

# Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition

Iris A. Bermejo,<sup>†,□</sup> Imanol Usabiaga,<sup>‡,□</sup> Ismael Compañón,<sup>†,□</sup> Jorge Castro-López,<sup>§</sup> Aran Insausti,<sup>‡,¶</sup> José A. Fernández,<sup>‡,Ⓢ</sup> Alberto Avenzoa,<sup>†</sup> Jesús H. Busto,<sup>†</sup> Jesús Jiménez-Barbero,<sup>||,Ⓢ,∇</sup> Juan L. Asensio,<sup>Ⓢ</sup> Jesús M. Peregrina,<sup>†</sup> Gonzalo Jiménez-Osés,<sup>†</sup> Ramón Hurtado-Guerrero,<sup>\*,§,#</sup> Emilio J. Cocinero,<sup>\*,‡,¶</sup> and Francisco Corzana<sup>\*,†,Ⓢ</sup>

<sup>†</sup>Departamento de Química, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain

<sup>‡</sup>Departamento de Química Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV-EHU), 48080 Bilbao, Spain

<sup>§</sup>Institute of Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, BIFI-IQFR (CSIC), 50018 Zaragoza, Spain

<sup>||</sup>CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain

<sup>Ⓢ</sup>Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain

<sup>∇</sup>Department of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Spain

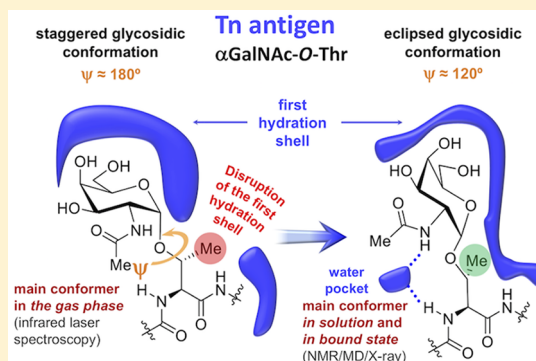
<sup>Ⓢ</sup>Instituto de Química Orgánica General, IQOG-CSIC. 28006 Madrid, Spain

<sup>#</sup>Fundación ARAID, 50018 Zaragoza, Spain

<sup>¶</sup>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,  $\alpha$ GalNAc-O-Thr and  $\alpha$ GalNAc-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  $\alpha$ GalNAc-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.<sup>1,2</sup> In general, the aggressiveness of the carcinoma and the occurrence of these antigens has a clear direct correlation,<sup>3</sup> promoting their use as biomarkers and potential therapeutic targets against cancer.<sup>4</sup> 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.<sup>5</sup>

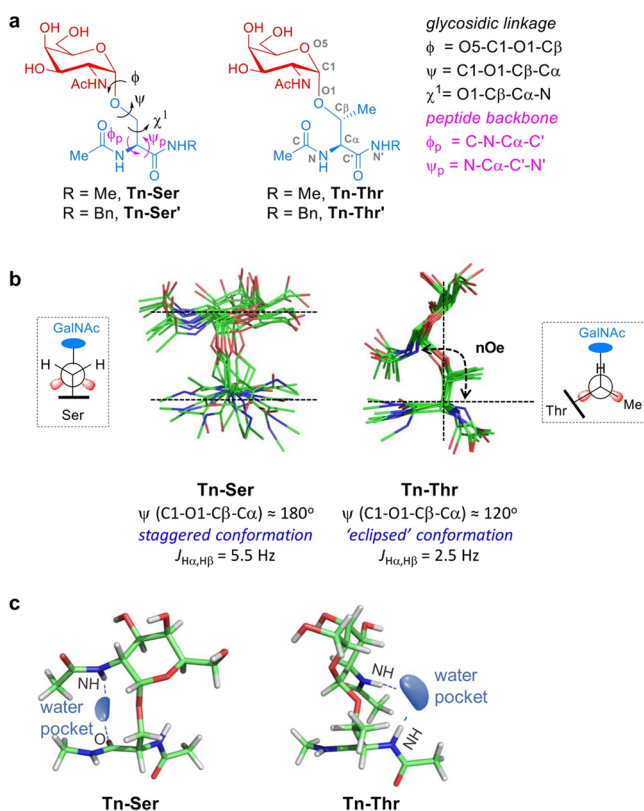
In general, the Tn antigen is referred to as GalNAc  $\alpha$ -O-linked to a serine or a threonine residue (Tn-Ser and Tn-Thr,

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.<sup>6–12</sup>

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.<sup>6</sup> On the other hand, several anti-Tn antibodies show a clear preference for glycopeptides containing the Tn-Ser antigen.<sup>13</sup> 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.<sup>7,8</sup> The values of the  $\psi$  torsion angle of the glycosidic linkage and of the associated coupling constant  $J_{\text{H}\alpha,\text{H}\beta}$  are shown. The Newman projections for the  $\text{C}\beta\text{-O}1$  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.<sup>7,8</sup> 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.<sup>9</sup> In the context of antifreeze glycoproteins, which consist of the tandem repeating  $(\text{Thr-Ala-Ala})_n$  polypeptide glycosylated with  $\text{Gal}\beta(1,3)\text{-GalNAc}$ , the replacement of a threonine by a serine residue eliminates the antifreeze activity.<sup>14</sup> 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.<sup>4</sup>

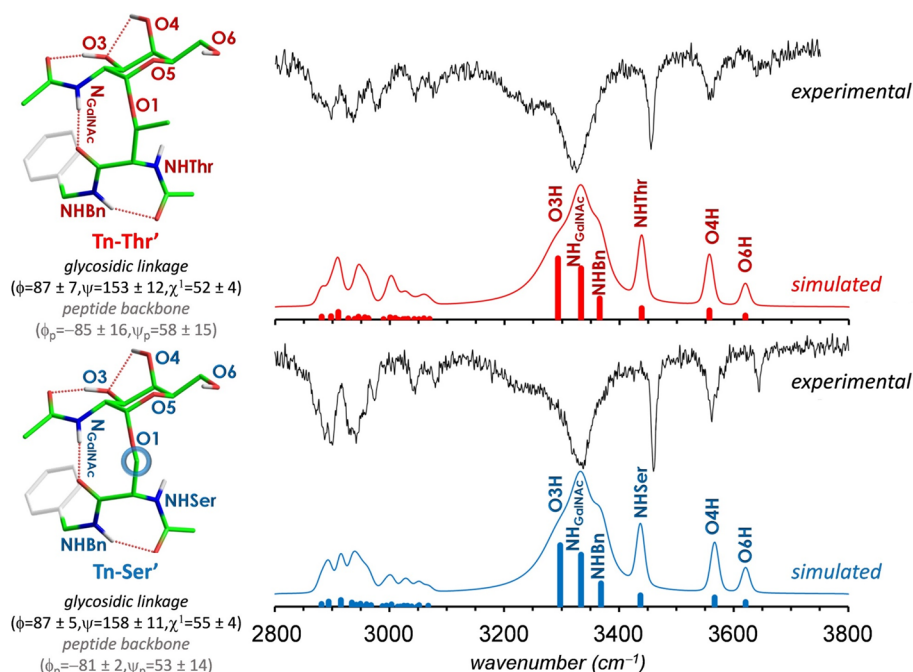
Our conformational analysis based on NMR data combined with experiment-guided Molecular Dynamics (MD) simulations,<sup>15</sup> showed that **Tn-Thr** (Figure 1b) is rather rigid in solution,<sup>8</sup> with its *O*-glycosidic linkage in the so-called “eclipsed” conformation ( $\phi \approx 80^\circ$ ,  $\psi \approx 120^\circ$ ), and its side chain fixed at a  $\chi^1$  torsional angle value around  $60^\circ$  (Figure 1a). Conversely, **Tn-Ser** (Figure 1b) displays the typical *exo*-anomeric/*syn* conformation for the glycosidic linkage with  $\phi$  and  $\psi$  values  $\sim 80^\circ$  and  $\sim 180^\circ$ , respectively. This latter system is more flexible and exhibits the three possible staggered conformers for the side chain in solution.<sup>7</sup> 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  $\psi$  torsion angle.<sup>8</sup> 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 <sup>15</sup>N- and <sup>13</sup>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.<sup>8</sup> 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.<sup>16–18</sup> To facilitate their detection through mass-selected ultraviolet (UV) photoionization, the Tn antigens were synthesized as benzamide 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<sup>15</sup> (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  $\gamma$ -turn stabilized by a strong hydrogen bond between the amino acid C-terminal amide and the N-terminal acetamide carbonyl group (band at  $\sim 3370 \text{ cm}^{-1}$  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



**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.<sup>20</sup> 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).

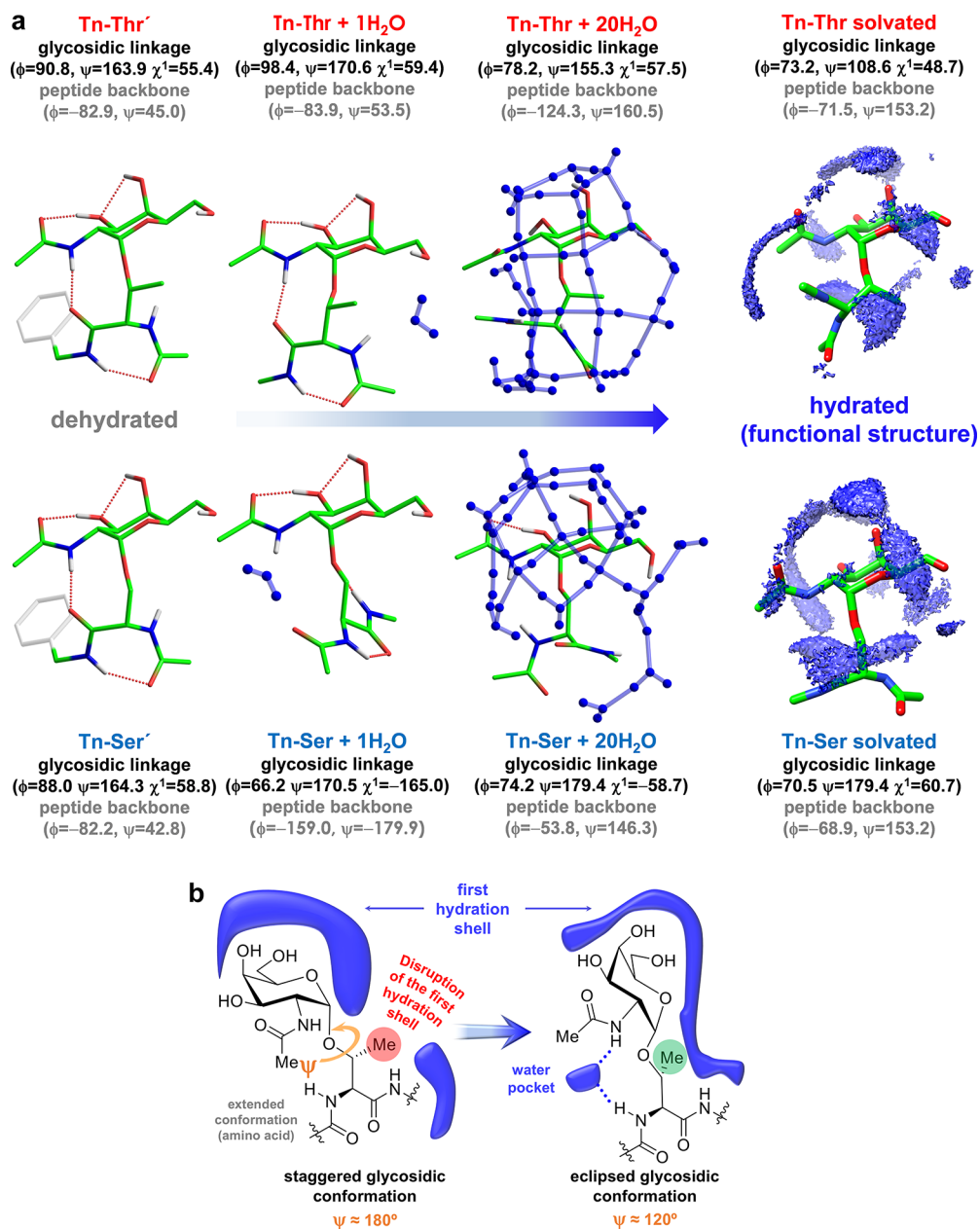
O3H hydroxyl group and the acetamide NH acts as a H-bond donor to the underlying amino acid C-terminal carbonyl group. As a result, the  $\psi$  torsion angle of the glycosidic linkage is locked to the value  $158 \pm 11^\circ$  for the **Tn-Ser'** antigen and  $153 \pm 12^\circ$  for the **Tn-Thr'** 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  $\sim 3200$ – $3350$   $\text{cm}^{-1}$  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 $\cdots$ O=C) is displaced about  $300$   $\text{cm}^{-1}$  and it is broadened around  $200$   $\text{cm}^{-1}$ . Although NH groups display similar behavior, they display narrower bands and smaller displacements than OH groups.<sup>16b,c</sup> In this case, the acetamide NH of GalNAc (strong NH<sub>GalNAc</sub> $\cdots$ O=C) is displaced around  $100$   $\text{cm}^{-1}$ , and it is broadened about  $100$   $\text{cm}^{-1}$ . The remaining hydroxyl groups of the sugar are engaged in weak hydrogen bonds, characterized by bands within the  $3500$ – $3700$   $\text{cm}^{-1}$  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<sup>19</sup> 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  $\beta$ -methyl group of the threonine derivative.<sup>8</sup> 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  $\psi \approx 165^\circ$ ), 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  $\gamma$ -turn form observed *in vacuo* toward the extended arrangement populated in solution. In contrast, the **Tn-Thr** $\cdots$ H<sub>2</sub>O 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  $\psi \approx 120^\circ$  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





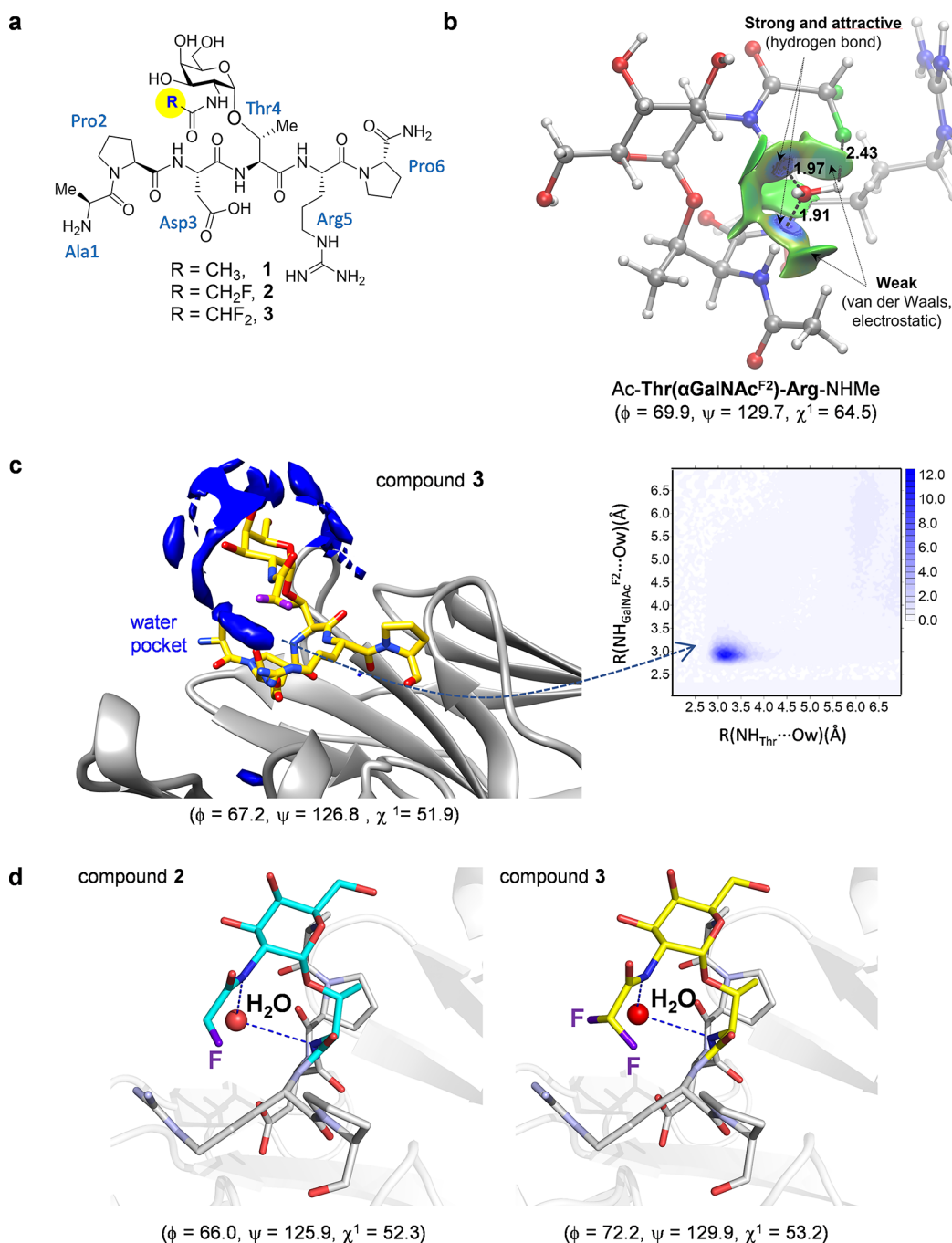
**Figure 3.** Gradual solvation and first hydration shell of the Tn antigens. (a) Lowest energy conformers calculated at the M06-2X/6-31+G(d) level<sup>20</sup> for **Tn-Thr** (upper panel) and **Tn-Ser** (lower panel) with discrete water molecules (1 and 20), together with the averaged first hydration shell derived from the experiment-guided 1  $\mu$ s 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  $\gamma$ -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  $\beta$ -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  $\beta$ -methyl group of the threonine residue (Figure S8), forcing the  $\psi$  glycosidic torsion to rotate around  $60^\circ$  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.

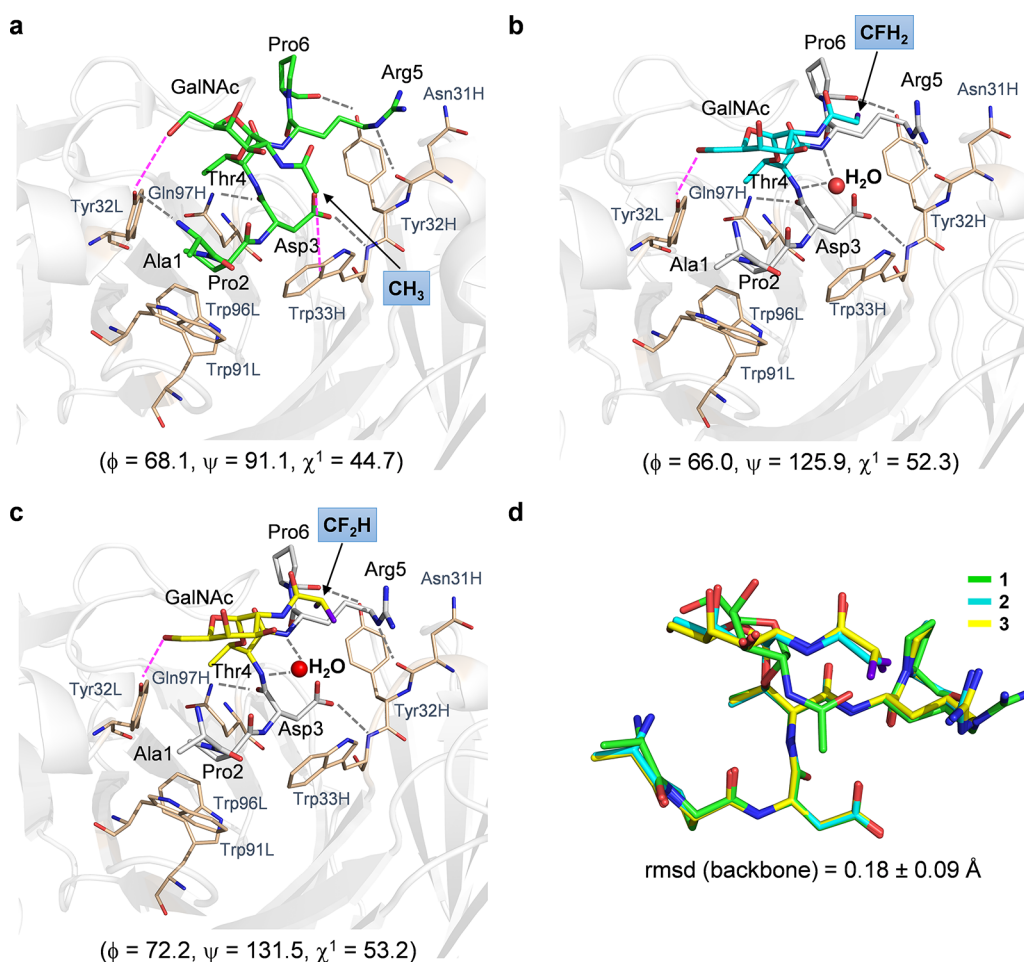
**Analysis of the Crystal Structures of Fluorinated Glycopeptides Bound to an Anti-MUC1 Antibody.** As previously



**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.<sup>22</sup> (b) Non-covalent interactions for the water molecule bound to the hydrophilic pocket of model glycopeptide Ac-Thr( $\alpha$ GalNAc<sup>F2</sup>)-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.<sup>20,23</sup> (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<sup>24</sup> 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,<sup>6a</sup> the structure of the complex between the Thr-containing glycopeptide 1 (Figure 4a) when bound to an anti-MUC1 antibody (SM3)<sup>21</sup> 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.<sup>8</sup> 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<sup>6b</sup> (compounds 2 and 3 in Figure 4a; see



**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.

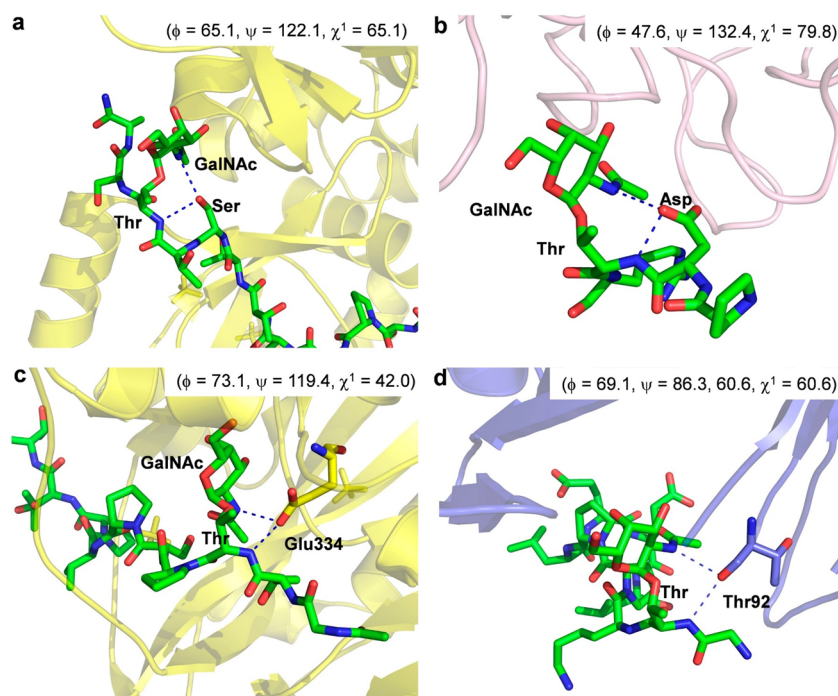
also [Supporting Information](#), Scheme S3 and Figures S43–S56). In line with this idea, quantum mechanical calculations performed on a reduced model of **3** indicate that the water molecule is stabilized not only by two hydrogen bonds with the NH groups of the sugar and Thr residues, but also through 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<sup>6,22</sup> 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<sup>22</sup> 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.<sup>8,25</sup> 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 residue in the fluorinated variants, and is similar to that found in the crystal structure reported for the naked peptide<sup>21</sup> ([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





**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 MUCSAC-13, showing the lectin domain (PDB ID: 5AJP).<sup>27</sup> (b) Zoom-in of the crystal structure of soybean agglutinin from *Glycine max* in complex with the glycopeptide PDT( $\alpha$ GalNAc)R (PDB ID: 4D69).<sup>30</sup> (c) Zoom-in of the crystal structure of the inactive form of GalNAc-T2 in complex with UDP and the glycopeptide MUCSAC-3,13, showing the catalytic domain (PDB ID: 5AJO).<sup>27</sup> (d) Zoom-in of the crystal structure of the antibody 237 in complex with its glycopeptide epitope (PDB ID: 3IET).<sup>28</sup> In all cases, the geometry of the glycosidic linkage is shown in parentheses.

in the solid state for glycopeptides **2** and **3**, with  $\psi$  close to  $120^\circ$  (Figure 5b,c), which is stabilized by the occurrence of the bridging water molecule. In the natural glycopeptide, however, the torsional angle  $\psi$  takes a value close to  $90^\circ$ . This orientation favors a CH- $\pi$  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  $\psi$  around  $120^\circ$  in solution.<sup>6</sup> Moreover, the glycosidic linkage of **Tn-Thr** antigen in complex with other proteins/enzymes can adopt variable values for  $\psi \approx 90^\circ$  to  $130^\circ$  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  $\text{CH}_2\text{F}$  and  $\text{CHF}_2$  groups. As a consequence, the crystal structures do not confirm the proposed additional  $\text{O}-\text{H}\cdots\text{F}$  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,<sup>26–28</sup> 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<sup>29</sup> 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  $\text{C}_\beta$  disturbs the proper solvation of the “native” gas-phase

geometry and  $\psi$  rotates around  $60^\circ$  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,<sup>6,13</sup> enzymes (GalNAc-transferases),<sup>26,27</sup> and lectins.<sup>30</sup> In contrast, for the Tn-Ser antigen, different arrangements of the glycosidic linkage may occur in the bound state since the lack of the  $\beta$ -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,<sup>31–33</sup> the extension of this mechanism to the  $\beta$ -O-GlcNAc-Ser and  $\beta$ -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.es

\*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 Cáncer 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 BIO-STRUCTX\_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-Molinero, 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-González, 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-González, 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.; Fukushima, 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.



(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.; Voadlo, 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.; Hiram, 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-González, 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.