experimental $^{3}J_{H,H}$ constants as time-averaged restraints (MD-tar simulations,[55] exploiting additional restraining potential terms added to the force field) successfully predicted ensembles of low-energy conformers that quantitatively reproduced NMR data (see Figure 4 and also Supplementary Tables 14 and 15 for $^{3}J_{H,H}$ constants).

These simulations revealed that in the furanosides *αf* and *βf* the C-5′ hydroxymethyl group is notably flexible in water (Figure 4b) and preferentially adopts *gg* and *gt* staggered rotamers for both. Of note, no significant intramolecular hydrogen bonds were detected in solution, with a population <10% in all cases. Analysis of the first hydration shell predicts that *αf* can accommodate more water molecules than *βf* (19 versus 15, Figure 4c). Together these solvent interactions drive scaffold-specific conformer selection from corresponding "naked" cores. Thus, while unnatural scaffold *αf* populates $^{2}T_{1}$ and ^{2}E conformations (Figure 4d) in both gas and solution phase ("unselected"), the natural scaffold *βf* is driven from ^{4}E in the gas phase *βf*-1 to E_{2} and $^{3}T_{2}$ by the loss of intramolecular bond O5′−H5′···O1′ enabled by the observed flexibility of the C-5′ hydroxymethyl group, in agreement with conformers derived by fitting the experimental homo- and heteronuclear ^{3}J coupling constants to a two-state model[56,57] (^{2}E and $^{3}T_{2}$ for *αf* and *βf*, respectively).

For the pyranosides *αp* and *βp*, characteristic $^{3}J_{H2a,H3}$ values confirmed strong propensity to remain in chair conformations[58] observed in the gas phase, yet with inverted conformation ($^{4}C_{1} \rightarrow {}^{1}C_{4}$ and $^{1}C_{4} \rightarrow {}^{4}C_{1}$, respectively). This result is also in line with previous studies.[59] In contrast to the furanosides, unrestrained MD simulations gave good agreement with experiment (Supplementary Table 16); extended analyses of total trajectory time (Supplementary Figures 7 and 8) suggested that *αp* is driven in solution to adopt only a 82(5)% partial $^{1}C_{4}$-to-$^{4}C_{1}$ equilibrium position. As expected, no significant intramolecular hydrogen bonds (population of <10%) were detected in solution through the MD simulations for either of the anomers. On the other hand, full solvation analysis performed by MD revealed that the conformational inversions are driven by an ability to properly accommodate the water in its first hydration shell, which more than offsets *endo*-anomeric preferences. Moreover, in the case of the *α*-anomer, the $^{1}C_{4}$ conformer features a larger dipole moment relative to the $^{4}C_{1}$ chair, which could also explain why this anomer preferentially adopts the $^{1}C_{4}$ chair in aqueous solution.

DISCUSSION

There are considered to be three broad conformational determinants in DNA residues: C4′−C5′ rotation (i.e., *g*/*t*); base position/"flip" (i.e., *syn*/*anti*); and ring pucker (i.e., "North"/"South") pseudorotation (Figure 4d). Although to some extent linked, sometimes even mutually "geared"[60] and accessible via typically relatively low barriers, it is the latter from these that appears to dominate function. In general terms, therefore many biomolecular interactions with such "DNA sugar scaffolds" are found to favor function with either North or South sugar ring conformers.

A North-vs-South (N-vs-S) delineation is an observation made well beyond the more obvious and static structures, such as in for example A- or B-form duplex DNA. Thus, through the use of structural biology, spectroscopy, and elegant probe molecules (e.g., conformationally restricted variants[26,61,62]), a clear role for the conformation of deoxynucleosides in

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

determining their biological function has emerged in recent decades.[63] For example, clear "N-vs-S" preferences have been seen in the function of base deaminases,[64] base C-methyltransferases,[60] kinases,[65] and even receptors[66] and transporters.[67] This can have profound effects; DNA-polymerizing enzymes, such as reverse transcriptase[68] or polymerases,[69] can have striking apparent selectivity; AZT, for example, is bound in essentially an exclusively North form by HIV reverse transcriptase, and this conformational mimicry appears critical to its potency.[68]

As a result, the delineation in this work of the inherent preferences of the available ribosides *αf*, *αp*, *βp*, *βf* and their behavior upon hydration allows consideration of the fundamental biases of such scaffolds. Notably, it is the natural scaffold *βf* that displays a unique "tipping point"/"knife-edge" conformational behavior that allows the greatest malleability by environment. Here, we observe population of kinetically trapped Southern (*βf*-1) and Northern conformers (*βf*-5, *βf*-6) as well as the "tipping" by water of *βf*-1 into "Northern" conformers. Associated assumptions have often been made around the central role of "sugar pucker". For example, the itinerary of the equilibrium between North and South has always been assumed to proceed "via the East".[62] Yet, here we see that *βf* sits in a semi-Western (indeed, SW tending to W) ^{4}E *βf*-1 pucker as its baseline conformer. Any engagement by "naked" *βf*-1 with the environment (e.g., added solvent or biomolecule) therefore starts from this point.

The partial population of clear Northern conformers *βf*-5 and *βf*-6 even when "naked" also suggests that *βf* is conformationally "primed". The unnatural deoxyriboside scaffold variants *αp*, *βp*, and even *αf* do not show this behavior. The switched loss of strength of the O5′−H5′··· aglycone hydrogen bond seen here (either through hydration, protein binding, or covalent capping, e.g., as O5′-P phosphoester) thus precludes the occurrence of this pucker leading to the adoption, in turn, of Northern (i.e., E_2 or $^1T_2$ structures upon hydration) highly similar to those of *βf*-5 and *βf*-6. In this way, *βf* is therefore uniquely suitable for a switchable (e.g., South-to-North) transition, unlike all of the other scaffolds studied here—it is "ready to switch". Notably, the C2′ hydroxylated scaffold of *βf* (the RNA scaffold equivalent of *βf*) is already tipped to North ($^3T_2$), further highlighting the uniqueness of the *βf* 2-deoxyriboside scaffold.

Notably, in *βf* the OH3′ hydroxyl group is not engaged in a hydrogen bond in the lowest lying energy conformer (*βf*-1, as well as in fact *βf*-6) found in the gas phase; *βf*-5 has interaction with OH5′ but contributes only 4% to the global population. Together, these combined factors ("tipping point" conformation and exposed OH3′) allow us to speculate further on the functional benefit of the *βf* scaffold over others. N ↔ S conformational transition allows strong modulation of the O5′-to-O3′ distance; a switchable system with OH3′ exposed at modulated distances for the reaction would allow a mechanism for ready positional alteration and hence the semiexpulsion of ligand that is thought to be critical for reducing product inhibition (and hence allowing more efficient turnover) in processive enzymes such as nucleotidyl polymerases and reverse transcriptases.[70] In other words, the ready positional manipulation of a "free OH3′" uniquely in the *βf* scaffold is potentially beneficial in many biocatalytic (and hence functional) scenarios. More generally, it might even explain the observed direction of such polymerizations of DNA in nature in growing from 5′-to-3′ (via such a modulated, free OH3′). This may not be restricted to polymerases given the wide N-vs-S preferences noted above. In several nucleoside/tide kinases, which operate at the O5′-site of ribosides, a critical and evolutionarily conserved role has been identified for OH3′: hydrogen bonding to homologous Tyr-Glu diad motifs is required in several to ensure catalysis.[71]

It is also tempting to speculate that, given the context of DNA polymerases as archetypal models for elegant but sometimes nonaligned "induced fit" hypotheses (for an excellent comparative discussion see ref 72), this may be part of broader conformational, "prechemistry" mechanisms—potential triggers for induced fit.

Regardless, it is clear that the unique flexibility that we observe for "naked" *βf* for the first time is striking and likely an important determinant in its utilization/selection by nature as a scaffold from many stereo- and constitutional sugar isomers. Although such conformational "fitness" has been the subject of insightful prior speculation,[73] our work provides the first direct experiment of such a flexible trigger based on inherent conformational tendencies in the *βf* scaffold.

Finally, given the low barriers for DNA conformational interchange and the discrepancies that we observed in the work presented here exploiting existing QM and MM methods, the structures we present here (nonreliant on QM) should prove valuable in benchmarking future quantum mechanical models. In this way, our combined approach opens up the exploration of DNA's mechanistic tendencies to even larger systems in the future.

METHODS

Synthesis. The 2-deoxyribosides *αf*, *βf*, *αp*, and *βp* (Figure 1) were synthesized by modification of a previously published literature procedure.[74] Briefly, for access to *αp* and *βp*, 2-deoxy-D-ribose was dissolved in 1% HCl in methanol and stirred for 18 h. After the reaction workup, pure *αp* was obtained through purification by column chromatography. The remaining 2-deoxyribosides *βp*, *αf*, and *βf* coeluted as an inseparable mixture, but recrystallization of this mixture in diethyl ether afforded pure *βp*. For access to *αf* and *βf*, 2-deoxy-D-ribose was dissolved in 0.1% HCl in methanol and stirred for 15 min. After the reaction workup, the mixture of *αf* and *βf* was acetylated using acetic anhydride in pyridine and pure *αf*-2OAc and *βf*-2OAc was obtained through purification by column chromatography. Deacetylation using K_2CO_3 in methanol yielded pure *αf* and *βf*. Full experimental details can be found in the Supporting Information.

Pulse Fourier Transform Microwave Spectrometers. To obtain the rotational spectra of the furanose species, we used a chirped-pulse Fourier transform microwave spectrometer (CP-FTMW) built at the Spectroscopy Group at the University of the Basque Country (UPV/EHU)[44] and following the design of Pate and co-workers (Supplementary Figure 1).[75] A short broadband pulse (1 μs, 12 GHz, frequency range 6−18 GHz) is generated by an Arbitrary Waveform Generator (AWG) and amplified by a Traveling Wave Tube Amplifier (TWTA). The pulse is broadcast into a high vacuum chamber (down to 10^{-6} mbar) through a Q-par horn antenna where it interacts with the molecular supersonic jet expansion. The time-domain signal of the molecular emission is collected by another horn antenna and delivered to a digital 20 GHz oscilloscope and Fourier-transformed to obtain the rotational spectrum in the frequency domain. The spectral resolution of this apparatus is ∼10 kHz.

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

Rotational spectra of pyranose conformers were recorded in a 4−18 GHz Fourier transform microwave spectrometer (FTMW) based on the Balle-Flygare design,[76] constructed at the UPV/EHU and described elsewhere (Supplementary Figure 2).[43] Appropriate excitation pulses create optimum $\pi/2$ polarization conditions on the sample, which expands in a supersonic jet coaxially within a Fabry-Pérot microwave resonator. The resulting transient spontaneous emission from the expanding molecular ensemble is amplified and down-converted to the radio frequency region, where it is digitized, and Fourier transformed to yield the frequency-domain spectrum. All transitions recorded in this spectrometer appear split into two Doppler components, because of the coaxial arrangement of the molecular jet and the resonator. The higher spectral resolution of this apparatus (better than 5 kHz) allowed us to resolve the hyperfine line splittings due to internal rotation of methyl tops, where present.

Samples in the CP-FTMW spectrometer were vaporized by a conventional heating method by wrapping a heating wire to a customized nozzle at 145 °C. The vacuum chamber hosts three pulsed solenoid valves running simultaneously in order to increase the signal-to-noise ratio. Gaseous samples were mixed with He carrier gas at 6 bar, and the resulting flow was injected into the vacuum chamber through a small orifice (1 mm diameter), producing the gas expansion and subsequent cooling (rotational temperature near 5 K). The duration of the molecular pulse was 0.5−1.2 ms. Two nozzles were enough for αf, whereas three nozzles were used for βf, at 30 chirps per molecular pulse. To obtain a good signal-to-noise ratio, a total of 20 and 40 Mcycles were collected for the αf and βf spectra, respectively. In the FTMW spectrometer equipped with UV ultrafast laser vaporization,[39] it consists of a mixture of the chemical compound and a commercial binder, mechanically pressed to get cylinder-type solid rods that are inserted in the apparatus and rotated by a stepper motor. A picosecond pulse from the third harmonic (355 nm) of a Nd:YAG laser hits the rod in the presence of Ne carrier gas (∼6 bar), delivering the molecules intact into the gas phase. This technique prevents sample decomposition, that often happens with biomolecules upon thermal heating.

Computational Details. *Conformational Search.* The conformational search used a molecular mechanics approach with MMFFs, OPLS, and AMBER force fields in a 20 kJ/mol energy window and was followed by geometry optimizations with quantum chemistry calculations [*ab initio* MP2 and DFT (B3LYP-D3)] with 6-311++G(d,p) basis functions. Gaussian 16 software[77] was used in all cases.

Unrestrained MD Simulations. Simulations were performed with the AMBER18 package,[53] implemented with a GLYCAM06j[54] force field. Each molecule was immersed in a water box with a 10 Å buffer of TIP3P water molecules. A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal mol^{-1} were applied to the solute, and the Andersen temperature coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects are modeled using the particle-mesh-Ewald method.[78] An 8 Å cutoff was applied to Lennard−Jones interactions. Each system was equilibrated for 2 ns with a 2 fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for an additional 5 μs under the same simulation conditions.

Experiment-Guided MD Simulations. The setup of the MD simulations with time-averaged restraints was identical, in terms of force fields, water model, and algorithms to that described above for the unrestrained MD simulations. The experimental $^3J_{H,H}$ coupling constants were imposed as a time-averaged restraint, applying a linear averaging. The equilibrium ^3J range was set to $^3J_{exp} - 0.2$ Hz $\leq {}^3J_{exp} \leq {}^3J_{exp} + 0.2$ Å. Trajectories were run at 300 K, with a decay constant of 20 ns and a time step of 1 fs. The force constants rk_2 and rk_3 used in each case were 0.2 kcal mol^{-1} Å^{-2}. The overall simulation length was 0.2 μs. The theoretical ^3J coupling constants were deduced from the dihedral values through the corresponding Altona equation and Sweet J software.[58,79]

Analysis of the First Hydration Shell from the MD Simulations. The water density properties were derived from the production trajectories using a cubic grid consisting of 150 × 150 × 150 bins with 0.5-Å spacing, through the *"grid"* command available in the *cpptraj* module of AMBER18. The surfaces represented in Figures 4 and Supplementary Figure 7 correspond to oxygen water density for the first hydration shell.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.9b01277.

Supplementary Figure 1: Scheme of the broadband chirped-pulse Fourier Transform microwave spectrometer. Supplementary Figure 2: Scheme of the molecular beam pulsed-jet Fourier Transform microwave spectrometer. Supplementary Figures 3−4: Interconversion paths between the lowest lying conformers. Supplementary Figure 5: Structural ensembles derived from unrestrained MD simulations in aqueous solution for αf and βf. Supplementary Figure 6: Distribution of torsional angle ω for αf and βf derived from unrestrained MD simulations in aqueous solution. Supplementary Figure 7: Conformational analysis of αp and βp in aqueous solution. Supplementary Figure 8: Evolution of the main conformers of αf (b) and βf (b) along experiment-guided MD simulations in aqueous solution. Supplementary Figures 9−20: Copies of ^{1}H and ^{13}C NMR spectra. Supplementary Tables 1−7: Experimental spectroscopic parameters and comparison with theoretical predictions of DNA Scaffolds. Supplementary Tables 8−10: Comparison between experimental (r_0) and calculated (r_e) structural parameters of DNA Scaffolds. Supplementary Tables 11−13: Experimental (r_0 and r_S) and calculated (r_e) Cartesian coordinates of DNA Scaffolds. Supplementary Table 14: Karplus equations used in the MD simulations. Supplementary Tables 15−16: Comparison between the experimental and theoretical coupling constants derived from MD simulations (PDF)

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

AUTHOR INFORMATION

Corresponding Authors

Benjamin G. Davis − *Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford OX1 3TA, United Kingdom; The Rosalind Franklin Institute, Oxfordshire OX11 0FA, United Kingdom;* orcid.org/0000-0002-5056-407X; Email: ben.davis@chem.ox.ac.uk, ben.davis@rfi.ac.uk

Francisco Corzana − *Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain;* orcid.org/0000-0001-5597-8127; Email: francisco.corzana@unirioja.es

Emilio J. Cocinero − *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain; Instituto Biofisika (CSIC, UPV/EHU), 48080 Bilbao, Spain;* orcid.org/0000-0001-7632-3728; Email: emiliojose.cocinero@ehu.es

Authors

Camilla Calabrese − *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain; Instituto Biofisika (CSIC, UPV/EHU), 48080 Bilbao, Spain;* orcid.org/0000-0003-4299-2098

Iciar Uriarte − *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain; Instituto Biofisika (CSIC, UPV/EHU), 48080 Bilbao, Spain*

Aran Insausti − *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain; Instituto Biofisika (CSIC, UPV/EHU), 48080 Bilbao, Spain*

Montserrat Vallejo-López − *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain*

Francisco J. Basterretxea − *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain*

Stephen A. Cochrane − *Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford OX1 3TA, United Kingdom;* orcid.org/0000-0002-6239-6915

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.9b01277

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank MINECO (Projects CTQ2017-89150 R, CTQ2015 68148-C2-2P), Ministerio de Ciencia, Innovación y Universidades (RTI2018-099592-B-C21), Basque Government (IT1162-19, PIBA 2018/11), the UPV/EHU (PPG17/10 and GIU18/207), CSIC (PIC2018, LINKA20249) Fundación BBVA and the Wellcome Trust (110270/A/15/Z) for financial support. C.C. and A.I. thank MINECO and Basque Government for Juan de la Cierva and predoctoral contracts, respectively. Laser and NMR resources of the UPV/EHU together with computational resources of BERONIA (Universidad de La Rioja) and UPV/EHU (SGIker and I2Basque) were used in this work.

REFERENCES

(1) Pabo, C. O.; Sauer, R. T. Protein-DNA recognition. *Annu. Rev. Biochem.* **1984**, *53*, 293−321.
(2) Pabo, C. O.; Sauer, R. T. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem.* **1992**, *61*, 1053−1095.
(3) Rohs, R.; et al. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.
(4) Higgs, P. G.; Lehman, N. The RNA World: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* **2015**, *16*, 7−17.
(5) Duzdevich, D.; Redding, S.; Greene, E. C. DNA Dynamics and Single-Molecule Biology. *Chem. Rev.* **2014**, *114*, 3072−3086.
(6) Barrangou, R.; Horvath, P. A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* **2017**, *2*, 17092.
(7) Rinaldi, C.; Wood, M. J. A. Antisense oligonucleotides: The next frontier for treatment of neurological disorders. *Nat. Rev. Neurol.* **2018**, *14*, 9−22.
(8) Wittrup, A.; Lieberman, J. Knocking down disease: A progress report on siRNA therapeutics. *Nat. Rev. Genet.* **2015**, *16*, 543−552.
(9) Herdewijn, P.; Marlière, P. Redesigning the leaving group in nucleic acid polymerization. *FEBS Lett.* **2012**, *586*, 2049−2056.
(10) Campbell, M. A.; Wengel, J. Locked vs. unlocked nucleic acids (LNA vs. UNA): Contrasting structures work towards common therapeutic goals. *Chem. Soc. Rev.* **2011**, *40*, 5680−5689.
(11) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* **2012**, *487*, 196−201.
(12) Arangundy-Franklin, S.; et al. A synthetic genetic polymer with an uncharged backbone chemistry based on alkyl phosphonate nucleic acids. *Nat. Chem.* **2019**, *11*, 533−542.
(13) Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure. *Science* **1999**, *284*, 2118−2124.
(14) Schöning, K.-U.; et al. Chemical Etiology of Nucleic Acid Structure: The α-Threofuranosyl-(3'→2') Oligonucleotide System. *Science* **2000**, *290*, 1347−1351.
(15) Gaweda, K.; Plazinski, W. The systematic influence of solvent on the conformational features of furanosides. *Org. Biomol. Chem.* **2019**, *17*, 2479−2485.
(16) Cheung, T., Ramesh, V. Biomolecular NMR Spectroscopy and Structure Determination of DNA. *Biomolecular and Bioanalytical Techniques* **2019**, 421..
(17) Watson, J. D.; Crick, F. H. C. Molecular Structures of Nucleic Acids. *Nature* **1953**, *171*, 737−738.
(18) Wang, A. H.; Fujii, S.; van Boom, J. H.; Rich, A. Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 3968−3972.
(19) Wang, A. H.-J.; Hakoshima, T.; van der Marel, G.; van Boom, J. H.; Rich, A. AT base pairs are less stable than GC base pairs in Z-DNA: The crystal structure of d(m5CGTAm5CG). *Cell* **1984**, *37*, 321−331.
(20) Larsen, T. A.; Kopka, M. L.; Dickerson, R. E. Crystal structure analysis of the B-DNA dodecamer CGTGAATTCACG. *Biochemistry* **1991**, *30*, 4443−4449.
(21) Shapiro, Y. E. NMR spectroscopy on domain dynamics in biomacromolecules. *Prog. Biophys. Mol. Biol.* **2013**, *112*, 58−117.
(22) Pérez, A.; Luque, F. J.; Orozco, M. Frontiers in Molecular Dynamics Simulations of DNA. *Acc. Chem. Res.* **2012**, *45*, 196−205.
(23) Imeddourene, A. B.; Xu, X.; Zargarian, L.; Oguey, C.; Foloppe, N.; Mauffret, O.; Hartmann, B. The intrinsic mechanics of B-DNA in solution characterized by NMR. *Nucleic Acids Res.* **2016**, *44*, 3432−3447.
(24) Schwieters, C. D.; Clore, G. M. A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* **2007**, *46*, 1152−1166.
(25) Tjandra, N.; Tate, S. I.; Ono, A.; Kainosho, M.; Bax, A. The NMR structure of a DNA dodecamer in an aqueous dilute liquid crystalline phase. *J. Am. Chem. Soc.* **2000**, *122*, 6190−6200.

(26) Taha, H. A.; Richards, M. R.; Lowary, T. L. Conformational analysis of furanoside-containing mono- and oligosaccharides. *Chem. Rev.* **2013**, *113*, 1851−1876.
(27) Olson, W. K.; Sussman, J. L. How Flexible Is the Furanose Ring? 1. A Comparison of Experimental and Theoretical Studies. *J. Am. Chem. Soc.* **1982**, *104*, 270−278.
(28) Talbot, F. O.; Simons, J. P. Sugars in the gas phase: the spectroscopy and structure of jetcooled phenyl beta-D-glucopyranoside. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3562−3565.
(29) Jockusch, R. A.; Talbot, F. O.; Simons, J. P. Sugars in the gas phase. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1502−1507.
(30) Barry, C. S.; et al. 'Naked' and Hydrated Conformers of the Conserved Core Pentasaccharide of N-linked Glycoproteins and Its Building Blocks. *J. Am. Chem. Soc.* **2013**, *135*, 16895−16903.
(31) Stanca-Kaposta, E. C.; et al. Solvent Interactions and Conformational Choice in a Core N-Glycan Segment: Gas Phase Conformation of the Central, Branching Trimannose Unit and its Singly Hydrated Complex. *J. Am. Chem. Soc.* **2008**, *130*, 10691−10696.
(32) Su, Z.; Cocinero, E. J.; Stanca-Kaposta, E. C.; Davis, B. G.; Simons, J. P. Carbohydrate-aromatic interactions: A computational and IR spectroscopic investigation of the complex, methyl α-l-fucopyranoside · toluene, isolated in the gas phase. *Chem. Phys. Lett.* **2009**, *471*, 17−21.
(33) Screen, J.; et al. IR-Spectral Signatures of Aromatic−Sugar Complexes: Probing Carbohydrate−Protein Interactions. *Angew. Chem., Int. Ed.* **2007**, *46*, 3644−3648.
(34) 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*, 76−80.
(35) Pérez, C.; et al. Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy. *Science* **2012**, *336*, 897−901.
(36) Patterson, D.; Schnell, M.; Doyle, J. M. Enantiomer-specific detection of chiral molecules via microwave spectroscopy. *Nature* **2013**, *497*, 475−477.
(37) Motiyenko, R. A.; Alekseev, E. A.; Dyubko, S. F.; Lovas, F. J. Microwave spectrum and structure of furfural. *J. Mol. Spectrosc.* **2006**, *240*, 93−101.
(38) Fritzsche, J. Rotationsspektroskopische Untersuchungen von Kohlehydraten im LASER-Plasma − Ablation von D-(−)-Ribose. Diplomarbeit, Gottfried-Wilhelm-Leibniz Univ. Hann., 2006.
(39) Cocinero, E. J.; et al. Ribose found in the gas phase. *Angew. Chem., Int. Ed.* **2012**, *51*, 3119−3124.
(40) Peña, I.; et al. Six pyranoside forms of free 2-deoxy-D-ribose. *Angew. Chem., Int. Ed.* **2013**, *52*, 11840−11845.
(41) Écija, P.; et al. Furanosic forms of sugars: Conformational equilibrium of methyl β-D-ribofuranoside. *Chem. Commun.* **2016**, *52*, 6241−6244.
(42) Calabrese, C.; et al. Conformational Behavior of D-Lyxose in Gas and Solution Phases by Rotational and NMR Spectroscopies. *J. Phys. Chem. Lett.* **2019**, *10*, 3339−3345.
(43) Cocinero, E. J.; et al. Conformational equilibria in vanillin and ethylvanillin. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12486−12493.
(44) Uriarte, I.; Perez, C.; Caballero-Mancebo, E.; Basterretxea, F. J.; Lesarri, A.; Fernandez, J. A.; Cocinero, E. J. Structural Studies of Nicotinoids: Cotinine versus Nicotine. *Chem. - Eur. J.* **2017**, *23*, 7238−7244.
(45) Gordy, W.; Cook, R. L. *Microwave Molecular Spectra*; Interscience: New York, 1984.
(46) Watson, J. K. D. *Vibrational Spectra and Structure*; Elsevier, 1977.
(47) Rudolph, H. D. Contribution to the systematics of r0-derived molecular structure determinations from rotational parameters. *Struct. Chem.* **1991**, *2*, 581−588.
(48) Kraitchman, J. Determination of Molecular Structure from Microwave Spectroscopic Data. *Am. J. Phys.* **1953**, *21*, 17−24.
(49) Costain, C. C. Determination of molecular structures from ground state rotational constants. *J. Chem. Phys.* **1958**, *29*, 864−874.
(50) Fraser, G. T.; Suenram, R. D.; Lugez, C. L. Rotational spectra of seven conformational isomers of 1-hexene. *J. Phys. Chem. A* **2000**, *104*, 1141−1146.
(51) Cocinero, E. J.; et al. Conformational choice and selectivity in singly and multiply hydrated monosaccharides in the gas phase. *Chem. - Eur. J.* **2008**, *14*, 8947−8955.
(52) Wiberg, K. B.; Bailey, W. F.; Lambert, K. M.; Stempel, Z. D. The Anomeric Effect: It's Complicated. *J. Org. Chem.* **2018**, *83*, 5242−5255.
(53) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. *Amber 2018*; University of California: San Francisco, 2018.
(54) Kirschner, K. N.; et al. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.
(55) Hendrickx, P. M. S.; Corzana, F.; Depraetere, S.; Tourwe, D. A.; Augustyns, K.; Martins, J. C. The use of time-averaged 3JHH Restrained molecular dynamics (tar-MD) simulations for the conformational analysis of five-membered ring systems: Methodology and applications. *J. Comput. Chem.* **2009**, *31*, 561−572.
(56) van Lieshout, H. C.; Raap, J.; van Boom, J. H.; Haasnoot, C. A. G. Conformations of Methyl 2'-Deoxy-α-D-ribofuranoside and Methyl 2'-Deoxy-β-D-ribofuranoside. A Proton Magnetic Resonance Spectroscopy and Molecular Mechanics Study. *J. Am. Chem. Soc.* **1988**, *110*, 2736−2743.
(57) Church, T. J.; Carmichael, I.; Serianni, A. S. 13C-1H and 13C-13C spin-coupling constants in methyl β-D-ribofuranoside and methyl 2-deoxy-β-D-erythro-pentofuranoside: Correlations with molecular structure and conformation. *J. Am. Chem. Soc.* **1997**, *119*, 8946−8964.
(58) Donders, L. A.; De Leeuw, F. A. A. M.; Altona, C. Relationship between proton—proton NMR coupling constants and substituent electronegativities. IV—An extended karplus equation accounting for interactions between substituents and its application to coupling constant data calculated by the Extended Hückel. *Magn. Reson. Chem.* **1989**, *27*, 556−563.
(59) Franks, F.; Lillford, P. J.; Robinson, G. Isomeric equilibria of monosaccharides in solution. Influence of solvent and temperature. *J. Chem. Soc., Faraday Trans. 1* **1989**, *85*, 2417−2426.
(60) Wang, P.; et al. Use of Oligodeoxyribonucleotides with Conformationally Constrained Abasic Sugar Targets To Probe the Mechanism of Base Flipping by HhaI DNA (Cytosine C5)-methyltransferase. *J. Am. Chem. Soc.* **2000**, *122*, 12422−12434.
(61) Marquez, V. E.; et al. Nucleosides with a twist. Can fixed forms of sugar ring pucker influence biological activity in nucleosides and oligonucleotides? *J. Med. Chem.* **1996**, *39*, 3739−3747.
(62) Mathé, C.; Périgaud, C. Recent Approaches in the Synthesis of Conformationally Restricted Nucleoside Analogues. *Eur. J. Org. Chem.* **2008**, *2008*, 1489−1505.
(63) Birnbaum, G. I.; Shugar, D. *Topics in Nucleic Acid Structure*; Macmillan Press, 1997.
(64) Marquez, V. E.; et al. Contrasting behavior of conformationally locked carbocyclic nucleosides of adenosine and cytidine as substrates for deaminases. *Nucleosides, Nucleotides Nucleic Acids* **2009**, *28*, 614−632.
(65) Sjuvarsson, E.; Marquez, V. E.; Eriksson, S. Selective Phosphorylation of South and North-Cytidine and Adenosine Methanocarba-Nucleosides by Human Nucleoside and Nucleotide Kinases Correlates with Their Growth Inhibitory Effects on Cultured Cells. *Nucleosides, Nucleotides Nucleic Acids* **2015**, *34*, 544−564.
(66) Jacobson, K. A.; et al. Methanocarba Analogues of Purine Nucleosides as Potent and Selective Adenosine Receptor Agonists. *J. Med. Chem.* **2000**, *43*, 2196−2203.
(67) Damaraju, V. L.; et al. Influence of Sugar Ring Conformation on the Transportability of Nucleosides by Human Nucleoside Transporters. *ChemBioChem* **2011**, *12*, 2774−2778.
(68) Marquez, V. E.; et al. HIV-1 reverse transcriptase can discriminate between two conformationally locked carbocyclic AZT triphosphate analogues. *J. Am. Chem. Soc.* **1998**, *120*, 2780−2789.

(69) Eoff, R. L.; et al. Selective modulation of DNA polymerase activity by fixed-conformation nucleoside analogues. *Angew. Chem., Int. Ed.* **2010**, *49*, 7481−7485.
(70) Majumdar, C.; Abbotts, J.; Broder, S.; Wilson, S. H. Studies on the mechanism of human immunodeficiency virus reverse transcriptase. Steady-state kinetics, processivity, and polynucleotide inhibition. *J. Biol. Chem.* **1988**, *263*, 15657−15665.
(71) Johansson, K.; et al. Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. *Nat. Struct. Biol.* **2001**, *8*, 616−620.
(72) Mulholland, A. J.; Roitberg, A. E.; Tuñón, I. Enzyme dynamics and catalysis in the mechanism of DNA polymerase. *Theor. Chem. Acc.* **2012**, *131*, 1286.
(73) Eschenmoser, A. Chemistry of potentially prebiological natural products. *Origins Life Evol. Biospheres* **1994**, *24*, 389−423.
(74) Deriaz, R. E.; Overend, W. G.; Stacey, M.; Wiggins, L. F. Deoxy-sugars. Part VI. The constitution of β-methyl-2-deoxy-L-ribopyranoside and of $\alpha\beta$-methyl-2-deoxy-L-ribofuranoside. *J. Chem. Soc.* **1949**, *0*, 2836−2841.
(75) Brown, G. G.; et al. A broadband Fourier transform microwave spectrometer based on chirped pulse excitation. *Rev. Sci. Instrum.* **2008**, *79*, 053103.
(76) Balle, T. J.; Flygare, W. H. Fabry-Perot cavity pulsed Fourier transform microwave spectrometer with a pulsed nozzle particle source. *Rev. Sci. Instrum.* **1981**, *52*, 33−45.
(77) Frisch, M. J. et al. *Gaussian 16*; Gaussian, Inc.: Wallingford, CT, 2016.
(78) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(79) Balacco, G. A desktop calculator for the Karplus equation. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 885−887.

https://dx.doi.org/10.1021/acscentsci.9b01277
ACS Cent. Sci. 2020, 6, 293−303

Methyl Jasmonate and Zingerone

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Cite This: J. Phys. Chem. Lett. 2018, 9, 5906−5914

Letter

pubs.acs.org/JPCL

Competing Dispersive Interactions: From Small Energy Differences to Large Structural Effects in Methyl Jasmonate and Zingerone

Iciar Uriarte,[‖] Aran Insausti, and Emilio J. Cocinero*

Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), Barrio Sarriena, 48940 Leioa, Spain

Biofisika Institute (CSIC, UPV/EHU), Universidad del País Vasco (UPV/EHU), Apartado 644, E-48080 Bilbao, Spain

Atef Jabri[‖] and Isabelle Kleiner*

Laboratoire Interuniversitaire des Systémes Atmosphériques CNRS/IPSL UMR 7583, Universités Paris-Est et Paris Diderot, 61 Avenue de General De Gaulle 94010 Créteil, France

Halima Mouhib

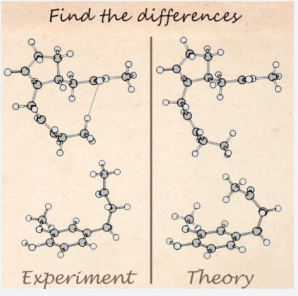
Laboratoire Modélisation et Simulation Multi Echelle MSME UMR 8208 CNRS, Université Paris-Est, 5 Boulevard Descartes, 77454 Marne-La-Vallée, France

Ibon Alkorta

Centro de Química Orgánica "Lora Tamayo", Instituto de Química Médica (IQM-CSIC), Juan de la Cierva, 3, 28006 Madrid, Spain

Supporting Information

ABSTRACT: Modern structural studies of biologically relevant molecules require an exhaustive interplay between experiment and theory. In this work, we present two examples where a poor choice of the theoretical method led to a misinterpretation of experimental results. We do that by performing a rotational spectroscopy study on two large and flexible biomolecules: methyl jasmonate and zingerone. The results show the enormous potential of rotational spectroscopy as a benchmark to evaluate the performance of theoretical methods.

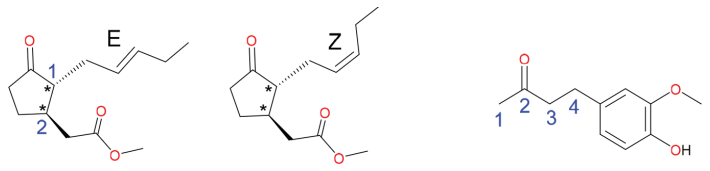


In the past decade, structural studies on large, flexible organic molecules of biological interest moved into the focus of interest. The three-dimensional conformation that a biomolecule adopts is intimately related to its function in the biological medium. Hence, a lot of effort is devoted to the structural determination of biomolecules of increasing size. Structural studies carried out in condensed media can provide an environment comparable to biological surroundings, but the solvent or crystal packing forces may mask some key interactions. High-resolution structural studies in a supersonic jet benefit from an interaction-free environment where biomolecules can be probed without any kind of interference. The intrinsic properties of the molecule under study can be analyzed and separated from the effects arising from solvent or crystal packing. Regardless of the high-resolution spectroscopic technique (laser spectroscopy, microwave, etc..), the general strategy in these studies employs an interplay between experimental results and theoretical calculations (molecular mechanics, molecular dynamics, and quantum chemistry). This symbiosis between theory and experiment is nowadays a cornerstone in structural studies. Because of the increasing size and degrees of freedom of the molecular systems targeted, there is a need for precise theoretical models that support the interpretation of experimental results. On the other hand, theoretical methods still require a strong and reliable experimental counterpart to safeguard the results of any implemented model. Especially the combination of rotational spectroscopy with highly accurate quantum chemistry is a

Received: July 30, 2018
Accepted: September 20, 2018
Published: September 20, 2018

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

Scheme 1. Chemical Structures of Methyl Jasmonate and Zingerone



Methyl jasmonate (1R, 2R) Zingerone

powerful tool to study the molecular structures of large and flexible organic molecules. The efficiency of this strategy was previously shown in biologically relevant cases such as amino acids,[1,2] sugars,[3−6] nucleobases,[7−9] odorants,[10−12] and some drugs.[13−15]

Recently, the unprecedented speed of rotational spectra acquisition obtained with broadband chirped-pulse Fourier transform microwave (CP-FTMW) spectrometers has opened new perspectives in structural chemistry.[16−20] Hence, it is possible to study larger molecular systems (15−25 heavy atoms) such as biomolecules[21] or large organic compounds,[22,23] and their weakly bound molecular complexes.[24−26] In addition, the implementation of laser vaporization sources allows one to study thermolabile systems, avoiding the undesirable thermal decomposition processes that occur when using conventional heating techniques.[1,5] The experimental progress brought by CP-FTMW spectroscopy now demands the development of more accurate, faster and efficient theoretical methods to analyze and fit the experimental data.

One of the crucial points in the predictive value of the key parameters (structure, energy, dipole moments, quadrupolar tensor, ...) of larger biomolecular systems is the choice of the theoretical method. There is a large array of quantum chemical methods for that purpose. For instance, B3LYP[27,28] was historically very important, its cost-effectiveness making it very commonly used. However, the need to characterize weakly bound systems forced the inclusion of empirical parametrizations such as Grimme's approach[29−32] (D, D2, D3, D4 and its subvariants) in order to account for dispersive interactions. Minnesota functionals[33] are more modern, more sophisticated, and highly parametrized, and Møller−Plesset perturbation theory[34] is very reasonable for small-to-medium systems but not without limitations. Hence, the choice of method is usually subject to a trade-off between accuracy and computational cost.

Systems with only covalent interactions or relatively strong hydrogen bonds (e.g., O−H···O, O−H···N, N−H···O, 15−25 kJ mol^{-1}) can be described at almost any reasonable level of theory. However, systems stabilized by O−H···halogen, C−H···O, C−H···N, C−H···π, halogen−halogen, π···π, lone-pair···π, van der Waals, or London dispersive interactions require specific approaches to be modeled properly. In many cases, the choice of the theoretical method is done routinely without taking into account the key interactions of the system. In addition, it is well-known that some methods are able to describe very well some weak interactions but fail at characterizing others.[35,36]

In the present study, we demonstrate that the choice of the appropriate quantum chemical method can be vital in the structural determination of biomolecules. In order to do so, we use the case studies of methyl jasmonate and zingerone, two of the main flavor components of jasmine and ginger, respectively (see Scheme 1 for chemical formulas).[37,38] The results reported in this work clearly demonstrate the importance of carefully choosing the appropriate theoretical method for the molecular system under study. We can anticipate that a poor choice of the theoretical method may compromise the interpretation of the experimental results.

Both substances are widely present in plants. Methyl jasmonate is a vital lipidic phytohormone produced by many plants in response to biotic stress (e.g., pathogens or pests) and abiotic stress (e.g., heat, cold, wind, precipitation). It is an important signaling molecule and plays a crucial role in the communication of plants.[39,40] Zingerone is the main component of ginger and it is responsible for its pungent scent. It has been shown to exhibit diuretic and neuroprotective functions, which are still not fully understood at a molecular scale.[41] Besides their interesting biological effects, these molecules represent limiting cases with respect to their size and molecular weights, for both the theoretical and the experimental studies presented in this work.

Case Study 1: Methyl Jasmonate. Methyl jasmonate (MJ) consists of a cyclopentanone ring with two large flexible side chains (see Scheme 1). There are two chiral centers in the molecule, meaning that there are two diastereoisomers to consider (*RR* and *RS*). The corresponding enantiomers (*SS* and *SR*) would be equivalent by symmetry and therefore, they would give rise to the same rotational spectra. Furthermore, the double bond in one of the side chains may adopt *Z* or *E* configurations. Therefore, there are four different structural isomers to consider (*RR-Z*, *RR-E*, *RS-Z*, and *RS-E*). Each one of them may adopt different conformations arising from rotations along single bonds. Moreover, there could be attractive intramolecular interactions (such as hydrogen bonds) between the two chains. We can infer from these structural and conformational degrees of freedom that the conformational casuistry of methyl jasmonate is very broad and complex.

The first step in this study was a fast conformational search with molecular mechanics (MM), using Merck's molecular force field (MMFF). About 100 plausible structures for each system were found in an energy window of 20 kJ mol^{-1}. Hence, 4 × ~100 ≈ 400 structures had to be reoptimized with more accurate theoretical calculations to predict the relevant rotational parameters (structures, energies, dipole moments, ...). At that stage, we chose the B3LYP method, a fast and reasonably accurate method, to guide the experiment and reoptimized all the structures.

In the meantime, the rotational spectrum of methyl jasmonate was recorded in the frequency region of 6−18 GHz using the CP-FTMW spectrometer available at the University of the Basque Country (UPV/EHU), see Figure S1 in the Supporting Information. The sample was heated up to 200 °C in a customized heating nozzle in order to build up enough concentration of methyl jasmonate in the gas phase. Two sets

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

Table 1. Microwave Spectroscopic Results of the Most Abundant Conformers of Methyl Jasmonate (RR, Z conformation of the double bond) and Zingerone Observed in the Molecular Jet

		Methyl Jasmonate		Zingerone
const.[a]	Units	Conf. MJ1	Conf. MJ2	Conf. Z1
A_0	MHz	634.08213 (87)	562.186 (74)	1023.7533 (22)
B_0	MHz	370.027517 (80)	382.806 (23)	434.84931 (53)
C_0	MHz	264.431520 (73)	260.36873 (80)	352.90045 (32)
Δ_J (D_J)	kHz	0.009815 (62)[b]	0.355 (37)[b]	0.06592 (70)[c]
Δ_{JK} (D_{JK})	kHz	0.3351 (21)[b]	3.55 (87)[b]	0.1088 (84)[c]
Δ_K (D_K)	kHz	0.1388 (16)[b]	−15.97 (2.51)[b]	0[h]
d_J (d_1)	kHz	0[h]	0.171 (19)[b]	−0.00648 (55)[c]
d_K (d_2)	kHz	0[h]	2.76 (45)[b]	−0.00173 (34)[c]
V_3[d]	cm^{-1}	446.03 (1.67)	427.88 (83)	336.474 (81)
σ[e]	kHz	17.1	18.7	22.2
N_{tot}[f]		580	215	183
N_A/N_E[g]		299/281	114/101	105/78

[a]The microwave data set was fitted using the program XIAM.[40,41] *A*, *B*, *C*: rotational constants. Δ_J (D_J), Δ_{JK} (D_{JK}), Δ_K (D_K), δ_J (d_1), δ_K (d_2): centrifugal distortion constants for Watson's A reduction for methyl jasmonate and (S) reduction for zingerone, *I*′ representation. Errors of the experimental data given in parentheses are in units of the last digit. [b]Watson's A reduction Δ_J, Δ_{JK}, Δ_K, δ_J, and δ_K are used for methyl jasmonate conformer 1 (Ray's kappa = −0.43) and for conformer 2 (Ray's kappa = −0.19). [c]Watson's S reduction D_J, D_{JK}, D_K, d_1 and d_2 are used for zingerone (Ray's kappa = −0.75). [d]Barrier to internal rotation of the methoxy methyl group and carbonyl methyl group in methyl jasmonate and zingerone, respectively. [e]Standard deviation of the fit. [f]Total number of fitted lines. [g]Number of fitted A and E species; [h]Parameters set to zero.

Table 2. Comparison of the Experimental Values of A_e, B_e and C_e with the Theoretical Values Calculated with Different Theoretical Methods (Basis Set: 6-311++G(d,p)) for Conf. MJ1 (a) and Conf. MJ2 (b) of Methyl Jasmonate[a]

a.

	Exp	B3LYP	TPSS	PBE	B2PLYP	B3LYP-D3	B3LYP-D3BJ	B2PLYP-D3BJ	MP2	M06-2X	B97-D3	B97-D3BJ
A_e	641.37313(87)	549.8	550.7	572.6	600.9	643.1	644.3	633.9	642.0	662.8	637.1	623.3
%error A_e		**14.3**	**14.1**	**10.7**	**6.3**	**-0.3**	**-0.5**	**1.2**	**-0.1**	**-3.3**	**0.7**	**2.8**
B_e	371.505517(80)	366.1	366.5	368.2	369.2	368.1	367.7	369.3	370.8	373.9	365.7	366.7
%error B_e		**1.4**	**1.3**	**0.9**	**0.6**	**0.9**	**1.0**	**0.6**	**0.2**	**-0.7**	**1.6**	**1.3**
C_e	267.38152(73)	245.1	245.1	250.4	257.4	265.1	265.7	264.3	266.7	269.6	263.1	260.9
%error C_e		**8.3**	**8.3**	**6.3**	**3.7**	**0.9**	**0.6**	**1.2**	**0.3**	**-0.8**	**1.6**	**2.4**
CPU time (h)/opt cycle		**0.8**	**3.2**	**2.3**	**15.8**	**0.9**	**0.9**	**13.8**	**5.1**	**1.0**	**1.7**	**0.9**

b.

	Exp	B3LYP	TPSS	PBE	B2PLYP	B3LYP-D3	B3LYP-D3BJ	B2PLYP-D3BJ	MP2	M06-2X	B97-D3	B97-D3BJ
A_e	560.931(74)	477.2	488.0	501.2	532.2	573.1	568.6	564.9	572.9	588.2	567.9	556.6
%error A_e		**14.9**	**13.0**	**10.6**	**5.1**	**-2.2**	**-1.4**	**-0.7**	**-2.1**	**-4.9**	**-1.2**	**0.8**
B_e	388.143(23)	382.5	381.2	382.0	382.4	380.0	380.9	381.6	382.5	382.5	378.0	379.3
%error B_e		**1.4**	**1.8**	**1.6**	**1.5**	**2.1**	**1.9**	**1.7**	**1.5**	**1.5**	**2.6**	**2.3**
C_e	261.00473(80)	237.4	240.5	243.9	253.8	261.5	261.1	262.2	266.7	267.6	259.7	258.3
%error C_e		**9.1**	**7.9**	**6.6**	**2.8**	**-0.2**	**0.0**	**-0.5**	**-2.2**	**-2.5**	**0.5**	**1.0**
CPU time (h)/opt cycle		**1.0**	**2.6**	**2.0**	**16.0**	**1.0**	**1.3**	**14.8**	**5.8**	**2.2**	**0.8**	**2.0**

[a]Note that the experimental values of B_e are calculated such that $B_e = B_0 - \Delta B_{vib}$, and the anharmonic correction ΔB_{vib} is estimated for each conformer separately at the B97D3/def2-SVP level of theory. Values shaded in orange correspond to Conf. MJ1′ (part a) and Conf. MJ2′ (part b).

of rotational transitions were spotted in the spectrum. They were first fitted separately to two semirigid asymmetric rotor Hamiltonians using Watson's A reduction.[42] The rotational transitions in the spectrum displayed splittings due to tunneling effects from internal rotation of a methyl top. Those splittings have to be accounted for in the fit by appropriate operators in the

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

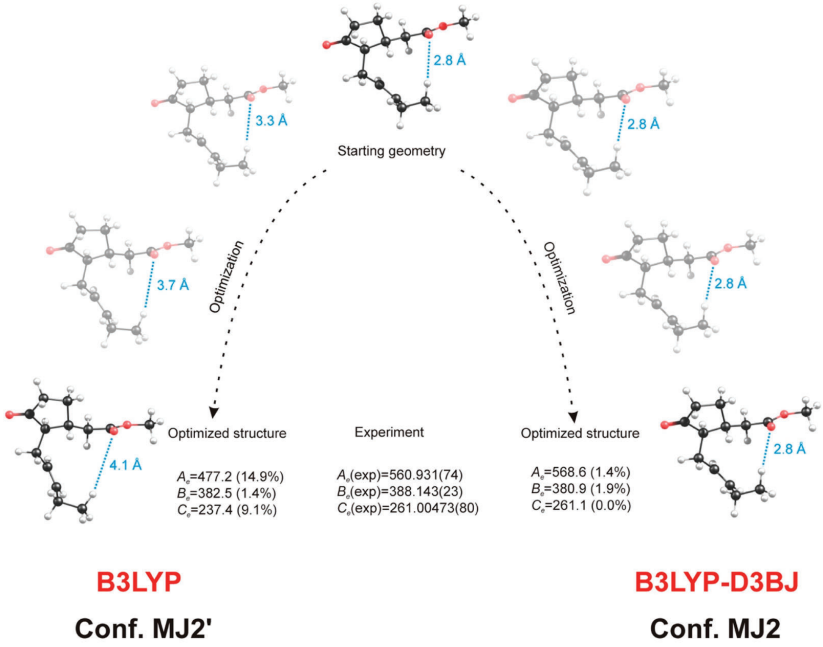


Figure 1. Steps of the optimization of Conf. MJ2′ and Conf. MJ2 of methyl jasmonate (RR, Z conformation of the double bond) using the B3LYP and B3LYP-D3BJ DFT methods, respectively, from the same starting geometry (basis set: 6-311++G(d,p)).

Hamiltonian, allowing for the estimation of the barrier for internal rotation of the methyl top (V_3). We used two different publicly available software programs (XIAM[43,44] and BELGI[45]) for fitting the experimental results. A number of previous studies have shown that the XIAM code and the BELGI code are complementary.[46−48] As described in the Supporting Information (effective Hamiltonian section), the XIAM code takes into account one torsional state at a time, and has a limited number of high order Hamiltonian terms implemented, whereas the BELGI-C1 code treats all torsional states together in one matrix and has many higher order terms. While XIAM benefits from the advantages of its ease of use and speed, in cases with rather low barriers, E species transitions are not always satisfactorily predicted because the interactions between torsional states can be large. Considering the importance of carefully deriving experimental rotational constants to compare with *ab initio* methods, we decided to use both methods in this work. In the present case, the results were similar (see Tables S1−S3), and in the following, we will only refer to the XIAM fits, which are summarized in Table 1. The experimental values of rotational and torsional parameters for the two conformers of MJ are presented in full in Table S1 of the Supporting Information.

The common strategy used in the microwave spectroscopy community to assign a molecular conformation to a given set of experimental rotational constants is to compare them to a predicted spectrum generated with the values arising from quantum chemical calculations. However, all attempts to assign the spectrum based on different structural models of methyl jasmonate failed. This spectrum did not correspond to any of the predicted conformations for methyl jasmonate using the B3LYP method. Finally, we considered the possibility of sample fragmentation resulting in decomposition products or internal reactions, but the considered species did not match the spectrum either. By this point, we had carried out more than 1000 calculations but none of them reproduced the experiment. Obviously, the B3LYP method failed to provide accurate structures of the most abundant conformers of methyl jasmonate. Therefore, we decided to redo the optimizations of methyl jasmonate using the MP2 *ab initio* method, which finally led to structures that were in agreement with our experimental results. Its lowest lying energy conformers (Conf. MJ1 and Conf. MJ2) had not been detected by the B3LYP method, and these two were the structures that reproduced the experimental data. Although optimizations at the B3LYP level may be largely sufficient for average sized systems, there is a limit, when intramolecular interactions become too important and cannot be described properly by this method. Indeed, for methyl jasmonate the B3LYP method was insufficient to provide accurate structures that are consistent with the highly accurate experimental results.

Finally, we carried out numerous calculations with different methods to compare their performance in reproducing experimental values. We always used the same basis set (6-311++G(d,p)). Theoretical methods yield the rotational constants at the bottom of the potential well (commonly known as B_e equilibrium structure) while the experimental rotational constants refer to the structure at the vibrational ground state v = 0 (B_0). We estimated $\Delta B_{vib} = B_0 - B_e$ with anharmonic calculations at the B97D3/def2-SVP level of theory and subtracted it from the experimental rotational constants in order to obtain the semiempirical values of B_e. The latter are directly comparable to the rotational constants obtained from quantum chemical calculations.[22,49] The results are summarized in Table 2 for the two observed conformers of methyl jasmonate.

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

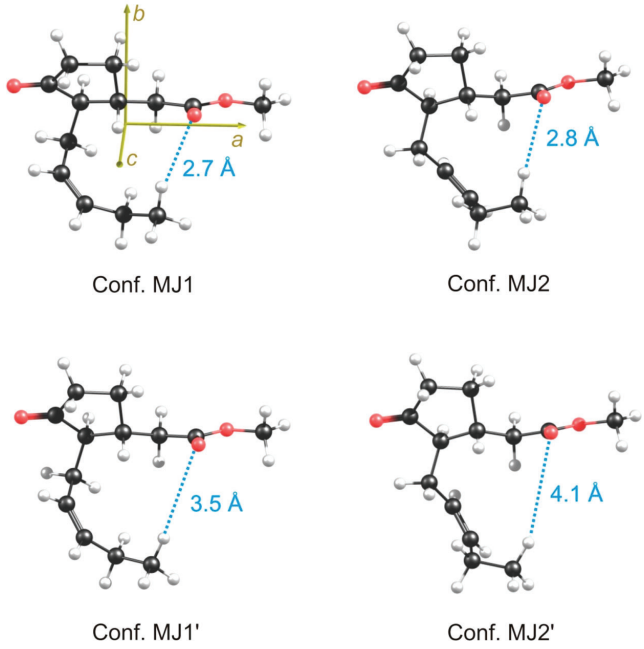


Figure 2. Two experimentally observed conformations of methyl jasmonate (*RR*, *Z* conformation of the double bond) (Conf. MJ1 and Conf. MJ2) as optimized at the B3LYP-D3BJ/6-311++G(d,p) level of theory, as well as the corresponding conformations Conf. MJ1′ and Conf. MJ2′ calculated at the B3LYP/6-311++G(d,p) level of theory. For Conf. MJ1, we also show the principal axes of inertia *a*, *b*, and *c*.

It is noteworthy that the B3LYP method, which does not account for dispersion effects, fails at reproducing the experimental structure (relative errors in *A* around 15%). This is also the case for other methods such as PBE, TPSS or B2PLYP (although B2PLYP provides significantly better results, with ∼6% error in the *A* rotational constant). Only *ab initio* and DFT methods that considered dispersion effects (either intrinsically or by adding a correction) were successful at reproducing experimental values. In this case, the Minnesota functional M06-2X gives worse results (relative errors in *A* around 4−5%) than DFT methods corrected by Grimme's corrections (relative errors 0−2%). Figure 1 displays the steps of the optimization of Conf. MJ2 with the B3LYP and B3LYP-D3BJ methods starting from the same geometry. We can clearly see that with B3LYP, as the optimization progresses, the two arms move away from each other, the final distance CH···O being 4.1 *Å* (conformer Conf. MJ2′). Hence, the CH···O interaction is not accounted for by the B3LYP method. On the contrary, using B3LYP-D3BJ, the two arms remain close to each other, held together by the CH···O weak interaction (final distance 2.8 Å, Conf. MJ2). We infer that the dispersion correction D3BJ has a dramatic effect on the rotational constants. In particular, the *A* rotational constant is strongly influenced by dispersion in the case of methyl jasmonate because the principal axis *a* lies along the two side chains. Hence, the moment of inertia I_a (to which *A* is inversely proportional) is very sensitive to the coordinates of the atoms in the side chains (see Figure 2 for the principal axes of inertia). The effect of dispersion is so pronounced in this molecule that the assignment to the molecular conformations was not possible with the B3LYP methodology.

To summarize, the two experimental conformations are assigned to Conf. MJ1 and Conf. MJ2: both of them are *RR* diasteroisomers (or *SS*), the configuration of the double bond in the side chain is *Z* and both conformations are stabilized by a C−H···O weak hydrogen bond (WHB). The corresponding 3D structures are depicted in Figure 2, along with the structures of Conf. MJ1′ and Conf. MJ2′, i.e., the conformations predicted from calculations where no dispersion is included.

Case Study 2: Zingerone. Zingerone (see Scheme 1) consists of an *o*-methoxyphenol moiety with a flexible chain in *para* position with respect to the hydroxy group. In order to guide the spectral search, we predicted the most favorable structures employing the same strategy as described above and commonly used in the literature: we performed MM to have an initial screening of the conformational scope (14 structures were found in 15 kJ mol^{-1}) and then we reoptimized all of them with the MP2 robust method.

Later, the rotational spectrum of zingerone in the 6−18 GHz frequency region was recorded using the FTMW and CP-FTMW spectrometers at the University of the Basque Country (UPV/EHU) see Figure S2 of the Supporting Information. The sample was heated to 200 °C in a customized heating nozzle in order to bring it into the gas phase. A set of rotational transitions corresponding to the A species of one conformation of the molecule was assigned and fitted to a semirigid asymmetric rotor Hamiltonian in Watson's S-reduction. Splittings between A and E species due to the internal rotation were accounted for in the fit by appropriate operators in the Hamiltonian. The results of the fit are summarized in Table 1.

Surprisingly, there was no correspondence between experiment and theory. The discrepancy in the rotational constants

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

was remarkable. Differences of ∼17% indicated that the predicted structure was not the experimentally observed structure, despite having used a method as reliable as MP2 theory.

In this case, we decided to reoptimize all the structures using DFT theory, in particular the B3LYP-D3BJ functional. Interestingly, the results generated with this functional were not comparable to those obtained with the MP2 method. With both methods, all the lowest lying conformations were stabilized by an OH···O interaction between the hydroxy and the methoxy group. However, the global minimum was different for the MP2 and B3LYP-D3BJ calculations. Additionally, the lowest lying structure predicted by B3LYP-D3BJ was not predicted using the MP2 method. The two conformations predicted as global minima in zingerone are shown in Figure 3. The first one (Conf. Z1, MP2 structure exhibits an attractive CH···π interaction between the terminal CH_3 in the flexible chain and the π cloud in the aromatic ring. This interaction forces the side chain to tilt toward the ring. It should be noted that this structure is not predicted by B3LYP (or TPSS, PBE, and B2PLYP, i.e., methods that do not account for dispersive interactions). In the second conformation (Conf. Z2, B3LYP-D3BJ structure), no such interaction occurs and the side chain is perpendicular to the ring. This conformer exhibits a C═O···H−C attractive interaction. This conformation was not predicted by the MP2 method. To summarize, we have two possible global minima, one of which was not predicted at the MP2/6-311++G(d,p) level of theory.

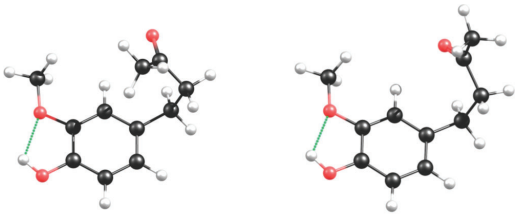


Figure 3. Conformations of zingerone predicted as global minima with the MP2 (Conf. Z1) and B3LYP-D3BJ methods (Conf. Z2), using the basis set 6-311++G(d,p).

In addition, we extended the predictions to other theoretical models. Table 3 compiles the results of those calculations. Clearly, the result of the optimizations highly depend on the applied theoretical method. The balance between the CH···π interaction (Conf. Z1) and the CH···O interaction (Conf. Z2) is very subtle. It is a challenging task for theoretical models to weigh the strength of the two interactions, since both are weak hydrogen bonds (WHB) with small stabilization energies below 4 kcal mol^{-1}.[50] Regarding the theoretical methods, three different trends are observed. They are summarized in Figure 4 with the three common methods MP2, B3LYP, and B3LYP-D3BJ. We can observe that B3LYP-D3BJ is capable of reproducing both conformations if the correct starting geometry is chosen. B3LYP always yields Conf. Z2 stabilized by CH···O interaction, that is, with the lateral chain perpendicular to the plane of the aromatic ring. Finally, MP2 calculations result always in Conf. Z1 with a CH···π interaction, where the chain is bent toward the benzene ring. Table 3 shows the differences in

Table 3. Comparison of the Experimental Values of A_e, B_e, and C_e with the Theoretical Values Calculated with Different Theoretical Methods (Basis Set: 6-311++G(d,p)) for Conf. Z2 (a) and Conf. Z1 (b) of Zingerone[a]

a.	Exp	B3LYP	TPSS	PBE	B2PLYP	B3LYP-D3	B3LYP-D3BJ	B2PLYP-D3BJ	MP2	M06-2X	B97-D3	B97-D3BJ
A_e	1029.5243(22)	1034.5	1021.8	1024.6	1021.2	1016.9	1017.8	1009.9	NA*	1028.5	1001.7	NA
%error A_e		**-0.5**	**0.8**	**0.5**	**0.8**	**1.2**	**1.1**	**1.9**	NA	**0.1**	**2.7**	NA
B_e	441.49935(53)	421.4	423.0	422.7	433.2	437.5	439.4	444.7	NA	441.7	436.9	NA
%error B_e		**4.6**	**4.2**	**4.3**	**1.9**	**0.9**	**0.5**	**-0.7**	NA	**-0.1**	**1.0**	NA
C_e	356.47486(32)	340.6	342.9	341.4	351.4	352.3	354.2	360.6	NA	357.6	351.7	NA
%error C_e		**4.5**	**3.8**	**4.2**	**1.4**	**1.2**	**0.6**	**-1.1**	NA	**-0.3**	**1.3**	NA
CPU time/opt cycle		**0.4**	**0.006**	**0.5**	**4.7**	**0.8**	**0.4**	**4.7**	NA	**0.5**	**0.004**	NA

b.	B3LYP	TPSS	PBE	B2PLYP	B3LYP-D3	B3LYP-D3BJ	B2PLYP-D3BJ	MP2	M06-2X	B97-D3	B97-D3BJ
A_e	NA*	NA	NA	NA	991.7	994.6	994.2	994.7	999.8	985.5	987.7
B_e	NA	NA	NA	NA	483.6	487.1	488.9	509.3	503.5	482.4	486.5
C_e	NA	NA	NA	NA	390.4	395.5	398.2	413.6	407.5	390.2	395.0
CPU time/opt cycle	NA	NA	NA	NA	0.6	1.4	7.5	2.5	1.9	0.6	0.006

[a]Note that the experimental values of B_e are calculated such that $B_e = B_0 - \Delta B_{vib}$, and the anharmonic correction ΔB_{vib} is estimated for each conformer separately at the B97D3/def2-SVP level of theory. NA = Not Applicable, the optimization leads either to Conf. Z1 (part a) or Conf Z2 (part b).

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

the rotational constants. Again, the choice of the theoretical method is crucial to success. The rotational constants are very sensitive to both conformers and we will only be able to discern unambiguously the structure of zingerone through microwave spectroscopy.

By comparing the experimental values of the rotational constants with the theoretical predictions, we can conclude that the experimental conformation of zingerone is Conf. Z2, which is stabilized by a CH···O interaction. This means that MP2/6 311++G(d,p), which is a very commonly used and most trusted method in the rotational spectroscopy community, fails to predict the experimentally observed conformation of zingerone as minimum energy structure. In fact, the observed conformer is not even predicted by this method. MP2 overestimates the CH···π interaction and, as the balance between WHBs is so subtle, the real structure is not predicted by the MP2 method. In order to visualize the subtlety of the competition between the two WHBs, we calculate the barrier between the two structures by performing a relaxed scan of the 1−2−3−4 dihedral angle (see Scheme 1 for atom numbering). Moreover, we perform this scan using two different methods (M06-2X and B3LYP-D3BJ) to evaluate the response of both of them in the WHBs competition. Figure 5 displays the two scans. We can observe the local minima corresponding to Conf. Z1 and Conf. Z2 in both scans. It is worth noting that these minima are almost isoenergetic. The energy difference between both of them is less than 0.5 kJ mol^{-1} (this value is probably within the error of the theoretical calculation). Moreover, M06-2X and B3LYP-D3BJ predict that the barrier between the two structures is very low (~1 kJ mol^{-1}), meaning that it can be easily overcome in the supersonic expansion and therefore, the system would relax to the most stable conformation. This behavior has been previously observed in rotational spectra and is well reported in the literature.[51,52] The most relevant difference between the two theoretical methods is precisely the energetic order of the conformers. While M06-2X predicts that the most stable conformation is Conf. Z1 (CH···π interaction), B3LYP-D3BJ predicts Conf. Z2 as the most stable structure (CH···O interaction). The small energy difference and the low barrier between the two conformations proves the subtlety of the competition between the two interactions, with different theoretical methods tipping the balance in favor of one structure or the other.

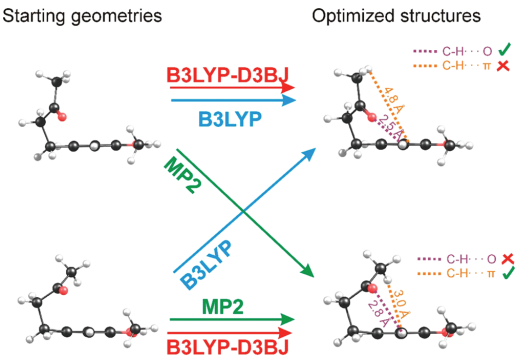


Figure 4. Schematic representation of the prototype behavior of theoretical methods in the conformational casuistry of zingerone.

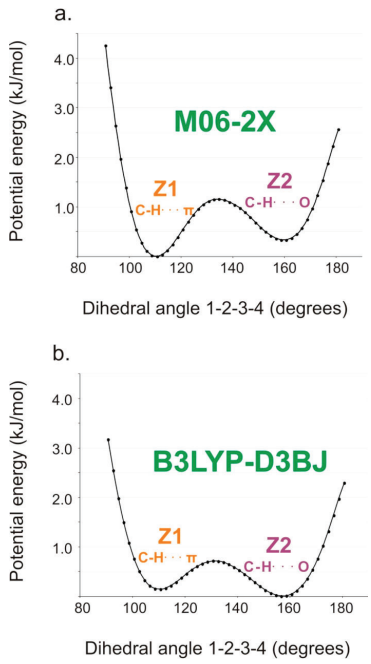


Figure 5. Potential energy scans of the 1−2−3−4 dihedral angle of zingerone to evaluate the energy barrier between Conf. Z1 (stabilized by a C−H···π interaction) and Conf. Z2 (stabilized by a C−H···O interaction). Two different theoretical methods have been used: M06-2X (a) and B3LYP-D3BJ (b), in order to evaluate their behavior with respect to the zingerone molecular system. The basis set is 6-311++G(d,p).

This study demonstrates that rotational spectroscopy combined with theoretical methods is a powerful strategy, which can be broadly applied to flexible organic compounds. However, one must be cautious in the use of different theoretical methods. There is currently no strategy to avoid brute force optimizations in this kind of molecular systems (for example, force fields are poorly parametrized for the accurate description of averaged sized biologically active ligands). Therefore, benchmark studies such as the ones presented here are extremely necessary. We can infer from the studies on methyl jasmonate and zingerone that using several distinct levels of theory allows one to get a good overview of the conformational landscape. This work demonstrates that microwave spectroscopy can benchmark new theoretical models very much in need to guide the experiments.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.8b02339.

Experimental and computational details, rotational spectra of methyl jasmonate and zingerone, experimental spectroscopic parameters of all observed species, and a list of experimental rotational transitions (PDF)

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

AUTHOR INFORMATION

Corresponding Authors

*(E.J.C.) E-mail: emiliojose.cocinero@ehu.es.
*(I.K.) E-mail: kleiner@lisa.u-pec.fr.

ORCID

Emilio J. Cocinero: 0000-0001-7632-3728
Ibon Alkorta: 0000-0001-6876-6211

Author Contributions

‖These authors contributed equally

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.J., H.M., and I.K. would like to acknowledge the CNRS InFinity program and the OSU/Efluve for funding; H.M. thanks the European COST initiative MOLIM for a STSM fellowship; E.J.C. and I.U. thank the MINECO (CTQ2017-89150-R) and the UPV/EHU (UFI11/23 and PPG17/10) for funds. Computational resources and laser facilities of the UPV-EHU were used in this work (SGIker). Computational resources of the MSME laboratory of the University Paris-Est were used in this work.

REFERENCES

(1) Lesarri, A.; Mata, S.; Cocinero, E. J.; Blanco, S.; López, J. C.; Alonso, J. L. The Structure of Neutral Proline. *Angew. Chem., Int. Ed.* **2002**, *41*, 4673−4676.

(2) Lesarri, A.; Sánchez, R.; Cocinero, E. J.; López, J. C.; Alonso, J. L. Coded Amino Acids in Gas Phase: The Shape of Isoleucine. *J. Am. Chem. Soc.* **2005**, *127*, 12952−12956.

(3) Cocinero, E. J.; Çarçabal, P.; Vaden, T. D.; Simons, J. P.; Davis, B. G. Sensing the Anomeric Effect in a Solvent-Free Environment. *Nature* **2011**, *469*, 76−80.

(4) Cocinero, E. J.; Gamblin, D. P.; Davis, B. G.; Simons, J. P. The Building Blocks of Cellulose: The Intrinsic Conformational Structures of Cellobiose, Its Epimer, Lactose, and Their Singly Hydrated Complexes. *J. Am. Chem. Soc.* **2009**, *131*, 11117−11123.

(5) Cocinero, E. J.; Lesarri, A.; Écija, P.; Basterretxea, F. J.; Grabow, J. U.; Fernández, J. A.; Castaño, F. Ribose Found in the Gas Phase. *Angew. Chem., Int. Ed.* **2012**, *51*, 3119−3124.

(6) Écija, P.; Uriarte, I.; Spada, L.; Davis, B. G.; Caminati, W.; Basterretxea, F. J.; Lesarri, A.; Cocinero, E. J. Furanosic Forms of Sugars: Conformational Equilibrium of Methyl β-D-Ribofuranoside. *Chem. Commun.* **2016**, *52*, 6241−6244.

(7) Nir, E.; Kleinermanns, K.; De Vries, M. S. Pairing of Isolated Nucleic-Acid Bases in the Absence of the DNA Backbone. *Nature* **2000**, *408*, 949−951.

(8) Alonso, J. L.; Peña, I.; López, J. C.; Vaquero, V. Rotational Spectral Signatures of Four Tautomers of Guanine. *Angew. Chem., Int. Ed.* **2009**, *48*, 6141−6143.

(9) Favero, L. B.; Uriarte, I.; Spada, L.; Écija, P.; Calabrese, C.; Caminati, W.; Cocinero, E. J. Solving the Tautomeric Equilibrium of Purine through Analysis of the Complex Hyperfine Structure of the Four ^{14}N Nuclei. *J. Phys. Chem. Lett.* **2016**, *7*, 1187−1191.

(10) Mouhib, H.; Stahl, W.; Lüthy, M.; Büchel, M.; Kraft, P. Cassis Odor through Microwave Eyes: Olfactory Properties and Gas-Phase Structures of All the Cassyrane Stereoisomers and Its Dihydro Derivatives. *Angew. Chem., Int. Ed.* **2011**, *50*, 5576−5580.

(11) Nguyen, H. V. L.; Mouhib, H.; Klahm, S.; Stahl, W.; Kleiner, I. A Touch of Lavender: Gas-Phase Structure and Dynamics of the Monoterpene Linalool Validated by Microwave Spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10012−10018.

(12) Loru, D.; Quesada-Moreno, M. M.; Avilés-Moreno, J. R.; Jarman, N.; Huet, T. R.; López-González, J. J.; Sanz, M. E. Conformational Flexibility of Limonene Oxide Studied By Microwave Spectroscopy. *ChemPhysChem* **2017**, *18*, 274−280.

(13) León, I.; Millán, J.; Cocinero, E. J.; Lesarri, A.; Fernández, J. A. Water Encapsulation by Nanomicelles. *Angew. Chem., Int. Ed.* **2014**, *53*, 12480−12483.

(14) León, I.; Millán, J.; Cocinero, E. J.; Lesarri, A.; Fernández, J. A. Shaping Micelles: The Interplay between Hydrogen Bonds and Dispersive Interactions. *Angew. Chem., Int. Ed.* **2013**, *52*, 7772−7775.

(15) Lesarri, A.; Vega-Toribio, A.; Suenram, R. D.; Brugh, D. J.; Grabow, J. U. The Conformational Landscape of the Volatile Anesthetic Sevoflurane. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9624−9631.

(16) Brown, G. G.; Dian, B. C.; Douglass, K. O.; Geyer, S. M.; Shipman, S. T.; Pate, B. H. A Broadband Fourier Transform Microwave Spectrometer Based on Chirped Pulse Excitation. *Rev. Sci. Instrum.* **2008**, *79*, 053103.

(17) Shipman, S. T.; Pate, B. H. New Techniques in Microwave Spectroscopy. *Handbook of High-resolution spectroscopy* **2011**, 801−827.

(18) Neill, J. L.; Shipman, S. T.; Alvarez-Valtierra, L.; Lesarri, A.; Kisiel, Z.; Pate, B. H. Rotational Spectroscopy of Iodobenzene and Iodobenzene-Neon with a Direct Digital 2−8 GHz Chirped-Pulse Fourier Transform Microwave Spectrometer. *J. Mol. Spectrosc.* **2011**, *269*, 21−29.

(19) Pérez, C.; Muckle, M. T.; Zaleski, D. P.; Seifert, N. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H. Structures of Cage, Prism, and Book Isomers of Water Hexamer from Broadband Rotational Spectroscopy. *Science* **2012**, *336*, 897−901.

(20) Pérez, C.; Lobsiger, S.; Seifert, N. A.; Zaleski, D. P.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H. Broadband Fourier Transform Rotational Spectroscopy for Structure Determination: The Water Heptamer. *Chem. Phys. Lett.* **2013**, *571*, 1−15.

(21) Uriarte, I.; Pérez, C.; Caballero-Mancebo, E.; Basterretxea, F. J.; Lesarri, A.; Fernández, J. A.; Cocinero, E. J. Structural Studies of Nicotinoids: Cotinine versus Nicotine. *Chem. - Eur. J.* **2017**, *23*, 7238−7244.

(22) Fokin, A. A.; Zhuk, T. S.; Blomeyer, S.; Pérez, C.; Chernish, L. V.; Pashenko, A. E.; Antony, J.; Vishnevskiy, Y. V.; Berger, R. J. F.; Grimme, S.; et al. Intramolecular London Dispersion Interaction Effects on Gas-Phase and Solid-State Structures of Diamondoid Dimers. *J. Am. Chem. Soc.* **2017**, *139*, 16696−16707.

(23) Domingos, S. R.; Cnossen, A.; Buma, W. J.; Browne, W. R.; Feringa, B. L.; Schnell, M. Cold Snapshot of a Molecular Rotary Motor Captured by High-Resolution Rotational Spectroscopy. *Angew. Chem., Int. Ed.* **2017**, *56*, 11209−11212.

(24) Seifert, N. A.; Zaleski, D. P.; Pérez, C.; Neill, J. L.; Pate, B. H.; Vallejo-Lõpez, M.; Lesarri, A.; Cocinero, E. J.; Castaño, F.; Kleiner, I. Probing the C-H···π Weak Hydrogen Bond in Anesthetic Binding: The Sevoflurane-Benzene Cluster. *Angew. Chem., Int. Ed.* **2014**, *53*, 3210−3213.

(25) Steber, A. L.; Pérez, C.; Temelso, B.; Shields, G. C.; Rijs, A. M.; Pate, B. H.; Kisiel, Z.; Schnell, M. Capturing the Elusive Water Trimer from the Stepwise Growth of Water on the Surface of the Polycyclic Aromatic Hydrocarbon Acenaphthene. *J. Phys. Chem. Lett.* **2017**, *8*, 5744−5750.

(26) Ghosh, S.; Thomas, J.; Huang, W.; Xu, Y.; Jäger, W. Rotational Spectra of Two Hydrogen-Bonded Methyl Salicylate Monohydrates: Relative Stability and Tunneling Motions. *J. Phys. Chem. Lett.* **2015**, *6*, 3126−3131.

(27) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(28) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(29) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(30) Grimme, S. Accurate Description of van Der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

(31) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(32) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 Dispersion Coefficient Model. *J. Chem. Phys.* **2017**, *147*, 034112.
(33) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(34) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622.
(35) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301−5337.
(36) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275−4284.
(37) Acree, T. E.; Nishida, R.; Fukami, H. Odor Thresholds of the Stereoisomers of Methyl Jasmonate. *J. Agric. Food Chem.* **1985**, *33*, 425−427.
(38) Asamitsu, Y.; Nakamura, Y.; Ueda, M.; Kuwahara, S.; Kiyota, H. Synthesis and Odor Description of Both Enantiomers of Methyl 4,5-Didehydrojasmonate, a Component of Jasmin Absolute. *Chem. Biodiversity* **2006**, *3*, 654−659.
(39) Cheong, J. J.; Choi, Y. Do. Methyl Jasmonate as a Vital Substance in Plants. *Trends Genet.* **2003**, *19*, 409−413.
(40) Avanci, N. C.; Luche, D. D.; Goldman, G. H.; Goldman, M. H. Jasmonates Are Phytohormones with Multiple Functions, Including Plant Defense and Reproduction. *GMR, Genet. Mol. Res.* **2010**, *9*, 484−505.
(41) Fay, H. A. C. A Highly Effective and Selective Male Lure for Bactrocera Jarvisi (Tryon) (Diptera: Tephritidae). *Aust. J. Entomol.* **2012**, *51*, 189−197.
(42) Watson, J. K. G. *Vibrational Spectra and Structure*; During, J. R., Ed.; Elsevier: Amsterdam, 1977.
(43) Hartwig, H.; Dreizler, H. The Microwave Spectrum of Trans-2,3-Dimethyloxirane in Torsional Excited States. *Z. Naturforsch., A: Phys. Sci.* **1996**, *51* (8), 923−932.
(44) Hansen, N.; Mäder, H.; Bruhn, T. A Molecular Beam Fourier Transform Microwave Study of O-Tolunitrile: ^{14}N Nuclear Quadrupole Coupling and Methyl Internal Rotation Effects. *Mol. Phys.* **1999**, *97*, 587.
(45) Kleiner, I.; Hougen, J. T. Rho-Axis-Method Hamiltonian for Molecules Having One Methyl Rotor and C1 Point-Group Symmetry at Equilibrium. *J. Chem. Phys.* **2003**, *119*, 5505−5509.
(46) Nguyen, H. V. L.; Jabri, A.; Van, V.; Stahl, W. Methyl Internal Rotation in the Microwave Spectrum of Vinyl Acetate. *J. Phys. Chem. A* **2014**, *118*, 12130−12136.
(47) Eibl, K.; Kannengießer, R.; Stahl, W.; Nguyen, H. V. L.; Kleiner, I. Low Barrier Methyl Rotation in 3-Pentyn-1-ol as Observed by Microwave Spectroscopy. *Mol. Phys.* **2016**, *114*, 3483−3489.
(48) Ferres, L.; Stahl, W.; Kleiner, I.; Nguyen, H. V. L. The Effect of Internal Rotation in P-Methyl Anisole Studied by Microwave Spectroscopy. *J. Mol. Spectrosc.* **2018**, *343*, 44−49.
(49) Vogt, N.; Demaison, J.; Cocinero, E. J.; Écija, P.; Lesarri, A.; Rudolph, H. D.; Vogt, J. The Equilibrium Molecular Structures of 2-Deoxyribose and Fructose by the Semiexperimental Mixed Estimation Method and Coupled-Cluster Computations. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15555−15563.
(50) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond in Structural Chemistry and Biology*; Oxford Univ. Press: Oxford, U.K., 1999.
(51) Ruoff, R. S.; Klots, T. D.; Emilsson, T.; Gutowsky, H. S. Relaxation of Conformers and Isomers in Seeded Supersonic Jets of Inert Gases. *J. Chem. Phys.* **1990**, *93*, 3142−3150.
(52) Godfrey, P. D.; Brown, R. D.; Rodgers, F. M. The Missing Conformers of Glycine and Alanine: Relaxation in Seeded Supersonic Jets. *J. Mol. Struct.* **1996**, *376*, 65−81.

DOI: 10.1021/acs.jpclett.8b02339
J. Phys. Chem. Lett. 2018, 9, 5906−5914

Butamben and Isobutamben

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2020, **56**, 6094

Received 31st January 2020,
Accepted 3rd April 2020

DOI: 10.1039/d0cc00760a

rsc.li/chemcomm

Conformational impact of aliphatic side chains in local anaesthetics: benzocaine, butamben and isobutamben†

A. Insausti,[ab] C. Calabrese,[ab] M. Parra,[ab] I. Usabiaga,[ac] M. Vallejo-López,[a] P. Écija,[a] F. J. Basterretxea,[a] J.-U. Grabow,[d] W. Caminati,[c] A. Lesarri[e] and E. J. Cocinero*[ab]

We studied the impact of aliphatic side chains on the stability and conformational landscape of the local anaesthetics benzocaine, butamben and isobutamben, combining high-resolution rotational spectroscopy in the microwave and millimetre regions and molecular modelling. The study reveals the connections between alkyl chain flexibility and molecular conformations.

Local anaesthetics (LAs) are essential medical agents used routinely by health professionals. The demand for drugs causing sensory and motor paralysis in local areas soon spurred interest in several families of compounds with action properties different from general anaesthetics. All LAs are membrane-stabilizing drugs that reversibly alter matter exchange rates, mainly by inhibiting Na^+ and Ca^{2+} ion channels of neuronal cell membranes, thus blocking the ions responsible for nerve conduction,[1–3] as illustrated in Fig. 1a. There is a high degree of variability found in LAs, but most of them can be classified according to their chemical structure as amino amides and amino esters (Fig. 1b).

The great number and diversity of substances that can act as LAs have prompted a debate on the mechanism of activity of these compounds and their potential optimization.[4] Several studies have pointed to steric dependences, in particular orientation preferences and the length of alkyl substituents, impacting the LA's ability to alter physical properties, conformations and interactions with receptors.[4,5]

In this context, we have focused on *p*-aminobenzoate (PABA), the precursor of one of the most common molecular families of amino ester-type LAs, which include benzocaine (**BZC**), butamben (**BTN**) and isobutamben (**BTI**) (Fig. 1c). In contrast to other conventional LAs, which include a protonable amino group in the side chain, they just exhibit a hydrocarbon chain (**BZC**: ethyl; **BTN**: *n*-butyl; **BTI**: isobutyl).

Many studies have been performed on PABA derivatives using different approaches,[6–10] *e.g.* low-resolution infrared, UV-Visible, Raman, and NMR spectroscopies and X-ray diffraction. They have shown clear evidence that membrane permeability, water solubility and potency of these compounds change drastically with the size and shape of the alkyl chain, altering the specific

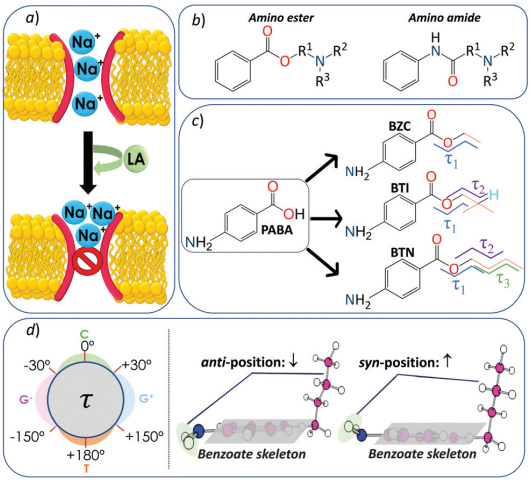


Fig. 1 (a) Inhibition of Na^+ ion flux due to the presence of local anaesthetics (LAs) in the Na^+ channels. (b) LA types classified according to their chemical structure. (c) Molecular structures of the analysed LAs: benzocaine (**BZC**), isobutamben (**BTI**) and butamben (**BTN**), defining the dihedral angles (τ_1, τ_2, and τ_3). (d) Labelling used to identify the orientation of the alkyl chains (left side) and amine protons with respect to the alkyl chain (right side). *E.g.*, the $TG^+\uparrow$ of **BTI** corresponds to the rotamer with: τ_1 = *trans* (178.1°), τ_2 = *gauche* + (+57.8°) and a *syn* orientation of the protons of the amine group.

[a] *Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), Leioa, 48940, Spain. E-mail: emiliojose.cocinero@ehu.es; Web: http://grupodeespectroscopia.es/MW/*
[b] *Instituto Biofísika (CSIC, UPV/EHU), Leioa, 48940, Spain*
[c] *Dipartimento di Chimica "G. Ciamician", Università di Bologna, via Selmi 2, Bologna 40126, Italy*
[d] *Institut für Physikalische Chemie und Elektrochemie, Gottfried Wilhelm Leibniz Universität Hannover, Hannover, 30167, Germany*
[e] *Departamento de Química Física y Química Inorgánica – IU CINQUIMA, Universidad de Valladolid, Valladolid, 47011, Spain*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cc00760a

action mechanisms.[10] However, there is a lack of information about the 3D-structure of these side chains. Hence, an accurate study of the conformational scope of the PABA-type LAs is necessary. It is of particular interest to evaluate how the alkyl chain length and branching affect the intrinsic conformation of PABA compounds.

Gas-phase studies offer several advantages over condensed-phase data by avoiding the masking effects observed in crystalline or condensed media, making them directly comparable with theoretical computations. Several experimental gas-phase approaches are available, most of them based on laser spectroscopy.[6,9] Here, we tackle this topic by using rotational spectroscopy, undoubtedly the experimental technique with unrivalled precision in structural gas phase determination. This technique, together with computational modelling, has been successfully used for biomolecules[11–15] and anaesthetics,[16–18] as well as to study anaesthetic–receptor simple models.[19]

Our research began with the creation and improvement of computational models for the free molecules of **BZC**, **BTI** and **BTN** using a two-step strategy. First, we predicted the lowest lying conformers of each system from a full conformational search applying fast molecular mechanics. Second, all the structures were fully optimized with three different computational methods of MP2, M06-2X and BL3YP-D3BJ (the theoretical protocol can be found in the section Computational methods of the ESI†). In addition, potential energy surfaces (PESs) were computed at the B3LYP-D3BJ/6-311++G(d,p) level for each system to explore the conformational landscape of each molecule with respect to its structural degrees of freedom (Fig. 2).

For **BZC**, which is the simplest system with only one dihedral angle (τ_1), three minima were identified. These minima correspond to the structures where the alkyl chain assumes G^+, T and G^- orientations, with G^+ and G^- being equivalent in symmetry (see Fig. 1d for the labelling system used). The PES built for **BZC** showed interconversion barriers of around 5 kJ mol^{-1} (Fig. 2a) between the different conformers. This computation agrees with previous experimental works performed by ion-dip and microwave spectroscopies in which *trans* (T) and *gauche* (G) orientations were observed, with the T conformation being the most stable.[6,9,20]

In **BTI**, the PES is more complex due to the more flexible alkyl side chain providing two degrees of torsional freedom with the dihedral angles τ_1 and τ_2. The computational methods predicted nine different plausible structures (Fig. 2b and Table S1, ESI†) for this LA. Possible conformers are reducted to five different structures with equivalent symmetries: $G^-G^+ = G^+G^-$, $TG^+ = TG^-$, $G^+G^+ = G^-G^-$, $G^-T = G^+T$ and TT. Finally, each of these structures, except for TT, can assume two plausible conformations, corresponding to the two orientations of the amino group with respect to the alkyl chain: *anti* (↓) and *syn* (↑). Different interconversion pathways can be obtained by varying either τ_1 or τ_2 coordinates. Moving along the τ_1 axis, energy barriers lower than 4 kJ mol^{-1} are predicted, which outlines the possibility of a conformational relaxation,[21] while the connection throughout the τ_2 axis is more hindered because of the bulkier terminal moiety (>15 kJ mol^{-1}, Fig. 2b).

The conformational analysis of **BTN** is the most complex: owing to three independent dihedral angles, a total of 27 global

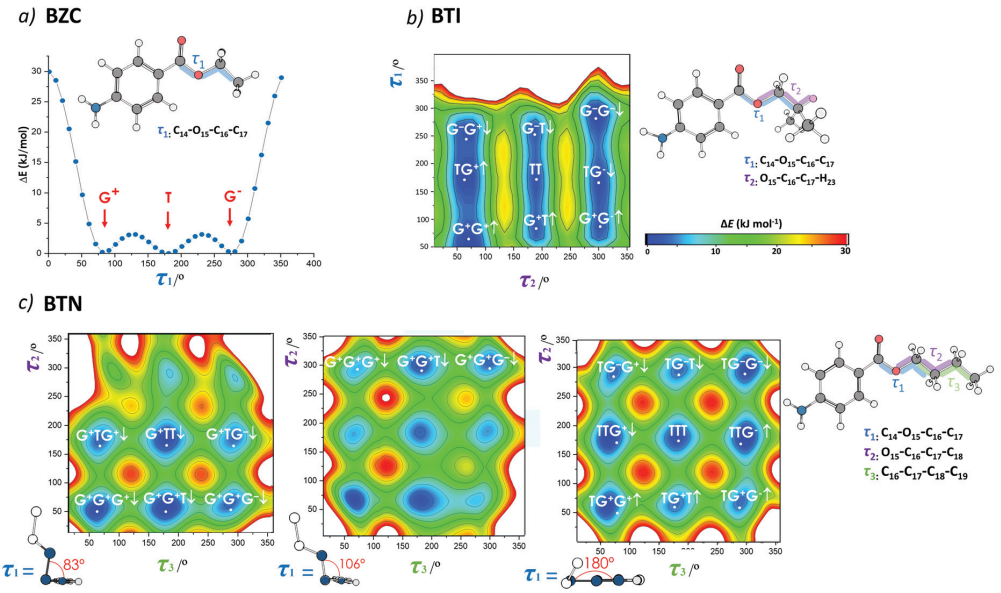


Fig. 2 Potential energy surfaces (PESs) of (a) benzocaine, (b) isobutamben (c) and butamben computed at the B3LYP-D3BJ/6-311++G(d,p) level of theory. For the latter system, three different PESs are shown where τ_1 has been fixed at different values. Above >30 kJ mol^{-1}, the energy values have been represented as white areas, since no minima have been detected. It should be pointed out that the conformation adopted by τ_2 in BTI is not equivalent to that in BTN, since the atoms involved in the two dihedral angles are different.

minima are predicted (Table S2, ESI†). To construct the PES of **BTN**, several full scans of τ_2 and τ_3 were performed (see Fig. S9, ESI†) with τ_1 (the torsionally most rigid angle) held at values from 60° to 180° varied in steps of 15°. Considering that the most stable structures are predicted for $\tau_1 = 83°$, an additional scan was done at this value. It should be noted that outside the investigated $60° < \tau_1 > 180°$ range, the molecule becomes strained due to steric hindrance. Finally, similar to **BTI**, the amine group can adopt two orientations (*anti*/*syn*). Three significant families arise from the τ_1 values of: (i) $\tau_1 = 83°$ (G), (ii) $\tau_1 = 106°$ (G), and (iii) $\tau_1 = 180°$ (T). Fig. 2c shows the most relevant PESs for **BTN** with an *anti*-arrangement. In particular, twelve minima belong to (i), where six of them assume a *syn* configuration and the other half belongs to the *anti*-arrangement. The same observation is made for (ii), where six minima are equally divided into *anti*/*syn*. In contrast, the (iii) family presents nine conformers, without *anti*/*syn* distinction by symmetry. All interconversion barriers are quite high with 15–20 kJ mol^{-1} values. Turning to the experiments, the simplest system, **BZC**, had been investigated by using laser[6,9] and microwave[20] spectroscopies. Unambiguously, the adoption of the G and T orientations of the alkyl chain in **BZC** was detected, with the T conformation being the most stable one. Covering different frequency ranges, the rotational spectra of **BTI** and **BTN** were recorded in three laboratories using different experimental set-ups (a free jet absorption millimetre-wave spectrometer in Bologna and two Fourier Transform Microwave spectrometers in Hannover and Bilbao). In addition, two vaporization sources were used: (a) heating methods using custom nozzles for the three laboratories; (b) a UV ultrafast laser vaporization system (in Bilbao) suppressing decomposition processes[14] (Experimental methods in the ESI†). The measured rotational transitions of both systems were fitted using a semi-rigid rotor Hamiltonian based on Watson's symmetric reduction in the I^r representation,[22] adding an additional term to account for nuclear quadrupole coupling with the molecular end-to-end rotation[23] (Tables S5–S8, ESI†). BTI exhibits a rather complex rotational spectrum, from which three different conformers could be identified, while BTN shows a simpler spectrum, a dominant structure was the only one observed. A direct comparison between experimental and predicted rotational constants (Tables S4 and S5, ESI†) unambiguously identified the observed conformers to be $G^-G^+ = G^+G^-$, $TG^+ = TG^-$, and TT for BTI and TG^-T for BTN.

Building on these results, an additional experiment was performed. We recorded the rotational spectrum of **BTI** replacing neon by argon as the carrier gas. With argon, the conformer $G^-G^+\downarrow$ was not observed (Fig. 3a), rationalized by more effective conformational relaxation processes occurring in argon expansion.[21] In particular, the conformer $G^-G^+\downarrow$ relaxes to the conformer $TG^-\downarrow$. This observation is quite informative, demonstrating that both conformers should be separated by a low barrier. This situation arises for conformers sharing the same τ_2, which implies barriers below <4 kJ mol^{-1} (Fig. 2b).

There is still one conformational degree of freedom remaining to be resolved: the position of the hydrogen atoms of the amino group. Adopting a pyramidal configuration, the functional group can be orientated *syn* (↑) or *anti* (↓) (Fig. 1d). Both orientations are

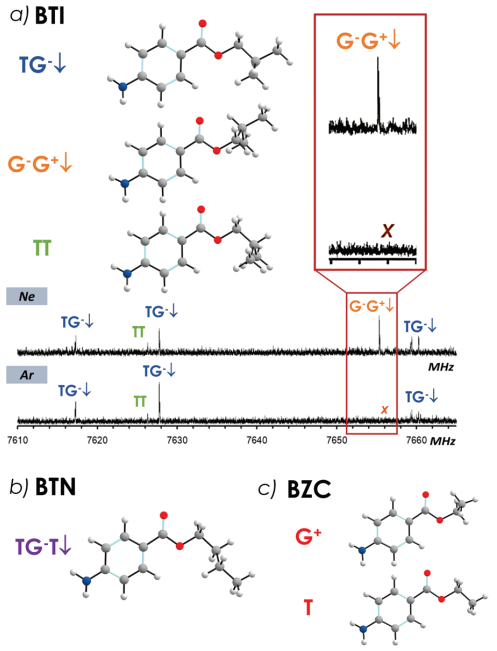


Fig. 3 (a) Section of the rotational spectra of **BTI** showing the 3 observed species and the conformational relaxation of $G^-G^+\downarrow$ to $TG^-\downarrow$. (b and c) Detected structures for **BTN** and **BZC**.

almost isoenergetic according to theoretical methods and their structural frames are almost identical (Fig. S4 and S5, ESI†). Hence, the corresponding rotational constants have similar values and cannot be used to distinguish between both structures unambiguously. However, we could resolve and exploit the hyperfine structure due to the ^{14}N nuclear quadrupole. This tensor is related to structural details on the electron distribution in the vicinity of the quadrupolar nucleus. We try to discern the ↑ or ↓ orientations (Tables S1 and S2, ESI†) related to this tensor together with the observed conformational relaxation, which implies the same τ_2 for the considered species. In this case, the distinction between *anti*/*syn* orientations was not so clear because of the similarity of the nuclear quadrupole constants of the following pairs: $TG^-\downarrow$ and $TG^-\uparrow$ in the case of **BTI**, and $TG^-T\downarrow$ and $TG^-T\uparrow$ in the case of **BTN**. However, as the assignment of the $G^-G^+\downarrow$ **BTI** conformer was more evident (Table S5, ESI†), we could hypothesize that the hydrogen atoms of the amine group will preferentially adopt an *anti*-position, and this chemical rationale can be applied to other systems. For this reason, we suggest with some confidence that the observed conformers are $TG^-\downarrow$, $G^-G^+\downarrow$ and TT for **BTI** and $TG^-T\downarrow$ for **BTN**.

Experimental relative populations could be estimated through relative intensities of rotational transitions.[13,24] This information is crucial since MP2 and B3LYP-D3BJ do not give consistent data regarding energy ordering (Tables S1 and S2, ESI†). We studied **BTI** with different carrier gases (He, Ne, Ar and Xe; Fig. S10, ESI†), confirming the conformational

relaxation of $G^-G^+\downarrow$ to $TG^-\downarrow$, observed previously in Ar. In fact, considering similar expansion velocities, noble gases of greater masses (Ar and Xe) produce greater kinetic energies, which promote the overcoming of low conformational barriers ($G^-G^+\downarrow \rightarrow TG^-\downarrow$: 3.3 kJ mol^{-1}).[21] Experimental population ratios are 54(5) : 30(5) : 16 and 82(5) : 0 : 18(5) for $TG^-\downarrow$: $G^-G^+\downarrow$: TT in Ne and Ar, respectively, which is equivalent to the energy ratios of 0 : 1.5(5) : 2.8(5) kJ mol^{-1} assuming a Boltzmann distribution at 298 K (see Table S3, ESI†). Actually, the small difference in population among the conformers of **BTI** makes their predictions challenging from a computational point of view. Our rotational data, with the moments of inertia specifically tied to conformations, are essential for clarifying the relative energy order for **BTI**.

Finally, comparing the results obtained for **BZC**, **BTN** and **BTI** (Fig. 3), we can hypothesize that PABA compounds with alkyl chains composed of less than three carbon atoms (like **BZC**) tend to maintain their alkyl chain in the plane of the benzoate skeleton. In contrast, alkyl chains with more than three carbon atoms (like **BTN** and **BTI**) make the molecules adopt an 'L-shape' conformation, where the alkyl chain is almost perpendicular to the plane of the benzoate skeleton. For **BTI**, which shows a branched alkyl chain, it is evident that the flexibility of the molecule decreases, leading to a small number of accessible conformations.

In fact, we compared the results obtained in the gas phase with the crystal structures. While **BTN** is a polymorphic system, its most stable crystal structure corresponds to the conformer detected in the gas phase ($TG^-T\downarrow$, CCDC code: EZAVIK 226561), where the alkyl chain adopts an 'L-shape'. Also for **BZC**, which exists in three different polymorphic forms,[25] the most stable crystal structure agrees with the T conformer that was found to be the global minimum also in the gas phase. No crystal structure of **BTI** has been isolated yet.[8] Even though one has to be careful in generalizing this finding, the PABA type anaesthetics exhibit similar molecular structures in both the gas phase and in the solid state.[26] Although these molecular ligands appear to be quite simple monomers, studying them in detail discloses the non-trivial conformational panorama of PABA derivatives, where computational works alone are not able to provide conclusive results as they are obtained here.

In conclusion, microwave spectroscopy enabled us to detect three conformers for **BTI** and one for **BTN**. The present work consolidates the use of rotational spectroscopy supported by computational modelling as a very powerful tool to provide precise conformational analyses of biomolecules. Moreover, the recent advances achieved by chirp FT-microwave spectroscopy have paved the way to study larger molecular systems.[27] By exploiting the higher sensitivity of this technique, the interaction between functional fragments able to mimic the receptor and LAs can be experimentally disclosed.

Financial support from the MINECO (CTQ2017-89150-R and PGC2018-098561-B-C22), Basque Government (IT1162-19 and PIBA 2018-11), the UPV/EHU (PPG17/10 and GIU18/207), CSIC (PIC2018 and LINKA20249), Fundación BBVA, Deutsche Forschungsgemeinschaft (DFG) and the Land Niedersachsen is acknowledged. A. I., C. C., and M. V. L. also thank the predoc and Juan de la Cierva contracts from the Basque Government and MECD. Computational resources and laser facilities of the UPV/EHU (SGIker) and CESGA were used in this work.

Conflicts of interest

There are no conflicts to declare.

References

1 D. S. Ragsdale, J. C. McPhee, T. Scheuer and W. A. Catterall, *Science*, 1994, **265**, 1724–1728.
2 D. E. Becker and K. L. Reed, *Anesth. Prog.*, 2012, **59**, 90–102.
3 A. Nardi, N. Damann, T. Hertrampf and A. Kless, *ChemMedChem*, 2012, **7**, 1712–1740.
4 H. Tsuchiya and M. Mizogami, *Anesthesiol. Res. Pract.*, 2013, 1–18.
5 J. Butterworth and G. Strichartz, *Anesthesiology*, 1990, **72**, 711–734.
6 E. Aguado, A. Longarte, E. Alejandro, J. A. Fernández and F. Castaño, *J. Phys. Chem. A*, 2006, **110**, 6010–6015.
7 Y. Kuroda, H. Nasu, Y. Fujiwara and T. Nakagawa, *J. Membr. Biol.*, 2000, **177**, 117–128.
8 A. C. Schmidt, *Int. J. Pharm.*, 2005, **298**, 186–197.
9 A. Longarte, J. A. Fernández, I. Unamuno and F. Castaño, *Chem. Phys.*, 2000, **260**, 83–93.
10 D. B. Tikhonov, I. Bruhova and B. S. Zhorov, *FEBS Lett.*, 2006, **580**, 6027–6032.
11 A. L. Steber, J. L. Neill, D. P. Zaleski, B. H. Pate, A. Lesarri, R. G. Bird, V. Vaquero-Vara and D. W. Pratt, *Faraday Discuss.*, 2011, **150**, 227–242.
12 J. L. Alonso and J. C. López, in *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, ed. A. M. Rijs and J. Oomens, Springer International Publishing, Cham, 2015, pp. 335–401.
13 I. Uriarte, S. Melandri, A. Maris, C. Calabrese and E. J. Cocinero, *J. Phys. Chem. Lett.*, 2018, 1497–1502.
14 E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J.-U. Grabow, J. A. Fernández and F. Castaño, *Angew. Chem., Int. Ed.*, 2012, **51**, 3119–3124.
15 T. Betz, S. Zinn and M. Schnell, *Phys. Chem. Chem. Phys.*, 2015, **17**, 4538–4541.
16 A. Lesarri, S. T. Shipman, J. L. Neill, G. G. Brown, R. D. Suenram, L. Kang, W. Caminati and B. H. Pate, *J. Am. Chem. Soc.*, 2010, **132**, 13417–13424.
17 A. Lesarri, A. Vega-Toribio, R. D. Suenram, D. J. Brugh, D. Nori-Shargh, J. E. Boggs and J. U. Grabow, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6610–6618.
18 N. A. Seifert, C. Pérez, J. L. Neill, B. H. Pate, M. Vallejo-López, A. Lesarri, E. J. Cocinero and F. Castaño, *Phys. Chem. Chem. Phys.*, 2015, **17**, 18282–18287.
19 N. A. Seifert, D. P. Zaleski, C. Pérez, J. L. Neill, B. H. Pate, M. Vallejo-López, A. Lesarri, E. J. Cocinero, F. Castaño and I. Kleiner, *Angew. Chem., Int. Ed.*, 2014, **53**, 3210–3213.
20 A. Lesarri, S. T. Shipman, G. G. Brown, L. Alvarez-Valtierra, R. D. Suenram and B. H. Pate, *OSU Int. Symp. Mol. Spectrosc.*, 2008, RH07.
21 R. S. Ruoff, T. D. Klots, T. Emilsson and H. S. Gutowsky, *J. Chem. Phys.*, 1990, **93**, 3142–3150.
22 J. K. G. Watson, *Vibrational Spectra and Structure*, Elsevier Science Publishers, New York, Amsterdam, 1977.
23 W. Gordy and R. L. Cook, *Microwave molecular spectra*, Wiley, New York, 1984.
24 G. T. Fraser, R. D. Suenram and C. L. Lugez, *J. Phys. Chem. A*, 2000, **104**, 1141–1146.
25 M. Paczkowska, G. Wiergowska, A. Miklaszewski, A. Krause, M. Mroczkowka, P. Zalewski and J. Cielecka-Piontek, *Molecules*, 2018, 23.
26 E. J. Cocinero, A. Lesarri, P. Ecija, J.-U. Grabow, J. A. Fernández and F. Castaño, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6076–6083.
27 C. Calabrese, W. Li, G. Prampolini, L. Evangelisti, I. Uriarte, I. Cacelli, S. Melandri and E. J. Cocinero, *Angew. Chem., Int. Ed.*, 2019, **58**, 8437–8442.

ISM prebiotic molecules

frontiers in Astronomy and Space Sciences

PERSPECTIVE
published: 15 March 2022
doi: 10.3389/fspas.2022.843766

The SKA as a Prebiotic Molecule Detector

Izaskun Jiménez-Serra[1]*, *Jesús Martín-Pintado*[1], *Aran Insausti*[2,3], *Elena R. Alonso*[2,3], *Emilio J. Cocinero*[2,3] *and Tyler L. Bourke*[4]

[1]Department of Astrophysics, Centro de Astrobiología (CSIC/INTA), Torrejón de Ardoz, Spain, [2]Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco, (UPV-EHU), Bilbao, Spain, [3]Biofisika Institute (CSIC, UPV/EHU), Leioa, Spain, [4]SKA Observatory, Jodrell Bank, Macclesfield, United Kingdom

One of the theories for the origin of life proposes that a significant fraction of prebiotic material could have arrived to Earth from outer space between 4.1 and 3.8 billion years ago. This suggests that those prebiotic compounds could have originated in interstellar space, to be later on incorporated to small Solar-system bodies and planetesimals. The recent discovery of prebiotic molecules such as hydroxylamine and ethanolamine in the interstellar medium, strongly supports this hypothesis. However, some species such as sugars, key for the synthesis of ribonucleotides and for metabolic processes, remain to be discovered in space. The unmatched sensitivity of the Square Kilometre Array (SKA) at centimeter wavelengths will be able to detect even more complex and heavier prebiotic molecules than existing instrumentation. In this contribution, we illustrate the potential of the SKA to detect simple sugars with three and four carbon atoms, using a moderate investment of observing time.

Keywords: square kilometre array, complex organic molecules, prebiotic chemistry, interstellar medium, astrochemistry

OPEN ACCESS

Edited by:
Ashraf - Ali,
University of Maryland, College Park, United States

Reviewed by:
Eric Herbst,
University of Virginia, United States

***Correspondence:**
Izaskun Jiménez-Serra
ijimenez@cab.inta-csic.es

Specialty section:
This article was submitted to Astrochemistry, a section of the journal Frontiers in Astronomy and Space Sciences

Received: *26 December 2021*
Accepted: *08 February 2022*
Published: *15 March 2022*

Citation:
Jiménez-Serra I, Martín-Pintado J, Insausti A, Alonso ER, Cocinero EJ and Bourke TL (2022) The SKA as a Prebiotic Molecule Detector. Front. Astron. Space Sci. 9:843766. doi: 10.3389/fspas.2022.843766

1 INTRODUCTION

The question of the origin of life has attracted the interest of researchers for centuries. Two main lines of thought have been proposed: i) Life may have emerged endogenously so that the building blocks of life could have formed *"in situ"* on Earth; or ii) Life could have originated somewhere else in the Universe and been transported to Earth in small Solar-system bodies. Alternatively, an intermediate theory contemplates the possibility that a fraction of the prebiotic material essential for the origin of life could have originated exogenously and been transferred to a young Earth *via* planetesimal impact on its surface (Anders, 1989; Chyba and Sagan, 1992). Prebiotic molecules such as amino acids, nucleobases and sugars have indeed been detected in meteorites (Cooper et al., 2001; Pizzarello et al., 2006; Callahan et al., 2011; Furukawa et al., 2019; Glavin et al., 2020) and in comets (Altwegg et al., 2016), which supports the latter hypothesis.

In the past decade, it has become clear that the interstellar medium (ISM) is an extraordinary chemical factory. About 250 molecules, including ringed-molecules (see e.g., Cernicharo et al., 2021; McGuire et al., 2021), have so far been reported in the ISM. In addition, the pace at which new molecules are detected not only seems steady but accelerating (McGuire, 2021). In particular, the so-called complex organic molecules (or COMs)[1] have attracted great interest in recent years since a

[1]COMs are defined as carbon-based molecules with ≥6 atoms in their structure (as e.g., methanol or CH_3OH; see Herbst and van Dishoeck, 2009).

subset of them could have been involved in the first biochemical reactions leading to the origin of life. This sub-set of COMs are typically called *prebiotic*. Some examples of prebiotic COMs include urea and hydroxylamine (Belloche et al., 2019; Jiménez-Serra et al., 2020; Rivilla et al., 2020) since they are possible precursors of ribonucleotides (see e.g., Becker et al., 2019; Menor Salvan et al., 2020); ethanolamine and n-propanol because they could have triggered the formation of phospholipids (Jiménez-Serra et al., 2022; Rivilla et al., 2021); or amino acetonitrile, vynil amine and ethyl amine since they are considered precursors of amino acids (Belloche et al., 2008; Zeng et al., 2021).

One of the most extended theories about the origin of life is the primordial RNA world. In this scenario, early forms of life relied solely on RNA to store genetic information and to catalyze chemical reactions. The basic units of RNA are ribonucleotides, which are composed of a phosphate group, a nitrogenous base, and a ribose sugar (a C_5 sugar with five carbon atoms). Interestingly, only small precursors of sugars such as glycolonitrile ($HOCH_2CN$; see Patel et al., 2015) or the simplest C_2 sugar molecule, glycolaldehyde (CH_2OHCHO), have been reported in the ISM (Hollis et al., 2000; Beltrán et al., 2009; Jørgensen et al., 2012; Zeng et al., 2019). Indeed, searches of C_3 sugars such as glyceraldehyde ($CHOCHOHCH_2OH$) or dihydroxyacetone (DHA, with the chemical formula $CH_2OHCOCH_2OH$), have not yielded any robust detection (Hollis et al., 2004; Widicus Weaver and Blake, 2005; Apponi et al., 2006; Jiménez-Serra et al., 2020). In contrast, larger sugar-like species such as C_3 sugars and ribose have been found in meteorites (de Marcellus et al., 2015; Furukawa et al., 2019), which opens the possibility that these species form in interstellar space.

The search of sugars in the ISM has partly been hindered by the lack of spectroscopic rotational data since these species are thermolabile and hygroscopic. This has triggered the development of new sample preparation techniques as well as the use of ultrafast laser vaporization sources to avoid their decomposition (Cocinero et al., 2012; Calabrese et al., 2020). Recently, the rotational spectroscopy of C_4 and C_5 sugars such as erythrulose and ribose has been obtained in the laboratory (Cocinero et al., 2012; Insausti et al., 2021), which enables their search in interstellar space.

2 SEARCHING FOR SUGARS IN SOURCES WITH LOW EXCITATION TEMPERATURES (T_{EX})

The first searches in the ISM of glyceraldehyde and dihydroxyacetone (DHA) targeted the massive star-forming region SgrB2 N-LMH. While an upper limit to the abundance of glyceraldehyde of $\leq(2.4–5.7) \times 10^{-11}$ was obtained by Hollis et al. (2004), Widicus Weaver and Blake (2005) reported a tentative detection of DHA toward this source with an abundance of 1.2×10^{-9}. The latter detection, however, was never confirmed (Apponi et al., 2006). For low-mass star-forming regions, Jiménez-Serra et al. (2020) also reported upper limits to the abundance of both species in the range $\leq(0.6–4)\times10^{-10}$ toward the low-mass hot corino IRAS16293-2422 B.

Massive hot cores and low-mass hot corinos are among the most chemically rich sources in the Galaxy and, traditionally, they have been the selected targets for searches of new prebiotic COMs in the ISM (see e.g., Jørgensen et al., 2012; Belloche et al., 2014; Jørgensen et al., 2016; Belloche et al., 2019). Their disadvantage, however, lies in the fact that their observed rotational spectra present hundreds (even thousands) of different molecular rotational transitions as a result of the high excitation temperatures (of T_{ex} = 100–300 K). These high temperatures populate a wide number of energy levels of a COM since its partition function is large. In addition, at temperatures of 100–300 K COM rotational spectra present their peak emission shifted towards millimeter and sub-millimeter wavelengths, which are heavily contaminated by the emission from smaller and lighter species such as CO. Consequently, the COM spectra observed in massive hot cores and low-mass hot corinos not only largely suffer from line blending and line confusion but also present weak lines due to the large partition functions expected at high excitation temperatures.

Alternatively, sources where COMs show low excitation temperatures (T_{ex}) represent better targets for the search and discovery of new large prebiotic species in the ISM (Jiménez-Serra et al., 2014)[2]. As a result of the low T_{ex}, the emission peak of the COM observed spectra is shifted towards lower frequencies, which are cleaner from the contribution from lighter molecules. In addition, the line intensities increase since only the lowest energy levels of the COMs can be populated at such low T_{ex} and so, the number of rotational transitions present in the measured spectra is significantly smaller than in hotter sources. The frequency span for the transitions between the lowest energy levels is also much larger than for those between the higher energy levels, which helps reducing significantly line blending and line confusion in the observed spectra (Jiménez-Serra et al., 2014).

Giant Molecular Clouds (GMCs) located in the Galactic Center such as the molecular cloud G+0.693-0.027 (hereafter G+0.693) have proven to be not only efficient chemical factories for the formation of complex organics (Requena-Torres et al., 2006; Requena-Torres et al., 2008; Zeng et al., 2018), but also excellent targets for the discovery of new prebiotic species (see e.g., Jiménez-Serra et al., 2020; Rivilla et al., 2020; Rivilla et al., 2021; Zeng et al., 2021). These clouds show low H_2 gas densities of $\sim10^4$ cm^{-3} and their gas and dust temperatures are decoupled (while $T_{dust} \leq$ 20K, T_{gas} ~70–150 K; see Rodríguez-Fernández et al., 2000; Zeng et al., 2018). The low H_2 gas densities also imply that for high-dipole moment molecules (such as COMs), and despite the high gas kinetic temperatures and broad linewidths of their emission (of ~ 20 km s^{-1}), their T_{ex} is low (between 5 and 20 K) and thus their observed spectra will be less affected by line blending and line confusion. Recent searches of prebiotic molecules toward the Galactic Center molecular cloud

[2]In massive hot cores and hot corinos, $T_{ex} \sim T_{kin}$ because of the high H_2 gas densities ($\geq10^6$ cm^{-3}) found in these star-forming regions

G+0.693 have yielded the discovery of several of these species for the first time in the ISM such as hydroxylamine (NH_2OH; Rivilla et al., 2020), ethanolamine ($NH_2CH_2CH_2OH$; Rivilla et al., 2021), thioformic acid (HC(O)SH; Rodríguez-Almeida et al., 2021) or *n*-propanol (n-C_3H_7OH; Jiménez-Serra et al., 2022). As a result, Galactic Center GMCs are prime targets for the search and discovery of large C_3 and C_4 sugars in the ISM.

3 THE POTENTIAL OF THE SKA TO DETECT NEW PREBIOTIC SPECIES IN THE ISM: THE CASE OF SUGARS

The Square Kilometre Array (SKA) will be the largest radio telescope in the world operating at centimeter and meter wavelengths. The Phase 1 of this observatory will consist of two radio interferometers located at two different sites: SKA1 LOW in Western Australia and SKA1 MID in South Africa (Braun et al., 2019). SKA1 LOW will have 512 stations of 256 log periodic dipole antennas each operating within the frequency range between 50 and 350 MHz, while SKA1 MID will have a total of 197 antennas (that include the 64 antennas of the MeerKAT array) operating at frequencies between 350 MHz and 15 GHz (Braun et al., 2019). Given the frequency coverage and how dramatically the strength of the rotational transitions of high-dipole molecules drops for frequencies ≤ 5GHz, SKA1 MID, and in particular its Band 5 receivers covering the frequency range between 4.8 and 15.3 GHz, are well suited for searches of prebiotic molecules in the ISM. On top of the advantages of observing at centimeter wavelengths noted in **Section 2**, it is important to stress that radio interferometers such as the SKA1 filter out any extended emission in the line of sight of the targeted source, significantly decreasing the linewidths of the observed molecular line emission and reducing the level of line blending and line confusion. In **Section 3.1**, we evaluate the feasibility of detecting large sugar-like species such as glyceraldehyde, dihydroxyacetone and erythrulose in the ISM using the Band 5 receivers of SKA1 MID.

3.1 SKA1 Simulations for C_3 and C_4 Sugars in Galactic Center GMCs: Glyceraldehyde, Dihydroxyacetone and Erythrulose

The emission from COMs in the Galactic Center GMCs is known to be extended (Martín-Pintado et al., 2001; Li et al., 2017; Li et al., 2020). This is particularly interesting for absorption studies in which COMs can be searched for against a bright continuum background source. This technique may allow the detection of low-abundance COMs whose emission spectra is expected to be very weak. Indeed, in the presence of a bright background continuum source, the predicted absorption line intensity (T_L) is given by (see e.g., Martín et al., 2019):

$$T_L = \left(T_{ex} - T_c - T_{bg}\right) \times \left[1 - exp(-\tau_\nu)\right] \quad (1)$$

where T_c is the temperature of a continuum source and T_{bg} is the temperature of the comic microwave background radiation (T_{bg} = 2.73 K). If $T_c \gg T_{ex}$, this will largely enhance the observed absorption line intensity over the emission for the T_{ex} expected to be close to T_{bg}.

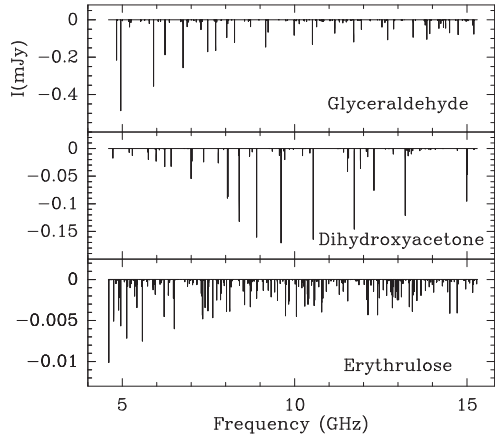


FIGURE 1 | Predicted absorption spectra of the C_3 and C_4 sugars glyceraldehyde, dihydroxyacetone and erythrulose obtained with the SLIM tool of the MADCUBA package considering the physical conditions of the G+0.693 molecular cloud (T_{ex} = 10K, linewidths of Δ_v = 20 km s^{-1} and extended emission across the beam) and a background source similar to the L source in the SgrB2 N molecular complex (measured flux of ~300 mJy within a beam of 1″ at 10 GHz). The assumed column densities for glyceraldehyde and dihydroxyacetone are 10^{13} cm^{-2} and 6.8× 10^{12} cm^{-2}, respectively, while for erythrulose we assume a column density one order of magnitude lower than that of glyceraldehyde (of 10^{12} cm^{-2}; see **Section 3.1**).

Therefore, we simulate the case of absorption spectra of glyceraldehyde, dihydroxyacetone and erythrulose ($C_4H_8O_4$) using the SLIM tool within the MADCUBA software package[3] (see Martín et al., 2019). We focus our simulations on C_3 and C_4 sugars because, as discussed in **Section 3.2**, C_5 sugars such as ribose are not expected to be abundant enough to be detected with SKA1 in a reasonable amount of observing time. We assume the physical conditions of the G+0.693 molecular cloud with a typical T_{ex} of ≤10K, linewidths of Δ_v = 20 km s^{-1} and extended morphology across the SKA1 beam for the COMs emission (Zeng et al., 2020). As background source, we consider the L source located in the SgrB2 N complex, which is a bright HII region found nearby the emission peak of G+0.693, and whose measured flux at 23 GHz is ~ 300 mJy within a beam of ~ 1" (de Pree et al., 1995). We have assumed that the emission from the HII region is optically thin and thus, that it shows an almost flat spectral energy distribution between 23 and 10 GHz. The simulations have been obtained considering a beam of 1", which will easily be reached by the SKA1 at the frequencies of the Band 5 receivers of SKA1 MID (see Braun et al., 2019). The spectroscopic data for glyceraldehyde have been extracted from

[3]MADCUBA, or MAdrid Data CUBe Analysis, is a software developed at the Center of Astrobiology in Madrid: https://cab.inta-csic.es/madcuba/index.html

TABLE 1 | List of transitions of glyceraldehyde, dihydroxyacetone and erythrulose that present the deepest features in the simulated spectra.

Molecule	Formula	Frequency (MHz)	Transition	Log [Int(300K)]	E_l (cm^{-1})	I_{depth} (mJy)
Glyceraldehyde	$CHOCHOHCH_2OH$	4,955.499 8	$6_{2,5} \rightarrow 6_{1,5}$	−7.700 2	3.845 8	−0.488
		5,907.255 6	$5_{2,4} \rightarrow 5_{1,4}$	−7.671 5	2.780 5	−0.360
		6,762.410 3	$4_{2,3} \rightarrow 4_{1,3}$	−7.693 1	1.888 8	−0.260
		4,835.648 0	$12_{3,10} \rightarrow 12_{2,10}$	−7.449 5	14.253 8	−0.223
		6,240.130 2	$11_{3,9} \rightarrow 11_{2,9}$	−7.312 7	12.128 5	−0.186
Dihydroxyacetone	$CH_2OHCOCH_2OH$	9,605.934 2	$4_{1,3} \rightarrow 4_{0,4}$	−7.686 9	1.258 4	−0.172
		10 540.639 1	$5_{1,4} \rightarrow 5_{0,5}$	−7.544 6	1.883 7	−0.166
		8,898.373 3	$3_{1,2} \rightarrow 3_{0,3}$	−7.842 7	0.756 3	−0.161
		11 731.758 5	$6_{1,5} \rightarrow 6_{0,6}$	−7.410 4	2.630 8	−0.147
		8,391.965 5	$2_{1,1} \rightarrow 2_{0,2}$	−8.025 3	0.378 6	−0.132
Erythrulose	$C_4H_8O_4$	4,609.671 9	$10_{3,8} \rightarrow 10_{2,8}$	−8.320 5	5.159 2	−0.010
		5,583.487 0	$9_{3,7} \rightarrow 9_{2,7}$	−8.244 2	4.262 4	−0.007 5
		5,131.326 5	$2_{2,1} \rightarrow 2_{1,1}$	−9.051 3	0.332 9	−0.0071[a]
		5,131.411 5	$2_{0,2} \rightarrow 1_{0,1}$	−8.806 1	0.086 1	−0.0071[a]
		6,508.324 0	$8_{3,6} \rightarrow 8_{2,6}$	−8.199	3.452 2	−0.006 1

[a]*Intensity of the feature resulting from the blending of the two indicated transitions.*

the CDMS catalogue (entry 090501; Endres et al., 2016), while the data for dihydroxyacetone is available in the JPL molecular catalogue (entry 90002; Pickett et al., 1998). The spectroscopic information for erythrulose can be found in Insausti et al. (2021).

As column densities, we have assumed that glyceraldehyde and dihydroxyacetone have column densities of 10^{13} cm^{-2} and 6.8×10^{12} cm^{-2}, respectively. These column densities correspond to the upper limits obtained toward G+0.693 in previous spectral surveys (see Table2 in Jiménez-Serra et al., 2020). For erythrulose, we consider that this C_4 sugar is a factor of 10 less abundant than the C_3 sugar glyceraldehyde, with an assumed column density of 10^{12} cm^{-2} for erythrulose. We note that this decrease by one order of magnitude in the column density of molecules from the same family when one carbon atom is added to the molecular structure, has been reported for alcohols (Jiménez-Serra et al., 2022) and thiols (Rodríguez-Almeida et al., 2021). The assumed spectral resolution considered in our simulations is 79.3 kHz (equivalent to a velocity resolution between ~1.6 and 5.0 km s^{-1}) across the full Band 5 frequency coverage.

Figure 1 presents the predicted absorption spectra of glyceraldehyde, dihydroxyacetone and erythrulose obtained for the extended GMC G+0.693 against the background source L. Depending on the structure of the molecule, the transitions that show the larger absorption depths lie in the range between 4.8 and 7.0 GHz for glyceraldehyde and erythrulose, and 8–14 GHz for dihydroxyacetone. The maximum bandwidth that will be covered simultaneously by the Band 5 receivers of SKA1 MID in its Phase 1 is 4 GHz at frequencies between 4.6 and 8.5 GHz, and 5 GHz for frequencies between 8.3 and 15.3 GHz. Given that the deepest absorption lines appear clustered within frequency ranges of ≤ 4-5 GHz-width, future detection experiments of these lines with SKA1 MID could be carried out in just two observing runs: one to simultaneously cover all the deepest features of glyceraldehyde and erythrulose, and a second one to simultaneously cover all the transitions of dihydroxyacetone. We stress that **Figure 1** includes all possible transitions of these C_3 and C_4 sugars available within the frequency range covered by the Band 5 receivers of SKA1 MID. The intensities of the deepest absorption features in our simulated spectra reach values of −0.47 mJy for glyceraldehyde, of −0.17 mJy for dihydroxyacetone, and −0.01 mJy for erythrulose. **Table1** lists the spectroscopic information of the deepest features found in our predicted spectra for each molecule. In **Section 3.2**, we evaluate the observing time required to perform these detection experiments with SKA1 MID during its Phase 1 of operations, which will have only 133 antennas equipped with Band 5 receivers.

3.2 Time Estimates and Key Science Projects

As seen from **Table1**, the predicted intensities for the deepest absorption features lie in the range between −0.19 and −0.5 mJy for glyceraldehyde, −0.13 and −0.2 mJy for dihydroxyacetone, and −0.006 and −0.01 mJy for erythrulose. According to Braun et al. (2019), a line sensitivity of 90 μJy will be achieved in 1 hour of observing time with the Band 5 receivers of SKA1 MID at frequencies of 4.6–8.5 GHz, and of 85 μJy at frequencies of 8.5–15.3 GHz, for a velocity resolution of ~ 30km s^{-1}.

Taking these numbers into account, the deepest absorption line of glyceraldehyde at 4955 MHz could be detected with a S/N ≥ 5 in just 8.5 h assuming a velocity resolution of 3 km s^{-1}. The weakest transitions of glyceraldehyde at 4835 MHz and at 6240 MHz, however, would require about 50 h of observing time for a similar velocity resolution and S/N. This is also the case of dihydroxyacetone, for which a total of ≥ 100hrs of integration time on-source would be needed to detect its weakest transitions at 8391 and 11 731 MHz with a S/N ≥ 5 and a velocity resolution of 3 km s^{-1}. A more time-consuming experiment would have to be performed for the discovery of the deepest absorption features of erythrulose. In order to detect the erythrulose transition at 4609 MHz with a S/N ≥ 3, and considering a velocity resolution of 20 km s^{-1} (i.e., the linewidth of the molecular emission in G+0.693; see Zeng et al., 2018), a total of ≥ 1100hrs of observing time on-source would be required.

The time estimates to perform the detection experiments of glyceraldehyde, dihydroxyacetone and erythrulose proposed above, may seem excessive. However, we note that these long

integration times are contemplated within the future Key Science Projects (KSPs) scheme that will be carried out by the SKA1. The KSPs are major surveys targeting ground-breaking transformational science in Astrophysics and Astrobiology that require considerable amount of time. Therefore, the expected integration time requests for these projects are typically of a few thousands of hours, which guarantees the detection of at least C_3 sugars such as glyceraldehyde and dihydroxyacetone, if present in the ISM. As discussed in **Section 4**, the discovery of C_4 sugars such as erythrulose may require the development of future instrumentation for SKA1 MID such as the high-frequency Band 6 receivers.

We note that, when performing deep integrations as proposed here, it is expected that the level of line confusion increases even at the low frequencies covered by the SKA1 as a consequence of weak features of low-abundance molecules becoming bright in the spectra. However, this potential problem is likely not an issue for the discovery of intermediate-sized species such as C_3 and C_4 sugars. As recently found by Jiménez-Serra et al. (2022) and by Rodríguez-Almeida et al. (2021), the addition of a $-CH_2$ group to the structure of alcohols, thiols, isocyanates, and carboxylic acids implies a decrease in their abundance by at least one order of magnitude (see also **Section 3.1**). This decrease in the abundance would imply an increase in the required SKA1 observing time by a factor of 100 and thus, for a reasonable integration time within the context of SKA1 Key Science Projects, the only sugars for which absorption features could be detected with SKA1 are the C_3 and the C_4 sugars together with their lowest-energy conformers. Indeed, for the case of *n*-propanol, only the *Ga* and *Aa* conformers with relative energies of E = 0 K and E = 40 K, have been found toward G+0.693 (the rest of conformers with energies > 50K do not show any detected features; see Jiménez-Serra et al., 2022). Therefore, although the conformers of C_3 and C_4 sugars as well as even larger sugars may be present in Galactic Center Giant Molecular Clouds such as G+0.693, these species likely do not confuse much the observed SKA1 spectrum because either their abundance is too low (as for C_5 sugars such as ribose) or their low-energy levels are not populated at the low excitation temperatures measured in these clouds (as for high-energy conformers).

4 FUTURE EXPANSION OF THE SKA: THE BAND 6 RECEIVERS

From **Section 3.2**, it is clear that, while relatively small molecules such as glyceraldehyde or dihydroxyacetone could be detected with a moderate investment of SKA1 MID observing time, larger prebiotic COMs such as erythrulose would still be well below the limit of what the SKA1 will be able to detect in its Phase 1. For this reason, one possible future expansion of SKA1 MID contemplates the development of higher-frequency receiver(s), the Band 6 receivers, which will increase the frequency coverage of SKA1 MID to higher frequencies from 15.3 up to 50 GHz. As reported in Memo 20-01 of SKA1 titled "SKA1 Beyond 15 GHz: The Science case for Band 6" (see Conway et al., 2021), high-dipole moment molecules such as COMs present rotational transitions at frequencies ≥20GHz that can be factors ≥10 brighter than those found at frequencies ≤15GHz (see Section 3.3.1 in this Memo). This is due to the fact that the Einstein A_{ul} coefficients increase as ν^3 with ν being the frequency, making them one order of magnitude higher at 36 GHz than at 12 GHz. This implies that observing times about ~100 times shorter would be needed to detect the C_4 sugar erythrulose in the ISM, reducing the total observing time needed for the discovery of this species from ≥1100hrs to a few tens of hours (or at most, to a few hundreds of hours). Therefore, if these receivers were finally included in the development program of the SKA1, they would make this observatory an unbeateable machine for the discovery of large prebiotic compounds in space.

5 CONCLUSION

In this contribution, we evaluate the feasibility of detecting small sugars in the ISM using the Band 5 receivers of SKA1 in its Phase 1. Our simulations show that Galactic Center Giant Molecular Clouds such as G+0.693 represent prime targets for future searches of these key prebiotic species in space. As shown in **Section 3.1**, the detection of small sugars could be achieved by carrying out broad spectroscopic surveys between 5 and 14 GHz in absorption against a bright continuum background source. By taking as template the L HII region source located in the SgrB2 N molecular complex, we estimate that the detection of C_3 sugars could be achieved in a few hundreds of hours. Larger C_4 sugars such as erythrulose would require thousands of hours of observing time. Future developments of the SKA such as the Band 6 receivers (which will increase the frequency coverage of SKA1 MID up to, at least, 24 GHz), will enable the search of these large sugars and other prebiotic COMs in just a few hundreds of hours.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

IJ-S has written the first version of the manuscript. IJ-S and JM-P have produced the MADCUBA-SLIM simulations of the spectra of glyceraldehyde, dihydroxyacetone and erythrulose reported in **Section 3.1**. All authors have contributed to the discussion of the results and have provided comments on the text.

FUNDING

IJ-S and JM-P acknowledge partial support from the Spanish State Research Agency (AEI) through project numbers PID 2019-105552RB-C41 and MDM-2017-0737 Unidad de Excelencia María de Maeztu-Centro de Astrobiología (CSIC-INTA). EC

thanks the support received from the MINECO (PID 2020-117892RB-I00), the Basque Government (IT1162-19 and PIBA 2018-11), the UPV/EHU (PPG17/10, GIU18/207), and CSIC (PIC 2018, LINKA20249). Computational resources and laser facilities of the UPV/EHU (SGIker) and CESGA were used in this work.

ACKNOWLEDGMENTS

We would like to thank an anonymous referee for his/her positive and constructive comments on an earlier version of the manuscript.

REFERENCES

Altwegg, K., Balsiger, H., Bar-Nun, A., Berthelier, J.-J., Bieler, A., Bochsler, P., et al. (2016). Prebiotic Chemicals-Amino Acid and Phosphorus-In the Coma of Comet 67P/Churyumov-Gerasimenko. *Sci. Adv.* 2, e16002855. doi:10.1126/sciadv.1600285

Anders, E. (1989). Pre-biotic Organic Matter from Comets and Asteroids. *Nature* 342, 255–257. doi:10.1038/342255a0

Apponi, A. J., Halfen, D. T., Ziurys, L. M., Hollis, J. M., Remijan, A. J., and Lovas, F. J. (2006). Investigating the Limits of Chemical Complexity in Sagittarius B2(N): A Rigorous Attempt to Confirm 1,3-Dihydroxyacetone. *ApJ* 643, L29–L32. doi:10.1086/504979

Becker, S., Feldmann, J., Wiedemann, S., Okamura, H., Schneider, C., Iwan, K., et al. (2019). Unified Prebiotically Plausible Synthesis of Pyrimidine and Purine Rna Ribonucleotides. *Science* 366, 76–82. doi:10.1126/science.aax2747

Belloche, A., Garrod, R. T., Müller, H. S. P., and Menten, K. M. (2014). Detection of a Branched Alkyl Molecule in the Interstellar Medium: Iso -propyl Cyanide. *Science* 345, 1584–1587. doi:10.1126/science.1256678

Belloche, A., Garrod, R. T., Müller, H. S. P., Menten, K. M., Medvedev, I., Thomas, J., et al. (2019). Re-exploring Molecular Complexity with ALMA (ReMoCA): Interstellar Detection of Urea. *A&A* 628, A10. doi:10.1051/0004-6361/201935428

Belloche, A., Menten, K. M., Comito, C., Müller, H. S. P., Schilke, P., Ott, J., et al. (2008). Detection of Amino Acetonitrile in Sgr B2(N). *A&A* 482, 179–196. doi:10.1051/0004-6361:20079203

Beltrán, M. T., Codella, C., Viti, S., Neri, R., and Cesaroni, R. (2009). First Detection of Glycolaldehyde outside the Galactic Center. *ApJ* 690, L93–L96. doi:10.1088/0004-637X/690/2/L93

Braun, R., Bonaldi, A., Bourke, T., Keane, E., and Wagg, J. (2019). *Anticipated Performance of the Square Kilometre Array - Phase 1 (SKA1). arXiv e-prints*arXiv:1912.12699.

Calabrese, C., Uriarte, I., Insausti, A., Vallejo-López, M., Basterretxea, F. J., Cochrane, S. A., et al. (2020). Observation of the Unbiased Conformers of Putative Dna-Scaffold Ribosugars. *ACS Cent. Sci.* 6, 293–303. doi:10.1021/acscentsci.9b01277PMID: 32123748

Callahan, M. P., Smith, K. E., Cleaves, H. J., Ruzicka, J., Stern, J. C., Glavin, D. P., et al. (2011). Carbonaceous Meteorites Contain a Wide Range of Extraterrestrial Nucleobases. *Proc. Natl. Acad. Sci.* 108, 13995–13998. doi:10.1073/pnas.1106493108

Cernicharo, J., Agúndez, M., Cabezas, C., Tercero, B., Marcelino, N., Pardo, J. R., et al. (2021). Pure Hydrocarbon Cycles in TMC-1: Discovery of Ethynyl Cyclopropenylidene, Cyclopentadiene, and Indene. *A&A* 649, L15. doi:10.1051/0004-6361/202141156

Chyba, C., and Sagan, C. (1992). Endogenous Production, Exogenous Delivery and Impact-Shock Synthesis of Organic Molecules: an Inventory for the Origins of Life. *Nature* 355, 125–132. doi:10.1038/355125a0

Cocinero, E. J., Lesarri, A., Écija, P., Basterretxea, F. J., Grabow, J.-U., Fernández, J. A., et al. (2012). Ribose Found in the Gas Phase. *Angew. Chem. Int. Ed.* 51, 3119–3124. doi:10.1002/anie.201107973

Conway, J., Beswick, R., Bourke, T., Coriat, M., Ferrari, C., Jimenez-Serra, I., et al. (2021). *Memo 20-01, SKA1 beyond 15GHz: The Science Case for Band 6 (Square Kilometre Array).*

Cooper, G., Kimmich, N., Belisle, W., Sarinana, J., Brabham, K., and Garrel, L. (2001). Carbonaceous Meteorites as a Source of Sugar-Related Organic Compounds for the Early Earth. *Nature* 414, 879–883. doi:10.1038/414879a

de Marcellus, P., Meinert, C., Myrgorodska, I., Nahon, L., Buhse, T., d'Hendecourt, L. L. S., et al. (2015). Aldehydes and Sugars from Evolved Precometary Ice Analogs: Importance of Ices in Astrochemical and Prebiotic Evolution. *Proc. Natl. Acad. Sci. USA* 112, 965–970. doi:10.1073/pnas.1418602112

de Pree, C. G., Gaume, R. A., Goss, W. M., and Claussen, M. J. (1995). The Sagittarius B2 Star-forming Region. II. High-Resolution H66 Alpha Observations of Sagittarius B2 North. *ApJ* 451, 284. doi:10.1086/176218

Endres, C. P., Schlemmer, S., Schilke, P., Stutzki, J., and Müller, H. S. P. (2016). The Cologne Database for Molecular Spectroscopy, CDMS, in the Virtual Atomic and Molecular Data Centre, VAMDC. *J. Mol. Spectrosc.* 327, 95–104. doi:10.1016/j.jms.2016.03.005

Furukawa, Y., Chikaraishi, Y., Ohkouchi, N., Ogawa, N. O., Glavin, D. P., Dworkin, J. P., et al. (2019). Extraterrestrial Ribose and Other Sugars in Primitive Meteorites. *Proc. Natl. Acad. Sci. USA* 116, 24440–24445. doi:10.1073/pnas.1907169116

Glavin, D. P., McLain, H. L., Dworkin, J. P., Parker, E. T., Elsila, J. E., Aponte, J. C., et al. (2020). Abundant Extraterrestrial Amino Acids in the Primitive Cm Carbonaceous Chondrite Asuka 12236. *Meteorit Planet. Sci.* 55, 1979–2006. doi:10.1111/maps.13560

Herbst, E., and van Dishoeck, E. F. (2009). Complex Organic Interstellar Molecules. *Annu. Rev. Astron. Astrophys.* 47, 427–480. doi:10.1146/annurev-astro-082708-101654doi:10.1146/annurev-astro-082708-101654

Hollis, J. M., Jewell, P. R., Lovas, F. J., and Remijan, A. (2004). Green Bank Telescope Observations of Interstellar Glycolaldehyde: Low-Temperature Sugar. *ApJ* 613, L45–L48. doi:10.1086/424927

Hollis, J. M., Lovas, F. J., and Jewell, P. R. (2000). Interstellar Glycolaldehyde: The First Sugar. *Astrophysical J.* 540, L107–L110. doi:10.1086/312881

Insausti, A., Alonso, E. R., Tercero, B., Santos, J. I., Calabrese, C., Vogt, N., et al. (2021). Laboratory Observation of, Astrochemical Search for, and Structure of Elusive Erythrulose in the Interstellar Medium. *J. Phys. Chem. Lett.* 12, 1352–1359. doi:10.1021/acs.jpclett.0c03050

Jiménez-Serra, I., Martín-Pintado, J., Rivilla, V. M., Rodríguez-Almeida, L., Alonso Alonso, E. R., Zeng, S., et al. (2020). Toward the RNA-World in the Interstellar Medium-Detection of Urea and Search of 2-Amino-Oxazole and Simple Sugars. *Astrobiology* 20, 1048–1066. doi:10.1089/ast.2019.2125

Jiménez-Serra, I., Rodríguez-Almeida, L. F., Martín-Pintado, J., Rivilla, V. M., Melosso, M., Zeng, S., et al. (2022). *Precursors of Fatty Alcohols in the ISM: Discovery of N-Propanol.*

Jiménez-Serra, I., Testi, L., Caselli, P., and Viti, S. (2014). Detectability of Glycine in Solar-type System Precursors. *ApJ* 787, L33. doi:10.1088/2041-8205/787/2/L33

Jørgensen, J. K., Favre, C., Bisschop, S. E., Bourke, T. L., van Dishoeck, E. F., and Schmalzl, M. (2012). Detection of the Simplest Sugar, Glycolaldehyde, in a Solar-type Protostar with ALMA. *ApJ* 757, L4. doi:10.1088/2041-8205/757/1/L4

Jørgensen, J. K., van der Wiel, M. H. D., Coutens, A., Lykke, J. M., Müller, H. S. P., van Dishoeck, E. F., et al. (2016). The ALMA Protostellar Interferometric Line Survey (PILS). *A&A* 595, A117. doi:10.1051/0004-6361/201628648

Li, J., Shen, Z., Wang, J., Chen, X., Li, D., Wu, Y., et al. (2017). Widespread Presence of Glycolaldehyde and Ethylene Glycol Around Sagittarius B2. *ApJ* 849, 115. doi:10.3847/1538-4357/aa9069

Li, J., Wang, J., Qiao, H., Quan, D., Fang, M., Du, F., et al. (2020). Mapping Observations of Complex Organic Molecules Around Sagittarius B2 with the ARO 12 M Telescope. *Monthly Notices R. Astronomical Soc.* 492, 556–565. doi:10.1093/mnras/stz3337

Martín, S., Martín-Pintado, J., Blanco-Sánchez, C., Rivilla, V. M., Rodríguez-Franco, A., and Rico-Villas, F. (2019). Spectral Line Identification and Modelling (SLIM) in the MAdrid Data CUBe Analysis (MADCUBA) Package. *A&A* 631, A159. doi:10.1051/0004-6361/201936144

Martín-Pintado, J., Rizzo, J. R., de Vicente, P., Rodríguez-Fernández, N. J., and Fuente, A. (2001). Large-Scale Grain Mantle Disruption in the Galactic Center. *Astrophysical J.* 548, L65–L68. doi:10.1086/318937

McGuire, B. A. (2021). *2021 Census of Interstellar, Circumstellar, Extragalactic, Protoplanetary Disk, and Exoplanetary Molecules. arXiv e-prints*arXiv:2109.13848.

McGuire, B. A., Loomis, R. A., Burkhardt, A. M., Lee, K. L. K., Shingledecker, C. N., Charnley, S. B., et al. (2021). Detection of Two Interstellar Polycyclic Aromatic

Hydrocarbons *via* Spectral Matched Filtering. *Science* 371, 1265–1269. doi:10.1126/science.abb7535

Menor Salván, C., Bouza, M., Fialho, D. M., Burcar, B. T., Fernández, F. M., and Hud, N. V. (2020). Prebiotic Origin of Pre-RNA Building Blocks in a Urea "Warm Little Pond" Scenario. *ChemBioChem* 21, 3504–3510. doi:10.1002/cbic.202000510

Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D., and Sutherland, J. D. (2015). Common Origins of RNA, Protein and Lipid Precursors in a Cyanosulfidic Protometabolism. *Nat. Chem* 7, 301–307. doi:10.1038/nchem.2202

Pickett, H. M., Poynter, R. L., Cohen, E. A., Delitsky, M. L., Pearson, J. C., and Müller, H. S. P. (1998). Submillimeter, Millimeter, and Microwave Spectral Line Catalog. *J. Quantitative Spectrosc. Radiative Transfer* 60, 883–890. doi:10.1016/S0022-4073(98)00091-0

Pizzarello, S., Cooper, G. W., and Flynn, G. J. (2006). *The Nature and Distribution of the Organic Material in Carbonaceous Chondrites and Interplanetary Dust Particles*, 625–652. doi:10.2307/j.ctv1v7zdmm.36

Requena-Torres, M. A., Martin-Pintado, J., Martín, S., and Morris, M. R. (2008). The Galactic Center: The Largest Oxygen-bearing Organic Molecule Repository. *ApJ* 672, 352–360. doi:10.1086/523627

Requena-Torres, M. A., Martín-Pintado, J., Rodríguez-Franco, A., Martín, S., Rodríguez-Fernández, N. J., and de Vicente, P. (2006). Organic Molecules in the Galactic center. *A&A* 455, 971–985. doi:10.1051/0004-6361:20065190

Rivilla, V. M., Jiménez-Serra, I., Martín-Pintado, J., Briones, C., Rodríguez-Almeida, L. F., Rico-Villas, F., et al. (2021). Discovery in Space of Ethanolamine, the Simplest Phospholipid Head Group. *Proc. Natl. Acad. Sci. USA* 118, e2101314118. doi:10.1073/pnas.2101314118

Rivilla, V. M., Martín-Pintado, J., Jiménez-Serra, I., Martín, S., Rodríguez-Almeida, L. F., Requena-Torres, M. A., et al. (2020). Prebiotic Precursors of the Primordial RNA World in Space: Detection of NH2OH. *ApJ* 899, L28. doi:10.3847/2041-8213/abac55

Rodríguez-Almeida, L. F., Jiménez-Serra, I., Rivilla, V. M., Martín-Pintado, J., Zeng, S., Tercero, B., et al. (2021). Thiols in the Interstellar Medium: First Detection of HC(O)SH and Confirmation of C2H5SH. *ApJL* 912, L11. doi:10.3847/2041-8213/abf7cb

Rodríguez-Fernández, N. J., Martín-Pintado, J., de Vicente, P., Fuente, A., Hüttemeister, S., Wilson, T. L., et al. (2000). Non-equilibrium H_2 Ortho-To-Para Ratio in Two Molecular Clouds of the Galactic Center. *Giant Mol. Clouds Galaxy* 356, 695–704.

Widicus Weaver, S. L., and Blake, G. A. (2005). 1,3-Dihydroxyacetone in Sagittarius B2(N-LMH): The First Interstellar Ketose. *ApJThe First Interstellar Ketose* 624, L33–L36. doi:10.1086/430407

Zeng, S., Jiménez-Serra, I., Rivilla, V. M., Martín, S., Martín-Pintado, J., Requena-Torres, M. A., et al. (2018). Complex Organic Molecules in the Galactic Centre: the N-Bearing Family. *Monthly Notices R. Astronomical Soc.* 478, 2962–2975. doi:10.1093/mnras/sty1174

Zeng, S., Jiménez-Serra, I., Rivilla, V. M., Martín-Pintado, J., Rodríguez-Almeida, L. F., Tercero, B., et al. (2021). Probing the Chemical Complexity of Amines in the ISM: Detection of Vinylamine (C2H3NH2) and Tentative Detection of Ethylamine (C2H5NH2). *ApJL* 920, L27. doi:10.3847/2041-8213/ac2c7e

Zeng, S., Quénard, D., Jiménez-Serra, I., Martín-Pintado, J., Rivilla, V. M., Testi, L., et al. (2019). First Detection of the Pre-biotic Molecule Glycolonitrile (HOCH2CN) in the Interstellar Medium. *Monthly Notices R. Astronomical Soc. Lett.* 484, L43–L48. doi:10.1093/mnrasl/slz002

Zeng, S., Zhang, Q., Jiménez-Serra, I., Tercero, B., Lu, X., Martín-Pintado, J., et al. (2020). Cloud-cloud Collision as Drivers of the Chemical Complexity in Galactic Centre Molecular Clouds. *Galactic Centre Mol. clouds* 497, 4896–4909. doi:10.1093/mnras/staa2187

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.