

Algorithmic Biosynthesis: Staged Glycan Assembly

Anjali Jaiman and Mukund Thattai

February 28, 2017

Introduction

O-glycans are a common post-translational modification which attaches a branched oligosaccharide to a serine or threonine amino acid residue on a folded protein. O-glycan biosynthesis is ubiquitous [4] and well-conserved [DATA] across eukaryotes, and yet, sequencing efforts have revealed tremendous structural innovation and diversity. Mammalian o-glycans are composed of a small set of conserved monosaccharides and sugar linkages [1, ?] [Supplemental Fig1], and mainly decorate the extracellular surface of membrane bound and secreted proteins.

O-glycan assembly is processive; a glycosyltransferase shifts a single specific monosaccharide-type from a nucleotide sugar to a particular “acceptor” monosaccharide on the growing glycan structures, catalyzing a carbon-carbon bond between two particular carbons on the two sugar monomers [Figure 1B]. Thus the minimal specificity of a glycosyltransferase is the acceptor monosaccharide, the donor monosaccharide and the carbon linkage.

It is possible, and in some cases clear that there are additional specificities [cite]. A glycosyltransferase may have post-node specificity, recognizing additional branches already attached to the acceptor monomer. It may have pre-node specificity, recognizing a pattern of monosaccharides before the acceptor monosaccharide. Or a GTase may have off-node specificity, recognizing a motif on an entirely separate branch of the glycan [Supplemental Figure 1]. These additional specificities can be either activating or inhibitory, and from early and incomplete in vitro studies, many GTases appear to have hierarchical substrate specificity—meaning that they recognize many different substrates with differing strengths—thus as a GTase acts on different substrates, it reacts with different kinetic rates [?].

But unlike proteins, which are encoded in DNA, glycans are not constructed from a template.

How are the set of glycans structures produced by a cell limited and controlled? Many branched glycan structures, as they are being assembled have more than one acceptor monosaccharide available for glycosylation. Given that many GTases are locally acting and promiscuous, the

web of reactions forms a complex network which, when drawn out graphically, indeed resembles a network of diverging and converging paths leading to a number of different fully-extended N-glycan structures.

The compartments of the Golgi are not biochemically homogeneous.

microheterogeneity of N-glycans on glycoproteins can thus be attributed to the nature of the glycosylation reactions and the mixing characteristics of the Golgi apparatus as proteins traverse the system.

Usefulness:

References

- [1]
- [2] Structure determination of the disialylated poly-(n-acetyllactosamine)-containing o-linked carbohydrate chains of equine chorionic gonadotropin. *Glycoconjugate Journal*, 11, 1993.
- [3] Recombinant muc1 probe authentically reflects cell-specific o-glycosylation profiles of endogenous breast cancer mucin. high density and prevalent core 2-based glycosylation. *Journal of Biological Chemistry*, 277, 2002.
- [4] Rolf Apweiler, Henning Hermjakob, and Nathan Sharon. On the frequency of protein glycosylation, as deduced from analysis of the swiss-prot database. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1473, 1999.

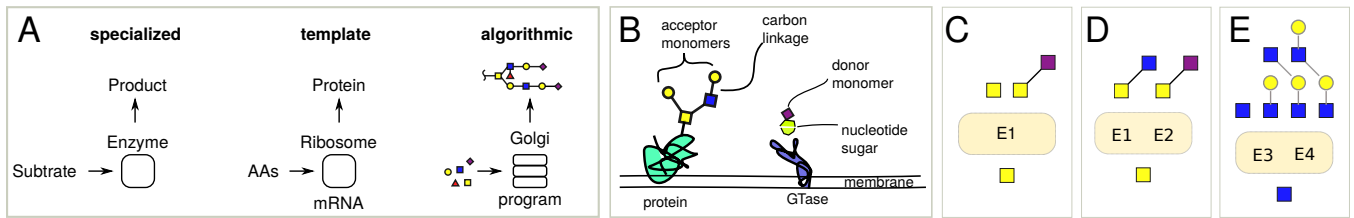


Figure 1: A) Schematic comparing different biosynthesis mechanisms: a specialized enzyme contains the exact information needed to catalyze the synthesis of a particular molecule, a metabolite for example; a ribosome uses the mRNA as a template use to “copy” the information into a peptide; the Golgi apparatus acts as a program, the output of which produces a particular set of glycans. B) GTases transfer a single donor monosaccharide from a nucleotide sugar onto an acceptor monosaccharide on the growing glycan. Most GTases have a minimal specificity of donor, acceptor and carbon pair linkages. Thus, the NeuAc (purple diamond) can be attached to either one of the galactoses (yellow circles). C) Stochasticity inherent in these reactions leads to a diversity of products. D) Competition between enzymes for the same substrate/acceptor monosaccharide also results in a heterogeneous set of structures. E) Polymerization also significantly increases the total set of structures by producing an infinite set of chains containing differing numbers of polymeric repeats.]

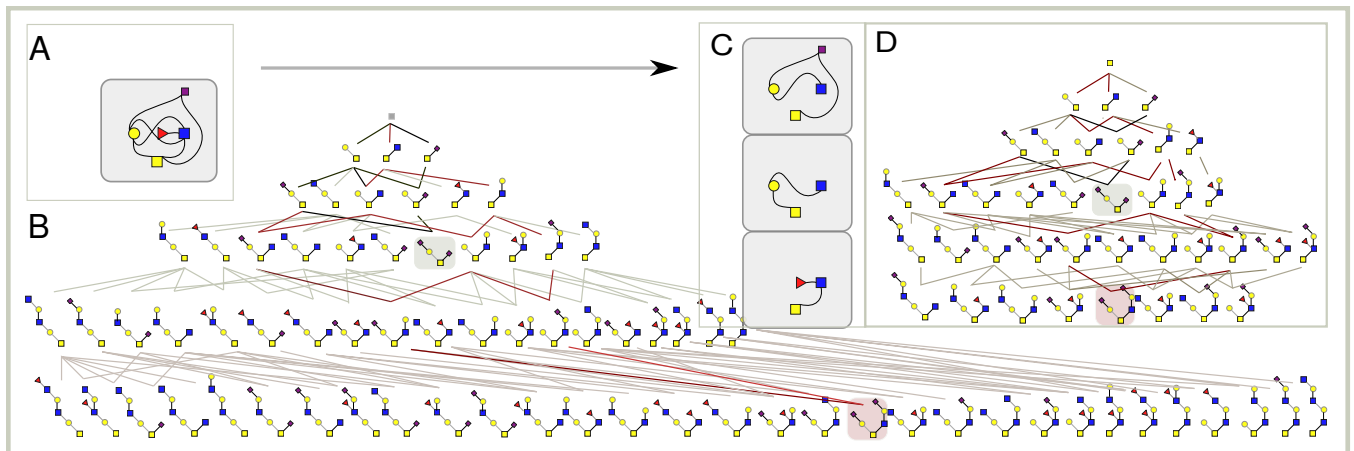


Figure 2: A) Sources of diversity: Incompleteness B) Sources of diversity: Competition. C) Sources of diversity: Polymerization.

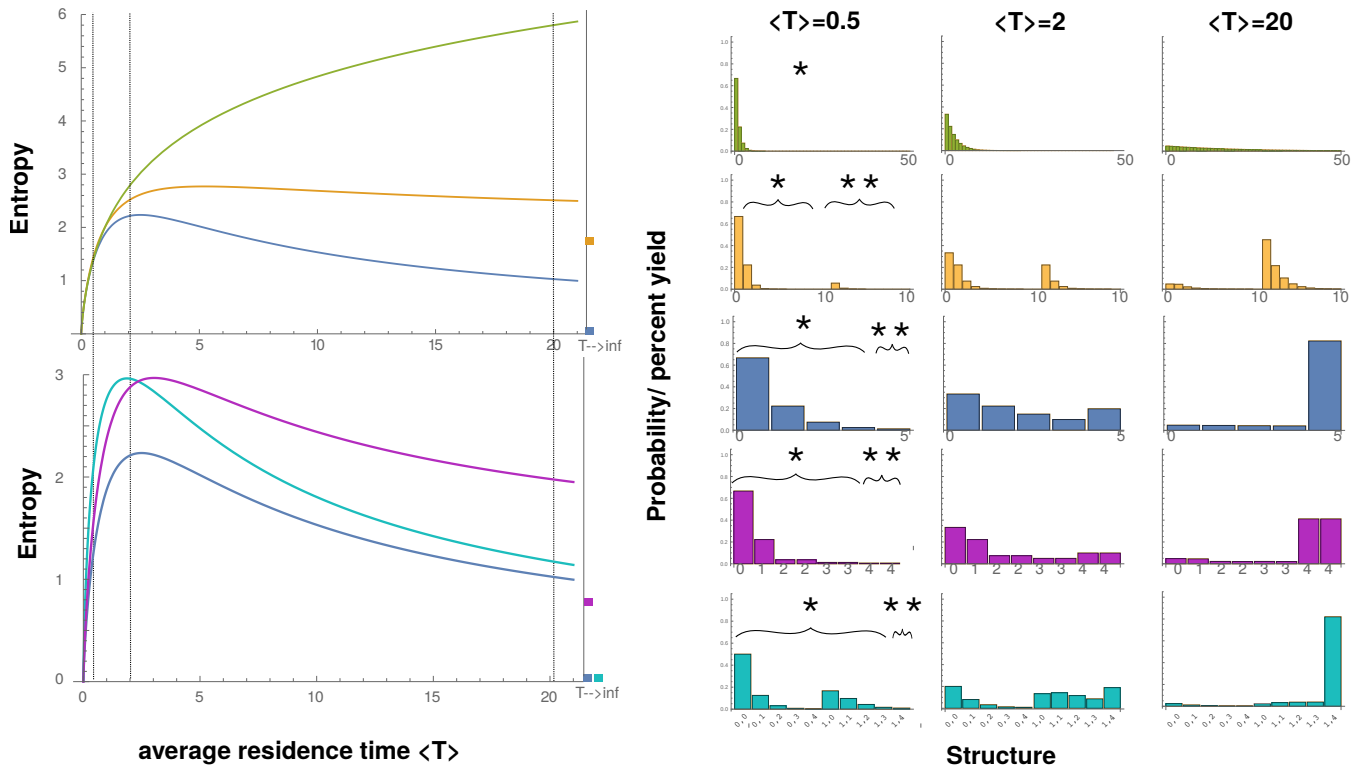
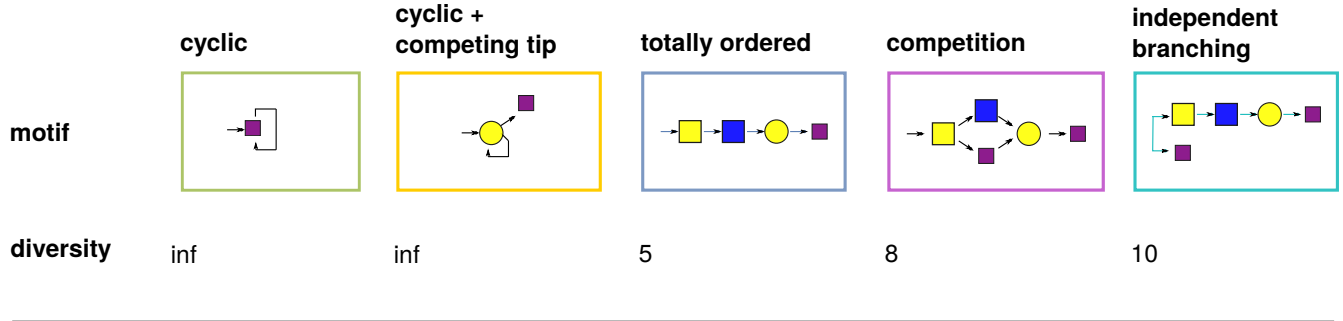


Figure 3: A) Entropy as a function of residence time comparing a cyclic reaction network (blue), a cyclic reaction network with a competing tip (red) and a totally ordered reaction network (orange). B) Entropy as a function of average residence time comparing a totally ordered reaction network (orange), a partially ordered reaction network (light blue), and totally ordered reaction network with competition (green).

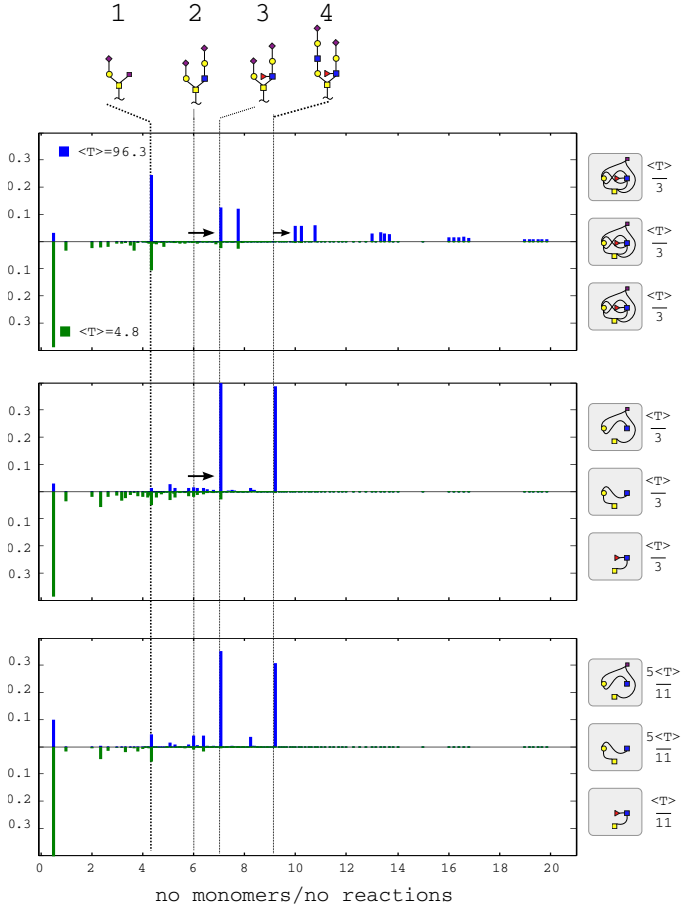


Figure 4: A) Distribution of glycosylation products formed by putting all required locally acting enzymes in each of the three compartments. B) Distribution of glycosylation products when the required enzymes are compartmentalized according to a "best" partition. C) Distribution of glycosylation products when the average residence times in each compartment of the "best" partition are not taken as equal. Instead, they are taken 1:5:5. In blue, Gillespie simulations were run with total average time is taken at 96.3 for each of the three programs. In mirrored plot (green), the three simulations were run with total average time at 4.8. D) Compares the KL divergences of the three programs above as the distance between a uniform distribution of the four structures and the program distribution as a function of average residence time.

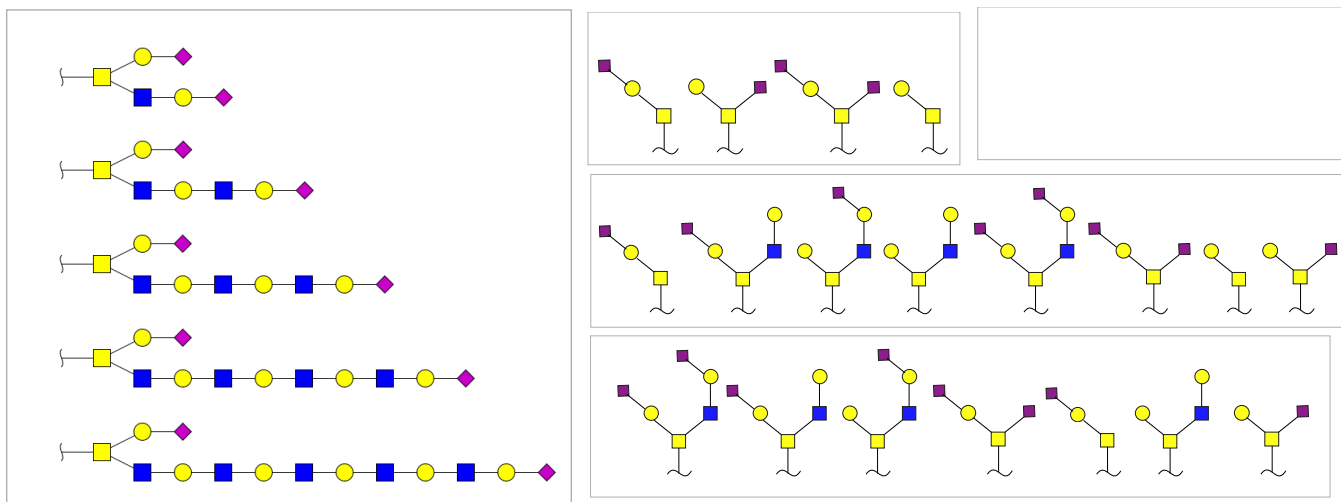


Figure 5: A) O-glycans found on equine chorionic gonadotropin [2]. O-glycans found on a recombinant MUC1 expressed in B) T47D, C) MDA-MB231, D) ZR-75-1 human breast adenocarcinoma cell lines [3]. All structure sets contain one or more of the glycan structures found in hCG o-glycan set described above and contain a subset of the linkages found. They are ordered left to right (or top to bottom) in terms of greatest to lowest yields. Trace amounts of other species reported by [3] were not included.