# Expanding the Conformational Landscape of Minimalistic Tripeptides by Their *O*-Glycosylation

Alexandra Brito,[a-c]‡ Dhwanit Dave,[c-e]‡ Ayala Lampel,[c]† Vânia I. B. Castro,[a,b] Daniela Kroiss,[c-e] Rui L. Reis,[a,b] Tell Tuttle,[f] Rein V. Ulijn,[c-e, g]* Ricardo A. Pires[a,b] and Iva Pashkuleva[a,b]*

[a]3B's Research Group, I3Bs - Research Institute on Biomaterials, Biodegradables and Biomimetics, University of Minho, Headquarters of the European Institute of Excellence on Tissue Engineering and Regenerative Medicine, AvePark, Parque de Ciência e Tecnologia, Zona Industrial da Gandra, 4805-017 Barco, Guimarães, Portugal; [b]ICVS/3Bs - PT Government Associate Laboratory, Braga/Guimarães, Portugal; [c]Advanced Science Research Center (ASRC) at the Graduate Center, City University of New York (CUNY), 85 St Nicholas Terrace, New York, New York 10031, USA; [d]Department of Chemistry, Hunter College, City University of New York, 695 Park Avenue, New York 10065, USA; [e]Ph.D. program in Chemistry, The Graduate Center of the City University of New York, New York 10016, USA; [f]Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK; [g]Ph.D. program in Biochemistry, The Graduate Center of the City University of New York, New York 10016, USA

*Supporting Information Placeholder*

**ABSTRACT:** We report on the supramolecular self-assembly of tripeptides and their *O*-glycosylated analogues, in which the carbohydrate moiety is coupled to a central serine or threonine flanked by phenylalanine residues. The substitution of serine with threonine introduces differential side-chain interactions, which results in the formation of aggregates with different morphology. *O*-glycosylation decreases the aggregation propensity due to rebalancing of the π interactions. The glycopeptides form aggregates with reduced stiffness but increased thermal stability. Our results demonstrate that the designed minimalistic glycopeptides retain critical functional features of glycoproteins and therefore are promising tools toward elucidation of molecular mechanisms involved in glycoproteins interactome. They can also serve as an inspiration for the design of functional glycopeptide-based biomaterials.

## INTRODUCTION

Protein aggregation is a supramolecular process often associated with pathological conditions.[1-2] The propensity of a protein to aggregate is primarily coded by the intrinsic properties of the amino acids sequence but also depends on multiple contributing factors from the crowded cellular milieu and post-translational modifications, *e.g.* glycosylation.[1, 3-4] Because even the simplest protein is typically composed of hundreds of amino acids, the experimental study and computational modelling of this process is challenging due to the associated combinatorial complexity.[5]

Some years ago, Gazit proposed a reductionist biodesign, which uses intermolecular self-assembly of minimalistic (less than five amino acids) peptide sequences that can code specific protein bioinformation and transfer it to the assembled system (Chart 1A).[6-7] Such molecular biomimetics are simpler in composition, thus, allowing rational and systematic experimental and computational studies to establish connections between the peptide sequence and supramolecular functionality.[5, 8-9] Moreover, their simplicity makes them attractive candidates as building blocks for supramolecular materials with designed functions, which may be useful for a variety of biomedical and technological applications.[9-11]

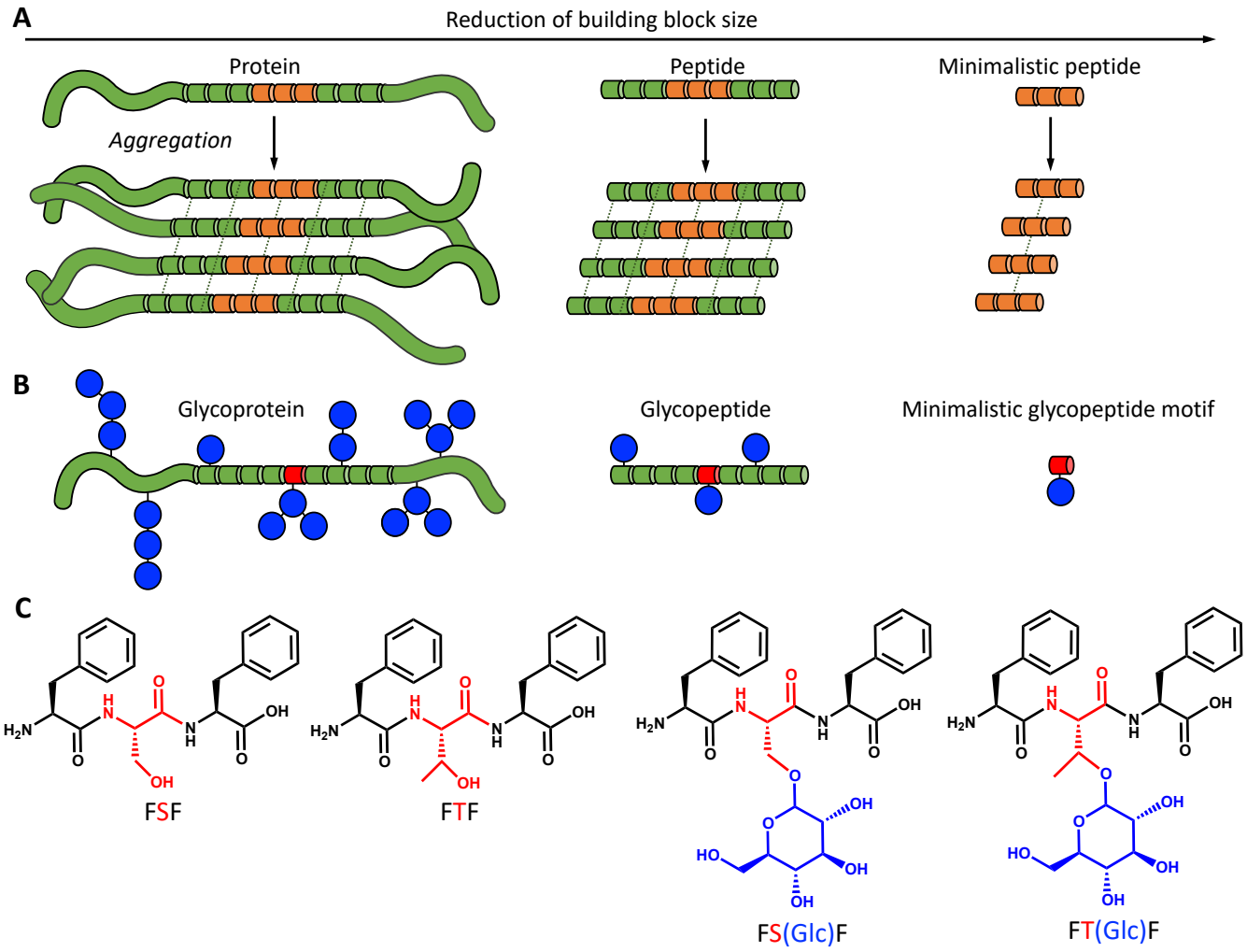
The main rationale of this study is to demonstrate that the reductionist approach proposed for proteins is extendable to glycoproteins (Chart 1B), *i.e.* that short glycotripeptides can be used as simplified analogues of complex *O*-glycoproteins to study and model molecular mechanisms of fundamental properties such as conformational changes and aggregation, and the obtained insights can be applied to rationally modify properties of supramolecular materials based on these motifs.

Glycosylation is a common post-translational modification that effectively enriches the protein repertoire beyond the bioactivities coded by the amino acids sequence and alter the energy landscape associated with the protein aggregation.[4, 12-15] However, the exact mechanism of this process is poorly understood and mainly based on *in silico* models.[4] In eukaryotic cells, *O*-glycosylation takes place at the endoplasmic reticulum or Golgi, where a monosaccharide (usually *N*-

This is a peer-reviewed, accepted author manuscript of the following article: Brito, A., Dave, D., Lampel, A., Castro, V. I. B., Kroiss, D., Reis, R. L., Tuttle, T., Ulijn, R. V., Pires, R. A., & Pashkuleva, I. (2021). Expanding the conformational landscape of minimalistic tripeptides by their O-glycosylation. Journal of the American Chemical Society , 143(47), 19703-19710. https://doi.org/10.1021/jacs.1c07592

acetylgalactosamine but also fucose and glucose) is coupled to the hydroxyl of serine (S) or threonine (T) of newly synthesized polypeptides.[4, 13, 15] Previous studies have demonstrated that the torsion angle (Ψ) of the glycosidic linkage that determines the orientation of the carbohydrate chain is different for S and T glycopeptides but the consequence of this difference for distinct biological functions is not clear.[16-17] In here, we applied a reductionist approach to study the effect of S *vs* T and their glycosylation on conformer selection and molecular aggregation.

**Chart 1. Schematic presentation of (A) the reductionist approach using short peptides as analogues of proteins in aggregation studies and (B) the herein proposed approach that uses minimalistic glycopeptide motifs; (C) chemical structure of the peptides and glycopeptides used in this study.***




*orange: amino acids involved in the aggregation, green: other amino acids; blue: carbohydrates, red: glycosylated amino acid

## RESULTS AND DISCUSSION

Our design of minimalistic *O*-glycoproteins mimics is based on the simplest *O*-glycoprotein motif[16] - S or T, which are functionalized with glucose (Glc). While our study is focused on a deliberately simple model system, *O*-glucosylation is biologically relevant: it is essential for Notch trafficking/signaling and has been associated with defects in neurogenesis, cardiovascular remodeling, somitogenesis, and aberrant gastrulation.[18-19] To promote the aggregation of this motif, we have also included phenylalanine (F) in the peptide sequence (Chart 1C) because previous experimental and computational approaches have shown that the presence of aromatic amino acids enhances the aggregation propensity of short sequences (*i.e.* sequences with limited number of H-bonding between backbone elements).[5, 9, 20-21]

F has an ability to aggregate alone or when inserted into short (di- and tri-) peptides.[22-24] Tripeptides with flanked aromatic amino acids, *e.g.* FXF, where X is a hydrophilic amino acid, adopt conformations that allow

intramolecular stacking of the two aromatic rings thus, exposing the central amino acid to water.[25] Such sequences self-assemble in water due to formation of aromatic zippers and a hydrophobic collapse.[23-24, 26] In our molecular design the hydrophilic amino acid is also introduced in the middle of the peptide sequence and thus, the *O*-glycosylation of the short peptide chains at S or T generates minimalistic *O*-glycopeptides, which differ from previously described self-assembling glycopeptides that are end-on glycosylated.[27-29]

We used all-atom molecular dynamics simulations (MDS) with explicit water to investigate the conformational space of the designed peptides and their glycosylated analogs. In agreement with previous studies that include the Aromatic-X-Aromatic motif,[24-26] we found that in the predominant conformations of FSF and FTF the aromatic amino acids adopt arrangement that allow their intramolecular stacking (Fig. 1A).

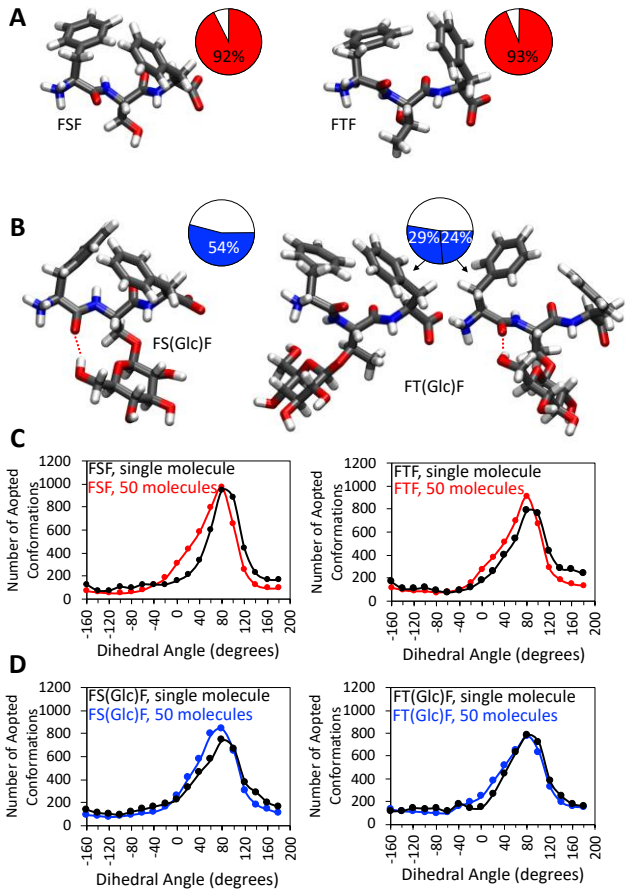



Figure 1. Molecular dynamics analyses showing (A, B) the representative central structures of the largest clusters obtained for (A) model tripeptides and (B) their *O*-glycosylated analogues (the pie-charts and percentages show the fraction of 5000 structures adopting these conformations; a root mean square deviation (RMSD) linkage cutoff of 0.1 nm was used for the analyses; supplementary data are provided in Fig. S17); (C, D) Comparative dihedral analysis (CZ(F)-CA(F)-CA(F)-CZ(F)) of single molecule and 50 molecules simulations of (C) tripeptides and (D) glycotripeptides. Data for the alpha anomers are presented in Fig. S28.

The F/F dihedral distribution showed similar peaks in the 90º region for FSF and FTF corresponding to intramolecular stacking interactions and was suggestive of molecular reorganization for supramolecular self-assembly (Fig. 1C, black). When MDS was applied to 50 molecules instead of one, we observed a small shift in the dihedrals' distribution to lower angles for both peptides (Fig. 1C, red *vs* black). This shift is indicative of reorganization of the F/F intramolecular stacks to allow formation of intermolecular aromatic zippers (Fig. 2A *vs* 2B), *i.e.* the stacking of interdigitated F side chains from cross strand peptides leading to self-assembly.

MDS of the glycopeptides revealed that the glycosylation widens the conformational landscape (Figs. 1B, S25, S26 and S27). The flexibility of the glycosidic bond provides additional modes of interactions with contributions from H-bonding, CH-π and electrostatic salt-bridge type interactions[30], leading to different conformer distributions. A comparison of the F/F dihedral distributions in the tripeptides and the glycopeptides (Fig. 1D) reveals a reduction in the mean dihedral angle, indicating a wider distribution of glycopeptides conformations that are stabilized by non π-π type interactions. Of note, the data obtained for the alpha and beta anomers were very similar (Fig. S28). Ramachandran plots of these (glyco)peptides (Fig. S27) showed that the backbone conformations are similar to the reported for FXF peptides[26] and a conformational diversity arises from side-chains and their glycosylation. Additionally, MDS showed that Glc anomers are involved in different intramolecular H-bonding, *e.g.* the alpha anomer forms H-bond with the carboxylate oxygen, whereas in beta stereochemistry the Glc interacts with the amide oxygen in FS(Glc)F (Fig. S28), thus, influencing the glycopeptides conformation.

The computational results were verified experimentally. The aggregation of the tripeptides and their glycosylated analogs was studied in water at 40 mM, *i.e.* above the critical aggregation concentration, giving rise to transparent viscous liquids. The S to T exchange in these peptides introduces a methyl group into the structure, which affects the morphology of the generated assemblies: FSF forms nano-tapes while nanofibrils are observed for FTF (Figs. 3, S33). These results are consistent with previous observations on S/T substitution in self-assembling Fmoc-dipeptides, where planar structures were observed for Fmoc-SF-OMe and an extended network of twisted fibers was obtained for Fmoc-TF-OMe.[31]

The MDS data provided insights in the supramolecular interactions dictating the organization within these structures. A substantial decrease of the solvent-accessible surface area (SASA) for the aromatic F groups over time (Fig. 2E, black) was observed, indicating that, as expected, these groups participate in the self-assembly and are mostly buried in the core of the assembled structures (Fig. 2D, SI movies 1,2). We also observed a change in SASA for the amino acid residues during the self-assembly process: SASA reduced less for S and T compared to F (Fig. 2E, red), confirming greater exposure of these amino acids on the surface of the peptides' assemblies (Fig. 2D, SI movies 1,2). A comparison between S and T revealed differences: S has higher propensity to form hydrogen bonds with water (Table 1, Fig. S30A), while T has a higher tendency to interact with phenyl rings *via* CH-π interactions (Table 1, Figs. 2B, S31A radial distribution peak at 4.5 Å).

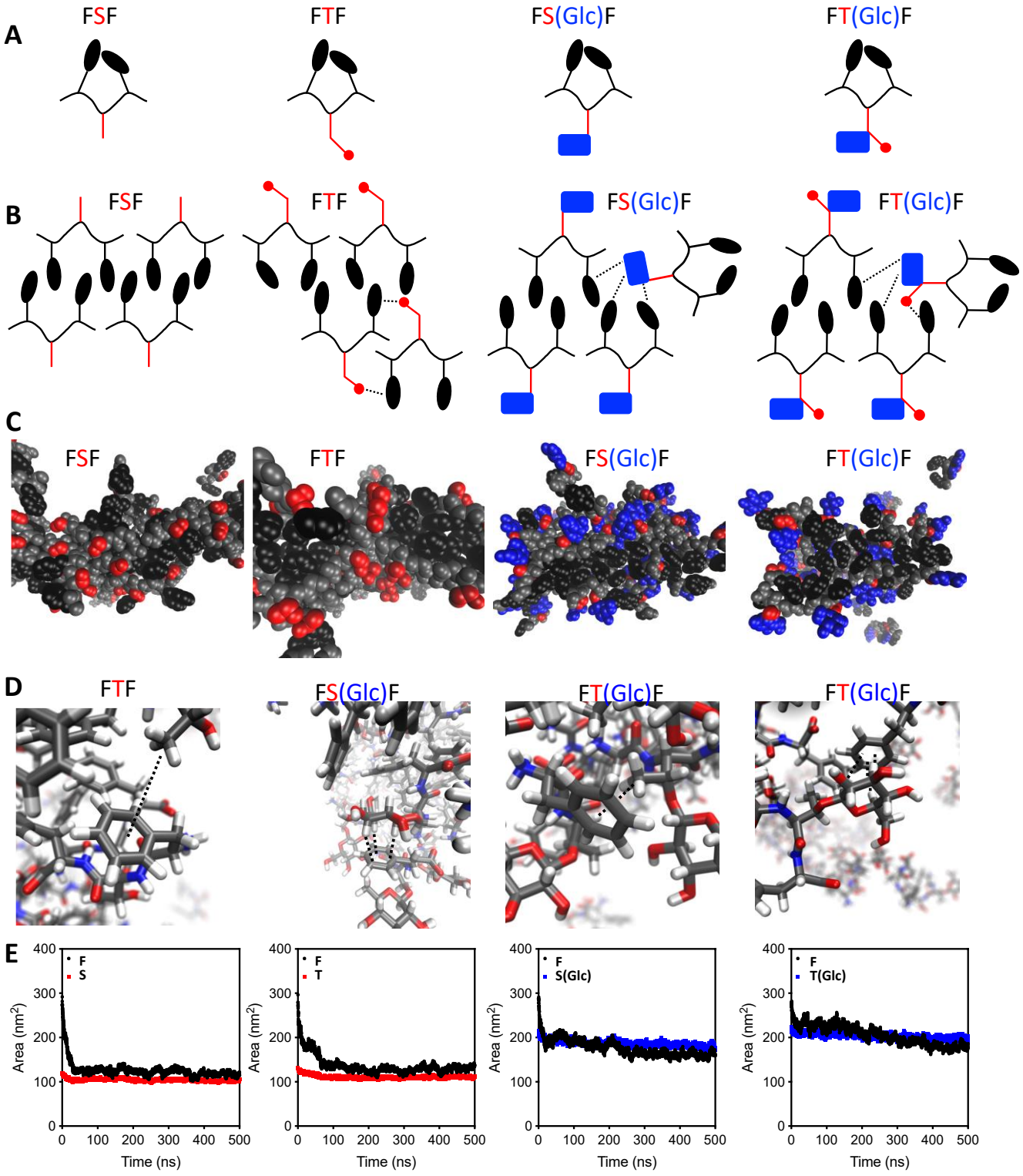



Figure 2. Schematic presentation of (A) predominant conformations of single (glyco)tripeptides based on computational modeling and (B) the CH-π interactions (black dotted lines) involved in the formation of aggregates as shown by the calculated probability g(r); F are shown in black, S/T in red (the red circles represent the methyl group of T) and glucose in blue. (C) Representative van der Waals structures of the (glyco)peptide aggregates observed during MDS: F forms the core of the aggregates with the backbone shown in grey while the polar amino acid (S/T) and glucose are primarily water exposed with some incorporation into aggregates due to (D) CH-π interactions (black dotted lines). (E) Solvent-accessible surface area (SASA) analysis. More details are provided in Table 1 and the SI.

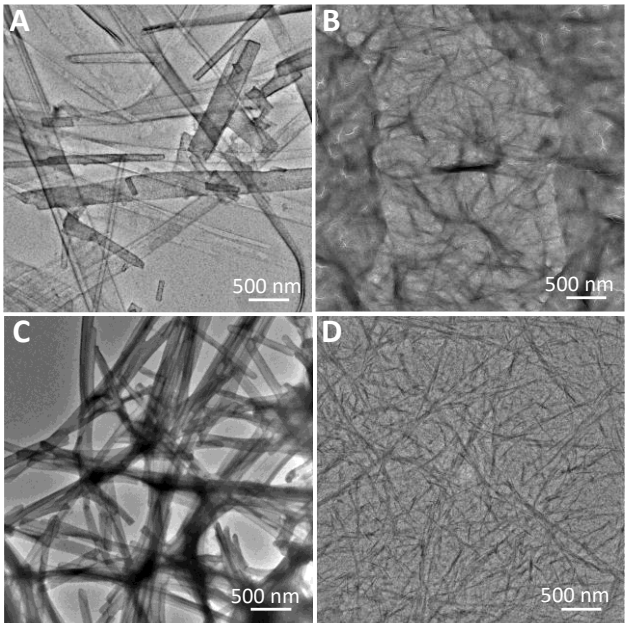



Figure 3. Transmission electron microscopy (TEM) images of (A) FSF, (B) FTF, (C) FS(Glc)F, and (D) FT(Glc)F assemblies formed in water (40 mM, room temperature, 24 hrs).

Despite the greater hydrophobicity of FTF compared to FSF due to the additional methyl group, the MDS showed a counterintuitive reduction in aggregation propensity for FTF. This result indicates possible disruption of the primarily π-π driven aggregation by formation of CH-π interactions in FTF aggregates (Fig. 2D), *i.e.* the methyl groups protruding from the FTF peptide chains disturb the assembly of the aromatic zippers (Fig. 2B, FSF *vs* FTF) and can explain the different morphology of FSF and FTF assemblies.

**Table 1. Computational data for aggregation propensity (50 molecules simulations, average of three 500 ns runs), number of hydrogen bonds formed between the aggregates and the solvent, the probability (g(r)) of CH-π interactions between the aromatic F and S, T, and Glc and β-sheet like H-bonds in the aggregates of the studied (glyco)peptides.**

| | Aggregation propensity* | H-bonds with water | CH-π interactions (g(r))** | β-sheet like H-bonds |
|---|---|---|---|---|
| FSF | 2.23±0.26 | 313±22 | 2.8±0.3 | 28±2 |
| FTF | 2.18±0.05 | 277±20 | 4.8±1.7 | 22±6 |
| FS(βGlc)F | 1.90±0.14 | 776±37 | 3.5±0.3 | 10±4 |
| FS(αGlc)F | 1.82±0.17 | 740±38 | 4.1±0.2 | 14±7 |
| FT(βGlc)F | 1.62±0.08 | 784±37 | 5.6±1.1 | 4±2 |
| FT(αGlc)F | 1.76±0.06 | 797±30 | 6.6±0.4 | 8±1 |

*SASAinitial/SASAfinal; **peak g(r) was measured between heavy atoms of the polar amino acid and the aromatic side chain.

Circular dichroism (CD) data further supported the MDS data. The CD spectra of the peptides have an intensive, positive signal at ~220 nm for the n-π* transition (Fig. 4B).[32-33] In the FTF spectrum there is an additional positive peak at ~200 nm that was assigned to π-π* transition and confirms that the methyl group of T affects the relative spatial orientation of F and thus, the supramolecular interactions and the aggregation process.

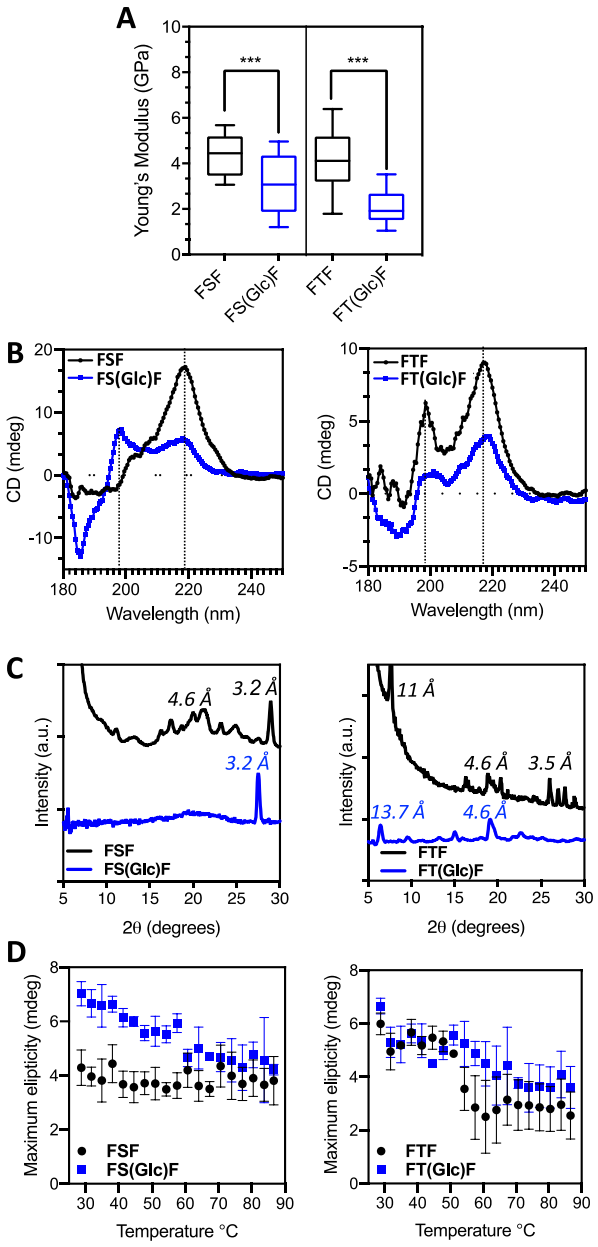



Figure 4. Effect of glycosylation on the aggregation: (A) Young's modulus of the aggregates determined by atomic force microscopy (***p<0.0005); (B) circular dichroism (CD) spectra; (C) X-ray diffraction patterns; and (D)

thermal stability of the generated aggregates obtained from the CD spectra at different temperatures.

The *in silico* models showed lower aggregation propensity for *O*-glycotripeptides (Table 1). Experimentally, we used mixtures of alpha and beta anomers (alpha:beta was 58:42 for FS(Glc)F and 47:53 FT(Glc)F, Figs. S10 and S18). We observed that the glycosylation affected the fiber diameter as observed by TEM (Fig. 3), as well as the mechanical properties of the aggregates (Fig. 4A), while the overall morphologies between peptides and the respective glycosylated analogues appeared similar.

Native *O*-glycoproteins have a high capacity to capture water, which is essential for their viscoelastic properties and physiological functions. At the molecular level, protein glycosylation usually causes higher hydration that can lead to enhanced steric bulk, *i.e.* steric hindrance around the protein backbone, which can prevent aggregation, including β-sheet formation.[34] Thus, the decreased Young's modulus of *O*-glycopeptides (Fig. 4A, 2-fold as compared with the non-glycosylated tripeptides) is likely due to the increased hydration capacity and/or structural changes caused by the conformational distortions and supramolecular forces, such as H-bonding and π-interactions, impaired by the introduced Glc.[30, 35] As discussed, the carbohydrate moiety is predominantly exposed on the surface of the assemblies contributing to their increased hydration when compared with the respective tripeptides. The MDS showed that as expected, the glycosylated peptides have higher SASA (Fig. 2E) combined with an enhanced propensity to form hydrogen bonding with water (Table 1, Fig. S29B,C). However, the results also indicated the presence of carbohydrate moieties in the core of the aggregates (Fig. 2C-D, SI movies 3,4), suggesting their involvement in the aggregation process (Fig. 2B), beyond simple hydration. Indeed, the glycosylation led to a ~1.2-fold increase of the propensity to form CH-π interactions (Table 1), which in turn affects the n-π and π-π interactions. This rebalancing of the π-interactions was confirmed by the CD spectra (Fig. 4B), where a decrease of the 220 nm signal intensity for both glycopeptides was observed. The introduction of the carbohydrate group at S has a similar stereochemical effect as the T's methyl group (Fig. 2B, FS(Glc)F *vs* FTF) shown by the appearance of the π-π* signal in the FS(Glc)F spectrum (FSF *vs* FS(Glc)F in Fig. 4B).

X-ray diffraction (XRD, Fig. 4C) corroborated the rebalance of the supramolecular interactions upon *O*-glycosylation of the tripeptides that was observed by CD: in the case of FSF the peak associated with ß-sheet formation (4.6 Å) vanishes for the respective glycosylated analogues and this change can be explained with the above-mentioned steric bulk, while in the case of FTF, the peak associated with aromatic interactions (3.5 Å) disappears upon glycosylation. MDS analysis showed a reduction of H-bonds between the peptide backbones (β-sheet like interactions, Table 1) upon glycosylation. Moreover, a significant decrease of the FT(Glc)F *vs* FTF aggregation propensity was also observed.

Together these data confirm the disruption of the aromatic zippers in the glycotripeptides due to the formation of stronger CH-π interactions and explain the disappearance of the aromatic peak. In case of FS(Glc)F, the aromatic interactions are preserved as evidenced by the smaller decrease in the aggregation propensity but the significant reduction in the backbone H-bonds explains the disappearance of the beta-sheet peak upon glycosylation.

The performed *O*-glycosylation also affected the thermal stability of the aggregates (Fig. 4D, Table 2).[36] Upon heating, the aggregates of the glycosylated FS(Glc)F were more stable with a melting temperature that was 27 °C higher compared to that of FSF. The difference was less pronounced for the FT(Glc)F/FTF couple (~10 °C). These results agree with previous studies with glycoproteins showing that the glycosylation generally improves the thermal stability of the proteins and the magnitude of this effect depends on the size of the carbohydrate chain, the position of glycosylation and the protein crystallinity.[37-39]

**Table 2. Midpoint transition ($T_M$), enthalpy (ΔH) and heat capacity change (ΔCp) of the disassembly calculated from the CD spectra (signal at ~220nm) at different temperatures.**

| | $T_M$, °C | ΔH, kJ mol$^{-1}$ | ΔCp, J °C$^{-1}$ |
|---|---|---|---|
| FSF | 52.1±3.1 | 18.7±2.3 | 84.9±22.9 |
| FTF | 57.2±3.1 | 16.7±2.9 | 100.5±32.3 |
| FS(Glc)F | 70.0±3.6 | 28.3±3.5 | 165.3±28.2 |
| FT(Glc)F | 66.2±3.0 | 20.8±2.8 | 122.2±29.1 |

Finally, we also studied the aggregation of mixtures of peptides and the respective glycosylated analogues at different molar ratios to simulate a scenario in which proteins and glycoproteins co-exist. The morphology of the assemblies obtained from the mixtures was different from the single-component systems and we observed the formation of entangled nanofibers for all mixtures (Fig. 5A, B). The Young's modulus gradually decreased upon addition of the glycopeptides, reached a minimum at ratio 1:1 and further enrichment of the mixtures with *O*-glycopeptides did not affect the modulus significantly (Fig. 5C, D). These results are consistent with co-assembly and indicate that in a crowded environment glycosylation affects not only the aggregation of the protein to which the

carbohydrate unit is bound but also to the close neighbors.

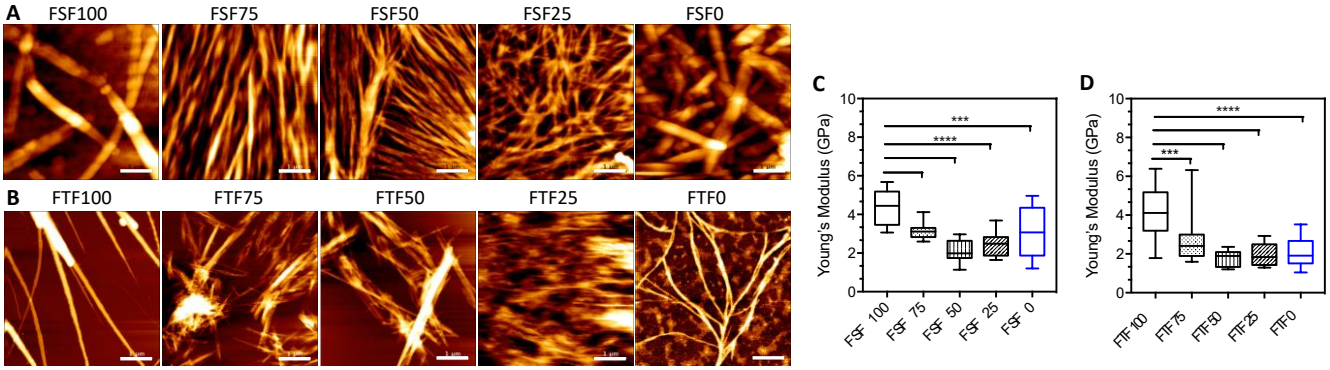



Figure 5. Aggregation of peptide/glycopeptide mixtures at different ratios: (A, B) Atomic force microscopy (AFM) images of the assemblies obtained from mixtures at different ratios and (C, D) Young's modulus for these assemblies measured by AFM. FXF100 (FXF:FX(Glc)F=1:0); FXF75 (FXF:FX(Glc)F=3:1); FXF50 (FXF:FX(Glc)F=1:1); FXF25 (FXF:FX(Glc)F=1:3); FXF0 (FXF:FX(Glc)F=0:1). Scale bares: 1 μm.

## CONCLUSIONS

In summary, we showed a distinct role of F, S/T, and Glc in the glycopeptides supramolecular interactome and consequently in the characteristics of the generated aggregates. The introduced glycosylation clearly influenced the aggregation, giving rise to enhanced disorder and dynamics in the assembled structures, due to the introduction of CH-π interactions. Such interactions are often challenging to quantify because they are usually inaccessible in native glycoproteins but they are crucial for protein synthesis, trafficking, and function.

## ASSOCIATED CONTENT

**Supporting Information**. Description of the used experimental procedures; Data on (glyco)peptides characterization: $^{1}$H NMR, $^{13}$C NMR, MS, HPLC; Supplementary data obtained from the molecular dynamic simulations; Supplementary microscopy images (TEM and SEM). SI movies show representative 500ns molecular dynamics trajectories of the self-assembly of 50 (glyco)peptide molecules in a periodic cubic box starting from a random dispersed configuration with the F residues colored grey, the polar amino acids, S (movie 1) and T (movie 2) colored red and the beta-D-Glc (movies 3 and 4) colored blue. Explicit TIP3 water molecules are not shown for clarity.

## AUTHOR INFORMATION

### Corresponding Author

Iva Pashkuleva: e-mail pashkuleva@i3bs.uminho.pt
Rein V Ulijn: e-mail rulijn@gc.cuny.edu

### Present Addresses

†Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 9978, Israel.

### Author Contributions

‡These authors contributed equally.

### Funding Sources

No competing financial interests have been declared.

## ACKNOWLEDGMENT

We acknowledge the EU's H2020 program (Forecast 668983) and the Portuguese FCT (BD/113794/2015; PTDC/BTM-MAT/28327/2017 CARDIOHEAL; M-ERA-NET2/0001/2016 INCIPIT) for the financial support. Part of this research was supported by National Science Foundation (NSF) grant CHE-1808143 and a grant of computer time from the City University of New York High Performance Computing Center under NSF grants CNS-0855217, CNS-0958379 and ACI-1126113. DD would like to thank Maithreyi Ramakrishnan for provision of a dihedral analysis script and Mateusz Marianski for discussions.

## REFERENCES

1. Vendruscolo, M.; Zurdo, J.; MacPhee, C. E.; Dobson, C. M., Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Philos T R Soc A* **2003,** *361* (1807), 1205-1222.
2. Gazit, E., The "Correctly folded" state of proteins: Is it a metastable state. *Angewandte Chemie-International Edition* **2002,** *41* (2), 257-259.
3. White, D. A.; Buell, A. K.; Knowles, T. P. J.; Welland, M. E.; Dobson, C. M., Protein Aggregation in Crowded Environments. *Journal of the American Chemical Society* **2010,** *132* (14), 5170-5175.
4. Shental-Bechor, D.; Levy, Y., Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (24), 8256-8261.
5. Pappas, C. G.; Shafi, R.; Sasselli, I. R.; Siccardi, H.; Wang, T.; Narang, V.; Abzalimov, R.; Wijerathne, N.; Ulijn, R. V., Dynamic peptide libraries for the discovery of supramolecular nanomaterials. *Nature Nanotechnology* **2016,** *11* (11), 960-967.

6. Brahmachari, S.; Arnon, Z. A.; Frydman-Marom, A.; Gazit, E.; Adler-Abramoyich, L., Diphenylalanine as a Reductionist Model for the Mechanistic Characterization of beta-Amyloid Modulators. *Acs Nano* **2017,** *11* (6), 5960-5969.

7. Gazit, E., Reductionist Approach in Peptide-Based Nanotechnology. In *Annual Review of Biochemistry*, Kornberg, R. D., Ed. 2018; Vol. 87, pp 533-553.

8. Yoshida, M.; Hinkley, T.; Tsuda, S.; Abul-Haija, Y. M.; McBurney, R. T.; Kulikov, V.; Mathieson, J. S.; Reyes, S. G.; Castro, M. D.; Cronin, L., Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides. *Chem-Us* **2018,** *4* (3), 533-543.

9. Frederix, P. W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T., Exploring the sequence space for (tri-) peptide self-assembly to design and discover. *Nature chemistry* **2015,** *7* (1), 30-37.

10. Clemons, T. D.; Stupp, S. I., Design of materials with supramolecular polymers. *Progress in Polymer Science* **2020,** *111*, 101310.

11. Levin, A.; Hakala, T. A.; Schnaider, L.; Bernardes, G. J. L.; Gazit, E.; Knowles, T. P. J., Biomimetic peptide self-assembly for functional materials. *Nature Reviews Chemistry* **2020,** *4* (11), 615-634.

12. da Costa, D. S.; Reis, R. L.; Pashkuleva, I., Sulfation of Glycosaminoglycans and Its Implications in Human Health and Disorders. *Annual Review of Biomedical Engineering* **2017,** *19*, 1-26.

13. Xu, C. C.; Ng, D. T. W., Glycosylation-directed quality control of protein folding. *Nature Reviews Molecular Cell Biology* **2015,** *16* (12), 742-752.

14. Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884-890.

15. Pratt, M. R.; Bertozzi, C. R., Synthetic glycopeptides and glycoproteins as tools for biology. *Chemical Society Reviews* **2005,** *34* (1), 58-68.

16. Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; de Luis, M. G.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A., Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *Journal of the American Chemical Society* **2007,** *129* (30), 9458-9467.

17. 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., Principles of mucin architecture: Structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains. *Journal of the American Chemical Society* **2002,** *124* (33), 9833-9844.

18. Yu, H. J.; Takeuchi, H., Protein O-glucosylation: another essential role of glucose in biology. *Current Opinion in Structural Biology* **2019,** *56*, 64-71.

19. Acar, M.; Jafar-Nejad, H.; Takeuchi, H.; Rajan, A.; Ibrani, D.; Rana, N. A.; Pan, H.; Haltiwanger, R. S.; Bellen, H. J., Rumi is a CAP10 domain glycosyltransferase that modifies notch and is required for notch signaling. *Cell* **2008,** *132* (2), 247-258.

20. Pawar, A. P.; DuBay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M., Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *Journal of Molecular Biology* **2005,** *350* (2), 379-392.

21. Gazit, E., A possible role for pi-stacking in the self-assembly of amyloid fibrils. *Faseb Journal* **2002,** *16* (1), 77-83.

22. Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caflisch, A.; Frenkel, D.; Gazit, E., Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nature Chemical Biology* **2012,** *8* (8), 701-706.

23. Shlomo, Z.; Vinod, T. P.; Jelinek, R.; Rapaport, H., Stacking interactions by two Phe side chains stabilize and orient assemblies of even the minimal amphiphilic beta-sheet motif. *Chemical Communications* **2015,** *51* (15), 3154-3157.

24. Vargiu, A. V.; Iglesias, D.; Styan, K. E.; Waddington, L. J.; Easton, C. D.; Marchesan, S., Design of a hydrophobic tripeptide that self-assembles into amphiphilic superstructures forming a hydrogel biomaterial. *Chemical Communications* **2016,** *52* (35), 5912-5915.

25. Lampel, A.; McPhee, S. A.; Park, H. A.; Scott, G. G.; Humagain, S.; Hekstra, D. R.; Yoo, B.; Frederix, P.; Li, T. D.; Abzalimov, R. R.; Greenbaum, S. G.; Tuttle, T.; Hu, C. H.; Bettinger, C. J.; Ulijn, R. V., Polymeric peptide pigments with sequence-encoded properties. *Science* **2017,** *356* (6342), 1064-1068.

26. Garcia, A. M.; Iglesias, D.; Parisi, E.; Styan, K. E.; Waddington, L. J.; Deganutti, C.; De Zorzi, R.; Grassi, M.; Melchionna, M.; Vargiu, A. V.; Marchesan, S., Chirality Effects on Peptide Self-Assembly Unraveled from Molecules to Materials. *Chem-Us* **2018,** *4* (8), 1862-1876.

27. Restuccia, A.; Seroski, D. T.; Kelley, K. L.; O'Bryan, C. S.; Kurian, J. J.; Knox, K. R.; Farhadi, S. A.; Angelini, T. E.; Hudalla, G. A., Hierarchical self-assembly and emergent function of densely glycosylated peptide nanofibers. *Communications Chemistry* **2019,** *2*, 53.

28. Lee, S. S.; Fyrner, T.; Chen, F.; Alvarez, Z.; Sleep, E.; Chun, D. S.; Weiner, J. A.; Cook, R. W.; Freshman, R. D.; Schallmo, M. S.; Katchko, K. M.; Schneider, A. D.; Smith, J. T.; Yun, C. W.; Singh, G.; Hashmi, S. Z.; McClendon, M. T.; Yu, Z. L.; Stock, S. R.; Hsu, W. K.; Hsu, E. L.; Stupp, S. I., Sulfated glycopeptide nanostructures for multipotent protein activation. *Nature Nanotechnology* **2017,** *12* (8), 821-829.

29. Liu, R. Y.; Zhang, R.; Li, L.; Kochovski, Z.; Yao, L. T.; Nieh, M. P.; Lu, Y.; Shi, T. F.; Chen, G. S., A Comprehensive Landscape for Fibril Association Behaviors Encoded Synergistically by Saccharides and Peptides. *Journal of the American Chemical Society* **2021,** *143* (17), 6622-6633.

30. Brito, A.; Kassem, S.; Reis, R. L.; Ulijn, R. V.; Pires, R. A.; Pashkuleva, I., Carbohydrate amphiphiles for supramolecular biomaterials: Design, self-assembly, and applications. *Chem-Us* **2021**, *in press*, DOI: 10.1016/j.chempr.2021.04.011.

31. Hughes, M.; Frederix, P. W. J. M.; Raeburn, J.; Birchall, L. S.; Sadownik, J.; Coomer, F. C.; Lin, I. H.; Cussen, E. J.; Hunt, N. T.; Tuttle, T.; Webb, S. J.; Adams, D. J.; Ulijn, R. V., Sequence/structure relationships in aromatic dipeptide hydrogels formed under thermodynamic control by enzyme-assisted self-assembly. *Soft Matter* **2012,** *8* (20), 5595-5602.

32. Brito, A.; Abul-Haija, Y. M.; da Costa, D. S.; Novoa-Carballal, R.; Reis, R. L.; Ulijn, R. V.; Pires, R. A.; Pashkuleva, I., Minimalistic supramolecular proteoglycan mimics by co-assembly of aromatic peptide and carbohydrate amphiphiles. *Chemical Science* **2019,** *10* (8), 2385-2390.

33. Naganagowda, G. A.; Gururaja, T. L.; Satyanarayana, J.; Levine, M. J., NMR analysis of human salivary mucin (MUC7) derived O-linked model glycopeptides: comparison of structural features and carbohydrate-peptide interactions. *Journal of Peptide Research* **1999,** *54* (4), 290-310.

34. Kramer, J. R.; Onoa, B.; Bustamante, C.; Bertozzi, C. R., Chemically tunable mucin chimeras assembled on living cells. *Proceedings of the National Academy of Sciences of the United States of America* **2015,** *112* (41), 12574-12579.

35. Birchall, L. S.; Roy, S.; Jayawarna, V.; Hughes, M.; Irvine, E.; Okorogheye, G. T.; Saudi, N.; De Santis, E.; Tuttle, T.; Edwards, A. A.; Ulijn, R. V., Exploiting CH-pi interactions in supramolecular hydrogels of aromatic carbohydrate amphiphiles. *Chemical Science* **2011,** *2* (7), 1349-1355.

36. Greenfield, N. J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* **2006,** *1* (6), 2527-2535.

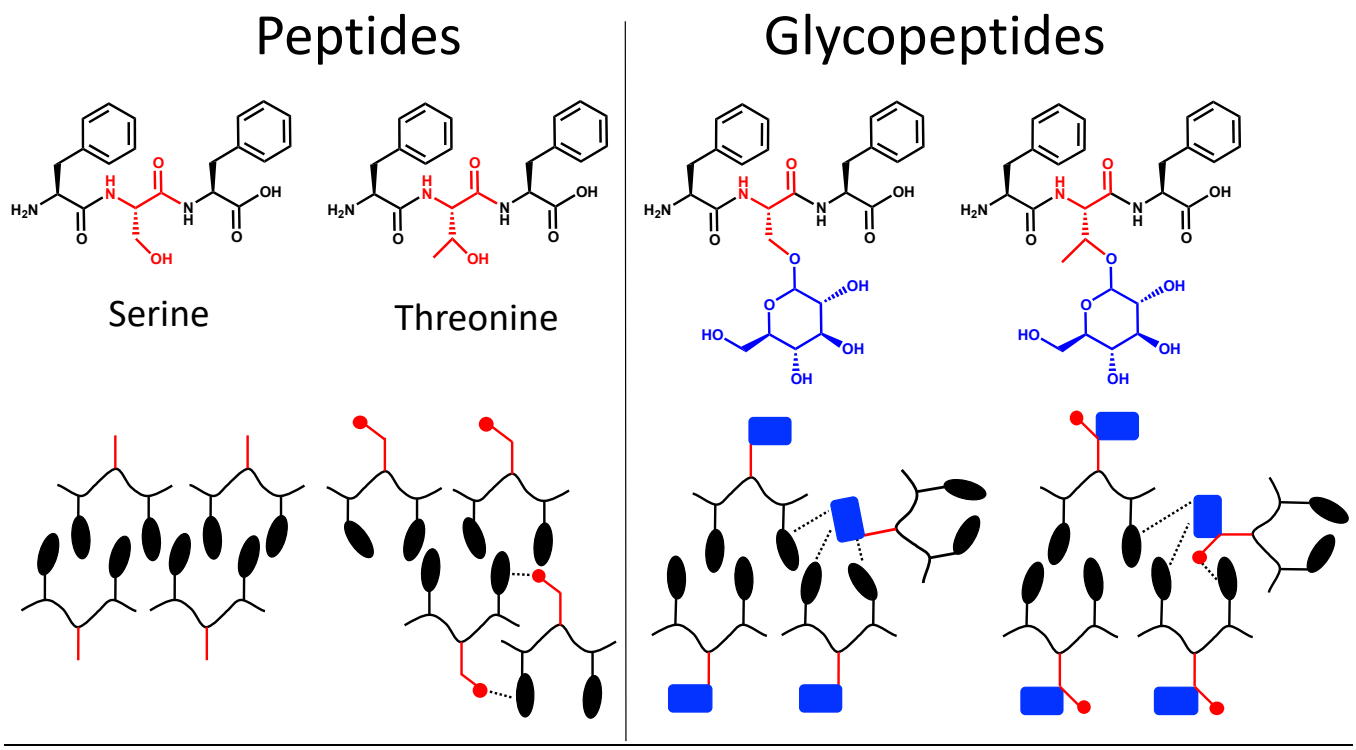
37. Chen, W. T.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C. H.; Dyson, H. J.; Kelly, J. W., Structural and Energetic Basis of Carbohydrate-Aromatic Packing Interactions in Proteins.

*Journal of the American Chemical Society* **2013,** *135* (26), 9877-9884.

38. Wang, C. Q.; Eufemi, M.; Turano, C.; Giartosio, A., Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **1996,** *35* (23), 7299-7307.

39. Sola, R. J.; Al-Azzam, W.; Griebenow, K., Engineering of protein thermodynamic, kinetic, and colloidal stability: Chemical glycosylation with monofunctionally activated glycan. *Biotechnology and Bioengineering* **2006,** *94* (6), 1072-1079.

Table of Contents



# Supporting information

# Expanding the Conformational Landscape of Minimalistic Tripeptides by Their *O*-Glycosylation

Alexandra Brito,[a-c]‡ Dhwanit Dave,[c-e]‡ Ayala Lampel,[c]† Vânia I. B. Castro,[a,b] Daniela Kroiss,[c-e] Rui L. Reis,[a,b] Tell Tuttle,[f] Rein V. Ulijn,[c-e, g]* Ricardo A. Pires[a,b] and Iva Pashkuleva[a,b]*

[a]3B's Research Group, I3Bs - Research Institute on Biomaterials, Biodegradables and Biomimetics, University of Minho, Headquarters of the European Institute of Excellence on Tissue Engineering and Regenerative Medicine, AvePark, Parque de Ciência e Tecnologia, Zona Industrial da Gandra, 4805-017 Barco, Guimarães, Portugal; [b]ICVS/3Bs - PT Government Associate Laboratory, Braga/Guimarães, Portugal; [c]Advanced Science Research Center (ASRC) at the Graduate Center, City University of New York (CUNY), 85 St Nicholas Terrace, New York, New York 10031, USA; [d]Department of Chemistry, Hunter College, City University of New York, 695 Park Avenue, New York 10065, USA ; [e]Ph.D. program in Chemistry, The Graduate Center of the City University of New York, New York 10016, USA; [f]Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK; [g]Ph.D. program in Biochemistry, The Graduate Center of the City University of New York, New York 10016, USA

*Supporting Information Placeholder*

## Experimental procedures

*Synthesis and purification of peptides and glycopeptides:* Amino acids and amide resin were purchased from Novabiochem. Fmoc-S-[β-Glc(OAc)4]-OH and Fmoc-T-[β-Glc(OAc)4]-OH were purchased from Aapptec. N,N,N′,N′-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU), N,N-Diisopropylethylamine (DIPEA), dimethylformamide (DMF), dichloromethane (DCM), trifluoroacetic acid (TFA), diethyl ether, and methanol were purchased from Fisher Scientific. Piperidine, triisopropylsilane (TIS), 3,6-Dioxa-1,8-octane-dithiol (DODT) and sodium methoxide were purchased from Sigma-Aldrich. Peptides (FSF and FTF) and glycopeptides (FS(Glc)F and FT(Glc)F) were synthesized following standard Fmoc solid phase peptide synthesis protocols with DIPEA/HBTU activation. Peptides were cleaved with mixture of TFA/TIS/water (95:2.5:2.5) and precipitated in diethyl ether. The obtained products were dried, resuspended in ultrapure water, and freeze dried. FS(Glc)F and FT(Glc)F were further deacetylated with sodium methoxide and precipitated in methanol. The obtained peptides and glycopeptides were purified by HPLC using a Macherey-Nagel C18 column (250 mm length, 4.6 mm internal diameter and 3 μm particle size) on Dionex P780 HPLC system with UVD170U UV-Vis detector at 282 nm wavelength. To each sample (50 μL in water) was added 950 μL of a mixture water:acetonitrile (50:50) containing 0.1 % of trifluoroacetic acid. Gradient of water:acetonitrile (20:80 to 80:20 for 20 min, flow rate of 1 mL min-1) was used as a mobile phase. Identity and purity of all peptides was confirmed by LC-ESI using Bruker's maXis-II ETD ESI-QqTOF instrument equipped with the Dionex Ulti-mate-3000 LC system. Peptides and glycopeptides with purity above 95 % were used for all studies. NMR analysis was used to confirm the structure of the purified (glyco)peptides. The two anomers of the (glyco)peptides (alpha and beta) are distinguishable by $^1$H-NMR: we have used the signals of the protons at 4.6-4.4 ppm (identified from the HMQC-NMR spectra) to calculate the ratio between the two anomers. At RT, these signals are overlapping themselves and with the signal of $D_2O$ and thus, we performed the analyses at different temperatures to overcome this issue.

*Aggregation of peptides and glycopeptides:* F(Glc)F and FS(Glc)F were dissolved in distilled water to 65 mM stock concentration. Then, diluted to 40 mM final concentration and incubated for 24 h at room temperature. FTF and FSF peptides were dissolved in distilled water to 40 mM and incubated for 24 h. All characterization was performed after this period.

*Atomistic Molecular Dynamics Simulations (MDS)*: Tripeptide structures were generated using Avogadro (v 1.2.0)[1] and the CHARMM36m[2] force-field was used. GROMACS 2020.1[3] implementation of CHARMM[4] was used for the simulations. First, the peptide molecule was centered in a box such that it was 1.5 nm from the edge of the box and this box was solvated with TIP3P water. These systems were then relaxed for 50000 steps using the steepest descent integrator. This box was then subjected to NVT and NPT equilibration for 100 ps each using a 2 fs timestep while holding the peptides with a position restraint. The position restrains were released for the production runs where the systems were equilibrated for 500 ns. The LINCS constraint algorithm was used to constraint bonds and PME was used for electrostatics. Temperature was set to 300 K using the modified Berendsen thermostat (V-rescale) and the pressure coupling was achieved using the Parinello-Rahman algorithm with a reference pressure of 1.0 bar. The glycosylated tripeptide structures were generated using CHARMM-GUI[5-8] and a protocol as the above-described was used for the simulations. Structures were saved every 100 ps leading to 5000 structures per simulation. They were then clustered using gmx cluster with a Root Mean Square Deviation (RMSD of the position of all the (glyco)peptide atoms) cutoff of 0.1 nm. The clusters containing more than 100 structures were used for conformational analysis. Self-assembly simulations were repeated with similar protocols as described above with 50 peptides/glycopeptides added to a 7x7x7 $nm^3$ box (=0.24 M) and were subjected to triplicate production runs for 500 ns for statistical analysis. Some simulations were rerun for 2000 ns to ensure equilibration of self-assembled structures. The trajectories generated were visualized using VMD.[9] Hydrogen bonding, Solvent-Accessible Surface Area (SASA), Ramachandran plot and radial-distribution function (RDF) analysis was also carried out using inbuilt GROMACS tools with default settings. For RDF g(r) measurements, CH-π interactions were measured from heavy atoms such as CB for S and CG2 for T to the CG for F (Fig. S22) to correspond with experimental XRD data and to avoid rotational variance with C-H atoms. Therefore, the g(r) was peaked at r∼0.45 nm, which is longer than typical CH-π distances.

*Transmission electron microscopy (TEM):* 5 μL of the sample solution in water (24 hrs after dissolving to allow aggregation) was dropped onto a carbon-coated grid (Electron Microscopy Sciences) and incubated for one minute. The excess of solution was removed using filter paper (blotting), followed by staining with 5 μL of 2 % (w/v) uranyl acetate solution for 30 sec. The excess staining was removed by blotting and the grid was left to dry in air at room temperature. Samples were then imaged in a FEI TITAN Halo TEM operating at 300 kV and images were recorded using a FEI CETA 16M camera.

*Scanning electron microscopy (SEM):* A drop of a solution of the self-assembled systems was deposited onto a silicon wafer and left to dry in air at room temperature. Dried samples were coated with platinum (Leica ACE600 sputter coater) and imaged using a FEI Helios Nanolab660 Dualbeam FIB-SEM, recorded at 5 kV.

*Atomic Force Microscopy (AFM):* A fresh cleaved mica sheet was functionalized with 200 μL of (3-Aminopropyl)triethoxysilane (APTES) for 30 min at room temperature. The mica was rinsed with deionized water and dried under nitrogen flow. A drop of the self-assembled sample was added on the functionalized mica sheet. All AFM images were acquired with a JPK Nanowizard 3 in air at room temperature. The morphological features and mechanical properties of the nanofibers were assessed under JPK quantitative imaging mode using RTESPA-525 probes (k∼200 N/m, Bruker, Germany), calibrated by the JPK contact free method. All approaching force curves were fitted using Hertz/Sneddon model to obtain the Young's modulus. The reported Young's moduli are averaged over 30 measurements.

*Wide-angle-X-ray Powder Diffraction (XRD)*: Glycopeptides and peptides were dissolved in water (40 mM) 24 hrs prior to the measurements and maintained at room temperature. The formed aggregates were freeze-dried and the obtained powder was placed on a quartz substrate. X-ray intensity data were recorded on a PANalytical X'Pert Pro Powder Diffraction instrument with an x-ray source of 3 kW and working at a wavelength of 1.5405 Å.

*Circular dichroism (CD):* 28 μL of the sample (7.5-15 mM in water) was placed in a 0.1 mm demountable quartz cuvette and spectra were recorded on a JASCO J-1500 spectrometer (20 °C, 2 s integration, a step size of 1 nm, single acquisition with a slit width of 1 nm). Continuous ramp temperature measurements were performed for the stability studies.[10] The samples (300 μL, 2 mM aggregated glycopeptides and peptides in water) were added in a 1 mm cuvette placed in a temperature controlled module. Spectra were recorded as a function of temperature starting at 25 °C and increasing to 90 °C (2 °C steps) and then decreasing back to 25 °C (5 °C steps) for a complete cycle. The intensity of the signal at ~220 nm was measured and it is presented as a function of the temperature in the main manuscript (Fig. 4D).

*Calculation of $T_M$:* We have adopted a procedure applied to full length proteins, in which changes in CD as a function of temperature is used to determine the thermodynamics of unfolding and the respective descriptors (equation 1): the vant's Hoff enthalpy (ΔH) and entropy (ΔS) of unfolding, the midpoint of the unfolding transition TM and the free energy of folding (ΔG).[10] In our case, the studied molecules undergo transition between two

states: assembled (A) and free (F) one and at any temperature T the assembly constant is given by K (equation 2). The fraction of the assembled molecules at any temperature is given by $\alpha$ (equations 3 and 4), where $\theta_T$ is the observed ellipticity at any temperature, $\theta_A$ is the ellipticity of the assemblies and $\theta_F$ is the ellipticity of the free form. $T_M$ is the temperature at which $\alpha$ = 0.5, and $\Delta$Cp is the heat capacity of the transition assembly/free state.

$$\Delta G = \Delta H(1 - T/T_M) \text{-} \Delta \mathcal{C}_P\ ((T_M - T) + T\ln(T/T_M)) \qquad \textit{Equation 1}$$

$$K = \exp(-\Delta G/(RT)) \qquad \textit{Equation 2}$$

$$\alpha = K/(1 + K) \qquad \textit{Equation 3}$$

$$\alpha = (\theta_T - \theta_F)/(\theta_A - \theta_F) \qquad \textit{Equation 4}$$

The intensity of the signal at 220 nm in CD spectra was determined as a function of temperature (T Kelvin), and fitted to the Gibbs Helmholtz equation (equation 1) that describes the folding as a function of temperature. To determine the midpoints of the disassembly curves we applied the van't Hoff equation (equation 5):

$$\ln(K_2/K_1) = \Delta H/R(1/T_1 - 1/T_2) \qquad \textit{Equation 5}$$

Where R is the gas constant equal to 1.98 cal $mol^{-1}$, $K_2$ and $K_1$ are the assembly constants at $T_2$ and $T_1$, respectively. To determine the enthalpy of unfolding from van't Hoff analysis, the natural logarithm (ln) of the folding constant K was plotted as a function of 1/T, where T is the absolute temperature (Fig. S26). The data was truncated to the region giving a linear curve with the highest slope. $\Delta$Cp is usually set at 0 for the initial calculations of the thermodynamics of folding of a monomeric protein. Initial values of $\Delta$H, $T_M$, $\theta_A$ and $\theta_F$ were estimated and used as initial parameters for nonlinear least squares fitting routines using Prism8 software. $\Delta$Cp of unfolding was used as fixed parameter for determining the enthalpy and $T_M$ of unfolding using equation 1.

**Characterization of the used (glyco)peptides**

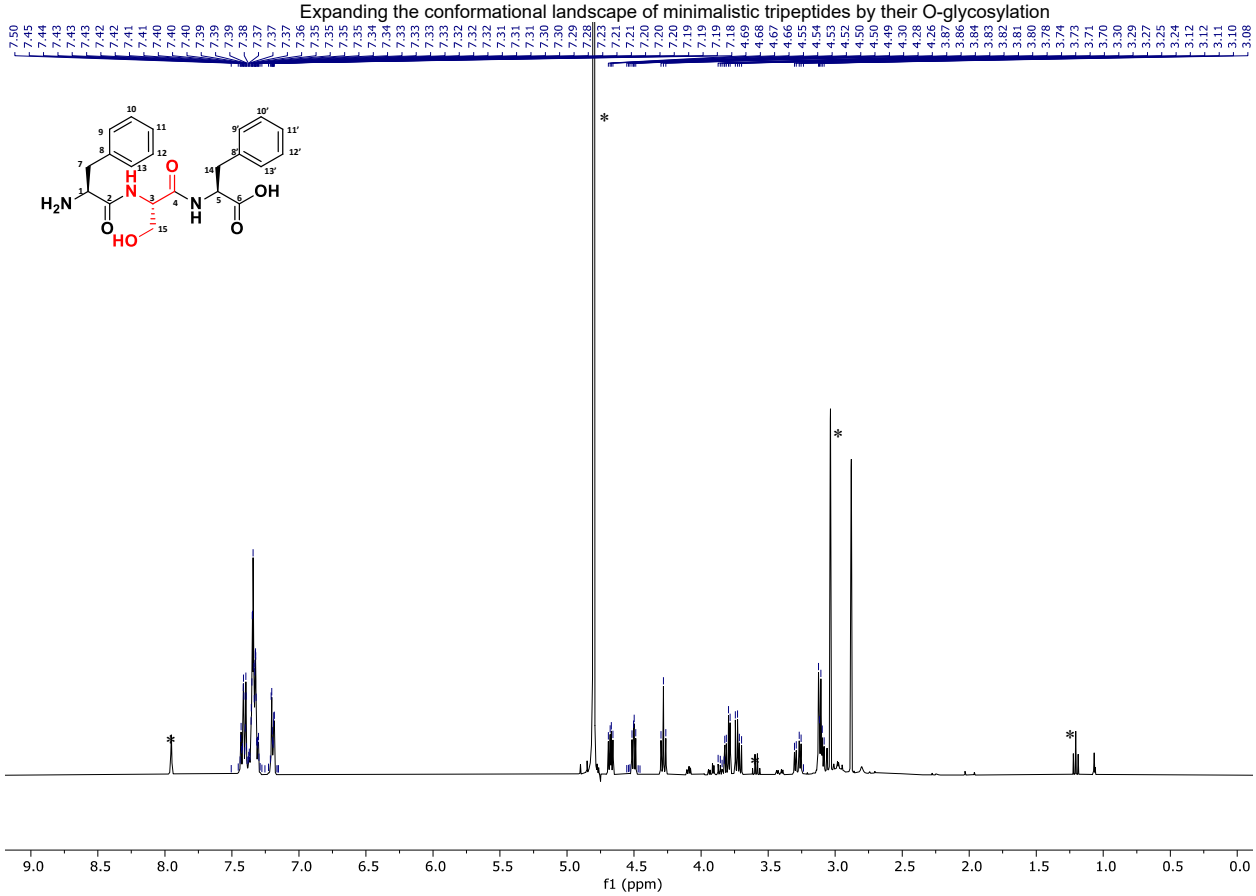



Figure S1. $^{1}$H NMR spectrum of FSF (400 MHz, $D_2O$, 298 K): δ 7.44-7.40 (m, 2H, H12', H10'), 7.38-7.30 (m, 6H, H10, H9, H13, H12, H9', H13'), 7.22-7.19 (m, 2H, H11 H11') 4.75-4.66 (dd, 1H ,H3) 4.52-4.49 (dd, 1H, H1), 4.30-4.27 (t, 1H, H5) 3.83- 3.70 (dd, 2H, H15, H15'), 3.31-3.26 (dd, 2H, H7, H14) 3.16-3.06 (m, 2H, H7', H14'). * Solvent

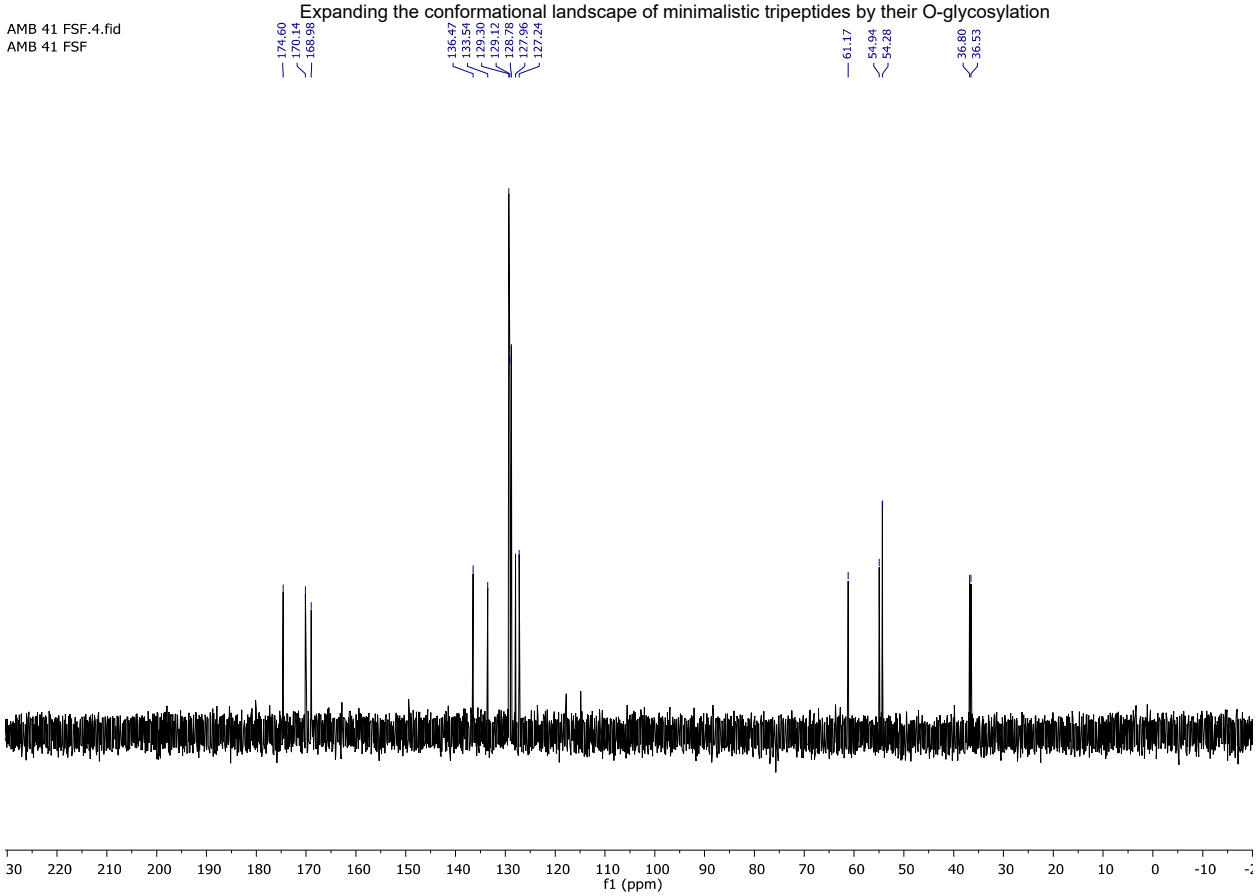



Figure S2. $^{13}C$ NMR spectrum of FSF (101 MHz, $D_2O$): δ 174.60 (C6), 170.14 (C2), 168.98 (C4), 136.47 (C8'), 133.54 (C8), 129.30 (C13', C9'), 129.12 (C13, C9), 128.78 (C12, C10), 127.96 (C12', C10'), 127.24 (C11, C11'), 61.17 (C15), 54.94 (C5), 54.28 (C3, C1), 36.80 (C7), 36.53 (C14).

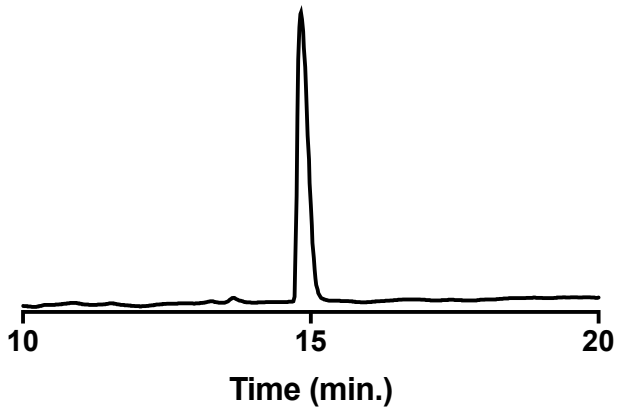



Figure S3. HPLC of purified FSF.

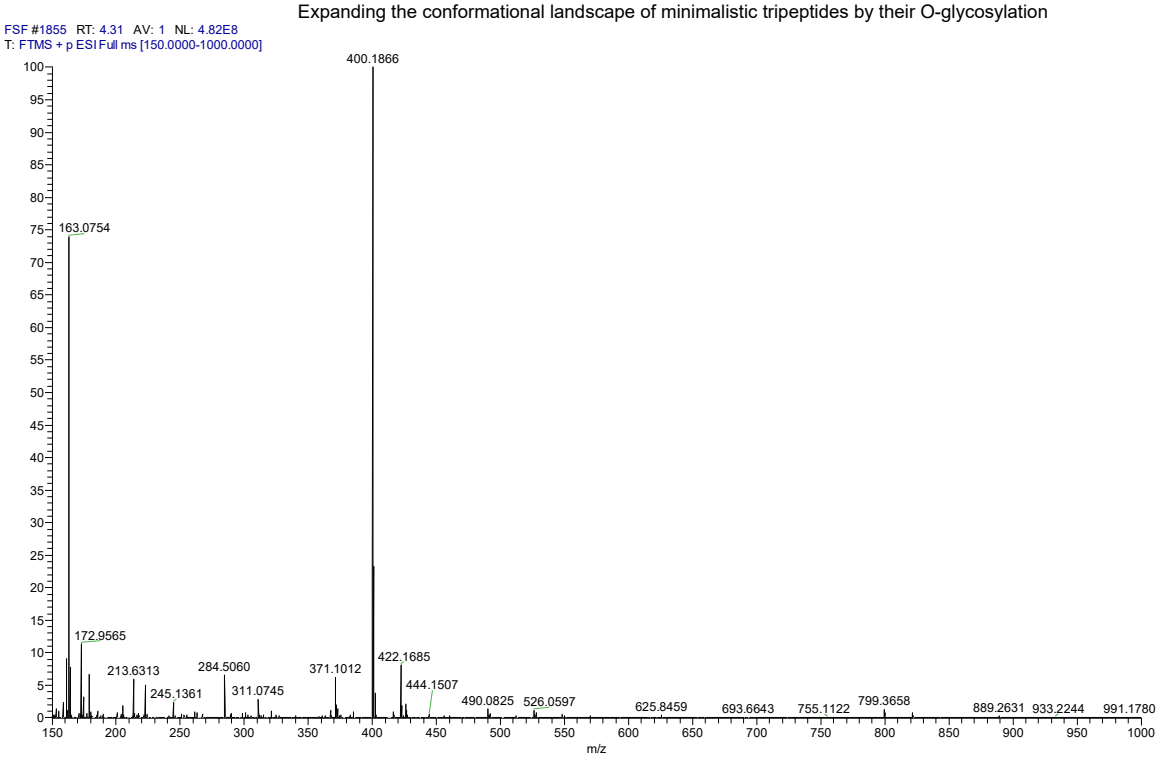



Figure S4. ESI-MS of FSF (m/z): $[M - H]^-$ 400.18.

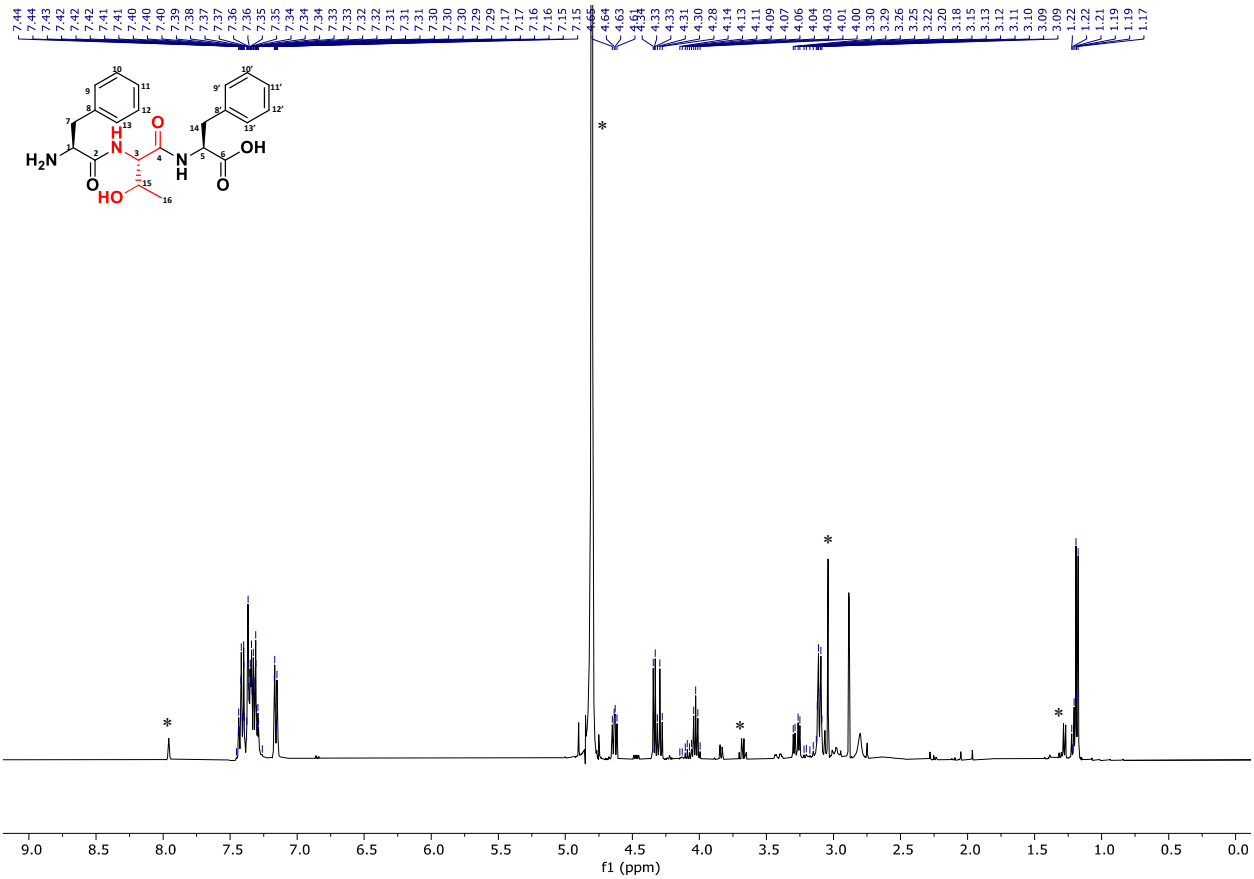



Figure S5. $^1$H NMR spectrum of FTF (400 MHz, $D_2O$, 298 K): δ 7.44-7.39 (M, 4H, H12, H10, H12', H10'), 7.38-7.33 (m, 4H, H9, H13, H9', H13'), 7.33-7.30 (m, 2H, H11', H11), 4.63 (m, 1H, H5) , 4.34-4.28 (m, 1H, H3), 4.08 (m, 1H, H1), 3.30 (m, 1H, H15), 3.29-2.80 (m, 4H, H7, H7', H14, H14'), 1.19 (m, 3H, H16). * Solvent

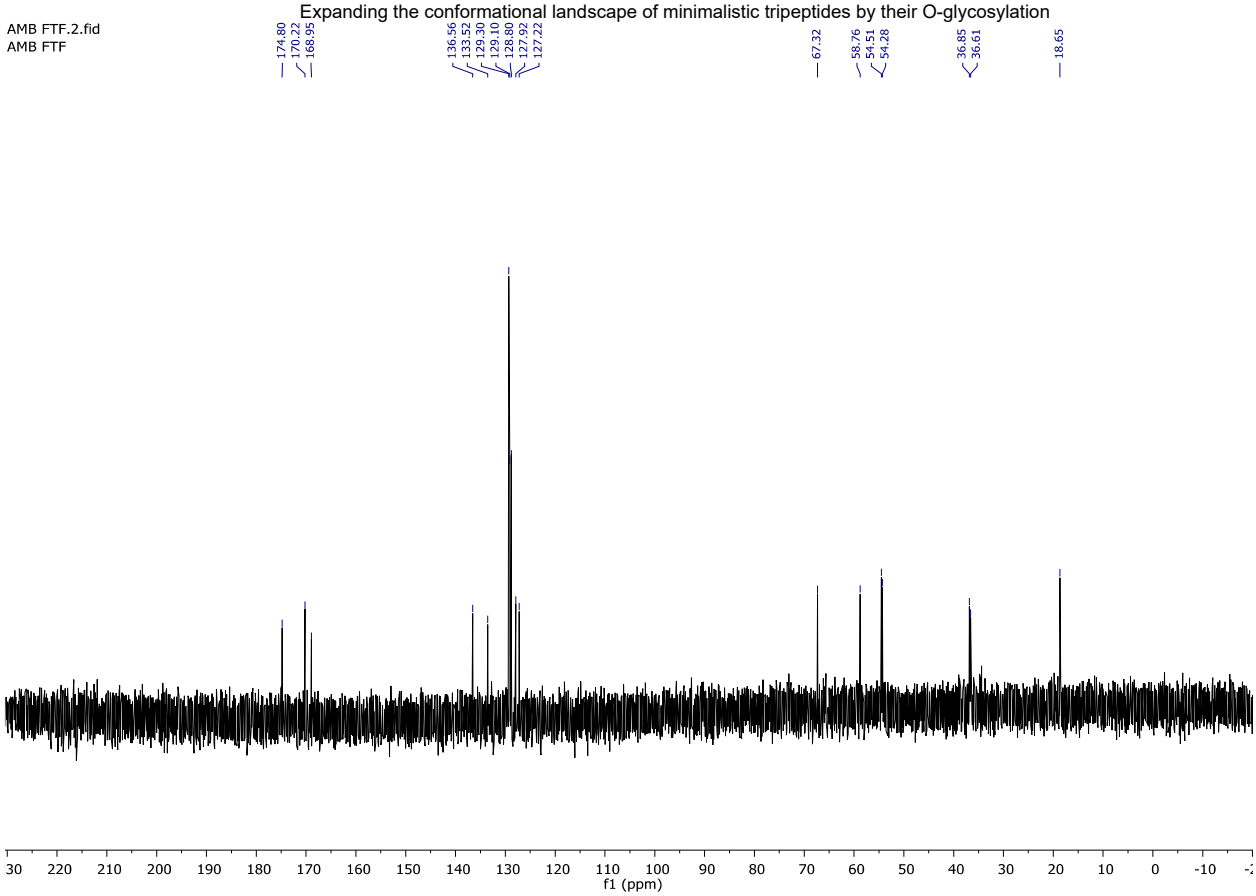



Figure S6. $^{13}C$ NMR spectrum of FTF (101 MHz, $D_2O$) δ 174.80 (C6), 170.22 (C2), 168.95 (C4), 136.56 (C8'), 133.52 (C8), 129.30 (13', C9'), 129.10 (C13, C9), 128.80 (C12, C10), 127.92 (C12', C10'), 127.22 (C11, C11'), 67.32 (C15), 58.76 (C5), 54.51 (C3), 54.28 (C1), 36.85 (C7), 36.61 (C14), 18.65 (C16).

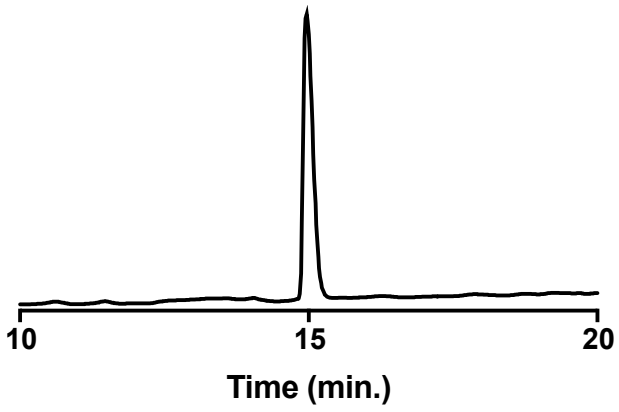



Figure S7. HPLC of purified FTF.

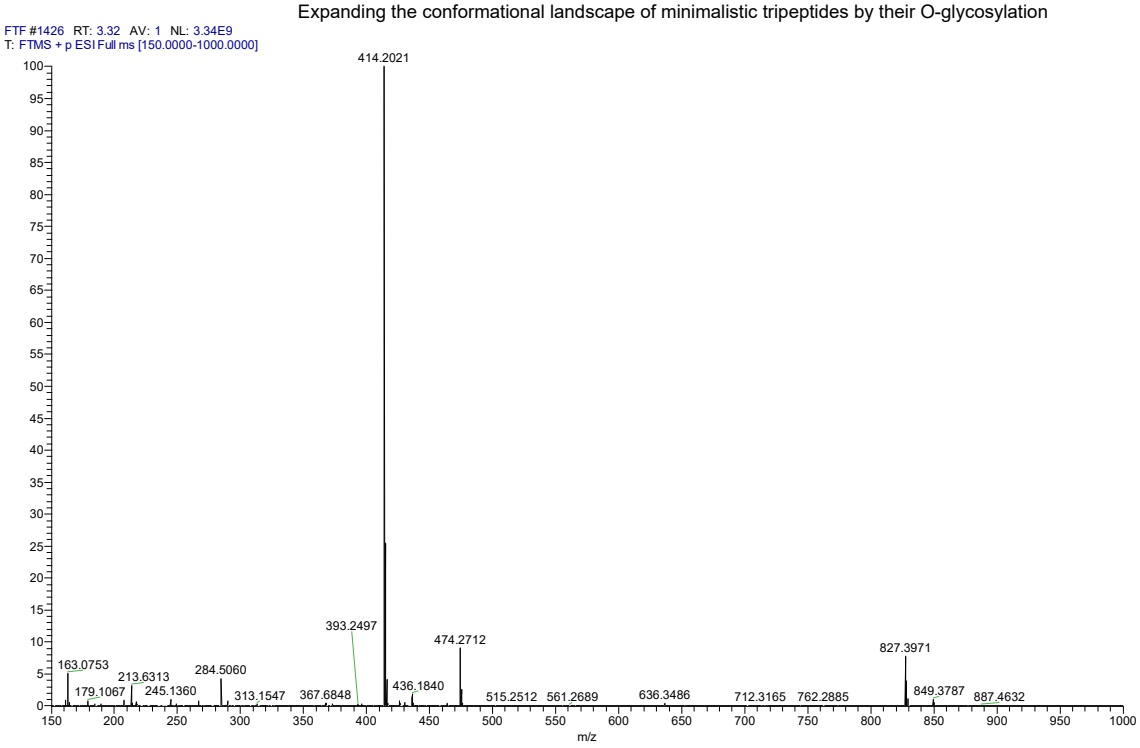



Figure S8. ESI-MS of FTF (m/z): $[M - H]^-$ 414.20.

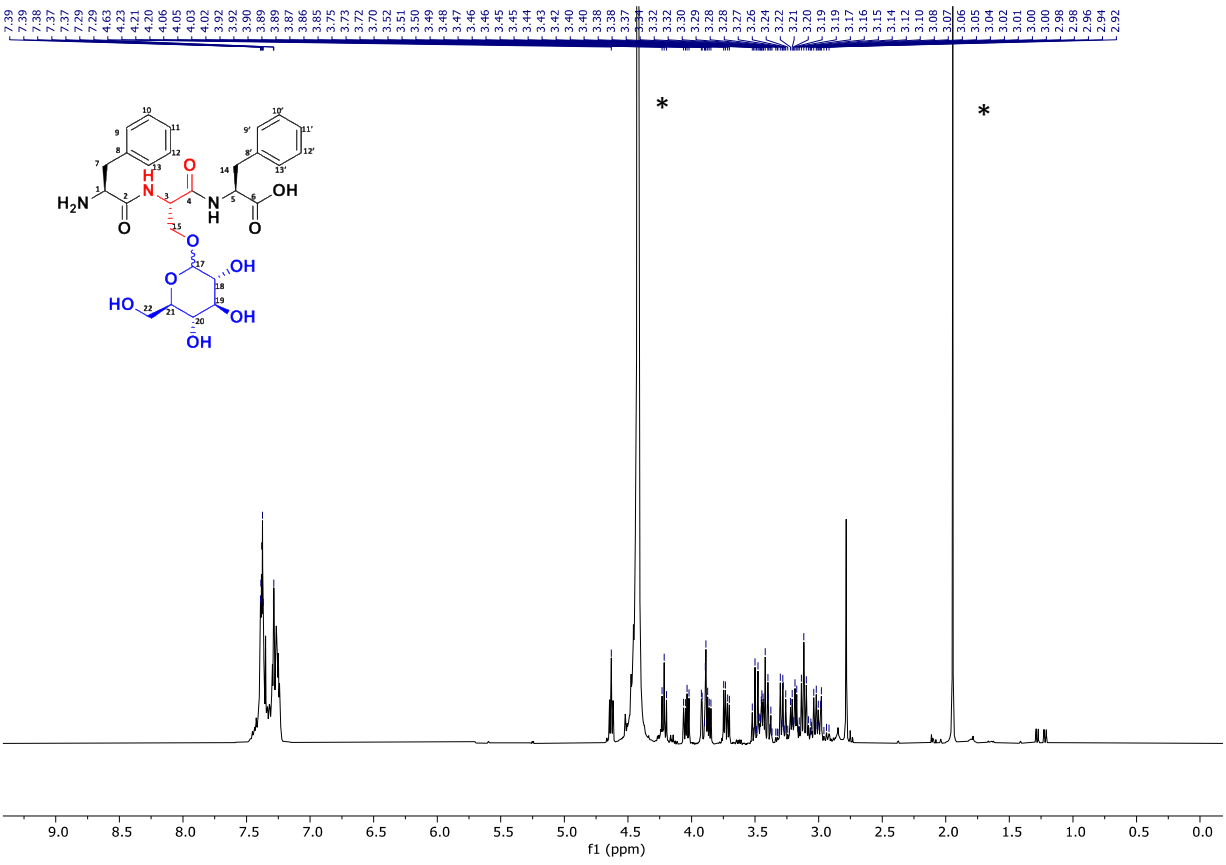



Figure S9. $^1$H NMR spectrum of FS(Glc)F (400 MHz, $D_2O$, 298 K): δ 7.45-7.29 (m, 10H, H10', H12', H9, H9', H10, H10', H13, H13', H11, H11'), 4.63 (d, 1H, H17α, *J*= 4Hz), 4.61 (d, 1H, H17β, *J*= 8Hz), 4.49 (m, 1H, H5), 4.25 (m, 1H, H3), 4.10 (m, 1H, H15), 3.90 (m, 1H, H19), 3.73 (m, 1H, H23), 3.50-3.10 (m, 9H, H22, H20, H28, H22', H21, H7, H7', H14, H14'). * Solvent

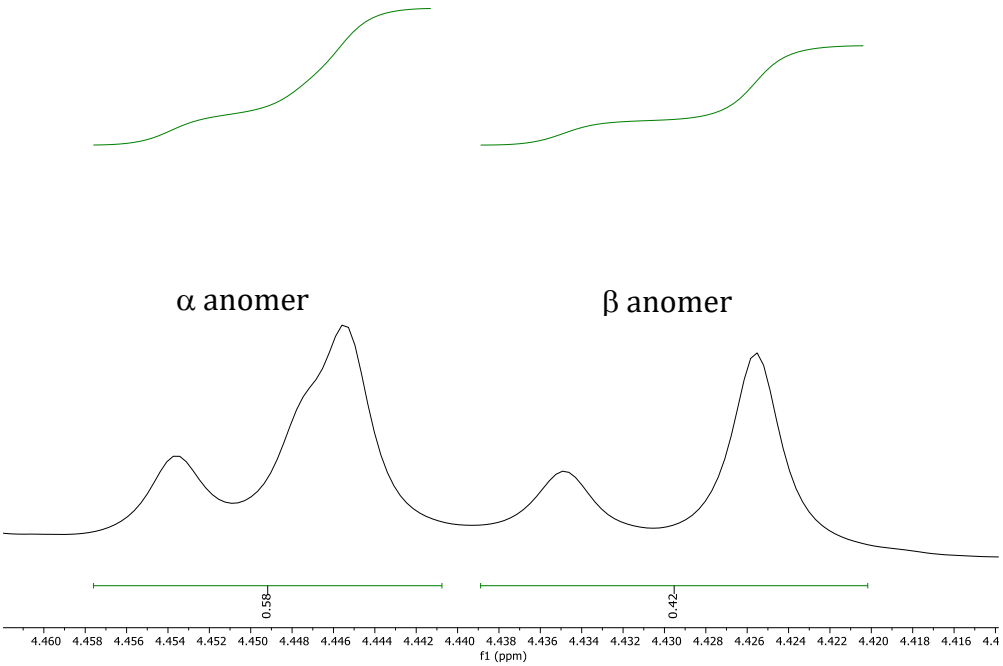



Figure S10. $^{1}H$ NMR spectrum of FS(Glc)F (400 MHz, $D_2O$, 313K) showing the signals used for calculation of the ratio between the two anomers (α:β=0.58:0.42).

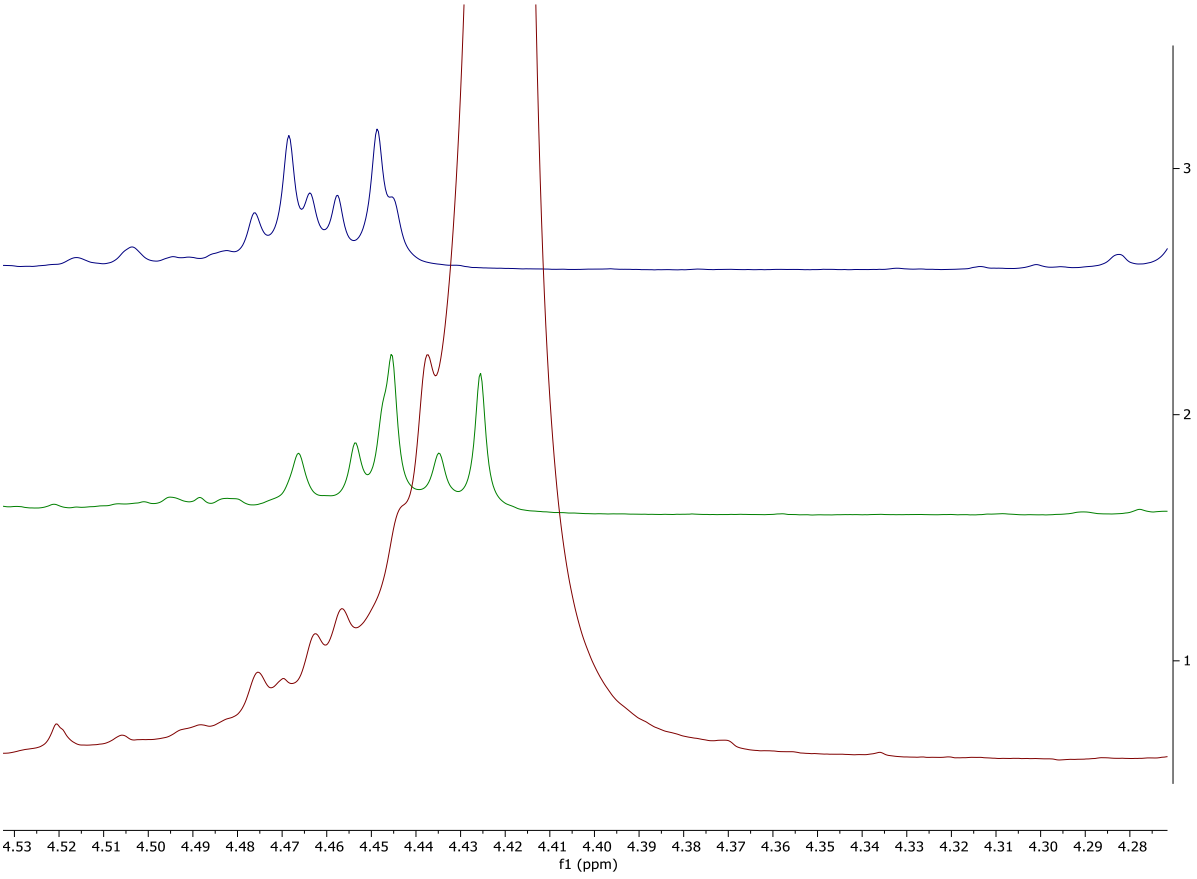



Figure S11. $^{1}H$ NMR spectra of FS(Glc)F (400 MHz, $D_2O$) showing the anomers signal at different temperatures: 298 K in blue, 313 K in green, and 363 K in red.

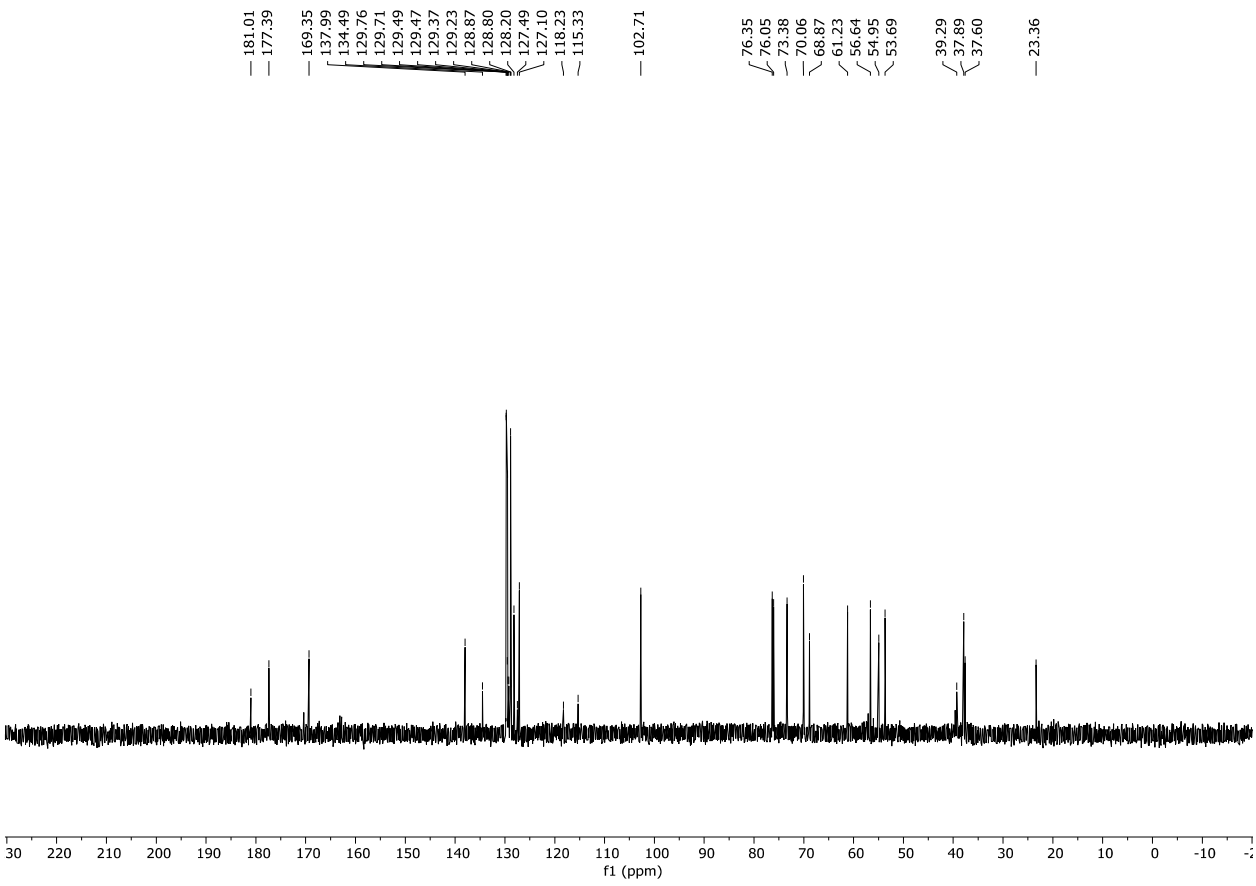



Figure S12. $^{13}C$ NMR spectrum of FS(Glc)F: (101 MHz, $D_2O$, 298K) δ 181.01 (C6), 177.39 (C2), 169.35 (C4), 137.99 (C8), 134.49 (C8'), 129.76 (C13'), 129.49 (C9'), 129.47 (C9), 129.37 (C13), 128.87 (C12'), 128.20 (C10'), 127.49 (C12), 127.10 (C10), 118.23 (C11), 115.33 (C11'), 102.71 (C17α, C17β), 76.35 (C21), 76.05 (C19), 73.38 (C18), 70.06 (C15), 68.87 (C20), 61.23 (C22), 56.64 (C3), 54.95 (C5), 53.69 (C1), 39.29 (C7), 37.7 (C14), 23.36 (C16).

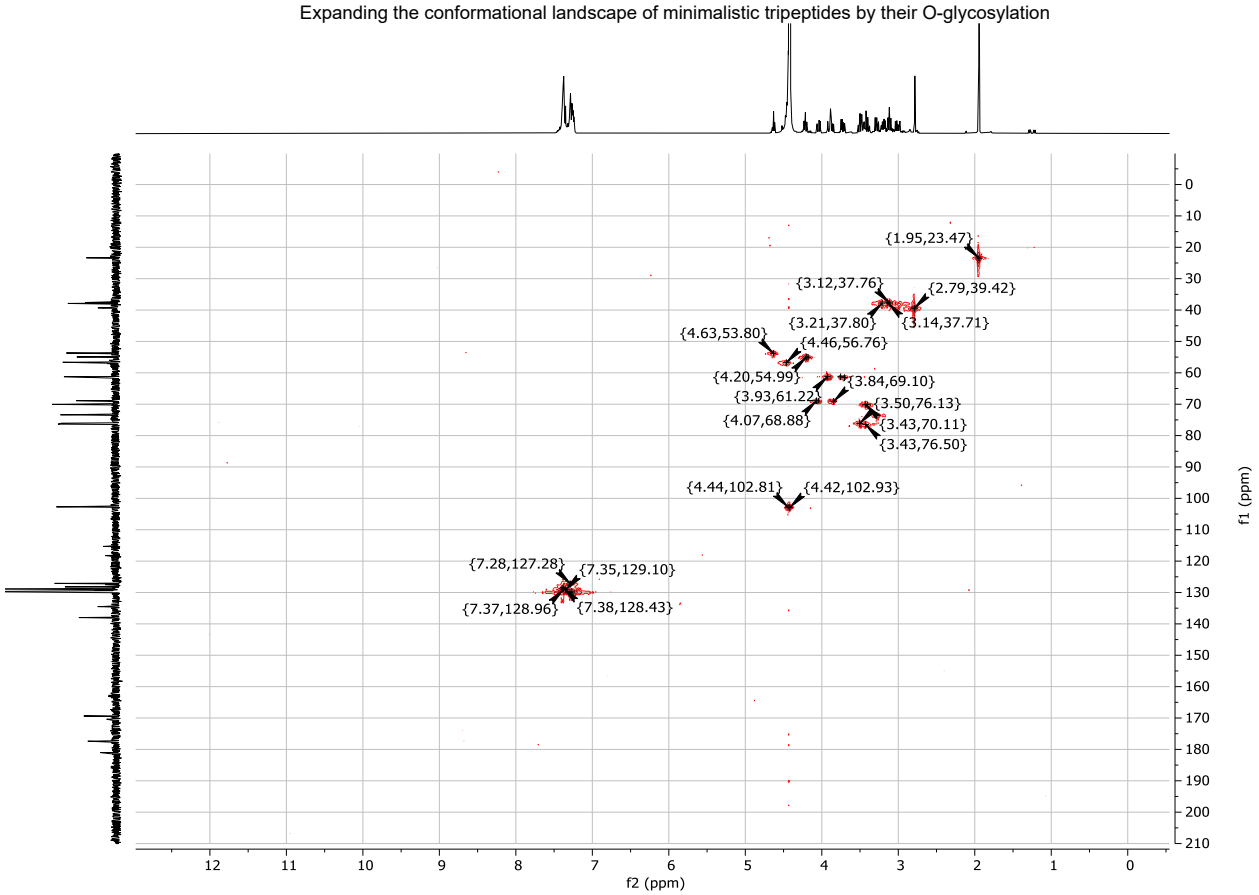



Figure S13. HMQC-NMR spectrum of FS(Glc)F (400 MHz, $D_2O$, 313K).

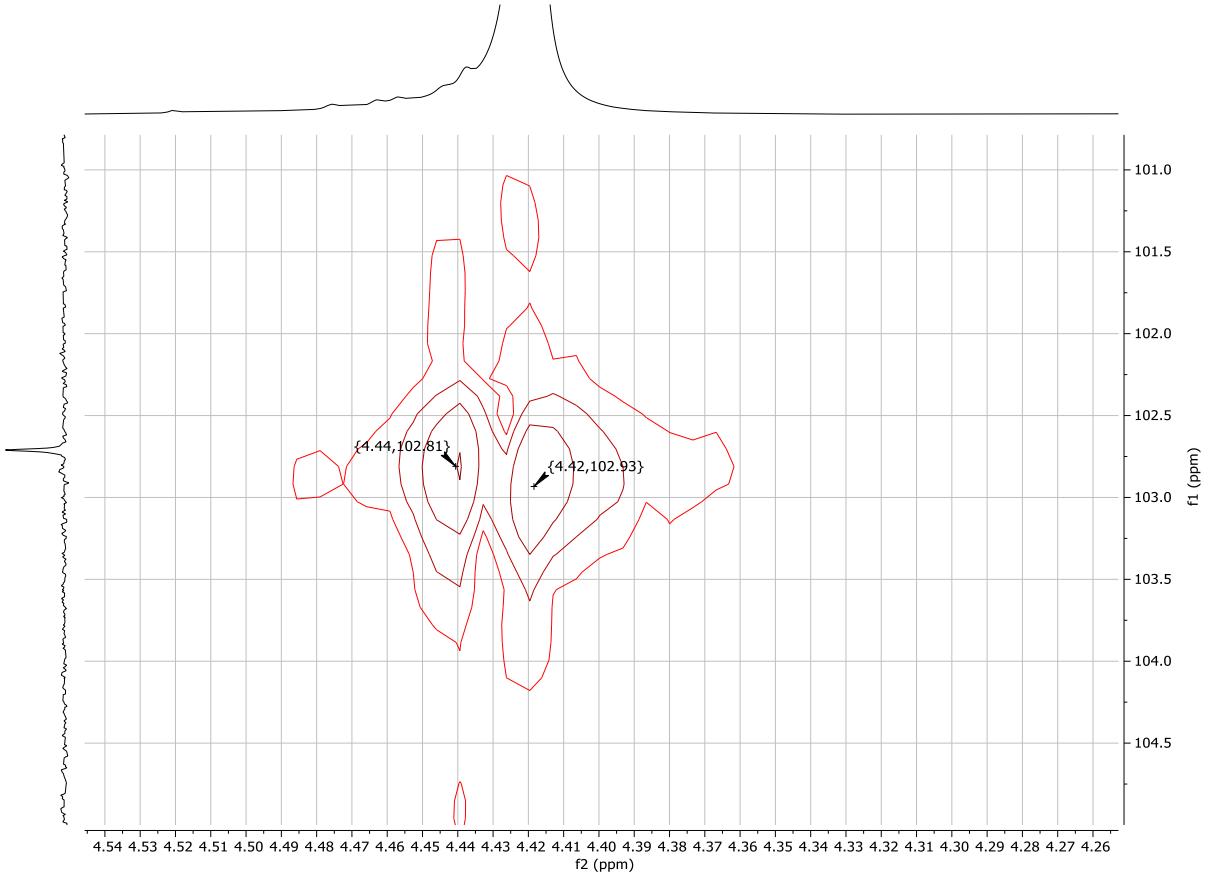



Figure S14. HMQC-NMR spectrum of FS(Glc)F (400 MHz, $D_2O$, 313K) showing the anomeric signals.

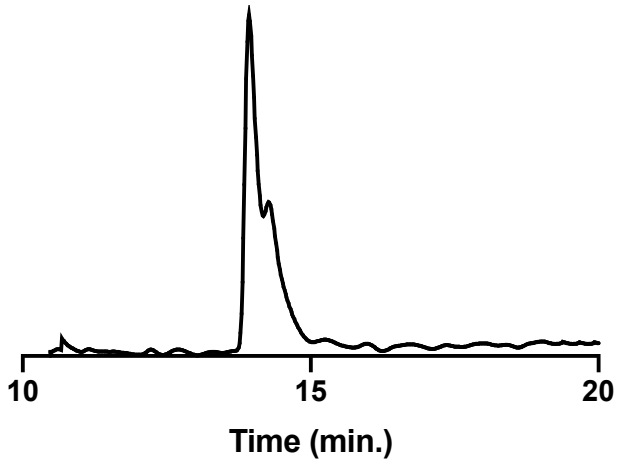



Figure S15. HPLC of purified FS(Glc)F.

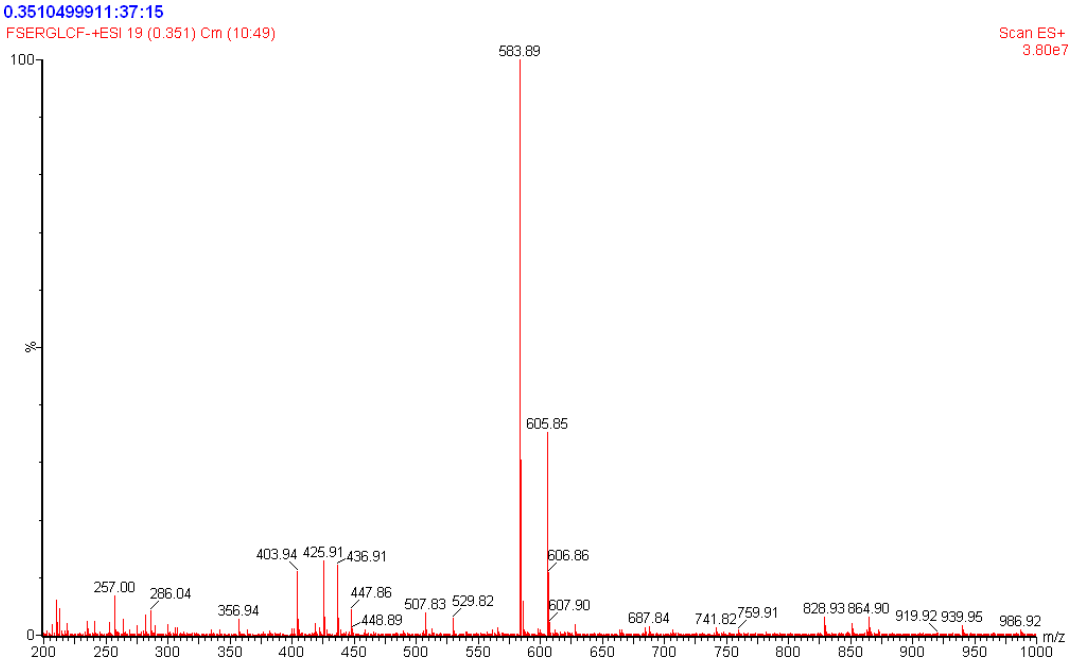



Figure S16. ESI-MS of FS(Glc)F (m/z): $[M - H]^-$ 583.89.

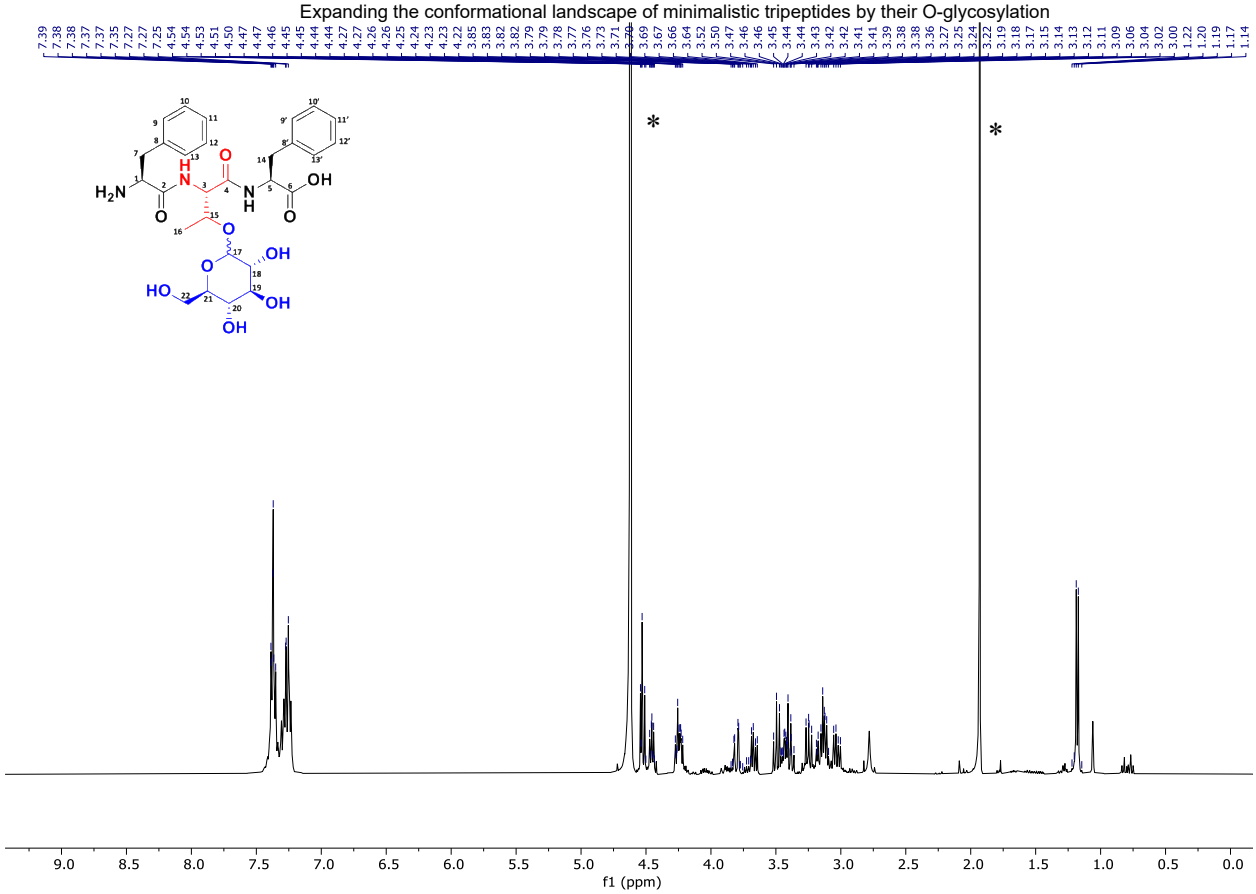



Figure S17. $^1$H NMR spectrum of FT(Glc)F (400 MHz, $D_2O$, 298K) δ 7.39-7.25 (m, 10H, H10', H12', H9, H9', H10, H10', H13, H13', H11, H11'), 4.53 (d, 1H, H17α, *J*= 4Hz), 4.51 (d, 1H, H17β, *J*= 8Hz), 4.47 (m, 1H, H5), 4.27 (m, 1H, H3), 4.22 (m, 1H, H15), 3.85 (m, 1H, H19), 3.81 (dd, 1H, H1), 3.73 (m, 1H, H22), 3.52-3.02 (m, 9H, H20, H18, H22', H21, H7, H7', H14, H14'), 1.19 (d, 3H, H16). * Solvent

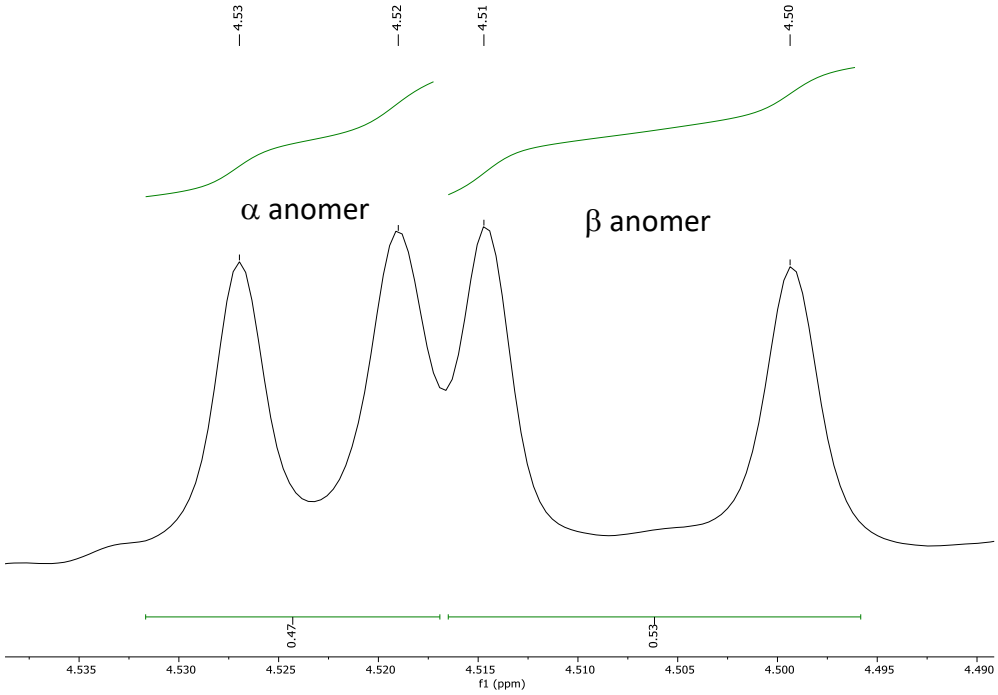



Figure S18. $^1$H NMR spectrum of FT(Glc)F (400 MHz, $D_2O$, 363K) showing the signals of the two anomers used for the calculation of the ratio α:β (α:β=0.47:0.53)

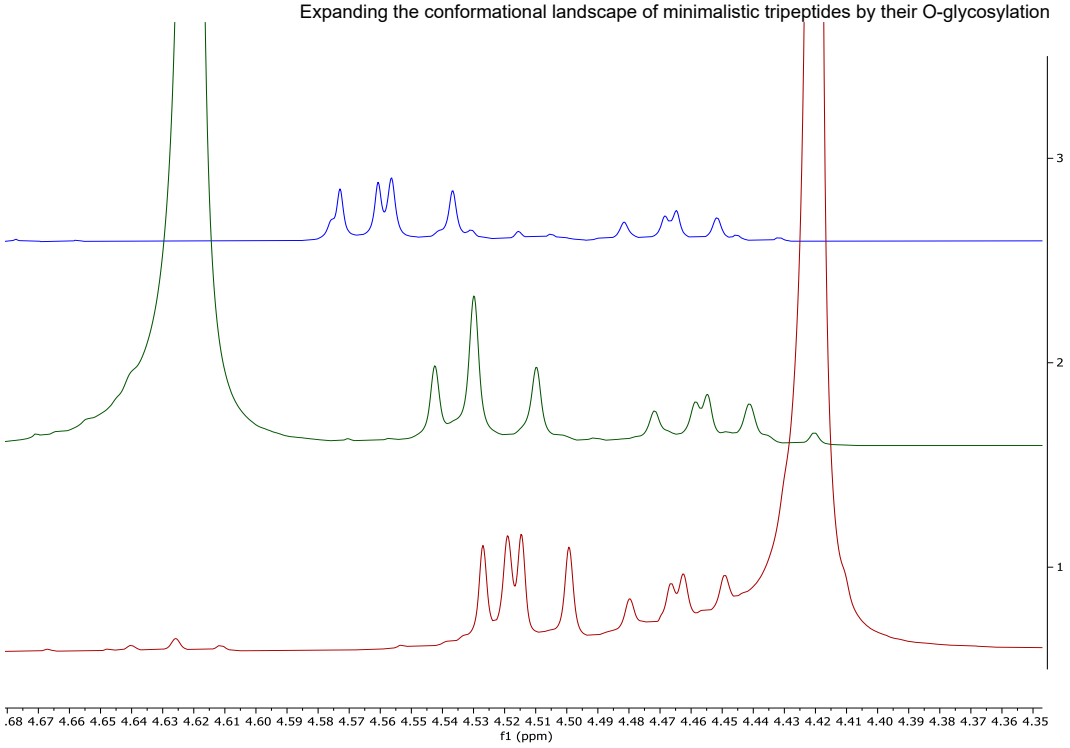



Figure S19: $^{1}H$ NMR spectra of FT(Glc)F anomers peaks (400 MHz, $D_2O$) at different temperatures: 298 K in blue, 313 K in green, and 363 K in red.

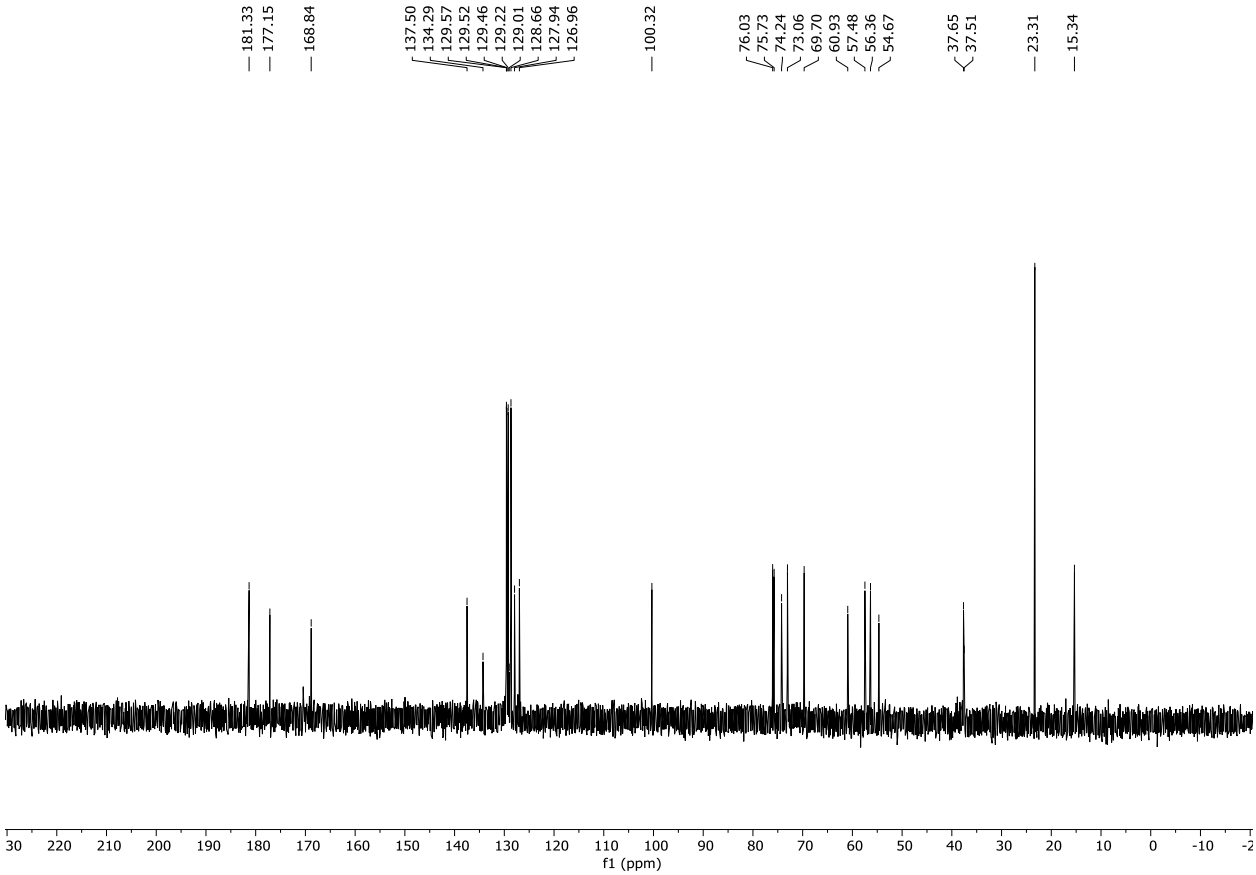



Figure S20: $^{13}C$ NMR spectrum of FT(Glc)F (101 MHz, $D_2O$, 298K) δ 181.33 (C6), 177.15 (C2), 168.84 (C4), 137.50 (C8), 134.29 (C8'), 129.57 (C13'), 129.46 (C9'), 129.22 (C9), 129.01 (C13), 128.66 (C12', C10'), 127.94 (C12, C10), 126.96 (C11, C11'), 100.32 (C17α and C17β), 76.03 (C21), 75.73 (C19), 74.24 (C18), 73.06 (15), 69.70 (C20), 60.93 (C22), 57.48 (C3), 56.36 (C5), 54.67 (C1), 37.65 (C7), 37.51 (C14), 15.34 (C16).

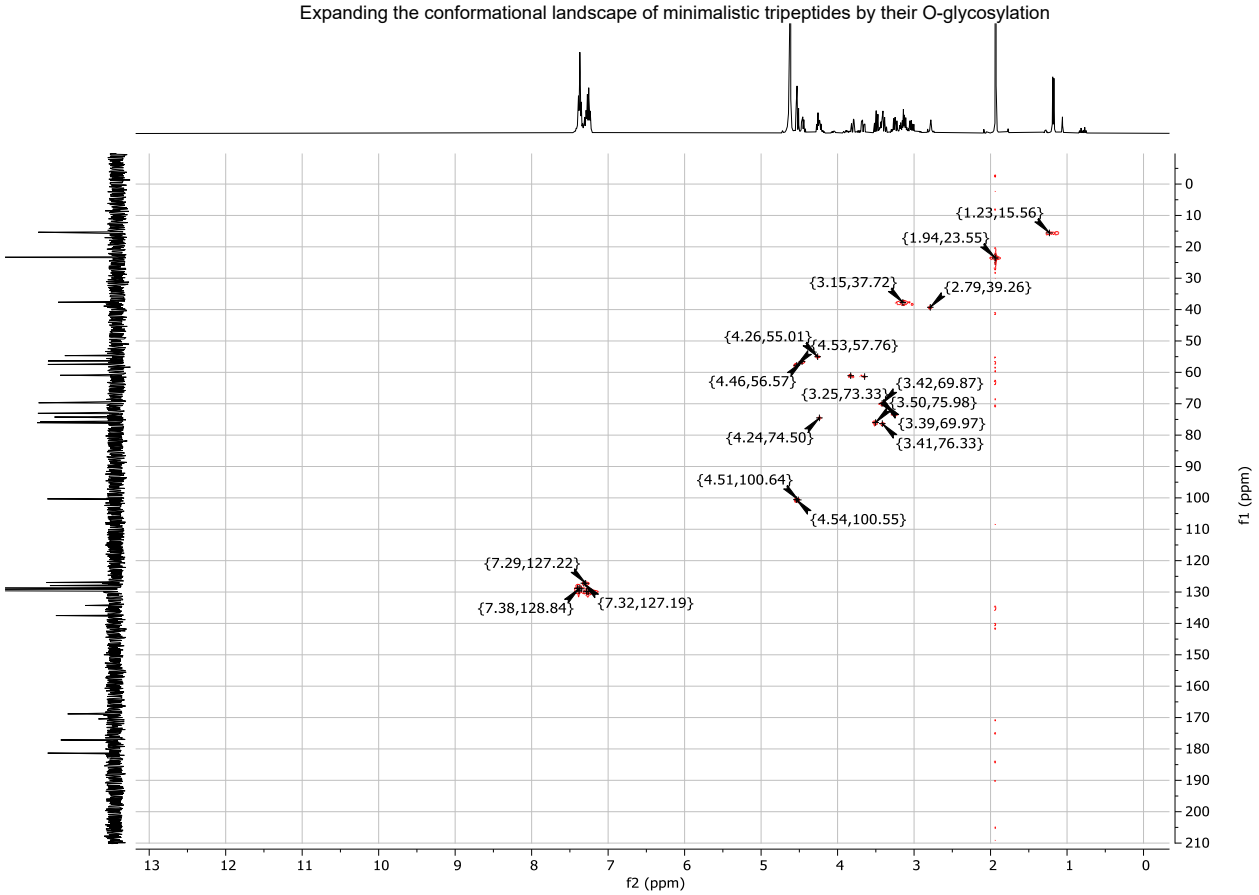



Figure S21. HMQC-NMR spectrum of FT(Glc)F (400 MHz, $D_2O$, 363K).

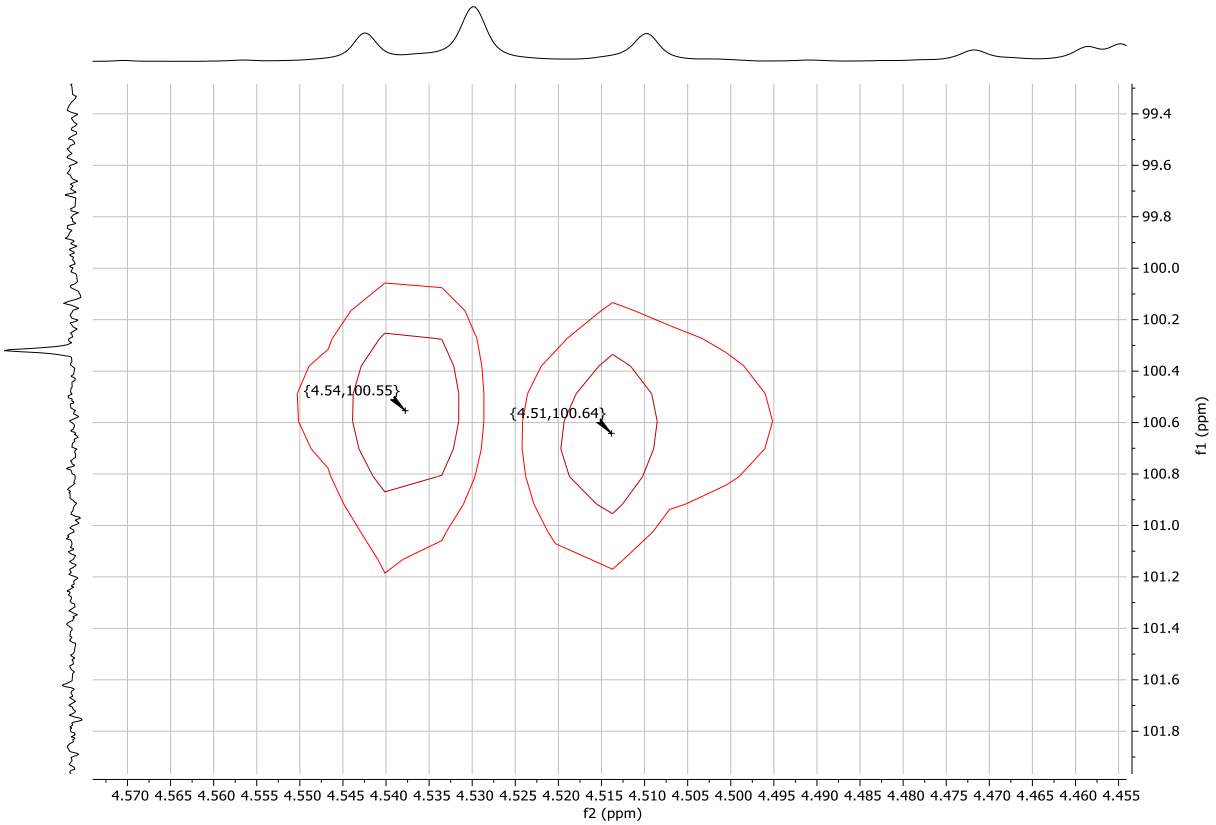



Figure S22. The anomeric peaks in HMQC-NMR spectrum of FT(Glc)F (400 MHz, $D_2O$, 363K).

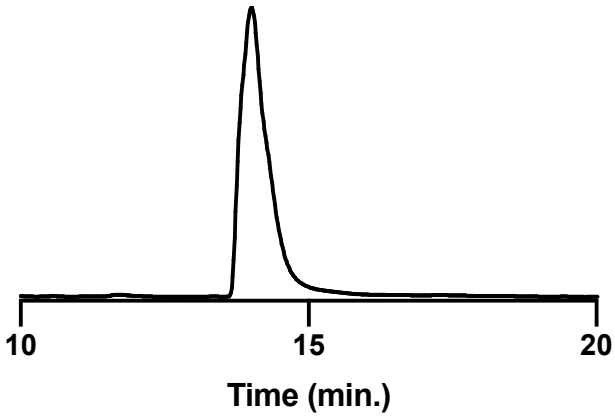



Figure S23. HPLC of purified FT(Glc)F.

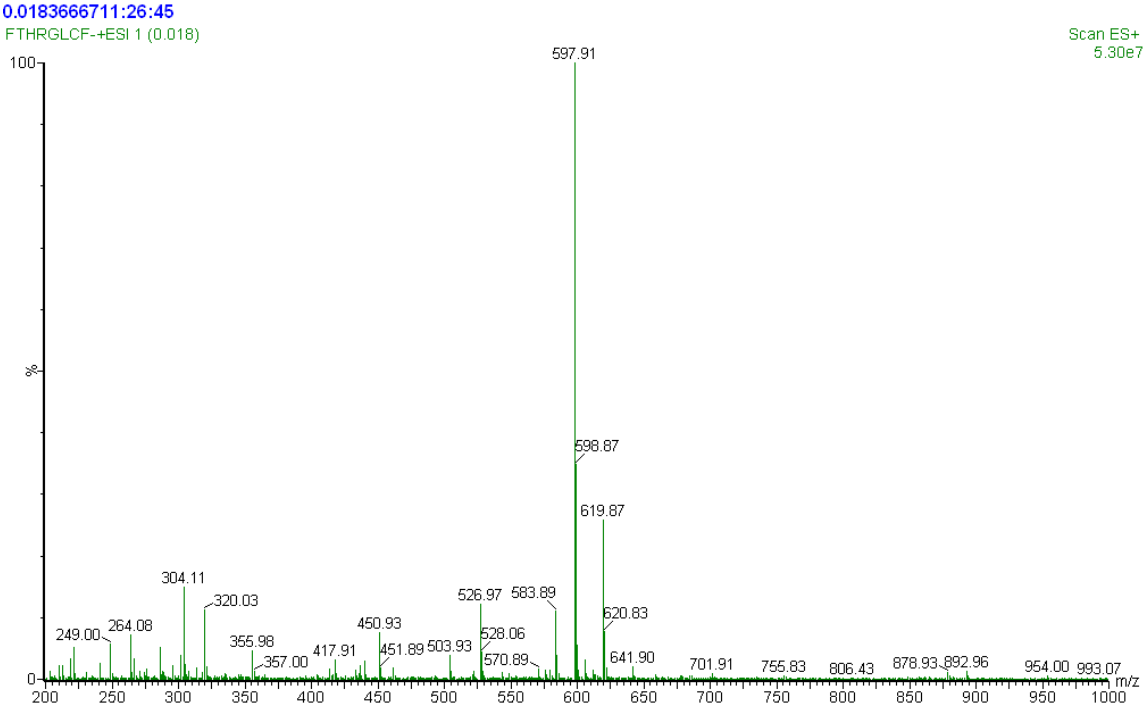



Figure S24. ESI-MS of FT(Glc)F (m/z): $[M - H]^-$ 597.91.

## Supplementary computational data

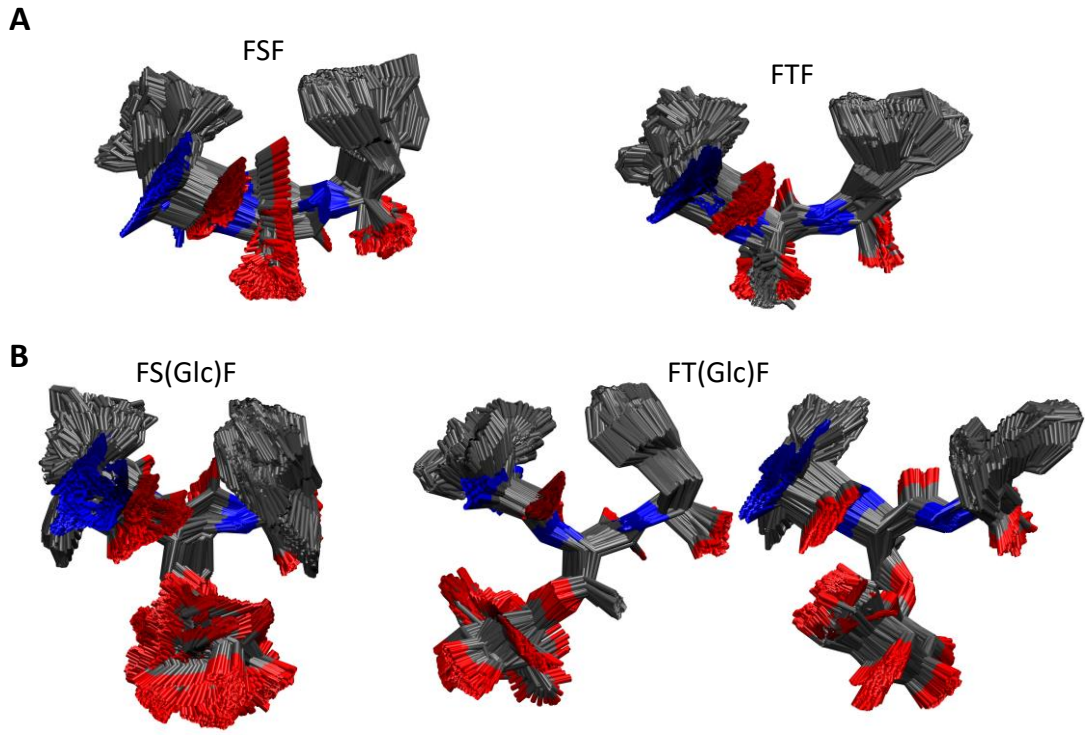



Figure S25. Ensemble of conformers present in the largest cluster of each (glyco)peptide. The respective central conformers are shown in Figure 1 within an RMSD cutoff 0.1 nm. Hydrogens are removed for clarity.

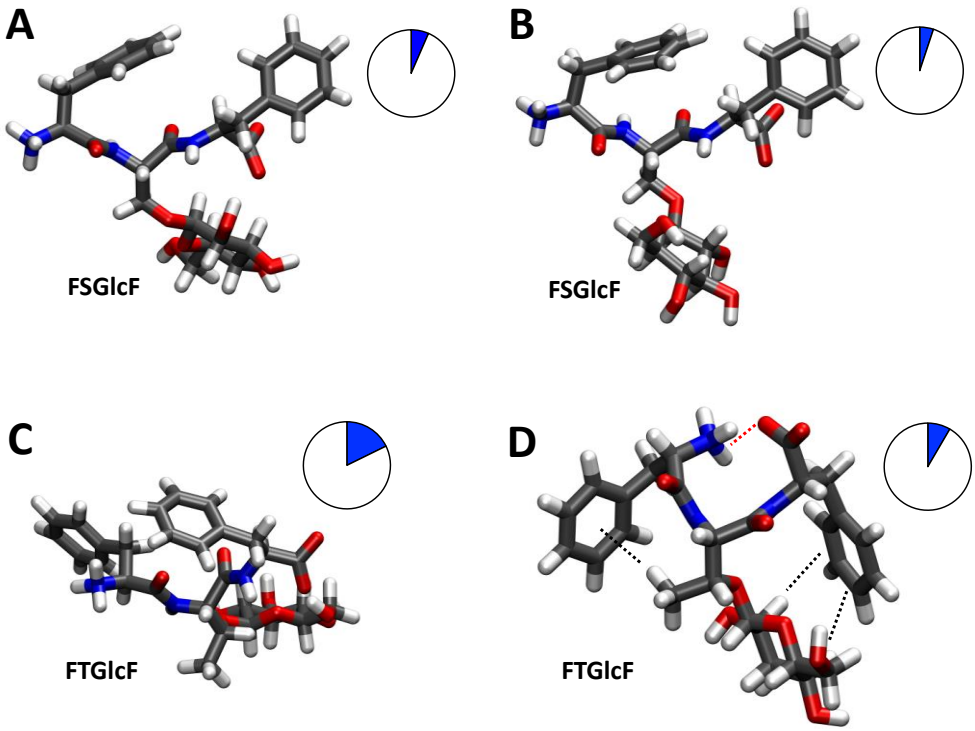



Figure S26. Representative central structures of the clusters containing more than 100 structures out of 5000 structures generated from a 500ns MD simulation for (A, B) FS(Glc)F and (C, D) FT(Glc)F at an RMSD cutoff of 0.1 nm showing a

wider diversity in conformations compared to the tripeptides (Fig. 1). Red dotted lines represent H-bonding and black dotted lines represent CH-π interactions.

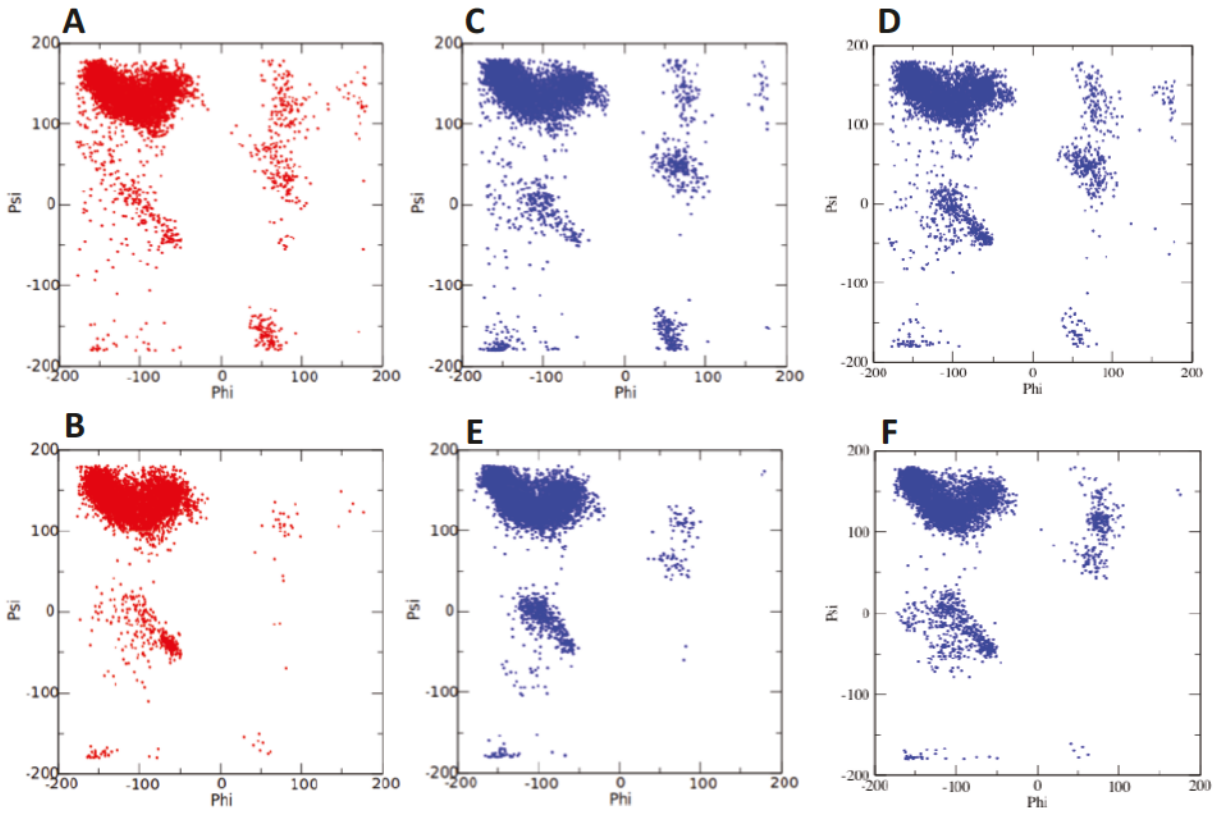



Figure S27. Ramachandran plots for (A) FSF, (B) FTF, (C) FS(βGlc)F, (D) FS(αGlc)F, (E) FT(βGlc)F, and (F) FT(αGlc)F.

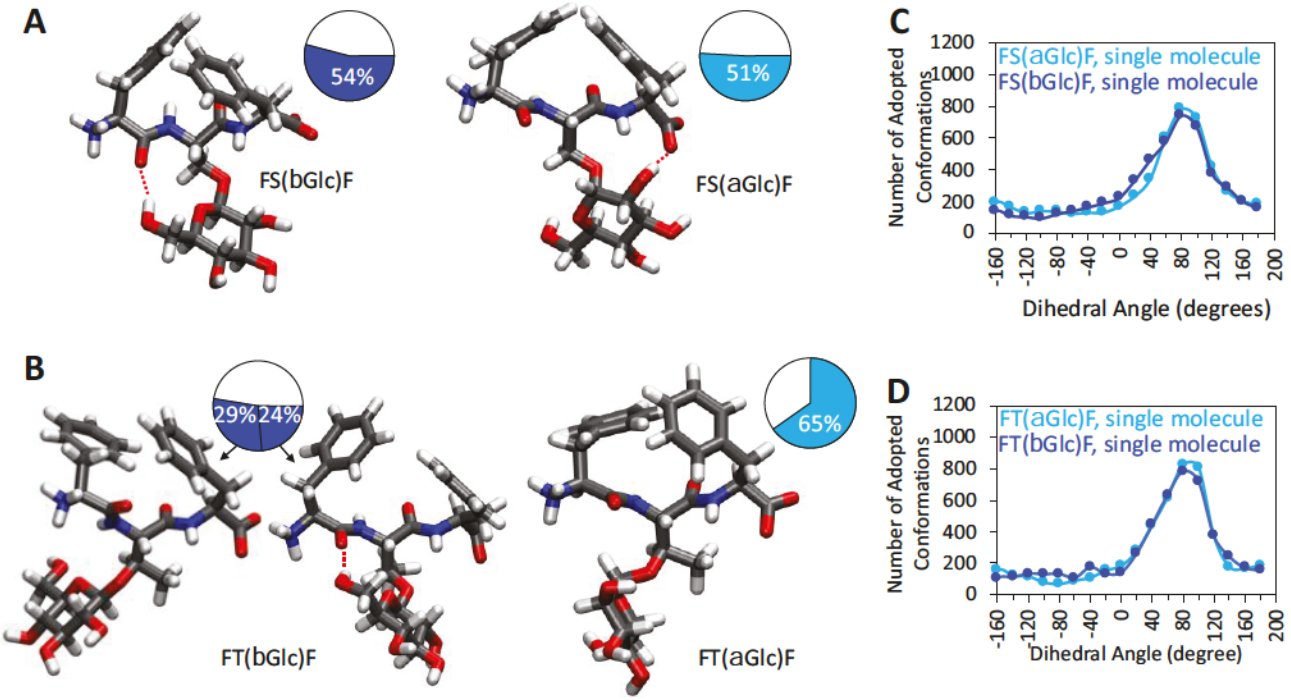



Figure S28. Representative central structures of the largest clusters obtained for the two anomers of (A) FS(Glc)F and (B) FT(Glc)F and (C, D) the respective dihedral analysis of single molecule for the anomers. Data were generated from

500 ns MD simulations within a 0.1 nm RMSD cluster cut-off. The pie-charts show the fraction of all conformers represented by the central structure. Red dotted lines represent H-bonding.

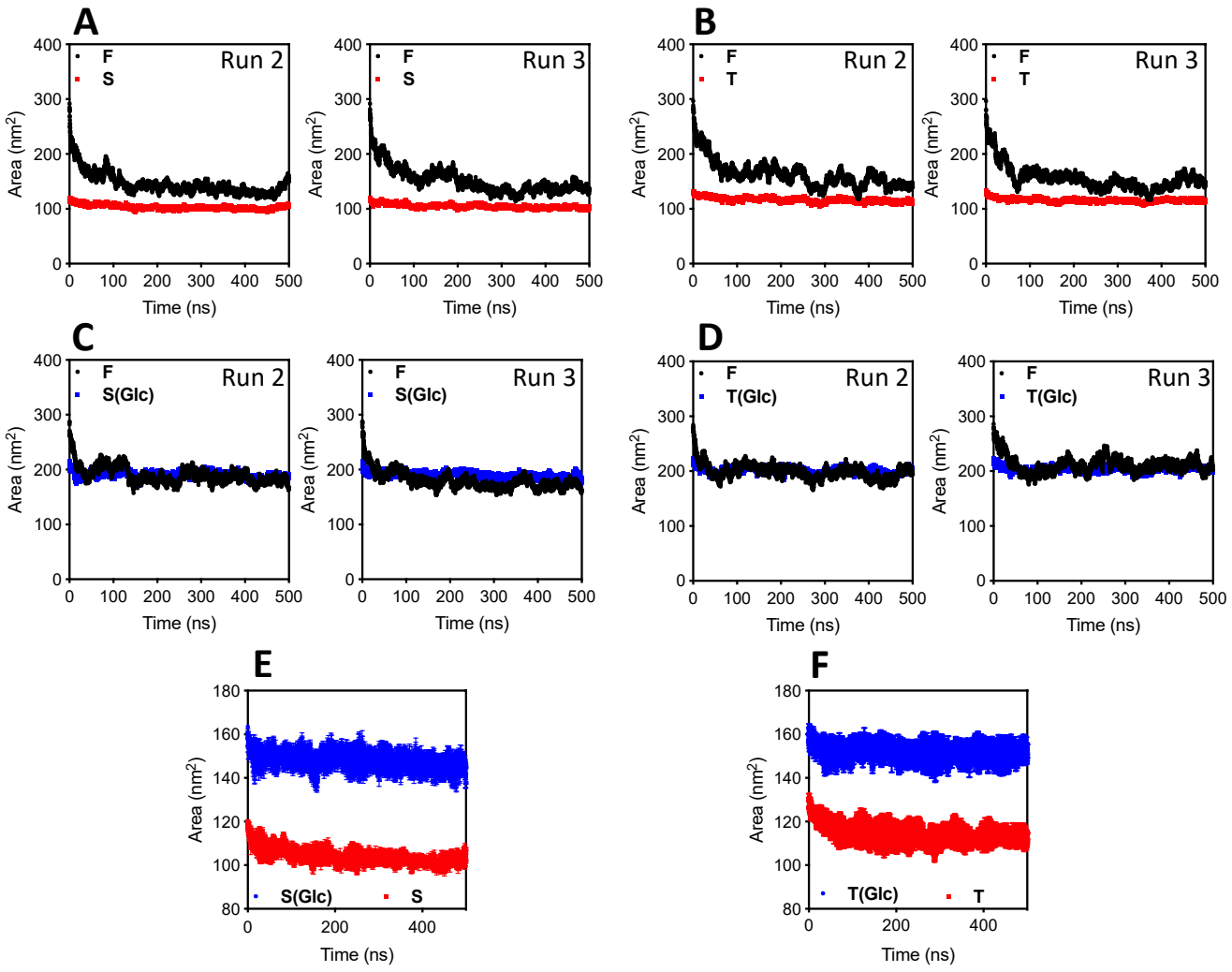



Figure S29. Computational data generated from different independent runs and showing the solvent-accessible surface area (SASA) analysis for (A) FSF, (B) FTF, (C) FS(Glc)F, and (D) FT(Glc)F. (E, F) SASA plots for sugar vs peptide (blue vs red) showing a rebalance of interactions with water. Data for aromatic amino acids are presented in black, hydrophilic amino acids in red and glucose in blue.

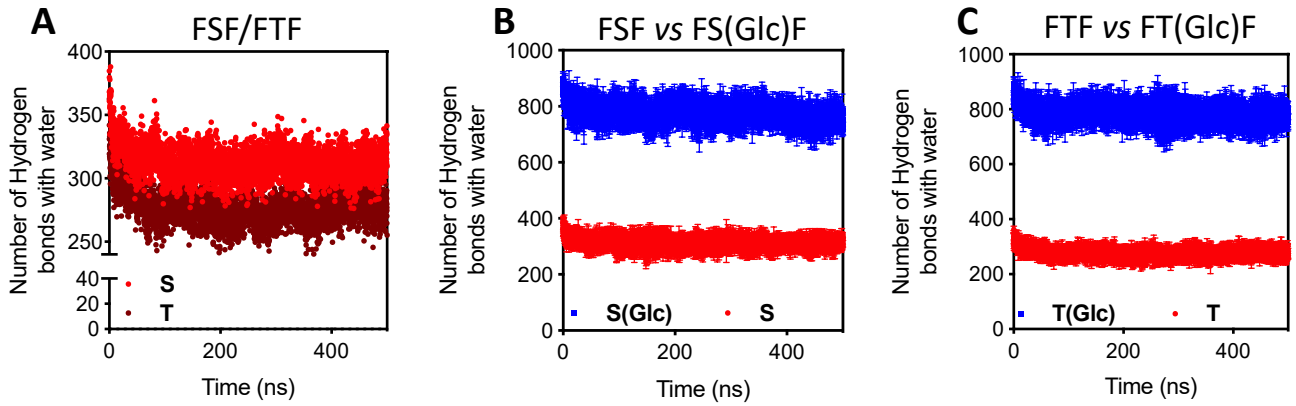



Figure S30. Computational modelling studies showing comparative analysis of the number of hydrogen bonds with the solvent (water) between (A) the tripeptides and (B, C) the tripeptides and the respective *O*-glycosylated analogues.

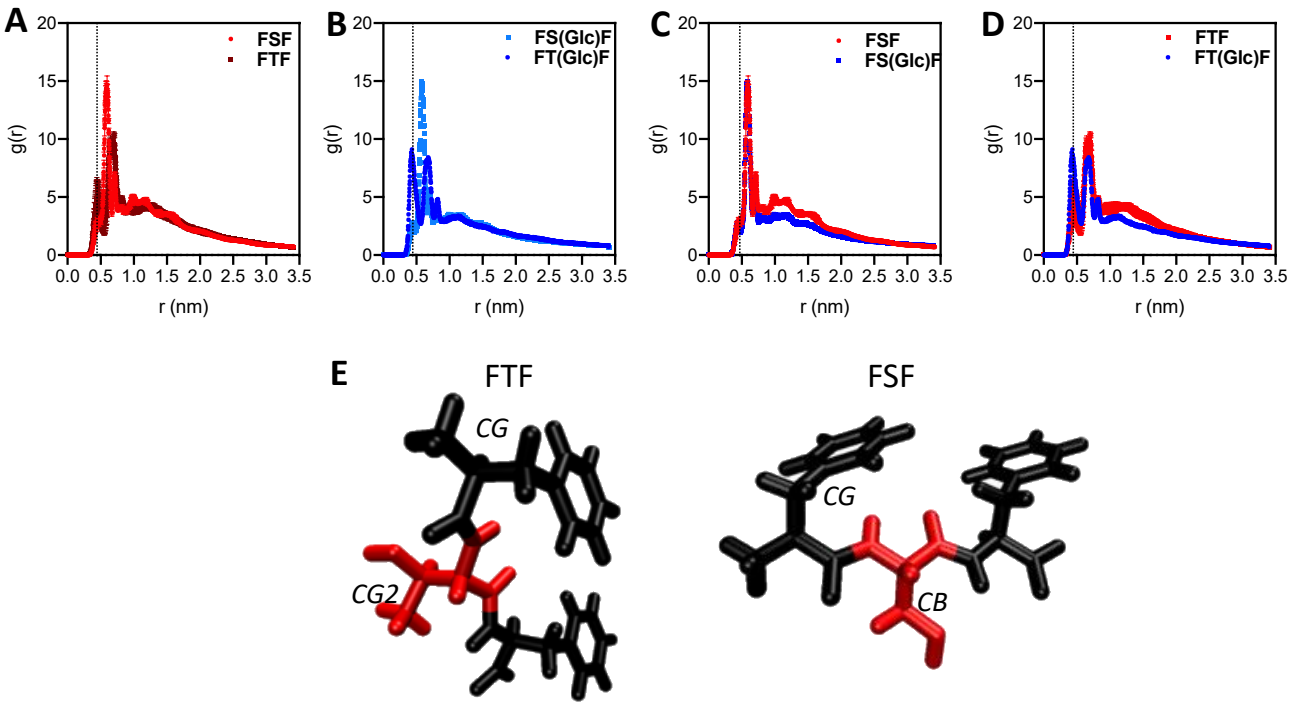



Figure S31. Statistical average of radial distribution function (g(r)) plotted as function of distance (r) between (A) S and T, (B) S(Glc) and T(Glc), (C) S and S(Glc), (D) T and T(Glc), and the aromatic F and (E) schematic presentation of the atoms used for the radial distribution studies: CG of the F, CG2 of T and CB of S. The distance limit was set to 0.45 nm between the CG2 or CB and CG.

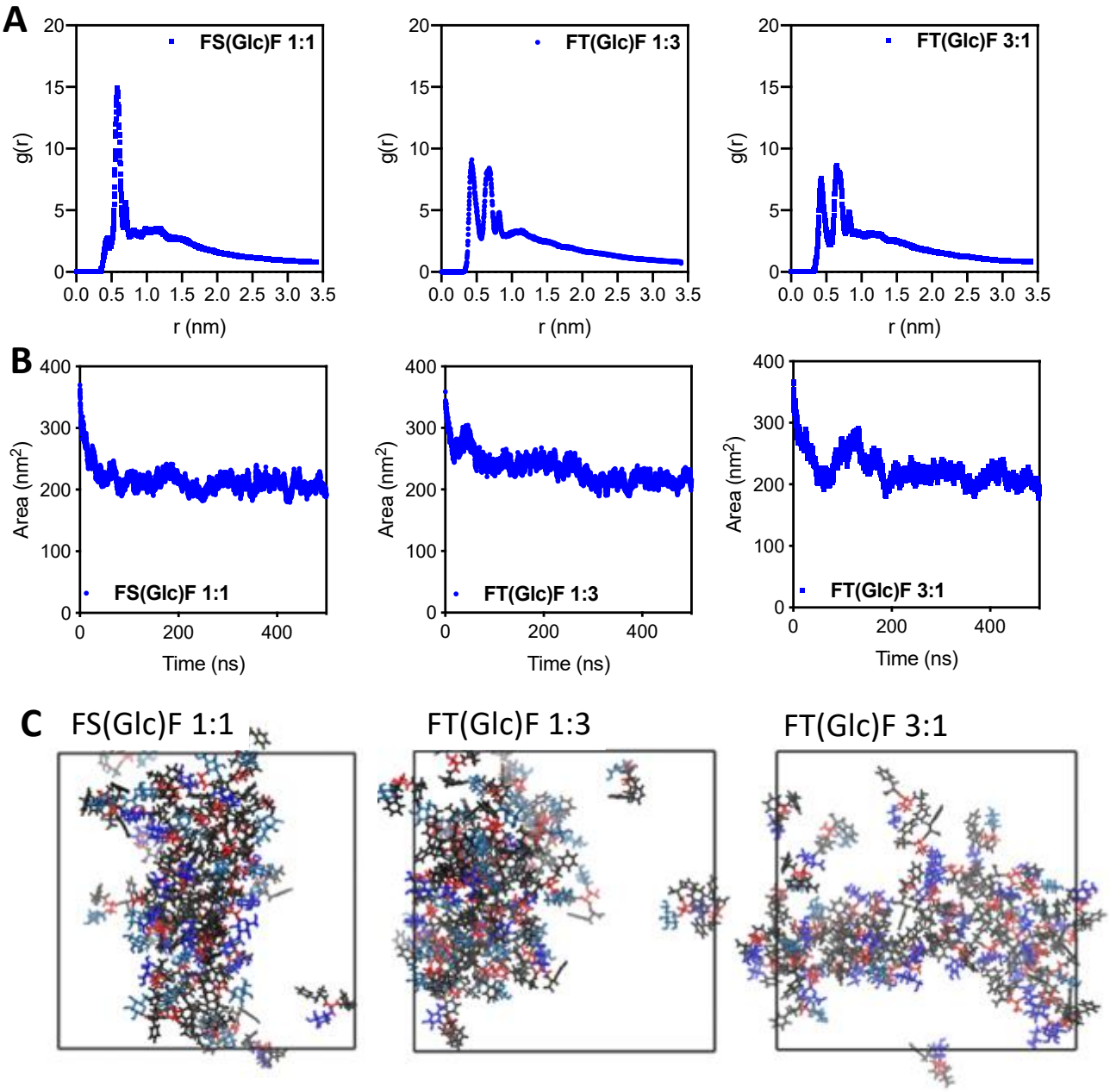



Figure S32. Computational studies for the aggregation of mixtures of glycopeptide anomers at different ratios, namely alpha:beta = 1:1, 1:3 and 3:1. (A) Radial distribution function between S/T and F; (B) Solvent accessible surface area (SASA); (C) screenshot at 500 ns of the simulation.

## Supplementary microscopy data

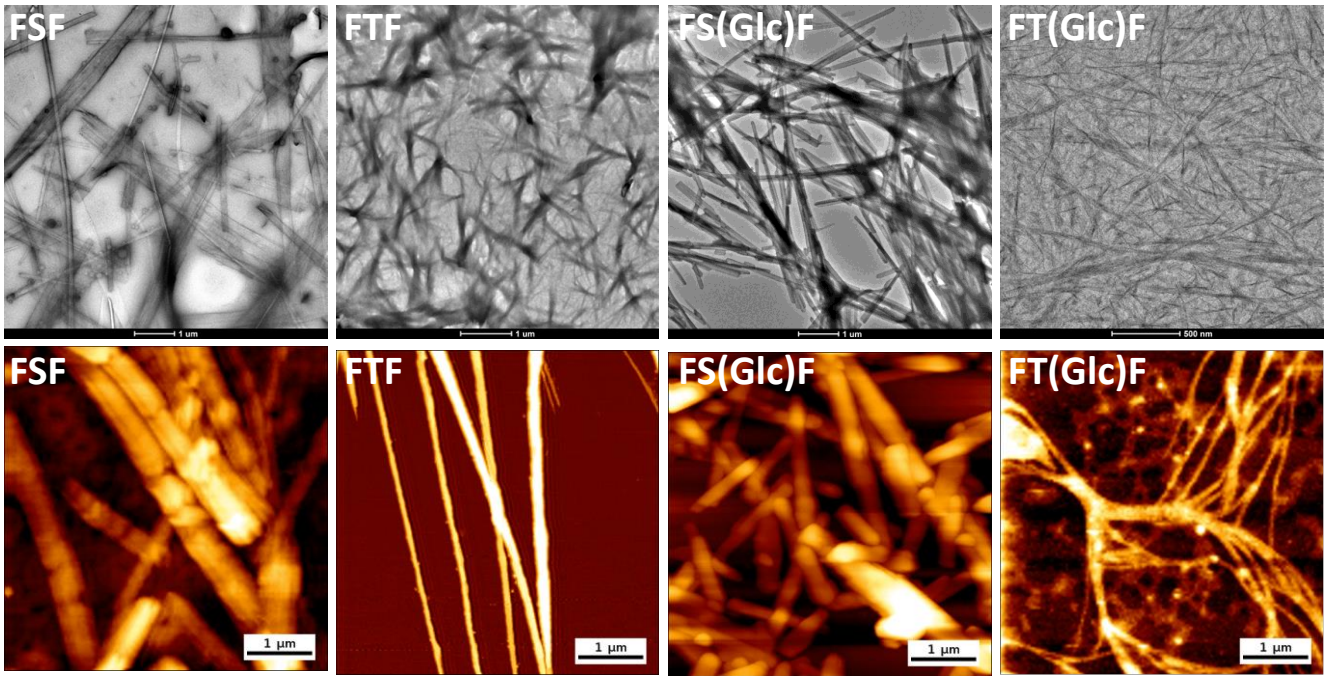



Figure S33. Transmission electron microscopy (TEM, up) and atomic force microscopy (AFM, down) images of the aggregates formed in water (40 mM) by the peptides FSF and FTF and the respective glycopeptides FS(Glc)F and (D) FT(Glc)F.

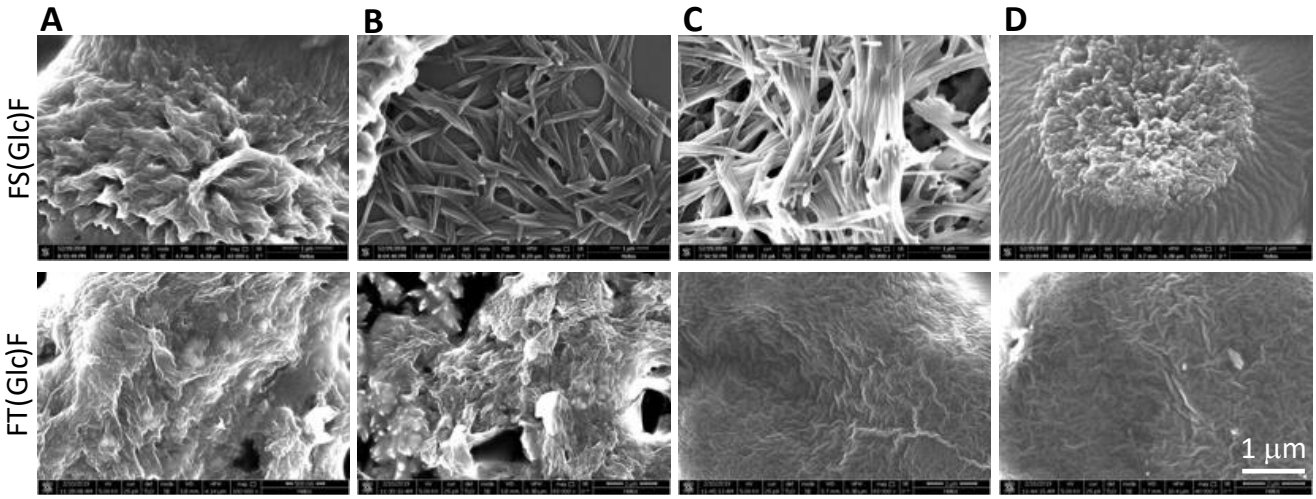



Figure S34. Scanning electron microscopy (SEM) images of the glycopeptides aggregates generated at different concentrations: (A) 10 mM, (B) 20 mM, (C) 40 mM, (D) 65 mM.

**Supplementary data for the aggregates' thermal stability**

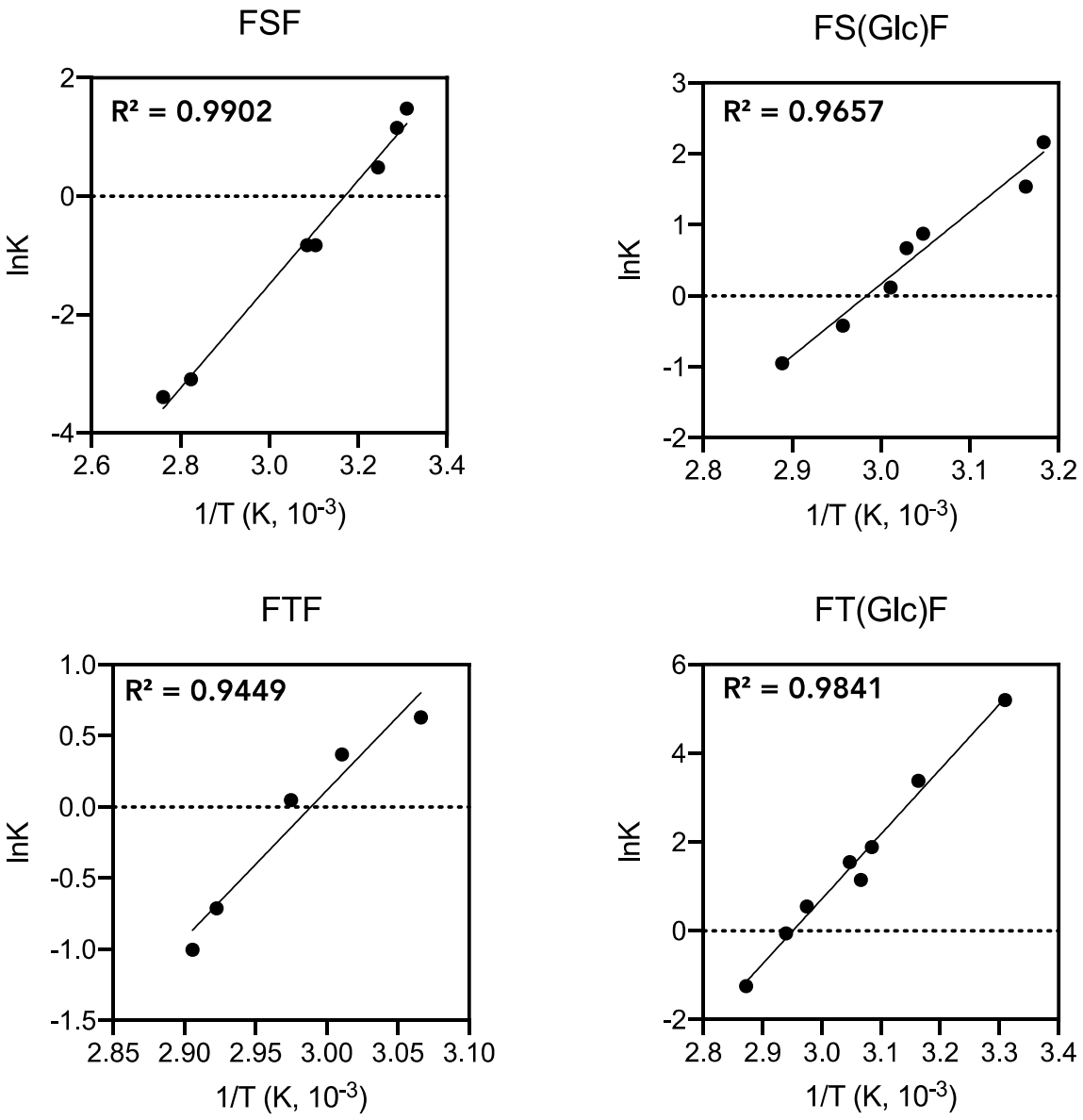



Figure S35. Representative Vant's Hoff plots of the data of the ellipticity curves as a function of temperature used to determine the thermodynamics of aggregation (Figure 4 in the main manuscript).

## REFERENCES

1. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **2012,** *4*.

2. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *Journal of Chemical Theory and Computation* **2012,** *8* (9), 3257-3273.

3. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

4. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E., Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* **2010,** *6* (2), 459-466.

5. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **2008,** *29* (11), 1859-65.

6. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **2009,** *30* (10), 1545-1614.

7. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. F.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 405-413.
8. Park, S. J.; Lee, J.; Qi, Y. F.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W., CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019,** *29* (4), 320-331.
9. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996,** *14* (1), 33-38.
10. Greenfield, N. J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* **2006,** *1* (6), 2527-2535.