Supplementary Material for:

# Glycoprofile analysis of an intact glycoprotein as inferred by NMR spectroscopy

Luca Unione,*[1] Maria Pia Lenza,[1] Ana Ardá,[1] Pedro Urquiza,[1] Ana Laín,[1] Juan Manuel Falcón-Pérez,[1] Jesús Jiménez-Barbero,*[1,2,3] Oscar Millet*[1]

[1]CIC bioGUNE, Bizkaia Technology Park, Bld 800, 48170 Derio, Spain
[2]Basque Foundation for Science IKERBASQUE, 48009 Bilbao, Spain
[3]Dept. Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country, Leioa, Spain

E-mail: jjbarbero@cicbiogune.es, omillet@cicbiogune.es, lunione.atlas@cicbiogune.es

## Methods

**Expression and purification of FcεRIα glycoprotein.** The FcεRI extracellular domain (FcεRIα) was transiently transfected into HEK293T cells, cultured in 20mL serum-free growth medium uniformly $^{13}C$,$^{15}N$ labeled at the time of transfection (U-$^{13}C$,$^{15}N$ Bioexpress 6000, CIL) and grown for 72h[1]. A pool of 4.4 x $10^6$ cells were used for transfection and 25 μg of endotoxin-free DNA, cloned in the pHLsec plasmid[2], were complexed with 50μg of PEI into 5mL of cell culture media. The remaining 15 mL of media, supplemented with 1x penicillin-streptomycin-fungizone, were added to the plate. More than 90% of the cultured cells were efficiently transfected, as determined by fluorescence confocal microscopy of cell culture transfected with GFP plasmid and using the same transfection conditions. Cells were grown at 37 ºC in a humidified atmosphere containing 5% $CO_2$. After harvesting ~2.0 x $10^7$ cells from the confluent p150 plate and sterile filtration on a 0.22 μM filter (Sartorius), the supernatant was purified using a His-tag purification protocol: N-terminally His6-tagged FcεRI was purified through Ni-NTA agarose column (Invitrogen). Up to five washing cycles with 20mM Imidazole in PBS1x buffer were applied, followed by three elution cycles with 200mM Imidazole and a final 500mM Imidazole elution. The purity was assessed by coomassie gel. Imidazole was removed and buffer exchanged through 10 KDa filter (Sartorius). Finally, we checked FcεRI structural/functional properties by Western blot detection and circular dichroism (Figure S1).

**NMR experiments.** FcεRI was resuspended in phosphate saline buffer pH 6.8, 10% $D_2O$ or pure $D_2O$ for NMR analysis. TSP at 1mM concentration was used as internal reference. 280 μL of glycoprotein concentrated up to 60 μM were transferred to a 5 mm Shigemi NMR tube. All NMR experiments were recorded between 283 K and 323 K using an 800 MHz Bruker Avance III spectrometer equipped with a cryoprobe. Standard TOCSY-HSQC (from 20 to 100 ms mixing times), NOESY-HSQC (200 and 300 ms mixing times) and $^1H$-$^{13}C$ HSQC experiments were used to characterize the glycoprofile of FcεRI glycoprotein. Two points of the proton-carbon relaxation-dispersion curve (with $\nu_{CPMG}$ values of 100 and 2000 Hz) were compared to discard the presence of chemical-conformational exchange. Best-TROSY experiments were recorded to get structural data for the peptidic moiety of the glycoprotein or interaction studies with lectin. A 1:1 lectin:glycoprotein molar ratio was used in the molecular interaction studies. Deuterated urea-d4 at 7M concentration was used for structural studies of unfolded FcεRI glycoprotein. All data were analyzed using Bruker Topspin 3.2 and CcpNmr Analysis programs.

The relative concentration of the different glycoforms respect to the protein concentration (60 μM) was estimated by the integration of the unfolded FcεRI HSQC peaks of the branched Man residues, and compared to internal reference TSP (Trimethylsilylpropanoic acid, 1mM). The relative quantification of the N-glycan types is then achieved by dividing each volume by the sum of all the volumes obtained. Error bars are calculated by standard deviation analysis from triplicates. The quantification of the hybrid and triantennary glycans was carried out by subtraction of the integrals arising from branch C to those of branches D, and H to G respectively (Figure 2). The same quantification performed on the folded HSQC of FcεRI was used as a reporter for the segmental dynamics of the N-glycans.

**Semiquantitative analysis N-glycans.** The relative concentration of the different glycoforms was estimated by the integration of the volumes of the anomeric C-H NMR signals of unique monosaccharides in the different N-glycans species. Specifically, the branched Man residues (A-H in figure 1) are unique reporters of the different N-glycans species. The normalized value of Man A has been used as reporter of pauci-mannose specie, Man C as reporter of high-mannose N-glycans, F for bi-antennary and H for tetra-antennary complex type. The contribution for hybrid type N-glycans and tri-antennary has been calculated as the difference of (D-C) and (G-H) because not unique resonance can report for them. The relative quantification (%) of the N-glycan types is then achieved by dividing each volume by the sum of all the volumes obtained. Error bars are calculated by standard deviation analysis from duplicates and error propagation has been calculated.

**Molecular Dynamics Simulation.** The starting structure for the MD simulation of the heterogeneously glycosylated FcεRI was generated using the reported X-ray structure (PDB code 1J88) as template[4]. The only GlcNAc residues observed at each glycosylation site were replaced by the full N-glycan chains, built with the GLYCAM builder in GLYCAM-Web (http://glycam.org/). The starting structure was first minimized within the same web server, using the ff14SB protein force field (for the protein part) and the Glycam_06j-1 force field (for the glycan residues). The generated structure was used as starting geometry for 1 μs MD in explicit water, using the Amber16 program with the same force fields. In particular, the starting structure was solvated in a 12 Å octaedral box of explicit TIP3P waters, and counterions were added to maintain electroneutrality. In order to fill all of the protein cavities with water molecules, a previous minimization for only solvent and ions was made. To reach a low energy starting structure, the entire system was minimized with a higher number of cycles, using the accurate steepest descent algorithm. For the MD run the system was subjected to two fast MD simulations (heating and equilibration) before starting the real one: (1) 200 ps of MD heating the whole system from 0 to 300 K, using NVT ensemble and a cutoff of 10 Å, followed by (2) equilibration of the entire system over 200 ps at 300 K using NPT ensemble and a cutoff of 10 Å. A relaxation time of 2 ps was used in order to equilibrate the entire system in each step. The equilibrated structure was the starting point for the final MD simulations at constant temperature (300 K) and pressure (1 atm). The MD was recorded without constraints, using an NPT ensemble with periodic boundary conditions, a cutoff of 10 Å, and the particle mesh Ewald method. Coordinates and energy values were recorded every 50000 steps (20 ps) for a total of 50000 MD models. The data analysis was performed using 1000 frames extracted from the MD simulation. The first 100 ns of the simulation data were discarded for thermalization and the remaining 900 ns were used for the data analysis.

**Enzymatic reactions with glycosidases.** The samples for the molecular interaction studies in between FcεRI and hGal3 were subjected to specific enzymatic deglycosylations. First, the terminal 3'-sialic acid residues were removed by treating the glycoprotein with the specific α2-3 neuraminidase S (New England Biolabs). Second, the galactose residues were removed with β1-4 Galactosidase (New England Biolabs). 40 units of each enzyme were added to 280 μL buffer (PBS1x 5mM $CaCl_2$, pH 6.8) containing 250 μg of glycoprotein. The reactions were incubated at 37 ºC O.N.

**REFERENCES**


(1) Barbieri, L.; Luchinat, E.; Banci, L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc* **2016,** *11* (6), 1101-1111.

(2) Aricescu, A. R.; Lu, W.; Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica* **2006,** *62* (Pt 10), 1243-1250.

(3) Urquiza, P.; Lain, A.; Sanz-Parra, A.; Moreno, J.; Bernardo-Seisdedos, G.; Dubus, P.; Gonzalez, E.; Gutierrez-de-Juan, V.; Garcia, S.; Erana, H.et al. Repurposing ciclopirox as a pharmacological chaperone in a model of congenital erythropoietic porphyria. *Sci Transl Med* **2018,** *10,* eaaat7467.

(4) Garman, S. C.; Sechi, S.; Kinet, J. P.; Jardetzky, T. S. The analysis of the human high affinity IgE receptor Fc epsilon Ri alpha from multiple crystal forms. *Journal of molecular biology* **2001,** *311* (5), 1049-1062.

**Table S1. Glycan $^{13}C/^{1}H$ chemical shifts assignment (referenced to TSP).** Residue letter code corresponds to that defined in Fig.1 of the main text.

| Residue | Residue letter code | C1/H1 | C2/H2 | C3/H3 | C4/H4 | C5/H5 | C6/H6, H6' | C7/H7 | C8/H8 | C9/H9, H9' |
|---|---|---|---|---|---|---|---|---|---|---|
| GlcNAc | η | 81.3/5.05 | 56.7/3.83 | 74.5/3.71 | 81.6/3.74-81.8/3.92[c] | 78.0/3.74 | 68.8/3.94, 3.79 | | | |
| Fuc | Q | 102.3/4.88 | 71.2/3.83 | 72.3/3.91 | 74.8/3.79 | 69.8/4.16 | 18.2/1.27 | | | |
| GlcNAc | J | 104.1/4.68 | 57.9/3.79 | 76.1/3.77 | 82.6/3.72 | 77.3/3.60 | 62.8/3.89, 3.77 | | | |
| GlcNAc | I | 104.4/4.59 | 57.9/3.79 | 76.1/3.77 | 82.3/3.71 | 77.3/3.60 | 62.9/3.89, 3.77 | | | |
| Man | N | 103.4/4.76 | 73.0/4.25 | 83.5/3.79 | 68.4/3.83-68.6/3.89[b] | 76.5/3.77 | 68.3/3.96, 3.77 | | | |
| Man | E | 102.5/5.17 | 79.7/4.21 | 72.4/3.93 | 70.3/3.53 | 76.6/3.79 | 65.6/3.87, 3.66 | | | |
| Man | F | 100.0/4.92 | 79.8/4.09 | 74.6/3.88 | 70.3/3.53 | 75.9/3.61 | 64.5/3.91, 3.64 | | | |
| GlcNAc | O | 102.6/4.59 | 57.8/3.73 | 74.9/3.74 | 81.9/3.73 | 77.4/3.58 | 63.1/4.0, 3.85 | | | |
| GlcNAc | O' | 102.8/4.59 | 57.8/3.73 | -/- | 72.8/3.48 | 78.9/3.46 | | | | |
| GlcNAc | P | 104.5/4.55 | 57.8/3.73 | -/- | 81.9/3.70 | 77.4/3.58 | 63.1/4.0, 3.85 | | | |
| Man | G | 100.2/4.87 | 80.0/4.08 | 74.5/3.90 | 70.4/3.43 | 75.7/3.62 | 68.8/3.95, 3.79 | | | |
| Man | H | 102.0/5.16 | 79.1/4.24 | 72.3/3.91 | 80.9/3.65 | 74.8/3.75 | | | | |
| Gal | K | 106.3/4.47 | 73.9/3.59 | 75.4/3.67 | 71.5/3.95 | 78.2/3.74 | 63.8/3.78, 3.75 | | | |
| Gal | L | 105.8/4.53 | 72.2/3.61 | 78.5/4.11 | 70.3/3.97 | 78.2/3.74 | 63.8/3.78, 3.75 | | | |
| Gal | M | 106.4/4.45 | 73.9/3.59 | 75.4/3.67 | 71.5/3.95 | 76.5/3.81 | 66.1/3.98, 3.59 | | | |
| Neu5Ac | Z | | | 42.9/1.70, 2.68 | 71.1/3.71 | 54.8/3.79 | 75.0/3.72 | 71.4/3.60 | 74.6/3.89 | 63.0/3.67, 3.82 |
| Neu5Ac | X | | | 42.5/1.79, 2.77 | 71.1/3.71 | 54.6/3.83 | 75.4/3.72 | 71.1/3.62 | 74.6/3.89 | 63.0/3.67, 3.82 |
| Man | A | 105.1/5.12 | 72.9/4.07 | 73.3/3.85 | 69.7/3.66 | 68.5/3.72 | 62.8/3.83,3.63 | | | |
| Man | B | 102.3/4.92 | 72.8/3.99 | 73.3/3.85 | 69.7/3.69 | 68.5/3.72 | 62.8/3.83,3.63 | | | |
| Man | C | 103.7/5.35 | 81.3/4.12 | 72.8/3.99 | 71.0/3.79 | 76.3/3.77 | 63.9/3.89,3.77 | | | |
| Man | D | 102.7/4.89 | 72.4/4.14 | 81.4/3.93 | 68.6/3.85 | 73.5/3.85 | 68.8/3.95, 3.79 | | | |
| Man | R | 101.1/5.12 | 81.4/4.03 | 72.8/3.99 | 71.0/3.79 | 75.8/3.68 | 63.9/3.89,3.77 | | | |
| Man | S | 103.5/5.39 | 81.3/4.12 | 72.8/3.99 | 71.0/3.79 | 76.3/3.77 | 63.9/3.89,3.77 | | | |
| Man | T | 103.5/5.29 | 81.3/4.12 | 73.0/4.01 | 71.0/3.79 | 76.3/3.77 | 63.9/3.89,3.77 | | | |
| Man | U,V,W | 105.0/5.07 | 73.0/4.09 | 73.5/3.89 | 71.1/3.72 | 76.3/3.77 | 63.9/3.89,3.77 | | | |
| GlcNAc | Y | 105.4/4.75 | 58.5/3.91 | 77.4/3.56 | 78.7/3.91 | 77.3/3.56 | 62.4/4.00, 3.87 | | | |
| Fuc | Θ | 101.3/5.13 | 70.6/3.71 | 72.4/3.8 | 75.4/3.78 | 69.5/4.80 | 18.1/1.18 | | | |
| Gal | Ω | 104.8/4.46 | 74.6/3.65 | 73.9/3.50 | 71.1/4.15 | 77.7/3.72 | 63.9/3.74,3.71 | | | |

[a] Chemical shifts that could not be determined accurately due to overlap (-/-).
[b] When present in high-Mannose type N-glycans.
[c] When the residue is not-fucosylated.

**Figure S1. Structural functional study of FcεRIα.** a) Circular dichroism spectrum of soluble FcεRIα protein. Measurement was performed at 25 °C in a 0.02 cm path length cell. Protein concentration was 5 μM, in 10 mM phosphate saline buffer pH 6.8. The spectrum is essentially identical to that reported in inhibition and binding studies of the FcεRIα receptor. b) Western blot analysis of the FcεRIα induced expression in human HEK293T cells. 1th line corresponds to the supernatant from transfected cell growth in uniformly $^{13}C$, $^{15}N$ labeled medium. 2th line corresponds to the supernatant from not transfected cells (negative control). 3th line corresponds to the supernatant from transfected cell growth in common unlabeled medium. 4th and 5th lines correspond to the cell pellet from labeled and unlabeled growth media, respectively. The degree of glycosylation for this glycoprotein is essentially independent from the cell growth media, whereas a significant difference in protein molecular weight is evident depending from its localization, inside cell or secreted. This difference reflects the higher degree of glycosylation for a protein that has passed through the Golgi and the one that is decorated with only the minimal glycosylation from the ER.

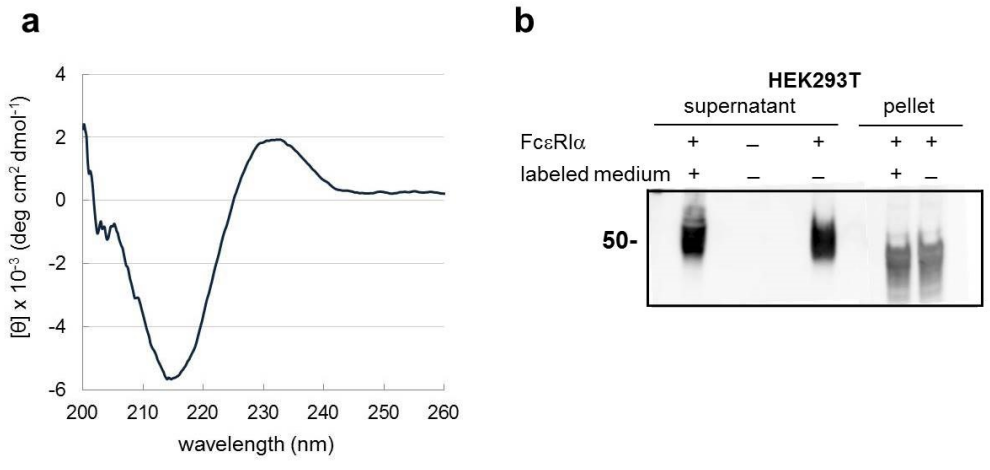

**Figure S2. Validation of NMR assignment through the use of trimming enzymes.** a) Glycans are assigned using the one-letter code, as referred in the panel. b) FcεRIα glycoprotein was sequentially treated with Mannosidase, Fucosidase, Neuraminidase S, Neuraminidase A, Galactosidase and N-Acetylglucosaminidase. A $^{1}H$-$^{13}C$-HSQC spectrum was acquired after each enzymatic reaction. Intensity quantification was done by denaturing the protein after all the enzymatic reactions. The reaction trimming percentages are indicated in red, while the new signals corresponding to the trimmed glycans are shown in green. All the spectra other than the β-1,4-galactosidase and the β-glucosaminidase were acquired with 1536 x 984 points. These two spectra were recorded using 1536 x 430 points. Signals for these two spectra may look slightly broader in the carbon dimension, without affecting the chemical shift nor the quantification.

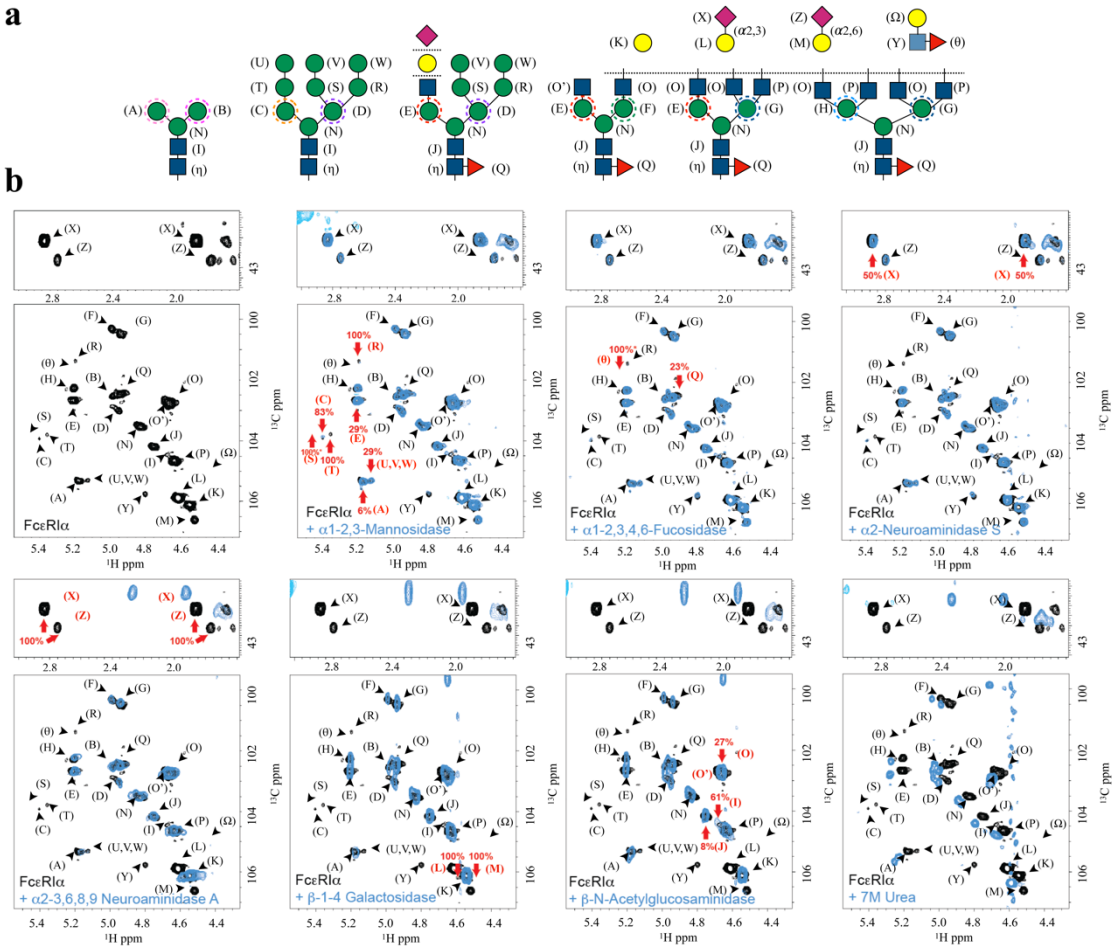

**Figure S3. Assigned NMR spectrum of the glycan moieties.** a) Glycans are assigned using the one-letter code, as referred in the panel. b) $^{1}H$-$^{13}C$-HSQC spectra comparing the chemical shift and magnitude intensity signals for selected sugar residues in the N-glycans, acquired in either, folded (red) or unfolded (blue) protein states. Sugar NMR signals which are sensitive to protein structure experience much pronounced chemical shift perturbation as well as magnitudes intensity gains in the unfolded states because of the increased molecular tumbling of an unstructured molecule. The red colored dashed traces highlight the differences among the NMR signals originating from the Man residues in the ($\alpha$1-3) arms with respect to those on the ($\alpha$1-6). The green colored dashed traces highlight the very pronounced differences for Man residues from the high mannose N-glycans, suggesting that these N-glycans are effectively packed within structured proteins domains in the native state, whereas solvent exposed in the unfolded state. Finally, NMR signals from inner–core residues, which also experience substantial intensity gain, are also indicated by the arrows.

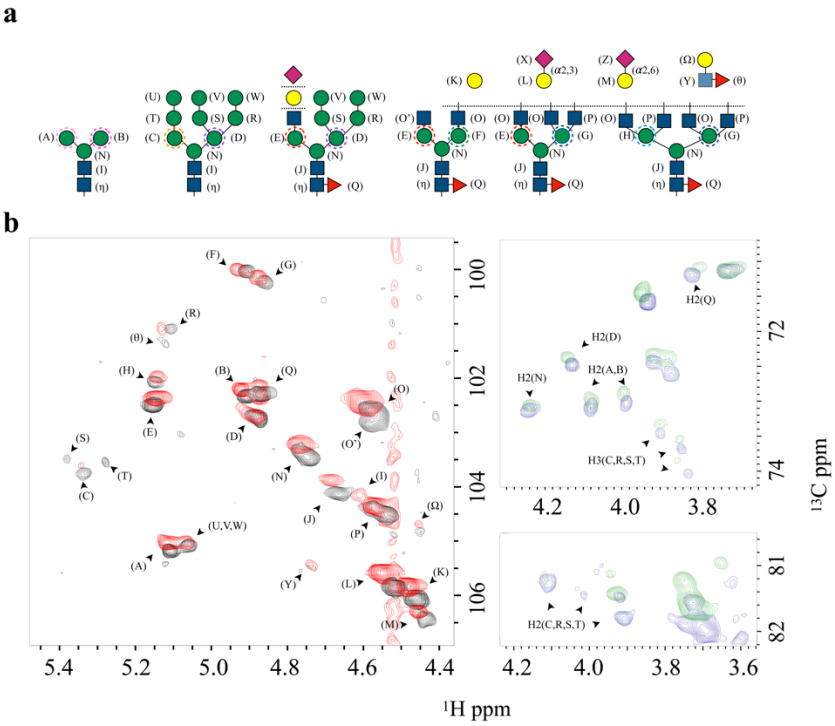


**Figure S4. Front and back view of the last frame from the MD simulation taken as representative structure.** The glycan structures consider the intrinsic heterogeneity for this protein from HEK 293 cells. Each N-glycan type is colored as follows: high mannose N-glycan (green); hybrid N-glycan (orange); complex N-glycans (red).

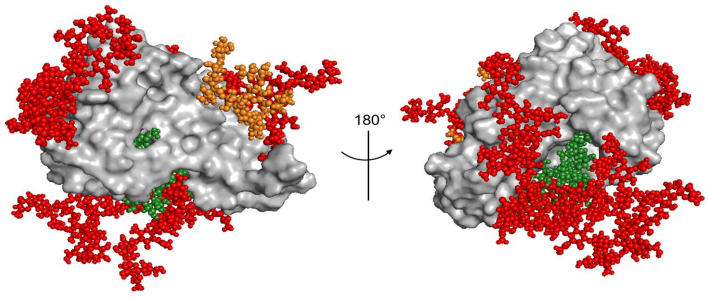

**Figure S5. Comparison of the anomeric region between WT FcεRIα and N132A mutant.** a) Signals that disappear or decrease intensity upon mutation are shown by a red asterisk or a red arrow respectively. b) Anomeric region of the $^{1}H,^{13}C$-HSQC spectrum for WT and N132A FcεRIα. Glycans are assigned using the one-letter code, as referred in the top panel. The analysis of the glycan content for this mutant demonstrates a significative reduction in the population of high mannose species

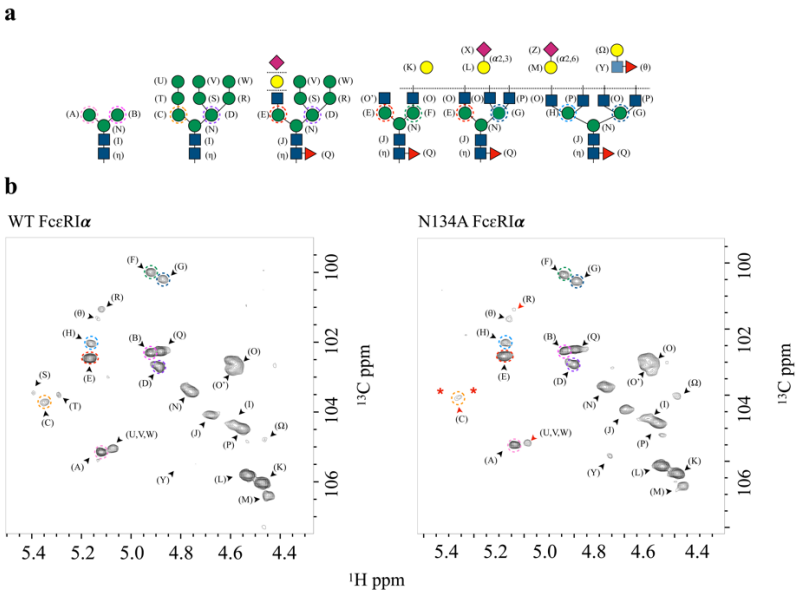


**Figure S6. Model for glycoprotein-hGal3 interaction.** The model for the lectin-glycoprotein interaction was built by manual docking the N-glycan at Asn137 of the glycoprotein obtained from the MD simulation approach to the crystal structure of the human galectin 3 bound to N-Acetyllactosamine (pdb code 1A3K). Here the proteins are represented in cartoon with green surface (hGal3) and grey surface (FcεRIα). The interaction is mediated by the tetra-antennary N-glycan at Asn137. All N-glycans are represented as sticks with unique color code: bi-antennary N-glycan at Asn18 (blue); bi-antennary N-glycan at Asn39 (light-blue); hybrid N-glycan at Asn47 (light-orange); tetra-antennary N-glycan at Asn71 (red); high-mannose N-glycan at Asn132 (green); tetra-antennary N-glycan at Asn137 (pink); tetra-antennary N-glycan at Asn163 (orange).