Pareto-optimal synthesis of multiple glycans in Golgi compartments

Aashish Satyajith[1,2,*] and Mukund Thattai[2]

[1]Chennai Mathematical Institute, Chennai, India
[2]Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, TIFR, Bangalore, India
[*]Corresponding author. Email: aashish@cmi.ac.in

***Abstract*—Glycans are tree-shaped sugar chains covalently bonded to proteins in the plasma membrane of eukaryotic cells. They are assembled step by step through monomer addition reactions in different compartments of the Golgi apparatus. Glycans engage in multiple functions at the cell-surface, making their correct manufacture to be of vital importance.**

The enzyme distribution across Golgi compartments and the residence time of a glycan in each compartment have a marked effect on the amount of correct glycans manufactured, or the glycan yield. We model the synthesis of glycans in the eukaryotic Golgi apparatus, with a focus on optimizing this glycan yield. In particular, we explore strategies for the simultaneous synthesis of multiple glycans. Each strategy corresponds to a particular choice of how enzymes are split across Golgi compartments with defined residence times. We explicitly include the effect of undesirable reactions, which inevitably occur due to enzyme promiscuity. We find there is often a tradeoff in the achievable yield of distinct glycans. We explore this tradeoff using a Pareto optimality framework, and demonstrate that Pareto optimality can be a valuable aid in studying glycan manufacture.

I. Introduction

Proteins and lipids across all domains of life, particularly those on the cell surface, have branched sugar polymers known as glycans covalently attached to them [1–3]. These structures play important roles in cellular and organismal biology. For example, glycans influence the biochemical and biophysical properties of cell surfaces, mediate cell-cell interactions in tissues [3], and underlie host-pathogen specificity [4]. Incorrect glycan formation has also been linked to many non-pathogenic diseases in humans [5–7]. Having a clear understanding of glycan synthesis is therefore of prime importance.

In eukaryotic cells, proteins that are produced in the endoplasmic reticulum (ER) are subjected to further chemical modifications called glycosylation inside the ER and Golgi apparatus [8]. This results in the covalent addition of sugar monomers atop the protein substrates, finally culminating in the tree-like structure that is the glycan. The additions are performed in presence of enzymes called glycosyltransferases [8] and they are distributed across the ER and Golgi. The Golgi itself is organized as an ordered series of compartments with each compartment containing distinct sets of enzymes [9]. In this work, we explore why enzymes are distributed across compartments the way they are.

Numerous attempts have been made at answering this question from different standpoints; for example some works consider the Golgi as trying to distribute the enzymes in such a way that the Kullbach-Leibler divergence (KL-divergence) of a target glycan distribution is minimized [10]; previous work has shown that compartmentalized Golgi leads to controlled biomolecular assembly [11].

The Golgi manufactures a vast repertoire of glycans at a given time [12]. Glycan substrates undergoing glycosylation in the same cisterna are exposed to its environment for the same amount of time, before the cistera matures [13]. The amount of time a glycan resides in a compartment is called its residence time. A short residence time might not suffice for the manufacture of a complex glycan to run to completion. A long residence time might be detrimental for fully grown glycans, since promiscuous enzymes might execute undesirable reactions [14]. Both of these events result in the loss of glycoprotein substrate, and incur significant biosynthetic costs to the cell [15].

We therefore assume that the enzyme distribution across the Golgi compartments and their residence times must be designed to minimize loss of glycoprotein substrate, or equivalently, maximize glycan yield. The constraint to maximize glycan yield limits the space of allowed residence time-enzyme distribution configurations: for a given configuration, no other configuration should be able to generate more of one glycan without decreasing the yields of other glycans. The space of such configurations is called the Pareto set of configurations. Studying the Pareto set has practical applications; in the pharmaceutical industry there is great value in controlling glycosylation to achieve better antibody efficacy, selective targeting of effector cells, etc. [16, 17]. The notion has been employed with success in engineering [18], social sciences [19], and several areas of biology such as ecological studies [20–22], neuroscience [23], etc.

We devise a simple chemical kinetic model for the synthesis of a glycan. This allows us to infer the optimal choice of parameter values (enzyme distribution, residence times) for that glycan. The requirement for multiple glycans to be manufactured – each of which can have differing transit time requirements for their synthesis – brings in trade-offs. Pareto-optimal solutions [24] are a class of solutions that reconcile this trade-off. In this paper, we study the Pareto set of enzyme distribution and residence times in the context of multiple glycan manufacture.

II. Mathematical Modeling of Glycan Synthesis

A. Model assumptions

Our model is based on the following assumptions: (1) The residence time in a compartment is the same for all glycoproteins that transit that compartment; in other words, we we adopt the cisternal maturation model. (2) The residence times of different compartments can be different. (3) Glycans can only be grown by requisite (desirable) monomer addition reactions on the structure that contains the root. (4) There is a constant supply of donor sugar nucleotides. (5) Glycosylation enzymes are not saturated; therefore the monomer addition reactions obey first-order kinetics [25]. (6) All enzymes catalyze reactions at roughly the same rate, so we can set the rate of all reactions to 1 unit by selecting the appropriate time unit. (7) Enzymes are promiscuous; they can catalyze reactions other than their primary reaction, albeit at a much smaller rate [14]. We denote q to be this smaller rate. If there are $|Z_j|$ number of enzymes in compartment j that manufacture a glycan, there are $|Z_j|$ ways for the undesirable reaction to manifest on that glycan. (8) Each glycan has its own set of enzymes, and enzymes of one glycan do not catalyze the reactions needed for another glycan, be it desirable or undesirable. Both desirable and undesirable reactions occur independently. (9) Reactions are irreversible. (10) Independent components of a glycan tree can grow at the same time in a compartment (i.e. reactions can proceed in parallel). These components could be different connected components altogether or be different branches on the root of the same connected component. Fig. 1B shows an example of two branches of the glycan in Fig. 1A (right) growing independently. (11) In case of multiple glycan manufacture, if more of a glycan can be manufactured without decreasing the amounts of other glycans, that is preferable for the cell. (12) Aberrant proteins are mostly dealt with by degradation, leading to loss of substrate [26]. (13) The bulk of reactions leading to loss of substrate are reactions that do not run to completion [27]. We ignore other types of reactions in this work.

B. One glycan synthesis

We formalize two aspects that control how glycans get made: the distribution of enzymes across compartments, and residence times of each compartment.

Enzyme splits. Let $\mathcal{E}_B$ be the set of enzymes that are required to synthesize a glycan tree B. For an N-compartment Golgi apparatus, a split denotes the enzymes present in each of the compartments. Formally, a split $\mathcal{S}$ is an element of $S_N(\mathcal{E}_B) = 2^{\mathcal{E}_B} \times \cdots \times 2^{\mathcal{E}_B}$, product taken N times, where $2^{\mathcal{E}_B}$ denotes the power set of $\mathcal{E}_B$. Thus the enzyme split of a glycan can be represented as $(Z_1, \ldots, Z_n)$ where Z_j denotes the set of enzymes in compartment j. In our work, we do not consider redundant presence of enzymes across multiple compartments; only partitions of $\mathcal{E}_B$ are considered for candidate splits. An example of enzyme splits is given in Fig. 2A.

Residence times. For an N-compartment Golgi apparatus, residence times are specified by a tuple in $\mathbb{R}_+^N$. The units of time is such that the desirable monomer additions proceed at a rate of 1 unit.

A *strategy* denotes the splits and the residence times assigned to each compartment. They are represented as a two-tuple $(\mathcal{S}, \mathcal{T})$ where $\mathcal{S}$ and $\mathcal{T}$ are elements of $S_N(\mathcal{E}_B)$ and $\mathbb{R}_+^N$ respectively. Once the splits and residence times have been fixed, we can calculate the chance that a given glycan is correctly formed. We call this the *yield* of that glycan.

Calculating glycan yield: Given our assumption that monomer addition reactions follow first-order kinetics, the waiting time for reactions follows an exponential distribution. For the complete development of a branch, a monomer addition reaction has to first take place, initiating the creation of that branch, followed by all the other reactions needed to complete that branch. We use this observation to calculate yields in an inductive fashion. Let $a_1, \ldots, a_K$ be the nodes attached to the root node, in an independent tree B that is being grown in a compartment. Denote the subtree rooted at $a_1, \ldots, a_K$ by $B_{a_1}, \ldots, B_{a_K}$ respectively. Suppose the i-th branch is to be made in time t. Let t_1 be the time at which the a_i monomer addition to the root takes place, starting the growth of that branch. This time can be anywhere from 0 to t, i.e. $0 \leq t_1 \leq t$. The rest of the branch has time $t - t_1$ to develop, and thus the *yield of formation* of that branch can be written recursively as

$$\int_0^t e^{-t_1} \Big(\int_0^{t-t_1} \textit{yield-formation}(B_i; t - t_1) \, \mathrm{d}t_2 \Big) \, \mathrm{d}t_1$$

Recall that independent components of a glycan can grow in parallel. The yield of formation of the entire glycan B is then given by the products of yields of formation of independent components:

$$\textit{yield-formation}(B; t) =$$

$$\prod_{i=1}^{K} \int_0^t e^{-t_1} \Big(\int_0^{t-t_1} \textit{yield-formation}(B_i; t - t_1) \, \mathrm{d}t_2 \Big) \, \mathrm{d}t_1.$$

As an example, consider the reactions needed to form a glycan with two branches on the substrate with the left subtree needing one reaction, and the right branch needing two serial reactions (i.e. happening one after another; see Figure 1(A) for an example of a glycan needing four serial reactions). The yield of formation of required reactions can then be written down as

$$(1 - e^{-t}) \times \int_0^t e^{-t_1} \Big(\int_0^{t-t_1} e^{-t_2} \, \mathrm{d}t_2 \Big) \, \mathrm{d}t_1$$

The term $(1 - e^{-t}) = \int_0^t e^{-t} \, \mathrm{d}t$ on the left corresponds to the yield of the left branch with one reaction. The term on the right corresponds to the yield of two serial reactions. We also note that in case of trees that have only serial reactions or parallel reactions, simpler expressions for yield can exist.

Recall that undesirable reactions occur at rate q. The chance of an undesirable reaction occurring within residence time t is $1 - e^{-qt}$, therefore the chance of it *not* occurring is e^{-qt}. As mentioned earlier, if there are $|Z_j|$ enzymes in compartment j, then there are as many ways for this promiscuous effect to manifest during the compartment residence time t_j. Thus, the probability that an undesirable reaction does not occur is $e^{-q \cdot t_j \cdot |Z_j|}$.

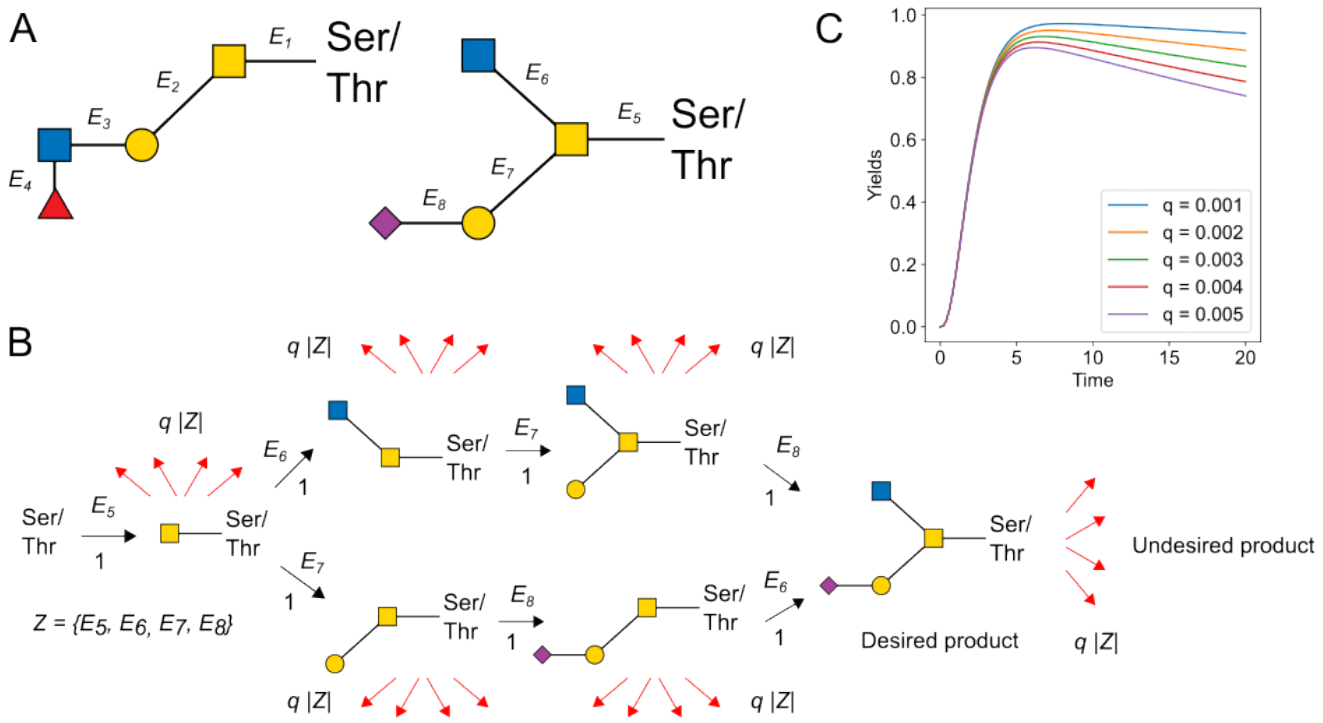

Fig. 1. Enzymes catalyze desirable and undesirable reactions, at different rates. **(A)** Examples of glycan structures manufactured atop a Ser/Thr glycosylation site. The colored shapes denote sugar monomers. The labels on the edges denote enzymes that are required to catalyze the linkage reaction. The glycan to the left needs four serial reactions (i.e. reactions occurring one after another) to form correctly. In case of the glycan to the right, the reaction catalyzed by E_6 can proceed independently of the other branch, once the E_5 reaction has taken place. **(B)** Different growth orders that can give rise to the glycan to the right of panel A. The numbers beside the arrows indicate the rate of reaction. The red arrows denote an undesirable reaction taking place. **(C)** Yield of glycan to the right of (A) at different values of undesirable reaction rate q and residence times.

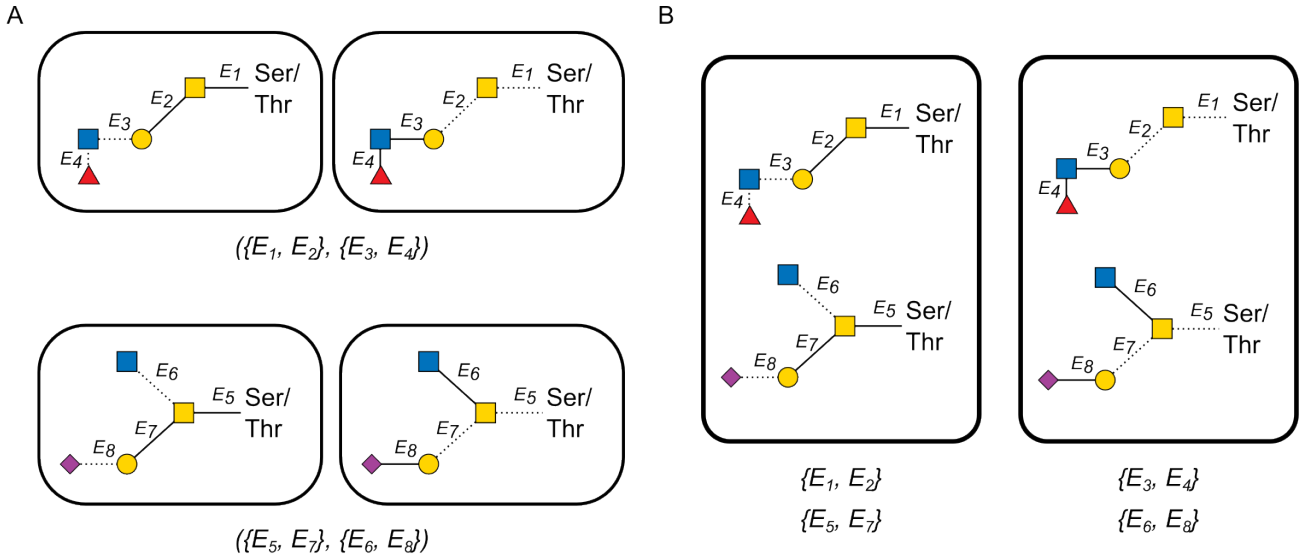

Fig. 2. Trade-offs arise during multiple glycan manufacture. **(A)** Examples of splits and their (text) notation. The top two boxes in panel A denote a Golgi configuration with enzymes E_1 and E_2 in the first compartment, and E_3 and E_4 in the second compartment. The thick lines in a structure inside a compartment denote reactions that are to happen in that compartment, while dotted lines denote reactions that happen elsewhere. Another example of a split is given at the bottom. Splits can be represented pictorially or via text, as shown. **(B)** Golgi configurations needed to produce the different glycans shown in 1A. In the second compartment, two serial and two parallel reactions need to happen for the correct manufacture of the respective glycans. The reactions will have different optimal residence times, creating trade-offs.

The correct formation of a glycan depends on desirable reactions occuring and undesirable reactions not occuring. Since they are independent events, the yield of that glycan is given by the product of these two events. In other words, if $(Z_1, \ldots, Z_N)$ denotes the enzyme split of a glycan B with residence times $(t_1, \ldots, t_N)$, the yield of that glycan is $\prod_{j=1}^{N} \textit{yield-formation}(Z_j; t_j) \cdot e^{-q \cdot t_j \cdot |Z_j|}$. Thus, the objective function for the optimal yield of a glycan B can be formulated as:

$$\underset{\substack{(Z_1,\ldots,Z_N)\in S_N(B)\\(t_1,\ldots,t_N)\in\mathbb{R}_+^N}}{\text{maximize}} \quad \prod_{j=1}^{N} \textit{yield-formation}(Z_j; t_j) \cdot e^{-q \cdot t_j \cdot |Z_j|} \tag{1}$$

Finally, define the yield of a strategy $(\mathcal{S}, \mathcal{T}) = ((Z_1, \ldots, Z_N), (t_1, \ldots, t_N))$ as

$$\textit{yield}(\mathcal{S}, \mathcal{T}) = \prod_{j=1}^{N} \textit{yield-formation}(Z_j; t_j) \cdot e^{-q \cdot t_j \cdot |Z_j|}$$

A glycoprotein population can include many glycoforms that have varying functional activity [28], all of which are produced by the same Golgi apparatus. We discuss this next.

C. Multiple glycan synthesis

Consider that the Golgi produces K glycan trees $B_1, \ldots, B_K$. Let $\mathcal{S}^1, \ldots, \mathcal{S}^K$ denote the splits of the K glycans. As mentioned earlier, we assume that the enzymes needed to catalyze the manufacture of one glycan do not catalyze the reactions needed for another glycan, be it desirable or undesirable. We will write the collection of splits $\{\mathcal{S}^1, \ldots, \mathcal{S}^K\}$ as $\boldsymbol{\mathcal{S}}$, in bold type, and write $\boldsymbol{\mathcal{S}}^i$ to denote the split corresponding to the ith glycan. Once we fix the splits for each of the K glycans and the residence times of each compartment, denoted $(\boldsymbol{\mathcal{S}}, \mathcal{T})$, we can calculate the yields $y_1, \ldots, y_K$ for each of the glycans.

This process of synthesising multiple glycans can create tradeoffs, arising from the fact that the residence times are fixed for each compartment. A short residence time might not give enough time for the reactions of a complex glycan to run to completion. A long residence time might lead to undesirable reactions on a fully-grown glycan. Figure 2(B) gives an example where such a scenario arises.

Another aspect of multiple glycan manufacture is that the cell will have its own fitness function $f(y_1, \ldots, y_K)$ based on the yields $y_1, \ldots, y_K$ of the K glycans, which we assume is what the cell is optimized for. As mentioned previously, if the yield of a glycan can be improved without decreasing the yield of other glycans, that is advantageous for the cell. In other words, we assume $f(y_1, \ldots, y_K)$ is a monotonically increasing function of y_i $\forall i$, keeping $y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_K$ fixed. With this assumption, we can use the notion of Pareto optimality. We now define Pareto dominance of strategies.

Pareto optimality: A strategy is said to *weakly Pareto dominate* another strategy if all the glycans that are being synthesised have either same or better yields using the former strategy than the latter strategy [24]. Formally, let $(u_1, \ldots, u_K)$ and $(v_1, \ldots, v_K)$ be the yield vectors corresponding to strategies $(\boldsymbol{\mathcal{S}}, \mathcal{T})$ and $(\boldsymbol{\mathcal{S}'}, \mathcal{T}')$ respectively. $(\boldsymbol{\mathcal{S}'}, \mathcal{T}')$ weakly Pareto dominates $(\boldsymbol{\mathcal{S}}, \mathcal{T})$, written $(\boldsymbol{\mathcal{S}}, \mathcal{T}) \preceq (\boldsymbol{\mathcal{S}'}, \mathcal{T}')$, if $u_i \leq v_i, \forall i \in [K]$.

The set of all strategies that cannot be Pareto dominated is called the Pareto set, and the set of their corresponding yields is called the Pareto front. We will use Pareto domination in the weak sense throughout this chapter. Given our assumption earlier that $f(.)$ is monotonically increasing, we can restrict ourselves to only considering the set of Pareto-optimal solutions for further study. Pareto-optimal solutions for glycan manufacture can be obtained by maximizing convex weightings of the yield functions; see Appendix B.

III. Results

We will restrict ourselves to the study of two-compartment Golgi systems, since that suffices to capture the aspects we want to focus on.

A. One glycan synthesis

There is an optimal residence time per compartment: Suppose the glycan to the right of Fig. 1A is to be grown in a single compartment. Fig. 1C shows its yield for different residence times and at different values of q. As is clear, there is an optimal residence time for the synthesis of a given glycan, and other residence times lead to suboptimal yields. A global maximum is guaranteed because log-concave functions have a global maximum, and the yield function is a log-concave function of residence times (see Appendix A for proof).

Some enzyme splits are better than others: We demonstrate this for a glycan that needs four serial reactions (i.e. occurring one after another) for its correct synthesis.

Fig. 3A plots the yield contours under a 1-3 split (first reaction in the first compartment, three remaining reactions in the next compartment) and a symmetric (2-2) split. The symmetric split achieves a higher yield. The yield function is anisotropic around the peak; deviations from optimum have direction-dependent effects. The contours also indicate that the optimal residence times for the first and second compartments depends on the split.

Fig. 3B shows the optimal yield of the same glycan under all possible splits, for various values of q. The symmetric split achieves the highest yield among all the splits.

B. Multiple glycan synthesis

Pareto-optimal solutions for two glycan manufacture: We consider the Pareto set of trees 1 and 2 shown in Fig. 4A. Candidate solutions can be obtained by fixing the split of both trees and finding the optimal residence times for a range of convex weightings of their yields (Appendix B, [29]). Some of the candidate solutions generated will be Pareto optimal. Fig. 4B sketches the yields of both the trees under a range of convex weightings of their yields. The blue and red colors denote the different splits of the second tree that are used when calculating the yields (the first tree has one optimal split). One can see that some segments of the blue curve Pareto-dominate

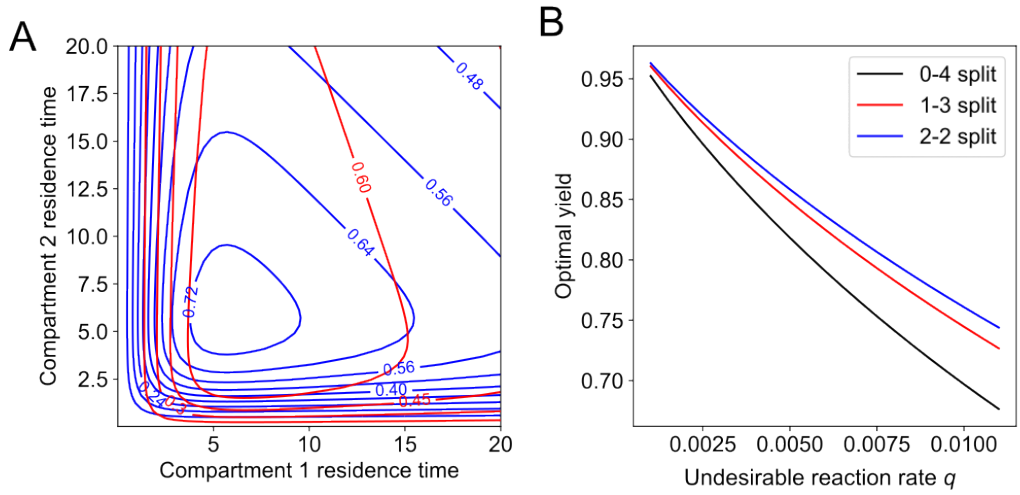

Fig. 3. Optimal splits and residence times for single glycan manufacture. **(A)** Values of yields w.r.t. residence times in two compartments for different splits of four serial reactions, at $q = 0.01$. The blue lines denote contours for a 2-2 split (first two reactions in the first compartment, next two in the second compartment) while the red lines denote contours for a 1-3 split. **(B)** Optimal yield vs q curve for all possible splits of four reactions across two compartments.

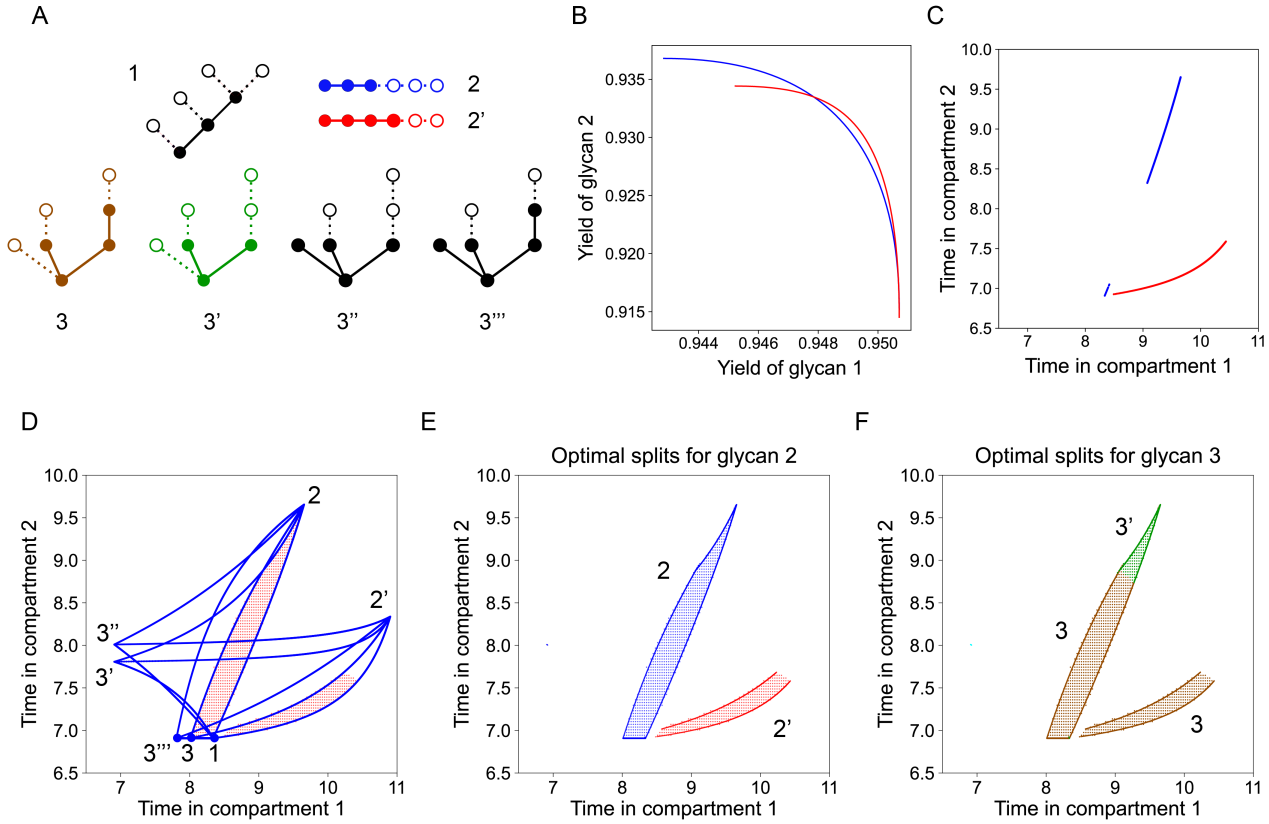

Fig. 4. Pareto-optimal splits and residence times for multiple glycan manufacture. **(A)** The tree structures that have been used to study Pareto-optimal solutions, and their splits. Numbers adjacent to the structures and their prime versions (e.g. 2 and 2') represent different splits of the same tree. The nodes with filled-in colours (along with unbroken lines) represent monomers that will be linked in the first compartment, and the empty nodes with borders (and dotted lines) represent monomers that will be added in the second compartment. The colours are used to indicate splits in subsequent plots. **(B)** Yields corresponding to two different solution sets generated by a range of convex weightings of yields of trees 1 and 2. For this and all subsequent plots, we set $q = 0.001$. The red (blue) curve is the set of candidate solutions assuming only splits 1 and 2 (splits 1 and 2'), are allowed (see Appendix C for details). **(C)** The Pareto set corresponding to (B). Colours indicate the splits that generated those solutions. **(D)** The labeled points denote optimal residence times for each split. Each blue curve going from one labeled point to another represents Pareto-optimal solutions assuming only those two splits are allowed (and disregarding the third glycan). The red dotted region denotes the Pareto-optimal solutions for the three glycans under study. **(E)** and **(F)** Optimal splits of the second and third glycan respectively corresponding to the Pareto set in (D).

segments of the red curve, and vice versa [22]. The crossover point represents the point at which the (Pareto-) optimal split of second tree changes. To obtain a Pareto-optimal solution with a higher yield of the second glycan than at crossover point, the blue split must be used. If the yield of the first glycan is to be slightly higher than at crossover point, the red split must be used. The Pareto set is sketched in Fig. 4C.

Pareto-optimal solutions for three or more glycan manufacture: Since generating Pareto-optimal solutions using the convex weighting method is computationally demanding, we sampled the set of Pareto-optimal solutions by brute force. For a fixed residence time (t_1, t_2), we evaluated the maximum yield at (t_1, t_2) across different splits. This was repeated for a grid of (t_1, t_2) values. Finally, we chose the yield tuples which were not Pareto-dominated.

Fig. 4D shows the Pareto set for the three glycans in Fig. 4A; it is now a region in the transit time space. The region is enclosed by boundaries corresponding to candidate solutions generated by considering fixed splits of only two glycan trees at a time (Fig. 4D, blue lines). We posit that the Pareto-optimal solutions for three glycan manufacture will be contained in the boundaries so obtained, but this requires a formal proof. Figs. 4E, F show the optimal splits of second and third glycans and transit times that form the Pareto set (there is only one optimal split for glycan 1).

In the three-dimensional yield space, there is a yield manifold corresponding to every combination of splits of the three glycans. The manifold is obtained by convex weightings of the three glycan yields under their respective splits. Two manifolds may intersect along a curve. The intersection curve represents points at which the optimal split changes to achieve differing Pareto-optimal yields; it plays a role similar to the crossover point in two-glycan manufacture (Fig. 4B).

There are two possibilities that arise when a fourth glycan is added to the mix: either there are no new Pareto-optimal solutions, or new Pareto solutions emerge continuous to existing regions in the Pareto set or by formation of new regions.

IV. Discussion

The tree shape and non-template driven growth of glycans raises interesting questions regarding the efficacy of its manufacturing process. The abstract growth template of glycan manufacture (topological ordering of growth, exposure to different environments across time, ceasing to exist with a small probability) is widespread and can be observed in areas such as ecology (growth of a microorganism [30], speciation [31]), finance [32], disease progression [33], etc. We have shown that under certain conditions, the survival function of an entity following the template is a concave function of the parameters influencing its growth. The concavity can be exploited to study the Pareto set of solutions.

In the particular context of glycan manufacture, the Pareto set has interesting properties. The Pareto set comprises disconnected regions of the continuous parameter space (residence times; Fig. 4E). More interestingly, small changes in the continuous parameter (residence times) can lead to different optimal discrete parameters (splits) (Fig. 4F). This indicates that yields of different splits are comparable. A tradeoff arises here, between the cost of maintaining records of optimal splits at various residence times, and the survival advantage a small increase in yield confers to a cell. Given that glycan manufacture is replicated across trillions of cells, even a small increase in yield conferred by changing the enzyme distribution might be advantageous.

We have computed the optimal configuration for multiple glycan manufacture assuming that the objective function the cell optimizes is monotonically increasing in the yields of the glycans. If this is not true, there might be presence of local maxima.

The crossover of yields in Fig. 4B raises another interesting possibility. If a mix of different Golgi configurations are allowed in a population, the yields of glycans weighted by their fraction of occurrence in the population might Pareto dominate yields achieved by a single split. We have not explored this possibility of bet hedging in the context of glycan manufacture [22].

Acknowledgements

We would like to thank Shaon Chakraborty, Sachit Daniel and Nandita Chaturvedi for useful discussions.

MT acknowledges support from the Simons Foundation (287975).

Author Contributions

MT conceived the project. AS carried out the analysis. AS wrote the paper.

Appendix

A. Finding optimal splits and residence times for single glycan manufacture

The yield objective is a log-concave function of residence times: A strictly positive function $f : \mathbb{R} \to \mathbb{R}_+$ is log-concave if the logarithm of the function is concave, i.e. $\log\ f(\theta x + (1-\theta)y) \geq \theta\ \log\ f(x) + (1-\theta)\ \log\ f(y)$ for $0 \leq \theta \leq 1$; $x, y \in \mathbb{R}$ [35]. A differentiable function f is concave in an interval if its first derivative is monotonically decreasing in that interval. From the definition of log-concavity, it follows that if f can be decomposed into a product of log-concave functions, then f is log-concave.

A function $f(x_1, \ldots, x_N)$ is said to be pointwise log-concave if for all fixed $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N$, the function is log-concave in $x_i, \forall i$.

We prove that the yield function *yield*$(\mathcal{S}, \mathcal{T})$ for a glycan, with fixed $\mathcal{S}$, is point-wise log-concave. Since the yield function comprises products of yields corresponding to splits in each compartment, it suffices to prove that the expression for yield in a single compartment is a log-concave function of the residence time for that compartment. We do this by induction.

Base case: The yield corresponding to a single reaction is $(1 - e^{-t}) \cdot e^{-qt}$ for some q (t is the residence time). Both the terms can be shown to be log-concave, and therefore their product is also log-concave.

Inductive case: Section II-B discussed how to decompose a glycan tree to calculate its yield. A split can have several independent components, and the final yield is calculated as the product of yields of individual components (multiplied by the undesirable reaction term, which is log-concave). We show that yield of each component is a log-concave function of the residence time.

Let us assume that the yield of a sub-tree B_i attached to a root node is log-concave w.r.t. the residence time t, i.e. *yield-formation*$(B_i; t)$ is log-concave. The overall yield of the branch containing B_i is calculated as

$$\int_0^t e^{-t_1} \Big(\int_0^{t-t_1} \textit{yield-formation}(B_i; t_2) \ \mathrm{d}t_2 \Big) \ \mathrm{d}t_1$$

Write $\int_0^\tau$ *yield-formation*$(B_i; t_2)$ $\mathrm{d}t_2$ as $Y(\tau)$. The above expression represents the convolution of two log-concave functions $F(\tau) = e^{-\tau}$ and $Y(\tau)$, and is therefore log-concave [35].

The yield function is a log-concave function of residence times assuming splits are fixed. To pick the optimal yields we brute-force over all possible splits and pick the residence times giving the highest yield.

B. Generating Pareto-optimal solutions for multiple glycan manufacture

Several methods exist in literature to generate Pareto-optimal solutions in case of convex objectives (maximizing a function is equivalent to minimizing its negative; if the function is concave, its negative is convex). If there are multiple convex objectives $f_1(\boldsymbol{x}), \ldots, f_K(\boldsymbol{x}); f_i(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ for some n, then it is known that solutions of

$$\min_{\boldsymbol{x}} \sum_{i=1}^{K} c_i \ f_i(\boldsymbol{x}); \sum_{i=1}^{K} c_i = 1 \text{ and } c_i \geq 0$$

i.e. minimizers of convex combinations of the given convex functions, are Pareto optimal [36, 37].

We enumerate over all possible splits for all glycans (fixing which makes the yield functions log-concave functions of residence times, as discussed in Appendix A) and use brute force to pick the Pareto-optimal solutions. In fact, this forms the motivation behind labeling a family of strategies (set of solutions) using splits.

C. Generating a set of solutions from splits

Suppose there are two glycans being synthesised across N compartments, with yield functions *yield*$_1(\cdot)$ and *yield*$_2(\cdot)$ and splits $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. Then a family of solutions can be generated by solving the program

$$\max_{\mathcal{T} \in \mathbb{R}^N} c \cdot \textit{yield}_1(\mathcal{S}_1, \mathcal{T}) + (1-c) \cdot \textit{yield}_2(\mathcal{S}_2, \mathcal{T})$$

for $0 \leq c \leq 1$. The residence times generated for different values of c can be collected together and forms the set of solutions indexed by splits $(\mathcal{S}_1, \mathcal{S}_2)$.

D. Code availability

A Jupyter Notebook generating the plots can be found at https://gitlab.com/aashishsatya/pareto-optimal-glycan-synthesis.

References

1. Adibekian, A. *et al.* Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chemical Science* **2,** 337–344 (2011).
2. Wetzel, L. A. *et al.* Predicted glycosyltransferases promote development and prevent spurious cell clumping in the choanoflagellate *S. rosetta*. *eLife* **7** (2018).
3. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nature Reviews Molecular Cell Biology* **13,** 448–462 (2012).
4. Zhao, X., Chen, H. & Wang, H. Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Frontiers in Molecular Biosciences* **8.** ISSN: 2296-889X (2021).
5. Lloyd, K. O., Burchell, J., Kudryashov, V., Yin, B. W. T. & Taylor-Papadimitriou, J. Comparison of O-Linked Carbohydrate Chains in MUC-1 Mucin from Normal Breast Epithelial Cell Lines and Breast Carcinoma Cell Lines. *Journal of Biological Chemistry* **271,** 33325–33334 (1996).
6. Lo-Guidice, J.-M. *et al.* Sialylation and sulfation of the carbohydrate chains in respiratory mucins from a patient with cystic fibrosis. *Journal of Biological Chemistry* **269,** 18794–18813 (1994).
7. Brockhausen, I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO reports* **7,** 599–604 (2006).
8. Alberts, B. *Molecular biology of the cell* (Garland Science, 2015).
9. *Essentials of Glycobiology, 3rd Edition* (eds Varki, A. *et al.*) ISBN: 978-1-621821-32-8 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2017).
10. Yadav, A., Vagne, Q., Sens, P., Iyengar, G. & Rao, M. Glycan processing in the Golgi: optimal information coding and constraints on cisternal number and enzyme specificity. *eLife* **11** (2022).
11. Jaiman, A. & Thattai, M. Golgi compartments enable controlled biomolecular assembly using promiscuous enzymes. *eLife* **9** (2020).
12. Fisher, P., Spencer, H., Thomas-Oates, J., Wood, A. J. & Ungar, D. Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation. *Cell Reports* **27,** 1231–1243.e6 (2019).
13. Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116,** 153–66. ISSN: 0092-8674 (2 2004).
14. Biswas, A. & Thattai, M. Promiscuity and specificity of eukaryotic glycosyltransferases. *Biochemical Society Transactions* (2020).

15. Gutierrez, J. M. *et al.* Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nature Communications* **11** (2020).
16. Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery* **8,** 226–234 (2009).
17. Zhong, X., D'Antona, A. M., Scarcelli, J. J. & Rouse, J. C. New Opportunities in Glycan Engineering for Therapeutic Proteins. *Antibodies* **11,** 5. ISSN: 2073-4468 (1 2022).
18. Deb, K. *Evolutionary Algorithms for Multi-Criterion Optimization in Engineering Design* in (1999).
19. Aziz, H., Brandt, F. & Harrenstein, P. Pareto optimality in coalition formation. *Games and Economic Behavior* **82,** 562–581 (2013).
20. Shoval, O. *et al.* Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336,** 1157–1160 (2012).
21. Tendler, A., Mayo, A. & Alon, U. Evolutionary trade-offs, Pareto optimality and the morphology of ammonite shells. *BMC Systems Biology* **9,** 12. ISSN: 1752-0509 (2015).
22. Levins, R. Theory of fitness in a heterogeneous environment. I. The fitness set and adaptive function. *The American Naturalist* **96,** 361–373 (1962).
23. Jedlicka, P., Bird, A. D. & Cuntz, H. Pareto optimality, economy–effectiveness trade-offs and ion channel degeneracy: Improving population modelling for single neurons. *Open Biology* **12** (2022).
24. Miettinen, K. *Nonlinear multiobjective optimization* (Kluwer Academic Publishers, 1998).
25. Hirschberg, K. *et al.* Kinetic analysis of secretory protein traffic and characterization of Golgi to plasma membrane transport intermediates in living cells. *The Journal of Cell Biology* **143,** 1485–1503 (1998).
26. Schwabl, S. & Teis, D. Protein quality control at the Golgi. *Current Opinion in Cell Biology* **75,** 102074. ISSN: 0955-0674 (2022).
27. Varki, A. Biological roles of glycans. *Glycobiology* **27,** 3–49 (2016).
28. Stanley, P. Golgi glycosylation. *Cold Spring Harbor Perspectives in Biology* **3** (2011).
29. Emmerich, M. T. M. & Deutz, A. H. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural Computing* **17,** 585–609. ISSN: 1567-7818 (3 2018).
30. Austin, D., White, N. & Anderson, R. The Dynamics of Drug Action on the Within-host Population Growth of Infectious Agents: Melding Pharmacokinetics with Pathogen Population Dynamics. *Journal of Theoretical Biology* **194,** 313–339. ISSN: 00225193 (3 1998).
31. Ovaskainen, O. & Meerson, B. Stochastic models of population extinction. *Trends in Ecology and Evolution* **25,** 643–652. ISSN: 01695347 (11 2010).
32. Fu, D. *et al.* The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences* **102,** 18801–18806. ISSN: 0027-8424 (52 2005).
33. Chowell, G., Sattenspiel, L., Bansal, S. & Viboud, C. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews* **18,** 66–97. ISSN: 15710645 (2016).
34. Mehta, A. Y. & Cummings, R. D. GlycoGlyph: A glycan visualizing, drawing and naming application. *Bioinformatics* **36,** 3613–3614 (2020).
35. Boyd, S. P. & Vandenberghe, L. *Convex optimization* (Cambridge Univ. Pr., 2011).
36. Marler, R. T. & Arora, J. S. The weighted sum method for multi-objective optimization: New insights. *Structural and Multidisciplinary Optimization* **41,** 853–862 (2009).
37. Chankong, V. & Haimes, Y. Y. *Multiobjective Decision making: Theory and methodology* (Dover, 2008).