The chicken-and-egg problem in glycobiology

Exploring the conformational phase space of N-linked glycans using enhanced MD and sketch-map analysis

I. L. Grothaus, L. Colombi Ciacchi*, Hybrid Materials Interfaces Group, Faculty of Production Engineering and Bremen Center of Computational Materials Science, University of Bremen*

In Short

- N-linked glycans are diverse post-translational protein modifications, existing in many different structures
- The core structural feature is conserved in all N-linked glycans
- We wonder if the conserved glycan core structure is also the energetically most favourable one
- Therefore we constructed many different core configurations varying in linkage type
- Finally, the usage of enhanced sampling techniques is necessary to sample all relevant conformers in order to assess the free energy difference between configurations

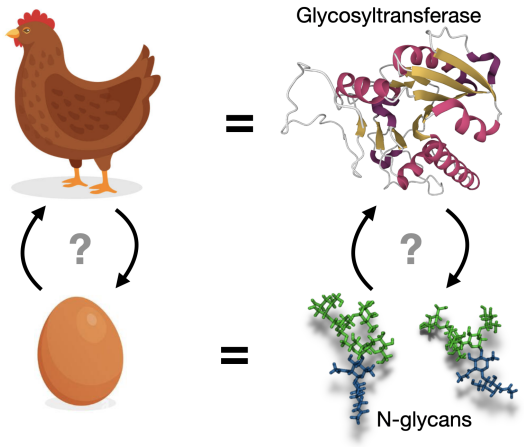

Figure 1: *Schematic representation of the analogy between the classical "chicken-and-egg" problem to our research question if glycosyltransferases shaped the N-glycan structures or if energetically favourable structures were favoured, shaping the catalytic sides of glycosyltransferases.*

N-glycans represent the most divers post-translational modification of proteins but have long been overlooked in computer simulations, despite their omnipresence in the cell occuring for over 50% of mammalian proteins [1]. These polysaccharides are the most abundant form of glycosylation of proteins and are covalently connected to asparagine residues in the polypeptide chain via glycosidic linkages [2]. The process of glycosylation starts after the translation of proteins by ribosomes into the interior of the Endoplasmatic reticulum (ER), where N-glycans can be attached with the help of enzymes called glycosyltransferases (GTs). By different type of GTs, N-glycans are further processed to diverse structures in the ER and Golgi Apparatus. The structure of N-glycans are completely dependent on which GTs are acting and at which time point, leading to branched structures built up of single monosaccharide units connected by glycosidic linkages. They exhibit flexible dynamic behaviours, also depending on the type of linkage and which monosaccharides are connected. In principle, one can imagine many differently coloured beads on a string (representing the different monosaccharides chained together), forming a tree-like structure through branches and junctions. Over the years, different biological functions have been assigned to N-glycans like involvement in folding and stability of proteins or as target structures for lectins and antibodies, illustrating their importance in the majority of cellular processes [2].

As already mentioned, GTs play a fundamental role in the construction of N-glycans, as each GT can only catalyse one transfer reaction. A specific GT can only transfer a specific monosaccharide type and only create a specific linkage type. All this presupposes that in order to create a specific N-glycan, all necessary GTs need to be present in the cell in the right location at the right time. It is conspicuous that certain linkage types are recurring and the stem of each N-glycan is even conserved among all structures. We therefore raise the question "Where does this conservation originates from?" Is it solely depending on the evolution of GTs that shaped the N-glycan structures, or are the conserved structural features simultaneously the most energetically favourable ones and GTs evolved to construct exactly these? This foundational research question can be compared to the classical "chicken-and-egg" problem, where it is let unanswered if the chicken or the egg existed first (Figure 1). Within this analogy, one can compare the GTs to the chicken and the N-glycans represent the egg, leaving the structural conservation unanswered.

In order to address this question on a computational level, we aim to compare the free energy differences between different N-glycan configurations to

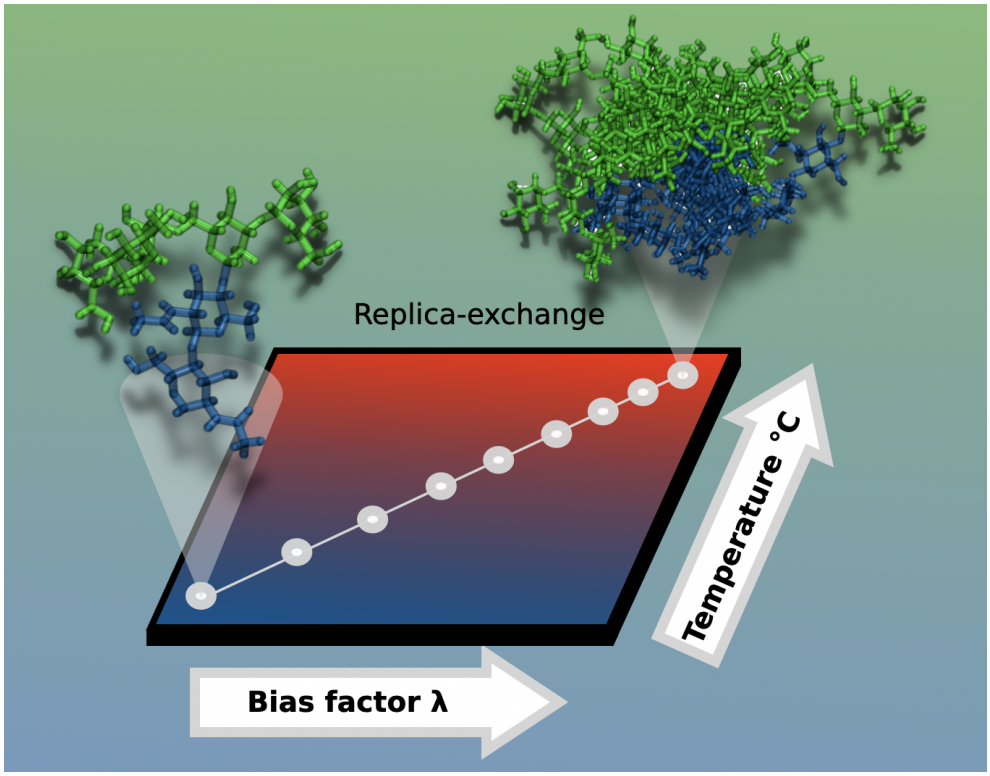

Figure 2: *Depiction of the employed enhanced sampling simulation technique, where N-glycan conformations are sampled via heating of the system and application of additional biases to explore more conformers.*

asses if the conserved structure represents the minimum energy structure. The different N-glycan configurations are altered in their linkage and monosaccharide type to test all possible combinations that are available within biochemical constraints. One can employ molecular dynamics simulations to sample different conformation of each configuration over the trajectory length and use the potential energy associated to each conformer for the calculation of the free energy of each configuration. We envision to not only use plain molecular dynamics, but rather apply our previously developed enhanced-sampling molecular dynamics scheme called REST-RECT [3]. It combines two replica-exchange methodologies, where one heats the N-glycan in order to overcome any energy barrier within the whole system and the second simultaneously enhances transitions of all specific energy barriers that are already known for the system to be critical to overcome (Figure 2. The second part uses concurrent one-dimensional energy potentials in the framework of metadynamics, which can in fact capture effectively all biologically relevant global conformers of branched glycans [3].

The outlined approach highly requires High Performance Computing (HPC) infrastructures as each of the simulations require several hundreds of cores for calculation. Furthermore, enhanced sampling techniques require the performance of multiple simulations in parallel and therefore run on even more processors. This setup can be easily parallelized on HPC infrastructures and results are much faster obtained when comparing the actual running time.

Additionally, the overall aim of this study is to test all biochemically possible N-glycan configurations for one hand-selected specific N-glycan structure, summing up to 30 different structural setups. This ensures that are strong and clear statement can be made about the most energetically favourable structures.

WWW

http://www.hmi.uni-bremen.de

More Information

[1] R Apweiler, H Hermjakob, and N Sharon. *Biochimica et biophysica acta* **1999** 1473(1):48. doi:10.1016/s0304-4165(99)00165-8

[2] Ajit Varki. *Essentials of Glycobiology* **2009**.

[3] I Grothaus, G Bussi, and L Colombi Ciacchi. *bioRxiv* **2022**. doi:10.1101/2022.06.17.496605

DFG Subject Area

310-01