Chem Soc Rev

TUTORIAL REVIEW

View Article Online
View Journal | View Issue

Cite this: *Chem. Soc. Rev.*, 2017, **46**, 7154

Principles of mucin structure: implications for the rational design of cancer vaccines derived from MUC1-glycopeptides

Nuria Martínez-Sáez, [ab] Jesús M. Peregrina [a] and Francisco Corzana *[a]

Received 29th November 2016

DOI: 10.1039/c6cs00858e

rsc.li/chem-soc-rev

Cancer is currently one of the world's most serious public health problems. Significant efforts are being made to develop new strategies that can eradicate tumours selectively without detrimental effects to healthy cells. One promising approach is focused on the design of vaccines that contain partially glycosylated mucins in their formulation. Although some of these vaccines are in clinical trials, a lack of knowledge about the molecular basis that governs the antigen presentation, and the interactions between antigens and the elicited antibodies has limited their success thus far. This review focuses on the most significant milestones achieved to date in the conformational analysis of tumour-associated MUC1 derivatives both in solution and bound to antibodies. The effect that the carbohydrate scaffold has on the peptide backbone structure and the role of the sugar in molecular recognition by antibodies are emphasised. The outcomes summarised in this review may be a useful guide to develop new antigens for the design of cancer vaccines in the near future.

Key learning points

(1) Main features of tumour-associated mucin-1 (MUC1). Scope and limitations of MUC1-based vaccines to treat cancer.
(2) Structural properties of the most immunogenic fragment of MUC1 in solution. Stabilising factors that explain why α-*O*-glycosylation prompts a 'stiffening effect' and favours extended conformations of the peptide backbone of MUC1.
(3) Conformational differences between α-*O*-GalNAc-serine and α-*O*-GalNAc-threonine pairs.
(4) Structural analysis of MUC1 derivatives bound to anti-MUC1 antibodies and factors that govern the molecular recognition process.
(5) Description of the structural key points that can be useful for the rational design of MUC1-derived cancer vaccines.

1. Introduction

1.1. Mucins

Mucins are the most abundant macromolecules in mucus and are responsible for its biochemical and biophysical properties. They are a closely related family of high molecular weight *O*-glycoproteins that play a critical role in the renewal and differentiation of the epithelium, cell adhesions, immune response, and cell signalling.[1–3] These glycoproteins—designated MUC1 to MUC21—are divided into two groups by subcellular localization; secreted and membrane-associated mucins. For instance, MUC2, MUC5AC, MUC5B, MUC6-9 and MUC19 are secreted into the extracellular region and their biological roles are related to epithelial protection, whereas MUC1, MUC3A, MUC3B, MUC4, MUC11-18, MUC20 and MUC21 are transmembrane glycoproteins characterised by an N-terminal extracellular domain and a C-terminal cytoplasmic tail, which are connected by a unique transmembrane domain (Fig. 1). While secreted mucins interact only with proteins outside the cells, membrane-bound mucins can interact with proteins present outside or inside the cell through their extracellular domain or their cytoplasmic tail, respectively. They can also interact with other transmembrane proteins through their transmembrane region. The interaction of mucins with other proteins, as well as their biological functions and their role in diseases, are described in various reviews.[1–3]

Despite these differences, the primary structure of these proteins shares some similarities. For example, they have a variable number of tandem repeat (VNTR) domains, which are

[a] *Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, 26006 Logroño, Spain. E-mail: francisco.corzana@unirioja.es; Tel: +34 941299632*
[b] *Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Universiteitsweg 99, Utrecht, The Netherlands*

rich in proline (Pro or P), threonine (Thr or T), and serine (Ser or S) residues, and are also known as PTS domains. Both Ser and Thr residues are possible sites for *O*-glycosylation, whereas Pro has a structural role, which fixes specific conformations. The protein backbone of mucins is decorated with complex branched *O*-linked oligosaccharides, which features an α-*N*-acetylgalactosamine (GalNAc) as the first carbohydrate attached to Ser/Thr side chains, and a few *N*-glycan chains that mask the protein backbone. The carbohydrate counterparts constitute more than 60% of the mucin molecular mass. It is known that protein glycosylation blocks proteases from accessing the peptide backbone, which results in resistance to enzymatic degradation. Glycosylation also modulates the intracellular trafficking of proteins and are involved in cell-adhesion, host-microbial interactions and viral infections.

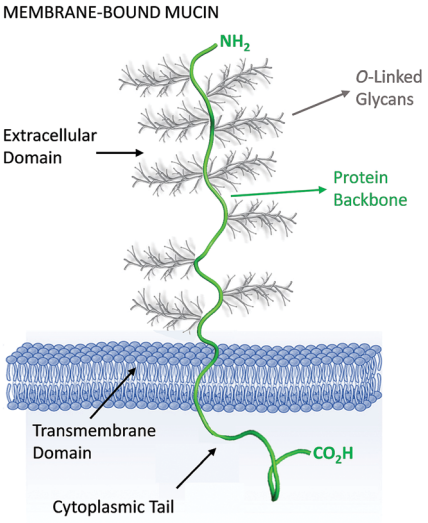


Fig. 1 Structure of membrane-associated mucins, which serve as ligands for lectins, siglecs, selectins, adhesion molecules, bacteria, and other biologically important molecules.

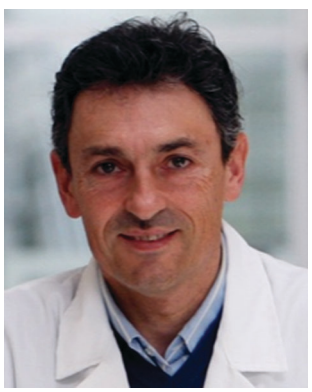
Nuria Martínez-Sáez

Nuria Martínez-Sáez was born in Logroño, Spain. She studied Chemistry at the University of La Rioja and received her BSc in 2008 and the PhD in 2013 under the supervision of Prof. Jesús Manuel Peregrina and Dr Francisco Corzana. In 2014, she moved to Cambridge where she was working as a postdoctoral researcher in the field of protein modification. Currently, she is Marie Curie Fellow at the group of Prof. Geert-Jan Boons at the University of Utrecht, The Netherlands. Her research interests focus on various aspects of Chemical Biology and Glycobiology to develop novel therapeutics and vaccines modalities related to life-threatening diseases.

The overexpression of mucins and alterations in their glycosylation are associated with the development and progression of malignant diseases, such as cancer. Therefore, mucins can be used as valuable markers to distinguish between normal and disease conditions.[1–3] For instance, MUC1 is a marker of poor prognosis in breast, prostate and lung cancer, whereas MUC4 and MUC16 are used to detect pancreatic and ovarian cancer, respectively. In general, overexpression of MUC2, MUC5AC, MUC5B and MUC6 is found in colorectal cancers.

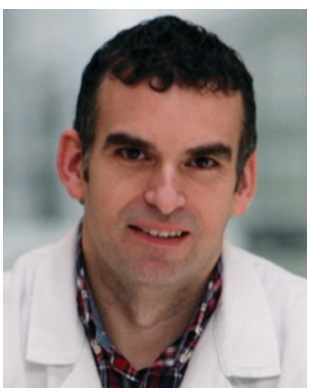
1.2. Biological relevance of MUC1

Among the mucins, MUC1 is by far one of the most studied. It consists of an extracellular domain that comprises a VNTR region of 20 amino acids (His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala) that can be repeated



Jesús M. Peregrina

Jesús M. Peregrina graduated in Chemistry at University of Zaragoza (1988). He received a PhD in Organic Chemistry from this university in 1992 under the supervision of Prof. C. Cativiela and Prof. A. Avenoza. After that, he joined the research group of A. Avenoza in the Department of Chemistry at University of La Rioja (UR) in 1993. He became an Associate Professor of Organic Chemistry at UR in 1997 and was promoted to Full Professor in 2009. His research focuses on stereoselective synthesis of amino acids and (glyco)peptides, conformational analysis and applications of nuclear magnetic resonance to wine chemistry and antimicrobial drug susceptibility.



Francisco Corzana

Francisco Corzana received a PhD in Organic Chemistry from the University of La Rioja (UR) in 2001 under the supervision of Prof. Avenoza and Dr Zurbano. He trained then as a postdoctoral fellow with Prof. Engelsen in Computational Chemistry (University of Copenhagen, 2001–2002). In 2003, he moved to the Biological Research Centre (CSIC, Madrid) to work on conformational analysis of glycopeptides by NMR and MD simulations under the supervision of Prof. Jiménez-Barbero and Dr Asensio. Since 2005, he is independent researcher at UR. His research is focused on the synthesis and conformational analysis of glycopeptides and their uses in clinical applications.

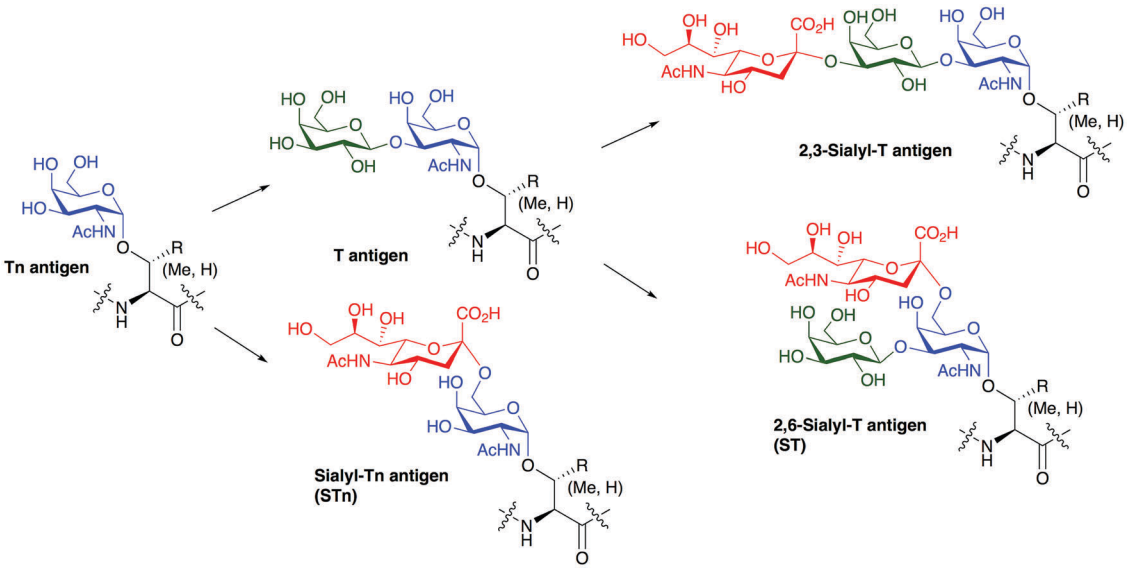


Fig. 2 Chemical structures of representative tumour-associated antigens derived from mucins.

from 20 to 125 times. This VNTR region includes five potential *O*-glycosylation sites, with three Thr and two Ser residues.

The assembly of *O*-glycans in MUC1 is initiated by the addition of GalNAc to the Ser or Thr in the PTS region to form the Tn antigen (Fig. 2). This reaction is orchestrated by a battery of enzymes named αGalNAc transferases. After the incorporation of the first carbohydrate, β1,3-galactosyltransferase catalyses the attachment of galactose to the Tn antigen to generate Core 1 *O*-glycan (also called T antigen). The residue α-*N*-acetyl glucosamine (GlcNAc) is then added by β1,6-GlcNAc transferase to obtain Core 2 *O*-glycan, which undergoes further chain elongation, followed by termination upon the insertion of a fucose or a sialic acid to the terminal sugar.[4]

Interestingly, although MUC1 is densely glycosylated in normal cells, it is aberrantly glycosylated in cancer cells and carries simple and truncated carbohydrates. As a result of the down-regulation of β1,6-GlcNAc transferase in tumour cells, the transformation of Core 1 into Core 2 is drastically reduced. Hence, immature Tn and T antigens are expressed on malignant cells. In parallel, α2,6- and α2,3-sialyltransferases are normally up-regulated, which causes premature sialylation of these antigens, and results in the formation of the sialyl-Tn antigen or the 2,3-sialyl-T and 2,6-sialyl-T determinants (Fig. 2).[4] All these antigens are referred to as tumour-associated carbohydrate antigens (TACAs). Also, incomplete glycosylation in tumour cells leads to the exposure of peptide epitopes, which are masked in healthy cells. A set of monoclonal antibodies (mAbs) recognise these structures and specifically bind to malignant cells. In this context, it is well known that the Pro-Asp-Thr-Arg (PDTR) fragment of MUC1 is an immunogenic epitope that is recognized by these anti-MUC1 antibodies.[5,6]

1.3. Applications of MUC1 to the synthesis of therapeutic vaccines to treat cancer

Tumour-associated MUC1 (TA-MUC1) is aberrantly glycosylated in around 80% of human cancers. Hence, specific antigens, such as the peptide backbone or TACAs, are exposed to the immune system. In fact, humoral and cellular immune responses against TA-MUC1 have been observed in cancer patients. The presence of circulating antibodies against MUC1 at the time of cancer diagnosis has been correlated with a favourable disease outcome in breast cancer patients. This feature makes TA-MUC1 an attractive target for cancer immunotherapy.[7–9] However, to achieve high-affinity IgG antibodies, the involvement of helper T-cells (T_h cells) is crucial. This pathway requires the processing and presentation of antigens by class II major histocompatibility complex (MHC-II) molecules, which are found in antigen presenting cells, such as B cells, followed by their recognition by T_h cells[7] (Fig. 3).

The combined approach of MUC1 glycopeptides with an immunostimulant has found widespread uses to activate the T-cell dependent pathway. Potential immunostimulants are T-cell epitope peptides, carrier proteins or Toll-like-receptor ligands based on immunologically active lipopeptides. Different combinations of these components have been used to prepare multicomponent vaccines.[7–9]

Furthermore, the immune system needs an alert signal to generate adaptive immune responses, which, in general, can be accomplished by using an adjuvant in the vaccine preparation, such as the incomplete Freund's adjuvant or the saponin adjuvant QS-21.[10]

1.4. Structure-based design of vaccines

One of the first fully synthetic two-component vaccine, consisting of a glycopeptide from TA-MUC1 and a T-cell epitope peptide from tetanus toxoid, was described more than 15 years ago by Kunz's group.[8] Although high titers of specific antibodies were induced in animals, the overall immune response was weak. More recently, the same group has confirmed robust immunogenicity of TA-MUC1-glycopeptides conjugated to carrier proteins. The induced antibodies discriminate between normal and

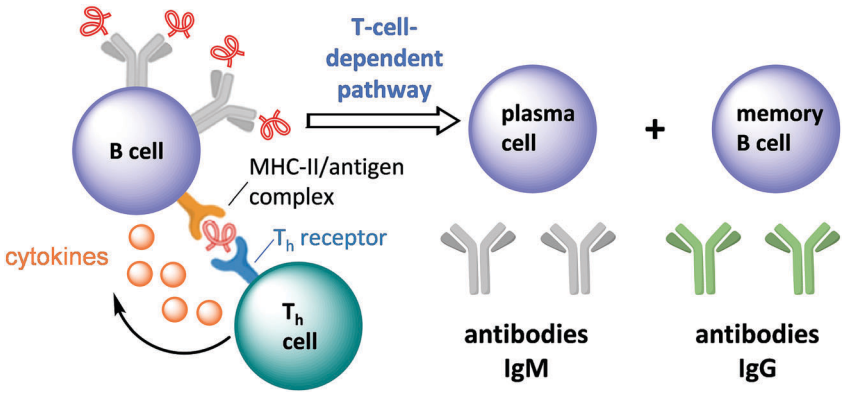


Fig. 3 Activation of T_h cells through antigen presenting cells, such as B-cells, produces a T-cell dependent pathway with production of IgG antibodies.

tumour mammary cells and permit diagnosis of human pancreatic cancer.[11] A number of research groups have pursued "multicomponent" approaches aiming at eliminating the carrier protein and boosting the immune response. For instance, Boons and co-workers have reported a three-component vaccine,[9] combining a partial sequence of the TA-MUC1 tandem repeat with the Tn antigen in the PDTR region, a T-cell epitope peptide from Poliovirus and a lipopeptide that acts as a Toll-like receptor ligand. Notably, the vaccine induces strong IgG immune responses in mice and the elicited antibodies are able to reduce the tumour growth.

Despite the advances achieved in recent years by using this sort of MUC1-based vaccines in the development of stronger immune response, it is important to note that the therapeutic effects of most of the vaccines that undergo clinical trials are not ideal or require subset analysis to prove efficacy.[12] The main drawback with therapeutic vaccines is that cancer cells can generate immune escape mechanisms,[13] which results in an increased tolerance of the antigens by the immune system. An attractive approach to tackle this issue may be the utilization of unnatural derivatives. In principle, these derivatives will be more resistant to enzymatic degradation, which could be translated into stronger and longer-lasting immunogenicity and protective efficacy.[8] For example, Kunz and co-workers[8] synthesised a vaccine carrying a fluorinated T antigen and observed that this vaccine was able to elicit strong response in mice and the resulting antibodies showed a similar structural selectivity to those obtained with the natural vaccine. Corzana and co-workers[14] developed a tripartite vaccine containing the unnatural amino acid α-methylserine glycosylated with GalNAc in the most immunogenic domain of MUC1. Although the resulting glycopeptide was more resistant to enzymatic degradation than the natural counterpart, the elicited immune response was not improved compared to a vaccine candidate containing a natural MUC1 fragment. These findings suggest that vaccine candidates should mimic the structure, dynamics and epitope presentation of the tumour-associated mucins. In addition, as described in Section 3, the elicited antibodies can interact directly with the carbohydrate moiety, which may indicate that the effect of the sugar cannot be reduced to a conformational effect.

The results achieved to date with some synthetic antitumour vaccines reveal that the use of unglycosylated MUC1 or MUC1 peptides glycosylated with complex saccharides does not elicit antibodies against cancer-expressed MUC1.[9] Other factors, including the length and the sequence of the MUC1 peptide, the tumour-associated carbohydrate antigen, the spacer used to separate the epitopes from the protein carrier, the coupling ratio of the antigen or the immunogenic carrier are crucial for the efficiency of the vaccines.[8]

Therefore, although the design of glycopeptide vaccines is based on NMR, X-ray and molecular modelling studies performed over the last 15–20 years (Sections 2 and 3), there is still much to be learned about the factors that govern antigen presentation and the structural elements required to achieve optimal antigen–antibody interactions. To really advance the design of structure-based vaccines, the effects that the carbohydrates located in the most immunogenic fragment and in distant regions, which affects antigen arrangement, need to be rationalised. It is also critical to ascertain whether the first GalNAc attached to the backbone is responsible for modification of the peptide structure, or whether the extra linked sugars could also have an impact on the conformation of the peptide and, consequently, the antigen presentation. It is also important to be able to determine which sites should be selected for glycosylation at MUC1 to boost specific immunity.

With this background information in mind, this review will focus on the most significant milestones achieved to date in the conformational analysis of tumour-associated MUC1 derivatives. Firstly, we will cover the study of free antigens in solution by emphasising the structure of the most immunogenic fragment of MUC1, and by describing the effect that glycosylation has on modulating its conformational equilibrium. We will complete this section by describing the conformational differences between the two Tn antigens (α-*O*-GalNAc–Ser *versus* α-*O*-GalNAc–Thr). We will then continue with an overview of the bound states in solution and the solid state by analysing the different geometries found for the antigen in complex to anti-MUC1 antibodies and by stressing the differences between the bound-structures and those displayed in solution. Finally, we will summarise the key elements that contribute to enhancement of the antigen–antibody binding. All these outcomes may well

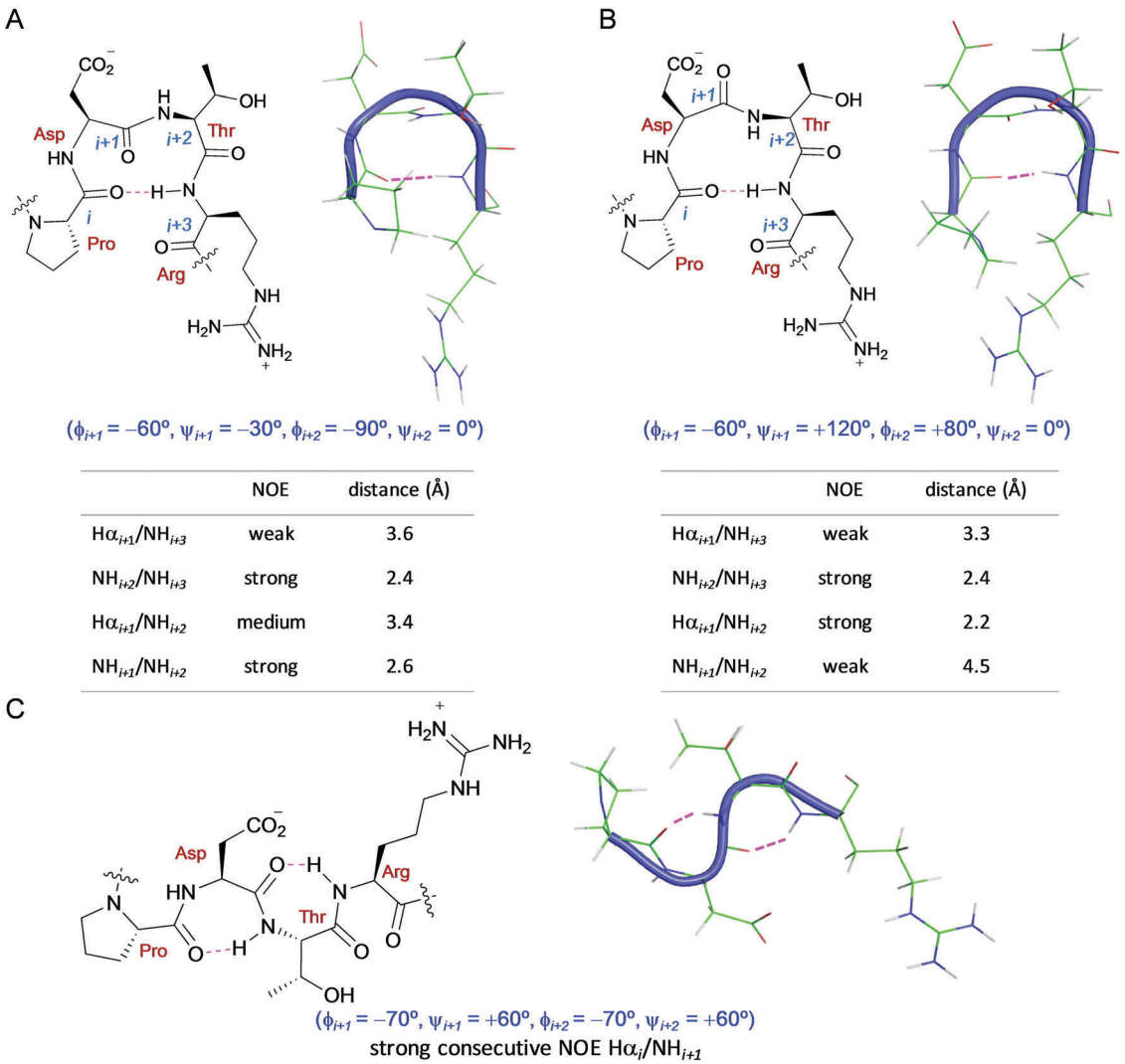
be useful as a guide to develop new antigens for the design of cancer vaccines in the near future.

2. Conformational analysis of MUC1 derivatives in solution

2.1. Conformation of the naked peptide

Most of the studies performed on MUC1-related peptides are based on NMR spectroscopic experiments, combined with molecular modelling and circular dichroism analysis. NMR spectroscopic data provides insights into the properties of peptides and glycopeptides in solution, although the results do not give direct structural outcomes. A widespread method used in NMR spectroscopy is the measurement of ^3J coupling constants. Dihedral values obtained from these constants, along with mean distances derived from nuclear Overhauser effect (NOE) data, are then used as constraints for molecular dynamics (MD) simulations. However, when molecules are rather flexible, as occurs with short glycopeptides, these calculations might generate virtual or incorrect structures. An alternative way to tackle conformational analysis based on NMR spectroscopy is to combine NMR spectroscopic data with MD simulations that are time-averaged restrained (MD-tar).[15] In this approach, the experimental data (distances from NOEs or dihedral angles from ^3J) are used as experimental restraints with the aim of obtaining a distribution of low-energy conformers (dependent on the force field) that can quantitatively reproduce NMR spectroscopic data. Hence, this procedure affords a simple and robust way to consider the flexibility of the molecule in the interpretation of NMR spectroscopic data.

Three main structures have been proposed for the sequence PDTR of MUC1; a type I β-turn, a type II β-turn or two-consecutive inverse γ-turns (Fig. 4), which can be characterised by NMR spectroscopy. A β-turn is characterised by the observation of a weak $H\alpha_{i+1}/NH_{i+3}$ cross-peak, which corresponding to a



	NOE	distance (Å)
$H\alpha_{i+1}/NH_{i+3}$	weak	3.6
NH_{i+2}/NH_{i+3}	strong	2.4
$H\alpha_{i+1}/NH_{i+2}$	medium	3.4
NH_{i+1}/NH_{i+2}	strong	2.6

	NOE	distance (Å)
$H\alpha_{i+1}/NH_{i+3}$	weak	3.3
NH_{i+2}/NH_{i+3}	strong	2.4
$H\alpha_{i+1}/NH_{i+2}$	strong	2.2
NH_{i+1}/NH_{i+2}	weak	4.5

Fig. 4 Main structures proposed in solution for the most immunogenic domain of MUC1 (PDTR). Type I β-turn (A), type II β-turn (B) and two consecutive inverse γ-turns (C), together with characteristic torsional angles (ϕ/ψ) and the most significant NOE cross-peaks that support these structures in solution.

distance of 3.6 Å for a type I β-turn, and 3.3 Å for type II β-turn, and a strong NH_{i+2}/NH_{i+3} cross-peak, which corresponding to a distance of 2.4 Å for both types of turns. A type I β-turn displays a medium $H\alpha_{i+1}/NH_{i+2}$ NOE cross-peak and a strong NH_{i+1}/NH_{i+2} peak, which corresponds to distances of 3.4 and 2.6 Å, respectively; whereas a type II β-turn shows the opposite trend, with a strong $H\alpha_{i+1}/NH_{i+2}$ NOE cross-peak and a weak NH_{i+1}/NH_{i+2} signal, for which the corresponding distances are 2.2 and 4.5 Å, respectively.

The inverse γ-turn involves a strong $H\alpha_i/NH_{i+1}$ and weak $H\alpha_i/NH_{i+2}$ cross-peaks, which corresponds to a distance of 2.4 and 4.3 Å, respectively.[16]

Initial studies on MUC1-like peptides were conducted in non-aqueous solvents or in mixtures of water and organic solvents. Under these conditions, type I β-turns appear to be the predominant conformer in solution for the most immunogenic peptide motif (PDTR). Scanlon *et al.*[17] performed NMR spectroscopic studies on 20- and 11-residue MUC1 fragments in dimethyl sulfoxide (DMSO). The authors concluded that the type I β-turn was largely populated by the PDTR fragment, which is stabilized in DMSO by the presence of a salt-bridge interaction between the aspartic and arginine side chains. This structure was also proposed by Kunz and co-workers[18] for a glycopeptide that comprised the tandem repeat sequence under near physiological conditions (phosphate buffer at pH 6.5 at 25 °C) and by using ROESY-derived constraints and MD simulations (Fig. 5A). The structure proposed by Kunz's group is in good agreement with the chemical shift deviations and temperature coefficients measured for the glycopeptide. The lack of ROESY contacts between the side chains of Asp (D15) and Arg (R17) residues indicates that the salt-bridge interaction is rarely populated in water.

Liu *et al.*[20] studied the 16-residue peptide $G^1VT^3S^4$APDTRPAPGSTA[16] and its partially glycosylated forms with GalNAc attached to Thr (T3) or Ser (S4). The study combined NMR spectroscopic experiments performed in CD_3OH/H_2O (6 : 4) at pH 5.6 with MD simulations. The distal sugar does not impose any conformational change in the PDTR region and both glycosylated and unglycosylated peptides shown a type I β-turn for the most immunogenic motif, which coexists with random coil conformations. The salt bridge between the side chains of Asp and Arg residues is not dominant.

Fontenot *et al.*[21,22] reported the structural analysis of one-, two-, and three-tandem repeats of MUC1 by NMR spectroscopy conducted in H_2O/D_2O (9 : 1) at pH 5.9, and by circular dichroism analysis. They proposed a type II β-turn conformation for the PDTR epitope and a rod-like structure for the protein core of MUC1 when two- or three-tandem repeats are displayed. The electrostatic interaction between Asp and Arg residues was also found in the structure of the 60-residue triple repeat of the MUC1 peptide. A different conformation was proposed by Kirnarsky *et al.*,[19] who analysed the conformational behaviour of a 15-residue peptide P^1PAHGVT^{7}SAPDTRPA[15] in water (Fig. 5B). The study reported NOE-derived distances, which were obtained at low temperature (5–10 °C) in water, and MD simulations. According to this investigation, the main conformations of PDTR epitope include mostly extended structures combined with two consecutive inverse γ-turns, formed by the PDT and DTR fragments and stabilised in water by weak hydrogen bonds. The authors indicate that this structure resembles an S-shape bend and is close to the conformation defined in the crystal structure with the antibody SM3.[23,24] This monoclonal antibody, induced with partly deglycosylated mucin from human milk, binds to the sequence PDTRP of TA-MUC1 present in tumour tissues.[6] A comparable structure was recently reported by Nishimura and co-workers[25] for a naked MUC1 peptide in water (a mixture H_2O/D_2O) at pH 5.5 and with a sample temperature of 5 °C.

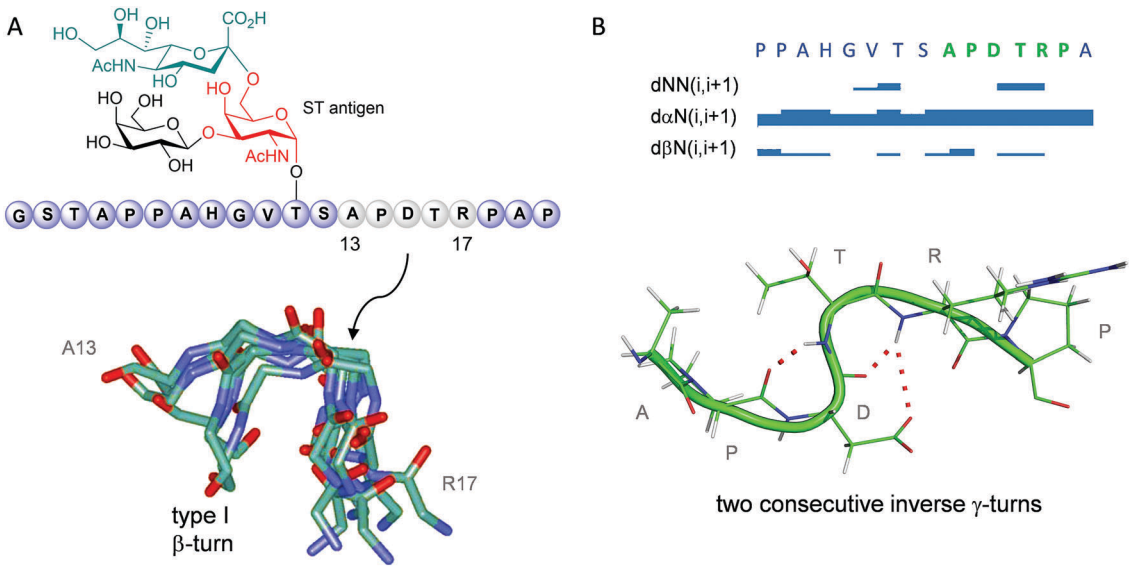
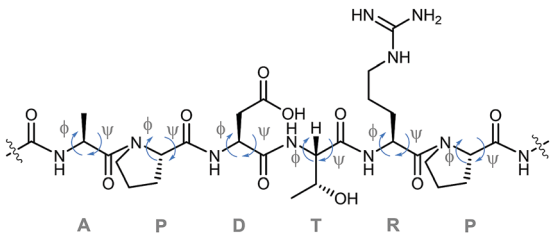


Fig. 5 (A) A β-turn-like structure proposed in water (pH 6.5, 25 °C) for the APDTR sequence of a 20-residue glycopeptide that bears the ST antigen.[18] (B) Proposed conformation of two consecutive inverse γ-turns in water (pH 4.5, 5 °C) for the same sequence incorporated into a 15mer MUC1 derived peptide.[19] In this later case, the structure is stabilised by weak hydrogen bonds (shown as red dashed lines).

Table 1 Comparison of the main reported NMR-derived structures of the APDTRP fragment



Res.	Torsion angle[a]	Type I β-turn[20]	Type II β-turn[22]	Inverse γ-turn[19]
A	ϕ	—	−159	−126
	ψ	—	148	139
P	ϕ	−107	−66	−75
	ψ	−13	**161**	114
D	$\boldsymbol{\phi}$	**−65**	**57**	**−82**
	$\boldsymbol{\psi}$	**−5**	**40**	**64**
T	$\boldsymbol{\phi}$	**−106**	**−110**	**−122**
	$\boldsymbol{\psi}$	**−3**	**35**	**68**
R	ϕ	−127	43	−103
	ψ	119	67	120
P	ϕ	—	−64	−72
	ψ	—	163	152

[a] Values of the ϕ/ψ torsion angles are given in degrees.

Campbell and co-workers[26] studied a short MUC1 peptide with a TSAPDTRPA sequence, and the Tn-glycosylated version of this peptide, TSAPDT(α-*O*-GalNAc)RPA. This thorough study involved the use of NOESY experiments in water (pH 5.1 and 5 °C) interpreted by MD simulations. The authors concluded that a type I β-turn conformation is the most sampled in water (around 45%) for the naked peptide. The type II β-turn and inverse γ-turns conformers were also populated, which indicates the presence of a complex conformational equilibrium for the free peptide. This outcome is in line with a study performed by the same group on a 16-residue peptide (GVT^{3}S^{4}APDTRPAPGSTA) and Tn-glycosylated versions of this peptide at either Thr3, Ser4 in water.[27]

As Campbell[27] and others suggest,[18,20] it is likely that the different experimental conditions used in each NMR spectroscopic study might lead to a different favoured conformation. Table 1 shows the torsion angle values displayed by the main structures proposed in solution for the APDTRP motif.

2.2. Effect of α-*O*-glycosylation on peptide backbone conformation

There are at least two plausible mechanisms by which the carbohydrate can alter the shape and dynamics of the underlying peptide backbone. On the one hand, nonspecific steric effects exerted by the bulky carbohydrate could limit the conformational space available to the underlying peptide. On the other hand, specific peptide–sugar hydrogen bonds could 'lock' the orientation of the carbohydrate with respect to the peptide and favour an extended structure for the backbone.

To shed some light on the influence of the sugar on peptide conformation, Kirnarsky *et al.*[19,28] scrutinised peptides and the corresponding Tn-glycopeptides by NMR spectroscopy (Fig. 6A). They proved that the sugar affects the conformational equilibrium of the peptide to favour a more extended structure near the glycosylated points. Of note, a medium-size NOE cross-peak contact between the amide proton of the glycosylated Thr residue and the *N*-acetyl group of GalNAc (Fig. 6B) was observed. Molecular modelling results indicate that the monosaccharide is positioned almost perpendicular to the peptide backbone.

Campbell's group[26] studied the effect of Tn glycosylation at the central Thr within the PDTRP core epitope region of the 9-amino acid MUC1 sequence. The smaller temperature coefficients of the NH protons, together with the increase in coupling constants ($^3J_{H\alpha,NH}$), and the positive chemical shift deviations of Hα protons observed for the glycopeptide corroborates that glycosylation shifts the conformational equilibrium away from a type I β-turn conformation toward a more rigid and extended structure. In addition, chemical shift perturbations analysis and ^{13}C relaxation experiments reveal that glycosylation prompts a 'stiffening effect' that is transmitted from the site of glycosylation

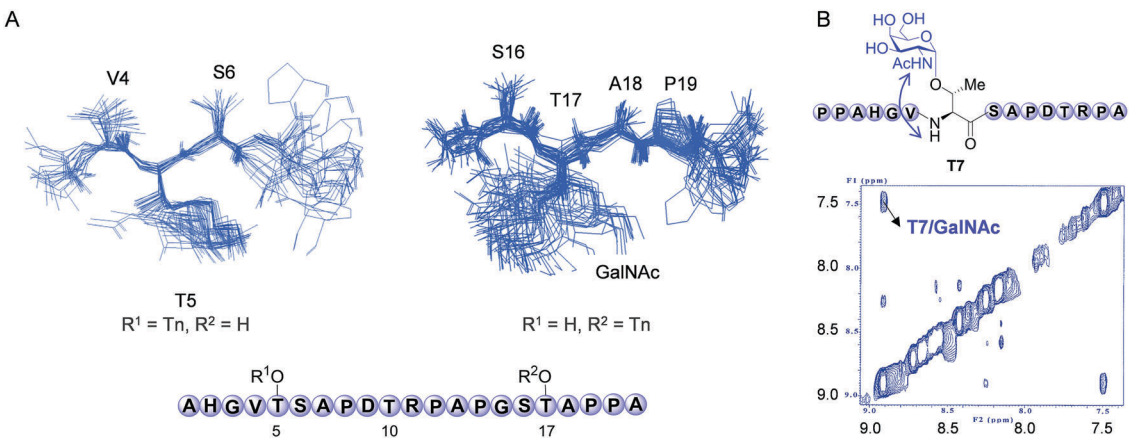


Fig. 6 (A) Most populated clusters obtained for two MUC1 glycopeptides by using NMR spectroscopy based constraints. A clear extended conformation of the peptide backbone is observed in both compounds.[28] Only the residues close to the glycosylation site are shown for clarity. (B) Amide region of the NOESY spectrum in H_2O/D_2O (9 : 1) at 5 °C (200 ms mixing time) for a 15-residue glycopeptide. A clear NOE cross-peak between T7 and the *N*-acetyl group of the GalNAc is observed.[19]

only to adjacent residues. As in the previous study, a strong NOE contact between the *N*-acetyl group of GalNAc and the NH proton of the glycosylated Thr was observed, which suggests an interaction between the peptide backbone and GalNAc (Fig. 6B). A hydrogen bond between the NH proton of the glycosylated Thr and the carbonyl group of the sugar, along with a steric effect provoked by this residue, could be responsible for the conformational change in the peptide backbone. A similar deduction was reached for clusters of the Tn[29,30] or STn antigens.[31] Strikingly, the study conducted by Campbell's group on glycopeptides derived from the sequence GVT^{3}S^{4}APDTRPAPGSTA, with singly or doubly Tn-glycosylated versions at T3 or/and S4, points out that the hydrogen bond interaction should be stronger for the αGaNAc–Thr pair than for the α-GalNAc–Ser set, as suggested by the temperature coefficient data.[27]

Danishefsky and co-workers performed an exhaustive and pioneering study of various glycopeptides derived from the mucin-like fragment STTAV of human T-cell surface glycoprotein CD43. The study involved MD simulations with NOE-derived distances and coupling constants as constraints.[32,33] Glycopeptides that have clustered Tn, T, and STn carbohydrate antigens, and a glycopeptide with a β-*O*-linked T antigen were analysed (Fig. 7A).

The study shows that compounds with α-GalNAc exhibit largely an extended structure of the peptide regardless of the extent and nature of glycosylation beyond the initial GalNAc residue (Fig. 7B), which highlighting the significance of the first α-GalNAc on the conformational preferences of mucin-like glycopeptides. In sharp contrast, the β-GalNAc derivative exhibited conformational behaviour similar to that observed for the naked peptide, which is dynamic and relatively unrestrained. A hydrogen bond, that involves the *N*-acetyl amide proton of α-GalNAc and the carbonyl group of the underlying amino acid, appears to lock the orientation of glycan with respect to the peptide backbone that stabilises the extended structure of the peptide core (Fig. 7C). The atoms involved in this hydrogen bonding differ from those proposed by Campbell's group.[26] However, the average values calculated for the NH···O distance (3.6 Å) and for the N–H–O angle (121 degrees) imply that it is difficult to assert whether these interactions are more appropriately viewed as electrostatic, or should be conventional hydrogen bonds. In this respect, the study emphases that the GalNAc amide protons are participants in bifurcated hydrogen bonds, which interact with the peptide backbone and with the oxygen of the glycosidic linkage (Fig. 7C).

In addition, the structure appears to be further stabilised by close hydrophobic contacts between several methyl groups (Fig. 7D). The molecular model proposed for one of the glycopeptides shows that the *N*-acetyl methyl groups of GalNAc units interact with methyl groups on the amino acid residues at position $i + 2$. Nishimura and co-workers[34] suggested similar pre-organization for α-GalNAc residues at consecutive Thr–Thr fragments in the formation of the extended conformations of glycopeptide MUC5AC. In this case, the *N*-acetyl methyl group of GalNAc interacts with the methyl group of a valine residue located two residues down the peptide chain.

Kunz and co-workers[18] studied the structural effects of the ST antigen located at T11 on a MUC1 peptide (Fig. 8A). The chemical shift deviation, in combination with the *J* coupling

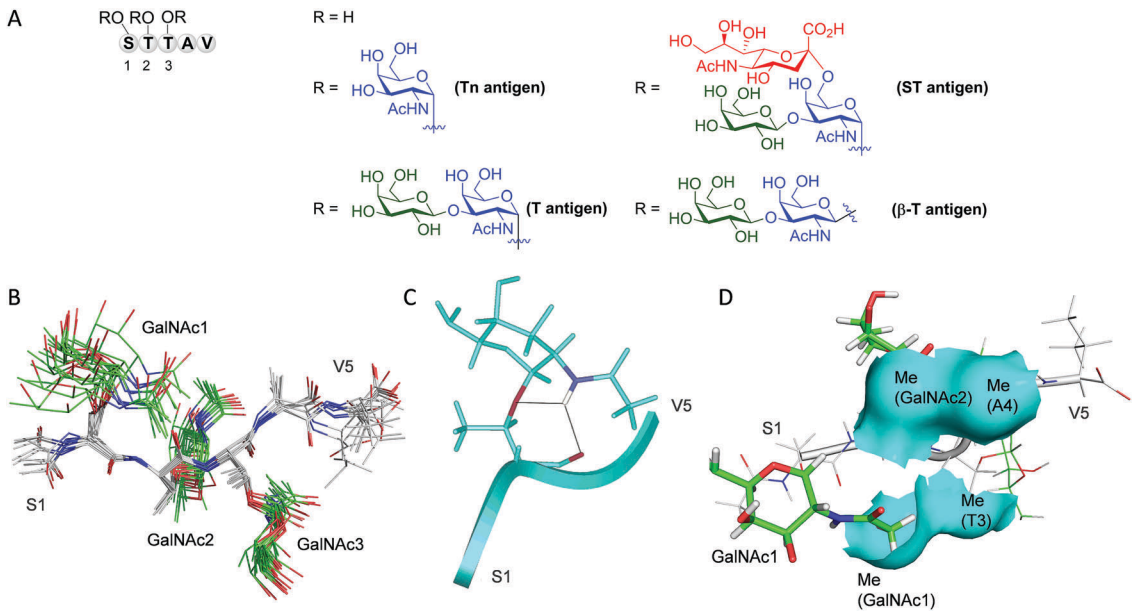


Fig. 7 (A) Mucin glycopeptides derived from cell surface glycoprotein CD43 and studied by NMR spectroscopy.[33] (B) Superposition of the peptide backbone and first sugar residue of the best structures calculated from NMR spectroscopy based constraints for the glycopeptide with the ST antigen. The carbon atoms of the peptide backbone are shown in grey and the carbon atoms of GalNAc units in green. (C) Proposed hydrogen-bonding-type interactions of the *N*-acetyl amide proton of GalNAc, shown by black lines, and the glycosylated Thr residue in mucin glycopeptides. The relevant oxygen atoms are highlighted in red and nitrogen atoms in blue. The backbone trace of the rest of the peptide portion is shown as a ribbon. (D) Molecular model proposed for the glycopeptide that has the ST antigen, which shows van der Waals interactions between the *N*-acetyl methyl groups of GalNAc and the methyl groups on amino acid residues at position $i + 2$.

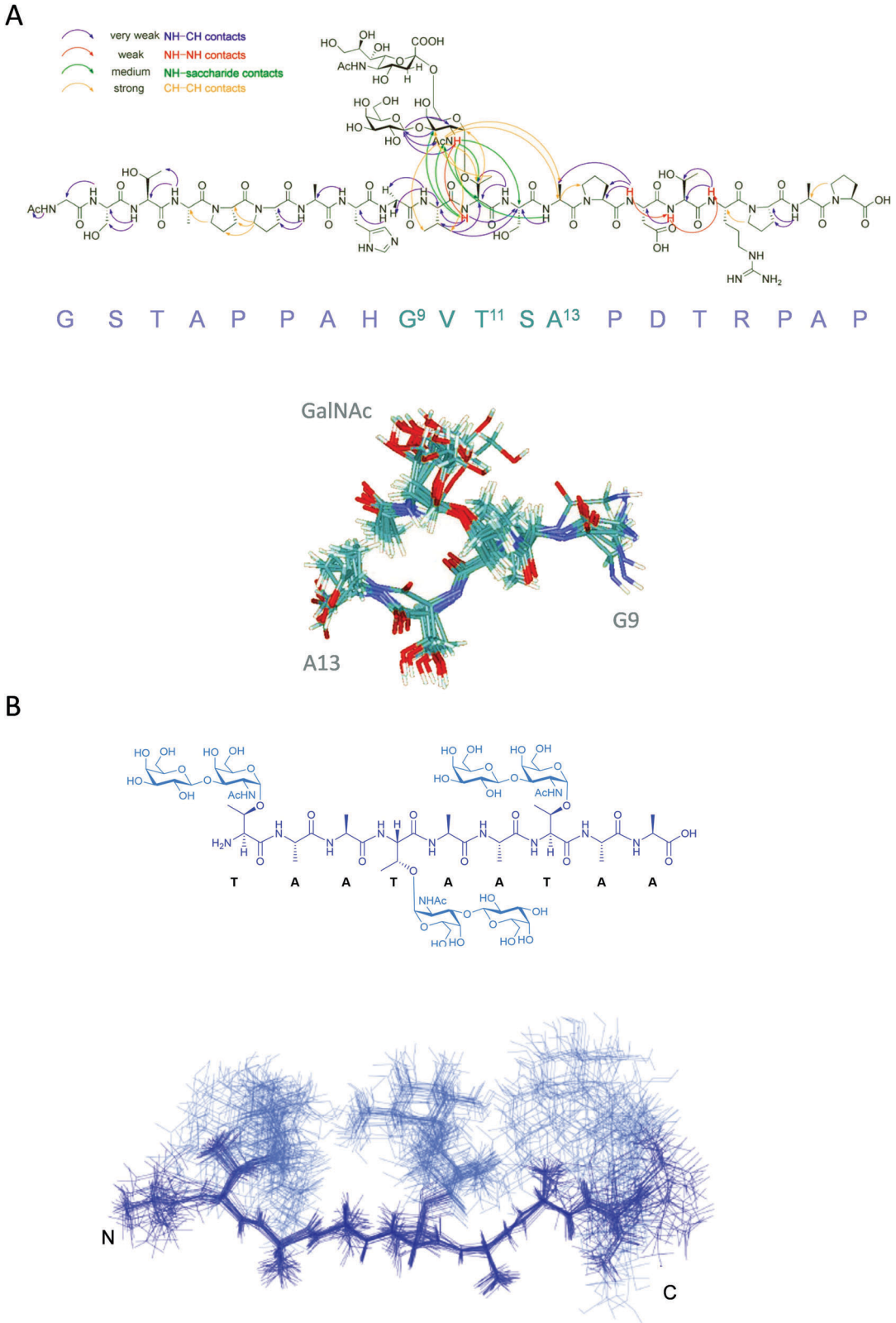


Fig. 8 (A) Top panel: Relevant ROESY contacts for the conformational analysis observed for the 20mer glycopeptide in water with the ST antigen (pH 6.5, 25 °C).[18] A weak ROESY contact between the amino group of GalNAc and the NH of the glycosylated Thr residue is observed. Bottom panel: Main structural ensemble deduced for the GVTSA fragment from NMR spectroscopy based constraints. Only the first GalNAc moiety is shown for clarity. (B) Superposition of the 25 lowest energy structures of a representative antifreeze glycopeptide calculated from NMR spectroscopy based constraints.[35] The peptide backbone adopts an extended structure in both glycopeptides.

and temperature coefficients, reinforces the finding that glycosylation shifts the conformational equilibrium of the underlying peptide backbone toward a more extended and rigid state. In addition, the MD simulations in combination with ROESY-derived distances as restraints, indicate the existence of a hydrogen bond between the carbonyl oxygen of T11 and the amide proton of GalNAc. In this way, the GalNAc is 'locked' and positioned towards the *N*-terminal part of the peptide. The atoms involved in this stabilizing interaction match those proposed previously by the Danishefsky group.[32,33]

Nishimura and co-workers demonstrated with CD studies and NMR/MD simulations that these results are not exclusive of mucin-like glycopeptides. Indeed, they observed that the hydrogen bonds between the GalNAc residue and the NH group of Thr are crucial for the presentation of the active conformation of a synthetic antifreeze glycoprotein.[35] The structure of these molecules consists of the tandem repeating polypeptide (Ala–Ala–Thr)$_n$ glycosylated with the T antigen. The replacement of Thr by a Ser residue obstructs the antifreeze activity (Fig. 8B).

Corzana and co-workers have conducted conformational analyses of short glycopeptides derived from MUC1.[30,36–38] As a result of the extra flexibility displayed by the peptide motif in this kind of compounds, the NOE-derived distances and coupling constants are used with MD-tar. In stark contrast to classical constraints, this technique provides a distribution of low-energy conformers that can quantitatively reproduce the NMR spectroscopic data. As with long glycopeptides, the α-*O*-GalNAc forces the peptide to adopt largely extended conformations in water solution, irrespective of the peptide sequence. However, the 'stiffening effect' is more relevant for the PDTR fragment relative to other glycosylation points present in the tandem repeat sequence of MUC1 (Fig. 9).[37] Besides, as reported by Campbell *et al.*,[27] different conformational behaviour between the two Tn antigens (α-*O*-GalNAc–Ser/Thr) was observed in solution for various investigated glycopeptides (see Section 2.3).[15,30]

We conclude this section by highlighting that α-*O*-glycosylation forces the peptide backbone into a stable conformation, which minimises the steric effects by further modification, and serves as the preferred peptide conformation of the epitopes recognised by some mAbs.[22] It seems that the sugar elongation from the first GalNAc residue does not influence peptide conformation.

Nishimura and co-workers[25] demonstrated that the conformational equilibrium of these systems could be more complex. For instance, they observed that although peptide **1** and glycopeptide **2** are not recognised by anti-KL-6 antibody (Fig. 10), the sialylated peptide **3** exhibits a high binding affinity to anti-KL-6. This is a promising probe for monitoring KL-6/MUC1 levels in the bloodstream,[39] which in turn is a biomarker of lung, breast, and colorectal cancer, among others. Encouraged by these results, they performed a comprehensive NMR spectroscopic study on these glycopeptides to determine their conformational preferences in solution.

As expected, glycosylation at the GVTSA and GSTA regions with the tetrasaccharide shown in Fig. 10 (glycopeptide **2**) contributed to the stabilization of the extended and a rigid conformation; no significant conformational change in these fragments was observed upon further sialylation (glycopeptide **3**). In contrast, attaching the tetrasaccharide at T12 in glycopeptide **2** induces a conformational shift of the PDTR region from a γ-turn-like (found in naked peptide **1**) into a β-turn-like structure. Further sialylation, as for compound **3**, afforded again a γ-turn-like form. This result demonstrates that the structural impact of bulky glycans can be different from that observed for simple glycosylation with GalNAc.[32]

The Nishimura group proved that glycosylation at adjacent Ser/Thr residues in the tandem repeat MUC1 glycopeptides induces significant conformational changes in the PDTR motif

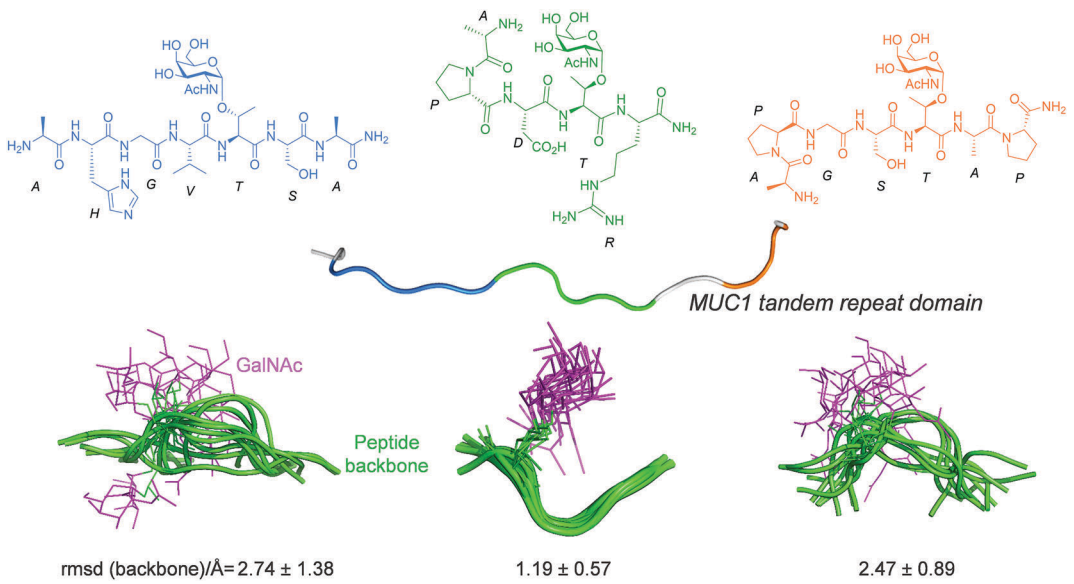


Fig. 9 Structural ensembles obtained from 20 ns MD-tar in water for three mucin glycopeptides.[37] The root-mean-squared deviation (rmsd) values for the peptide backbone are also shown. The peptide backbone is shown in green and the GalNAc unit is shown in pink.

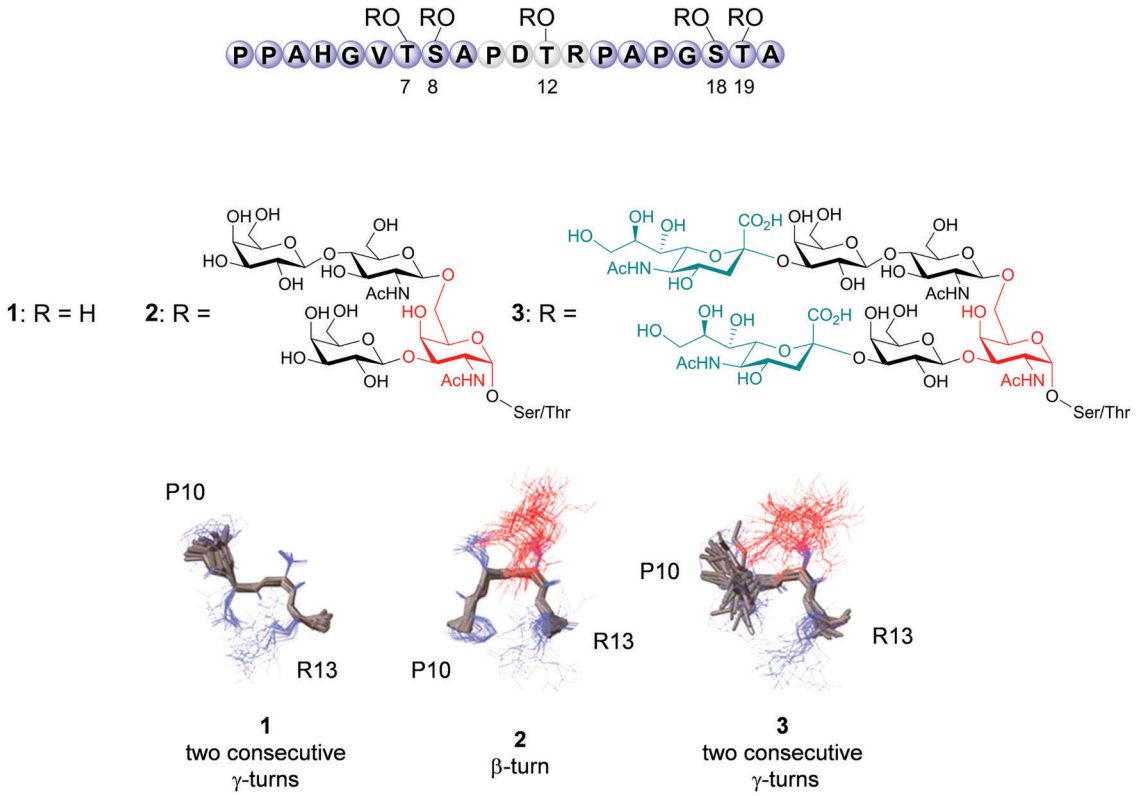


Fig. 10 Conformational impact of multiple *O*-glycosylations of the MUC1 tandem repeat.[25] The 30 best structures calculated from NMR spectroscopy based constraints of MUC1 naked peptide **1** (left), glycopeptides **2** (middle), and **3** (right) are shown. Only the peptide fragment $P^{10}DTR^{13}$ and the first GalNAc is shown for clarity. The main chain of the peptide backbone is coloured grey; the side chain is shown as a blue line, and the GalNAc moiety attached at the Thr residue is shown as a red line.

in a glycoform-dependent manner.[40] For instance, the multiple modifications by the Tn antigen at five Ser/Thr residues (compound **4**, Fig. 11) can convert the extended peptide backbone of this epitope region into a β-turn-like conformation. In contrast, multiple modifications by the 2,3-sialyl-T antigen (compound **5**, Fig. 11) did not have any conformational impact on the PDTR motif. In this respect, the authors highlight that the conformational impacts of other important *O*-glycans remain to be clarified and indicate that a better understanding of the significance of post-translational *O*-glycosylation of mucin in the generation of disease-relevant antigenic structures may contribute to the development of more adequate strategies toward new antibodies that target new epitopes with clinical potential.

As already mentioned, an interesting strategy to improve the effectiveness of the vaccines may be the use chemical modifications of the antigens to generate non-natural determinants. In this regard, Kunz and co-workers[8] developed an efficacious vaccine in mice containing a fluorinated T antigen with fluoro substituents instead of the 6- and 6′-hydroxy groups. Alternatively, Corzana *et al.*[14,41,42] have synthesised and performed conformational analysis of MUC1-derived glycopeptides that contain unnatural amino acid (*S*)-α-methylserine (MeSer) linked to α-*O*-GalNAc. This kind of α,α-disubstituted residue allows the design of MUC1-like glycopeptides with a folded structure, which is inaccessible through natural amino acids (Fig. 12A). The engineered glycopeptides can modulate the proximity between the peptide backbone and a given receptor, which resulting in an improvement in the binding affinity in some cases.[42]

Along similar lines, the crystal structure of antibody SM3 complexed with a small glycopeptide reveals that the region that contains the Thr residue adopts a helix-like conformation to properly fit in the binding pocket.[24] Therefore, a tripartite cancer vaccine candidate with a quaternary (*S*)-α-methylserine in the most immunogenic domain of MUC1 was synthesised and examined for antigenic properties in transgenic mice.[14] As expected, the unnatural residue exhibits mainly a helix-like conformation in solution. Conversely, both the peptide backbone and the glycosidic linkage display a higher flexibility relative to the natural epitope (Fig. 12B). This fact appears to negatively impact molecular recognition by the immune system. In fact, the response provoked by the unnatural vaccine (total IgG antibodies) was not improved relative to an identical vaccine that contained natural antigen (Fig. 12C). This result provides evidence that engineered MUC1-based vaccines with unnatural amino acids should emulate the conformational properties of the GalNAc–Thr pair to trigger a strong immune response.

The same group analysed numerous sulfa-Tn antigens that feature a variable linker between the carbohydrate (GalNAc) and the peptide backbone[43] (Fig. 12D). The linker blocks direct contact between the sugar moiety and the backbone and promotes a helix-like conformation for the glycosylated residue.

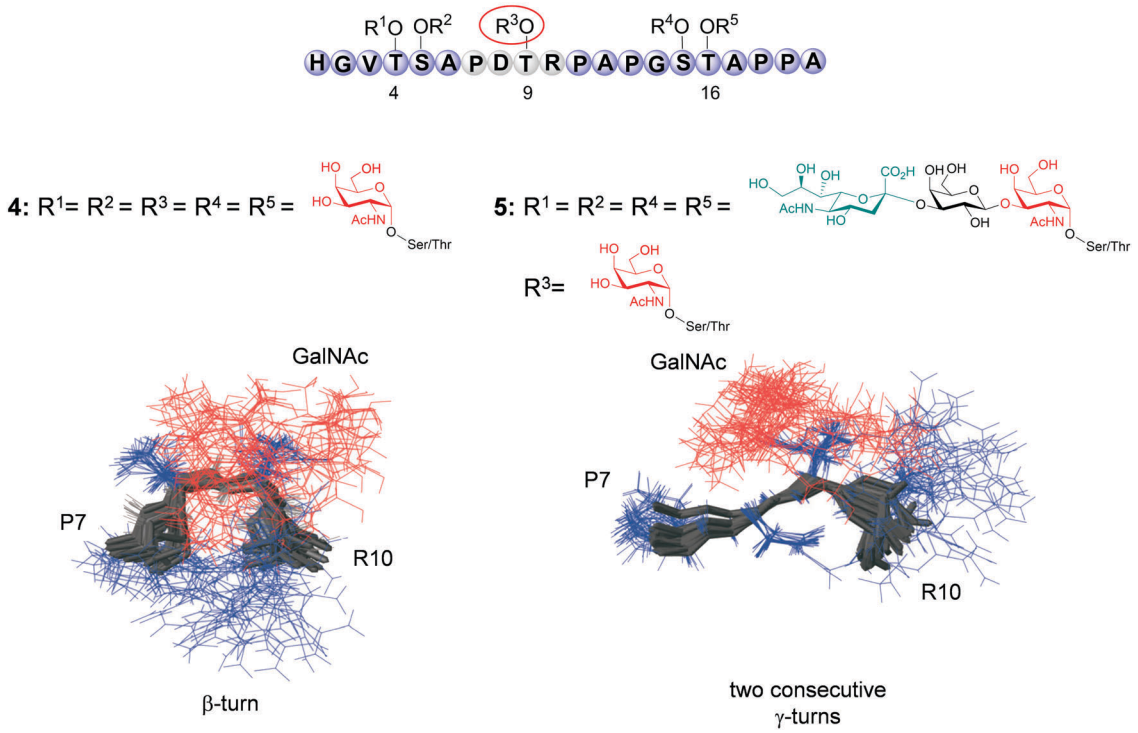


Fig. 11 Conformational impact of multiple *O*-glycosylations at adjacent Ser/Thr residues on the immunodominant PDTR region of MUC1 fragments.[40] The 30 best calculated structures in the P^7DTR^{10} region from NMR spectroscopic data for MUC1 fragments **4** and **5**. The main chain of the glycopeptide backbone is coloured grey, the side chains are shown as blue lines, and the GalNAc moiety attached at the Thr residue is shown as a red line.

These results highlight the pivotal role that the first sugar unit plays in the conformational preferences of the peptide motif. Difficulties that arise from the *exo*-anomeric effect for sulfur-for-oxygen substitution result in a flexible glycosidic linkage that can explore conformations that are absent in natural glycopeptides.

2.3. α-*O*-Glycosylation of serine *versus* α-*O*-glycosylation of threonine

In general, the Tn antigen is referred to as GalNAc α-*O*-linked to a Ser or a Thr residue without specifying which of the two amino acids the GalNAc is linked to. However, Campbell[27] and others have observed differences in terms of conformational behaviour between α-GaNAc–Thr and α-GalNAc–Ser pairs.[15,24,38,44–47]

Leading on from this result, Corzana and co-workers[15,48] carried out a thorough study of these antigens in solution by combining NMR spectroscopic data with MD-tar simulations (Fig. 13). Interestingly, α-*O*-GalNAc–Thr (**Tn-Thr**, Fig. 13) is rather rigid in solution with its *O*-glycosidic linkage in an eclipsed conformation,[15] with ϕ = 80°, ψ = 120°, and the side chain, characterised by the χ^1 torsional angle, close to 60°. This structure is in good agreement with the medium-size NOE cross-peak between the NH protons of the Thr and the GalNAc residues, and with the small $J_{H\alpha,H\beta}$ coupling value. Conversely, α-*O*-GalNAc–Ser (**Tn-Ser**, Fig. 13) is more flexible and has a typical *exo*-anomeric/*syn* conformation for the glycosidic linkage, with ϕ and ψ values ≈80° and 180°, respectively, and the three likely staggered conformers for the side chain are in agreement with the larger $J_{H\alpha,H\beta}$ coupling constant.[25] Owing to the different geometry of the glycosidic linkage, the orientation of the carbohydrate moiety varies in both molecules. Thus, although in **Tn-Thr** the sugar is almost perpendicular to the peptide, in **Tn-Ser** the GalNAc unit adopts a parallel arrangement. These conformational differences between both Tn antigens can be attributed to steric repulsions between the carbohydrate moiety (endocyclic oxygen) and the β-methyl group of the Thr. This steric clash forces the centre of mass of GalNAc to be located distant to the Thr, which leads to an eclipsed conformation of the ψ torsion angle.

A different behaviour was also observed for these derivatives in the bound state.[24] In fact, the lack of a β-methyl group in **Tn-Ser** results in a decrease in the binding affinity of this compound. This decrease is due to both the loss of the interaction between the β-methyl and the hydrophobic pocket found in **Tn-Thr**, and selection by the antibody of a conformation for the glycosidic linkage that is rarely populated in solution for the free antigen.

In contrast to previous reports[33] (Fig. 14A), this study discloses that structures that contain intramolecular hydrogen bonds between the sugar and the peptide residues are not well populated in solution, therefore they are insufficiently strong to maintain the well-defined conformation for this type of molecule. The structure of the Tn antigens can be satisfactorily explained by the presence of water pockets/bridges between the GalNAc and the peptide moieties (Fig. 14B).[15,48,49] The atoms involved in these water pockets differs from **Tn-Ser** to **Tn-Thr**,

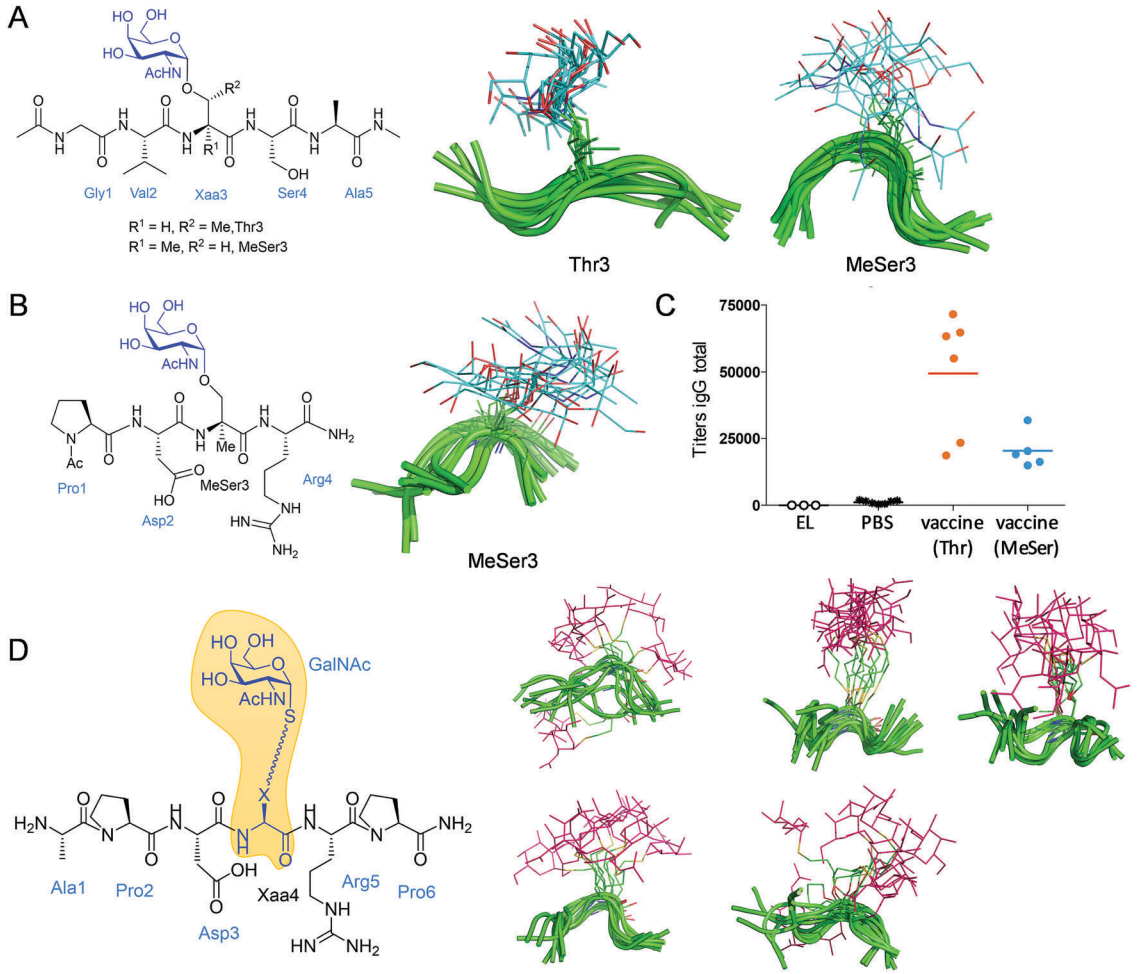


Fig. 12 (A) Structural ensembles obtained for mucin Tn-glycopeptides with Thr or unnatural residue α-methylserine (MeSer) through 20 ns MD-tar simulations in water.[41] (B) Structural ensembles obtained for Tn glycopeptide within the sequence PDTR, in which the amino acid Thr has been replaced by a MeSer residue. As a result, the peptide backbone (in green) adopts folded-like conformations and the glycosidic linkages become more flexible.[14] (C) Despite the peptide backbone of the new vaccine presenting the bioactive conformation in solution and it being more resistant to enzymatic degradation than the natural one, the immune response elicited by this unnatural vaccine is not superior to the Thr analogue.[14] (D) Design of new glycopeptides that contain a linker between the peptide backbone and the carbohydrate moiety that may favour the presentation of the latter, together with structural ensembles obtained for various glycopeptides. Notably, the linker avoids interactions between the peptide and the carbohydrate, which allows the underlying peptide to adopt largely folded conformations.[43]

which highlights the fact that these two systems structure the first hydration shell in a different way. This result could be relevant for the antifreeze activity observed exclusively in GalNAc–Thr-containing glycopeptides.[35]

3. Molecular recognition of MUC1 derivatives by anti-MUC1 antibodies

Despite a plethora of anti-MUC1 mAbs having been developed,[6] both the molecular mechanism by which MUC1 specific antibodies recognise their target and the role of glycosylation within the epitope are not well understood. Initially, it was suggested that the non-glycosylated PDTR sequence constituted the epitope of the anti-MUC1 antibodies, which could offer better access to the peptide epitope because of the lack of interference by glycans located near the epitope. However, various studies have provided evidence that glycosylation of the PDTR sequence with short glycans improves antibody binding and evokes a stronger immune response in vaccine candidates. These results agree with vaccine studies, in which incorporation of MUC1 glycopeptides has shown promising results. In contrast, the inclusion of unglycosylated MUC1 epitopes does not result in an effective immune response.[9]

Several hypotheses have been put forward to try to explain how glycosylation influences binding affinity: the carbohydrate moiety may form part of the epitope,[50] the sugar may form specific, but non-essential contacts with the antibody,[24] or the carbohydrate influences the structure of the peptide fragment of the epitope (as described in the previous section), which could favour the shape complementarity with the antibody.[51]

Enzyme-linked immunosorbent assays (ELISA) and microarrays are two methodologies that allow for quick identification of tumour-specific epitopes, which simplifies their practical

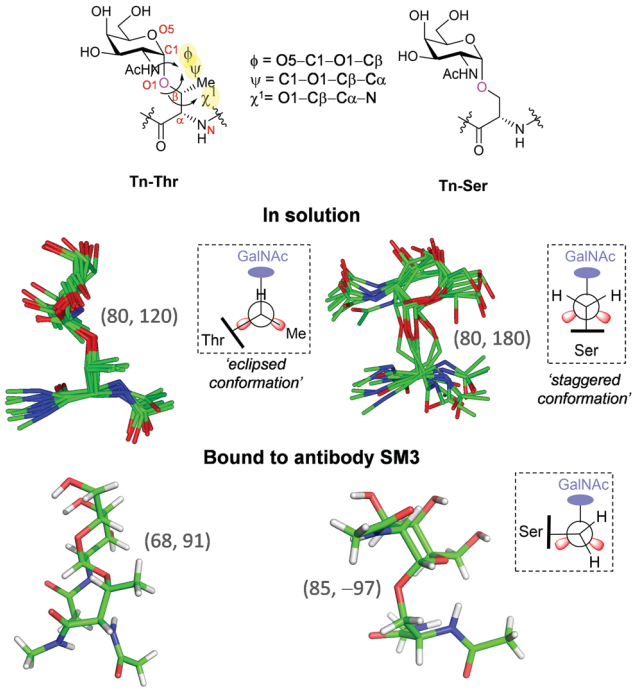


Fig. 13 Main conformation in solution[15,48] and conformation when bound to antibody SM3[24] for the Tn antigen with a Thr (**Tn-Thr**) or a Ser (**Tn-Ser**) amino acid moiety. The values of the torsional angles ϕ/ψ of the glycosidic linkage (in parentheses), together with the definition of the torsional angles, are shown. A Newman projection for the Cβ–O1 bond is also given.

application in the generation of improved antibodies for immunotherapeutic use, diagnostic tools, and design of MUC1-based cancer vaccines. For instance, Nishimura and co-workers performed an exhaustive study to determine the specific epitope structure of anti-KL-6 MAb.[39] To this end, they combined a tailored-made glycopeptide library of synthetic glycopeptides with a conventional ELISA protocol. The study demonstrated that the smallest antigenic structure recognised by the antibody is the sequence PDTRPAP, in which the Thr residue is modified by 2,3-ST antigen. It was also reported that glycosylation at neighbouring Thr/Ser residues, outside the PDTRPAP motif, strongly influences the interaction between anti-KL-6 MAb and MUC1 glycopeptides.

Nevertheless, for an in depth understanding of the mechanisms that govern the antibody–antigen interaction, a more accurate description is required. X-ray crystallography and NMR spectroscopy provide a full picture of how proteins and ligands interact on the atomic scale, characterises their interfaces and documents the contributing interactions of each residue. In the next section, we will describe how these two techniques have been used to provide a detailed understanding of the interactions that govern recognition by tumour-derived MUC1 by anti-MUC1 antibodies. This information could be essential for the design of MUC1-based vaccines with improved immune response and the development of therapeutic anti-MUC1 antibodies.

3.1. Conformational studies in solution

NMR spectroscopy has been extensively used to assess binding properties of ligands with their biological targets, and to determine their conformations under near physiological conditions in solution. Transferred NOE (trNOE) spectra give valuable information that could be used to elucidate the three-dimensional structure of the ligand in the bound state. In a complementary way, saturation transfer difference (STD) NMR spectroscopy is a technique that can be used to identify the binding epitope in a ligand–protein interaction to give detailed information about the contribution of each residue of the ligand.[52]

In recent years, these NMR spectroscopy techniques have been widely used to identify the recognition epitope of MUC1 derivatives and elucidate its conformation in the bound state with several mAbs, as well as to analyse the contribution of the carbohydrate to the recognition process.

One of the anti-MUC1 antibodies that has attracted considerable attention is SM3. Meyer and co-workers[53] performed STD NMR and trNOE spectroscopic experiments to determine the epitope mapping and the conformation of MUC1 pentapeptide P^1DTRP^5 and Tn-glycopeptide [PDT(α-*O*-GalNAc)RP] in complex with the antibody (Fig. 15). The NMR experiments proved that the naked peptide adopts an extended conformation to fit into the binding pocket (Fig. 15B). The structure is stabilized by a salt bridge between side chains of the Asp2 and Arg4 residues. These experiments indicate that the epitope is mostly located in the N-terminal region, whereas the C-terminal segment has fewer

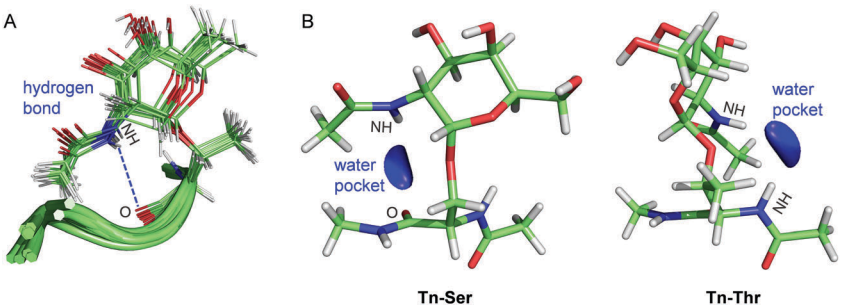


Fig. 14 Stabilizing interactions between the first GalNAc motif and the underlying amino acid favour extended conformations of the peptide backbone in mucin glycopeptides. (A) A hydrogen bond with the *N*-acetyl group of GalNAc and the carbonyl group of the glycosylated residue is formed.[33] (B) Water-bridging molecules between the peptide fragment and the sugar may explain the interactions between these entities. It is important to note that antigens **Tn-Ser** and **Tn-Thr** accommodate different water pockets owing to their distinct conformational behaviour in solution.[15,48]

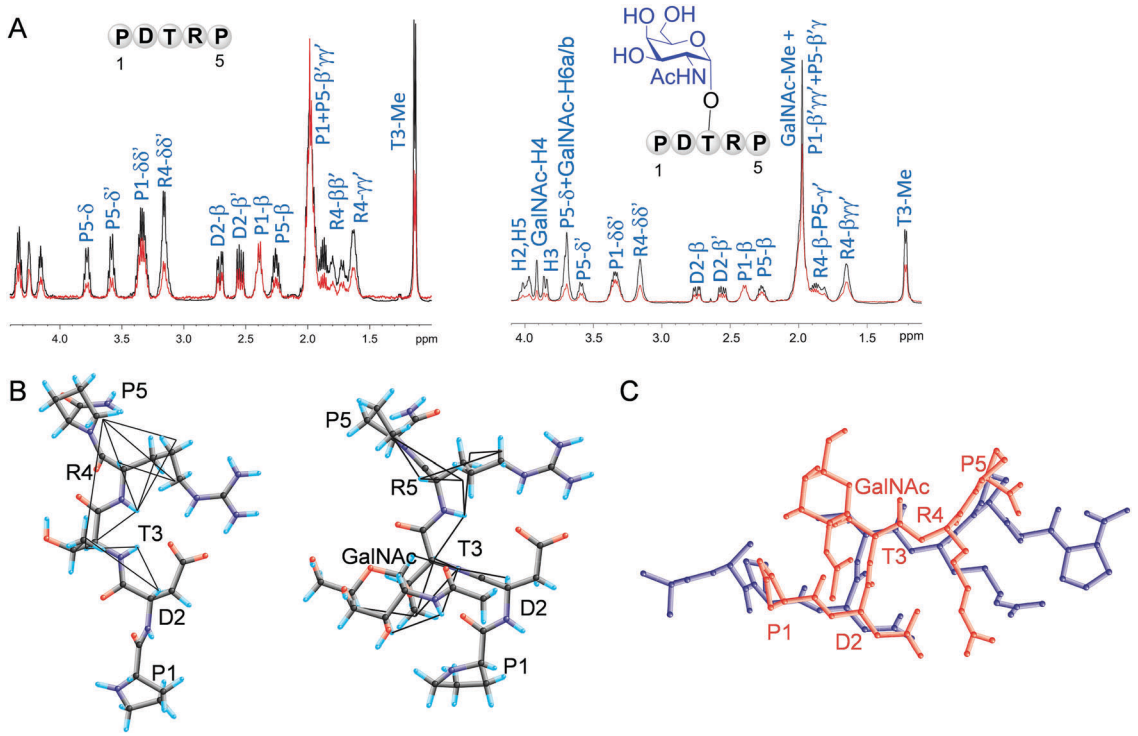


Fig. 15 (A) Superposition of the STD spectrum (red) and the reference ^{1}H NMR spectrum (black) of peptide PDTRP (left panel) and the corresponding Tn-glycopeptide bound to antibody SM3 (right panel).[53] (B) trNOE-derived structure of the peptide PDTRP (left panel) and the corresponding Tn-glycopeptide (right panel). (C) Superposition of the Tn-glycopeptide (red) with peptide SAPDTRPAP (blue) as determined from the X-ray crystal structure.[23]

interactions with the surface of the antibody (Fig. 15A). Conversely, the glycopeptide, which displays a similar conformation to the naked peptide, is engaged in contacts with SM3 through all its amino acids (Fig. 15B). The GalNAc residue interacts mainly through the *N*-acetyl residue, whereas the rest of the protons exhibit weak contacts (Fig. 15A). With regard to the dynamics of the complex, there is different flexibility in each of the N- and the C-terminal regions of the naked peptide, in which Pro5 has more flexibility and less contact with the protein. In the glycopeptide, these large differences in the flexibility of the backbone were not observed, which reflects conformational stabilisation of the peptide by the GalNAc moiety or a binding contribution by the GalNAc residue. These findings are in good agreement with the information provided by the X-Ray structure of various MUC1-like peptides and glycopeptides bound to the SM3 antibody (Fig. 15C and Section 3.2),[23,24] with the exception of the salt bridge, which is not observed in the solid state.

According to the NMR data, Pro1 seems to play a pivotal role in the recognition of MUC1 peptides and glycopeptides by SM3, which has the strongest STD signals (only comparable with the *N*-acetyl of GalNAc). In this line of research, a recent study performed by Yang and co-workers[54] analysed the effect of shortening the PDTRP epitope on the affinity to a MUC1 specific monoclonal antibody (isotype IgG1, clone 6A4). The study shows that the sequence GVTSA**PD**, which incorporates only two residues of the epitope (shown in bold letters), retains the binding affinity. However, substitution of the Pro by an Asp residue in this sequence (GVTSA**DD**) results in no interaction with the antibody.

Another significant anti-MUC1 antibody that has been widely studied is monoclonal antibody B27.29, which was raised against ovarian tumour cell derived mucin and displays specificity for MUC1 found in tumours of the ovaries and breast.[29] A thorough study to determine the binding epitope of MUC1 peptides to this antibody was performed by Campbell *et al.*[27,55] These authors use isotope-edited NMR spectroscopic methods, which include ^{13}C and ^{15}N relaxation measurements and HSQC-monitored titrations. The study was conducted with two derivatives: a 16-residue peptide sequence G^{1}VT^{3}S^{4}APDTRPAPGSTA^{16} and a two-repeat MUC1 peptide sequence (V^{1}TSAPDTRPAPGSTAPPAHG^{20})$_2$. The NMR spectroscopic data confirmed that all the residues within the PDTRPAP sequence are involved in binding (Fig. 16). Conversely, the residues outside of the epitope do not show any interaction with the antibody and remain very flexible in the bound state. These authors also studied the effect that glycosylation outside the PDTRPAP epitope produces in the recognition of MUC1 derivatives by the antibody. Double Tn-glycosylation within the GVTSA sequence enhances the binding affinity, which suggesting that B27.29 maps two separate parts of the glycopeptide; the PDTRPAP epitope region and a second carbohydrate epitope that comprises both Tn motifs at Thr3 and Ser4. Analysis by trNOE in the bound state indicated that, as with the free antigen,[26] glycosylation at Thr3 and Ser4 does not affect the PDTRP epitope conformation, which displays a type I β-turn found within the B27.29 combining site.

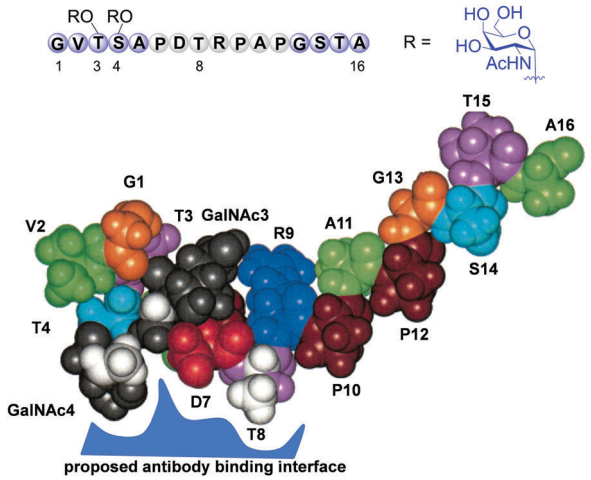


Fig. 16 Space-filled model of the Tn3,Tn4-glycosylated MUC1 peptide that shows a proposed antibody binding interface, which is based on peptide-antibody and GalNAc-antibody NOEs observed in the presence of antibody B27.29.[27] Clear contacts between the antibody and *N*-acetyl groups of Tn3 and Tn4 and antibody-T8 are observed. These atoms are shown in white in the model.

Recently, Coelho *et al.*[47] have used a combination of microarrays, NMR spectroscopic techniques and computational methods to characterise the fine epitope mapping of a MUC1 library, which incorporated peptides and glycopeptides with different GalNAc glycosylation points, towards two different families of cancer-related mAbs. The first family comprises two anti-MUC1 antibodies, VU-3C6 and VU-11E2, which recognize TA-MUC1 present in breast cancer.[6] The second group consists of two anti-Tn mAbs, 8D4 and 14D6, generated by using a synthetic Tn antigen-based vaccine (Fig. 17).[47] In anti-MUC1 antibodies, it was found that both bind to the non-glycosylated MUC1 peptides. The STD analysis identified the TR peptide moiety of the PDTRP region to be the main structural motif for the recognition of VU-3C6 mAb, whereas VU-11E2 needs a more complex epitope that involves all the amino acids in the sequence PDTRP. A similar result was obtained for glycopeptides that bear the Tn antigen within the epitope region. GalNAc recognition occurs through the H2 proton and the *N*-acetyl moiety. However, no STD response was observed for the GalNAc moiety when it is located at the GSTAP or GVTSAP regions. In terms of binding affinity, glycosylation with GalNAc at the PDTRP region improves the recognition of both antibodies with the MUC1 derivatives in a strict peptide-sequence-dependent manner, with a specific binding profile with respect to the glycosylation position. These studies also determined that glycosylation at the PDTRP sequence does not significantly modify the conformations in the bound state, because the epitope deduced by STD NMR experiments is the same as that observed for the naked peptides.

In the case of 14D6 and 8D4 mAbs, the STD-NMR approach, together with microarray studies, clearly demonstrate that the Tn motif in the MUC1 sequence is required for binding. Interestingly, both antigens show higher affinity for glycopeptides with Ser. STD NMR experiments proved that the antibodies recognise the

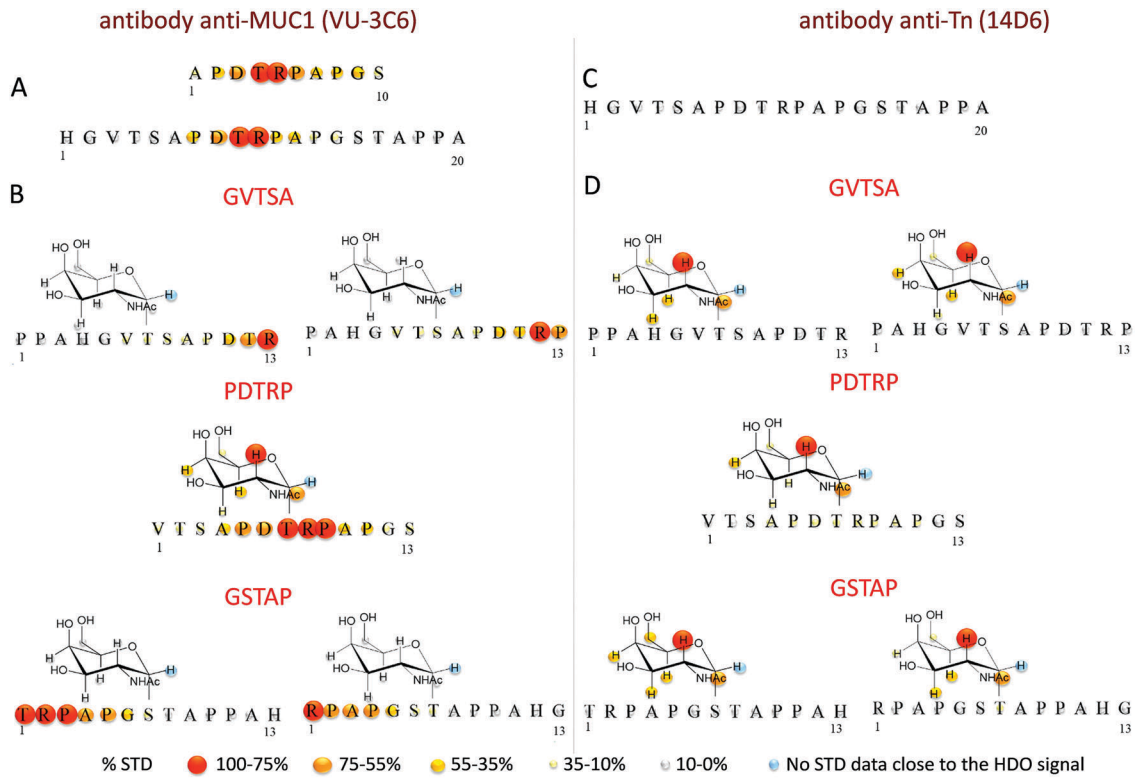


Fig. 17 STD-derived epitope for MUC1 (glyco)peptides with antibodies VU-3C6 and 14D6.[47] Naked peptides (A) and glycopeptides (B) with VU-3C6. Naked peptide (C) and glycopeptides (D) with 14D6.

GalNAc residue, with clear participation of the *N*-acetyl group and H2 for binding. These protons received much more saturation when GalNAc is attached to a Ser residue. The authors conclude that the α-*O*-GalNAc–Ser pair is more flexible, which allows the glycopeptide to display the required conformation in the bound state, without a major entropy penalty. This flexibility was not observed in the case of glycosylated Thr derivatives.

3.2. Conformational studies in the solid state

Conformational studies in the solid state with X-ray crystallography allow for an accurate picture of antibody–antigen interactions to be revealed. Nevertheless, to date, only a few X-ray structures of MUC1 antigens complexed to antibodies have been reported.[23,24,51]

The first high-resolution crystal structure of a MUC1 peptide (T^1SAPDTRPAPGST^{13}) bound to the Fab fragment of antibody SM3 was reported by Freemont *et al.*[23] This structure provided the first molecular details into the recognition of a cell surface peptide of a tumour antigen by a monoclonal antibody (Fig. 18A). The study confirmed that the conformation of the bound antigen is mainly extended, with no discernible standard secondary structure. This conformation is similar to that proposed by NMR experiments in the bound state.[53] The nine resolved residues of the antigen (Ser2 to Pro10) make a total of 114 interactions with 15 residues of the antibody, of which 93 are van der Waals contacts and 21 are either direct or water-mediated hydrogen bonds. These interactions were in total agreement with SM3 epitope mapping experiments,[53] which showed that the recognition epitope of this antibody comprises the residues P^4DTRP.[8]

Upon detailed examination of the structure, it was revealed that Pro4 stacks with Trp91L, Trp96L, and Tyr32L, which could explain why this residue is required for the binding of anti-MUC1 antibodies.[54] In addition, the side-chains of Asp5 and Arg7 are engaged in CH–π contacts with Trp33H and Tyr32H, respectively, and Thr6 is exposed to solvent and forms hydrogen bonds with surrounding water molecules and with Gln97H. Remarkably, the β-methyl group of Thr6 is engaged in a hydrophobic interaction with Tyr32L.

The first high-resolution X-ray structure that shed some light on the role of the sugar in recognition processes between antibodies and tumour-associated glycopeptides was reported by Brooks *et al.*[50] (Fig. 18B). The structure presented 237mAb, an anti-Tn antibody, in a complex with a Tn-glycopeptide derived from the tumour-associated protein podoplanin ($E^1RG^3TKP^4P^5LEEL^{11}$), and disclosed that both the peptide and the carbohydrate make specific contacts with the antibody and, consequently, form part of the epitope. The carbohydrate lies in a hydrophobic pocket that establishes at least one interaction through every single hydroxyl group. The *N*-acetyl is engaged in a hydrogen bond with the carbonyl group of a Ser and in a CH–π interaction with a tyrosine residue. With respect to the peptide backbone, it fits tightly to allow the antibody to interact through several hydrogen bonds, as well as through CH–π interactions that involve Pro6 and Pro7 residues with a histidine and a tryptophan residue, respectively.

In 2015, the Corzana group[24] provided the first crystal structure of Tn-glycosylated MUC1 epitope with antibody SM3 ([A^1PDT^4(α-*O*-GalNAc)RP^6], Fig. 19), together with the X-ray structures of Tn derivatives with Ser and Cys [APDS(α-*O*-GalNAc)RP and APDC(α-*O*-GalNAc)RP]. Bio-layer interferometry studies revealed that the substitution of Thr by Ser residue within the recognition epitope of the MUC1 glycopeptides [APDT(α-*O*-GalNAc)RP and APDS(α-*O*-GalNAc)RP] reduces drastically the binding affinity of SM3. However, glycosylation of Thr produces around a 3-fold enhancement in affinity relative to the naked peptide (APDTRP). Consequently, the study demonstrates the non-equivalence of

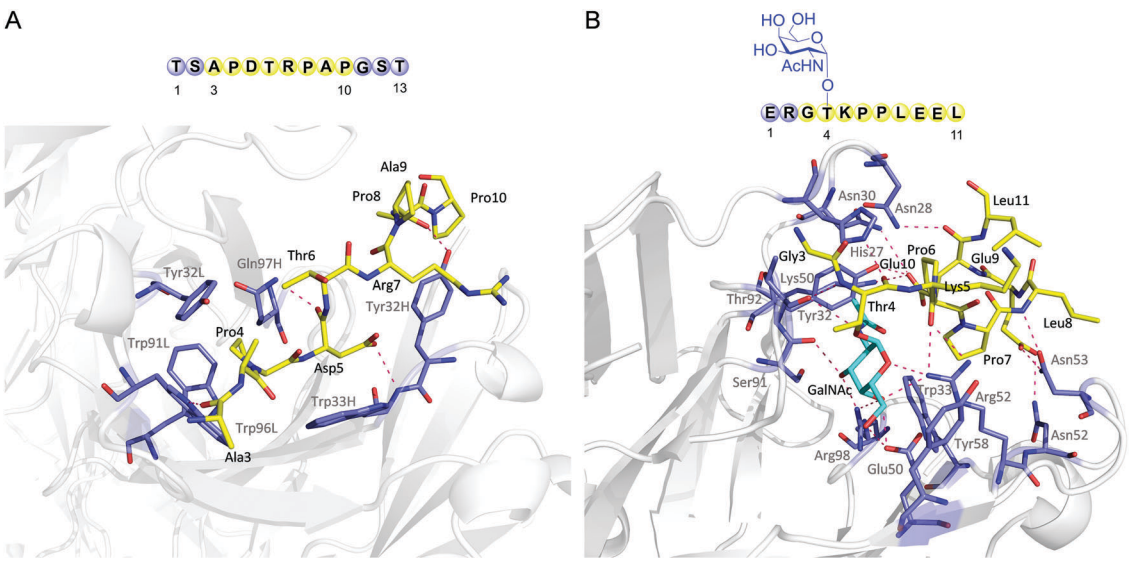


Fig. 18 (A) X-ray structure of 13-residue MUC1 peptide TSAPDTRPAPGST with Fab of antibody SM3 (PDB entry: 1SM3).[23] (B) X-ray structure of Tn-glycopeptide ERGT(α-*O*-GalNAc)KPPLEELS with the Fab region of 237mAb (PDB entry: 3IET).[50] The antigens are shown as sticks with carbon atoms in yellow. The GalNAc moiety is in blue. The antibodies are shown as white cartoons and those residues that interact with the antigen are shown as sticks with carbon atoms in purple. Red dashed lines indicate antigen–antibody hydrogen bonds.

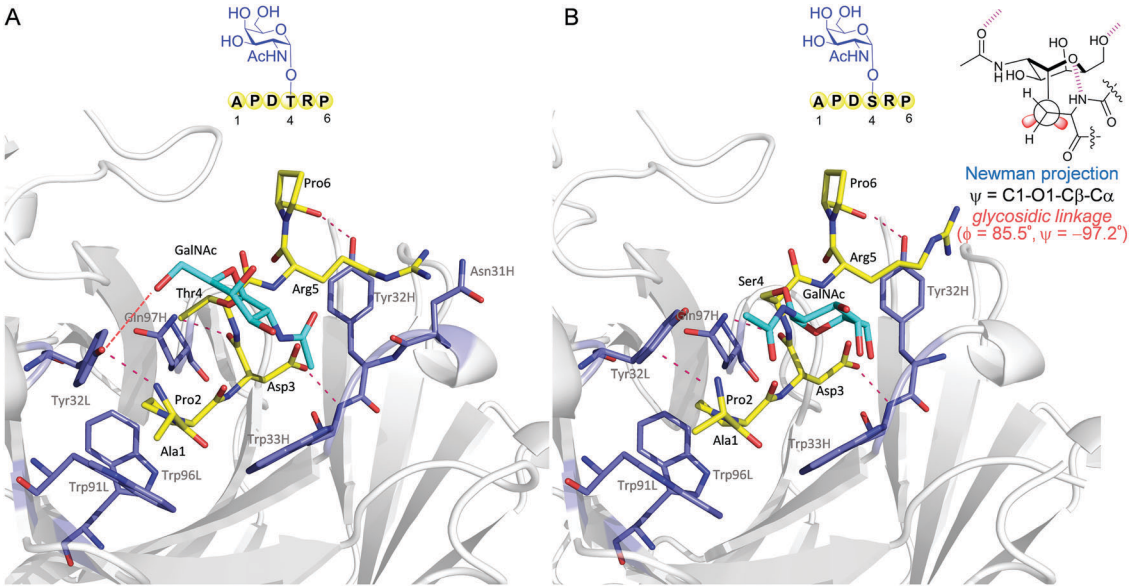


Fig. 19 (A) Binding interactions of Tn-glycopeptide derived from APDTRP with the scFV-SM3 region of SM3 revealed by X-ray crystallography (PDB entry: 5A2K). (B) Binding interactions of Tn-glycopeptide derived from APDSRP with the scFV-SM3 region of SM3 revealed by X-ray crystallography (PDB entry: 5A2I). The geometry of the glycosidic linkage for the glycopeptide with a Ser residue is shown. The antigens are shown as sticks with carbon atoms in yellow. The GalNAc moiety is in blue. The antibody is shown as white cartoons and those residues that interact with the antigen are shown as sticks with carbon atoms in purple. Red dashed lines indicate antigen–antibody hydrogen bonds.[24]

Ser and Thr *O*-glycosylation points in molecular recognition processes by antibody SM3.

Crystallographic analysis revealed that the overall conformation of the peptide fragment of the studied MUC1 variants was nearly identical to that found in the crystal structure reported for the naked peptide[23] (Fig. 18A), which indicates that the GalNAc motif, regardless of the attached amino acid (Ser, Thr or Cys), does not significantly affect the conformation of the peptide backbone in the SM3-bound state. Even though most of the interactions between the peptide fragment and the antibody are identical to those reported by Freemont *et al.*,[23] in glycopeptides with Ser or Cys, the presentation of the side chain of Arg5 differs slightly from that found for Thr analogues, and results in the lack of a hydrogen bond between the side chain of this residue and the carbonyl group of Asn31H.

The main difference between the glycopeptides that incorporate Ser and Thr in the bound state is in the geometry of the glycosidic linkage. In the Thr derivative, this linkage adopts the expected *exo*-anomeric/*syn* conformation, with ϕ and ψ values $\approx$68 and 91 degrees, respectively.[15] This conformation is similar to that found for a Tn-glycopeptide bound to 237-mAb[50] and to that disclosed for the same glycopeptide in complex with lectin SBA.[37] This geometry allows the formation of an intermolecular hydrogen bond between the hydroxymethyl group of GalNAc and the side chain of Tyr32L of the antibody. Moreover, the *N*-acetyl group of the sugar stacks with the aromatic ring of Trp33H, which provides the drive for the observed selectivity of SM3 for GalNAc-containing antigens. This interaction pattern is compatible with the above-mentioned STD NMR spectroscopic study conducted by the Meyer's group.[53] In sharp contrast, in the bound state of the Ser and Cys glycopeptides, the GalNAc unit does not establish stabilizing contacts with the antibody (Fig. 19B).

In these cases, the glycosidic linkage shows a high-energy conformation, with a value of the dihedral ψ close to -97 degrees. This conformation is barely populated in solution (around 20%), as demonstrated by MD simulations performed on the free antigens. This unusual conformation of the glycosidic linkage is stabilised by two intramolecular hydrogen bonds with the peptide backbone; one between the endocyclic oxygen of GalNAc and the NH group of the attached Ser (or Cys), and the second between the hydroxymethyl group and the side-chain of Asp3. This geometry of the glycosidic linkage has not been previously observed for protein-bound Tn-containing peptides.

These differences in the presentation of the Tn antigen (GalNAc–Thr *versus* GalNAc–Ser) when bound to antibody SM3 could explain the variations in the affinity values. The improvement in the affinity between the glycosylated and unglycosylated derivatives with Thr is explained by understanding the interactions that the GalNAc motif forms with the protein.

Very recently, Boons *et al.*[51] reported the X-ray structure of a MUC1 peptide (A^1PDTRPAP^8) and the corresponding Tn-glycopeptide bound to antibody AR20.5, which is an anti-MUC1 monoclonal antibody produced by immunization with MUC1 from an ovarian cancer patient. This antibody is currently under investigation for its therapeutic potential and has undergone a successful phase I clinical trial with no observed toxicity[51] (Fig. 20A). The epitope recognised by AR20.5 comprises the six-amino acid region DTRPAP and the antibody binds to the glycopeptide with a 20-fold increase in affinity relative to the naked peptide.

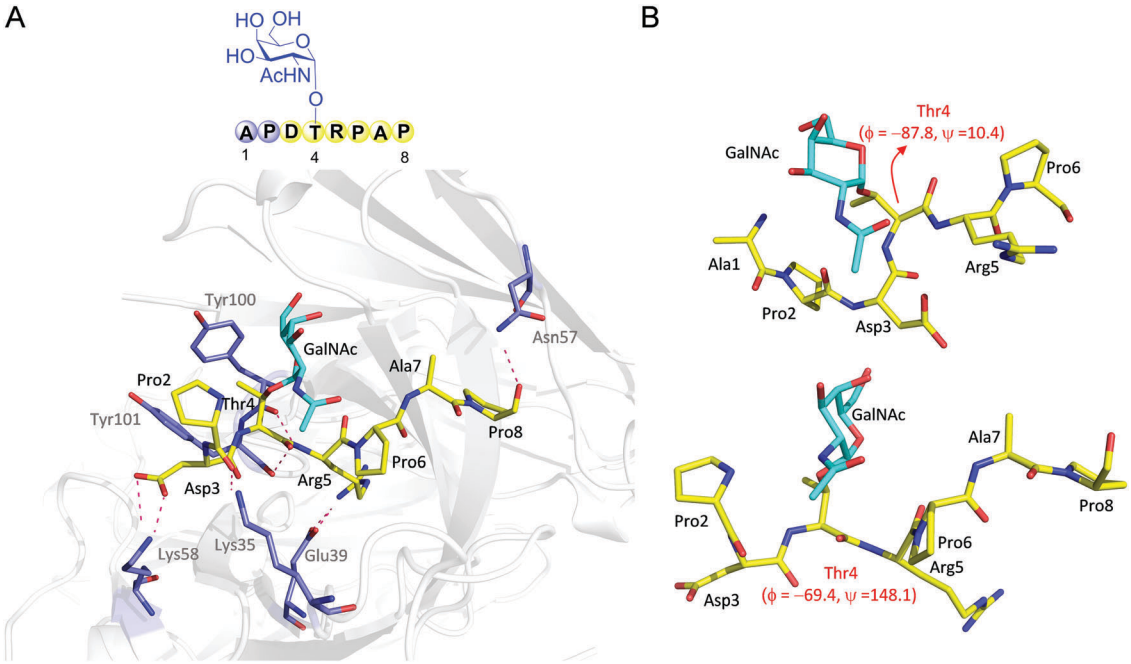


Fig. 20 (A) X-ray structure of antibody AR20.5 in a complex with a MUC1-derived Tn-glycopeptide (PDB entry: 5T78).[51] The antigen is shown as sticks with carbon atoms in yellow. The GalNAc moiety is in blue. The antibody is shown as white cartoons, and those residues that interact with the antigen are shown as sticks with carbon atoms in purple. Red dashed lines indicate antigen–antibody hydrogen bonds. (B) Conformations of MUC1 Tn-glycopeptide bound to antibody SM3 (top panel)[24] and antibody AR20.5 (bottom panel).[51] Glycosylation does not influence the conformation of the MUC1 epitope for either antibody (bottom and top panel). The conformation of the antigens is not identical for the two antibodies, but both antigens display an extended conformation with no evidence of a turn.

The X-ray structure of the unglycosylated peptide shows that both Asp3 and Arg5 interact with the antibody through salt-bridges. In addition, Arg5 is engaged in a hydrogen bond with TyrH100, and two other hydrogen bonds are formed for Thr4 and Pro8 with TyrH101 and AsnH57, respectively.

The peptide backbone of the glycopeptide presents an identical structure and the same interactions with the antibody as the naked derivative with the exception of an extra hydrogen bond established by Pro8. Unexpectedly, the carbohydrate faces away from the binding site towards the solvent and does not form any specific hydrogen bonds. In fact, only a van der Waals contact between the GalNAc and Tyr100H residue is observed. This interaction may contribute to an enhance binding affinity with the carbohydrate, however, it is unlikely sufficient to explain the 20-fold increase in affinity relative to the non-glycosylated peptide. Therefore, the results suggest that glycosylation alters the conformational equilibrium of the antigen, which allows the antibody to select the correct conformation.

3.3. Key elements involved in molecular recognition of the antigen–antibody state

Given the aforementioned results, it is clear that antibody B27.29 behaves in a different way to SM3 and AR20.5. Thus, although B27.29 interacts with two different parts of the MUC1 glycopeptide, SM3 and AR20.5 recognise only the PDTRP region. However, the PDTRP fragment is not recognised in a single conformation in the bound state. In fact, although a type I β-turn appears to be present in the complex with B27.29,[27]

Table 2 ϕ/ψ dihedral angles of the APDTRPAP region of MUC1 as determined by X-ray crystallography and NMR spectroscopy

	Peptide[a]			Glycopeptide[a]		
Res.	AR20.5[51]	SM3[23]	SM3[24]	AR20.5[51]	SM3[24]	Free antigen[26]
	ϕ, ψ	ϕ, ψ	ϕ, ψ	ϕ, ψ	ϕ, ψ	ϕ, ψ
A		−92.3, 160.0				−66.3, 145.8
P		−80.5, −164.6	−81.4, −157.5		−78.9, −157.4	−64.3, 142.2
D		−78.5, 100.6	−76.0, 102.7	−100.1, 26.4	−77.5, 885	−81.8, 148.8
T/Tn	**−76.1, 154.7**	**−97.2, 16.0**	**−94.7, 8.7**	**−69.4, 148.1**	**−87.8, 10.4**	**−92.3, 115.4**
R	−63.2, 132.4	−70.7, 143.4	−65.4, 129.1	−66.8, 128.7	−63.1, 131.9	−76.1, 145.9
P	−69.0, 151.2	−65.0, 134.8		−67.5, 153.8		−68.0, 152.5
A	−69.2, 153.5	−65.7, 153.1		−69.0, 149.6		
P						

[a] Values of the ϕ/ψ torsion angles are given in degrees.

SM3 and AR20.5 recognize an extended structure of the backbone.[23,24,51] However, a thorough examination of the torsional angles ϕ/ψ of each residue showed that antibody AR20.5 requires an extended-like PPII conformation for the Thr to properly fix the antigen into the binding site (Fig. 20B and Table 2). As already mentioned, this geometry is promoted by α-*O*-glycosylation in the free antigen. In sharp contrast, the Thr residue adopts a typical helix-like form in a complex with SM3 (Table 2). This significant variation could explain the higher improvement in affinity observed upon glycosylation with antibody AR20.5 relative to SM3.

With regard to the influence of the sugar, the three proposed mechanisms feasible depend on the antibody. Thus, although with antibodies B27.29 and anti-KL-6 the carbohydrate is required in the binding process and is part of the epitope, for antibody SM3, the sugar establishes specific interactions that favour binding. Conversely, in antibody AR20.5, the sugar moiety forces the peptide of the free antigen to adopt the bioactive conformation, which allows better shape complementarity that, in turn, promotes binding. Moreover, in AR20.5, salt bridges are key to molecular recognition, whereas, in SM3, hydrogen bonds and CH–π interactions play a crucial role. In the context of rational design, chemical modifications that improve these stabilizing contacts will provoke an enhancement in affinity, which will lead to more efficient antigens.

4. Conclusions and prospects

To date, several cancer vaccines based on partially glycosylated mucins have been developed and some of them are in clinical trials. One of the main obstacles with these vaccines is that cancer cells can generate immune escape mechanisms, which results in increased tolerance to tumour-associated antigens by the immune system. Although the design of synthetic antigens is a widespread strategy to overcome these issues, such chemical modifications could induce structural changes and alter the presentation of antigens to the immune system and, consequently, weaken the immune response.

To advance the design of structure-based vaccines, it is crucial that we clearly understand the conformational behaviour of the tumour-associated mucins in both the free and bound states by identifying those elements that play a pivotal role in their complex conformational equilibrium. This review highlights the most significant achievements reached to date in this field.

It is well known that the immunogenic PDTR epitope must be glycosylated with simple carbohydrates to elicit a robust immune response. This glycosylation favours an extended conformation of the peptide fragment, which likely favours its presentation to its biological targets. Contrary to what was initially thought, distant sugars at a given glycosylation site and glycosylation in remote regions can alter the presentation of the antigen and consequently its effectiveness to provoke a significant response.

The carbohydrate moiety can be directly involved in molecular recognition by the immune system. Indeed, in the context of antigen–antibody interactions, X-ray structures and some NMR studies indicate that truncated glycans can be part of the epitope or can contribute to the stability of the complex by forming precise contacts with the surface of the antibodies. These results help us to understand and rationalize how glycosylation can modify the structure of an antigen and the effects that those modifications can have on binding affinity. Moreover, studies in the bound state can contribute significantly to the development of a generation of high-affinity therapeutic antibodies.

The outcomes described in this review serve as a starting point towards innovative structure-based design of MUC1 mimics that can be part of a new generation of cancer vaccines.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Universidad de La Rioja and the Ministerio de Economía y Competitividad (project CTQ2015-67727-R). They also thank all present and past QUIBI group members for their contribution to the development of part of the work described herein. N. M.-S. thanks to Víctor J. Somovilla for his help and patience. F. C. thanks Alberto Corzana Ortega for inspiration.

Notes and references

1 M. A. Hollingsworth and B. J. Swanson, *Nat. Rev. Cancer*, 2004, **4**, 45–60.
2 D. W. Kufe, *Nat. Rev. Cancer*, 2009, **9**, 874–885.
3 S. Senapati, S. Das and S. K. Batra, *Trends Biochem. Sci.*, 2010, **35**, 236–245.
4 S. Nath and P. Mukherjee, *Trends Mol. Med.*, 2014, **20**, 332–342.
5 S. Gendler, J. Taylor-Papadimitriou, T. Duhig, J. Rothbard and J. Burchell, *J. Biol. Chem.*, 1988, **263**, 12820–12823.
6 U. Karsten, N. Serttas, H. Paulsen, A. Danielczyk and S. Goletz, *Glycobiology*, 2004, **14**, 681–692.
7 R. M. Wilson and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2013, **135**, 14462–14472.
8 N. Gaidzik, U. Westerlind and H. Kunz, *Chem. Soc. Rev.*, 2013, **42**, 4421–4442.
9 M. A. Wolfert and G.-J. Boons, *Nat. Chem. Biol.*, 2013, **9**, 776–784.
10 G. Ragupathi, J. R. Gardner, P. O. Livingston and D. Y. Gin, *Expert Rev. Vaccines*, 2011, **10**, 463–470.
11 B. Palitzsch, N. Gaidzik, N. Stergiou, S. Stahn, S. Hartmann, B. Gerlitzki, N. Teusch, P. Flemming, E. Schmitt and H. Kunz, *Angew. Chem., Int. Ed.*, 2016, **128**, 2944–2949.
12 V. Lakshminarayanan, N. T. Supekar, J. Wei, D. B. McCurry, A. C. Dueck, H. E. Kosiorek, P. P. Trivedi, J. M. Bradley, C. S. Madsen, L. B. Pathangey, D. B. Hoelzinger, M. A. Wolfert,

G.-J. Boons, P. A. Cohen and S. J. Gendler, *PLoS One*, 2016, **11**, e0145920.
13 G. L. Beatty and W. L. Gladney, *Clin. Cancer Res.*, 2015, **21**, 687–692.
14 N. Martínez-Sáez, N. T. Supekar, M. A. Wolfert, I. A. Bermejo, R. Hurtado-Guerrero, J. L. Asensio, J. Jiménez-Barbero, J. H. Busto, A. Avenoza, G.-J. Boons, J. M. Peregrina and F. Corzana, *Chem. Sci.*, 2016, **7**, 2294–2301.
15 F. Corzana, J. H. Busto, G. Jiménez-Oses, M. García de Luis, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina and A. Avenoza, *J. Am. Chem. Soc.*, 2007, **129**, 9458–9467.
16 Y. Inai, S. Kurashima, T. Hirabayashi and K. Yokota, *Biopolymers*, 2000, **53**, 484–496.
17 M. J. Scanlon, S. D. Morley, D. E. Jackson, M. R. Price and S. J. Tendler, *Biochem. J.*, 1992, **284**, 137–144.
18 S. Dziadek, C. Griesinger, H. Kunz and U. M. Reinscheid, *Chem. – Eur. J.*, 2006, **12**, 4981–4993.
19 L. Kirnarsky, O. Prakash, S. M. Vogen, M. Nomoto, M. A. Hollingsworth and S. Sherman, *Biochemistry*, 2000, **39**, 12076–12082.
20 X. Liu, J. Sejbal, G. Kotovych, R. R. Koganty, M. A. Reddish, L. Jackson, S. S. Gandhi, A. J. Mendonca and B. M. Longenecker, *Glycoconjugate J.*, 1995, **12**, 607–617.
21 J. D. Fontenot, N. Tjandra, D. Bu, C. Ho, R. C. Montelaro and O. J. Finn, *Cancer Res.*, 1993, **53**, 5386–5394.
22 J. D. Fontenot, S. V. Mariappan, P. Catasti, N. Domenech, O. J. Finn and G. Gupta, *J. Biomol. Struct. Dyn.*, 1995, **13**, 245–260.
23 P. Dokurno, P. A. Bates, H. A. Band, L. M. Stewart, J. M. Lally, J. M. Burchell, J. Taylor-Papadimitriou, D. Snary, M. J. Sternberg and P. S. Freemont, *J. Mol. Biol.*, 1998, **284**, 713–728.
24 N. Martínez-Sáez, J. Castro-López, J. Valero-Gónzalez, D. Madariaga, I. Compañón, V. J. Somovilla, M. Salvadó, J. L. Asensio, J. Jiménez-Barbero, A. Avenoza, J. H. Busto, G. J. L. Bernardes, J. M. Peregrina, R. Hurtado-Guerrero and F. Corzana, *Angew. Chem., Int. Ed.*, 2015, **54**, 9830–9834.
25 T. Matsushita, N. Ohyabu, N. Fujitani, K. Naruchi, H. Shimizu, H. Hinou and S.-I. Nishimura, *Biochemistry*, 2013, **52**, 402–414.
26 J. Schuman, A. P. Campbell, R. R. Koganty and B. M. Longenecker, *J. Pept. Res.*, 2003, **61**, 91–108.
27 J. S. Grinstead, R. R. Koganty, M. J. Krantz, B. M. Longenecker and A. P. Campbell, *Biochemistry*, 2002, **41**, 9946–9961.
28 L. Kinarsky, G. Suryanarayanan, O. Prakash, H. Paulsen, H. Clausen, F.-G. Hanisch, M. A. Hollingsworth and S. Sherman, *Glycobiology*, 2003, **13**, 929–939.
29 A. Borgert, J. Heimburg-Molinaro, X. Song, Y. Lasanajak, T. Ju, M. Liu, P. Thompson, G. Ragupathi, G. Barany, D. F. Smith, R. D. Cummings and D. Live, *ACS Chem. Biol.*, 2012, **7**, 1031–1039.
30 F. Corzana, J. H. Busto, M. García de Luis, J. Jiménez-Barbero, A. Avenoza and J. M. Peregrina, *Chem. – Eur. J.*, 2009, **15**, 3863–3874.
31 J. Schuman, D. Qiu, R. R. Koganty, B. M. Longenecker and A. P. Campbell, *Glycoconjugate J.*, 2000, **17**, 835–848.
32 D. H. Live, L. J. Williams, S. D. Kuduk, J. B. Schwarz, P. W. Glunz, X. T. Chen, D. Sames, R. A. Kumar and S. J. Danishefsky, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3489–3493.
33 D. M. Coltart, A. K. Royyuru, L. J. Williams, P. W. Glunz, D. Sames, S. D. Kuduk, J. B. Schwarz, X.-T. Chen, S. J. Danishefsky and D. H. Live, *J. Am. Chem. Soc.*, 2002, **124**, 9833–9844.
34 R. Hashimoto, N. Fujitani, Y. Takegawa, M. Kurogochi, T. Matsushita, K. Naruchi, N. Ohyabu, H. Hinou, X. D. Gao, N. Manri, H. Satake, A. Kaneko, T. Sakamoto and S.-I. Nishimura, *Chem. – Eur. J.*, 2011, **17**, 2393–2404.
35 Y. Tachibana, G. L. Fletcher, N. Fujitani, S. Tsuda, K. Monde and S.-I. Nishimura, *Angew. Chem., Int. Ed.*, 2004, **43**, 856–862.
36 F. Corzana, J. H. Busto, M. García de Luis, A. Fernández-Tejada, F. Rodríguez, J. Jiménez-Barbero, A. Avenoza and J. M. Peregrina, *Eur. J. Org. Chem.*, 2010, 3525–3532.
37 D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, H. Coelho, J. Valero-Gónzalez, J. Castro-López, J. L. Asensio, J. Jiménez-Barbero, J. H. Busto, A. Avenoza, F. Marcelo, R. Hurtado-Guerrero, F. Corzana and J. M. Peregrina, *ACS Chem. Biol.*, 2015, **10**, 747–756.
38 D. Madariaga, N. Martínez-Sáez, V. J. Somovilla, L. García-García, M. Á. Berbis, J. Valero-Gónzalez, S. Martín-Santamaría, R. Hurtado-Guerrero, J. L. Asensio, J. Jiménez-Barbero, A. Avenoza, J. H. Busto, F. Corzana and J. M. Peregrina, *Chem. – Eur. J.*, 2014, **20**, 12616–12627.
39 N. Ohyabu, H. Hinou, T. Matsushita, R. Izumi, H. Shimizu, K. Kawamoto, Y. Numata, H. Togame, H. Takemoto, H. Kondo and S.-I. Nishimura, *J. Am. Chem. Soc.*, 2009, **131**, 17102–17109.
40 S. Rangappa, G. Artigas, R. Miyoshi, Y. Yokoi, S. Hayakawa, F. Garcia-Martin, H. Hinou and S.-I. Nishimura, *MedChemComm*, 2016, **7**, 1102–1122.
41 F. Corzana, J. H. Busto, F. Marcelo, M. García de Luis, J. L. Asensio, S. Martín-Santamaría, J. Jiménez-Barbero, A. Avenoza and J. M. Peregrina, *Chem. – Eur. J.*, 2011, **17**, 3105–3110.
42 F. Corzana, J. H. Busto, F. Marcelo, M. G. de Luis, J. L. Asensio, S. Martín-Santamaría, Y. Sáenz, C. Torres, J. Jiménez-Barbero, A. Avenoza and J. M. Peregrina, *Chem. Commun.*, 2011, **47**, 5319–5321.
43 V. Rojas-Ocáriz, I. Compañón, C. Aydillo, J. Castro-López, J. Jiménez-Barbero, R. Hurtado-Guerrero, A. Avenoza, M. M. Zurbano, J. M. Peregrina, J. H. Busto and F. Corzana, *J. Org. Chem.*, 2016, **81**, 5929–5941.
44 M. A. Brister, A. K. Pandey, A. A. Bielska and N. J. Zondlo, *J. Am. Chem. Soc.*, 2014, **136**, 3803–3816.
45 D. Mazal, R. Lo-Man, S. Bay, O. Pritsch, E. Dériaud, C. Ganneau, A. Medeiros, L. Ubillos, G. Obal, N. Berois, M. Bollati-Fogolin, C. Leclerc and E. Osinaga, *Cancer Immunol. Immunother.*, 2013, **62**, 1107–1122.
46 Y. Zhang, Q. Li, L. G. Rodriguez and J. C. Gildersleeve, *J. Am. Chem. Soc.*, 2010, **132**, 9653–9662.
47 H. Coelho, T. Matsushita, G. Artigas, H. Hinou, F. J. Cañada, R. Lo-Man, C. Leclerc, E. J. Cabrita, J. Jiménez-Barbero,

S.-I. Nishimura, F. Garcia-Martin and F. Marcelo, *J. Am. Chem. Soc.*, 2015, **137**, 12438–12441.

48 F. Corzana, J. H. Busto, G. Jiménez-Oses, J. L. Asensio, J. Jiménez-Barbero, J. M. Peregrina and A. Avenoza, *J. Am. Chem. Soc.*, 2006, **128**, 14640–14648.

49 S. S. Mallajosyula and A. D. MacKerell, *J. Phys. Chem. B*, 2011, **115**, 11215–11229.

50 C. L. Brooks, A. Schietinger, S. N. Borisova, P. Kufer, M. Okon, T. Hirama, C. R. Mackenzie, L.-X. Wang, H. Schreiber and S. V. Evans, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10056–10061.

51 M. Movahedin, T. M. Brooks, N. T. Supekar, N. Gokanapudi, G.-J. Boons and C. L. Brooks, *Glycobiology*, 2017, **27**, 677–687.

52 B. Meyer and T. Peters, *Angew. Chem., Int. Ed.*, 2003, **42**, 864–890.

53 H. Möller, N. Serttas, H. Paulsen, J. M. Burchell, J. Taylor-Papadimitriou and B. Meyer, *Eur. J. Biochem.*, 2002, **269**, 1444–1455.

54 C. Her, W. M. Westler and T. Yang, *JSM Chem.*, 2013, **1**, 1004.

55 J. S. Grinstead, J. T. Schuman and A. P. Campbell, *Biochemistry*, 2003, **42**, 14293–14305.