JOC The Journal of Organic Chemistry

Subscriber access provided by NEW YORK UNIV

Article

Design of alpha-S-neoglycopeptides derived from MUC1 with a flexible and solvent exposed sugar moiety

Víctor Rojas-Ocáriz, Ismael Compañón, Carlos Aydillo, Jorge Castro-López, Jesús Jimenez-Barbero, Ramon Hurtado-Guerrero, Alberto Avenoza, María M Zurbano, Jesús M. Peregrina, Jesus H Busto, and Francisco Corzana

J. Org. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.joc.6b00833 • Publication Date (Web): 15 Jun 2016

Downloaded from http://pubs.acs.org on June 18, 2016

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

The Journal of Organic Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

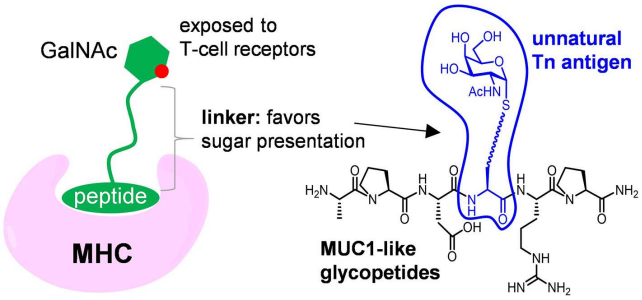
Design of α-*S*-neoglycopeptides derived from MUC1 with a flexible and solvent exposed sugar moiety

Víctor Rojas-Ocáriz,[‡,1] Ismael Compañón,[‡,1] Carlos Aydillo,[#,1] Jorge Castro-Lopez,[2] Jesús Jiménez-Barbero,[3] Ramón Hurtado-Guerrero,[2] Alberto Avenoza,[1] María M. Zurbano,[1] Jesús M. Peregrina,[*,1] Jesús H. Busto,[*,1] and Francisco Corzana[*,1]

[1]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química. Madre de Dios 53, 26006 Logroño, Spain. [2](i) BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D, Zaragoza, Spain. (ii) Fundación ARAID, 50018, Zaragoza, Spain. [3](i) Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia Building 801 A, 48160 Derio (Spain). (ii) IKERBASQUE, Basque Foundation for Science 48011 Bilbao (Spain). (iii) Department of Chemical and Physical Biology Centro de Investigaciones Biológicas, CSIC Ramiro de Maeztu 9, 28040 Madrid (Spain).

[#]Current address: CECB, Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark.

TOC



Abstract

The use of vaccines based on MUC1 glycopeptides is a promising approach to treat cancer. We present herein several sulfa-Tn antigens incorporated in MUC1 sequences that possess a variable linker between the carbohydrate (GalNAc) and the peptide backbone. The main conformations of these molecules in solution have been evaluated by combining NMR experiments and molecular dynamics simulations. The linker plays a key role in the modulation of the conformation of these compounds at different levels, blocking a direct contact between the sugar moiety and the backbone, promoting a helix-like conformation for the glycosylated residue and favoring the proper presentation of the sugar unit for molecular recognition events. The feasibility of these novel compounds as mimics of MUC1 antigens has been validated by X-ray diffraction structure of one of these unnatural derivatives complexed to an anti-MUC1 monoclonal antibody. These features, together with potential lack of immune suppression, render these unnatural glycopeptides promising candidates for designing alternative therapeutic vaccines against cancer.

Introduction

Chemistry-based approaches are of paramount importance to battle disease. In the quest of chemical weapons against cancer, MUC1 is a glycoprotein over-expressed in most of tumors.[1-4] While in healthy cells, the MUC1 backbone presents complex oligosaccharides, in tumor cells the peptide backbone is decorated with simple and truncated carbohydrates. As a consequence, different tumor-associated carbohydrate antigens (TACAs), such as the Tn determinant (α-*O*-GalNAc-Ser/Thr),[5] are exposed to the immune system and can be recognized by different antibodies.[6] For this reason, MUC1 derivatives are attracting great interest as a potential tool in developing therapeutic vaccines for the treatment of cancer.[4,7,8] However, to date, none of these vaccines have succeeded in clinical trials.[9] On the one hand, natural TACAs are tolerated by the immune system.[10] To overcome this problem, chemical modifications of these antigens, generating non-natural determinants have been proposed.[7,11-20] On the other hand, most glycan antigens are B-cell epitopes and produce only short-

lived and low-affinity antibodies.[4] Induction of a robust immune response that produces high affinity IgG antibodies is essential and may be achieved through initiation of a T-cell-dependent pathway.

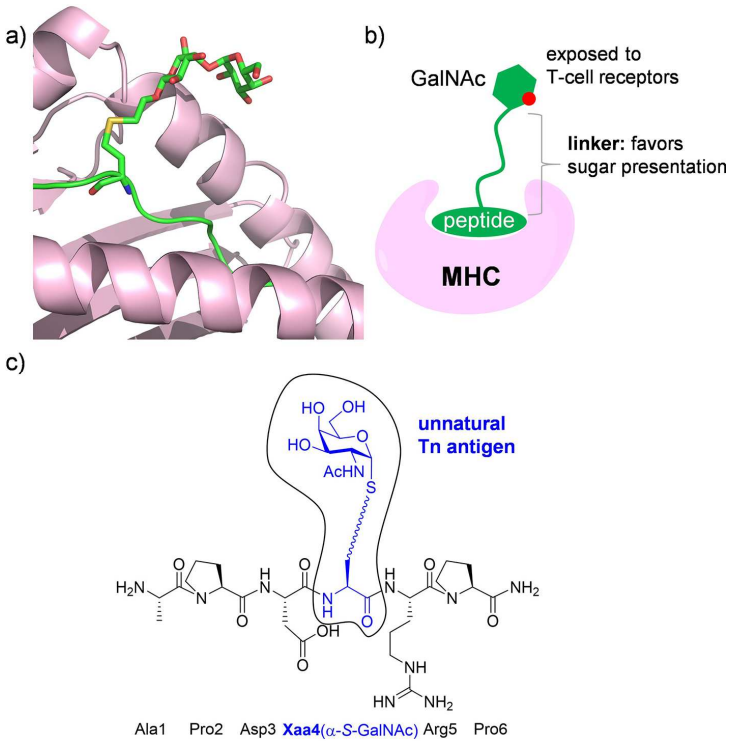


Figure 1. (a) Crystal structure of a MHC-I molecule in complex with a glycopeptide containing a linker between the peptide backbone and the sugar moiety (PDB id: 1KBG, ref. 21e). (b) Design of novel glycopeptides containing a linker that favors the presentation of the sugar moiety and facilitates the interaction with T-cell receptors. (d) General structure of the neoglycopeptides synthesized and studied in this work. The peptide sequence PDTRP constitutes the most immunogenic domain of mucin MUC1.[6]

A key step in this process is the presentation of the antigen through the major histocompatibility (MHC) molecule to T-cell receptors. In that context, recent studies indicate that glycopeptides can arbitrate classical MHC-mediated immune response.[21] The crystal structure of an MHC-I molecule

with a glycopeptide containing a linker between the peptide and the sugar[21e] showed that while the aglycone part of the antigen binds to MHC molecule, the carbohydrate moiety can facilitate the recognition of T cell receptors (TCR) and therefore can stimulate immune response (Figures 1a and 1b).

In an effort to combine these two concepts, we present herein a set of unnatural Tn antigens characterized by the incorporation of a flexible and variable linker in their structures, placed between a peptide fragment and the GalNAc moiety (Figure 1c). These new sulfa-Tn derivatives have been prepared as Fmoc-protected glycosyl amino acids, ready to use in solid phase peptide synthesis (SPPS). They have been included in the Ala-Pro-Asp-Xaa(α-*S*-GalNAc)-Arg-Pro peptide sequence, which constitutes the key immunogenic epitope of MUC1[6] when Xaa is a threonine. In particular, in this work, we have synthesized and studied six neoglycopeptides, being Xaa a cysteine, homocysteine, homohomocysteine, *O*-(mercaptopropyl)serine, *S*-(mercaptopropyl)cysteine, or *O*-(mercaptopropyl)threonine residue. The main conformations of these molecules in solution have been evaluated. The linker plays a key role in the conformational behavior of these molecules, from the global 3D structure to the presentation of the glycan part of these compounds. Indeed, the linker blocks the direct contact between the sugar moiety and the backbone, promotes a helix-like conformation for the glycosylated residue, provides extra flexibility to the molecule and favors the presentation of the sugar unit for molecular recognition events. These novel compounds are indeed mimics of MUC1 antigens as demonstrated by the analysis of the X-ray diffraction data obtained for one of these unnatural glycopeptides complexed to an anti-MUC1 monoclonal antibody. The global behavior of this complex has also been analyzed by MD simulations.

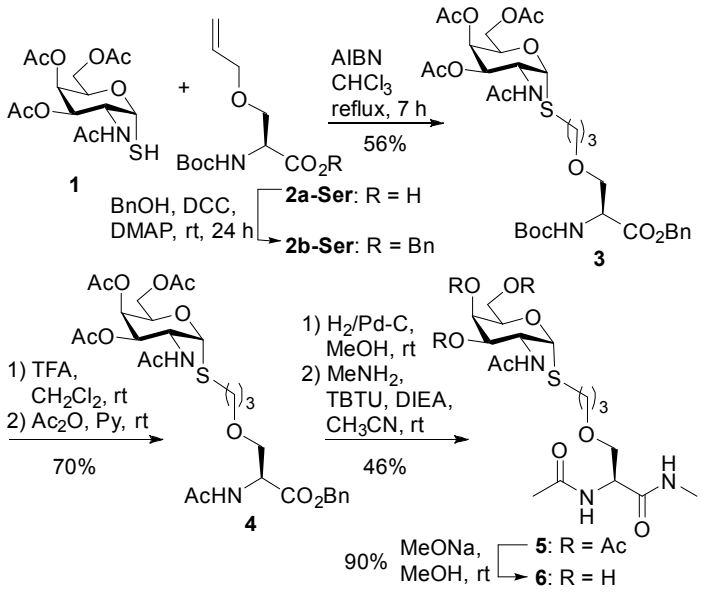
Results and Discussion

From a synthetic perspective, the existence of *S*-glycopeptides with important biological properties has stimulated the development of a variety of glycosylation methods. These methods generally use

protected carbohydrates as electrophiles and Cys derivatives as nucleophiles. Alternatively, dehydroalanine (Dha)[22-24] containing peptides and nucleophilic thiocarbohydrate derivatives[25] have been used in 1,4-conjugate additions with poor results in diastereoselectivity.[26] Recently, we have reported a new chiral Dha derivative,[27] which has been used as acceptor in stereoselective sulfa-Michael additions of appropriately protected thiocarbohydrates. This procedure has been applied to the synthesis of *S*-(2-acetamido-2-deoxy-α-D-galactopyranosyl)-L-cysteine, which could be considered as a mimic of the Tn antigen.[28] A method based on the alkene hydrothiolation reaction,[29] which is called thiol–ene coupling (TEC), has also been developed. The thermally or photochemically induced TEC versions occur through a radical mechanism in anti-Markovnikov regioselective fashion.[29] These methods have been applied, in general, to the synthesis of β-*S*-glycosyl amino acids and β-*S*-glycopeptides.[29] Nevertheless, the α-S-glycosidic bond formation in these motifs,[29c] especially the α-*S*-GalNAc moiety of the Tn antigen,[30,31] has received little attention. In that context, the synthesis of α-thio-linked glycosyl amino acids and glycopeptides by an anomeric thiol group *S*-alkylation with β-iodo(or bromo)alanine has been described.[30a] Later, this methodology was modified to a "two-step one-pot" reaction, generating α-thio-linked glycosyl amino acids by condensation of a β-bromoalanine derivative with the corresponding thiolate salt.[31]

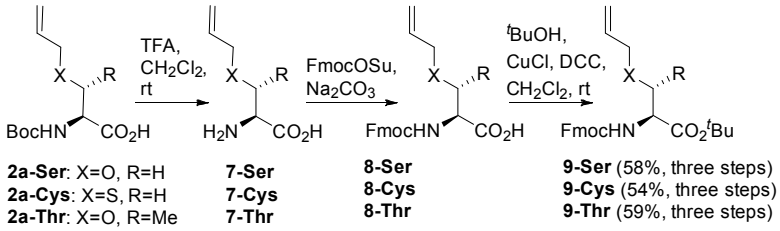
We report herein new stereocontrolled entries to adequately protected *S*-glycosyl amino acids, mimics of the Tn antigen, as building blocks for solid phase peptide synthesis of *S*-linked glycopeptides, exploiting the reactivity of the sulfhydryl group of tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-α-D-galactopyranose (**1**, abbreviated as GalNAc-α-SH) in two types of reactions. The first one comprises the hydrothiolation reaction of double bonds included in the side chains of amino acid derivatives. This reaction has been extensively used in glycochemistry.[32] Moreover, the mild reaction conditions, atom economy, and regioselectivity of this process satisfy essential requirements of the click concept.[33] The second type of reaction assayed involves the anomeric thiol group *S*-alkylation with bromo-amino acid derivatives.

To know the viability of this project, we started with the synthesis of the shortest glycopeptide (*S*-glycosyl amino acid ended in amide groups) that includes the largest distance between carbohydrate and peptide backbone moieties. We selected GalNAc-α-SH (**1**) as a glycosyl donor, whose synthesis was made following the procedure described in the literature[34a] but properly modified to carry out in a gram scale (Supporting Information). GalNAc-α-SH (**1**) was subjected to an alkene hydrothiolation reaction with *O*-allyl serine derivative **2b-Ser**, which was prepared according to the procedure described[34b] from commercially available *N*-Boc-serine derivative **2a-Ser**. The hydrothiolation reaction was carried out using azobisisobutyronitrile (AIBN) as a radical initiator. The best conditions involved the use of 1.7 equivalents of thiol **1**, 1.2 equivalents of AIBN and 1.0 equivalent of alkene **2b-Ser** to afford compound **3** in a 56% yield (Scheme 1). This derivative was transformed into the required diamide **6**. Deprotection of Boc group, followed by acetylation of the amino group gave compound **4**. This compound was subjected to hydrogenolysis of benzyl ester and the corresponding carboxylic acid was transformed into methyl amide. Diamide derivative **5** was deacetylated to give target compound **6** (Scheme 1).



Scheme 1. Synthesis of *S*-glycosyl amino acid **6** as a diamide.

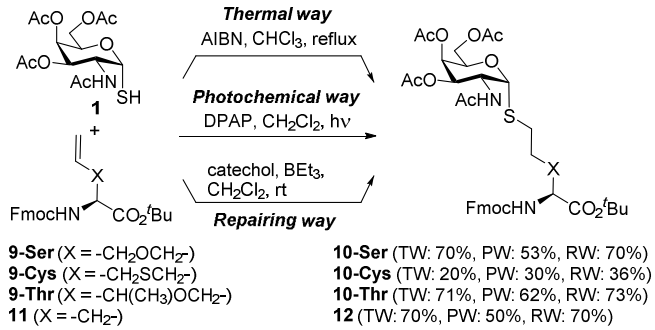
Encouraged by this result, we synthesized the corresponding Fmoc-protected *S*-glycosyl amino acids **10-Ser**/**Cys**/**Thr**, ready to use in SPPS and in a gram scale. To this aim, it was necessary to prepare the (*O* or *S*)-allyl *N*-Fmoc-serine/cysteine/threonine *tert*-butyl esters **9-Ser**/**Cys**/**Thr**, which had not been described to date. Their synthesis started from commercially available *N*-Boc-amino acids **2a-Ser**/**Cys**/**Thr**, whose Boc groups were removed, giving amino acids **7-Ser**/**Cys**/**Thr**. The amino groups of these compounds were then protected as Fmoc carbamates, affording compounds **8-Ser**/**Cys**/**Thr**. Esterifications of acids **8-Ser**/**Cys**/**Thr** with *tert*-butyl alcohol gave the required allylic derivatives **9-Ser**/**Cys**/**Thr** (Scheme 2).



Scheme 2. Synthesis of *N*-Fmoc-Ser/Cys/Thr(*O*- or *S*-allyl)-*O*tBu derivatives **9-Ser**/**Cys**/**Thr**.

As a next step, we assayed hydrothiolation reactions between glycosyl-thiol **1** and alkenes **9-Ser**/**Cys**/**Thr**, following the conditions above described for compound **2b-Ser**, obtaining good yields of the required building blocks **10-Ser** and **10-Thr** (70% and 71%, respectively). However, compound **10-Cys** was obtained only in a 20% yield (Scheme 3). To increase these yields, the photochemical version of the hydrothiolation reaction[29a,b] was tested because it is considered as a light-induced click reaction,[35] with applications to diverse fields of study of biomolecular systems. We used a photochemical reactor equipped with 16 UVA (350 nm) lamps of 8 W. The reaction was initiated with 2,2-dimethoxy-2-phenylacetophenone (DPAP) and after testing several conditions, the best yields involved the use of 1.0 equiv. of alkene, 1.2 equiv. of compound 1 and 0.4 equiv. of DPAP in dichloromethane (1 mL). In this way, building blocks **10-Ser**/**Cys**/**Thr** were obtained in a

53%, 30% and 62%, respectively (Scheme 3). It is well-known that this type of reactions involving *O*- or *S*-allyl derivatives works with moderate to low yields.[36] We also tested a recent method with **9-Ser**/**Cys**/**Thr**, involving the use of BEt_3 as an initiator and catechol as a co-reagent.[37] Although, only a slight improvement of the yield was achieved, this method provided an improved way to purify the desired compounds (Scheme 3).

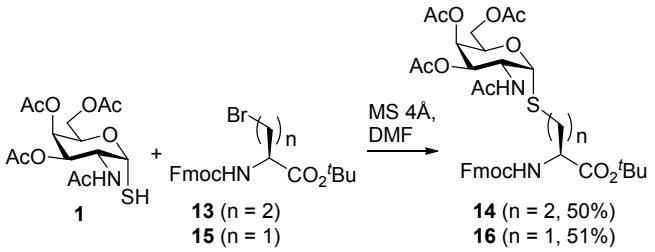


Scheme 3. Synthesis of *S*-glycosyl amino acid building blocks **10-Ser**/**Cys**/**Thr** and **12** via hydrothiolation. TW, PW and RW stand for *thermal way*, *photochemical way* and *repairing way*, respectively.

Compound **12**, which contains a three-methylene bridge between *S*-α-GalNac and amino acid moieties, was synthesized following the same hydrothiolation protocol, tested again by both thermal and photochemical ways, starting from α-allylglycine **11**. This protected compound was obtained according to the protocol previously published[38] and it was reacted with glycosyl donor **1**, giving a 70% and 50% yield of desired building block **12** in thermal and phochemical ways, respectively (Scheme 3). We also tested the repairing way using catechol/BEt_3 achieving compound **12** in a better yield than photochemical way (70%).

To access to GalNAc-*S*-glycosylated building blocks **14** and **16**, L-homocysteine (hCys) and L-cysteine (Cys) derivatives, respectively, we used a different methodology based on the anomeric thiol group *S*-alkylation with bromo-amino acid derivatives (Scheme 4). This methodology has been above-commented[30a,31] as well as their difficulties to carry out the nucleophilic substitution in a

basic medium in the presence of Fmoc protecting groups. We followed a recently published mild synthetic strategy[39] that simply requires activated molecular sieves to selectively promote *S*-alkylation. Therefore, activated molecular sieves (4Å) were employed as a base to promote *S*-alkylation of GalNAc-α-SH **1** with appropriate protected γ-bromohomoalanine[40] **13** and β-bromoalanine[30,41] **15** using dry DMF as a solvent (Scheme 4).

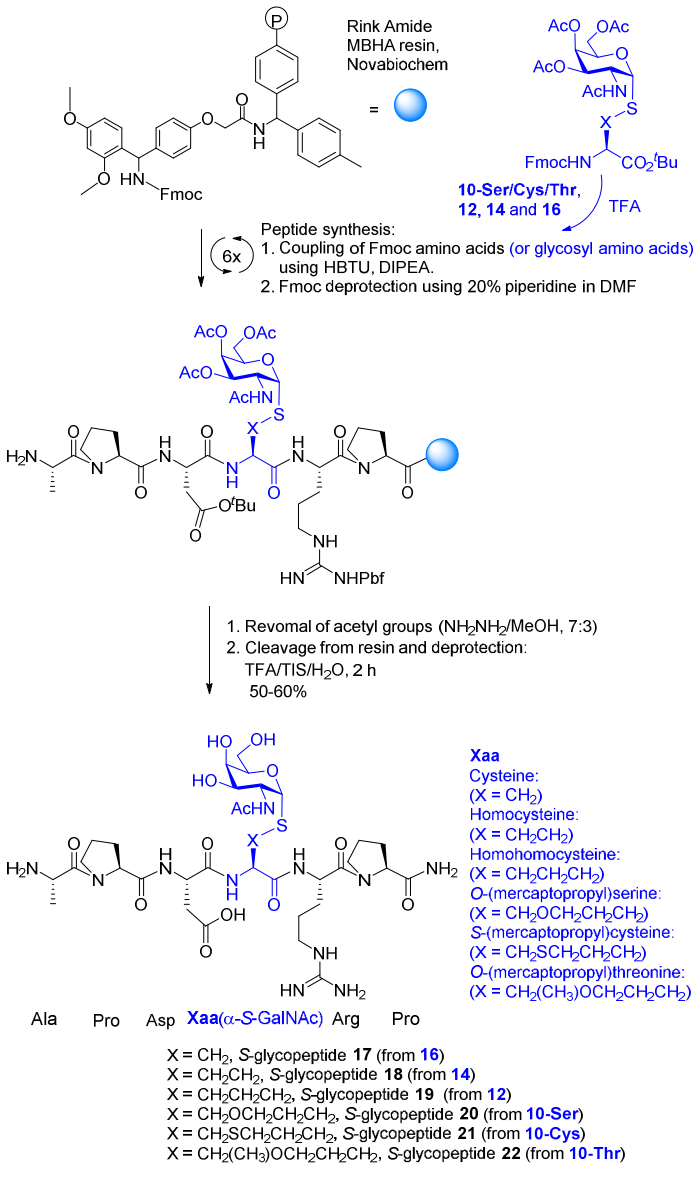


Scheme 4. Synthesis of *S*-glycosyl amino acid building blocks **14** and **16** via nucleophilic substitution.

Although the reactions proceeded with moderate yields, the procedure was easy to perform, fast, and provided similar overall yields when it was compared with other methods published recently.[28] These known bromo-derivatives **13** and **15** were obtained from L-homoserine and L-serine, respectively, following the procedure described in the literature, that involved *N*-Fmoc protection, *tert*-butyl ester formation and bromination under standard conditions.[30]

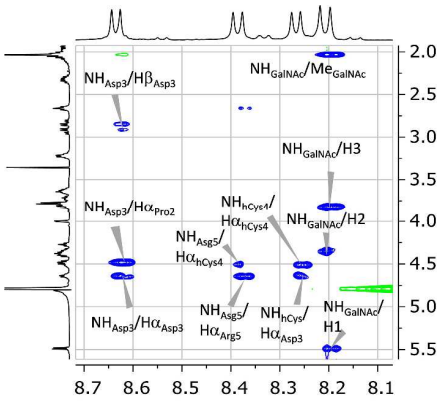
These six mimics of the Tn antigen (**10-Ser**/**Cys**/**Thr**, **12**, **14** and **16**), all of *S*-glycosyl amino acids whose amino, carboxylic acid and hydroxyl groups are protected as *N*-Fmoc, *tert*-butyl esters and acetates, respectively, are adequate building blocks for SPPS of glycopeptides. Consequently, they were used to synthesize different α-*S*-neoglycopeptides shown in Scheme 5. The synthesis of these compounds was performed using the automated SPPS protocol with a Rink amide MBHA resin and Fmoc and side chain protected amino acids. α-*S*-Glycosylamino acid building blocks **10-Ser**/**Cys**/**Thr**, **12**, **14** and **16** were manually coupled to the peptide sequence to improve the yield of

this step. After removal of the acetyl groups of the carbohydrate with a mixture of NH_2NH_2/MeOH, α-*S*-neoglycopeptides were released from the resin using trifluoroacetic acid (TFA). Under these conditions, all the acid-labile side-chain protecting groups were also removed. Purification by preparative HPLC and subsequent lyophilization gave the target α-*S*-neoglycopeptides **17**-**22** in good overall yields.



Scheme 5. α-*S*-neoglycopeptides synthesized in this work.

The conformational behavior of the synthesized neoglycopeptides in solution was then deduced by using NMR. In particular, NOE data, supported by molecular dynamics (MD) simulations with time-averaged restraints (MD-tar), were employed.[42,43] In a first step, full assignment of the protons in all derivatives was carried out using standard COSY and HSQC experiments. 2D-NOESY experiments in H_2O/D_2O (9/1) (25 °C, pH = 5.5) were then performed (Figure 2 and Supporting Information). Distances involving NH protons were semi-quantitatively deduced by integrating the volume of the corresponding cross-peaks. This experimental data was used as restraints in MD-tar simulations (Figure 2 and Supporting Information).



Distance	Exp.[a]	MD-tar[a]
$NH_{Asp3} - H\alpha_{Pro2}$	2.6	2.4
$NH_{Asp3} - H\alpha_{Asp3}$	3.0	2.9
$NH_{hCys} - H\alpha_{Asp3}$	3.0	2.6
$NH_{hCys} - H\alpha_{hCys}$	3.0	2.9
$NH_{Arg5} - H\alpha_{hhSer4}$	3.0	3.0
$NH_{Arg5} - H\alpha_{Arg5}$	3.1	2.9

[a] Distances are given in Å.

Figure 2. Section of the 500 ms 2D NOESY spectrum (400 MHz) in H_2O/D_2O (9:1) at 25 °C of derivative **18**, showing the amide region. Diagonal peaks and exchange cross-peaks connecting NH protons and water are negative (green color). The NOE contacts are represented as positive cross-peaks (blue color). Comparison of the experimental and MD-tar derived distances for this glycopeptide is also shown.

The structure ensembles obtained through the MD-tar simulations indicate that the neoglycopeptides are rather flexible in solution (Figure 3 and Supporting Information).

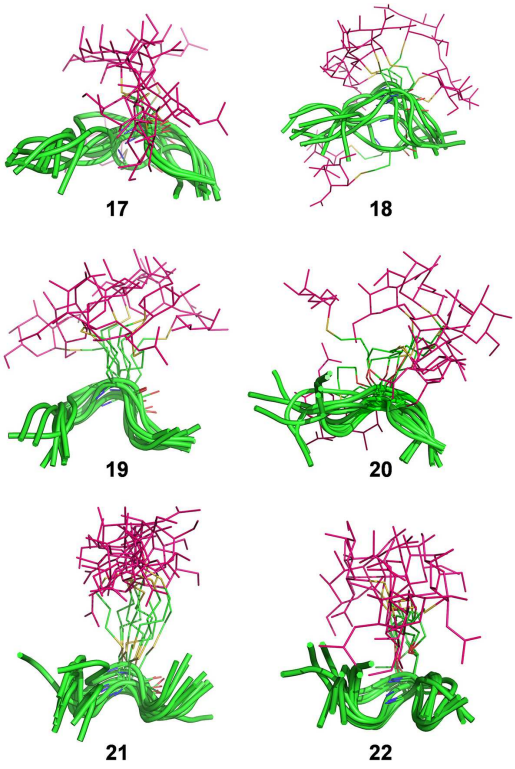


Figure 3. Ensembles obtained through 20 ns MD-tar simulations for neoglycopeptides **17**-**22**. The backbone is shown in green and the GalNAc in purple.

In general, the studied neoglycopeptides exhibit a random-coil conformation for the peptide backbone. This result was corroborated by circular dichroism (CD) spectroscopic studies (Supporting Information). The CD spectra of these derivatives differed from the one observed for the natural APDT(α-*O*-GalNAc)RP glycopeptide, which adopts mainly a PPII conformation for the peptide backbone in water[44] (Supporting Information). As a point of interest, the contacts between the sugar and the peptide fragment are almost insignificant. As a consequence, the glycosylated residue exhibits a helix-like conformation in all studied compounds (Figure 4a for compound **22** and Supporting Information for the other compounds). This feature is corroborated by the medium-

size NOE intensities observed for the NH_{Xaa4}–$H\alpha_{Xaa4}$, and NH_{Arg5}–$H\alpha_{Xaa4}$ cross-peaks, where Xaa is the glycosylated residue.[45] Conversely, it has been shown that GalNAc-glycosylation of the natural Ser or Thr residues forces these amino acids to display extended conformations[46-52] due to the presence of stabilizing contacts between the sugar and the peptide moieties.[46a,50]

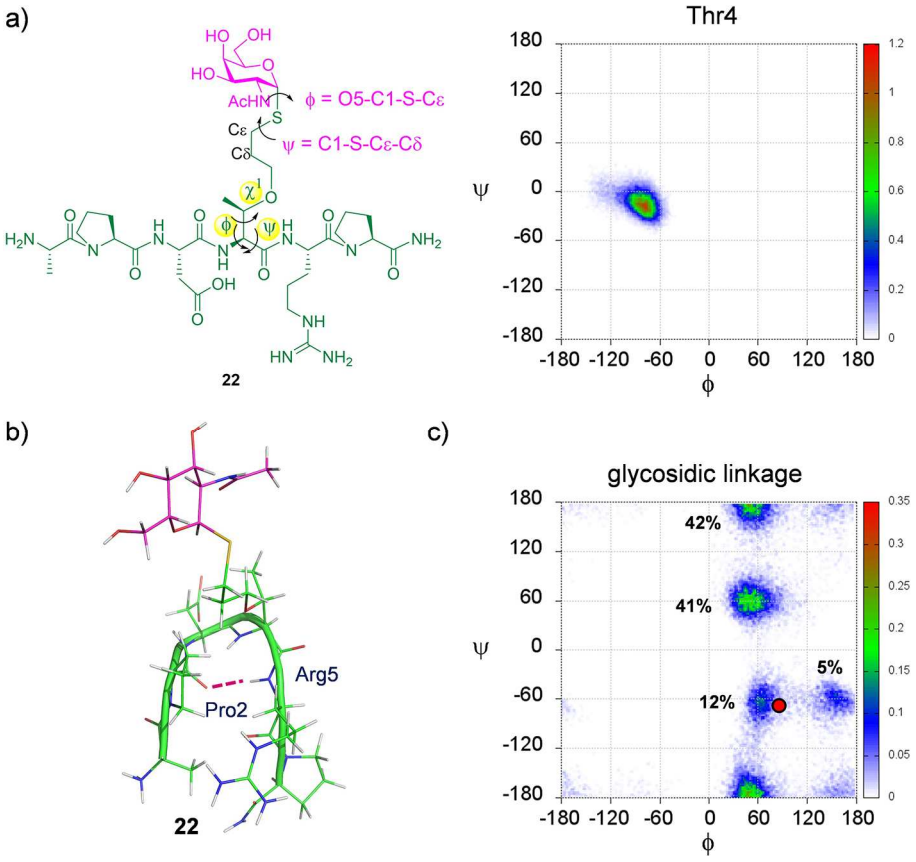


Figure 4. (a) Definition of the glycosidic linkage in this work, together with ϕ/ψ distribution obtained through 20 ns MD-tar simulations for amino acid Thr4 in compound **22**. (b) Type I β-turn conformer found in water for glycopeptides **19**-**22**. (c) ϕ/ψ distribution obtained through 20 ns MD-tar simulations for the glycosidic linkage in compound **22**. The conformation found in the crystal structure for the glycosidic linkage of APDC(GalNAc)RP in complex to SM3 antibody (PDB id: 5A2L, ref. 44) is represented as a red circle.

Of note, compounds **19**-**22**, which are bearing a longer linker, adopt a low population (<25%) of a type I β-turn-like, stabilized by a hydrogen bond between Pro2 (CO) and Arg5 (NH), which has been reported for natural MUC1-like derivatives[52] (Figure 4b).

Concerning the *S*-glycosidic linkage, as stated by us for other derivatives,[34] it is rather flexible in all compounds, with two main conformations, characterized by ψ *ca*. 180º or 60º, corresponding to the local minima calculated for methyl 4-thio-α-maltoside.[53] The studied glycopeptides display a conformation centered at $\phi/\psi \approx 60^{o}/-60^{o}$ degrees, also close to that observed in the crystal structure of a mucin-like glycopeptide incorporating GalNAc-Cys[44] (Figure 4c and Supporting Information).

Regarding the lateral chain, derivatives carrying a serine, homoserine or homohomoserine (compounds **18**, **19** and **20**, respectively) present a flexible dihedral angle χ^1, with two main values (Supporting Information). This result is similar to that found for other glycopeptides bearing a serine residue.[49] For the other derivatives, χ^1 is fixed around 60 degrees. Glycopeptides **20**, **21** and **22** present a similar behavior for the side chain, being derivative **22**, with a threonine amino acid, the one with the more rigid linker in solution (Supporting Information). The result is in agreement with our previous studies on Thr-derived glycopeptides.[50] In general, and according to our simulations, the linker of all glycopeptides is located almost perpendicularly to the peptide backbone (Figure 3 and Supporting Information).

The viability of these novel compounds as mimics of MUC1 antigen was validated by the analysis of the crystal structure of one of these unnatural glycopeptides, compound **22**, complexed to the anti-MUC1 monoclonal antibody SM3 produced in our lab.[44] The analysis of the X-ray structure (PDB id: 5fxc, Figure 5a and Supporting Information) revealed that the hydrogen bonds, as well as the hydrophobic contacts present in the natural glycopeptide APDT(GalNAc)RP bound to SM3,[44] are also conserved in the scFv-SM3:**22** complex. In detail, Pro2 stacks with Trp91L, Trp96L and Tyr32L. The side-chains of Asp3 and Arg5 are engaged in hydrophobic contacts with Trp33H and

Tyr32H, respectively. NH group of Ala1 and carbonyl group of Thr4 are involved in hydrogen bonds with Tyr32L and Gln97H, respectively. A hydrogen bond between side chain of Arg5 and carbonyl group of Asn31H was also observed.

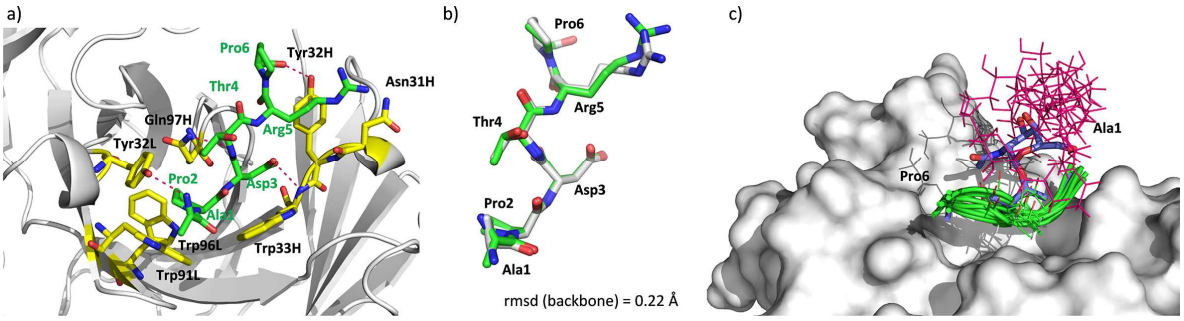


Figure 5. (a) Crystal structure of glycopeptide **22** in complex with scFv-SM3, showing the stabilizing interactions antigen-antibody. Carbon atoms of glycopeptide **22** are in green. Carbon atoms of key residues of SM3 are coloured in yellow. Interactions between the peptidic fragment and SM3 surface are in pink. (b) Superposition of the peptide backbone of glycopeptide APDT(GalNAc)RP and glycopeptide **22** in complex with scFv-SM3, showing the root-mean-square deviation (rmsd) value for the peptide backbone. (c) Ensembles obtained from the unrestrained 100 ns MD simulations performed on complex scFv-SM3:**22**. The antibody is shown as a grey surface and the peptide backbone of glycopeptide **22** is shown in green. The carbon atoms of GalNAc unit are in pink. The conformation of glycopeptide APDT(GalNAc)RP found in the X-ray structure is also shown in purple-blue (PDB id: 5A2K, ref. 44).

The conformation of the peptide fragment of **22** bound to the mAb is almost identical to those observed for natural glycopeptides in complex with this antibody, indicating the minor impact that the linker has on the peptide conformation in the bound state (Figure 5b). Unfortunately, the structure of the linker and the GalNAc, could not be solved due to the lack of electron density. This

is presumably due to the high degree of flexibility of the linker. Therefore, molecular dynamics (MD) simulations (100 ns) were performed. All interactions between glycopeptide **22** and the mAb observed in the X-ray structure are also observed though MD simulations. As deduced from these calculations, the linker is rather flexible and the GalNAc residue is exposed to the solvent (Figure 5c).

Conclusions

Several sulfa-Tn antigen derivatives with variable lengths between the glycan and the peptide backbone have been synthesized using two types of reactions with GalNAc$(OAc)_3$-α-SH. The first one comprised a thiol-ene coupling of alkenyl-α-amino acids and the second one involved an anomeric thiol group S-alkylation of bromo-α-amino acids. The presence of a linker between the peptide backbone and the sugar precludes the intramolecular interactions between these two moieties, favoring the presentation of the sugar for molecular recognition events. This property, together with the potential lack of immunosuppression suffered by the natural derivatives, makes the studied compounds promising candidates for the design of novel therapeutic vaccines to treat tumors.

Experimental Section

Reagents and general procedures. Commercial reagents were used without further purification. Analytical thin layer chromatography (TLC) was performed on *Macherey-Nagel* precoated aluminium sheets with a 0.20 mm thickness of silica gel 60 with fluorescent indicator UV_{254}. TLC plates were visualized with UV light and by staining with phosphomolybdic acid (PMA) solution (5 g of PMA in 100 mL of absolute ethanol) or sulfuric acid-ethanol solution. Column chromatography was performed on silicagel (230−400 mesh). ^{1}H and ^{13}C NMR spectra were measured with a 400 MHz spectrometer with TMS as the internal standard. Multiplicities are quoted as singlet (s), broad singlet (br s), doublet (d), doublet of doublets (dd), triplet (t), or multiplet (m). Spectra were assigned using COSY and HSQC experiments. All NMR chemical shifts (δ) were recorded in ppm and coupling constants (J) were reported in Hz. The results of these experiments were processed with MestreNova software. High resolution electrospray mass (ESI) spectra

were recorded on a microTOF spectrometer; accurate mass measurements were achieved by using sodium formate as an external reference.

General procedure for the thiol-ene coupling by thermal way. To a solution of alkene-amino acid derivative (1.0 equiv) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-α-D-galactopyranose **1** (1.7 equiv) in $CHCl_3$ (5 mL), α,α′-azoisobutyronitrile (AIBN or VAZO 64, Molekula 78-67-1) (0.6 equiv) was added and reaction mixture was then heated at reflux. After stirring for 3 h, another portion of AIBN (0.6 equiv) was added and the mixture was stirred for 4 h more. The reaction was concentrated and the residue was purified by a silica gel column chromatography to give the corresponding *S*-galactosaminyl amino acid derivative.

General procedure for the thiol-ene coupling by photochemical way. The reaction was carried out in a glass vial (diameter: 2 cm), sealed with a natural rubber septum, located 2.5 cm away from the UVA lamp apparatus (LUZCHEM - LZC-4V-, 16 UVA-lamps of 350 nm and 8 W). To a solution of alkene-amino acid derivative (1.0 equiv) and tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-α-D-galactopyranose **1** (1.2 equiv) in CH_2Cl_2 (2.5 mL), 2,2-dimethoxy-2-phenylacetophenone (DPAP) (0.2 equiv) was added. After stirring for 3 h, another portion of DPAP (0.2 equiv) was added and the mixture was stirred for 4 h more. The reaction was concentrated and the residue was purified by a silica gel column chromatography to give the corresponding *S*-galactosaminyl amino acid derivative.

General procedure for the thiol-ene coupling by repairing way. To a solution of tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-α-D-galactopyranose **1** (2.0 equiv), catechol (1.2 equiv) and alkene-amino acid derivative (0.75 M in CH_2Cl_2, 1.0 equiv), BEt_3 (1.0 M in hexane, 1.2 equiv) was added. The reaction was stirred at room temperature and BEt_3 was regularly added until complete consumption of the alkenyl-amino acid derivative (0.1 equiv every 45 min). The reaction was then concentrated and the residue was purified by a silica gel column chromatography to give the corresponding *S*-galactosaminyl amino acid derivative.

Solid-phase (glyco)peptide synthesis (SPPS). All glycopeptides were synthesized by a stepwise solid-phase peptide synthesis using the Fmoc strategy on Rink Amide MBHA resin (128 mg, 0.1 mmol, loading of the resin: 0.78 mmol/g). The glycosylated amino acid building block (2.0 equiv) was manually coupled using HBTU, while the other Fmoc amino acids (10.0 equiv) were automatically coupled on an Applied Biosystems 433A peptide synthesizer using HBTU. The *O*-acetyl groups of $(AcO)_3$GalNAc moiety were removed in a mixture of NH_2NH_2/MeOH (7:3). The glycopeptides were then released from the resin and all acid sensitive side-chain protecting groups were simultaneously removed using TFA 95%, TIS 2.5%, H_2O 2.5%, followed by precipitation with diethyl ether. Finally, all glycopeptides were purified by HPLC on a Waters Delta Prep 4000 reverse phase HPLC and Waters 2987 Dual Absorbance Detector, using a Phenomenex Luna

C18(2) column (10 μ, 250 mm × 21.2 mm), 2% (v/v) CH_3CN in H_2O (containing 0.1% v/v TFA) gradient to 13% CH3CN (t = 27 min) and then to 60% CH_3CN (t = 33 min), 10 mL/min and registered at λ = 212 nm.

3,4,6-Tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-α-D-galactopyranose (1)

A stirred suspension of a mixture of α,β-anomers of 1,3,4,6-tetra-*O*-acetyl-2-acetamido-2-deoxy-D-galactopyranose (20.5 g, 52.65 mmol) and Lawesson's reagent (18.1 g, 44.7 mmol) in (1:1) toluene/1,2-dichloroethane solution (200 mL) was heated at 80 °C for 12 h. The reaction was then evaporated and the crude used directly in the next reaction. This crude was dissolved in 200 mL of methanol and stirred at 0 °C. TFA (10 mL) and water (10 mL) were then added, and the reaction was allowed to warm to room temperature and stirred for 4 h. The reaction was concentrated and the residue was purified by column chromatography using EtOAc/CH_2Cl_2 (8:2) as eluent, to give compound **1** (9.55 g, 50%) as a white foam. Spectroscopic data are consistent with the literature (ref. 26b and 34a).

***N*-Boc-*O*-(*S*-((2-acetamido-2-deoxy)-(3,4,6-tri-*O*-acetyl)-α-D-galactopyranosyl)mercaptopropyl)-L-serine benzyl ester (3)**

A solution of *N*-Boc-L-Ser(*O*-allyl)-OBn (**2b-Ser**) (90 mg, 0.27 mmol), 3,4,6-tri-*O*-acetyl-2-acetamido-2-deoxy-1-thio-α-D-galactopyranose (**1**) (166 mg, 0.46 mmol) and AIBN (27 mg, 0.16 mmol) in $CHCl_3$ (3 mL) was heated at reflux. After stirring for 3 h, another portion of AIBN (27 mg, 0.16 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was cooled, concentrated, and then chromatographed on silica gel using EtOAc/CH_2Cl_2 (8:2) as eluent to give compound **3** (104 mg, 56%) as a white foam. $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +71.4. **HRMS (ESI+) (m/z)** 699.2791 $[M+H]^+$; calculated $C_{32}H_{47}N_2O_{13}S^+$: 699.2793. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.39–7.29 (m, 5H, arom), 5.58 (d, *J* = 8.6 Hz, 1H, NH sugar), 5.45 (d, *J* = 5.1 Hz, 1H, H1), 5.41–5.34 (m, 2H, NH, H4), 5.25 (d, *J* = 12.4 Hz, 1H, CH_{2Bn}), 5.13 (d, *J* = 12.4 Hz, 1H, CH_{2Bn}), 5.03 (dd, *J* = 11.6, 2.7 Hz, 1H, H3), 4.81–4.71 (m, 1H, H2), 4.51 (t, *J* = 6.4 Hz, 1H, H5), 4.49–4.43 (m, 1H, Hα), 4.16–4.03 (m, 2H, 2H6), 3.84 (dd, *J* = 9.4, 2.6 Hz, 1H, Hβ), 3.64 (dd, *J* = 9.4, 3.2 Hz, 1H, Hβ), 3.51–3.35 (m, 2H, 2Hδ), 2.67–2.47 (m, 2H, 2Hζ), 2.15 (s, 3H, $COCH_3$), 2.03–1.95 (m, 9H, $COCH_3$, $COCH_3$, $NHCOCH_3$), 1.87–1.71 (m, 2H, 2Hε), 1.44 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.1, 170.7, 170.5, 170.4, 170.2, 155.6 (CO), 135.6, 128.7, 128.5, 128.3 (arom), 85.3 (C1), 80.2 ($\underline{C}(CH_3)_3$), 71.0 (Cβ), 69.7 (Cδ), 68.6 (C3), 67.5 (C4, C5), 67.2 (CH_{2Bn}), 62.0 (C6), 54.3 (Cα), 48.5 (C2), 29.8 (Cε), 28.5 ($C(\underline{C}H_3)_3$), 28.0 (Cζ), 23.5 ($NHCO\underline{C}H_3$), 20.8 (3 $CO\underline{C}H_3$).

***N*-Acetyl-*O*-(*S*-((2-acetamido-2-deoxy)-(3,4,6-tri-*O*-acetyl)-α-D-galactopyranosyl)mercaptopropyl)-L-serine benzyl ester (4)**

Trifluoroacetic acid (378 μL, 4.91 mmol) was added over a solution of compound **3** (104 mg, 0.15 mmol) in CH_2Cl_2 (8 mL) and the resulting solution was stirred for 2 h at room temperature and the solvent was removed under reduced pressure. The residue was dissolved in Et_2O (20 mL) and then evaporated. This operation was repeated several times to obtain the corresponding compound, which was used directly in the next reaction. Acetic anhydride (2.0 mL) and pyridine (4 mL) were added to previous compound, and the solution was stirred for 12 h at room temperature. The solvent was removed under reduced pressure and the residue was chromatographed on silica gel using CH_2Cl_2/MeOH (9:1) as eluent to give compound **4** (67 mg, 70%) as a white foam. $[\alpha]_D^{20}$ (c=1.01, $CHCl_3$): +81.9. **HRMS (ESI+) (m/z)** 641.2382 $[M+H]^+$; calculated $C_{29}H_{41}N_2O_{12}S^+$: 641.2375. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.43–7.22 (m, 5H, arom), 6.42 (d, *J* = 8.0 Hz, 1H, NH), 5.57 (d, *J* = 8.7 Hz, 1H, NH sugar), 5.48 (d, *J* = 5.2 Hz, 1H, H1), 5.38 (s, 1H, H4), 5.25 (d, *J* = 12.3 Hz, 1H, CH_{2Bn}), 5.16 (d, *J* = 12.3 Hz, 1H, CH_{2Bn}), 5.04 (dd, *J* = 11.7, 2.5 Hz, 1H, H3), 4.83–4.71 (m, 2H, Hα, H2), 4.51 (t, *J* = 6.4 Hz, 1H, H5), 4.17–4.05 (m, 2H, 2H6), 3.86 (dd, *J* = 9.3, 1.5 Hz, 1H, Hβ), 3.69–3.63 (m, 1H, Hβ), 3.52–3.46 (m, 1H, Hδ), 3.43–3.35 (m, 1H, Hδ), 2.66–2.58 (m, 1H, Hζ), 2.55–2.48 (m, 1H, Hζ), 2.15 (s, 3H, $COCH_3$), 2.08–1.95 (m, 12H, $COCH_3$, $COCH_3$, $NHCOCH_3$, $NHCOCH_3$), 1.86–1.72 (m, 2H, 2Hε). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.1, 170.5, 170.1, 170.2, 170.3 (6 CO), 135.5, 128.7, 128.5, 128.2 (arom), 85.1 (C1), 70.7 (Cβ), 69.4 (Cδ), 68.5 (C3), 67.4 (C4, C5), 67.3 (CH_{2Bn}), 61.8 (C6), 52.8 (Cα), 48.5 (C2), 29.7 (Cε), 27.6 (Cζ), 23.3 ($NHCOCH_3$), 23.2 ($NHCOCH_3$), 20.8, 20.8 (3 $COCH_3$).

***N*-Acetyl-*O*-(*S*-((2-acetamido-2-deoxy)-(3,4,6-tri-*O*-acetyl)-α-D-galactopyranosyl)mercaptopropyl)-L-serine methyl amide (5)**

A solution of compound **4** (67 mg, 0.10 mmol) in MeOH (10 mL) was treated with hydrogen under atmospheric pressure at room temperature, using Pd/C 10% as a catalyst (67 mg). This reaction mixture was stirred for 24 h. The solution was then filtered over celite, evaporated and this crude was used directly in the next reaction. It was dissolved in dry CH_3CN (20 mL) under an inert atmosphere. Methylamine hydrochloride (21 mg, 0.31 mmol), TBTU (44 mg, 0.14 mmol) and DIEA (91 μL, 0.52 mmol) were then added and the reaction was stirred at room temperature for 24 h. After evaporation, the residue was purified by a silica gel column chromatography, eluting with CH_2Cl_2/MeOH (9:1), to give compound **5** (27 mg, 46%) as a yellow oil. $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +101.4. **HRMS (ESI+) (m/z)** 564.2234 $[M+H]^+$; calculated $C_{23}H_{38}N_3O_{11}S^+$: 564.2222. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 6.71–6.60 (m, 2H, $NHCOCH_3$, $NHCH_3$), 6.00 (t, *J* = 8.4 Hz, 1H, NH sugar), 5.53 (dd, *J* = 8.0, 5.4 Hz, 1H, H1), 5.38–5.36 (m, 1H, H4), 5.04 (dt, *J* = 11.8, 3.1 Hz, 1H, H3), 4.75–4.66 (m, 1H, H2), 4.55–4.47 (m, 2H, H5, Hα), 4.14–4.07 (m, 2H, 2H6), 3.77 (ddd, *J* = 14.4, 9.4, 4.6 Hz, 1H, Hβ), 3.61–3.43 (m, 3H, Hβ, 2Hδ), 2.81 (d, *J* = 4.8 Hz, 3H, $NHCH_3$), 2.73–2.58 (m, 2H, 2Hζ), 2.14 (s, 3H, $COCH_3$), 2.06–1.95 (m, 12H, $COCH_3$, $COCH_3$, $NHCOCH_3$, $NHCOCH_3$), 1.92–1.77 (m, 2H, 2Hε). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm) (*Duplication of signals are observed, probably due to the coexistence of various conformers, since when the spectra were*

registered at higher temperature some of them collapse in one signal): 171.2, 170.6, 170.6, 170.5, 170.3, 170.3 (CO), 85.4, 85.4 (C1), 70.1, 70.1 (Cβ), 69.3, 69.2 (Cδ), 68.5 (C3), 67.6, 67.5, 67.4, 67.4 (C4, C5), 61.9, 61.9 (C6), 52.8, 52.8 (Cα), 48.8, 48.7 (C2), 29.7, 29.6 (Cε), 28.0, 27.9 (Cζ), 26.5 ($NHCH_3$), 23.4, 23.4 ($NHCOCH_3$), 23.3, 23.3 ($NHCOCH_3$), 20.9, 20.8, 20.8 (3 $COCH_3$).

_N_-Acetyl-_O_-(_S_-((2-acetamido-2-deoxy)-α-D-galactopyranosyl)mercaptopropyl)-L-serine methyl amide (6)

To a solution of compound **5** (27 mg, 0.05 mmol) in MeOH (2 mL), MeONa (0.5 M, 150 μL) was added and the reaction was stirred at room temperature for 2 h. Then, Dowex 50W-X8 (40 mg) was added to the reaction and it was stirred for 10 min. After that, the mixture was filtered and evaporated obtaining compound **6** in a 90% yield. In order to perform ELLA test, the residue was dissolved in H_2O (0.5 mL) and compound **6** was purified by semipreparative HPLC analysis using the following procedure: 0.5 mL injection loop was conducted on a Waters Delta Prep 4000 reverse phase HPLC and Waters 2987 Dual Absorbance Detector, using a Phenomenex Luna C18(2) column (10 μ, 250 mm × 21.2 mm), 5% (v/v) CH_3CN in H_2O (containing 0.1% v/v TFA) gradient to 30% CH_3CN (t = 30 min), 10 mL/min, λ = 212 nm to furnish **6** (10 mg, 50%, t_R 25.3 min) as a white foam after lyophilization. $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +85.0. **HRMS (ESI+) (m/z)** 438.1911 $[M+H]^+$; calculated $C_{17}H_{32}N_3O_8S^+$: 438.1905. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.20 (d, *J* = 5.5 Hz, 1H, $NHCOCH_3$), 8.13 (d, *J* = 7.6 Hz, 1H, NH sugar), 7.91 (s, 1H, $NHCH_3$). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.50 (d, *J* = 5.5 Hz, 1H, H1), 4.46 (t, *J* = 4.9 Hz, 1H, Hα), 4.35 (dd, *J* = 11.4, 5.5 Hz, 1H, H2), 4.28 (t, *J* = 6.1 Hz, 1H, H5), 4.00 (d, *J* = 2.4 Hz, 1H, H4), 3.84 (dd, *J* = 11.4, 3.1 Hz, 1H, H3), 3.80–3.71 (m, 4H, 2H6, 2Hβ), 3.69–3.55 (m, 2H, 2Hδ), 2.75 (s, 3H, $NHCH_3$), 2.72–2.61 (m, 2H, 2Hζ), 2.07 (s, 3H, $NHCOCH_3$), 2.05 (s, 3H, $COCH_3$), 1.94–1.81 (m, 2H, 2Hε). **^{13}C NMR (100 MHz, D_2O)** δ (ppm): 174.6, 174.6, 172.2 (CO), 83.9 (C1), 71.7 (C5), 69.2 (Cβ, Cδ), 68.6 (C4), 67.8 (C3), 61.3 (C6), 54.0 (Cα), 50.3 (C2), 28.5 (Cε), 26.8 (Cζ), 26.1 ($NHCH_3$), 22.0 ($NHCOCH_3$), 21.9 ($NHCOCH_3$).

_N_-Fmoc-_O_-(allyl)-L-serine *tert*-butyl ester (9-Ser)

To a solution of compound **2a-Ser** (1.02 g, 4.16 mmol) in CH_2Cl_2 (5 mL), trifluoroacetic acid (5 mL) was added. The resulting solution was stirred for 2 h at room temperature and the solvent was removed under reduced pressure. The residue was dissolved in Et_2O (10 mL) and then evaporated. This operation was repeated several times to obtain compound **7-Ser**, which was used directly in the next reaction. A solution of acetone/H_2O (1:1) (30 mL) was added to the previous compound and it was then treated with Na_2CO_3 (882 mg, 8.32 mmol) and FmocOSu (1.4 g, 4.16 mmol). The mixture was stirred for 18 h at room temperature and concentrated. The crude mixture was diluted with H_2O and extracted with Et_2O. The aqueous layer was then acidified to pH 3-4 with a 10% $KHSO_4$ solution. This aqueous solution was extracted with CH_2Cl_2 (3 x 50 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered,

and concentrated to give compound **8-Ser** (1.50 g, 98%) as a white solid, which was used directly in the next reaction. ^{t}Butyl alcohol (1.6 mL, 16.73 mmol) was added over DCC (2.72 g, 13.07 mmol) and CuCl (40 mg, 0.41 mmol) under inert atmosphere, light protected, and the resulting mixture was stirred for 5 days and dry CH_2Cl_2 (6 mL) was then added. After that, compound **8-Ser** (1.50 g, 4.08 mmol) dissolved in dry CH_2Cl_2 (6 mL) was added over the previous solution and the mixture was stirred for 4 h. The dicyclohexylurea byproduct was removed by filtration over celite, and the solution was washed with saturated $NaHCO_3$ (2 x 50 mL). The organic layer was dried, filtered and concentrated, and the residue was purified by silica gel column chromatography using hexane/EtOAc (8:2) as eluent, to give compound **9-Ser** (1.0 g, 58%) as a colorless oil. $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +16.5. **HRMS (ESI+) (m/z)** 446.1943 $[M+Na]^+$; calculated $C_{25}H_{29}NNaO_5^+$: 446.1938. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.80–7.73 (m, 2H, arom), 7.65–7.59 (m, 2H, arom), 7.44–7.36 (m, 2H, arom), 7.34–7.28 (m, 2H, arom), 5.86 (m, 1H, Hε), 5.67 (d, *J* = 8.3 Hz, 1H, NH), 5.27 (d, *J* = 17.3 Hz, 1H, Hζ*trans*), 5.20 (d, *J* = 10.4 Hz, 1H, Hζ*cis*), 4.47–4.33 (m, 3H, Hα, CH_{2Fmoc}), 4.24 (t, *J* = 7.2 Hz, 1H, CH_{Fmoc}), 4.04 (dd, *J* = 12.8, 5.3 Hz, 1H, Hδ), 3.96 (dd, *J* = 12.8, 5.4 Hz, 1H, Hδ), 3.86 (dd, *J* = 9.4, 2.5 Hz, 1H, Hβ), 3.69 (dd, *J* = 9.3, 2.2 Hz, 1H, Hβ), 1.49 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 169.5 (CO), 156.2 (CO), 144.1, 144.0, 141.4 (arom), 134.2 (Cε), 127.8, 127.2, 125.3, 120.1 (arom), 117.5 (Cζ), 82.4 ($\underline{C}(CH_3)_3$), 72.4 (Cδ), 70.4 (Cβ), 67.3 (CH_{2Fmoc}), 55.1 (Cα), 47.3 (CH_{Fmoc}), 28.2 ($C(\underline{C}H_3)_3$).

_N_-Fmoc-_S_-(allyl)-L-cysteine _tert_-butyl ester (9-Cys)

Following the same procedure above described for compound **9-Ser** and starting from **2a-Cys** (0.76 g, 2.91 mmol), compound **9-Cys** (679 mg) was obtained as a white solid with a 54% overall yield.

Mp: 61–63 ºC. $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): -6.0. **HRMS (ESI+) (m/z)** 440.1896 $[M+H]^+$; calculated $C_{25}H_{30}NO_4S^+$: 440.1890. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.78–7.76 (m, 2H, arom), 7.64–7.60 (m, 2H, arom), 7.43–7.38 (m, 2H, arom), 7.35–7.29 (m, 2H, arom), 5.82–5.71 (m, 1H, Hε), 5.62 (d, *J* = 7.7 Hz, 1H, NH), 5.16–5.08 (m, 2H, 2Hζ), 4.50 (m, 1H, Hα), 4.40 (d, *J* = 7.1 Hz, 1H, CH_{2Fmoc}), 4.25 (t, *J* = 7.1 Hz, 1H, CH_{Fmoc}), 3.18–3.15 (m, 2H, 2Hδ), 2.97 (dd, *J* = 13.8, 4.5 Hz, 1H, Hβ), 2.87 (dd, *J* = 13.8, 5.6 Hz, 1H, Hβ), 1.50 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 169.9 (CO), 155.8 (CO), 144.0, 143.9, 141.4 (arom), 133.9 (Cε), 127.8, 127.2, 125.2, 120.1 (arom), 117.9 (Cζ), 82.9 ($\underline{C}(CH_3)_3$), 67.2 (CH_{2Fmoc}), 54.2 (Cα), 47.2 (CH_{Fmoc}), 35.5 (Cδ), 33.3 (Cβ), 28.1 ($C(\underline{C}H_3)_3$).

_N_-Fmoc-_O_-(allyl)-L-threonine _tert_-butyl ester (9-Thr)

Following the same procedure above described for compound **9-Ser** and starting from **2a-Thr** (2.24 g, 8.64 mmol), compound **9-Thr** (2.03 g) was obtained as a white solid with a 59% overall yield.

Mp: 58–60 ºC. $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): -0.3. **HRMS (ESI+) (m/z)** 438.2292 $[M+H]^+$; calculated $C_{26}H_{32}NO_5^+$: 438.2275. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.78–7.74 (m, 2H, arom), 7.68–7.63 (m, 2H, arom), 7.43–7.38 (m, 2H, arom), 7.35–7.30 (m, 2H, arom), 5.94–5.82 (m, 1H, Hε), 5.60 (d, *J* = 9.6 Hz, 1H, NH), 5.29 (d, *J* = 17.0 Hz, 1H, Hζ*trans*), 5.19 (d, *J* = 10.4 Hz, 1H, Hζ*cis*), 4.48–4.38 (m, 2H, CH_{2Fmoc}), 4.34 (dd, *J* = 9.6, 1.3 Hz, 1H, Hα), 4.26 (t, *J* = 7.2 Hz, 1H, CH_{Fmoc}), 4.14–4.06 (m, 2H, Hβ, Hδ), 3.91 (dd, *J* = 12.6, 5.5 Hz, 1H, Hδ), 1.51 (s, 9H, $C(CH_3)_3$), 1.25 (d, *J* = 6.2 Hz, 3H, CH_3β). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 169.6 (CO), 156.7 (CO), 144.0, 143.8, 141.2 (arom), 134.5 (Cε), 127.6, 127.0, 127.0, 125.1, 125.1, 119.9 (arom), 116.9 (Cζ), 81.8 ($\underline{C}(CH_3)_3$), 74.7 (Cβ), 70.0 (Cδ), 67.1 (CH_{2Fmoc}), 59.1 (Cα), 47.1 (CH_{Fmoc}), 28.0 ($C(\underline{C}H_3)_3$), 16.2 (CH_3β).

***N*-Fmoc-*O*-(*S*-((2-acetamido-2-deoxy)-(3,4,6-tri-*O*-acetyl)-α-D-galactopyranosyl)mercaptopropyl)-L-serine *tert*-butyl ester (10-Ser)**

THERMAL WAY: A solution of **9-Ser** (195 mg, 0.46 mmol), compound **1** (284 mg, 0.78 mmol) and AIBN (46 mg, 0.28 mmol) in $CHCl_3$ (5 mL) was heated at reflux. After stirring for 3 h, another portion of AIBN (46 mg, 0.28 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was cooled, concentrated and then chromatographed on silica gel using EtOAc/CH_2Cl_2 (8:2) as eluent to give protected glycosyl amino acid **10-Ser** (253 mg, 70%) as a white solid.

PHOTOCHEMICAL WAY: A solution of compound **9-Ser** (232 m, 0.55 mmol), compound **1** (240 mg, 0.66 mmol) and DPAP (30 mg, 0.11 mmol) in CH_2Cl_2 (2.5 mL) was irradiated at λ_{max} 365 nm. After stirring for 3 h, another portion of DPAP (30 mg, 0.11 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was then concentrated and the residue was chromatographed on silica gel using EtOAc/CH_2Cl_2 (8:2) as eluent to give protected glycosyl amino acid **10-Ser** (227 mg, 53%) as a white solid.

REPAIRING WAY: A solution of compound **9-Ser** (205 mg, 0.48 mmol), compound **1** (352 mg, 0.97 mmol)**,** catechol (64 mg, 0.58 mmol) and BEt_3 (84 μL, 0.58 mmol) in CH_2Cl_2 (0.65 mL) was stirred at room temperature. After stirring for 1 h, another portion of BEt_3 (7 μL) was added (until complete consumption of compound **9-Ser**). The reaction mixture was then concentrated and the residue chromatographed on silica gel using EtOAc/CH_2Cl_2 (8:2) as eluent, to give protected glycosyl amino acid **10-Ser** (265 mg, 70%) as a white solid.

Physical data: mp: 63–65 ºC, $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +95.5. **HRMS (ESI+) (m/z)** 787.3119 $[M+H]^+$; calculated $C_{39}H_{51}N_2O_{13}S^+$: 787.3106. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.78–7.74 (m, 2H, arom), 7.65–7.59 (m, 2H, arom), 7.42–7.37 (m, 2H, arom), 7.34–7.29 (m, 2H, arom), 5.70–5.60 (m, 2H, N$\underline{H}$Ac, N$\underline{H}$Fmoc), 5.49 (d, *J* = 5.2 Hz, 1H, H1), 5.38 (s, 1H, H4), 5.04 (dd, *J* = 11.6, 2.6 Hz, 1H, H3), 4.82–4.72 (m, 1H, H2), 4.53 (t, *J* = 6.3 Hz, 1H, H5), 4.46–4.33 (m, 3H, CH_{2Fmoc}, Hα), 4.25 (t, *J* = 7.1 Hz, 1H, CH_{Fmoc}), 4.17–4.03 (m, 2H, 2H6), 3.82 (dd, *J* = 9.4, 2.8 Hz, 1H, Hβ), 3.67 (dd,

J = 9.3, 2.3 Hz, 1H, Hβ), 3.60–3.43 (m, 2H, 2Hδ), 2.75–2.57 (m, 2H, 2Hζ), 2.14 (s, 3H, $COCH_3$), 2.06–1.95 (m, 9H, 2 $COCH_3$, $NHCOCH_3$), 1.91–1.83 (m, 2H, 2Hε), 1.48 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.1, 170.4, 170.4, 170.2, 169.4, 156.1 (6 CO), 144.0, 144.0, 141.4, 127.8, 127.2, 125.3, 120.0 (arom), 85.2 (C1), 82.4 ($C(CH_3)_3$), 71.2 (Cβ), 69.7 (Cδ), 68.6 (C3), 67.5, 67.4 (C4, C5), 67.2 (CH_{2Fmoc}), 61.9 (C6), 55.0 (Cα), 48.5 (C2), 47.3 (CH_{Fmoc}), 29.8 (Cε), 28.1 ($C(CH_3)_3$), 28.0 (Cζ), 23.4 ($NHCOCH_3$), 20.8, 20.8 (3 $COCH_3$).

***N*-Fmoc-*S*-(*S*-((2-acetamido-2-deoxy)-(3,4,6-tri-*O*-acetyl)-α-D-galactopyranosyl)mercaptopropyl)-L-cysteine *tert*-butyl ester (10-Cys)**

THERMAL WAY: Following the same procedure above described for compound **10-Ser** and starting from **9-Cys** (191 mg, 0.43 mmol), compound **10-Cys** (70 mg, 20%) was obtained as a white solid.

PHOTOCHEMICAL WAY: Following the same procedure above described for compound **10-Ser** and starting from **9-Cys** (203 mg, 0.46 mmol), compound **10-Cys** (111 mg, 30%) was obtained as a white solid.

REPAIRING WAY: Following the same procedure above described for compound **10-Ser** and starting from **9-Cys** (310 mg, 0.71 mmol), compound **10-Cys** (201 mg, 36%) was obtained as a white solid.

Physical data: mp: 69–71 ºC, $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +82.9. **HRMS (ESI+) (m/z)** 803.2878 $[M+H]^+$; calculated $C_{39}H_{51}N_2O_{12}S_2^+$: 803.2878. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.77–7.74 (m, 2H, arom), 7.62–7.58 (m, 2H, arom), 7.41–7.36 (m, 2H, arom), 7.33–7.28 (m, 2H, arom), 5.69 (d, *J* = 8.0 Hz, 2H, NHAc, NHFmoc), 5.49 (d, *J* = 5.3 Hz, 1H, H1), 5.36 (d, *J* = 2.4 Hz, 1H, H4), 5.02 (dd, *J* = 11.7, 3.1 Hz, 1H, H3), 4.79–4.71 (m, 1H, H2), 4.54–4.44 (m, 2H, H5, Hα), 4.43–4.34 (m, 2H, CH_{2Fmoc}), 4.22 (t, *J* = 7.0 Hz, 1H, CH_{Fmoc}), 4.15–4.03 (m, 2H, 2H6), 2.99 (dd, *J* = 13.7, 4.8 Hz, 1H, Hβ), 2.92 (dd, *J* = 13.7, 5.1 Hz, 1H, Hβ), 2.75–2.13 (m, 4H, 2Hζ, 2Hδ), 2.13 (s, 3H, $COCH_3$), 2.02–1.93 (m, 9H, 2 $COCH_3$, $NHCOCH_3$), 1.91–1.82 (m, 2H, 2Hε), 1.48 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.0, 170.4, 170.3, 170.2, 169.8, 155.8 (6 CO), 143.9, 143.9, 141.4, 127.8, 127.2, 125.2, 120.1 (arom), 85.2 (C1), 83.0 ($C(CH_3)_3$), 68.6 (C3), 67.5 (C5), 67.4 (C4), 67.2 (CH_{2Fmoc}), 61.9 (C6), 54.4 (Cα), 48.4 (C2), 47.2 (CH_{Fmoc}), 34.8 (Cβ), 31.7 (Cδ), 29.9 (Cζ), 29.3 (Cε), 28.1 ($C(CH_3)_3$), 23.4 ($NHCOCH_3$), 20.8, 20.8 (3 $COCH_3$).

The MS spectrum of the crude reaction mixture, before to be purified by column chromatography, showed signals corresponding to the expected product **10-Cys** and side-products named **24**, **26**, **27**, **28** and **29** that probably are formed by radical couplings as shown in the scheme shown in the Supporting Information. This feature may explain the low yield obtained in this reaction, following the tree methodologies, when it is compared with the formation of **10-Ser**, **10-Thr** or **12**.

***N*-Fmoc-*O*-(*S*-((2-acetamido-2-deoxy)-(3,4,6-tri-*O*-acetyl)-α-D-galactopyranosyl)mercaptopropyl)-L-threonine *tert*-butyl ester (10-Thr)**

THERMAL WAY: Following the same procedure above described for compound **10-Ser** and starting from **9-Thr** (206 mg, 0.47 mmol), compound **10-Thr** (268 mg, 71%) was obtained as a white solid.

PHOTOCHEMICAL WAY: Following the same procedure above described for compound **10-Ser** and starting from **9-Thr** (160 mg, 0.37 mmol), compound **10-Thr** (181 mg, 62%) was obtained as a white solid.

REPAIRING WAY: Following the same procedure above described for compound **10-Ser** and starting from **9-Thr** (242 mg, 0.55 mmol), compound **10-Thr** (322 mg, 73%) was obtained as a white solid.

Pysical data: mp: 76–78 ºC, $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +78.3. **HRMS (ESI+) (m/z)** 801.3271 $[M+H]^+$; calculated $C_{40}H_{53}N_2O_{13}S^+$: 801.3263. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.76–7.73 (m, 2H, arom), 7.64–7.59 (m, 2H, arom), 7.40–7.35 (m, 2H, arom), 7.33–7.27 (m, 2H, arom), 5.75 (d, *J* = 8.7 Hz, 1H, NHAc), 5.54–5.46 (m, 2H, H1, NHFmoc), 5.37 (s, 1H, H4), 5.04 (dd, *J* = 11.7, 2.7 Hz, 1H, H3), 4.71–4.82 (m, 1H, H2), 4.53 (t, *J* = 6.4 Hz, 1H, H5), 4.43–4.34 (m, 2H, CH_{2Fmoc}), 4.20–4.19 (m, 2H, Hα, CH_{Fmoc}), 4.14–4.04 (m, 2H, 2H6), 3.98 (d, *J* = 6.1 Hz, 1H, Hβ), 3.62–3.54 (m, 1H, Hδ), 3.41–3.30 (m, 1H, Hδ), 2.74–2.55 (m, 2H, 2Hζ), 2.12 (s, 3H, $COCH_3$), 2.02–1.95 (m, 9H, 3 $COCH_3$, $NHCOCH_3$), 1.89–1.78 (m, 2H, 2Hε), 1.46 (s, 9H, $C(CH_3)_3$), 1.18 (d, *J* = 6.2 Hz, 3H, 3Hγ). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.0, 170.4, 170.3, 170.2, 169.8, 156.8 (6 CO), 144.0, 143.9, 141.4, 127.8, 127.1, 125.2, 120.0 (arom), 85.1 (C1), 82.1 ($C(CH_3)_3$), 75.6 (Cβ), 68.6 (C3), 67.4 (C4, C5, Cδ), 67.2 (CH_{2Fmoc}), 61.9 (C6), 59.1 (Cα), 48.4 (C2), 47.3 (CH_{Fmoc}), 30.2 (Cε), 28.2 ($C(CH_3)_3$), 28.0 (Cζ), 23.3 ($NHCOCH_3$), 20.8, 20.8 (3 $COCH_3$), 16.4 (Cγ).

_N_-Fmoc-_S_-((2-acetamido-2-deoxy)-(3,4,6-tri-_O_-acetyl)-α-D-galactopyranosyl)-L-homohomocysteine _tert_-butyl ester (12)

THERMAL WAY: A solution of *N*-Fmoc-L-allylglycine *tert*-butyl ester **11** (100 mg, 0.25 mmol), compound **1** (157 mg, 0.43 mmol) and AIBN (25 mg, 0.15 mmol) in $CHCl_3$ (2.5 mL) was heated at reflux. After stirring for 3 h, another portion of AIBN (25 mg, 0.15 mmol) was added and the reaction was stirred for 4 h more. The reaction mixture was cooled, concentrated and then chromatographed on silica gel using EtOAc/CH_2Cl_2 (8:2) as eluent to give protected glycosyl amino acid **12** (140 mg, 70%) as a white solid.

PHOTOCHEMICAL WAY: A solution of compound **11** (142 mg, 0.36 mmol), compound **1** (156 mg, 0.43 mmol) and DPAP (19 mg, 0.07 mmol) in CH_2Cl_2 (2.5 mL) was irradiated at λ_{max} 365 nm. After stirring for 3 h, another portion of DPAP (19 mg, 0.07 mmol) was added and the reaction was stirred for an additional time (4 h). The reaction mixture was concentrated and the residue was chromatographed on silica gel using EtOAc/CH_2Cl_2 (8:2) as eluent to give protected glycosyl amino acid **12** (137 mg, 50%) as a white solid.

REPAIRING TEC WAY: A solution of compound **11** (214 mg, 0.54 mmol), compound **1** (395 mg, 1.09 mmol), catechol (72 mg, 0.65 mmol) and BEt_3 (95 μL, 0.65 mmol) in CH_2Cl_2 (0.72 mL) was stirred at room temperature. After

stirring for 1 h, another portion of BEt_3 (8 μL, 0.05 mmol) was added (until complete consumption of compound **11**). The reaction mixture was concentrated and the residue chromatographed on silica gel using EtOAc/CH_2Cl_2 (7:3) as eluent to give protected glycosyl amino acid **12** (288 mg, 70%) as a white solid.

Physical data: mp: 59–61 ºC, $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +85.5. **HRMS (ESI+) (m/z)** 757.3012 $[M+H]^+$; calculated $C_{38}H_{49}N_2O_{12}S^+$: 757.3001. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.76–7.71 (m, 2H, arom), 7.61–7.56 (m, 2H, arom), 7.40–7.30 (m, 2H, arom), 7.33–7.27 (m, 2H, arom), 5.80 (d, *J* = 8.7 Hz, 1H, NHAc), 5.54–5.48 (m, 2H, NHFmoc, H1), 5.35 (s, 1H, H4), 5.03 (dd, *J* = 11.6, 2.3 Hz, 1H, H3), 4.79–4.71 (m, 1H, H2), 4.51 (t, *J* = 6.4 Hz, 1H, H5), 4.37 (d, *J* = 6.9 Hz, 2H, CH_{2Fmoc}), 4.29–4.18 (m, 2H, CH_{Fmoc}, Hα), 4.13–4.03 (m, 2H, H6), 2.66–2.54 (m, 2H, 2Hδ), 2.12 (s, 3H, $COCH_3$), 2.01–1.85 (m, 10H, 2 $COCH_3$, $NHCOCH_3$, Hβ), 1.78–1.59 (m, 3H, Hβ, 2Hγ), 1.46 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.4, 171.2, 170.5, 170.4, 170.2, 156.1 (6 CO), 144.0, 141.4, 127.9, 127.2, 125.2, 120.1 (arom), 85.3 (C1), 82.6 ($C(CH_3)_3$), 68.7 (C3), 67.4, 67.4 (C4, C5), 67.1 (CH_{2Fmoc}), 61.8 (C6), 53.6 (Cα), 48.5 (C2), 47.3 (CH_{Fmoc}), 32.2 (Cβ), 30.9 (Cδ), 28.7 ($C(CH_3)_3$), 25.7 (Cγ), 23.5 ($NHCOCH_3$), 20.9, 20.8 (3 $COCH_3$).

_N_-Fmoc-_S_-((2-acetamido-2-deoxy)-(3,4,6-tri-_O_-acetyl)-α-D-galactopyranosyl)-L-homocysteine _tert_-butyl ester (14)

Under an inert atmosphere, a solution of compound **1** (1.21 g, 3.36 mmol) in dry DMF (6 mL) was transferred into a Schlenk, which contains 4 Å molecular sieves, previously activated at 280 °C for 4 h under vacuum. After stirring for 5 min, a solution of compound **13** (500 mg, 1.12 mmol) in dry DMF (12 mL) was dropped into the Schlenk. The reaction was stirred for 24 h at room temperature, filtered over celite (to discard the molecular sieves) and the solvent was concentrated. The residue was purified by gradient column chromatography on silica gel using the following set of proportions: EtOAc/hexane, (1:1), (6:4), (7:3) to give protected glycosyl amino acid **14** (416 mg, 50%) as a white solid. Mp: 54–56 ºC, $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +68.2. **HRMS (ESI+) (m/z)** 743.2844 $[M+H]^+$; calculated $C_{37}H_{47}N_2O_{12}S^+$: 743.2846. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.77–7.75 (m, 2H, arom), 7.60–7.58 (m, 2H, arom), 7.42–7.38 (m, 2H, arom), 7.33–7.30 (m, 2H, arom), 5.56 (d, *J* = 8.7 Hz, 1H, NHAc), 5.48 (d, *J* = 5.0 Hz, 1H, H1), 5.43 (d, *J* = 7.9 Hz, 1H, NHFmoc), 5.39–5.38 (m, 1H, H4), 5.06–5.02 (m, 1H, H3), 4.82–4.75 (m, 1H, H2), 4.56–4.53 (m, 1H, H5), 4.40 (d, *J* = 6.7, 2H, CH_{2Fmoc}), 4.35–4.31 (m, 1H, Hα), 4.22 (t, *J* = 6.2 Hz, 1H, CH_{Fmoc}), 4.17–4.04 (m, 2H, 2H6), 2.74–2.55 (m, 2H, 2Hγ), 2.16–1.97 (m, 14H, 3 $COCH_3$, $NHCOCH_3$, 2Hβ), 1.48 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.0, 170.4, 170.3, 170.2, 167.8, 155.9 (6 CO), 143.9, 143.8, 141.3, 127.8, 127.1, 125.0, 120.0 (arom), 85.7 (C1), 82.9 ($C(CH_3)_3$), 68.5 (C3), 67.5 (C5), 67.3 (C4), 67.0 (CH_{2Fmoc}), 61.7 (C6), 53.6 (Cα), 48.3 (C2), 47.2 (CH_{Fmoc}), 33.5 (Cβ), 28.0 ($C(CH_3)_3$), 27.5 (Cγ), 23.3 ($NHCOCH_3$), 20.7, 20.7, 20.7 (3 $COCH_3$).

_N_-Fmoc-_S_-((2-acetamido-2-deoxy)-(3,4,6-tri-_O_-acetyl)-α-D-galactopyranosyl)-L-cysteine _tert_-butyl ester (16)

Under an inert atmosphere, a solution of compound **1** (1.01 g, 2.784 mmol) in dry DMF (6 mL) was transferred into a Schlenk, which contains the 4 Å molecular sieves, previously activated at 280 °C for 4 h under vacuum. After stirring for 5 min, a solution of compound **15** (400 mg, 0.928 mmol) in dry DMF (12 mL) was dropped into the Schlenk. The reaction was stirred for 24 h at room temperature, filtered over celite (to discard the molecular sieves) and the solvent was concentrated. The residue was purified by gradient column chromatography on silica gel using the following set of proportions: EtOAc/hexane, (1:1), (6:4), (7:3) to give protected glycosyl amino acid **16** (330 mg, 51%) as a white solid. Mp: 59–61 ºC, $[\alpha]_D^{20}$ (c=1.00, $CHCl_3$): +72.8. **HRMS (ESI+) (m/z)** 729.2686 $[M+H]^+$; calculated $C_{36}H_{45}N_2O_{12}S^+$: 729.2688. **^{1}H NMR (400 MHz, $CDCl_3$)** δ (ppm): 7.77–7.75 (m, 2H, arom), 7.64–7.62 (m, 2H, arom), 7.42–7.29 (m, 4H, arom), 6.10 (d, *J* = 8.5 Hz, 1H, NHAc), 5.65 (d, *J* = 8.9 Hz, 1H, NHFmoc), 5.38–5.37 (m, 2H, H1, H4), 4.98–4.94 (m, 1H, H3), 4.85–4.78 (m, 1H, H2), 4.59–4.55 (m, 1H, Hα), 4.48–4.38 (m, 3H, H5, CH_{2Fmoc}), 4.23–4.19 (m, 2H, H6, CH_{Fmoc}), 4.02–3.98 (m, 1H, H6), 3.29–3.20 (m, 1H, Hβ), 3.05–2.98 (m, 1H, Hβ), 2.18 (s, 3H, $COCH_3$), 2.01 (s, 3H, $NHCOCH_3$), 1.99 (s, 3H, $COCH_3$), 1.97 (s, 3H, $COCH_3$), 1.47 (s, 9H, $C(CH_3)_3$). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 171.0, 170.5, 170.2, 170.0, 168.8, 155.8 (6 CO), 143.9, 143.8, 141.4, 127.7, 127.1, 125.1, 120.1 (arom), 87.3 (C1), 82.9 ($C(CH_3)_3$), 68.3 (C3), 68.2 (C5), 67.4 (C4), 66.9 (CH_{2Fmoc}), 62.0 (C6), 54.5 (Cα), 48.4 (C2), 47.2 (CH_{Fmoc}), 36.2 (Cβ), 28.0 ($C(CH_3)_3$), 23.3 ($NHCOCH_3$), 20.7, 20.7, 20.6 (3 $COCH_3$).

Ala-Pro-Asp-Cys(α-*S*-GalNAc)-Arg-Pro-NH_2 (17)

Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Cys(α-*S*-$(AcO)_3$GalNAc)-OH (137 mg, 0.2 mmol), Fmoc-Asp(O^tBu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Ala-OH (311 mg, 1 mmol), peptide **17** was obtained with a 58% yield after purification by reversed phase HPLC and lyophilization (t_R 20.7 min). **HRMS (ESI+) (m/z)** 860.3936 $[M+H]^+$; calculated $C_{34}H_{58}N_{11}O_{13}S^+$: 860.3938. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.67 (t, J = 6.9 Hz, 1H, NHAsp), 8.40 (d, *J* = 7.5 Hz, 1H, NHCys), 8.36 (d, *J* = 7.4 Hz, 1H, NHArg), 8.20 (d, *J* = 8.2 Hz, 1H, $NHCOCH_3$), 7.70 (s, 1H, $CONH_2$), 7.20 (t, *J* = 5.3 Hz, 1H, NHεArg), 7.00 (s, 1H, $CONH_2$), 6.67 (br s, 2H, NH_2ηArg). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.58 (d, *J* = 5.4 Hz, 1H, H1), 4.73 (t, *J* = 6.7 Hz, 1H, HαAsp),4.68 (d, *J* = 3.9 Hz, 1H, HαArg), 4.65–4.57 (m, 1H, HαCys), 4.54–4.46 (m, 1H, HαPro), 4.44–4.35 (m, 3H, H2, HαPro, HαAla), 4.27–4.17 (m, 1H, H5), 4.00 (d, *J* = 2.9 Hz, 1H, H4), 3.85–3.59 (m, 7H, 2HδPro, 2HδPro, 2H6, H3), 3.30–3.18 (m, 2H, 2HδArg), 3.16–2.83 (m, 4H, 2HβAsp, 2HβCys), 2.45–2.25 (m, 2H, HβPro, HβPro), 2.14–1.91 (m, 9H, $NHCOCH_3$, 2HβArg, HβPro, HβPro, HγPro, HγPro), 1.82–1.60 (m, 4H, HγPro, HγPro, 2HγArg), 1.55 (d, *J* = 7.0 Hz, 3H, 3HβAla). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 176.8, 174.6, 173.9, 173.6, 171.9, 171.4, 171.1, 169.2 (8 CO), 156.8 (CζArg), 85.4 (C1), 71.9 (C5), 68.4 (C4), 67.6 (C3), 61.1 (C6), 60.3 (CαPro ,

CαPro), 53.7 (CαCys), 51.1 (CαArg), 50.1 (CαAsp), 50.0 (C2), 48.0 (CαAla), 47.9 (CδPro), 47.7 (CδPro), 40.6 (CδArg), 35.2 (CβAsp), 32.6 (CβCys), 29.6 (CβPro), 29.3 (CβPro), 27.7 (CγPro), 24.7(CγPro), 24.6 (CβArg), 24.1 (CγArg), 21.9 ($NHCO\underline{C}H_3$), 15.1 (CβAla).

Ala-Pro-Asp-hCys(α-*S*-GalNAc)-Arg-Pro-NH_2 (18)

Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-hCys(α-*S*-$(AcO)_3$GalNAc)-OH (135 mg, 0.2 mmol), Fmoc-Asp(O^{t}Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Ala-OH (311 mg, 1 mmol), peptide **18** was obtained with a 67% yield after purification by reversed phase HPLC and lyophilization (t_R 20.8 min). **HRMS (ESI+) (m/z)** 874.4084 $[M+H]^+$; calculated $C_{35}H_{60}N_{11}O_{13}S^+$: 874.4087. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.63 (d, *J* = 7.0 Hz, 1H, NHAsp), 8.39 (d, *J* = 7.6 Hz, 1H, NHArg), 8.27 (d, *J* = 7.2 Hz, 1H, NHhCys), 8.21 (d, *J* = 8.2 Hz, 1H, $NHCOCH_3$), 7.71 (s, 1H, $CONH_2$), 7.21 (t, *J* = 5.2 Hz, 1H, NHεArg), 7.00 (s, 1H, $CONH_2$), 6.67 (br s, 2H, NH_2ηArg). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.50 (d, *J* = 5.4 Hz, 1H, H1), 4.68–4.63 (m, 2H, HαAsp, HαhCys), 4.54–4.47 (m, 2H, HαPro, HαArg), 4.42–4.34 (m, 3H, H2, HαPro, HαAla), 4.30–4.27 (m, 1H, H5), 4.00 (s, 1H, H4), 3.89–3.64 (m, 7H, 2HδPro, 2HδPro, 2H6, H3), 3.23 (t, *J* = 6.7 Hz, 2H, 2HδArg), 3.01–2.85 (m, 2H, 2HβAsp), 2.67 (t, *J* = 6.9 Hz, 2H, 2HγhCys), 2.39–2.28 (m, 2H, HβPro, HβPro), 2.23–1.67 (m, 15H, HβPro, HβPro, 2HγPro, 2HγPro, 2HβhCys, 2HβArg, 2HγArg, $NHCO\underline{CH_3}$), 1.55 (d, J = 6.9 Hz, 3H, 3HβAla). **^{13}C NMR (100 MHz, $CDCl_3$)** δ (ppm): 175.8, 174.5, 173.8, 173.5, 172.8, 172.1, 171.3, 169.3 (8 CO), 156.8 (CζArg), 84.8 (C1), 71.8 (C5), 68.5 (C4), 67.7 (C3), 61.2 (C6), 60.3 (CαPro, CαPro), 52.4 (CαArg), 51.2 (CαhCys), 50.1 (CαAsp), 50.0 (C2), 48.0 (CαAla), 47.9 (CδPro), 47.7 (CδPro), 40.5 (CδArg), 35.0 (CβAsp), 31.3 (CγPro), 29.6 (CβPro), 29.4 (CβPro), 27.4 (CγPro), 27.1 (CγhCys), 24.7 (CβArg), 24.7 (CβhCys), 24.1 (CγArg), 21.9 ($NHCOCH_3$), 15.1 (CβAla).

Ala-Pro-Asp-hhCys(α-*S*-GalNAc)-Arg-Pro-NH_2 (19)

Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-hhCys(α-*S*-$(AcO)_3$GalNAc)-OH (140 mg, 0.2 mmol), Fmoc-Asp(O^{t}Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Ala-OH (311 mg, 1 mmol), peptide **19** was obtained with a 59% yield after purification by reversed phase HPLC and lyophilization (t_R 24.4 min). **HRMS (ESI+) (m/z)** 888.4251 $[M+H]^+$; calculated $C_{36}H_{62}N_{11}O_{13}S^+$: 888.4249. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.63 (d, *J* = 7.0 Hz, 1H, NHAsp), 8.32 (d, *J* = 7.3 Hz, 1H, NHArg), 8.27 (d, *J* = 7.1 Hz, 1H, NHhhCys), 8.20 (d, *J* = 8.1 Hz, 1H, $NHCOCH_3$), 7.69 (s, 1H, $CONH_2$), 7.20 (t, *J* = 5.2 Hz, 1H, NHεArg), 7.00 (s, 1H, $CONH_2$), 6.67 (br s, 2H, NH_2ηArg). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.49 (d, *J* = 5.4 Hz, 1H, H1), 4.68–4.64 (m, 2H, HαAsp, HαArg), 4.50–4.46 (m, 1H, HαPro), 4.41–4.30 (m, 4H, H2, HαPro, HαAla,

HαhhCys), 4.25 (t, *J* = 5.9 Hz, 1H, H5), 3.98 (d, *J* = 2.8 Hz, 1H, H4), 3.83 (dd, *J* = 11.4, 3.1 Hz, 1H, H3), 3.81–3.58 (m, 6H, 2H6, 2HδPro, 2HδPro), 3.27–3.16 (m, 2H, 2HδArg), 2.95 (dd, *J* = 17.0, 6.5 Hz, 1H, HβAsp), 2.85 (dd, *J* = 17.0, 7.1 Hz, 1H, HβAsp), 2.70–2.55 (m, 2H, 2HδhhCys), 2.38–2.27 (m, 2H, HβPro, HβPro), 2.10–1.81 (m, 11H, NHCO$\underline{C}H_3$, 2HβhhCys, 2HγhhCys, HβPro, HβPro, HγPro, HγPro), 1.80–1.58 (m, 6H, 2HγArg, 2HβArg, HγPro, HγPro), 1.54 (d, *J* = 7.0 Hz, 3H, 3HβAla). **^{13}C NMR (100 MHz, D_2O)** δ (ppm): 176.8, 174.5, 173.8, 173.5, 173.1, 172.0, 171.3, 169.2 (8 CO), 156.7 (CζArg), 83.6 (C1), 71.7 (C5), 68.5 (C4), 67.7 (C3), 61.2 (C6), 60.3 (CαPro), 60.3 (CαPro), 53.2 (CαhhCys), 51.0 (CαArg), 50.1 (CαAsp), 50.0 (C2), 48.0 (CαAla), 47.9 (CδPro), 47.7 (CδPro), 40.6 (CδArg), 35.1 (CβAsp), 30.0 (CγPro), 29.6 (CβPro, CβhhCys), 29.4 (CβPro), 29.4 (CδhhCys), 27.4 (CγPro), 25.0 (CβArg), 25.0 (CγhhCys), 24.1 (CγArg), 21.9 (NHCO$\underline{C}H_3$), 15.1 (CβAla).

Ala-Pro-Asp-Ser(*O*-mercaptopropyl-α-*S*-GalNAc)-Arg-Pro-NH_2 (20)

Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Ser(mercaptopropyl-α-*S*-$(AcO)_3$GalNAc)-OH (146 mg, 0.2 mmol), Fmoc-Asp(O^{t}Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Ala-OH (311 mg, 1 mmol), peptide **20** was obtained with a 67% yield after purification by reversed phase HPLC and lyophilization (t_R 25.7 min). **HRMS (ESI+) (m/z)** 459.7210 $[M+2H]^{+2}$; calculated $C_{37}H_{65}N_{11}O_{14}S^{+2}$: 459.7211. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.60 (d, *J* = 6.8 Hz, 1H, NHAsp), 8.31–8.20 (m, 3H, NHSer, NHArg, $NHCOCH_3$), 7.69 (s, 1H, $CONH_2$), 7.25 (d, *J* = 6.8 Hz, 1H, NHεArg), 7.01 (s, 1H, $CONH_2$), 6.68 (br s, 2H, NH_2ηArg). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.50 (d, *J* = 5.4 Hz, 1H, H1), 4.76–4.68 (m, 2H, HαAsp, HαArg), 4.53 (t, *J* = 5.0 Hz, 1H, HαSer), 4.51–4.47 (m, 1H, HαPro), 4.42–4.32 (m, 3H, H2, HαPro, HαAla), 4.26 (t, J = 6.0 Hz, 1H, H5), 3.99 (d, *J* = 2.8 Hz, 1H, H4), 3.83 (dd, *J* = 11.4, 3.0 Hz, 1H, H3), 3.81–3.70 (m, 6H, 2H6, HδPro, HδPro, 2HβSer), 3.69–3.55 (m, 4H, HδPro, HδPro, 2HδSer), 3.27–3.17 (m, 2H, 2HδArg), 2.98 (dd, *J* = 17.1, 6.4 Hz, 1H, HβAsp), 2.89 (dd, *J* = 17.1, 6.9 Hz, 1H, HβAsp), 2.72–2.56 (m, 2H, HζSer), 2.39–2.26 (m, 2H, HβPro, HβPro), 2.11–1.80 (m, 12H, NHCO$\underline{C}H_3$, 2HβArg, 2HεSer, HβPro, HβPro, 2HγPro, HγPro), 1.79–1.59 (m, 3H, HγPro, 2HγArg), 1.55 (d, *J* = 7.0 Hz, 3H, 3HβAla). **^{13}C NMR (100 MHz, D_2O)** δ (ppm): 176.8, 174.5, 173.8, 173.6, 172.2, 171.1, 170.0, 169.3 (8 CO), 156.7 (CζArg), 83.9 (C1), 71.7 (C5), 69.5 (CδSer), 69.1 (CβSer), 68.5 (C4), 67.7 (C3), 61.2 (C6), 60.4 (CαPro), 60.4 (CαPro), 53.6 (CαSer), 51.0 (CαArg), 50.2 (C2), 50.0 (CαAsp), 48.1 (CαAla), 47.9 (CδPro), 47.8 (CδPro), 40.6 (CδArg), 35.2 (CβAsp), 29.6 (CβPro), 29.4 (CβPro), 28.7 (CεSer), 27.7 (CγPro), 26.9 (CζSer), 24.7, 24.6 (CβArg, CγPro), 24.0 (CγArg), 21.9 (NHCO$\underline{C}H_3$), 15.1 (CβAla).

Ala-Pro-Asp-Cys(*S*-mercaptopropyl-α-*S*-GalNAc)-Arg-Pro-NH_2 (21)

Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Cys(mercaptopropyl-α-*S*-$(AcO)_3$GalNAc)-OH (149 mg, 0.2 mmol), Fmoc-Asp(O^{t}Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Ala-OH (311 mg, 1 mmol), peptide **21** was obtained with a 63% yield after purification by reversed phase HPLC and lyophilization (t_R 30.3 min). **HRMS (ESI+) (m/z)** 934.4123 $[M+H]^+$; calculated $C_{37}H_{64}N_{11}O_{13}S_2^+$: 934.4121. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.56 (d, *J* = 7.3 Hz, 1H, NHAsp), 8.31 (d, *J* = 7.5 Hz, 1H, NHArg), 8.18 (d, *J* = 7.4 Hz, 1H, NHCys), 8.13 (d, *J* = 8.0 Hz, 1H, $NHCOCH_3$), 7.59 (s, 1H, $CONH_2$), 7.09 (t, *J* = 5.4 Hz, 1H, NHεArg), 6.91 (s, 1H, $CONH_2$), 6.57 (br s, 1H, NH_2ηArg). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.51 (d, *J* = 5.4 Hz, 1H, H1), 4.74–4.66 (m, 2H, HαAsp, HαArg), 4.55–4.47 (m, 2H, HαCys, HαPro), 4.41–4.31 (m, 3H, H2, HαPro, HαAla), 4.26 (t, *J* = 6.0 Hz, 1H, H5), 3.99 (d, *J* = 2.4 Hz, 1H, H4), 3.86–3.79 (m, 2H, H3, HδPro), 3.77 (d, *J* = 6.2 Hz, 2H, 2H6), 3.75–3.69 (m, 1H, HδPro), 3.69–3.60 (m, 2H, 2HδPro), 3.26–3.18 (m, 2H, 2HδArg), 3.02–2.94 (m, 2H, HβCys, HβAsp), 2.91–2.84 (m, 2H, HβAsp, HβCys), 2.77–2.63 (m, 4H, 2HζCys, 2HδCys), 2.39–2.27 (m, 2H, HβPro, HβPro), 2.10–1.92 (m, 9H, $NHCOCH_3$, 2HβArg, 2HγPro, 2HβPro), 1.91–1.82 (m, 3H, 2HεCys, HγPro), 1.80–1.61 (m, 3H, HγPro, 2HγArg), 1.54 (d, *J* = 7.0 Hz, 3H, 3HβAla). **^{13}C NMR (100 MHz, D_2O)** δ (ppm): 176.8, 174.5, 173.8, 173.5, 171.9, 171.5, 171.1, 169.3 (8 CO), 156.7 (CεArg), 83.9 (C1), 71.7 (C5), 68.5 (C4), 67.6 (C3), 61.2 (C6), 60.4 (CαPro), 60.3 (CαPro), 52.9 (CαCys), 51.1 (CαArg), 50.2 (C2), 49.9 (CαAsp), 48.0 (CαAla), 47.9 (CδPro), 47.7 (CδPro), 40.6 (CδArg), 35.1 (CβAsp), 32.5 (CβCys), 30.3 (CδCys), 29.6 (CβPro), 29.4 (CβPro), 29.1 (CζCys), 28.6 (CεCys), 27.6 (CγPro), 24.7, 24.6 (CβArg, CγPro), 24.1 (CγArg), 21.9 ($NHCOCH_3$), 15.1 (CβAla).

Ala-Pro-Asp-Thr(*O*-mercaptopropyl-α-*S*-GalNAc)-Arg-Pro-NH_2 (22)

Following SPPS methodology with Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Arg(Pbf)-OH (649 mg, 1 mmol), Fmoc-Thr(mercaptopropyl-α-*S*-$(AcO)_3$GalNAc)-OH (149 mg, 0.2 mmol), Fmoc-Asp(O^{t}Bu)-OH (412 mg, 1 mmol), Fmoc-Pro-OH (337 mg, 1 mmol), Fmoc-Ala-OH (311 mg, 1 mmol), glycopeptide **22** was obtained with a 62% yield after purification by reversed phase HPLC and lyophilization (t_R 29.7 min). **HRMS (ESI+) (m/z)** 932.4506 $[M+H]^+$; calculated $C_{38}H_{66}N_{11}O_{14}S^+$: 932.4506. **^{1}H NMR (400 MHz, H_2O/D_2O 9:1)** δ (ppm) amide protons: 8.71 (d, *J* = 7.0 Hz, 1H, NHAsp), 8.23 (d, *J* = 7.9 Hz, 1H, $NHCOCH_3$), 8.18 (d, *J* = 7.4 Hz, 1H, NHArg), 8.06 (d, *J* = 8.2 Hz, 1H, NHThr), 7.68 (s, 1H, $CONH_2$), 7.21 (t, *J* = 4.5 Hz, 1H, NHεArg), 7.01 (s, 1H, $CONH_2$), 6.67 (br s, 2H, NH_2ηArg). **^{1}H NMR (400 MHz, D_2O)** δ (ppm): 5.45 (d, *J* = 5.5 Hz, 1H, H1), 4.78–4.74 (m, 1H, HαAsp), 4.65 (dd, *J* = 8.1, 5.3 Hz, 1H, HαArg), 4.46 (dd, *J* = 8.1, 6.3 Hz, 1H, HαPro), 4.40–4.28 (m, 4H, H2, HαThr, HαPro, HαAla), 4.22 (t, *J* = 6.0 Hz, 1H, H5), 4.01–3.94 (m, 2H, HβThr, H4), 3.79 (dd, *J* = 11.4, 3.1 Hz, 1H, H3), 3.76–3.55 (m, 7H, 2H6, 2HδPro, 2HδPro, HδThr), 3.46–3.38 (m,

1H, HδThr), 3.23–3.15 (m, 2H, 2HδArg), 2.97 (dd, *J* = 17.1, 7.0 Hz, 1H, HβAsp), 2.87 (dd, *J* = 17.1, 6.6 Hz, 1H, HβAsp), 2.69–2.50 (m, 2H, 2HζThr), 2.36–2.22 (m, 2H, HβPro, HβPro), 2.08–1.71 (m, 13H, NHCO$\underline{CH_3}$, 2HβArg, 2HεThr, HβPro, HβPro, 2HγPro, 2HγPro), 1.70–1.57 (m, 2H, 2HγArg), 1.51 (d, *J* = 7.0 Hz, 3H, 3HβAla), 1.13 (d, *J* = 6.3 Hz, 3H, 3HγThr). **^{13}C NMR (100 MHz, D_2O)** δ (ppm): 176.7, 174.5, 173.8, 173.5, 172.5, 171.2, 171.1, 169.2 (8 CO), 156.7 (CζArg), 83.9 (C1), 74.8 (CβThr), 71.7 (C5), 68.4 (C4), 67.6 (C3), 67.4 (CδThr), 61.1 (C6), 60.2 (CαPro, CαPro), 57.8 (CαThr), 51.1 (CαArg), 50.2 (C2), 49.9 (CαAsp), 48.0 (CαAla), 47.9 (CδPro), 47.6 (CδPro), 40.6 (CδArg), 35.0 (CβAsp), 29.6 (CβPro), 29.4 (CβPro), 29.0 (CεThr), 27.6 (CγPro), 26.9 (CζThr), 24.7, 24.6 (CβArg, CγPro), 24.1 (CγArg), 21.9 (NHCO$\underline{C}H_3$), 15.7 (CγThr), 15.0 (CβAla).

2D NMR experiments. Spectra were assigned using COSY and HSQC experiments. NOESY experiments were recorded on a 400 MHz spectrometer at 298 K and pH 5.5 in H_2O/D_2O (9:1). The experiments were conducted by using phase-sensitive ge-2D NOESY with WATERGATE for H_2O/D_2O (9:1) spectra. Distances involving NH protons were semi-quantitatively determined by integrating the volume of the corresponding cross-peaks. The number of scans used was 16 and the mixing time was 500 ms.

Molecular dynamics (MD) simulations with time averaged restraints (MD-tar). The simulations were carried out with AMBER 12 package[54] with ff14SB[55], GAFF[56] and GLYCAM06j[57] force fields. The parameters and charges for the unnatural amino acids were generated with the *antechamber* module of AMBER 12, using GAFF force field and AM1-BCC method for charges.[58] Prior to MD-tar productive simulations, we performed an equilibration protocol consisting of an initial minimization of the water box of 5000 steps, followed by a 2500-step minimization of the whole system. Then, the TIP3P water[59] box was heated at constant volume until 298 K using a time constant for the heat bath coupling of 1 ps. The equilibration finished with 200 ps of MD simulation without restraints, at a constant pressure of 1 bar and turning on the Langevin temperature scaling with a collision frequency of 1 ps. Furthermore, non-bonded interactions were cut-off at 8.0 Å and updated every 25 steps. Periodic boundary conditions and the Particle Mesh Ewald method[60] were turned on in every step of the equilibration protocol to evaluate the long-range electrostatic forces, using a grid spacing of approximately 1 Å. The NOE-derived distances shown in Figure 2 and in the Supporting Information were imposed as time-averaged constraint, applying an r^{-6} averaging. The equilibrium distance range was set to $r_{exp} - 0.2$ Å $\leq r_{exp} \leq r_{exp} +$ 0.2 Å. Trajectories were run at 298 K, with a decay constant of 2000 ps and a time step of 1 fs. The force constants r_{k2} and r_{k3} used in each case were 10 kcal·mol^{-1}·Å^{-2}. The overall simulation length for the simulations was 20 ns. The coordinates were saved each 1 ps, obtaining MD trajectories of 20000 frames each. A convergence within the equilibrium distance range was obtained in the simulations.

Molecular dynamics (MD) simulations on the complex SM3:22. The crystal structure of glycopeptides **22** in complex with scFv-SM3 (PDB id: 5fxc) was used as starting coordinates. The structure was conveniently modified to contain the missing linker as well as the GalNAc unit for the glycopeptide. The complex was then immersed in a water box with a 10 Å buffer of TIP3P water molecules. A two-stage geometry optimization approach was performed. The first stage minimizes only the positions of solvent molecules and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then gently heated by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions. Harmonic restraints of 30 kcal·mol^{-1} were applied to the solute, and the Andersen temperature-coupling scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method. An 8 Å cut-off was applied to Lennard-Jones and electrostatic interactions. Each system was equilibrated for 2 ns with a 2 fs time step at a constant volume and temperature of 300 K. Production trajectories were then run for additional 100 ns under the same simulation conditions.

Crystallization. Expression and purification of scFv-SM3 has been described previously by us.[53] Crystals were grown by sitting drop diffusion at 18 ºC. The drops were prepared by mixing 0.5 µL of protein solution, containing 15 mg/mL of scFv-SM3 and 10 mM of glycopeptide **22** with 0.5 µL of the mother liquor. Crystals of scFv-SM3 with the peptide above were grown in 20% PEG 3350, 0.2 M disodium hydrogen phosphate. The crystals were cryoprotected in mother liquor containing 15% ethylenglycol and frozen in a nitrogen gas stream cooled to 100 K.

Structure determination and refinement. The data was processed and scaled using the XDS package[61] and CCP4software,[62] relevant statistics are given in Supplementary Table S1. The crystal structures were solved by molecular replacement with Phaser[62] and using the PDB entry 1SM3 as the template. Initial phases were further improved by cycles of manual model building in Coot[63] and refinement with REFMAC5.[64] The final models were validated with PROCHECK,[65] model statistics are given in the Supporting Information. Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB). PDB id: 5fxc

Supporting Information

2D NOESY spectra of the neoglycopeptides, conformational analysis details, NMR spectra for all new compounds, CD spectra, HPLC chromatograms and X-ray structure details. This material is available free of charge via the Internet at http://pubs.acs.org.

Author Information

Corresponding authors

*E-mail: francisco.corzana@unirioja.es; jesusmanuel.peregrina@unirioja.es;

hector.busto@unirioja.es

Author contributions

‡These authors contributed equally.

Notes

The authors declare no competing financial interest.

Acknowledgments

We thank *the Ministerio de Economía y Competitividad* (projects CTQ2015-67727-R, CTQ2015-64597-P-C02-01, CTQ2013-44367-C2-2-P and UNLR13-4E-1931). V. R.-O. And I. C. thank Gobierno de La Rioja and *Universidad de La Rioja* for the FPI grants, respectively.

References

1. (a) Gaidzik, N.; Westerlind, U.; Kunz, H. *Chem. Soc. Rev.* **2013**, *42*, 4421–4442. (b) Pinho, S. S.; Reis, C. A. *Nat. Rev. Cancer* **2015**, *15*, 540–555.

2. Nativi, C.; Renaudet, O. *ACS Med. Chem. Lett.* **2014**, *5*, 1176–1178.

3. Buskas, T.; Thompson, P.; Boons, G. J. *Chem. Commun.* **2009**, 5335–5349.

4. (a) Wilson, R. M.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2013**, *135*, 14462–14472. (b) Keding, S. J.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11937–11942.

5. Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem. Int. Ed.* **2011**, *50*, 1770–1791.

6. Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycobiology* **2004**, *14*, 681–692.

7. Martínez-Sáez, N.; Supekar, N. T.; Wolfert, M. A.; Bermejo, I. A.; Hurtado-Guerrero, R.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Boons, G.-J.; Peregrina, J. M.; Corzana, F. *Chem. Sci.* **2016**, *7*, 2294–2301.

8. Palitzsch, B.; Gaidzik, N.; Stergiou, N.; Stahn, S.; Hartmann, S.; Gerlitzki, B.; Teusch, N.; Flemming, P.; Schmitt, E. Kunz, H. *Angew. Chem. Int. Ed.* **2016**, *55*, 2894–2898.

9. Miles, D.; Roché, H.; Martin, M.; Perren, T. J.; Cameron, D. A.; Glaspy, J.; Dodwell, D.; Parker, J.; Mayordomo, J.; Tres, A.; Murray, J. L.; Ibrahim, N. K. *Oncologist* **2011**, *16*, 1092–1100.

10. Ohyama, C. *Int. J. Clin. Oncol.* **2008**, *13*, 308–313.

11. Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. *Angew. Chem. Int. Ed.* **2014**, *53*, 11917–11920.

12. Dondoni, A.; Marra, A. *Chem. Rev.* **2000**, *100*, 4395–4422.

13. Kuberan, B.; Sikkander, S. A.; Tomiyama, H.; Linhardt, R. J. *Angew. Chem. Int. Ed.* **2003**, *42*, 2073–2075.

14. Awad, L.; Madani, R.; Gillig, A.; Kolympadi, M.; Philgren, M.; Muhs, A.; Gérard, C.; Vogel, P. *Chem. Eur. J.* **2012**, *18*, 8578–8582.

15. Wu, X.; Lipinski, T.; Paszkiewicz, E.; Bundle, D. R. *Chem. Eur. J.* **2008**, *14*, 6474–6482.

16. Bundle, D. R.; Rich, J. R.; Jacques, S.; Yu, H. N.; Nitz, M.; Ling, C.-C. *Angew. Chem. Int. Ed.* **2005**, *44*, 7725–7729.

17. Huo, C.-X.; Zheng, X.-J.; Xiao, A.; Liu, C.-C.; Sun, S.; Lv, Z.; Ye, X.-S. *Org. Biomol. Chem.* **2015**, *13*, 3677–3690.

18. Oberbillig, T.; Mersch, C.; Wagner, S.; Hoffmann-Röder, A. *Chem. Commun.* **2012**, *48*, 1487–1489.

19. Johannes, M.; Reindl, M.; Gerlitzki, B.; Schmitt, E.; Hoffmann-Röder, A. *Beilstein J. Org. Chem.* **2015**, *11*, 155–161.

20. (a) Vichier-Guerre, S.; Lo-Man, R.; Huteau, V.; Dériaud, E.; Leclerc, C.; Bay, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3567–3570. (b) Norgren, A. S.; Arvidsson, P. I. *Org. Biomol. Chem.* **2005**, *3*, 1359–1361.

21. (a) Wolfert, M. A.; Boons, G.-J. *Nat. Chem. Biol.* **2013**, *9*, 776–784. (b) Apostolopoulos, V.; Yuriev, E.; Lazoura, E.; Yu, M.; Ramsland, P. A. *Human Vaccines* **2008**, *4*, 400–409. (c) Glithero, A.; Tormo, J.; Haurum, J. S.; Arsequell, G. *Immunity* **1999**, *10*, 63–74. (d) Apostolopoulos, V.; Yuriev, E.; Ramsland, P. A.; Halton, J.; Osinski, C.; Li, W.; Plebanski, M.; Paulsen, H.; McKenzie, I. F. C. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15029–15034. (e) Speir, J. A.; Abdel-Motal, U. M.; Jondal, M.; Wilson, I. A. *Immunity* **1999**, *10*, 51–61.

22. Bonauer, C.; Walenzyk, T.; König, B. *Synthesis* **2006**, 1–20.

23. (a) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477–501. (b) Jack, R. W.; Jung, G. *Curr. Opin. Chem. Biol.* **2000**, *4*, 310–317. (c) Jung, G. *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1051–168.

24. (a) Levengood, M. R.; van der Donk, W. A. *Nat. Protoc.* **2007**, *1*, 3001–3010. (b) Wang, J.; Schiller, S. M.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2007**, *46*, 6849–851. (c) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053. (d) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2008**, *47*, 6399–6401. (e) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. Asian J.* **2009**, *4*, 630–640.

25. Cohen, S. B.; Halcomb, R. L. *J. Am. Chem. Soc.* **2002**, *124*, 2534–2543.

26. (a) Zhu, Y.; van der Donk, W. A., *Org. Lett.* **2001**, *3*, 1189–119. (b) Galonic, D. P.; van der Donk, W. A.; Gin, D. Y. *Chem. Eur. J.* **2003**, *9*, 5997–6006.

27. Aydillo, C.; Avenoza, A.; Busto, J. H.; Jiménez-Osés, G.; Peregrina, J. M.; Zurbano, M. M. *Org. Lett.* **2012**, *14*, 334–337.

28. Aydillo, C.; Compañón, I.; Avenoza, A.; Busto, J. H.; Corzana, F.; Peregrina, J. M.; Zurbano, M. M. *J. Am. Chem. Soc*. **2014**, *136*, 789–800.

29. (a) Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 7798–7802. (b) Fiore, M.; Lo Conte, M.; Pacifico, S.; Marra, A.; Dondoni, A. *Tetrahedron Lett.* **2011**, *52*, 444–447. (c) Dondoni, A. *Angew. Chem. Int. Ed.* **2008**, *47*, 8995–8997. (d) Griesbaum, K. *Angew. Chem., Int. Ed. Engl.* **1970**, *9*, 273–287. (e) Lázár, L.; Csávás, M.; Herczeg, M.; Herczegh, P.; Borbás, A. *Org. Lett.* **2012**, *14*, 4650–4653. (f) Markey, L.; Giordani, S.; Scanlan, E. M. *J. Org. Chem.* **2013**, *78*, 4270–4277. (g) Dénès, F.; Pichowicz, M.; Povie, G.; Renaud, P. *Chem. Rev.* **2014**, *114*, 2587–2693.

30. (a) Zhu, X.; Schmidt, R. R. *Chem. Eur. J.* **2004**, *10*, 875–887. (b) Galonic, D. P.; van der Donk, W. A.; Gin, D. Y. *J. Am. Chem. Soc.* **2004**, *126*, 12712–12713. (c). Galonic, D. P.; Ide, N. D.; van der Donk, W. A.; Gin, D. Y. *J. Am. Chem. Soc.* **2005**, *127*, 7359–7369. (d) Zhu, X.; Schmidt, R. R. *Tetrahedron Lett.* **2003**, *44*, 6063–6067.

31. Thayer, D. A.; Yu, H. N.; Galan, M. C.; Wong, C.-H. *Angew. Chem. Int. Ed.* **2005**, *44*, 4596–4599.

32. (a) Marra, A.; Staderini, S.; Berthet, N.; Dumy, P.; Renaudet, O.; Dondoni, A. *Eur. J. Org. Chem.* **2013**, 1144–1149. (b) Hoyle, C. E.; Lowe, A. B.; Bowman, C. N. *Chem. Soc. Rev.* **2010**, *39*, 1355–1387. (c) Hoyle, C. E.; Bowman, C. N. *Angew. Chem. Int. Ed.* **2010**, *49*, 1540–1573.

33. Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

34. (a) Knapp, S.; Myers, D. S. *J. Org. Chem.* **2002**, *67*, 2995–2999. (b) Rojas, V.; Carreras, J.; Corzana, F.; Avenoza, A.; Busto, J. H.; Peregrina, J. M. *Carbohydr. Res.* **2013**, *373*, 1–8.

35. (a) Tasdelen, M. A.; Yagci, Y. *Angew. Chem. Int. Ed.* **2013**, *52*, 5930–5938. (b) Saada, M.-C.; Ombouma, J.; Montero, J.-L.; Supuran, C. T.; Winum, J.-Y. *Chem. Commun.* **2013**, *49*, 5699–5701. (c) Dondoni, A.; Marra, A. *Chem. Soc. Rev.* **2012**, *41*, 573–586.

36. (a) Triola, G.; Brunsveld, L.; Waldmann, H. *J. Org. Chem.* **2008**, *73*, 3646–3649. (b) Merbouh, N.; Wallner, F. K.; Cociorva, O. M.; Seeberger, P. H. *Org. Lett.* **2007**, *9*, 651–653.

37. Povie, G.; Tran, A. T.; Bonnaff, D.; Habegger, J.; Hu, Z.; Le Narvor, C.; Renaud, P. *Angew. Chem. Int. Ed.* **2014**, *53*, 3894–3898.

38. (a) Tadd, A. C.; Meinander, K.; Luthman, K.; Wallén, E. A. A. *J. Org. Chem.* **2011**, *76*, 673–675. (b) O'Donnell, M. J.; Delgado, F. *Tetrahedron* **2001**, *57*, 6641–6650.

39. Calce, E.; Leone, M.; Monfregola, L.; De Luca, S. *Org. Lett.* **2013**, *15*, 5354–357.

40. Knerr, P. J.; Tzekou, A.; Ricklin, D.; Qu, H.; Chen, H.; van der Donk, W. A.; Lambris, J. D. *ACS Chem. Biol.* **2011**, *6*, 753–760.

41. Knerr, P. J.; van der Donk, W. A. *J. Am. Chem. Soc.* **2012**, *134*, 7648–7651.

42. Pearlman, D. *J. Biomol. NMR* **1994**, *4*, 1–16.

43. 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.

44. 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.; Peregrina, J. M.; Hurtado-Guerrero, R.; Corzana, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 9830–9834.

45. Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 519–538.

46. (a) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. *J. Am. Chem. Soc.* **2002**, *124*, 9833–

9844. (b) Live, D. H.; Kumar, R. A.; Beebe, X.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 12759–12761. (c) O'Connor, S. E.; Imperiali, B. *Chem. Biol.* **1998**, *5*, 427–437.

47. Kinarsky, L.; Suryanarayanan, G.; Prakash, O.; Paulsen, H.; Clausen, H.; Hanisch, F.-G.; Hollingsworth, M. A.; Sherman, S. *Glycobiology* **2003**, *13*, 929–939.

48. Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. *Chem. Eur. J.* **2006**, *12*, 4981–4993.

49. 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.

50. 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.

51. Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-Gónzalez, J.; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ACS Chem. Biol.* **2015**, *10*, 747−756.

52. Matsushita, T.; Ohyabu, N.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Hinou, H.; Nishimura, S.-I. *Biochemistry* **2013**, *52*, 402–414.

53. Mazeau, K.; Tvaroska, I. *Carbohydr. Res.* **1992**, *225*, 27–41.

54. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12 (2012).

55. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

56. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

57. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

58. Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

59. Kiyohara, K.; Gubbins, K.; Panagiotopoulos, A. *Mol. Phys.* **1998**, *94*, 803–808.

60. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

61. Kabsch, W. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.

62. (a) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242. (b) Collaborative Computational Project, Number 4 *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760–763.

63. Emsley, P.; Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.

64. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367.

65. Laskowski, R. A.; Moss, D. S.; Thornton, J. M. *J. Mol. Biol.* **1993**, *231*, 1049−1067.