Algorithmic Biosynthesis: Staged Glycan Assembly

Anjali Jaiman and Mukund Thattai

March 21, 2017
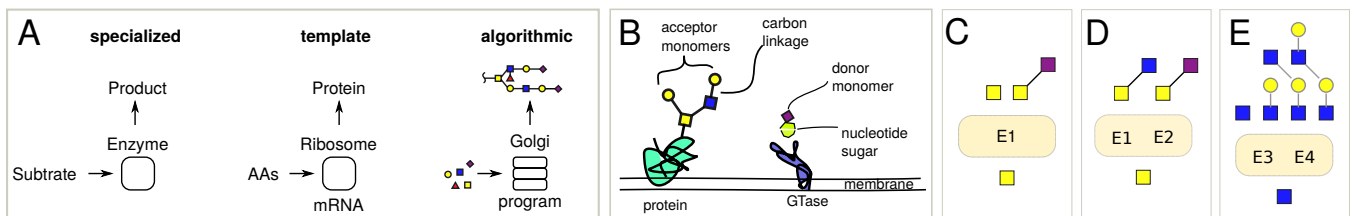

Figure 1: A) O-glycosylation mainly takes place in the Golgi apparatus. Proteins from the ER are shipped to the Golgi apparatus, where they are glycosylated before they are transported to the plasma membrane. B) Locally acting enzymes. B) A glycan is an oligosaccharide. A GalNac monosaccharide is covalently attached to a serine or threonine residue on a glycoprotein. Each monosaccharide ring has multiple carbons that can bind to other monosaccharides. This allows glycans to branch. C) In the Golgi apparatus 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. D) Some GTases common to mammals [see Supplemental Figure 1 for full reaction network]
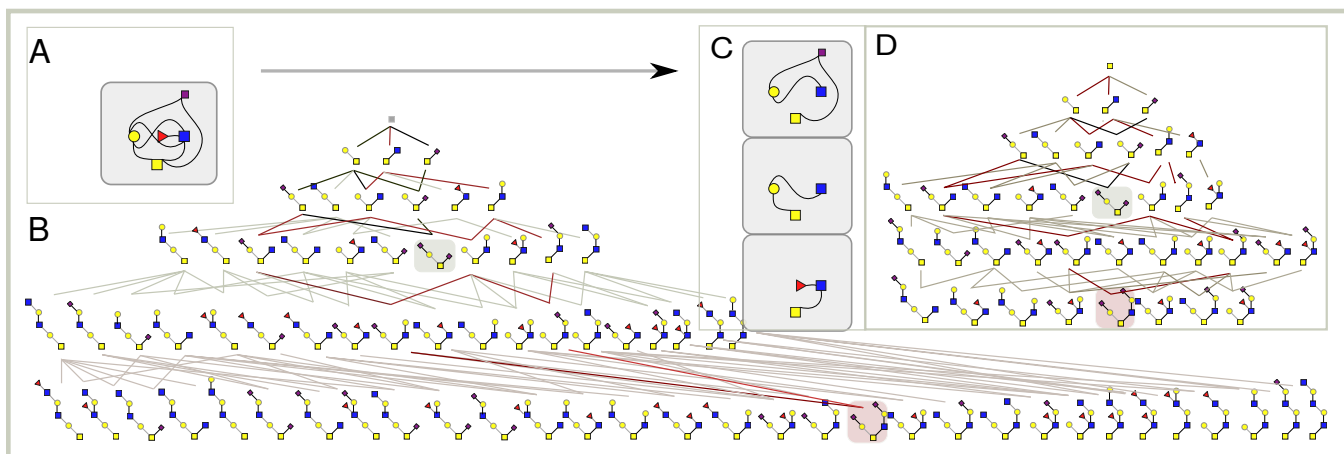
Figure 2: A) Sources of diversity: Incompleteness B) Sources of diversity: Competition. C) Sources of diversity: Polymerization.

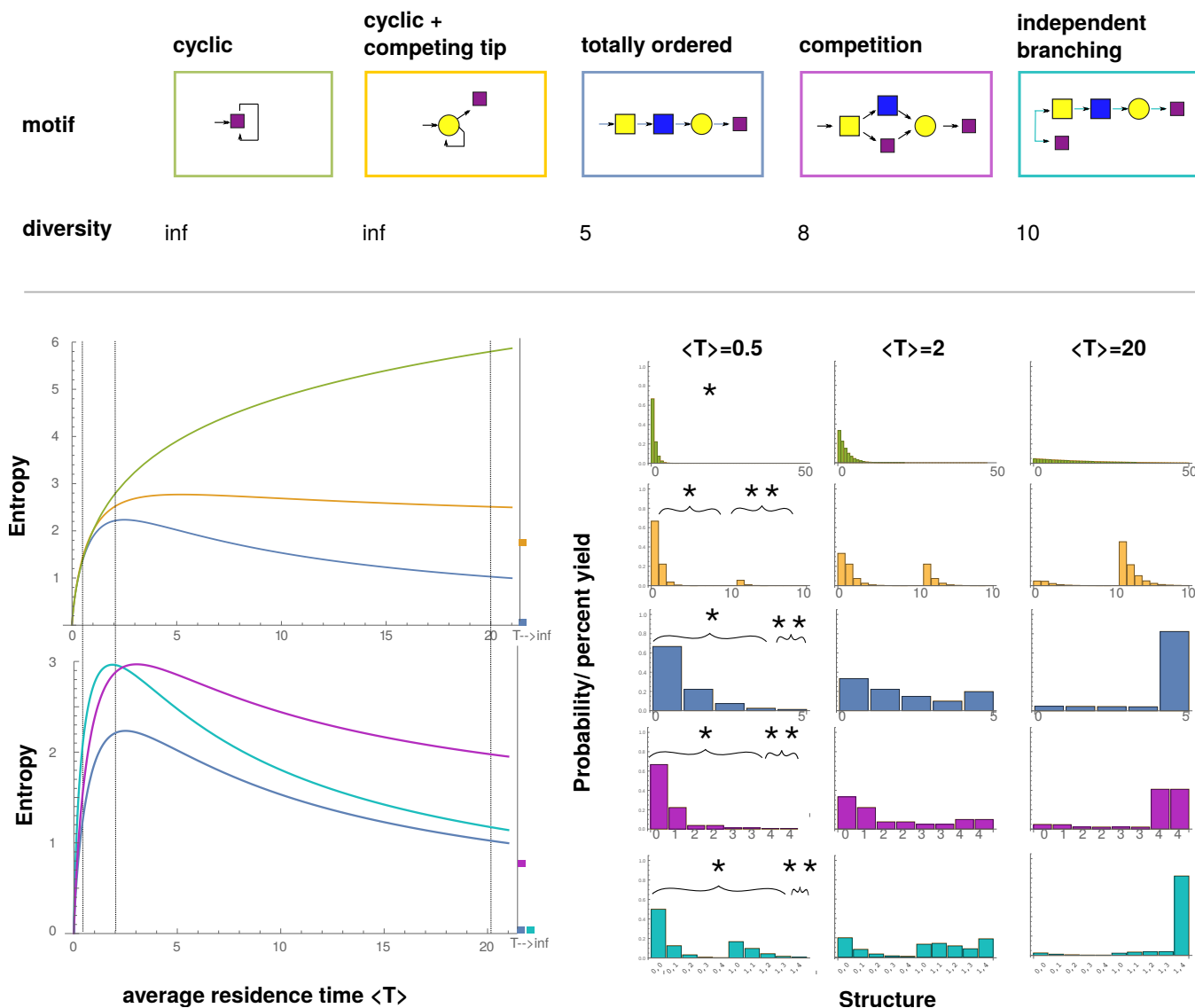|	cyclic	cyclic + competing tip	totally ordered	competition	independent branching
motif					
diversity	inf	inf	5	8	10

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).
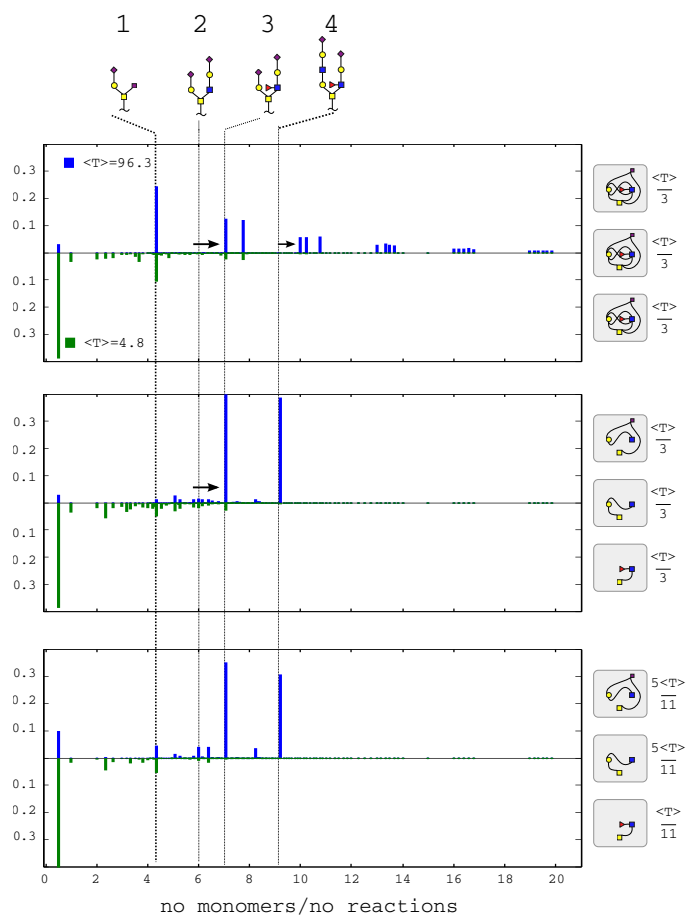
2

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, gillespe 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.
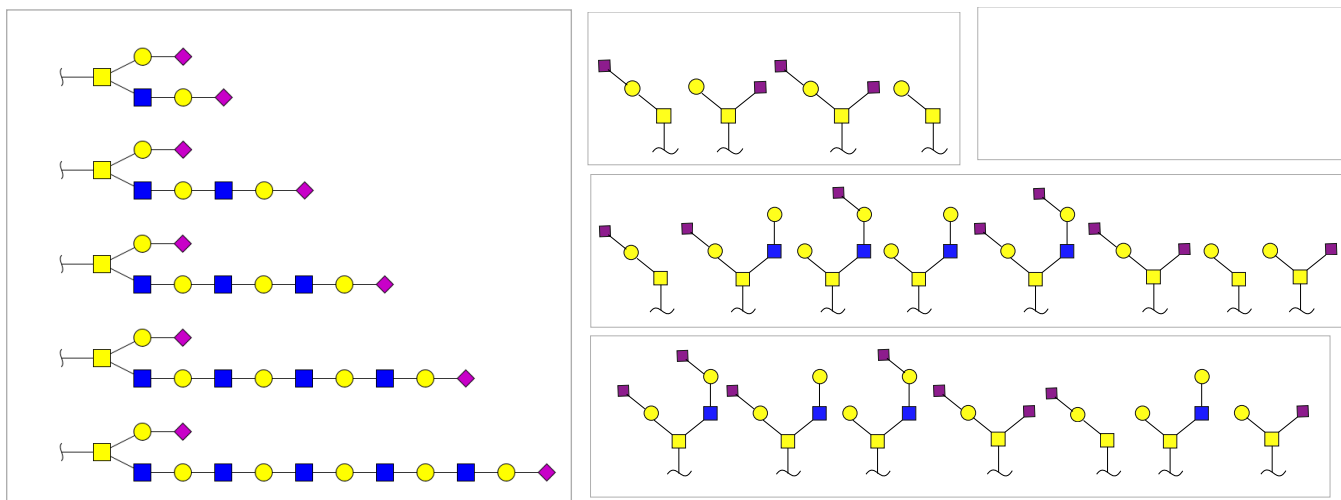
Figure 5: A) O-glycans found on equine chorionic gonadotropin [**?**]. O-glycans found on a recombinant MUC1 expressed in B) T47D, C) MDA-MB231, D)ZR-75-1 human breast adenocarcinoma cell lines [**?**]. All structure sets contain one or more of the glycan structures found in hCG o-glycan set described 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 [**?**] were not included.

cell

o-glycan

golgi

ER

Legend

Symbol	Name
◆	N-Glycolylneuraminic acid (Neu5Gc)
◆	N-acetylneuraminic acid
▲	Fucose (Fuc)
☐	N-Acetylgalactosamine (GalNAc)
○	Galactose (Gal)
■	N-Acetylglucosamine (GlcNac)
●	Glucose (Glc)
◇	Glucuronic acid (GlcA)
●	Mannose (Man)
★	Xylose (Xyl)
◇	Iduronic acid (IdoA)

human GalNac o-glycans (5)

human base entities (9)

mammalian basetypes (14)

mammalian monosacharides (35)

prokaryotic monosaccharides ( )

acceptor product Unipro

acceptor	product	Unipro
○	■–○	E7
☐	☐–■(purple)	E6
☐	☐–■	E5
○	■–○	E4
■	○–■	P15291
■	▲–■	P22083
☐	○–☐	E1

D

⊱R—■
       R'

⊱R—■³—¹○
       R'

E

☐R — r — ■
              R'

⊱R — r — ■³—¹○
                  R'

F

⊱R—■
       r' — R'

⊱R—■³—○
       r'
            R'

B4GalT5

A4GNT

C1GalT-1

ST3Gal-1

ST6GalNac

ST3Gal-1

GCNT1

C2GnT