
Pareto-Optimal Golgi Design and Decoding Cell-Surface Glycans

By

Aashish Satyajith

*A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy*

to

Chennai Mathematical Institute

01 March 2025



Plot-H1, SIPCOT IT Park,
Siruseri, Kelambakkam,
Tamilnadu - 603103
India



CHENNAI
MATHEMATICAL
INSTITUTE

Aashish SATYAJITH

Plot-H1, SIPCOT IT Park
Padur Post, Siruseri
Tamil Nadu, India-603103
E-mail: aashish@cmi.ac.in

DECLARATION

I declare that the thesis entitled **Pareto-Optimal Golgi Design and Decoding Cell-Surface Glycans** submitted by me for the degree of **Doctor of Philosophy in Computer Science** is the record of academic work carried out by me under the guidance of Prof. Mukund THATTAI and this work has not formed the basis for the award of any degree, diploma, associateship, fellowship or other titles in this University or any other University or Institution of Higher Learning.

Aashish SATYAJITH
Chennai Mathematical Institute
01 March 2025.



CHENNAI
MATHEMATICAL
INSTITUTE

Prof. Mukund THATTAI

Plot-H1, SIPCOT IT Park
Padur Post, Siruseri
Tamil Nadu, India-603103
E-mail: thattai@ncbs.res.in

CERTIFICATE

I certify that the thesis entitled **Pareto-Optimal Golgi Design and Decoding Cell-Surface Glycans** submitted for the degree of **Doctor of Philosophy in Computer Science** by Aashish SATYAJITH is the record of research work carried out by him under my guidance and supervision, and that this work has not formed the basis for the award of any degree, diploma, associateship, fellowship or other titles in this University or any other University or Institution of Higher Learning. I further certify that the new results presented in this thesis represent his independent work in a very substantial measure.

Chennai Mathematical Institute
Date: 01 March 2025.

Prof. Mukund THATTAI
Supervisor.

CHENNAI MATHEMATICAL INSTITUTE

Department of Theoretical Computer Science

Abstract

Doctor of Philosophy

**Pareto-Optimal Golgi Design and
Decoding Cell-Surface Glycans**

by Aashish SATYAJITH

A forest of tree-shaped sugar chains called glycans coat the surfaces of eukaryotic cells, embedded to proteins in their plasma membrane. They are manufactured step by step in enzyme-catalyzed (stochastic) reactions as they traverse the multiple compartments of the Golgi apparatus.

Of the billions of glycans that can theoretically be manufactured, cells must manufacture only those glycans that are required for their function. The amount of each glycan made is regulated in two ways: the distribution of enzymes across the compartments of the Golgi apparatus, and the amount of time a growing glycan spends in a compartment.

We posit that the enzyme distribution and the residence times must be such that the chances of producing useful glycans is maximized. When multiple glycans are to be manufactured, competing manufacturing requirements can create trade-offs. We study the solution space reconciling this trade-off using a Pareto-optimality framework. We show that enzyme splits change discontinuously as residence times are varied, creating an interesting space of growth strategies.

A major function of glycans is to mediate communication between living cells. How information gets encoded by one cell and decoded by other cells remains unclear. This question becomes especially relevant considering that there is significant variability across cells in their surface glycan distribution due to the stochastic nature of glycan manufacture.

We answer this question in the specific context of Golgi dysfunction. We perturb the Golgi to various extents and measure the corresponding cell-surface glycans. Then we ask if we can infer the amount of perturbation that was made by measuring cell-surface glycans alone. To do this, we first model the dynamics of cell-surface glycans and fit the model to experimental data. We use the model to construct an optimal Bayesian decoder that predicts Golgi dysfunction given single-cell glycan distribution. We show that having multiple types of glycans on the cell surface helps in faster, more accurate information dissipation.

In summary, this thesis forms a bridge between the form and function of the Golgi apparatus.

Acknowledgements

As I ponder the list of people to thank for my thesis, I am astounded by how interconnected our lives really are. A full list of acknowledgements would feature the driver of the van who brings in the NCBS canteen supplies at 3 AM in the morning, the friendly canteen staff, NCBS cleaning staff, administrative staff at both CMI and NCBS who ensure our stipends get disbursed on time, engineers of large language models (LLMs), and so on. In the interest of brevity, I will restrict my attention to people who have helped technically with my project, and at a deeply personal level. I apologize for the glaring omissions.

With Mukund as the charioteer, there should have been no doubt that the thesis will reach this far. I truly lucked into this opportunity to work with a world-class scientist. I know I still have a long way to go as a researcher, but I'm thankful for the strong initial push he gave. Every meeting with him continues to be revelation. As a teacher, Mukund's courses have been some of the best I've sat through, and I've learned a lot from them. Nidhi, Kaadambari and Raghav have been wonderful hosts whenever we visited their house.

Chennai Mathematical Institute (CMI) has managed to set up an environment where there's a lot of academic freedom and you're mostly left to you and your research. I hope CMI stays that way for a long time to come. The CMI faculty were encouraging of my renegade plan to work at NCBS. I thank Madhavan Mukund for extending my tenure on request.

I am thankful to Govind R. and Suriya Selvarajan for behind-the-scenes support. Chats with them have often offered me blinding clarity about science, and career. Shoutout to my friends from undergrad and master's for staying in touch despite the distance.

Signposts that have led my journey to where I am today came by way of teachers; in particular, K. Murali Krishnan of NIT Calicut has my gratitude for making me think about a career in research.

Mind cannot function without a healthy body, and in this regard full credit goes to *Kalari Gurukkal* Ranjan Mullarat for his *kalaripayattu* training. I would like to thank Shaileshanand Jha and Anton Swaminathan Iyer for the many conversations about science, *kalari*, and miscellany, and for splitting auto rickshaw fare. Anton helped me save a lot of time by pointing me to the Cellpose software for cell boundary identification.

I would like to thank Alkesh Yadav and Amit Kumar for useful discussions regarding my first and second project respectively. I've also profited from talking to Ayan Roychowdhury, Shubham Shinde, and other simonellas about my projects. Physics of Life 2023 batch has been an energetic bunch of people to hang out with. A chunk of biology I know is thanks to my GS2020 "batchmates", who held journal clubs every once a while and gave consistently high-quality talks. Nivedita Mukherjee while claiming to be naïve has always asked sharp questions about my projects, and I've also enjoyed all of our random conversations.

I would like to thank K. V. Subrahmanyam, Mukund, Shachi Gosavi, Shaon Chakrabarti, and Shruthi Viswanath for providing me an opportunity to explore the joys of teaching by allowing me to TA their respective courses. Bhanu Priya S. has my sincerest gratitude for taking much of my TA load when I had to be away during Mukund's STOCH2023 offering.

My thesis committee members Sandeep Krishna and K. V. Subrahmanyam have always asked insightful questions that made me think more deeply about my projects. K. V. has, as always, been very accommodating of my last-minute requests. My thesis reviewers Felix Campelo (ICFO–Institute of Photonic Sciences, Barcelona) and Mohit Kumar Jolly (IISc Bangalore) raised incisive questions about my projects, and offered constructive criticism. It resulted in a more complete and coherent thesis, and helped fix errors.

A major chunk of my thesis is thanks to Prof. Nagaraj Balasubramanian and his lab (specifically Prachi Joshi, Debiprasad Panda, Radhika Malaviya and Rajeshwari B. R.) at IISER Pune. Nagaraj was kind enough to host me for a couple of weeks at IISER. Prachi has been an ideal colleague to work with, and I thank her for patiently explaining the logic behind every step of the data generation process. Anshuman Biswas has our thanks for key inputs during the conceptualization stage of the decoding project. The collaboration has furthered my conviction that confluence of experimentalists and theorists progresses biology in directions that either team cannot achieve, on their own.

Despite the lab being as compartmentalized as the Golgi, Sahana Shridhar, Sachit Daniel, Kritika Kumari and Nandita Chaturvedi have been incredibly helpful at various points during my time at NCBS, both in terms of technical and moral support. Each of their personalities has offered me something to learn and grow from. Ex-labmate Nida gets brownie points for sending relevant articles (and songs!) she thought I'd like my way.

I would like to thank Saptarshi Dasgupta, Abhishek Sonke and Rohini Subrahmanyam for keeping me alive during COVID-19 by providing me food and shelter, and them and Ayanthi Bhattacharya and Anupama Bharadwaj for keeping me alive in spirit during the same time. Ayanthi Bhattacharya taught me everything I know about genetics and fly pushing, and I will strive to achieve the standard of her talks. I've pestered Abhishek and Saptarshi with many a differential equations doubts, all answered with aplomb. Both of them have introduced me to a variety of entities, both academic and non-academic, for which I will forever be grateful. It's been fun talking to Suvranil Ghosh about science and literature, and decoding what he meant to say. (Baby) Shaon Chakrabarti, Riya Bhatia, Julian Weninger, Simon, and Fazli Yousufzai were great to hang out with. Many people mentioned herein have been victims of my PJs; they have my deepest sympathies.

From discussing the technicalities of my project to hearing my rants and even suggesting templates for my thesis, Janaky has been an amazing companion and a constant support throughout whatever challenges that cropped up during my PhD. I cannot think of more ways she could have helped with the process. The extended family has also been quite fun to hang out with. It would be remiss not to mention close aunts, uncles, and cousins and everyone else who kept checking on my PhD progress and gave their well-wishes. My sister Aishwarya's personality could not be more different from mine, and therefore her perspectives and opinions have provided me with an often much-needed independent perspective. She's an inspiration to me more than she's aware, or I've let on.

My parents have been swell throughout this journey. As someone who started out with a lucrative B. Tech. in CSE, then switched to *theory*, and *then* to biology, I've given my parents many facepalm moments. Whenever I made an unconventional choice they would ask me if I knew what I was doing; I would lie and say yes, after which they would continue to support me no matter what. I couldn't have asked for better parents.

Contents

Abstract	vii
Acknowledgements	ix
0 Preface	1
1 Introduction	3
1.1 Aspects of the triumvirate pathway	3
1.1.1 Cargo transport	3
1.1.2 Endoplasmic reticulum	6
1.1.3 Golgi apparatus	6
1.1.4 Endosomes, endocytosis	8
1.1.5 Lysosomes	9
1.2 Glycans, Glycosylation	9
1.2.1 Michaelis-Menten kinetics	10
1.2.2 Reactions involving multiple substrates	12
1.2.3 Drug inhibition kinetics	12
1.2.4 The multifaceted roles of glycans	13
1.2.5 Measuring cell-surface glycans	14
1.3 Probing glycan manufacture	14
1.4 Probing glycan function	15
1.4.1 Elements of information theory	16
2 Pareto optimal glycan synthesis via enzyme compartmentalization	21
2.1 Modeling	21
2.1.1 One glycan manufacture	22
2.1.2 Multiple glycan manufacture	25
2.2 Pareto optimality, convexity, and all that	26
2.2.1 Connecting Pareto optimality and concavity	27
2.3 Results	29
2.3.1 One glycan synthesis	29
2.3.2 Multiple glycan synthesis	30
2.4 Discussion	32
3 Dynamics inside the cell	35
3.1 cis-Golgi dynamics	35
3.1.1 Experimental procedure	35
3.1.2 Birth-death processes	36
3.1.3 Results	38
3.2 Endocytosis dynamics	41
3.2.1 Experimental Procedure	41
3.2.2 Image Analysis	41
3.2.3 Results	43
3.3 Discussion	44

4	Dynamics at the cell surface	47
4.1	Experimental procedure	47
4.2	Modeling	48
4.2.1	Model with internal pools in quasi-steady state	51
4.3	Data analysis pipeline	51
4.4	Results	53
4.5	Discussion	57
5	Decoding cell-surface signals	65
5.1	Decoding to metapopulations	65
5.2	Advantage when using more than one glycan type	68
5.3	Information conveyed by cell-surface glycans	68
5.4	Discussion	69
6	Outlook	73
6.1	Probing glycan manufacture	73
6.2	Probing glycan function	73
6.3	Concluding remarks	74
	Bibliography	77

List of Figures

1.1	Biosynthetic, endocytic, and retrieval pathways	4
1.2	GEF-GAP cycle	5
1.3	Organization of the Golgi apparatus.	7
1.4	ConA and WGA bind to glycans.	14
1.5	Shannon's weaver model of communication	18
1.6	Graphical abstract of the Golgi fragmentation decoding study.	19
2.1	Example of enzyme compartmentalization.	22
2.2	Glycan assembly, and yields.	24
2.3	Different splits give different yields.	30
2.4	The space of Pareto optimal solutions have non-trivial structure.	31
3.1	BFA gives a gradient of Golgi phenotypes.	37
3.2	Distribution of Golgi phenotypes can be explained by a birth-death model.	39
3.3	Number of Golgi units as a function of time and concentration.	40
3.4	BFA-GCA calibration.	40
3.5	BFA does not affect endocytosis for $t \leq 30$ mins.	43
3.6	Bootstrap estimates for cis-Golgi fragmentation and endocytosis parameters.	45
3.7	Dynamics of endocytosis.	46
4.1	Schematic of the lectin labelling experiment.	47
4.2	Model schematic.	49
4.3	Shape parameters of the ellipse corresponding to a bivariate normal distribution	52
4.4	Comparison of WGA and ConA signal as a function of time.	54
4.5	Model fit is agnostic to glycoprotein half-lives.	55
4.6	Cell-surface glycan dynamics can be explained by Golgi fragmentation dynamics.	56
4.7	Noise in single-cell measurements.	58
4.8	Batch-to-batch normalization on single-cell data.	59
4.9	CGWR covariances are nearly constant across conditions.	60
4.10	CRWG covariances are nearly constant across conditions.	61
4.11	Bootstrap estimates for cell-surface glycan dynamics.	62
4.12	BFA does not appear to impact trans-Golgi morphology. We image the trans-Golgi in non-adherent WT-MEF cells using a fluorescently labeled marker (GalTase-RFP). Bottom: representative images showing the three trans-Golgi morphology classes. Top: Distribution of trans-Golgi morphologies in cell populations ($n > 100$) suspended for 60 min, then incubated for 30 min with varying concentrations of BFA. Error bars show the SEM over N replicate populations.	63
5.1	Parabolic trajectory of the means in log space, at different concentrations and times.	66

5.2	Golgi phenotypes can be decoded from single-cell measurements with accuracy far greater than chance.	67
5.3	ConA-WGA combination conveys more information than other lectin combinations.	68
5.4	ConA-WGA combination conveys more accurate information faster than other lectin combinations.	70
5.5	Single-lectin decoders.	72

List of Tables

1.1	Compartmentalization of glycosylation processing steps in the secretory pathway	11
3.1	Experimental procedure for generating a gradient of Golgi phenotypes. . .	36
3.2	cis-Golgi fragmentation parameter values	40
3.3	Experimental procedure for testing endocytosis of cell-surface lectins. . . .	42
3.4	Endocytosis parameter values	43
4.1	Experimental procedure for lectin labeling and flow cytometry analysis. . .	48
4.2	Golgi-to-PM and recycling parameter values	55

List of Abbreviations

BFA	Brefeldin A
CGWR	ConA-Green WGA-Red
CHO	Chinese Hamster Ovary
CRN	Chemical Reaction Network
CRWG	ConA-Red WGA-Green
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
ER GIC	ER-Golgi intermediate compartment
FBS	fetal bovine serum
PBS	phosphate-buffered saline
PFA	paraformaldehyde
PM	plasma membrane
TGN	trans-Golgi network

List of Symbols

$\text{Dom}(f)$	Domain of the function f
$\mathcal{E}_B$	Enzymes required to build glycan tree B
$[k], k \in \mathbb{N}$	the set $\{1, 2, \dots, k\}$
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$ S , S \text{ is a set}$	the cardinality of the set S
$[S]$	concentration of chemical species S
$\mathcal{S}$	split; partition of enzymes
$\vec{\mathcal{S}}$	set of splits corresponding to multiple glycans
$(\vec{\mathcal{S}}, \mathcal{T})$	strategy denoting splits and residence times assigned to each compartment
$S_N(\mathcal{E}_B)$	set of all N -compartment splits of $\mathcal{E}_B$
$\mathbf{x} < \mathbf{y}$	$\mathbf{y}$ Pareto dominates $\mathbf{x}$, see Definition 2.1.1

To Achan, Amma, Janu, Aish and my other sisters.

Chapter 0

Preface

A couple billion years ago, a series of chance events happened in a rather unassuming part of the universe we call the solar system. One of its members, the very planet we live in, played host to an as-yet unique event: it had created objects that could consume, store, replicate. *Live*.

Then slowly, these individual objects "realized" that often, their best chance to continue existence of their kind is to play nice with each other. That realization led to a proliferation of all kinds of coordination among individual molecules. Groups of molecules that coordinated, began to coordinate with other such systems, leading to *emergence* of coordination, *across multiple scales*. This space of meta-coordinations that "worked" comparably with their immediate neighborhood turned out to be extremely large: zillions of such instances began proliferating on Earth, resulting in entities that walk, crawl, fly, swim, slide, float, disperse, stand still. Life became even more beautiful.

This thesis is about one of the key ingredients that make this cooperation possible at the cellular level: **glycans**, the underrated [\[Var13\]](#), tree-shaped sugar chains covalently attached to proteins in the plasma membrane that perform so many functions where they are, essentially making multicellularity possible.

The first part of the thesis deals with how glycans get made. The second part of the thesis deals with how they perform one of their key functions – information transfer – after they are made.

We begin by laying out the philosophy of the thesis, and introduce its players.

Chapter 1

Introduction

The defining feature of eukaryotic cells is membrane-bound organelles, dividing the cell into structural and functional units. These organelles work in concert creating pathways that eventually result in an operational cell [Alb+22]. Of these pathways, the exchange of materials between cells is enabled by the secretory pathway, the endocytic pathway, and the retrieval pathway, together which we shall call the triumvirate pathway. These pathways comprise a common set of organelles: endoplasmic reticulum (ER), pre-Golgi intermediates, Golgi apparatus, trans-Golgi network, and so on. Figure 1.1 shows how the pathways work in tandem inside the cell and forms a graphical abstract of this chapter. We begin with a description of cargo transport and the organelles involved in these pathways.

1.1 Aspects of the triumvirate pathway

1.1.1 Cargo transport

Cargo is transported using carriers called vesicles that bud off organelle membranes. Vesicles are categorized by the coats that cover them. The four most prominent examples of vesicle coats are COPI, COPII, clathrin, and retromer coats. COPI vesicles are responsible for retrograde transport from the Golgi to the ER, and transport between different compartments of the Golgi. They are divided into COPIa and COPIb vesicles. COPIa vesicles transport material between Golgi and ER, and COPIb vesicles perform intra-Golgi transport (Figure 1.3). COPII vesicles enable anterograde transport from the ER to the Golgi. Clathrin-coated vesicles mediate transport between trans-Golgi, plasma membrane (PM), and endosomes [Alb+22].

Vesicle budding is an intricate process, and a high-level description of it is as follows:

1. Cargo binds to cargo receptors in the membrane of the source organelle.
2. Cargo receptors bind to adaptors.
3. Adaptors recruit coats. The region of the membrane around the bound cargo starts bending to obtain a spherical shape. During this time, more cargo and coats are loaded onto the vesicle.
4. Vesicles pinch off of membranes.

In the case of COPI vesicles, COPI coat recruitment is regulated by a coat recruitment GTPase known as Arf1p. GTPases are a large family of hydrolase¹ enzymes that bind to guanosine triphosphate (GTP) and hydrolyze it to guanosine diphosphate (GDP)

¹enzymes that use water to break a chemical bond.

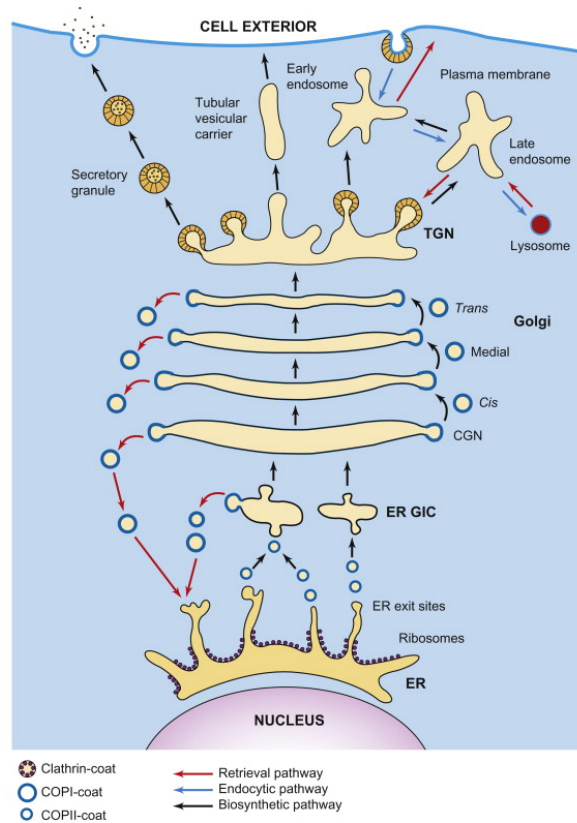


FIGURE 1.1: Biosynthetic, endocytic, and retrieval pathways [PMB12]. Secretory pathway exports materials from a cell, while the endocytic pathway imports them from the extracellular surroundings. Retrieval pathways ensure correct localization of proteins that have been transported elsewhere. ER GIC: ER-Golgi intermediate compartment.

[Sto+93]. GTPases act as switches for several processes in eukaryotic cells such as signal transduction, protein biosynthesis, and in our particular case, formation of vesicle coats [Gil87]. The switch itself is regulated by two classes of proteins called guanine exchange factors (GEFs) and GTPase-activating proteins (GAPs) [Alb+22]. GEFs catalyze the exchange of GDP for GTP, activating the now GTP-bound protein. GAPs have the opposite effect: they trigger hydrolysis of bound GTP to GDP inactivating the now GDP-bound protein. A schematic of this process is shown below in Figure 1.2. These details are important in understanding the effect of brefeldin A, a reagent used in our work that we will introduce shortly.

Note that cargo transport is selective with regard to both the entity that is transported and the destination to which it is transported. Cargo adapters ensure that the appropriate cargo is loaded onto the vesicles. Rab proteins direct vesicles to the correct spot on the target membrane for fusion [HN11]. Soluble NSF association receptor proteins, or SNAREs, mediate vesicle membrane fusion [BG04].

SNAREs are divided into two categories: v-SNAREs, which are embedded into vesicle membranes, and t-SNAREs, which are present on target membranes. Vesicle fusion and cargo release only happen when v-SNAREs recognize their corresponding t-SNAREs and bind. More recently, a broader classification scheme has been proposed for SNAREs taking their structure into account. In this regimen, vesicles are divided into Q-SNAREs and

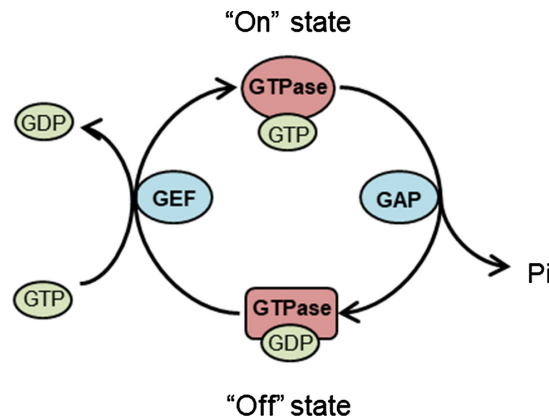


FIGURE 1.2: A canonical GEF-GAP cycle to regulate the activity of GTPases. GEF: guanine exchange factor; GAP: GTPase activating protein. [Wes+21].

R-SNAREs, with Q- and R-SNAREs acting as t- and v-SNAREs respectively [MS11].

Much of our understanding of the process of secretion and cargo transport comes from transport inhibition experiments. Experiments typically involve inhibiting a component of the cell and looking at the effect on secretion, thereby elucidating the role of the component in the secretory pathway. Some of the earlier results in cargo transport – from the 1970s and the 1980s – were obtained using a reagent called Brefeldin A. Since BFA is a central reagent in this thesis, we discuss it in more detail next.

Brefeldin A, “an enigmatic inhibitor of secretion”

Brefeldin A (BFA), a lactone antiviral produced by the fungus *Penicillium brefeldianum* [Hut+83], was initially isolated as a potential antiviral drug [Tam+68], but has since been repurposed to study protein transport. Despite three decades of research, a complete picture of the action of brefeldin A on various cell types continues to elude us. This is in part due to its widely varying effects depending on the cell type [Ito+12]. We lay out known effects.

In murine embryonic fibroblasts (MEFs), BFA targets Arf1 GEFs GBF1 and BIG1/2 [Raj+23]. They regulate GTPase Arf1, and its proper functioning is necessary for the recruitment of COPI and clathrin coats. If their activity is inhibited by the application of BFA, the following events occur [NRR02]:

- Recruitment of Arf1 GTPase is inhibited. This prevents the formation of COPI vesicles that perform retrograde transport between the Golgi and the ER.
- SNAREs, which normally would have been present on COPI vesicles to help them fuse with the ER, begin to accumulate on the cis-Golgi.
- The accumulation of SNAREs causes aberrant fusion of the cis-Golgi with the ER.

After the fallback of cis-Golgi to the ER, the glycoprotein modification reactions that should have happened in the cis-Golgi continue to happen inside the ER. However, anterograde vesicles may not be able to fuse with the remaining Golgi cisterna [NRR02], and the rate of delivery of the modified glycoprotein substrate to the trans-Golgi falls.

Next, we look at where all cargo gets transported from, and to.

1.1.2 Endoplasmic reticulum

The endoplasmic reticulum consists of membranous tubules and sacs called cisternae. It is made up of two subunits, the smooth endoplasmic reticulum (SER) and the rough endoplasmic reticulum (RER). The ER spans the volume of the cell: RER is found in the periphery of the nucleus and the SER near the plasma membrane.

The RER cisternae are studded with ribosomes² giving it a rough appearance, and its name. Ribosomes are responsible for protein synthesis and folding. They link amino acid chains in the order specified by messenger RNAs (mRNAs) in a process called translation. The rate of amino acid synthesis varies across organisms; in yeast, about 6-9 amino acids are linked per second [RO82].

Some of these manufactured proteins are secreted by the cell. Most secretory proteins are glycoproteins – proteins with carbohydrates covalently bonded to them – and the central entity of this thesis. The membrane of rough ER separates the proteins synthesized by it from the proteins synthesized by free-floating ribosomes in the cytosol. The rough ER is also the membrane factory of the cell; membrane proteins grow from ribosomes which are inserted into the ER membrane itself. Then, as the ER membrane expands, membrane proteins are transferred via vesicles to the other components of the endomembrane system.

The smooth ER participates in various metabolic processes, although the extent varies by cell type. Its functions include the synthesis of lipids, the metabolism of carbohydrates, the detoxification of drugs and poisons, and the storage of calcium ions. The outer surface of the SER lacks ribosomes.

1.1.3 Golgi apparatus

After protein synthesis and some modifications in the ER, protein substrates enter the cis-Golgi network and eventually into the Golgi apparatus. The Golgi consists of a series of flattened membrane-enclosed compartments called cisternae. The cisternae are ordered and labeled as cis, medial, and trans, the cis face being close to the ER and the trans face to the plasma membrane. Enzymes that act early on the protein substrate are mostly found in the cis-Golgi, and late-acting enzymes are found in the trans-Golgi. The enzymes broadly shape the function of each of the compartments as follows [Alb+22]:

- cis-Golgi network: phosphorylation of oligosaccharides on lysosomal proteins.
- cis cisterna: removal of mannose residues.
- medial cisterna: removal of mannose residues, addition of GlcNAc residues.
- trans cisterna: addition of galactose, sialic acid.
- trans Golgi network (TGN): sulfation of tyrosines and carbohydrates.

A more elaborate description of enzymes involved, their localization, and function is given in Table 1.1. The Golgi thus performs an ordered series of covalent modifications on the protein substrate. The amount of time a protein spends in a compartment is called the residence time of that compartment. If a compartment has a long residence time, then a

²They are not bound to the RER at all times; they keep binding and unbinding.

MAMMALIAN GOLGI

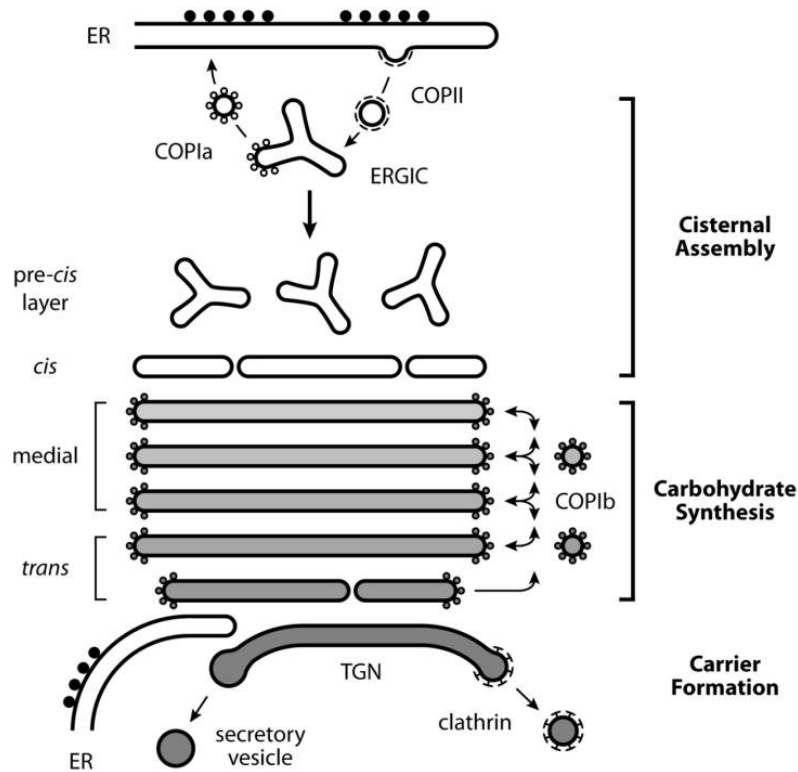


FIGURE 1.3: Three-stage model of the Golgi [DSG13]. Cargo traverses through cis-, medial- and trans compartments, and the trans-Golgi network before being carried off elsewhere. COPIa vesicles carry out retrograde (Golgi-to-ER) transport, and COPIb vesicles carry out intra-Golgi transport. COPII vesicles perform anterograde transport from ER to Golgi. Clathrin-coated vesicles transport material from Golgi to PM.

protein substrate in that compartment is expected to undergo more of the types of reactions ascribed to that compartment. The time it takes for a substrate to traverse the entire Golgi stack is estimated to be around 15 to 20 minutes [UB97].

The transport of cargo to and from these compartments was discussed in Section 1.1.1, and a schematic of the three-compartment model together with transport carriers is shown in Figure 1.3. After glycosylation is completed in the trans-Golgi (a process we will look at shortly), the Golgi sorts the modified protein for transport to various destinations such as lysosomes, plasma membrane, or secretory vesicles. Due to its assembly line nature and sorting functions, similes to manufacturing factories have been ascribed to the Golgi.

Several models have been proposed to explain how the Golgi matures [GL11]. These models have been proposed based on our interpretation of results of experiments blocking vesicle formation [Lip+89], visualizing cargo movements [Pat+08] and tracking Golgi-resident enzymes [Rab+95]. We review the two most prominent of the Golgi maturation models.

Vesicle transport model

The vesicle transport model was the widely accepted model of Golgi maturation from the early 1980s to the late 1990s, and remains in use. The Golgi is viewed as having a set of stable compartments operating in tandem. Newly synthesized secretory cargo is delivered to the cis-Golgi via COPII vesicles. Intra-Golgi transport is mediated via COPI vesicles. There are both anterograde and retrograde COPI vesicles; the former carries secretory cargo forward, and the latter recycles trafficking components. Since each cargo is brought in by its own carrier, the cargo can have differing residence times in the same compartment. The enzymes in the Golgi are not loaded into the intra-Golgi COPI vesicles; the compartments thus retain their characteristics.

However, the transport model cannot explain how large cargo traverses the Golgi, since the vesicles are too small for large cargo to be loaded. In addition to this, evidence for anterograde COPI vesicles is inconclusive.

Cisternal maturation model

In this model, cis-Golgi compartments form by homotypic fusion of ER-derived vesicles. COPI vesicles recycle Golgi proteins and enzymes from older cisternae thereby "maturing" the younger cisternae [GL11]. The system thus behaves as though cargoes in the same compartments are subject to the same residence times. This is a key difference from the vesicle transport model, where each cargo could have its own residence time.

Some observations are yet to be reconciled with the cisternal maturation model. Significant levels of Golgi enzymes have not been detected in COPI vesicles. The kinetics of cargo exit is expected to be linear under this model, since cisternae are created at a constant rate at the cis-face, mature, and are consumed by vesicles at the trans-face at a constant rate. This has not yet been conclusively observed.

Despite its shortcomings, the cisternal maturation model is currently the most accepted model of Golgi maturation [HMH07; GL11; MT16].

1.1.4 Endosomes, endocytosis

The endosomes are the sorting station of the endocytic pathway, just as the Golgi functions as the sorting station in the secretory pathway. Endosomes exist in three variants: early, late, and recycling endosomes [Mel96].

Materials to be ingested into the cell are progressively enclosed by a small portion of the plasma membrane. This is followed by the enclosure budding off and forming an endocytic vesicle. The endocytic vesicle fuses with early endosomes (EE), and EEs are the first point of entry into the cell. Here the endocytosed cargo is sorted, after which two things can happen: either cargo remains in the EE, or it is returned to the PM directly or via recycling endosomes. The EE then matures into late endosomes, before they fuse with lysosomes.

Mechanisms of endocytosis

Two modes of endocytosis have been well characterized, although there are several others [Alb+22]:

- **Clathrin-mediated.** In this mode of endocytosis, clathrin (discussed in Section 1.1.1) induces the formation of a clathrin-coated pit on the plasma membrane. This pit later buds off as a clathrin vesicle, taking along with it material from outside the cell that "fell" into the pit. Often, receptors are embedded into these clathrin-coated pits, which (1) enhances the selectivity of endocytosed material and (2) increases the efficiency of endocytosis. Note that the AP1 adapter protein is involved in clathrin formation, so the effect of BFA on endocytosis will have to be tested for a comprehensive understanding of cell-surface glycan dynamics.
- **Caveolae-mediated.** Caveolae are pits in the plasma membrane whose structural is maintained by cholesterol-binding proteins called caveolins, and by cavin proteins, which stabilize the caveolae. Caveolins are integral membrane proteins (but not trans-membrane). BFA does not affect caveolin-mediated endocytosis.

1.1.5 Lysosomes

Lysosomes are membranous sacs consisting of hydrolytic enzymes that many eukaryotic cells use to hydrolyze³ or break down macromolecules. The lysosomes have an acidic environment that is conducive to the functioning of these hydrolytic enzymes. The cytosol has a near-neutral pH at which the enzymes are not active, so when a lysosome breaks, it is not detrimental to the cell (excessive leakage can destroy a cell by self-digestion). The enzymes and membranes of lysosomes are manufactured by rough ER, which are then transferred to the Golgi for further modification. Some lysosomes derive from budding at the trans face of the Golgi.

The ER, Golgi apparatus, endosomes, and lysosomes form a minimal set of organelles in any model of cargo transport.

1.2 Glycans, Glycosylation

Once proteins are synthesized in the ER, some of these proteins undergo a variety of post-translational modifications [WGG05]. A part of these post-translational modifications comprises glycosidical⁴ linkage of monosaccharides atop the protein substrates. The reaction is called glycosylation. The sugar chain structure that results is a glycan [IUP14], and the entire structure is called a glycoprotein.

Of all the characteristics of living cells, glycans are arguably the most diverse in terms of chemical complexity and evolutionary diversification [Var17]. This diversity stems from two factors: (1) diversity among monosaccharides, of which there are commonly nine; and (2) diversity in terms of the position where glycosidic linkage takes place. A monosaccharide might be capable of forming multiple bonds, and so at a particular site one may find two, three, or even four branches. Because of this, glycans are tree-shaped.

Glycans are classified into two types based on how sugar chains are linked to the polypeptide backbone:

- N-glycans, where sugar chains are attached to the Asn residue in a sequon⁵ of type Asn-X-Ser or Asn-X-Thr, where X can be any amino acid except proline; and

³any chemical reaction in which a molecule of water breaks down chemical bonds.

⁴a type of bond that joins carbohydrates or sugar molecules to other groups which may or may not be carbohydrates.

⁵sequence of consecutive amino acids.

- O-glycans, where sugar chains are linked via the oxygen atom of a serine or threonine residue of a peptide chain.

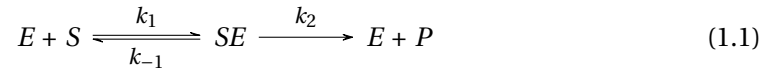
N-glycosylation is more common than O-glycosylation [Apw99]. Before the glycan can be assembled on a protein substrate, the substrate is readied in a carefully orchestrated manner proceeding as follows [Rot02]:

1. A lipid-linked oligosaccharide precursor is assembled step by step in a highly conserved process across evolution.
2. The preassembled oligosaccharide, whose structure is $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-Dol-P-P}$ is transferred to an asparagine in a nascent protein.
3. The structure in the previous step has three glucose and mannose residues; these three glucose residues and some mannose residues are removed from the structure⁶.

Many reactions in biology, including the formation of glycans, comprise two chemical species that join together in the presence of an enzyme to form a product. The simplest of these reactions involves a single substrate and a product. In the case of glycans, the class of enzymes responsible for their manufacture are glycosyltransferases and glycosidases. The enzymes responsible for various glycosylation reactions, and their site of (relative) abundance, are given in Table 1.1.

1.2.1 Michaelis-Menten kinetics

The dynamics of enzyme-catalyzed reaction systems, such as the ones that manufacture glycans, was quantitatively studied by Maud Menten and Leonor Michaelis [JG11; Mur02]. They imagined a substrate S reacting with an enzyme E to form a complex SE which is then converted into a product P and the enzyme:



The Law of Mass Action states that the rate of an elementary reaction is proportional to the product of concentrations of the reactants raised to their stoichiometric coefficients. Letting $s = [S]$, $e = [E]$, $c = [SE]$ and $p = [P]$ denote the concentrations⁷ of substrate, enzyme, complex, and product respectively, the Law of Mass Action applied to 1.1 yields the following set of differential equations:

$$\frac{ds}{dt} = -k_1 es + k_{-1} c, \quad \frac{de}{dt} = -k_1 es + (k_{-1} + k_2) c \quad (1.2)$$

$$\frac{dc}{dt} = k_1 es - (k_{-1} + k_2) c, \quad \frac{dp}{dt} = k_2 c \quad (1.3)$$

with the initial conditions $s(0) = s_0$, $e(0) = e_0$, $c(0) = 0$, $p(0) = 0$. The k 's are called rate constants. Note that the total amount of enzyme is conserved since it is not consumed during the reaction; this means $e(t) + c(t) = e_0$ (say). Substituting $e = e_0 - c$ in Equations 1.2 and 1.3, we get:

⁶This step is actually a key event in protein quality control; structures that retain the mannose residues are sent to lysosomes for degradation [Alb+22].

⁷ $[\]$ denotes concentration.

<i>Compartment</i>		<i>Processing Step</i>	<i>Enzymes</i>
Endoplasmic reticulum		N-linked glycosylation	ER mannosidase
		Removal of mannose	Glucosidase I
		Removal of 2 glucoses	Glucosidase II
cis-Golgi network		Phosphorylation of oligosaccharides in lysosomal luminal proteins O-linked glycosylation	N-acetylgalactosaminyl transferase
cis-Golgi terna	cis-	Removal of mannose	Golgi mannosidase I
medial Golgi cisterna		Removal of mannose Addition of N-acetylglucosamine	N-acetyl glucosamine transferase I
trans-Golgi terna	cis-	Addition of galactose	Golgi mannosidase II
trans-Golgi network		Addition of NANA Sulfation of tyrosines and carbohydrates Addition of sialic acid Selective recognition of M6P by M6P receptor	α 2,6-sialyltransferase

TABLE 1.1: Compartmentalization of glycosylation processing steps in the secretory pathway [PMB12]

$$\frac{ds}{dt} = -k_1 s e_0 + (k_1 s + k_{-1}) c \quad (1.4)$$

$$\frac{dc}{dt} = k_1 e_0 s - (k_1 s + k_{-1} + k_2) c \quad (1.5)$$

It is typically assumed that the complex formation is a fast step, post which the complex concentration attains equilibrium. Setting $\frac{dc}{dt} = 0$ in equation 1.5 yields the steady-state complex concentration as

$$c = \frac{k_1 e_0 s}{k_1 s + k_{-1} + k_2} = \frac{e_0 s}{s + \frac{k_{-1} + k_2}{k_1}}.$$

The fraction in the denominator $\frac{k_{-1} + k_2}{k_1}$ is called the Michaelis constant and is denoted as K_m . Now the expression for c can be substituted in equation 1.4, and with some algebra one gets:

$$\frac{ds}{dt} = -\frac{k_2 e_0 s}{s + K_m}. \quad (1.6)$$

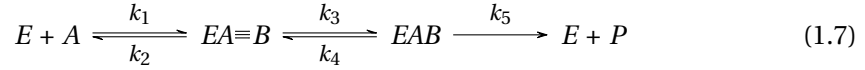
If one assumes that the substrate concentration $s \ll K_m$, this permits an approximation to the Michaelis-Menten equation 1.6:

$$\frac{ds}{dt} = -\frac{k_2 e_0}{K_m} s.$$

In other words, the reaction obeys first-order kinetics. Such conditions are often met in case of glycosylation reactions [LD08].

1.2.2 Reactions involving multiple substrates

More broadly, one can imagine a substrate reacting with another entity (such as a sugar donor) to produce a product:



In our context, E, A, and B corresponds to the glycosyltransferase enzyme, the sugar donor (such as UDP-GlcNAc), and the growing glycan substrate respectively (the first intermediate usually comprises the monomer and the enzyme). EA and EAB denote the complexes formed, and P denotes the glycan with the monomer added. It can be shown that under steady-state assumption of the intermediates, the rate of formation of product is given by [CC07]:

$$v = \frac{k_1 \cdot k_3 \cdot k_5 \cdot [A][B][E]_t}{k_2(k_4 + k_5) + k_1(k_4 + k_5)[A] + k_3 k_5 [B] + k_1 k_3 [A][B]}$$

where $[E]_t = [EA] + [EAB] + [E]$. When the monomer A is in excess, this equation reduces to

$$v = \frac{k_5 [B][E]_t}{\frac{k_4 + k_5}{k_3} + [B]}$$

Comparing this with the Michaelis-Menten (Briggs-Haldane) equation $v = \frac{k_2 [B][E]_t}{\frac{k_{-1} + k_2}{k_1} + [B]}$

for the reaction schematic $E + S \xrightleftharpoons[k_{-1}]{k_1} SE \xrightarrow{k_2} E + P$, we see that the parameters above match those of the Michaelis-Menten equation with $k_2 = k_5$ and $\frac{k_4 + k_5}{k_3} = \frac{k_{-1} + k_2}{k_1}$.

1.2.3 Drug inhibition kinetics

The kinetics of a substrate when an inhibitory drug is added can be studied in a similar manner. The drugs bind to their targets and form a complex that prevents the target from performing its function. Consider the following chemical equation:



where D is the drug and R is the target. At equilibrium, the forward and reverse reaction rates are equal:

$$k_{on}[D][R] = k_{off}[DR].$$

The ratio of k_{off} and k_{on} is called the dissociation constant K_D :

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[D][R]}{[DR]} \quad (1.9)$$

The fraction of target that is unbound (and capable of functioning) is given by

$$f_{\text{unbound}} = \frac{[R]}{[R] + [DR]}$$

Substituting for $[R]$ from equation 1.9, we get

$$f_{\text{unbound}} = \frac{1}{1 + [D]/K_D}.$$

1.2.4 The multifaceted roles of glycans

Given that glycans are the first point of contact for any foreign object that interacts with a cell, it is only natural that they participate in multiple functions. We sample some of them here [Var17]:

- **Protein folding.** The function of a glycan begins with the protein it is attached to. Chaperone proteins bind to glycans and help in proper folding of its host protein.
- **Structural functions.** Glycans provide strength and rigidity to the cell walls of plants and fungi, and the exoskeletons of arthropods. Thick layers of glycans perform critical barrier functions in many epithelial surfaces such as the inner lining of airways and intestines.
- **Immune response.** Glycans are sensed by a variety of glycan-binding proteins that help initiate pro- and anti-inflammatory immune response. Cells that are affected by pathological conditions have abnormal glycans on their surface that help immune cells identify and remove affected cells. For instance, tumor cells with aberrant glycosylation are identified by receptors on Natural Killer (NK) cells. The glycans on an immune cell also help it migrate to its site of action.
- **Self-nonsel self discrimination.** Glycans help form signatures that help one entity of an organism recognize another entity of the same organism. Glycans are the reason why cross-species fertilization does not occur.
- **Cell adhesion.** Bulky glycoproteins in cell surface glycocalyx⁸ can indirectly promote cell adhesion and signaling which are necessary to support cell growth and survival. In some cases, bulky and negatively charged glycoproteins can have anti-adhesive actions which inhibit cell-cell and cell-matrix interactions, e.g. this plasticity is important for neuronal migration, reorganization following injury etc.
- **Degradation.** N-linked glycans with a mannose-6-phosphate residue serve as signals that the glycoprotein be sent to lysosomes for degradation.
- **Storage.** Glycans help in long-term energy storage; glycogen in animals and starch in plants are examples.
- **Signaling.** Glycans mediate crosstalk [OMT22], direct trafficking and membrane localization of receptors [Pas+14]. They facilitate many functions including receptor activation and ligand binding [RKO91], cell-cell interaction during early development [Zho11], signal modulation [Reb+96], immune signaling [Zho+21], regulation of cell death [LR13], tumor microenvironment signaling [CCR19], and so on.

The signaling ability of glycans to convey information about the cell will be the focus of the second part of the thesis.

⁸layer of glycoproteins and glycolipids which surround the cell membranes of bacteria, epithelial cells, and other cells.

1.2.5 Measuring cell-surface glycans

The dynamics of cell-surface glycans in single cells can be tracked using lectins, or proteins that bind to carbohydrates. We use Concanavalin A (ConA), which binds to oligo-mannose glycans, and wheat germ agglutinin (WGA), which binds to glycans with sialic acid and N-GlcNAc terminal units [Sta11]; see Figure 1.4.

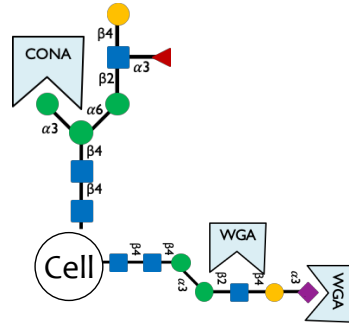


FIGURE 1.4: Fluorescently tagged lectins can be used to measure cell-surface glycan distributions. ConA binds to oligo-mannose glycans, and WGA binds to glycans with sialic acid and N-GlcNAc terminal units. (Figure credits: Prachi Joshi.)

ConA-binding glycans are transported from the cis-Golgi to the trans-Golgi, where a fraction of them are further glycosylated to create WGA-binding glycans [Fis+19a]. For our experiments, we tagged the lectins with green and red fluorophores (Alexa488 and Alexa647); the exact matching of ConA and WGA to the fluorophores is different in different experiments and will be detailed when the experiments are discussed. The lectins were allowed to bind to glycans on MEF cells, and flow cytometry was performed. Flow cytometry outputs the light intensity of these fluorophores at a single-cell resolution, and is reflective of the glycan distribution on the surface of each cells. By probing glycans processed in different Golgi compartments, we can distinguish the effects of trafficking and glycosylation.

1.3 Probing glycan manufacture

In most works that study glycan manufacture, the broad template followed is to model glycosylation and fit the model to an observed glycan distribution. The models usually have *chemical parameters* (e.g. rates of glycosylation enzymes) and *physical parameters* (e.g. number of cisternae, residence times).

The first efforts to model glycan manufacture were by Umana and Bailey in 1997 [UB97]. They calculated qualitative trends in N-linked oligosaccharide distribution resulting from changes in the levels of one or more enzymes involved in the chemical reaction network (CRN) of N-linked oligosaccharide biosynthesis. They viewed the Golgi apparatus as a system comprising four reactors in series, corresponding to cis, medial, and trans-Golgi cisternae and trans-Golgi network. Their model was derived from experimental data on Chinese hamster ovary (CHO) cells and had mass balances for 33 different oligosaccharides. At that time, other models for glycosylation by specific enzymes, such as sialyl-transferases [MAG97], were also developed.

Krambeck and Betenbaugh [KB05] extended Umana and Bailey's model to include further galactosylation, fucosylation, extension of antennae by N-acetylglucosamine repeats,

and sialylation. They adopted the vesicle transport model of the Golgi, and their model considered a total of 7,565 oligosaccharide structures in a network of 22,871 reactions. Enzyme activities were adjusted to match glycoform distributions for CHO. The model predicted downstream effects that were hard to guess intuitively, such as an increase in antennarity on enhanced ManI activity. A further extension in 2009 (KB2009) enabled a direct fit of enzyme localization to mass spectrometric data rather than relying on human annotation of peaks in the data [Kra+09]. In summary, there has been a surge of interest in modeling glycosylation, driven largely by pharmacological motivation [HMH07; LN08; Jef09; VNK11; Spa+16; Kra+17; Fis+19a; Fis+19b; Kot+23].

The Golgi manufactures a vast repertoire of glycans at a given time [Fis+19a]. Glycan substrates undergoing glycosylation in the same cisterna are exposed to its environment for the same amount of time, before the cisterna matures [BG04]. 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 could execute undesirable reactions [BT20]. Both of these events result in the loss of glycoprotein substrate, and incur significant biosynthetic costs to the cell [Gut+20].

We therefore assume that the enzyme distribution across the Golgi compartments and their residence times must be designed to minimize loss of glycoprotein substrate and 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. [Jef09; Zho+22]. The notion has been employed with success in engineering [Deb99], social sciences [ABH13], and several areas of biology such as ecological studies [Sho+12; TMA15; Lev62], neuroscience [JBC22], 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 tradeoffs. Pareto optimal solutions [Mie98] are a class of solutions that reconcile this tradeoff. In Chapter 2, we study the Pareto set of enzyme distribution and residence times in the context of multiple glycan manufacture.

1.4 Probing glycan function

As mentioned in the previous section, glycans are the channels through which cells dissipate information to the extracellular world. This is surprising considering that there is significant cell-to-cell variability in glycan types, even among the same developmental lineage [Min+21]. How then do glycans function as molecular reporters despite such variability [BS05]?

To concretize this question, let us take the particular case of Golgi dysfunction. Disruption of Golgi organization and intra-Golgi trafficking impacts the types and levels of glycans on the cell surface [Fis+19a; Fis+19b; Pot+21; Bek+17; ZW16]. Aberrant glycosylation is implicated in cancers [Llo+96; HU22], pathogenic infections [Jon+16; Ols02],

autoimmune disorders [Jon+16; Zho+21], and genetic diseases [FN11]. Clearly Golgi dysfunction is correlated with changes in cell-surface glycan distribution.

In this thesis we quantify the extent to which information about Golgi dysfunction is conveyed by cell-surface glycans. Previous related works quantify the information glycan *receptors* are able to convey to a cell, on exposure to the same glycan [Fuc+23]. We discuss preliminaries [Mac02; CT06].

1.4.1 Elements of information theory

The ensemble is the basic data structure:

Definition 1.4.1 (Ensemble). An ensemble is a triple $(x, \mathcal{A}_X, \mathcal{P}_X)$ where

- x is the outcome, or value of a random variable
- $\mathcal{A}_X$ is the set of values x can take; $\mathcal{A}_X = \{a_1, \dots, a_I\}$
- $\mathcal{P}_X$ is the set of probabilities $\{p_1, \dots, p_I\}$ with $P(x = a_i) = p_i, p_i \geq 0$ and

$$\sum_{a_i \in \mathcal{A}_X} P(x = a_i) = 1$$

From the ensemble we can calculate probabilities of events:

Definition 1.4.2 (Probability of a subset). If $T \subseteq \mathcal{A}_X$, then

$$P(T) = P(x \in T) = \sum_{a_i \in T} P(x = a_i)$$

Information decreases uncertainty. One can think of obtaining information as reducing the set of possible outcomes of an experiment. Surprising events tell us more about a process or an event; for instance, saying that a word starts with `xy1 . . .` helps us narrow down the communicated word than if the word had started with `pro . . .`. The Shannon information content of an outcome is a quantity that captures this idea that rare events convey more information about a system:

Definition 1.4.3 (Shannon information content of an outcome). The Shannon information content of an outcome x is defined as

$$h(x) = \log_2 \frac{1}{P(x)}.$$

(In this section, logarithm is taken with base 2 unless otherwise noted.) The definition of $h(x)$ has extremely nice theoretical properties [Mac02] that we won't go into. We can now define the Shannon entropy of a random variable:

Definition 1.4.4 (Entropy of an ensemble). The entropy $H(X)$ of an ensemble X is defined as

$$H(X) = \sum_{x \in \mathcal{A}_X} P(x) \log \frac{1}{P(x)}$$

In other words, it is the average Shannon information content associated with the outcomes of an experiment. For example, the entropy associated with the outcome of a fair coin toss is one bit.

What if you want to study pairs of events? The data structure analogous to the ensemble has a name:

Definition 1.4.5 (Joint ensemble). A joint ensemble XY is an ensemble in which each outcome is an ordered pair x, y with $x \in \mathcal{A}_X, y \in \mathcal{A}_Y$. $P(x, y)$ denotes the joint probability of x and y .

The joint entropy of X, Y is defined as

Definition 1.4.6 (Entropy of a joint ensemble).

$$H(X, Y) = \sum_{x \in \mathcal{A}_X, y \in \mathcal{A}_Y} P(x, y) \log \frac{1}{P(x, y)}.$$

One can recover the individual ensemble probabilities from the joint like so:

Definition 1.4.7 (Marginal probabilities). The marginal probability $P(x)$ can be obtained from the joint probabilities $P(x, y)$ by summation:

$$P(x = a_i) = \sum_{y \in \mathcal{A}_Y} P(x = a_i, y).$$

Knowing whether an event has taken place can change the probability associated with another event. This notion is captured by conditional probabilities:

Definition 1.4.8 (Conditional probability).

$$P(x = a_i | y = b_j) = \frac{P(x = a_i, y = b_j)}{P(y = b_j)}.$$

provided $P(y = b_j) \neq 0$ (it is undefined if $P(y = b_j) = 0$).

How much does one's uncertainty decrease given another event has taken place? Conditional entropy quantifies this notion:

Definition 1.4.9 (Conditional entropy). The conditional entropy of X given $y = b_k$ is the entropy of the probability distribution $P(X|y = b_k)$, i.e.

$$H(X|Y = b_k) = \sum_{x \in \mathcal{A}_X} P(x|y = b_k) \log \frac{1}{P(x|y = b_k)}.$$

This can be generalized to a random variable:

Definition 1.4.10. The conditional entropy of X given Y is the average over y , of the conditional entropy of X given y , i.e.

$$H(X|Y) = \sum_{y \in \mathcal{A}_Y} P(y) \left(\sum_{x \in \mathcal{A}_X} P(x|y = b_k) \log \frac{1}{P(x|y = b_k)} \right).$$

We are finally ready to quantify how much information one random variable conveys about another:

Definition 1.4.11 (Mutual information between two random variables). The mutual information $I(X; Y)$ between random variables X and Y is defined as

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

We won't prove the equality; see [Mac02]. To give labels to X and Y above, one can let X correspond to the glycan distribution and Y correspond to a cellular property of interest; mutual information then quantifies how much information glycans convey about the

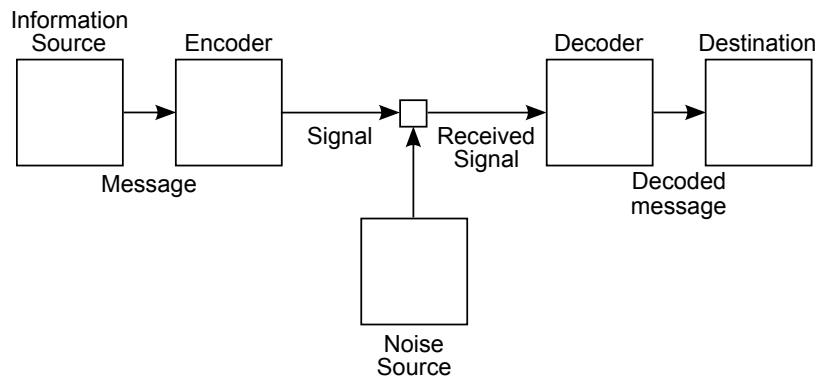


FIGURE 1.5: Shannon's weaver model of communication [Wan08]. The information source selects a message to send, which the encoder converts into a signal which can be sent through a channel. It is received at the decoder end, which tries to reconstruct the (error-prone) message. Adapted from [Phl].

cellular property of interest.

Communication between two entities was modeled by Shannon [Sha48] as shown in Figure 1.5: the information source has a message to send, which is encoded via the transmitter into bits (or another entity such as light pulses, molecules, etc.) that can be sent across a (noisy) channel. The bits (some of which might have flipped due to noise) reach a receiver, who then decodes the bits to try to obtain the original message. In our context, glycans can be viewed as the channel through which the Golgi communicates.

[Fuc+23] calculated the mutual information between glycan dosage and $\text{NF-}\kappa\text{B}$ ⁹ expression when glycans are read by different receptors; in this sense, it becomes a measure of the information sensing capacity of the receptor. [Fuc+23] thus focuses on the receiver side of Figure 1.5. We ask the opposite question: how much information about the cell state is encoded in glycans?

There are two ways to answer this question: (1) quantify the accuracy with which a reader cell is able to decode the underlying cell state and (2) quantify the mutual information between different glycan readouts and the underlying cell state. We do both in the thesis. For this task, we take the extent of Golgi fragmentation to be the cell state, and ask the accuracy with which a reader cell is able to infer the extent of Golgi fragmentation. Golgi morphology can be perturbed by inhibiting the vesicle trafficking regulator Arf1 using brefeldin A (BFA) [Sci+97; Str+93; War+01] and golgicide A (GCA) [S  e+09; Raj+23]. We developed a chemical kinetic model of the action of BFA on cell surface glycan levels and fit it to experimental data. The experimental data came from quantifying endocytosis rates, cis-Golgi fragmentation dynamics, and cell-surface glycan levels at different concentrations of BFA and GCA and the various time durations they were applied for. The fitted model is then used to construct the decoder, whose accuracy we finally quantify.

Chapter 2 explores enzyme compartmentalization for optimal glycan manufacture.

⁹a transcription factor that regulates multiple aspects of innate and adaptive immune functions [Liu+17].

Chapter 3 discusses the dynamics of cis-Golgi fragmentation and endocytosis on BFA action. Chapter 4 fleshes out the full model of glycosylation and describes how the cell surface glycan dynamics is influenced by the dynamics of cis-Golgi fragmentation and endocytosis. The model is then used to decode Golgi fragmentation in Chapter 5. We take stock of what has been achieved in Chapter 6, and list some broad ideas for future work. A graphical abstract summarising our full approach, which explains how Golgi morphology data and endocytosis data are fed into the model, how predictions of the model are compared to cell-surface lectin measurements, and how ConA and WGA measurements can be used to decode the Golgi phenotype, is given in Figure 1.6. The GitLab repo corresponding to the decoding study can be found at <https://gitlab.com/aashishsatya/decoding-golgi-perturbations>.

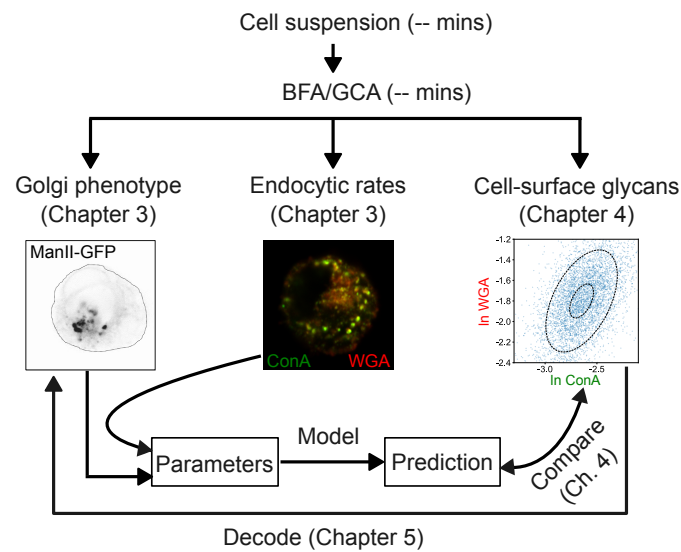


FIGURE 1.6: Graphical abstract of the Golgi phenotype decoding study. We use BFA and GCA to inhibit the Arf1 GEFs GBF1 and BIG1/2 and disrupt the Golgi. We measure three properties: Golgi phenotype (Chapter 3), endocytosis rates (Chapter 3), and cell-surface lectin levels (Chapter 4). We use the cell-surface lectin levels to decode the Golgi phenotype (Chapter 5).

Chapters 3, 4 and 5 have been adapted from the following preprint [Jos+24]:

Prachi Joshi, Aashish Satyajith, Debiprasad Panda, BR Rajeshwari, Mukund Thattai, Nagaraj Balasubramanian. Cell-surface glycans are quantitative reporters of Golgi dysfunction in single cells. *bioRxiv* 2024.06.17.599374; doi: [doi:10.1101/2024.06.17.599374](https://doi.org/10.1101/2024.06.17.599374)

Chapter 2

Pareto optimal glycan synthesis via enzyme compartmentalization

If one zooms out and looks at growth processes, one realizes that some of them follow the recipe below:

1. There is an entity that grows.
2. There is a topological ordering which the growth process must respect.
3. The entity is exposed to different environments that influence the rate of this growth, at different times.
4. The entity ceases to exist with a small probability.

For example, such behavior can be seen in ecology (growth of a microorganism [AWA98], speciation [OM10]), finance [Fu+05], disease progression [Cho+16], etc. We use glycan manufacture to study the optimal control parameters of such processes, particularly when competing interests are involved.

The stochastic nature of the glycan manufacturing process implies that desirable reactions occasionally do not happen. In addition to this, enzymes are capable of catalyzing reactions other than their "main" reaction, albeit at a much slower rate [TS00; BT20]. The production of a glycoprotein substrate is a complicated process, and both events can result in the loss of a painstakingly created glycoprotein substrate. This observation forms the crux of our argument: we posit that the compartment residence times and the enzyme distribution in the Golgi must be such that the least amount of the glycoprotein substrate goes to waste.

The glycoprotein substrate wastage is minimized if the substrate becomes any one of the "correct" glycans. Consider a given configuration of enzyme distribution and compartment residence times. If there exists another configuration that produces more of at least one glycan without decreasing the amounts of any other glycan, then it is preferable for the cell. This notion is best expressed in terms of Pareto optimality, which we discuss next.

2.1 Modeling

Definition 2.1.1 (Pareto optimality [ED18]). A vector $\mathbf{y}$ is said to *Pareto dominate* the vector $\mathbf{x}$ (denoted $\mathbf{x} < \mathbf{y}$) if and only if

$$\forall i \in \{1, \dots, m\} : \mathbf{x}_i \leq \mathbf{y}_i \text{ and } \exists j \in \{1, \dots, m\} : \mathbf{x}_j < \mathbf{y}_j$$

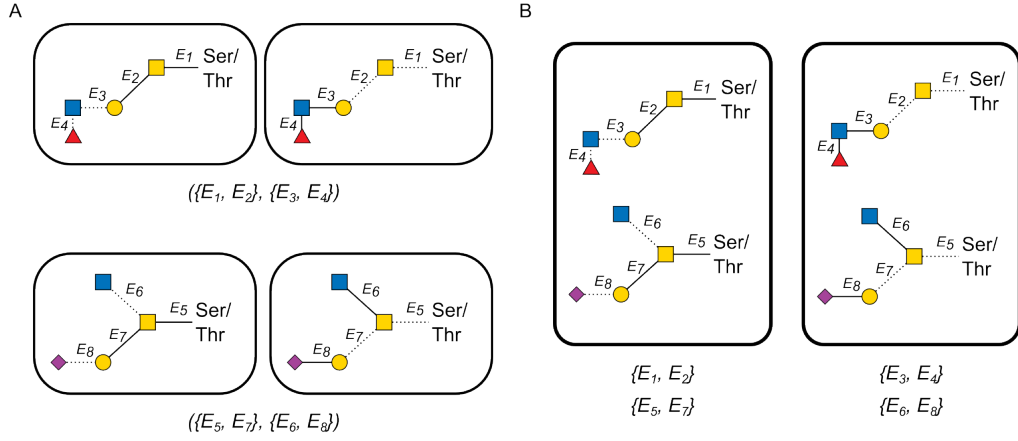


FIGURE 2.1: **(A)** Examples of splits and their (text) notation. The top two boxes in panel A denote a Golgi configuration with enzymes E_1 and E_2 in the first compartment, and E_3 and E_4 in the second compartment. The thick lines denote reactions that are to happen in that compartment, while dotted lines denote reactions that happen elsewhere. Another example of a split is given at the bottom. **(B)** Example of a Golgi configuration needed to produce the different glycans shown in 2.1(A). The ordered tuple of splits is $((\{E_1, E_2\}, \{E_3, E_4\}), (\{E_5, E_7\}, \{E_6, E_8\}))$. In the second compartment, two serial and two parallel reactions need to happen, which will have different optimal residence times, creating tradeoffs.

In our case, the real-valued vectors are the *yields* of the glycans of interest, which denote the fraction of the initial protein substrate that is correctly modified into that glycan. It can also be interpreted to mean the *chance* that the initial protein substrate is correctly modified into that glycan. Let's formalize glycan synthesis and define yields.

2.1.1 One glycan manufacture

We need to 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 $S_N(\mathcal{E}_B) = 2^{\mathcal{E}_B} \times \dots \times 2^{\mathcal{E}_B}$, 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 $(Z_1, \dots, Z_n)$ 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 as candidate splits. An example of enzyme splits is given in Fig. 2.1A.

Residence times. For an N -compartment Golgi apparatus, residence times are specified by a tuple in $\mathbb{R}_+^N$. The units of time will be made precise shortly.

A *strategy* is defined as a two-tuple $(\vec{\mathcal{S}}, \mathcal{T})$ where $\vec{\mathcal{S}}$ and $\mathcal{T}$ are elements of $S_N(\mathcal{E}_B)$ and $\mathbb{R}_+^N$ respectively, and denote the splits and the residence times assigned to each compartment.

We make several simplifying assumptions in our model of glycan manufacture:

- We adopt the cisternal maturation model; therefore the residence time in a compartment is the same for all glycoproteins that transit that compartment.
- The residence times of different compartments can be different.
- Glycans can only be grown by monomer addition reactions on the structure that contains the root.
- Glycosylation enzymes are not saturated; therefore the monomer addition reactions obey first-order kinetics [Hir+98].
- 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.
- Enzymes are promiscuous; they can catalyze reactions other than their primary reaction, albeit at a much smaller rate [BT20]. We denote q to be this smaller rate. If there are $|Z_j|$ enzymes in compartment j that manufacture a glycan, there are $|Z_j|$ ways for the undesirable reaction to manifest.
- 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.
- 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. 2.2B shows an example of two branches of the glycan in Fig. 2.2A (right) growing independently.
- Aberrant proteins are mostly dealt with by degradation, leading to loss of substrate [ST22].
- 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.
- The bulk of reactions leading to loss of substrate are reactions that do not run to completion [Var16]. We ignore other types of reactions in this work.

For the complete development of a branch, a monomer addition reaction first takes place (starting the creation of that branch) and is followed by all the other reactions needed to complete that branch. We use this observation to calculate yields in an inductive fashion. Let $a_1, \dots, a_K$ be the nodes attached to the root node, in a tree B that is growing in a compartment. Denote the subtree rooted at $a_1, \dots, a_K$ by $B_{a_1}, \dots, B_{a_K}$ respectively. Suppose the i -th branch is to be made within 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 t_1 can be anywhere from 0 to t , that is, $0 \leq t_1 \leq t$. The rest of the branch has time $t - t_1$ to develop, and thus the *yield of formation* of that branch, which denotes the chances of correct formation of that branch, can be written recursively as

$$\text{yield-formation}(B; t) = \int_0^t e^{-t_1} \left(\int_0^{t-t_1} \text{yield-formation}(B_i; t - t_1) dt_2 \right) dt_1$$

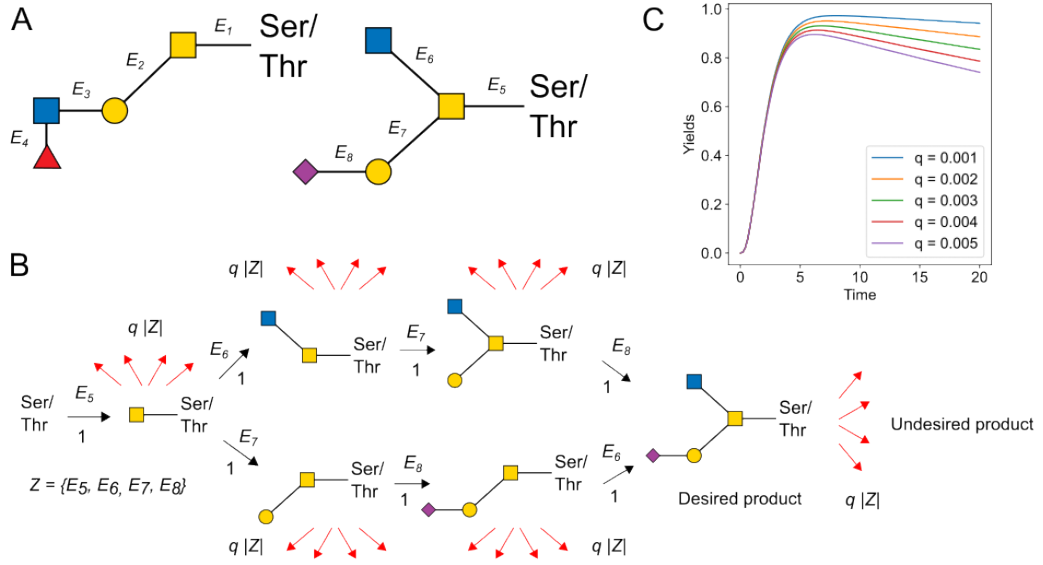


FIGURE 2.2: Glycan assembly, and yields. **(A)** The glycan to the left needs four reactions to happen one after another to form correctly. The edges are labeled by the enzymes needed to catalyze that reaction. The different shapes and colors denote sugar monomers [Nee+19]. In case of the glycan to the 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 red arrows denote an undesirable reaction taking place. The numbers near the arrows denote the rate at which the reactions happen. **(C)** Yields of glycan to the right of (A) at different value of q and residence times.

Since we assume the subtrees can form independently, the yield of formation of the entire glycan B is

$$\text{yield-formation}(B; t) = \prod_{i=1}^K \int_0^t e^{-t_1} \left(\int_0^{t-t_1} \text{yield-formation}(B_i; t-t_1) dt_2 \right) dt_1.$$

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 reactions in series (i.e. happening one after another; these types of reactions will be called as serial reactions). The yield of formation of required reactions can then be written down as

$$(1 - e^{-t}) \times \int_0^t e^{-t_1} \left(\int_0^{t-t_1} e^{-t_2} dt_2 \right) dt_1$$

The term $(1 - e^{-t}) = \int_0^t e^{-t} dt$ 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 note that for 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 is $1 - e^{-qt}$, therefore the chance of it *not* happening is e^{-qt} (t is the compartment residence time). If Z_j denotes the split of glycan in compartment j , then there are $|Z_j|$ ways for this promiscuous effect to manifest during the compartment residence

time t_j , and thus the probability that an undesirable reaction does not occur is $e^{-q \cdot t_j \cdot |Z_j|}$.

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

$$\underset{\substack{(Z_1, \dots, Z_N) \in S_N(B) \\ (t_1, \dots, t_N) \in \mathbb{R}_+^N}}{\text{maximize}} \quad \prod_{j=1}^N \text{yield-formation}(Z_j; t_j) \cdot e^{-q \cdot t_j \cdot |Z_j|} \quad (2.1)$$

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

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

A glycoprotein population can include many glycoforms that have varying functional activity [Stal1], all of which are produced by the same Golgi. This is the aspect we bring into our model next.

2.1.2 Multiple glycan manufacture

Consider that the Golgi produces K glycan trees $B_1, \dots, B_K$. Let $\mathcal{S}^1, \dots, \mathcal{S}^K$ denote the splits of the K glycans. We will write the ordered tuple of splits $(\mathcal{S}^1, \dots, \mathcal{S}^K)$ as $\vec{\mathcal{S}}$, and write $\vec{\mathcal{S}}^i$ to denote the split corresponding to the i^{th} glycan. Once we fix the splits for each of the K glycans and the residence times of each compartment, together denoted $(\vec{\mathcal{S}}, \mathcal{T})$, we will be able to obtain the yields $y_1, \dots, y_K$ for each of the glycans.

The varied structures of glycans imply different optimal residence times for their manufacture. Since the residence times of compartments are fixed, the process of multiple glycan manufacture brings in trade-offs. A short residence time might not give enough time for a complex glycan to run its reactions to completion. A long residence time might lead to undesirable reactions on a fully-grown glycan. Fig. 2.1B gives an example where such a scenario arises.

A second aspect of multiple glycan manufacture is that the cell will have its own fitness function $f(y_1, \dots, y_K)$ based on the yields $y_1, \dots, y_K$ of the K glycans, which we assume is what the cell is optimized for. Of course, it is impossible to know what this objective function f is.

As mentioned earlier, we assume that if the yield of a glycan can be improved without hurting the yield of other glycans, that is better for the cell. In other words, $f(y_1, \dots, y_K)$ is a monotonically increasing function of $y_i \forall i$, keeping $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_K$ fixed. With this assumption, we can use the notion of Pareto optimality to study the structure of the solution space of processes similar to glycan manufacture, which involve discrete and continuous parameters for optimization. We now define Pareto dominance of strategies.

2.2 Pareto optimality, convexity, and all that

Definition 2.2.1 (Weak Pareto dominance of strategies). Let $(u_1, \dots, u_K)$ and $(v_1, \dots, v_K)$ be the yield vectors corresponding to strategies $(\vec{\mathcal{S}}, \mathcal{T})$ and $(\vec{\mathcal{S}}', \mathcal{T}')$ respectively. $(\vec{\mathcal{S}}', \mathcal{T}')$ is said to *weakly Pareto dominate* $(\vec{\mathcal{S}}, \mathcal{T})$ with respect to the yield function $\text{yield}(\cdot)$, if $u_i \leq v_i, \forall i \in [K]$ and $\exists i | u_i < v_i$. We will denote this by $(\vec{\mathcal{S}}, \mathcal{T}) \preceq_{\text{yield}} (\vec{\mathcal{S}}', \mathcal{T}')$.

Informally, a strategy weakly Pareto dominates another strategy if all glycans that are being synthesized have the same or better yields using the former strategy than the latter [Mie98]. 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 $f(\cdot)$ is monotonically increasing, we can restrict ourselves to only considering the set of Pareto optimal solutions for further study.

How can we generate Pareto optimal solutions for this objective? In our case, we got lucky: the yield functions are from a class of functions called concave functions whose behavior we know well. The additive inverse of the yield functions is then a convex function. As it turns out, there is a simple recipe to generate Pareto-optimal solutions for convex functions. We outlay all this next.

Definition 2.2.2 (Convexity). A function $f(x)$ is said to be convex if $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$ for $0 \leq \lambda \leq 1; x_1, x_2 \in \text{Dom}(f)$.

Geometrically, this means that if you pick two points on the abscissa, the function value between them must lie below the line joining the corresponding ordinate points.

In the same vein, one can define concave functions:

Definition 2.2.3 (Concavity). A function $f(x)$ is said to be concave if $f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda f(x_1) + (1 - \lambda)f(x_2)$ for $0 \leq \lambda \leq 1; x_1, x_2 \in \text{Dom}(f)$.

Definition 2.2.4 (Log concavity). A strictly positive function $f(x)$ is said to be log concave if the logarithm of the function is concave, i.e. if $\log f(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda \log f(x_1) + (1 - \lambda) \log f(x_2)$ for $0 \leq \lambda \leq 1; x_1, x_2 \in \text{Dom}(f)$.

This can be extended to a scalar function involving many variables:

Definition 2.2.5 (Pointwise log concavity). A function $f(x_1, \dots, x_N)$ is said to be pointwise log-concave if for all fixed $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N$, the function is log-concave in $x_i, \forall i$.

Now let's connect our yield function and convexity:

Theorem 2.2.1. Given a fixed split, the yield objective function is log-concave in residence times.

Proof. 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 log-concave in the residence time for that compartment. Since the product of log-concave functions is log-concave, the yield function is log-concave. We prove log-concavity of yield in a single compartment by induction.

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

Inductive case: Let us assume $\text{yield-formation}(B_i; t)$ is log-concave. The yield of formation of the branch containing B_i is calculated as

$$\int_0^t e^{-t_1} \left(\int_0^{t-t_1} \text{yield-formation}(B_i; t_2) dt_2 \right) dt_1$$

Write $\int_0^\tau \text{yield-formation}(B_i; t_2) dt_2$ as $Y(\tau)$. The above expression then represents the convolution of two log-concave functions $F(\tau) = e^{-\tau}$ and $Y(\tau)$, and is therefore log-concave by [BV11]. The term corresponding to detrimental reactions not happening, $e^{-(1+|B_i|) \cdot q \cdot t}$, is also log-concave in t . The product of two log-concave functions is log-concave, and thus, the yield function is log-concave. $\square$

2.2.1 Connecting Pareto optimality and concavity

Definition 2.2.6 (Convex combination of functions). We call $\sum_i w_i f_i(\mathbf{x})$ such that $w_i \geq 0$ and $\sum_i w_i = 1$ a convex combination of the f_i 's.

Pareto-optimal solutions of multi-objective concave functions turn out to be maximizers of convex combinations of the individual concave functions. Since the proof is scattered across literature, we compile it here. We will prove lemmas and theorems in terms of convex functions, keeping in mind that the negative of a convex function is concave. To restate the first line in terms of convex functions, Pareto-optimal solutions of multi-objective convex functions are *minimizers* of convex combinations of the individual convex functions. Note that when results are expressed in terms of convexity, the inequality sign is reversed in our Definition 2.1.1 of Pareto optimality.

Before we can prove the statement, we will need some reformulations and lemmas.

Let $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^K$ be a multi-objective function, $\mathbf{f} = (f_1(\mathbf{x}), \dots, f_K(\mathbf{x}))$ and let $S = \text{Dom}(\mathbf{f})$. Consider the following weighting problem:

$$\text{minimize } \sum_i w_i f_i(\mathbf{x}) \quad \text{such that } w_i \geq 0 \text{ and } \sum_i w_i = 1.$$

Now we can formally state the theorem mentioned at the beginning of the section:

Theorem 2.2.2 (Pareto optimal solutions are minimizers of convex combination of convex functions). Let f_i be convex in $\mathbf{x} \forall i$. Then $\mathbf{x}^*$ is a Pareto optimal solution $\iff$ there exists a weight vector $\mathbf{w}$, $w_i \geq 0; \sum_i w_i = 1$ such that $\mathbf{x}^*$ is a solution of the weighting problem.

$\Leftarrow$ is easily proven by contradiction. To prove $\Rightarrow$, let's consider another formulation of the (convex) minimization problem. An ε -constraint problem (ε -CP) is a problem of the form:

$$\text{minimize } f_l(\mathbf{x}) \quad \text{subject to } f_j(\mathbf{x}) \leq \varepsilon_j, \forall j, j \neq l.$$

where $l \in [K]$.

Lemma 2.2.3 (Pareto-optimal solutions are solutions to the ε -CP [Mie98]). A decision vector $\mathbf{x}^* \in S$ is Pareto optimal $\iff \mathbf{x}^*$ is a solution to the ε -constraint problem with $\varepsilon_j = f_j(\mathbf{x}^*) \forall l \in [k]$.

Proof. $\Rightarrow$: Let a solution $\mathbf{x}^*$ to the ε -CP not be Pareto-optimal. This means $\exists \mathbf{x} \in S | f_l(\mathbf{x}) < f_l(\mathbf{x}^*)$ and $f_j(\mathbf{x}) \leq \varepsilon_j = f_j(\mathbf{x}^*) \forall j \neq l$. This contradicts the Pareto optimality of $\mathbf{x}^*$.

$\Leftarrow$: Since $\mathbf{x}^*$ is a solution of ε -CP $\forall l$, this means there is no $\mathbf{x} \in S$ such that $f_l(\mathbf{x}) < f_l(\mathbf{x}^*)$ and $f_j(\mathbf{x}) \leq f_j(\mathbf{x}^*)$. This is the definition of Pareto optimality, and $\mathbf{x}^*$ is Pareto optimal. $\square$

Note that the above lemma holds regardless of convexity. But what if the objectives were convex?

Theorem 2.2.4. Let the multi-objective optimization problem be convex. If $\mathbf{x}^*$ is a solution to ε -CP for any given f_l to be minimized and $\varepsilon_j = f_j(\mathbf{x}^*) \forall j \neq l$, then there exists a weighing vector $\mathbf{0} \leq \mathbf{w} \in \mathbb{R}^K, \sum_i w_i = 1$ such that $\mathbf{x}^*$ is also a solution of the weighting problem.

To prove this, we need the separating hyperplane theorem:

Theorem 2.2.5 (Separating hyperplane theorem). A subset Ω of a vector space over $\mathbb{R}$ is a convex set if $\forall x, y \in \Omega, \lambda x + (1 - \lambda)y \in \Omega, 0 \leq \lambda \leq 1$. Let Ω be a convex set in $\mathbb{R}^n$ that does not contain the origin $\mathbf{0}$. Then there exists a hyperplane $\{\mathbf{c} | \forall \mathbf{x} \in \Omega, \mathbf{c}^T \mathbf{x} = 0\}, \mathbf{c} \neq \mathbf{0}$ separating Ω and $\mathbf{0}$, i.e. $\mathbf{x} \in \Omega \implies \mathbf{c}^T \mathbf{x} \geq 0$.

Proof. The theorem is quite well known in convex optimization literature; see [Man94; BV11]. $\square$

We also need the following theorem:

Theorem 2.2.6 (Generalized Gordon Theorem [Man94]). Let $\mathbf{f}(\mathbf{x}) = \{f_1(\mathbf{x}), \dots, f_K(\mathbf{x})\}$ be a K -dimensional convex function on the convex set $\Gamma \subseteq \mathbb{R}^N$. Then either:

1. $\mathbf{f}(\mathbf{x}) < \mathbf{0}$ has a solution $\mathbf{x} \in \Gamma$, or
2. $\exists \mathbf{p} \geq \mathbf{0}, \mathbf{p} \in \mathbb{R}^N | \mathbf{p}^T \mathbf{f}(\mathbf{x}) \geq 0 \forall \mathbf{x} \in \Gamma$

but not both.

Proof. (1) $\implies \neg(2)$: Assume $\exists \mathbf{x} \in \Gamma | \mathbf{f}(\mathbf{x}) < \mathbf{0}$. Then $\forall \mathbf{p} \geq \mathbf{0}, \mathbf{p} \in \mathbb{R}^M, \mathbf{p}^T \mathbf{f}(\mathbf{x}) < \mathbf{0}$ and hence (2) cannot hold.

$\neg(1) \implies (2)$: Define $G(\mathbf{x}) = \{\mathbf{y} | \mathbf{y} \in \mathbb{R}^M, \mathbf{f}(\mathbf{x}) < \mathbf{y}\}$ for $\mathbf{x} \in \Gamma$. Let $G = \bigcup_{\mathbf{x} \in \Gamma} G(\mathbf{x})$. Now, by hypothesis, $\mathbf{f}(\mathbf{x}) < \mathbf{0}$ has no solution, $\therefore \mathbf{0} \notin G$.

Claim 1: G is convex.

Proof: Let $\mathbf{y}_1, \mathbf{y}_2 \in G$. Then $\mathbf{y}_1 > \mathbf{f}(\mathbf{x}_1), \mathbf{y}_2 > \mathbf{f}(\mathbf{x}_2)$ for some $\mathbf{x}_1, \mathbf{x}_2 \in \Gamma$.

And so $\lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 > \lambda \mathbf{f}(\mathbf{x}_1) + (1 - \lambda) \mathbf{f}(\mathbf{x}_2) \geq \mathbf{f}(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2)$ by convexity of $\mathbf{f}$, for some $\lambda \in [0, 1]$.

$\therefore \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \in G$ and G is convex.

Now, $\mathbf{0} \notin G$ and G is convex. Therefore by the separating hyperplane theorem (Theorem 2.2.5), $\exists \mathbf{p} \in \mathbb{R}^M | \mathbf{p}^T \mathbf{u} \geq 0, \mathbf{u} \in G$. Since each u_i can be made arbitrarily large, it follows that $\mathbf{p} \geq \mathbf{0}$.

Claim 2: $\mathbf{p}^T \mathbf{f}(\mathbf{x}) \geq 0 \forall \mathbf{x} \in \Gamma$.

Proof: Assume otherwise. Then

$$\inf_{\mathbf{x} \in \Gamma} \mathbf{p}^T \mathbf{f}(\mathbf{x}) = -\delta \text{ for some } \delta \in \mathbb{R}_+$$

Let $\varepsilon > 0$ and set $\mathbf{u} = \mathbf{f}(\mathbf{x}) + \mathbf{e}\varepsilon$ where $\mathbf{e}$ is a vector of 1's in $\mathbb{R}^M$, $\mathbf{x} \in \Gamma$.

$\therefore \mathbf{u} \in G(\mathbf{x})$ and $\mathbf{u} \in G$.

$\therefore \mathbf{p}^T \mathbf{u} \geq 0 \implies \mathbf{p}^T \mathbf{f}(\mathbf{x}) + \mathbf{p}^T \mathbf{e}\varepsilon \geq 0 \implies \mathbf{p}^T \mathbf{f}(\mathbf{x}) \geq -\mathbf{p}^T \mathbf{e}\varepsilon$.

Set ε small enough so that $\mathbf{p}^T \mathbf{e}\varepsilon < \delta$.

$\therefore -\mathbf{p}^T \mathbf{e}\varepsilon > -\delta$ and $\mathbf{p}^T \mathbf{f}(\mathbf{x}) > -\delta$ for $\mathbf{x} \in \Gamma$, contradicting that the infimum is δ .

$\therefore \mathbf{p}^T \mathbf{f}(\mathbf{x}) \geq 0$. □

Now we can finish the proof of Theorem 2.2.2 that Pareto optimal solutions are minimizers of convex combinations of convex functions.

Proof of Theorem 2.2.2. It is given that each f_i is convex and that the feasible region is convex. Since $\mathbf{x}^*$ is a solution to the ε -CP, the system of inequalities

$$f_k(\mathbf{x}) < f_k(\mathbf{x}^*), f_j(\mathbf{x}) \leq f_j(\mathbf{x}^*) = \varepsilon_j, j = 1, \dots, N, j \neq k$$

has no solution. Applying Generalized Gordon Theorem (Theorem 2.2.6) with $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}) - f_1(\mathbf{x}^*), \dots, f_N(\mathbf{x}) - f_N(\mathbf{x}^*))$, $\exists \mathbf{p} \in \mathbb{R}^N$, $\mathbf{p} \geq 0$ such that

$$\sum_j p_j (f_j(\mathbf{x}) - f_j(\mathbf{x}^*)) \geq 0 \quad \forall \mathbf{x} \in S \text{ (the feasible region)}.$$

Set

$$w_j = \frac{p_j}{\sum_j p_j}$$

Thus $w_j \geq 0 \quad \forall j$ and $\sum_j w_j = 1$.

Now $\forall \mathbf{x} \in S$,

$$\sum_j w_j (f_j(\mathbf{x}) - f_j(\mathbf{x}^*)) \geq 0 \text{ or}$$

$$\sum_j w_j (f_j(\mathbf{x})) \geq \sum_j w_j (f_j(\mathbf{x}^*))$$

i.e. $\mathbf{x}^*$ is also a solution to the weighting problem for a particular combination of weights. □

To summarize, if $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$ are convex functions of $\mathbf{x}$ (such as additive inverse of the log yields), then the set below is the Pareto set:

$$\left\{ \arg \min_{\mathbf{x}} c \cdot y_1(\mathbf{x}) + (1 - c) \cdot y_2(\mathbf{x}); 0 \leq c \leq 1 \right\}$$

2.3 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.

2.3.1 One glycan synthesis

There is an optimal residence time per compartment.

Suppose we grow the glycan to the right in Fig. 2.2A, in a single compartment. Fig. 2.2C 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, as we proved earlier, is log-concave with respect to residence times.

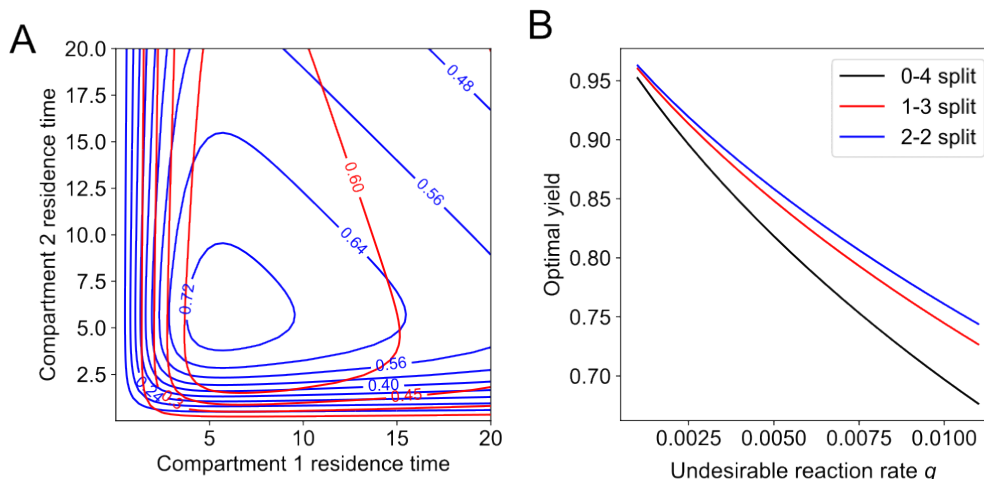


FIGURE 2.3: Different splits give different yields. (A) Values of yields w. r. t. residence times in two compartments for different splits of four serial reactions. The blue lines denote contours for a 2-2 split (first two reactions in the first compartment, next two in the second compartment) while 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.

Some enzyme splits are better than others.

We demonstrate this for a glycan that needs four serial reactions to occur for its correct synthesis.

Fig. 2.3A 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. One can immediately see that there are optimal residence times for the first and second compartments, depending on the split. The yield function is anisotropic around the peak; deviations from optimum have direction-dependent effects.

Fig. 2.3B shows the optimal yield of the same glycan under all possible splits, for various values of q . In this case, the symmetric split gives a better yield than any other split.

2.3.2 Multiple glycan synthesis

Pareto-optimal solutions for two glycan manufacture

We consider the Pareto set of trees 1 and 2 shown in Fig. 2.4A. Candidate solutions can be obtained by fixing the split of each tree and finding the optimal residence times for a range of convex weightings of the yields of both glycans ([ED18]). Some of the candidate solutions generated will be Pareto-optimal. Fig. 2.4B sketches the yields of trees 1 and 2 under optimal residence times corresponding to a range of convex weightings of their yields. The blue and red colors denote the different splits of tree 2 that are used when calculating the yields (tree 1 has one optimal split). One can see that some segments of the blue curve Pareto-dominate segments of the red curve, and vice versa [Lev62]. The crossover point represents the point at which the (Pareto) optimal split of tree 2 changes. To obtain a Pareto optimal solution with a higher yield of glycan 2 than at crossover point, the blue split must be used. If the yield of glycan 1 is to be slightly higher than at crossover point, the red split must be used. The Pareto set is sketched in Fig. 2.4C.

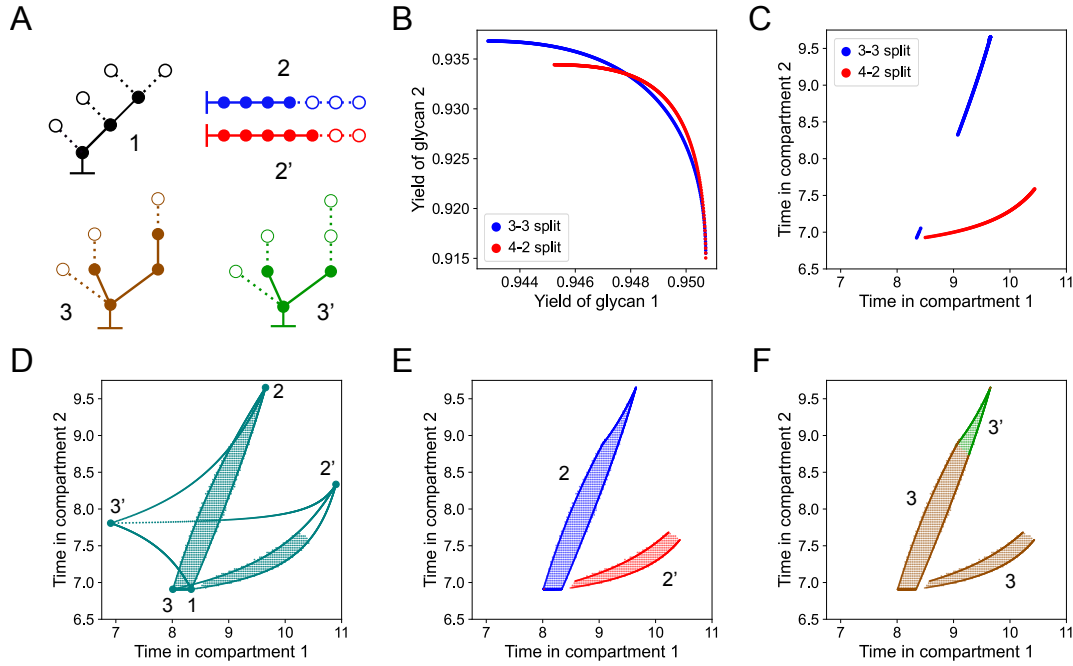


FIGURE 2.4: Pareto-optimal splits and residence times for multiple glycan manufacture. **(A)** Pareto-optimal splits of the three trees. Numbers adjacent to the structures and their prime versions (e.g. 2 and 2') represent different splits of the same tree. The nodes with filled-in colors (along with 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 same colors are used to indicate splits in subsequent plots. **(B)** Yields of trees 1 and 2 obtained by maximizing different convex weightings of trees (Appendix B), using split 2 and 2' (blue and red curves respectively). Tree 1 has only one optimal split. For this and all subsequent plots, $q = 0.001$. **(C)** The Pareto set corresponding to (B). Colors indicate the splits that generated those solutions. **(D)** The labeled points denote optimal residence times for each split. Each thick 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 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).

Pareto-optimal solutions for three or more glycan manufacture

We take examples of three glycans with differing structures and study their Pareto set. Since generating Pareto-optimal solutions using the convex weighting method described above 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 of the trees. This was repeated for a grid of (t_1, t_2) values, and the maximum yield of each tree was stored. Finally, we chose the yield tuples which were not Pareto-dominated.

Fig. 2.4D shows the Pareto set for the three glycans in Fig. 2.4A; 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. 2.4D, blue lines). Figs. 2.4E, F show the optimal splits of glycans 2 and 3 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 when we considered manufacturing two glycans at a time (Fig. 2.4B).

2.4 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 several areas, as mentioned before. 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.

We have considered an abstract formulation of glycan manufacture, valid for both N- and O-glycans. The feature of enzyme splits adds an extra layer of complexity to finding the Pareto-optimal solutions, as the splits have to be enumerated as well. The Pareto set comprises disconnected regions of the continuous parameter space (residence times; Fig. 2.4D). We conjecture that the Pareto-optimal solutions for three glycan manufacture will be contained in the boundaries obtained by assuming that only two splits are allowed. Notably, small changes in the residence times can lead to different optimal splits (Fig. 2.4F); this behaviour is analogous to first-order phase transitions reported in other Pareto-optimized systems [SS15]. The observation also indicates that yields of different splits are comparable. However, even modest increases in yields may confer a survival advantage as glycan biosynthesis is scaled across a large number of cells.

A Pareto optimality-based perspective has another advantage: we can make general statements about the enzyme distributions and residence times without invoking a specific objective function [Yad+22]. We make only the reasonable assumption that the objective function the cell optimizes is monotonically increasing in the yields of the glycans. Thus, in our framework, reconciling the intrinsic trade-offs of glycan biosynthesis corresponds to operating on a Pareto front; we therefore propose that the Golgi might balance

competing glycan outputs by operating near such Pareto-optimal compromises.

Our model makes three predictions that can be probed *in vitro* using immobilized enzyme cascades [Mak+24] or droplet-based microfluidic platforms [Ise+25], originally meant to achieve targeted glycosylation. In these systems, each reactor module or incubation chamber can be viewed as a Golgi compartment, with reaction or chamber time representing residence time and the allocation of enzymes in early and late modules representing an effective enzyme split. First, for a fixed enzyme split, the yield of a given glycan is predicted to be unimodal as a function of compartment residence times (Fig. 2.2C). This can be tested by sampling a range of effective residence times – for example by changing reaction time in immobilized cascades or adjusting flow rates and the number of incubation chambers in microfluidics – while keeping other factors such as enzyme composition, substrate and donor supply, temperature etc. fixed. The yields at a fixed residence time can be measured using established glycomics workflows [Mak+24; Ise+25]. Second, when multiple glycans are produced simultaneously, the model predicts trade-offs in their yields under different residence times (Fig. 2.4B). The absence of a non-trivial front (for example, a single configuration that dominates all others) would falsify this prediction. Third, within the Pareto front, small changes in residence times are predicted to cause switches in the enzyme split that optimizes the joint yields (Fig. 2.4F). Our kinetic framework can be applied to the enzyme sets used in these platforms to identify a small number of candidate enzyme splits whose predicted yields are similar and have a crossover (Fig. 2.4B). The residence times can be varied in a small neighborhood around the model-predicted crossover. Observing a change in the Pareto-optimal split between these two conditions would support the split-switching behavior illustrated in Fig. 2.4F. Performing analogous tests in biological Golgi stacks is more challenging because compartment-specific residence times remain difficult to measure *in vivo*.

We have only considered partitions of enzymes as candidate splits for tractability. Redundant localization of enzymes will increase the rate of undesirable reactions and its net effect on yield will be context-dependent; a systematic exploration of this possibility is left for subsequent studies. The trade-offs presented here are contingent on the cisternal maturation model of Golgi traffic; although the model is widely believed to be true [MT16] there is still lack of consensus [GL11]. We have assumed for simplicity that enzyme splits can be set independently for each glycan, when in fact enzymes executing desirable reactions can be shared across different glycans [JT20; BT20]. This aspect becomes relevant while considering other models of Golgi trafficking, where cargo can have different residence times in the Golgi [Tie+25]. The presence of common enzymes involved in the manufacture of different glycans will again create trade-offs. In summary, regardless of the underlying model of Golgi trafficking used, some trade-offs must arise. Our model can be adapted to study the trade-offs under these scenarios, and will be explored in a future work. This will also enable direct comparisons to existing glycan distribution and residence time data [Tie+25].

Critical quality attributes have glycan fractions to evaluate a biological product's performance or for testing biosimilars [YI21]. Given a glycosylation reaction network and the appropriate parameter values, yields for specific glycans of interest can be computed [FKK25]. This can be used to generate Pareto-optimal solutions for the glycosylation reaction network under the required constraints. Studying the Pareto set and Pareto front will provide insights into the sensitivity of yields to parameter changes (Fig. 2.3A), parameter regions where modest adjustments can cause non-trivial changes in optimizer (Fig. 2.4F), and so on. A robust solution can then be chosen from the Pareto set. Thus, a Pareto

optimality-based framework for glycan manufacture has broad applicability, ranging from addressing fundamental questions about how cells resolve biosynthetic trade-offs to guiding the design of efficient glycosylation processes for next-generation biotherapeutics.

Chapter 3

Dynamics inside the cell

In the previous chapter, we looked at how the enzymes in the Golgi gets distributed to produce glycans efficiently. In this chapter, we look at how cis-Golgi numbers change, with the idea of connecting the cis-Golgi numbers to cell-surface glycan distribution in the next chapter. We also study another cellular process that can change the distribution of cell-surface glycans – endocytosis. We will understand how the dynamics of both processes change when BFA is added.

3.1 cis-Golgi dynamics

Given that we are interested in studying Golgi *dysfunction*, it is necessary to have a model system that can be used to generate a gradient of dysfunctional Golgi in a controlled manner. As mentioned before, BFA is used for this purpose, and we employ the system developed in [SEB18]. We also use another inhibitor, Golgicide A (GCA), which generates a gradient of Golgi phenotypes [Raj+23] to cross-validate our model and data. GCA affects only Arf GEF GBF1 [São+09], unlike BFA, which affects GBF1 and BIG1/2.

To generate a gradient of phenotypes, we treat suspended MEF cells with different concentrations of BFA and GCA, sample the phenotypes obtained, and image them using confocal microscopy.

3.1.1 Experimental procedure

The experimental procedure for suspension of cells in BFA or GCA is common to all experiments and is given in Table 3.1. These steps are performed after standardization experiments that determine the correct concentration of lectins, amount of BFA to use, etc. The data for GCA comes from [Raj+23], which has an additional 60 minutes of suspension time after Step 4.

Once cells were suspended for the required amount of time, they were fixed. The equatorial plane of cells was shot on a Zeiss confocal microscope (LSM710), as identified by the largest perimeter. A minimum of 100 cells and a maximum of 200 cells were counted and manually classified according to their Golgi organization phenotype: (i) organized Golgi, compact; (ii) disorganized Golgi, opened up; (iii) partially fragmented, opened up with some fallback to the ER; (iv) completely fragmented, complete fallback to the ER [Raj+23]. Examples of cells with these phenotypes are shown in Fig. 3.1A. The distribution of these phenotypes at different BFA concentrations and times after treatment is shown in Fig. 3.1B.

<i>Step</i>	<i>Procedure</i>	<i>Rationale</i>
1	Cells were serum-deprived using low serum DMEM (0.2% fetal bovine serum), for 14.5 hours.	Serum components might sustain signaling pathways that are otherwise controlled by adhesion; serum deprivation prevents this from happening. The entire experiment is performed under low-serum conditions.
2	Cells were detached using Trypsin or Accutase and washed with low-serum medium.	Trypsin or Accutase dissociates adherent cells from their growth surface.
3	Cells were counted and distributed equally between all time points. One aliquot was kept aside to be processed as the 5 min time point (if required).	The cells in the kept-aside aliquot are used as control and also for gauging autofluorescence (see Table 4.1).
4	Cells were gently mixed with low-serum medium containing 1% methylcellulose.	Methylcellulose makes the solution more viscous. This prevents cells from settling down and adhering to each other, resulting in a single-cell suspension.
5	Different concentrations of BFA or GCA (DMSO for control) were added to the suspensions and mixed gently.	To perturb the Golgi; gentle mixing prevents formation of clumps which can induce cell-cell signaling.
6	Cells were incubated at 37°C with 5% CO ₂ for the required times.	To maintain ideal cell-culture conditions and to track time-specific effects of BFA.
7	The suspension was diluted and washed with cold, low-serum medium.	To wash away methylcellulose, as otherwise it might form aggregates between cells and interfere with read-outs. Cold temperature slows down metabolic activity and causes minimal change to the physiological state of the cell at a given time point.

TABLE 3.1: Experimental procedure for generating a gradient of Golgi phenotypes.

3.1.2 Birth-death processes

To understand the cis-Golgi dynamics that gives rise to these phenotypes, we developed a stochastic birth-death model for Golgi biosynthesis and fragmentation (Fig. 3.2B). We assume that the cis-Golgi is made up of discrete units, each of which is created by material from the ER [MO14; Har+24]. These units can fuse back to the ER at a rate proportional to their v-SNARE concentration (recall that v-SNAREs help vesicles dock on to their target

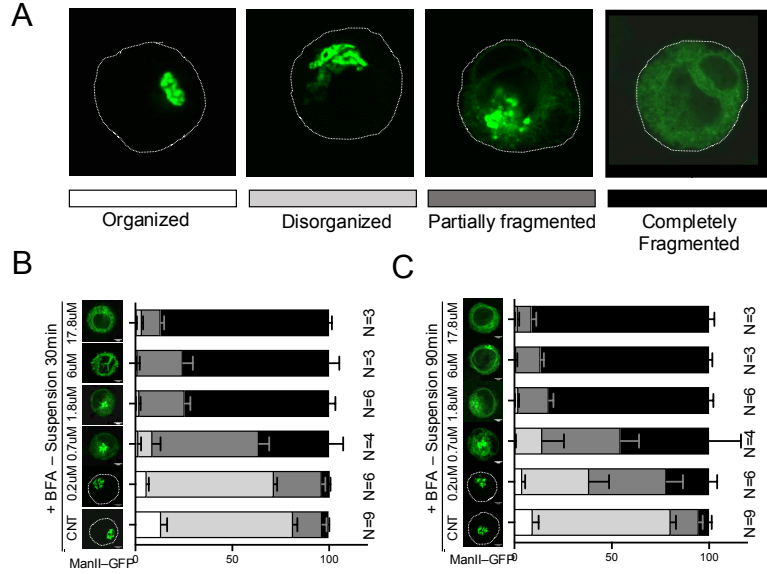


FIGURE 3.1: BFA action gives a gradient of Golgi phenotypes. **(A)** Representative images showing the four *cis*-Golgi morphology classes: organized, disorganized, partially fragmented, and completely fragmented. Green colour shows the localization of ManII-GFP. The completely fragmented class corresponds to the fallback of the entire *cis*-Golgi into the ER. **(B,C)** Distribution of *cis*-Golgi morphology in cell populations ($n = 100$) incubated for 30 min (B) or 90 min (C) with varying concentrations of BFA.

membrane). We assume the per-membrane-area concentration of these v-SNAREs on the *cis*-Golgi obeys

$$\frac{dS(t; c)}{dt} = \theta_{\text{COPII}} - \frac{\theta_{\text{COPI}}}{(1 + c/c_0)} S(t; c) \quad (3.1)$$

where the first term corresponds to the arrival of v-SNAREs from the ER on COPII vesicles and the second to their removal from the *cis*-Golgi by COPI vesicles. The rate of COPI vesicle formation is inhibited by increasing BFA or GCA concentrations (c). The hyperbolic dependence on concentration is commonly used to model inhibitory effects of a drug, and is reminiscent of the fraction of unbound receptor concentration on drug action described in Section 1.2.3. c_0 is the inhibitor concentration at which the Golgi-to-ER COPI flux reaches its half-maximal level. In reality there is a small dependence of c on θ_{COPII} as well since COPII vesiculation relies on COPI-dependent recycling of the components [NRR02]. However, we choose to ignore this dependence.

We assume v-SNAREs rapidly equilibrate, reaching a quasi-steady-state level $S(c) = (\theta_{\text{COPII}}/\theta_{\text{COPI}}) (1 + c/c_0)$. The units of *cis*-Golgi are created by material from the ER, and they can fuse back to the ER at a rate proportional to the v-SNARE concentration. The expected number of *cis*-Golgi units $n_{\text{cis}}(t; c)$ at time t and inhibitor concentration c is therefore obtained by solving:

$$\begin{aligned} \frac{dn_{\text{cis}}(t; c)}{dt} &= \phi_+ - \phi_- S(c) n_{\text{cis}}(t; c) \\ &= \phi_+ - \phi_- \frac{\theta_{\text{COPII}}}{\theta_{\text{COPI}}} \left(1 + \frac{c}{c_0}\right) n_{\text{cis}}(t; c) \end{aligned} \quad (3.2)$$

where ϕ_+ and $\phi_- S(c)$ denote the rate of cis-Golgi creation and rate constant of cis-Golgi fallback (loss of cis-Golgi) respectively. The solution to this equation is given by

$$\begin{aligned} n_{\text{cis}}(t; c) &= \frac{\phi_+}{\phi_-} \cdot \frac{\theta_{\text{COPI}}}{\theta_{\text{COPII}}(1 + c/c_0)} + \left(\frac{\phi_+}{\phi_-} \cdot \frac{\theta_{\text{COPI}}}{\theta_{\text{COPII}}} - \frac{\phi_+}{\phi_-} \cdot \frac{\theta_{\text{COPI}}}{\theta_{\text{COPII}}(1 + c/c_0)} \right) e^{-\phi_- \frac{\theta_{\text{COPII}}}{\theta_{\text{COPI}}} \left(1 + \frac{c}{c_0}\right) t} \\ &= \frac{n_0}{1 + c/c_0} + \left(n_0 - \frac{n_0}{1 + c/c_0} \right) e^{-(1 + c/c_0) (t/t_0)} = n_0 \cdot \frac{c_0 + c \cdot e^{-(1 + c/c_0) (t/t_0)}}{c_0 + c} \end{aligned} \quad (3.3)$$

with parameters $t_0 = (\theta_{\text{COPI}}/\theta_{\text{COPII}})/\phi_-$, and $n_0 = \phi_+ t_0$. n_0 is the average number of cis-Golgi units in steady state in the absence of the inhibitor; c_0 is the inhibitor concentration at which the Golgi-to-ER COPI flux reaches its half-maximal level; and t_0 is the characteristic timescale on which cis-Golgi levels change. Since birth and death are stochastic, there will be fluctuations around this expected value. It is well known that the steady-state counts of entities (such as the number of cis-Golgi units) under a stochastic birth-death process with a constant rate of creation and a rate of destruction proportional to the instantaneous counts follows a Poisson distribution. Interestingly, the counts of entities during the *transient* phase also follows a Poisson distribution with a time-dependent mean [Tha16]; for a proof, see [Kal94]. In our case, the mean number of cis-Golgi units at time t and BFA concentration c is Poisson distributed with mean given by the $n_{\text{cis}}(t; c)$ above. Birth-death models are also applicable for other organelles like vacuoles and peroxisomes [MO14].

We used the the data in Fig. 3.1 to fit the parameters of this model. To do this, we needed to link n_{cis} to the different Golgi phenotypes (completely fragmented, partially fragmented, etc.). The solution to Equation 3.2 dictates a (discrete) Poisson distribution; how does one connect this Poisson distribution to the distribution of phenotypes?

We imagined the number of cis-Golgi units n_{cis} to take on real values, as permitted by using a continuous approximation to the Poisson distribution [Ili13]. Then, three n_{cis} thresholds are used to divide the space of (real-numbered) n_{cis} values into four phenotypes. For instance, if the n_{cis} threshold separating the completely fragmented phenotype from the other phenotypes is 0.6, then any cell with $n_{\text{cis}} \leq 0.6$ corresponds to a completely fragmented phenotype (note that the more extreme the phenotype, the lesser the n_{cis} value). We fit the two thresholds of completely fragmented and partially fragmented phenotypes (data not shown). The values obtained suggested that the completely fragmented and partially fragmented phenotypes correspond to $n_{\text{cis}} = 0$ and 1 respectively (Fig. 3.2C). We took $n_{\text{cis}} > 1$ to correspond to the disorganized phenotype.

3.1.3 Results

Since n_{cis} is Poisson distributed, the fraction of cells with $n_{\text{cis}} = 0$ is $e^{-n_{\text{cis}}(t; c)}$. We fit the parameters of equation to the observed fraction of cells with completely fragmented cis-Golgi (Fig. 3.2, dark purple). Remarkably, the model precisely predicts the observed fraction of cells with partially fragmented or disorganized cis-Golgi at all inhibitor concentrations and times, with no additional fitting parameters (Fig. 3.2D-G, orange and light purple). The fitted values of c_0^{BFA} and n_0 are comparable across all treatments, further demonstrating the robustness of our model (Table 3.2).

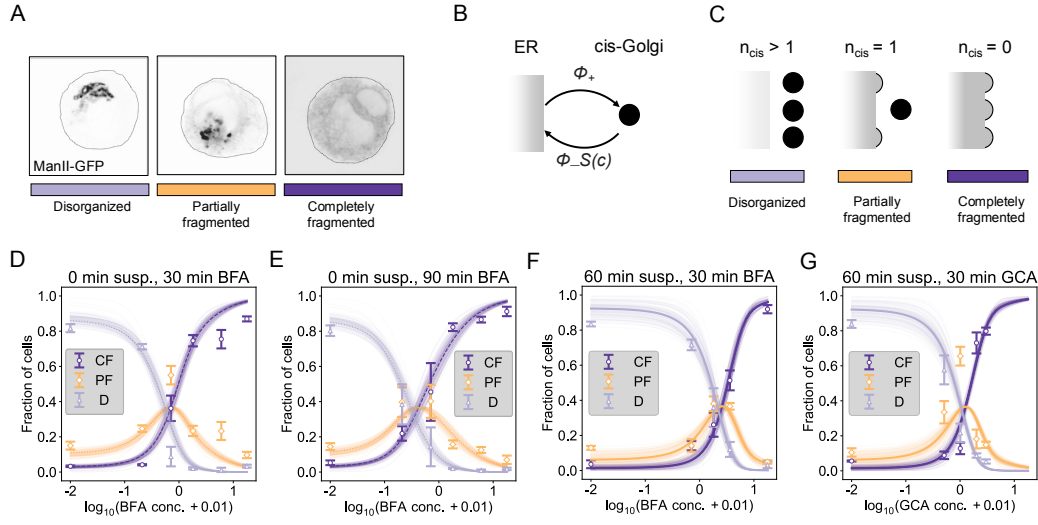


FIGURE 3.2: cis-Golgi fragmentation by BFA and GCA. **(A)** We monitor the cis-Golgi in non-adherent WT-MEF cells using a fluorescently labeled cis-medial marker (ManII-GFP). Representative inverse grayscale images of fluorescence intensity show the three cis-Golgi morphology classes: disorganized (light purple), partially fragmented (orange), and completely fragmented (dark purple). (We merged the organized class with the disorganized class, as was done in [Raj+23].) The completely fragmented class corresponds to the fallback of the entire cis-Golgi into the ER; the background appears gray because the fluorescence level is low, and the image has been scaled by maximum intensity. **(B)** Birth-death model of Golgi biogenesis. We assume the cis-Golgi is made up of discrete units (black circles) which are created by material from the ER with rate ϕ_+ , and fuse back into the ER with rate constant $\phi_-S(c)$. $S(c)$ is the per-area concentration of COPI-specific v-SNAREs on the cis-Golgi. **(C)** We posit that cells with $n_{\text{cis}} = 0$, $n_{\text{cis}} = 1$ and $n_{\text{cis}} > 1$ cis-Golgi units correspond to the completely fragmented, partially fragmented and disorganized phenotypes. Fusion of the cis-Golgi with the ER (bumps) leads to the transfer of the cis-Golgi marker to the ER (gray). **(D-G)** The fraction of cells in a population ($n \geq 100$) with cis-Golgi that are disorganized (light purple), partially fragmented (orange) or completely fragmented (dark purple), for varying inhibitor concentrations (x-axes) and varying suspension and inhibitor exposure times (separate panels). Data points and error bars show mean $\pm$ SEM of the observed fractions over $N \geq 3$ replicate populations. Dotted lines show the predictions of the cis-Golgi birth-death model (Eq. 3.3, Table 3.2). Light solid lines show variations in model predictions over 100 bootstrap replicates (Fig. 3.6A,B). The model is only fit to the completely-fragmented fraction, so the prediction to the partially-fragmented and disorganized fraction is parameter-free. (D,E) Data from this study, for 30 mins or 90 mins of BFA treatment following 0 mins of cell suspension (bootstraps in Fig. 3.6A). (E,G) Data from [Raj+23], for 30 mins of BFA or GCA treatment following 60 mins of cell suspension (bootstraps in Fig. 3.6B).

Parameter	Description	Value
n_0 (0,60)	number of cis-Golgi units	3.6 ± 0.4 , 4.4 ± 0.8
c_0^{BFA} (0,60)	Golgi→ER half-max BFA conc.	$0.15 \pm 0.03 \mu\text{M}$, $0.16 \pm 0.05 \mu\text{M}$
c_0^{GCA} (60)	Golgi→ER half-max GCA	$0.079 \pm 0.025 \mu\text{M}$
t_0 (0,60)	cis-Golgi timescale	$83 \pm 25 \text{ min}$, $359 \pm 225 \text{ min}$

TABLE 3.2: cis-Golgi fragmentation parameter values (mean $\pm$ SD across bootstrap replicates). Values in brackets show cell suspension time in mins.

From the inferred fits, one can see how $n_{\text{cis}}(t; c)/n_0$ changes as a function of BFA concentration and time (Fig. 3.3).

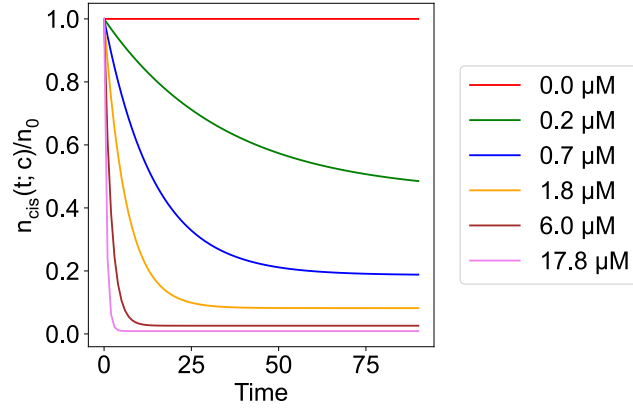


FIGURE 3.3: The number of $n_{\text{cis}}(t; c)$ as a function of t for different c . As the concentration value increases, the drop in n_{cis} begins to saturate.

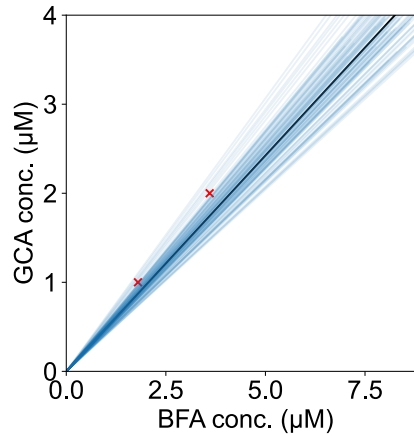


FIGURE 3.4: Calibration between BFA and GCA concentrations in terms of their effect on cis-Golgi morphology. Concentrations of BFA and GCA that obtain the same fraction of fragmented cis-Golgi are plotted against each other. The thick line denotes the calibration obtained from best fit to the data, and the thin blue lines denote calibrations obtained from bootstrap replicates. The red crosses denote BFA and GCA concentrations that were taken to be equivalent and used for the experiments.

The ratio $c_0^{\text{BFA}}/c_0^{\text{GCA}} = 2.07 \pm 0.18$ allows us to calibrate BFA and GCA concentrations that cause equivalent disruption to the cis-Golgi. Among the inhibitor concentrations tested in [Raj+23], the closest equivalent pairs are $([\text{BFA}], [\text{GCA}]) = (1.8 \mu\text{M}, 1 \mu\text{M})$ and $([\text{BFA}], [\text{GCA}]) = (3.6 \mu\text{M}, 2 \mu\text{M})$ (Fig. 3.4). This calibration will be useful when we compare the effects of BFA and GCA on cell-surface glycan levels. We note that the fitted values of c_0^{BFA} and n_0 are comparable between the two protocols, demonstrating the robustness of our model.

3.2 Endocytosis dynamics

The cell surface is not static; in addition to entities being brought to the cell surface from inside the cell, entities are *endocytosed*, or pulled into the cell [Alb+22]. This will naturally change the amount of materials at the cell surface [Poz+05; Paw+16; Dam+91; Pry+92]. Since the distribution of glycans is also influenced by endocytosis, it is imperative to consider if BFA has any effect on endocytosis rates.

Recall that the distribution of cell-surface glycans is measured using the fluorescence intensities of ConA and WGA lectins; see Section 1.2.5. For the endocytosis experiments, ConA was tagged with the green fluorophore and WGA with the red fluorophore. We bound glycans with lectins and quantified the rate at which glycans became internalized at a given amount of BFA concentration and time after BFA addition.

3.2.1 Experimental Procedure

Table 4.1 stops at tagging the cell-surface glycans with lectins. For the endocytosis experiments, the steps were the same until Step 10 of Table 4.1, after which the experiment proceeded as outlined in Table 3.3. We only considered BFA concentrations of 0 μM , 0.1 μM and 1.8 μM , acting for 0, 15 and 30 minutes.

The output of the endocytosis experiments are confocal images showing the amount of internalized lectin. Ten cells were selected from cell populations corresponding to the nine configurations $\{0.0 \mu\text{M}, 0.1 \mu\text{M}, 1.8 \mu\text{M}\} \times \{0', 15', 30'\}$, producing 90 images. A sample of the largest Z stack of confocal images obtained at different BFA concentrations and times is given in Fig. 3.5A. The experiment was repeated for $N = 3$ times producing a dataset of total 270 images.

3.2.2 Image Analysis

Using this data, we quantified how much cell-surface glycans were endocytosed. There were many choices to make this quantification; a natural candidate was to look at total pixel intensities at the surface before and after endocytosis. This approach presented a few problems:

1. Ideally, 100% of the pixel intensities should come from the area of the cell designated as the rim when checking cells in the control population (the population obtained immediately after treating with 0 μM BFA). This was not observed; only around 80% of the pixel intensities are in the area of the cell designated as the rim. This means that the cells have some background fluorescence...
2. ...which we discovered is not constant across cells.

Both of these facts called for a quantity *independent of confounding entities* such as background fluorescence. We defined a quantity called the endocytic index to be the fraction of pixels in the periphery, whose intensity is greater than the median intensity of the interior of the cell. This index produces a fraction close to 1.0 for many cells in the control population and is also robust to changes in background fluorescence. The full image analysis pipeline that does this quantification is outlined below:

1. The cell boundary was identified using the Python module Cellpose [Str+21]. The computed boundaries were manually inspected and found to be well determined for 234 out of 270 cells; these were retained for further analysis.
2. The saved boundaries were imported into Fiji's image analysis software as .roi files [Sch+12]. The cell periphery was defined as the region obtained by scaling the original boundary from 85% to 105%, and the interior was defined as the region within the 85% limit.
3. We computed an endocytic index E for each cell and labeled lectin, defined as the fraction of pixels in the periphery whose intensity is greater than the median intensity of the interior of the cell. This index is robust to changes in background fluorescence and overall fluorescence scales.
4. The endocytic indices of ConA and WGA, pooled across all BFA concentrations, were fit to first-order exponential decay by least-squares minimization: $EI_X(t) = e^{-\lambda_X t}$, $EI_Y(t) = e^{-\lambda_Y t}$.

Step	Procedure	Rationale
11	Labeled cells were equally distributed for three time points: 0 min, 15 min and 30 min.	
12	These cells were kept in 500 μ L low-serum media (with DMSO/BFA) at 37°C with 5% CO ₂ .	
13	Cells were washed with cold PBS, spun down with a centrifuge and fixed with 3.5% PFA after the respective time-points.	PBS wash removes traces of media. Centrifuge spin settles the cells down. PFA cross-links proteins and biomolecules essentially freezing cell structure, making it easier to perform downstream tasks such as confocal imaging.
14	All samples were mounted on slides using Fluoramount-G, allowed to dry and then imaged using confocal microscopy.	

TABLE 3.3: Experimental procedure for testing endocytosis of cell-surface lectins.

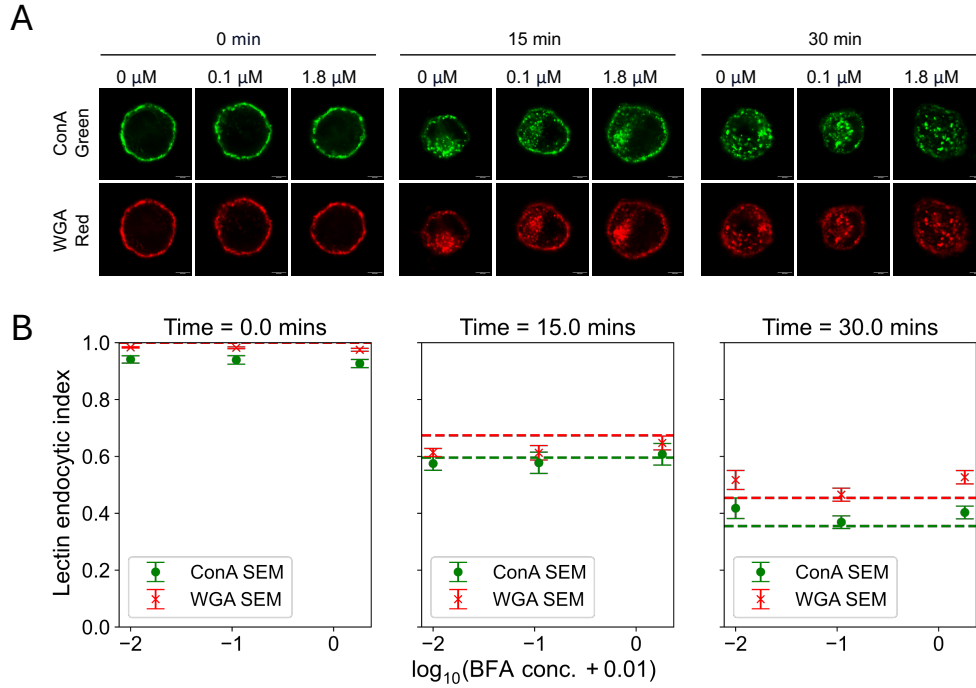


FIGURE 3.5: Endocytosis of cell-surface glycans. **(A)** WT-MEFs labeled with both ConA and WGA are fixed immediately, or maintained at 37°C for 15 min or 30 min before fixing. Confocal sections show representative images of lectin internalization in cells following each treatment. **(B)** We use a normalized endocytic index (EI) to measure the fraction of internalized lectins across time and BFA concentrations ($N \geq 18$, median $N = 28$ cells per concentration, at each time point). Cells at 0 mins with 0 μM BFA (untreated) is control, and the lectin endocytic indices of those cells have a normalized value of 1.0. Dashed lines show the exponential fit to $EI = e^{-\lambda t}$. Endocytic kinetics do not vary with BFA concentration (Fig. 3.7) but are different for ConA and WGA: $1/\lambda_X = 29$ min, $1/\lambda_Y = 38$ min (Fig. 3.6C; Table 3.4).

3.2.3 Results

<i>Parameter</i>	<i>Description</i>	<i>Value</i>
$1/\lambda_X$	ConA endocytic timescale	29 ± 0.9 min
$1/\lambda_Y$	WGA endocytic timescale	38 ± 1.4 min

TABLE 3.4: Endocytosis parameter values (mean $\pm$ SD across bootstrap replicates)

Cell-surface lectin levels showed an exponential decay (Fig. 3.7). The endocytic rates of ConA- and WGA-binding glycans obtained by fitting to $EI = e^{-\lambda t}$ are presented in Table 3.4. The standard deviations were obtained using 100 bootstrap replicates; see Fig. 3.6C.

We found that BFA had no effect on observed endocytic indices (Fig. 3.5B). However, the endocytosis rate of ConA was consistently higher than that of WGA (Table 3.4, Fig. 3.6C).

3.3 Discussion

The stochastic model of cis-Golgi biosynthesis provides a mechanistic basis for the BFA-driven fallback of the cis-Golgi into the ER (Fig. 3.1). The effect of BFA on the trans-Golgi is more subtle. Upon loss of adhesion, the trans-Golgi becomes dispersed, with a loss of active Arf1 and GBF1 [Raj+23] but BFA does not further perturb its morphology.

An interesting parameter that we obtain from our fits is n_0 , the number of cis-Golgi units, with a mean around 3.6. This would need to be verified by independent measurements. The number of stacks is known to range from zero to a few to dozens in animal cells, depending on the nature of the cell type [Ben+24].

One must be careful with the interpretation of the c_0 parameter, the Golgi to ER half-max BFA concentration. Variations of our model, for example, those that assume a fixed supply of cis-Golgi material, or loss of cis-Golgi units to separate units in first-order reactions (which remain unperturbed on inhibitor action), will produce identical fits, but possibly with a different c_0 value.

Regarding endocytosis, we have shown that endocytosis rates do not depend on BFA concentrations for $t \leq 30$ minutes when BFA concentrations are small. Endocytosis rates can increase at higher concentrations or longer times, as shown in human fibroblasts [Dam+91] and MDCK cells [Pry+92]. The functional consequence of endocytosis rates of WGA-binding glycans in MEFs being faster than ConA-binding glycans, if any, is yet to be explored.

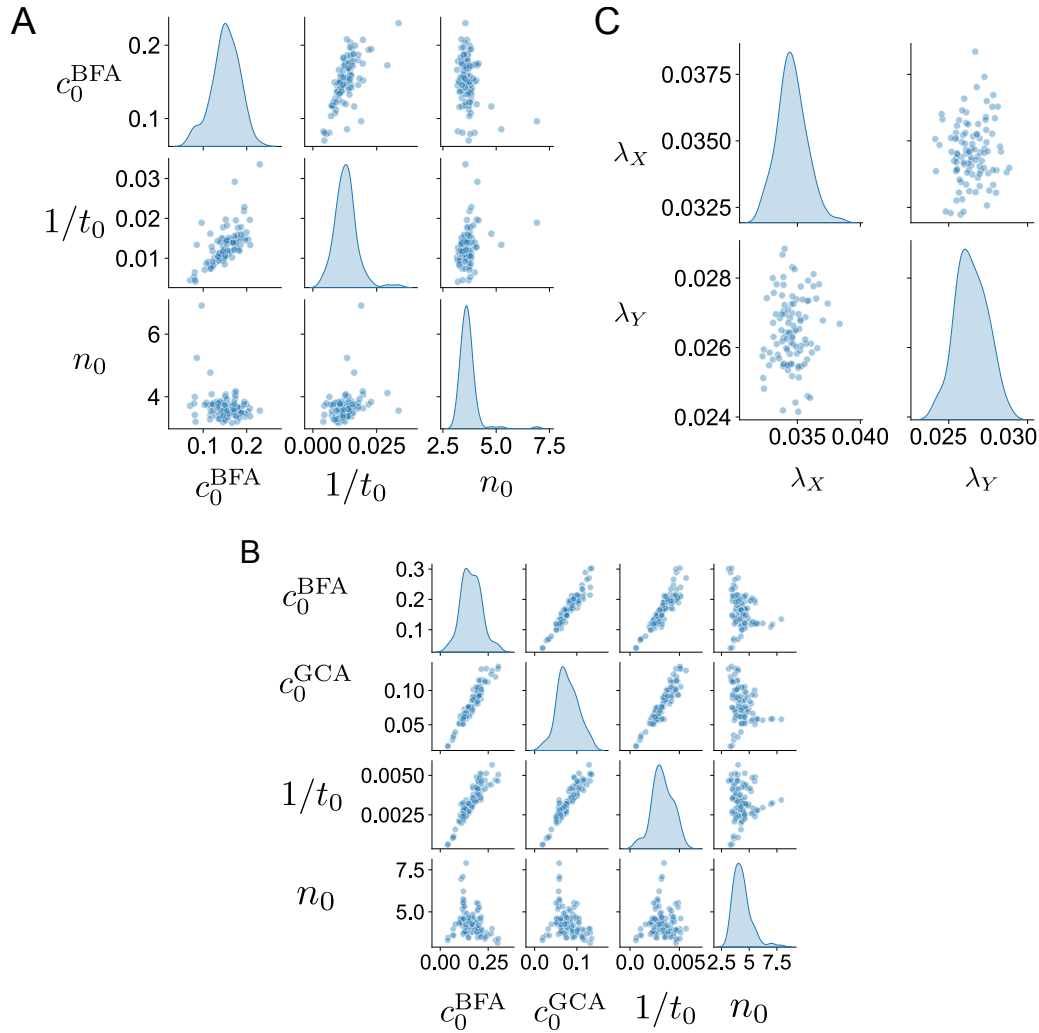


FIGURE 3.6: To check for non-identifiability issues in our models and to obtain confidence intervals for our parameters, we performed bootstrap replicates where we sampled the data with replacement and fit the sampled data to the model. A unimodal peak and a tight distribution of the parameter fits indicates that the model is not overparameterized. Bootstrap estimates of fitted parameter uncertainties from 100 bootstrap replicates for (A) cis-Golgi fragmentation dynamics for 30 and 90 minutes of BFA (B) cis-Golgi fragmentation dynamics for BFA and GCA with common n_0 and t_0 and BFA- and GCA- specific c_0 (a single outlier with $1/t_0 < 0.006$ was removed), and (C) endocytosis parameters.

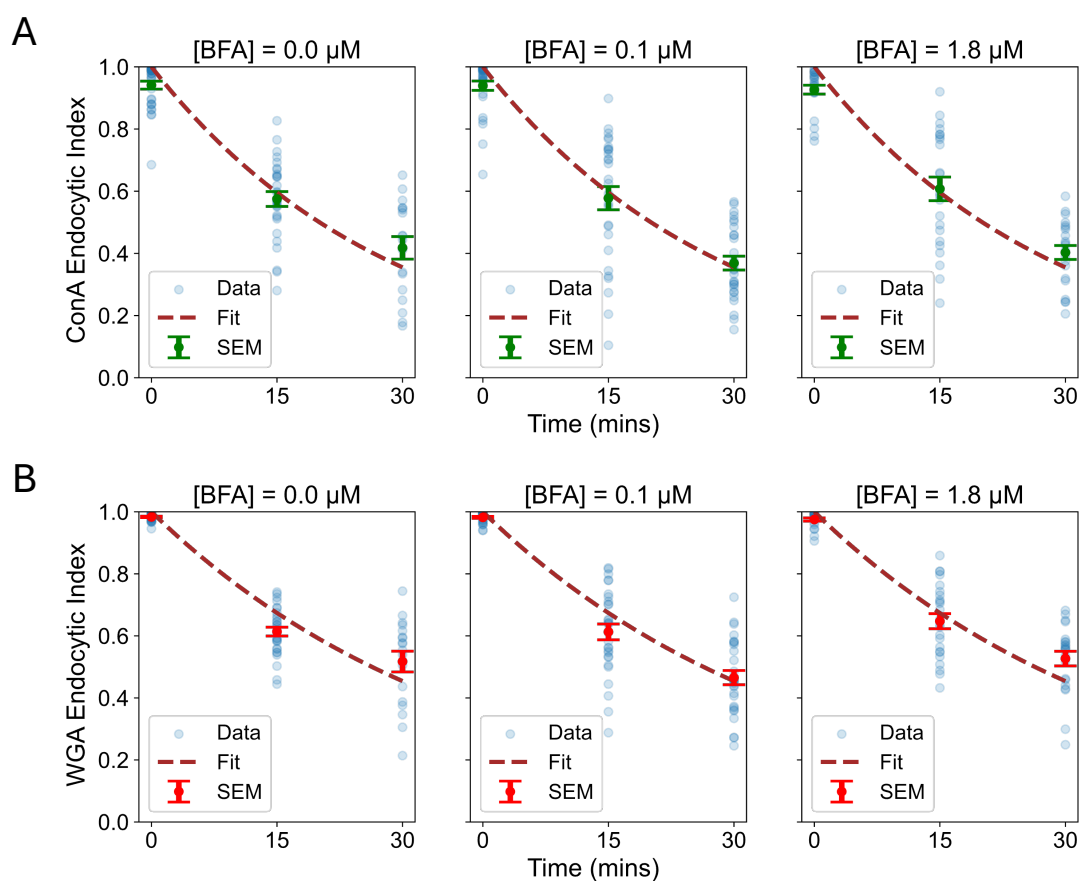


FIGURE 3.7: Dynamics of endocytosis. Blue points show single-cell values, and green or red points show mean $\pm$ SEM of (A) ConA and (B) WGA endocytic indices. Dotted lines show the exponential fit to $EI = e^{-\lambda t}$. There are $N \geq 18$ datapoints at every concentration, for a fixed time.

Chapter 4

Dynamics at the cell surface

In this chapter, we finally integrate the cis-Golgi fragmentation model, results from endocytosis experiments, and changes in Golgi trafficking from BFA action to characterize the cell-surface glycan distribution at a given BFA concentration and time.

4.1 Experimental procedure

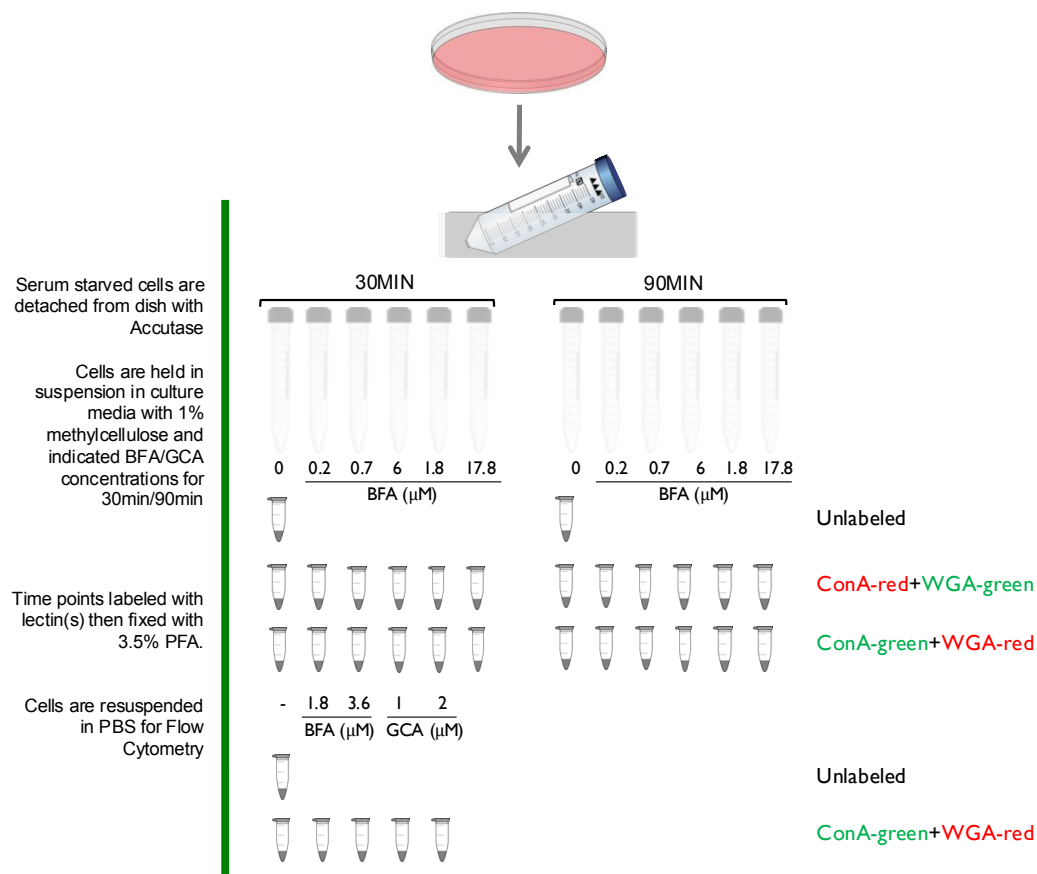


FIGURE 4.1: Schematic of the lectin labelling experiment. (Figure credits: Prachi Joshi.)

The schematic of the experimental procedure is given in Fig. 4.1, and the steps are outlined in Table 4.1. The table picks up from Step 7 of Table 3.1, at which point we obtain a suspension of cells treated with BFA. After the cells were tagged with lectins, samples were

run on BD Celesta flow cytometer. All flow cytometry experiments (except BFA/GCA comparison) were performed using fluorophore-swapped versions of the lectins as controls (CGWR, Con-Green/WGA-Red; CRWG, ConA-Red/WGA-Green). For clarity, we report our main results in CGWR units (we will provide details of the conversion shortly). The experiments result in single-cell fluorescent intensity measurements along with some cell size measurements at different BFA and time configurations.

<i>Step</i>	<i>Procedure</i>	<i>Rationale</i>
8	Cells were given a PBS wash and then reconstituted in a defined volume of PBS.	To remove residual media.
9	Cells were incubated with optimal concentrations for ConA and WGA tagged with different fluorophores (Alexa488, Alexa647).	For lectin labeling.
10	The labeling reaction was kept in ice in the dark for 15 minutes, followed by two washes with cold PBS.	To prevent photobleaching of the fluorophores and to allow for the labeling to happen.
11	PFA (3.5%) was used to fix the lectin-labeled cell samples and resuspended in cold PBS (350 μ L).	PFA cross-links proteins and biomolecules, essentially freezing cell structure. This makes it easier to perform downstream tasks such as flow cytometry, since cell structure is preserved.
12	Cells were run through a flow cytometer. The cell population, as defined morphologically using their forward and side scatter (FSC and SSC) plot, was selected using polygon gating for data acquisition.	To remove outliers and signals arising due to foreign objects, and to also gauge autofluorescence (gating is done on the unlabeled population).
13	More than 5,000 events generated in the last step were further analyzed using FlowJo software to get single-cell data for each sample.	

TABLE 4.1: Experimental procedure for lectin labeling and flow cytometry analysis.

4.2 Modeling

Cell-surface glycan levels are determined by an interplay between synthesis and secretion from the Golgi, endocytosis, and recycling. We combined all these processes within a

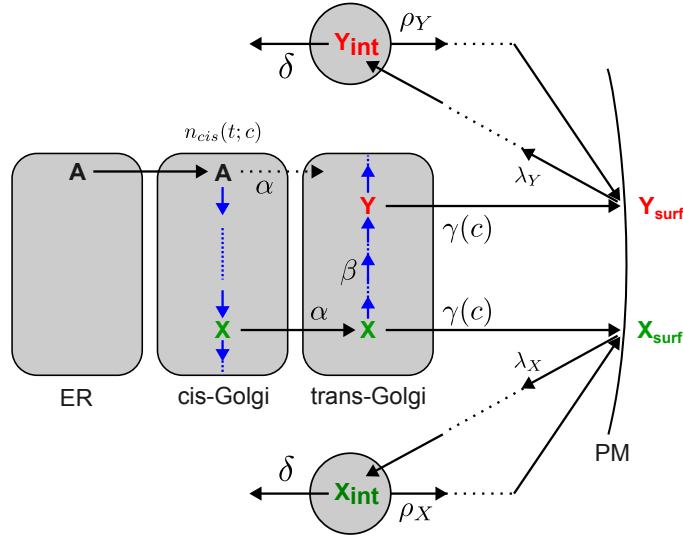


FIGURE 4.2: Modeling the synthesis and trafficking of glycans. For simplicity, we assume that the Golgi consists of two enzymatically distinct compartments, cis and trans. Oligo-mannose glycans (A) are synthesized in the ER and transported to the cis-Golgi, where they are further processed to form ConA-binding glycans (X). Glycans from the cis-Golgi are transported to the trans-Golgi with rate-constant α , so the flux of glycans emerging from the cis-Golgi is $\alpha n_{cis}(t; c)$ (c represents the BFA concentration). X can be directly trafficked from the trans-Golgi to the plasma membrane with rate-constant $\gamma(c)$, or it may be first glycosylated with rate-constant β to create WGA-binding glycans (Y). The fraction of X converted to Y is therefore $\beta/(\gamma(c) + \beta)$. The trans-Golgi residence time is proportional to $1/\gamma(c)$. ConA- and WGA-binding glycans at the cell surface (X_{surf} and Y_{surf}) are endocytosed at rates λ_X and λ_Y . We assume that there are stable internal pools X_{int} and Y_{int} from which glycans are recycled to the plasma membrane with rate constants ρ_X and ρ_Y , or degraded with rate constant δ .

unified model of intracellular traffic and glycosylation (Fig. 4.2). This model is based on the following assumptions, supported by published data and our own observations:

- The time taken for a glycan to go from the cis-Golgi to the plasma membrane is much shorter than the timescale over which the mean number of cis-Golgi units changes. This is justified since Golgi residence times range from 5 to 15 minutes [Sun+21; Hir+98], whereas the observed timescale over which cis-Golgi activity falls is $t_0 \sim 83$ minutes (Fig. 3.2, Table 3.2).
- Endocytic rate constants are independent of time and BFA concentrations, for the time scale our experiments. This is consistent with our endocytosis measurements (Fig. 3.5).
- Intracellular pools of endocytosed glycans are stable over the time course of the experiment, so the recycling rates are time-invariant [ME19; XW17].
- Glycosylation enzymes are not saturated [Hir+98], so enzymatic reactions obey first-order kinetics.
- BFA does not directly affect enzymatic rates [KDL92].

- BFA reduces the flux of glycans to the trans-Golgi via its influence on the number of cis-Golgi units [NRR02] (Fig. 3.2).
- BFA can inhibit the formation of clathrin-coated vesicles from the trans-Golgi to the plasma membrane, since these are regulated by Arf1 activity [NRR02; SG20].

The dynamical variables X_{surf} and Y_{surf} below correspond to the surface levels of the ConA and WGA substrates, and X_{int} and Y_{int} correspond to internal pools from which they are recycled to the cell surface. Fig. 4.2 explains how the dynamics of X_{surf} and X_{int} depends on intracellular traffic and glycosylation. The dynamics can be written as:

$$\begin{aligned} \frac{dX_{\text{surf}}}{dt} &= \underbrace{\alpha \cdot n_{\text{cis}}(t; c)}_{\text{synthesis}} \cdot \underbrace{\frac{\gamma(c)}{\gamma(c) + \beta}}_{\text{modification}} - \underbrace{\lambda_X X_{\text{surf}}}_{\text{endocytosis}} + \underbrace{\rho_X X_{\text{int}}}_{\text{recycling}} \\ \frac{dX_{\text{int}}}{dt} &= \lambda_X X_{\text{surf}} - (\delta + \rho_X) X_{\text{int}} \end{aligned} \quad (4.1)$$

and that of Y_{surf} and Y_{int} can be written as

$$\begin{aligned} \frac{dY_{\text{surf}}}{dt} &= \alpha \cdot n_{\text{cis}}(t; c) \cdot \frac{\beta}{\gamma(c) + \beta} - \lambda_Y Y_{\text{surf}} + \rho_Y Y_{\text{int}} \\ \frac{dY_{\text{int}}}{dt} &= \lambda_Y Y_{\text{surf}} - (\delta + \rho_Y) Y_{\text{int}}. \end{aligned} \quad (4.2)$$

We use two parameters to define how inhibitors affect the rate constant of Golgi-to-PM transport: $\gamma(c) = \gamma_0 / (1 + c/c_1)$.

At 0 μM BFA, one can solve for the steady-state concentration of X_{int} and Y_{int} , in terms of the steady-state concentration of surface glycan levels:

$$X_{\text{int}}^0 = \frac{\lambda_X}{\delta + \rho_X} X_{\text{surf}}^0$$

and

$$Y_{\text{int}}^0 = \frac{\lambda_Y}{\delta + \rho_Y} Y_{\text{surf}}^0.$$

Substituting these values into Equations 4.1 and 4.2 gives the steady-state value of X_{surf} and Y_{surf} at control:

$$\begin{aligned} X_{\text{surf}}^0 &= \frac{1}{\lambda_X} \alpha n_0 \frac{\gamma_0}{\gamma_0 + \beta} \frac{\delta + \rho_X}{\delta} \\ Y_{\text{surf}}^0 &= \frac{1}{\lambda_Y} \alpha n_0 \frac{\beta}{\gamma_0 + \beta} \frac{\delta + \rho_Y}{\delta}. \end{aligned} \quad (4.3)$$

We normalize Equations 4.1 and 4.2 with X_{surf}^0 and Y_{surf}^0 to allow a direct comparison with normalized lectin signals. The dynamics of normalized lectin signals X_{surf} and Y_{surf} can thus be obtained by solving the formidable equations (we relabel X_{surf} , Y_{surf} , X_{int} and Y_{int} as their normalized quantities):

$$\begin{aligned} \frac{dX_{\text{surf}}}{dt} &= \frac{n_{\text{cis}}(t; c)}{n_0} \cdot \frac{1 + \beta/\gamma_0}{1 + (\beta/\gamma_0)(1 + c/c_1)} \frac{\delta}{\delta + \rho_X} - \lambda_X X_{\text{surf}} + \rho_X X_{\text{int}} \\ \frac{dX_{\text{int}}}{dt} &= \lambda_X X_{\text{surf}} - (\delta + \rho_X) X_{\text{int}} \end{aligned} \quad (4.4)$$

and

$$\begin{aligned}\frac{dY_{\text{surf}}}{dt} &= \frac{n_{\text{cis}}(t; c)}{n_0} \left(1 + \frac{c}{c_1}\right) \frac{1 + \beta/\gamma_0}{1 + (\beta/\gamma_0)(1 + c/c_1)} \frac{\delta}{\delta + \rho_X} - \lambda_Y Y_{\text{surf}} + \rho_Y Y_{\text{int}} \\ \frac{dY_{\text{int}}}{dt} &= \lambda_Y Y_{\text{surf}} - (\delta + \rho_Y) Y_{\text{int}}.\end{aligned}\quad (4.5)$$

with the initial conditions $X_{\text{surf}} = 1$, $Y_{\text{surf}} = 1$, $X_{\text{int}} = \lambda_X / (\delta + \rho_X)$ and $Y_{\text{int}} = \lambda_Y / (\delta + \rho_Y)$. Notice that α has been absorbed during normalization. The terms corresponding to cis-Golgi fragmentation ($n_{\text{cis}}(t; c)/n_0$; Fig. 3.3) and endocytosis (λ_X, λ_Y ; Fig. 3.7) have already been determined in the previous chapter.

The half-life of glycans is estimated to be in the range of 10 to 100 hours [XW17]. A half-life of $t_{1/2}$ corresponds to a degradation rate of $\ln 2 / t_{1/2}$. We shall see that varying the degradation rate of the ConA and WGA substrates over this broad range has no discernible effect on lectin levels over a 90 min timecourse. This justifies our choice to treat internal glycan pools as being in quasi-steady state throughout the time of our experiment, and we discuss this next.

4.2.1 Model with internal pools in quasi-steady state

Recall the equations for (unnormalized) X_{surf} and Y_{surf} :

$$\begin{aligned}\frac{dX_{\text{surf}}}{dt} &= \alpha \cdot n_{\text{cis}}(t; c) \cdot \frac{\gamma(c)}{\gamma(c) + \beta} - \lambda_X X_{\text{surf}} + \rho_X X_{\text{int}} \\ \frac{dY_{\text{surf}}}{dt} &= \alpha \cdot n_{\text{cis}}(t; c) \cdot \frac{\beta}{\gamma(c) + \beta} - \lambda_Y Y_{\text{surf}} + \rho_Y Y_{\text{int}}\end{aligned}\quad (4.6)$$

Since internal glycan pools are assumed to be in a quasi-steady state, $X_{\text{int}} = X_{\text{surf}}^0 \frac{\lambda_X}{\delta + \rho_X}$, $Y_{\text{int}} = Y_{\text{surf}}^0 \frac{\lambda_Y}{\delta + \rho_Y}$. We can repeat the same steps as in the previous section to obtain the equations for normalized cell-surface lectin levels:

$$\begin{aligned}\frac{dX_{\text{surf}}}{dt} &= \lambda_X \cdot \left(\frac{n_{\text{cis}}(t; c)}{n_0} \cdot \frac{\frac{\gamma(c)}{\gamma(c) + \beta}}{\frac{\gamma_0}{\gamma_0 + \beta} + \frac{\rho_X X_{\text{int}}}{\alpha n_0}} - X_{\text{surf}} + \frac{\rho_X X_{\text{int}}}{\alpha n_0 (\frac{\gamma_0}{\gamma_0 + \beta} + \frac{\rho_X X_{\text{int}}}{\alpha n_0})} \right) \\ \frac{dY_{\text{surf}}}{dt} &= \lambda_Y \cdot \left(\frac{n_{\text{cis}}(t; c)}{n_0} \cdot \frac{\frac{\beta}{\gamma(c) + \beta}}{\frac{\beta}{\gamma_0 + \beta} + \frac{\rho_Y Y_{\text{int}}}{\alpha n_0}} - Y_{\text{surf}} + \frac{\rho_Y Y_{\text{int}}}{\alpha n_0 (\frac{\beta}{\gamma_0 + \beta} + \frac{\rho_Y Y_{\text{int}}}{\alpha n_0})} \right)\end{aligned}\quad (4.7)$$

with the initial conditions $X_{\text{surf}}(0) = 1$ and $Y_{\text{surf}}(0) = 1$. There are four parameters that need to be specified: γ_0/β , the Golgi-to-PM rate constant; c_1 , the Golgi-to-PM half-max BFA concentration; and $\rho_X X_{\text{int}}/\alpha n_0$ and $\rho_Y Y_{\text{int}}/\alpha n_0$, the scaled ConA and WGA recycling rates. This four-parameter model is what we adopt throughout the rest of the thesis.

4.3 Data analysis pipeline

The raw data of a flow cytometry experiment for a particular BFA concentration and time are single cell ($n > 5,000$) red and green fluorescence intensity readings and cell size measurements. We processed the raw data as follows to fit the above model to data:

1. Red and green intensities of single cells were consistently proportional to their forward scatter (FSC) in the flow-cytometric measurements. We interpreted this as a cell size effect and corrected for it by scaling all intensities by the FSC (Fig. 4.7).

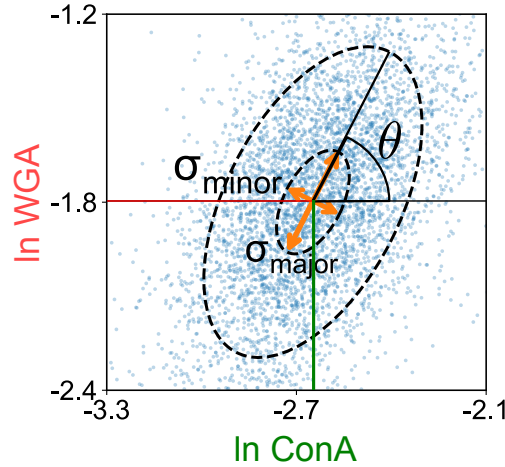


FIGURE 4.3: Graphical representation of the shape parameters of an ellipse corresponding to a bivariate normal distribution: angle with respect to the uncorrelated frame, θ ; scale, σ_{major} ; and the aspect ratio, $\sigma_{\text{minor}}/\sigma_{\text{major}}$.

2. The signals were observed to form an approximate bivariate log-normal distribution, which has a characteristic elliptical scatter (Fig. 4.3, 4.7C). The distribution means and shape parameters of this ellipse (angle with respect to the uncorrelated frame, θ ; major axis, σ_{major} ; and the ratio of minor to major axis, $\sigma_{\text{minor}}/\sigma_{\text{major}}$) were robustly estimated using the Python module astroML [Van+12] (Fig. 4.3). The shape parameters were used to calculate the covariances of the distribution.
3. For the BFA-only data, we checked if 95% of the cell population fell within the 6σ contour of the fitted distribution. Of the 237 flow cytometry experiments performed (across time, BFA concentrations and replicates) 207 satisfied this condition and were retained for further analysis. For the batch of experiments comparing signals due to BFA and GCA (Fig. 4.6D), we checked if 91.5% of the population fell within 6σ contour of the fitted distribution. 10 on 19 experiments satisfied this condition, and were retained for further analysis.
4. Sample preparation was carried out in batches, as shown in Fig. 4.1. There were large differences in absolute fluorescence levels across batches (Fig. 4.8, left column). To compensate for this in the case of BFA-only data, we scaled all fluorescence signals for each batch by the geometric mean of the 5 min, 30 min and 90 min $0\ \mu\text{M}$ BFA controls in the same batch. Fluorescence values were tightly distributed following this normalization (Fig. 4.8, right column). We refer to this final normalized intensity as the red or green signal. In case of experiments comparing BFA and GCA signals, we scaled all fluorescence signals of an experiment by the 30 min $0\ \mu\text{M}$ BFA or GCA control of that experiment.
5. For each condition (combination of BFA concentration and time), there were at least nine biological replicates and at least 20 replicates for the $0\ \mu\text{BFA}$ controls. We computed the geometric mean of red signals and the geometric mean of green signals, across replicates for each condition and fluorophore configuration. We then compared the mean red and green signal across equivalent experiments (i.e. red of CRWG against green of CGWR; or red of CGWR against green of CRWG). These showed a tight linear relationship (Fig. 4.4A). We therefore pooled all the data across both fluorophore-swapped controls for subsequent analysis, using the best-fit line

from Fig. 4.4A to convert CRWG measurements into CGWR units. For clarity, we present our results in CGWR units.

In summary, the ConA signal is defined as the median green fluorescence intensity of a cell population, normalized by the green fluorescence intensity of the zero BFA control from the same batch. The WGA signal is the median *red* fluorescence intensity of a cell population normalized by the red fluorescence intensity of the zero BFA control from the same batch. Note that in Step 2 of the above pipeline, we calculate the center (mean) of the bivariate log-normal distribution; this is equivalent to the log of *median* ConA and WGA signals in the linear scale.

We solved Eq. 4.6 with initial conditions $X_{\text{surf}} = 1$ and $Y_{\text{surf}} = 1$, to allow a direct comparison with normalized lectin signals. We inferred model parameters by least-squares fitting, using Python's SciPy module [Vir+20]. To estimate confidence intervals for each parameter, we repeated the fitting procedure over 100 bootstrap replicate datasets. These datasets were synthesized by randomly sampling the ConA and WGA signals with replacement for each time and BFA concentration (except the 0 μM BFA controls which were earlier used to normalize intensities).

4.4 Results

The pooled CGWR and CRWG data (after conversion into CGWR units) gave ≥ 7 replicates across BFA concentrations and timepoints. At each BFA concentration, both ConA and WGA signals were seen to decrease from the 30 min to the 90 min time point, as expected when glycan synthesis is reduced (Fig. 4.4B,C). ConA signals showed a steep and consistent decrease as a function of BFA (Fig. 4.6A), while WGA signals showed a pronounced increase at low BFA (0.2 μM) followed by a subsequent decrease (Fig. 4.6B).

There was a high degree of cell-to-cell variability within cell populations at each condition (Fig. 4.7). Cell populations were observed to form a bivariate log-normal distribution. The elliptical 1σ contour in log-log space provides a geometric summary of the covariance of a bivariate log-normal distribution. The center of the ellipse is the logarithm of the median ConA and WGA signals in linear space. The shape of the ellipse can be quantified by three parameters: the angle of the uncorrelated frame with respect to the X-axis (specified by θ); scale, specified by the major axis σ_{major} ; and aspect ratio, specified by $\sigma_{\text{minor}}/\sigma_{\text{major}}$. We found that the shape parameters of the ellipse were essentially independent of time and BFA concentration for the CGWR and CRWG constructs (Figs. 4.9, 4.10). We therefore modeled the covariance as constant and averaged the angle, major axis, and minor axis across conditions. This covariance matrix will be used to build the decoder.

Next, we fit the data to the solutions of Equations 4.4 and 4.5. As mentioned before, we evaluated two scenarios: one where glycoprotein half-life was 10 hours and another where glycoprotein half-life was 100 hours. This captures the wide range of reported glycoprotein half-lives [XW17]. There is no discernible effect under the different degradation rates in a 90-minute timecourse (Fig. 4.5A, B), but the effect is observable at much longer timescales (Fig. 4.5C, D). Even when the contribution of glycoproteins from cis-Golgi to the cell-surface drops, cell-surface glycan levels will remain relatively high because of replenishment from internalized pools. This gives rise to the intermediate phase observed in Figs. 4.5C and D. When the internal pools are finally depleted, the relative levels of cell-surface glycans drops to zero (Fig. 4.5C and D). (Note that Fig. 4.5C and D plot the

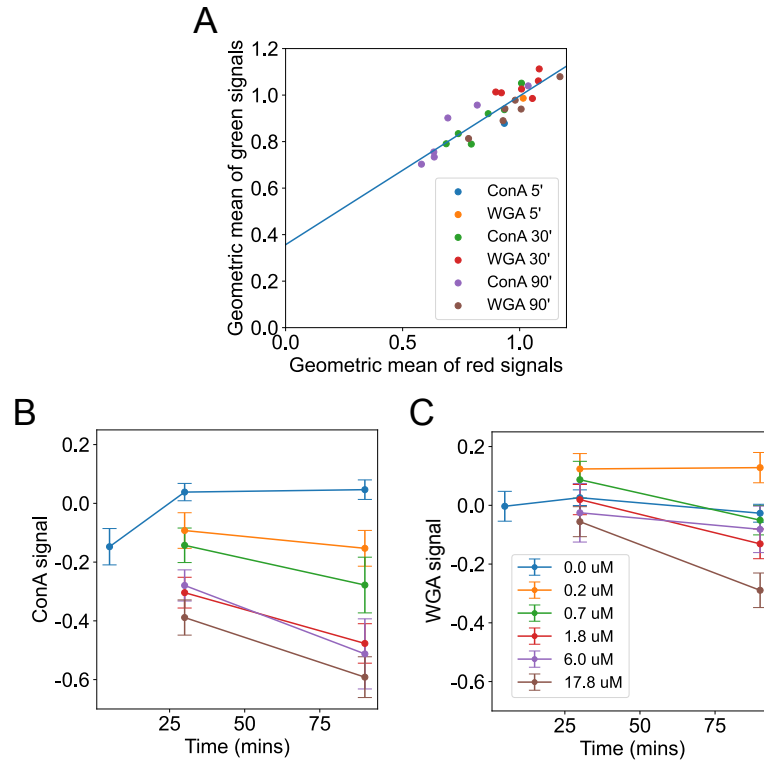


FIGURE 4.4: **(A)** Comparison of fluorophore-swapped lectin measurements. Each point corresponds to a specific BFA concentration, time, and fluorophore configuration, and shows the geometric mean of the signal across replicates. **(B,C)** Dynamics of ConA and WGA signals. We convert CRWG experiments to CGWR units by using the linear relationship in Fig. 4.4A, then pool all replicates. Points show mean $\pm$ SEM across replicates in log space (natural logs). For 0 μ M BFA the ConA signal (but not the WGA signal) transiently drops at the 5 min time before recovering. This may be due to reconfiguration of intra-Golgi trafficking upon cell de-adhesion.

normalized fluorescence intensity levels and will fall to the background fluorescence levels at long times.)

Since recycling rates only change significantly long after the experiment timepoints, we treated them to be in quasi-steady state (Equation 4.7) and fit this model to data. The observed ConA and WGA dynamics were well captured by the trafficking model Eq. 4.6, with just four free parameters (Fig. 4.6A-C). Most fitted parameter values fell within a tight uncertainty range (Table 4.2; Fig. 4.11).

We next compared the impact of different inhibitors on ConA and WGA dynamics. We found that, at equivalent inhibitor concentrations (Fig. 3.4), the effect of BFA and GCA on cell surface lectins was identical (Fig. 4.6D). In particular ConA signals showed a consistent decrease while WGA signals showed a consistent increase at the 30 min time point, under both BFA and GCA treatments. The consistent effect of two distinct inhibitors strongly suggests that the observed variation of cell-surface glycan levels is due to disruption of the cis-Golgi.

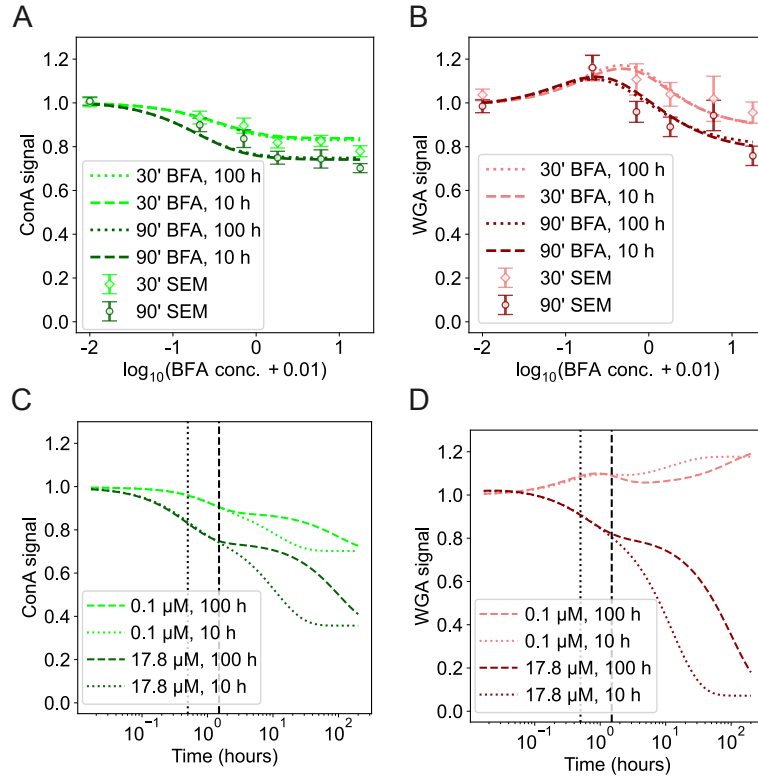


FIGURE 4.5: Model fit to experimental data is agnostic to glycoprotein half-lives, since the timepoints of experiment are much shorter than the half-lives. (A) and (B) denote the fits to ConA and WGA signals respectively, under glycoprotein half-lives of 10 h and 100 h (legend). (C) and (D) denote long-term behaviour of ConA and WGA signals respectively, under the same half-lives. The difference in ConA and WGA signals under different half-lives cannot be discerned at the 30 min or 90 min observation time of our experiments (vertical lines), only becoming apparent at much longer times (the time axis is in log scale). Note that the signals shown here are in fluorescence units and not glycan units, and therefore fall to the background fluorescence level at long times (Fig. 4.4A).

Parameter	Description	Value
γ_0/β	Golgi-to-PM rate constant	7.9 ± 2.0
c_1	Golgi-to-PM half-max BFA conc.	$0.08 \pm 0.03 \mu\text{M}$
$\rho_X X_{\text{int}}/\alpha n_0$	ConA recycling rate	1.50 ± 0.17
$\rho_Y Y_{\text{int}}/\alpha n_0$	WGA recycling rate	3.86 ± 1.13

TABLE 4.2: Golgi-to-PM and recycling parameter values (mean $\pm$ SD across bootstrap replicates)

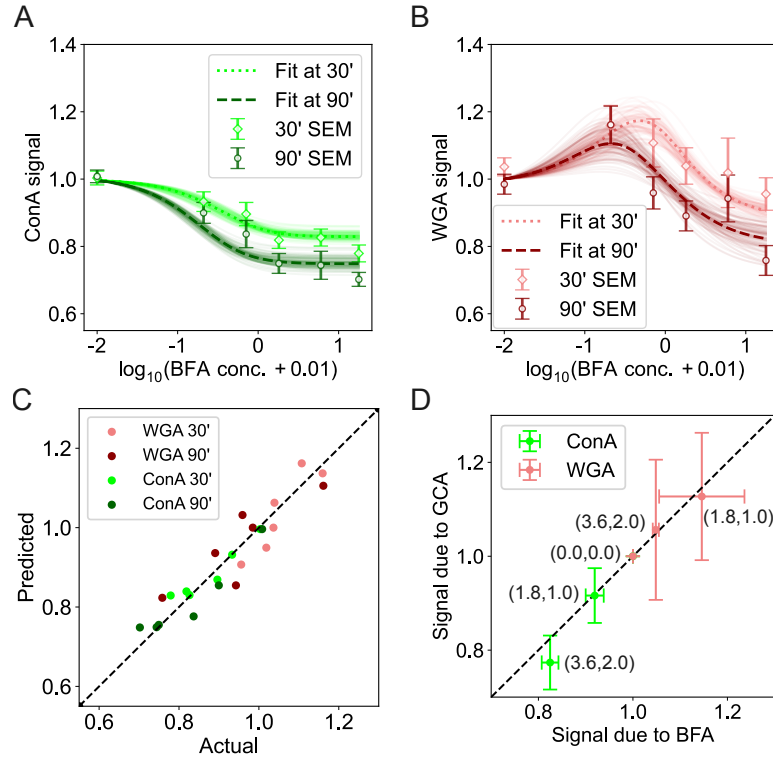


FIGURE 4.6: Cell-surface glycan dynamics can be explained by Golgi fragmentation dynamics. **(A,B)** Dynamics of cell-surface ConA (A) and WGA (B) signals, represented as a function of time and BFA concentration. The signal is defined as the median fluorescence intensity of a cell population, normalized by the intensity of the zero BFA control from the same batch. Data points and error bars show mean $\pm$ SEM of signals across $N \geq 7$ replicate populations, after conversion into CGWR units (Fig. 4.4A). The dotted and dashed lines show the predictions of the 4-parameter trafficking model. Light solid lines show variations in model predictions over 100 bootstrap replicates **(C)** There is a strong agreement between predicted and actual values. **(D)** ConA and WGA signal at 30 min, under treatment with concentrations of BFA and GCA that cause similar disruption of the cis-Golgi (Fig. 3.4). In parentheses, we show (BFA, GCA) concentrations corresponding to each experiment ($N = 3$ replicates, except for the highest inhibitor levels which have $N = 2$ replicates; measurements were done using the CGWR constructs). Both inhibitors produce consistent changes in ConA and WGA levels.

4.5 Discussion

Several works [SEB18; Raj+23] point to the fact that the effect of perturbing a cell by loss of adhesion can be mimicked by a careful experiment involving an Arf1 inhibitor, such as brefeldin A. Given that is the case, a model such as ours can be repurposed to study Golgi dysfunction on loss of adhesion. We have demonstrated that equivalent levels of BFA and GCA cause identical effects on cell-surface glycans at the single-cell level, strongly bolstering the idea that these cell-surface changes are indeed mediated by changes in the cis-Golgi. This hints at the broad applicability of our framework.

The value γ_0/β indicates that most of the substrates produced in cis-Golgi are transported to the plasma membrane rather than being modified in the cis-Golgi, and is a testable prediction emerging from the model. The contrast in the response of ConA and WGA is unexpected and counterintuitive, has never been previously observed, and we have provided a plausible mechanistic explanation for it.

The effect of inhibitors on the trans-Golgi is more subtle. Upon loss of adhesion, the trans-Golgi becomes more dispersed, with a loss of active Arf1 [Raj+23] (Fig. 4.12). The distribution of different trans-Golgi phenotypes suggests that the trans-Golgi is not affected on BFA perturbation given that cells are already suspended. However, it appears that there are more punctate structures in the BFA-treated cells than the cells at CNT ([YSS95], Fig. 4.12). Our synthesis/trafficking model suggests that glycosylation continues to occur in the trans-Golgi following inhibitor addition, but the increased compartmental residence time changes the proportion of different glycan types that are synthesized and delivered to the plasma membrane (Fig. 4.2). This increase is captured by the two parameters which specify the BFA dependence of the Golgi-to-PM transport rate $\gamma(c) = \gamma_0/(1 + c/c_1)$, and result in the “bump” in WGA signals. However, this effect would need to be checked by independent measurements.

The scale and timing of the observed ConA and WGA signal variations are essentially fixed by $n_{\text{cis}}(t; c)$ and λ_X, λ_Y , showing that the overall decrease in ConA and WGA signals is driven by the reduced flux from the cis-Golgi and ongoing endocytosis. The recycling rates $\rho_X X_{\text{int}}$ and $\rho_Y Y_{\text{int}}$ set the baseline level of surface glycans at high BFA concentration, where synthesis is expected to be negligible (note that recycling may be overestimated due to background fluorescence). This explains what possibly was an unexpected result: that complete disruption of the cis-Golgi by BFA or GCA only causes a partial (on the scale of 25%) drop in cell-surface glycan levels, as measured by lectin labeling. This is likely due to recycling from internal pools over the 90 minute time frame of our measurements. Over time we do expect the internal pools to themselves be depleted (Fig. 4.5C,D) when we would expect to see further drops in cell-surface lectins. The fits to recycling rates are another testable prediction emerging from the model; around 60% of the ConA substrate ($= 1.5/(1.5 + 1)$) and 80% of the WGA substrate ($= 3.86/(3.86 + 1)$) on the cell surface is expected to be contributed by recycling.

Next, we return to the question posed at the beginning of the thesis: decoding Golgi dysfunction from cell-surface glycans.

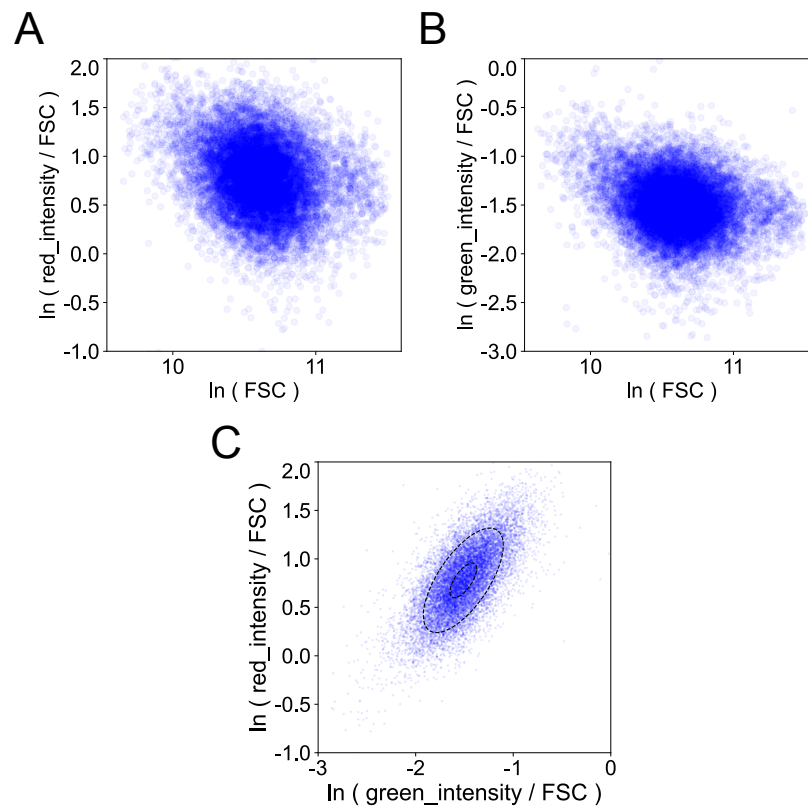


FIGURE 4.7: Example of flow-cytometric single-cell lectin data at 0 μM BFA and 30 min, for ConA-Green / WGA-Red lectins ($n = 13,882$ cells). **(A,B)** Lectin signal intensities after scaling by forward scatter (FSC). **(C)** Scaled red vs. green intensities. We use robust estimation to find the parameters of the best-fit bivariate log-normal distribution. Ellipses show the 1σ and 3σ contours of the distribution.

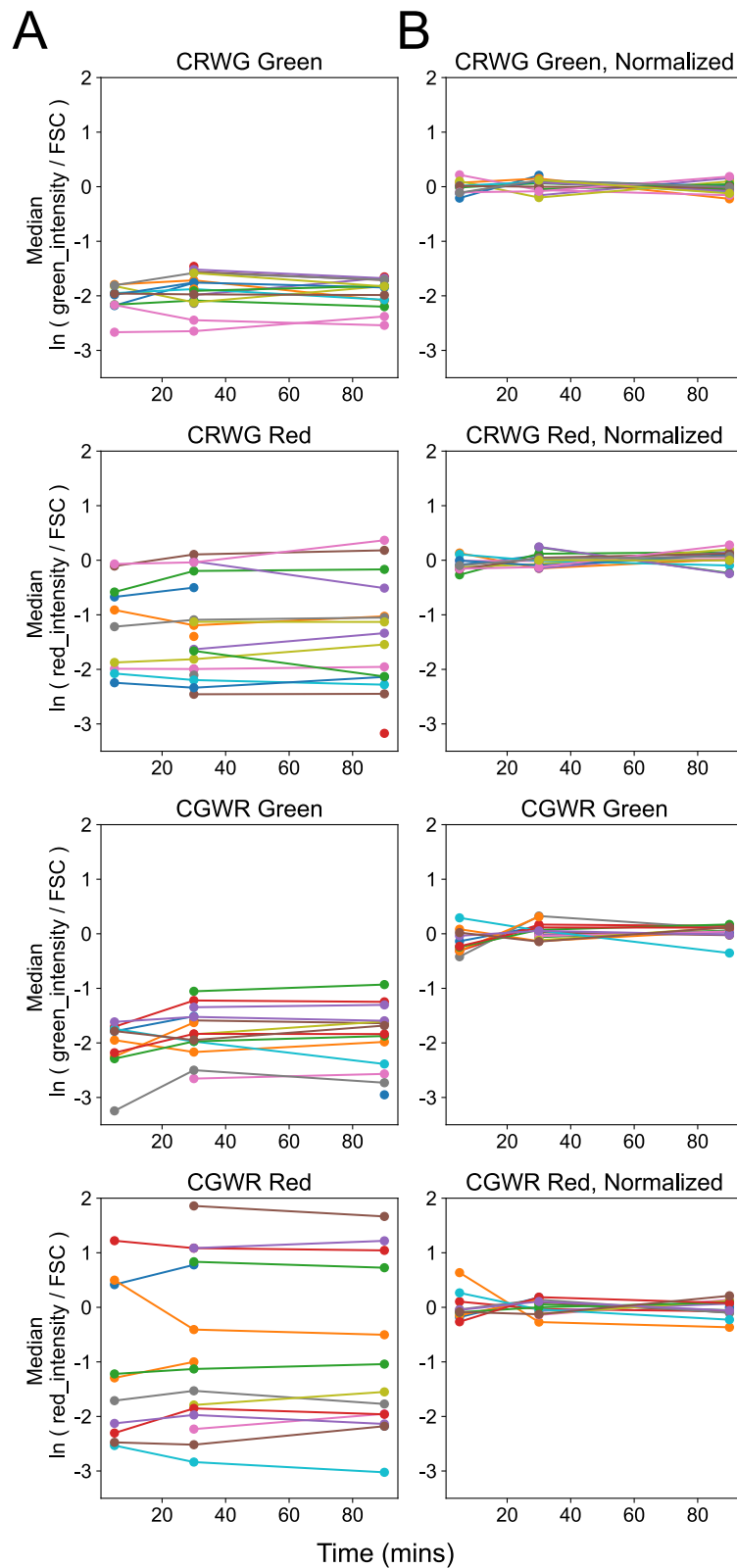


FIGURE 4.8: Batch-to-batch normalization. We show median red and green intensities of cell populations at 0 μM BFA and various time points. (A) Raw intensities. (B) Intensities after normalizing by the geometric mean of the 5 min, 30 min and 90 min values. Batches were retained if the BFA 0 μM control was available for at least one time point.

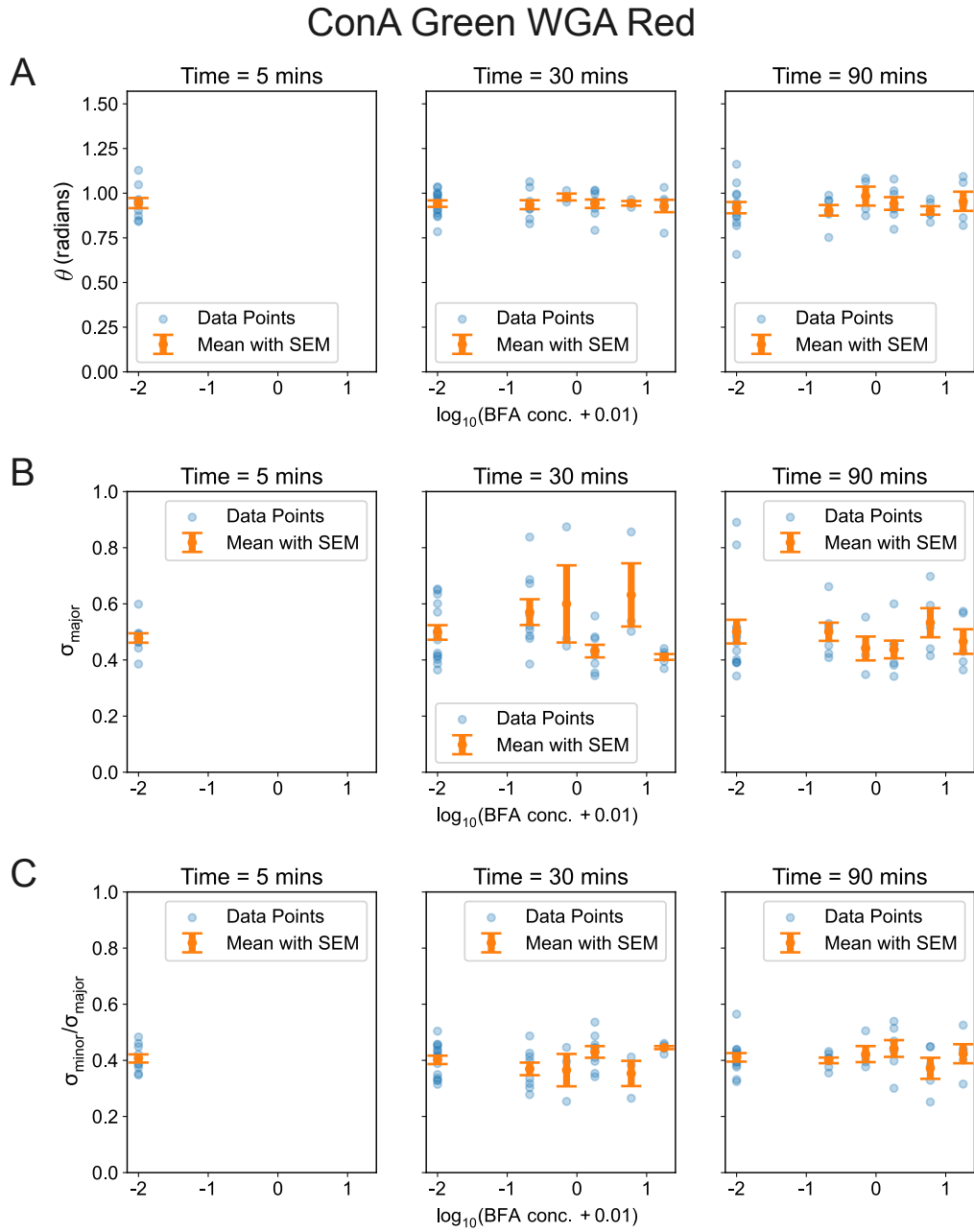


FIGURE 4.9: Covariances of the ConA-Green WGA-Red lectins as a function of time and BFA concentration. We show parameters for the 1σ elliptical contours of log-normal distributions in log-log space (natural logs), with green signals on the X-axis and red signals on the Y-axis (Fig. 4.7C). **(A)** Angle in radians of the uncorrelated frame of the ellipse, with respect to X-axis (θ). **(B)** Length of the major axis. **(C)** Ratio of the minor to major axes. There are at least 3 data points for every non-zero BFA concentration, and at least 10 datapoints for 0 μM BFA.

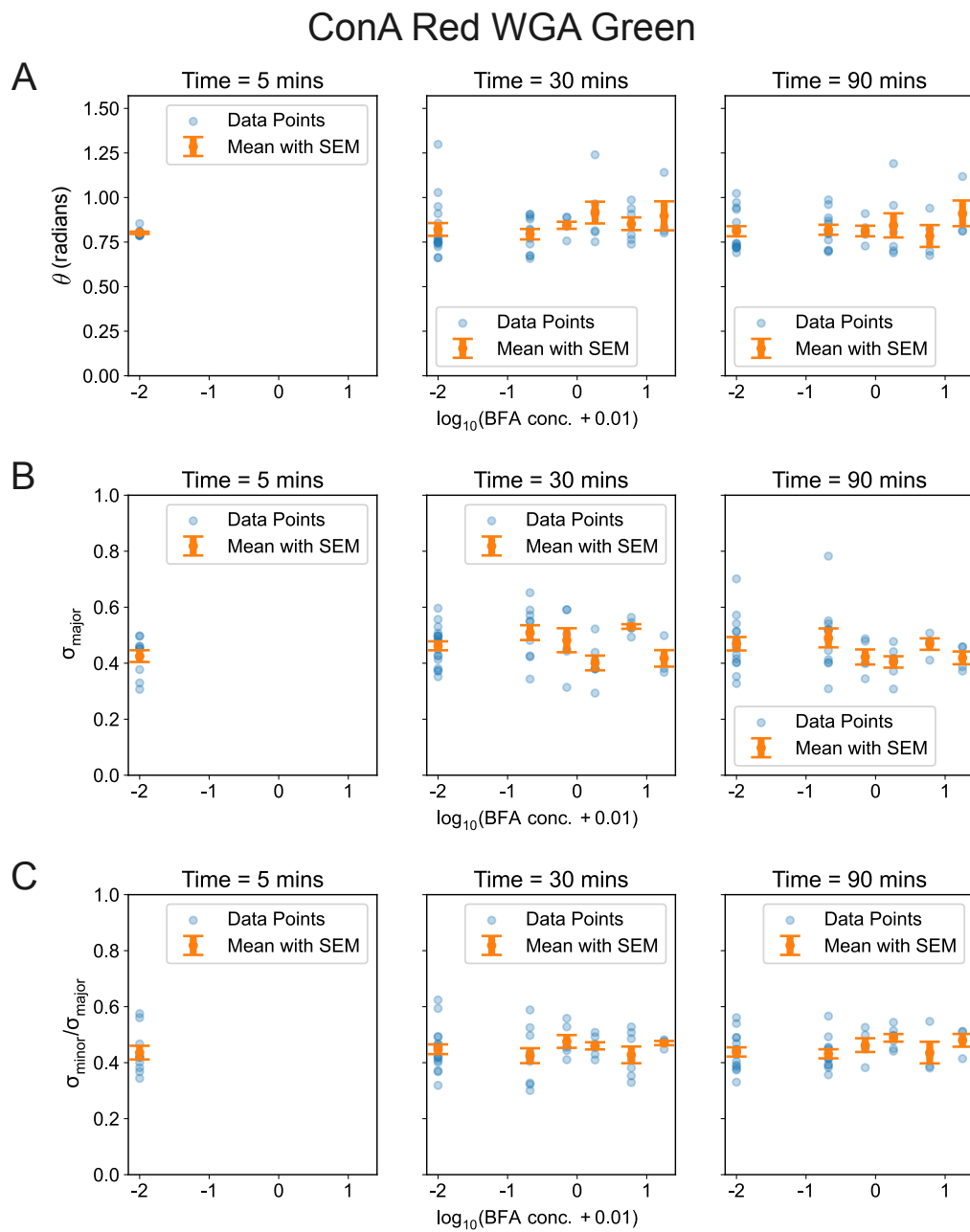


FIGURE 4.10: Covariances of the ConA-Red WGA-Green lectins as a function of time and BFA concentration. We show parameters for the 1σ elliptical contours of log-normal distributions in log-log space (natural logs), with green signals on the X-axis and red signals on the Y-axis (Fig. 4.7C). **(A)** Angle in radians of the uncorrelated frame of the ellipse, with respect to X-axis (θ). **(B)** Length of the major axis. **(C)** Ratio of the minor to major axes. There are at least 3 data points for every non-zero BFA concentration, and at least 10 datapoints for 0 μM BFA.

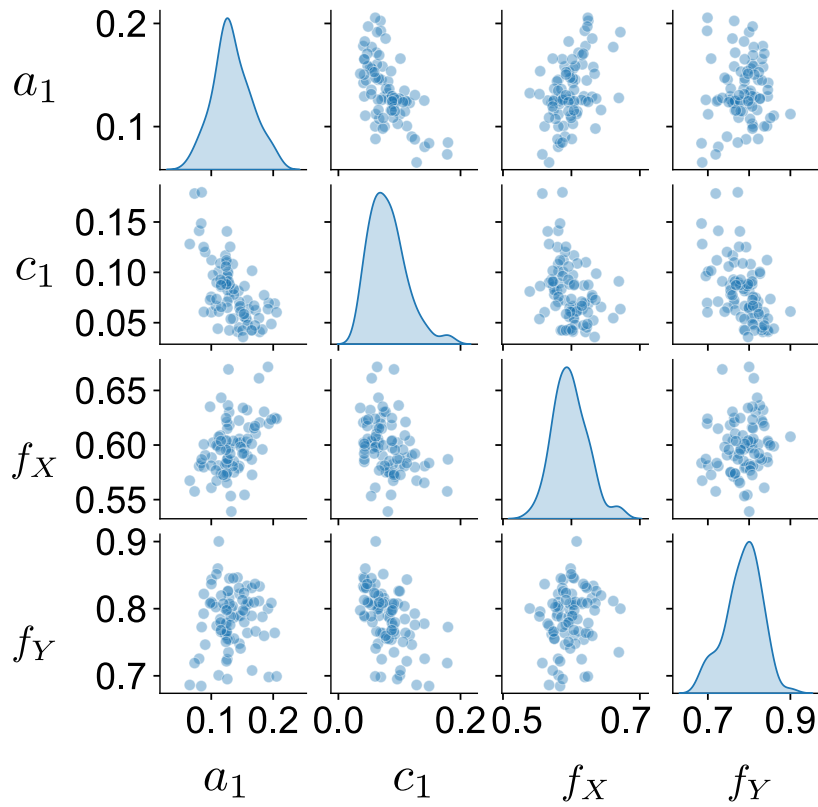


FIGURE 4.11: Bootstrap estimates of fitted parameter uncertainties from 100 bootstrap replicates for cell-surface glycan dynamics (outliers with $a_1 < 0.005$ were removed). For comparing with parameter values in Table 4.2, use: $f_X^{-1} = 1 + \alpha n_0 / \rho_X X_{\text{int}}$, $f_Y^{-1} = 1 + \alpha n_0 / \rho_Y Y_{\text{int}}$.

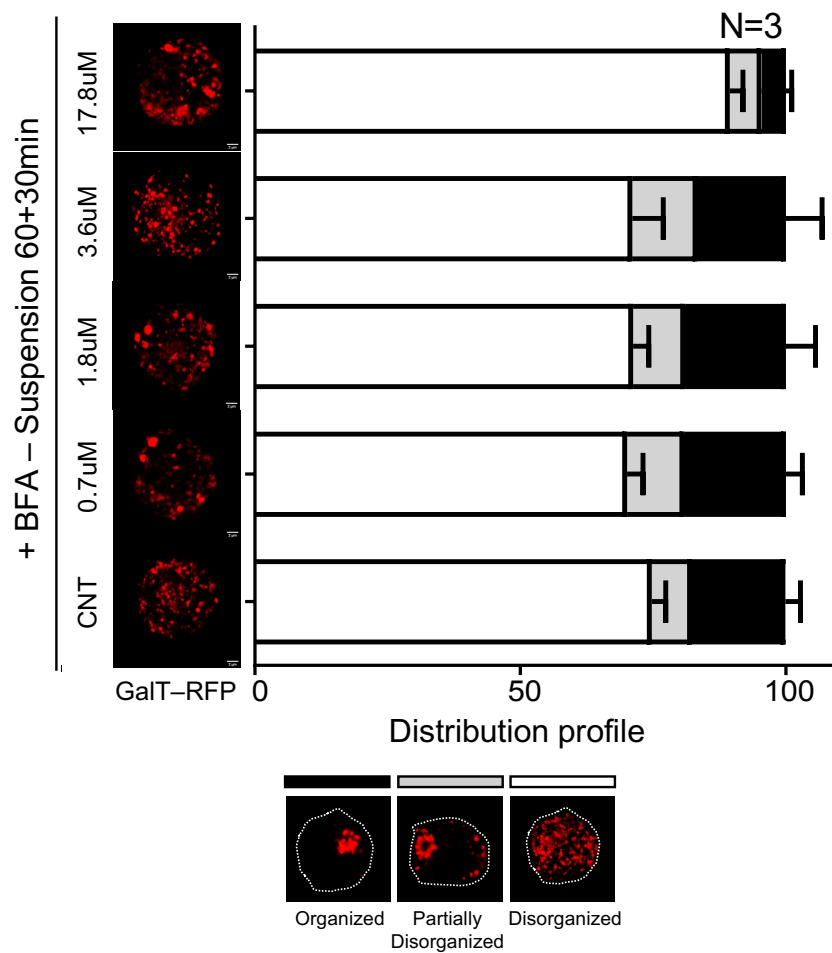


FIGURE 4.12: BFA does not appear to impact trans-Golgi morphology. We image the trans-Golgi in non-adherent WT-MEF cells using a fluorescently labeled marker (GalTase-RFP). Bottom: representative images showing the three trans-Golgi morphology classes. Top: Distribution of trans-Golgi morphologies in cell populations ($n > 100$) suspended for 60 min, then incubated for 30 min with varying concentrations of BFA. Error bars show the SEM over N replicate populations.

Chapter 5

Decoding cell-surface signals

According to our birth-death model, (Eq. 3.2) cis-Golgi units are gained or lost every $t_0/(2n_0) \sim 10$ min; in comparison, cell-surface glycans have a lifetime of over $1/\lambda_X = 30$ min before being endocytosed (Table 3.3). Thus, it is plausible that surface glycan levels in a single cell reflect the average number of functional cis-Golgi units in which these glycans were processed over the past 30 minutes. The stochasticity in the number of Golgi units explains why lectin signals form a unimodal distribution (Fig. 4.7): at a given BFA concentration, all cells in a population have similar time-averaged cis-Golgi levels, though they may have very different instantaneous cis-Golgi morphologies. Suppose cells could have been drawn from one of several populations, each characterized by a distinct value of n_{cis} (Fig. 5.2A). By measuring lectin levels on a single cell, can we infer which population it belongs to?

The optimal Bayesian decoder provides the best answer to this question, in the sense that it minimizes the probability of error [Bis06; Pet+19]. A fixed value of n_{cis} results in a probability distribution over the ConA and WGA signals. The population is inferred to be the one most likely to have generated the observed signal. We have previously shown that the signals form a bivariate log-normal distribution (Fig. 4.7). The likelihood then depends on the median ConA and WGA signals, and covariances of the signals.

In the last chapter, we determined how the median ConA and WGA signals vary with time and BFA concentrations, using the best-fit parameters of the synthesis/trafficking model. We also observed that the covariance matrix remains essentially the same across different conditions of time and BFA concentrations. We used the average CGWR covariance for the decoder analysis. We construct decoders in log-log space. From the behavior of the medians and covariances, we know the population distribution $p(g, r|c, t)$, where g and r represent the natural logs of the ConA and WGA signals.

5.1 Decoding to metapopulations

To set up the decoder, we start with an ordered series of cell populations, with decreasing levels of n_{cis} . These are generated according to predicted means (in log space) and covariances for 12 BFA concentrations, uniformly distributed in log space (these concentrations are sufficient to capture the full range of ConA and WGA variation, see Fig. 5.1). We pool these into N metapopulations, each corresponding to a contiguous range of BFA concentrations. N sets the resolution of the inference task (Fig. 5.2A, left). We assume the BFA concentration is chosen from a uniform prior across 12 values in μM : $\{0.0, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28, 2.56, 5.12, 10.24, 32.0\}$. We divide these into N equiprobable bins of size $12/N$, for $N = 2, 3, 4, 6, 12$. We then superpose distributions for all the concentrations in each bin to generate N metapopulation distributions. In this way, each metapopulation can be thought of as a probability distribution over the four morphology classes. Note that

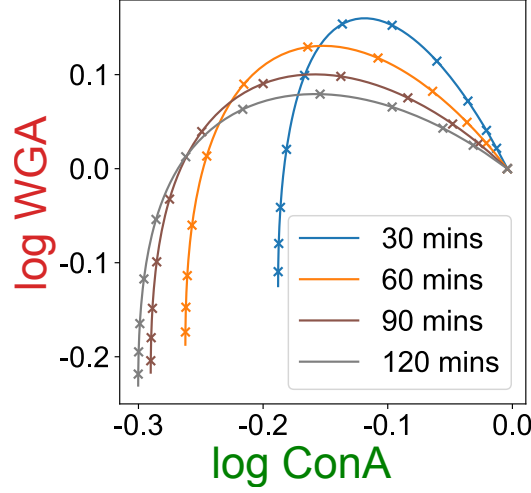


FIGURE 5.1: Predicted ConA and WGA signals as a function of time and BFA concentration in a 2D log-log plot (natural logs). The points marked 'x' correspond to a uniform grid of BFA concentrations in log space, used as a prior for decoding. This prior is chosen because it evenly samples the parabolic trajectories.

the decoder resolution is not related to the number of Golgi morphology classes, which is always four (Fig. 3.1A). The probability that meta-population B results in ConA signal g and WGA signal r is $p_B(g, r|t) = \frac{1}{|B|} \sum_{c \in B} p(g, r|c, t)$.

For optimal Bayesian decoding we must find the most likely metapopulation that generated a given g, r signal:

$$\hat{B}_{g,r} = \operatorname{argmax}_{B \in \mathcal{B}_N} p_B(g, r|t)$$

where $\mathcal{B}_N$ denotes the set of N metapopulations. The decoder splits the space of fluorescence values into N regions corresponding to different meta-populations. If a single cell falls into a certain region, our best guess is that it was drawn from the corresponding metapopulation. In Figs. 5.2A,B we show dual lectin decoders for the 30 min time point, for $N = 2, 3, 4, 6$; these decoders have a non-trivial polar structure due to the parabolic trajectory of the means in log space (Fig. 5.1). Interestingly, the decoders typically split into three large regions and several smaller regions, suggesting that the $N = 3$ decoder is the natural choice for this specific task. This means we can efficiently decode to three metapopulations, corresponding to high, intermediate, and low levels of n_{cis} .

We can determine the accuracy of the decoder by repeating the inference task over many cells (Fig. 5.2A, right). The accuracy of a decoder for meta-population B at time t is

$$\int_g \int_r p_B(g, r|t) \delta(B, \hat{B}_{g,r}) dr dg$$

where $\hat{B}_{g,r}$ is the inferred meta-population at signal g, r , and $\delta(B, \hat{B}_{g,r})$ returns 1 if its inputs match, 0 otherwise. To evaluate this integral numerically, we considered the range of g and r from -2 to 2 , which covered more than 99.999% of the signal probability distribution $p(g, r)$. The overall accuracy of the decoder is

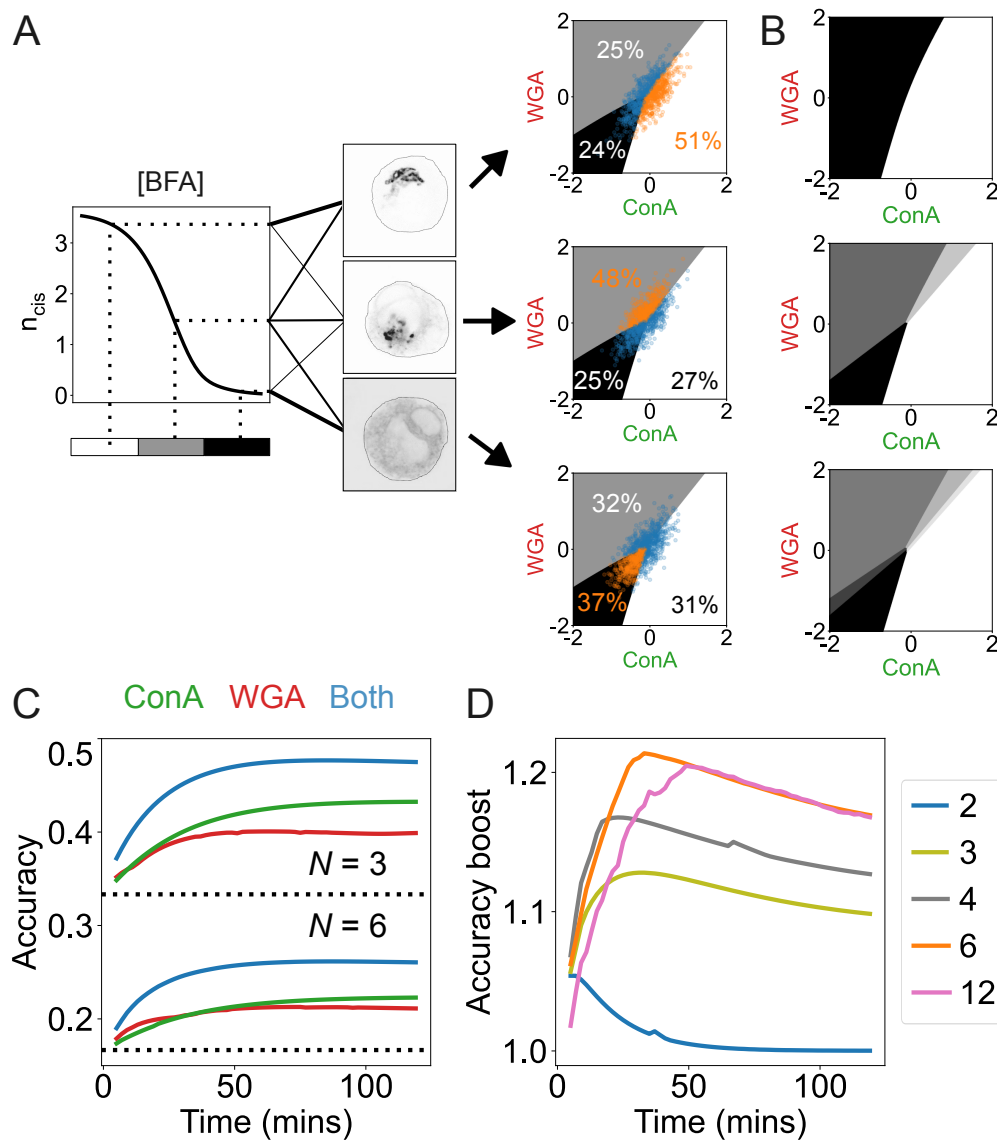


FIGURE 5.2: Decoding Golgi phenotypes from single-cell measurements. **(A)** Left: Populations at distinct BFA concentrations contain distinct mixtures of Golgi morphologies (illustrated by lines of varying thickness). We pool all populations corresponding to a range of BFA concentrations into $N = 3$ meta-populations (white, grey and black). Right: The $N = 3$ decoder for the 30 min time point, showing the regions of fluorescence space that map to each meta-population. The scatterplots show synthetic datasets of 1,000 cells generated according to the predicted meta-population distributions. Percentage values show the fraction of cells that lie in each region. Orange points show the cells that decode to the correct meta-population. Note that the decoders do not predict the Golgi morphology of a single cell, they predict that the cell was drawn from a meta-population with a specific mixture of Golgi morphologies. **(B)** Decoders for $N = 2, 4, 6$. All decoders are shown in the log-log space of ConA and WGA signals (natural logs). **(C)** Accuracy of decoding (averaged across N meta-populations) as a function of time, using only ConA, only WGA or both lectins together. Black dotted lines indicate the accuracy $1/N$ expected by chance. **(D)** The boost in accuracy from using two lectins, compared to using only ConA, for various values of N . The small spikes seen in Fig. 5.2(C) and (D) are an artifact of the discrete concentration prior.

$$\frac{1}{|N|} \sum_{B \in \mathcal{B}_N} \int_g \int_r p_B(g, r|t) \delta(B, \hat{B}_{g,r}) dr dg \quad (5.1)$$

We find that the decoders predict the correct metapopulation far better than chance (Fig. 5.2C).

5.2 Advantage when using more than one glycan type

The marginal of a bivariate (log-)normal distribution gives a univariate (log-)normal distribution [Pis14]. For decoding using a single glycan, we proceed exactly as described above, replacing the joint distribution $p(g, r|c, t)$ with the marginal distribution $p(g|c, t)$ or $p(r|c, t)$. Accuracies of the single-lectin decoder are also computed in the same manner as Equation 5.1, except that the prediction is calculated from red or green signals alone, as the case may be.

Decoders that integrate information from two lectins perform significantly better than those that use a single lectin (Fig. 5.2C). The performance of the decoders tends to increase with time since the addition of BFA. Dual-lectin decoders perform better than single-lectin decoders, with a boost in accuracy around the 30 min point (Fig. 5.2D). Using two lectins thus alleviates the speed-accuracy tradeoff: for $N = 3$, the accuracy achieved by a single-lectin decoder at 50 min is matched by that of a dual-lectin decoder at 15 min.

5.3 Information conveyed by cell-surface glycans

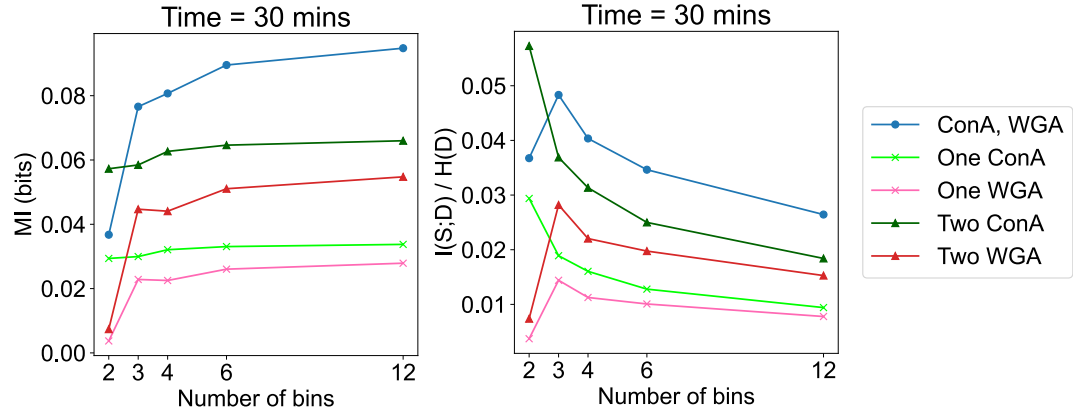


FIGURE 5.3: ConA-WGA combination conveys more information than other lectin combinations. (A) shows the mutual information between different combination of lectin signal and bins, for different number of bins. (B) shows the mutual information scaled by the number of bits required to fully specify a bin, under those number of bins.

The same conclusions as above – that multiple glycan types confer an advantage to the cell in terms of conveying information – can be established by using the concept of mutual information [CT06; Pet+19]. We use our framework to quantify the amount of information glycans convey about Golgi dysfunction. Let S be the random variable that denotes the (ConA, WGA) signals, and D be the random variable corresponding to the bins (D for dysfunction). Then the signal entropy at time t is given by

$$H(S) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(g, r|t) \log p(g, r|t) dr dg$$

where $p(g, r|t)$ is calculated as

$$p(g, r|t) = \frac{1}{N} \sum_{B \in \mathcal{B}_N} p_B(g, r|t).$$

The conditional entropy $H(S|D)$ is obtained as

$$H(S|D) = \frac{1}{|N|} \sum_{B \in \mathcal{B}_N} H(S|B) = - \frac{1}{|N|} \sum_{B \in \mathcal{B}_N} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_B(g, r|t) \log p_B(g, r|t) dr dg$$

Then the mutual information between S and D can be computed as $I(S; D) = H(S) - H(S|D)$. We calculated the mutual information conveyed by five combinations of lectins: both ConA and WGA, two ConA signals, two WGA signals, one ConA signal, and one WGA signal. Note that obtaining two ConA (or two WGA) signals is the same as obtaining a single ConA (or WGA) signal from a distribution with half the variance as the original.

Fig. 5.3A plots the mutual information under various numbers of bins. Since mutual information is calculated under different numbers of bins, we also plot mutual information scaled by $H(D)$, the number of bits required to specify a bin (Fig. 5.3B). If the information conveyed by the signal suffices to identify the bin, this quantity would be equal to 1. Note that plotted in this manner, there is a peak at $N = 3$ bins, the number of bins we designated as a natural choice for the decoding task in the previous section.

Next, we plot how mutual information varies as a function of time for a fixed number of bins. We see that in most cases, for around the first 50 minutes, the combination of ConA and WGA provide more information than any other combination of lectins. This is especially relevant in the context of entities like glycans that have to dissipate information in a timely manner. Beyond the 50-minute mark, obtaining two ConA signals conveys slightly more information.

The alleviation of the speed-accuracy trade-off observed when using multiple glycan types also holds when mutual information is the quantity used to assess performance; one can see several instances supporting this point in Fig. 5.4.

5.4 Discussion

The data processing inequality [CT06] says that processing cannot increase information. If X, Y and Z were random variables with a causal structure $X \rightarrow Y \rightarrow Z$, then $I(X; Y) \geq I(X; Z)$. Given that the WGA substrate is a processed version of the ConA substrate, it is reasonable to ask how the WGA substrate conveys any information about BFA perturbation at all.

The conversion rate from the ConA substrate to the WGA substrate at the trans-Golgi is dictated by the concentration of BFA, so the (relative) WGA amounts *do* contain information about BFA concentration. The mutual information between (ConA, WGA) signals and Golgi dysfunction peaks out at 0.15 bits.

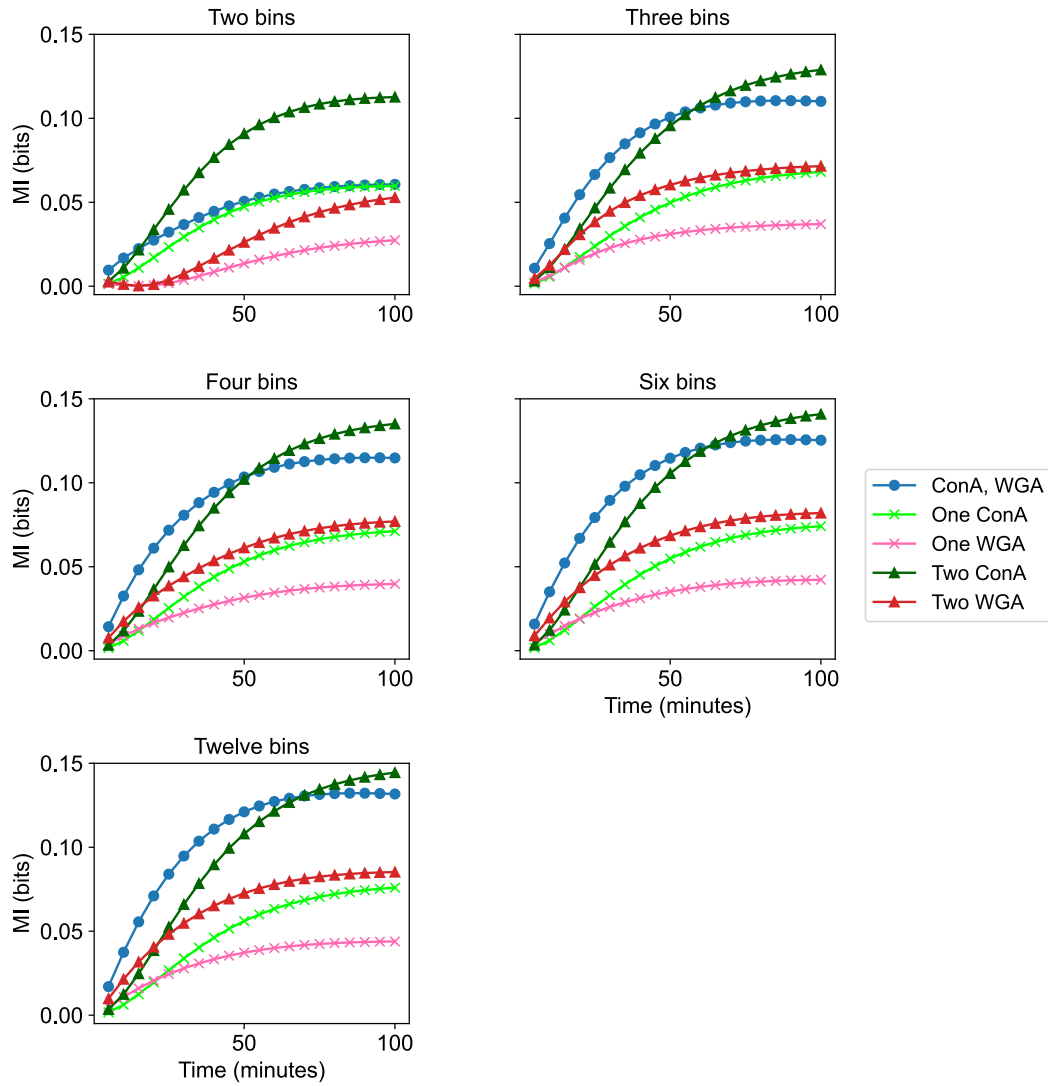


FIGURE 5.4: ConA-WGA combination conveys more accurate information faster than other lectin combinations. Figure shows mutual information as a function of time, for different number of bins. In most cases, for around the first 50 minutes, ConA-WGA combination conveys more information than other lectin combinations, including multiple ConA or WGA signals.

The fact that a noisy WGA signal can convey more information than a second ConA signal (which would have made the overall ConA signal less noisy) is interesting. We can quantify how beneficial both signals together are by calculating the difference in mutual information when two measurements are used simultaneously and the sum of mutual information of individual measurements. This quantity $I(D; (ConA, WGA)) - I(D; ConA) - I(D; WGA)$ is called *synergy* between measurements [HWM23]; in our case, the maximum synergy is around 0.03 bits.

There could be many reasons for the low mutual information between (ConA, WGA) and Golgi dysfunction. Firstly, we only measured two lectins! There can be hundreds of glycan types on the cell-surface, augmenting the information-conveying ability of cells. Secondly, errors can creep in from several directions: glycosylation level, technical noise, experimental noise, and so on. Third, the Golgi system is not designed to convey information about disruption due to BFA!

We can also see why two ConA signals convey more information about the bins than ConA and WGA signals, in some cases. Recall that the mean behavior (in log space) of ConA and WGA follows a parabolic trajectory. At lower concentrations, the ConA signal drops steeply making concentrations distinguishable, and a reduction in noise in this regime is beneficial to identify the perturbation using the ConA signal. Higher concentrations produce roughly the same ConA signal, making it less useful while decoding higher concentrations. In the interest of uniformly sampling the mean curve, we may have inadvertently included more "low" concentrations – the regime ConA performs better – which might be why we see ConA performing better in some cases.

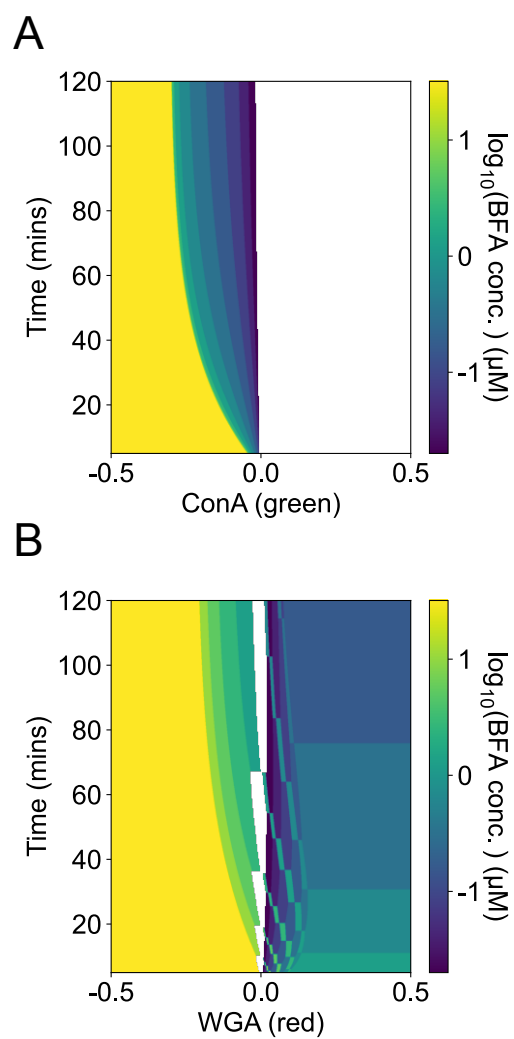


FIGURE 5.5: Single-lectin decoders for $N = 12$ populations, across time. (A) Decoder using ConA alone. (B) Decoder using WGA alone. Colors show the best-guess BFA concentration value at each level of green or red. White represents 0 μM BFA.

Chapter 6

Outlook

To recap, we have demonstrated the following:

- The space of Pareto-optimal solutions is non-trivial even for well-understood classes of functions like point-wise convex functions when discrete controls are involved (Chapter 2).
- Dynamics of cis-Golgi, endocytosis, and transport regulate cell-surface glycan distribution (Chapters 3, 4).
- Glycans convey information far better than chance, and a combination of lectins is better than an equivalent amount of fewer lectins (Chapter 5).

There remains much to be done, of course.

6.1 Probing glycan manufacture

The central question in the area remains to devise a model of the Golgi that is able to explain all observations as opposed to a large subset, which is what the current models do. Quantitative data on intra-Golgi transport [Luj+23; Tie+24] will greatly aid in this process. As we push the boundaries of experimental techniques and generate more fine-grained data, newer models of Golgi and glycan manufacture will need to be developed concurrently, refining our understanding of both.

There were aspects left understudied in the enzyme compartmentalization project due to lack of physiological relevance, which may be of interest in the abstract setting. We observed that in some settings, cell type having Golgi with slightly different enzyme distributions has a higher "average yield". This phenomenon, called bet hedging, has been widely studied in other contexts [Gro+23; Lev62] but has not been explored in a setting similar to ours.

An interesting alternative framing would be to explicitly impose a global time budget for Golgi transit and ask how close physiological enzyme distributions lie to a speed–error frontier, in analogy with recent work that characterizes optimal speed–accuracy–dissipation trade-offs in other biochemical systems [CMP23; RP15]. Although we do not pursue this here, our kinetic and Pareto framework is directly compatible with such an analysis.

6.2 Probing glycan function

Several people have pointed out that amusingly, there is a disconnect between scientists who study Golgi organization and scientists who study glycan function. Our work is an

attempt to bridge the gap between these two aspects. We have shown that glycans convey nontrivial amounts of information about Golgi organization, and how such information might be encoded and decoded.

Although this was demonstrated in an artificial setting of disruption due to BFA, the skeleton we developed is widely applicable with some tweaks. For starters, the Golgi fragments during cell division [TW13] altering the glycan distribution; in theory, this means outside entities should be able to infer the stage of division of the cell using cell-surface glycans. Methods to reliably measure Golgi function and dysfunction in single cells could in principle be used to develop disease biomarkers [Yar+20; HU22]. Cancer cells with distinctly altered Golgi organization [Pet15; MFS22; Mar+23] and glycosylation signatures [RB15; Zha21; Kho+22] could be direct candidates to test this hypothesis.

In our work, we have focused on how glycan information gets decoded, *assuming an optimal decoder*. At the cellular level, many factors might prevent glycan information from being optimally decoded. The cell surface decoding could be studied in the context of a realistic decoding mechanism. As a starting point, we note that receptors capable of measuring log-fold changes in the concentration of ConA- and WGA-binding glycans, with a linear activation mechanism, will be able to distinguish low and high perturbations. Adding more such receptors will allow for more fine-grained decoding of cell state, and even opens up the possibility of decoding *time* since a perturbation has taken place.

We also haven't touched upon the subject of information *encoding*; we have taken the ConA-WGA encoding to be a given. It will be interesting to study how dysfunction *should* have been encoded to allow for the highest chances of information recovery.

It's also worthwhile to reflect the role of mathematical models in our analysis. They play two distinct roles. At one level, they are merely curve-fitting tools that allow us to interpolate between and compress a large amount of experimental data into a few parameter values. Our results on decoder accuracy only depend on the analytic curves being good representations of the data, regardless of the ontological content of the underlying models. At another level, the fact that we can parsimoniously explain the observations using just a handful of parameters (Tables 3.2, 3.4 and 4.2) suggests that the model does capture aspects of the underlying cell-biological processes.

6.3 Concluding remarks

There is a lot to be answered along the following axes:

- **Speed.** Given that responses cannot take arbitrarily long, glycan-binding proteins must somehow ensure that glycans are "read" within the requisite time [Den15]. How is this speed achieved? (We have shown that information can be *conveyed*, the question is about how the information might be *read*.)
- **Error.** Given that glycosylation is stochastic, there must be some margin of error that is tolerated by the reading entity. What is this margin of error? Is it that the errors in some glycans are more detrimental than errors in other glycans? Or are all glycans created equal? In the former scenario, how can one rigorously think about the margin of error that is tolerated?
- **Diversity.** What subset of the glycan diversity actually gets "read" by different entities? What are the information-conveying capacities of such combinations? Does

the information-conveying capacity correlate with the known functions of such interactions? Existing works [Son+20] and unpublished work from Satyajit Mayor's lab provide details about the density of different types of glycans on the surface. This data can potentially answer whether cells that are associated with conveying information have sparse but many different types of glycans, or dense but similar kinds of molecules on the cell surface.

- **Cost.** There are several interesting questions to probe in this regard. Different cell types have different glycan distributions generated owing to (1) different enzymes (2) different distributions of the same enzymes; what is the cost to maintaining records of such enzymes distributions? Does this cost place limits on how many different cell types there can be? Do similar or related cell types have similar enzyme distributions? How close are the enzyme distributions to each other? An understanding of "enzyme distribution lineage" is virtually nonexistent.

In other words, there's much to look forward to.

Bibliography

- [ABH13] Haris Aziz, Felix Brandt, and Paul Harrenstein. “Pareto optimality in coalition formation”. In: *Games and Economic Behavior* 82 (2013), pp. 562–581. DOI: [10.1016/j.geb.2013.08.006](https://doi.org/10.1016/j.geb.2013.08.006).
- [Alb+22] Bruce Alberts et al. *Molecular biology of the cell*. W. W. Norton & Company, 2022.
- [Apw99] R Apweiler. “On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1473 (1 Dec. 1999), pp. 4–8. ISSN: 03044165. DOI: [10.1016/S0304-4165\(99\)00165-8](https://doi.org/10.1016/S0304-4165(99)00165-8).
- [AWA98] D.J. Austin, N.J. White, and R.M. Anderson. “The Dynamics of Drug Action on the Within-host Population Growth of Infectious Agents: Melding Pharmacokinetics with Pathogen Population Dynamics”. In: *Journal of Theoretical Biology* 194 (3 Oct. 1998), pp. 313–339. ISSN: 00225193. DOI: [10.1006/jtbi.1997.0438](https://doi.org/10.1006/jtbi.1997.0438).
- [Bek+17] Michael E. Bekier et al. “Knockout of the Golgi stacking proteins GRASP55 and GRASP65 impairs Golgi structure and function”. In: *Molecular Biology of the Cell* 28 (21 Oct. 2017), pp. 2833–2842. ISSN: 1059-1524. DOI: [10.1091/mbc.e17-02-0112](https://doi.org/10.1091/mbc.e17-02-0112).
- [Ben+24] Giovanna Benvenuto et al. “Evolution of the ribbon-like organization of the Golgi apparatus in animal cells”. In: *Cell Reports* 43 (3 Mar. 2024), p. 113791. ISSN: 22111247. DOI: [10.1016/j.celrep.2024.113791](https://doi.org/10.1016/j.celrep.2024.113791).
- [BG04] Juan S Bonifacino and Benjamin S Glick. “The mechanisms of vesicle budding and fusion.” In: *Cell* 116 (2 Jan. 2004), pp. 153–66. ISSN: 0092-8674. DOI: [10.1016/s0092-8674\(03\)01079-1](https://doi.org/10.1016/s0092-8674(03)01079-1).
- [Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN: 0387310738.
- [BS05] William Bialek and Sima Setayeshgar. “Physical limits to biochemical signaling”. In: *Proceedings of the National Academy of Sciences* 102.29 (July 2005), pp. 10040–10045. ISSN: 1091-6490. DOI: [10.1073/pnas.0504321102](https://doi.org/10.1073/pnas.0504321102). URL: <http://dx.doi.org/10.1073/pnas.0504321102>.
- [BT20] Ansuman Biswas and Mukund Thattai. “Promiscuity and specificity of eukaryotic glycosyltransferases”. In: *Biochemical Society Transactions* (June 2020). DOI: [10.1042/bst20190651](https://doi.org/10.1042/bst20190651). URL: <https://doi.org/10.1042/bst20190651>.
- [BV11] Stephen P. Boyd and Lieven Vandenbergh. *Convex optimization*. Cambridge Univ. Pr., 2011.
- [CC07] Paul F. Cook and W.W. Cleland. *Enzyme Kinetics and Mechanism*. Garland Science, Mar. 2007. ISBN: 9780203833575. DOI: [10.4324/9780203833575](https://doi.org/10.4324/9780203833575).

- [CCR19] Kevin Brown Chandler, Catherine E Costello, and Nader Rahimi. “Glycosylation in the Tumor Microenvironment: Implications for Tumor Angiogenesis and Metastasis.” In: *Cells* 8 (6 June 2019). ISSN: 2073-4409. DOI: [10.3390/cells8060544](https://doi.org/10.3390/cells8060544).
- [Cho+16] Gerardo Chowell et al. “Mathematical models to characterize early epidemic growth: A review”. In: *Physics of Life Reviews* 18 (Sept. 2016), pp. 66–97. ISSN: 15710645. DOI: [10.1016/j.plrev.2016.07.005](https://doi.org/10.1016/j.plrev.2016.07.005).
- [CMP23] Davide Chiuchiu, Shrabani Mondal, and Simone Pigolotti. “Pareto optimal fronts of kinetic proofreading”. In: *New Journal of Physics* 25 (4 Apr. 2023), p. 043007. ISSN: 1367-2630. DOI: [10.1088/1367-2630/acc757](https://doi.org/10.1088/1367-2630/acc757).
- [CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006. ISBN: 0471241954.
- [Dam+91] H. Damke et al. “Effects of Brefeldin A on the endocytic route. Redistribution of mannose 6-phosphate/insulin-like growth factor II receptors to the cell surface.” In: *Journal of Biological Chemistry* 266 (36 Dec. 1991), pp. 24829–24833. ISSN: 00219258. DOI: [10.1016/S0021-9258\(18\)54303-9](https://doi.org/10.1016/S0021-9258(18)54303-9).
- [Deb99] Kalyanmoy Deb. “Evolutionary Algorithms for Multi-Criterion Optimization in Engineering Design”. In: 1999.
- [Den15] James W. Dennis. “Many Light Touches Convey the Message”. In: *Trends in Biochemical Sciences* 40 (11 Nov. 2015), pp. 673–686. ISSN: 09680004. DOI: [10.1016/j.tibs.2015.08.010](https://doi.org/10.1016/j.tibs.2015.08.010).
- [DSG13] Kasey J. Day, L. Andrew Staehelin, and Benjamin S. Glick. “A three-stage model of Golgi structure and function”. In: *Histochemistry and Cell Biology* 140 (3 Sept. 2013), pp. 239–249. ISSN: 0948-6143. DOI: [10.1007/s00418-013-1128-3](https://doi.org/10.1007/s00418-013-1128-3).
- [ED18] Michael T. M. Emmerich and André H. Deutz. “A tutorial on multiobjective optimization: fundamentals and evolutionary methods”. In: *Natural Computing* 17 (3 Sept. 2018), pp. 585–609. ISSN: 1567-7818. DOI: [10.1007/s11047-018-9685-y](https://doi.org/10.1007/s11047-018-9685-y).
- [Fis+19a] Peter Fisher et al. “Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation”. In: *Cell Reports* 27.4 (2019), 1231–1243.e6. DOI: [10.1016/j.celrep.2019.03.107](https://doi.org/10.1016/j.celrep.2019.03.107).
- [Fis+19b] Peter Fisher et al. “Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation”. In: *Cell Reports* 27 (4 Apr. 2019), 1231–1243.e6. ISSN: 22111247. DOI: [10.1016/j.celrep.2019.03.107](https://doi.org/10.1016/j.celrep.2019.03.107).
- [FKK25] Konstantinos Flevaris, Pavlos Kotidis, and Cleo Kontoravdi. “GlyCompute: towards the automated analysis of protein N-linked glycosylation kinetics via an open-source computational framework”. In: *Analytical and Bioanalytical Chemistry* 417 (5 Feb. 2025), pp. 957–972. ISSN: 1618-2642. DOI: [10.1007/s00216-024-05522-3](https://doi.org/10.1007/s00216-024-05522-3).
- [FN11] H. H. Freeze and B. G. Ng. “Golgi Glycosylation and Human Inherited Diseases”. In: *Cold Spring Harbor Perspectives in Biology* 3.9 (2011), a005371–a005371. DOI: [10.1101/cshperspect.a005371](https://doi.org/10.1101/cshperspect.a005371).

- [Fu+05] Dongfeng Fu et al. "The growth of business firms: Theoretical framework and empirical evidence". In: *Proceedings of the National Academy of Sciences* 102 (52 Dec. 2005), pp. 18801–18806. ISSN: 0027-8424. DOI: [10.1073/pnas.0509543102](https://doi.org/10.1073/pnas.0509543102).
- [Fuc+23] Felix F Fuchsberger et al. "Information transfer in mammalian glycan-based communication". In: *eLife* 12 (Feb. 2023). ISSN: 2050-084X. DOI: [10.7554/eLife.69415](https://doi.org/10.7554/eLife.69415).
- [Gil87] Alfred G. Gilman. "G PROTEINS: TRANSDUCERS OF RECEPTOR-GENERATED SIGNALS". In: *Annual Review of Biochemistry* 56 (1 June 1987), pp. 615–649. ISSN: 0066-4154. DOI: [10.1146/annurev.bi.56.070187.003151](https://doi.org/10.1146/annurev.bi.56.070187.003151).
- [GL11] Benjamin S Glick and Alberto Luini. "Models for Golgi traffic: a critical assessment." In: *Cold Spring Harbor perspectives in biology* 3 (11 Nov. 2011), a005215. ISSN: 1943-0264. DOI: [10.1101/cshperspect.a005215](https://doi.org/10.1101/cshperspect.a005215).
- [Gro+23] Daan H. de Groot et al. "Effective bet-hedging through growth rate dependent stability". In: *Proceedings of the National Academy of Sciences* 120 (8 Feb. 2023). ISSN: 0027-8424. DOI: [10.1073/pnas.2211091120](https://doi.org/10.1073/pnas.2211091120).
- [Gut+20] Jahir M. Gutierrez et al. "Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion". In: *Nature Communications* 11.1 (2020). DOI: [10.1038/s41467-019-13867-y](https://doi.org/10.1038/s41467-019-13867-y).
- [Har+24] Akihiro Harada et al. "Dynamic movement of the Golgi unit and its glycosylation enzyme zones". In: *Nature Communications* 15 (1 May 2024), p. 4514. ISSN: 2041-1723. DOI: [10.1038/s41467-024-48901-1](https://doi.org/10.1038/s41467-024-48901-1).
- [Hir+98] Koret Hirschberg et al. "Kinetic analysis of secretory protein traffic and characterization of Golgi to plasma membrane transport intermediates in living cells". In: *The Journal of Cell Biology* 143.6 (1998), pp. 1485–1503. DOI: [10.1083/jcb.143.6.1485](https://doi.org/10.1083/jcb.143.6.1485).
- [HMH07] Patrick Hossler, Bhanu Chandra Mulukutla, and Wei-Shou Hu. "Systems Analysis of N-Glycan Processing in Mammalian Cells". In: *PLoS ONE* 2.8 (2007). Ed. by Mariko Hatakeyama, e713. DOI: [10.1371/journal.pone.0000713](https://doi.org/10.1371/journal.pone.0000713).
- [HN11] Alex H. Hutagalung and Peter J. Novick. "Role of Rab GTPases in Membrane Traffic and Cell Physiology". In: *Physiological Reviews* 91 (1 Jan. 2011), pp. 119–149. ISSN: 0031-9333. DOI: [10.1152/physrev.00059.2009](https://doi.org/10.1152/physrev.00059.2009).
- [HU22] Yoshimi Haga and Koji Ueda. "Glycosylation in cancer: its application as a biomarker and recent advances of analytical techniques". In: *Glycoconjugate Journal* 39.2 (Feb. 2022), pp. 303–313. ISSN: 1573-4986. DOI: [10.1007/s10719-022-10043-1](https://doi.org/10.1007/s10719-022-10043-1). URL: <http://dx.doi.org/10.1007/s10719-022-10043-1>.
- [Hut+83] C. Richard Hutchinson et al. "Comparative biochemistry of fatty acid and macrolide antibiotic (brefeldin A). Formation in penicillium brefeldianum". In: *Tetrahedron* 39 (21 Jan. 1983), pp. 3507–3513. ISSN: 00404020. DOI: [10.1016/S0040-4020\(01\)88660-9](https://doi.org/10.1016/S0040-4020(01)88660-9).
- [HWM23] Lauritz Hahn, Aleksandra M. Walczak, and Thierry Mora. "Dynamical Information Synergy in Biochemical Signaling Networks". In: *Physical Review Letters* 131 (12 Sept. 2023), p. 128401. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.131.128401](https://doi.org/10.1103/PhysRevLett.131.128401).
- [Ili13] Andrii Iliencko. "Continuous counterparts of Poisson and binomial distributions and their properties". In: *Annales Univ. Sci. Budapest., Sect. Comp* 39 (2013), pp. 137–147. DOI: <https://doi.org/10.48550/arXiv.1303.5990>.

- [Ise+25] Florin N. Isenrich et al. “Microfluidic mimicry of the Golgi-linked N -glycosylation machinery”. In: *Lab on a Chip* 25 (8 2025), pp. 1907–1917. ISSN: 1473-0197. DOI: [10.1039/D5LC00005J](https://doi.org/10.1039/D5LC00005J).
- [Ito+12] Yoko Ito et al. “cis-Golgi proteins accumulate near the ER exit sites and act as the scaffold for Golgi regeneration after brefeldin A treatment in tobacco BY-2 cells.” In: *Molecular biology of the cell* 23 (16 Aug. 2012), pp. 3203–14. ISSN: 1939-4586. DOI: [10.1091/mbc.E12-01-0034](https://doi.org/10.1091/mbc.E12-01-0034).
- [IUP14] IUPAC. “glycans”. In: International Union of Pure and Applied Chemistry (IUPAC), Feb. 2014. DOI: [10.1351/goldbook.G02645](https://doi.org/10.1351/goldbook.G02645).
- [JBC22] Peter Jedlicka, Alexander D. Bird, and Hermann Cuntz. “Pareto optimality, economy–effectiveness trade-offs and ion channel degeneracy: Improving population modelling for single neurons”. In: *Open Biology* 12.7 (2022). DOI: [10.1098/rsob.220073](https://doi.org/10.1098/rsob.220073).
- [Jef09] Roy Jefferis. “Glycosylation as a strategy to improve antibody-based therapeutics”. In: *Nature Reviews Drug Discovery* 8.3 (2009), pp. 226–234. DOI: [10.1038/nrd2804](https://doi.org/10.1038/nrd2804).
- [JG11] Kenneth A. Johnson and Roger S. Goody. “The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper”. In: *Biochemistry* 50 (39 Oct. 2011), pp. 8264–8269. ISSN: 0006-2960. DOI: [10.1021/bi201284u](https://doi.org/10.1021/bi201284u).
- [Jon+16] Sanne E. de Jong et al. “IgG1 Fc N-glycan galactosylation as a biomarker for immune activation.” In: *Nature: Scientific Reports* 6 (2016).
- [Jos+24] Prachi Joshi et al. “Cell-surface glycans are quantitative reporters of Golgi dysfunction in single cells”. In: *bioRxiv* (2024). DOI: [10.1101/2024.06.17.599374](https://doi.org/10.1101/2024.06.17.599374). eprint: <https://www.biorxiv.org/content/early/2024/06/18/2024.06.17.599374.full.pdf>. URL: <https://www.biorxiv.org/content/early/2024/06/18/2024.06.17.599374>.
- [JT20] Anjali Jaiman and Mukund Thattai. “Golgi compartments enable controlled biomolecular assembly using promiscuous enzymes”. In: *eLife* 9 (2020). DOI: [10.7554/elife.49573](https://doi.org/10.7554/elife.49573).
- [Kal94] Vladimir V. Kalashnikov. *Mathematical Methods in Queuing Theory*. Springer Netherlands, 1994. ISBN: 978-90-481-4339-9. DOI: [10.1007/978-94-017-2197-4](https://doi.org/10.1007/978-94-017-2197-4).
- [KB05] Frederick J. Krambeck and Michael J. Betenbaugh. “A mathematical model of N-linked glycosylation”. In: *Biotechnology and Bioengineering* 92 (6 Dec. 2005), pp. 711–728. ISSN: 0006-3592. DOI: [10.1002/bit.20645](https://doi.org/10.1002/bit.20645).
- [KDL92] R D Klausner, J G Donaldson, and J Lippincott-Schwartz. “Brefeldin A: insights into the control of membrane traffic and organelle structure.” In: *The Journal of cell biology* 116 (5 Mar. 1992), pp. 1071–1080. ISSN: 0021-9525. DOI: [10.1083/jcb.116.5.1071](https://doi.org/10.1083/jcb.116.5.1071).
- [Kho+22] Elham Khosrowabadi et al. “Altered glycosylation of several metastasis-associated glycoproteins with terminal GalNAc defines the highly invasive cancer cell phenotype”. In: *Oncotarget* 13.1 (Jan. 2022), pp. 73–89. ISSN: 1949-2553. DOI: [10.18632/oncotarget.28167](https://doi.org/10.18632/oncotarget.28167). URL: <http://dx.doi.org/10.18632/oncotarget.28167>.
- [Kot+23] Pavlos Kotidis et al. “CHOGlycoNET: Comprehensive glycosylation reaction network for CHO cells”. In: *Metabolic Engineering* 76 (Mar. 2023), pp. 87–96. ISSN: 10967176. DOI: [10.1016/j.ymben.2022.12.009](https://doi.org/10.1016/j.ymben.2022.12.009).

- [Kra+09] Frederick J Krambeck et al. “A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data”. In: *Glycobiology* 19 (11 Nov. 2009), pp. 1163–1175. ISSN: 1460-2423. DOI: [10.1093/glycob/cwp081](https://doi.org/10.1093/glycob/cwp081).
- [Kra+17] Frederick J. Krambeck et al. “Model-based analysis of N-glycosylation in Chinese hamster ovary cells”. In: *PLOS ONE* 12 (5 May 2017), e0175376. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0175376](https://doi.org/10.1371/journal.pone.0175376).
- [LD08] K. S Lau and J. W Dennis. “N-Glycans in cancer progression”. In: *Glycobiology* 18 (10 June 2008), pp. 750–760. ISSN: 0959-6658. DOI: [10.1093/glycob/cwn071](https://doi.org/10.1093/glycob/cwn071).
- [Lev62] Richard Levins. “Theory of fitness in a heterogeneous environment. i. the fitness set and adaptive function”. In: *The American Naturalist* 96.891 (1962), pp. 361–373. DOI: [10.1086/282245](https://doi.org/10.1086/282245).
- [Lip+89] Jennifer Lippincott-Schwartz et al. “Rapid redistribution of Golgi proteins into the ER in cells treated with brefeldin A: Evidence for membrane cycling from Golgi to ER”. In: *Cell* 56 (5 Mar. 1989), pp. 801–813. ISSN: 00928674. DOI: [10.1016/0092-8674\(89\)90685-5](https://doi.org/10.1016/0092-8674(89)90685-5).
- [Liu+17] Ting Liu et al. “NF-κB signaling in inflammation”. In: *Signal Transduction and Targeted Therapy* 2 (1 July 2017), p. 17023. ISSN: 2059-3635. DOI: [10.1038/sigtrans.2017.23](https://doi.org/10.1038/sigtrans.2017.23).
- [Llo+96] Kenneth O. Lloyd et al. “Comparison of O-Linked Carbohydrate Chains in MUC-1 Mucin from Normal Breast Epithelial Cell Lines and Breast Carcinoma Cell Lines”. In: *Journal of Biological Chemistry* 271.52 (1996), pp. 33325–33334. DOI: [10.1074/jbc.271.52.33325](https://doi.org/10.1074/jbc.271.52.33325).
- [LN08] Gang Liu and Sriram Neelamegham. “In silico Biochemical Reaction Network Analysis (IBRENA): a package for simulation and analysis of reaction networks.” In: *Bioinformatics (Oxford, England)* 24 (8 Apr. 2008), pp. 1109–11. ISSN: 1367-4811. DOI: [10.1093/bioinformatics/btn061](https://doi.org/10.1093/bioinformatics/btn061).
- [LR13] R G Lichtenstein and G A Rabinovich. “Glycobiology of cell death: when glycans and lectins govern cell fate.” In: *Cell death and differentiation* 20 (8 Aug. 2013), pp. 976–86. ISSN: 1476-5403. DOI: [10.1038/cdd.2013.50](https://doi.org/10.1038/cdd.2013.50).
- [Luj+23] Pablo Lujan et al. *Sorting of secretory proteins at the trans-Golgi network by TGN46*. Oct. 2023. DOI: [10.7554/eLife.91708.1](https://doi.org/10.7554/eLife.91708.1).
- [Mac02] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. USA: Cambridge University Press, 2002. ISBN: 0521642981.
- [MAG97] Thomas J. Monica, Dana C. Andersen, and Charles F. Goochee. “A mathematical model of sialylation of N-linked oligosaccharides in the trans-Golgi network”. In: *Glycobiology* 7 (4 1997), pp. 515–521. ISSN: 0959-6658. DOI: [10.1093/glycob/7.4.515](https://doi.org/10.1093/glycob/7.4.515).
- [Mak+24] Elli Makrydaki et al. “Immobilized enzyme cascade for targeted glycosylation”. In: *Nature Chemical Biology* 20 (6 June 2024), pp. 732–741. ISSN: 1552-4450. DOI: [10.1038/s41589-023-01539-4](https://doi.org/10.1038/s41589-023-01539-4).
- [Man94] Olvi L. Mangasarian. *Nonlinear Programming*. Society for Industrial and Applied Mathematics, Jan. 1994. ISBN: 978-0-89871-341-1. DOI: [10.1137/1.9781611971255](https://doi.org/10.1137/1.9781611971255).

- [Mar+23] Marta Martins et al. “The Golgi Apparatus as an Anticancer Therapeutic Target”. In: *Biology* 13.1 (Dec. 2023), p. 1. ISSN: 2079-7737. DOI: [10.3390/biology13010001](https://doi.org/10.3390/biology13010001). URL: <http://dx.doi.org/10.3390/biology13010001>.
- [ME19] Sebastian Montealegre and Peter M. van Endert. “Endocytic Recycling of MHC Class I Molecules in Non-professional Antigen Presenting and Dendritic Cells”. In: *Frontiers in Immunology* 9 (Jan. 2019). ISSN: 1664-3224. DOI: [10.3389/fimmu.2018.03098](https://doi.org/10.3389/fimmu.2018.03098).
- [Mel96] Ira Mellman. “Endocytosis and Molecular Sorting”. In: *Annual Review of Cell and Developmental Biology* 12 (1 Nov. 1996), pp. 575–625. ISSN: 1081-0706. DOI: [10.1146/annurev.cellbio.12.1.575](https://doi.org/10.1146/annurev.cellbio.12.1.575).
- [MFS22] Marta Martins, Ana Sofia Fernandes, and Nuno Saraiva. “Golgi: Cancer cell fate control”. In: *The International Journal of Biochemistry & Cell Biology* 145 (Apr. 2022), p. 106174. ISSN: 1357-2725. DOI: [10.1016/j.biocel.2022.106174](https://doi.org/10.1016/j.biocel.2022.106174). URL: <http://dx.doi.org/10.1016/j.biocel.2022.106174>.
- [Mie98] Kaisa Miettinen. *Nonlinear multiobjective optimization*. Kluwer Academic Publishers, 1998.
- [Min+21] Fumi Minoshima et al. “Integrated analysis of glycan and RNA in single cells”. In: *iScience* 24.8 (Aug. 2021), p. 102882. ISSN: 2589-0042. DOI: [10.1016/j.isci.2021.102882](https://doi.org/10.1016/j.isci.2021.102882). URL: <http://dx.doi.org/10.1016/j.isci.2021.102882>.
- [MO14] Shankar Mukherji and Erin K O’Shea. “Mechanisms of organelle biogenesis govern stochastic fluctuations in organelle abundance”. In: *eLife* 3 (June 2014). ISSN: 2050-084X. DOI: [10.7554/elife.02678](https://doi.org/10.7554/elife.02678).
- [MS11] J. Malsam and T. H. Sollner. “Organization of SNAREs within the Golgi Stack”. In: *Cold Spring Harbor Perspectives in Biology* 3 (10 Oct. 2011), a005249–a005249. ISSN: 1943-0264. DOI: [10.1101/cshperspect.a005249](https://doi.org/10.1101/cshperspect.a005249).
- [MT16] Somya Mani and Mukund Thattai. “Stacking the odds for Golgi cisternal maturation”. In: *eLife* 5 (2016). DOI: [10.7554/elife.16231](https://doi.org/10.7554/elife.16231).
- [Mur02] J. D. Murray. *Mathematical Biology*. Ed. by J. D. Murray. Vol. 17. Springer New York, 2002. ISBN: 978-0-387-95223-9. DOI: [10.1007/b98868](https://doi.org/10.1007/b98868).
- [Nee+19] Sriram Neelamegham et al. “Updates to the Symbol Nomenclature for Glycans guidelines”. In: *Glycobiology* 29 (9 Aug. 2019), pp. 620–624. ISSN: 1460-2423. DOI: [10.1093/glycob/cwz045](https://doi.org/10.1093/glycob/cwz045).
- [NRR02] Andreas Nebenführ, Christophe Ritzenthaler, and David G. Robinson. “Brefeldin A: Deciphering an enigmatic inhibitor of secretion”. In: *Plant Physiology* 130.3 (2002), pp. 1102–1108. DOI: [10.1104/pp.011569](https://doi.org/10.1104/pp.011569).
- [Ols02] Fredrik J. Olson. “Blood Group A Glycosyltransferase Occurring as Alleles with High Sequence Difference Is Transiently Induced during a *Nippostrongylus brasiliensis* Parasite Infection”. In: *Journal of Biological Chemistry* 277.17 (2002).
- [OM10] Otso Ovaskainen and Baruch Meerson. “Stochastic models of population extinction”. In: *Trends in Ecology and Evolution* 25 (11 Nov. 2010), pp. 643–652. ISSN: 01695347. DOI: [10.1016/j.tree.2010.07.009](https://doi.org/10.1016/j.tree.2010.07.009).
- [OMT22] Lalhaba Oinam, Fumi Minoshima, and Hiroaki Tateno. “Glycan profiling of the gut microbiota by Glycan-seq”. In: *ISME Communications* 2 (1 Dec. 2022). ISSN: 2730-6151. DOI: [10.1038/s43705-021-00084-2](https://doi.org/10.1038/s43705-021-00084-2).

- [Pas+14] Matthew J. Paszek et al. "The cancer glycocalyx mechanically primes integrin-mediated growth and survival". In: *Nature* 511 (7509 July 2014), pp. 319–325. ISSN: 0028-0836. DOI: [10.1038/nature13535](https://doi.org/10.1038/nature13535).
- [Pat+08] George H. Patterson et al. "Transport through the Golgi Apparatus by Rapid Partitioning within a Two-Phase Membrane System". In: *Cell* 133 (6 June 2008), pp. 1055–1067. ISSN: 00928674. DOI: [10.1016/j.cell.2008.04.044](https://doi.org/10.1016/j.cell.2008.04.044).
- [Paw+16] Archana Pawar et al. "Ral-Arf6 crosstalk regulates Ral dependent exocyst trafficking and anchorage independent growth signalling". In: *Cellular Signalling* 28 (9 Sept. 2016), pp. 1225–1236. ISSN: 08986568. DOI: [10.1016/j.cellsig.2016.05.023](https://doi.org/10.1016/j.cellsig.2016.05.023).
- [Pet+19] Mariela D Petkova et al. "Optimal Decoding of Cellular Identities in a Genetic Network." In: *Cell* 176 (4 Feb. 2019), 844–855.e15. ISSN: 1097-4172. DOI: [10.1016/j.cell.2019.01.007](https://doi.org/10.1016/j.cell.2019.01.007).
- [Pet15] Armen Petrosyan. "Onco-Golgi: Is Fragmentation a Gate to Cancer Progression?" In: *Biochemistry & Molecular Biology Journal* 01.01 (2015). ISSN: 2471-8084. DOI: [10.21767/2471-8084.100006](https://doi.org/10.21767/2471-8084.100006). URL: <http://dx.doi.org/10.21767/2471-8084.100006>.
- [Phl] Phlsph7. *Shannon's weaver model of communication*. https://web.archive.org/web/20250106060318/https://en.wikipedia.org/wiki/Shannon-Weaver_model/. Accessed: 2025-01-06.
- [Pis14] H. Pishro-Nik. *Introduction to probability, statistics, and random processes*. Kappa Research LLC, 2014. ISBN: 978-09-906-3720-2. DOI: [doi: 10.25334/Q40H8J](https://doi.org/10.25334/Q40H8J). URL: <https://www.probabilitycourse.com>.
- [PMB12] Gustavo Pigino, Gerardo A. Morfini, and Scott T. Brady. "Intracellular Trafficking". In: Elsevier, 2012, pp. 119–145. DOI: [10.1016/B978-0-12-374947-5.00007-9](https://doi.org/10.1016/B978-0-12-374947-5.00007-9).
- [Pot+21] Prathyush Pothukuchi et al. "GRASP55 regulates intra-Golgi localization of glycosylation enzymes to control glycosphingolipid biosynthesis". In: *The EMBO Journal* 40.20 (Sept. 2021). ISSN: 1460-2075. DOI: [10.15252/embj.2021107766](https://doi.org/10.15252/embj.2021107766). URL: <http://dx.doi.org/10.15252/embj.2021107766>.
- [Poz+05] Miguel A. del Pozo et al. "Phospho-caveolin-1 mediates integrin-regulated membrane domain internalization". In: *Nature Cell Biology* 7 (9 Sept. 2005), pp. 901–908. ISSN: 1465-7392. DOI: [10.1038/ncb1293](https://doi.org/10.1038/ncb1293).
- [Pry+92] K Prydz et al. "Effects of brefeldin A on endocytosis, transcytosis and transport to the Golgi complex in polarized MDCK cells." In: *The Journal of cell biology* 119.2 (1992), pp. 259–272. DOI: [10.1083/jcb.119.2.259](https://doi.org/10.1083/jcb.119.2.259).
- [Rab+95] Catherine Rabouille et al. "Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides". In: *Journal of Cell Science* 108 (4 Apr. 1995), pp. 1617–1627. ISSN: 0021-9533. DOI: [10.1242/jcs.108.4.1617](https://doi.org/10.1242/jcs.108.4.1617).
- [Raj+23] B R Rajeshwari et al. "Kinetics of Arf1 inactivation regulates Golgi organisation and function in non-adherent fibroblasts." In: *Biology open* 12 (4 Apr. 2023). ISSN: 2046-6390. DOI: [10.1242/bio.059669](https://doi.org/10.1242/bio.059669).

- [RB15] Nathan E. Reticker-Flynn and Sangeeta N. Bhatia. “Aberrant Glycosylation Promotes Lung Cancer Metastasis through Adhesion to Galectins in the Metastatic Niche”. In: *Cancer Discovery* 5.2 (Feb. 2015), pp. 168–181. ISSN: 2159-8290. DOI: [10.1158/2159-8290.cd-13-0760](https://doi.org/10.1158/2159-8290.cd-13-0760). URL: <http://dx.doi.org/10.1158/2159-8290.cd-13-0760>.
- [Reb+96] Abdelhadi Rebbaa et al. “Ganglioside GM3 inhibition of EGF receptor mediated signal transduction”. In: *Glycobiology* 6 (4 1996), pp. 399–406. ISSN: 0959-6658. DOI: [10.1093/glycob/6.4.399](https://doi.org/10.1093/glycob/6.4.399).
- [RKO91] Alan C. Rapraeger, Alison Krufka, and Bradley B. Olwin. “Requirement of Heparan Sulfate for bFGF-Mediated Fibroblast Growth and Myoblast Differentiation”. In: *Science* 252 (5013 June 1991), pp. 1705–1708. ISSN: 0036-8075. DOI: [10.1126/science.1646484](https://doi.org/10.1126/science.1646484).
- [RO82] J F Ross and M Orlowski. “Growth-rate-dependent adjustment of ribosome function in chemostat-grown cells of the fungus *Mucor racemosus*”. In: *Journal of Bacteriology* 149 (2 Feb. 1982), pp. 650–653. ISSN: 0021-9193. DOI: [10.1128/jb.149.2.650-653.1982](https://doi.org/10.1128/jb.149.2.650-653.1982).
- [Rot02] Jürgen Roth. “Protein N -Glycosylation along the Secretory Pathway: Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions”. In: *Chemical Reviews* 102 (2 Feb. 2002), pp. 285–304. ISSN: 0009-2665. DOI: [10.1021/cr000423j](https://doi.org/10.1021/cr000423j).
- [RP15] Riccardo Rao and Luca Peliti. “Thermodynamics of accuracy in kinetic proof-reading: dissipation and efficiency trade-offs”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2015 (6 June 2015), P06001. ISSN: 1742-5468. DOI: [10.1088/1742-5468/2015/06/P06001](https://doi.org/10.1088/1742-5468/2015/06/P06001).
- [Sáe+09] José B Sáenz et al. “Golgicide A reveals essential roles for GBF1 in Golgi assembly and function”. In: *Nature Chemical Biology* 5 (3 Mar. 2009), pp. 157–165. ISSN: 1552-4450. DOI: [10.1038/nchembio.144](https://doi.org/10.1038/nchembio.144).
- [Sch+12] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis”. In: *Nature Methods* 9 (7 July 2012), pp. 676–682. ISSN: 1548-7091. DOI: [10.1038/nmeth.2019](https://doi.org/10.1038/nmeth.2019).
- [Sci+97] Noah Sciaky et al. “Golgi Tubule Traffic and the Effects of Brefeldin A Visualized in Living Cells”. In: *The Journal of Cell Biology* 139 (5 Dec. 1997), pp. 1137–1155. ISSN: 0021-9525. DOI: [10.1083/jcb.139.5.1137](https://doi.org/10.1083/jcb.139.5.1137).
- [SEB18] Vibha Singh, Chaitanya Erady, and Nagaraj Balasubramanian. “Cell-matrix adhesion controls Golgi organization and function through ARF1 activation in anchorage-dependent cells”. In: *Journal of Cell Science* (2018). DOI: [10.1242/jcs.215855](https://doi.org/10.1242/jcs.215855).
- [SG20] Danièle Stalder and David C. Gershlick. “Direct trafficking pathways from the Golgi apparatus to the plasma membrane”. In: *Seminars in Cell & Developmental Biology* 107 (Nov. 2020), pp. 112–125. ISSN: 10849521. DOI: [10.1016/j.semcdb.2020.04.001](https://doi.org/10.1016/j.semcdb.2020.04.001).
- [Sha48] C. E. Shannon. “A Mathematical Theory of Communication”. In: *SIGMOBILE Mob. Comput. Commun. Rev.* 5.1 (Jan. 1948), pp. 3–55. ISSN: 1559-1662. DOI: [10.1145/584091.584093](https://doi.org/10.1145/584091.584093). URL: <http://doi.acm.org/10.1145/584091.584093>.
- [Sho+12] O. Shoval et al. “Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space”. In: *Science* 336.6085 (2012), pp. 1157–1160. DOI: [10.1126/science.1217405](https://doi.org/10.1126/science.1217405).

- [Son+20] Sungmin Son et al. “Molecular height measurement by cell surface optical profilometry (CSOP)”. In: *Proceedings of the National Academy of Sciences* 117 (25 June 2020), pp. 14209–14219. ISSN: 0027-8424. DOI: [10.1073/pnas.1922626117](https://doi.org/10.1073/pnas.1922626117).
- [Spa+16] Philipp N. Spahn et al. “A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering”. In: *Metabolic Engineering* 33 (2016), pp. 52–66. DOI: [10.1016/j.ymben.2015.10.007](https://doi.org/10.1016/j.ymben.2015.10.007).
- [SS15] Luís F. Seoane and Ricard Solé. “Phase transitions in Pareto optimal complex networks”. In: *Physical Review E* 92 (3 Sept. 2015), p. 032807. ISSN: 1539-3755. DOI: [10.1103/PhysRevE.92.032807](https://doi.org/10.1103/PhysRevE.92.032807).
- [ST22] Sinead Schwabl and David Teis. “Protein quality control at the Golgi”. In: *Current Opinion in Cell Biology* 75 (2022), p. 102074. ISSN: 0955-0674. DOI: <https://doi.org/10.1016/j.ccb.2022.02.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0955067422000205>.
- [Sta11] P. Stanley. “Golgi glycosylation”. In: *Cold Spring Harbor Perspectives in Biology* 3.4 (2011). DOI: [10.1101/cshperspect.a005199](https://doi.org/10.1101/cshperspect.a005199).
- [Sto+93] Pieter F.W. Stouten et al. “How does the switch II region of G-domains work?” In: *FEBS Letters* 320 (1 Mar. 1993), pp. 1–6. ISSN: 0014-5793. DOI: [10.1016/0014-5793\(93\)81644-F](https://doi.org/10.1016/0014-5793(93)81644-F).
- [Str+21] Carsen Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. In: *Nature Methods* 18 (1 Jan. 2021), pp. 100–106. ISSN: 1548-7091. DOI: [10.1038/s41592-020-01018-x](https://doi.org/10.1038/s41592-020-01018-x).
- [Str+93] G.J. Strous et al. “Differential effects of Brefeldin A on transport of secretory and lysosomal proteins.” In: *Journal of Biological Chemistry* 268 (4 Feb. 1993), pp. 2341–2347. ISSN: 00219258. DOI: [10.1016/S0021-9258\(18\)53781-9](https://doi.org/10.1016/S0021-9258(18)53781-9).
- [Sun+21] Xiuping Sun et al. “A quantitative study of the golgi retention of glycosyltransferases”. In: *Journal of Cell Science* 134.20 (2021). DOI: [10.1242/jcs.258564](https://doi.org/10.1242/jcs.258564).
- [Tam+68] G Tamura et al. “Antiviral activity of brefeldin A and verrucarin A.” In: *The Journal of antibiotics* 21 (2 Feb. 1968), pp. 160–1. ISSN: 0021-8820. DOI: [10.7164/antibiotics.21.160](https://doi.org/10.7164/antibiotics.21.160).
- [Tha16] Mukund Thattai. “Universal Poisson Statistics of mRNAs with Complex Decay Pathways”. In: *Biophysical Journal* 110 (2 Jan. 2016), pp. 301–305. ISSN: 00063495. DOI: [10.1016/j.bpj.2015.12.001](https://doi.org/10.1016/j.bpj.2015.12.001).
- [Tie+24] Hieng Chiong Tie et al. “Quantitative intra-Golgi transport and organization data suggest the stable compartment nature of the Golgi”. In: *eLife* (June 2024). DOI: [10.7554/eLife.98582.1](https://doi.org/10.7554/eLife.98582.1).
- [Tie+25] Hieng Chiong Tie et al. “Quantitative intra-Golgi transport and organization data suggest the stable compartment nature of the Golgi”. In: *eLife* 13 (July 2025). ISSN: 2050-084X. DOI: [10.7554/eLife.98582](https://doi.org/10.7554/eLife.98582).
- [TMA15] Avichai Tendler, Avraham Mayo, and Uri Alon. “Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells”. In: *BMC Systems Biology* 9.1 (Mar. 2015), p. 12. ISSN: 1752-0509. DOI: [10.1186/s12918-015-0149-z](https://doi.org/10.1186/s12918-015-0149-z). URL: <https://doi.org/10.1186/s12918-015-0149-z>.

- [TS00] Robert H. Tukey and Christian P. Strassburg. “Human UDP-Glucuronosyltransferases: Metabolism, Expression, and Disease”. In: *Annual Review of Pharmacology and Toxicology* 40 (1 Apr. 2000), pp. 581–616. ISSN: 0362-1642. DOI: [10.1146/annurev.pharmtox.40.1.581](https://doi.org/10.1146/annurev.pharmtox.40.1.581).
- [TW13] Danming Tang and Yanzhuang Wang. “Cell cycle regulation of Golgi membrane dynamics”. In: *Trends in Cell Biology* 23 (6 June 2013), pp. 296–304. ISSN: 09628924. DOI: [10.1016/j.tcb.2013.01.008](https://doi.org/10.1016/j.tcb.2013.01.008).
- [UB97] Pablo Umana and James Bailey. “A Mathematical Model of N-Linked Glycoform Biosynthesis”. In: *Biotechnology and Bioengineering* 1 (2 1997), pp. 890–908. ISSN: 1027-3662. DOI: [10.1080/10273669708833015](https://doi.org/10.1080/10273669708833015).
- [Van+12] Jacob VanderPlas et al. “Introduction to astroML: Machine learning for astrophysics”. In: IEEE, Oct. 2012, pp. 47–54. ISBN: 978-1-4673-4627-6. DOI: [10.1109/CIDU.2012.6382200](https://doi.org/10.1109/CIDU.2012.6382200).
- [Var13] Ajit Varki. “Account for the ‘dark matter’ of biology”. In: *Nature* 497 (7451 May 2013), pp. 565–565. ISSN: 0028-0836. DOI: [10.1038/497565a](https://doi.org/10.1038/497565a).
- [Var16] Ajit Varki. “Biological roles of glycans”. In: *Glycobiology* 27.1 (2016), pp. 3–49. DOI: [10.1093/glycob/cww086](https://doi.org/10.1093/glycob/cww086).
- [Var17] Ajit Varki. “Biological roles of glycans”. In: *Glycobiology* 27 (1 Jan. 2017), pp. 3–49. ISSN: 0959-6658. DOI: [10.1093/glycob/cww086](https://doi.org/10.1093/glycob/cww086).
- [Vir+20] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nature Methods* 17 (3 Mar. 2020), pp. 261–272. ISSN: 1548-7091. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [VNK11] Ioscani Jimenez del Val, Judit M. Nagy, and Cleo Kontoravdi. “A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus”. In: *Biotechnology Progress* 27 (6 Nov. 2011), pp. 1730–1743. ISSN: 8756-7938. DOI: [10.1002/btpr.688](https://doi.org/10.1002/btpr.688).
- [Wan08] Wanderingstan. *Shannon's diagram of a general communications system, which shows the process that produces a message*. 2008. URL: https://en.wikipedia.org/wiki/A_Mathematical_Theory_of_Communication.
- [War+01] Theresa H. Ward et al. “Maintenance of Golgi structure and function depends on the integrity of ER export”. In: *The Journal of Cell Biology* 155 (4 Nov. 2001), pp. 557–570. ISSN: 1540-8140. DOI: [10.1083/jcb.200107045](https://doi.org/10.1083/jcb.200107045).
- [Wes+21] Christian A. E. Westrip et al. “Developmentally regulated GTPases: structure, function and roles in disease”. In: *Cellular and Molecular Life Sciences* 78 (23 Dec. 2021), pp. 7219–7235. ISSN: 1420-682X. DOI: [10.1007/s00018-021-03961-0](https://doi.org/10.1007/s00018-021-03961-0).
- [WGG05] Christopher T. Walsh, Sylvie Garneau-Tsodikova, and Gregory J. Gatto. “Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications”. In: *Angewandte Chemie International Edition* 44 (45 Nov. 2005), pp. 7342–7372. ISSN: 1433-7851. DOI: [10.1002/anie.200501023](https://doi.org/10.1002/anie.200501023).
- [XW17] Haopeng Xiao and Ronghu Wu. “Quantitative investigation of human cell surface N-glycoprotein dynamics”. In: *Chemical Science* 8 (1 2017), pp. 268–277. ISSN: 2041-6520. DOI: [10.1039/C6SC01814A](https://doi.org/10.1039/C6SC01814A).
- [Yad+22] Alkesh Yadav et al. “Glycan processing in the Golgi: Optimal information coding and constraints on cisternal number and enzyme specificity”. In: *eLife* 11 (2022). DOI: [10.7554/elife.76757](https://doi.org/10.7554/elife.76757).

- [Yar+20] Rebecca Yarwood et al. “Membrane trafficking in health and disease”. In: *Disease Models & Mechanisms* 13 (4 Apr. 2020). ISSN: 1754-8411. DOI: [10.1242/dmm.043448](https://doi.org/10.1242/dmm.043448).
- [YI21] Ou Yang and Marianthi Ierapetritou. “mAb Production Modeling and Design Space Evaluation Including Glycosylation Process”. In: *Processes* 9 (2 Feb. 2021), p. 324. ISSN: 2227-9717. DOI: [10.3390/pr9020324](https://doi.org/10.3390/pr9020324).
- [YSS95] H Yasuhara, S Sonobe, and H Shibaoka. “Effects of brefeldin A on the formation of the cell plate in tobacco BY-2 cells.” In: *European journal of cell biology* 66 (3 Mar. 1995), pp. 274–81. ISSN: 0171-9335.
- [Zha21] Xiaoyan Zhang. “Alterations of Golgi Structural Proteins and Glycosylation Defects in Cancer”. In: *Frontiers in Cell and Developmental Biology* 9 (May 2021). ISSN: 2296-634X. DOI: [10.3389/fcell.2021.665289](https://doi.org/10.3389/fcell.2021.665289). URL: <http://dx.doi.org/10.3389/fcell.2021.665289>.
- [Zho+21] Xing Zhou et al. “Antibody glycosylation in autoimmune diseases”. In: *Autoimmunity Reviews* 20.5 (May 2021), p. 102804. ISSN: 1568-9972. DOI: [10.1016/j.autrev.2021.102804](https://doi.org/10.1016/j.autrev.2021.102804). URL: <http://dx.doi.org/10.1016/j.autrev.2021.102804>.
- [Zho+22] Xiaotian Zhong et al. “New Opportunities in Glycan Engineering for Therapeutic Proteins”. In: *Antibodies* 11 (1 Jan. 2022), p. 5. ISSN: 2073-4468. DOI: [10.3390/antib11010005](https://doi.org/10.3390/antib11010005).
- [Zho11] W. Zhong. “Golgi during Development”. In: *Cold Spring Harbor Perspectives in Biology* 3.9 (July 2011), a005363–a005363. ISSN: 1943-0264. DOI: [10.1101/cshperspect.a005363](https://doi.org/10.1101/cshperspect.a005363). URL: <http://dx.doi.org/10.1101/cshperspect.a005363>.
- [ZW16] Xiaoyan Zhang and Yanzhuang Wang. “GRASPs in Golgi Structure and Function”. In: *Frontiers in Cell and Developmental Biology* 3 (Jan. 2016). ISSN: 2296-634X. DOI: [10.3389/fcell.2015.00084](https://doi.org/10.3389/fcell.2015.00084).